Development of Y-STR genotyping systems suitable for sexual assault cases 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

Kevin Wesley Cloete

March 2010

Supervisor: ***Dr. Mongi Benjeddou (PhD)***

Keywords

Forensics

Short Tandem Repeat (STR)

Y-loci

Minimal Haplotype (MinHt)

YHRD

Polymerase Chain Reaction (PCR)

Electrophoresis

Multiplex PCR

Genotyping

Gene Diversity

Polymorphism

Abstract

Development of Y-STR genotyping systems suitable for sexual assault cases in South Africa

K W Cloete

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

24/02/2010

Sexual assault is a significant problem facing the South African society. In this context, efficient but also affordable genotyping systems are needed for positive identification of criminals in incidences of sexual violence. The aim of this study was therefore to develop non-commercial Y-STR genotyping systems suitable for sexual assault cases in South Africa.

Y-chromosome STR loci constituting the minimal haplotype are still the most widely used loci in investigating sexual assault cases despite the fact that DYS391 and DYS392 have shown low levels of polymorphism in Xhosa populations in Cape Town. The minimal haplotype was, therefore, further investigated in the Cape Muslim population.

The Cape Muslim population generally exhibited high GD values among all the South African populations. These values were higher than 0.5 for most loci, and ranged from 0.447 for DYS391 to 0.957 for DYS385. The highest number of alleles in most loci was also recorded in this population. The overall assessment of the minimal haplotype has shown that this system is still a useful in investigating sexual assault case in many South African subpopulations. Therefore the exercise of internal validation of the minimal haplotype system was successfully carried out in the laboratory.

The properties of additional novel and widely used STRs were also investigated in this study. Loci were successfully sequenced and allele nomenclature was assigned to them according to the ISFG guidelines.

UWC Y-plex 2 was originally constructed as an 11 Y-STR multiplex, comprised of Y-STRs that show gene diversity values of above 0.7 in South African sub-populations. Only Loci DYS449, DYS481, DYS518, DYS557, DYS570, DYS607, DYS612 and DYS614 were used in a single multiplex reaction and named the eight Y-STR's system.

Population data from the Cape Muslim Community was generated using the eight Y–STR's system. The GD values ranged between 0.789 (DYS607) and 0.884 (DYS449) in this population. The population data generated from this study, as well as previous studies, clearly indicated that the eight Y-STR's genotyping system could be useful in forensic casework in South Africa. The internal validation exercise was therefore successfully carried out for this genotyping system.

List of Publications

1. **Cloete, K., Ehrenreich, L., D'Amato M.E., Leat, N., Davison, S. and Benjeddou, M. (2009).** Analysis of seventeen Y-chromosome STR loci in the Cape Muslim population of South Africa. *Legal Medicine*, doi:10.1016/j.legalmed.2009.10.001

2. **D'Amato, M.E., Ehrenreich, L., Cloete, K., Benjeddou, M. and Davison, S. (2009).** Characterization of the highly discriminatory loci DYS449, DYS481, DYS518, DYS612, DYS626, DYS644 and DYS710, *Forensic Science International*: *Genetics*, doi:10.1016/j.fsigen.2009.06.011

3. **Leat, N., Ehrenreich, L., Benjeddou, M., Cloete, K. and Davison, S. (2007).** Properties of novel and widely studied Y-STR loci in three South African populations. *Forensic Science International* **168 (2- 3),** 154 – 161

Declaration

I declare that 'Development of Y-STR genotyping systems suitable for sexual assault cases 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 or quoted have been indicated and acknowledged by complete references.

Full Name: Kevin Wesley Cloete

Signed:

Date: 24/02/2010

Acknowledgements

I would like to thank my research supervisor Dr Mongi Benjeddou for his support, patience and guidance throughout this project.

I would like to acknowledge and thank the National Research Foundation (NRF) for financial assistance.

Many thanks also go to all the Cape Muslim individuals who so generously donated biological material for the study.

I would also like to thank Dr Neil Leat, Dr Maria Eugenia D'Amato and Liezle Ehrenreich for their assistance in the lab work.

My gratitude also goes to all the students and staff who made working in the Department of Biotechnology such a pleasant experience.

Thanks also go to family and friends who have supported me during this endeavour.

List of Abbreviations

DNA	Deoxyribonucleic Acid
STR	Short Tandem Repeat
Y-STR	Y-chromosome STR
PAR	Pseudo-Autosomal Region
NRY	Non-recombining Y-chromosome region
MSY	Male-specific Y-chromosome region
MH	Minimal Haplotype
ISFG	International Society of Forensic Genetics
YHRD	Y-Chromosome Haplotype Reference Database
SWGDAM	Scientific Working Group on DNA Analysis Methods
YCC	Y-chromosome Consortium
NIST	National Institute of Standards and Technology
PCR	Polymerase Chain Reaction
$MgCl_2$	Magnesium Chloride
dNTPs	deoxi Nucleotide TriPhosphates
EDTA	Ethylene Diamine Tetra Acetic Di-Sodium Salt
TEMED	N, N, N', N' Tetramethyl-EthyleneDiamine
NaCl	Sodium Chloride
RFU	Relative Fluorescent Units

List of Tables

Table 1.1.

Expected DNA amounts extracted from biological samples (Lee and Ladd, 2001) **3**

Table 2.1.

Primer sequences, dye labels and final concentrations used in the MinHt multiplex. **26**

Table 2.2.

Allele frequencies and gene diversity at 9 Y-chromosome STRs in the Cape Muslim population (n=98) typed with the 'minimal haplotype' system. **30**

Table 2.3.

List of 9 Y chromosome STR haplotypes observed in the Cape Muslim population (n=98) typed with the 'minimal haplotype' sytem. **31**

Table 2.4.

Nine Y Chromosome STRs haplotypes not unique typed in 98 individuals from the Cape Muslim population using the minimal haplotype system. **34**

Table 2.5.

Comparison of Gene Diversity values for the MinHt loci among South African sub-populations. **35**

Table 3.1.

Dye labels, PCR product sizes, repeat motives, and Gene Bank accession numbers for the MinHt Y-STR (Butler *et al.*, 2002). **38**

Table 3.2.

Y-loci base pair sizes generated during amplification reactions using 0.5ng DNA template. The amplifications were performed in 3 occasions with a time interval of one month. **43**

Table 3.3.

Y-loci base pair sizes generated during amplification reactions from four different DNA samples. **43**

Table 3.4.

Minor and major component genotypes at non-overlapping Y-STR loci from two male DNA mixture amplifications. **46**

Table 4.1.

Primers sequences of Y-STRs generated by *Primer3* software. **49**

Table 4.2.

DNA used during amplification and sequencing of selected Y-STRs. **52**

Table 4.3.

Sequence Data for selected Y-STR loci. **55**

Table 5.1.

Primer sequences and final concentration for the eight Y-STRs multiplex. **60**

Table 5.2.

Allele frequencies and gene diversity values of eight Y-STR loci from 100 Cape Muslim males. **63**

Table 5.3.

Comparison of GD values from eight Y-STR loci among five South African sub- populations. **64**

Table 5.4.

Haplotypes for eight Y-STR loci observed from 100 Cape Muslim males. **65**

Table 5.5.

Haplotypes shared by more than one Cape Muslim male (n=100). **68**

Table 6.1.

General information of the Y-STRs selected from Leat *et al*. (2007). **70**

Table 6.2.

Y-loci base pair sizes generated during amplification reactions using 0.5ng DNA template. **72**

Table 6.3.

Y-loci base pair sizes generated during amplification reactions from four different DNA samples. **72**

Table 6.4.

Minor and major component genotypes at non-overlapping Y-STR loci from two male DNA mixture amplifications. **75**

List of Figures

Figure 1.1.

Simplified Y-chromosome structure representing the 3 regions: pseudo- autosomal (PAR1 and PAR2), heterochromatic and the non-recombining (Jobling & Tyler-Smith, 2000). **6**

Figure 1.2.

Primer design-based approach to multiplex optimization (Butler *et al*., 2002). **14**

Figure 1.3.

Predicted allele size range for Y-STRs amplified in (a) Y-PLEX™ 12 kit from *Reliagene*, (b) MenType Argus Y-MH from *Biotype*, (c) PowerPlex Y from *Promega* and (d) AmpFlSTR® Yfiler™ from *Applied Biosystems*. **16**

Figure 1.4.

Predicted amplicon size range for loci amplified in 20-plex (Butler *et al*., 2002). **17**

Figure 1.5.

Predicted allele size range for Y-STRs amplified in 21 Y-STR megaplex (Hanson and Ballantyne, 2004). **18**

Figure 1.6.

Predicted allele size range for Y-loci amplified in (a) UWC-Yplex 1, (b) UWC-Yplex 2 and (c) UWC-Yplex 3. **19**

Figure 2.1.

An example of a typical electropherogram of the MinHt multiplex reaction. **29**

Figure 3.1.

Y-STR profiles generated from varying amounts of template DNA, from panels top to bottom: 0.005ng, 0.05ng, 0.125ng, 0.25ng, and 0.5ng. **42**

Figure 3.2.

Amplification of male DNA in the presence of female DNA. Y-STR profiles shown in the panels from top to bottom: 0.5ng of male DNA, 0.5ng male DNA with 0.5ng female DNA (1:1), 0.5ng male DNA with 1ng female DNA (1:2), 0.5ng male DNA with 2.5ng female DNA (1:5) , 0.5ng male DNA with 10ng female DNA (1:10), and 0.5ng female DNA. **45**

Figure 3.3.

Amplification of mixed male DNA samples. Y-STR profiles shown in panels from top to bottom: 0.5ng male DNA 01 with 0.5ng male DNA 02 (1:1) and 0.5ng male DNA 01 with 2.5ng male DNA 02 (1:5). **47**

Figure 5.1.

An example of an eight Y-STRs multiplex reaction electropherogram. **62**

Figure 6.1.

Y-STR profiles generated from varying amounts of template DNA, from panels top to bottom: 0.005ng, 0.05ng, 0.125ng, 0.25ng, and 0.5ng. **71**

Figure 6.2.

Amplification of male DNA in the presence of female DNA. Y-STR profiles shown in the panels from top to bottom: 0.5ng of male DNA, 0.5ng male DNA with 5ng female DNA (1:10) and 0.5ng female DNA. **74**

Figure 6.3.

Amplification of mixed male DNA samples. Y-STR profiles shown in panels from top to bottom: 0.5ng male DNA 01 with 0.5ng male DNA 02 (1:1) and 0.5ng male DNA 01 with 2.5ng male DNA 02 (1:5). **76**

Table of Contents

Keywords		i
Abstract		ii
List of Publications		iv
Declaration		v
Acknowledgements		vi
List of Abbreviations		vii
List of Tables		viii
List of Figures		xi
Chapter 1 – General Introduction		**1**
1.1	Sexual assault crime in South Africa	1
1.2	Investigation procedures for sexual assault cases	2
1.3	Genetic identity testing	3
1.3.1	Restriction Fragment Length Polymorphism (RFLP)	4
1.3.2	Variable Number of Tandem Repeats (VNTR)	4
1.3.3	Short Tandem Repeats (STR)	4
1.4	Y-Chromosomal Short Tandem Repeats (Y-STR)	5
1.4.1	Y-Chromosome	5
1.4.2	Discovery of Y-STR loci	7
1.4.3	Uses of Y-STR loci	9
1.5	Multiplex amplification of Y-STRs	11
1.5.1	Analysis of Y-STR loci	11
1.5.2	Development of Y-STR multiplexes	11
1.5.3	Optimization of Y-STR multiplexes	12

1.6	Survey of Y-STR multiplexes	15
1.6.1	Commercial Multiplexes	15
1.6.2	Non-commercial Multiplexes	17
1.6.3	UWC Y-Plexes	18
1.7	Validation of Y-STR multiplexes	19
1.7.1	Sensitivity studies	20
1.7.2	Reproducibility studies	20
1.7.3	Species specificity studies	21
1.7.4	Environmental studies	21
1.7.5	DNA mixture studies	21
1.8	Population databases and population studies	22
1.9	Objectives of the study	23

Chapter 2 – Analysis of the "minimal haplotype" loci in the Cape Muslim population of South Africa **25**

2.1	Introduction	25
2.2	Material and Methods	26
2.2.1	DNA samples	26
2.2.2	PCR multiplex amplification	26
2.2.3	PCR amplicon analysis	27
2.2.4	Allelic ladder construction	28
2.2.5	Statistical data analysis	28
2.3	Results and Discussion	29
2.3.1	MinHt multiplex performance	29
2.3.2	Analysis of allele and haplotype frequencies	30

2.3.3	Comparison of allele frequencies among South African sub-populations	34
2.3.4	Comparison of common haplotypes with those recorded in the YHRD	35
2.4	Summary	36
Chapter 3 – Internal validation of the MinHt multiplex		**37**
3.1	Introduction	37
3.2	Material and Methods	38
3.2.2	PCR amplification conditions	38
3.2.3	PCR amplicon analysis	39
3.2.4	Validation procedures	39
3.2.4.1	Sensitivity studies	39
3.2.4.2	Reproducibility studies	39
3.2.4.3	Species specificity studies	39
3.2.4.4	Environmental studies	40
3.2.4.5	DNA mixture studies	40
3.2.4.5.1	Male/female mixture	41
3.2.4.5.2	Male/male mixture	41
3.3	Results and Discussion	41
3.3.1	Minimal Sensitivity	41
3.3.2	Reproducibility studies	43
3.3.3	Species Specificity Studies	44
3.3.4	Environmental studies	44
3.3.5	DNA Mixtures studies	44
3.3.5.1	Male/female mixtures	44

3.3.5.2 Male/male mixtures 46

3.4 Summary 47

Chapter 4 - The properties of twelve novel and widely studied Y-STR loci 48

4.1 Introduction 48

4.2 Material and Methods 48

4.2.1 Selection of Y-STRs 48

4.2.2 DNA samples 49

4.2.3 Primers 49

4.2.4 PCR single amplification 50

4.2.5 DNA Sequencing 51

4.3 Results and Discussion 53

4.3.1 Allele nomenclature 53

4.4 Summary 58

Chapter 5 - Analysis of eight Y-STR loci in the Cape Muslim population of South Africa 59

5.1 Introduction 59

5.2 Material and Methods 59

5.2.1 PCR multiplex amplification 59

5.2.2 Allelic ladder construction 61

5.3 Results and Discussion 61

5.3.1 Performance of the Eight Y-STR's system 61

5.3.2 Analysis of allele and haplotype frequencies for Cape Muslim males 62

5.3.3 Comparison of allele frequencies among six South African sub-populations 64

5.3.4 Haplotype analysis of the Cape Muslim population 68

5.3.5 Haplotype comparison of Cape Muslim males with South African sub-populations 68

5.4 Summary 68

Chapter 6 – Internal validation of the eight Y-plex genotyping system 69

6.1 Introduction 69

6.2 Material and Methods 69

6.2.1 DNA samples 69

6.2.2 PCR amplification conditions 69

6.2.3 PCR amplicon analysis 70

6.2.4 Validation procedures 70

6.3 Results and Discussion 70

6.3.1 Minimal Sensitivity 70

6.3.2 Reproducibility studies 72

6.3.3 Species Specificity Studies 73

6.3.4 Environmental studies 73

6.3.5 DNA mixtures studies 74

6.3.5.1 Male/female mixtures 74

6.3.5.2 Male/male mixtures 75

6.4 Summary 76

Chapter 7 – Conclusion 77

References 79

Electronic Supplementary Resources 88

Appendix 89

Chapter 1

General Introduction

1.1. Sexual assault crime in South Africa

Violent crimes in South Africa involve murder, attempted murder, physical assault, assault with the intent to do grievous bodily harm, indecent assault, and rape. Rape is the third most common violent crime in South Africa. The available statistics from the South African Police Service shows a total of 276 828 rape cases were reported between 2001 and 2007. The South African Police Service (http://www.saps.org.za) reported between April 2007 and December 2007 that the highest number of rape cases was recorded in Gauteng Province (8073 cases), while the highest rape case ratio per 100 000 individuals was recorded in the Northern Cape Province reaching 92 reported rape cases. Due to a change in the definition of sexually motivated crime resulting from the implementation of the Sexual Assault Amendment Bill (ACT 32) of 2007 on 16 December, rape and indecent assault case figures are only provided for the period April to December 2007. Although these figures are high, there are other organizations, such as the National Institute of Crime Rehabilitation and Rape Crisis (http://www.rapecrisis.org.za), which speculates that these figures represent merely the one out of twenty cases that do get reported.

Law enforcement experience has shown that in most sexual assault crimes (a) victims are acquainted with the perpetrator/s in some way, (b) males are less likely to report their sexual assault to the police, (c) victims may lie about specifics of the sexual assault because they fear that their actions may have contributed to the sexual assault and (d) the police are more likely to be notified of sexual assaults that are committed by strangers than by someone the victim knows (www.theiacp.org). In addition to the fact that these rape crimes are highly under-reported, the conviction rate of these crimes is also very low. According to a recent police

study, it was estimated that of the reported rape cases only 15% reach a conviction. The very effort to report a sexual assault case is a battle in overcoming shame, embarrassment and public scrutiny (www.theiacp.org).

Sexual assault is indeed a significant problem facing South African society, and a real effort should be made to (1) properly investigate these violent crimes by law enforcement agencies, and (2) develop efficient systems needed for the positive identification of criminals in incidences of sexual violence.

1.2. Investigation procedures for sexual assault cases

Sexual assault victims undergo a physical examination by a forensic medical doctor as a routine and standard procedure. The purpose of the examination is to record and collect all physical evidence that can de used to identify the perpetrator(s). All foreign biological samples found on the victim are carefully sampled to prevent contamination, and sent to police DNA forensic laboratory. Biological evidence sampled includes blood and blood stains, semen and semen stain, bone, tissues, organs, teeth, fingernail clippings, saliva, urine and other biological fluids (Lee and Ladd, 2001). DNA is then extracted from the biological samples, and specific polymorphic regions of the DNA are examined as part of the genetic identity testing. The success of the DNA analysis is dependent on the quantity, purity and degradation of the DNA extracted from the biological evidence. Table 1.1 shows the expected amounts of DNA that can be extracted from a variety of biological sample types. However, it should be noted that the quantities and quality of the DNA extracted may be slightly altered by environmental exposure to light, temperature and humidity changes (Bender *et al*., 2004). DNA extraction from environmental contaminates (soil, clothing dyes and grease) can also influence the DNA purity which in turn compromises DNA testing efficiency (Lee and Ladd,

2001). In addition, it is also important to use a male genetic identification system that is suitable for this type of investigation.

Table 1.1. Expected DNA amounts extracted from biological samples (Lee and Ladd, 2001).

Type of sample	Amount of DNA
Liquid blood	20 000 - 40 000 ng/mL
Blood stain	250 - 500 ng/cm2
Liquid semen	150 000 - 300 000 ng/mL
Post- vaginal sexual intercourse swab	10 - 3000 ng/swab
Plucked hair (with root)	1 - 750 ng/root
Shed hair (with root)	1 - 10 ng/root
Liquid saliva	1 000 - 10 000 ng/mL
Oral swab	100 - 1500 ng/swab
Urine	1 - 20 ng/mL
Bone	3 - 10 ng/mg
Tissue	50 - 500ng/mg

1.3. Genetic identity testing

Genetic identity testing is achieved by examining polymorphic regions of DNA. Typically sets of polymorphisms are examined together to provide a genetic profile. A variety of identification systems making use of these polymorphisms are used for different forensic casework, each with different limitations such as DNA fingerprinting (Kirby, 1992). There are three main types of DNA fingerprints: Restriction fragment length polymorphism (RFLP), Variable number of tandem repeats (VNTR) and Short Tandem Repeats (STR) (Kirby, 1992). The polymorphic markers most commonly used are on the autosomal (1-22) chromosomes. While they have an excellent capacity to distinguish between individuals, they do also have disadvantages. In sexual assault cases it is often difficult to separate the female victim's profile from the rapist's profile. Analysis of Y-chromosome markers overcomes this by generating male specific profiles.

1.3.1. Restriction fragment length polymorphism (RFLP)

RFLPs were the first type of DNA fingerprinting used during the nineteen eighties (1980s) examining sequences of base pairs within regions of DNA with a high probability of being completely unique to the suspect. The analysis links the suspect to the evidence collected. Unfortunately, the RFLP requires a large amount of genetic material from the crime scene such as several strands of hair or large amount of blood. Limitations of the technique include DNA degradation, with too much of it inhibiting the process. The method is also very lengthy taking from three weeks to 3 months to complete (Kirby, 1992).

1.3.2. Variable number of tandem repeats (VNTR)

A Variable Number Tandem Repeat (or VNTR) is a short nucleotide sequence, known also as a motif, repeated a variable number of times at a specific chromosomal locus. The number of repeats varies from individual to individual, from as few as 7 to more than 40 repeats. Each variant acts as an inherited allele, allowing them to be used for personal or parental identification. These polymorphic regions within DNA have been successfully used for genotyping purposes for more than 10 years (Jeffreys *et al*., 1985). VNTR patterns have been used to establish both paternity and maternity. Parent-child VNTR pattern analysis has been used to solve standard father-identification cases as well as more complicated cases of confirming adoption and biological parenthood. However, VNTR analysis has similar disadvantages to those of the RFLPs analysis such as the fact that analysis of VNTR loci requires large amounts of DNA and that the methodology is inept for typing degraded DNA and samples with very small amounts of DNA (Góes *et al*., 2002).

1.3.3. Short Tandem Repeats (STRs)

Short tandem repeats (STRs) were first observed during the late 1980's (Litt and Lutty, 1989). They are brief lengths of the non-coding region of the human genome consisting of less than

400 base pairs (hence 'short') in which are 3 to 15 repeated units, each repeated unit ranging between 3 to 7 base pairs (hence 'tandem repeats'). STRs occur in all 22 pairs of autosomal chromosomes and the X and Y sex chromosomes. Repeated numerous times in a head-to-tail manner, the STRs occur at an estimated one every 10 000 base pairs within the human genome (Edwards *et al*., 1991). The STRs are recombined during each mitotic event because half of the genetic information comes from each parent. A person is homozygous if the STR alleles at a specific location are identical and heterozygous if the alleles are different (Kirby, 1992).

In the last decade, hundreds of STRs have been identified, analysed and selected for forensic casework, specifically in human identification (Moretti *et al*., 2001). Autosomal STRs, although useful for human identity testing, have limitations that decrease their effectiveness for use in sexual assault cases. In sexual assault cases, biological material available for analysis is often a mixture of material from the male assailant and the female victim. The availability of Y chromosome specific STRs has in indeed helped to overcome this problem.

