Y-STR profiling of four South African populations using the University of the Western Cape 10 locus set.

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

Kebareng Jacobeth Tsiana
November, 2015

Supervisors: Prof. Maria Eugenia D'Amato (PhD)
Prof. Sean Davison (PhD)

Keywords

South Africa

Population

Y-chromosomal Short Tandem Repeats (Y-STRs)

Loci

University of the Western Cape 10 locus set (DYS710, DYS518, DYS385a/b, DYS644, DYS612, DYS626, DYS504, DYS447, DYS449, and DYS481)

Genotyping

Database

Multiplex Polymerase Chain Reaction

Electrophoresis

Allele frequency

Gene diversity

Haplotype Diversity

Discrimination capacity

Polymorphism

Abstract

Y-STR profiling of four South African populations using the University of the Western Cape 10 locus set

Kebareng Jacobeth Tsiana

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

In this study the 10 Y-specific loci of the University of the Western Cape (DYS710, DYS518 385a/b, DYS644, DYS612, DYS626, DYS504, DYS447, DYS447, and DYS481) were analysed in 492 individuals from South African population groups. Four different populations namely; Zulu, Coloured, Afrikaner and Asian Indian were sampled. A total of 488 haplotypes were observed, 412 of which were unique. Haplotype diversity was 0.9981. Gene Diversity values ranged from 0.8075 for DYS447 to 0.9209 for DYS710. The discriminatory capacity was 0.9106 which is high.

The study showed that the University of the Western Cape 10 locus is a powerful discrimination tool for routine forensic applications and could be used in genealogical investigations as compared to other commercial kits when used on the South African populations (Zulu, Coloured, Afrikaner and Asian Indian) considering its high discriminatory capacity. This data will be used for the establishment of a Y-STR DNA databases for South African population which would aid law enforcement authorities in the investigation and resolution of crimes.

AMOVA computed using haplotype frequencies showed that when male haplotypes from the four different populations were compared, 0.22 % of the total genetic variation was due to the variability among populations and 99.78 % of the total variation is found within populations. However AMOVA computed using distance matrix showed that 5.97 % of the total variation

was due to variability among populations and 94.07 % of the total variation is found within populations. Genetic substructure was found among the four studied South African population groups.

All the six population pairwise comparisons using AMOVA were significant .Therefore Y-STRs are very useful in comparing closely related populations. It should be noted that their utility for evolutionary purposes, they need to be combined more stable Y-DNA markers such as single nucleotide polymorphisms (SNPs).

Factorial Correspondence Analysis (FCA) showed that the Coloured population has large genetic contribution from Afrikaner population and lesser contribution from the Zulu and Asian Indian population groups.

Declaration

I declare that *Y-STR profiling of four South African populations using the University of the Western Cape 10 locus set* is my own work and it has not been submitted for any degree or examination from any other university and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Full name: Kebareng Jacobeth Tsiana

Signed:

Date: 15 November 2015

Acknowledgements

First of all, I would like to thank God, my heavenly Father who gave me the strength and the ability to accomplish this study.

I express my gratitude to my supervisors Professor Maria Eugenia D'Amato and Professor Sean Davison for their academic guidance in helping me bring this work to a higher academic level.

I thank all the individuals who gave their biological samples for this study.

I want to specially thank my entire family or their continuous support, love and encouragement; this thesis couldn't have been accomplished without them.

I would like to acknowledge and thank National Research Foundation (NFR) and Inqaba Biotech for funding this research project.

My gratitude also goes to all student and staff who made my experience at the University of the Western Cape in particular the Department of Biotechnology and UWC-Forensic DNA laboratory an enjoyable one.

Dedication

I dedicate this thesis to my mother, Mme Sefedile Tsiana for all her care and support throughout my life.

List of abbreviations

DNA	Deoxyribonucleic acid
STR	Short tandem repeat
Y-STR	Y-chromosome short tandem repeat
SNP	Single nucleotide polymorphism
PCR	Polymerase chain reaction
UWC	University of the Western Cape
bp	base pair
$MgCl_2$	Magnesium chloride

List of tables

Table 3.8

Haplotypes shared by more than one Afrikaner male (*n*=101).

Table 3.9

Haplotypes shared by more than one Asian Indian male (*n*=93).

Table 3.10

Forensic parameters of the four South African populations. N = sample size, *n* haplotypes = number of different haplotypes in the sample, HD = Haplotype Diversity, DC = Discrimination Capacity.

Table 3.11

Haplotype sharing information.

Table 3.12

AMOVA results computed using haplotype frequencies, based on 9 Y-STR loci (locus DYS385a/b was excluded in the analysis).

Table 3.13

AMOVA results computed using distance matrix, based on 9 Y-STR loci (locus DYS385a/b was excluded in the analysis).

Table 3.14

Y- STR haplotype pairwise Fst values for four South African populations (locus DYS385 a/b was excluded in the analysis. All pairwise comparisons are significant.

List of figures

Figure 3.3

Allele frequency distribution of DYS644 within each population group. The observed allele frequency is indicated on the Y-axis, allelic designation (in number of repeats) for each locus is indicated on the X-axis.

Figure 3.4

Allele frequency distribution of DYS612 within each population group. The observed allele frequency is indicated on the Y-axis, allelic designation (in number of repeats) for each locus is indicated on the X-axis.

Figure 3.5

Allele frequency distribution of DYS626 within each population group. The observed allele frequency is indicated on the Y-axis, allelic designation (in number of repeats) for each locus is indicated on the X-axis.

Figure 3.6

Allele frequency distribution of DYS504 within each population group. The observed allele frequency is indicated on the Y-axis, allelic designation (in number of repeats) for each locus is indicated on the X-axis.

Figure 3.7

Allele frequency distribution of DYS447 within each population group. The observed allele frequency is indicated on the Y-axis, allelic designation (in number of repeats) for each locus is indicated on the X-axis.

Figure 3.8

Allele frequency distribution of DYS449 within each population group. The observed allele frequency is indicated on the Y-axis, allelic designation (in number of repeats) for each locus is indicated on the X-axis.

Figure 3.9

Allele frequency distribution of DYS481 within each population group. The observed allele frequency is indicated on the Y-axis, allelic designation (in number of repeats) for each locus is indicated on the X-axis.

Figure 3.10

Gene diversity values for each 10 Y-STR loci, within each population group. All Y-axes indicate the gene diversity values. All X-axes indicate ten Y-STR loci within each population group.

Figure 3.11

Gene diversity values among all 492 male individuals combined. All Y-axes indicate the gene diversity values. All X-axes indicate ten Y-STR within the four South African population groups.

Figure 3.12

Factorial Correspondence Analysis of the UWC 10 locus data set for the different samples in four South African populations.

Table of contents

Chapter 1 - Literature review

Introduction

Research in DNA technologies has helped law enforcement agencies such as the police in the investigation of crimes such as murder, attempted murder, physical assault, and sexual assault. Sexual assault such as rape is one of the most violent crimes and is a serious problem which is faced by many countries and South Africa is no exception. According to the crime statistics, from the South African Police Services between April 2014 and March 2015, (www.saps.gov.za/resource.../statistics/crimestats/2015/crime_stats), 114 270 cases of crime had been reported; 4 691 of such cases were of sexual assault.

In a sexual assault case, investigators collect different types of evidence when possible, such as victim testimony; physical evidence, such as items from the crime scene; and biological evidence (California State Auditor Report, 2014, p. 7). Identification of spermatozoa from the sexual assault victims is the main biological evidence looked for in investigating sexual assault crimes. However this method is limited to semen containing samples. As a result it cannot be used on sample mixtures involving (i)vasectomized or azoospermic men, and the (2) presence of other body–fluid mixtures (for example saliva–skin, skin–sweat) from victims and suspects of different sex (Redd *et al*., 2002; Hanson and Ballantyne, 2007).

Short tandem repeats on the Y chromosome have proven to be valuable DNA markers particularly for resolving such sexual assaults cases where the spermatozoa could be absent (Mc Donald *et al*., 2015; Purps *et al.*, 2015). Therefore targeting male specific polymorphism of the Y chromosome is particularly helpful as it enables the sensitive detection of the presence of male DNA despite the absence of spermatozoa.

1.1 DNA markers used in forensic testing

Tandemly repeated DNA sequences are widely used in forensic DNA testing due to their high rate of mutation (Dieringer and Schlötterer, 2003; Ellegren, 2004) as compared biallelic markers such as single nucleotide polymorphisms (SNPs). Tandemly repeated DNA sequences have multiple alleles while biallelic markers have only two kinds of alleles.

The two groups of tandemly repeated DNA sequences namely; minisatellites and microsatellites differ according to the size of the repeat unit and the overall length of the repeat array.

1.1.1 DNA fingerprinting methods

Minisatellites (VNTRs) consist of repeated sequences that can vary in unit length from 6 to 100 bases repeat. VNTRs were the first polymorphism to be used in DNA-based identification (Jeffreys *et al*, 1988). VNTRs play an important role in forensics because of their high polymorphism, which is a result of very high levels of allele length variability (Tamaki and Jeffreys, 2005). VNTRs are detected by a technique called restriction fragment length polymorphism (RFLP) which involves the use of a restriction enzymes to cut the regions of DNA surrounding the polymorphic region (Jeffreys *et al*, 1985). The resulting DNA fragments are then detected by gel electrophoresis using radioactively labelled DNA probes where a pattern of bands distinctive to the individual will be produced. This method is highly suitable for the detection of alleles in high molecular weight DNA such as in the case in the analysis of DNA from fresh blood samples in a paternity testing scenario or in some sexual assault evidence such as semen (Chimera and Dyer, 1992).

However in forensic case work samples, DNA may be collected from environmentally stressed samples or from old samples, which mostly consist of degraded DNA or low copy DNA which is insufficient for VNTR analysis due to the size of VNTR genetic markers (Chimera and Dyer, 1992). The other disadvantage is the low resolution of gels used for

determining the size of the DNA restriction fragments (Decorte and Cassiman, 1993), resulting inability to determine the alleles discretely. The multilocus VNTRs were more difficult than the unilocus VNTRs for interpretation.

The introduction of the Polymerase Chain Reaction (PCR) provided a significant methodological improvement, in terms of specificity and sensitivity (Mullis *et al,* 1986). Despite all this VNTR analysis was not amenable to PCR based techniques due to the large size of the amplicons. Shorter alleles were found to be preferably amplified as compared to longer alleles (Jeffreys *et al,* 1988; Decorte *et al,* 1990). Because of all these limitations their utility in forensics genetics has now been replaced by STRs.

1 .1.2 Short tandem repeats (STRs)

Short tandem repeats are repeated in tandems of 1-7 base pairs (Goodwin *et al,* 2007; Butler, 2007). They are usually repeated up to 30 times (Tamaki and Jeffreys, 2005). STRs are found on both types of chromosomes: autosomal and the sex chromosome X and Y. They exist as dinucleotide, tri-nucleotide, tetra-nucleotide, penta-nucleotide and hexa-nucleotide. All these different kinds of STR can be utilized for forensic DNA testing. However, dinucleotides and some tri-nucleotides are no longer used because they form stutter artefacts during PCR (Decorte and Cassiman, 1993). As a result penta- and tetra-nucleotides repeats are generally ideal in forensic investigation because they are less susceptible to the formation of these additional products.

STRs are presently the most analysed polymorphism in forensic genetics. This is due to the fact that most of the microsatellite loci can be efficiently amplified by standard PCR since the repeats regions are shorter than a 100 base pairs (Tamaki and Jeffreys, 2005). Their small size is specifically suitable for the analysis of degraded DNA and limited amount of DNA which is common in forensic casework. The polymorphic information content of these

markers is lower compared to VNTR regions (Tamaki and Jeffreys, 2005; Decorte and Cassiman, 1993). But because of their high abundance in the genome and their amenability to PCR based techniques, the same level of information can be obtained by looking at a slightly increased number of markers which is attainable through multiplex PCR (Decorte and Cassiman, 1993).

Initially the research, on the use of STRs for forensic purposes focused only on characterising and evaluating autosomal markers (Urquhart *et al*, 1994; Kimpton *et al*, 1993; Hammond *et al*, 1994). The reason for this was because autosomal markers can be used to identify individuals from both sexes with a high level of discriminatory capacity (Kayser *et al*, 1997). It was also suggested that it was because STRs in Y chromosome occur at a lesser ratio as compared to autosomes (Lutz *et al*, 1992). The difficulty of autosomal genetic markers in resolving biological evidence in sexual assault cases, such as rape, led to interest in research of Y chromosome polymorphisms for application in forensic casework. As a result the use of Y chromosome specific STRs overcame this limitation.

1.1.3 Detection of STRs

Historically STR loci were detected using silver polyacrylamide gels (Lins *et al*, 1996). The limitation of this system was low resolution. The number of loci to be included in the multiplex was restricted because loci with overlapping size ranges could not be co-amplified (Goodwin *et al*, 2007). The Forensic community has now shifted to using a fluorescent detection method which at first used gel electrophoresis (Swerdlow *et al*, 1991) and then capillary electrophoresis (Buel *et al*, 1998; Butler *et al*, 2004). Capillary electrophoresis employs capillary filled with polymer instead of polyacrylamide. Capillary electrophoresis systems are far superior than gels electrophoresis system as they are easy to use, data collection is automated and there is less labour (Butler, 2011).

The advantage of fluorescent detection over other methods is that many loci with overlapping sizes can be detected by attaching different fluorescent dyes to their primers, thus increasing the level of resolution (Kimpton *et al*, 1993).

1.1.4 Challenges of typing DNA with STRs

1.1.4.1 Degraded DNA

Detection of degraded DNA using STRs is a challenge. The size of the amplicon has an effect on the typing of degraded DNA (Senge *et al*, 2011). When the DNA is degraded there is a greater chance of amplifying amplicons shorter than the size ranges found in commercial kits because there is a high chance that longer repeats are fragmented (Chung *et al*, 2004). As a result there will be no intact template DNA.

1.1.4.1.1 Approaches to successfully type degraded DNA

1.1.4.1.1 (a) Mini-STRs

Commercial STR kits amplify fragments in the size ranges of 100-450 base pairs (Mulero *et al*, 2006; Thompson, 2012) . Profiles of such degraded samples may have artefacts such as allele drop out and allele drop in (Schneider *et al*, 2004; Dixon *et al*, 2006). Allele dropout creates a false homozygote.

To successfully amplify such samples, the size of the amplicons has been reduced by redesigning the primers as close as possible to the repeat region (Wiegand and Kleiber, 2001**;** Tsukada *et al*, 2002). This approach has shown to improve the chances of successfully typing degraded DNA (Butler, 2007; Abrahams and Benjeddou, 2011).

However, the regions close to the polymorphic region are susceptible to mutations (Rolf *et al*, 2011). The presence of polymorphism within the primer binding site could result in the occurrence of a null allele as the primers will mismatch the target or the priming site. Therefore concordance studies are very necessary with any newly developed miniSTRs.

1.1.4.1.1 (b) SNPs

Single nucleotide polymorphisms (SNPs) are another type of marker which have been analysed for applicability in resolving degraded DNA samples (Sobrino *et al*, 2005; Inagaki *et al*, 2004). They are more ideal than miniSTRs, since they produce amplicons much smaller than miniSTRs. But due to their low level of polymorphism large numbers of SNPs are required in order to obtain the same discrimination power as the STRs. Approximately twice as many SNP markers were needed to provide the same information as STRs (Fernández1 *et al*, 2013).

1.1.4.2 Stutters

Another problem associated with typing STRs is artefacts called stutters. These products have been reported to be caused by the slippage of the *Taq* DNA polymerase enzyme (Levinson and Gutman, 1987; Klintschar and Wiegand, 2003). High concentrations of magnesium have been found to have an effect on the fidelity of *Taq* polymerase (Eckert and Kunkel, 1990). An increase in magnesium concentration has been directly correlated with the increase in *Taq* polymerase slippage which results in the formation of artefacts called stutter bands (Mulero *et al*, 2006; Viguera *et al*, 2001). Because these additional PCR products are mostly one repeat unit smaller than the expected main allele peak; it can be difficult to distinguish an allele belonging to a minor contributor from a stutter product of an allele of the major contributor.

1.2 Polymerase Chain Reaction (PCR)

Since its first description in 1985, the Polymerase Chain Reaction (PCR) has gained popularity in the field of molecular biology and has significantly aided molecular biology by enabling fast and sensitive analysis of DNA information (Mullis *et al*, 1986). This technique enables the amount of DNA to be increased to a level where it be can manipulated with ease.

PCR is based on the principle of DNA complementarity which was discovered by Watson and Crick (Watson and Crick, 1953). It is an *in-vitro* enzymatic process is similar to the replication process used by the cells to copy their own DNA (Goodwin *et al*, 2007). During each cycle the amount of target DNA is duplicated. In each PCR cycle, the two DNA strands are separated by heat in the denaturation step and are incubated with DNA polymerase, deoxyribonucleotide triphosphate (dNTPs) and primers. In the next step annealing, the primers bind to the target DNA specifically at the complementary sequence on the DNA. The DNA polymerase catalyses the synthesis of the complementary strand in the extension step. The newly synthesised DNA commonly referred as the PCR product or amplicon will act as the template DNA sequence in subsequent cycles.

Initially when PCR was developed the Klenow fragment of *E.coli* was used for the extension step (Saiki *et al*, 1988). Due to the thermo sensitivity of this enzyme, it meant fresh enzyme had to be added at the beginning of every elongation/extension step due to the loss in activity of the enzyme (Saiki *et al*, 1988) .

Later the replacement with the thermostable DNA polymerase from the hot spring *Thermus Aquaticus* for its Klenow fragment became the most significant development in PCR technology (Saiki *et al*, 1988). The thermostability of the enzyme meant that it did not have to be added every time after every cycle. This feature of the enzyme greatly enabled easy automation of the temperature cycling conditions (Reynolds and Sensabaugh, 1991).

The use of the PCR for human DNA profiling is one of the most important developments in forensics (Reynolds and Sensabaugh, 1991). The limitations associated with typing of DNA in forensic such as low copy number DNA and degraded DNA were solved through the use of PCR. Furthermore DNA based typing is highly sensitive and are fast (Decorte *et al*, 1990).

However the sensitivity of the PCR raises major concerns about contamination of the DNA samples or the PCR reactions since even minute DNA contamination will be amplified leading false positive results (Decorte and Cassiman, 1993). This could lead to incorrect conclusions and in extreme cases might exclude or include suspects from being included in the crime (Decorte and Cassiman, 1993).

1.2.1 Multiplex PCR

Multiplex PCR is the most commonly used type of PCR in forensics. Multiplex PCR involves the amplification of more than one target, by incorporating more than one primer pair sequence in the reaction tube.

1.2.1.1 Advantages of multiplex PCR

Multiplex PCR has emerged as a crucial assay as it is cheaper than singleplex since there is a reduction in the cost of analysis and labour from assaying multiple markers simultaneously (Butler *et al*, 2001). The expense of reagents is far less in multiplex PCR than any other type of PCR where several tubes for singleplex are used (Edwards and Gibbs, 1994). Time and effort in the laboratory is saved as the amount of knowledge obtained per unit time in the laboratory is increased (Elnifro *et al*, 2000). Lastly in instances where the test evidence sample volume is inadequate, multiplexing enables information to be obtained from inadequate samples (Butler *et al*, 2001) without compromising the test utility (Elnifro *et al*, 2000).

1.2.1.2 Importance of multiplex PCR in forensic testing

Multiplex PCR in forensic testing has enabled more information to be obtained from a test sample with limited availability which is usually the common case in forensic case work. High degree of discrimination between individuals is obtainable in a relatively short period of time. The amplification of multiple STR loci in the same tube is achieved through the use of oligonucleotide primers that bind to various regions of the genome. Multiplex PCR analysis

has greatly increased the power of discrimination with the ability to analyse multiple regions of DNA at the same time. Multiple markers from the individual are analysed in order to determine the person's DNA profile. The higher number of markers analysed the greater chances of obtaining a unique profile STR profile which is the goal of forensic testing.

1.2.1.3 Challenges of multiplex PCR

PCR primer design and optimization of a multiplex PCR is a greater challenge than for singleplex PCR because the annealing of multiple primers needs to take place under the same annealing conditions without the primers interfering adversely with each other (Schoske *et al*, 2003). Hence the success of multiplex PCR is dependent on the compatibility of primers to be included in the multiplex. Extensive optimization is normally required to obtain a good balance between amplicons of the various loci being amplified (Butler *et al*, 2001; Henegariu *et al*, 1997). These assays can be laborious and time-consuming to establish, requiring lengthy optimization procedures with the adjustment of a number of variables such as primer concentrations (Butler *et al*, 2001; Schoske *et al*, 2003; Henegariu *et al*, 1997).

Several challenges are associated with multiplexing include poor sensitivity and specificity, and/or preferential amplification of certain specific loci (Polz and Cavanaugh, 1998; Markoulatos *et al*, 2002). Having more than one primer pair in a PCR increases the chances of obtaining spurious amplification products, mainly due to the formation of primer dimers (Brownie *et al*, 1997). These nonspecific products may be amplified more efficiently than the desired target, consuming reaction components and generating impaired rates of annealing and extension (Elnifro *et al*, 2000).

Primers in multiplex are labelled with a fluorescent dye which is attached to the 5' (non-reactive) end of a primer that is incorporated through NHS-esters and amine linkages (Schoske *et al*, 2003; Butler *et al*, 2004). These primers use fluorescence energy transfer to

increase the absorption and emission properties of the dye label (Ju *et al*, 1995). Each of the primers has its own unique strong excitation wavelength and fluorescence emission maximum (Ju *et al*, 1995; Schoske *et al*, 2003).

1.3 Features of a forensic testing system

Any DNA marker system to be considered for forensic testing should have certain characteristics. Firstly it should be robust; enabling the effective recovery of DNA information from biological samples collected from a crime scene .The system should enable the DNA profiles to be detected from a wide range of environmentally challenged samples which could be encountered in forensic casework. This enables the system to be easily employed by different forensic laboratories worldwide (Hedman, 2011). The system should have high discriminatory capacity in order to distinguish DNA profiles from different individuals with high statistical confidence. Lastly the testing system should exhibit a fairly even distribution of diversity within and among different populations (Hedman, 2011). This will enable the same marker system to be used in different countries and populations (Hedman, 2011). This allows for the universal standardization of the system which will facilitate comparison of data between different forensic laboratories worldwide (Hedman, 2011).

