Reconstruction of Major Male and Female Lineages of the Kensington Muslim Community

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

Shafieka Isaacs

September 2010

Supervisor: Prof. Mongi Benjeddou (PhD)

Keywords

Genetic Variation

Mitochondrial DNA (mtDNA)

Single Nucleotide Polymorphisms (SNPs)

Y-Chromosome

Haplogroups

Lineages

Polymerase Chain Reaction (PCR)

Restriction Fragment Length Polymorphism (RFLP)

Electrophoresis

SNaPshot Minisequencing Assay

Genotyping

Maternal Ancestry

Paternal Ancestry

Abstract

Reconstruction of Major Male and Female Lineages of the Kensington Muslim Community

S. Isaacs

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

The earliest Cape Muslims were brought to the Cape from Africa and Asia from 1652 to 1834. They were part of an involuntary migration of slaves, political prisoners and convicts, and they contributed in the ethnic diversity of the present Cape Muslim population. In addition to migrations, this population has grown by combined and continuous multiplex processes of intermarriage, natural increase, conversion and blending with other communities. Previous studies on the Cape Muslim population and local Cape Muslim communities have taken advantage of the huge amount of archival material covering the past three hundred years of Cape Muslim history in South Africa.

The aim of this study, however was to investigate the genetic diversity and origins of one of these local Muslim communities, by taking advantage of the availability of modern molecular genetics tools. A pilot study conducted with Muslim volunteers residing all over the Cape Metropolitan area was used as a point of reference for the Kensington Muslim community study. Samples were investigated for DNA polymorphisms from both maternal (mitochondrial DNA) and paternal (Y-chromosome) lineages. Variations in these two types of DNA are grouped into continent-specific haplogroups or lineages. Six mtDNA and eight Y- chromosome SNP markers were screened using polymerase chain reaction- restriction fragment length polymorphisms (PCR-RFLP).

Maternal lineages of Cape General Muslims were 34% African, 33% Asian and 33% European. The highest frequency for African lineages was observed in Cape Coloured Muslims (47%) and Cape Malay Muslims (45%). Cape Indian Muslims however displayed a higher frequency for Asian lineages (45%), while Cape Other Muslims for European lineages (50%). Paternal lineages indicated that 5% were of African descent, 75% Asian, 13% European and 7% of lineages remained undefined. Asian lineages were the most dominant paternal lineage with the highest frequency observed in Cape Indian Muslims (97%).

In the second part of the study, contributions of major maternal African, Asian and European haplogroups to the Kensington Muslim community gene pool was investigated. This was achieved by examining mtDNA variations found within this community using twelve mtDNA coding region SNPs genotyped by the SNaPshot minisequencing assay. Maternal lineages of Kensington General Muslims indicated that 55% were of African descent, 34% Asian, 1% European and 10% were of Eurasian descent. African lineages were the most frequent lineages of Kensington Coloured Muslims (56%), Kensington Malay Muslims (67%) and Kensington Other Muslims (100%). Kensington Indian Muslims maternal lineages were mainly of Eurasian origin (43%).

The last part of the study investigated the contribution of the major male continent-specific lineages in the Y-chromosome diversity of the Kensington Muslim community by typing thirteen Y-chromosome SNP markers. Paternal lineages of Kensington General Muslims indicated that 4% were of African origin, 51% Asian and 45% European. The highest frequency for Asian lineages was observed in Kensington Malay Muslims (56%), Kensington Indian Muslims (71%) and Kensington Other Muslims

(100%). However, Kensington Coloured Muslims paternal lineages were mainly of European origin (49%).

The data obtained for Cape Coloured Muslims and Cape Malay Muslims and their Kensington Muslim counterpart's maternal and paternal gene pool were in agreement with historical findings concerning the origins of earlier Cape Muslims. Cape Indian Muslims and Kensington Indian Muslims maternal and paternal gene pool however demonstrated a strong accordance to mtDNA and Y-chromosome lineages observed in indigenous Indian populations.

Declaration

I, Shafieka Isaacs, hereby declare that this dissertation is my own work. This thesis is submitted in partial fulfilment of the requirement for the degree of Magister Scientiae in the Department of Biotechnology at the University of the Western Cape. It has not been submitted before for any degree or examination at this or any other University and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Full Name: Shafieka Isaacs

Signed:

Date: September 2010

Acknowledgements

I would like to thank all the donors particularly the Kensington Muslim Community for their participation in the study and for making this research possible. I especially would like to thank Prof. Yusuf Da Costa for making available his publications and his daughter Galiema, for her assistance in arranging collections of biological samples from the Kensington Muslim Community. I would also like to express my gratitude to Mrs M. Oliver and Mr D. Essack for their assistance in the collection of samples.

I thank my supervisor Prof. Mongi Benjeddou for his patience and support throughout the years in the preparation of this dissertation.

Special thanks to Prof. Jasper Rees for making resources available in his laboratory to conduct part of my research experiments.

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

I am truly grateful to my family and friends for their love, patience and encouragement throughout my years of studying and research.

List of Abbreviations

A, G, C, T	Adenine, Guanine, Cytacine, Thymine
BP, KBP, MBP	Base pairs, Kilo base pairs, Million base pairs
COI	Cytochrome Oxidase Subunit
D-LOOP	Displacement loop (control region) of mtDNA
DNA	Deoxyribonucleic Acid
dNTPs	Deoxynucleotide Triphosphates
ddNTPs	Dideoxynucleotide Triphosphates
EXO I	Exonuclease I
HVS-I/HVS II	Hypervariable segment one and two
$MgCl_2$	Magnesium Chloride
MtDNA	Mitochondrial DNA
NP, NT	Nucleotide position, Nucleotide
NEB	New England Biolabs
NRY	Non-Recombining Y-chromosome
PCR	Polymerase Chain Reaction
PNG	Papua New Guinea
RFLP	Restriction Fragment Length Polymorphism
rCRS	Revised Cambridge Reference Sequence
RFU	Relative Fluorescent Units
SAP	Shrimp Alkaline Phosphatase
SNPs	Single Nucleotide Polymorphisms
STRs	Short Tandem Repeats
$tRNA^{Lys}$	Lysyl transfer RNA
YCC	Y Chromosome Consortium

Definitions of Basic Terms used in Thesis

CLADE	A group of species (or, sometimes, individuals) that share a closer common ancestry with one another than with any other such group; a monophyletic assemblage (Avise 2006).
HAPLOTYPE	A sequence type that comprises all identical sequences (Serk 2004).
HAPLOGROUP	A group of haplotypes that share a common ancestor defined by an array of synapomorphic substitutions (Serk 2004).
LINEAGE	Any array of characters/mutations shared by more than one haplotype (Serk 2004).
MONOPHYLETIC	Groups which include all the descendants of the most common ancestor (Avise 2006)
PARAPHYLETIC	An artificial assemblage that includes a common ancestor and some but not all of its evolutionary descendents (Avise 2006)

List of Figures

List of Tables

Table of Contents

Chapter 1- Review of the Literature

Chapter 2 - Reconstruction of Major Maternal and Paternal Lineages of the Cape Muslim population

Chapter 1: Review of the Literature

1.1. Introduction

Genetic variation among the human population have significantly impacted studies of medical genetics, DNA forensics, human evolution and molecular genealogy (Hammer et al. 2001; Hutchinson et al. 2004; Tishkoff and Kidd 2004; Lind et al. 2007). Protein polymorphisms and ABO blood groups were among the first genetic markers initially screened for variations (Jorde et al. 1998; Hammer et al. 2001; Wells et al. 2001; Cavalli-Sforza and Felman 2003). These markers provided valuable insights into population's genetic structure however had limitations, as these systems were few in number, uniform among populations and were occasionally affected by natural selection (Jorde et al. 1998). However, advancements in molecular technologies during the 80's and 90's ushered in a new era allowing genetic variation to be examined on a DNA level. This has allowed scientist for the first time to characterise different aspects of human evolution, migration and genetic divergence among populations (Jorde et al. 1998; Cavalli-Sforza and Felman 2003, Hutchinson et al. 2004). These molecular technologies have also provided modern molecular genetics with the necessary tools to reconstruct unknown genealogies of human populations (Hutchinson et al. 2004).

1.2. Types of Genetic Data

Autosomal DNA, X Chromosomal DNA, Mitochondrial (mtDNA) and Y-chromosome DNA (Y-DNA) represent four classes of DNA that are routinely used to study population genetics (Jorde et al. 1998; Cavalli-Sforza and Felman 2003, Hutchinson et al. 2004). Their differing mode of inheritance has made it possible to either trace recent or ancient genealogies of any population. This provides knowledge into a population's genetic structure indicating founder events, sex-biased gene flow and the ongoing history of a population. However, only two DNA types namely mitochondrial DNA

and the Y-chromosome are extensively used by molecular genetists to reconstruct ancestral genealogies of a population (Zerjal et al. 1997; Underhill 2000; Behar et al. 2007; Lind et al. 2007; Hutchinson et al. 2004). Since only one variant of mtDNA and Y-chromosome DNA can be transmitted in contrast to autosomal and X chromosomal DNA making these loci ideal candidates to determine the origins of a contemporary population's diversity (Jorde et al. 1998; Underhill 2003; Cavalli-Sforza and Feldman 2003; Hutchison et al. 2004).

1.2.1. Mitochondrial DNA (MtDNA)

The human mitochondrial DNA (mtDNA) genome is a closed double stranded circular molecule, 16,569 base pairs (bp) in length (Fig 1.1) (Wallace et al. 1999; Ballard and Whitlock 2004; Parkendorf and Stoneking 2005). The genome is located in the mitochondria, which is involved in both cellular respiration and oxidative phosphorylation. There are hundreds of mitochondria in cells each containing an abundance of mtDNA molecules. This feature alongside a lack of recombination, a high mutation rate and the mtDNA mode of inheritance has made it popular in tracing a population's maternal line (Sunnucks 2000; Budowle et al. 2003; Ballard and Whitlock 2004).

Mother's pass their mitochondrial DNA directly to both male and female children. Males however, are unable to transmit their mitochondria to their offspring. MtDNA is therefore transmitted unaltered to each female and carriers with it information regarding their ancestral maternal line (Wallace et al. 1999; Sunnucks 2000; Lell and Wallace 2000; Hutchinson et al. 2004). This maternal line is grouped into diverse mtDNA haplogroups according to distinct polymorphisms. These polymorphisms are mainly identified by detecting variations occurring within the mtDNA coding region,

and sequencing the two-hypervariable regions in the displacement loop (D-loop) (Wallace et al. 1999; Van Oven and Kayser 2008).

The D-loop is a 1.1kbp segment situated in the control region containing elements responsible for replication and transcription but is unable to code for DNA (Hutchinson et al. 2004). Yet it represents the most variable region of the mtDNA with an estimated mutation rate of 0.075-0.165 $x10^{-6}$ substitution/site/per year according to pedigree studies. The D-loop has been influenced by genetic drift and natural selection and has acquired mutations at an appreciable frequency (Lell and Wallace 2000; Hutchinson et al. 2004; Parkendorf and Stoneking 2005). Nevertheless, it offers limited discriminating powers as haplogroups identified in this region are poorly resolved. Analysis of single nucleotide polymorphisms (SNPs) in the coding region however has resolved this major drawback and subsequently increased mtDNA haplogroup discrimination (Quintáns et al. 2004; Álvarez-Iglesias et al. 2007).

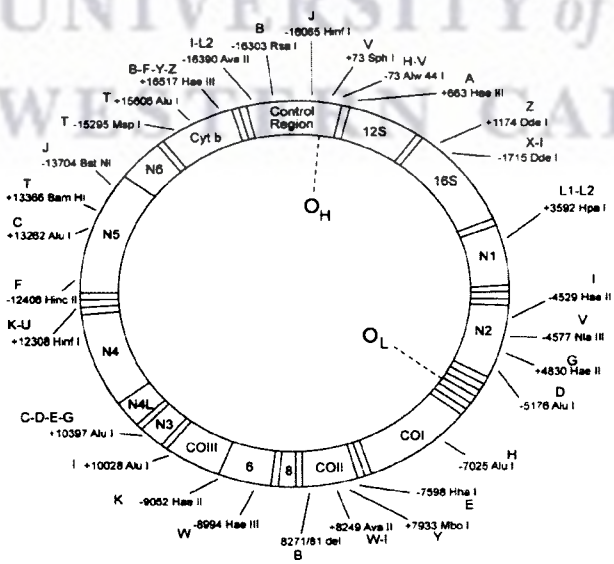


Figure 1.1: The human mitochondrial DNA genome with major mtDNA haplogroups identified in the control and coding region using RFLP analysis (adapted from Rubicz et al. 2007).

1.2.2. Y-Chromosome DNA

The Y-chromosome constitutes the largest non-recombining block of nucleotide (NRY), or haploid (non-paired) DNA in the human genome (Fig 1.2). The region consist of 60 million base pairs (Mbp) of DNA representing 95% of the entire genome (Jobling 2001; Butler 2003; Hammer and Zegura 2002; Rootsi 2004). Only two segments known as the pseudoautosomal regions undergoes recombination with the X chromosome which represents less than 3Mb of the genomes 60Mbp length (Jobling and Tyler Smith 2003; Rootsi 2004).

A large portion of the genome is unaltered by meiotic recombination and therefore preserves the allelic state of the Y-chromosome DNA (Underhill et al. 2001; Hammer and Zegura 2002; Underhill and Kivisild 2007). This will be inherited only by males from their fathers due to the sex-determining role of the Y-chromosome. Thus like its maternal counterpart from the mitochondria, the Y-chromosome DNA may be used to trace the paternal line of a population (Jobling and Tyler Smith 2003; Hammer et al. 2006; Onofri et al. 2006; Kayser et al. 2005).

Single nucleotide polymorphisms (SNPs) and short tandem repeats (STRs) markers are the main types of markers usually used in male genealogy studies. Both of these markers offer insights into the paternal ancestry of a population. However, they reflect different time scales of a population's history. Y-STRs indicate recent events while Y-chromosome SNPs reveal more ancient events (Hutchison et al. 2004; Onofri et al. 2006; Kaiser et al. 2005 ; Karafet et al. 2008).

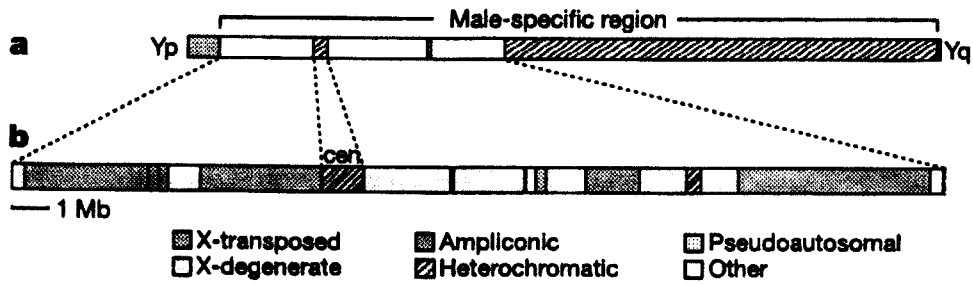


Figure 1.2: The male-specific region of the Y-chromosome. A) Is a representation of the entire Y-chromosome. B) Is an enlarged view of the euchromatic region with various euchromatic sequence classes being shown (adapted from Skaletsky et al. 2003).

1.3. Genetic Genealogy Markers

1.3.1. Short tandem repeats (STR)

Short tandem repeats (STRs) are short DNA sequences consisting of repeating units ranging from 1-6 base pairs (Hammer and Zegura 2002; Hutchinson et al. 2004). The number and sequence of these individual repeat units may vary, and are abundant in its distribution throughout the euchromatic parts of genomes (Hutchinson et al. 2004). Pedigree studies estimate STR's mutation rate at 2×10^{-3} per generation which is higher than SNP's thereby allowing for high levels of allelic variation (Hammer and Zegura 2002; Hutchinson et al. 2004). STRs can therefore have a dozen or more alleles whereas SNPs will only have two allelic states. This feature has therefore made STR's a useful tool for the elucidation of human population history and in forensic analysis (Hammer and Zegura 2002; Hutchinson et al. 2004).

1.3.2. Single Nucleotide Polymorphisms (SNPs)

Single Nucleotide polymorphisms (SNPs) are the simplest and most frequent class of genetic polymorphisms in the human genome (Sobrino et al. 2005). These polymorphisms occur as insertions, deletions or base substitutions at single positions in

the genome (Butler 2003; Budowle 2004). The latter represent the most common form of the polymorphism occurring either between purines (A>G) and pyrimidines (C>T) or between these two groups (A>T) (Budowle 2004; Hutchinson et al. 2004).

Single nucleotide polymorphisms in both the mitochondrial and Y-chromosome DNA provide insight into the history and evolution of their respective lineages. This is mainly attributed to the low mutation rate of SNPs, which enables a low-resolution view of maternal and paternal lineages (Butler 2003). Maternal (MtDNA) and paternal (Y-chromosome) lineages are also highly geographic and population specific, and are therefore routinely employed when examining the genealogical origins of populations (Sanchez et al. 2003; Hutchinson et al. 2004; Hammer et al. 2006).

A number of methods are currently available to genotype mtDNA and Y-chromosome SNPs this includes conventional techniques such as restriction fragment length polymorphism (RFLP) or one of the more automated techniques. The newly developed automated typing technologies allow for the rapid screening of multiple SNPs using multiplex amplification and detection formats. This provides a more cost effective and less time-consuming alternative to conventional techniques (Budowle 2004; Vallone and Butler 2004; Álvarez-Iglesias et al. 2007).

1.4. SNP Typing Methodologies

The growing interest in SNPs and their utilization in a variety of applications such as medical diagnostics, human identity testing and population genetics have lead to the development of a number of new typing technologies and assays (Vallone and Butler 2004; Budowle 2004; Sobrino et al. 2005). These were developed according to allelic discrimination and detection formats. Products of allelic discrimination can be detected

with various methods, which analyze reactions performed with different assay formats (Budowle 2004; Sobrino et al. 2005).

The main technologies used in SNP typing include allele specific hybridization, primer extension, oligonucleotide ligation and invasive cleavage. Each of these technologies can use several detection methods such as electrophoresis fluorescence, fluorescence resonance energy transfer (FRET), luminescence and mass measurement (Budowle 2004; Sobrino et al. 2005).

This review will focus mainly on two of these technologies namely Luminex and SNaPshot due to their simultaneous evaluation of multiple SNP markers (Butler 2003; Vallone and Butler 2004). This review will also cover some of the conventional methods as a number of laboratories still frequently utilize these techniques.

1.4.1. Conventional Typing Methods

1.4.1.1. Restriction Fragment Length Polymorphism (RFLP)

Restriction Fragment Length Polymorphisms (RFLP) analysis identifies variations in the mtDNA and Y-chromosome using restriction endonucleases (Parker et al. 1998; Schurr 2000). The region containing the variation or SNP is first amplified using polymerase chain reaction (PCR) (Navajas and Fenton 2000; Schlotterer 2004). This product is then analysed for the polymorphism using a restriction endonuclease that cleaves at a specific nucleotide sequence known as a recognition site. Since each polymorphism is unique, it will therefore have a unique recognition site. An individual will therefore be assigned to a haplogroup based on the RFLP pattern generated from restriction digestion reactions after agarose gel electrophoresis (Fig1.3.) (Parker et al. 1998; Navajas and Fenton 2000; Schlotterer 2004). A number of polymorphic restriction sites are screened prior to the haplogroup assignment using either low

resolution RFLP (using three or six endonuclease) or high-resolution RFLP (using twelve to fourteen endonuclease enzymes) analysis (Richards and Macaulay 2001; Torroni et al. 1996).

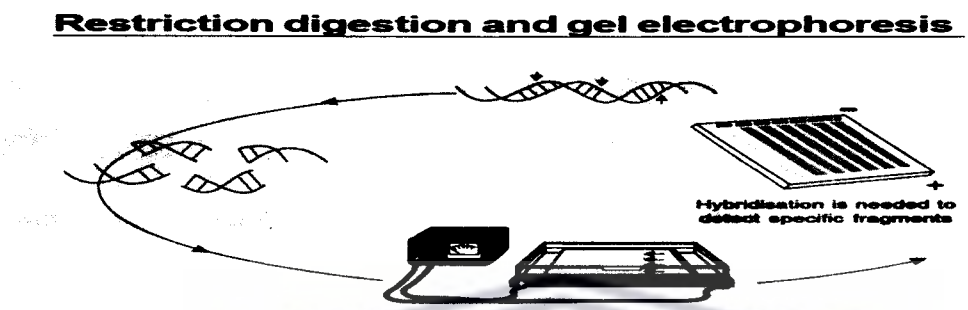


Figure 1.3: An overview of a restriction digestion reaction. The restriction enzyme cleaves the DNA fragment at the recognition site resulting in various size fragments. These fragments are analysed by agarose gel electrophoresis allowing for a haplogroup assignment (adapted from de Vincente and Fulton 2005).

1.4.1.2. Sequencing

While being one of the more expensive and time-consuming techniques, sequencing yet provides complete information regarding the entire region of analysis (Quintás et al. 2004; Schlötterer et al. 2004). This analysis is usually performed using a DNA fragment amplified by PCR. However, a number of processes are involved prior to the actual sequencing reaction. This includes an initial PCR reaction, which undergoes purification and thereafter is amplified using the forward then reverse sequencing primer. This product is then analyzed on a Genetic Analyzer indicating the nucleotide sequence and substitution events contained within the DNA (Parker et al. 1998; Butler 2003; Budowle 2004).

Sequencing technology has been used for the identification of haplogroups in the HV1/HV2 region of the mtDNA genome. This was achieved by comparing sequences obtained from the HV1/HV2 region to the revised Cambridge Reference Sequence

(rCRS). The rCRS was the first mitochondrial genome to be completely sequenced and serves as a reference standard to identify any mtDNA insertions/deletions and SNP polymorphisms (Budowle et al. 2003; Van Oven and Kayser 2008). Sequencing technology has also been used to identify variations in the Y-chromosome (Butler 2003).

This technology has also allowed researchers to identify genes that affect health, cause disease, and genetic variations (Sobrino et al. 2005). However, the technique is not practical when determining a large number of individuals sequence data and therefore more cost –effective and high throughput technologies are being utilized in population studies (Brandstätter et al. 2003; Hutchinson et al. 2004; Vallone et al. 2004).

1.4.2. Automated Typing Technology

1.4.2.1. Luminex Assay

Along with SNaPshot, the Luminex assay allows for the rapid analysis of SNP markers in parallel using multiplex PCR (Butler 2003; Budowle 2004; Vallone and Butler 2004). The assay is based on the Luminex platform using allele-specific hybridization and involves the amplification of the SNP region then the labelling thereof using a fluorescent dye. The fluorescently labelled product is hybridized to allele-specific probes attached to latex beads (Butler 2003; Sobrino et al. 2005). Each bead varies in colour and is assigned an oligonucleotide probe for each possible SNP allele. This enables up to 50 markers to be processed simultaneously as a hundred different bead colours are possible. These beads are then analysed through Flow cytometry using two different lasers, one measuring the fluorescence of the labelled product while the other examines the bead colour passing the detector (Butler 2003; Sobrino et al. 2005). This

information is then arranged into bins based on bead colour, SNP marker, and allele state thereby permitting the assignment of a SNP from the two possible allelic states.

This process takes the Luminex 100 flow cytometer instrument 30 seconds to complete for each individual sample. Therefore, a 96 well reaction plate run will take less than an hour to complete (Butler 2003). The only disadvantage with this system is the use of specialised equipment. This has therefore prompted many laboratories especially forensic laboratories to rather use the SNaPshot assay. Since products from this assay can be detected on a capillary electrophoresis instrument, which is routinely used in forensic laboratories when analyzing short tandem repeats for human identification and lineage based studies (Budowle 2004; Vallone et al. 2004).

1.4.2.2. SNaPshot Assay

The SNaPshot ™ kit from Applied Biosystems is one of the most common commercial SNP genotyping technologies (Sobrino et al. 2005). It is based on the allele specific primer extension (Fig 1.4.) or minisequencing approach (Butler 2003; Vallone and Butler 2004; Sobrino et al. 2005). During the reaction, a primer that anneals to its target DNA immediately adjacent to the SNP is extended with a single nucleotide complementary to the polymorphic site (Butler 2003). Since the reaction is highly specific, it therefore requires the purification of the initial PCR product (region flanking the polymorphic site). This is done to remove any excess primers and dNTP's preventing interference during the primer extension reaction. This procedure is repeated following the extension reaction prior to product analysis.

In the SNaPshot ™ assay, primer extension is facilitated by the use of fluorescently labelled ddNTP's, which is assigned a particular fluorescent dye. The incorporation of

a ddNTP can therefore be detected by electrophoresis and fluorescence using an automated capillary Genetic Analyzer (Butler 2003; Budowle 2004; Sobrino et al. 2005). A number of other technologies are also available for analyzing primer extension products this includes microarrays, matrix assisted laser desorption/ionisation and mass spectrometry (Quintáns et al. 2004).

Primer extension reactions are primarily performed as multiplex reactions and are separated by the addition of tails at the 5 ' end of the extension primers using varying lengths of non-human sequences. Similar to the Luminex assay, 96 well reactions may also be analyzed allowing for a number of SNP markers to be processed (Sobrino et al. 2005). This is especially important when resolving a number of individual's genotypes as in the case of population genealogy studies.

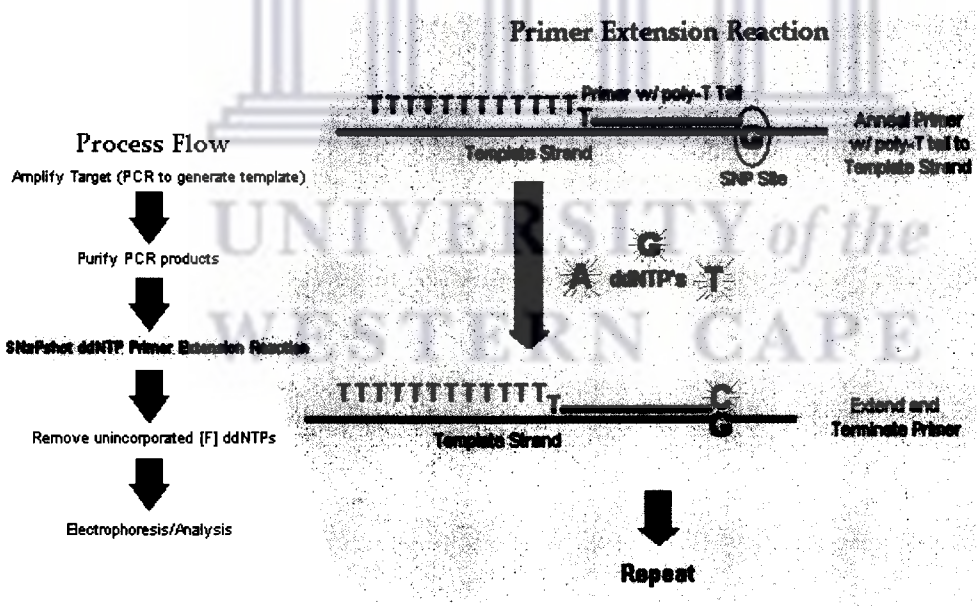


Figure 1.4: Overview of the SNaPshot assay. The five steps of the assay include amplification, purification of PCR product, extension, purification of extension product and detection/analysis. The diagram gives a more detailed description of this process indicating the extension of the target DNA with an extension primer using fluorescently tagged ddNTP's. Once the ddNTP is incorporated, the extension reaction is terminated leaving a specific size product (based on the length of the poly-T tail). This product in addition to the incorporated ddNTP is used to determine the SNP state (adapted from Parsons 2006).