1.4. Y-Chromosomal Short Tandem Repeats (Y-STR)

1.4.1. Y-chromosome

The Y-chromosome is one of the smallest chromosomes in the human genome with an average size of ~60Mb (Buhler, 1980). The tips of the Y-chromosome arms consist of the pseudo-autosomal regions (PARs). These PARs represent ~5% of the Y-chromosome and contain sequences that are homologous to sequences on the X chromosome. During male meiosis, these PARs are the regions where the Y-chromosome pairs and exchanges genetic material with the X chromosome (Quintana-Murci and Fellous, 2001). Hence genetic elements located within these PARs, are inherited in an autosomal manner. The non-recombining region of the Y-chromosome (NRY) is made up of heterochromatin (30Mb) and

euchromatin (24Mb) and is always in a haploid state. Unless a mutation occurs, the NRY is inherited intact through paternal lineages (Quintana-Murci *et al*., 2001). For this reason the NRY is useful for several male identity testing applications. Some of theses applications include: (1) paternity testing, (2) identification of male remains after disasters, (3) investigation of male lineages for anthropological purposes, and (4) identification of male perpetrators in criminal cases, such as sexual assault (Kurihara *et al*., 2004; Koyama *et al*., 2002; Jobling and Tyler-Smith, 1995; Ploski *et al*., 2002; Dettlaff-Kakol and Pawlowski, 2002).

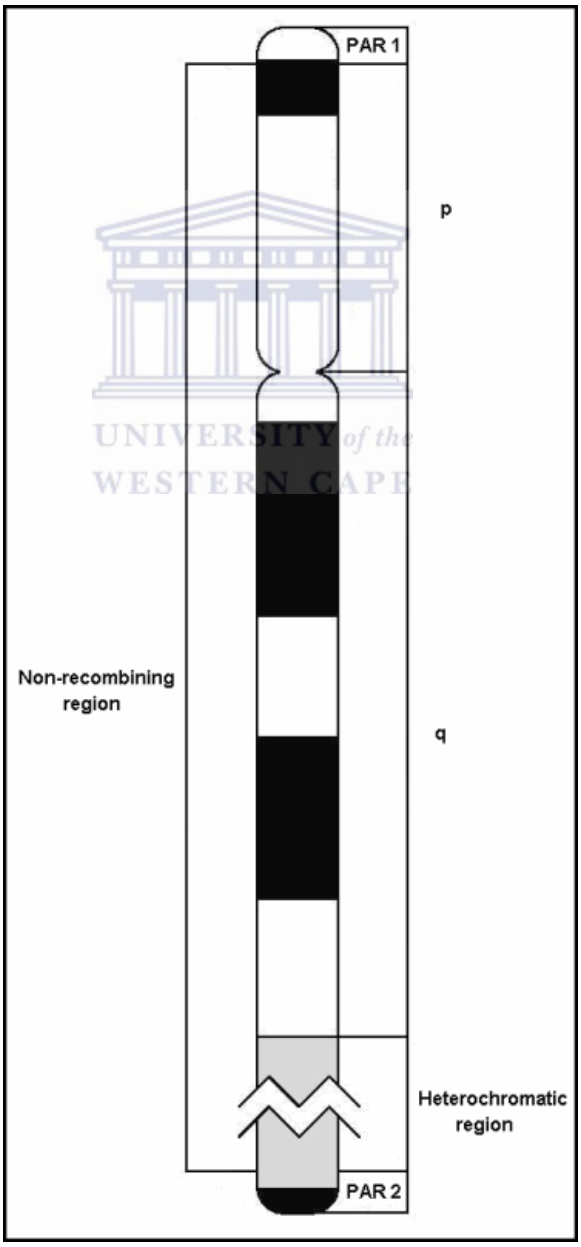

Figure 1.1. Simplified Y-chromosome structure representing the 3 regions: pseudo- autosomal (PAR1 and PAR2), heterochromatic and the non-recombining (Jobling and Tyler-Smith, 2000).

1.4.2. Discovery of Y-STR loci

By the mid 1990's only a few Y-STR loci with the potential for use in forensic studies, were known. These included the dinucleotide Y-STR loci YCA I, YCA II, and YCA III, the trinucleotide Y-STR loci DYS388 and DYS392, and the tetranucleotide Y-STR loci DYS19, DYS288, DYS385, DYS389, DYS390, DYS391, DXYS156Y, and DYS393 (Kayser *et al*., 1997). While some of these are single-copy loci, YCA I, YCA II, YCA III and DYS385 are duplicated on the Y-chromosome. Duplicated loci generate two fragments when subjected to PCR using one set of primers. The forward primer-binding site of DYS389 is also duplicated, yielding two products (DYS389 I and DYS389 II) that are approximately 100bp apart.

A collaborative study was undertaken to assess the suitability of these loci for forensic studies (Kayser *et al*., 1997). A total of 3825 males from 48 different population groups from Europe, America, Asia, Africa and Oceania, were typed. Gene diversity values ranged from low (DYS288, DYS388, DXYS156Y and YCA I), moderate (DYS391, DYS392 and DYS393), high (DYS19, DYS390, DYS389I/II and YCA II) to very high (DYS385 and YCA III). As a result, it was suggested that seven Y-STR loci (DYS19, DYS389I/II, DYS390, DYS391, DYS392, and DYS393) be used routinely in forensic applications. The inclusion of DYS385 with these loci constitutes what is commonly referred to as the 'minimal haplotype' (http://www.yhrd.org). Inclusion of YCA II to this core set was referred to as the 'extended haplotype'.

Prior to the release of large amounts of Y-chromosome sequence, attempts to identify novel Y-STR loci depended on a series of cloning and hybridization steps. In the most recent and almost certainly the last example of this approach, a cosmid library was constructed from flow-sorted human Y-chromosomes (White *et al*., 1999). Probes containing the repeated element [GATA]10 or [TATC]10 were used to select cosmids and subclones containing

repeated GATA elements. This approach led to the identification of seven loci with the potential for use in human identity testing (Y-GATA-A4, Y-GATA-A8, Y-GATA-A10, Y-GATA-C4, Y-GATA-H4, and Y-GATA-A7.1 and Y-GATA-A7.2). At the time this work practically doubled the number of known tetranucleotide Y-STR loci.

The release of substantial amounts of Y-chromosome sequence data allowed for a more straightforward approach to the identification of novel Y-STR-loci. Ayub *et al*. (2000) surveyed 1.22Mb of Y-chromosome sequence, identifying 25 STR sequences. A subset of six loci was selected for further analysis in a sample of 278 Pakistani males. While gene diversity values for several of the loci were low (DYS436 – 0.064; DYS435 – 0.070; DYS434 – 0.222), three loci appeared to be relatively polymorphic (DYS437 – 0.664; DYS438 – 0.684; DYS439 – 0.728). These loci have been incorporated into commercial Y-STR typing systems and are widely accepted by the forensic community. The Scientific Working Group on DNA Analysis Methods (SWGDAM) has therefore recommended that DYS438 and DYS439 replace the YCA II locus in the Forensic Y User Group's 'extended haplotype'. This is largely due to technical difficulties often encountered when typing dinucleotide STR loci such as YCA II.

Recently a substantial number of STR loci have been identified from Y-chromosomal sequence data. Using this approach, Iida and co-workers (2001 and 2002) characterized five more loci (DYS441, DYS442, DYS443, DYS444, and DYS445). A more thorough survey was conducted by Redd *et al*. co-workers (2002), resulting in the identification of 14 novel Y-STR loci (DYS446, DYS447, DYS448, DYS449, DYS450, DYS452, DYS453, DYS454, DYS455, DYS456, DYS458, DYS459, DYS463 and DYS464). By far the most comprehensive analysis of this kind has been conducted by Kayser and co-workers (2004).

The 23Mb of sequence surveyed by Kayser *et al*. (2004) represents almost all the euchromatic sequence of the Y-chromosome. Loci were selected with repeated elements ≥3 and ≤6 bp in size. Loci with dinucleotide repeated elements were avoided since they have a propensity to generate PCR 'stutter' products or 'shadow' bands. These 'stutter' products are often generated by the PCR amplification of STR loci and are generally one repeat unit shorter than the primary product. This approach resulted in the identification of 475 potential Y-STR loci of which 45 had previously been identified. PCR primers were successfully designed for 281 loci in silico. Of those, 166 primer-sets generated male-specific amplicons and 139 loci were demonstrated to be polymorphic in a group of eight individuals representing different binary-marker haplogroups. Using this new sequence-based approach, the pool of Y-STR loci available for evolutionary, paternity testing and forensic casework has expanded considerably.

1.4.3. Uses of Y-STR loci

STR loci are useful for several male identity testing applications. Some of theses applications include: (1) paternity testing, (2) identification of male remains after disasters, (3) investigation of male lineages for anthropological purposes, and (4) identification of male perpetrators in criminal cases, such as sexual assault (Kurihara *et al*., 2004; Koyama *et al*., 2002; Jobling and Tyler-Smith, 1995; Ploski *et al*., 2002; Dettlaff-Kakol and Pawlowski, 2002).

In sexual assault cases the most common means by which a suspect is identified is by the extraction and analysis of genetic information from semen. Differential extraction methods enable separation of sperm and female epithelial fractions, but fresh biological samples are required. Differential extraction is best done within 48 hours after the sexual assault (vaginal penetration). The likelihood for a successful differential extraction decreases significantly after 48 hours (Hall and Ballantyne, 2003). The study by Hall and colleagues showing the

success rate of differential and non-differential DNA extraction using STR genotyping over a period of time was informative. Results showed that DNA obtained from differential extraction and used during Y-STR genotyping could determine a full Y-STR profile for a maximum of 12 hours post-vaginal penetration. In contrast the non-differential extracted DNA showed a full Y-STR profile up to 48 hours post-vaginal penetration. The same non-differential extracted DNA when used in autosomal genotyping could not however separate the male suspects STR profile from the female victim's. It should be noted that a successful autosomal genotyping was possible from DNA extracted in a differential manner within 12 hours after the sexual assault.

Research on STR genotyping systems by Shewale *et al.* (2004) summarizes that Y-chromosomal genotyping is superior to autosomal genotyping in the following ways (1) Differential extraction for male and female components are unnecessary, (2) Y-STR genotyping is possible from azoospermic or oligospermic males, including vasectomized or orchidectomized males, (3) a full Y-STR profile can be generated from minuscule male DNA found within abundant female DNA, (4) the non-recombinant properties of the NRY region within male DNA allows the identification of multiple unrelated male assailants (5) and the rapid and straightforward analysis of Y-STR profiling allowed by the 'single allele per locus' concept. Moreover, the usefulness of Y-STRs is also due to the following factors: (1) the high variability of certain Y-loci in the human populations (2) the small size in base pairs (100 – 400 bp) and (3) the ability to type multiple Y-STRs in a single PCR reaction. While useful the Y-loci do have certain limitations such as the difficulty in separating paternally related individuals (due to lack of recombination in the NRY) (Gusmão *et al*., 1999; Ali and Hasnain, 2002) and the construction of a robust and effective Y-STR multiplex (Henegariu *et al*., 1997).

1.5. Multiplex amplification of Y-STRs

1.5.1. Analysis of Y-STR loci

Analysis of Y-STR loci is achieved by PCR amplification and fragment analysis to determine the differences in sizes. As a direct result of its uni-parental pattern of inheritance, the product rule used for estimating polymorphism in autosomal chromosomes cannot be applied to Y-chromosome (Butler *et al*., 2002). More Y-STR loci would therefore be needed to obtain a haplotype with the same power of discrimination as an autosome of the same size and with the same density of loci (Bosch *et al*., 2002). It would therefore be advantageous to amplify as many polymorphic Y-STR loci so as to increase the power of discrimination of the resultant haplotype. Multiplex PCR is an effective means of achieving this objective.

1.5.2. Development of Y-STR multiplexes

A multiplex assay is the simultaneous amplification of one or more DNA sequences using multiple primers in one reaction. Multiplexing is therefore useful in saving time and costs. This is important especially in situations where limited or degraded DNA samples are available it is of great value. The procedure generates multiple Y-STRs simultaneously enabling rapid construction and examination of a specific person's STR profile (Butler, 2003).

Creating a useful and effective STR multiplex does not come without some difficulties. At times DNA sequences are not amplified in equal amounts and the mix of multiple primer pairs increases the likelihood of generating more non-specific artifacts. While these difficulties do not diminish the effectiveness of the PCR, they do make STR data analysis problematic. During development of a multiplex the following steps are to be considered (1) create compatible and specific primers that share identical melting temperatures, (2) ensure that primers do not share any complementary sequences to each other or any other human

DNA sequence other than its target region and (3) test each primer pair individually as uniplex before addition to the multiplex (Henegariu *et al*., 1997).

1.5.3. Optimization of Y-STR multiplexes

Henegariu and co-investigators (1997) were one of the first groups to work on finding the best approaches to multiplex PCR optimization. They investigated factors that could affect the performance of multiplex PCR (Henegariu *et al*., 1997). The group selected 22 primer pairs on chromosome 12, and 24 primer pairs on the Y-chromosome. A comprehensive analysis of PCR parameters was undertaken in order to determine whether these parameters could be used to improve multiplex PCR performance. They investigated cycling conditions in terms of cycle numbers, annealing temperature and annealing duration as well as extension temperatures and extension duration. The ideal concentration of PCR reagents such as Taq polymerase, PCR buffer, dNTPs and $MgCl_2$ as well as the ideal dNTP/ $MgCl_2$ ratio were investigated. The effect of individual primer concentrations on the yield of individual PCR products and the effect of additives such as BSA, glycerol and DMSO were also studied. Based on the results of the study, a generalized protocol was designed, with guidelines on how to approach the optimization of a multiplex PCR (Henegariu *et al*., 1997).

Another, more rigorous, optimization protocol was established by a group attempting to construct universal multiplex PCR systems for comparative genotyping (Wallin *et al*., 2002). The three areas of investigation were primer selection, PCR amplification and fluorescent allele detection. A primer selection process was based on extensive testing of numerous primers in uniplex and multiplex PCRs, as well as testing for amplification specificity and primer binding site mutations. Vast amounts of candidate primers were screened in uniplex PCRs for signal strength, and only those with the highest signal strength were chosen. Primer concentrations were varied to establish a performance window in which signal strength

reached a plateau. These primers were combined in a multiplex PCR at equimolar concentrations and the balance between the PCR products for each locus was investigated. By adjusting primer concentrations, signal strength was maximized so that all amplified products labelled with the same fluorescent dye would have similar peak heights (different fluorescent dyes have different emission wavelengths and some dyes will therefore appear brighter than others). Primers that could successfully be combined in this way were then tested for amplification specificity. A series of parameters were changed that would reduce the stringency of the multiplex PCR including reducing annealing temperatures, increasing DNA template, using single stranded Chelex extracted DNA and increased $MgCl_2$ concentrations were used to lower the stringency. Primers that produced non-specific artifacts at a lower stringency were rejected from being used in the multiplex. The primers were then investigated for primer binding site (pbs) mutations by conducting database searches. Where pbs mutations were found, primers were either re-designed to avoid mutations or degenerate primers added to the multiplex to compensate for the mutation. Optimal PCR cycling parameters and the most appropriate composition of PCR reagents was established empirically.

More recently a multiplex design strategy focusing on primer sequence analysis was used to construct a multiplex capable of amplifying up to 20 Y-STR loci (Butler *et al*., 2002). The reasoning behind the approach is as follows: Since multiplex PCR uses one set of cycling conditions, primers in a multiplex PCR should possess similar annealing temperatures. These primers should also show no significant interactions within themselves, between one another or with undesirable sequences in the DNA template used. This would ensure the efficiency of the multiplex PCR by limiting the occurrence of both primer-dimers and non-specific amplified artifacts. The approach was presented in a flow diagram (Figure 1.2) divided in two parts; the first part described an in silico primer design and, testing and the second part

described experimental amplification optimization. In this way, a successfully optimized multiplex PCR was created to amplify all the desired loci and achieve similar yields between the amplicons of all loci. The resulting profiles were free from non-specific artifacts and included amplicons that were easy to distinguish from other loci in the multiplex.

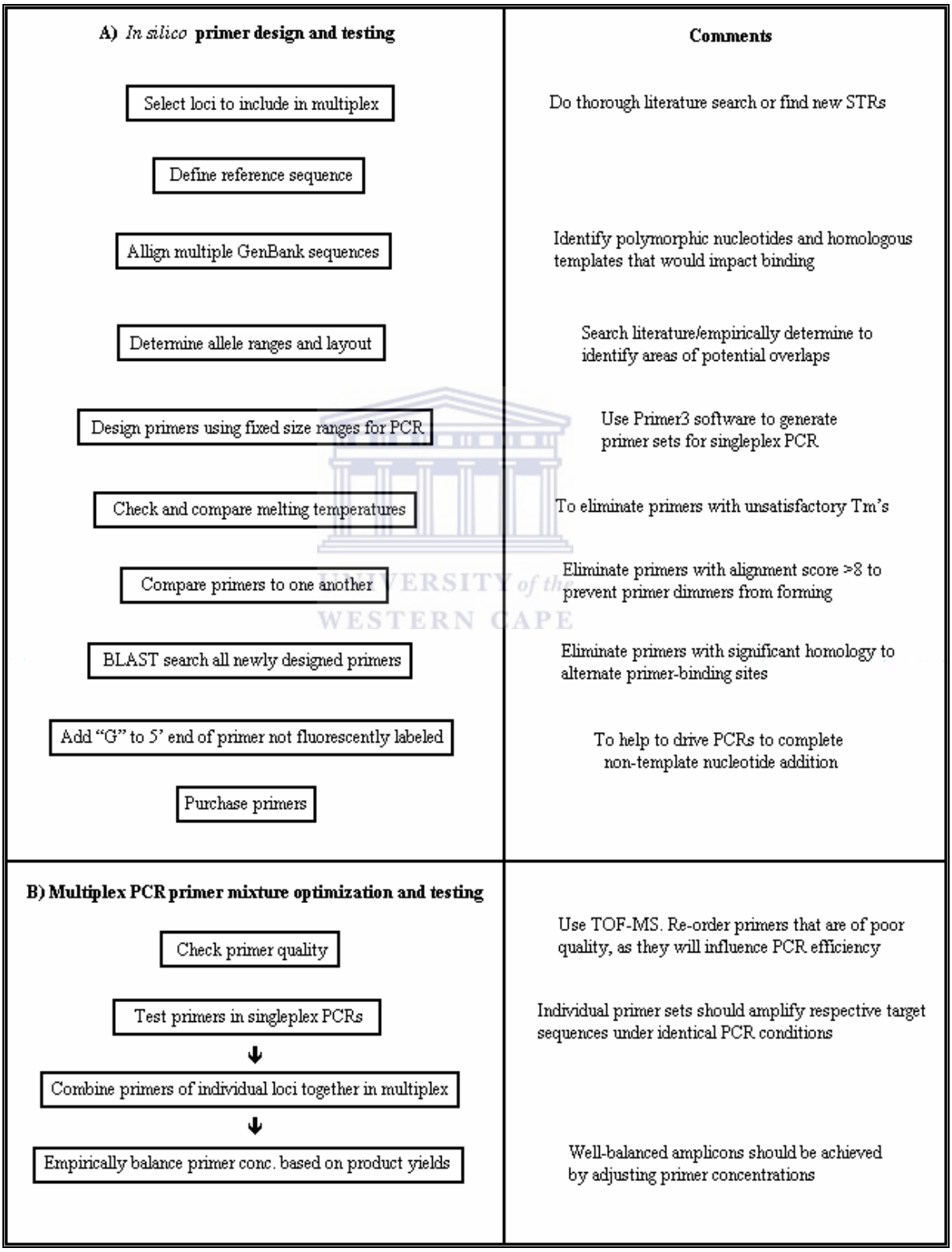

Figure 1.2. Primer design-based approach to multiplex optimization (Butler *et al.*, 2002)

1.6. Survey of Y-STR multiplexes

1.6.1. Commercial Multiplexes

A number of commercial Y-STR multiplexes have been developed throughout the years and made available to the forensic community as ready to use kits. Commercial Y-STR multiplex kits such as AmpFlSTR® Yfiler™ *(Applied Biosystems*), PowerPlex® Y (*Promega*), Y-PLEXTM12 (*Reliagene*) and Mentype Argus® Y-MH (*Biotype*) go through exacting tests and quality control to ensure that all genetic data obtained are reliable, accurate and precise. Commercial genotyping kits are provided with allelic ladders that facilitate accurate typing. These kits are designed, optimized and validated to be sensitive, male-specific and robust, needing only minuscule male DNA to obtain a complete profile. However, these kits could be very expensive and well beyond the means of many forensic laboratories especially in the developing countries such as South Africa.

The Y-plex™ 6 and Y-plex™ 5 (*ReliaGene Technologies, Inc. New Orleans, LA*) were some of the first commercial Y-STR multiplexes that became available. Y-plex™ 6 included DYS19, DYS390, DYS391, DYS393, DYS389II and DYS385, while Y-plex™ 5 amplifies DYS389 I/II, DYS392, DYS438 and DYS439 loci. Y-PLEXTM12, PowerPlex®Y and AmpFlSTR® Yfiler™ kits included the loci making up the MinHt, as well as DYS438 and DYS439. These Y-STR loci make up the extended haplotype as recommended by the SWGDAM. In the case of Y-PLEXTM12, however, the sex-determination locus Amelogenin was added to the multiplex, which serves as an internal control (Shewale *et al*., 2004). The expected sizes of amplicons generated with some of these commercial kits are presented in the Figure 1.3

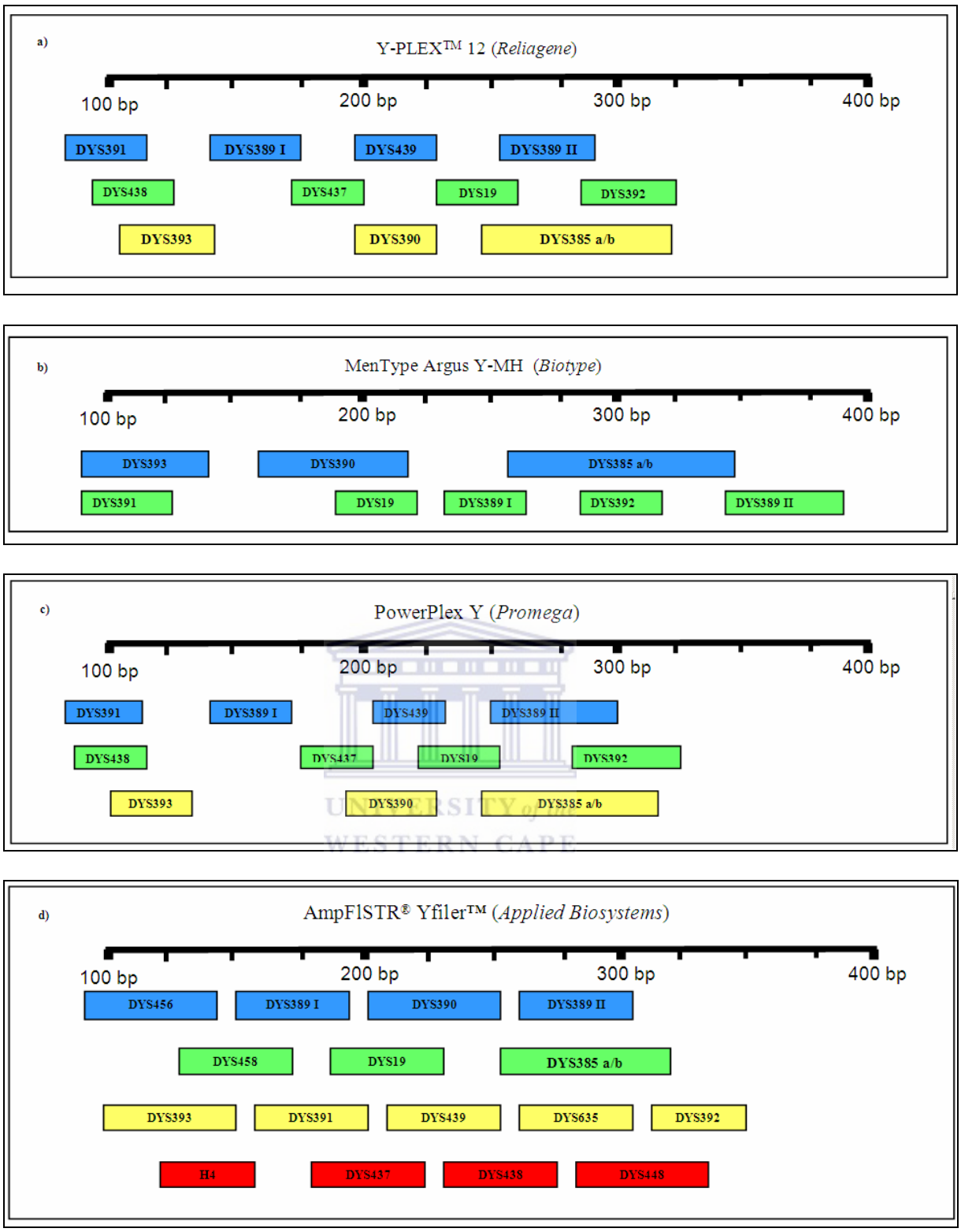

Figure 1.3. Predicted allele size range for Y-STRs amplified in (a) Y-PLEX™ 12 kit from *Reliagene*, (b) MenType Argus Y-MH from *Biotype*, (c) PowerPlex Y from *Promega* and (d) AmpFlSTR® Yfiler™ from *Applied Biosystems*.