Commercial companies have developed DNA testing kits for both autosomal and Y chromosome. These kits have been optimized and validated for a variety of forensic casework. STR multiplex kits are developed in such a way that they enable robust analysis of markers investigated.

1.4 Y chromosome

1.4.1 Structure of Y chromosome

The Y chromosome which is considered to be one of the smallest human chromosome consists of distinctive segments: the pseudo autosomal regions (PARs) which is about 5% of

the chromosome and the non-recombining portion (NRY) which is about 95% of the chromosome as reviewed by b Gusmão *et al*, 1999. The pseudo autosomal region consists of two portions; PAR1 and PAR2. PAR1 is situated at the extreme end of the short arm while the PAR2 is situated at the end of the long arm (Quintana-Murci and Fellous, 2001) . The pseudo autosomal region is similar to the X chromosome and undergoes meiosis. During meiosis, one of the pseudo autosomal regions, especially the PAR2 exchanges genetic material with the pseudo autosomal region of the X chromosome (Jobling and Tyler-Smith, 2003). As a result genes found within the PAR region are passed on to the next generation in the same way as the autosomal genes (Quintana-Murci and Fellous, 2001).

NRY doesn't exchange genetic material during meiosis, and as result it is inherited in its intact form. It consists of two major regions the euchromatin which is about 23 megabases (Mb) and the hypervariable heterochromatin (Skaletsky *et al*, 2003; Willard, 2003). The length of the euchromatin is constant among the males contrary to heterochromatin, which could be undetected in some males as reviewed by Gusmão *et al*, 1999.

The euchromatin consists mostly of genes which are transcriptionally active including genes for male sex cell development, the sex–determining region Y (SRY) (Quintana-Murci and Fellous 2001). The SRY gene causes the foetus to become a male through the development of the testis in early stages of embryogenesis (Jobling *et al*, 1997). Lastly the heterochromatin which includes the distal long arm is composed mostly of two highly variable repetitive DNA sequences and is commonly seen as non-functional (Gusmão *et al*, 1999; Willard, 2003) , because it is regarded as non-transcriptable (Quintana-Murci and Fellous, 2001). It consists of sequence families DYZ1 and DYZ2, containing about 5000 and 2000 copies of each respectively (Quintana-Murci and Fellous, 2001). Polymorphic regions in the heterochromatin have been found in several male populations (Quintana-Murci and Fellous, 2001; Iida and Kishi, 2005). Polymorphism found in this region include short tandem repeats

(STRs) and single nucleotide polymorphisms (SNPs). Hence STRs which are suitable for forensic applications are found in this part of the genome. This feature has enabled the application of Y chromosome polymorphism in human population and evolutionary studies (Jobling *et al*, 1997). NRY contains many blocks which are homologous to the X chromosome (Jobling and Tyler-smith, 2000) and therefore has to be excluded for male specific identification.

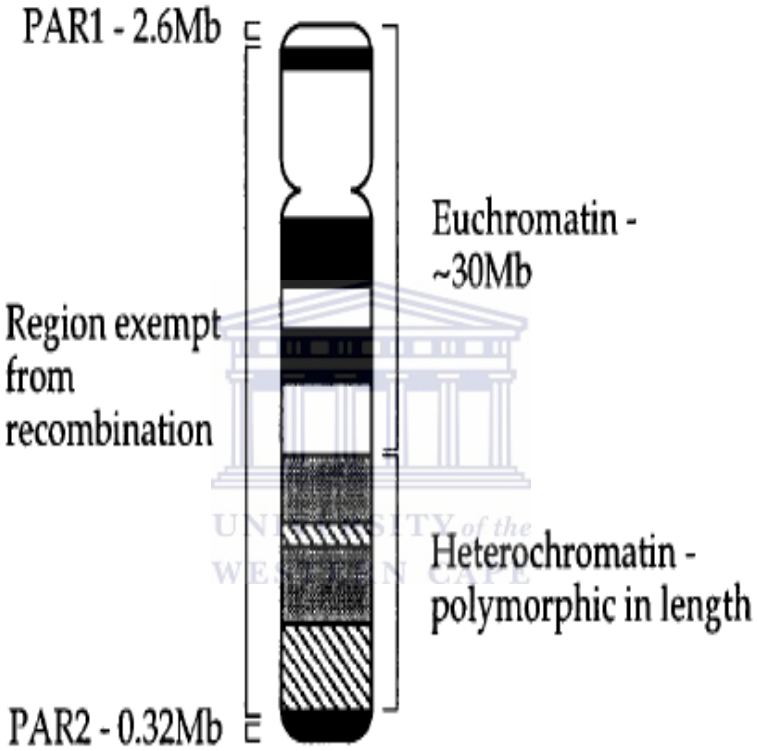

Figure 1.1. Schematic diagram of the human Y chromosome (Jobling *et al*. 1997).

1.4.2 Forensic applications of Y -STRs

1.4.2.1 Forensic casework on sexual assault evidence

Y chromosome markers greatly aid in resolving the majority of violent crimes which are mostly committed by males. The male specificity of the Y chromosome has enabled the discrete detection of the male profile from the crime scene sample when typed with Y specific markers (Jobling *et al*, 1997).

Autosomal DNA tests often fail to provide conclusive results when used in gender testing or in sexual assaults cases. The autosomal kits include the amelogenin locus which is a gender

determining marker. The problem is that the amelogenin test is not reliable as it is prone to errors. Some males are wrongly detected as females because they have a deletion in the AZF gene (Kayser and Schneider, 2009). In some population groups the deletion occurs at low rate (Steinlechner *et al*, 2002), while in others it occurs at high rate (Thangaraj *et al*, 2002). Therefore autosomal testing is not reliable for gender testing.

Y-STR analyses are highly valuable for investigations of sexual assault cases in which the minor male proportion in DNA mixtures frequently remains undetected in standard analyses (Purps *et al*, 2015). Sexual assault evidence in vaginal swabs typically contains a larger proportion of female contribution relative to male contribution (Roewer, 2009). PCR for determining the autosomal DNA profile of male DNA for help identify the perpetrator is often not successful (Hanson and Ballantyne, 2007), due to the preferential amplification of the major component of female cells (Roewer, 2009). Therefore the use of Y-STR markers targets only the male DNA fraction of the sample.

The other advantage of using Y-specific markers is that they give conclusive results in cases where spermatozoa are undetectable or absent in the biological evidence. Reasons for undetectable or absent spermatozoa, include azoospermic or vasectomized perpetrators, lack of ejaculation (Kayser, 2007). The biological evidence can contain admixture of body fluids other than semen, such as in saliva/saliva mixtures, saliva/vaginal secretion mixtures, or fingernail scrapings comprising cells from the (female) victim and cells from the perpetrator (Hanson and Ballantyne, 2007). In the above mentioned cases Y-STR analyses enables the male DNA profile from the perpetrator to be obtained. Y-STR testing is also beneficial in resolving gang rape cases as it enables the determination of number of DNA profiles in the mixture (Hall and Ballantyne, 2003) as the conventional autosomal method cannot be used in sample involving more than one male semen donor (Redd *et al*, 2002).

1.4.2.2 Paternity testing

The inheritance of the Y chromosome along the male lineages enables Y-STR polymorphisms to be used in paternity testing especially in deficiency cases where the father is not available for DNA testing. Inferences are made available in reference to the male relative, who either is a brother or a cousin (Hall and Ballantyne, 2003). For this reason it is possible to determine the complete Y chromosome information of the absent father through the use of the DNA from any paternal male relative.

A classic example is a paternity testing that linked the third US president Thomas Jefferson to the child of one of his slaves, Sally Hemings (Foster *et al*, 1998). However, it should be considered that a result based only on the Y chromosome does not exclude as the father, any male relative in the same patrilineage and thus whenever possible autosomal markers should be used to avoid or reduce this possibility.

1.4.2.3 Human remains identification in mass disasters and missing person's investigations

Mass disasters can be classified into major categories namely natural and manmade. Natural disasters include floods, earthquakes, tsunami, volcanic eruptions, and droughts while man made include transportation disasters such as road accidents, rail, air and maritime disasters, construction disasters, fires and wars (Ziętkiewicz *et al*, 2012). All of these result in the loss of many lives.

In both mass disasters and missing person cases, it is the responsibility of law enforcement agencies to try and identify the human remains of the victims so they can be brought to the victims' families. Identification of victim's remains is necessary for legal and administrative purposes such as obtaining a death certificate which can be used by the victim's relatives to claim money from insurance companies (Butler, 2011; Caenazzo *et al*, 2013). Secondly it brings closure to the victims' families and they can bury their loved ones properly.

The conventional methods used in mass disasters rely on physical characteristics to aid in the identification of human remains (Caenazzo *et al*, 2013; Brenner and Weir, 2003). Often in mass disasters the remains are extremely defragmented, decomposed and the remains of the victims are mixed up (Ziętkiewicz *et al*, 2012). As a result conventional methods become limited in their use as they can only identify an intact body. Using physical characteristics for identification is also not reliable as there is great risk of ambiguity where a person can be wrongly identified (Brenner and Weir ,2003).

DNA profiling has become a superior method in the identification of human remains not only because it helps in the identification of extremely decomposed remains which are beyond recognition, but also helps in the association of fragmented remains (Caenazzo *et al*. 2013; Budowle *et al*. 2005).

Y-STR haplotyping has proved to be useful in tracing paternal relationships in mass disaster identification (Gojanović and Sutlović, 2007; Marjanović *et al*, 2009) and it helps in the identification of missing persons (Davison *et al*, 2008). Due to its lack of recombination most of the Y chromosome is paternally inherited as a block of linked haplotype markers from one generation to the next. Therefore males in the same lineage will have the identical Y-STR haplotype allowing for the determination of the missing person's male haplotype by typing male relatives (Hanson and Ballantyne, 2007).

DNA profiles from the victim's remains are compared with DNA profiles from direct reference samples belonging to the victim such as tooth brush or unwashed clothing. This provide the simplest way to obtain a match thus identification (Butler, 2011). In instances where direct reference samples is unavailable relatives of the victims 's DNA profile is

compared to the victims for indirect identification using kinship analysis such as Y chromosomal testing.

Identification of human remains in mass disasters is problematic as compared to paternity testing because many comparisons have to be made depending on the number of victims (Butler, 2011). The recovery of DNA information from the recovered remains could be limited depending of the type of the disaster and as a result partial or mixed profiles could result after typing these kinds of samples (Butler, 2011).

1.4.3 Importance of Y-STRs in human evolutionary studies

Since the beginning of time humankind have always been concerned about their origins. The Out of Africa theory was hypothesised to explain the origin of mankind (Cann *et al*, 1987; Jobling and Tyler-smith, 1995) . According to the theory there was one genesis for modern humans in Africa which occurred less than 200 000 years ago (Jobling and Tyler-smith, 1995). Men and women moved out of Africa to the rest of the world when climate conditions were unfavourable (Jobling and Tyler-Smith, 2003). The theory has received support from different disciplines such as history, archaeology, paleontobiology as well as linguistics (Stringer and Andrews, 1988). However because of inconsistent evidence supporting the theory it was there scepticism around it.

The improvements of molecular technologies in the late 80's to the early 90's enabled the invention of molecular DNA approaches to address some of the question surrounding human evolution. DNA can aid in explaining origin because it is inherited; transmitted from one generation to the next and it maintains mutations which took place along the way. Although DNA is unique for each individual expect monozygotic twins, this uniqueness or polymorphism shows a record of individual relatedness and genetic history (Jobling and

Tyler-smith, 1995). The Y chromosome in particular can be used to reconstruct paternal human history and provide information on paternal migration distribution due to its unique features; non-recombination and male specificity (Wang and Li, 2015; Nayak *et al*, 2014).

The NRY region of the Y chromosome is passed from father to son unaltered provided no mutation occurs. Once a mutation occurs, modern Y chromosome holds a record of all mutations that happened in the past. This basically means information on ancient Y DNA, which is the origin of mankind, can be known or inferred by using the information or data collected from modern Y chromosome, using DNA polymorphism (Gusmão *et al*, 1999; Jobling *et al*, 1997; Underhill and Kivisild, 2007).

Binary markers such as SNPs have a slower rate of mutation on the order of $3.0 \times 10{-}8$ mutations/nucleotide/generation (Xue *et al*, 2009), as to the compared to the faster multiallelic markers STRs. The mutation rates of STRs are about 4 to 5 orders of magnitude higher than SNPs (Wang *et al*, 2010). SNP markers are used in constructing a phylogeny tree connecting all the Y chromosome lineages from world populations (Y Chromosome Consortium, 2002; Karafet *et al*, 2008; van Owen *et al*, 2013).

STRs on the other hand besides being used in forensic identification, it used in the estimation of population diversity (Wang *et al*, 2013). Y-STR has also been used in time estimations for SNP lineages (Wang and Li, 2013).

The combination of data from binary markers namely; haplogroups and data from STRs namely haplotypes, called lineages, is used to construct phylogenetic trees (Underhill and Kivisild, 2007). The phylogenetic tree consists of main branches, nodes and leaves. The

haplogroups which are the main branches are defined by SNPs; the nodes represent the common ancestors and haplotypes which are individual profiles are the leaves

The latest human Y chromosome tree constructed by van Owen *et al*, 2013 confirms the origin of the modern man in Africa. From the phylogenetic trees information about the level of population structure, pattern and time of population evolution can be deduced (Underhill & Kivisild, 2007; Karafet *et al*,2008; Shi *et al*, 2010).

Studies such as the one done at the University of Cambridge confirmed the 'Out Of Africa" hypothesis that all modern humans stem from a single group of Homo sapiens who emigrated from Africa 2,000 generations ago and spread throughout Eurasia over thousands of years . DNA in particular Y-STRs and other markers such as mitochondrial DNA were analysed and the data was compared with various DNA patterns associated with early humans (Science Daily, 2007).

1.4.4 Limitations of Y-STR typing

The haploid and patrilineal inheritance nature of the Y chromosome makes it difficult to interpret Y-STR data, because male relatives of several generations will have the same Y-STR profile (Roewer, 2009). These two features makes the discriminatory capacity of Y-STR markers much lower when compared with the autosomal STRs. Y-STRs usually differentiate unrelated Y–chromosomes (i.e. paternal lineages) while autosomal STRs can discriminate between any individuals with high statistical confidence (Redd *et al*, 2002). When a crime sample matches the Y-STR profile of a suspect, the patrilineal relative of the suspect cannot be excluded as being the contributor of the biological evidence, more especially if crime runs in the family (Jobling *et al*, 1997; Roewer, 2009). In spite of this, Y-STR markers provide a valuable addition to the forensic DNA tester's tool kit.

However, Ballantyne *et al*, 2010, discovered 13 rapidly mutating (RM) Y-STR from a library of 186 Y-STRS and they are anticipated to change Y-chromosome application from paternal lineage differentiation to male individualisation (Kayser and de Knijff, 2011). According to pedigree studies, conventionally used sets of Y-STRs such as Y-filer, have a lower mutation rate of about $1x\ 10^{-3}$ per generation while the RM-STRs have a higher mutation rate 1×10^{-2} per generation (Ballantyne *et al*, 2012). It has been shown that the RM-STRs can distinguish over 70 % of relatives in comparison with conventional Y-filer Y-STRs which can only distinguish 13% of relatives (Ballantyne *et al*, 2010). Because of their high mutation rate these markers are expected to be useful in cases where the current systems have failed to be informative. Firstly they can be used to resolve cases such as gang rape in which there are several suspects involved (Ballantyne *et al*, 2010, Ballantyne *et al*. 2012). Another application of RM Y-STRs could be instances where conventional Y-STRs was used but did not provide conclusive results about whether a man or his relatives are the stain donors (Ballantyne *et al*, 2010; Ballantyne *et al*. 2012).

According to Kayser and de Knijff (2011), despite all their superiority over the current systems, RM-STRs are unlikely to replace the current Y-STR in human identity testing, even if they are commercialised. It is because the current Y-STRs are useful in other forensic applications such as kinship testing or disaster victim identification, which involve testing relatives using the current Y-STRs which have lower mutation rate to reconstruct family lineages (Kayser and de Knijff, 2011). RM –STRs have been analysed using data from 111 worldwide populations which was generated by 13 centres, and have shown exceptional value in differentiating paternal lineages (Ballantayne *et al*, 2014).

1.4.5 Y-STR testing kits

Several kits are currently available, which contain additional Y-STR loci to supplement the Scientific Working Group on DNA Analysis Methods (SWGDAM) set. The SWGDAM set

include core loci which are DYS19, DYS385a/b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS438 and DYS439. They belong to the basic set of markers which have been selected for use in resolving forensic casework (Kayser *et al*, 1997). It has been shown that analysis with this core set of Y-STRs can discriminate most of the male individuals in various populations around the world (Kayser *et al*, 2002).

The PowerPlex® Y System has been was developed to type these same 11 core loci plus the DYS437 locus (*Promega*) (Krenke *et al*, 2005). The AmpF*l*STR® Yfiler® system (*Applied Biosystems*) includes the loci of the PowerPlex® Y System and the highly polymorphic loci DYS448, DYS456, DYS458, YGATA H4 and DYS635 (Mulero *et al*, 2006).

PowerPlex®Y23 System *(Promega)* includes all of the loci from PowerPlex® Y as well as AmpF*l*STR® Yfiler® as well as six Y-STR loci that are not in any other commercially available kit. The six additional Y-STR loci are DYS481, DYS533, DYS549, DYS570, DYS576 and DYS643. DYS576 and DYS570 are rapidly mutating.

Yfiler® Plus Kit *(Thermo Fisher Scientific)* is based on a multiplex that amplifies 27 Y-STR loci in a single PCR amplification reaction. It includes the loci of the AmpF*l*STR® Yfiler® plus 10 highly polymorphic markers namely; DYS576, DYS627, DYS460, DYS518, DYS570, DYS449, DYS481, DYF387S1a, DYS387S1b and DYS533. Of these 10 markers, 7 are rapidly mutating (DYS576, DYS627, DYS518, DYS570, DYS449, DYF387S1a and DYF387S1b).

The numbers of polymorphic markers included in the commercial kits has been increasing over the years, so as to increase their ability to differentiate paternal lineages. More studies of genetic diversity at non-core Y-STRs on the local populations have also been increasing over the years to search for loci with high forensic significance in local populations (Cloete *et al*,

2010; D'Amato *et al*, 2010; D'Amato *et al*, 2009). A more recent study was undertaken to analyses 36 Y-STR loci, which included all the loci found in the commercial kits (Westen *et al*, 2015).

1.4.6 Reference Y-STR databases

Reference databases which consist of anonymous DNA profiles from specific or different populations were established for the sole purpose of facilitating accurate estimation of haplotype frequencies (Kayser *et al*, 2001; Kayser *et al*, 2002). Y chromosomal Short Tandem Repeat Haplotype Reference Database (YHRD) is the largest and the most publicly accessed Y-STR database. It consists of the male –specific DNA profiles of the loci which constitutes the Y-STR commercial kits (Willuweit and Roewer, 2007). This database currently stores Y-STR profiles from 917 sampling locations in 128 countries which have been submitted by more than 250 institutions and laboratories (Willuweit and Roewer, 2015).

Haplotype frequency value enables the rate of coincidental matches between the samples that do not originate from the same person to be determined. This value provides a measure of the discrimination ability of any forensic testing system (Foreman and Evett, 2001).Without reference databases, it is not possible to assess the evidential weight of the crime and as such the evidence cannot stand in court. This helps in the exclusion and elimination of suspects.

However, the best criterion on how to best use frequency database for estimation of statistical weight of Y-STR haplotype match, is yet to be decided/unknown (de Knijff, 2003; Brenner, 2010) . According to Kayser and de Knijff (2011) insufficient empirical data from the frequency databases makes it difficult to determine the reliability of Y-STR haplotype frequencies, from the current Y-STR databases, because they contain unrelated males. They suggested the problem could be overcome through the establishment of Y-STR haplotype

frequency reference databases which contains DNA profiles from related and unrelated male. In that way the amount of male population's substructure in a region could be reflected.

The information from reference databases can also be used in human evolutionary and population studies of specific or different populations Kayser *et al*, 2002; Willuweit and Roewer, 2015).

1.5 Historical background of South Africa

Introduction

South Africa is known for its rich cultural diversity and diverse people. The diversity of South African population is attributed to its dynamics of colonial history, historical governance and South African situation with regards to main trade routes from the fifteenth to the ninetieth century (Campbell, 1897; The Encyclopedia of World History, 2001)

The population of South Africa is estimated to be 50 586 757 million and is dispersed through all nine provinces (www.statssa.gov.za/publications/populationstats.asp). The people are classified as 79.5 % Black Africans, 9.0 % Coloureds, 2.5 % Asians and Indians and 9 % White (www.statssa.gov.za/publications/populationstats.asp).

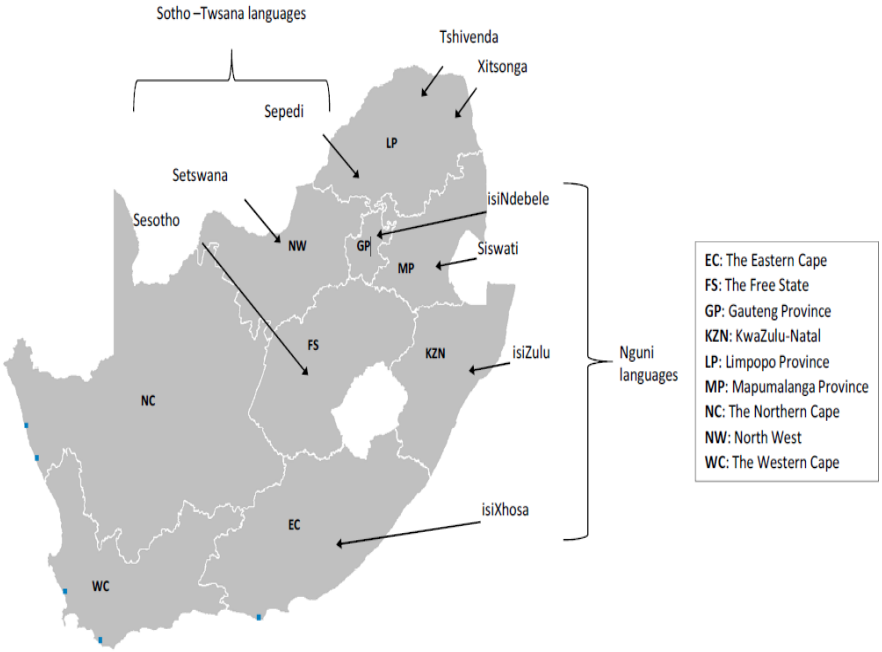

Figure 1.2. Map of South Africa showing the nine provinces and the approximate historical location of the Bantu linguistic population groups in South Africa (Warnich *et al*, 2011) adapted from (Lane *et al*, 2002).