1.5. Genealogy Studies

The ongoing research into genetic variation and the demographical history of populations have provided genealogists with valuable markers in tracing ancestral lineages. MtDNA and Y-chromosome haplogroups are highly informative when reconstructing the prehistory of populations (Cavalli-Sforza and Felman 2003; Underhill and Kivisild 2007; Maji et al. 2009). Moreover, they are defined by unique polymorphisms, which can be represented by phylogenetic trees.

Phylogenetic trees depict the evolutionary history of mtDNA and Y-chromosome haplogroups and indicate the sequential accumulation of mutations along maternal and paternal inherited lineages (Hutchison et al. 2004; Underhill and Kivisild 2007; Van Oven and Kayser 2008). Maternal and paternal lineages can therefore be traced following the mutations indicated by phylogenetic trees. This is usually achieved by screening one or more polymorphisms defining major monophyletic haplogroups. Thereafter additional downstream mutations are screened indicating haplogroups derived from major haplogroups (Hutchison et al 2004; Cox 2006; Athey 2005; Maji et al. 2009).

Over the years, a number of polymorphism have been discovered and have further increased the phylogeographic resolution of both Y-chromosome and mtDNA lineages. Furthermore, has lead to the development of nomenclature systems for both mtDNA and Y-chromosome haplogroups (Hammer and Zegura 2002; Santos et al. 2004; Cox 2006; Underhill and Kivisild 2007; Van Oven and Kayser 2008).

1.5.1. MtDNA Haplogroup Nomenclature

The nomenclature of mtDNA haplogroups is designated according to letters of the alphabet except the letter O (Fig 1.5). The first haplogroups identified were titled A, B, C and D thereafter all other subsequent haplogroups were named after other alphabet letters (Schurr and Wallace 2002; Van Oven and Kayser 2008). A set of rules were later proposed by Torroni et al (1996) and Richards et al (1998) in naming haplogroups to provide a unified nomenclature system (Van Oven and Kayser 2008). This was based on data generated from RFLPs and sequencing of the HV1/HV2 region in the control region.

This system however, is not always strictly followed as some researchers use different nomenclatures to describe haplogroups (Budowle et al. 2003; Van Oven and Kayser 2008). This has been an ongoing issue within the mtDNA community however, no universal nomenclature system exist for mtDNA phylogeny as seen with the Y-chromosome (Butler 2003; Budowle et al. 2003; Van Oven and Kayser 2008). This in part may be attributed to the existing phylogenetic trees available. The most recent attempts to provide a complete phylogenetic tree of global human mtDNA variation have been by Van Oven and Kayser (2008). However, the adaptation of a universal nomenclature is only possible through the consistent use and recognition of one nomenclature system within the mtDNA scientific community (Budowle et al. 2003; Van Oven and Kayser 2008).

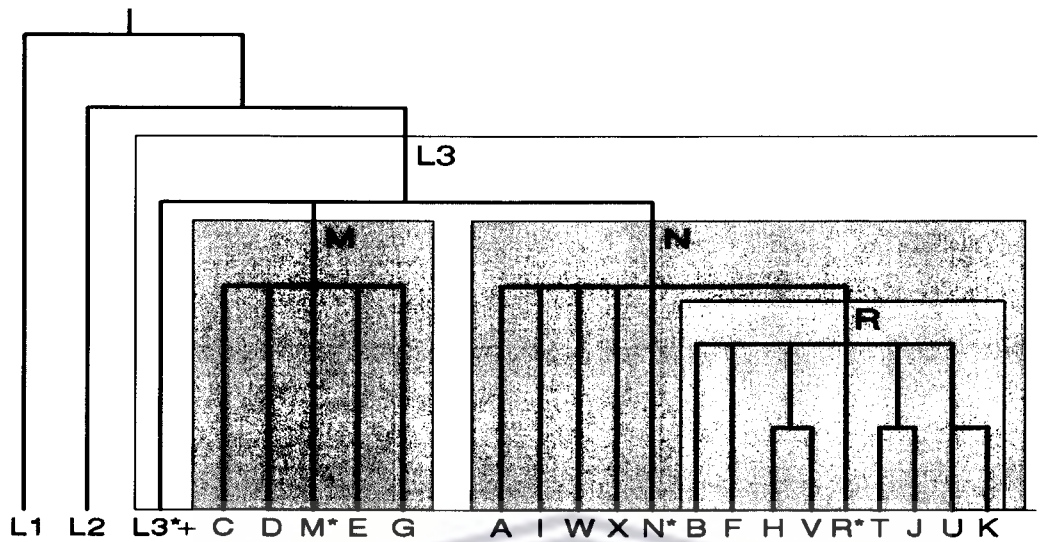


Figure 1.5: A world tree of mtDNA variation indicating the major macrohaplogroups and haplogroups radiating from these groups (adapted from Jobling et al. 2003).

1.5.1.2. MtDNA Haplogroup Variation

The migration of women out of Africa about 150 000 years before present to various continents have resulted in unique base substitutions known as polymorphisms in the mitochondrial genome. These have been shown to be continent specific and are linked to specific mtDNA haplotypes and groups of related haplotypes (haplogroups) (Wallace et al. 1999; Budowle et al. 2003; Maji et al. 2009). Initial studies using RFLP and later sequencing analysis indicated that mtDNA haplogroups were mainly derived from three macrohaplogroups L, M and N. These branch off into other haplogroups, which can further be subdivided into smaller branches or subgroups (Fig 1.6). In addition, macrohaplogroups distribution extends into various regions of the globe displaying characteristic patterns that are specific and unique to a region and population (Mishmar et al. 2003; Underhill and Kivisild 2007; Maji et al. 2009).

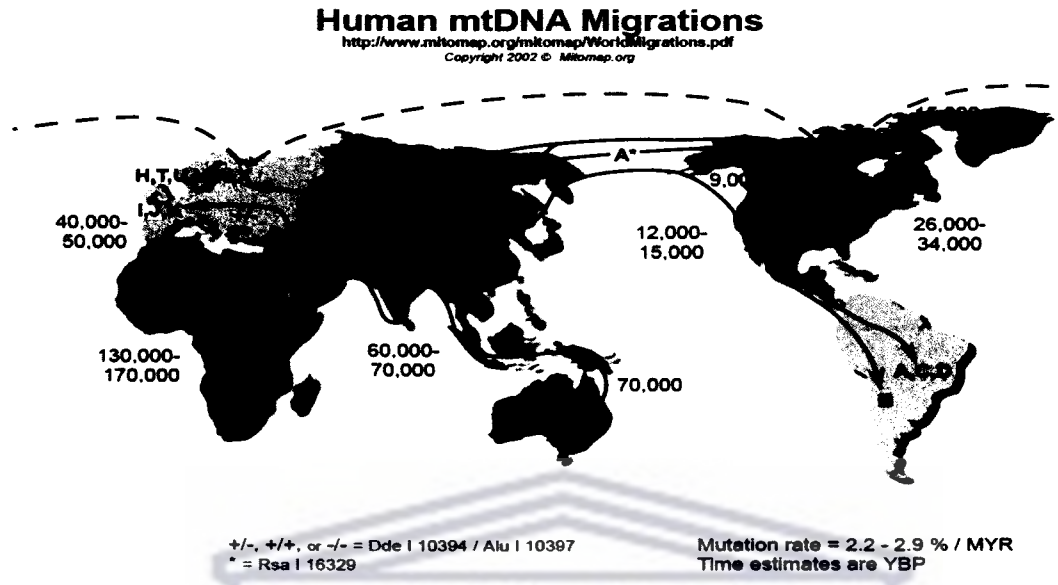


Figure 1.6: Distribution of human mtDNA haplogroups and the migration routes to various continents (adapted from www.mitomap.org/mitomap/WorldMigrations.pdf).

1.5.1.3. African MtDNA Variation

The most ancient mtDNA haplogroup is macrohaplogroup L consisting of haplogroups L0, L1 and L2. Haplogroup L is restricted to Africa, particularly to Sub-Saharan African populations. Haplogroup L3 however radiates out of Africa to form non-African macrohaplogroups M and N (Mishmar et al. 2003; Underhill and Kivisild 2007; Van Oven and Kayser 2008; Maji et al. 2009).

Haplogroups L0, L1 and L2 harbour a specific *Hpa*I site at np 3592 (C to T at np 3594) together with a *Dde*I site at np 10394 (A to G at np 10398) (Chen et al. 2000; Wallace et al. 1999). L1 is defined by an additional np insertion at *Hinfl* site 10806 (T to C at np 10810) and represents 52% of L haplotypes and 29% of all African mtDNA. Haplogroup L2 is defined by an additional mutation at *Hinfl* site 16389 and an *Ava*II site deletion at np 16390 (G to A at np 16390). Haplogroup L2 represents 48 % of L haplotypes and 34% of all African mtDNA (Salas et al. 2002; Wallace et al. 1999).

Haplogroups L0, L1 and L2 are further divided into subgroups, which show affinities for various African populations (Wallace et al. 1999; Gonder et al. 2007). Two length mutations have also been observed for haplogroup L1, a nine np COII/tRNA ^{Tyr} deletion at nps 8272 and 8289 and a 10-12bp insertion of the cytosine's between the tRNA ^{Tyr} and COI gene (nps 5895-5899) (Wallace et al. 1999).

Haplogroup L3 is defined by the absence of the *Hpa*I site at np 3592 and *Dde* I 10394 site. This haplogroup represents only a small fraction of African mtDNA diversity but represents a large proportion of European and Asian haplogroup diversity (Gonder et al. 2007; Maji et al. 2009).

1.5.1.4. Asian MtDNA Variation

All Asian mtDNA haplogroups stem from two macrohaplogroups N and M having unique distributions within Southeastern and Central Asia (Bermishiva et al. 2002; Mishmar et al. 2003; Wen et al. 2005; Underhill and Kivisild 2007). Macrohaplogroup M, has the most frequent and diverse distribution in Asia, Melanesia and Native American populations (particularly South Amerindians) (Kivisild et al. 1999; Bermisheva et al. 2002; Maji et al. 2009). Macrohaplogroup N haplogroups is widely distributed among East Asian, South Asians, Native American (mainly Central and North Amerindians) and Polynesian populations.

Macrohaplogroup M is defined by a *Dde*I site at np 10394 (A to G at np 10398) and an *Alu*I *site* at np 10397 (C to T at np 10400) (Wallace et al. 1999; Roychoudhury et al. 2000; Yonggang and Yaping 2003). Haplogroups derived from M have indicated independent origins among Asian populations. Since certain haplogroups of M are only associated with, Indian populations while others with South East and East Asian

populations. M2 is the most ancient M haplogroup (estimated coalescence time around 60,200 $^{+/-}$ 8600 yrs) and is exclusively found in India among the Dravidian castes and Mundari speaking Austro-Asiatic tribes (Berniell-Lee et al. 2008; Maji et al. 2009).

Haplogroups derived from macrohaplogroup M include C, D, E and G. These haplogroups all harbour the *Dde*I site at np 10394 and *Alu*I site at np 10397. In addition, haplogroup C has a deletion at Hinc II 13259 site (A to G at np 13263) while haplogroup D at the *Alu*I site at np 5176 (C to A at np 5178). Haplogroup E is defined by the absence of an *Hpa*I site at np 7598 (G to A at np 7600), haplogroup G by the presence of a *Hae*III at np 4830 (A to G at np 4833) (Wallace et al. 1999; Roychoudhury et al. 2000; Edwin et al. 2002; Cordaux et al. 2003; Nelson et al. 2007). These haplogroups are absent in Southern Asian populations but has significant frequencies in East Asian populations (Tibetans, Koreans and Han Chinese). Haplogroups C and D also have considerable frequencies among Siberian and Native American populations (Wallace et al. 1999; Comas et al. 2004).

Haplogroups A, B and F originate from macrohaplogroup N and lack the *Dde*I site at np 10394 and the *Alu*I site at np 10397. Haplogroups B and F belong to a subhaplogroup of macrohaplogroup N known as haplogroup R (Wallace et al. 1999; Tolk et al. 2001; Underhill and Kivisild 2007). Haplogroup A is defined by a *Hae* III site at np 663 (A to G at np 663), haplogroup B by a 9bp deletion between COII and tRNA ^{Lys} genes occurring between np 8272-8289 (C to G at np 8272-8280) (Wallace et al. 1999; Edwin et al. 2002; Schurr and Wallace 2002; Nelson et al. 2007). Haplogroup F is defined by the deletion of a *Hinc*II site at np 12406 (G to A at np 12406) (Tolk et al. 2001; Nelson et al. 2007). The distribution of haplogroup A is mainly restricted to East Asian, Siberian and Native American populations (Wallace et al. 1999; Comas et

al. 2004). Haplogroup B is widespread among Central, Southern, and coastal Asian populations (Wallace et al. 1999; Merriwether et al.1999; Schurr and Wallace 2002). Haplogroup F is common in Southeast Asia with the Vietnamese (32%) having the highest frequency. It is also frequent among the Chinese (10.8%), Mongolians (8.7%), Koreans (5.1%), and Japanese (5.1%) but rare in Northeastern Asians (Wallace et al. 1999; Tolk et al. 2001; Schurr and Wallace 2002).

1.5.1.5. European MtDNA Variation

The mtDNA haplogroups of Europeans are assessed using data from both RFLP analysis of the coding region and the sequencing of HV1 region (Bermisheva et al. 2002; Piechota et al. 2004; Vallone et al. 2004). Approximately 98% of European mtDNAs belong to one of nine haplogroups: H, J, K, T, and U, V, W and I and X (Torroni et al. 1996; Wallace et al. 1999; Lell and Wallace 2000, Mishmar et al. 2003; Loogväli et al. 2004; Manwaring et al. 2006). All of these nine haplogroups derive from macrohaplogroup N and haplogroup R (Mishmar et al. 2003; Nelson et al. 2007; Tripathy et al. 2008).

Phylogenetic analysis has revealed that European mtDNA's form two distinct groups based on the presence or absence of the *Dde*I site at np 10394 (A to G at np 10398). Haplogroups lacking the site are H, T, U, and V, W and X while those retaining the site are I, J and K (Torroni et al. 1996; Wallace et al. 1999; Lell and Wallace 2000).

Haplogroup H is the most prevalent mtDNA in Europe with the highest frequency (40-50%) in Western and Northern Europe. An intermediate frequency (20-40%) is observed in Southern, South western and Eastern Europe, North Africa and Turkey. The lowest frequency for haplogroup H is reported in the Middle East, India, and

Siberia (Wallace et al. 1999; Alves-Silva et al. 2000; Bermisheva et al. 2002; Cordaux et al. 2003). Haplogroup H is defined by the absence of the *Dde*I site 10394 and an *Alu*I site at np 7025 (C to T at np 7028) (Wallace et al. 1999, Nelson et al. 2007).

Haplogroups J, T, K and U makeup the next most common haplogroups representing 20 % of European mtDNA (Torroni et al. 1996; Piechota et al. 2004). Haplogroup J is defined by the absence of a *BstN*I site at np 13708 (G to A at np 13078), haplogroup T by the presence of a *BamH*I site at np 13366 and an *AluI* site at np 15606 (Torroni et al. 1996; Wallace et al. 1999). Haplogroup K is defined by the deletion of the *Hae*II site at 9052 and an insertion of a *Hinfl* site at np 13708 (Torroni et al.1996; Macaulay et al. 1999; Wallace et al. 1999). Haplogroup U is identified by the presence of a *Hinfl* site at np 12308 (Torroni et al. 1996; Wallace et al. 1999; Alzualde et al. 2005).

Haplogroups I, K, V, W and X frequencies are relatively low and do not exceed beyond 10% (Torroni et al. 1996; Wallace et al. 1999; Piechota et al. 2004). Haplogroup I is characterized by the deletion of the *Dde*I site at np 1715 (G to A at np 1719) and the insertion of a *AluI* site at np 10032 and is believed to be a European specific haplogroup (Macaulay et al. 1999; Santos et al. 2004; Nelson et al. 2007). The absence of a *Hae*II site at np 9052 and the presence of a *Hinfl* site at np 12308 identifies haplogroup K. Haplogroup V is defined by the deletion of a *Nle* III site at np 4577 (Torroni et al. 1996; Piechota et al. 2004). Haplogroup W is defined by the insertion of a *Ava*II site at 8249 and the deletion of the *Hae*III site at np 8994 and haplogroup X by the deletion of a *Dde*I site at np 1715 (G to A at np 1719). Haplogroup X is also present among Native American populations (Torroni et al. 1996; Piechota et al. 2004; Nelson et al. 2007).

1.5.2. Y-Chromosome Nomenclature

In February 2002 the Y Chromosome Consortium (YCC) published a set of rules to classify the various Y-chromosome haplogroups (Hammer and Zegura 2002; Butler 2003; Cox 2006; Karafet et al. 2008). This was based on genotyped markers representative of a global set of samples identified from 153 haplogroups. The terminology proposed by de Kniff (2000) was used to describe phylogenetic relationships (Hammer & Zegura 2002, Kayser et al. 2003; Underhill 2005). Terms such as lineage, basal lineage, sublineage, clade and subclade were used to describe hierarchical levels within branches of the Y-chromosome phylogenetic tree. NRY lineages defined by SNPs were classified as haplogroups while sublineages of haplogroups defined by STR variation were classified as haplotypes. In addition, monophyletic haplogroups consisting of basal lineages and monophyletic subclades were potentially paraphyletic haplogroups. They represent underived internal nodes of the haplogroup tree and form distinct clusters from lineages defined by a derived state (Hammer and Zegura 2002, Underhill 2005).

Capital letters A-R were used to denote the 18 major clades, while subclades were named after the initial of a major clade and paragroups were distinguished by *. There are two complementary nomenclature systems used to define the relation between major clades and their subclade. An alphanumeric system which easily conveys the relationship between a clade and subclade (e.g. G, G1, etc) and a mutation system defining haplogroups according to their terminal mutation (eg, G-M201, G-P20) (Hammer and Zegura 2002; Underhill 2005; Berniell-lee et al. 2007; Underhill and Kivisild 2007). The YCC hierarchical nomenclature system has since 2002 been widely accepted by the scientific community and has been modified by Jobling and Tyler-Smith in 2003 and most recently by Karafet et al in 2008 (Karafet et al. 2008).

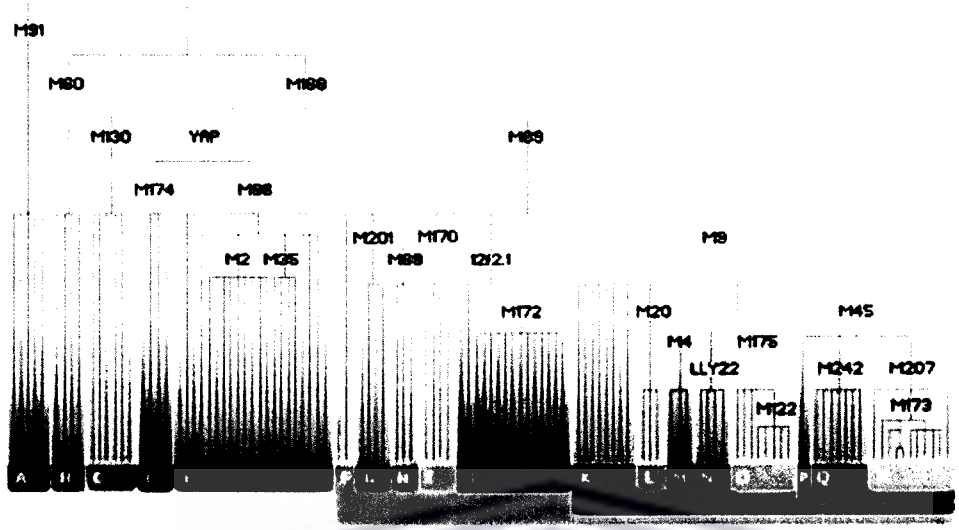


Figure 1.7: A phylogenetic tree of NRY haplogroups A-R indicating the mutations defining each haplogroup. (adapted from genographic.nationalgeographic.com/ genographic/index.html).

1.5.2.1. Y-Chromosome Variation

The Y-chromosome variation seen in worldwide populations has been attributed to past migrations of males. Initial expansions occurred out of Africa then Asia with the latter heavily influencing global patterns of NRY variation (Hammer et al. 2001). In addition, NRY variation indicate geographical specificity thought to be influenced by patrilocality as mtDNA variation show a reduction in haplogroup geographical clustering in comparison to Y-chromosome lineages (Underhill et al. 2000; Hammer et al. 2001; Cavalli-Sforza and Felman 2003; Jobling and Tyler-Smith 2003).

The most ancient Y-chromosome lineages, haplogroup A and B are restricted to Africa. They are distinct from other African and non-African haplogroups due to the absence of the M168 mutation (Underhill et al. 2001; Cavalli-Sforza and Felman 2003; Jobling and Tyler-Smith 2003; Underhill and Kivisild 2007). The M168 mutation represents an important central node in the Y-chromosome phylogenetic tree as three subclusters coalesce at the root of the M168 node. This includes haplogroups C, D/E

and F from which both African and non-African lineages are derived. Haplogroup F represents an internal node of all other haplogroups (G to R) that encompass both the M9 and M45 lineages (Fig 1.8) (Underhill 2000; Underhill et al. 2001; Cavalli-Sforza and Felman 2003; Rootsi 2004; Underhill and Kivisild 2007).

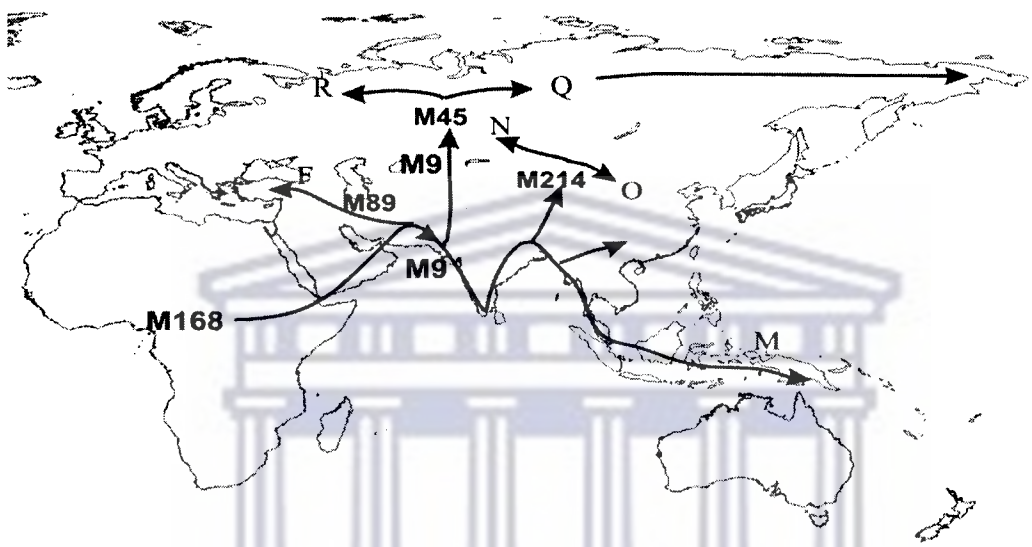


Figure 1.8: The global distribution of Y-chromosome haplogroups derived from the M168 mutation and their migration routes (adapted from Underhill 2003).

1.5.2.2. African Y-Chromosome Variation

African Y-chromosome lineages are mainly characterized by three NRY haplogroups A, B and E. Haplogroup A is defined by the M91 mutation and haplogroup B by mutations M60 and M181. Furthermore, represents the most ancient African and NRY lineages (Underhill et al. 2001; Cruciani et al. 2002; Hammer and Zegura 2002; Cavalli-Sforza and Felman 2003; Jobling and Tyler-Smith 2003; Karafet et al. 2008). Haplogroup A and B are distributed across Africa at low frequencies, present in South African Khoisan and Bantu speakers. These haplogroups are also present among Central African pygmies and populations from Sudan, Ethiopia and Mali. The lowest

frequency is observed among North western Africans (Underhill et al. 2001; Cruciani et al. 2002; Karafet et al. 2008).

Haplogroup E is distinguished from haplogroups A and B, as it has evolved from the M168 mutation. It is derived from the haplogroup DE that is defined by a YAP element (Alu insertion occurring on the long arm of the Y-chromosome) and the equivalent M145/M203 base substitutions (Hammer and Horai 1995; Underhill et al. 2001). Haplogroup E is prevalent in Africa with a moderate frequency in the Middle East and Southern Europe and is defined by the mutations M40/M96. The haplogroup also has a sporadic distribution in Central and South Asia (Underhill et al. 2001; Underhill 2003; Semino et al. 2004; Karafet et al. 2008).

Haplogroup E is comprised of subgroups E-M33, E75 and E-P2. These haplogroups are mainly restricted to Sub-Saharan Africa with a low distribution in Europe. Haplogroup E-P2 has the greatest distribution consisting of three subgroups. However, only two E-M2 and E-M35 are the most widely distributed (Underhill et al. 2001; Luis et al. 2004; Semino et al. 2004). Haplogroup E-M2 is mainly observed in Sub-Saharan Africa and indicates Bantu expansion across Africa with the highest frequency observed in Senegal and a low incidence in North Africa and Iraq. Haplogroup E-M35 is found in Africa, Near East and Europe. Among the subgroups E-M78 is found in Europe, E-M123 in the Near East, North Africa and Europe while the remaining subgroups are mainly restricted to Ethiopians (Underhill et al. 2001; Luis et al. 2004; Semino et al. 2004; Karafet et al. 2008). A small percentage of African lineages also belong to haplogroups J and R frequently found in Europe and the Middle East (Hammer and Zegura 2002).

1.5.2.3. Asian Y-Chromosome Variation

The Asian continent has the highest geographical distribution of NRY haplogroups, as 15 of 18 haplogroups/paragroups are found within this region. This has been attributed to genetic drift and displays distinct region specificity among populations (Hammer and Zegura 2002; Rootsi 2004). Furthermore, Asian lineages also indicate deep structure within the NRY phylogeny as haplogroups C and D both stem from the M168 mutation (Underhill et al. 2001; Karafet et al. 2008).

Haplogroup C is defined by the M216 mutation and is absent in African populations and is mainly restricted to Asian populations. It is distributed in South eastern Asia, East Asia extending into Central and Northern Asia as well as Australo-Melanesia and the Americas (Underhill et al. 2001; Jobling and Tyler-Smith 2003; Underhill and Kivisild 2007; Karafet et al. 2008).

Haplogroup D is derived from haplogroup DE and is restricted to Asia. This haplogroup is defined by the M174 mutation and is prevalent in Central Asia (Tibet) and Japan with low frequencies observed in Southeast Asia and the Andaman Islanders (Underhill et al. 2001; Hammer and Zegura 2002, Karafet et al. 2008).

Other Asian haplogroups originate from the M89 mutation and subsequent lineages. Therefore, further characterization of Asian haplogroups will be classed according to the following mutations: A) haplogroups derived from M89 b) haplogroups derived from M89 and M9 and C) haplogroups derived from M89 and M45.

1.5.2.3.1. Haplogroups derived from M89

This includes haplogroups F and H, which are mainly restricted to the Indian subcontinent. Macrohaplogroup F is defined by the M89 mutation and has an extensive

global distribution. However, few populations posses this mutation alone but, rather downstream mutations defining M89 lineages. Yet in Asia, the haplogroup and sublineages have been reported which is primarily prevalent in Indian populations (Athey 2005; Karafet et al. 2008).

Haplogroup H is defined by the M69 mutation and consists of a paragroup and nine additional lineages. The haplogroup is restricted to the Indian subcontinent and was initially defined by the marker M52 however, recent studies indicate the marker defines a subgroup H1 (Hammer and Zegura 2002, Sahoo et al. 2006; Karafet et al. 2008; Tripathy et al. 2008).