1.6.2. Non-commercial multiplexes

The availability of the genome sequence, the understanding of the principles that affect multiplex PCR and the strategies used for effective optimization of multiplex PCR, has made the development of these multiplexes easier (Henegariu *et al*., 1997; Wallin *et al*., 2002; Butler *et al*., 2002). In fact, the number of non –commercial Y-STR multiplexes developed by forensic genetics laboratories is unknown as very little data is published about these multiplexes. Yet among of the largest Y-STR multiplexes available to the forensic community is the non-commercial megaplex developed by Butler and co-workers (2002), and the 21 Y-STRs megaplex developed by Hanson and Ballantyne (2004).

Butler and colleagues' megaplex, also know as the NIST Y-STR 20plex, amplifies 20 loci, using 17 primer pairs in a 5-dye detection system (dyes 6-FAM, VIC, NED and PET). The NIST Y-STR 20plex included the current extended haplotype Y-loci in addition to some the Y-STRs identified by White *et al*. (1999), Ayub *et al*. (2000) and Redd *et al*. (2002). The loci included in this megaplex are shown in Figure 1.4.

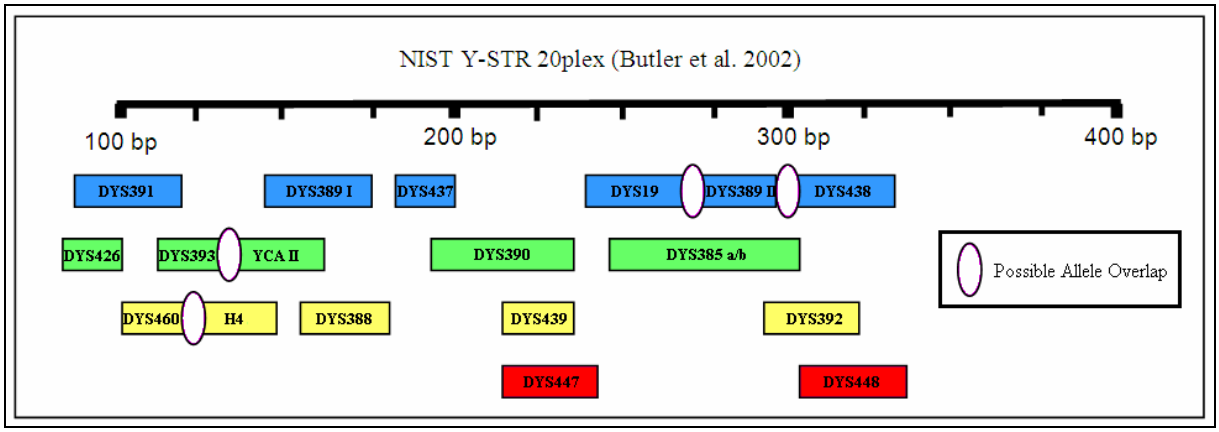

Figure 1.4. Predicted amplicon size range for loci amplified in 20-plex (Butler *et al*., 2002)

The largest Y-STR multiplex developed is the 21 Y-STRs megaplex from Hanson and Ballantyne (2004). The "minimal haplotype" was not included this multiplex, which made it significantly different from most multiplexes designed. The inclusion of the multi-copy loci

DYS527 and DYS464 made allele designation very challenging because of overlapping alleles.

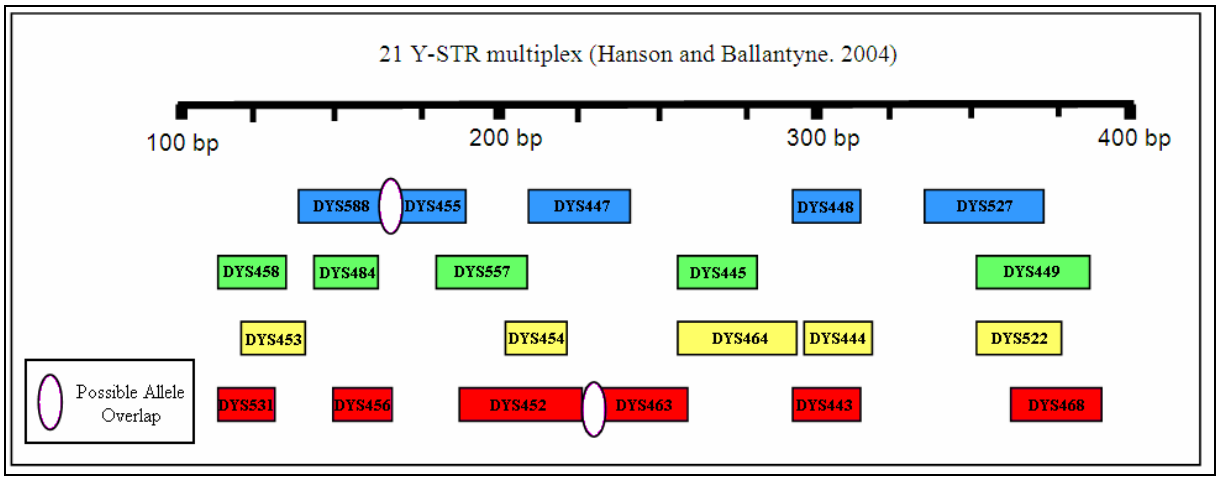

Figure 1.5. Predicted allele size range for Y-STRs amplified in 21 Y-STR megaplex (Hanson and Ballantyne, 2004)

1.6.3. UWC Y-Plexes

The Y-STR multiplexes developed by Leat *et al*. (2007) represents the culmination of a survey from Y-chromosome sequence data, identifying at least 191 polymorphic Y-STR loci (Ayub *et al*., 2000; Iida *et al*., 2001; Iida *et al*., 2002; Redd *et al*., 2002; Kayser *et al*., 2004). Several factors were considered during the selection of these Y-STRs. These included variability, the ease with which male specific PCR primers were designed, the extent to which stutter artifacts were generated during PCR, and copy number. It resulted in the development of the following three UWC-YPlex multiplexes, with the addition of DYS449 in each multiplex to ensure that consistent data is generated by each throughout population study genotyping.

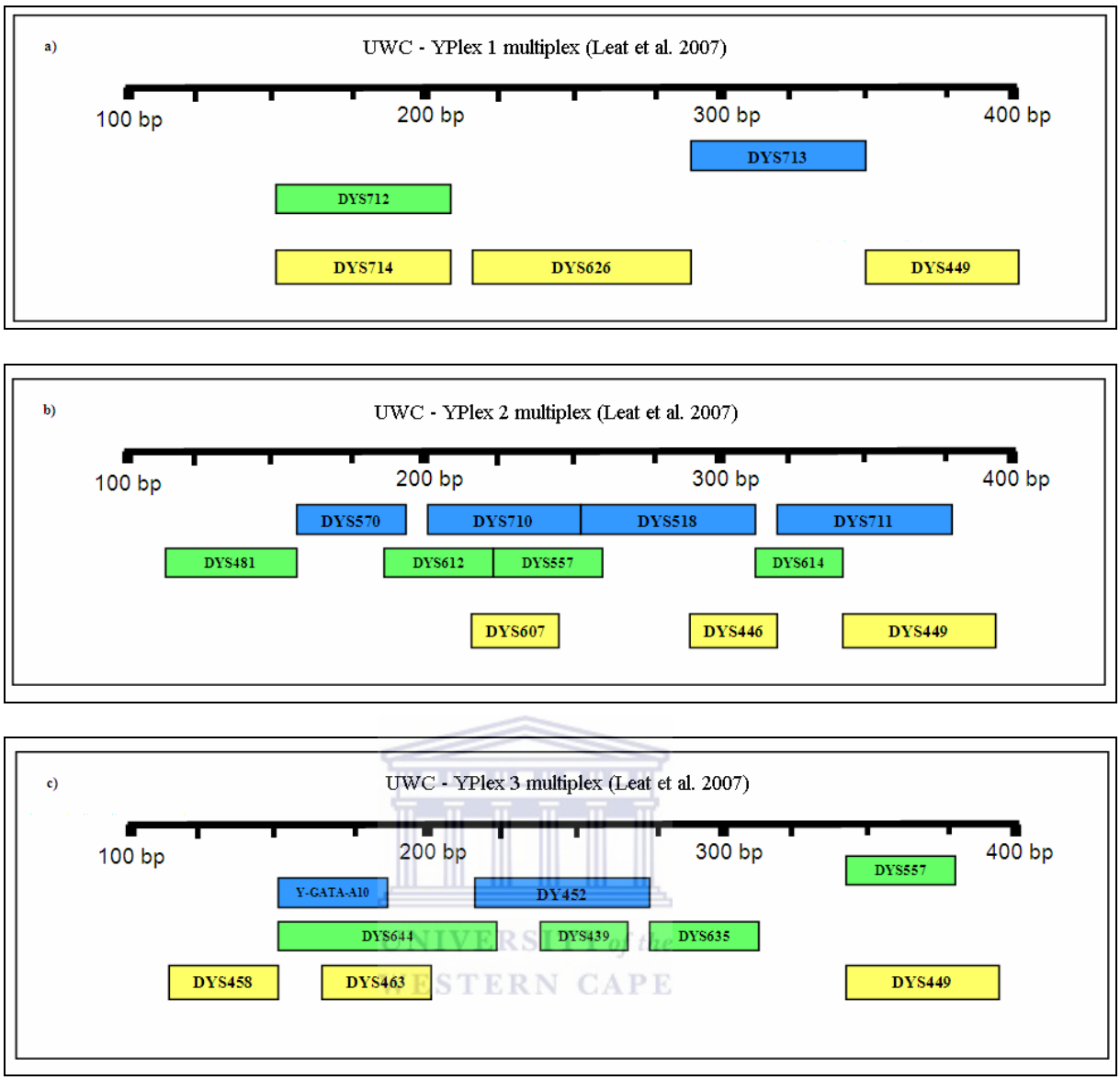

Figure 1.6. Predicted allele size range for Y-loci amplified in (a) UWC-Yplex 1, (b) UWC-Yplex 2 and (c) UWC-Yplex 3

1.7. Validation of Y-STR multiplexes

Y-chromosome genotyping multiplexes must be optimized to the point where it meets certain performance standards. Unless these standards are met or exceeded, its limitations are clearly defined; all data acquired by the particular multiplex will be questioned regarding its use in any study. There are several governing bodies that ensure that high genotyping 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). The SWGDAM organization has proposed

revised validation guidelines for the use of multiplex PCR genotyping systems in forensic casework. The guidelines define developmental validation as the demonstration of the accuracy, precision, and reproducibility of a procedure by the manufacturer and the internal validation as the in-house demonstration of the reliability and limitations of the procedure by a forensic DNA testing laboratory (SWGDAM, 2004; Daniels *et al*., 2004). The Internal validation procedures summarized in the following steps:

1.7.1. Sensitivity studies

The sensitivity procedure determines the minimum and maximum range of DNA quantities able to produce reliable, interpretable genotyped data. Using DNA of various quantities the PCR multiplex system is investigated by amplifying and genotyping each sample. The amount of DNA template used ranging from 0.005ng to 2.5ng. The procedure is repeated again after a minimum time interval of 7 days has passed, and the genetic profile results obtained are documented at each interval. Documentation of each success and failure of the PCR multiplex establishing the minimal level of DNA needed to generate a useful genetic profile. The repeated testing over time establishes the precision and accuracy of the results generated from the minimum DNA amount used (SWGADAM, 2004; Shewale *et al*., 2004)

1.7.2. Reproducibility studies

The reproducibility procedure evaluates the ability of the PCR multiplex system to reproduce complete, interpretable genetic profile data consistently on known genotyped samples after a minimum time interval of 7 days. The genetic profile of each known DNA sample documented after amplification, determining that identical genetic data is generated multiple times using the same DNA samples and PCR multiplex system (Prinz *et al*., 2001; SWGADAM, 2004).

1.7.3. Species specificity studies

The species specificity procedure is applied to human DNA multiplex systems to determine if any genotyped data can be generated by non-human DNA. The Non-human DNA is added to the PCR multiplex at an amount of usually 10-fold to that determined by the sensitivity studies. Adding such an abundant amount of non-human DNA to the PCR multiplex system serves two purposes, the first is to determine if any amplification occurs and the second is to see whether high amounts of animal DNA can possibly generate artifact peaks during the genotyping procedure (Prinz *et al*., 2001.; Shewale *et al*., 2004).

1.7.4. Environmental studies

The environmental studies records the genotyped results obtained from DNA extracted from biological samples deposited on various substrates and exposed to different environmental conditions. Determining if samples exposed to rain, air, sand, wind, etc. and extracted from different materials such as cotton, denim, leather, silk, nylon, wool, paper etc. will negatively influence the PCR multiplex in generating a useful, interpretable genetic profile (Prinz *et al*., 2001; Daniels *et al*., 2004)

1.7.5. DNA mixture studies

The DNA mixture procedure involves the use of different ratios of male-to-male and female-to-male DNA mixtures. In an attempt to mimic DNA mixtures typically found during forensic casework, the purpose of the study is to determine the range of detectable and forensically useful DNA mixtures. The female-to-male DNA mixture procedure adds equal to plentiful amounts of female DNA together with equal to minuscule amount of male DNA. The procedure documents the female-to-male DNA ratio range at which a useful genetic profile can be generated from a small amount of male DNA found within an abundant amount of female DNA. The second DNA mixture procedure uses DNA of two unrelated males at 1:1

and 1:100 of male DNA ratios. Determining if a useful genetic profile can be generated from any of the male-to-male DNA mixtures, the purpose is to differentiate one male DNA from the other male DNA found in the mixture (Daniels *et al*., 2004; Shewale *et al*., 2004; Mulero *et al*., 2006).

1.8. Population databases and population studies

Population databases are important for comparison purposes to understand how frequent or how rare a crime scene DNA profile may be in a particular population. Hundreds of publications in the literature contain information on DNA profile generated with genotype information from common STR loci across tens of thousands of individuals collected from various populations around the world (Butler *et al.,* 2006). Population comparison DNA databases are often generated by individual forensic laboratories to assess variation in common local populations. This is particularly important to locales that may have an isolated and inbred population within its jurisdiction (Butler *et al.,* 2006). South Africa has a multiracial and multiethnic society, and generating population data for its different subpopulations and ethnic groups for different forensic STR loci is indeed of extreme importance.

The primary goal of generating a population database is to find all ‘common’ alleles and sample these alleles multiple times in order to reliably estimate the alleles present in the population under consideration (Butler, 2005). A listing of observed alleles for Y-chromosome STRs that have been reported in numerous population studies from around the world can be obtained from the Y-chromosome Haplotype Reference Database (YHRD) (www.yhrd.org).

The Y-chromosome Haplotype Reference Database (YHRD) is a freely accessible website that aims for the collection of Y-STR data from as many different regions across the world. The website database contains 57208 Y-haplotypes with release 24 (as of September 2008) that represents the geographical and ethnical structure of diverse populations worldwide and assists in the development of Y-STRs capable of discriminating between unrelated males within a population. Typing of the MinHt in different populations and submissions of Y-STR data to the YHRD website has resulted in the formation of the most comprehensive database available, replacing three earlier database versions that collected European, Asian and US American Y chromosomes separately (Roewer *et al*., 2001). The pooling of Y-STR data supports presentation of Y-marker evidence in court cases (murder, assault, rape etc) by allowing fast and easy access to frequency estimates in a comprehensive documented way (Willuweit and Roewer, 2007).

1.9. Objectives of the study

Sexual assault is a significant problem facing the South African society. In this context, efficient but also affordable genotyping systems are needed for positive identification of criminals in incidences of sexual violence. In developing countries such as South Africa, the use and development of non-commercial genotyping systems is extremely important to cut the cost of the forensic DNA procedures. For this reason, the Forensic DNA Laboratory (FDL) at the University of the Western Cape has been involved in the validation as well as the development of non-commercial genotyping systems since 2002. The research groups of the FDL have also been generating population data for the different subpopulations and ethnic groups in the country for different loci and in particular those included in the minimal haplotype system. This was warranted by the fact that South Africa has a multiracial and multiethnic society.

Y-chromosome STR loci constituting the minimal haplotype are still the most widely used loci in investigating sexual assault cases despite the fact that DYS391 and DYS392 have shown low levels of polymorphism in Xhosa populations in Cape Town (Leat *et al.,* 2004). The minimal haplotype was, therefore, further investigated in the Cape Muslim population. The Cape Muslim population makes up a significant part of the population in the Cape Town metropolitan area. According to the last census (2001) (http://www.statssa.gov.za/census01), this population was estimated at 10% of the total population of the metropolitan area of Cape Town. Other unofficial estimates go as high as 15%.

The overall assessment of the minimal haplotype system has shown that it is still a useful genotyping system in investigating sexual assault case in many South African subpopulations. Therefore the exercise of internal validation of the minimal haplotype system was successfully carried out in the laboratory. The properties of additional novel and widely used STRs were investigated. Loci were successfully sequenced and allele nomenclature was assigned to them according to the ISFG guidelines.

UWC Y-plex 2 was originally constructed as an 11 Y-STR multiplex, comprised of Y-STRs that show gene diversity values of above 0.7 in South African sub-populations. Despite repeated attempts reliable sequences could only be generated for 8 of the Y-STRs found within the UWC Y-plex 2. Repeated attempts at sequencing of DYS446, DTS710 and DYS711 were made but no reliable sequences were obtained. Loci DYS449, DYS481, DYS518, DYS557, DYS570, DYS607, DYS612 and DYS614 were then used in a single multiplex reaction and named the eight Y-STR's system. Population data from Cape Muslim Community was generated for the eight Y–STR's system. The exercise of internal validation of this system was also successfully carried out in the laboratory.

Chapter 2

Analysis of the 'minimal haplotype' loci in the Cape Muslim population of South Africa

2.1. Introduction

Y-chromosome STR loci constituting the minimal haplotype (MinHt) are still the most widely used loci in investigating sexual assault cases despite the fact that DYS391 and DYS392 have shown low levels of polymorphism in Xhosa populations in Cape Town (Leat *et al.,* 2004). The minimal haplotype was, therefore, further investigated in the Cape Muslim population. The Cape Muslim population makes up a significant part of the population in the Cape Town metropolitan area. According to the last census (2001) (http://www.statssa.gov.za/census01), this population was estimated at 10% of the total population of the metropolitan area of Cape Town. Other unofficial estimates go as high as 15%. The coming to existence of this community is linked to the slavery trade, migration, colonialism, ancient trade routes in Africa, and the spread of Christianity and Islam in the sub-Saharan Africa. The earliest Cape Muslims were part of involuntary migration of slaves, political prisoners and criminals from Africa and Asia that lasted from 1652 to 1834 (Da Costa, 1994). In addition to migrations, this community has grown by combined and continuous multiplex processes of intermarriage, natural increase, conversion and blending with other communities (Shell, 1993; 2000). As a result, this community is characterized by a complex population structure and high level of mixed ancestries. The availability of population data for this community is valuable for a number of human genetics areas of research such as forensic genetics, genetic association studies, and anthropological projects.

2.2. Materials and Methods

2.2.1. DNA samples

Biological samples in the form of buccal (oral) swabs were collected from 98 healthy unrelated males randomly selected from the Cape Muslim population living in the Western Cape Province of South Africa. Total DNA from buccal cells was extracted using the Buccal Amp ™ Quick Extract DNA extraction kit (*Epicentre Technologies*) as per the manufacturers' recommendations.

2.2.2. PCR multiplex amplification

The MinHt Y-STR loci were amplified in a single multiplex reaction. The amplifications were performed in a final volume of 10µl containing 5-10 ng male DNA, 10X SuperTherm PCR buffer (with 15mM $MgCl_2$) (*Southern Cross Biotechnology*), 200µM dNTPs (*Roche*), and 5U/ µl of Super-Therm Taq polymerase (*Southern Cross Biotechnology*). Primers were synthesized by *Applied Biosystems* using previously reported sequences (Butler *et al*. 2002). The primer sequences, corresponding fluorescent dye labels and the final concentration, at which each primer was used, are shown in table 2.1.

Table 2.1. Primer sequences, dye labels and final concentrations used in the MinHt multiplex

Y-STR	Primer Sequences (5' to 3')	[Primer]
DYS 19 F	**NED** act act gag ttt ctg tta tag tgt ttt t	1.8 µM
DYS 19 R	gtc aat ctc tgc acc tgg aaa t	1.8 µM
DYS 385 F	**VIC** agc atg ggt gac aga gct a	0.6 µM
DYS 385 R	gcc aat tac ata gtc ctc ctt tc	0.6 µM
DYS389 F	**6FAM** cca act ctc atc tgt att atc tat g	1.3 µM
DYS389 R	gtt atc cct gag tag tag aag aat g	1.3 µM
DYS390 F	**VIC** tat att tta cac att ttt ggg ccc	0.2 µM
DYS390 R	gtg aca gta aaa tga aaa cat tgc	0.2 µM

Table 2.1. (*Continued*)

Y-STR	Primer Sequences (5' to 3')	[Primer]
DYS391 F	**6FAM** ttc aat cat aca ccc ata tct gtc	0.2 μM
DYS391 R	gat aga ggg ata ggt agg cag gc	0.2 μM
DYS392 F	**NED** tag agg cag tca tcg cag tg	1.8 μM
DYS392 R	gac cta cca atc cca ttc ctt	1.8 μM
DYS393 F	**VIC** gtg gtc ttc tac ttg tgt caa tac	0.4 μM
DYS393 R	gaa ctc aag tcc aaa aaa tga gg	0.4 μM

PCR thermocycling was conducted using a GeneAmp 2700 thermocycler (*Applied Biosystems*). The multiplex cycling parameters were as follow: denaturation at 94°C for 10 min, followed by 33 cycles of 94°C for 1 min, 60°C for 1 min and 72°C for 1 min, and a final step at 60°C for 45 minutes remaining at a holding temperature of 25°C . PCR products were stored in the dark at 4°C.