1.5.1 Bantu Speakers/Black Africans/Bantu migration

The Black Africans are the present descendants of the Bantu speaking people who originated from North Western Cameroon/Southern Nigeria and spread throughout Sub-Sahara Africa and the KhoiSan who originated from the eastern Africa (Berniell-Lee *et al*, 2009; Sirugo *et al*, 2008).The Khoisan speakers were the first inhabitants in Southern Africa about 20 000 years ago while the Bantu speakers migrated to Southern Africa about 1000- 2000 yb p(Huffman, 2007; Mitchell, 2002). Of the eleven official languages spoken in South Africa, nine are Bantu languages (Patterson *et al*, 2010). The Bantu languages in the order of the most commonly spoken language namely Zulu, Xhosa, Pedi, Tswana, Southern Sotho, Tsonga, Swati, Venda, and Ndebele belong to Niger Congo phylum. *(http://www.statssa.gov.za/Census2011/Products/Census_2011_Census_in_brief.pdf):* The Niger Congo phylum which is the largest of the four major language families spoken in Africa is spoken mainly by agriculturalist populations across a wide geographic distribution in Africa (Sirugo *et al*, 2008) as indicated by the map in figure 4:2. The South African Bantu

ethnic groups belong to the southern branch of the Eastern Bantu speaking group (http://www.ethnologue.com).

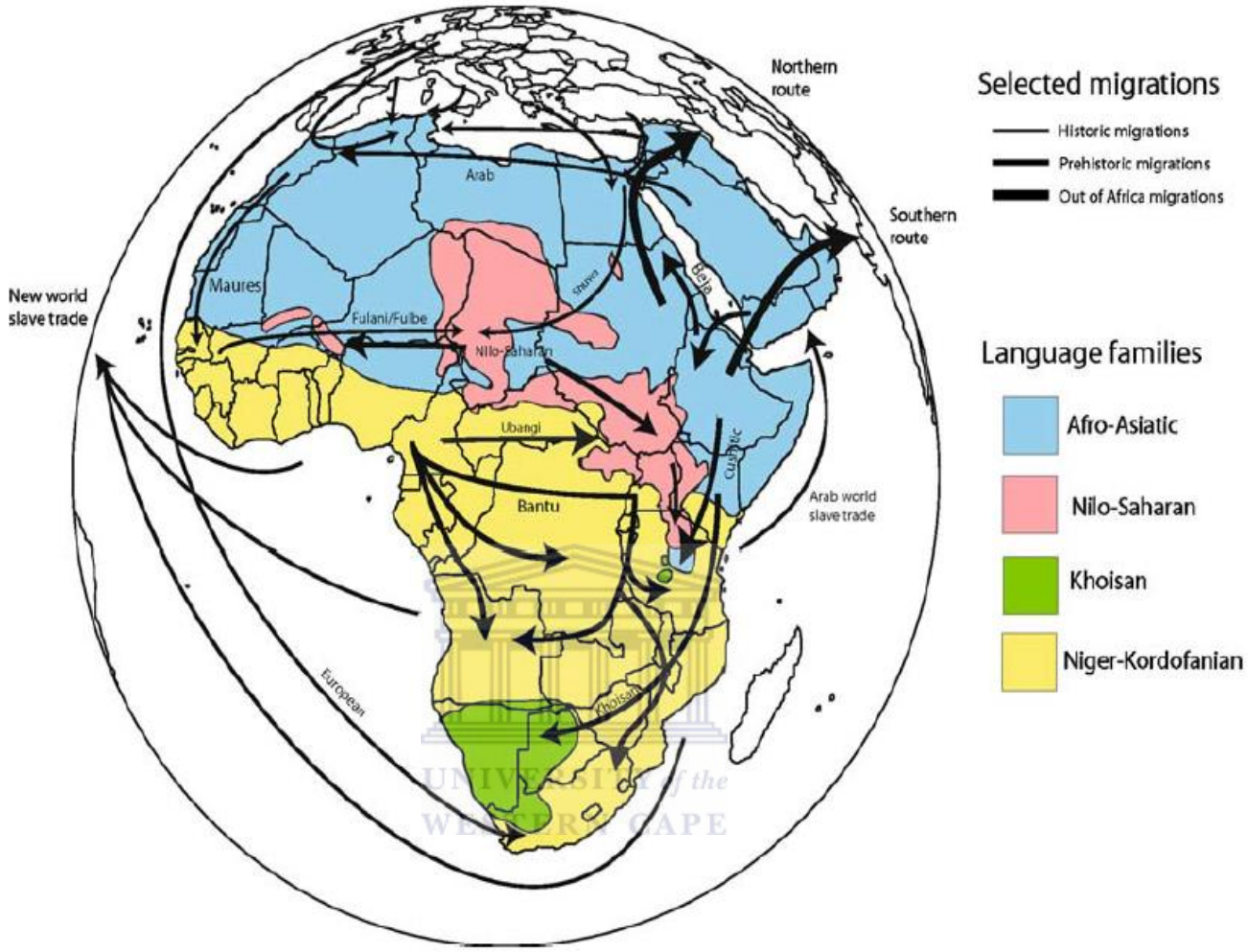

Figure 1.3. A map of selected migrations and language family distributions in Africa (adapted from Reed andTishkoff , 2006). More recent migrations in historical times are represented by *thin arrows* and inferred prehistoric migrations are represented by *medium arrows* (Sirugo *et al*, 2008).

The Bantu migration or Bantu expansion is the most significant and the highly documented migration ever in the history of humankind (Berniell-Lee *et al*, 2009; Peires, 1986). The migration is believed to be influenced by increase in population size, which was as a result of introduction of farming and at a later stage the Iron Age (Vansina, 1995). It occurred in at least two major routes; the western and the eastern route (Berniell-Lee *et al*, 2009).

1.5.2 Bantu speaking people and the Khoisan

Upon their arrival in Southern Africa the Bantu people who were mixed farmers rearing cattle and sheep and growing crops settled on the eastern part of the region which had good soils and came across the Khoikhoi who were also pastoralists and had settled in the same region (Stearns, 2001; Thompson, 2002). They also encountered, the San who were hunters/gatherers who lived of the on the eastern arid parts of Southern Africa (Thompson, 2002; Lewis-Wiiliams, 1986). Overtime many KhoiSan people became part of the Bantu communities either as business partners or customers with the trading of iron tools and weapons (Thompson, 2002). The Bantu also adopted some of the cultures of the Khoisan which is evident by the click sound found in some of the Bantu languages; for example the click sound found in the Xhosa (Southern Nguni)(_Güldemann and Stoneking, 2008). This was originally spoken by the KhoiSan. This illustrates a long period of being close to each other which included inter-marriages (Denoon and Nyeko, 1987).

1.5.3 Diversification of South African Bantu speakers

As time went by South African Bantu speakers started branching from each other. Two main language groups arose, the Nguni and the Sotho Tswana with the smaller groups, the Venda and Tsonga (Saunders and Bundy, 1992). The Nguni such as Zulu, Xhosa settled along the coast while the Sotho Tswana settled on the interior (Stearns, 2001). Those speaking different Nguni languages could understand each other but struggled to understand South-Tswana languages. It is because the different languages belonging to the same group (for example Xhosa and Zulu or Tswana and Pedi) had similar syntax and shared most of the vocabulary (Thompson, 2002). Despite their segregation they all led the same way of life style, they practiced agriculture; growing crops and keeping of livestock (Thompson, 2002; Maggs, 1986). The Nguni practiced agriculture and pastoralism on a small scale while, the Sotho-Tswana relied more on pastoralism than growing crops (Denoon and Nyeko, 1987). They also formed chiefdoms, with the Sotho –Tswana forming bigger kingdoms which expanded

as far as Botswana (Stearns, 2001). Moreover they had homestead systems which were related to the chiefdom though paternal lineages and clans (Peires, 1986). People of the same clan were related either through a common family, a common set of praise songs or a common totemic animal (Peires, 1986). How the modern day ethnic groups such as Xhosa, Zulu came about is still unknown. According to Peires (1986) around 1600 certain Nguni clans began to increase their powers, and took over smaller clans, for example Tshawe founded the Xhosa kingdom after conquering the brothers Cirha and Jwarha in battle and integrated the clans into his clans.

1.5.4 Arrival of the Europeans

Between 100 and 300 years later, the European traders settled in South Africa. The Dutch were the first Europeans to arrive under the leadership of Van Riebeeck 1652. They formed the Dutch East Africa or Dutch Verenidge Oostindischee Compangnie (VOC) which served as a refreshment station for traders between Europe and Asia (Boucher, 1986; Terreblanche, 2002). The company had an economic relationship with the Khoikhoi where they relied on the Khoikhoi for meat (Stearns, 2001). However this relationship was short lived because of two reasons: the high demand for meat by the growing population was not always met and the conversion of grazing land by the incoming settlers into cultivated land interfered with traditional grazing land rights of the Khoikhoi (Boucher, 1986).

Later the British came at the beginning on the 19 Centrury, bringing slaves alongside them from Asia and in Eastern Africa, Madagascar (Campbell, 1897). All these historic events contributed to the rich genetic and cultural diversity of South Africa.

1.5.5 South African ethnic groups

1.5.5.1 Zulu

According to history there were three major Northern Nguni chiefdoms namely: the Ndwandwe under Zwide, the Ngwane under Sobhuza, and the Mthwethwa under Dingiswayo

(Edgecombe, 1986; Denoon and Nyeko, 1987). The Ndwadwe were in the north, the Ngwane settled in the far north while the Mthwethwa were in the south (Saunders and Bundy, 1992). Under Mthwethwa chiefdom was a smaller chiefdom Zulu, under the leadership Senzangakhona, who fathered a child Shaka out of wedlock in about 1787 (Denoon and Nyeko, 1987). Of this three major leaders, Zwide was known for cruelty among his enemies unlike Dingiswayo who incorporated the conquered people into his chiefdom, Zwide took the livestock and scattered people all over place (Edgecombe, 1986). Towards the end of his rule in 1818, Dingiswayo had effectively changed his chiefdom into a fairly large multi-chiefdom alliance that expanded from Mfolozi River in the North to the Tukela in the South (Denoon and Nyeko, 1987).

After sometime they were wars between the Northern Nguni chiefdoms due to disputes over land ownership due to the enlargement of the settlement. As the chiefdoms expanded, their territories adjoined (Saunders and Bundy, 1992). The first clash was between Zwide and Sobhuza. This was won by Zwide and Sobhuza was forced to flee inland and settled in the present day Swaziland (Saunders and Bundy, 1992). After the death of the Zulu chief Senzangakhona in 1816, a conspiracy at Dingiswayo court resulted in the dismissal of Sigujana, the legitimate heir to the Zulu throne, and the imposition of Shaka as the new chief of the Zulu; Sigujana was Shaka's half-brother (Denoon and Nyeko, 1987; Thompson, 2002). A war broke between Zwide and Dingiswayo which lead to defeating of the Mthwethwa and the death of Dingiswayo (Edgecombe, 1986). Following his death Shaka incorporated Mthethwa to the Zulu chiefdom as result increasing his chiefdom (Denoon and Nyeko, 1987). Shaka later conquered the Ndwedwes. His conquest of Ndwandwe was one of the most remarkable events in Zulu history. According to Edgercombe (1986) by 1819 Shaka has defeated all his enemies and integrated them into his kingdom making one united kingdom.

By 1820's Shaka had created the most powerful kingdom in South –east Africa (Edgecombe 1986).

The rise of the powerful Zulu kingdom under Shaka is believed to be the cause of Mfecane wars. The period was characterised by waves of migrations, periodic raiding, series of wars and frequent period of migration for people in Southern Africa (Edgecombe, 1986). It was felt more by the Southern Sotho who had settled of the west of the Drakensberg (Terreblanche 2002).

The impact of Mfecane was felt over most part of Africa from the southern, central and eastern regions of Africa (Edgecombe, 1986; Saunders and Bundy, 1992). It led to the establishment of new kingdoms as far as north as present day Malawi, Zambia and Tanzania (Thompson, 2002). Shaka died in 1828 after he was assainated by his brothers. Today the Zulu is one of the largest ethnic groups in South Africa.

1.5.5.2 Afrikaner

The Afrikaners are the descendant of the small number of Dutch immigrants who settled in the Cape in the 16 Centrury. They had their distinct language knows as Afrikaans, which was derived from the Dutch (Smith, 1952; Thompson, 2002). Towards the end of the 18 Centrury and at the start of the 19 Centrury British settlers arrived in the Cape (Thompson, 2002). The arrival of the British contributed negatively on the independence and the culture of Afrikaner the people. They also disrupted the culture of the indigenous people, such as the Khoisan and the Bantu (Terreblanche, 2002; Thompson, 2002).

Between 1834 and 1840, the Afrikaners left the Eastern Cape and settled into the interior. The immigration of which approximately 15 000 farmers left with their families is known as the Great Trek. Many reasons were proposed for what drove those farmers and their families

into the interior. Lack of adequate land for grazing and they were dissatisfied with the British governance (Saunders and Bundy, 1992; du Bruyn, 1986).

In the Eastern Cape both the Xhosa and the Afrikaners were relying for their survival on the grazing land. As the population grew the land became scarse and they were compelled to move interior northwards and southwards to looks for better pastures (Carruthers, 2007; Saunders and Bundy, 1992). The Boers in the Eastern colony were subjected to constant San raids from the northern part of the Eastern Cape (Denoon and Nyeko, 1987). They were also subjected to frontier wars of the Xhosa attacks (du Bruyn, 1986; Thompson, 2002; Denoon and Nyeko 1987). The Boers felt unhappy and were insecure that even though the British were in power they had failed to protect them from all these attacks.

The idea of moving away from the colonial limits will give them the freedom of running their own affairs without interference from the British governance, however the Boer trek could be seen as of expansion of colonisation (Denoon and Nyeko, 1987). Afrikaners and their families or Voortrekkers as they were called left the Cape Colony in large groups under different trek leaders (Denoon and Nyeko, 1987). The great Trek was not a single combined movement, rather a series of movement by different groups under various trek leaders. The leaders were Andries Hendrik Potgieter, Gert Maritz, Piet Retief and Piet Uys (Thompson, 2002). Potgieter finally settled in the north of Vaal River and established Potchefstroom (Carruthers, 2007). Retief and others moved to Natal because of its good rainfall and its prospective harbour in the hope that they get land from the Zulu King Dingaan (Thompson, 2002).

Today Afrikaners are spread inland in places such as Tshwane (formerly Pretoria), Limpopo (formerly Northern Transvaal), Mpumalanga (Eastern Transvaal) (Hall *et al*, 2002).

Afrikaners are often regarded as a homogeneous white population of Dutch origin (Greeff, 2007). This could be attributed to their geographical isolation, which is further influenced by language difference and religious affiliation (Hall *et al*, 2002). Afrikaner population exhibit high frequency of number of disease genes, due to founder effect (Jenkins, 1990). The arrival of the British settlers did not have any significant impact on the reduction of these disease associated genes since there was little integration with British settlers (Jenkins, 1990).

1.1.1.3 Asian Indian

South African Asian Indians are the descendants of Indian slaves who came to South African during the end of the nineteenth century to work in the sugar cane plantations in Natal (Motala *et al*, 1993). They belong to the two ethnic groups which are Aryans and Dravidians from northern and southern India, respectively (Mistry, 1965). Most of the Indians reside in the Durban area (Omar *et al*, 1994).

In 1845 Natal was annexed by the British colonisers (Terreblanche, 2002). The British commercial farmers in Natal owned sugar plantations. They experienced shortage of cheap labour (Saunders and Bundy, 1992). At first they had expected the Zulu people to work in the sugar plantations, but it wasn't the case (Saunders and Bundy, 1992). Unlike in the Eastern Cape, in Natal the Africans had easy access to land which made it possible for nearly all Africans to become peasant farmers (Terreblanche, 2002). To combat this problem of shortage of labour they decided to import indentured Indians labours to work on the sugar plantations (Terreblanche, 2002).

Naturally Indians are homogenous, as there is small amount of gene flow from African has been found in them (Soodyall and Jenkins, 1992). This phenomenon is likely to be as a result of strict religious practice which usually did not permit marriage outside the Asian Indian ethnic group (Warnich *et al*, 2011).

1.1.1.4 Coloured

The South African Coloured (SAC) is an admixed population mostly reside in the Western Cape Province. This self-identified population group begun at Cape of Good Hope area present Cape Town after the Dutch East India Company set up a refreshment station in 1652. History shows that over generations that the KhoiSan, the Bantu speakers, European settlers, and slaves originating from India, Java, Madagascar and Mozambique have contributed to the emergence of this group (Patterson *et al*, 2010). Autosomal DNA studies have revealed that the Coloured population is a result of integration of diverse people from Africa, Europe and Asia (Quintana-Murci *et al*, 2010). Most of the Coloured people are Afrikaans speakers.

1.6 Aim of the study

The aim of this study is to further test the level of polymorphism of the University of the Western Cape (UWC) 10 locus set or UWC 10 plex in four South African populations namely; Zulu, Coloured, Afrikaner and Asian Indian. The UWC 10 plex was designed from 45 loci found on the Y chromosome (D'Amato *et al*, 2011). The UWC 10 plex consists of the markers DYS710, DYS518, DYS385a/b, DYS644, DYS612, DYS626, DYS504, DYS447, DYS449, and DYS481. Of the 10 markers, 4 loci namely; DYS612, DYS626, DYS449, DYS518 are rapidly mutating.

The selection of the UWC 10 plex was encouraged by studies of minimal haplotype loci which showed low level of polymorphism and poor resolution in the local populations (Leat *et al*, 2007; D'Amato *et al*, 2008). As a result the search for loci with higher forensic significance was carried out (D'Amato *et al*, 2010). The selection criterion of 10 plex loci was based on high discriminatory capacity, and the amenability to analysis when multiplexed.

The Y –STR system was tested in three local South African populations namely; Xhosa, Asian Indian, and English Caucasian in parallel with other commercial kits. (D'Amato *et al*, 2011). The system showed higher power of discriminatory capacity as compared to other

commercially used kit namely; PowerPlex ® System (*Promega*) and AmpFlSTR ® Yfiler™ (*Applied Biosystems*) at the time of the design of this multiplex. The superiority of the UWC-plex in terms of the resolution as compared to other commercially available kit is attributed to the inclusion of RM Y-STRs which yield high-resolution paternal lineage differentiation (Ballantyne *et al*, 2010; 2012; 2014).

It is important to understand the genetic diversity of the UWC 10 plex in various population groups in South Africa. This work forms part of the project which involves the establishment of DNA reference databases for various ethnic population groups in South Africa. The information from these databases will have many applications for example; it will used to fight crime, especially in sexual assault cases, as well as in disease association studies.

Chapter 2 - Materials and Methods

2.1 DNA extraction

DNA extraction was previously done (D'Amato *et al*, 2008; Leat *et al*, 2007). Of the DNA previously extracted; 41 were Zulu, 79 were Asian Indian, 108 were Coloureds and 100 were Afrikaner. More DNA samples were collected from males of the above population groups living in the Western Cape Province. They were collected as buccal swabs. These swabs were stored in envelopes at -20 ° C until the time of use.

The DNA was extracted using salting out technique adapted from (Medrano *et al,* 1990) as indicated in the Appendix. The technique employs high salt concentration to precipitate proteins from the DNA. The cell membrane was disrupted using detergents in order to recover nuclei. The protein was then precipitated from the nuclei using a high salt concentration. The DNA was recovered using ethanol precipitation.

Of the DNA extracted; 54 were Zulu, 14 were Asian Indian, 95 were Coloured and 1 was Afrikaner. A total of 492 samples were analysed altogether (Zulu; 95, Asian Indian; 93, Afrikaner; 101 and Coloured; 203 samples).

2.2 PCR amplification

During the initial study in which the loci for the UWC 10-locus set were selected some of the primers were redesigned in order to reduce the amplicon size (DYS449, DYS481,DYS612) and avoiding tracts homologous to X-chromosome (DYS644, DYS710) to accommodate all markers according to size, colour (DYS504) and Tm in a single multiplex reaction (D'Amato *et al*,2011). The following *Applied Biosystem* fluorescent dyes were used for the Y-STR 10 plex: 6 FAM (DYS710, DYS518), VIC (DYS385a/b, DYS644), NED (DYS612, DYS626, DY504) and PET (DYS481, DYS447, DYS449) as indicated in table 2.1.

The reactions were optimized in a final volume of 10 µl, using 2 ng male DNA, 2 mM deoxynucleoside triphosphates (dNTPs: dATP, dCTPs, dGTP, dTTP) (*Roche*), 4 mg/ml bovine serum albumin (*AEC Amersham*), 5 U/0.5 Super- Therm Gold Taq polymerase (*Southern Cross Biotechnology*) with 10 X reaction buffer (with $MgCl_2$). Primer sequences, the final concentration at which each primer was used have been is indicated in table 2.1.

PCR amplification was conducted using a GeneAmp 2700 thermocycler (*Applied Biosystems*) as follows: 10 min at 94 ºC, (2) 2 cycles: 94 ºC for 30 s, 66 ºC for 1 min, 72 ºC for 1 min, (3) 2 cycles: 65.5 ºC (4) 2 cycles 65 ºC, (5) 2 cycles 64.5 ºC , (6) 2 cycles 64 ºC , (7) 24 cycles 62, ºC (8) final extension 68 ºC for 75 mins and a holding temperature of 4 ºC if the PCR products were to remain in the thermal cycler or if using Arktik thermocycler *(Inqaba Biotech*) ,5 U/0.1 5 Super- Therm Gold Taq polymerase (*Southern Cross Biotechnology*) was used (D'Amato *et al*, 2011) .

Table 2.1

The 10 Y-specific loci of the University of the Western Cape Primer Sequences, dye labels and final concentrations of primers at which each primer was used in the multiplex reaction.