1.5.2.3.2. Haplogroups derived from M89 and M9

Haplogroups harbouring the M9 mutation defined by a C to G base substitution are common in Asia. Yet, only few individuals in some populations harbour the M9G mutation alone (Kayser et al. 2003; Karafet et al. 2005). Haplogroups derived from the M9 mutation have moderate distributions in the Phillipines, Indonesia, Melanesia, Papua New Guinea (PNG) and Micronesia (Kayser et al. 2003, Karafet et al. 2005; Karafet et al. 2008). This includes haplogroups L, M, N and O defined by the following downstream mutations M11, M5, M214 and M175 respectively. Haplogroup L is mainly found in India however, is also present in the Middle East, Central Asia, Northern Africa, Europe and the Mediterranean coast (Underhill et al. 2001; Jobling and Tyler-Smith 2003; Karafet et al. 2008). Haplogroup M is restricted to near and remote Oceania and Eastern Indonesia with males from PNG and Melanesia having the highest frequency (Jobling and Tyler-Smith 2003; Karafet et al. 2005; Karafet et al. 2008).

Haplogroup N is prevalent in North western Asia whilst haplogroup O in East Asia. Haplogroup O can also be found in Central Asia and Oceania with moderate to low frequencies. The haplogroup is also highly diverse as subgroups show regional specificity. Haplogroup O-M95 is present mainly among Southeast Asians while O-M122 in Austronesians (Hammer and Zegura 2002; Jobling and Tyler-Smith 2003; Kayser et al. 2003; Underhill and Kivislid 2007; Kumar et al. 2007; Karafet et al. 2008).

1.5.2.3.3. Haplogroups derived from M89 and M45

The few remaining haplogroups representing Asian diversity include haplogroup P defined by the mutation M45. This haplogroup is found at a very high frequency in Central Asia. Later expansions of the haplogroup is observed in the Americas, Europe and India (Wells et al. 2001; Bortolini et al. 2003). Giving rise to haplogroups Q defined by the M242 mutation and haplogroup R defined by the M207 mutation.

Haplogroup Q is frequent in Central and Northern Asia with low frequencies in East Asia, Europe and the Middle East. This haplogroup also represents a major Native American lineage as haplogroup Q3 defined by the M3 mutation is strictly associated with Native American populations (Jobling and Tyler-Smith 2003; Cox 2006; Karafet et al. 2008).

Haplogroup R represent a small percentage of Asian lineages and is widely distributed across Europe but to a lesser extent across North Africa and West Asia (Cavalli-Sforza and Felman 2003; Karafet et al. 2005; Underhill and Kivislid 2007; Karafet et al. 2008).

1.5.2.4. European Y-Chromosome Variation

European Y-chromosome haplogroups are well studied and characterized, indicating 95% of European Y-chromosome variation belonging to 10 phylogenetic haplogroups. Haplogroups I, N3, R1a and R1b represent at least 70-80% of European Y-chromosome diversity while 20% is defined by haplogroups E3b, G and J2 (Rootsi 2004; Karafet et al. 2008). Haplogroup I is defined by the mutation M170 and is widely distributed across Europe with a low frequency in the Middle East. In addition, haplogroup I along with R represents major European Y-chromosome haplogroups (Hammer and Zegura 2002; Rootsi et al. 2004; Underhill and Kivislid 2007; Karafet et al. 2008). N3 is a sublineage of haplogroup N defined by markers M45/M178 and has a high frequency among East Siberian ethnic groups (Rootsi 2004; Kharkov et al. 2007). Haplogroup R1 is defined by the M173 mutation and is derived from haplogroup R-M207. Haplogroup R1 is widespread across Europe with the highest frequency occurring in Western Europe among the Basque and British populations (Wells et al. 2001; Kayser et al. 2005; Sahoo et al. 2006).

Haplogroups E, G and J have a low frequency in Europe attributed to the demic diffusion of Neolithic farmers into Europe (Underhill et al. 2001; Beleza et al. 2005). Furthermore, the occurrence of haplogroup E subgroups E-M35 and E-M123 in West Europe has been attributed to late demographic expansions from the Balkans along the Mediterranean Sea (Rootsi 2004; Beleza et al. 2005). Haplogroup G is defined by M201 and is common in the Middle East, Mediterranean and the North Caucus. It also appears in Southeast Europe (Cinnioğlu et al. 2004; Monteil et al. 2005; Karafet et.al. 2008).

The distribution of haplogroup J and lineages occurs in the Middle East, North Africa, Europe, Central Asia, Pakistan and India. However, a subgroup defined by the M172 mutation mainly occurs in Europe while subclade J-M267 is present in the Middle East, North Africa and Ethiopia (Underhill et al. 2001; Cavalli-Sforza and Felman 2003; Cinnioğlu et al. 2004; Underhill and Kivislid 2007; Karafet et al. 2008).

1.4. Aims of the present study

Cape Muslims were brought during the 16th and 18th century from Africa and Asia. They were slaves, convicts and political exiles and they contributed in the ethnic diversity of the present Cape Muslim population. The history of the Muslims in the Cape has attracted researchers from different disciplines such as historians, anthropologists and sociologist. No comprehensive genetic data is available to give insight into the genetic structure of the present Muslim population in the greater Cape Metropolitan area. This study therefore attempts to contribute to the existing knowledge about the Muslim population in Cape Town but from a genetics perspective. Past demographic events were therefore inferred through the analysis of the maternal and paternal gene pools of Kensington Muslims and Cape Muslims. The specific aims of the present study were as follows:

1. To study the genetic history and diversity of Muslims living in the Cape Metropolitan area (Cape Muslims) from a maternal and paternal perspective. MtDNA and Y-chromosome molecular genetic markers were analyzed to reconstruct the major male and female lineages of Cape Muslims (Chapter II).

2. To determine the contributions of Major African, Asian and European female lineages to the maternal gene pool of Kensington Muslims through the analysis of mtDNA genetic markers (Chapter III).

3. To assess which male lineages mainly contributed towards the paternal gene pool of Kensington Muslims by analyzing Y-chromosome genetic markers (Chapter III).

Chapter 2

Reconstruction of Major Maternal and Paternal Lineages of the Cape Muslim population

2.1. Introduction

South African Muslims represent a prominent multiethnic minority. The foremost ethnic groups are the Indian Muslims and the Cape Muslims. The latter representing the largest and oldest Muslim population in the city of Cape Town (Matthee, 2008). The establishment of the community represents one of the important and relatively unexplored chapters in South African history. The community's existence can be attributed to slave trade, migration, colonialism, ancient trade routes in Africa, inter-community marriages and the spread of Christianity and Islam in the region. The earliest Cape Muslims were brought by the Dutch East Indian Company (DEIC) as slaves, political exiles, and convicts. Originating from various regions in Africa and Asia, they contributed in the diverse heterogeneity of the present Muslim population at the Cape (Bradlow and Cairns 1978; Davids 1980; Mahida 1993; Shell 2000; Da Costa and Davids 1994; Benjeddou et al. 2006). Currently however, no significant studies exist regarding the genetic structure of the population. The history of the community has been extensively researched yet is inadequate to validate the genetic and geographical origins of the population. The aim of the study described in this chapter was to investigate the genetic genealogy of the Muslim population living in the Cape Metropolitan area. Mitochondrial and Y-chromosome DNA variations were used to infer the geographic/ethnic origins of the population. This was achieved by screening single nucleotide polymorphisms (SNP) using polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP). Single nucleotide polymorphisms are geographical and population specific and arranged into distinct phylogenetic groups

known as haplogroups, which have extensively been used in studies to infer past demographic events and genetic variation in numerous population studies (Torroni et al. 1996; Zerjal et al. 1997; Jorde et al. 1998; Wallace et al. 1999; Lell and Wallace 2000; Bermisleva et al. 2002; Edwin 2002; Budowle et al. 2003; Jobling and Tyler-Smith 2003; Kayser et al. 2003; Martinez Marignac et al. 2003; Quintáns et al. 2004; Hurles et al. 2005; Wood et al. 2005; Underhill and Kivisild 2007; Maji et al. 2009).

2.2. Material and Methods

2.2.1. DNA Samples

Buccal samples were collected from a hundred unrelated Muslim males born and residing within the Cape Town Metropolitan area. Ethics clearance for the study was obtained from the University of the Western Cape Ethics Committee. Donors signed an informed consent and completed an anonymous questionnaire pertaining to their own as well as their maternal and paternal family history regarding place of birth, religious affiliation, home language, and ethnicity. Each sample was then catalogued using the assigned reference number and stored at -20°C.

2.2.2. Questionnaire

The information obtained from the population through the questionnaire was used together with the results of the genetic study in an integrated manner. This was done to draw more conclusive answers regarding the population history and explore past demographic events of this population. The questionnaire was used to deduce some demographic events such as conversion, migration and inter-ethnic marriages.

Cape General Muslims: this term was used throughout the thesis and refers to Muslim individuals born and residing within the Cape Metropolitan area.

Conversion: The religious denominations of the donor mother and father were compared to that of their parents to establish if either parent had converted to Islam.

Internal Immigrants: This refers to individuals who migrated from areas such as East London, Durban, Johannesburg, and Port Elizabeth to the Cape Metropolitan area.

External Immigrants: These are individuals who had immigrated from foreign countries such as India, and other parts of the world to the Cape Metropolitan area.

Inter-ethnic marriages: This refers to the marriages of individuals of different ethnic identities.

2.2.3. DNA Extraction

Total genomic DNA was extracted from buccal swabs using the BuccalAmp™ DNA Extraction kit (*Epicentre Technologies*) according to the manufacturer's specifications (Appendix I). Extracted DNA was further purified with a standard phenol chloroform protocol to improve the quality of the DNA as well as the sensitivity of the PCR. The comprehensive protocol is listed in Appendix I.

2.2.4. Quantification of DNA

Extracted DNA was quantified using the Nanodrop ND 1000 UV-Vis spectrophotometer. Thereafter 20µl aliquots were prepared as working stock while the remaining stock was stored at -20°C.

2.2.5. MtDNA Markers and Typing Strategy

A hierarchical strategy (Figure 2.1) was employed to genotype samples using SNP's for the identification of haplogroups B, H, J, L, and M. The SNP's investigated in the study included: an *Alu*I deletion at np 7025, *Bst O*I deletion at np 13704, *Hpa*I

insertion at np 3592, *Dde*I insertion at np 10394 and a *Alu*I insertion at 10397. Furthermore, a 9bp intergenic length mutation defining haplogroup B was screened for the absence of one of two copies of the tandem repeat sequence (CCCCCTCTA). The tandem sequence is located between the cytochrome oxidase subunit and tRNALYS region (region V) on the mitochondrial DNA molecule. Samples were initially screened for nucleotide positions 10394 *Dde*I and 10397 *Alu*I. Published data using restriction fragment length polymorphisms have indicated that the absence or presence of these positions (motif) separate haplogroups into two major categories (Torroni et al. 1996; Merriweather et al. 1999; Wallace et al. 1999; Bermisheva et al. 2002; Schurr and Wallace 2002; Loogväli et al. 2004). This reduced the potential haplogroups to which an unknown mtDNA might belong. These sites are situated adjacent to each other and was determined using a single amplicon. Following the result from the motif, samples were then tested for haplogroup specific markers sharing the same motif.

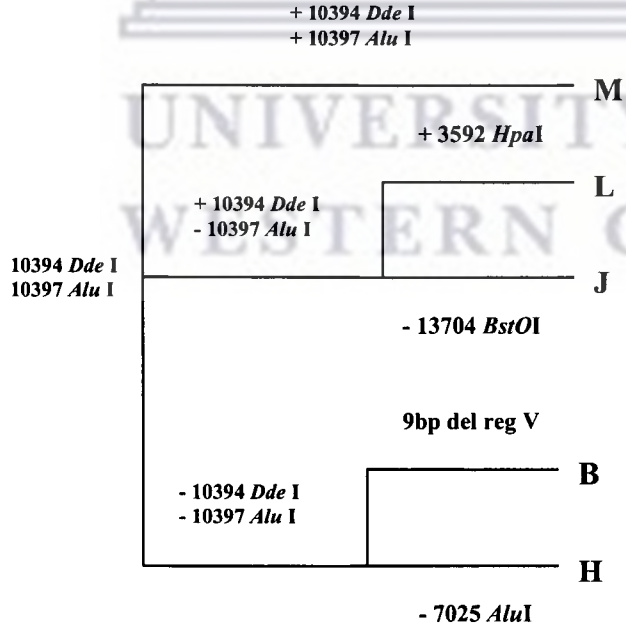


Figure 2.1. A simplified phylogenetic tree to illustrate the hierarchical manner in which mtDNA haplogroups were typed (adapted from Torroni et al. 1996).

2.2.6. Y-Chromosome Markers and Typing Strategy

The Y-chromosomal binary markers typed (Table 2.2) included eight SNP's M9, M170, M173, M35, M52, M2, M175 and one YAP insertion polymorphism. All the markers had been previously published (Underhill et al. 1997; Hammer and Horai 1995, Flores et al. 2003; Kayser et al. 2005). Samples were genotyped using the PCR-RFLP methodology, following a phylogenetic hierarchical approach (Figure 2.2). The nomenclature for the Y-chromosome haplogroups was according to the Y Chromosome Consortium.

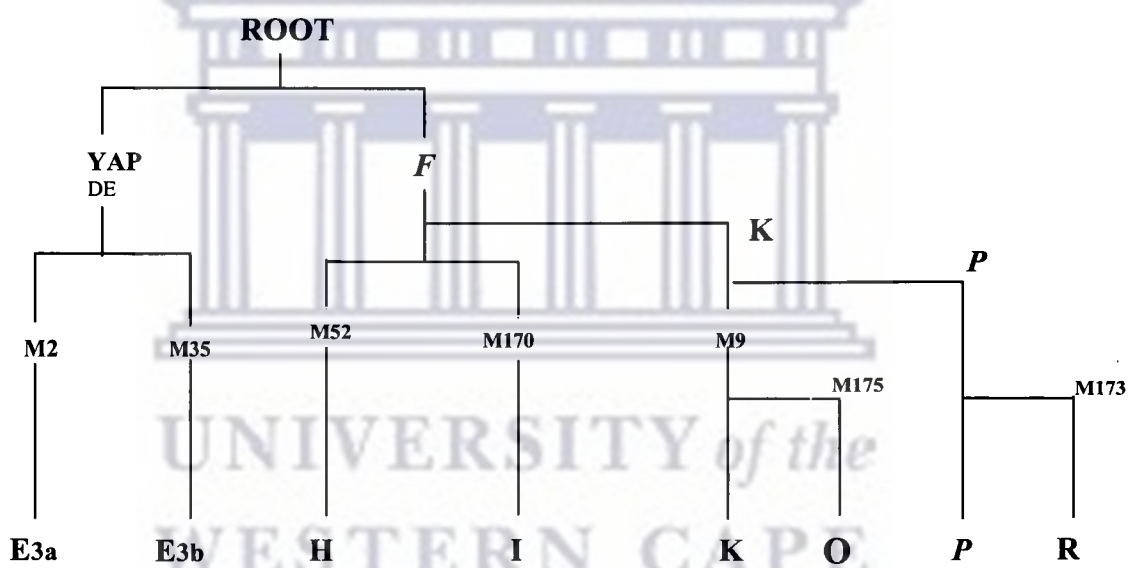


Figure 2.2. A phylogenetic tree of Y-chromosome haplogroup markers typed in this study (modified from Jobling and Tyler-Smith 2003). Italicized haplogroups were not typed.

2.2.7. PCR Amplification

2.2.7.1. Mitochondrial DNA

PCR amplifications were performed in a final reaction volume of 25µl, containing 1x PCR Buffer with 2.5mM $MgCl_2$ (*NEB*), 1U of *Taq* DNA polymerase (*NEB*), 0.2mM dNTPs (*Roche*), 0.2µM of each primer and 10ng of template DNA. Primers were synthesized by *Whitehead Scientific* using previously described sequences obtained from (Martinez-Cruzado et al. 2001 and Santos et al. 2004).

Amplifications were performed using a GeneAmp PCR System 2700 thermal cycler (*Applied Biosystems*) and Perkein Elmer GeneAmp PCR System 2400 (*Applied Biosystems*). Thermal cycling conditions for primer sequences from Santos et al. 2004 were as follows: 1 cycle at 94°C for 5 min, 40 cycles at 94°C for 50 s, 1 min at primer specific annealing temperature and 72°C for 1min (Table 2.1), followed by a final cycle at 72°C for 5 min completing the amplification.

Thermal cycling conditions for Martinez-Cruzado et al. 2001 primer sequence was: 1 cycle at 94°C for 5 min, 30 cycles at 94°C for 1 min, 55°C for 1 min and 72°C for 1.5 min. The final extension was at 72°C for 10 min.

Table 2.1: Primer sequences, annealing temperatures and PCR fragment sizes of mtDNA markers used in this study

Motif	Haplogroup	Defining marker	Primer sequence 5'-3'	Annealing Temperature °C	Fragment Size (bp)
(+/-)	J, L, M	10394 *Dde*I	Fw TTGATCTAGAAATTGCCCTC	48.2	276
			Rv GTATTCCTAGAAGTGAGATG		
			Santos et al. 2004		
(+/+)	M	10397 *Alu*I	Fw TTGATCTAGAAATTGCCCTC	48.2	276
			Rv GTATTCCTAGAAGTGAGATG		
			Santos et al. 2004		
(+/-)	L	3592 *Hpa*I	Fw CTAGGCTATATACAACTACGC	50.9	330
			R v GGCTACTGCTCGCAGTG		
			Santos et al. 2004		
(+/-)	J	13704 *BstO*I	Fw TCGAATAATTCTTCTCACCC	47	137
			Rv TAGTAATGAGAAATCCTGCG		
			Santos et al. 2004		
(-/-)	B	9bp del	Fw ACAGTTTCATGCCCATCGTC	55	121 deletion absent
			Rv ATGCTAAGTTAGCTTTACAG		111 deletion present
			Martinez-Cruzado et al. 2001		
(-/-)	H	7025 AluI	Fw CCGTAGGTGGCCTGACTGGC	56	123
			Rv TGATGGCAAATACAGCTCCT		
			Santos et al. 2004		

* (+/+) indicates the presence of both 10394 *Dde* I and *Alu* I 10397 sites respectively
* (+/-) indicates the absence for *Alu* I 10397 site only
* (-/-) indicates the absence for both 10394 *Dde* I and *Alu* I 10397 sites respectively

2.2.7.2. Y-Chromosome

Standard and Touch down PCR amplifications (Table 2.2) were performed in a final volume of 25µl, containing 1X PCR Buffer with 1.5-2.5mM $MgCl_2$ (*Bioline*), 1U of *Taq* DNA polymerase (*NEB*), 0.2mM of dNTPs (*Roche*), 1.8mg/ml of BSA (*NEB*), 0.4µM (M9, M170, M173, M175, M35 and Yap), 0.8µM (M52) and 2µM (M2) of

each primer and 20-50ng of template DNA. Primers were synthesized by *Whitehead Scientific* using previously described sequences obtained from (Hammer and Horai 1995; Underhill et al. 2000; Flores et al. 2003; Kayser et al. 2005; Monteil et al. 2005). Amplifications were performed using a GeneAmp PCR System 2700 thermal cycler (*Applied Biosystems*) or Perkein Elmer GeneAmp PCR System 2400 (*Applied Biosystems*). The primer sequences for Y-chromosome markers M9, M170, M52, M175, M2 and YAP element were amplified according to standard PCR protocols. All the primer sequences except M2 and the YAP element had the following thermal cycling conditions: 1 cycle at 95°C for 4 min, 40 cycles at 94°C for 30s, 30s at primer specific annealing temperature (described in Table 2.2), 72°C for 45s followed by a final cycle at 72°C for 10 min.

Thermal cycling conditions for the YAP element were: 1 cycle at 94°C for 2 min, 40 cycles at 94°C for 1 min, 55°C for 30s and 72°C for 1min.

Thermal cycling conditions for M2 were: 1 cycle at 95^0C for 4min 40 cycles starting at 94°C for 30s, 62^{0^0}C for 30s, 72°C for 45s followed by a final cycle at 72°C for 10 min.

Touchdown PCR was used for M173 and M35 using a GeneAmp PCR System 2700 thermal cycler (*Applied Biosystems*). The touchdown PCR cycling conditions were as follows: 1 cycle at 95°C for 4 min, 14 cycles at 94°C for 30s, 30s at 56-59°C respectively decreasing annealing temperature by 0.5°C per cycle (Table 2.2), 72°C for 45s and a final cycle at 72 °C for 10 min this was followed with 40 cycles at a constant annealing temperature of 63°C.

Table 2.2: Primer sequences, annealing temperatures and PCR fragment sizes of Y-chromosome markers

Haplogroup	Defining marker	Primer sequence 5′-3′	Annealing Temperature °C	Fragment Size (bp)
DE	Yap element	Fw CAGGGGAAGATAAAGAAATA	55	150/455
		Rv ACTGCTAAAAGGGGATGGAT		
		Hammer and Horai 1995		
E3a	M2	Fw AGGCACTGGTCAGAATGAAG	62	209
		Rv AATGGAAAATACAGAGCTCCCC		
		Underhill et al. 1997		
E3b	M35	Fw TAAGCCTAAAGAGCAGTCAGAG	59/63#	513
		Rv CCAATTACTTTCAACATTTAAGACC		
		Kayser et al. 2005		
H	M52	Fw GCAGCATATAAAACTTTCAGG	56	88
		Rv GACGAAGCAAACATTTCAAG3		
		Flores et al. 2006		
I	M170	Fw TATTTACTTAAAAATCATGGTTC	49	99
		Rv CAATTACTTTCAACATTTAAGACC		
		Kayser et al. 2005		
K	M9	Fw GCAGCATATAAAACTTTCAGG	54	164
		R v GAAATGCATAATGAAGTAAGCG		
		Kayser et al. 2005		
O	M175	Fw TTGAGCAAGAAAAATAGTACCCA	56	439
		Rv CTCCATTCTTAACTATCTCAGGGA		
		Underhill et al. 2000		
R1	M173	Fw TTTCTGAATATTAACAGATGACAACG	56/63 #	128
		R v CAGTACTCACTTTAGGTTTGCCA		
		Kayser et al. 2005		

# indicates touchdown PCR

2.2.7.3. Verification of PCR product

PCR amplicons of mtDNA and Y-chromosome were verified by agarose gel electrophoresis. An aliquot of 9µl PCR product was run on a 2% agarose gel w/v *(Whitehead Scientific)* stained with 0.1mg/ml Ethidium Bromide and photodocumentated under UV light. A PCR marker (*NEB*) was run alongside PCR products to ensure that the correct amplicon size was observed (Table 2.1 and 2.2).

2.2.8. Restriction Digestion

Following PCR verification, respective PCR products were then subjected to restriction digestion. Restriction digestion reactions were performed in a final reaction volume of 25µl and 20µl for the mtDNA and Y-chromosome respectively. These reactions contained 10µl of the unpurified amplified product incubated with the haplogroup specific restriction enzyme from *NEB* (Table 2.3.) according to conditions specified by the manufacturer.

Aliquots of 9µl digestion products were electrophoresed on a 3% w/v *Whitehead Scientific* agarose gel previously stained with 0.1mg/ml ethidium bromide and photodocumentated under UV light. The banding patterns of the digested fragments were compared to a PCR marker (*NEB*) to determine the ancestral and derived polymorphic states.

Table 2.3: Restriction enzymes, optimal temperatures and banding patterns used to genotype mtDNA and Y-chromosome haplogroups.

Haplogroup	Defining marker	Enzyme	Temperature (°C)	PCR-RFLP fragments (bp)	
				Ancestral	Derived
MtDNA					
J, L, M	10394 *Dde* I	*Dde* I	37	171,105	133, 105, 38
M	10397 *Alu*I	*Alu*I	37	276	147, 129
L	3592 *Hpa*I	*Hpa*I	37	330	207, 123
J	13704 *Bst*OI	*Bst*OI	60	137	98, 39
B	9bp del region V	N/A	N/A	120	111
H	7025 *Alu*I	*Alu*I	37	108, 15	78, 30, 15
Y chromosome					
DE	YAP element	N/A	N/A	150	455
E3a	M2	*Nla* III	37	102, 68, 39	68, 141
E3b	M35	*Bts*I	55	513	351, 162
H	M52	*Sty* I	37	63, 25	82
I	M170	*NlaIV*	37	99	23, 76
K	M9	*Hinf*I	37	100, 64	164
O	M175	*Ear* I	37	87, 357	439
R1	M173	*HpyCh4*IV	37	102, 26	128

2.3. Results and Discussion

2.3.1. Self-Perceived Ethnic Classification

A hundred samples were included in this part of the study. DNA was extracted and analyzed from these samples. The questionnaire revealed that most donors classified themselves as Coloured, Malay, or Indian. A small percentage of donors, however, were reluctant to indicate their ethnicity. The term Other Muslims was used for this group. A study conducted by Da Costa (1994) concerning Muslim identity reported that a number of Cape Muslims consider being Muslim as their only and most important classification. They therefore do not recognise any other racial or ethnic classification (Da Costa 1994). Other Muslims represented 10% (10) of donors (10%) while 34% (34) were Coloured, 29 % (29) Indian and 27% (27) Malay (Figure 2.3).

Figure 2.3. Ethnic groups (as defined by donors) observed in the study.

2.3.2. PCR-RFLP Analysis

All the respective SNP markers used in the study were screened using PCR-RFLP except mtDNA haplogroup B and Y-chromosome haplogroup DE. These haplogroups were characterized by either an insertion or deletion polymorphism thereby rendering restriction digestion non-essential. The banding pattern of the amplified fragment was therefore sufficient to discern a genotype. None of the donor's harboured the 9bp deletion for haplogroup B while three individuals harboured the YAP element defining the haplogroup DE.

Amplicons amplified with mtDNA defining marker 10394 *Dde*I (Figure 2.4) contained a non-polymorphic restriction site. This produced cleaved fragments even in the absence of the polymorphism serving as an internal control preventing false positive results (Figure 2.5). This was particularly relevant as marker 10394 *Dde*I was used in the hierarchical approach to differentiate haplogroups according to their motif. Seventy –six donors were found to harbour this polymorphism.

Haplogroups defined by the absence of cleaved digested fragments were verified using both positive and negative controls. The Y-chromosome haplogroup K genotypes were verified in this manner. Figure 2.6 shows the amplified product of Y-chromosome haplogroup K while Figure 2.7 indicates the digested fragments. This polymorphism was found among forty-three individuals in the study, making it the dominant Y-chromosome haplogroup. The optimal resolution of amplified fragments and digestion products was achieved using Whitehead Scientific agarose gels but fragments smaller than 50bp however, could not be detected.

Figure 2.4: PCR amplification of defining marker 10394 *Dde*I. Lane 1 contained the PCR marker and Lane 2 the negative control. The remaining lanes indicate amplified fragments of 276bp.

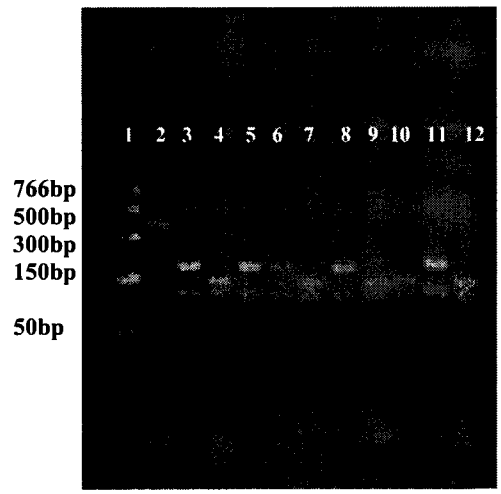


Figure 2.5: PCR-RFLP screening for the 10394 *Dde*I site. Lane 1 contained the PCR marker and Lane 2 the undigested PCR product (negative control). Lanes 3-12 indicate *Dde*I digested PCR products. Five samples harboured the derived state (lanes 4, 7, 9, 10 and 12) for the polymorphisms (133bp, 105bp, and 38bp) while the remaining samples were negative (171bp, 105bp).