2.2.3. PCR amplicon analysis

Amplified DNA fragments were analyzed using an ABI 377 Genetic Analyzer (*Applied Biosystems*). Each sample was prepared for electrophoresis by mixing 1 μl unpurified PCR products with 1 μl of loading mixture. The loading mixture consists of 5 μl Hi-Di formamide, 1 μl Dextran Blue and 1.5μl GS500™ ROX™ internal size standard (*Applied Biosystems*). The mixture was denatured at 95°C for 5 minutes and then immediately snap-cooled on ice for 3 minutes until the amplicons were loaded on the 377 Genetic Analyzer and separated on polyacrylamide gels as per the manufacturer's instructions. DNA sequence data was collected with the ABI 377 collection software (*Applied Biosystems*). Analysis of the data was done using GeneScan® 3.7 (*Applied Biosystems*) software. Amplicons were then converted to allele numbers using the Genotyper® 3.7 (*Applied Biosystems*).

2.2.4. Allelic ladder construction

DNA samples from a previous study by conducted by Leat and co-workers (2004) were used to construct the allelic ladder for the present study. Samples representing the most common alleles were mixed and amplified for each MinHt primer set. This resulted in a single completed PCR reaction containing a range of reference amplicons for each MinHt locus. The allele designations were matched to amplicon base pair sizes for each locus based upon amplification of reference alleles. The allelic ladder was constructed using these combinations of alleles at each locus to facilitate consistent allele typing using Genotyper® 3.7 (*Applied Biosystems*).

2.2.5. Statistical data analysis

Allele frequencies were calculated by direct counting. Gene diversity was calculated as (Nei, 1987)

$$GD = n\,(1- \sum p_i^2) / (n-1)$$

where n is the sample size and pi is the frequency of the i th allele. Haplotype diversity (HD) was calculated with the same equation using haplotype frequencies. Standard errors for HD estimates were calculated according to the following equation (Nei, 1987)

$$SE = \sqrt{2[\sum p_i^3 - (\sum p_i^2)^2] / n}$$

Discrimination capacity was determined by dividing the number of observed haplotypes to the number of sampled individuals. Gene diversity values smaller than 0.5 were considered low and greater than 0.7 were considered high.

2.3. Results and Discussion

2.3.1 MinHt multiplex performance

A total of 100 buccal swab samples were collected from male donors from the Cape Muslim community. Ninety eight samples were successfully genotyped. The concentration of DNA extracted from these samples ranged from 20ng/ul to 150ng/ul. Amplified fragments, run on an ABI 377 instrument, typically generated electropherograms that could easily be interpreted. An example of such an electropherogram for the MinHt multiplex reaction is presented in Figure 2.1.

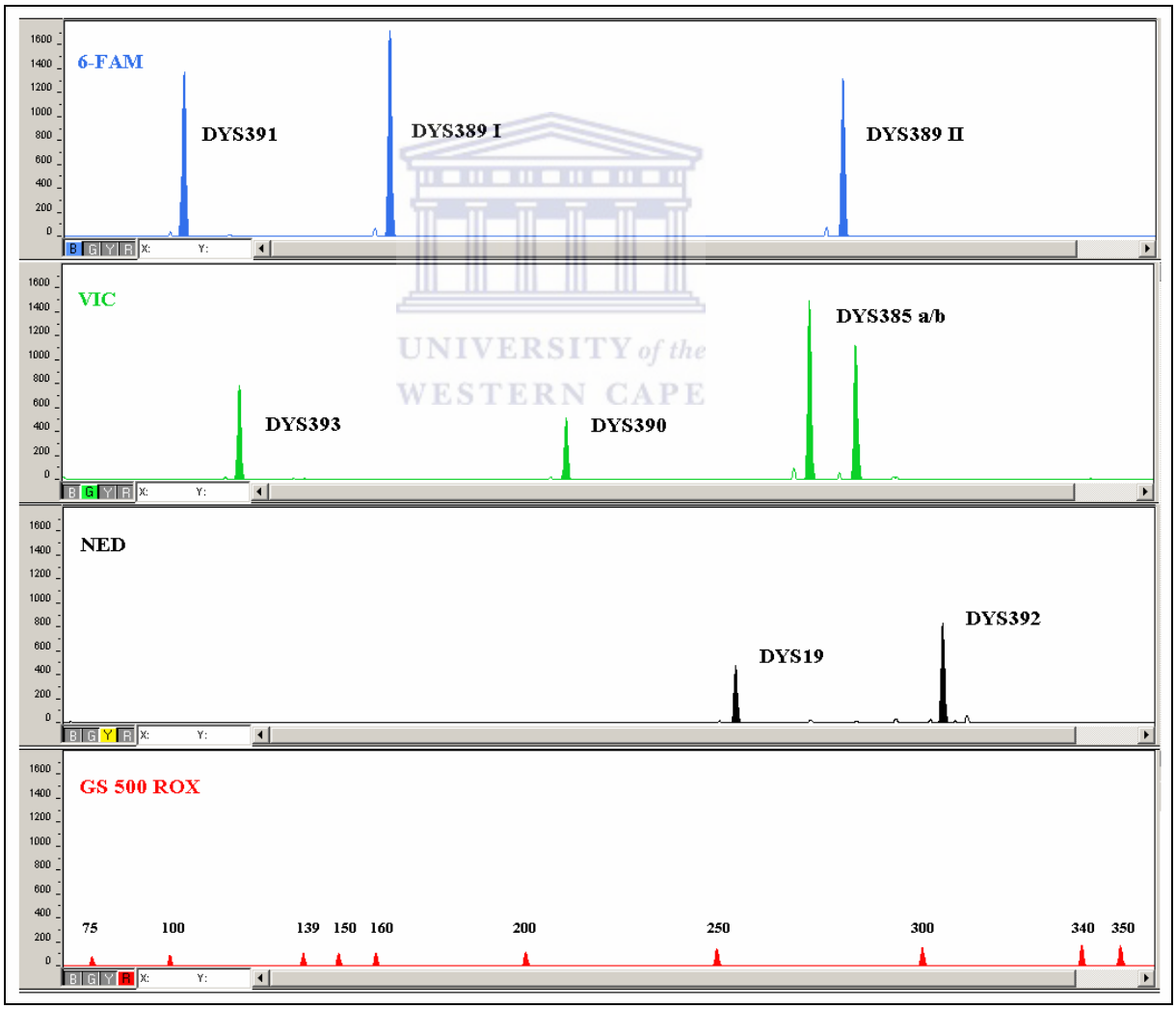

Figure 2.1. An example of a typical electropherogram of the MinHt multiplex reaction

2.3.2 Analysis of allele and haplotype frequencies

The GD values for the population ranged from 0.447 for DYS391 to 0.957 for DYS385 (Table 2.2). The highest GD value for a single-copy locus was recorded for DYS390 (0.802). The lowest gene diversity was obtained for DYS391 (0.448). This is to be expected as 70% of the samples shared DYS391's allele 10 (Table 2.2). Similar results were obtained for DYS391 in other South African populations (Leat *et al*., 2004; 2007). The proportion on individuals sharing DYS391's allele 10 reached 83% for Asian Indian, 60% for Mixed Ancestry and as high as 93% for Xhosa. Unique haplotypes were obtained for 70 individuals, 13 haplotypes were found to be shared by more than one individual in the population (Table 2.2 and Table 2.3), and the haplotype diversity was 99.64%. The discrimination capacity was 84.69% (Table 2.3).

Table 2.2. Allele frequencies and gene diversity at 9 Y-chromosome STRs in the Cape Muslim population (n=98) typed with the 'minimal haplotype' system

Allele	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	H	DYS385a/b
09					0.010	0.010		08-14	0.010
10					**0.704**	0.041		09-16	0.010
11		0.020			0.245	**0.592**	0.041	10-13	0.010
12		0.153			0.041	0.082	0.327	10-15	0.010
13	0.041	**0.612**				0.194	**0.459**	11-11	0.020
14	**0.459**	0.214				0.071	0.133	11-12	0.010
15	0.245					0.010	0.041	11-13	0.041
16	0.153							11-14	**0.163**
17	0.102							11-15	0.020
18								11-16	0.010
19				0.010				11-17	0.010
20				0.010				12-12	0.020
21				0.071				12-14	0.031
22				0.184				12-17	0.031
23				0.245				12-18	0.041
24				**0.286**				12-19	0.010
25				0.153				12-20	0.010
26				0.041				13-13	0.051
27			0.010					13-16	0.061

Table 2.2. (*Continued*)

Allele	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	H	DYS385a/b
28			0.153					13-17	0.020
29			0.276					13-18	0.041
30			**0.398**					13-19	0.010
31			0.102					13-21	0.010
32			0.061					14-14	0.051
								14-15	0.020
								14-16	0.020
								14-17	0.020
								14-18	0.020
								15-15	0.010
								15-16	0.031
								15-17	0.010
								15-18	0.020
								15-19	0.010
								16-16	0.010
								16-17	0.051
								16-18	0.031
								16-19	0.010
								17-17	0.010
								17-18	0.020
GD	0.701	0.561	0.736	0.802	0.447	0.605	0.668		0.957

Highest frequencies are shown in bold

GD = Gene Diversity Index

H = Haplotypes

Table 2.3. List of 9 Y chromosome STR haplotypes observed in the Cape Muslim population (n=98) typed with the 'minimal haplotype' sytem

H	DYS19	DYS389 I	DYS389 II	DYS390	DYS391	DYS392	DYS393	DYS385 a,b	*n*
H01	**14**	**13**	**30**	**24**	**10**	**13**	**13**	**11,14**	**2**
H02	**15**	**13**	**31**	**25**	**10**	**11**	**13**	**11,14**	**2**
H03	14	13	29	24	10	11	13	14,16	1
H04	14	14	30	24	10	14	13	11,14	1
H05	15	14	32	23	10	11	13	11,11	1
H06	14	12	28	23	10	11	12	14,14	1
H07	14	13	29	24	11	13	14	11,12	1
H08	17	13	30	21	10	11	13	12,20	1
H09	**17**	**13**	**29**	**25**	**11**	**13**	**13**	**14,14**	**2**

Table 2.3. ***(Continued)***

H	DYS19	DYS389 I	DYS389 II	DYS390	DYS391	DYS392	DYS393	DYS385 a,b	*n*
H10	17	14	30	22	10	10	13	16,16	1
H11	**14**	**12**	**28**	**22**	**10**	**11**	**13**	**14,15**	**2**
H12	15	13	30	24	11	11	14	11,14	1
H13	16	13	30	24	10	11	14	11,14	1
H14	16	13	29	22	11	11	12	16,18	1
H15	14	13	30	23	10	11	12	12,18	1
H16	16	13	29	25	11	11	13	11,15	1
H17	15	13	31	24	11	11	13	08,14	1
H18	15	11	30	20	10	12	14	15,15	1
H19	**14**	**14**	**30**	**23**	**10**	**11**	**14**	**12,17**	**3**
H20	**14**	**13**	**30**	**24**	**11**	**11**	**13**	**13,18**	**2**
H21	16	13	30	26	11	11	13	11,14	1
H22	16	13	30	25	12	11	13	11,14	1
H23	**17**	**13**	**30**	**26**	**11**	**11**	**13**	**11,14**	**3**
H24	15	13	30	25	11	11	14	11,14	1
H25	14	13	29	24	10	13	13	10,15	1
H26	**14**	**12**	**28**	**22**	**10**	**14**	**11**	**13,16**	**2**
H27	17	14	31	21	10	9	13	16,17	1
H28	15	12	28	21	10	11	12	11,17	1
H29	15	12	29	22	10	11	13	14,14	1
H30	16	13	29	22	10	11	12	15,17	1
H31	14	13	29	25	10	14	12	17,17	1
H32	14	14	30	21	10	10	15	13,21	1
H33	14	13	29	25	10	11	12	12,18	1
H34	15	14	32	25	10	11	13	12,14	1
H35	16	12	28	24	10	11	13	11,16	1
H36	15	13	29	23	10	14	12	09,16	1
H37	14	14	31	22	10	11	14	10,13	1
H38	**14**	**13**	**29**	**24**	**11**	**13**	**13**	**11,13**	**2**
H39	14	13	29	19	11	13	13	13,13	1
H40	14	13	30	23	10	11	12	13,13	1
H41	16	13	29	23	11	10	12	12,18	1
H42	15	14	32	23	10	12	14	15,16	1
H43	14	13	29	23	9	11	12	13,16	1
H44	16	12	28	24	11	11	12	14,18	1
H45	15	13	30	23	10	12	15	15,16	1
H46	**14**	**13**	**30**	**24**	**10**	**13**	**13**	**12,14**	**2**
H47	16	13	29	22	11	11	12	16,17	1

Table 2.3. ***(Continued)***

H	DYS19	DYS389 I	DYS389 II	DYS390	DYS391	DYS392	DYS393	DYS385 a,b	*n*
H48	15	14	31	23	10	11	12	13,17	1
H49	14	13	29	22	10	11	12	14,17	1
H50	14	13	29	22	10	14	11	15,16	1
H51	**14**	**13**	**30**	**23**	**10**	**11**	**12**	**13,16**	**2**
H52	14	12	28	24	10	11	13	14,16	1
H53	14	12	28	22	10	14	11	14,18	1
H54	16	11	27	24	10	11	12	13,18	1
H55	14	13	29	23	10	12	12	13,19	1
H56	15	14	30	22	10	11	12	16,19	1
H57	14	13	30	23	10	11	12	15,19	1
H58	16	13	30	21	10	12	15	16,18	1
H59	**15**	**14**	**32**	**25**	**12**	**11**	**13**	**11,13**	**2**
H60	16	13	31	22	11	11	12	17,18	1
H61	17	14	31	25	10	11	13	11,11	1
H62	14	14	30	23	10	11	12	13,13	1
H63	14	12	29	22	10	11	14	14,14	1
H64	15	14	30	22	10	10	15	12,18	1
H65	16	13	29	24	10	15	12	11,15	1
H66	15	12	28	24	11	13	13	12,19	1
H67	15	13	28	24	10	13	13	15,18	1
H68	**13**	**13**	**30**	**24**	**10**	**13**	**13**	**13,13**	**2**
H69	13	13	31	24	10	11	12	16,17	1
H70	14	13	29	24	12	13	13	11,14	1
H71	14	13	30	23	11	13	13	11,14	1
H72	13	14	29	23	10	13	12	16,17	1
H73	14	13	30	23	10	11	12	13,18	1
H74	14	13	29	23	10	11	12	13,16	1
H75	17	12	28	23	10	12	14	12,12	1
H76	14	12	28	24	10	12	12	13,17	1
H77	14	13	30	25	10	11	13	16,18	1
H78	15	13	29	22	11	12	12	14,17	1
H79	15	13	30	21	10	11	14	17,18	1
H80	15	14	32	24	10	13	13	12,12	1
H81	16	13	31	25	10	11	13	11,14	1
H82	15	13	28	21	10	11	12	16,17	1
H83	14	14	30	23	10	11	12	15,18	1

Haplotype Diversity= 99.64%; Discrimination Capacity= 84.69%

Standard Error for HD is 0.00086

n = number of individuals observed for each haplotype

Table 2.4. Nine Y Chromosome STRs haplotypes not unique typed in 98 individuals from the Cape Muslim population using the minimal haplotype system

Haplotypes **DYS19, DYS389-I, DYS389-II , DYS390, DYS391, DYS392, DYS393, DYS385**[a]	**n**	**frequency**
17-13-30-26-11-11-13-(11,14)	3	0.031
14-14-30-23-10-11-14-(12,17)	3	0.031
17-13-29-25-11-13-13-(14,14)	2	0.020
15-13-31-25-10-11-13-(11,14)	2	0.020
15-14-32-25-12-11-13-(11,13)	2	0.020
14-13-30-24-10-13-13-(11,14)	2	0.020
14-12-28-22-10-11-13-(14,15)	2	0.020
14-13-30-24-11-11-13-(13,18)	2	0.020
14-12-28-22-10-14-11-(13,16)	2	0.020
14-13-29-24-11-13-13-(11,13)	2	0.020
14-13-30-24-10-13-13-(12,14)	2	0.020
14-13-30-23-10-11-12-(13,16)	2	0.020
13-13-30-24-10-13-13-(13,13)	2	0.020

[a] Haplotypes for the duplicated locus DYS385 are presented in parenthesis.

n = number of individuals observed for each haplotype.

2.3.3. Comparison of allele frequencies among South African sub-populations

For the single-copy loci, the highest GD values were recorded for DYS390 in most of the South African sub-populations (Table 2.5). The DYS19 showed the highest GD value in the Xhosa population instead. The lowest GD values were recorded for DYS393 in the Caucasian populations (Afrikaner Caucasian [0.322] and English Caucasian [0.325]), and for DYS 391 in the Asian Indian (0.279), the Mixed Ancestry (0.510) and the Cape Muslim population (0.447). For the Xhosa population, the GD value for DYS392 as low as 0.080. In case of the Cape Muslim population, most loci showed GD values higher than 0.5, which suggests the usefulness of the minimal haplotype system in this population.

Table 2.5. Comparison of Gene Diversity values for the MinHt loci among South African sub-populations

Population Group	*n*	DYS19	DYS389I	DYS389II	DYS393	DYS392	DYS391	DYS390	DYS385
Afrikaner Caucasian [b]	108	0.583	0.497	0.589	0.335	0.635	0.518	0.712	0.834
Asian Indian [a]	88	0.690	0.620	0.616	0.692	0.479	0.279	0.734	0.935
Mixed Ancestry [b]	114	0.653	0.611	0.625	0.509	0.652	0.510	0.768	0.905
English Caucasian [a]	100	0.460	0.530	0.437	0.325	0.574	0.544	0.662	0.820
Xhosa [a]	99	0.700	0.680	0.683	0.550	0.080	0.130	0.610	0.918
Cape Muslim	98	0.701	0.561	0.736	0.668	0.605	0.447	0.802	0.957

[a] Populations investigated by Leat *et al*. 2004; 2007

[b] Populations investigated by Ehrenreich *et al*. 2008

2.3.4. Comparison of common haplotypes with those recorded in the YHRD

The YHRD website contains 65165 Y-chromosomal haplotypes and is the 25th database version. The website database has become an important tool for use in comparing MinHt Y-STRs from different populations across the world. The website standardizes Y-STR nomenclature by ensuring the successful completion of a quality assurance exercise. All Y-STR population data accepted is done so with the understanding that a population is defined as a group of more than 50 individuals living in the same area. When using Y-STRs it is advisable to understand the data obtained with a look towards the origin of the population (Willuweit and Roewer, 2007).

The haplotypes observed in the Cape Muslim population were compared with those recorded in the YHRD. Haplotype 38 (indicated as H38 in table 2.2), which was shared by 2 individuals from this population, matched 138 males in the YHRD including 3 individuals from South Africa (one English-speaking Caucasian and two Mixed Ancestry males) (Leat *et al*., 2004). An additional 10 haplotypes from this population matched haplotypes reported by Leat and co-workers in their studies of local populations in Cape Town (Leat *et al*., 2007).This includes matches for H03, H11, H27, H34, H64 and H73 found in the Mixed Ancestry population, matches for H51 and H71 in English-speaking Caucasians, and finally matches for H11, H45, H71 and H77 in Afrikaner Caucasians.

The YHRD analysis and comparison with haplotypes reported worldwide showed that 19% of the Cape Muslim community matched haplotypes previously reported from the Sub-Saharan Africa (South Africa and Angola). Twenty two percent of these haplotypes were also matched to haplotypes reported from all over Africa (South Africa, Angola, Ethiopia, Reunion, Guinea-Bissau, Tunisia, Algeria, Equatorial Guinea, and Morocco). Matches were also found for 17% of Cape Muslim haplotypes with haplotypes from the Middle East (Egypt, Iran, Iraq, Turkey, Syria, and Armenia). Final matches were found for 16% of the haplotypes with East Asians (predominately with Indian-and Han Chinese heritage in countries Nepal, Singapore, Taiwan, India, China, Vietnam and Malaysia).

2.4. Summary

The Cape Muslim population generally exhibited high GD values among all the South African populations. These values were higher than 0.5 for most loci, and ranged from 0.447 for DYS391 to 0.957 for DYS385. The highest number of alleles in most loci was also recorded in this population. This suggests the usefulness of all these loci in forensic casework in this population. It is therefore recommended that the minimal haplotype system should go through the validation procedures in the laboratory and consequently be used in forensic casework.

Chapter 3

Internal validation of the MinHt multiplex

3.1. Introduction

Y-chromosome genotyping multiplexes must be optimized to a point where it meets certain performance standards. Unless these standards are met or exceeded, its limitations are clearly defined; all data acquired by the particular multiplex will be questioned regarding its use in any study. There are several governing bodies that ensure that high genotyping 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) (SWGDAM. 2004; Daniels *et al*., 2004).

The internal validation procedure refers to a less extensive study whose purpose is to demonstrate the laboratory's capability in performing accurate and precise genotyping using pre-existing multiplexes. The in-house validation exercises include: (1) establishing that the genotyping system is sensitive and performs consistently using freshly prepared and stored DNA, (2) that the multiplex assays yield consistent, accurate and precise results over a broad time period, and (3) that the limitations of the genotyping system, used to analyze samples similar to those encountered in forensic casework, are identified (Gill *et al*., 2001; Prinz *et al*., 2001; Shewale *et al*., 2004).

The study of the minimal haplotype system described in chapter 1, as well as previous studies by Leat and co-investigators (2004 and 2007) has shown the usefulness of this non-commercial genotyping system in investigating sexual assault cases in many South African subpopulations. The in-house validation of this system was therefore successfully carried out, and it is described in this chapter.

3.2. Materials and Methods

3.2.1. DNA samples

Biological samples used in this study were obtained in the form of buccal swabs or whole blood samples and donated by volunteers from diverse South African subpopulations. Total DNA from buccal cells was extracted using the Buccal Amp ™ Quick Extract DNA extraction kit (*Epicentre Technologies*) as per the manufacturers' recommendations. In case of whole blood samples, DNA was extracted by salting out proteins as described by Lahiri and Nurnberger, (1991) (see Appendix).

3.2.2. PCR amplification conditions

The MinHt Y-STR loci were amplified in a single multiplex reaction as described previously (section 2.2.2).

Table 3.1. Dye labels, PCR product sizes, repeat motives, and Gene Bank accession numbers for the MinHt Y-STR (Butler *et al.*, 2002)

Y-STRs	Dye	PCR Product size (bp)	Repeat motif	GenBank® accession
DYS 19	NED	233 - 269	TAGA	AC017019
DYS 385 a/b	VIC	242 - 306	GAAA	AC022486
DYS389 I	6FAM	143 - 175	[TCTG][TCTA]	AC004617
DYS389 II	6FAM	263 - 295	[TCTG][TCTA]	AC004617
DYS390	VIC	189 – 233	[TCTA][TCTG]	AC011289
DYS391	6FAM	93 – 121	TCTA	AC011302
DYS392	NED	290 – 320	TAT	AC011745
DYS393	VIC	109 – 133	AGAT	AC006152

3.2.3. PCR amplicon analysis

Amplified DNA fragments were analyzed as previously described (section 2.2.3).