Locus	Primer sequences	PCR conc.(µM)
DYS710	F ACTTTTCTGAATCCTGGACAAGTG	0.3
	R FAM-TTCCTCATACTCTCTCCCTCCC	0.3
DYS518	F FAM-CACAAGTGAAACTGCTTCTCG	0.192
	R CATCTTCAGCTCTTACCATGG	0.192
DYS385ab	F VIC-AGCATGGGTGACAGAGCTA	0.24
	R GCCAATTACATAGTCCTCCTTTC	0.24
DYS644	F GGAAGAAGCTGATTTCAATCTCC	0.145
	R VIC-CAGGAGACTGAGGCAGAAAGTC	0.145
DYS612	F GAAGTTTCACACAGGTTCAGAGG	0.102
	R NED-AAAAAGGGAACTGAGGGAAGG	0.102
DYS626	F GCAAGACCCCATAGCAAAAG	0.228

	R NED-AAGAAGAATTTTGGGACATGTTT	0.228
DYS504	F NED-CTAAGCTGCAAGAAAAAGTCC	0.168
	R GAATCACTTGAACCCAAGATG	0.168
DYS481	F GAATGTGGCTAACGCTGTTC	0.3
	R PET-TCACCAGAAGGTTGCAAGAC	0.3
DYS447	F GGGCTTGCTTTGCGTTATCTCT	0.72
	R PET-GGTCACAGCATGGCTTGGTT	0.72
DYS449	F PET-GAATATTTTCCCTTAACTTGTGTG	0.72
	R CACTCTAGGTTGGACAACAAGAG	0.72

2.3 Detection of PCR products

The separation and detection of the PCR product was performed using a Genetic Analyser 3500 (*Applied Biosystems*) using the G5 matrix filter set to determine the five dyes, 6FAM™(blue), VIC™ (green), NED™ (yellow), PET™ (red), and LIZ™ (orange), available from *Applied Biosystems*.

Samples were prepared for electrophoresis by adding 1 µl PCR product with 9 µl of loading mixture. The loading mixture consists of 0.3 µl LIZ500 size standard (*Applied Biosystems*) and 8.7 µl Hi-di Formaldehyde (*Applied Biosystems*). The solution was denatured at 95ºC for 5 minutes on a GeneAmp thermocycler (*Applied Biosystems*) and then immediately snap – cooled on ice slurry for about 2 -3 minutes. Virtual filter –set G5 was used to analyse the fragments. In house sequenced allelic ladder was used to genotype the samples.

2.4 Analysis

2.4.1 Summary statistics

Statistical parameters such as allelic frequencies and gene diversity values were calculated using the software Genepop software version 4.2.2 (Rousset, 2008).

Relative allele frequencies for each population and the overall population were calculated by dividing the number of occurrence for each allele by the sample population size. The allele frequency of the multicopy locus DYS385a/b was analysed as combination of both alleles. A single primer pair of DYS385a/b locus amplifies more than one region in the genome and hence they are analysed as a haplotype. These regions cannot be differentiated using the any

of the primers and methods described here and hence are analysed as a phenotype. For example, if a particular male DNA sample has two alleles with 12 and 13 repeats for the DYS385a/b it is genotyped as either 12-13 or 13-12 as it is not known which region has respective length variation as the primers target two regions of the chromosome. Therefore that particular sample is analysed as phenotype and is given allele designation of 12-13.

Gene diversity, of each locus was computed using the formula $D= (n/n\text{-}1) (1- \Sigma xi_2)$ (Nei, 1987), where *n* is the sample size and xi is the relative allele frequency. Gene diversity among populations occurs if there are differences in allele frequencies between those populations.

Haplotype frequencies were computed using the software Arlequin software version 3.5. 1.3 (Excoffier and Lischer, 2010). Haplotype diversity (HD) was calculated using the same equation as calculating gene diversity using haplotype frequencies instead of allele frequencies.

Discriminatory capacity (DC) was determined by dividing the number of different haplotypes observed in a given population by the total number of sampled individuals in that population.

2.4.2 Genetic relationships between populations

The level of genetic relationship among the populations was tested using descriptive statistics parameters namely; Analysis of molecular variance (AMOVA), Population pair-wise genetic distance (Fst) and Factorial Correspondence Analysis (FCA) were computed.

2.4.2.1 Analysis of molecular variance (AMOVA) using non hierarchic analysis

The analysis of molecular variance (AMOVA) developed for genetic analysis was conducted in order to look into mutational differences between the loci in different populations. A comparison of allele frequencies (as in the case of computing population pair-wise genetic

distance (Fst) test below) does not take into account the structural relationships between haplotypes and thus does not allow one to infer evolutionary relatedness between populations hence AMOVA is used (Roewer *et al*, 2006).

AMOVA estimated the level of genetic differences among the ethnic populations applying conventional Fst statistics with and without haplotype frequencies using distance matrix between haplotypes as implemented in Arlequin version 3.5.1.3 (Excoffier and Lischer, 2010).

Non hierarchical AMOVA analysis was used to evaluate genetic differences (1) among population and (2) within populations (Excoffier and Lischer, 2010) based on ethnicity. Unlike its counterpart hierarchical AMOVA is used for testing the homogeneity among groups composed of different populations. These can be defined according to language geographic location, or any other criteria to be tested.

2.4.2.2 Population pair-wise genetic distance (Fst)

Pair-wise genetic distance is an analogue of the commonly used Fst that measures the evolutionary distance between individual haplotypes .Fst (Wright, 1951) which is derived from population genetic theory describes the present state of population structure, and it is driven by past evolutionary processes such as mutation or natural selection (Holsinger and Weir, 2009; Meirmans and Hedrick, 2011). Population differentiation is also driven by genetic drift and levels of gene flow.

Fst and associated probability values (P values) were calculated using analysis of molecular variance (AMOVA) as found in Arlequin version 3.5.1.3 (Excoffier and Lischer, 2010). Number of permutations were set at 1 000, while the significant level was set at $P<0.05$.

The DYS385a/b marker was excluded in the population comparison because two alleles were not assigned to the individual locus (ie, DYS385a or DYS385b).

2.4.2.3 Factorial Correspondence Analysis (FCA)

The level of genetic relationship was further tested using Factorial Correspondence Analysis (FCA) as implemented in Genetix version. 4.05.2 (Belkhir, 1999). FCA like methods discussed before is also used to infer population genetic structure from the genetic data. It is based on establishing the relationship between the different populations by employing axes, called principal components in 3 dimensions, using discrete variables (loci) which could have different states (alleles) (Jombart *et al*, 2009).

The method summarizes multivariate genetic information into a few synthetic variables whilst at the same time describing as much of the variability between individuals or populations as possible (Jombart *et al*, 2009; Jombart *et al*, 2010).The analysed populations or individuals appear as dots in a hyperspace. The distance between the individuals or populations along the factorial axis gives are indication of the degree of the genetic differentiation. The FCA method has an added advantage as they are computational efficient and can be clearly applied to huge data set(Patterson *et al*, 2006) unlike Bayesian clustering and can clearly differentiate between even relatively "similar" subpopulations and admixed groups (Odong *et al*, 2013).

Chapter 3 - Results

3.1 Summary statistics

3.1.1 Allele and haplotype frequencies

Table 3.1

Allele and haplotype frequencies of 10 Y-STR loci among Zulu males (*n*=95).

Allele	DYS710	DYS518	DYS644	DYS612	DYS626	DYS504	DYS447	DYS449	DYS481	Haplotype	DYS385a/b
12						0.137				11_11	0.053
13						0.632				15-15	0.011
14			0.011			0.232				11_16	0.011
15										15-16	0.147
16										16-16	0.063
17										15-17	0.063
18										16-17	0.116
19										17-17	0.074
20										14-18	0.011
20.4										16-18	0.074
21				0.168						17-18	0.032
21.4			0.305							15-19	0.021
22				0.042						16-19	0.042
22.4			0.347							17-19	0.042
23							0.211		0.074	19-19	0.011
23.4			0.126							14-20	0.042
24					0.021		0.032		0.095	15-20	0.053
24.4			0.042							16-20	0.095
25				0.011	0.2		0.316	0.0105	0.252	17-20	0.021
25.4			0.147							14-21	0.011
26				0.021	0.326		0.284		0.273	17-21	0.011
26.4			0.021								
27				0.232	0.189		0.158	0.158	0.232		
28				0.189	0.147			0.389	0.063		
28.2											

29				0.126	0.105		0.074	0.011
29.2								
30				0.137	0.01		0.063	
30.2	0.032							
31	0.011	0.011		0.011			0.063	
31.2								
32	0.095	0.011		0.053			0.116	
32.2	0.053							
33	0.095	0.032		0.011			0.074	
33.2	0.421							
34		0.168						
34.2	0.053							
35	0.021	0.137					0.011	
35.2	0.074							
36	0.116	0.295					0.042	
36.2	0.011							
37	0.011	0.137						
37.2								
38	0.011	0.053						
38.2								
39		0.053						
39.2								
40		0.053						
40.2								
41		0.032						
41.2								
42		0.021						
42.2								
43								

43.3										
44										
Gene diversity	0.7868	0.8452	0.7541	0.8511	0.7926	0.5344	0.7572	0.7975	0.7975	0.8185

*Table shows allele frequencies for each Y-STR locus except DYS385a/b for which genotype frequencies are reported. The genotype frequencies for DYS385a/b were calculated for the combination of two alleles.

Table 3.2

Allele and haplotype frequencies of 10 Y-STR loci among Coloured males (*n*=203).

Allele	DYS710	DYS518	DYS644	DYS612	DYS626	DYS504	DYS447	DYS449	DYS481	Haplotype	DYS385a/b
12			0.005			0.044				10_11	0.005
13			0.015			0.167				11_11	0.01
14			0.089			0.143				12_12	0.015
15			0.128			0.158				11_13	0.03
16			0.207			0.187				12_13	0.015
17			0.241			0.271				13_13	0.005
18			0.015			0.03				08_14	0.005
19			0.01							10_14	0.015
20										11_14	0.207
20.4			0.005							12_14	0.015
21				0.025			0.02		0.054	13_14	0.02
21.4			0.079							14_14	0.02
22				0.01	0.044		0.005		0.281	10_15	0.005
22.4			0.069				0.02			11_15	0.039
23					0.01		0.108		0.148	12_15	0.044
23.4			0.039				0.005			13_15	0.015
24				0.01	0.02		0.153	0.01	0.094	14_15	0.034
24.4			0.064							15_15	0.025

25					0.079	0.35	0.01	0.138	11_16	0.015
25.4			0.015						13_16	0.02
26	0.005			0.01	0.291	0.202	0.025	0.099	14_16	0.02
26.4			0.02						15_16	0.044
27				0.054	0.227	0.094	0.133	0.084	16_16	0.039
28		0.005		0.069	0.167	0.034	0.123	0.069	12_17	0.005
28.2	0.005								13_17	0.02
29	0.005			0.108	0.108	0.01	0.2	0.01	14_17	0.02
29.2	0.02								15_17	0.025
30	0.015	0.01		0.222	0.034		0.167	0.015	16_17	0.059
30.2	0.02								17_17	0.015
31	0.074	0.025		0.212	0.015		0.158		12_18	0.005
31.2	0.049								13_18	0.01
32	0.059	0.044		0.167	0.005		0.089	0.01	14_18	0.01
32.2	0.049								16_18	0.015
33	0.069	0.143		0.099			0.054		17_18	0.03
33.2	0.133								13_19	0.005
34	0.118	0.167		0.015			0.025		14_19	0.015
34.2	0.054								15_19	0.015
35	0.118	0.207					0.005		16_19	0.01
35.2	0.0245								17_19	0.015
36	0.069	0.163					0.005		19_19	0.005
36.2	0.0245								12_20	0.005
37	0.025	0.113							13_20	0.005
37.2	0.02								14_20	0.005
38	0.015	0.069							15_20	0.03
38.2	0.005								16_20	0.01
39	0.005	0.02							15_21	0.005
39.2	0.005								12_22	0.005

40	0.001	0.03									
40.2											
41	0.005	0.005									
41.2											
42											
42.2											
43											
43.3											
44											
Gene diversity	0.929	0.865	0.8611	0.8519	0.8183	0.8194	0.7949	0.8674	0.8509	0.8721	

*Table shows allele frequencies for each Y-STR locus except DYS385a/b for which genotype frequencies are reported. The genotype frequencies for DYS385a/b were calculated for the combination of two alleles.

Table 3.3

Allele and haplotype frequencies of 10 Y-STR loci among Afrikaner males (*n*=101).

Allele	DYS710	DYS518	DYS644	DYS612	DYS626	DYS504	DYS447	DYS449	DYS481	Haplotype	DYS385a/b
12										11_11	0.01
13			0.06			0.05				11_12	0.01
14			0.1			0.178				11_13	0.02
15			0.069			0.1				13_13	0.02
16			0.416			0.149				10_14	0.01
17			0.257			0.475				11_14	0.347
18			0.01			0.05				12_14	0.01
19										13_14	0.079
20										14_14	0.02
20.4										11_15	0.129
21					0.01				0.069	12_15	0.01

21.4								13_15	0.01
22				0.059	0.02		0.436	14_15	0.119
22.4			0.059					15_15	0.03
23				0.02	0.228		0.1	9_16	0.02
23.4			0.02					12_16	0.01
24				0.03	0.178		0.079	13_16	0.01
24.4			0.01					14_16	0.01
25				0.059	0.356		0.119	15_16	0.01
25.4								16_16	0.01
26				0.01	0.267	0.188	0.02	0.08	11_17	0.01
26.4									15_17	0.02
27				0.01	0.2	0.03	0.05	0.08	13_18	0.01
28				0.02	0.139		0.208	0.04	16_18	0.04
28.2	0.01								18_18	0.01
29	0.02			0.089	0.158		0.188		16_19	0.02
29.2	0.01									
30		0.01		0.366	0.03		0.228			
30.2	0.059									
31	0.04			0.248	0.01		0.118			
31.2	0.1									
32	0.05	0.04		0.158	0.01		0.1			
32.2	0.03									
33	0.109	0.218		0.069	0.01		0.04			
33.2	0.089									
34	0.1	0.218		0.23			0.02			
34.2	0.089									
35	0.119	0.208					0.03			
35.2	0.059									
36	0.079									

36.2	0.02									
37	0.01	0.129								
37.2										
38	0.01	0.04								
38.2										
39										
39.2										
40		0.02								
40.2										
41										
41.2										
42										
42.2										
43										
43.3										
44										
Gene diversity	0.9283	0.8358	0.7459	0.7729	0.8438	0.7127	0.7602	0.8483	0.7687	0.792

*Table shows allele frequencies for each Y-STR locus except DYS385a/b for which genotype frequencies are reported. The genotype frequencies for DYS385a/b were calculated for the combination of two alleles.

Table 3.4

Allele and haplotype frequencies of the 10 Y-STR loci among Asian Indian males (*n*=93).

Allele	**DYS710**	**DYS518**	**DYS644**	**DYS612**	**DYS626**	**DYS504**	**DYS447**	**DYS449**	**DYS481**	**Haplotype**	**DYS385a/b**
12										11_13	0.022
13			0.043			0.011				13_13	0.011
14			0.344			0.075				14_13	0.011
15			0.226			0.452				11_14	0.151

16			0.215		0.312			12_14	0.011
17			0.086		0.14			14_14	0.022
18			0.032		0.011			7_15	0.011
19			0.032		0.022			11_15	0.032
20					0.011			13_15	0.022
20.4								14_15	0.022
21					0.011		0.043	15_15	0.011
21.4								7_16	0.011
22					0.032		0.108	9_16	0.011
22.4					0.161			13_16	0.022
23					0.14		0.473	14_16	0.011
23.4					0.043			15_16	0.032
24				0.032	0.161	0.011	0.204	16_16	0.043
24.4			0.022					12_17	0.011
25				0.043	0.172	0.011	0.086	13_17	0.075
25.4								14_17	0.086
26				0.065	0.183	0.054	0.032	15_17	0.1
26.4								16_17	0.032
27			0.022	0.237	0.043	0.1	0.043	17_17	0.011
28			0.086	0.108	0.011	0.032	0.011	13_18	0.054
28.2								14_18	0.043
29			0.151	0.269	0.011	0.086		15_18	0.011
29.2	0.011							16_18	0.011
30	0.032		0.258	0.14		0.14		17_18	0.011
30.2	0.032							13_19	0.043
31	0.054	0.011	0.247	0.086		0.11		14_19	0.011
31.2	0.11							17_19	0.011
32	0.097	0.065	0.172	0.022		0.258		13_20	0.032
32.2	0.086							14_20	0.011

33	0.054	0.118		0.054				0.118		
33.2	0.151									
34	0.043	0.172		0.011				0.032		
34.2	0.054									
35	0.075	0.097						0.043		
35.2	0.022									
36	0.011	0.226						0.011		
36.2	0.075									
37	0.022	0.215								
37.2	0.054									
38		0.032								
38.2	0.011									
A 39		0.054								
39.2										
40		0.011								
40.2										
41										
41.2										
42										
42.2										
43										
43.3										
44										
Gene diversity	0.9308	0.8506	0.781	0.8179	0.8347	0.6807	0.8691	0.8738	0.7183	0.871

*Table shows allele frequencies for each Y-STR locus except DYS385a/b for which genotype frequencies are reported. The genotype frequencies for DYS385a/b were calculated for the combination of two alleles.

Table 3.5

Allele and haplotype frequencies and gene diversity values of the overall population (*n*= 492).

Allele	DYS710	DYS518	DYS644	DYS612	DYS626	DYS504	DYS447	DYS449	DYS481	Haplotype	DYS385a/b
12			0.002			0.045				10_11	0.002
13			0.026			0.203				11_11	0.016
14			0.124			0.154				11_12	0.002
15			0.11			0.171				12_12	0.006
16			0.211			0.167				11_13	0.02
17			0.169			0.236				12_13	0.006
18			0.014			0.024				13_13	0.008
19			0.01				0.004			8_14	0.002
20							0.002			10_14	0.01
20.4			0.002							11_14	0.185
21				0.043	0.002		0.01		0.045	12_14	0.01
21.4			0.091							13_14	0.024
22				0.012	0.03		0.012		0.226	14_14	0.016
22.4			0.108				0.039			7_15	0.002
23					0.008		0.159	0.002	0.185	10_15	0.002
23.4			0.045				0.01			11_15	0.049
24				0.004	0.024		0.136	0.006	0.111	12_15	0.02
24.4			0.041							13_15	0.012
25				0.002	0.091		0.311	0.008	0.146	14_15	0.04
25.4			0.035							15_15	0.052
26	0.002			0.01	0.25		0.211	0.024	0.116	7_16	0.002
26.4			0.012							9_16	0.006
27				0.073	0.215		0.083	0.114	0.104	11_16	0.008
28		0.002		0.085	0.146		0.016	0.175	0.051	12_16	0.002
28.2	0.004									13_16	0.014

29	0.006		0.116	0.148	0.006	0.15	0.006	14_16	0.012
29.2	0.012							15_16	0.055
30	0.012	0.006	0.242	0.049		0.154	0.006	16_16	0.039
30.2	0.032							11_17	0.002
31	0.051	0.014	0.187	0.024		0.122		12_17	0.004
31.2	0.061							13_17	0.022
32	0.071	0.041	0.144	0.008		0.126	0.004	14_17	0.024
32.2	0.053							15_17	0.045
33	0.079	0.132	0.067	0.002		0.067		16_17	0.053
33.2	0.183							17_17	0.022
34	0.077	0.179	0.014			0.02		12_18	0.002
34.2	0.061							13_18	0.016
35	0.091	0.172				0.018		14_18	0.014
35.2	0.041							15_18	0.002
36	0.069	0.191				0.012		16_18	0.03
36.2	0.03							17_18	0.02
37	0.018	0.14						18_18	0.002
37.2	0.018							13_19	0.01
38	0.01	0.053						14_19	0.008
38.2	0.004							15_19	0.01
39	0.002	0.028						16_19	0.016
39.2	0.002							17_19	0.016
40	0.004	0.028						19_19	0.004
40.2								12_20	0.002
41	0.002	0.008						13_20	0.008
41.2	0.002							14_20	0.012
42		0.004						15_20	0.022
42.2								16_20	0.022
43								17-20	0.004

43.3										14_21	0.002
44										15_21	0.002
										17_21	0.002
										12_22	0.002
Gene diversity	0.9203	0.8599	0.8752	0.8546	0.8363	0.8214	0.8075	0.8749	0.8539	0.8713	

*Table shows allele frequencies for each Y-STR locus except DYS385a/b for which genotype frequencies are reported. The genotype frequencies for DYS385a/b were calculated for the combination of two alleles.

3.1.2 Haplotypes shared by more than one male in different population groups

Table 3.6

Haplotypes shared by more than one Zulu male (*n*=95).

Haplotype DYS710, DYS518, DYS644, DYS612, DYS626, DYS504, DYS447, DYS449, DYS481,DYS385a/b	n	Frequency
33.2-34-22.4-28-26-13-25-28-26-(15/15)	2	0.02
36-40-25.4-21-25-14 -23-28- 25-(16/20)	2	0.02
32-37- 23.4-30-29-12-26-36- 25-(11/11)	2	0.02
33-35- 21.4-29-26-13-27-29-26-(15/16)	2	0.02
33.2-36-22.4-28-26-13-26-28-28-(16/17)	2	0.02
36-39-25.4-21-25-14-23-28- 25-(16/19)	2	0.02
35.2-38-25.4- 21-25-14-23-27-27(14/20)	2	0.02
33.2-35-22.4-27-26-13-25-28-26-(15/16)	2	0.02
33.2-36-23.4-27-27-13-25-30-27-(16/17)	2	0.02
33.2-36-21.4-30-28-13-26-33-23-(17/19)	3	0.03
33.2-34-22.4-27-26-13-25-28-26-(15/16)	5	0.05

*Haplotypes from duplicated locus DYS385 are presented in parenthesis.

Table 3.7

Haplotypes shared by more than one Coloured male (*n*=203).

Haplotype DYS710, DYS518, DYS644, DYS612, DYS626, DYS504, DYS447, DYS449, DYS481,DYS385a/b	n	Frequency
30.2-36-16-32-26-120-25-29-27-(15/16)	2	0.01
35-37-16-30- 27-17- 25-29- 21-(12/15)	2	0.01
35-33-17-31-26-16-25-30-22-(11/14)	2	0.01
31-34-16-30-27-15-26-27-27-(14/16)	2	0.01
33-34-21.4-31-31-13- 26-31- 25-(17/19)	2	0.01
31.2-33-17-30-28-17-23-32-25-(14/15)	2	0.01
34-37-17-28-26-14-23-34-28-(13/17)	2	0.01
31.2-36-26.4-22-25-14-23-27-25-(14/19)	2	0.01
33.2-35-22.4-27-26-13-25-27-27-(14/17)	2	0.01
35-33-16-30-27-16-24-29-22-(12/15)	2	0.01
33.2-35-17-30-28-17- 23-27- 25-(14/14)	2	0.01
40-38-19-29-27-15-25-27-24-(15/17)	2	0.01
29.2-36-24.4-31-29-16-24-25-260-(12/15)	2	0.01
35-33-17-33-26-17-24-28-22-(11/14)	3	0.015

*Haplotypes from duplicated locus DYS385 are presented in parenthesis.