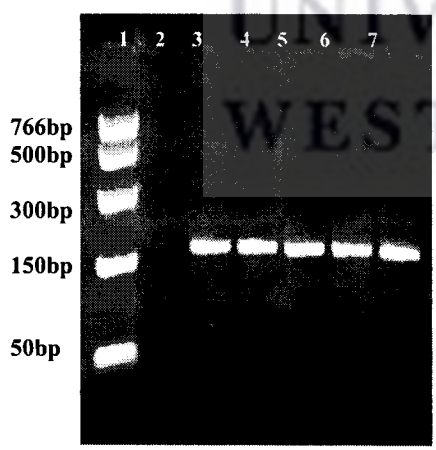


Figure 2.6: PCR amplification of M9 defining Haplogroup K. Lane 1 contained the PCR marker and Lane 2 the negative control. Lanes 3-7 indicate amplified fragments of 164bp.

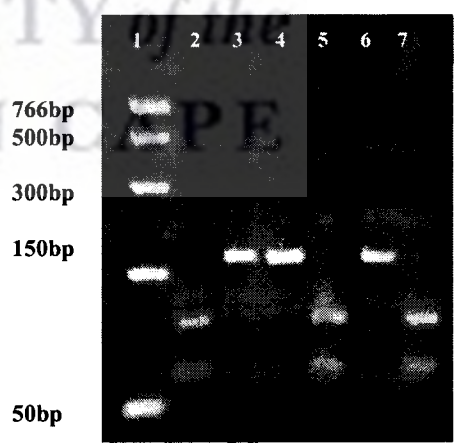


Figure 2.7: PCR-RFLP screening of marker M9. Lane 1 contained the PCR marker. Lane 2 contained a negative control (100bp, 64b) and Lane 3 a positive control (164bp). Lane 4 and 6 indicate samples harbouring the polymorphism. Lane 5 and 7 indicate samples negative for the polymorphism.

2.3.3. MtDNA Haplogroup Variation

The mtDNA haplogroup frequencies obtained in the study is reported in Figure 2.8 A-E. The haplogroups L (34%), M (33%), H (24%), and J (9%) characterized the mtDNA diversity present among the General Cape Muslim population. Haplogroup L was the most common haplogroup present among all the self-perceived ethnic groups. Cape Coloured Muslims (47%) and Cape Malay Muslims (44%) showed the most significant frequencies while Cape Other Muslims (20%) and Cape Indian Muslims (14%) the lowest (refer to Figure 2.8 B, C, D and E).

Haplogroup L is widely distributed among African populations especially in Sub-Saharan Africa (Wallace et al. 1999; Gonder et al. 2007; Maji et al. 2009). The presence of this haplogroup therefore suggests Khoisan and Bantu females as likely contributors of this lineage. However, African slaves originating from Mozambique and Madagascar may also have been contributors as haplogroup L, represents a common mtDNA among the Malagasy population (Hurles et al. 2005).

Factors such as conversion, internal immigrants, and inter-ethnic marriage listed in Table 2.4. were shown to have significantly contributed to haplogroup L frequency in Cape Coloured Muslims. The combined frequency of these factors represents 24 % of L mtDNA in Cape Coloured Muslims. In Cape Malay Muslims 15% of haplogroup L, frequency was due to internal immigrants and inter-ethnic marriages. In Cape Indian Muslims haplogroup L accounts for 14% of their mtDNA however, 7% could possibly be attributed to geneflow with the Siddis. The Siddis are tribal groups in India that were brought as slaves during the 16th and 17th century to India from East Africa and other regions in Africa such as Mozambique (Lodhi 1992; Gauniyal et al. 2008). This is highly plausible given that maternal grandmothers were recent immigrants from

India (Table 2.4). These results also suggest that the remaining L mtDNA was mainly derived through recent admixture with either Cape Coloured Muslims or Cape Malay Muslims.

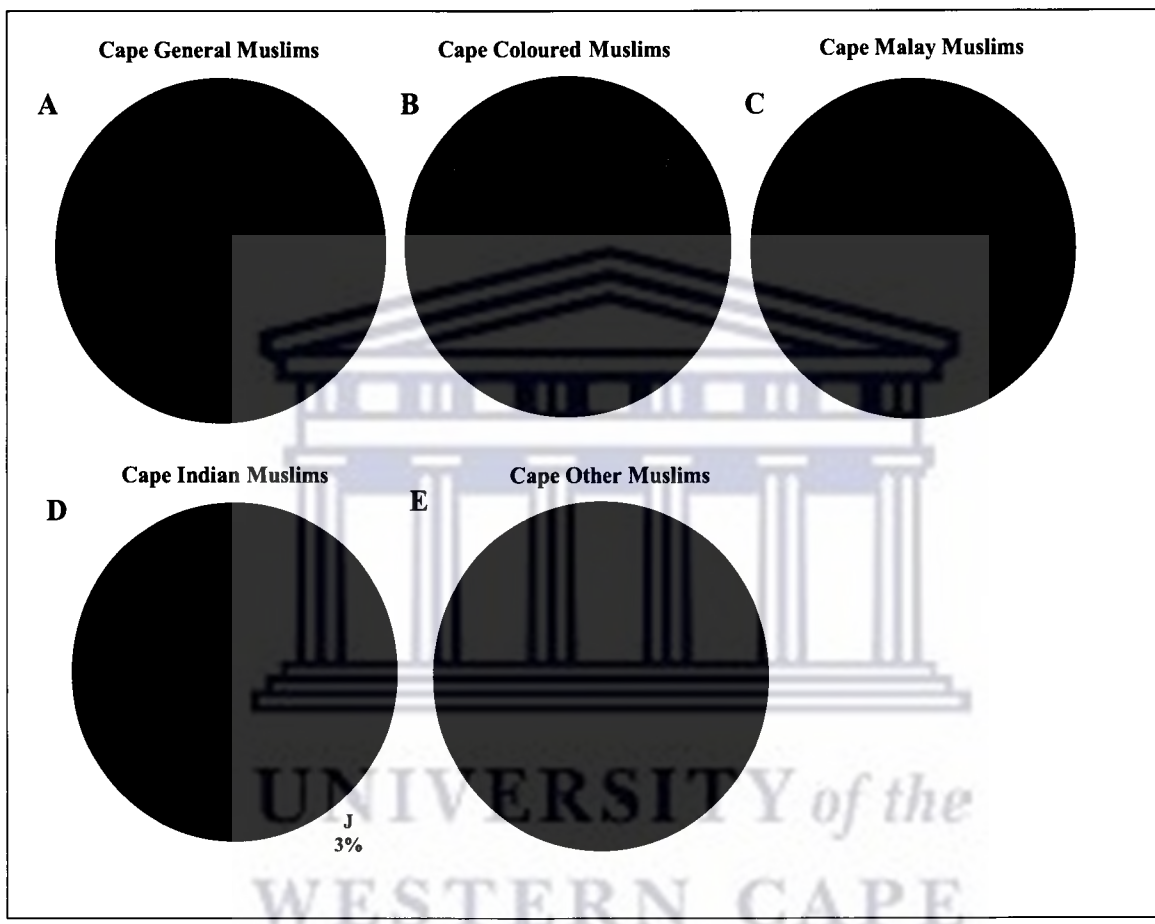


Figure 2.8. (A) MtDNA haplogroups observed in the Cape General Muslims. (B-E) MtDNA haplogroups observed in each self-perceived ethnic group.

Table 2.4. Distribution of mtDNA haplogroups in the different ethnic groups deduced from the questionnaire data

Factor	Ethnic groups	No of individuals	No of haplogroups	B	No of individuals in each Haplogroup H	J	L	M
Conversion	Coloured	4	3		1		2	1
	Cape Malay	0						
	Indian	0						
	Other	0						
Factor	Ethnic groups	No of individuals	No of haplogroups	B	No of individuals in each Haplogroup H	J	L	M
Internal Immigration	Coloured	3	3		1		2	
	Cape Malay	3	3			2	1	
	Indian	1	1		1			
	Other							
Factor	Ethnic groups	No of individuals	No of haplogroups	B	No of individuals in each Haplogroup H	J	L	M
External Immigrants	Coloured	0						
	Cape Malay	0						
	Indian	20	4		7	1	2	10
	Other	1	1		1			
Factor	Ethnic groups	No of individuals	No of haplogroups	B	No of individuals in each Haplogroup H	J	L	M
Inter-ethnic marriage	Coloured	8	3		2		4	2
	Cape Malay	5	2				3	2
	Indian	2	2					2
	Other							

* Ethnic identity is listed as perceived by donor

Haplogroup M was the second most frequent haplogroup in the study. It is common among Asian populations and is the most dominant mtDNA haplogroup in India (Schurr and Wallace 1999; Kivislid et al. 1999; Roychoudhury et al. 2000; Torroni et al. 2001; Edwin et al. 2002; Cordaux et al. 2003; Tripathy et al. 2008, Maji et al. 2009). A study by Cordaux et al (2003) reported the haplogroup's frequency among Northeast tribes as being 51% while East, Central and South tribes shared an equal frequency of 76%. This result was consistent with findings in the study as Cape Indian Muslims displayed the highest frequency (45%). Furthermore, nearly 34% of

haplgroup M frequency could be attributed to recent external immigrants from India (Table 2.4).

A moderate frequency for haplogroup M was observed among the Cape Coloured Muslims (32%), Cape Malay Muslims (22%), and Cape Other Muslims (30%). Only a small fraction of their mtDNA was derived through conversion and inter-ethnic marriages (Table 2.4). The origin of their mtDNA was likely mainly derived from female slaves originating from India and South East Asia as a significant number of Asian slaves originated from India and the Indonesian Archipelago (Bradlow and Cairns 1978; Davids 1980; Da Costa and Davids 1994). Although, haplogroup M may even have been introduced via slaves from Madagascar due to their mixed Indonesian ancestry and recently through admixture with Indian and other Asian populations (Hurles et al. 2005).

West Eurasian mtDNA, haplogroup H and J were both observed in the study albeit at different frequencies. Haplogroup H was the third most frequent mtDNA (24%) in the study (Figure 2.8 A). Among the self-perceived ethnic groups, Cape Other Muslims displayed the highest frequency (50%) followed by Cape Indian Muslims (38%). The high frequency of haplogroup H in Cape Indian Muslims was mainly due to internal and external immigrants their contributions makeup 28% of the total haplogroup H diversity. A lower frequency for haplogroup J was observed mainly prevalent among the Cape Malay Muslims (15%) and Cape Coloured Muslims (12%). Cape Indian Muslims had the lowest frequency (3%), while Cape Other Muslims harboured no J mtDNA.

Haplogroup H and J may have been introduced via recent and historical admixture with European and Indian populations. Since haplogroup, H represents nearly 50% of European mtDna while haplogroup J represents 11.3% of European mtDNA (Torroni et al. 1996; Wallace et al. 1999; Lell and Wallace 2000; Maca-Meyer et al. 2001; Richards and Macaulay 2001; Bermisheva et al. 2002; Loogväli et al. 2004; Piechota et al. 2004; Alzualde et al. 2005; Manwaring et al. 2006). In addition, West Eurasian haplogroups makeup 20% to 30% of Indian mtDNA with Northern Indians possessing a higher frequency of West Eurasian mtDNA than Southern Indians (Roychoudury et al. 2000; Tripathy et al. 2008).

Haplogroup B was the only mtDNA not observed in the study. The 9bp deletion defining this haplogroup is routinely used to infer Asian and Polynesian ancestry (Ballinger et al. 1992; Stone and Stoneking 1998; Edwin et al. 2002; Berniell-Lee et al. 2008). Haplogroup B has a high to moderate frequency among East Indonesians and Malays (Stone and Stoneking 1998; Merriweather et al. 1999; Schurr and Wallace 2002). Therefore, the absence of this haplogroup was unexpected since a number of slaves originated from the Indonesian Archipelago (Bradlow and Cairns 1978; Davids 1980; Da Costa and Davids 1994). The absence of this haplogroup may possibly be attributed to the low sample number investigated in this study. Given that, only a hundred samples were screened in the study. On the other hand may indicate that few Cape Muslims are from Indonesian descent.

2.3.4. Y-Chromosome Haplogroup Variation

All eight bi-allelic markers analysed in this study were found among the Cape General Muslim population. Haplogroup frequencies is reported in Figure 2.9 (A), belonging to haplogroups K (43%), O (22%), H (10%), I (8%), R1 (5%), DE (3%), E3a (1%), E3b

(1%) while 7% remained undefined. Haplogroups K, O, and H were the most frequent in the study. Cape Malay Muslims displayed the highest frequency (52%) for haplogroup K (Figure 2.9 C). At least 27% of haplogroup K frequency was attributed to conversion, external-immigrants and inter-ethnic marriages (Table 2.5). Cape Indian Muslims though demonstrated a high frequency for haplogroups O (38%) and H (17%) (refer to Figure 2.9 D). Internal and external immigrants were responsible for at least 24 % of their haplogroup O frequency (Table 2.5). Other self-perceived ethnic groups also exhibited a high frequency for haplogroups K, O and H (refer to Figure 2.9 B, C and E).

Males originating from Asia probably introduced haplogroups K, O, and H given their geographic distribution. Haplogroup K is distributed throughout Asia having a substantial presence in East Asia (Basu et al. 2003; Kumar et al. 2007). Haplogroup O represents a major Y-chromosome lineage in East Asia and has a moderate or low frequency in Central Asia, South East Asia, and Oceania (Cordaux et al. 2004; Karafet et al. 2005; Sahoor et al. 2006; Kumar et al. 2007; Karafet et al. 2008). Haplogroup H is restricted to India and is frequently found among tribal and lower caste population groups (Krithika et al. 2007; Karafet et al. 2008).

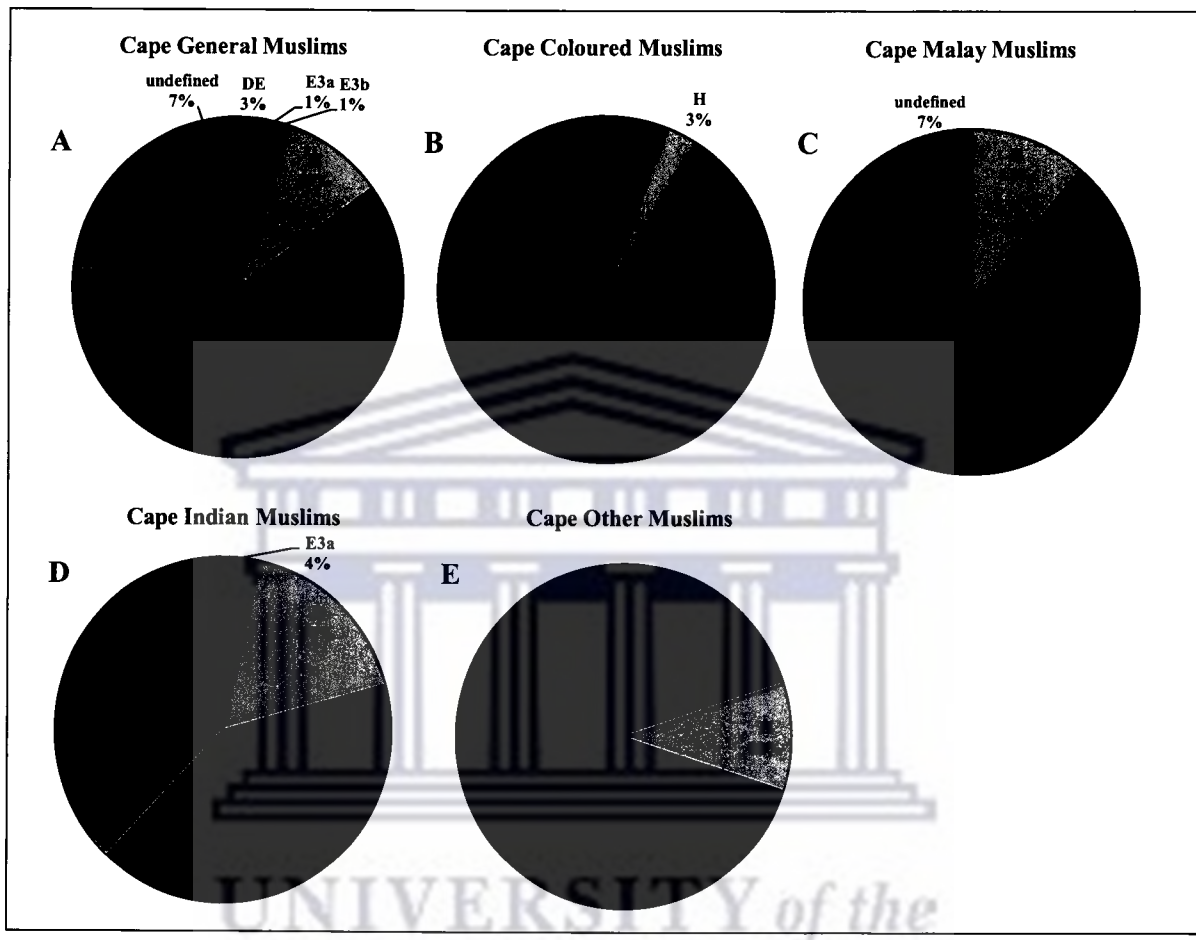


Figure 2.9. (A) Y-chromosome haplogroups observed in the Cape General Muslims. (B-E) Y-chromosome haplogroups observed in each self-perceived ethnic group

Table 2.5. Distribution of Y-chromosome haplogroups in the different ethnic groups deduced from the questionnaire data

Factor	Ethnic groups	No of individuals	No of haplogroups	DE	No of individuals in each Haplogroup: E3a	E3b	H	I	K	O	R1	undefined
Conversion	Coloured Cape Malay	4	2						3		1	
	Indian	0										
	Other	2	2						1		1	

Factor	Ethnic groups	No of individuals	No of haplogroups	DE	No of individuals in each Haplogroup: E3a	E3b	H	I	K	O	R1	undefined
Internal Immigrants	Coloured Cape Malay	0										
	Indian	3	3		1		1		1			
	Other	5	3			1			3		1	

Factor	Ethnic groups	No of individuals	No of haplogroups	DE	No of individuals in each Haplogroup: E3a	E3b	H	I	K	O	R1	undefined
External Immigrants	Coloured Cape Malay	5	2				1		4			
	Indian	2	2				1		1			
	Other	17	3				1		9	7		
		4	4	1			1		1	1		

Factor	Ethnic groups	No of individuals	No of haplogroups	DE	No of individuals in each Haplogroup: E3a	E3b	H	I	K	O	R1	undefined
Inter-ethnic marriage	Coloured Cape Malay	12	5					1	6	3	1	1
	Indian	4	2						3	1		
	Other	1	1				1					

* Ethnic identity is listed as perceived by donor

The European Y-chromosome haplogroups observed in the study were R1 and I. These haplogroups are widespread across Europe and are found at high frequencies in West Europe (Cavalli-Sforza and Feldman 2003; Rootsi et al. 2004; Fregel et al. 2009). Haplogroup I was the most frequent in the study (Figure 2.9 A), it observes its highest frequency (17%) in Cape Coloured Muslims (Figure 2.9 B), while R1 in Cape Other Muslims (10%) (Figure 2.9 E). These haplogroups were however, absent in Cape Indian Muslims. Early male European settlers possibly introduced haplogroup I and

R1. This result is common and frequently observed in populations that originate from slavery and colonization (Torres et al. 2007; Berniell-Lee et al. 2008).

Haplogroups DE, E3a and E3b represented the remaining Y-chromosome haplogroups. These haplogroups had the lowest frequencies in the study. The YAP element defining haplogroup DE was observed only in Cape Coloured Muslims (6%) and Cape Other Muslims (10%). This haplogroup defines both Asian (haplogroup D) and African (haplogroup E) Y-chromosome lineages. Nonetheless, observes its greatest frequency in Sub Saharan populations (Hammer and Horai 1995; Agrawal et al. 2005). It is therefore most likely that haplogroup DE was derived from an African source. Individual's harbouring the YAP element probably belonged to haplogroup E, considering that haplogroups, E3a and E3b were also found in the study. These haplogroups show a distinct geographical distribution in African populations.

Haplogroup E3a is common in Sub Saharan populations and is used to trace Bantu migrations (Hammer and Zegura 2002; Beleza et al. 2005; Fregal et al. 2009). The haplogroup was only observed in an individual classified as Cape Indian Muslim. It is probably the result of recent admixture considering that E3a was introduced through an internal immigrant from Durban (Table 2.5).

Haplogroup E3b was restricted to Cape Other Muslims defining 10% of their Y-chromosome haplogroup diversity (Figure 2.9 E). The haplogroup is frequent among East and North African populations (Hammer and Zegura 2002; Beleza et al. 2005). From the data listed in Table 2.5 it was established that this haplogroup was recently introduced possibly as result of recent admixture. However, could possibly be derived from an East African slave as many Muslims, originally from the Cape Metropolitan

area have migrated to areas such as Port Elizabeth, East London etc (Davids 1980; Da Costa and Davids 1994).

2.3.5. The overall geographic origins of the Cape Muslim population

The mtDNA and Y-chromosome haplogroups observed in the study indicated that Cape General Muslims maternal and paternal lineages originate in Africa, Asia and Europe. Thereby, demonstrating the highly diverse geographical origins of Cape General Muslims mtDNA and Y-chromosomes. From the 100 male samples analyzed in the study, 34% carried an African maternal lineage while the contribution of Asian and European lineages in the population was 33% in both cases (Figure 2.10 A).

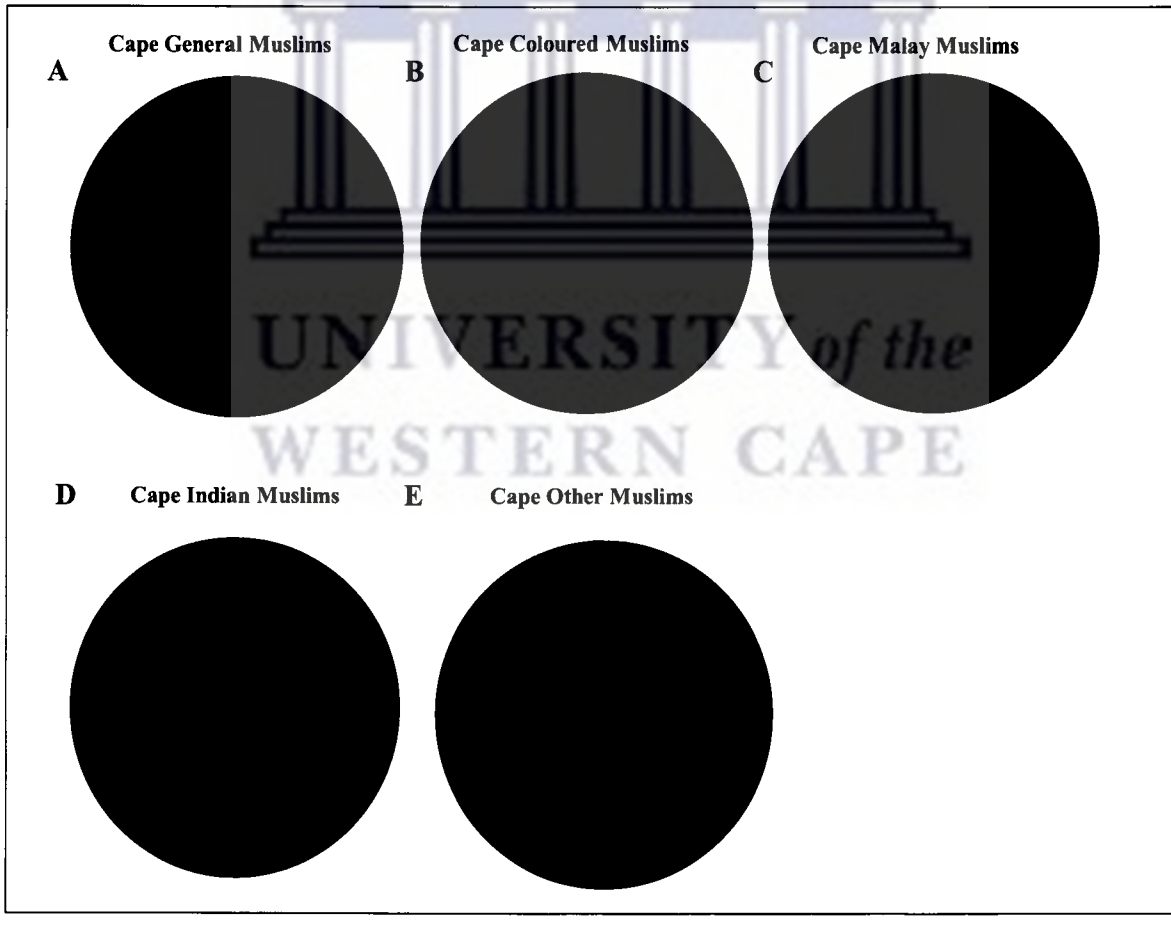


Figure 2.10 (A) Distribution of the major maternal lineages found in Cape General Muslims. (B-E) Distribution of the major maternal lineages found in each self-perceived ethnic group.

The continent-specific lineages were found at various frequencies in each of the self-perceived ethnic groups within the Cape General Muslim population. In the thirty-four donors classified as Cape Coloured Muslims 47% (16/34) belonged to an African lineage, 32% (11/34) to an Asian lineage and 21% (7/34) to European lineages (Figure 2.10 B). The twenty-seven Cape Malay Muslims in the study show a similar frequency for these lineages as 45% (12/27) harboured African mtDNA, 22% (6/27) Asian and 33% (9/27) European mtDNA (Figure 2.10 C).

In contrast to Cape Coloured Muslims and Cape Malay Muslims, Cape Indian Muslims maternal lineages significantly differed. Their mtDNA indicated that 14% (4/29) carried an African lineage, 45% (13/29) Asian and 41% (12/29) European lineages (Figure 2.10D). This correlates well with studies regarding mtDNA variation in India as Indian females harbour both Asian specific and West European mtDNA haplogroups (Kivislid et al. 1999; Cordaux et al. 2004). Therefore, it is quite possible that the high frequency of European lineages observed in Cape Coloured Muslims and Cape Malay Muslims were perhaps derived from Indian female slaves. The presence of African lineages in Cape Indian Muslims was probably the result of recent admixture with Cape Coloured Muslims and Cape Malay Muslims or African immigrant tribal groups in India. Considering, that indigenous Indian populations do not harbour African mtDNA and that external immigrants from India significantly contributed to mtDNA variation of Cape Indian Muslims.

Cape Other Muslims maternal lineages were similar to Cape Indian Muslims as 20% (2/10) belonged to an African lineage, 30% (3/10) to Asian and 50% (5/10) to European lineages (Figure 2.10 E).

The results for maternal lineages for the most part showed a good concordance between the self-perceived ethnic identity by donors and the origins of their mtDNA haplogroups with the exception of Cape Malay Muslims. Their mtDNA showed a higher frequency for African and European lineages rather than the expected Asian lineages.