3.2.4. Validation procedures

3.2.4.1. Sensitivity studies

Varying amounts of male DNA, ranging from 0.005ng to 2.5ng, were used as template in the MinHt multiplex in order to determine the minimum amount of DNA needed to obtain full interpretable Y-STR profiles. The quantities of male DNA used were 0.005ng, 0.05ng, 0.125ng, 0.25ng, 0.5ng and 2.5ng. Each quantity of DNA was tested three times; each test was performed after a minimum of 7 days from the previous test. The aim of these tests was to ensure the accuracy and consistency of Y-STR profiles generated each time for each DNA quantity. The absence of any 4-dye peak after amplification, which is also known as allele dropout, was recorded.

3.2.4.2. Reproducibility studies

The procedure involved the use of four male DNA samples amplified and run on the ABI 377 sequencer, followed by GeneScan® version 3.7 analyses. The process was repeated on three occasions with a minimum of 7 days elapsing between each experiment. Use of fresh PCR reagents at each separate experiment is recommended. The experiment was done in order to determine whether consistent full Y-STR profiles are generated for the multiplex each time over a period of time.

3.2.4.3. Species specificity studies

Non-human whole blood samples were tested with the MinHt multiplex. This included samples from canine (dog), feline (cat), bovine (cow), equine (horse) and ovis (sheep). The Lahiri and Nurnberger method (1991) was used to extract DNA from all the blood samples.

The aim of the study was to determine whether the MinHt would also amplify non-human DNA. In this procedure, the minimum detectable amount of human DNA (0.5ng) was used together with an abundant amount of non-human DNA (5ng) in the multiplex reactions.

3.2.4.4. Environmental studies

Samples were prepared by pipetting 50µl of male human blood onto different fabrics and allowing it to dry overnight. The samples were then exposed to environmental conditions in an enclosed garden area, for a period of time ranging from 7 to 30 days. DNA was extracted from fabrics using the QIAamp Tissue Kit (Qiagen, Valencia. CA. DNA extractions were performed according to the manufacturer's recommendations. DNA was quantified then used in the multiplex reactions to determine whether full Y-STR profiles were obtained after subjecting the samples to environmental conditions. This was done to imitate the environmental conditions that influence the integrity of biological evidence collected from a crime scene.

3.2.4.5. DNA mixture studies

Forensic casework in sexual assault crimes requires the differentiation between male/male and male/female mixtures that might be present in the biological evidence collected in the crime scene. DNA mixtures of varying ratios were used in order to evaluate the performance of MinHt multiplex to generate full and interpretable Y-STR profiles in conditions similar to those encountered in crime scene situations.

3.2.4.5.1. Male/female mixture

Varying amounts of female DNA were added to 0.5ng male DNA to make up the DNA mixtures at different ratios. Male/female ratios tested were 1:1 (0.5 ng male DNA/ 0.5 ng female DNA), 1:2 (0.5ng male DNA/ 1 ng female DNA), 1:5 (0.5ng male DNA/ 2.5ng female DNA) and 1:10 (0.5ng male DNA/ 5ng female DNA).

3.2.4.5.2. Male/male mixture

For the male/male mixtures, DNA from two males (male DNA 1 and male DNA 2) was combined at ratios of 1:1, 1:2, 1:5 and 1:10 followed by amplification. Male/male ratios tested were 1:1 (0.5 ng male DNA 1/ 0.5 ng male DNA 2), 1:2 (0.5ng male DNA 1 / 1 ng male DNA 2), 1:5 (0.5ng male DNA 1/ 2.5ng male DNA 2) and 1:10 (0.5ng male DNA 1/ 5ng male DNA 2).

3.3. Results and Discussion

3.3.1. Minimal Sensitivity

The amount of amplifiable DNA available during forensic casework is limited. The purpose of the sensitivity study is to determine the minimum amount of male DNA needed to generate a complete Y-STR profile. The sensitivity study was performed by using varying DNA amounts from one genomic source in a multiplex reaction. Amplification reactions generated no profiles for the 0.005ng DNA template and partial profiles from the 0.05ng, 0.125ng, and 0.25ng DNA amounts. Full Y-STR profiles were obtained from the 0.5ng and 2.5ng DNA templates used (figure 3.1). GeneScan® software analysis on sensitivity multiplex reactions with full Y-STR profiles showed consistent base pair sizes for each locus (as shown in table 3.2).The sensitivity study demonstrating that a minimal amount of 0.5ng DNA template will generate a full, accurate and consistent Y-STR profile.

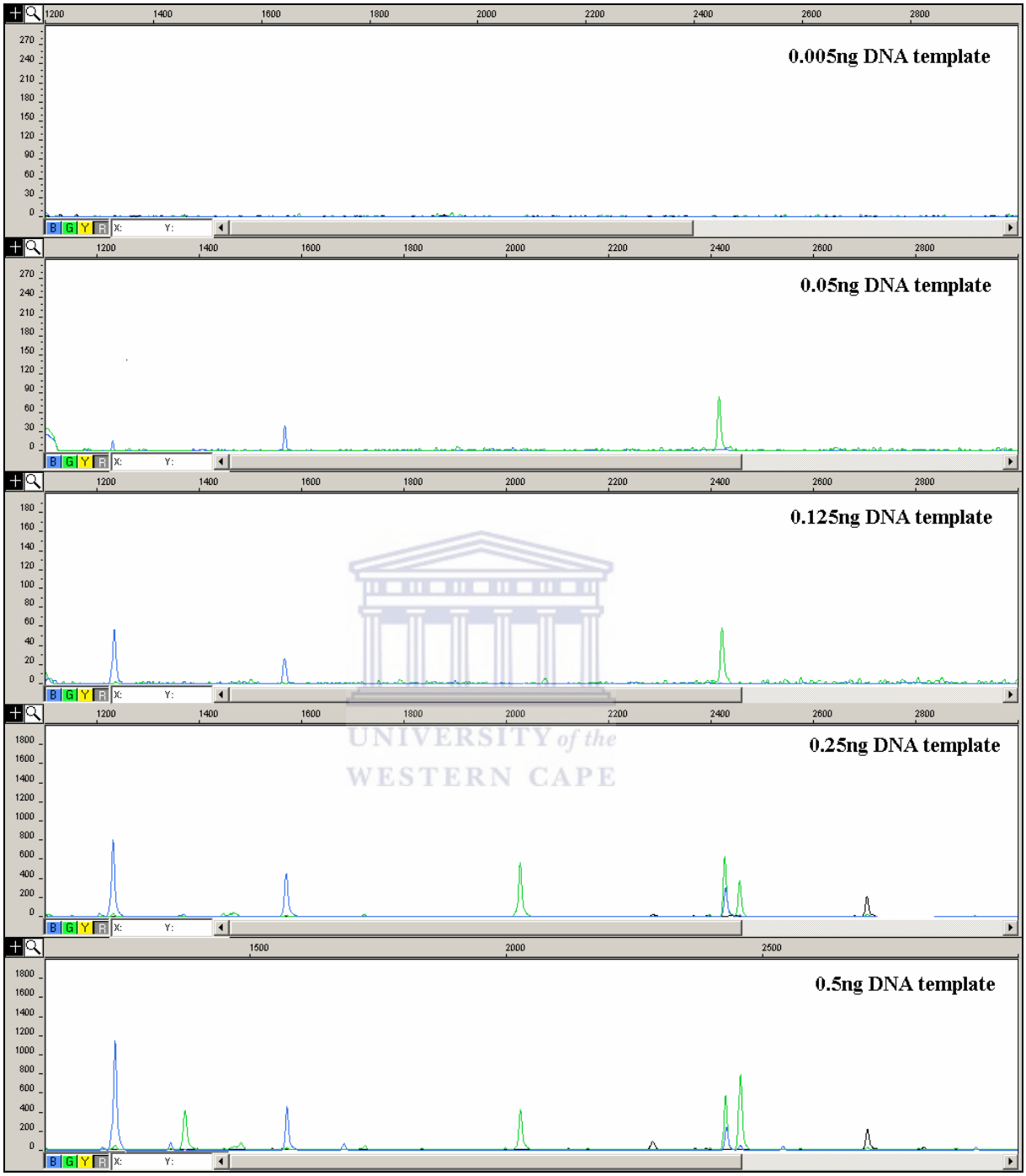

Figure 3.1. Y-STR profiles generated from varying amounts of template DNA, from panels top to bottom: 0.005ng, 0.05ng, 0.125ng, 0.25ng, and 0.5ng

Table 3.2. Y-loci base pair sizes generated during amplification reactions using 0.5ng DNA template. The amplifications were performed in 3 occasions with a time interval of one month.

Y-STRs	Amplification One (27 May 2006) *Base pairs	Amplification Two (04 July 2006) *Base pairs	Amplification Three (07 August 2006) *Base pairs
DYS 19	253.67	253.55	254.11
DYS 385 a	270.25	270.83	270.46
DYS 385 b	273.89	274.51	274.06
DYS389 I	154.90	154.79	155.27
DYS389 II	272.24	271.82	272.65
DYS390	219.12	218.48	219.32
DYS391	103.10	102.76	103.27
DYS392	307.15	306.97	307.25
DYS393	123.75	123.65	123.50

*As determined by GeneScan® 3.7 software

3.3.2 Reproducibility studies

The reproducibility study determines the consistency of Y-STR data generated from multiple male DNA samples using the MinHt multiplex. Four DNA samples were successfully amplified and analyzed on three different occasions. Analysis showed consistent results generated for each DNA sample at three different times, demonstrating the multiplex's capability in generating consistent Y-loci results (table 3.3).

Table 3.3. Y-loci base pair sizes* generated during amplification reactions from four different DNA samples.

	DYS19	DYS385a	DYS385b	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393
DNA 01 (Asian Indian 49)									
1	253.86	252,38	263,64	162,61	287.51	222,93	103,14	304,21	123,47
2	253.53	253,58	264,70	162,58	287,32	222,40	102,98	304,00	123,47
3	253.67	252.91	264.11	162.58	287.34	222.7	103.25	304	123.42
DNA 02 (Asian Indian 51)									
1	253,84	270,88	274,68	154,94	272,30	218,78	103,14	306,91	123,47
2	253,67	272,18	275,95	154,80	271,58	218,50	103,03	306,95	123,20
3	253.52	271.57	275.31	154.98	271.81	218.41	103.01	307.12	123.82
DNA 03 (Asian Indian 54)									
1	257,71	263,48	267,17	155,11	276.01	214,70	103,07	304,21	123,38
2	257.08	264,83	268,61	154,79	275,54	214,02	102,98	304,00	123,35
3	257.54	264.28	267.83	154.92	275.75	214.33	103.16	304	123.51

Table 3.3 (*Continued*)

	DYS19	DYS385a	DYS385b	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393
					DNA 04 (Asian Indian 64)				
1	250.3	271,03	278,42	158,75	283,98	214,72	103,11	304.21	123,51
2	249.69	272,27	279,69	158,54	283,35	214,09	103,00	304	123,44
3	249.78	271.47	278.95	158.77	283.53	214.34	103.17	303.88	123.61

*As determined by GeneScan® 3.7 software

3.3.3. Species Specificity Studies

The human-specificity study examines the effect a range of non-human DNA has on a multiplex reaction. The multiplex reaction using 5ng DNA from canine (dog), feline (cat), bovine (cow), equine (horse) and ovis (sheep) together with 0.5ng of human DNA showed no amplification. This demonstrates the multiplex's non-reactivity on the range of animal DNA used during this study.

3.3.4. Environmental studies

Studies examined the effect of environmental conditions on DNA used for multiplex reactions. The DNA used during the amplification reaction was extracted from blood stains left on fabrics (cotton, denim and leather), which were exposed to high levels of heat, light and humidity. No amplification occurred from any of the extracted DNA samples. This is possible due to excessive degradation of DNA due to exposure to environmental conditions and/or the insufficient amount of DNA needed for a full interpretable Y-STR profile.

3.3.5. DNA mixtures studies

3.3.5.1. Male/female mixtures

This procedure examines the amplification and analytical effect that male-female DNA mixtures have on male-specific multiplex reactions. Results obtained showed that male DNA is detectable and interpretable in the presence of an equal amount of female DNA. Male/female ratio mixtures 1:2, 1:5 and 1:10 did not generate clear, interpretable Y-STR

profiles. Figure 3.2 shows that in the presence of abundant female DNA the MinHt multiplex display unspecific background and amplification artefacts from the 1:2 ratio and higher.

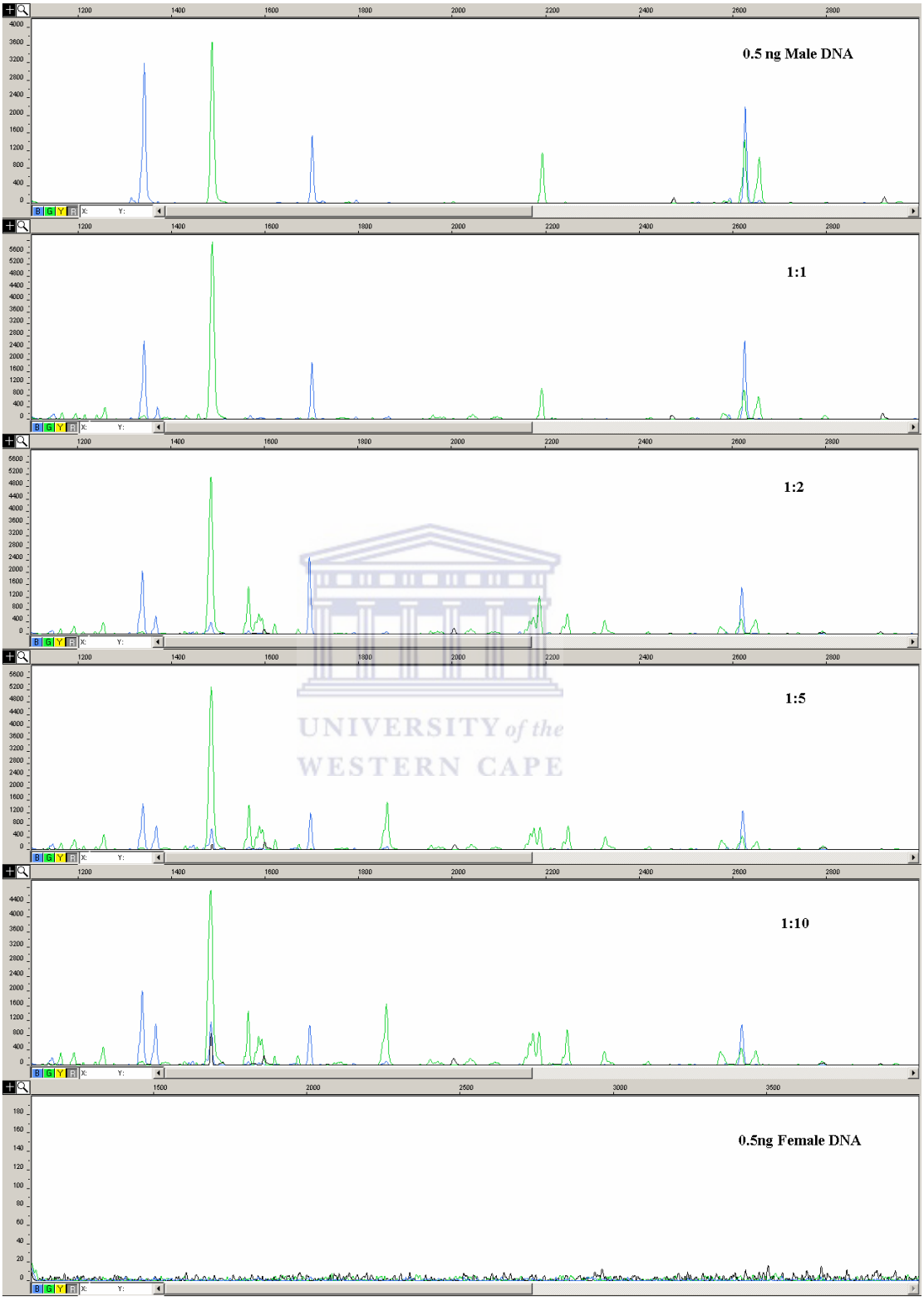

Figure 3.2. Amplification of male DNA in the presence of female DNA. Y-STR profiles shown in the panels from top to bottom: 0.5ng of male DNA, 0.5ng male DNA with 0.5ng female DNA (1:1), 0.5ng male DNA with 1ng female DNA (1:2), 0.5ng male DNA with 2.5ng female DNA (1:5) , 0.5ng male DNA with 10ng female DNA (1:10), and 0.5ng female DNA.

3.3.5.2. Male/male mixtures

This procedure examines the amplification and analytical effect that a mixed sample from two male donors have on male-specific multiplex reactions. The minor component genotype at non overlapping Y-STR loci from DNA mixture amplifications are shown in table 3.4. The male samples, Asian Indian 49 and Asian Indian 51, differ on base pair sizes of 6 of the 9 Y-loci generated (DYS385a/b. DYS389I/II, DYS390, and DYS392). The male/male mixtures were prepared, amplified and analyzed the results demonstrating that both male DNA profiles are discernible from one another up to ratio 1:2. The mixture ratio 1:5 profile shows allele dropout for DYS392 (figure 3.3). Coupled with the diminishing peak heights of the remaining Y-loci it would be reasonable to assume that at mixture ratios ranging from 1:10 to 1:1000 only one complete male profile would be generated by a multiplex reaction.

Table 3.4. Minor and major component genotypes at non-overlapping Y-STR loci from two male DNA mixture amplifications.

Mixture ratio	DYS385a	DYS385b	DYS389I	DYS389II	DYS390	DYS392
Male DNA 01 (Asian Indian 49) – Minor Component						
1:1	11	14	14	32	25	11
1:2	11	14	14	32	25	11
1:5	11	14	14	32	25	0[a]
Male DNA 02 (Asian Indian 51) – Major Component						
1:1	16	17	12	28	24	12
1:2	16	17	12	28	24	12
1:5	16	17	12	28	24	12

[a] No allele value could be detected.

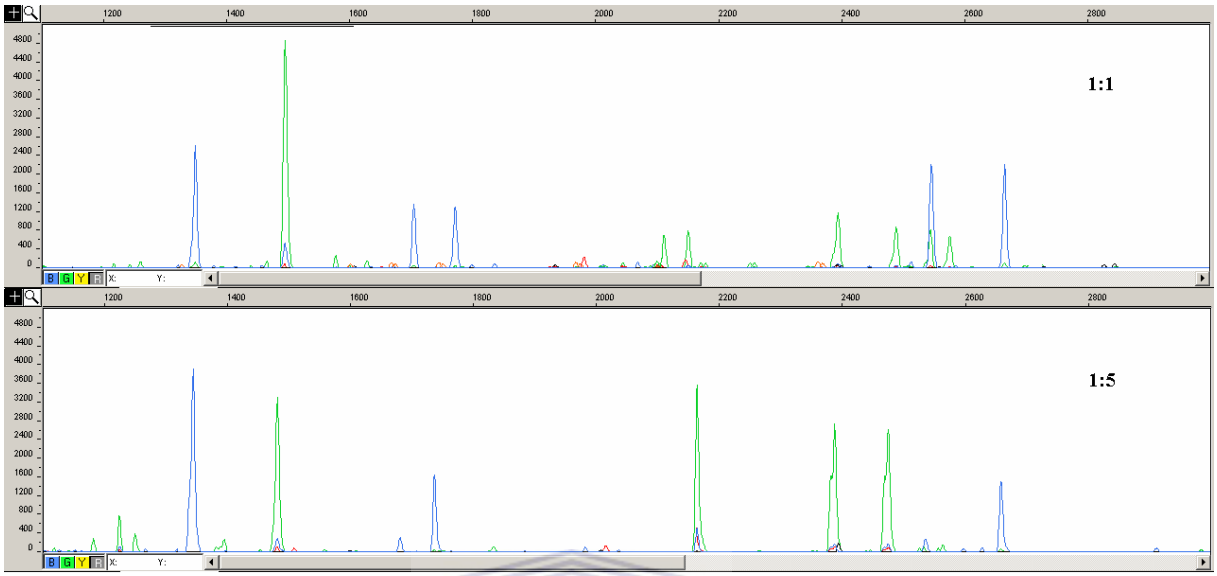

Figure 3.3. Amplification of mixed male DNA samples. Y-STR profiles shown in panels from top to bottom: 0.5ng male DNA 01 with 0.5ng male DNA 02 (1:1) and 0.5ng male DNA 01 with 2.5ng male DNA 02 (1:5).

3.4 Summary

The in-house validation of the minimal haplotype genotyping system was successfully carried out, following the ISFG and EDNAP guidelines. Although the MinHt has been shown to be a useful system in South Africa context, a comprehensive project aiming at the development of a more suitable genotyping system was initiated. As part of the project, an extensive study of the properties of a large number of Y-STR loci was carried out.

Chapter 4

The properties of twelve novel and widely studied Y-STR loci

4.1. Introduction

The overall objective of the study was to select markers which could augment those already in use and possibly replace those with limited polymorphism in South Africa populations. This was prompted by an initial study revealing extremely low level of polymorphism for two loci of the minimal haplotype (DYS391 and DYS392) in Xhosa population in Cape Town (Leat *et al*., 2004). Of the 99 individuals typed in the study only 47 unique haplotypes were observed and the most common haplotype was shared by 13 individuals. DYS 391, DYS392 and DYS437 were also found to be virtually monomorphic in a population surveyed in Maputo, Mozambique (Alves *et al*., 2003). As part of this study the properties of twelve novel and widely studied Y-STR loci were investigated. The result of this investigation is described in this chapter.

4.2. Materials and Methods

4.2.1. Selection of Y-STRs

Only twelve Y-STR loci were investigated in this part of the study and described in this chapter. Loci were fully sequenced and correct allele nomenclature was assigned to them. This included DYS449, DYS481, DYS518, DYS557, DYS570, DYS607, DYS612, DYS614, DYS626, DYS644, DYS712 and DYS714 loci.

4.2.2. DNA samples

Samples from diverse South African subpopulations were obtained as whole blood from blood transfusion services and DNA extracted by salting out proteins as described by Lahiri and Nurnberger, (1991) by Leat *et al*. (2004; 2007).

4.2.3. Primers

Unlabeled primers were designed utilizing Primer3 online-software and expected amplicon sizes as observed with contigs NT_011875.7, NT_011878.7, NT_011896.7, NT_011903.3. All the primers were submitted to the BLAST facility at NCBI to examine them for uniqueness and male-specificity. Primers were synthesized by *Inqaba Biotech*, *Applied Biosystems* and *Whitehead Scientific*. Primer sequences are presented in Table 4.1.

Table 4.1. Primers sequences of Y-STRs generated by *Primer3* software.