Table 3. 8

Haplotypes shared by more than one Afrikaner male (*n*=101).

Haplotype DYS710, DYS518, DYS644, DYS612, DYS626, DYS504, DYS447, DYS449, DYS481,DYS385a/b	n	Frequency
32-34-16-33-26-14-25-28-27-(14/15)	2	0.02
34.2-33-16-31-29-18- 25-30- 22-(11/14)	2	0.02
31.2-37-14-31-29-15- 24-31- 24-(11/14)	2	0.02
34.2-34-16-30-26-170-26-28- 25-(11/13)	2	0.02
33-37-13-30-28-14-24-32-23-(11/14)	2	0.02
31.2-33-16-30-29-17-23-30- 27-(14/15)	3	0.03

*Haplotypes from duplicated locus DYS385 are presented in parenthesis.

Table 3.9

Haplotypes shared by more than one Asian Indian male (*n*=93).

Haplotype DYS710, DYS518, DYS644, DYS612, DYS626, DYS504, DYS447, DYS449, DYS481,DYS385a/b	n	Frequency
33.2-33-15-31-28-16- 22.4-32-25-(15/17)	2	0.02
30-33-15-32-27-16-26-27-23-(13/17)	2	0.02
32-39-16-31-29-16-22-32-23-(14/17)	2	0.02
36.2-36-15-32-29-17- 22.4-32-24-(15/17)	2	0.02

*Haplotypes from duplicated locus DYS385 are presented in parenthesis.

3.1.3 Population comparisons

In this section the allele frequency distributions are shown for each population, while the gene diversity is shown for each population group and the overall population.

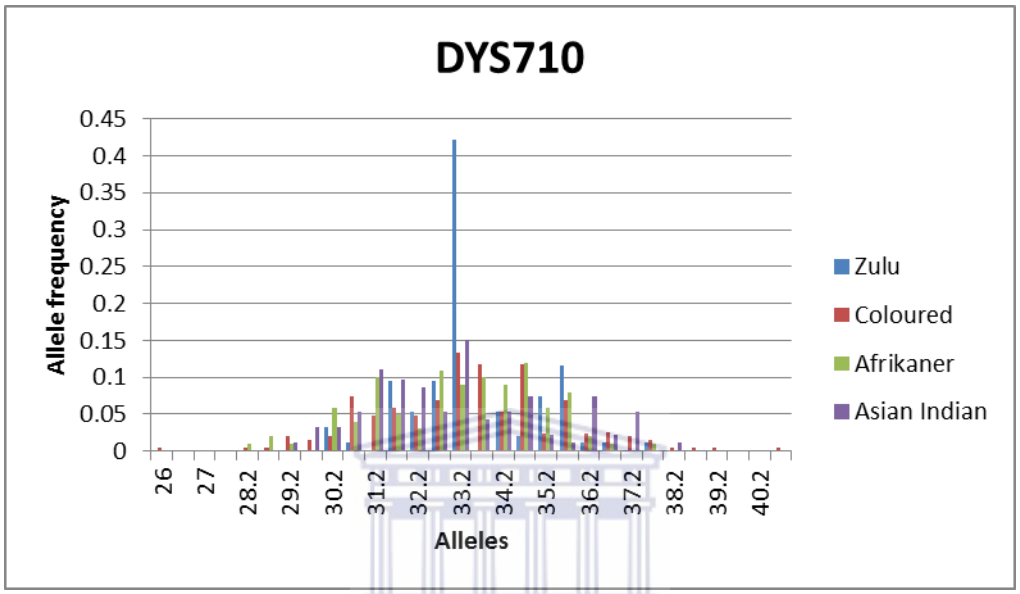

Figure 3.1. Allele frequency distribution of DYS710 within each population group. The observed allele frequency is indicated on the Y-axis, allelic designation (in number of repeats) for each locus is indicated on the X-axis.

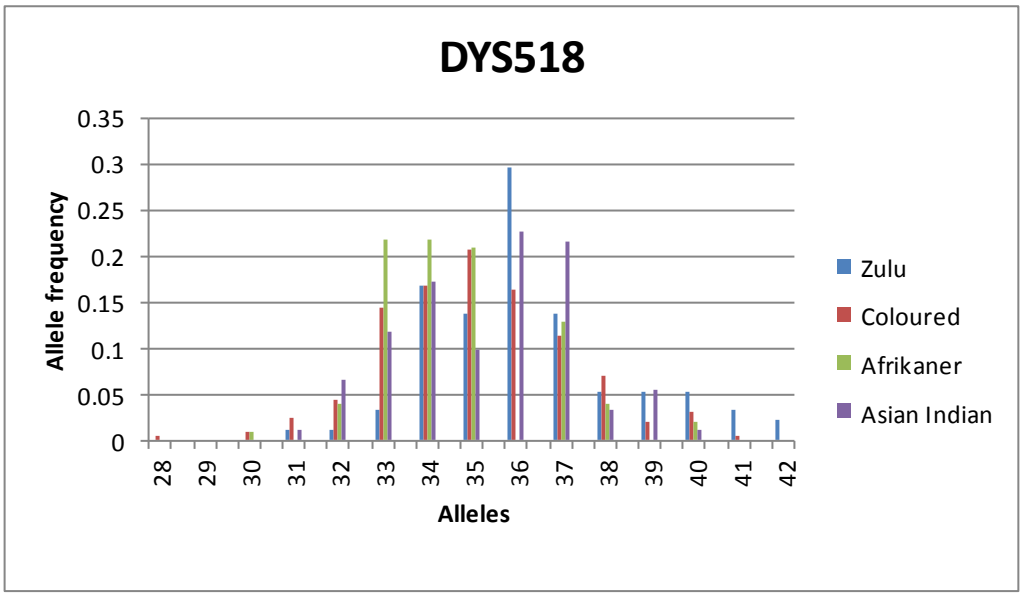

Figure 3.2. Allele frequency distribution of DYS518 within each population group. The observed allele frequency is indicated on the Y-axis, allelic designation (in number of repeats) for each locus is indicated on the X-axis.

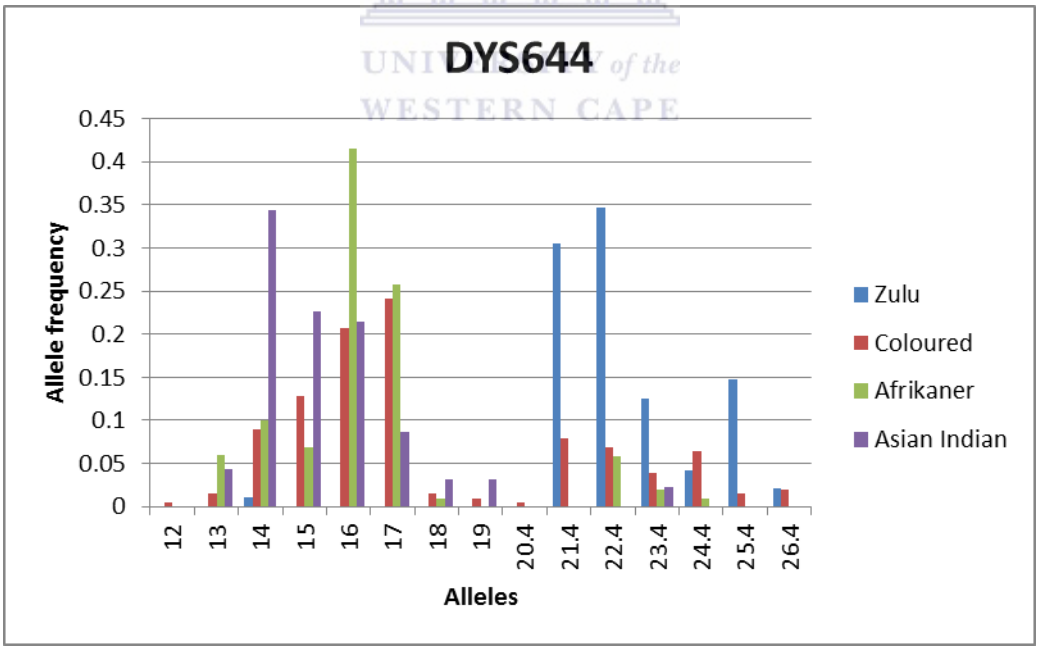

Figure 3.3. Allele frequency distribution of DYS644 within each population group. The observed allele frequency is indicated on the Y-axis, allelic designation (in number of repeats) for each locus is indicated on the X-axis.

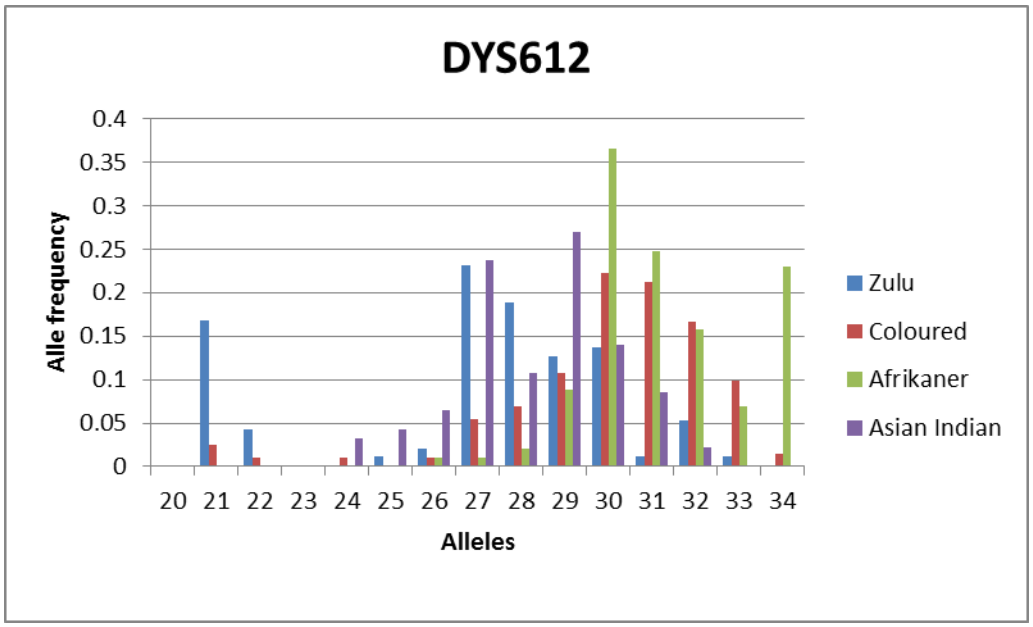

Figure 3.4. Allele frequency distribution of DYS612 within each population group. The observed allele frequency is indicated on the Y-axis, allelic designation (in number of repeats) for each locus is indicated on the X-axis.

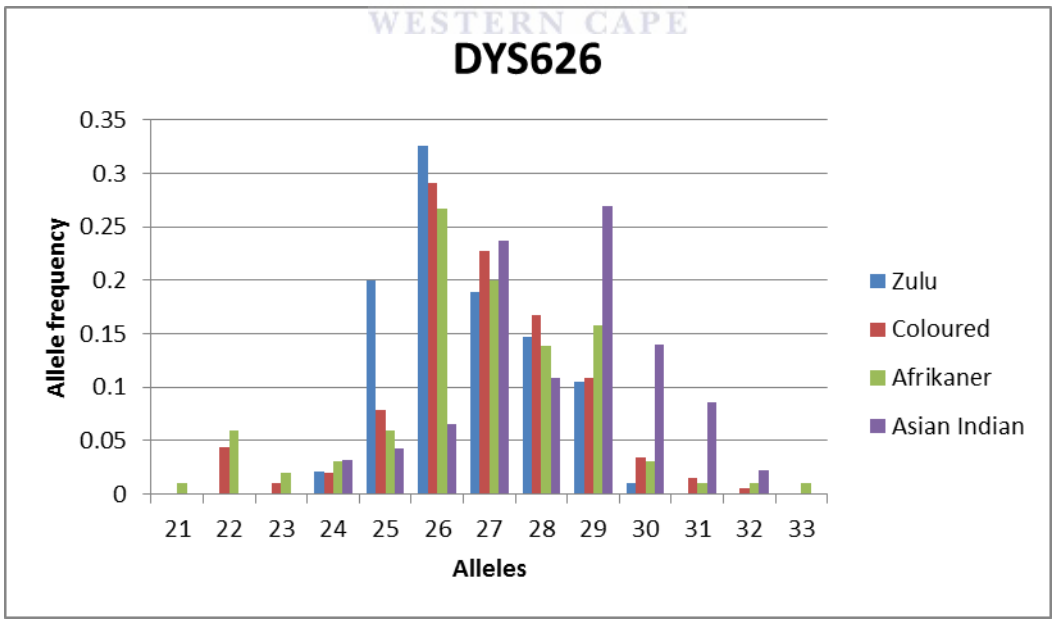

Figure 3.5. Allele frequency distribution of DYS626 within each population group. The observed allele frequency is indicated on the Y-axis, allelic designation (in number of repeats) for each locus is indicated on the X-axis.

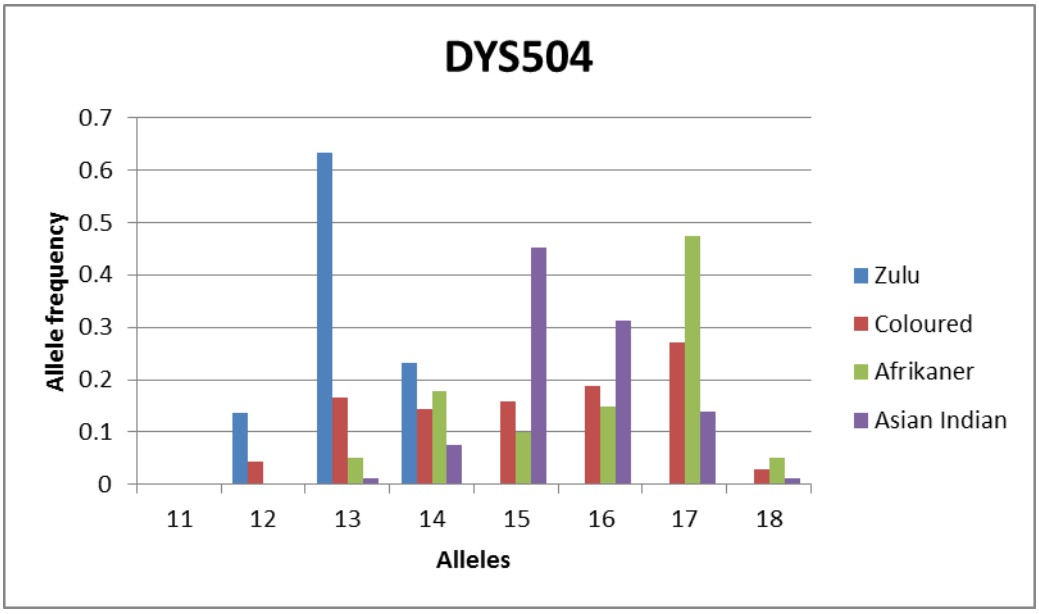

Figure 3.6. Allele frequency distribution of DYS504 within each population group. The observed allele frequency is indicated on the Y-axis, allelic designation (in number of repeats) for each locus is indicated on the X-axis.

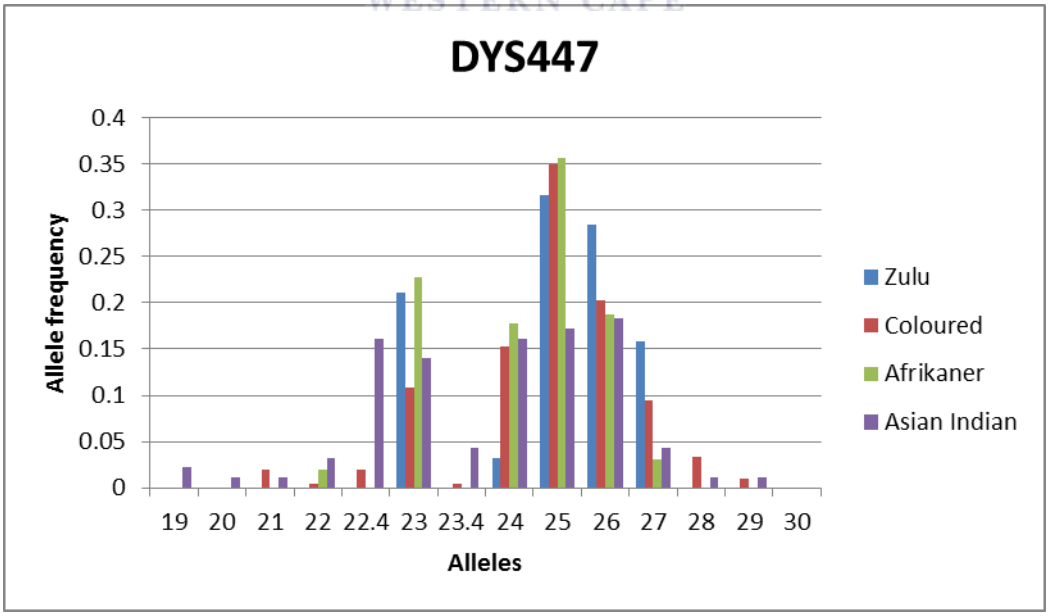

Figure 3.7. Allele frequency distribution of DYS447 within each population group. The observed allele frequency is indicated on the Y-axis, allelic designation (in number of repeats) for each locus is indicated on the X-axis.

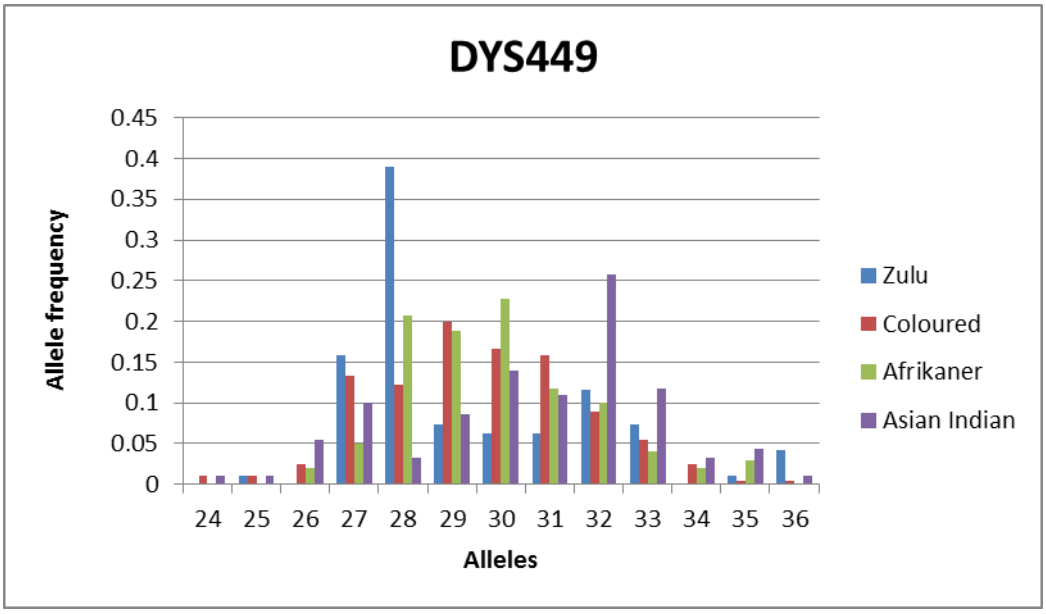

Figure 3.8. Allele frequency distribution of DYS449 within each population group. The observed allele frequency is indicated on the Y-axis, allelic designation (in number of repeats) for each locus is indicated on the X-axis.

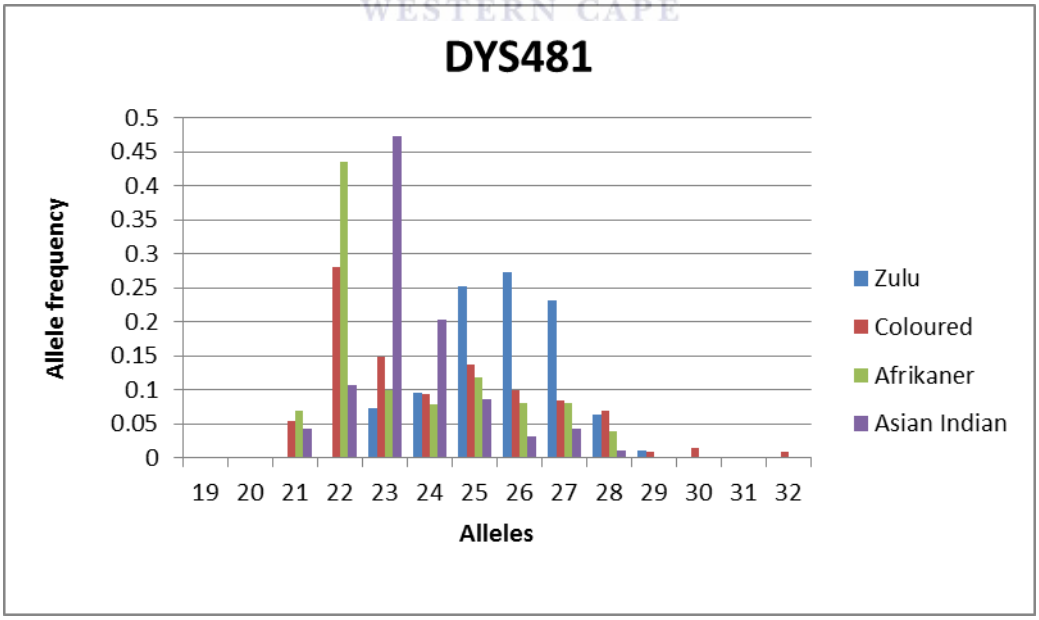

Figure 3.9. Allele frequency distribution of DYS481 within each population group. The observed allele frequency is indicated on the Y-axis, allelic designation (in number of repeats) for each locus is indicated on the X-axis.