Paternal lineages displayed a contrasting genetic structure to maternal lineages whereby 5% of male lineages were of African origin, 75% Asian, 13% European, and 7% of lineages remained undefined (Figure 2.11 A). The undefined male lineages in the study may have belonged to either Asian or European Y-chromosome haplogroups, as these paternal lineages were the most frequent in the study in comparison to African lineages.

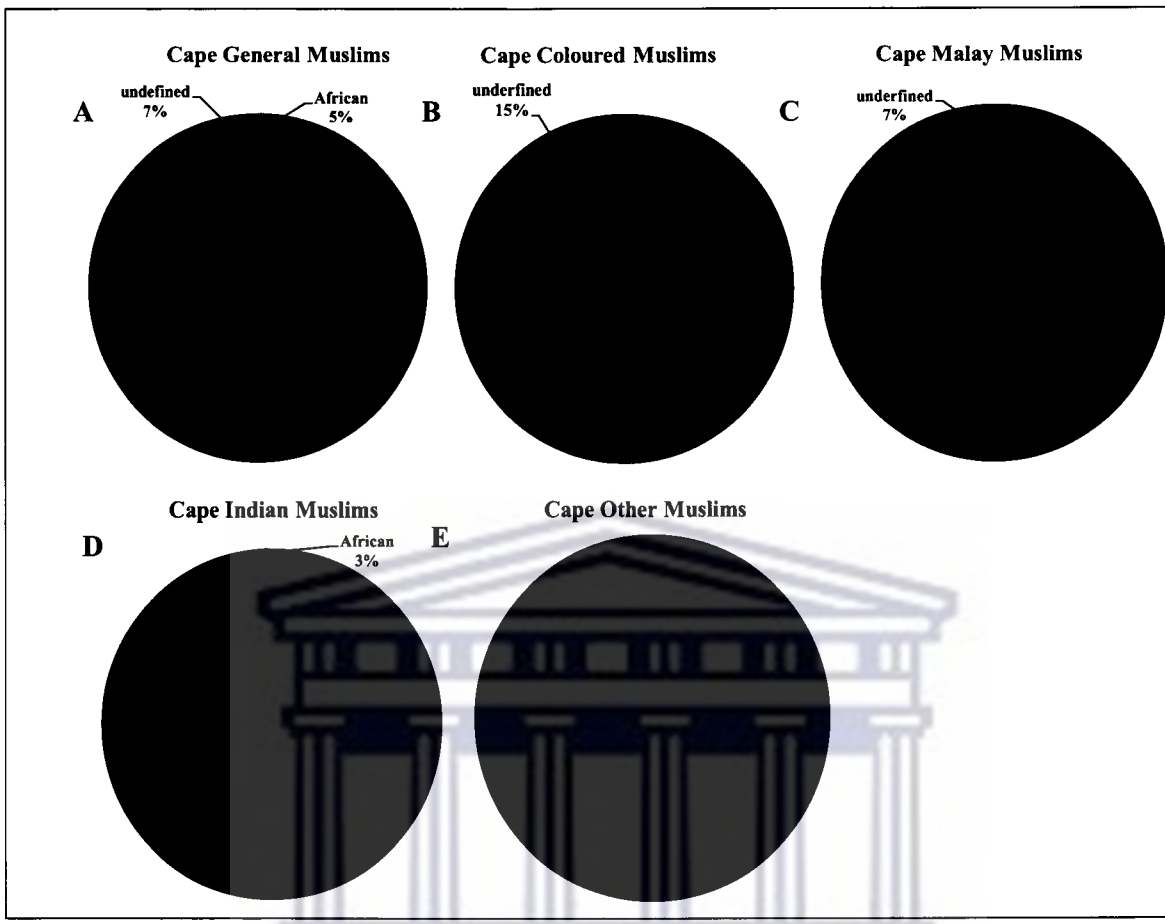


Figure 2.11 (A) Distribution of the major paternal lineages found in Cape General Muslims. (B-E) Distribution of the major paternal lineages found in each self-perceived ethnic group.

Similar to maternal lineages self-perceived ethnic groups demonstrated various frequencies for certain paternal lineages. Although, all the self-perceived ethnic groups shared a high frequency for Asian lineages. In Cape Coloured Muslims 6% (2/34) belonged to African lineages, 56% (19/34) to Asian, 23% (7/34) to European lineages and 15% of paternal lineages remained undefined (Figure 2.11 B). Cape Malay Muslims paternal lineages indicated that 78% (21/27) were of Asian descent, 15% (4/27) European and 7% (2/27) of lineages remained undefined (Figure 2.11 C). Although, notable differences exist between the Cape Coloured Muslims and Cape Malay Muslims paternal structures. These include a lower frequency of Asian lineages,

a higher frequency for European male lineages and the presence of African Y lineages. These two ethnic groups paternal structure yet seem more closely related in comparison to Cape Indian Muslims.

Cape Indian Muslims paternal lineages indicated that 3% (1/29) carried an African lineage and 97% (28/29) Asian lineages (Figure 2.11 D). These differences were attributed to the fact that Cape Indian Muslims were mainly the descendants of recent Indian immigrants as indicated by data listed in Table 2.5. Therefore have undergone considerable less admixture than Cape Coloured Muslims and Cape Malay Muslims.

Cape Other Muslims paternal lineages were 20% (2/10) African, 70% Asian (7/10) and 10% (1) European (Figure 2.11E). Their paternal lineage composition seems to be closely related to both Cape Coloured Muslims and Cape Malay Muslims (Figure 2.11 B and C). Given the presence of African Y-chromosome lineages and the high frequency of Asian lineages

Overall, paternal lineages seem to provide a more accurate correlation of self-perceived ethnic identity of donors and the origins of their Y-chromosome haplogroups in comparison to mtDNA haplogroups. This could possibly provide an explanation as to why certain donors perceive themselves as Cape Malay Muslims and Cape Indian Muslims even though their mtDNA belong to non-Asian haplogroups.

In conclusion, this study was able to identify the major maternal and paternal lineages present in the Cape General Muslim population. Maternal lineages of the Cape General Muslims indicated that females of African, Asian and European ancestry equally contributed to the mtDNA genepool. Whereas paternal lineages indicated males of

Asian ancestry mainly contributed to the Y-chromosome genepool. The results also demonstrated a higher contribution of African maternal lineages in comparison to paternal lineages. It is therefore likely that the slave population brought from Asia were mainly males, these males mainly married indigenous African females given the high frequency of African maternal lineages in Cape Coloured Muslims and Cape Malay Muslims.

Chapter 3

MtDNA haplogroup analysis of the Kensington Muslim community

3.1. Introduction

Kensington lies 8km from Cape Town and is situated in the northern suburbs between Maitland and Goodwood. The area is one of the oldest Coloured communities yet its geographical location has isolated it from the rest of the Cape Flats (Da Costa 1983). In addition, both the Cape Flats and Kensington are two of forty –four areas in Cape Town having a significant Muslim population (Matthee 2008). A significant proportion of this population is made up of descendants of slaves brought to Cape Town during the 16th and 18th century (Bradlow and Cairns 1978; Davids 1980; Da Costa and Davids 1994; Matthee 2008). The slaves came from various geographical regions in Asia and Africa and along with indigenous African people and colonizing Europeans contributed to the gene pool of the Cape Muslims (Bradlow and Cairns 1978; Davids 1980; Da Costa and Davids 1994; Matthee 2008). The extent of these contributions to the maternal ancestry of the Cape General Muslim population was examined in a study (Chapter 2). The study analysed mtDNA coding region polymorphisms (SNPs) defining five mtDNA haplogroups. Only four mtDNA haplogroups were observed in the study and revealed that the Cape Coloured Muslims and Cape Malay Muslims maternal ancestry were mainly of African origin while Cape Indian Muslims maternal ancestry were primarily of Asian origin. The aim of this study (described in this chapter) was to determine the contribution of major African, Asian and European haplogroups to the maternal gene pool of the Kensington Muslim population. This was achieved by examining mtDNA variation found within the Kensington Muslim Community using twelve mtDNA coding region SNPs genotyped by the SNaPshot minisequencing assay. The assay provides an alternative to the time consuming PCR-

RFLP method allowing for a rapid, robust and simultaneous genotyping of multiple SNP markers (Quintáns et al. 2004; Budowle 2004; Sobrina et al. 2004).

3.2. Material and Methods

3.2.1. Biological Samples and DNA extraction

DNA samples were obtained from unrelated Muslim males residing in Kensington. Informed consent was obtained for all the samples used in the study, which had been approved by the University of the Western Cape Ethics Committee. An anonymous survey regarding place of birth, religious affiliation, home language, and ethnicity was also completed by donors. Additional information about maternal and paternal family members was also provided.

DNA was extracted, quantified, and stored until further analysis as described in Chapter 2 (section 2.2.3). A hundred and ten samples were successfully genotyped.

3.2.2. Questionnaire

The data obtained from the Kensington Muslim population was used alongside the results from the genetic study as described in Chapter 2 (section 2.2.2).

Kensington General Muslims: this term was used throughout the thesis and refers to Muslim individuals residing within the Kensington area.

The terms **conversion**, **internal immigrants**, **external immigrants,** and **inter-ethnic marriages** are also used throughout the thesis. The definitions of these terms were as previously described in Chapter 2 (section 2.2.2).

3.2.3. MtDNA SNP Selection

The relevant markers were selected from a SNP assay developed by Nelson et al. (2007). Based on the data generated from Chapter 2 (section 2.3.3). The assay allowed

for the genotyping of multiple haplogroups using twelve specific mtDNA SNPs. The assay has a good discrimination power that allows distinguishing major African, Asian, and European haplogroups found in the population. A summary of the expected haplogroup genotypes is illustrated in (Table 3.1). Although the assay was developed to simultaneously type for the twelve polymorphic sites, markers however were procured at various stages resulting in two multiplexes. An unforeseen advantage of the two multiplexes allowed mtDNA haplogroups to be typed following a hierarchical strategy.

Multiplex I contained the following polymorphic sites: 8272-8280 del, 10398, 10400, 3594, 7028, 12406, and 7600. The presence of one or a combination of these markers led to the assignment of the following haplogroups: B, E, F, H, L1/L2, L3, M, and N representing more than 50% of the haplogroups typed using the complete assay (Figure 3.1).

Multiplex II contained polymorphic sites 13263, 1719, 5178, 663 and 4833 which were only typed following the outcome of Multiplex I results. Multiplex II primarily contained downstream markers exploring the more basal branches of major haplogroups present in Multiplex I. Haplogroups C, D, G, I, and X were determined using Multiplex II (Nelson et al. 2007).

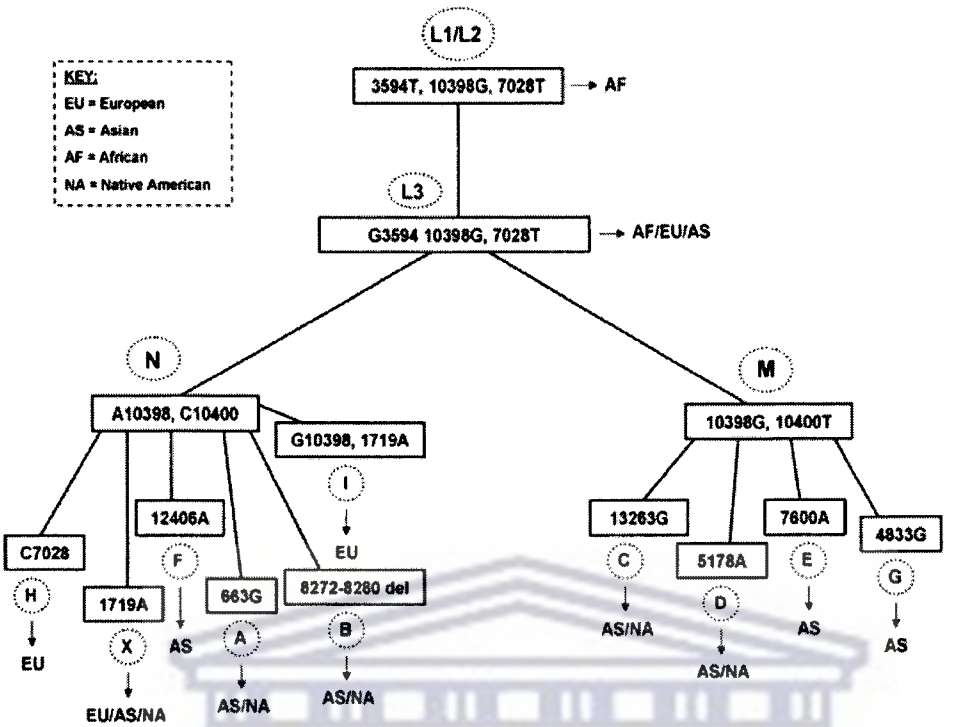


Figure 3.1. A schematic representation of the SNP typing strategy used in the study (adapted from Nelson et al. 2007).

Table 3.1: The expected base substitutions inferring maternal ancestry for each haplogroup

	Nucleotide position and base substitution												
Haplogroup	**8272-8280 del**	**13263**	**1719**	**5178**	**663**	**10398**	**10400**	**3594**	**7028**	**12406**	**4833**	**7600**	**Inferred ancestry**
A	C	A	G	C	**B**	A	C	C	T	G	A	G	Asian
B	**G**	A	G	C	**A**	A	C	C	T	G	A	G	Asian
C	C	**G**	G	C	A	**G**	T	C	T	G	A	G	Asian
D	C	A	G	**A**	A	**G**	T	C	T	G	A	G	Asian
E	C	A	G	C	A	**G**	T	C	T	G	A	**A**	Asian
F	C	A	G	C	A	A	C	C	T	**A**	A	G	Asian
G	C	A	G	C	A	**G**	T	C	T	G	**G**	G	Asian
H	C	A	G	C	A	A	C	C	C	G	A	G	European
I	C	A	**A**	C	A	**G**	C	C	T	G	A	G	European
L1/L2	C	A	G	C	A	**G**	C	T	T	G	A	G	African
L3	C	A	G	C	A	**G**	C	C	T	G	A	G	African
M	C	A	G	C	A	**G**	T	C	T	G	A	G	Asian
N	C	A	G	C	A	A	C	C	T	G	A	G	European/Asian
X	C	A	**A**	C	A	A	C	C	T	G	A	G	European/Asian

Bold letters indicate the expected base substitution (adapted from Nelson et al. 2007)

3.2.4. Multiplex PCR Amplification of Mitochondrial SNP sites

Multiplex PCR was performed using the Qiagen Multiplex kit in a final volume of 10 µl containing 15ng template DNA ,1X QIAGEN Multiplex PCR Master Mix , 0.5 X Q-Solution (*Qiagen*) and the respective multiplex primer premix. A detailed guide of the primer mix can be found in Appendix I. Primer concentrations for each Multiplex is given in (Table 3.2). Primers were synthesized by *Whitehead Scientific* using previously reported sequences (Nelson et al. 2007).

Amplifications were performed in a GeneAmp^{R} 2700 PCR System thermal cycler (*Applied Biosystems*). The thermal cycling conditions were as follows: 1 cycle at 95°C for 15 min, 40 cycles of 94°C for 30 sec, 60°C for 1 min and 60 sec, and 72°C for 1 min and 60 sec, followed by a final cycle at 72°C for 10 min.

PCR products were verified by agarose gel electrophoresis to ensure that the correct amplicon size was obtained. An aliquot of 2µl of PCR product was run on a 3% agarose gel stained with 0.1mg/ml ethidium bromide.

Table 3.2: MtDNA multiplex I and II amplicon product size, and final primer concentrations

Nucleotide position	Primer sequence (5'-3')	Primer length (bp)	Amplicon length (bp)	Final (µM)
Multiplex I				
8272-8280 del	Fw TAAAAATCTTTGAAATAGGGCCC	23	89 (del) 80	0.35
	Rv GTTAATGCTAAGTTAGCTTTACAGTGG	27		
10398/10400	Fw AGTCTGGCCTATGAGTGACTAC	22	86	0.35
	Rv AATGAGTCGAAATCATTCGTTT	22		
3594	Fw CTTAGCTCTCACCATCGCTCT	21	90	0.35
	Rv AGAATAAATAGGAGGCCTAGGTTG	24		
7028	Fw TATTAGCAAACTCATCACTAGACATCGT	28	96	0.35
	Rv TGGCAAATACAGCTCCATTGA	22		
12406	Fw AATTCCCCCCATCCTTACC	19	78	0.35
	Rv GCGACAATGGATTTTACATAATG	23		
7600	Fw GGCTAAATCCTATATATCTTAATGGCA	27	64	0.35
	Rv GGGAAGTAGCGTCTTGTAGACC	22		
Multiplex II				
13263	Fw CAAAAAAATCGTAGCCTTCTCC	22	67	0.35
	Rv GTTGATGCCGATTGTAACTATTATG	25		
1719	Fw CCCACTCCACCTTACTACCAGA	22	84	0.35
	Rv TGCGCCAGGTTTCAATTT	18		
5178	Fw TAAACTCCAGCACCACGACC	20	79	0.35
	Rv GTGGATGGAATTAAGGGTGTTAG	23		
663	Fw ACATCACCCCATAAACAAATAGG	23	108	0.35
	Rv TGGTGATTTAGAGGGTGAACTCA	23		
4833	Fw AATAGCCCCCTTTCACTTCTG	21	72	0.35
	Rv AGAAGAAGCAGGCCGGA	17		

Table and primers were adapted from Nelson et al. 2007

3.2.5. SNaPshot Minisequencing Reactions

Following verification, purification of PCR products were performed using *USB* Shrimp Alkaline Phosphatase (SAP) and Exonuclease I (ExoI) to remove excess primers and un-incorporated dNTPs. Aliquots of 5µl PCR product were incubated with 2µl (1unit/µl) of SAP and 2µl (10 units/µl) Exo I enzymes at 37˚C for 60 min followed by the inactivation of the enzymes at 75˚C for 15 min.

Minisequencing reactions were carried out in a final reaction volume of 13.5µl containing 2µl SNaPshot ™ Multiplex Kit (*Applied Biosystems)*, 2µl purified PCR product and 5.35µl (Multiplex I) or 2.77µl (Multiplex II) extension primers (final concentrations listed in Table 3.3). The orientation of each extension primer is given according to the SNP location (forward or reverse). Primers were manufactured by *Whitehead Scientific* using previously reported sequences (Nelson et al. 2007).

Minisequencing reactions were performed in a GeneAmp^{R} 2700 PCR System thermal cycler (*Applied Biosystems*) according to the manufactures specifications: 25 cycles of 96ºC for 10 seconds, 50ºC for 5 s, and 60º C for 30 s, and storage at 4ºC. In addition, a positive and negative control (provided with the kit) was performed for every minisequencing batch of samples. Minisequencing products were purified with the addition of 2µl of SAP (USB) to remove the unincorporated fluorescently labelled ddNTP's. Reactions were briefly mixed then incubated at 37˚C for 60 min followed by enzyme inactivation at 75˚C for 15 min.

Table 3.3. MtDNA minisequencing primer sequences, length, base substitution and final primer concentration

Nucleotide position	Primer sequence (5'-3')	Primer length (bp)	Final (μM)	Base substitution
Multiplex I				
8272-8280 del	**Fw CCCTATAGCACCCCCTCTA**	**19**	**0.37**	**C>G**
10398	**Fw (21-POLY-T-TAIL) GAGTGACATACAAAAAGGATTAGACTGA**	**48**	**0.259**	**A>G**
3594	**Rv (29-POLY-T-TAIL) TAGGAGGCCTAGGTTGAGGTT**	**50**	**0.259**	**C>T**
10400	**Rv (24-POLY-T-TAIL) TTCGTTTTGTTTAAACTATATACCAATTC**	**53**	**0.296**	**C>T**
7028	**Rv (33-POLY-T-TAIL) CCTATTGATAGGACATAGTGGAAGTG**	**59**	**0.277**	**C>T**
12406	**Fw (50-POLY-T-TAIL) CCCATCCTTACCACCCTC**	**68**	**0.259**	**G>A**
7600	**Fw (51-POLY-T-TAIL) TATCTTAATGGCACATGCAGC**	**72**	**0.259**	**G>A**
Multiplex II				
13263	**Fw (3-POLY-T-TAIL) TAGCCTTCTCCACTTCAAGTCA**	**25**	**0.129**	**A>G**
1719	**Fw (7-POLY-T-TAIL) CTCCACCTTACTACCAGACAACCTTA**	**33**	**0.259**	**G>A**
5178	**Fw (13-POLY-T-TAIL) CTACTATCTCGCACCTGAAACAAG**	**37**	**0.149**	**C>A**
663	**Fw (19-POLY-T-TAIL) CCATAAACAAATAGGTTTGGTCCT**	**43**	**0.155**	**A>G**
4833	**Fw (54-POLY-T-TAIL) TTTCCAGAGGTTACCCAAGGC**	**72**	**0.333**	**A>G**

Minisequencing primers and table adapted from Nelson et al. 2007

3.2.6. Analysis on the ABI 3130xl Genetic Analyzer

The minisequencing products (1.5 μl) were mixed with 9.6μl HiDi ™ Formamide and 0.4μl of Genescan ™ -120 LIZ ™ internal sizing standard (*Applied Biosystems*) and capillary electrophoresis was performed on an ABI Prism 3130XL ®Genetic Analyzer (*Applied Biosystems*). Following denaturation of samples at 95°C for 5 minutes in a GeneAmp® PCR System 9600 thermal cycler (*Applied Biosystems).* Using a 36cm capillary array filled with Performance Optimum Polymer 7 (POP 7) (*Applied Biosystems).* Typical run module parameters were run temperature 60°C, pre-run voltage 15kV, pre-run time 180 s, injection voltage 1.2kV, and injection time 23s, run voltage 15kV, and run time 1200 s. The filter set E5 was used to process data from the

five dyes dR110, dR6G, dTAMRA, dROX and LIZ 120 while data analysis was performed using Genemapper v4.0 software (*Applied Biosystems*).

3.3. Results and Discussion

3.3.1. Multiplex PCR

Initial multiplex PCR reactions included reverse touch down (as stated by Nelson et al. 2007), gradient, and hot-start PCR techniques but repeatedly failed to amplify all of the SNP fragments in each multiplex. This was only achieved once multiplex PCR was performed using the QIAGEN Multiplex PCR Master Mix kit. However, analyses of both multiplex PCR products revealed the presence of artifacts. These artifacts may have been the result of using inexpensive non-HPLC purified oligonucleotides instead of the recommended purified oligonucleotides. Since the reduction of DNA concentration, primer concentration and the increase in annealing temperature, reaction volume showed no significant reduction in the amount of artifacts. This nonetheless did not interfere with the SNaPshot minisequencing reactions. Electropherograms of these profiles showed clear distinct peaks of the expected SNP without the presence of these artifacts.

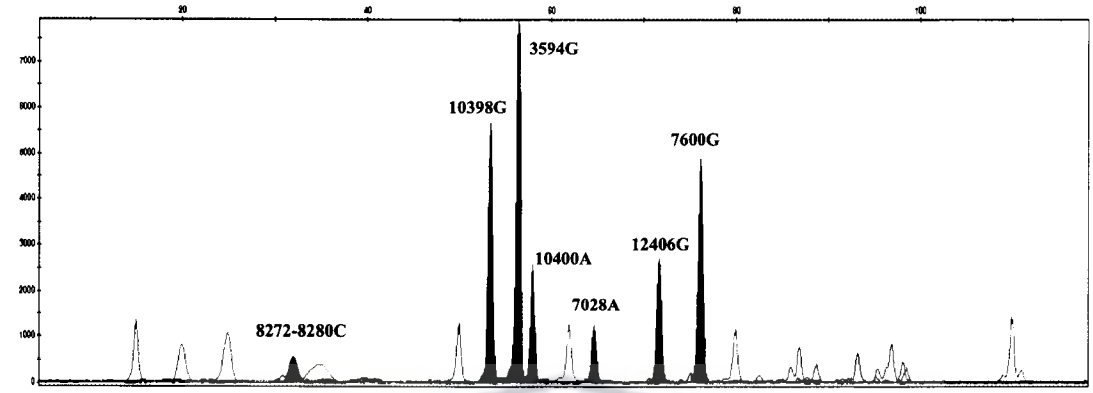
3.3.2. MtDNA Genotyping Procedure

All 110 samples were successfully genotyped using multiplex I and II with Figures 3.2. and 3.3. representing electropherograms of samples positive for haplogroup D and L3 respectively. The SNaPshot multiplex kit provides the following assignment of fluorescent dyes for each individual ddNTP: dR6G= ddATP (Green), dTAMRA = ddCTP (Black), dR110 = ddGTP (Blue) and dROX = ddTTP (Red). This indicates which base substitution occurs for each SNP during the minisequencing reaction. If however, an extension primer was designed in the reverse orientation such as 10400 (C>T) the reverse complement nucleotide was incorporated during the minisequencing

reaction, this is illustrated in (Figure 3.2.). Therefore, a simple conversion to the correct nucleotide base was applied when typing these SNPs. During genotyping, it was also observed that the detected size of minisequencing products by the automated sequencer and those listed by Nelson et al. (2007) did not correspond. Factors influencing size variation as reported by Quintáns et al. (2004) include differences in electrophoresis mobility determined by polymer used, length, sequence and the dye used to label the minisequencing primer. A greater size variation occurs for short oligonucleotides rather than longer ones. This may have been the cause as minisequencing primers listed in the published article by Nelson et al. (2007) were longer than the actual size of primers. In addition, electrophoresis of minisequencing fragments was performed using POP7 instead of the recommended POP4. Both systems show good resolution however, POP4 is known to have a shorter resolution time (Quintáns et al. 2004). Peaks above 400 relative fluorescence units were considered positive signals and a SNP haplogroup was assigned accordingly.