Y-STR	Primer Sequences (5' to 3')	Estimated Amplicon size*
DYS449 F	tgg agt ctc tca agc ctg ttc	
DYS449 R	g cct gga agt gga gtt tgc tgt	356 bp
DYS481 F	gtg aga gtg ttgc gag agt tag atg	
DYS481 R	ctc ttc ctt cta gag gag tgg ctt t	369 bp
DYS518 F	ttg aac ctg gga aac aga gg	
DYS518 R	gct ctt acc atg ggt gat ttc ttt c	335 bp
DYS557 F	gcc cag cat gtg ttt tga cta ttt	
DYS557 R	ggt cct gta ggc agg gtt aag	282 bp
DYS570 F	tgt gac atc aag gtt atg aaa cg	
DYS570 R	tca gca tag tca aga aac cag aca ac	214 bp
DYS607 F	agc ata cag cgt aat cac agc tca c	
DYS607 R	gct caa aac cca taa ttc agg tcc ag	234 bp
DYS612 F	ccc cca tgc cag taa gaa ta	
DYS612 R	gct cta gat acc cat ggc aag tgt c	241 bp

Table 4.1 (*Continued*)

Y-STR	Primer Sequences (5' to 3')	Estimated Amplicon size*
DYS614 F	gtg tgt ggc att gtg tgt ttg t	
DYS614 R	ctg gca ttc caa ttc act gt	500 bp
DYS626 F	ttg cag aga gct gag att gc	
DYS626 R	tcc tag agg ctt cta atg gag aga	386 bp
DYS644 F	gca act tgc acc ttc aaa gc	
DYS644 R	agg aaa tca agg ctg cag tg	246 bp
DYS712 F	taa cct agt ttg gaa ggt caa gga g	
DYS712 R	tta tat ggt aca gcc cat gaa cac tt	220 bp
DYS714 F	cac ctt ctg cat cga tct ttc	300 bp
DYS714 R	ctg act ttg ggg tat gga ttc tgt	

*From sequences on contigs NT_011875.7, NT_011878.7, NT_011896.7, and NT_011903.3

4.2.4. PCR single amplification

Single PCR amplification of the loci were preformed in a 25µl volume. The uniplex reaction included: 10ng of male DNA, 2.5µl of 10X PCR Buffer (10mM Tris-HCl [pH 8.5], 50 mM KCl), 0.2mM of each dNTP (*Roche*), 1.5mM $MgCl_2$ (*Applied Biosystems*), 0.5µM of each primer and 0.25µl Amplitaq Gold DNA polymerase (*Applied Biosystems*). Thermocycling parameters were as follow: an initial denaturation period of one cycle at 94˚C for 10 minutes, followed by 30 cycles of denaturation at 94˚C for 1 minute, annealing at 60˚C for 1 minute and an extension of 72˚C for 1 minute, the final cycle for 1 minute at 72˚C. The amplification process was done utilizing the GeneAmp 2400 thermocycler (*Applied Biosystems*). Single amplification for each Y-STR was done to allow that only one Y-STR was identified during each DNA sequence procedure.

4.2.5. DNA Sequencing

The PCR amplicons were purified using ExoSAP-IT (*USB Corp., Cleveland, Ohio*). Protocol for DNA purification indicated 2µl of Exo-SAP IT per every 5µl of PCR amplicon used. The samples were placed into a GeneAmp 2400 thermocycler (*Applied Biosystems*) and incubated for 30 minutes at two distinct temperatures (an initial 15 minutes at 37˚C and 15 minutes at 80˚C). Following Exo-SAP IT amplicon treatment (removing unincorporated primers and dNTPs), the sequencing reactions were prepared using the BigDye® Terminator version 3.1 sequencing kit (P/N 4336917) from *Applied Biosystems*.

DNA sequencing reactions were performed in reaction volumes of 10µl with 4µl of BigDye® Terminator mix, 1µl of either forward or reverse primer (0.5µM) of the selected Y-STR, 1.5µl of distilled water and 3.5µl of purified PCR amplicon (with a recommended 10–15ng of DNA). Cycle sequencing was performed as follows: 1 cycle at 96˚C for 1 minute, 25 cycles of 96˚C for 10 seconds, 50˚C for 5 seconds and 60˚C for 4 minute, the final cycle step hold at 4˚C.

Gel electrophoresis was performed following the ABI 377 manufacturer's instructions. Gel plates were assembled using 0.2mm spacers. A polyacrylamide gel solution was prepared containing 1 X TBE (pH 8.3), 36% w/v urea and a final concentration of 5% Long Ranger® polyacrylamide (*Applied Biosystems*). A gel solution of 25 ml was passed through a 0.22-micron filter and the gel solution de-gassed to prevent bubble formation within gel. Added to the gel solution mixture was 125 µl of 10% Ammonium Persulphate (APS) and 17.5 µl of N, N, N', N' – Tetramethyl – Ethylenediamine (TEMED). The gel was poured and allowed to polymerize for a minimum of two hours.

Table 4.2. DNA used during amplification and sequencing of selected Y-STRs

DNA utilized*	Sequenced Y-loci
AC 128	DYS449
MA 16	DYS449
EC 08	DYS557 & DYS570
EC 19	DYS612
EC 22	DYS607 & DYS626
EC 35	DYS607 & DYS626
EC 43	DYS557 & DYS570
EC 57	DYS612
AI 02	DYS518
AI 03	DYS712
AI 05	DYS570
AI 07	DYS644
AI 08	DYS481 & DYS644 & DYS714
AI 11	DYS714
AI 14	DYS626
AI 16	DYS481
AI 17	DYS607
AI 19	DYS570
AI 23	DYS614
AI 34	DYS712
AI 36	DYS614
AI 55	DYS518 & DYS626
AI 56	DYS612
AI 63	DYS557

*****As supplied by Leat *et al*. (2004, 2007)

AC= Afrikaner Caucasian
EC =English-speaking Caucasian
MA = Mixed Ancestry
AI = Asian Indian

Extension products resulting from the DNA sequencing reactions were prepared for electrophoresis. Preparation of samples included the Ethanol/EDTA/Sodium Acetate precipitation (see Appendix). After Ethanol/EDTA/Sodium Acetate precipitation the extension products were mixed with loading buffer (consisting of 5µl Hi-Di formamide [*Applied Biosystems*] and 1µl Dextran Blue loading dye). Loading mixture of 2µl each was loaded and electrophoresis on a polyacrylamide gel of the ABI 377 DNA sequencer started. Electrophoresis conditions consisting of 5% LongRanger gel, 1 M Urea, 1 X TBE (0.09 Tris, 0.09 M Boric acid,), 0.001 M EDTA, pH 8.0) as gel and electrophoresis buffer on a 48cm ABI 377 plate. The ABI 377 data collection samples were analyzed with Sequence Analysis

version 3.7 (*Applied Biosystems*) with the generated forward and reverse sequence of each Y-STR aligned and edited using BioEdit Sequence Alignment Editor Version 7 (Hall, 1999).

4.3. Results and Discussion

4.3.1 Allele nomenclature

Twelve Y- STRs selected from the 27 Y-loci previously investigated by Leat *et al.* (2007) were successfully sequenced. The sequencing included the determination of both forward and reverse DNA sequences for each Y-STR. The reverse and forward sequences were aligned and a consensus sequence was constructed using DNA sequencing software [BioEdit Ver.7 and/or Sequence Analysis 3.7 (*Applied Biosystems*)]. Sequencing data from two alleles per locus determined allele calling for each of the Y-STRs shown in Table 4.3.

Sequencing these Y-STR loci achieved two purposes. Firstly it enabled the assignment of correct allele nomenclature for each Y-STR and secondly to identify any possible polymorphism that may be found within flanking regions of the locus repeat motif and/or within the repeat motif itself (Gusmão *et al*., 2006). Repeat structures for each Y-locus, based on sequences as reported by Kayser *et al*. (2004), Leat *et al*. (2007), and Gusmão *et al*. (2006) are presented in table 4.3.

Sequencing confirmed the correct allele nomenclature for widely-studied Y-STRs DYS449, DYS481 DYS557, DYS570, DYS712 and DYS714 as shown by various publications (Redd *et al*., 2002; Leat *et al*., 2007; Butler *et al*., 2006; Gusmão *et al*., 2006). The sequenced novel Y-STRs DYS518, DYS607, DYS612, DYS614, DYS626 and DYS644 repeat structures matched those described by Kayser *et al*. (2004), Leat *et al*. (2007) and the NIST (SRM2395). DNA sequences of Y-STRs DYS518, DYS626 and DYS644 showed differences in their repeat structures from what was reported in some publications. DYS518 is a complex

tetranucleotide STR, with an additional repeat unit [AAAG] located downstream from the reported repeat motif, following a "GAAGAG" nucleotide sequence. The allele nomenclature consistent with ISFG guidelines would give DYS518's repeat structure as $[AAAG]_n[GGAG]_1AAAG]_4N_6[AAAG]_m$. Allele designation was determined as (n+1+4+m). For example the repeat structure $[AAAG]_{17}[GGAG]_1$ $[AAAG]_4$ GAAGAG$[AAAG]_{11}$, representing allele 33 and $[AAAG]_{19}$ $[GGAG]_1$ $[AAAG]_4$ GAAGAG $[AAAG]_{15}$, representing allele 39.

The sequenced Y-STRs DYS626 and DYS644 showed repeat motifs with the following allele nomenclature $[GAAA]_n[GAGA]_1[AAGA]_2[GGAA]_3$ and $[TTTTA]_n$ $\mathbf{N_7}$ $[TTTTA]_{1.}$ This does not match the reported repeat structures previously identified as DYS626 and DYS644 by Kayser *et al*. (2004). The journal publication reported DYS626 and DYS644 as complex STRs with repeat structures $[GAAA]_{16\text{-}27}[GGAA]_{5\text{-}6}$ and $[TTTTA]_{11}TTTA$ $[TTTTA]_{11}$. The sequence data shows a simple STR structure with diversity found only in the first repeat unit for each Y-STR. Allele designation was obtained by only counting the number of [GAAA] and [TTTTA] repeat units.

Table 4.3. Sequence Data for selected Y-STR loci

DYS449 - AC051663 (5’ to 3’)

Allele	Repeat structure variable of sequenced DNA
Repeat Structure nomenclature [b]	$[TTTC]_nN50[TTTC]_n$
26	$[TTTC]_{12}N_{50}\ [TTTC]_{14}$
29	$[TTTC]_{15}N_{50}\ [TTTC]_{14}$
34	$[TTTC]_{16}N_{50}\ [TTTC]_{18}$
DYS481 - AC016991 (5’ to 3’)	
Repeat Structure nomenclature [b]	$[CTT]_n$
21	$[CTT]_{21}$
22	$[CTT]_{22}$
27	$[CTT]_{27}$
DYS518 - AC010972 (5’ to 3’)	
Repeat Structure nomenclature [b]	$[AAAG]_n\ [GGAG]_1[AAAG]_4\ GAAGAG\ [AAAG]_m$
33	$[AAAG]_{17}[GGAG]_1[AAAG]_4\ GAAGAG\ [AAAG]_{11}$
34	$[AAAG]_{16}[GGAG]_1[AAAG]_4\ GAAGAG\ [AAAG]_{13}$
39	$[AAAG]_{19}[GGAG]_1[AAAG]_4\ GAAGAG\ [AAAG]_{15}$
DYS557 - AC007876 (5’ to 3’)	
Repeat Structure nomenclature [c]	$[TTTC]_4[TTC]_1[TTTC]_n$
15	$[TTTC]_4[TTC]_1[TTTC]_{15}$
16	$[TTTC]_4[TTC]_1[TTTC]_{16}$
21	$[TTTC]_4[TTC]_1[TTTC]_{21}$

Table 4.4 (*Continued*)

Allele	Repeat structure variable of sequenced DNA
DYS570 - AC012068 (5’ to 3’)	
Repeat Structure nomenclature [c]	$[TTTC]_n$
15	$[TTTC]_{15}$
17	$[TTTC]_{17}$
21	$[TTTC]_{21}$
DYS607 - AC053516 (5’ to 3’)	
Repeat Structure nomenclature [d]	$[GAAG]_n[GAAA]_1[GAAG]_1\,[GAAA]_1[GAAG]_1$
16	$[GAAG]_{12}\,[GAAA]_1[GAAG]_1\,[GAAA]_1[GAAG]_1$
19	$[GAAG]_{15}\,[GAAA]_1[GAAG]_1\,[GAAA]_1[GAAG]_1$
DYS612 - AC006383 (5’ to 3’)	
Repeat Structure nomenclature [b]	$[CCT]_5[CTT]_1[TCT]_4[CCT]_1[TCT]_n$
27	$[CCT]_5[CTT]_1[TCT]_4[CCT]_1[TCT]_{22}$
30	$[CCT]_5[CTT]_1[TCT]_4[CCT]_1[TCT]_{25}$
33	$[CCT]_5[CTT]_1[TCT]_4[CCT]_1[TCT]_{28}$
DYS614 - AC064829 (5’ to 3’)	
Repeat Structure nomenclature [b]	$[CTT]_2[CCT]_3[CTT]_3[CTTCTTT]_2[CTT]_2[CTG]_1[CTT]_2[CTGCTT]_2[CTGCT]_1[CTT]_n$
14	$[CTT]_2[CCT]_3[CTT]_3[CTTCTTT]_2[CTT]_2[CTG]_1[CTT]_2[CTGCTT]_2[CTGCT]_1[CTT]_{14}$
18	$[CTT]_2[CCT]_3[CTT]_3[CTTCTTT]_2[CTT]_2[CTG]_1[CTT]_2[CTGCTT]_2[CTGCT]_1[CTT]_{18}$
21	$[CTT]_2[CCT]_3[CTT]_3[CTTCTTT]_2[CTT]_2[CTG]_1[CTT]_2[CTGCTT]_2[CTGCT]_1[CTT]_{21}$

Table 4.4 (*Continued*)

DYS626 - NT011903 (5’ to 3’)

Allele	Repeat structure variable of sequenced DNA
Repeat Structure nomenclature [b]	$[GAAA]_n\ [GGAA]_n$
14	$[GAAA]_{14}[GAGA]_1[AAGA]_2[GGAA]_3$
17	$[GAAA]_{17}[GAGA]_1[AAGA]_2[GGAA]_3$
21	$[GAAA]_{21}[GAGA]_1[AAGA]_2[GGAA]_3$
DYS644 - AC006462 (5’ to 3’)	
Repeat Structure nomenclature [b]	$[TTTTA]_n$…TTTA $[TTTTA]_n$
13	$[TTTTA]_{13}$TTTTTA $[TTTTA]_1$
16	$[TTTTA]_{16}$TTTTTA $[TTTTA]_1$
19	$[TTTTA]_{19}$TTTTTA $[TTTTA]_1$
DYS712 - AC006376 (5’ to 3’)	
Repeat Structure nomenclature [a]	$[AGAT]_n\ [AGAC]_n$
21	$[AGAT]_{16}\ [AGAC]_5$
22	$[AGAT]_{17}\ [AGAC]_5$
27	$[AGAT]_{18}\ [AGAC]_9$
DYS714 - AC009233 (5’ to 3’)	
Repeat Structure nomenclature [a]	$[TTTCT]_n[CTTCT]_n[TTTCT]_2[CTTCT]_2[TTTCT]_2[T]_{11}$
21	$[TTTCT]_{13}[CTTCT]_2[TTTCT]_2[CTTCT]_2[TTTCT]_2[T]_{11}$
27	$[TTTCT]_{19}[CTTCT]_2[TTTCT]_2[CTTCT]_2[TTTCT]_2[T]_{11}$
30	$[TTTCT]_{19}[CTTCT]_5[TTTCT]_2[CTTCT]_2[TTTCT]_2[T]_{11}$

Allele nomenclatures for sequenced Y-STRs followed International Society of Forensic Genetics (ISFG) guidelines.

Reference allele nomenclature sequences in bold

[a] Repeat structure as reported by Leat *et al*. (2007)

[b] Repeat structure as reported in the online-only supplementary material of Kayser *et al*. (2004)

[c] Repeat structure as reported by Gusmao *et al*. (2006).

[d] Repeat structure as reported by NIST (SRM 2395)(http://www.cstl.nist.gov/biotech/strbase/srm2395.htm)

4.4 Summary

Investigation into the widely-studied and novel Y-loci sequences enabled allele nomenclature to be assigned to each Y-locus, following recommended ISFG guidelines. Eight of the STRs included in the present study are routinely amplified in a single multiplex originally named UWC Y-plex 2 and included a total of 11 loci (Leat *et al*., 2007). Despite the repeated attempts, to sequence the remaining 3 loci of the 11-plex, a reliable sequence could not be obtained. Loci DYS449, DYS481, DYS518, DYS557, DYS570, DYS607, DYS612 and DYS614 were then used in a single multiplex reaction and named the eight Y-STR's system. Initial population data in South Africa subpopulations for these eight Y-STRs showed gene diversity values above 0.7 (Leat *et al*., 2007). A further study was undertaken to obtain additional population data for the loci included the eight Y-STR's system. The following chapter reports the results of the investigation if these genotyping systems in the Cape Muslim population.

Chapter 5

Analysis of the loci of the eight Y-STR system in the Cape Muslim population of South Africa

5.1. Introduction

In this chapter, the findings of the investigation into the eight Y-STR genotyping system in the Cape Muslim population are reported in detail. The eight Y-STR loci included DYS449, DYS481, DYS518, DYS557, DYS570, DYS607, DYS612 and DYS614.

5.2. Materials and Methods

Materials and methods were essentially as described in chapter 2, and exceptions as described in the following sections.

5.2.1. PCR multiplex amplification

Eight Y-STRs were amplified in a single multiplex reaction. Amplifications performed in a final volume of 10µl containing 5-10 ng male DNA, 10X SuperTherm PCR buffer (with 15mM $MgCl_2$) (*Southern Cross Biotechnology*), 200µM dNTPs (Roche), and 5U/ µl of Super-Therm (*Southern Cross Biotechnology*). Primers were synthesized by *Applied Biosystems* using Leat *et al*. (2007) sequences. The primer sequences, corresponding fluorescent dye labels and the final concentration, at which each primer was used is shown in table 5.1.

Table 5.1. Primer sequences and final concentration for the eight Y-STRs multiplex.

Y-STR	Primer Sequences (5' to 3')	[Primer]
DYS570 F	g tga cta ggt aga aat cct ggc tgt g	0.07μM
DYS570 R	**FAM** tca gca tag tca aga aac cag aca ac	
DYS518 F	ggc aac aca agt gaa act gct tct	0.5 μM
DYS518 R	**FAM** gct ctt acc atg ggt gat ttc ttt c	
DYS481 F	**VIC** aaa agg aat gtg gct aac gct gtt c	0.2 μM
DYS481 R	gct cac cag aag gtt gca aga ctc a	
DYS612 F	**VIC** aag ttt cac aca ggt tca gag gtt g	0.3 μM
DYS612 R	gac act tgc cat ggg tat cta gag c	
DYS557 F	gcc cag cat gtg ttt tga cta ttt	0.15μM
DYS557 R	**VIC** ggt gtt tga ccc agt gat atg ttc t	
DYS614 F	gaa tat ggc taa ctc gat tca gag	0.5 μM
DYS614 R	**VIC** cca cca aaa ggt ttt cag act ca	
DYS607 F	**NED** agc ata cag cgt aat cac agc tca c	0.1 μM
DYS607 R	gct caa aac cca taa ttc agg tcc ag	
DYS449 F	**NED** tgg agt ctc tca agc ctg ttc ta	0.5 μM
DYS449 R	g cct gga agt gga gtt tgc tgt	

PCR amplification was performed in a GeneAmp 2400 thermocycler (*Applied Biosystems*) as follows: 1 cycle at 95°C for 10 minutes, 30 cycles of 94°C for 1 minute, 60°C for 1.5 minutes, 72°C for 1.5 minutes, followed by 1 final cycle at 72°C for 60 minutes. PCR products were stored in the dark at 4°C.

5.2.2. Allelic ladder construction

DNA samples of Leat *et al*. (2004) representing the most common alleles were mixed and amplified for each successfully sequenced Y-STR. This resulted in a single completed PCR reaction containing a range of reference amplicons for each Y-locus. The allele designations for each sequenced Y-STR were matched to amplicon base pair sizes for each locus based upon amplification of reference alleles. The allelic ladder was constructed using these combinations of alleles at each locus to facilitate consistent allele typing using Genotyper® 3.7 (*Applied Biosystems*).

5.3. Results and Discussion

5.3.1. Performance of the Eight Y-STR's system

DNA was extracted from 100 samples, genotyped and analyzed. The DNA extraction kit used on buccal swabs proved to be efficient yielding stock DNA concentrations ranging from 20ng/µl to 150ng/µl. Electrophoresis on multiplex amplicons, using the ABI 377 DNA sequencer, generated eletropherograms that can be easily interpreted. An example of such an electropherogram for the eight Y-STRs multiplex reaction is presented in figure 5.1.

Figure 5.1. An example of an eight Y-STRs multiplex reaction electropherogram

5.3.2 Analysis of allele and haplotype frequencies for Cape Muslim males

Gene diversity (GD) values were calculated for eight Y-STR loci in the Cape Muslim population. Table 5.2 shows the allele and haplotype frequencies for these eight Y-loci among the Cape Muslim males. The GD value for DYS449 among the Cape Muslim males was 0.884, representing the highest gene diversity observed for a single-copy Y-locus. The lowest GD value from the eight Y-STRs multiplex was 0.789 for DYS607. Gene diversity values for all eight Y-loci investigated among the Cape Muslim population were greater than 0.75.

Table 5.2. Allele frequencies and gene diversity values of eight Y-STR loci from 100 Cape Muslim males.

Allele	DYS449	DYS481	DYS518	DYS557	DYS570	DYS607	DYS612	DYS614
10								0.010
11						0.040		
12						0.030		
13				0.080	0.010	0.200		0.040
14				**0.220**	0.030	0.280		0.020
15				0.210	0.080	**0.290**		**0.260**
16				0.120	0.160	0.110		0.190
17				0.160	**0.280**	0.040		0.160
18				0.120	0.130	0.010		0.110
19		0.030		0.060	0.200			0.200
20				0.020	0.060			0.010
21		0.030		0.010	0.040			
22		**0.280**			0.010			
23		0.230						
24		0.180						
25	0.030	0.110						
26	0.050	0.090					0.040	
27	0.040	0.030					0.050	
28	0.090	0.020					0.130	
29	0.140						0.260	
30	0.160						**0.310**	
31	0.160		0.040				0.110	
32	**0.180**		0.060				0.060	
33	0.060		0.140				0.040	
34	0.070		0.130					
35	0.020		**0.220**					
36			**0.220**					
37			0.090					
38			0.040					
39			0.040					
40			0.010					
41			0.010					
GD	0.884	0.821	0.859	0.851	0.835	0.789	0.806	0.825

In bold are highest allele frequencies

GD = Gene Diversity Index

HD = Haplotype Diversity Index

5.3.3. Comparison of allele frequencies among six South African sub-populations

Table 5.3 compares the observed gene diversity (GD) values of the eight Y-loci among six South African sub-populations including data from previous studies by Leat and colleagues (2007; unpublished). Among the six South African sub-populations DYS449 exhibited the highest GD value of 0.884, with DYS607 showing the lowest of 0.570. The highest GD value observed was from the Cape Muslim population and the lowest from the Xhosa population. Cape Muslim males exhibit gene diversity values above 0.8 for seven loci, the same number of loci observed for the Asian Indian population with GD values above 0.8. The highest variability for the eight loci is therefore found in the Cape Muslim and Asian Indian populations. The Xhosa population, with GD values equal or below 0.7 for four loci (DYS557, DYS570, DYS607 and DYS614) exhibits the lowest variability.