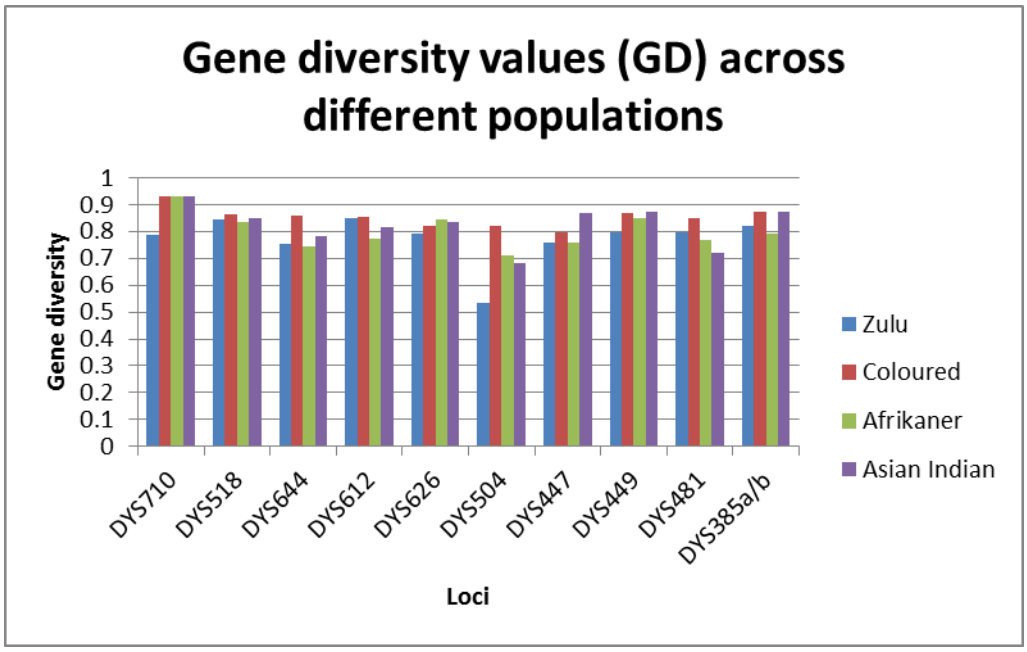

Figure 3.10. Gene diversity values for each 10 Y-STR loci, within each population group. All Y-axes indicate the gene diversity values. All X-axes indicate ten Y-STR loci within each population group.

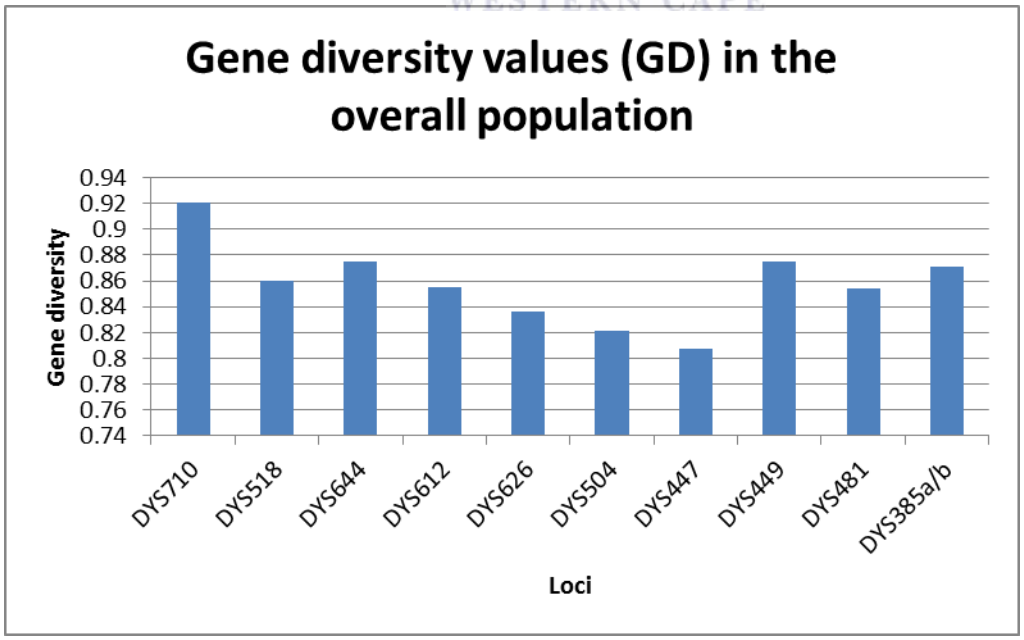

Figure 3.11. Gene diversity values among all 492 male individuals combined. All Y-axes indicate the gene diversity values. All X-axes indicate ten Y-STR within the four South African population groups.

3.1.4 Forensic parameters

Table 3.10

Forensic parameters of the four South African populations. N = sample size, *n* haplotypes = number of different haplotypes in the sample, HD = Haplotype Diversity, DC = Discrimination Capacity.

	Zulu	**Coloured**	**Afrikaner**	**Asian Indian**	**Overall**
N	95	203	101	93	492
n haplotypes	80	187	94	89	448
HD	0.9943	0.999	0.9982	0.9981	0.9981
DC	0.8421	0.9212	0.9307	0.9106	0.9106
Unique haplotypes	86%	92%	94%	96%	92%
Shared haplotypes	14%	8%	6%	4%	8%

Table 3:11

Haplotype sharing information.

Population	Zulu n=95	Coloured n= 203	Afrikaner n=101	Asian Indian n= 93
Zulu	11	none	none	none
Coloured	none	14	none	1
Afrikaner	none	none	6	none
Asian Indian	none	1	none	4

3.2 Genetic relationship between groups

3.2.1 AMOVA results: Non-hierarchical analyses

Table 3.12

AMOVA results computed using haplotype frequencies, based on 9 Y-STR loci (locus DYS385a/b was excluded in the analysis).

Source of variation	**df**	**Sum of squares**	**Variance components**	**Percentage of variation**
Among populations	3	1.884	0.00110 Va	0.22
Within populations	488	243.494	0.49896 Vb	99.78
Total	491	245.378	0.50006	

Fixation index FST: 0.00220

Significance tests (1023 permutations)

Va and FST: P (rand. value > obs. value) = 0.0000

P (rand. Value= obs. Value) = 0.0000

P-value = 0.00000+- 0.00000

Table 3.13

AMOVA results computed using distance matrix, based on 9 Y-STR loci (locus DYS385a/b was excluded in the analysis).

Source of variation	**df**	**Sum of squares**	**Variance components**	**Percentage of variation**
Among populations	3	93.292	0.23390 Va	5.97
Within populations	488	1 798.253	3.68494 Vb	94.03
Total	491	1891.545	3.91885	

Fixation index FST: 0.05969

Significance tests (1023 permutations)

Va and FST: P (rand. value > obs. value) = 0.00000

P (rand. value = obs. value) = 0.00000

P-value = 0.00000+-0.00000

3.2.2 Populations pairwise –genetic distance (Fst)

Table 3.14

Y- STR haplotype pairwise Fst values for four South African populations (locus DYS385 a/b was excluded in the analysis. All pairwise comparisons are significant.

	Zulu	**Coloured**	**Afrikaner**	**Asian Indian**
		(Significance of Fst)		
Zulu	**0.00000**	**-**	**-**	**-**
Coloured	0.00353	0.00000	-	-
Afrikaner	0.00405	0.00118	0.00000	-
Asian Indian	0.00373	0.00080	0.00126	0.00000
		(Fst)		

3.2.3 Factorial Correspondence Analysis (FCA)

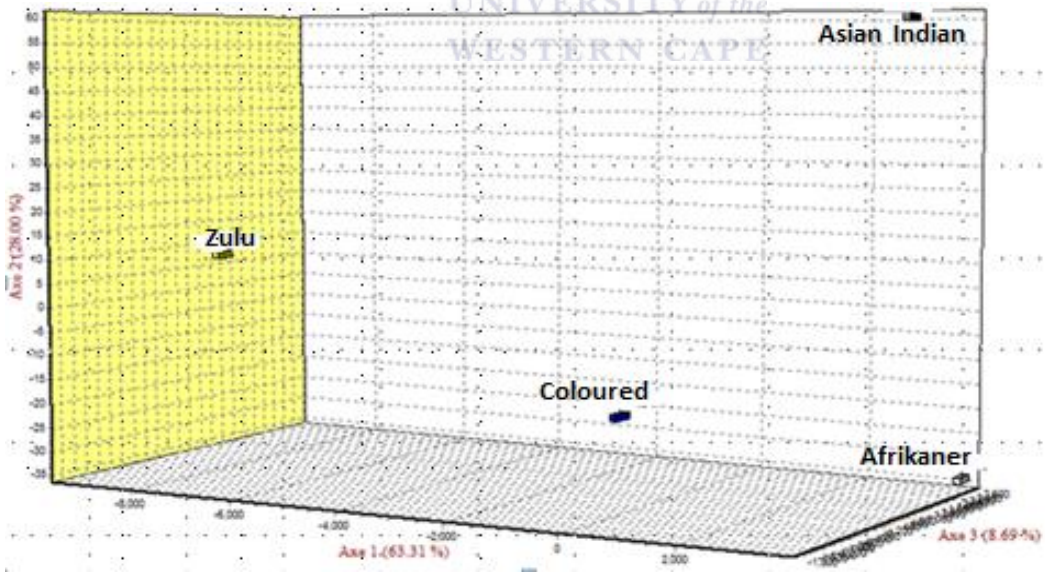

Figure 3.12. Factorial Correspondence Analysis of the UWC 10 locus data set for the different samples in four South African populations.`

Chapter 4 - Discussion

4.1 Population parameters; allele and haplotype frequencies

The allelic frequency defines the frequency of a given variant at a locus where more than one allele has been detected. The allelic frequencies of any particular locus depict the gene diversity value of such loci in different populations. Loci with high gene diversity are highly polymorphic. Therefore in forensic testing allelic frequency is used to evaluate the level of polymorphism of the loci.

In contrary the haplotype frequency of Y-STRs is essential in the calculation of match probability in forensic cases. According to Foreman and Evert, 2001 once a match has been found between the DNA profile of a perpetrator, *x* say, and that of a crime scene sample, it is now standard practice to report the weight of the DNA evidence in terms of a match probability. Given that *x* is not the source of the crime stain, this represents the probability that another individual in the relevant population would share the matching profile. The probability of obtaining a match between two distinct and unrelated individuals (PM) provides a measure of the discriminating power of the profiling system.

4.1.1 Allele and haplotype frequencies for Zulu males

Gene diversity values for the Zulu males ranges from 0.5344 for DYS504 to 0.8511 for DYS612 as shown in table 3.1.The gene diversity values are high indicating that all the loci are highly polymorphic; as a result they can be useful in forensic casework in this population.

4.1.2 Allele and haplotype frequencies for Coloured males

Gene diversity values for the Coloured males ranges from 0.7949 for DYS447 to 0.929 for DYS710 as shown in table 3.2. The gene diversity values are high indicating that all the loci are highly polymorphic; as a result they can be used in forensic casework in this population.

4.1.3 Allele and haplotype frequencies for Afrikaner males

Gene diversity values for the Afrikaner males ranges from 0.7127 for DYS504 to 0.9283 for DYS710 as shown in table 3.3. High gene diversity value implies that all the loci are highly polymorphic; hence they can be used in forensic casework in this population.

4.1.4 Allele and haplotype frequencies for Asian Indian males

Gene diversity values for the Asian Indian males ranges from 0.6807 for DYS504 to 0.9308 for DYS710 as shown in table 3.4. High gene diversity value implies that all the loci are highly polymorphic; hence they can be used in forensic casework in this population

4.1.5 Allele and haplotype frequencies for the overall population

According to figure 3.10 the gene diversity values per locus for the overall population range from 0.8075 for DYS447 to 0.9203 for DYS710. The gene diversity values indicate that all the markers are highly polymorphic; therefore they are suitable for forensic case work in South African population. These findings are consistent with the one of the initial study where all the 10 plex loci displayed high level of polymorphism, not only for each population group but for the overall population (D'Amato *et al*, 2011).

The most informative marker is DYS710.This is attributed to its high allelic ranges, hence higher gene diversity value. It was also the most informative marker in each South African population studied and in the overall population. (D'Amato *et al*, 2011).

The UWC 10 plex consists of two unique loci which have intermediate alleles; DYS447 and DYS644. Intermediate alleles of locus DYS447 were most frequently found in Indian Asian population. On the other hand the intermediate alleles of locus DYS644 were more common in Zulu and Coloured populations. They are less common in Afrikaner and Asian Indian groups. Though intermediate alleles occur in low frequencies, their presence could be valuable for better understanding for better understanding of diversity within the Y

chromosome gene pool (Myres *et al*, 2007). The occurrence of intermediate alleles also increase the discrimination power forensic DNA evidence (Serin *et al*, 2011).

4.2 Population comparisons

Figure 3.1 to figure 3.9 are graphical representations of allele frequency distributions among the populations studied for each locus in the four populations.

All the population have high number of allelic ranges nearly in all loci. High allelic ranges results in high gene diversity value which has been observed across all the population groups and in the overall populations as indicated in figure 3.10 and 3.11.

All Y-STR loci show a unimodal distribution with one frequent allele and with less frequent adjacent alleles, differing by one repeat unit from the most frequent allele. Therefore, the distribution of chromosome Y STR alleles appears to be in accordance with the often proposed step-wise mutation model whereby new alleles of a STR-locus descend from the ancestral allele by deletions or expansions of one repeat unit. (de Knijff *et al*, 1997).

Similar unimodal distribution was found in other studies (de Knijff *et al*, 1997; Kayser *et al*, 2001). Across the regions, the most frequent allele differs for nearly all loci. For example DYS710, allele 33.2 is the most frequent allele in all the four populations except in Afrikaner. In Zulu allele 33.2 is found in 42 % of the population, in Coloured allele 33.2 is found in 13 % of the population, and in Asian Indian it in 15 % of the population .Allele 35 is most frequently allele in Afrikaner population found in 12 % of the population.

4.3 Forensic parameters

The levels of genetic diversity for the haplotypes and the Discrimination Capacity per population group and the overall population and other forensic parameters are indicated in table 3.10. As expected, the most diverse population is Coloured (0.999), followed by Afrikaner (0.9982), Asian Indian (0.9981) and Zulu (0.9943). The highest genetic diversity

of the Coloured population is due to its complex history of admixture. The Coloured came about as a result of enslavement by European settlers at the Cape. According to Patterson *et al,* 2010, the settlers were mostly Dutch, with some French and German. Major ethnic contributions to the Cape slaves were made from East Africa (Mozambique), Madagascar (and surrounding islands), India (a variety of populations including Coromandel, Malabar, Bengal and Ceylon, now Sri Lanka) and substantial numbers from Indonesia (more specifically Java, Batak, Bali and Bugis and Makassar from the Celebes). Together with the Khoisan they contributed to the Coloured gene pool making hence making the population to be the most diverse human population in South Africa.

Haplotype diversity and discrimination capacity for the studied overall population were found to be 0.999128 and 0.95588 respectively, which imply that the 10Y-STR loci studied in the South African population are highly polymorphic. The high level polymorphism has also been found in each population group studied. The results of this study support the forensic case work utility of these loci in South African population groups. Previously the performance in DC of the UWC 10 plex was carried out in comparison with other commercially used Y-STR typing systems on South African populations (D'Amato *et al*., 2011). It was tested on the Xhosa, European Caucasian and Asian Indians. The UWC 10 plex outperformed the other typing system, as it showed the highest DC not only for each population group but also for the overall population.

In comparison to other Y-STR genotyping systems, the discriminatory capacity of 10 plex were higher than those reported for AmpF*l*STR® Yfiler® for in the Xhosa populations in South Africa (D'Amato *et al*, 2011). However the discriminatory capacity of the UWC 10

plex in comparison with AmpF*l*STR® Yfiler® is lower than those reported in Botswana (unpublished data).

The advantage of 10 plex over other Y-STR typing system is that sufficient degree of informativeness can be achieved using less number of loci which is cheaper. UWC 10 plex consists of 10 loci which is a much lesser number of loci as compared to currently used commercially Y-STRs typing system such as Yfiler® Plus Kit which has 27 loci and the PowerPlex Y23 System which has 23 loci.

4.4 Genetic relationships between populations

Genetic relationship between the populations was tested by AMOVA, population pairwise Fst and Factorial Correspondence Analysis.

4.4.1 AMOVA analysis: Non-hierarchical analyses

4.4.1.1 AMOVA computed from haplotype frequencies

According to the AMOVA computed results using haplotype frequencies as indicated in table 3.12 approximately 0.22 % of the total genetic variation between the four populations was due to the variability among populations, with the highest variability of approximately 99.78 % within populations. This is an indication that genetic most variation is contained within populations.

The non significant percentages of variance of (0.22 %) among groups indicate that all the four tested South African population are similar. This similarity is probably as a result of haplotypes that are shared by different populations. However the sharing of haplotypes between different populations occurred only between the Coloured and the Zulu as indicated in table 3.11; which is a sign of gene flow. There is a lot of admixture in between the coloured population and other groups. This is in good according with historical documents.

Genetic studies have also shown that the coloured population group has genetic contributions from different groups such as Xhosa, Khoisan, European and Asians (Jenkin *et al,* 1990: De wet *et al*, 2010) hence shared haplotypes with other population groups.

This result is only true for the coloured populations, as the rest of other population groups had no shared haplotypes among them due to restricted gene flow. This is an indication that AMOVA using haplotype frequencies is not sensitive enough detect levels of differences among the population groups. Considering the high haplotype diversity of the overall population, it can be argued that each sample contains nearly all single, non repeated elements, AMOVA using haplotype frequencies reads extremely high level of variation hence it fails to detect such levels of variation.

The corresponding value of this variation was evaluated by the coancestory coefficient, Fst which measures the amount of differentiation of two or more populations which originate from the common ancestor as a result of genetic drift. As indicated in table 3.12, the Fst (0.00220) is significant indicating that there is genetic differentiation at Y-STRs among the studied South African population groups; hence there is restricted gene flow. The estimated Fst results are accordance with one of the earlier findings, using the markers of the minimal haplotypes on various population groups of South Africa (D' Amato *et al*, 2008). D'Amato *et al*, 2008 used similar populations as the ones in this study; however Xhosa was used instead of Zulu which is the case in this study.

4.4.1.2 AMOVA computed using distance matrix

According to the AMOVA results computed using distance matrix as indicated in table 3.13, approximately 5.97 % of the total genetic variation between the four populations was due to the variability among populations, while 94.03 % of is variability is within populations.

This shows significant differences between populations. The differences among populations detected by AMOVA analysis using distance matrix is also being supported by population pairwise results. Genetic differentiation among South African populations is supported by the limited level of admixture (gene flow) between them. This could be most probably due to differences in social, cultural and ethnical backgrounds. For example Asian Indians practice strict religion which does not permit marriage outside this ethnic group. The Zulu is also genetically different from Asian Indian, despite that they both reside in the Kwa-Zulu Natal Province; there was restricted gene flow between the two populations.

The coancestory coefficient, Fst (0.05969) indicates genetic differentiation. The Fst computed using distance matrix (0.05969) is more sensitive in terms differentiation detection as compared with Fst computed using haplotype frequencies (0.00220). In spite of that they both show genetic differentiation among populations, though in varying magnitude.

From the findings of AMOVA it has been found that AMOVA works differently if use distance matrix. This is due to the fact that when you just use haplotype frequencies, nearly every single element is different from the other, while when you use distance matrix, it picks up elements which are closer to each other, and see whether the distance between elements within a sample is different from the distance between them and any other sample.

4.4.2 Population pair-wise genetic distance (Fst)

The 0.22 % of the total genetic variation among populations found in AMOVA was explained further by pairwise Fst values as shown in table 3.14. All the pairwise comparisons are significant, implying that the populations are different from each other. The highest genetic differentiation is found between Zulu and Afrikaner males, followed by Zulu and Asian Indian males. This could be explained by restricted level of gene flow between the populations.

The observation in population pairwise genetic distance has also been observed in AMOVA using distance matrix, as well as in coancestory coefficient and Fst computed using both kinds AMOVA; haplotype frequencies and distant matrix.

4.4.3 Factorial Correspondence Analysis (FCA)

The scatter plot of FCA of the Y-STR genotypes by population groups in figure 3.12 showed that the coloured population group is in the intermediate position of the FCA. The Zulu is located on the extreme left of the FCA plot, while the Asian Indian is on the extreme top right. The intermediate position of the Coloured population indicates they are admixed. Historical documents have shown that the Coloured population came about as a result from contribution from different population groups such as the KhoiSan (Boonzaaier *et al* ,1996), Bantu, Indian, Malay, Malagasy and Dutch (Peterson *et al*, 2013; Quintana-Murci *et al*, 2010).

The intermediate position of the Coloured population in the scatter plot of FCA has also been observed by D'Amato *et al*, 2008 using the markers of the minimal haplotype. The trend in the scatter plot of FCA results has also been observed in other studies using different kinds of markers such as mtDNA and SNPs respectively (Quintana-Murci *et al*, 2010; Patterson *et al*, 2010).

The Afrikaner population group is close to the Coloured population in the scatter plot of FCA. This is in good accord with historical records because the Afrikaner originates from Dutch settlers who were colonisers at the Cape of Good Hope, now Cape Town (Shell, 1994). The Afrikaner population is also more closely related to the Asian Indian population because of their common indoeuropean ancestry (Chang *et al*, 2015).

Zulus are distant from the rest of other population groups. They are no shared haplotypes between this population group and all the other tested groups. There is only one shared

haplotype between the Coloured and Asian Indian population groups. The shared haplotype is expected between these two population groups because of the admixture nature of the Coloured population group.

FCA is in accord with Fst, AMOVA and historical records.

Chapter 5 - Conclusion

The study showed that the University of the Western Cape 10 locus is a powerful discrimination tool for routine forensic applications and could be used in genealogical investigations as compared to other commercial kits when used on the South African populations (Zulu, Coloured, Afrikaner and Asian Indian). This confirms the previously published results on other South African populations such as the Xhosa and English Caucasian(D'Amato *et al*, 2011). The data from this study will be used to establish a Y-STR reference DNA database hosted at the University of the Western Cape (UWC).

This study is part of the project which is on-going to optimise and validate UWC 10 –plex by using new dyes other than the *Applied Biosystems* ones which were used in the original version of the 10 plex.

Genetic substructure has been found in the studied population groups. The more closely related are Coloured and the all other tested groups due to their admixed nature, having genetic components form all the other tested groups. Asian Indian is more closely related to Afrikaner due to their common indoeuropean ancestry.