Multiplex I

Haplogroup D



Multiplex II

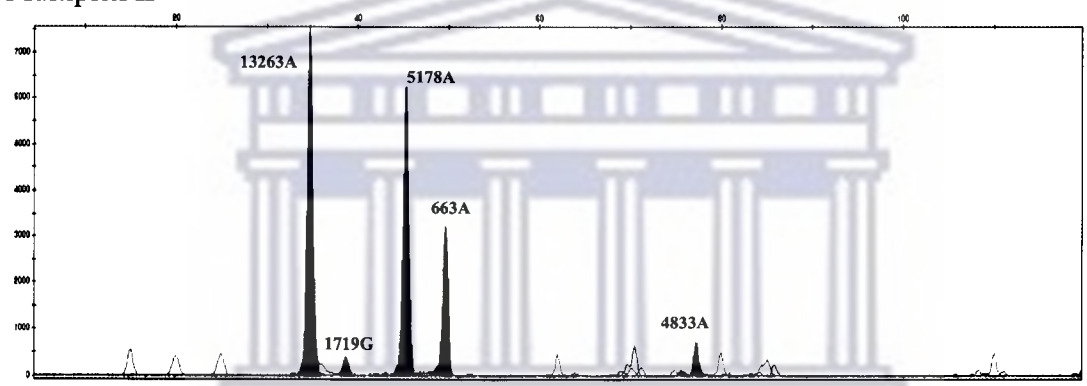
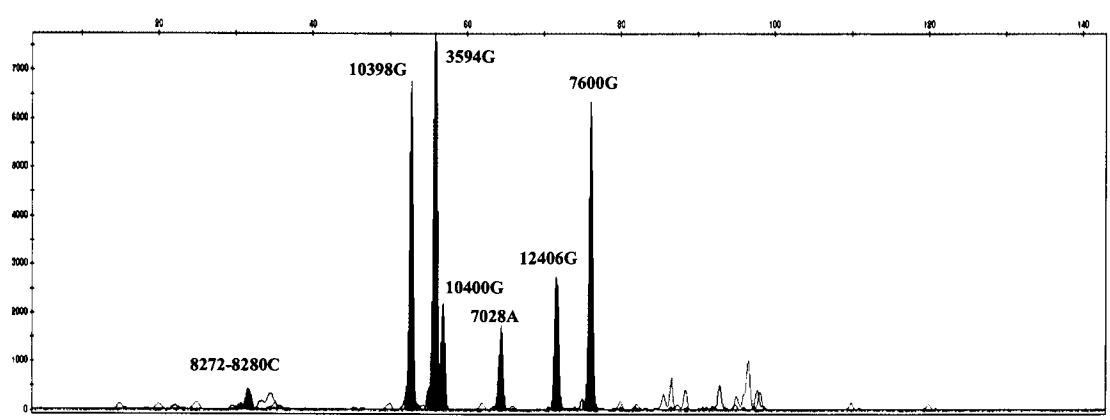


Figure 3.2. Electropherogram of multiplex I and II showing the derived state for haplogroup D defined by SNP sites 10398,10400,7028 and 5178

Multiplex I

Haplogroup L3



Multiplex II

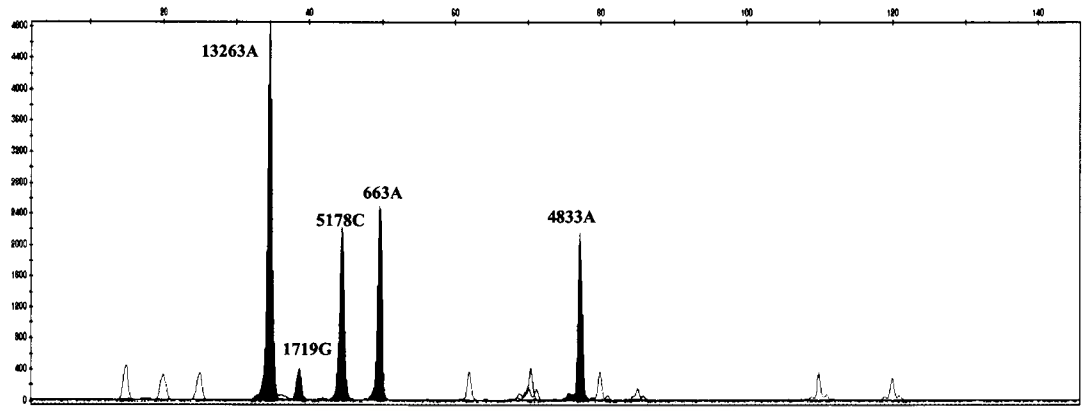


Figure 3.3. Electropherograms of multiplex I and II showing the derived state for haplogroup L3 defined by SNP sites 10398 and 7028.

3.3.3. Self-Perceived Ethnic Classification

Samples of a hundred and ten donors were used in the present study. The questionnaire indicated that donors mainly classified themselves as belonging to three ethnic classifications: 76% (84/110) were Coloured while 17% (18/110) Malay and 6% (7/110) Indian Muslims (Figure 3.4). Only 1% (1/110) was reluctant to assign an ethnic classification the term Other Muslims was later used for this individual.

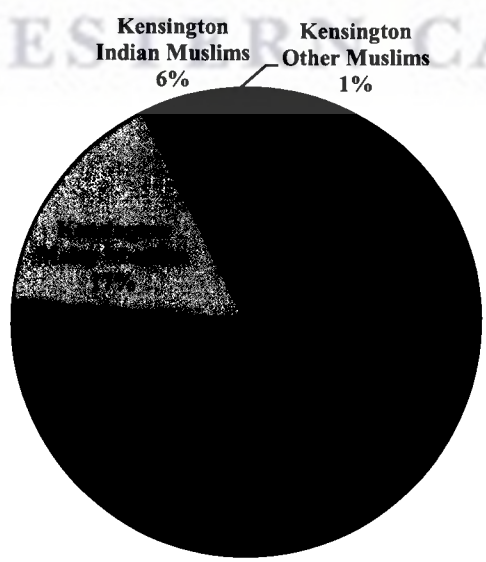


Figure 3.4. Ethnic groups identified in Kenisington General Muslims according to donor classification.

3.3.4. MtDNA Haplogroup Variation

MtDNA haplogroups observed in the Kensington General Muslims and in each self-perceived ethnic group is reported in Figure 3.5 A-D. Haplogroup L1/L2 was the most dominant mtDNA haplogroup in the study comprising 47% of the total sample (Figure 3.5 A). In the self-perceived ethnic groups, Kensington Malay Muslims had the highest frequency (61%) followed by Kensington Coloured Muslims (48%), Kensington Indian Muslims however had the lowest frequency (14%) (Figure 3.5 B, C and D).

Haplogroup L1 and L2 are restricted to Africa with Sub-Saharan populations having the highest frequency. Furthermore, haplogroup L1/L2 share the distinct SNP site 3594 (C>T) characterizing haplogroups of the major African macrohaplogroup L (Chen et al. 2000; Torroni et al. 2001; Salas et al. 2002). Given the high frequency of L1/L2 in Kensington General Muslims, it is quite possible that these haplogroups were introduced to the population as a result of inter-ethnic marriages between Asian male slaves and indigenous African females (Du Pre 1994, Da Costa and Davids 1994). Since Bantu and Khoisan populations, mtDNA mainly belong to subgroups of L1 and L2. Further characterization of L1/L2 subgroups would undoubtedly provide a more accurate account of the contributions made by different African populations.

A small percentage (17%) of L1/L2 mtDNA was contributed by converts, internal and external immigrants and inter-ethnic marriages (Table 3.4.). These factors however only influenced the frequency of L1/L2 mtDNA in Kensington Coloured Muslims. Three variants of haplogroup L1/L2 were also detected in the study, which was re-confirmed by re-extracting donor samples then subjecting it to SNaPshot analysis. The nine base deletion between 8272-8280 was observed in 11% (6/52) of L1/L2 mtDNA this has been previously reported by both Cann and Wilson (1983) and Wrischnik et al.

(1987) (Wallace et al. 1999). The 9bp deletion was initially thought to be Asian specific but later studies have indicated the independent and multiple origin of this marker (Soodyall et al. 1996; Cordaux et al. 2003). In Southern Africa, the 9bp deletion has only been observed in Bantu –speaking populations (Chen et al. 2000). The remaining L1/L2 variants were defined by an additional SNP site at 12406 and the absence of the 10398 site. The additional SNP site at 12406 defining haplogroup F may be a result of heteroplasmy. This event is extremely rare but may arise due to the following circumstances a) when more than one mtDNA is present in a tissue b) when an individual is heteroplasmic in one tissue but homoplasmic in another c) when one mtDNA type is expressed in one tissue while a different type in another (Budowle et al. 2003). The absence of the SNP site 10398 could possibly be the result of a back mutation known to occur for coding-region SNPs considered stable mutations (Behar et al. 2007).

The L3 haplogroup displayed a low frequency (8%) in the study mainly present in Kensington Coloured Muslims (8%) and Kensington Malay Muslims (5%) refer to Figure 3.5 B and C. Kensington Other Muslims mtDNA belonged to this haplogroup (data for this ethnic group not shown). This haplogroup observes its highest frequency in East Africa (Salas et al. 2002). Therefore may have been introduced by East African slaves as a significant number of African slaves originated from East Africa. However, were predominately male, which could possibly account for the low frequency of haplogroup L3 in the study (Du Pre 1994). The haplogroup may also have been introduced through recent admixture, as 2% of L3 mtDNA was introduced through converts and internal immigrants in Kensington Coloured Muslims while an external immigrant contributed the L3 mtDNA observed in the Kensington Malay Muslims (Table 3.4).

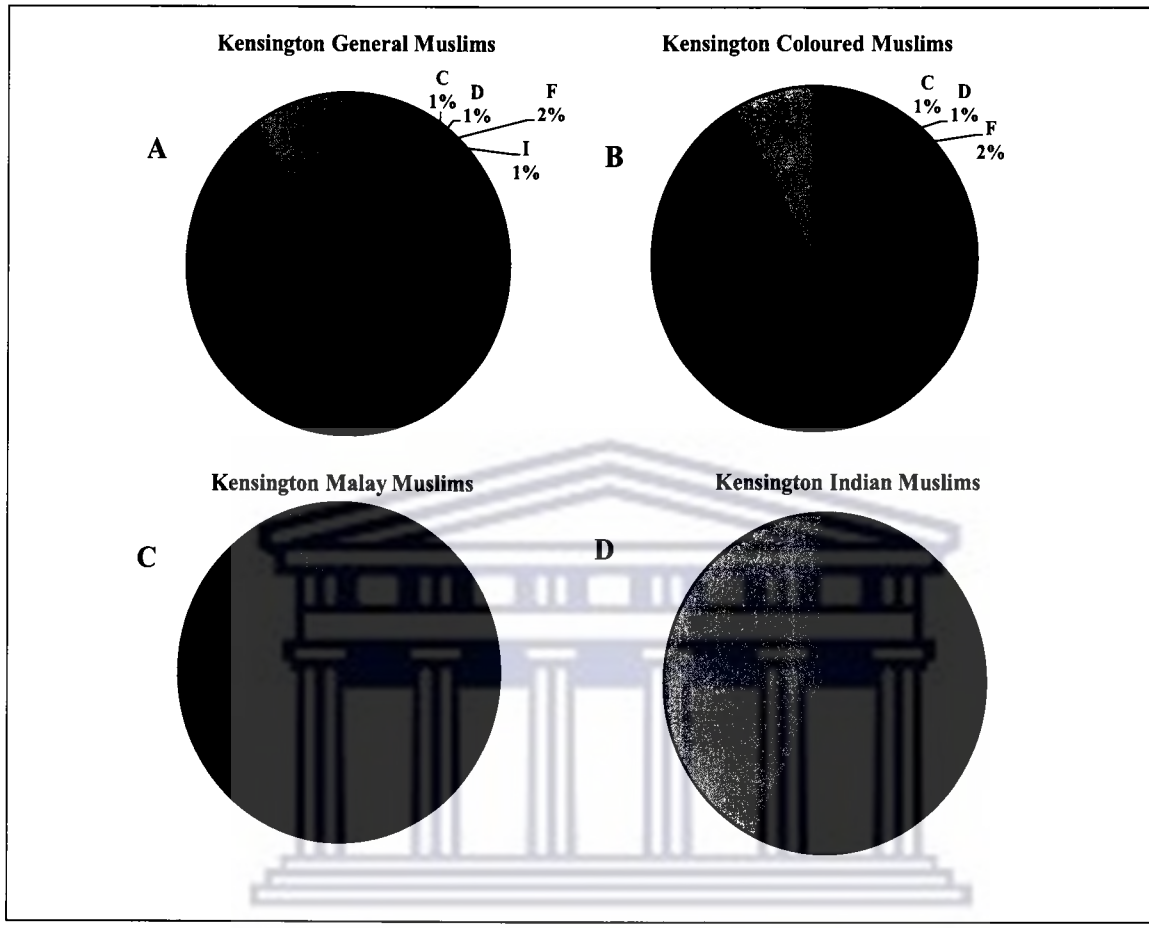


Figure 3.5. (A) MtDNA haplogroups observed in Kensington General Muslims. (B-D) MtDNA haplogroups observed in each self-perceived ethnic group.

Table 3.4. Distribution of mtDNA haplogroups in the different ethnic groups of Kensington General Muslims deduced from the questionnaire data

				No of individuals in each Haplogroup								
Factor	Ethnic groups	No of individuals	No of haplogroups	B	C	D	F	I	L1/L2	L3	M	N
Conversion	Coloured Cape	10	6				1		5	1	2	1
	Malay	0										
	Indian	0										
	Other	0										

				No of individuals in each Haplogroup								
Factor	Ethnic groups	No of individuals	No of haplogroups	B	C	D	F	I	L1/L2	L3	M	N
Internal Immigrants	Coloured Cape	8	8				1		3	1	2	1
	Malay	2	2	1								
	Indian	0				1						
	Other	0										

				No of individuals in each Haplogroup								
Factor	Ethnic groups	No of individuals	No of haplogroups	B	C	D	F	I	L1/L2	L3	M	N
External Immigrants	Coloured Cape	1	1						1			
	Malay	0	0							1		
	Indian	4	3					1			2	1
	Other	0										

				No of individuals in each Haplogroup								
Factor	Ethnic groups	No of individuals	No of haplogroups	B	C	D	F	I	L1/L2	L3	M	N
Inter-ethnic marriage	Coloured Cape	3	3						1			
	Malay	0	0									
	Indian	0										
	Other	0										

* Ethnic identity is listed as perceived by donor

The second most frequent haplogroup in the study was M (22%) which is commonly found throughout Asia and South East Asia (SEA). The Malays and Sabah Aborigines have the highest frequency in SEA. In India caste and tribal populations share a uniform distribution for the haplogroup (Roychoudery et al. 2000; Schurr and Wallace 2002; Maji et al. 2009). The haplogroup had a high frequency in Kensington Indian Muslims (29%) and Kensington Coloured Muslims (21%). In Kensington Indian Muslims haplogroup, M was attributed to recent external immigrants originating from

India (Table 3.4). However, in Kensington Coloured Muslims and Kensington Malay Muslims was likely derived from Asian female slaves from either India or Indonesia. Given that, only 6% of haplogroup M frequency in Kensington Coloured Muslims was contributed by converts and internal-immigrants (Table 3.4).

Other Asian haplogroups found in the study included haplogroup B, C, D, and F. Haplogroup B was observed in Kensington Coloured Muslims (9%) and Kensington Malay Muslims (11%) refer to (Figure 3.5 B and C). This haplogroup is frequent in SEA and Polynesia (Wallace et al. 1999; Kivisild et al. 1999; Ballinger et al. 2001; Tripathy et al. 2008). Therefore is likely attributed to slaves brought from the Indonesian archipelago. Considering that, north and south Indian populations have an extremely low 9bp deletion frequency (Tripathy et al. 2008). The low frequency of haplogroup B possibly suggests that few SEA slaves belonged to this haplogroup or that few Asian slaves originated from this region.

Haplogroups C and D collectively represented 2% of total mtDNA diversity and occur frequently among Tibetans, Koreans and Han Chinese (Wallace et al. 1999; Comas et al. 2004). These haplogroups are possibly the result of recent admixture with East Asian populations.

Haplogroup F was only observed in Kensington Coloured Muslims representing 2% of their total mtDNA diversity. Chinese in East Asia show a high frequency for the haplogroup while Vietnamese and the Orang Asli of Malaysia in Southeast Asia (Wallace et al. 1999; Kivisild et al. 1999; Tolk et al. 2001; Schurr and Wallace 2002). The haplogroup was introduced by a convert and an internal immigrant (Table 3.4) therefore is possibly the result of recent gene flow with either East or Southeast Asians.

The remaining haplogroups observed in the study were I and N. Haplogroup I was only found in Kensington Indian Muslims representing 14% of their total mtDNA while 2% of mtDNA diversity in the study. This haplogroup is probably the result of historical admixture as Western Eurasian haplogroups has a moderate frequency in India due to historical migration and invasions of West European populations (Lell and Wallace 2000; Piechota et al. 2004; Berniell-Lee et al. 2008).

Haplogroup N was carried by 43% of Kensington Indian Muslims, 10% Kensington Coloured Muslims and 6% Kensington Malay Muslims. This macro haplogroup encompass both Asian and European haplogroups (Lalueza-Fox et al. 2004; Underhill and Kivisild 2007; Maji et al. 2009). Therefore, the screening of Asian haplogroups (A, B and F) nested within haplogroup N suggested that individuals with an N mtDNA probably belonged to a Western European haplogroup. This though requires the use of more informative markers to establish which European mtDNA were harboured by these individuals.

3.3.5. The overall geographic origins of the Kensington Muslim Community

The mtDNA haplogroups observed in Kensington General Muslims revealed the geographic origins of maternal lineages, which are distributed in Africa, Asia, Europe and Eurasia (Figure 3.6 A). African maternal lineages were the most frequent in the study as 55% of donors harboured this lineage while 34% carried Asian, 1% European and 10% an Eurasian lineage. Kensington General Muslims maternal lineages differed from those observed in Cape General Muslims. Cape General Muslims displayed a higher frequency for European lineages (33%) and a lower frequency for African lineages (34%) (refer to Chapter 2, Figure 2.10 A).

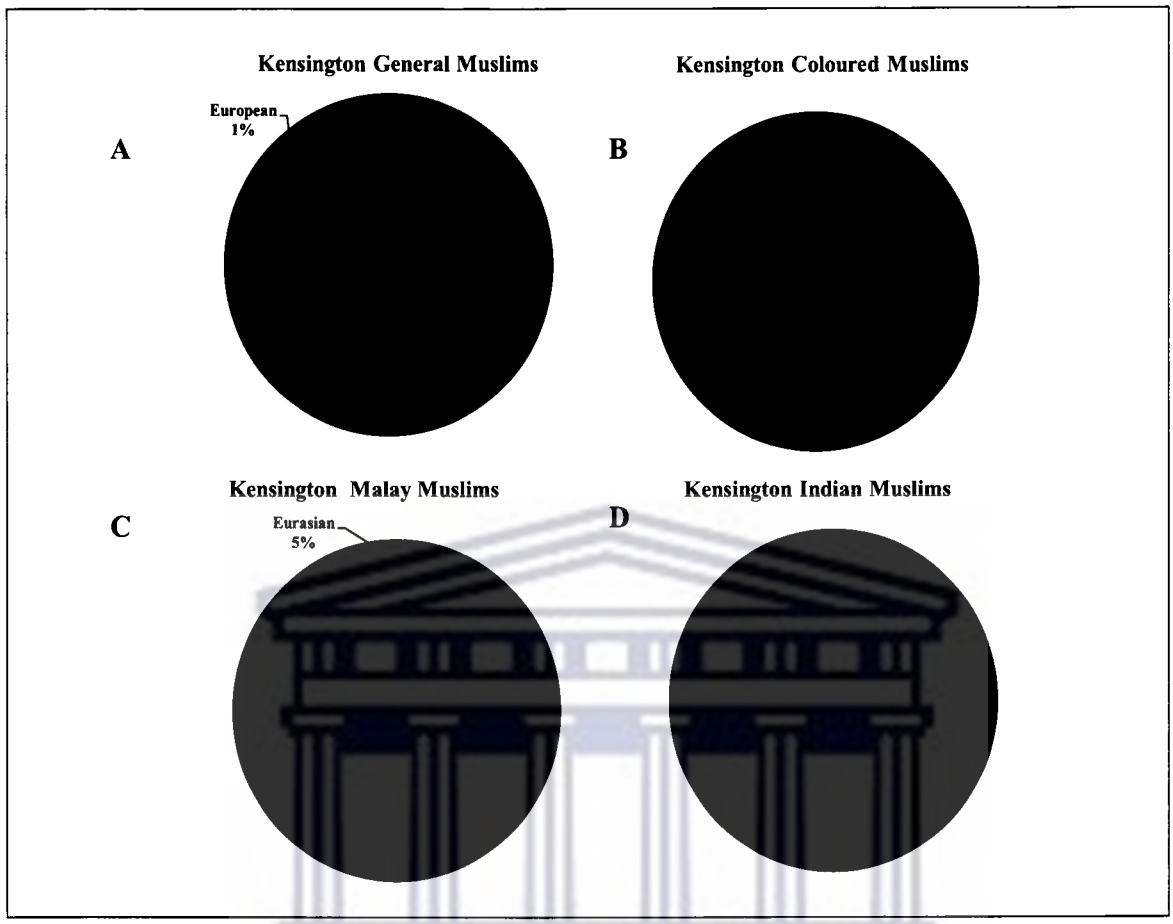


Figure 3.6: (A) Distribution of the major maternal lineages in Kensington General Muslims. (B-D) Distribution of the major maternal lineages found in each self-perceived ethnic groups.

A possible explanation for this result was the number of individuals associated with a particular self-perceived ethnic group in the respective studies. The study conducted on the Cape General Muslim population had an almost equal number of individuals representing each of the major self perceived ethnic groups which was made-up of 34% Cape Coloured Muslims, 27% Cape Malay Muslims, 29 % Cape Indian Muslims and 10% Cape Other Muslims. In the present study self-perceived ethnic groups, were comprised of 76% Kensington Coloured Muslims, 17 % Kensington Malay Muslims, 6% Kensington Indian Muslims and 1% Kensington Other Muslims. Kensington Coloured Muslims therefore mainly influenced the results of this study. It is important

to note however, that Kensington was previously a Coloured only residential area under the former Group Areas Act. Furthermore, the 1996 census data indicated that from the 4239 Muslims residing in Kensington, 86% were Coloured. The census data also indicated that different areas in the Cape Metropolitan area have a different composition of Muslim ethnic groups (Matthee 2008).

Kensington Coloured Muslims maternal lineages showed that 56 % (47/84) were of African origin, 36% (30/84) Asian and 8% (7/84) of Eurasian origin (Figure 3.6 B). Their counterpart the Cape Coloured Muslims shared a similar affinity for these lineages particularly for African and Asian lineages. Cape Coloured Muslims maternal lineages indicated that 47% (16/34) carried African lineages while 32% (11/34) Asian and 21% (7/34) European lineages (Chapter 2, Figure 2.10 B). This result was also observed between Kensington Malay Muslims and Cape Malay Muslims. Kensington Malay Muslims maternal lineages were 67% (12/18) African, 28% (5/18) Asian and 5% (1/18) Eurasian (Figure 3.6 C). Maternal lineages of Cape Malay Muslims were 45% (12/27) African, 22% (6/27) Asian and 33% (9/27) European (Chapter 2, Figure 2.10 C).

Based on the results it appears that Muslims perceived as Coloured and Malay share a uniform distribution of African and Asian lineages while European and Eurasian lineages may vary within these groups. These results also suggest that perhaps the geographical distribution of Muslims may possibly influence maternal lineages promoting admixture with certain ethnic groups. This seems probable considering the distribution of maternal lineages in Kensington Coloured Muslims and Kensington Malay Muslims were more similar in comparison to their Cape Muslim counterparts.

Moreover, Cape Coloured Muslims and Cape Malay Muslims maternal lineage distribution also show a closer relation in comparison to their Kensington counterparts.

Kensington Indian Muslims maternal lineages however showed little resemblance to either Kensington Coloured Muslims or Kensington Malay Muslims. Kensington Indian Muslims maternal lineages were 14% (1/7) African, 29% (2/7) Asian, 14% European (1/7) and 43% (4/7) Eurasian (refer to Figure 3.6 D). These lineages are characteristic of maternal lineages observed in indigenous populations of India with the exception of African lineages. This lineage was most likely introduced through recent admixture with Kensington Coloured Muslims, Kensington Malay Muslims or African Indian tribes such as the Siddis. Their counterpart the Cape Indian Muslims maternal lineages were almost identical as 14 % (4/29) were African, 45% (13/29) Asian and 41% (12/29) European. These results however were not unexpected as maternal grandmothers of both Kensington Indian and Cape Indian Muslims were primarily from India.

Kensington Other Muslims maternal lineage was of African origin while Cape Other Muslims maternal lineages were 20% (2/10) African, 30% (3/10) Asian and 50% (5/10) European. The small sample size for Kensington Other Muslims does not permit for a comparative analysis with Cape Other Muslim maternal lineage. It does however reveal that a higher percentage of Cape General Muslims (10%) identify by only their Muslim identity in comparison to Kensington General Muslims (1%).

Overall, mtDNA lineages of Kensington General Muslims showed a good correlation between ethnic identity and mtDNA haplogroups particularly for Kensington Coloured Muslims and Kensington Indian Muslims. Furthermore, revealed African and Asian

females as the primarily contributors of maternal lineages in Kensington General Muslims.

Chapter 4

Y-chromosome haplogroup analysis of the Kensington Muslim Community

4.1. Introduction

The male specific region of the Y-chromosome is suitable in tracing paternal ancestry as Y-chromosomal diversity has accumulated within lineages forming, male specific haplogroups distributed among human populations (Jobling 2000). These haplogroups are well characterized by single nucleotide polymorphisms (SNPs). Y-chromosome phylogeny is based on more than 200 SNPs defining 18 major haplogroups (Hammer and Zegura 2002; Cox 2006; Berniell et al. 2007; Karafet et al. 2008). It provides a detailed phylogeographic portrait of contemporary global population structure and past population movements and interactions (Underhill et al. 2001). This phylogeographic framework can therefore be used to infer the origin of the Kensington Muslims paternal ancestors, and provide insight into the Y-chromosomal structuring of this community. The Kensington community ancestry is derived from various parts of the world due the historical importation of slaves to Cape Town and through the of process of admixture with various population groups (Bradlow and Cairns 1978; Davids 1980; Da Costa and Davids 1994; Matthee 2008). This study therefore aims to determine the extent of admixture and assess which Y-chromosome lineages mainly contributed towards Y-chromosome diversity presently found among Kensington Muslims by typing thirteen binary markers in a hierarchical manner. Markers defining main haplogroups are typed first, and thereafter-defining haplogroups nested within the main haplogroups are genotyped (Berniell-Lee et al. 2007).

4.2. Material and Methods

4.2.1. Biological Samples and DNA extraction

The collection of biological samples, storage and DNA extraction thereof was as previously described in Chapter 3 (section 3.2.1).

4.2.2. Questionnaire

The data generated from the questionnaire and terms such as **Kensington General Muslims**, **conversion**, **internal immigrants**, **external immigrants**, and **inter-ethnic marriages** were used as previously described in Chapter 3 (section 3.2.2).

4.2.3. Y-Chromosome SNP Selection

A total of thirteen Y-chromosome binary markers were selected covering the major haplogroups representing African, Asian and European populations. The findings of Y-chromosome haplogroup diversity in Cape General Muslims (Chapter 2, section 2.3.4) was used as a reference for marker selection. Markers had previously been described by Onofri et al. (2006) and were initially arranged into two distinct multiplexes. However, in the present study, markers were arranged into four multiplexes (Table 4.2) due to the acquition of primers at various periods. Multiplex I contained the following SNPs: M173, M170, M45 and M89 when the derived state for M170 and M89 was observed samples were assigned to haplogroup I. However, when samples were only positive for either M45 or M89 further investigation was necessary to define haplogroups derived from these lineages. This was achieved by using Multiplex III covering M172 and M9 markers, as well as Multiplex IV covering M216, M201 and M214 markers. Haplogroups J2 and K could be assigned when using Multiplex III while haplogroups C, G and NO for Multiplex IV. The absence of polymorphic markers for Multiplex I was investigated in Multiplex II and IV. Multiplex II contained M91, M181, M52 and M96 defining haplogroups A, B H and E respectively.

The expected base substitutions for each marker and haplogroup assignment can be found in Table (4.1). The alteration of the multiplex however did not interfere with the hierarchical genotyping strategy using the Y-chromosome phylogenetic tree defining major haplogroups A to R (Figure 4.1). The nomenclature of Y-chromosome haplogroups were assigned according to the Y Chromosome Consortium.