Table 5.3. Comparison of GD values from eight Y-STR loci among five South African sub-populations

Population Group	*n*	DYS449	DYS481	DYS518	DYS557	DYS570	DYS607	DYS612	DYS614
Asian Indian [a]	77	0.860	0.670	0.830	0.850	0.830	0.810	0.800	0.850
Afrikaner Caucasian [b]	109	0.836	0.747	0.827	0.707	0.776	0.698	0.763	0.787
English Caucasian [a]	101	0.770	0.700	0.800	0.710	0.740	0.710	0.760	0.690
Xhosa [a]	88	0.770	0.800	0.860	0.670	0.700	0.570	0.810	0.650
Mixed Ancestry [b]	113	0.855	0.832	0.860	0.803	0.781	0.758	0.861	0.804
Cape Muslim	100	0.884	0.821	0.859	0.851	0.835	0.789	0.806	0.825

[a] Populations investigated by Leat *et al*. (2007)

[b] Populations investigated by Leat *et al*. (unpublished)

5.3.4. Haplotype analysis of the Cape Muslim population

The Y-STR haplotype data shown in table 5.4 was obtained during the genotyping of one hundred Cape Muslim males. Analysis identified 92 haplotypes of which eighty-four are unique and the remaining eight haplotypes are shared (shown in table 5.5). Statistical analysis of haplotype data shows a discrimination capacity of 0.920 from the eight Y-STR multiplex.

The eight Y-STRs multiplex exhibited a higher discrimination capacity than the MinHt system (0.847).

Table 5.4. Haplotypes for eight Y-STR loci observed from 100 Cape Muslim males.

H	DYS449	DYS481	DYS518	DYS557	DYS570	DYS607	DYS612	DYS614	*n*	*f*
H01	29	22	35	18	17	15	30	19	1	0.01
H02	**32**	**23**	**38**	**16**	**18**	**16**	**29**	**19**	**2**	**0.02**
H03	33	22	39	15	18	15	29	18	1	0.01
H04	28	22	33	16	18	15	30	17	1	0.01
H05	35	25	36	17	18	13	33	16	1	0.01
H06	29	27	35	15	19	15	29	14	1	0.01
H07	31	22	35	16	16	15	29	18	1	0.01
H08	31	26	34	17	15	15	28	18	1	0.01
H09	30	22	35	17	16	15	30	17	1	0.01
H10	**31**	**22**	**35**	**17**	**19**	**16**	**30**	**19**	**2**	**0.02**
H11	34	26	34	20	17	12	29	13	1	0.01
H12	31	25	34	15	21	14	29	15	1	0.01
H13	28	23	36	15	22	16	30	19	1	0.01
H14	32	23	36	14	17	16	31	19	1	0.01
H15	31	23	36	19	17	13	28	15	1	0.01
H16	30	27	33	21	17	13	32	16	1	0.01
H17	32	23	40	16	17	16	32	18	1	0.01
H18	**26**	**28**	**35**	**15**	**19**	**14**	**30**	**15**	**2**	**0.02**
H19	29	22	36	13	16	15	27	15	1	0.01
H20	25	25	34	18	18	14	29	13	1	0.01
H21	**32**	**19**	39	**15**	**19**	**14**	**30**	**19**	**2**	**0.02**
H22	34	23	37	15	21	17	31	19	1	0.01
H23	**29**	**22**	**36**	**13**	**16**	**15**	**26**	**15**	**2**	**0.02**
H24	33	23	37	14	19	15	28	19	1	0.01
H25	28	22	33	17	17	15	32	17	1	0.01
H26	30	22	39	17	19	13	31	16	1	0.01
H27	32	24	36	18	17	13	27	16	1	0.01
H28	27	25	36	14	14	13	30	19	1	0.01
H29	**32**	**21**	**31**	**14**	**17**	**13**	**26**	**15**	**2**	**0.02**
H30	33	23	34	18	21	11	30	17	1	0.01
H31	34	23	36	18	17	11	30	15	1	0.01

Table 5.4 (*Continued*)

H	DYS449	DYS481	DYS518	DYS557	DYS570	DYS607	DYS612	DYS614	*n*	*f*
H32	29	25	35	18	18	14	30	10	1	0.01
H33	30	26	35	13	16	14	29	17	1	0.01
H34	25	25	37	19	19	14	28	13	1	0.01
H35	33	23	37	15	19	16	30	19	1	0.01
H36	29	23	34	16	16	14	30	16	1	0.01
H37	27	23	33	13	13	14	29	18	1	0.01
H38	29	25	31	17	14	14	31	19	1	0.01
H39	30	22	33	16	19	15	32	18	1	0.01
H40	33	24	35	16	17	15	30	19	1	0.01
H41	30	24	34	14	17	13	28	15	1	0.01
H42	32	26	36	14	17	15	28	18	1	0.01
H43	26	27	35	15	18	14	29	15	1	0.01
H44	29	22	36	16	17	15	29	17	1	0.01
H45	31	24	33	15	16	14	29	15	1	0.01
H46	28	24	32	15	19	14	32	16	1	0.01
H47	26	26	35	16	19	15	30	15	1	0.01
H48	30	22	36	17	17	14	31	17	1	0.01
H49	28	23	35	14	18	14	28	15	1	0.01
H50	31	23	35	20	17	13	29	15	1	0.01
H51	30	22	34	14	16	14	29	16	1	0.01
H52	30	22	36	17	18	11	29	16	1	0.01
H53	28	25	32	14	15	13	30	20	1	0.01
H54	32	19	41	15	19	14	30	19	1	0.01
H55	30	24	36	14	16	14	29	15	1	0.01
H56	26	22	35	14	15	14	30	17	1	0.01
H57	27	26	36	15	19	16	29	15	1	0.01
H58	29	24	34	14	16	15	28	16	1	0.01
H59	**31**	**22**	**33**	**14**	**17**	**15**	**30**	**17**	**2**	**0.02**
H60	31	26	35	18	17	13	27	16	1	0.01
H61	32	23	34	16	15	18	30	19	1	0.01
H62	34	23	37	15	19	16	29	19	1	0.01
H63	30	24	33	15	18	13	29	15	1	0.01
H64	34	25	37	14	15	15	29	18	1	0.01
H65	29	23	31	16	17	14	30	17	1	0.01
H66	28	22	33	17	17	15	28	18	1	0.01
H67	28	23	35	15	16	14	31	16	1	0.01

Table 5.4 (*Continued*)

H	DYS449	DYS481	DYS518	DYS557	DYS570	DYS607	DYS612	DYS614	*n*	*f*
H68	30	24	32	14	15	17	30	15	1	0.01
H69	28	25	33	17	17	15	30	14	1	0.01
H70	31	24	34	18	16	13	31	15	1	0.01
H71	29	22	34	15	17	15	31	17	1	0.01
H72	**31**	**22**	**36**	**17**	**17**	**15**	**33**	**17**	**2**	**0.02**
H73	32	23	32	19	14	13	29	13	1	0.01
H74	33	24	33	14	19	13	31	15	1	0.01
H75	30	22	36	17	17	14	31	18	1	0.01
H76	31	23	37	15	20	17	30	19	1	0.01
H77	35	25	37	13	15	15	28	16	1	0.01
H78	30	22	32	14	21	14	30	16	1	0.01
H79	30	24	37	14	16	15	30	15	1	0.01
H80	32	24	33	13	20	14	32	17	1	0.01
H81	32	24	35	19	17	13	30	16	1	0.01
H82	27	26	35	13	20	14	28	15	1	0.01
H83	29	23	32	18	15	14	29	17	1	0.01
H84	34	22	36	17	19	12	33	16	1	0.01
H85	32	24	33	18	18	11	29	15	1	0.01
H86	25	24	38	18	16	13	27	16	1	0.01
H87	29	24	38	14	16	15	28	16	1	0.01
H88	32	24	35	19	18	13	29	16	1	0.01
H89	34	26	36	18	20	13	28	15	1	0.01
H90	30	23	36	15	20	17	31	19	1	0.01
H91	31	21	35	19	20	12	29	16	1	0.01
H92	32	22	34	14	19	16	27	18	1	0.01

Discrimination capacity is 0.920

HD is 0.9983838383838380; Standard Error for HD is 0.0005184592558726

Table 5.5. Haplotypes shared by more than one Cape Muslim male (n=100)

Haplotypes DYS449,DYS481,DYS518,DYS557,DYS570,DYS607,DYS612,DYS614	**H**	***n***	***f***
32-23-38-16-18-16-29-19	H02	2	0.020
31-22-35-17-19-16-30-19	H10	2	0.020
26-28-35-15-19-14-30-15	H18	2	0.020
32-19-39-15-19-14-30-19	H21	2	0.020
29-22-36-13-16-15-26-15	H23	2	0.020
32-21-31-14-17-13-26-15	H29	2	0.020
31-22-33-14-17-15-30-17	H59	2	0.020
31-22-36-17-17-15-33-17	H72	2	0.020

H = Haplotypes

n = number of individuals observed for each haplotype

f = frequency of haplotype in population sample

5.3.5. Haplotype comparison of Cape Muslim males with South African sub-populations.

From the ninety-two haplotypes identified, no shared haplotypes were detected among the Asian Indian, Xhosa, and English-speaking Caucasian sub-populations. Haplotype data from Leat and co-workers (unpublished) during comparison analysis of the Mixed Ancestry and Afrikaner-speaking Caucasian sub-populations however revealed four shared haplotypes in total. Cape Muslim haplotype number 50 (H50) was shared with one Afrikaner-speaking Caucasian sample, and the remaining three Cape Muslim haplotypes (H28, H37, H71) were found shared in the Mixed Ancestry sub-population (Leat *et al*. unpublished).

5.4. Summary

The population data generated from this study, as well as previous studies (Leat *et al*., 2004; 2007), clearly indicates that the eight Y-STR genotyping system could be useful in forensic casework in South Africa. For this reason, the internal validation exercise was carried out on this genotyping system and described in the next chapter.

Chapter 6

Internal validation of the eight Y-plex genotyping system

6.1. Introduction

This chapter reports the results of the in-house validation procedure that was carried out to test the eight Y-plex genotyping system.

6.2. Materials and Methods

6.2.1. DNA samples

Male DNA samples used were of individuals from diverse South African populations (Asian Indian, Mixed Ancestry, "Afrikaner"-and "English-speaking" Caucasian males and Xhosa). The male samples obtained by Leat *et al*. (2004; 2007) as whole blood from blood transfusion services. The single female DNA used was acquired as whole blood from a voluntary donor. DNA extracted from blood by salting out proteins as described by Lahiri and Nurnberger, (1991) (see Appendix).

6.2.2 PCR amplification conditions

Eight Y-STRs were amplified in a single multiplex reaction. Amplifications performed in a final volume of 10µl containing an initial 0.005 ng male DNA, 10X PCR buffer (with 15mM $MgCl_2$) (*Southern Cross Biotechnology*), 200µM dNTPs (*Roche*), and 5U/ µl of Super-Therm (*Southern Cross Biotechnology*). Primers were synthesized by *Applied Biosystems* using Leat *et al*. (2007) sequences. The primer sequences, corresponding fluorescent dye labels and the final concentration, at which each primer was used is shown in table 5.1. More information on selected Y-STRs is shown in table 6.1.

Table 6.1. General information of the Y-STRs selected from Leat *et al*. (2007).

Y-loci	Dye	PCR Product size (bp)	Repeat motif	GenBank® accession
DYS449	NED	340 - 408	[TTTC]	AC051663
DYS481	VIC	111 - 158	[CTT]	AC016991
DYS518	5FAM	251 - 311	[AAAG]	AC010972
DYS557	VIC	227 - 259	[TTTC]	AC007876
DYS570	5FAM	154 - 198	[TTTC]	AC012068
DYS607	NED	212 - 248	[GAAG]	AC053516
DYS612	VIC	192 - 230	[TCT]	AC006383
DYS614	VIC	312 - 348	[CTT]	AC064829

PCR amplification was performed in the GeneAmp 2400 thermocycler (*Applied Biosystems*) as follows: 1 cycle at 95°C for 10 minutes, 30 cycles of 94°C for 1 minute, 60°C for 1.5 minutes, 72°C for 1.5 minutes, followed by 1 final cycle at 72°C for 60 minutes. PCR products were stored in the dark at 4°C.

6.2.3. PCR amplicon analysis

Amplified DNA fragments were analyzed as previously described (section 2.2.3).

6.2.4. Validation procedures

Validation procedures followed as previously described (section 3.2.4)

6.3. Results and Discussion

6.3.1. Minimal Sensitivity

The minimal sensitivity study was designed to observe the Y-STR profiles generated during amplification of varying DNA amounts (0.005ng, 0.05, 0.125ng, 0.25ng, 0.5ng and 2.5ng) from one male genomic source GeneScan® software analysis on amplification results observed the generation of partial Y-STR profiles from DNA amounts between 0.005ng to 0.25ng. Full, interpretable Y-STR profiles were observed only during amplification of 0.5ng and 2.5ng DNA templates. GeneScan® software analysis on sensitivity multiplex reactions

with full Y-STR profiles showed consistent base pair sizes for eight loci (as shown in table 6.2).The sensitivity study demonstrating that a minimal amount of 0.5ng DNA template will generate a full, accurate and consistent Y-STR profile.

Figure 6.1. Y-STR profiles generated from varying amounts of template DNA, from panels top to bottom: 0.005ng, 0.05ng, 0.125ng, 0.25ng, and 0.5ng

Table 6.2. Y-loci base pair sizes generated during amplification reactions using 0.5ng DNA template.

Y-STRs	Sensitivity One *Base pairs	Sensitivity Two *Base pairs	Sensitivity Three *Base pairs
DYS449	357.12	356.98	356.75
DYS481	135.62	135.63	135.25
DYS518	272.16	271.38	271.40
DYS557	231.07	230.43	231.14
DYS570	192.92	192.36	192.81
DYS607	227.92	227.50	227.82
DYS612	214.50	213.99	213.87
DYS614	335.25	334.70	335.49

*As determined by GeneScan® 3.7 software

6.3.2 Reproducibility studies

The reproducibility study determines the consistency of Y-STR data generated from multiple male DNA samples using the eight Y-STRs multiplex. Four DNA samples were successfully amplified and analyzed on three different occasions. Analysis showed the consistent results generated for each DNA sample at three different times, demonstrating the multiplex's capability in generating consistent Y-loci results (presented in table 6.3)

Table 6.3. Y-loci base pair sizes* generated during amplification reactions from four different DNA samples.

	DYS449	**DYS481**	**DYS518**	**DYS557**	**DYS570**	**DYS607**	**DYS612**	**DYS614**
				DNA 01 (Asian Indian 49)				
1	372.43	132.05	283.35	235.43	189.58	239.87	210.34	338.53
2	373.47	132.3	283.03	234.23	188.3	239.63	210.36	338.82
3	373.07	132.37	283.22	234.75	189.01	239.63	210.79	338.29
				DNA 02 (Asian Indian 51)				
1	356.05	135.44	271.48	231.65	192.49	227.1	213.42	336.64
2	356.47	135.63	271.3	231.35	192.28	227.61	214.55	335.61
3	356.88	135.3	271.5	231.03	192.87	227.74	214.04	335.4

Table 6.3 (*Continued*)

	DYS449	DYS481	DYS518	DYS557	DYS570	DYS607	DYS612	DYS614
DNA 03 (Asian Indian 54)								
1	364.17	145.38	287.35	235.54	181.71	232	204.21	317.88
2	365.39	145.76	287.17	234.29	181.68	231.58	204.21	317.93
3	364.98	145.49	287.14	234.89	181.34	231.78	204.75	317.54
DNA 04 (Asian Indian 64)								
1	368.33	128.95	279.09	231.59	170.29	236.05	207.33	327.79
2	368.48	129.19	279.24	231.22	170.33	235.41	208.22	326.94
3	368.98	129.22	279.3	231.86	170.68	235.61	207.7	326.51

*As determined by GeneScan® 3.7 software

6.3.3. Species Specificity Studies

The species specificity study examines the effect a range of non-human DNA has on a multiplex reaction. The multiplex reaction using 5ng of DNA from canine (dog), feline (cat), bovine (cow), equine (horse) and ovis (sheep) showed no amplification. Demonstrating the multiplex's non-reactivity on the range of animal DNA used during this study.

6.3.4. Environmental studies

This study examines the effect environmental conditions have on DNA used for multiplex reactions. The DNA used during the amplification reaction was extracted from blood left on stained fabrics (cotton, denim and leather), which were exposed to high levels of heat, light and humidity. No amplification occurred from any of the extracted DNA. Possible factors that contributed towards none amplification was the degradation of DNA from environmental exposure and/or the insufficient amount of DNA needed for a full interpretable Y-STR profile.

6.3.5. DNA mixtures studies

6.3.5.1. Male/female mixtures

This procedure examines the amplification and analytical effect that male-female DNA mixtures have on male-specific multiplex reactions. Amplification reactions were prepared with male/female mixtures using 0.5ng of male DNA (Asian Indian 51) and corresponding amounts of female DNA to generate ratios of 1:0 to 1:10. Results obtained shows that male DNA is detectable and interpretable in a 1:10 male/female DNA mixture ratio (shown in figure 6.2). Female DNA was shown not to interfere with the amplification of male specific markers, nor generate amplification products of its own.

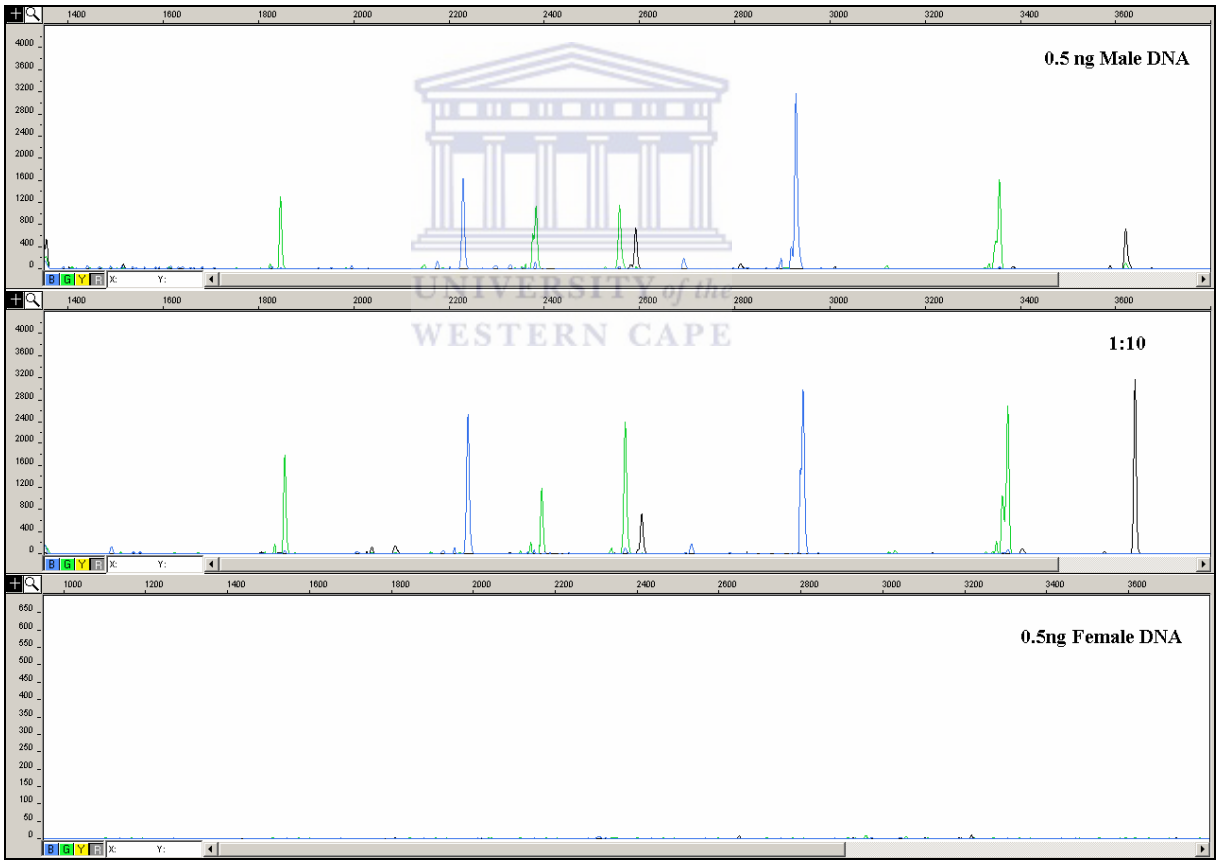

Figure 6.2. Amplification of male DNA in the presence of female DNA. Y-STR profiles shown in the panels from top to bottom: 0.5ng of male DNA, 0.5ng male DNA with 5ng female DNA (1:10) and 0.5ng female DNA.

6.3.5.2. Male/male mixtures

This procedure examines the amplification and analytical effect that a mixed sample from two male donors have on male-specific multiplex reactions. The selection criteria of male samples, Asian Indian 49 and Asian Indian 51, were based on the fact that each sample during amplification generates a different base pair size for each locus. The male/male mixtures were prepared, amplified and analyzed the results demonstrating that both male DNA profiles are discernible from each another up to ratio 1:5. The mixture ratio 1:5 profile shows diminishing peak heights (RFU) from each Y-locus on one male donor profile making it reasonable to assume that at a mixture ratios ranging from 1:10 to 1:1000 only one complete male profile would be generated by a multiplex reaction.

Table 6.4 Minor and major component genotypes at non-overlapping Y-STR loci from two male DNA mixture amplifications.