Though this study and the one done by (D'Amato *et al*, 2011) reflects virtually the genetic population substructure of South Africa further studies are needed to determine the genetic substructure of other South African Bantu –Speakers such as the Sotho-Tswana, the Venda and the Tsonga. In that manner the data can be used for forensic applications as well as for evolutionary studies. This will enable a clearer understanding of some of the key events in the demography and history of South Africa. It would be of be valuable in future to further analyse these population groups through identifying the Y-haplogroups using SNPs.

References

Abrahams, Z. and Benjeddou, M. (2011) The value of Non-CODIS miniSTR genotyping systems in forensic casework in South Africa .*African Journal of Biotechnology*. 10(86), pp. 19908-19912.

Alves, C., Gusma˜oa, L., Barbosaab, J. and Amorium, A . (2003). Evaluating the informative power of Y-STRs: a comparative study using European and new African haplotype data. *Forensic Science International*, 134(2-3), pp.126–133.

Bajaj, M. and Banerji, M.A. (2004) Type 2 diabetes in South Asians: a pathophysiologic focus on the Asian-Indian epidemic. *Current diabetes reports*, 4(3), pp.213–8.

Ballantyne, K.N., Keel, V., Wollstein, A., Choi, Y., Zuniga, S.B., Ralf, A., Vermeulen, M., de Kniff, P. and Kayser, M. (2012) A new future of forensic Y-chromosome analysis: rapidly mutating Y-STRs for differentiating male relatives and paternal lineages. *Forensic science international. Genetics*.6 (2), pp.208–18.

Ballantyne, K. N., Goedbloed, M., Fang, R., Schaap, O., Lao, O., Wollstein, A,. Choi, Y., van Duijn, K., Vermeulen, M., Brauer, S., Decorte, Poetsch, M., von Wurmb-Schwark, N., de Kniff, P., Labuda, D., Ve´zina, H., Knoblauch, H., Lessig, R., Roewer, R., Ploski, R., Dobosz, T., Henke, L., Furtado, M. and Kayser, M. (2010) Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications. *American journal of human genetics*. 87(3), pp.341–53.

Ballantyne , K.N. *et al*, (2014) Toward Male Individualization with Rapidly MutatingY-Chromosomal Short Tandem Repeats.*Human Mutatio*n. 35(8),pp.1021-1032.

Belkhir, K. (1999). GENETIX, logiciel sous WindowsTM pour la ge´ne´tique despopulations, Laboratoire Ge´nome et Populations, Universite´ de Montpellier II

Berniell-Lee, G., Calafell, F., Bosch,E., Heyer, E., Sica, L., Mouguiama-Daouda, P., van der Veen, L., Hombert, L., Quintana-Murci, L. and Comas, D. (2009) Genetic and demographic implications of the Bantu expansion: insights from human paternal lineages. *Molecular biology and evolution*. 26(7), pp.1581–9.

Boonzaaier, E., Malherbe, C., Smith, A. and Berens, P. (1996) *The Cape Herders: history of the Khoikhoi of Southern Africa*. Cape Town: David Philip Publishers.

Bosch, E., Calafell, F., Santos, F.R., Pe´rez-Lezaun, A., Comas, D., Benchems, N., Tyler-Smith, C. and Bertranpetit, J.(1999) Variation in Short Tandem Repeats Is Deeply Structured by Genetic Background on the Human Y Chromosome. *Am. J. Hum. Genet*. 65, pp.1623–1638

Boucher, M.(1986) The Age of Exploration. In: Cameron T. and Spies, S. (ed) *A New Illustrated history of South Africa*. Johannesburg and Cape Town: Southern Book Publishers (Pty) Ltd and Human and Rousseau, pp. 55–60.

Brenner, C.H. (2010) Fundamental problem of forensic mathematics--the evidential value of a rare haplotype. *Forensic science international. Genetics*. 4(5), pp.281–91.

Brenner, C.H. and Weir, B.S. (2003) Issues and strategies in the DNA identification of World Trade Center victims. *Theoretical Population Biology*. 63(3), pp.173–178.

Brownie, J., Shawcross, S.,Theaker, J., Whitcombe, D., Ferrie, R., Newton, C. and Little, S. **(**1997) The elimination of primer-dimer accumulation in PCR. *Nature reviews. Genetics*. 25(16), pp.3235–3241.

Budowle, B., Bieber, F.R. and Eisenberg, A.J.(2005) Forensic aspects of mass disasters: strategic considerations for DNA-based human identification. *Legal medicine.*7(4), pp.230–43.

Budowle, B., Eisenberg, A.F., van Daal, A. (2009) Validity of Low Copy Number Typing and Applications to Forensic Science. *Croatian Medical Journal*, 50, pp.207–217.

Buel, E., Schwartz , M.B. and LaFountain, M.J.(1998). Capillary Electrophoresis STR Analysis : Comparison to Gel- Based Systems. *Journal*. 43(1), pp.164–170.

Butler, J. (2007) Short tandem repeat typing technologies used in human identity testing. *BioTechniques*. 43(4), p.Sii–Sv.

Butler, J.M. (2011) *ADVANCED TOPICS IN FORENSIC DNA TYPING*, Amsterdam: Elsevier Academic Press.

Butler, J.M., Buel , E., Crivellente, F. and McCord, F. (2004) Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. *Electrophoresis*. 25(10-11), pp.1397–412.

Butler, J.M., Ruitberg, C.M. and Vallone, P.M. (2001) Capillary electrophoresis as a tool for optimization of multiplex PCR reactions. *Fresenius J Anal Chem*. 369, pp.200–205.

Butler, J.M., Shen, Y. and McCord, B. (2003) The Development of Reduced Size STR Amplicons as Tools for Analysis of Degraded DNA. *J Forenisc Sci*. 48(5), pp.1054–1064.

Caenazzo, L., Tozzo, P. and Rodriguez, D. (2013) Ethical issues in DNA identification of human biological material from mass disasters. *Prehospital and disaster medicine*. 28(4), pp.393–6.

California State Auditor Report. (2014) *Sexual Assault Evidence Kits*. (State Auditor: Elaine M. Howle.) California.

Campbell, C.T. (1897) *A History of the Colony of the Cape of Good Hope from its Conquest 1795 to the Settlement of Albany by the British Emigration of 1819* .Cape Town, Port Elizabeth, Johannesburg : J. C. JUTA & CO.

Carruthers, J. (2007) The Early Boer Republics: Changing Political Forces in the “Cradle of Humankind”, 1830s to 1890s. In P. Bonner, A. Esterhuysen, & T. Jenkins, eds. *A search for*

*origins: science, history and South Africa's "Cradle of Humankind."*Johannesburg: Wits University Press, pp. 180–191.

Chang, W., Cathcart, C., Hall, D. and Garrett, A. (2015). Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language*.91(1), pp. 194-244.

Chimera, J.A. and Dyer, D. (1992) DNA ANALYSIS FOR HUMAN IDENTIFICATION USING THE POLYMERASE CHAIN REACTION. *JOURNAL OF INSURANCEMEDIC1NE.* 24 (2), pp.109–113.

Chung, D.T., Dr´abek, J., Opel, *K.L.*, Butler, J.M. and McCord, B.R. (2004) A Study on the Effects of Degradation and Template Concentration on the Amplification Efficiency of the STR Miniplex Primer Sets. *J Forenisc Sci*.49(4), pp.733–740.

Cloete, K., Ehrenreich,L., D'Amato,M.E., Leat,N.,Davison,S. and Benjeddou, M (2010) Analysis of seventeen Y-chromosome STR loci in the Cape Muslim population of South Africa.*Legal Medicine*.12(11), pp.42-45.

D' Amato, E.M., Ehrenreich, L., Benjeddou, M., Davison,S. and Leat , N. (2008). Ancestry and genetic relationships between groups within the Cape Town metropolitan population inferred using Y-STRs genotyping. *Forensic science international. Genetics Supplement Series*. 1, pp.318–319.

D'Amato, M.E., Davison,S. and Benjeddou, M. (2009) Evaluation of 21 Y-STRs for population and forensic studies. *Forensic Science International: Genetics Supplement Series*

D'Amato, M.E., Ehrenreich, L., Cloete, K., Benjeddou, M. and Davison,S. (2010) Characterization of the highly discriminatory loci DYS449, DYS481, DYS518, DYS612, DYS626, DYS644 and DYS710. *Forensic science international. Genetics*.pp.104-110.

D'Amato, M.E., Bajic, V.B.and Davison, S. (2011) Design and validation of a highly discriminatory 10-locus Y-chromosome STR multiplex system. *Forensic science international. Genetics*. 5(2), pp.122–125.

Davison, S., Benjeddou, M. and D'Amato, M.E.(2008) Molecular genetic identification of skeletal remains of apartheid activists in South Africa. *African Journal of Biotechnology* . 7 (25), pp. 4750-4757

Decorte, R. and Cassiman, J. (1993) Forensic medicine and the polymerase chain reaction technique. *J Med Genet*. 30, pp.625–633.

Decorte, R., Cuppens, H. and Marynen, P. (1990) Rapid Detection of Hypervariable Regions by the Polymerase Chain Reaction Technique Hypervariable. *DNA AND CELL BIOLOGY*.9(6). pp.461–469.

Denoon, D. and Nyeko, B. 1987. *Southern Africa since* 1800. 2nd ed., Hong Kong: Longman Group (FE) Ltd.

de Knijff, P., Kayser, M., Caglià, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., Heidorn, F., Herrmann, S., Herzog, B., Hidding, M., Honda, K., Jobling, M., Krawczak, M., Lein, K., Meuser, S., Meyer, E., Oesterreich , W., Pandya, A., Parson, W., Penacino, M.G., Perez Lezaun, A., Piccinini, A., Prinz, M., Schmitt , C., Schneider, P.M., Szibor, R., Teifel., Greding, J., Weichhold, G., de Knijff, P. and Roewer, L. (1997) Chromosome Y microsatellites : population genetic and evolutionary aspects. *Int J Leg Med.*110, pp.134–140

de Knijff, P. (2003)Y-STRS Son , Give Up Your Gun : Presenting Y-STR Results in Court Y-STRS Promega[online].Available from www.dnatestingcentre.com/Reports/y-strs.pdf, pp.3–5. [Accessed 12 March 2016].

De Wit, E., Delport, W., Rugamika, C.E., Meintjes, A., Möller, M., van Helden, P.D., Seoighe, C. and Hoal, H.G. (2010) Genome-wide analysis of the structure of the South African Coloured Population in the Western Cape. *Human genetics*. 128(2), pp.145–53.

Dieringer, D. and Schlötterer, C. (2003) Two Distinct Modes of Microsatellite Mutation Processes : Evidence From the Complete Genomic Sequences of Nine Species. *Genome Research*.13, pp.2242–2251.

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

Du Bruyn, J.(1986) The Gret Trek. In: Cameron T., and Spies, S. (ed). *A New Illustrated history of South Africa*. Johannesburg and Cape Town: Southern Book Publishers (Pty) Ltd and Human and Rousseau, pp. 127–140.

Eckert, K.A.and Kunkel, T.A. (1990) High fidelity DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Nucleic Acids Research*. 18(13), pp.3739–3744.

Edgecombe, R.(1986) The Mfecane or Difaqane. In: Cameron T.and Spies, S. (ed). *A New Illustrated history of South Africa*. Johannesburg and Cape Town: Southern Book Publishers (Pty) Ltd and Human and Rousseau, pp. 115–126.

Edwards, M.C. and Gibbs, R.A. (1994) Manual Multiplex PCR : Advantages , Development , and Applications. *Genome Research*. 3, pp.SP65–S75.

Eldredge, E.A. (1992) Sources of Conflict in Southern Africa , C . 1800-30 : The ' Mfecane ' Reconsidered. *The Journal of African History*. 33(1), pp.1–35.

Ellegren, H. (2004) Microsatellites: simple sequences with complex evolution. *Nature reviews. Genetics*. 5(6), pp.435–45.

Elnifro, E.M., Ashshi, A.M. and Cooper, R.J. (2000) Multiplex PCR : Optimization and Application in Diagnostic Virology. *Clinical Microbilogy Reviews*. 13(4), pp.559–570.

Excoffier, L. and Lischer, H. (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows, *Mol. Ecol. Resour*. 10 564–567.

Fernández, M. E., Goszczynski, D.E., Lirón, J. P., Villegas-Castagnasso, E. E., Carino, M. H., Ripoli, M. V., Rogberg-Muñoz , A., Posik, D. M., Peral-García, P. and Giovambattista G. (2013) Comparison of the effectiveness of microsatellites and SNP panels for genetic identification, traceability and assessment of parentage in an inbred Angus herd. *Genetics and Molecular Biology* . 36 (2), pp 185-191.

Foreman, L. and Evett, I. (2001) Statistical analyses to support forensic interpretation for a new ten-locus STR profiling system. *Int J Leg Med*. 114, pp.147–155.

Foster, E. A., Jobling, M. A., Taylor, P. G., Donnelly P., de Knijff P., Mieremet, R., Zerja, T. and Tyler-Smith, C. (1998) Jefferson fathered slave ' s last child. *Nature*. 396, pp.4–5.

Gu¨ ldemann T. and Stoneking M. (2008) A historical appraisal of clicks: a linguistic and genetic population perspective. *Annu Rev Anthropol* ; 37, pp 93–109.

Gojanović, M.D. and Sutlović, D. (2007) Skeletal Remains from World War II Mass Grave : from Discovery to Identification. *Croat Med J*.48, pp.520–527.

Goodwin, W., Linancre, A. and Hadi, S. (2007) *An Introduction to Forenisc Genetics*, West Sussex: John Wiley and Sons Ltd.

Greeff, J.M. (2007) Deconstructing Jaco: genetic heritage of an Afrikaner. *Annals of human genetics*. 71(5), pp.674–88.

Gusmão, L., Brion, M., Gonzjalez- Neira A., Lareu, M. and Carracedo, A. (1999) Y Chromosome specific polymorphisms in forensic analysis. *Legal Medicine*. 1, pp.55–60.

Hall, A. and Ballantyne, J. (2003) The development of an 18-locus Y-STR system for forensic casework. *Analytical and bioanalytical chemistry*.376(8), pp.1234–46.

Hall, D., Wijsman, E. M., Roos, J. L., Gogos, J. A . and Karayiorgou, M. (2002) Extended Intermarker Linkage Disequilibrium in the Afrikaners Extended Intermarker Linkage Disequilibrium in the Afrikaners. *Genome Res.* 12 , pp.956–961.

Hammer, M.F. & Zegura, S.L., 2002. T HE H UMAN Y C HROMOSOME H APLOGROUP T REE : Nomenclature and Phylogeography of Its Major Divisions. *Annual Review of Anthropology*, 31(1), pp.303–321.

Hammond, H. A., Jin, L., Zhong, Y., Caskey, T. and Chakraborty, R. (1994). Evaluation of 13 Short Tandem Repeat Loci for Use in Personal Identification Applications. *Am. J. Hum. Genet*. 55, pp.175–189.

Hanson, E.K. and Ballantyne, J. (2007) An ultra-high discrimination Y chromosome short tandem repeat multiplex DNA typing system. *PloS one*.2(8), pp.e688.

Hedman, M., Neuvonen, A.M., Sajantila, A. and Palo, J.U. (2011) Dissecting the Finnish male uniformity: the value of additional Y-STR loci. *Forensic science international. Genetics*. 5(3), pp.199–201.

Hedman, M., 2011. *Uniparental DNA markers and Forensic Genetics in Finland*.Thesis, University of Helsinki

Henegariu, O., Heerema, S. R., Vance, G. H., Vogt, P. H. (1997) Multiplex PCR : Critical Parameters and. *BioTechniques*, 23, pp.504–511.

Holsinger, K. E. and Weir, B.S.(2009) Genetics in geographically structured populations: defining, estimating and interpreting F($_{ST}$). *Nat. Rev. Genet*. 10, pp.639–650.

Huffman, T,N. (2007) *Handbook to the iron age: the archaeology of precolonial farming societies in Southern Africa*. Scottsville (South Africa): University of KwaZulu-Natal Press.

Iida, R. and Kishi, K. (2005) Identification, characterization and forensic application of novel Y-STRs. *Legal medicine* .7(4), pp.255–8.

Inagaki, S., Yamamoto, Y., Doi, Y., Takata, T., Ishikawa, T., Imabayashi, K., Yoshitome, K., Miyaishi, S. and Ishizu, H. (2004) A new 39-plex analysis method for SNPs including 15 blood group loci. *Forensic science international*. 144(1), pp.45–57.

Jeffreys, A.J ., Wilson, V., and Thein, S. (1985) Individual-specific “fingerprints” of human DNA. *Nature*. 316(2023), pp.76–79.

Jeffreys, A.J., Wilson, V., Neumann, R. and Keyte, R.(1988) Amplification of human minisatellites by the polymerase chain reaction: towards DNA fingergerprinting of single cells. *Nucleic Acids Research*. 16(23), pp 10953-71.

Jenkins, T., 1990. Medical genetics around the world Medical genetics in South Africa. *J Med Genet*. 27, pp.760–779.

Jobling, M.A. and Tyler-smith, C. (1995) Fathers and sons: the Y chromosome and human evolution. *Elsevier Science Ltd*. 11(11), pp.449–456.

Jobling, M.A., Pandya, A. and Tyler-Smith, C.(1997) The Y chromosome in forensic analysis and paternity testing. *Int J Leg Med*. 110, pp.118–124.

Jobling, M.A. and Tyler-smith, C. (2000) New uses for new haplotypes the human Y chromosome , disease and selection. *TIG*, 16(8), pp.597–601.

Jobling, M. A. and Tyler-Smith, C.(2003.) The human Y chromosome: an evolutionary marker comes of age. *Nature reviews. Genetics*, 4(8), pp.598–612.

Jombart,T., Pontier D. and Dufour, A.B (2009) Genetic markers in the playground of multivariate analysis. *Heredity.* 102, 330–341.

Jombart,T., Devillard,S. and Balloux, F.(2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*.11(94).

Ju, J., Ruan, C., Fuller, C. W., Glazer, A. N. and Mathies, R. A.(1995) Fluorescence energy transfer dye-labeled primers for DNA sequencing and analysis. *Pro. Natl. Acad.Sci*. 92(10), pp.4347–4351.

Karafet, T.M,.Mendez, F. L., Meilerman, B.M.,Underhill, P. A., Zegura, S.L. and Hammer, M.F. (2008) New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Research* , pp.830–838.

Kayser, M., Caglià, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., Heidorn, F., Herrmann, S., Herzog, B., Hidding, M., Honda, K., Jobling, M., Krawczak, M., Lein, K., Meuser, S., Meyer, E., Oesterreich , W., Pandya, A., Parson, W., Penacino, M.G., Perez Lezaun, A., Piccinini, A., Prinz, M., Schmitt , C., Schneider, P.M., Szibor, R., Teifel., Greding, J., Weichhold, G., de Knijff, P. and Roewer, L. (1997) Evaluation of Y-chromosomal STRs : a multicenter study. *Int J Leg Med*. 110, pp.125–133.

Kayser, M., Krawczak, M., Excoffier, L., Dieltjes, P., Corach, D., Pascali, V., Gehrig, C., Bernini, L.F., Jespersen, J., Bakker, E., Roewer, L. and de Knijff, P. (2001) An Extensive Analysis of Y-Chromosomal Microsatellite Haplotypes in Globally Dispersed Human Populations. *Am. J. Hum. Genet*, 68, pp.990–1018.

Kayser, M., Brauer, S., Willuweit, S., Schädlich, H., Batzer , M.A., Zawacki, J., Prinz , M., Stoneking , M.(2002) Online Y-chromosomal Short Tandem Repeat Haplotype Reference Database (YHRD) for U. S. Populations. *Journal of Forensic Sciences*. 47(3), pp.513–519.

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. *Am. J. Hum. Genet*. 74, pp.1183–1197.

Kayser, M .(2007) Uni-parental markers in human identity testing including forensic DNA analysis. *BioTechniques*. 43(6), p.Sxv–Sxxi.

Kayser, M. and de Knijff, P.(2011) Improving human forensics through advances in genetics, genomics and molecular biology. *Nature reviews. Genetics*, 12(3), pp.179–92.

Kayser, M. and Schneider, P.M. (2009) DNA-based prediction of human externally visible characteristics in forensics: motivations, scientific challenges, and ethical considerations. *Forensic science international. Genetics*.3(3), pp.154–61.

Kimpton, C.P., Gill, P., Walton, A., Urquhart, A., Millican, E, S. and Adams (1993.) Automated DNA Profiling Employing Multiplex Amplification Short Tandem Repeat Loci. *Genome Res*. 3, pp.13–22.

Klintschar, M. and Wiegand, P.(2003) Polymerase slippage in relation to the uniformity of tetrameric repeat stretches. *Forensic Science International*.135(2), pp.163–166.

Krenke, B.E., Viculis, L., Richard, M.L., Prinz, M., Milne, S.C., Ladd, C., Gross, A., Gornall, T., Frappier, J.R., Eisenberg, A.J., Barna, C., Aranda, X.G., Adamowicz, M.S. and Budowle, B. (2005) Validation of a male-specific , 12-locus fluorescent short tandem repeat (STR) multiplex *Forenisc Science International*.48, pp.1–14.

Lahr, M.M.and Foley, R.A (1998) Towards a Theory of Modern Human Origins : Geography , Demography, and Diversity in Recent Human Evolution. *YEARBOOK OF PHYSICAL ANTHROPOLOGY*. 41, pp.137–176.

Lane, A.B., Soodyall, H., Arndt, S., Ratshikhopha, M.E., Jonker, E., Freeman , C., Yoing,L., Morar, B. and Toffie, L. (2002) Genetic substructure in South African Bantu-speakers: evidence from autosomal DNA and Y-chromosome studies. *American journal of physical anthropology*.119(2), pp.175–85.

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(1), pp.73–5

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. *Forenisc Science International*.168, pp.154–161.

Levinson, G. and Gutman, G.A. (1987) Slipped-Strand Mispairing : A Major Mechanism for DNA Sequence. *Evolution*. 4(3), pp.203–221.

Lewis-Wiiliams, D. (1986.)The San: Life, Belief and Art. In: Cameron, T. and Spies,S (eds) *A New Illustrated history of South Africa*. Johannesburg and Cape Town: Southern Book Publishers (Pty) Ltd and Human and Rousseau, pp. 31–36.