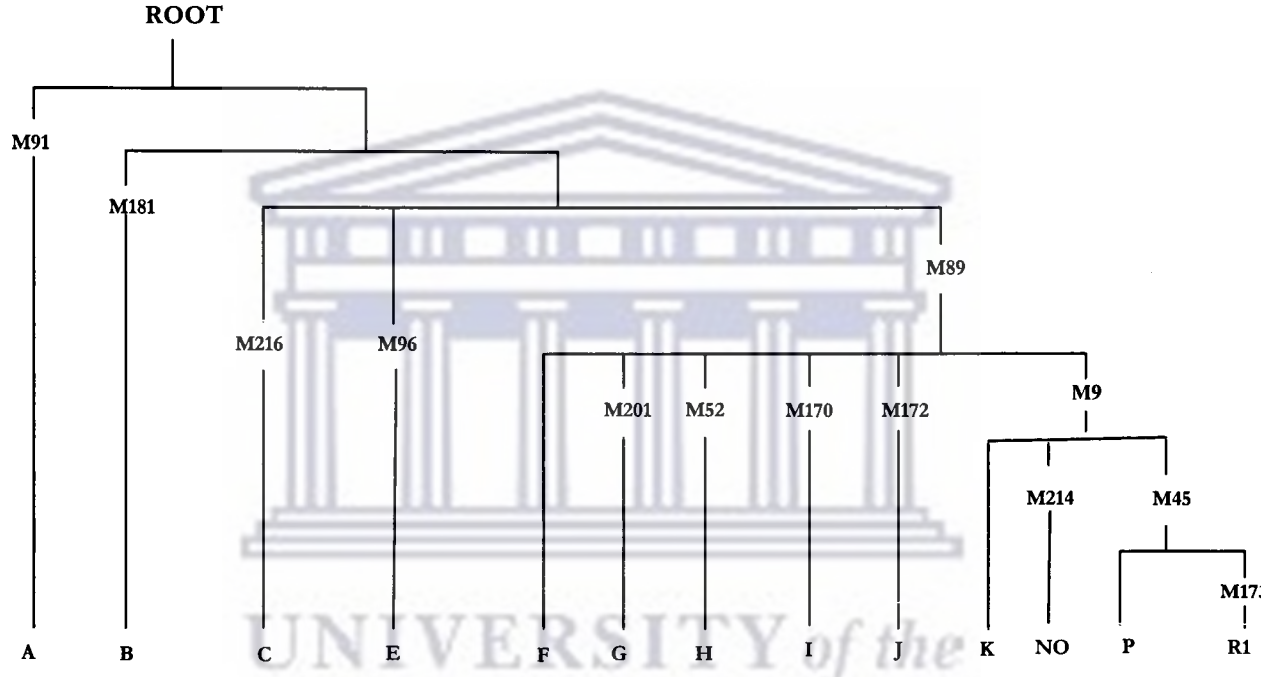


Figure 4.1. A phylogenetic tree of Y-chromosome haplogroups typed in the study (modified from Jobling and Tyler-Smith 2003).

Table: 4.1. Y-chromosome markers and base substitutions used to assign haplogroups and paternal ancestry

Markers typed and base substitutions defining each Y-Chromosome lineage														
Haplogroups	**M91**	**M181**	**M216**	**M96**	**M89**	**M201**	**M52**	**M170**	**M172**	**M9**	**M214**	**M45**	**M173**	**Inferred ancestry**
A	**8T**	T	C	G	C	G	A	A	T	C	T	G	A	African
B	9T	**C**	C	G	C	G	A	A	T	C	T	G	A	African
C	9T	T	**T**	G	C	G	A	A	T	C	T	G	A	Asian
E	9T	T	C	**C**	C	G	A	A	T	C	T	G	A	African
F	9T	T	C	G	**T**	G	A	A	T	C	T	G	A	Asian/European
G	9T	T	C	G	**T**	**T**	A	A	T	C	T	G	A	Asian/European
H	9T	T	C	G	**T**	G	**C**	A	T	C	T	G	A	European
I	9T	T	C	G	**T**	G	A	**C**	T	C	T	G	A	European
J2	9T	T	C	G	**T**	G	A	A	**G**	C	T	G	A	European
K	9T	T	C	G	**T**	G	A	A	T	**G**	T	G	A	Asia
NO	9T	**T**	C	G	**T**	G	A	A	T	**G**	C	G	A	Asia
P	9T	T	C	G	**T**	G	A	A	T	**G**	T	**A**	A	Asian
R1	9T	T	C	G	**T**	G	A	A	T	**G**	T	**A**	**C**	European

Bold letters indicate the expected base substitution (Table adapted from Nelson et al. 2007)

4.2.4. Multiplex PCR Amplification of Y-Chromosome SNP Sites

Multiplex PCR was performed using 20-25ng of template DNA and Qiagen[R] Multiplex PCR kit (*Qiagen*). The PCR reaction was setup in a final volume of 17µl comprised of 0.94x QIAGEN Multiplex PCR Master Mix, 0.5x Q-Solution. Primers were synthesized by *Whitehead Scientific* using previously described sequences (Onofri et al. 2006). A primer mix was setup for each of the respective multiplexes (listed in Appendix I). Primer sequences and final concentration of each primer is given in Table 4.2.

PCR amplifications were performed in a GeneAmp[R] 2720 PCR System thermal cycler (*Applied Biosystems*). The thermal cycling conditions were as follows: 1 cycle at 95˚C for 15 min, 40 cycles of 94˚C for 30 s, 60˚C for 1 min and 60 s, and 72˚C for 1 min and 60 s, followed by a final cycle at 72˚C for 10 min.

PCR products were verified by agarose gel electrophoresis to ensure that the correct amplicon sizes were obtained for each multiplex. An aliquot of 2µl of PCR product was run on a 3% agarose gel stained with 0.1mg/ml ethidium bromide.

Table 4.2. Y-chromosome multiplexes amplicon product size and final primer concentrations

Marker	Primer sequence (5'-3')	Primer length (bp)	Amplicon length (bp)	Final (µM)
Multiplex I				
M170	Fw GTTTTCATATTCTGTGCATTATACAAATTACTAT	34	96	0.21
	Rv CATTTTACAGTGAGACACAACCCAC	25		
M173	Fw AAAATTTTCTTACAATTCAAGGGCATT	27	104	0.21
	Rv GCTGCAGTTTTCCCAGATCCT	21		
M45	Fw GGTGTGGACTTTACGAACCAACCT	24	109	0.21
	Rv CCTGGACCTCAGAAGGAGCTTT3	22		
M89	Fw CTGCTCAGCTTCCTGGATTCA	21	110	0.21
	Rv CACTTTGGGTCCAGGATCACC	21		
Multiplex II				
M52	Fw CTCCCACCTCAACTTCCCAGAG	22	153	0.21
	Rv AGCAAACATTTCAAGAGAGAATGAAA	26		
M96	Fw TTCTCCATATCTGTGTAAAGGCAAGT	26	165	0.21
	Rv CCATAGGTTTTTAATATTATACCTGAGTG	29		
M181	Fw GCTAGCAAAGTTGGCTTGGG	20	166	0.21
	Rv GCACACTAGCTATAAGCAAAAGAAAT	26		
M91	Fw ATTGCGATGTTTTATTTCAAAACAAGATG	29	170	0.21
	Rv GCGTATTTTTCAAAAATATATGGAGAA	27		
Multiplex III				
M172	Fw TTTTATCCCCCAAACCCATTT	21	136	0.21
	Rv CATGTTGGTTTGGAACAGTTTATCC	25		
M9	Fw AGAACTGCAAAGAAACGGCCTA	22	127	0.21
	Rv TGCATAATGAAGTAAGCGCTACCT	24		
Multiplex IV				
M216	Fw AAGCCACTTAAATTCCAATGGA	22	171	0.21
	Rv CACTGCTAGTTATGTATACCTGTTGAA	27		
M201	Fw TATGCATTTGTTGAGTATATGTCAAAT	27	163	0.21
	Rv TCCAACACTAAGTACCTATTACGAAAA	27		
M214	Fw CAATTGTACAGCACAAATATATGCCTGTAAA	31	207	0.21
	Rv GAGGTCAAGGGTGTGGTGAG	20		

Primers were adapted from Onofri et al. 2006

4.2.5. SNaPshot Minisequencing Reactions and Analysis

Prior to minisequencing reactions, 15µl of PCR product was purified using 5µl (1unit/µl) SAP and 0.5µl (10 units/µl) ExoI enzymes following cycling conditions specified by manufacturer. Minisequencing reactions were performed as previously described in Chapter 3 (section 3.2.5) in a final reaction volume of 13.5µl using 0.15-0.30µM of extension primer. The primer sequences and final concentration of each primer is indicated in Table 4.3. *Whitehead Scientific* manufactured primers using sequences previously described by Onofri et al. (2006).

The purification and analysis of minisequencing primers was as previously described in Chapter 3 (section 3.2.5) using 1µl of SAP and 0.8µl SAP treated product for analysis.

Table 4.3 Y-chromosome minisequencing primer sequences, length, base substitution and final primer concentrations

Marker	Primer sequence (5'-3')	Primer length (bp)	Final (µM)	Base substitution
Multiplex I				
M173	Fw CAATTCAAGGGCATTTAGAAC	21	0.15	A>C
M170	Fw AAATTACTATTTTATTTACTTAAAAATCATTGTTC	35	0.15	A>C
M45	Fw (6-POLY-GACT-TAIL) AAATTGGCAGTGAAAAATTATAGATA	50	0.15	G>A
M89	Rv (34-POLY-T-TAIL) CAACTCAGGCAAAGTGAGAGAT	56	0.15	C>T
Multiplex II				
M52	Fw ATACCTATAAGAATATTGCCTGCA	24	0.15	A>C
M181	Fw (4-POLY-GACT-TAIL) GGACAACTTGATCATCTTTTTGA	39	0.15	T>C
M91	Rv (4-POLY-GACT-TAIL) GATACTACAGTAGTGAACTGATTAAAAAAAA	47	0.15	9T>8T
M96	Fw (6-POLY-GACT-TAIL) GTAACTTGGAAAACAGGTCTCTCATAATA	53	0.15	G>C
MultiplexIII				
M172	Rw AAGAAAATAATAATTGAAGACCTTTTAAGT	30	0.23	T>G
M9	Fw (6- POLY-GACT-TAIL) ACGGCCTAAGATGGTTGAAT	44	0.23	C>G
Multiplex IV				
M216	Rv CACTGCTAGTTATGTATACCTGTTGAAT	28	0.23	C>T
M201	Fw (4-POLY-GACT-TAIL) AGATCTAATAATCCAGTATCAACTGAGG	44	0.30	G>T
M214	Rv GA (8- POLY-GACT-TAIL) AGTGTGAGACACTGTCTGAAAACAAC	60	0.23	T>C

Minisequencing primers were adapted from Onofri et al. 2006

4.3. Results and Discussion

4.3.1. Multiplex PCR

The use of QIAGEN Multiplex PCR Master Mix kit simplified the task of optimising the respective multiplexes utilized in the study. Since the use of conventional PCR reagents and methods yielded faint PCR products and inconsistent results. In addition, frequent drop out for SNP markers was observed during genotyping analyses. This was rarely observed when amplifying SNP markers with the kit and only occurred for individual samples. This was solved by increasing the concentration of template DNA from 20ng to 25ng. None of the Y-chromosome PCR multiplexes had any artifacts as observed for mtDNA multiplexes.

4.3.2. Y-Chromosome SNP Genotyping

The thirteen Y-chromosome SNP markers typed in the study generated clear results using the SNaPshot Multiplex kit (Figure 4.2 and 4.3). This was achieved following the same guidelines as those previously described in Chapter 3 (section 3.3.2.). When extraneous peaks were observed within a multiplex, the following factors were responsible: insufficient purification of either PCR or minisequencing products. Size variations of minisequencing products were also observed for Y-chromosome multiplexes due to factors mentioned for mtDNA typing (Chapter 3, section 3.3.2.). A stronger fluorescent signal was observed for certain SNPs depending on the inserted ddNTP. This can be seen illustrated in Multiplex I for M89 (Figure 4.2 and 4.3). A higher fluorescence occurred when a ddGTP (Blue) was inserted rather than a ddATP nucleotide (Green). Haplogroups were assigned once peaks were above 400 relative fluorescence. When off-scale peaks occurred, the amount of minisequencing product was reduced for the ABI analysis.

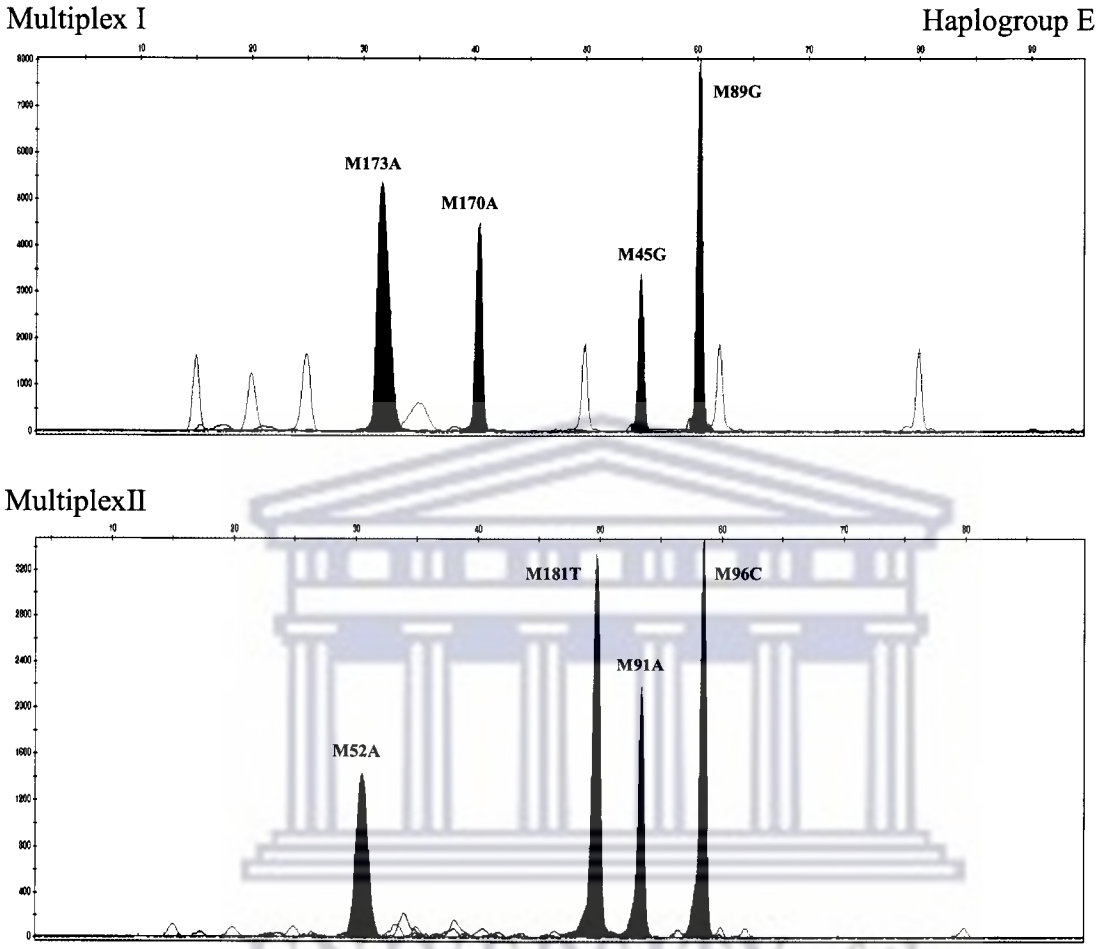
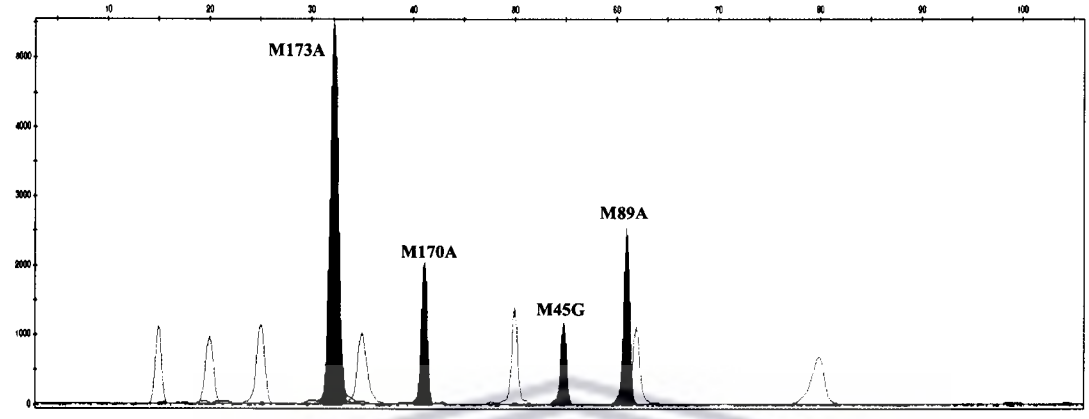


Figure 4.2. Electropherogram of multiplex I and II indicating the derived state for haplogroup E defined by M96. The colour of each peak indicates the incorporated ddNTP tagged with a unique fluorescent dye (dR6G= ddATP (Green), dTAMRA = ddCTP (Black), dR110 = ddGTP (Blue) and dROX = ddTTP (Red).

Multiplex I Haplogroup G



Multiplex IV

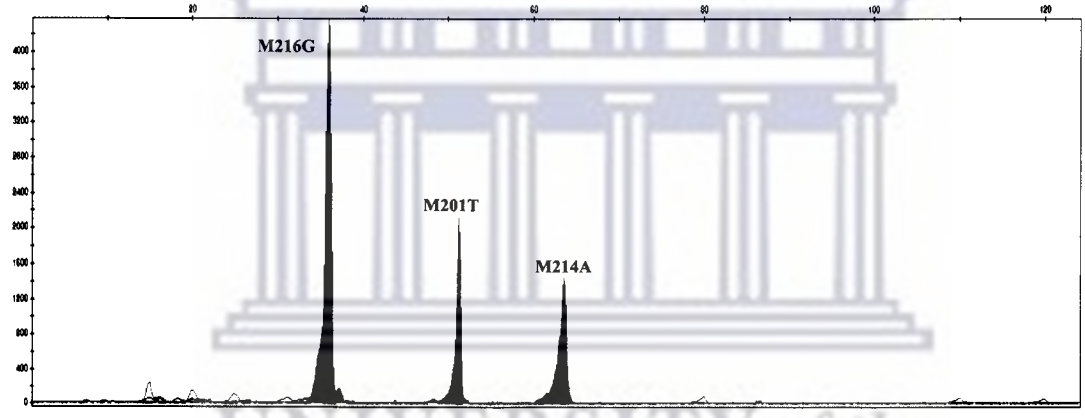


Figure 4.3. Electropherogram of multiplex IV and I showing the derived state for haplogroup G defined by the M89 and M201 polymorphism.

4.3.3. Self-Perceived Ethnic Classification

The ethnic affiliations indicated by Kensington Muslims and their frequency in the study were as previously described in Chapter 3 (section 3.3.3). Briefly, 76% (84/110) of donors perceived themselves as Coloured while 17% (18/110) Malay and 6% (7/110) Indian (refer to Chapter 3, Figure 3.4). One donor representing 1% of the total sample population however did not identify with any ethnic classification. The term Other Muslim was later used to describe this individual.

4.3.4. Y-Chromosome Haplogroup Variation

Analysis of the thirteen Y-chromosome SNP markers indicated that Kensington General Muslims Y-chromosomes belonged to haplogroups C (9%), E (4%), F (10%), G (1%), I (10%), J2 (9%), K (13%), NO (11%), P (8%) and R1 (25%). The frequency of each Y-chromosome haplogroup observed in the study and in each self-perceived ethnic group is illustrated in Figure 4.4 (A-D). Haplogroup R1 displayed the highest frequency in the study (25%) and in Kensington Coloured Muslims (26%). This haplogroup is common in European populations and observes its highest frequency in Britain and the Basque (Wells et al. 2001; Kayser et al. 2005; Karafet et al. 2008). Only 10% of haplogroup R1 frequency was derived from converts, internal and external immigrants and inter-ethnic marriage (Table 4.4). Therefore, a large percentage of haplogroup R1 frequency could possibly be derived from colonizing European males or their descendants.

The presence of other West European haplogroups I, G, and J2 suggest that this perhaps is the most likely source of these haplogroups. Given that only a small percentage of haplogroup I (2%) and J (1%) frequency was attributed to recent admixture introduced by external immigrants and inter-ethnic marriages (Table 4.4). Furthermore, West European haplogroups I, G, J2, and R1 were predominately found among the oldest Muslim ethnic groups the Kensington Coloured Muslims and Kensington Malay Muslims (Figure 4.4 B and C). Haplogroup R1 (29%) was the only West European haplogroup observed in Kensington Indian Muslims (Figure 4.4 D). This result however is not unexpected as West European Y-chromosome haplogroups are either absent or rarely present in Indian males. Their occurrence in Indian populations is a result of recent admixture with either Central Asians or European populations (Sahoo et al. 2006).

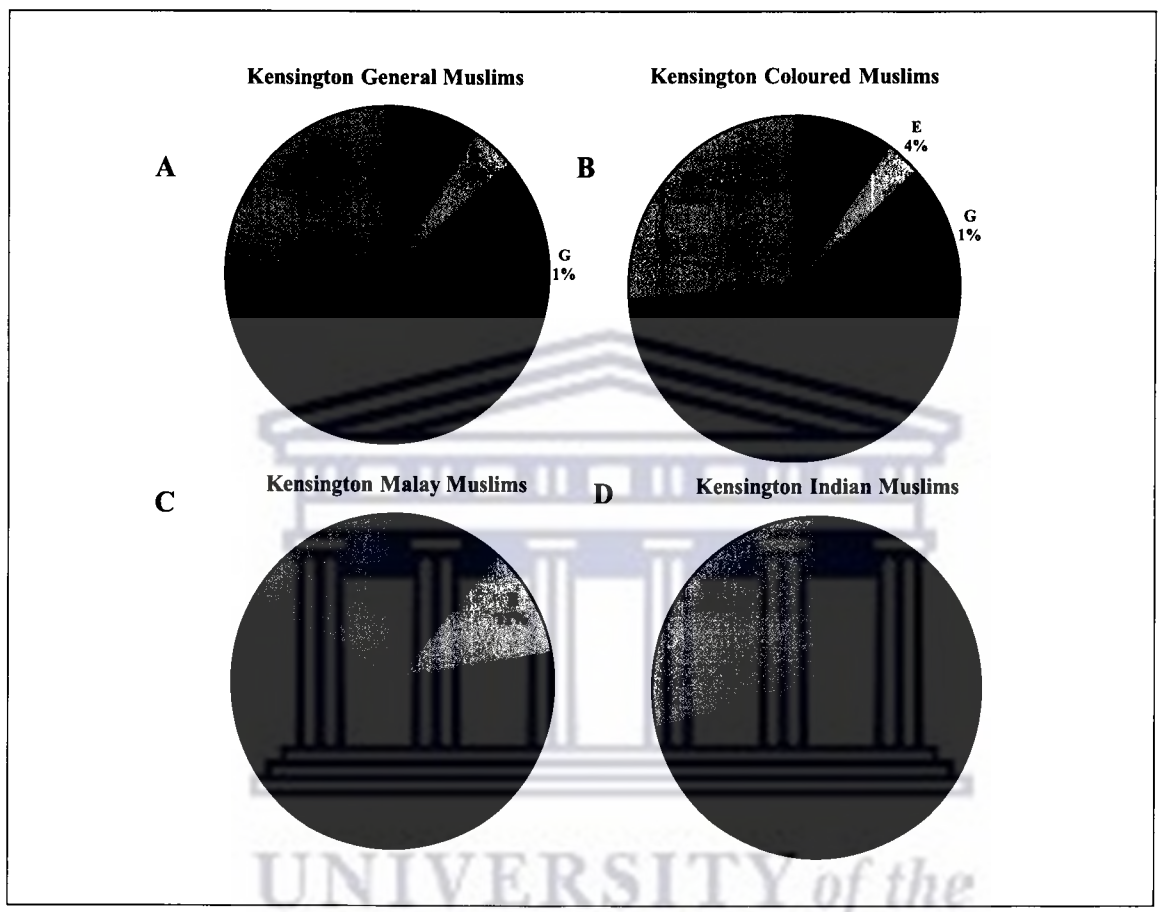


Figure 4.4. (A) Y-chromosome haplogroups observed in Kensington General Muslims. (B-D) Y-chromosome haplogroups observed in each self-perceived ethnic group

Table 4.4. Distribution of Y-chromosome haplogroups in the different ethnic of Kensington General Muslims deduced from the questionnaire data

				No of individuals in each Haplogroup									
Factor	Ethnic groups	No of individuals	No of haplogroups	C	E	F	G	I	J2	K	NO	P	R1
Conversion	Coloured	2	2							1			1
	Cape Malay	0											
	Indian	1	1			1							
	Other	0											

				No of individuals in each Haplogroup									
Factor	Ethnic groups	No of individuals	No of haplogroups	C	E	F	G	I	J2	K	NO	P	R1
Internal Immigrants	Coloured	6	3			2					1		3
	Cape Malay	1	1	1									
	Indian	1	1			1							
	Other	0											

				No of individuals in each Haplogroup									
Factor	Ethnic groups	No of individuals	No of haplogroups	C	E	F	G	I	J2	K	NO	P	R1
External Immigrants	Coloured	5	3					1		2			2
	Cape Malay	1	1							1			
	Indian	5	4			2					1	1	1
	Other	0											

				No of individuals in each Haplogroup									
Factor	Ethnic groups	No of individuals	No of haplogroups	C	E	F	G	I	J2	K	NO	P	R1
Inter- ethnic marriage	Coloured	12	8	2		2		1	1	1	2	1	2
	Cape Malay	1	1							1			
	Indian	0											
	Other	0											

* Ethnic identity is listed as perceived by donor

Haplogroups C (9%), F (10%), and K (13%), NO (11%), and P (8%) represent the majority of the remaining Y-chromosome haplogroups found in Kensington General Muslims (Figure 4.4 A-D). Furthermore, reflect haplogroups predominately associated with Asian Y-chromosome haplogroup diversity (Underhill 1997; Capelli et al. 2001; Cordaux et al. 2004; Kumar et al. 2007; Karafet et al. 2008). Thus given the history of early Muslims and their geographical origins these haplogroups can therefore mainly be attributed to the importation of Asian slaves, or from later Asian immigrants in the 19th century (Filipino's, Indians and Chinese) (Da Costa and Davids 1994; Du Pre

1994; van der Ross 2005). In recent years have mainly been introduced through internal immigrants and inter-ethnic marriages (Table 4.4).

Asian haplogroups and their high prevalence (51%) in the study also suggest that various Asian populations have contributed to the Y-chromosome haplogroup diversity of Kensington General Muslims. Since haplogroups NO and C are found frequently in Central and East Asia but show a higher diversity in South East Asia and Melanesia (Jobling and Tyler-Smith 2003; Sahoo et al. 2006; Underhill and Kivisild et al. 2007). Haplogroup F is common among Indian populations and haplogroup K is widespread in Asia, while haplogroup P observes its highest frequency in Central Asia (Wells et al. 2001; Al-Zackery et al. 2003; Bortolini et al. 2003; Kayser et al. 2003; Athey 2005; Karafet et al. 2005).

Haplogroup E had one of the lowest frequencies in the study present in Kensington Coloured Muslims (4%) and Kensington Malay Muslims (11%) (Figure 4.4 B and C). The haplogroup is widespread across Africa and may possibly be attributed to recent gene flow or from African slaves from East Africa.

4.3.5. The overall geographic origins of the Kensington Muslim Community

The analysis of 13 Y-chromosome SNP markers from the Y-chromosome of 110 Kensington General Muslims identified 10 paternal lineages that are distributed throughout African, Asian and European populations. The Y-chromosome lineages and their frequencies in the study and in each self-perceived ethnic group are reported in Figure 4.5 A - D. Kensington General Muslims paternal lineages were 4% African, 51% Asian and 45% European (Figure 4.5 A). Thereby, demonstrating a high degree

of diversity and indicated that males of Asian and European ancestry were mainly responsible for Y-chromosome diversity in Kensington General Muslims.