	DYS449	DYS481	DYS518	DYS557	DYS570	DYS607	DYS612	DYS614
	Male DNA 1 (Asian Indian 49) – Minor Component							
1:1	32	23	35	15	20	16	30	19
1:2	32	23	35	15	20	16	30	19
1:5	32	23	35	15	20	16	30	19
	Male DNA 2 (Asian Indian 51) – Major Component							
1:1	28	24	32	14	21	13	31	18
1:2	28	24	32	14	21	13	31	18
1:5	28	24	32	14	21	13	31	18

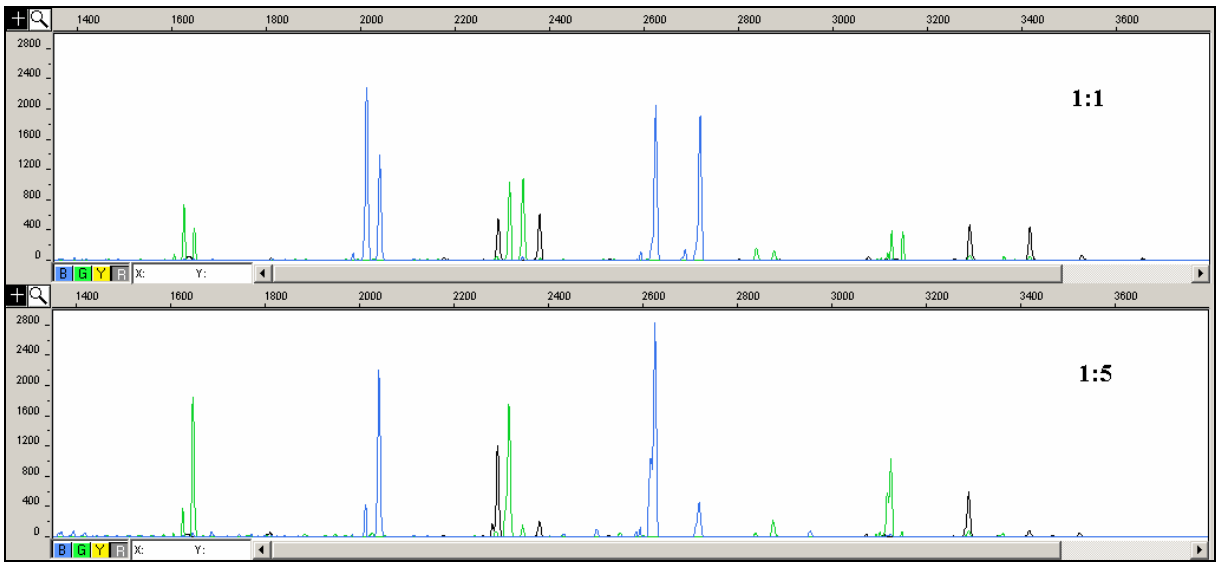

Figure 6.3. Amplification of mixed male DNA samples. Y-STR profiles shown in panels from top to bottom: 0.5ng male DNA 01 with 0.5ng male DNA 02 (1:1) and 0.5ng male DNA 01 with 2.5ng male DNA 02 (1:5).

6.4 Summary

The internal validation of the eight Y-plex genotyping system was carried out successfully. This system is therefore recommended for use in forensic casework in South Africa. Further improvement of the system can be achieved by carrying out additional optimization work as well as adding more loci.

Chapter 7

Conclusion

Sexual assault is a significant problem facing South African society. In this context, efficient but also affordable genotyping systems are needed for positive identification of criminals in incidences of sexual violence. In developing countries such as South Africa, the use and development of non-commercial genotyping systems is extremely important in lowering the financial cost when using genotyping systems.

The present study forms part of a comprehensive project aimed at the development of non-commercial genotyping systems suitable for sexual assault cases in South Africa. It has made a significant contribution in generating more population data for loci included and established male-specific genotyping systems as well as in systems still in the development process. The study has also contributed in investigating the properties of candidate loci for inclusion in male genotyping systems suitable for South Africa. Finally, the internal validation of two Y-STR genotyping systems was successfully carried out.

The overall assessment of the minimal haplotype system has shown that it is still a useful genotyping system when investigating sexual assault case in many South African sub-populations. Therefore the exercise of internal validation of the minimal haplotype system was successfully carried out in the laboratory.

As part of development process, the properties of additional novel and widely used STRs were investigated. Loci were successfully sequenced and allele nomenclature was assigned to them according to the ISFG guidelines.

Loci DYS449, DYS481, DYS518, DYS557, DYS570, DYS607, DYS612 and DYS614 were then selected for use in a new multiplex and named the eight Y-STR's system. Population data from Cape Muslim community was generated for the eight Y-STR's system. The exercise of internal validation of this system was also successfully carried out in the laboratory.

Future work can be carried out in order to further optimize the genotyping systems described in this study, as well as to investigate additional loci for possible inclusion in the newly developed 8 Y-STR system.

References

Ali, S and Hasnain, SE (2002). Molecular dissection of the human Y-chromosome. *Gene* **283 (1-2)**, 1-10

Ayub, Q., Mohyuddin, A., Qamar, R., Mazhar, K., Zerjal T., Mehdi, SQ., and Tyler-Smith, C. (2000). Identification and characterisation of novel human Y-chromosomal microsatellites from sequence database information. *Nucleic Acids Res* **28(2)**, e8

Álves C., Gusmao, L., Barbosa, J., and Amorim, A. (2003). Evaluating the informative power of Y-STRs: a comparative study using European and new African haplotype data. *Forensic Science International* **134**, 126-133

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

Bosch, E., Lee, A., Calafell, F., Arroyo, E,, Henneman, P., de Knijff, P. and Jobling, M.A. (2002). High resolution Y chromosome typing: 19 STRs amplified in three multiplex reactions. *Forensic Science International* **25**, 42-51

Buhler, E. (1980). A synopsis of the human Y Chromosome. *Human Genetics* **55**, 145-175

Butler, J., Schoske, R., Vallone, P., Kline, M.C., Redd, A.J. and Hammer, M.F. (2002). A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers. *Forensic Science International* **129**, 10-24

Butler, J. (2003). Recent Developments in Y-Short Tandem Repeats and single nucleotide polymorphism analysis. *Forensic Science Revised* **15**, 91-111

Butler, J. (2005). *Forensic DNA Typing: Biology and Technology behind STR Markers*, 2nd edn. London: Academic Press.

Butler, J., Decker, A., Vallone, P., and Kline M. (2006). Allele frequencies for 27 Y-STR loci with U.S. Caucasian, African–American and Hispanic samples. *Forensic Science International* **156**, 250-260

Da Costa, Y. and Davids, A. (editors) (1994). *Pages from Cape Muslim History*, Cape Town: Shuter and Shooter.

Daniels, D., Hall, A. and Ballantyne, J. (2004). SWGDAM developmental validation of a 19-locus Y-STR system for forensic casework. *Journal of Forensic Science* **49**, 1-16

Dettlaff-Kakol., A. and Pawlowski, R. (2002). First Polish DNA "manhunt" - an application of Y-chromosome STRs. *International Journal of Legal Medicine* **116**, 289-291

Edwards, A., Civitello, A., Hammond, H.A. and Caskey, C.T. (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *American Journal of Human Genetics* **49**, 746-756

Ehrenreich, L., Benjeddou, M., Davison, S., D'Amato. M.E. and Leat, N. (2008). Nine-locus Y-STR profiles of Afrikaner Caucasian and mixed ancestry populations from Cape Town, South Africa. *Legal medicine (Tokyo, Japan)* **10(4)**, 225-227

Gill, P., Brenner, C., Brinkmann, B,, Budowle, B., Carracedo, A., Jobling, M.A., de Knijff, P., Kayser, M., Krawczak, M., Mayr, W.R., Morling, N., Olaisen, B., Pascali, V., Prinz, M., Roewer, L., Schneider, P.M., Sajantila, A. and Tyler-Smith, C. (2001). DNA Commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome STRs. *Forensic Science International* **124**, 5-10

Góes, A.C.S., Silva, D., Domingues, C.S., Sobrinho, J.M. and Carvalho, E.F.C. (2002). Identification of a criminal by dna typing in a rape case in Rio de Janeiro, Brazil. *Sao Paulo Medical Journal* **120 (3)**, 77 - 80

Gusmão, L., Brion, M., González-Neira, A., Lareu, M. and Carracedo, A. (1999). Y-chromosome specific polymorphisms in forensic analysis. *International Journal of Legal Medicine* **1**, 55-60.

Gusmão,L., Butler, J., Carracedo, A., Gill, P., Kayser, M., Mayr, W.R., Morling, N., Prinz, M., Roewer, L., Tyler-Smith, C. and Schneider, P.M. (2006). DNA Commission of the International Society of Forensic Genetics (ISFG): An update of the recommendations on the use of Y-STRs in forensic analysis. *Forensic Science International* **157**, 187-197.

Hall, T. (1999). Bio Edit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 95-98.

Hall, A. and Ballantyne, J, (2003). Novel Y-STR typing strategies reveal the genetic profile of the semen donor in extended interval post-coital cervicovaginal samples. *Forensic Science International* **139**, 58-72

Hanson, E. and Ballantyne, J. (2004). A highly discriminating 21 locus Y-STR "megaplex" system designed to augment the minimal haplotype loci for forensic casework. *Journal of Forensic Science* **49**, 1-12

Hanson, E. and Ballantyne, J. (2006) Comprehensive annotated STR physical map of the human Y chromosome: Forensic implications. *International Journal of Legal Medicine* **8**, 110-120

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

Iida, R., Tsubota,E., and Matsuki,T. (2001). Identification and characterization of two novel human variableSTRs on the Y chromosome. *International Journal of Legal Medicine* **115**, 54-56.

Iida,R., Tsubota,E., Sawazaki,K., Masuyama,M., Matsuki,T., Yasuda,T. and Kishi, K. (2002). Characterization and haplotype analysis of the variable Y-STRs DYS443, DYS444 and DYS445 in a Japanese population. *International Journal of Legal Medicine* **116**, 191-194.

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

Jobling, M. and Tyler-Smith, C. (1995). Fathers and sons: the Y chromosome and human evolution. *Trends in Genetics* **11(11)**, 449-456

Jobling, M. and Tyler-Smith, C. (2000). New uses for new haplotypes: the human Y chromosome, disease and selection. *Trends in Genetics* **16 (8)**, 356 -362

Kayser, M., Caglia, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., Heidorn, F., Herrmann, S., Herzog, B., Hidding, M., Honda, K., Jobling, M., Krawczak, M., Leim, K., Meuser, S., Meyer, E., Oesterreich, W., Pandya, A., Parson, W., Penacino, G., Perez-Lezaun, A., Piccinini, A., Prinz, M., Schmitt, C., Roewer L., et al. (1997). Evaluation of Y-chromosomal STRs: a multicenter study. *International Journal of Legal Medicine* **110**, 125-133 and 141-149

Kayser, M., Kittler, R., Erler, A., Hedman, M., Lee, A.C., Mohyuddin, A., Mehdi, S.Q., Rosser, Z., Stoneking, M., Jobling, M.A., Sajantila, A. and Tyler-Smith, C. (2004). A comprehensive survey of human Y-chromosomal microsatellites. *American Journal of Human Genetics* **74**, 1183-1197.

Kirby, L. T. (1992). *DNA Fingerprinting. An Introduction*. New York: W. H. Freeman

Koyama, H., Iwasa, M., Tsuchimoshi, T., Maeno, Y., Isobe, I., Matsumoto, T. and Nagao, M. (2002). Utility of Y-STR haplotype and mtDNA sequence in personal identification of human remains. *American Journal of Forensic Medical Pathology* **23**, 181-185

Kurihara, R., Yamamoto, T., Uchihi, R., Li, S.L., Yoshimoto, T., Ohtaki, H., Kamiyama, K. and Katsumata, Y. (2004). Mutations in 14 Y-STR loci among Japanese father-son haplotypes. *International Journal of Legal Medicine* **118**, 125-131

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

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

Leat, N., Ehrenreich, L., Benjeddou, M., Cloete, K. and Davison, S. (2007). Properties of novel and widely studied Y-STR loci in three South African populations. *Forensic Science International* **168 (2- 3)**, 154 - 161

Lee, H. and Ladd, C. (2001). Preservation and collection of biological evidence. *Croatian Medical Journal* **42**, 225-228

Litt, M. and Lutty, J.A. (1989). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human Genetics* **44**, 397–401

Moretti, T.R., Baumstark, A.L., Defenbaugh, D.A., Keys, K.M., Smerick, J.B. and Budowle, B. (2001). Validation of short tandem repeats (STRs) for forensic usage: performance testing of fluorescent multiplex STR systems and analysis of authentic and simulated forensic samples. *Journal of Forensic Science* **46**, 647–660

Mulero, J.J., Chang, C.W., Calandro, L.M., Green, R.L., Li, Y., Johnson, C.L., and Hennessy, L.K. (2006). Development and validation of the AmpFłSTR® Yfiler PCR amplification kit : A male specific, single amplification 17 Y-STR multiplex system. *Journal of Forensic Science* **51(1)**, 64-75

Nei, M. (1987). *Molecular Evolutionary Genetics*. New York: Columbia University Press

Ploski, R., Wozniak, M., Pawlowski, R., Monies, D. M., Branicki, W., Kupiec, T., Kloosterman, A., Dobosz, T., Bosch, E., Nowak, M., Lessig, R., Jobling, M. A., Roewer, L., Kayser, M. (2002). Homogeneity and distinctiveness of Polish paternal lineages revealed by Y chromosome microsatellite haplotype analysis. *Human Genetics* **110**, 592-600

Prinz, M., Ishii, A., Coleman, A., Baum., H.J. and Shaler, R.C. (2001). Validation and casework application of a Y chromosome specific STR multiplex. *Forensic Science International* **120 (3)**, 177-188

Quintana-Murci, L. and Fellous, M. (2001). The Human Y Chromosome: The Biological Role of a 'Functional Wasteland'. *Journal of Biomedicine and Biotechnology* **1**, 18-24

Quintana-Murci, L., Krausz, C. and McElreavey, K. (2001). The human Y-chromosome: function, evolution and disease. *Forensic Science International* **118 (2-3)**, 169-181

Redd, A., Agellon, A., Kearney, V., Contreras, V.A., Karafet, T., Park, H., de Knijf, P., Butler, J.M. and Hammer, M.F. (2002). Forensic value of 14 novel STRs on the human Y chromosome. *Forensic Science International* **130**, 97-111.

Roewer, L., Krawczak, M., Willuweit, S., Nagy, M., Alves, C., Amorim, A., Anslinger, K., Augustin, C., Betz, A., Bosch, E., Cagliá, A., Carracedo, A., Corach, D., Dekairelle, A.F., Dobosz, T., Dupuy, B.M., Füredi, S., Gehrig, C., Gusmaõ, L., Henke, J., Henke, L., Hidding, M., Hohoff, C., Hoste, B., Jobling, M.A., Kärgel, H,J,, de Knijff, P., Lessig, R., Liebeherr, E., Lorente, M., Martínez-Jarreta, B., Nievas, P., Nowak, M., Parson, W., Pascali, V.L., Penacino, G., Ploski, R., Rolf, B., Sala, A., Schmidt, U., Schmitt, C., Schneider, P.M., Szibor, R., Teifel-Greding, J. and Kayser, M. (2001). Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes. *Forensic Science International* **118**, 106-113

Shell, R. (1993). From Rites to Rebellion: Islamic conversion, urbanization, and ethnic identities at the Cape of Good Hope, 1797 to 1904. *Can. J. Hist*. (December), 409-458.

Shell, R C.-H. In: Levtzon, N., Pouwels, R., (editors)(2000). *Islam In Africa*, Ohio University Press, 331-352.

Shewale J, Nasir H, Schneida E, Gross, A.M., Budowle, B. and Sinha, S.K. (2004). Y-chromosome STR system, Y-PLEX^{TM} 12, for forensic casework: development and validation. *Journal of Forensic Science* **49**, 1-13

Scientific Working Group on DNA Analysis Methods (SWGDAM). (2004). Revised validation guidelines. *Forensic Science Communications* **6**, 3

Wallin, J., Holt C, Lazaruk, K., Nguyen, T.H., and Walsh, P.S. (2002). Constructing universal multiplex PCR systems for comparative genotyping. *Journal of Forensic Science* **47**, 52-65

White, P., Tatum, O., Deaven, L. and Longmire, J.L. (1999). New, male-specific microsatellite markers from the human Y chromosome. *Genomics* **57**, 433-437

Willuweit, S. and Roewer, L, (2007). Y chromosomal haplotype reference database (YHRD): Update. *Forensic Science Internationa: Genetics* **1**, 83-87

Electronic Supplementary Resources

International Association of Chiefs of Police	http://www.theiacp.org
South African Police Service	http://www.saps.org.za
Rape Survivor Journey	http://www.rape.co.za
Rape Crisis:	http://www.rapecrisis.org.za
YHRD:	http://www.yhrd.org
Primer3:	http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi

Appendix

1 DNA Extraction, Quantification and Working Stock Solutions

1.1 Lahiri and Nurnberger (1991) DNA Extraction Protocol

1.1.1 Reagents (To be autoclaved immediately after preparation)

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 H_2O	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
Distilled H_2O	82ml
Nonidet P40 (*Sigma*)	2.25ml

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 H_2O	62ml

10% w/v SDS (100ml)

SDS (*Merck Laboratory Supplies*)	10g
Distilled H_2O	100ml

2M NaCl (100ml)

NaCl (*Merck Laboratory Supplies*)	11.67g
Distilled H_2O	100ml

1.1.2 DNA Extraction steps

1. 0.5ml of carefully mixed whole blood was transferred to a clean, dry, APPROPRIATELY LABELLED 1.5ml microfuge tube.
2. 0.5ml of sterile [TKM I + Nonidet P-40] was added to the 0.5ml blood and the contents of the tube mixed gently by inversion.
3. The tube was then placed in a bench-top centrifuge (*Eppendorf, 5415 D*) and the mixture centrifuged at 5000 rpm for 10 minutes to pellet the nuclei.
4. The supernatant was then carefully removed, paying attention to not disturb the pellet.
5. The pellet was then washed by adding 0.5ml sterile TKM I to it.
6. The tube was centrifuged for 10 minutes at 5000 rpm.
7. The supernatant was removed again. This time, extra care was taken to remove as much of it as possible, without disturbing the pellet.
8. 70µl of sterile TKM II was added to the pellets and tube vortexed (*Stuart Scientific, vortex mixer SA3*) until the pellet was completely re-suspended in the liquid.
9. 4.37µl of sterile 10% w/v SDS was added to this mix and the tube vortexed briefly.
10. The tube was then incubated in a water-bath (*Memmert*) at 55^{o}C for 10 minutes.
11. After incubation, 264µl of sterile 2M NaCl was added and the tube vortexed briefly.

12. The tube was then centrifuged at 13000 rpm for five minutes to pellet extra-cellular components.

13. After centrifugation, the supernatant was transferred to a clean, dry and CORRECTLY LABELLED, 1.5ml microfuge tube.

14. To this supernatant, 677µl of absolute ethanol (room temperature) was added and the contents of the tube mixed gently by inverting it a few times. At this point the DNA became visible.

15. The tube was then centrifuged at 13000 rpm for five minutes to pellet the DNA. (The DNA did not always end up at the bottom of the tube, but sometimes got stuck on the side of the tube).

16. The supernatant was then removed and the tube centrifuged again at 13000 rpm.

17. The supernatant was removed after centrifugation and 250µl of ice-cold 70% ethanol added to wash the DNA.

18. The tube was then centrifuged at 13000 rpm for five minutes and the supernatant removed.

19. Step 17 and 18 were repeated.

21. The DNA was dried at room temperature, re-suspended in 100µl of sterile distilled water and the DNA allowed to go into solution at room temperature overnight.

All the blood waste from the DNA extraction process were decanted into appropriate containers, sealed properly, clearly labelled as blood waste and discarded in an appropriate manner by *Waste Tech*, an accredited waste removal company.

1.1.3 Quantifying DNA and working stock dilutions

1. 5µl of each of the extracted DNA aliquots were diluted in 100µl of sterile distilled water.

2. Each of the dilutions was then placed in a quartz cuvette in a spectrophotometer (*Milton Roy, Genesis 5*) and the absorbance readings taken over a range of wavelengths from 240nm to 300nm.

3. The DNA concentration was calculated from the absorbance readings at 260nm.

4. Working stock dilutions at 2ng/µl were made from all the DNA samples.

5. The original DNA stocks as well as working stock solutions were stored at –20^{o}C

1.2 BuccalAmp™ Extraction (*Epicentre Biotechnologies*) Kit Protocol

1.2.1 Materials used during BuccalAmp™ Extraction

Cotton buccal swabs containing male DNA

BuccalAmp™ DNA Extraction 15 reactions Kit (Cat no. BQ0901S)

1.2.2 BuccalAmp™ Extraction DNA Extraction steps

1. Label the appropriate number of tubes containing QuickExtract DNA Extraction Solution 1.0.
2. Thoroughly rinse out the subject's mouth twice with water. We recommend that subjects abstain from drinking coffee before tissue collection. Alternatively, instruct subjects to gently brush the inside surface of both cheeks with a toothbrush (without toothpaste); follow with a thorough rinsing of the mouth with water.
3. Collect DNA (tissue) by rolling the cotton swab firmly on the inside of the cheek, approximately 20 times on each side, making certain to move the brush over the entire cheek. Yield is directly correlated with the starting amount of buccal cells. If yield is not a concern, use only one swab; if yield must be maximized, use a separate swab for each cheek surface, and if necessary, use a third swab, collecting tissue from both cheeks.
4. Place the swab end of the cotton swab into a tube containing QuickExtract DNA extraction solution and rotate the swab a minimum of 5 times. Press the swab against the side of the tube and rotate the swab while removing it from the tube to ensure most of the liquid remains in the tube.
5. Screw the cap on the tube tightly and vortex mix for 10 seconds. Incubate the tube at 65^{o}C for 1 minute.
6. Vortex mix for 15 seconds.
7. Transfer the tube to 98^{o}C and incubate for 2 minutes.

8. Vortex mix for 15 seconds.

9. Quantitate DNA yield using Nanodrop ND-1000.

10. Store the DNA at -20^{o}C, or at -70^{o}C for long term storage.

11. The yield of DNA is usually between 2-14 ng/ml.

1.2.1 Quantifying DNA and working stock dilutions

1. DNA concentration was calculated using the NanoDrop™ ND-1000 Spectrophotometer.

2. The NanoDrop^{TM} 1000 Spectrophotometer enables highly accurate UV/Vis analyses using as little as 1 µl samples. We recommend the using 2 µl.

3. DNA concentration for each of the extracted DNA aliquots were recorded and linked to corresponding DNA extracted stock solutions numbering.

4. Working stock dilutions at 5ng/µl were made from all the DNA samples.

5. The original DNA stocks as well as working stock solutions were stored at –20^{o}C

2 Short Tandem Repeat sequencing protocol

2.1. Ethanol/EDTA/Sodium Acetate precipitation protocol

2.1.1 Reagents require for DNA precipitation

125mM EDTA solution

EDTA (*Merck Laboratory Supplies*)	46, 53 g
Distilled H_2O	up to 100 ml

3M Sodium Acetate

$NaCH_3COO$ (*Merck Laboratory Supplies*)	24, 60g
Distilled H_2O	up to 100 ml

70% Ethanol

100 % Ethanol	70 ml
Distilled H_2O	30 ml

100% Ethanol

2.1.2 Ethanol/EDTA/Sodium Acetate DNA precipitation steps

1. Add 2 µl of 125mM EDTA to each eppendorf tube ensuring that the EDTA solution reaches the bottom of the tube.
2. Followed by 2 µl of 3 M sodium acetate also making sure the sodium acetate solution reaches the bottom of the tube.
3. Add 50 µl of 100% ethanol to each tube and mix by inverting the tubes 4 times.
4. Leave to incubate at room temperature (20^oC) for 15 min.

5 Tubes were then centrifuged at 13000 rpm for 30 minutes

6. It is important to proceed to the next step immediately. If this is not possible, then centrifuge the tubes for 2 minutes more immediately before performing the next step.
7. Remove supernatant and add 70 µl of 70% ethanol to each tube.
8. Centrifuge at 4000 rpm for 15 min, at low temperature if possible ($4^{o}C$)
9. Remove supernatant from each tube, while being careful not to disturbed the area where the pellet is.
10. Allow the ethanol to evaporate and after 20 minutes resuspend the samples in 2 µl ABI loading mixture (consisting of 5µl Hi-Di formamide [*Applied Biosystems*] and 1µl Dextrose Blue loading dye)
11. To store DNA samples, cover with aluminum foil, and store at 4 °C.