Lins, A.M., Sprecher , C.J., Puers, C.and Schumm, J.M. (1996) Multiplex Sets for the Amplification of Polymorphic Short Tandem Repeat Loci — Silver Stain and Fluorescence Detection. *BioTechniques*.20(5), pp.882–889.

McDonald, A., Jones, E., Lewis. J. and O'Rourke, P.(2015) Y-STR analysis of digital and/or penile penetration cases with no detected spermatozoa. *Forensic Sci Int Genet*. 15, pp. 84-9.

Maggs, T.(1986.)The Early History of the Black People in Southern Africa. In: Cameron, T.and Spies S. B.(ed). *A New Illustrated history of South Africa*. Johannesburg and Cape Town: Southern Book Publishers (Pty) Ltd and Human and Rousseau, pp. 37–43.

Marjanović, D., Durmić- Pašić1, A., Kovačević1, L., Avdić1, J., Džehverović1, M., Haverić1, S., Ramić1,J., Kalamujić1, B., Bilela1, L.L., Škaro, V., Projić, P., Bajrović, K., Drobnič, K., Dragan, J, D. and Primorac, D. (2009) Identification of Skeletal Remains of Communist Armed Forces Victims During and After World War II: Combined Y-chromosome Short Tandem Repeat (STR) and MiniSTR Approach. *Croatian Medical Journal*.50(3), pp.296–304.

Markoulatos, P., Siafakas, N. and Moncany, M. (2002) Multiplex Polymerase Chain Reaction : A Practical Approach. *Journal of Clinical Laboratory Analysis.* 51, pp.47 51.

Medrano, J.F., Aasen, E. and Sharrow, L. (1990) DNA extraction from nucleated red blood cells. *BioTechniques*. 8 ,43.

Meirmans, P.G. and Hedrick , P.W. (2011) Assessing population structure: F_{ST} and related measures. *Molecular Ecology Resources*.11,5-18.

Mitchell, P. (2002) *The archaeology of Southern Africa.* Cambridge (MA): Cambridge University Press.

Mistry, S.D. (1965) ETHNICS GROUPS OF INDIANS IN SOUTH AFRICA. *S.A. Medical Journal*.39, pp.691–694.

Motala, A.A., Omar, M. A. and Gouws, E.(1993) High risk of progression to NIDDM in South-African Indians with impaired glucose tolerance. *Diabetes*.42(4), pp.556–63. .

Mulero, J.J., Wei Chang, C., Calandro, L.M., Green, R.L., Li, Y., Johnson, C.L. and Hennessy, L.K. (2006) Development and Validation of the AmpF ‘ STR s Yfiler TM PCR Amplification Kit : A Male Specific , Single Amplification 17 Y-STR Multiplex System Ã. *J Forenisc Sci*.51(1), pp.64–75.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. and Erlich, H. (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor Symposium in Quantitative Biology*. 51, pp.263-73.

Myres, M,. Ekins, J.E., Lin, A. A., Cavalli-Sforza, L. L., Woodward, S.R. andUnderhill, PA (2007) Y-chromosome short tandem repeats DYS458.2 non-consensus alleles occur independently in both binary haplogroups J1-M267 and R1b3-M405. *Croat Med J*. 48, pp.450–9.

Nayak, B.P., Khajuria, H. and Gupta S. (2014) Genetic Analysis of Y-chromosomal STRs in Khandayat Population of Odisha, INDIA. *International Research Journal of Biological Sciences*.3(7),pp.26-28.

Nurse, G.T., Weiner, S. and Jenkins, T. (1985) *The Peoples of Southern Africa and Their Affinities* Oxford: Clarendon Press.

Odong, T. L., van Heerwaarden, J., van Hintum, T. J. L., van Eeuwijk, F. A. and Jansen, J. (2013) Improving Hierarchical Clustering of Genotypic Data via Principal Component Analysis. *Crop Science*.53, pp.1546-1554.

Omar, M.A.K., Seedat, M.A., Dyer, R.B., Motala , A.A.,Knight, L.T. and Becker, P.J . (1994) South African Indians Show a High Prevalence of NIDDM and Bimodality in Plasma Glucose Distribution Patterns. *DIABETES CARE*. 17(1), pp.70–73.

van Oven, M., Geystelen, A.V., Kayser, M., Decorte, R. and Larmuseau, M.H. (2013) Seeing the Wood for the Trees: A Minimal Reference Phylogeny for the Human Y Chromosome. *Human Mutation*. 35(2), pp.187–191.

Patterson, N., Petersen, D.C., van der Ross, R.E.,Sudoyo, H., Glashoff, R.H., Marzuki, S., Reich, D. and Hayes, V.M. (2010) Genetic structure of a unique admixed population : implications for medical research. *Human Molecular Genetics*.19(3), pp.411–419.

Peires, J.B. (1986) The Emergence of Black Communities. In: Cameron, T and Spies, S. B. (ed). *A New Illustrated history of South Africa*. Johannesburg and Cape Town: Southern Book Publishers (Pty) Ltd and Human and Rousseau, pp. 44–51.

Petersen, D. C. Libiger, O., Tindall, E.A., Hardie, R., Hannick, L.I., Glashoff, R. H., Mukerji, M., Indian Genome Variation Consortium, Fernandez, P., Haacke, W., Schork, N.J. and Hayes, V.M. (2013) Complex Patterns of Genomic Admixture within Southern Africa. *PLoS Genet*. 9(3) .

Polz, M.F. and Cavanaugh, C.M. (1998) Bias in Template-to-Product Ratios in Multitemplate PCR. *Applied and Environmental Microbiology*. 64(10), pp.3724–3730.

Purps, P., Geppert,M, M. Nagy, M. and Roewer, L. (2015) Validation of a combined autosomal/Y-chromosomal STR approach for analyzing typical biological stains in sexual-assault cases. *Forensic Science International: Genetics* 19, pp 238–242.

Quintana-Murci, L. and Fellous, M. (2001) The human Y chromosome : the biological role of a " functional wasteland ." *Journal of Biomedicine and Biotechnology*. 1, pp.18–24.

Quintana-murci, L., Harmant, C., Quach, H., Balanovsky, O., Zaporozhchenko, V., Bormans, C., van Helden, P.D., Hoal, E.G.and Behar, D.M. (2010) Strong Maternal Khoisan Contribution to the South African Coloured Population : A Case of Gender-Biased Admixture. *The American Journal of Human Genetics*. 86, pp.611–620.

Redd, A.J., Agellona, A.B., Kearneya, V.A., Contrerasa, V.A., Karafeta, T., Park, H., de Knijff, P., John. J.M. and Hamme, M.F. (2002) Forensic value of 14 novel STRs on the human Y chromosome. *Forenisc Science International*. 130, pp.97–111.

Reed, F. A and Tishkoff, S. A.(2006) African human diversity, origins and migrations. *Current opinion in genetics & development*. 16(6), pp.597–605.

Reynolds, R. and Sensabaugh, G. (1991) PERSPECTIVE : ANALYTICAL BIOTECHNOLOGY Analysis of Genetic Markers in Forensic DNA Samples Using the Polymerase Chain Reaction. *Anal. Chem*. 63, pp.2–15.

Roewer. L., Arnemann, J., Spurr, N.K., Grzeschik ,K.H. and Epplen, J.T (1992) Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts. *Hum Genet* .89,pp.389–39.

Roewer, L., Kayser, M., Dieltjes, P., Nagy, M., Bakker, E., Krawczak, M. and de Knijff, P. (1996) Analysis of molecular variance (AMOVA) of Y-chromosome-specific microsatellites in two closely related human populations. *Human Molecular Genetics*. 5(7), pp.1029–1033.

Roewer, L., Kayser, M., de Knijff, P., Anslinger,K., Betze, A., Caglia, A., Corach, D , Furedi, S., Henkei, L., Hidding, M., KaÈrgelk,H.J., Lessigl,R., Nagya, M., Pascalif, V.L., Parsonm, W., Rolfd,B., Schmitt, C., Sziborn,R., Teifel-Gredingo, J.and Krawczak,M. (2000) A new method for the evaluation of matches in non-recombining genomes : application to Y-chromosomal short tandem repeat (STR) haplotypes in European males. *Forenisc Science International*. 114, pp.31–43.

Roewer, L(2009) Y chromosome STR typing in crime casework. *Forensic Sci Med Pathol*.5, pp.77–84.

Rolf, B., Bulander, N. and Wiegand, P.(2011) Forensic Science International : Genetics Insertion- / deletion polymorphisms close to the repeat region of STR loci can cause discordant genotypes with different STR kits. *Forenisc Science International*: *Genetics* .5, pp.339–341.

Rousset, F., 2008. GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux, *Mol. Ecol. Resour*. 8 103–106.

Saiki, R.K., Gelfand, D.H., Stoffel, S.,Scharf, S.J., Higuchi, R., Horn, G,T., Karry, K.B. and Erlich, H.A.(1988)Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase. *Science*. 239(1986), pp.487–491.

Saunders, C. and Bundy, C. (1992) *Illustrated history of South Africa: the real story*, Cape Town: Readers D. Orem, D. E. (2009) *Nursing: concepts of practice*. 4th ed. St. Louis: Mosby-Year.

Schneider, P.M., Bendera,K., Mayrb,W.R., Parsonc, W., Hoste, B., Decortee, R., Cordonnierf, J., Vanekg, D., Morlingh,N., Karjalaineni,M., Carlottij,C.M., Sabatierk,M., Hohoffl, C., Schmitterm, H., Pflugn, W.,Wenzelo, R., Patzeltp,D.,Lessigq,R., Dobrowolskir,P., O'Donnells, G ., Garafanot,L., Doboszu,M., de Knijff, P., Bente Mevagw, Pawlowskix, R., Gusma˜o,L., Maria Conceicao Videz, Antonio Alonso Alonsoa1, Ferna´ndez, O.G., Nicola´s, P.S., Kihlgreena,A.,Ba¨ra,W., Meiera, V.,Teyssiera,A., Coquoza, R., Brandta, C., Germann,U., Gill, P., HallettaJ. and Greenhalgh, M. (2004) STR

analysis of artificially degraded DNA — results of a collaborative European exercise. *Forenisc Science International*. 139, pp.123–134.

Schoske, R., Vallone, P.M., Ruitberg, C.M. and Butler, J.M. (2003) Multiplex PCR design strategy used for the simultaneous amplification of 10 Y chromosome short tandem repeat (STR) loci. *Analytical and bioanalytical chemistry*. 375(3), pp.333–43.

Senge, T., Madea , B., Junge, A., Rothschild , M.A. and Schneider, P.M. (2011). STRs , mini STRs and SNPs – A comparative study for typing degraded DNA. *Legal Medicine*.13(2), pp.68–74.

Serin, A. Alper, B. and Sertdemir, Y. (2011) Haplotype frequencies of 17 Y-chromosomal short tandem repeat loci from the Cukurova region of Turkey. *Croat Med J*.52, pp.703–708.

Shewale, J.G., Shewale, J.G., Nasir, H., Schneida, E., Gross, A., .Budowle, B. and Sinha, K. (2004) Y-Chromosome STR System , Y-PLEX TM 12 , for Forensic Casework : Development and Validation . *J Forenisc Sci*. 49(6), pp.1–13.

Shi, W. Ayub, Q., Vermeulen , M., Shao, R., Zuinga S., van der Gaag,K.,de Kniff, P., Kayser, M., Xue, Y. and Tyler-Smith , C. (2010) A Worldwide Survey of Human Male Demographic History Based on Y-SNP and Y-STR Data from the HGDP – CEPH Populations Research article. *Mol. Bio. Evol*. 27(2). pp.385–393.

Sirugo, G., Hennig, B.J., Adeyemo,A.A, Matimba, A., Newport, M.J., Ibrahim, M.E., Ryckman, K.K.,Tacconelli, A.,Mariani-Costantini, R.,Novelli, G., Himla Soodyall, H., Charles N. Rotimi., C.N., Ramesar, R.S., Tishkoff, S.A . and Williams, S.M. (2008) Genetic studies of African populations : an overview on disease susceptibility and response to vaccines and therapeutics. *Hum Genet*.123, pp.557–598.

Skaletsky, H., Kuroda-Kawaguchi,T., Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G.,Repping, S., Pyntikova, T., Ali,T.J.Bieri, T.,Chinwalla, A.,Delehaunty, A., Kim Delehaunty, K., Du, H., Fewell, G., Fulton, L., Fulton, R.,Graves, T., Hou, S., Latrielle, P., Leonard, S., Mardis, E.,Maupin, R., McPherson, J.,Tracie Miner, T.,William Nash,W., Christine Nguyen,C., Ozersky,P.,Pepin,K. Rock, S., Rohlfing, T., Scott,K., Schultz, B. Strong, C., Tin-Wollam, A.,Yang, S.,Waterston, R.H. Wilson, R.K.,Steve Rozen, S. andPage, D.C (2003)The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*. 423, pp.825–838.

Smith, J.J . (1952) *Theories about the origin of the Afrikaans*. Johannesburg:Witwatersrand University Press.

Sobrino, B., Carracedo, A. and Brio´n, M. (2005) SNPs in forensic genetics : a review on SNP typing methodologies. *Forenisc Science International*. 154, pp.181–194.

Stearns, P.N. (2001)*The Encyclopedia of World History Ancient, Medieval, and Modern* Sixth., Boston: Houghton Mifflin Company.

Steinlechner, M., Berger, B., Niederst~itter, H. and Parson, W. (2002) Rare failures in the amelogenin sex test. *Int J Leg Med*, 116, pp.117–120.

Stewart, C. and Disotell, T.R.(1998) Primate evolution — in and out of Africa. *Current Biology*.8, pp.582–588.

Stringer, C.B. and Andrews, P. (1988). Genetic and fossil evidence for the origin of Modern Humans. *Science*, 239pp.1263-1268.

Stringer, C. (2002). Modern human origins : progress and prospects. *THE ROYAL SOCIETY*, 357, pp.563–579.

Swerdlow, H. et al., 1991. Three DNA Sequencing Methods Using Capillary Gel Electrophoresis and Laser- Induced Fluorescence. *Anal. Chem*, 63(24), pp.2835–2841.

Tamaki, K. and Jeffreys, A.J. (2005) Human tandem repeat sequences in forensic DNA typing. *Legal Medicine*, 7, pp.244–250.

Terreblanche, S., 2002. *A history of inequality in South Africa, 1652-2002* . Scottsville: University of Natal Press.

Thangaraj, K., Reddy, A.G. and Singh, L.(2002.)Is the anlelogenin gene reliable for gender identification in forensic casework and prenatal diagnosis ? *Int J Leg Med*. 116, pp.121–123.

Thompson, L., 2002. *A history of South Africa*, Johannesburg.

Thompson, J.(2012) *PowerPlex® Y23 Developmental Validation* . Promega

STriki-Fendri , S., Sa´nchez-Diz, P., Rey-Gonza´ lez ,D., Ayadi, I., Alfadhli, S.,Ahmed Rebai,A. and Carracedo, A.(2013) Population genetics of 17 Y-STR markers in West Libya (Tripoli region). *Forensic Science International: Genetics.*

Tsukada, K.,Takayanagi, K., Asamura,H. Ota, M. and Fukushima, H. (2002) Multiplex short tandem repeat typing in degraded samples using newly designed primers for the TH01 , TPOX , CSF1PO , and vWA loci. *Legal Medicine*. 4, pp.239–245.

Underhill, P.A. and Kivisild, T. (2007) Use of y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annual review of genetics*.41, pp.539–64.

University Of Cambridge. (2007, May 10). New Research Confirms 'Out Of Africa' Theory Of Human Evolution. *ScienceDaily*. Retrieved March 14, 2016 from www.sciencedaily.com/releases/2007/05/070509161829.h

Urquhart.A., Kimpton, C. and G. (1994)in Short Tandem Repeat sequences- a survey of twelve microsatellite loci for use as forensic identification markers. *Int J Leg Med*. 107, pp.13–20.

Vansina, J. (1995). New Linguistic Evidence and“ the Bantu Expansion.” *The Journal of African History*. 36(2), pp 173–195.

Vieira, T.C., Gigonzac, M.A.D., Silva, D.M., Rodovalho, R.G., Santos, G.S. and da Cruz, A.D. (2014) Y-STR haplotype diversity and population data for Central Brazil: implications for environmental forensics and paternity testing. *Genet. Mol. Res*. 13 (2), pp.3404-3410.

Viguera, E., Canceill, D. and Ehrlich, S.D.(2001) In vitro Replication Slippage by DNA Polymerases from Thermophilic Organisms. *J. Mol. Bio.*.312, pp.323–333.

Wang, C.C., Yan, S. and Li, H. (2010) Surnames and the Y chromosomes. *Commun Contemp Anthropol*. 4,pp 26‑33.

Wang, C.C. Li, H. (2013) Inferring human history in East Asia from Y chromosomes. *Investig Genet*.4:11.

Wang, C. and Li, H.(2015) Evaluating the Y chromosomal STR dating in deep-rooting pedigrees. *Investigative Genetics*, 6(8).

Warnich, L., Drögemöller, B., Pepper M.S., Dandara, C and Galen E.B. Wright, G.E.B (2011) Pharmacogenomic Research in South Africa : Lessons Learned and Future Opportunities in the Rainbow Nation. *Current Pharmacogenomics and Personalized Medicine*. 9, pp.191–207.

Watson, J.D. and Crick, F.H. (1953) THE STRUCTURE OF DNA. *Cold Spring Harb Symp Quant Biol*. 18, pp.123–131.

Westen, A.A., Kraaijenbrink , T.,Clarisse ,L., Grol ,L.J.W.,Willemse, P.,Zuniga , S.B.,Robles de Medina, E.A.,Schouten R.,van der Gaag,K.J., Weiler, N.E.C., Kal, A.J.Kayser , M.,Titia Sijen,T. and de Knijff,P.(2015) Analysis of 36 Y-STR marker units including a concordance study among 2085 Dutch males. *Forensic Science International: Genetic*s.14. pp.174–181

Wiegand, P. and Kleiber, M.(2001.)Less is more--length reduction of STR amplicons using redesigned primers. *International journal of legal medicine*, 114(4-5), pp.285–7.

Willard, H.F.(2003) Tales of the Y chromosome. *Genome Biology*. 423, pp.810–812.

Willuweit, S. and Roewer, L.(2007) Y chromosome haplotype reference database (YHRD): Update. *Forensic science international. Genetics*. 1, pp.83–87.

Willuweit , S. and Roewer, L. (2015) The new Y Chromosome Haplotype Reference Database *Forensic science international. Genetics Supplement Series*. 15, pp.43–48.

Wright, S.(1951) The general structure of populations. *Ann. Eugen*. 15: 323–354.

Xue ,Y., Wang, Q., Long, Q., Ng, B.L., Swerdlow, H., Burton , J., *et al*. (2009) Human Y chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree. *Curr Biol*. 19, 1453–7.

Y Chromosome Consortium (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res*. 2002;12,339–48.

Ziętkiewicz, E., Witt, M., Daca, P., Żebracka-Gala, J., Goniewicz,M, M.,Barbara Jarząb, B. and Witt, M. (2012) Current genetic methodologies in the identification of disaster victims and in forensic analysis. *J Appl Genetics*. 53, pp.41–60.

Electronic supplementary materials

South African Police Services:
www.saps.gov.za/resource.../statistics/crimestats/2015/crime_stats.

Statistics South Africa: www.statssa.gov.za/publications/populationstats.asp).

Ethnologue: http://www.ethnologue.com

Appendix

1. DNA extraction, quantification and working solution

1.1. Extraction of DNA from buccal swabs

For collection of samples buccal swabs were rubbed along cheeks on the inside of the mouth for approximately 30 seconds to collect cheek cells. This can be done by the collector or by the individual being sampled himself. The following salting out DNA extraction protocol was then performed for each sample.

1.1.1**Reagents**

a) Lysis buffer Stock solution

400 mM NaCl	2 M
10 mM Tris-ClH P^H =8	1 M
2 Mm EDTA	0.5 M
1 % SDS	MW=288.4

b) Proteinase K stock is kept at-20 ºC and working solution should be at final concentration 0.1 mg/ml.

c) Alcohol: Ethanol 99% or isopropanol at -20 º C and ethanol 70 % at -20 ºC

d) 5 M NaCl

Preparation of the lysis buffer

All ingredients were added in 80 % of their final volume in distilled water. SDS was added and the solution was left in the oven at 60 ºC until fully dissolved. The volume was transferred to a volumetric flask, and the necessary SABAX water was added to get the desired volume.

Procedure for DNA extraction

1. The surface of the swab was cut with a clean scapel/surgery blade on a clean surface (sterile petri dish or small plastic bag) on top of a tray.

2. 3 µl of proteinase K (20 mg/ml) and 600 µl of lysis buffer were added to either 1.5 or 2 ml tube Eppendorf tube.

3. Little pieces of excised swab were transferred to the Eppendorf tube and vortexed for 30 seconds. The swabs were incubated overnight at 56 ºC overnight.

4. The volume was transferred to a clean tube.

5. In order to recover the lysis solution with biological material which is still trapped between pieces of swab, the experiment was preceded as follows; the end of a 0.5 ml tube was perforated with a needle (21-22 gauge).

6. The perforated tube was placed inside of a 1.5 ml Eppendorf tube and spun for 1 min in a microcentrifudge. The collected volume was added to the previously separated lysis material.

7. Precipitation was done by adding 1/3 volume of 5 M NaCl and the tube(s) was shaken vigorously for 15 seconds.

8. The sample from the above step was centrifuged for 15 minutes at 5000rpm and the supernatant with DNA was transferred to another tube.

9 .**Equal** volume of cold isopropanol was added to the supernatant with DNA and left for 15 minutes at -150 ºC.

10. The DNA was pelleted by centrifugation at 14,000 rpm for 30 minutes.

11. The pellet was washed with 100 µl of 70 % ethanol to remove the salts and centrifuged at 14,000 rpm for 15 minutes.

12. The pellet was dried shortly in Speedy Vac or at 65 º C. Precaution was taken in order to prevent the DNA from becoming too dry since it will hardly dissolve.

13. The DNA was then dissolved in 30 µl SABAX water and stored at -20 ºC.

1.2. **Quantification**

1. 1µl of the extracted DNA from each sample was used for quantification using nanodrop ND 1000 spectrophotometer.

2. The DNA concentration was determined from the nanodrop readings.

3. Working stock dilutions of 2 ng/µl were made for all the DNA samples. Both the original stock and working stock solutions were stored at -20 ºC.