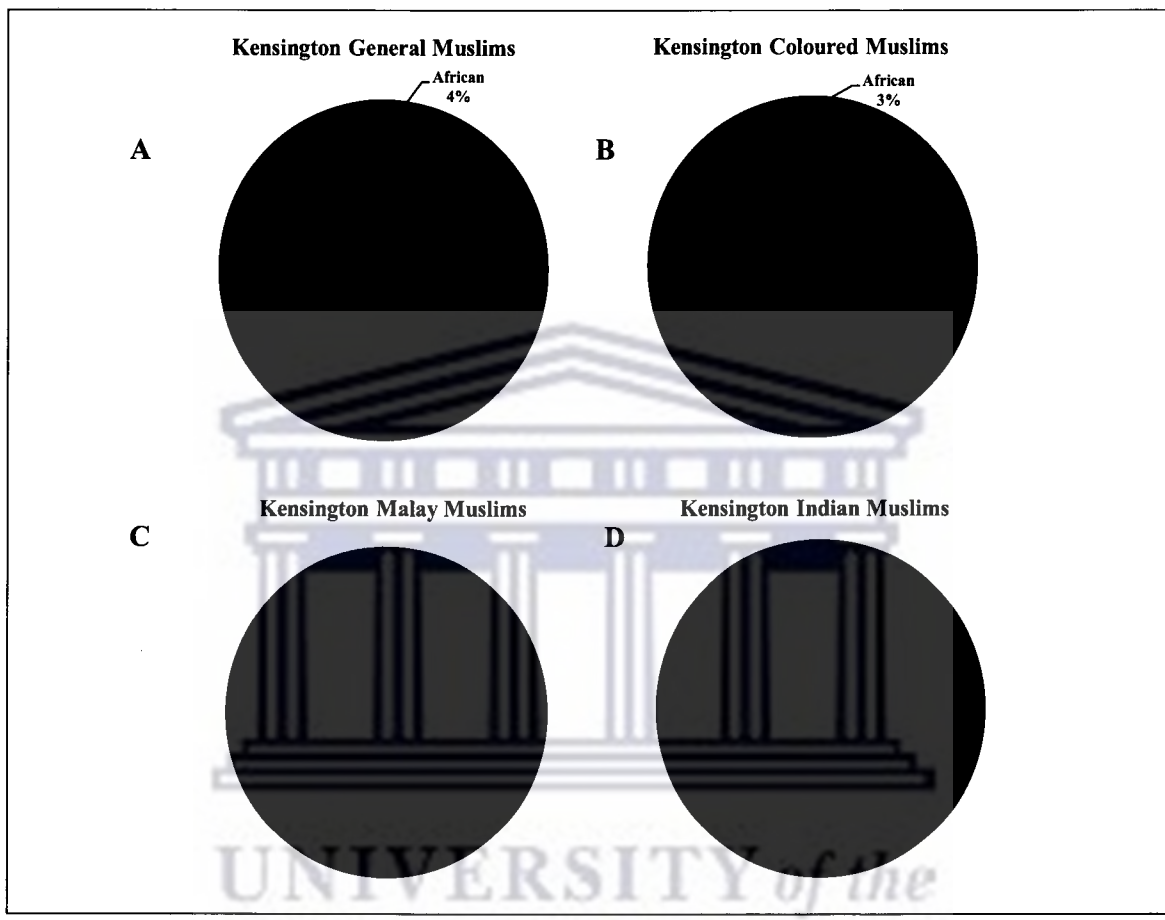


Figure 4.5. (A) Distribution of the major paternal lineages found in Kensington General Muslims. (B-D) Distribution of the major paternal lineages found in each self-perceived ethnic group.

Paternal lineages in Cape General Muslims showed a similar distribution, as African lineages were the least frequent (5%) while Asian (75%) and European lineages (13%) the most frequent (Chapter 2, Figure 2.11 A). Kensington General Muslims however demonstrated a higher frequency for European lineages. The result was possibly influenced by the sample group mainly consisting of Muslims perceived as Coloured (76%). This was previously discussed in Chapter 3, section 3.3.5, which indicated that Kensington was a former Coloured only residential area. The data from the 1996

census also indicated that from the 4239 Muslims residing in Kensington, 86% are Coloured (Matthee 2008).

Kensington Coloured Muslims paternal lineages were 3% (2/84) African, 48% (40/84) Asian and 49% (34/84) European (Figure 4.5 B). Cape Coloured Muslims paternal lineages show a very similar distribution, as 6% (2/34) were of African origin, 56% (19/34) Asian, and 23% (7/34) of European origin (Chapter 2, Figure 2.11 B). However, 15% of their Y-chromosome lineages remained undefined which was speculated to be either of Asian or European origin. This seems plausible considering the high frequencies of Asian and European lineages observed in Kensington Coloured Muslims.

Similarities between Kensington Malay Muslims and Cape Malay Muslims lineages were also observed. Kensington Malay Muslims paternal lineages were 11% (2/18) African, 56% (10/18) Asian and 33% (6/18) European (Figure 4.5 C). Paternal lineages of Cape Malay Muslims were 78% (21/27) Asian and 15% (4/27) European with 7% (2/27) of their lineages remaining undefined (Chapter 2, Figure 2.11 C). The absence of African lineages in Cape Malay Muslims could possibly be attributed to the few African markers screened in the study. In the present study all, the major African haplogroups were screened, however only one African haplogroup was found in Kensington General Muslims, haplogroup E. It is therefore quite possible that a percentage of the undefined lineages of Cape Malay Muslims could actually belong to haplogroup E. Considering that only markers M2 and M35 were screened defining sub lineages of haplogroup E whereas in the present study M96 was screened defining haplogroup E.

Kensington Indian Muslims paternal lineages indicated that 71% were of Asian origin and 29% of European origin (Figure 4.5 D). Cape Indian Muslims paternal lineages were 97% Asian and 3% African (Chapter 2, Figure 2.11 D). The African lineage was probably derived from recent admixture with Cape Coloured Muslims and Cape Malay Muslims.

Kensington Other Muslim paternal lineage was of Asian origin (data not indicated). Since only one individual belonged to this group comparative, analysis with Cape Other Muslims could not be achieved.

Overall, both Kensington Indian Muslims and Cape Indian Muslims paternal lineages correlated with findings of Y-chromosome diversity in India, as different Indian populations share distinct frequencies for certain Y-chromosome haplogroups. Tribal groups and lower caste Indian populations exhibit a higher frequency for Asian and Indian specific Y-chromosome haplogroups. Indian upper caste populations however show a higher frequency for non-Asian and Indian specific haplogroups. Their Y-chromosome lineages mainly belong to subgroups of haplogroup R (Sahoo et al. 2005). It is therefore possible that Kensington Indian Muslims and Cape Indian Muslims paternal lineages were derived from different populations in India. This is evident from the distribution of their paternal lineages and that both Kensington Indian Muslims and Cape Indian Muslims Y-chromosome lineages were mainly derived from recent Indian immigrants.

The results obtained in the study suggest that Kensington General Muslim's paternal lineages were mainly derived from males of Asian and European origin. African paternal lineages were also observed in the study but at a lower frequency present in

Kensington Coloured Muslims and Kensington Malay Muslims. Their Y-chromosome lineages show a similar construct in comparison to Kensington Indian Muslims. Furthermore, their paternal lineages were in agreement with historical data concerning early Muslims geographical origins and showed a higher degree of admixture in comparison to Kensington Indian Muslims. Overall, Y-chromosome lineages were also found to correlate with ethnic classification of donors.

Chapter 5

Conclusion

The present study forms part of a greater research project aimed at investigating the genetic diversity and origins of local Muslim communities living in the Cape Metropolitan area. These communities are investigated for DNA polymorphisms both from maternal (mitochondrial DNA) and paternal (Y-chromosome) lineages extensively. Variations in these two types of DNA are grouped into continent-specific haplogroups or lineages. Geographic origin is then assigned to each lineage, and consequently migration patterns are traced back to these communities. This also allows for the study of interactions with other communities living in the same residential areas. The contributions of these communities to the gene pool of the Cape Muslim communities through intermarriages, conversion and blending was analyzed.

The initial part of the study investigated the genetic genealogy of the Muslim population living in the Cape Metropolitan area. Mitochondrial and Y-chromosome DNA variations were used to infer the geographic/ethnic origins of the population. This was achieved by screening single nucleotide polymorphisms (SNPs) using polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP).

Maternal lineages of Cape Coloured Muslims and Cape Malay Muslims suggest that females mainly of African origin influenced their maternal gene pool. Whereas Asian and European females mainly influenced mtDNA diversity observed in Cape Indian Muslims. Cape Other Muslims maternal lineages revealed females of European and Asian ancestry as the main contributors of their lineages. Paternal lineages however

indicated Asian males as the major contributors of Y-chromosome diversity in Cape General Muslims. This finding was observed for all the self-perceived ethnic groups. The results of maternal and paternal lineages in Cape Coloured Muslims and Cape Malay Muslims correlated well with the historians findings concerning the geographical origins of earlier Muslims. This also indicated that the slave population brought from Asia were probably mainly Muslim and male thereby promoting inter-ethnic marriages with indigenous African females. The maternal and paternal lineages of Cape Indians correlated with findings concerning mtDNA and Y-chromosome studies performed on ethnic Indian populations. These results were also in agreement with the family history of Cape Indian Muslims as they were mainly the descendents of recent Indian immigrants.

In the second part of the study, contributions of Major African, Asian and European haplogroups to the Kensington General Muslim maternal gene pool was investigated. This was achieved by examining mtDNA variations found within this community using twelve mtDNA coding region SNPs genotyped by the SNaPshot minisequencing assay. The assay provided an alternative to the time consuming PCR-RFLP method allowing for the rapid, robust and simultaneous genotyping of multiple SNP markers.

Kensington Coloured Muslims, Kensington Malay Muslims and Kensington Other Muslims maternal lineages indicated African females as the main contributors of their female lineages. Kensington Indian Muslims maternal lineages however indicated Eurasian and Asian females mainly influencing their maternal gene pool.

This study also determined the extent of admixture and assessed which paternal lineages mainly contributed towards Y-chromosome diversity presently found among

Kensington General Muslims by typing thirteen binary markers in a hierarchical manner. Markers defining main haplogroups were typed first, and thereafter haplogroups nested within the main haplogroups were genotyped.

Paternal lineages of Kensington General Muslims indicated that both Asian and European males significantly contributed to Y-chromosome diversity within the community. Kensington Malay Muslims, Kensington Indian Muslims and Kensington Other Muslims paternal lineages were mainly derived from Asian males. Whereas, Kensington Coloured Muslims paternal lineages indicated European males then Asian males mainly influencing Y-chromosome diversity.

The Kensington-Factreton residential area in the Cape Peninsula was created to accommodate displaced Coloured people, Muslims and Christians, as a result of the Group Areas Act. It is therefore recommended to conduct a similar study on the Kensington Christian community in order to determine the extent of interactions between the two communities. The effect of inter-religious marriages and conversions on the genetic diversity of both populations can indeed be investigated.

Future studies of the local Muslim communities living in the Cape metropolitan area should include Y-chromosome STR and mitochondrial DNA sequence analysis. This would provide more detailed information on the genetic diversity within these communities and pinpoint more accurate geographic and ethnic origins.

References

Agrawal S, Khan F, Pandey A, Tripathi M, Herrera RJ (2005) YAP, signature of an African–Middle Eastern migration into northern India. Curr Sci 8:1977-1980

Álvarez-Iglesias V, Jaime JC, Carracedo A, Salas A (2007) Coding region mitochondrial DNA SNPs: Targeting East Asian and Native American haplogroups. Forensic Sci Int: Genet 1:44-45

Alves-Silva J, da Silva Santos M, Giumaraes PEM, Ferreira ACS, Bandelt HJ, Pena SDJ, Prado VF (2000) The Ancestry of Brazilian mtDNA lineages. Am J Hum Genet 67:444-461

Al-Zachery N, Semino O, Benuzzi G, Magri C, Pasarino G, Torroni A, Santachiara-Benerecetti AS (2003) Y-chromosome and mtDNA polymorphisms in Iraq, a crossroad of the early human dispersal and of post-Neolithic migrations. Mol Phylogenet Evol 28:458-72

Alzualde A, Izagirre N, Alonso S, Alonso A, De la Rúa C (2005) Temporal Mitochondrial DNA Variation in the Basque Country : Influence of Post-Neolithic Events. Annu Hum Genet 69:665-679

Athey TW (2005) Pitfalls in Determinations of Y Haplogroups F*. J Genetic Genealogy 1:35-39

Avise JC (2008) Evolutionary Pathways in Nature. Cambridge University Press, Cambridge, pp 240-247

Ballinger SW, Schurr TG, Torroni A, Gan YY, Hodge JA, Hassan K, Chen KH, Wallace DC(1992) Southeast Asian Mitochondrial DNA Analysis Reveals Genetic Continuity of Ancient Mongoloid Migrations. Genet 130: 139-152

Ballard JW and Whitlock MC (2004) The incomplete natural history of mitochondria. Mol Ecol 13:729-744

Basu A, Mukherjee N, Roy S, Sengupta S, Banerjee S, Chakraborty M, Dey B, Roy M, Roy B, Bhattacharyya NP, Roychoudhury S, Majumder PP (2003). Ethnic India: A Genomic View, With Special Reference to Peopling and Structure. Genome Res 13:2277-2290

Behar DM, Rosset S, Blue-Smith J, Balanovsky O,Tzur S, Comas D,Mitchell R J,Quintana-Murci L, Tyler-Smith C, Wells RS (2007) The Genographic Consortium,The Genographic Project Public Participation Mitochondrial DNA Database. PLoS Genet 3:1083-1095

Beleza S, Gusmão L, Lopes A, Alves C, Gomes I, Giouzeli M, Calafell F, Carracedo A, Amorim A (2005) Micro-Phylogeographic and Demographic History of Portuguese Male Lineages. Ann Hum Genet 70:181-194

Benjeddou M, Leat N, Davison S (2006) Use of molecular genetics and historical records to reconstruct the history of local communities. Afr. J. Biotechnol 5:2516-2519

Bermisheva MA, Tamberts K, Villems R, Khusnutdinova EK (2002) Diversity of Mitochondrial DNA Haplogroups in Ethnic Populations of the Volga-Ural Region. Mol Biol 36: 802-812

Berniell-Lee G, Sandoval K, Mendizabal I, Bosch E, Comas D (2007) SNPlexing the human Y-chromosome: A single-assay system for major haplogroup screening. Electrophoresis 28:3201-3206

Berniell-Lee G, Plaza S, Bosch E, Calafell F, Jourdan E, Césari M, Lefranc G, Comas D (2008) Admixture and Sexual Bias in the Population Settlement of La Réunion Island (Indian Ocean). Am J Physical Anthropol 136:100-107

Bortolini MC, Salzano FM, Thomas MG, Stuart S, Nasanen SPK, Bau CHD, Hutz MH, Layrisse Z, Petzl-Erler ML, Tsuneto LT, Hill K, Hurtado AM, Castro-de-Guerra D, Torres MM, Groot H, Michalski R, Nymadawa P, Bedoya G, Bradman N, Labuda

D, Ruiz-Linares A (2003) Y-Chromosome Evidence for Differing Ancient Demographic Histories in the Americas. Am J Hum Genet 73:524–539

Bradlow FR and Cairns M (1978) The early Cape Muslims: A study of their mosques, genealogy and origins. Printpak, Cape Town, pp 80-99

Brandstätter A, Parsons TJ, Parson W (2003) Rapid screening of mtDNA coding region SNPs for the identification of west European Caucasian haplogroups. Int J Legal Med 117:291-298

Budowle B, Allard MW, Wilson MR, Chakraborty R (2003) Forensics and Mitochondrial DNA: Applications, Debates and Foundations. Annu Rev Genomics Hum Genet 4:119-141

Budowle B (2004) SNP typing strategies. J. Forensic Sci 146S:S139-S142

Butler JM (2003) Recent Developments in Y-Short Tandem Repeat and Y-Single Nucleotide Polymorphism Analysis. Forensic Sci Rev 15:91-109

Capelli C, Wilson JF, Richards M, Stumpf MPH, Gratrix F, Oppenheimer S, Underhill P, Pascali VL, Ko TM, Goldstein DB (2001) A Predominantly Indigenous Paternal Heritage for the Austronesian- Speaking Peoples of Insular Southeast Asia and Oceania. Am J Hum Genet 68:432–443

Cavalli-Sforza LL and Felman MW (2003) The application of molecular genetic approaches to the study of human evolution. Nature Genetics 33:266-274

Chen Y, Olckers A, Schurr T, Kogelnik AM, Huoponen K, Wallace, DC (2000) MtDNA Variation in the South African Kung and Khwe - and Their Genetic Relationships to Other African Populations. Am J Hum Genet 66: 1362-1383

Comas D, Plaza S, Wells RS, Yuldaseva N, LaoO, Calafell F, Bertranpetit J (2004) Admixture, migrations and dispersals in central Asia: Evidence from maternal DNA lineages. Eur J Hum Genet 12:495–504

Cinnioğlu C, King R, Kivisild T, Kalfoğlu E, Atasoy S, Cavalleri GL, Lille AS, Roseman CC, Lin AA, Prince K, Oefner PJ, Shen P, Semino O, Cavalli-Sforza LL, Underhill PA (2004) Excavating Y-chromosome haplotype strata in Anatolia. Hum Genet 114: 127-148

Cordaux R, Saha N, Bentley GR, Aunger R, Sirajuddin SM, Stoneking M (2003) Mithchondrial DNA analysis reveals diverse histories of tribal populations from India. Eur J Hum Genet 11:253-264

Cordaux R, Weiss G, Saha N, Stoneking M (2004) The Northeast Indian Passageway: A Barrier or Corridor or Human Migrations. Mol Biol Evol 21:1525-1533

Cox MP (2006) Minimal hierarchical analysis of global human Y-chromosome SNP diversity by PCR-RFLP. Anthropol Sci 114:69-74

Cruciani F, Santolamazza P, Shen P, Macualay V, Moral P, Olckers A, Modiano D, Holmes S, Destro-Bisol G, Coia V, Wallace DC, Oefner PJ, Torroni A, Cavalli-Sforza LL, Scozzari R, Underhill PA (2002) A Back Migration from Asia to Sub-Saharan Africa Is Supported by: High-Resolution Analysis of Human Y-Chromosome Haplotypes. Am J Hum Genet 70:1197-1214

Da Costa Y (1983) A Politico-Geographical Image of the Kensington-Factreton Group Area in Cape Town. Masters Thesis, University of South Africa, South Africa

Da Costa Y and Davids A (1994) Pages from Cape Muslim history. Shuter and Shooter, Cape Town, pp 1-17

Da Costa Y (1994) Muslims in Greater Cape Town: A problem of identity. British J Sci.45:236-246

Davids A (1980) The mosques of the Bo-Kaap: a social history of Islam at the Cape. Cape and Transvaal Printers, South Africa, pp 31-41

Du Pre RH (1994) Separate but unequal. The 'Coloured'people of South Africa-A Political History. Jonathan Ball Publishers, Johannesburg, pp13-36

De Vincente MC and Fulton (2005) Using Molecular Marker Technology in Studies on Plant Genetic Diversity. Crop Sci 45:2676-2677

Edwin D, Vishwanathan H, Roy S, Usha Rani MV, Mujumder PP (2002) Mitochondrial DNA diversity among five tribal populations of southern India. Curr Sci 83:158-162

Flores C, Maca-Meyer N, Pérez JA, González AM, Larruga JM, Cabrera VM (2003) A Predominant European Ancestry of Paternal Lineages from Canary Islanders. Ann Hum Genet 67: 138-152

Fregel R, Gomes V, Gusmão L, González1 AM, Cabrera1 VM, Amorim A, Larruga JM (2009) Demographic history of Canary Islands male gene-pool: replacement of native lineages by European. BMC Evol Biol 9:1-14

Gonder MK, Mortensen HM, Reed FA, De Sousa A, Tishkoff SA (2007) Whole-mtDNA Genome Sequence Analysis of Ancient African Lineages. Mol Biol Evol 24:757–768

Gauniyal M, Chahal SMS, Kshatriya GK (2008) Genetic Affinities of the Siddis of South India: An Emigrant Population of East Africa. Hum Biol 80: 251–270

Hammer MF and Horai S (1995) Y Chromosomal DNA Variation and the Peopling of Japan. Am J Hum Genet 56: 951-962

Hammer MF, Karafet TM, Redd AL, Jarjanazi H, Santachiara-Benerecetti S, Soodyall H, Zegura SL (2001) Hierarchical Patterns of Global Human Y-Chromosome Diversity. Mol Biol Evol 18:1189-1203

Hammer MF and Zegura SL (2002) The Human Y Chromosome Haplogroup Tree: Nomenclature and Phylogeography of Its Major Divisions. Annu Rev Anthropol 31:303-321

Hammer MF, Chamberlin VF, Kearney V, Stover D, Zhang G, Karafet T, Walsh B, Redd AJ (2006) Population structure of Y chromosome SNP haplogroups in the United States and forensic implications for construction Y chromosome STR database. Forensic Sci Int 164: 45-55

Hurles ME, Sykes BC, Jobling MA, Forster P (2005) The Duel Origin of the Malagasy in Island Southeast Asia and East Africa: Evidence from Maternal and Paternal Lineages. Am J Hum Genet 76: 894–901

Hutchison LAD, Myres NM, Woodward, SR (2004) Growing the Family Tree: The Power of DNA in Reconstructing Family Relationships. Proceeding of the First Symposium on Bioinformatics and Biotechnology (BIOT-04, Colorado Springs)

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

Jobling MA and Tyler-Smith C (2003) The human Ychromosome: An evolutionary marker comes of age. Nat Rev Genet 4:598–612

Jobling MA, Hurles M, Tyler-Smith C (2003) Human evolutionary genetics: Origins, peoples and disease. Garland Science

Jorde LB, Bamshad M, Rogers AR (1998) Using mitochondrial and nuclear DNA markers to reconstruct human evolution. BioEssays 20:126-136

Karafet TM, Lansing JS, Redd AJ, Reznikova S, Watkins JC, Surata SPK, Arthawiguna WA, Mayer L, Bamshad M, Jorde JL, Hammer MF (2005) Balinese Y-Chromosome Perspective on the Peopling of Indonesia: Genetic Contributions from

Pre-Neolithic Hunter-Gatherers, Austronesian Farmers, and Indian Traders. Hum Biol 77: 93–114

Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF (2008) New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Genome Res 18:830-838

Kayser M, Brauer S, Weiss G, Schiefenhövel W, Underhill P, Shen P, Oefner P, Tommaseo-Ponzetta M, Stoneking M (2003) Reduced Y-chromosome, but Mitochondrial DNA, Diversity in Human Populations from West New Guinea. Am J Hum Genet.72: 281-302

Kayser M, Lao O, Anslinger K, Augustin C, Bargel G, Edelmann J, Elias S, Heinrich M, Henke J, Hohoff C, Illig A, Jonkisz A, Kizniar P, Lebioda A, Lessig R, Lewicki S, Maciejewska A, Monies DM , Pawlowski R, Poetsch M, Schmidt D, Schimdt U, Schneider PM, Stradmann-Bellinghausen B, Szibor R, Wegener R, Wozniak M, Zoledziewska M, Roewer L, Dobosz T, Ploski R (2005) Significant genetic differentiation between Poland and Germany follows present-day political borders, as revealed by Y-chromosome analysis. Hum Genet 117:428-443

Kharkov VN, Stepanov VA, Medvedeva OF, Spiridonova MG, Voevoda MI, Tadinova VN, Puzyrev VP (2007) Gene Pool Differences between Northern and Southern Altaians Inferred from the Data on Y-Chromosomal Haplogroups. Russ J Genet 43:551–562

Kivisild T, Kaldma K, Metspalu M, Parik J, Papiha S, Villems R (1999) The Place of the Indian mtDNA Variants in the Global Network of Maternal Lineages and the Peopling of the Old World. Genomic Diversity 135:152

Krithika S, Maji S, and Vasulu TS (2007) Molecular Genetic Perspective of Indian Populations: A Y-Chromosome Scenario. Anthropol Special 3: 385-392

Kumar V, Reddy ANS, Babu JP, Tipirisetti NR, Langstieh BT, Thangaraj K, Reddy AG, Singh L, Reddy BM (2007) Y-chromosome evidence suggests a common heritage of Austro-Asiatic populations. BMC Evol Biol 7:1-14

Lalueza Fox C, Sampietro ML, Gilbert MTP, Castri L, Facchini F, Pettener D, Bertranpetit J (2004) Unravelling migrations in the steppe: mitochondrial DNA sequences from ancient Central Asians. Proc. R Soc. Lond 271:941-947

Lell JT and Wallace DC (2000) The Peopling of Europe from the Maternal and Paternal Perspectives. Am J Hum Genet 67:1376-1381

Lind JM, Hutcheson-Dilks HB, Williams SW, Moore JH, Essex M, Ruiz-Pesini E, Wallace DC, Tishkoff SA, O'Brien SJ, Smith MW (2007) Elevated male European and female African contributions to the genomes of African American individuals. Hum Genet 120:713–722

Lodhi AY (1992) African Settlements in India. Nordic J Afr Studies 1: 83–86

Loogväli E, Roostalu U, Malyarchuk BA, Derenko MV, Kivisild T, Metspalu E, Tambets K, Reidla M, Tolk H, Parik J, Pennarun E, Laos S, Lunkina A, Golubenko M, Barać L,Peričić M, Balanosky OP, Gusar V, Khusnutdinova EK, Stepanov V, Puzyrev V, Rudan P, Balanovska EV, Grechanina E, Richard C, Moisan JP, Chaventré A, Anagnou NP, Pappa KI, Michalodimitrakis EN, Claustress M, Gölge M, Mikerezi I, Usanga E, Villems R (2004) Disuniting Uniformity: A Pied Cladistic Canvas of mtDNA Haplogroup H in Eurasia. Mol Biol Evol 21: 2012-2021

Luis JR, Rowold DJ, Regueiro M, Caeiro B, Cinnioğlu C, Roseman C, Underhill PA, Cavalli-Sforza LL, Herrera RJ (2004) The Levant versus the Horn of Africa: Evidence for Biodirectional Corridors of Human Migrations. Am J Hum Genet 74:532-544

Maca-Meyer N, Gonzalez AM, Larruga JM, Flores C, Cabrera VM (2001) Major genomic mitochondrial lineages delineate early human expansions. BMC Genet. 2:1471-2156

Macaulay V, Richards M, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonne-Tamir B, Sykes B, Torroni A (1999) The Emerging Tree of West Eurasian mtDNAs: A Synthesis of Control Region Sequences and RFLPs. Am J Hum Genet 64: 232-249

Mahida EM (1993) History of Muslims in South Africa: A Chronology. Kat Bros, Durban, pp 1-19

Maji S, Krithika S, Vasulu TS (2009) Phylogeographic distribution of mitochondrial DNA macrohaplogroup M in India. J Genet 88: 127–139

Manwaring N, Jones MM, Wang JJ, Rochtchina E, Mitchell P, Sue CM (2006) Prevalence of mitochondrial DNA haplogroup in an Australian population. Intl Med J 36:530-533

Martinez-Cruzado JC, Toro-Labrador G, Ho-Fung V, Estévez-Montero MA, Lobaina-Manzanet A, Padovani-Claudio DA, Sánchez-Cruz H, Ortiz-Bermúdez P, Sánchez-Crespo A (2001). Mitochondrial DNA Analysis Reveals Substantial Native American Ancestry in Puerto Rico. Hum Biol 73:491-511

Martinez Marignac VL, Bertoni B, Parra EJ, Bianchi NO (2004) Characterization of Admixture in an Urban Sample from Buenos Aires, Argentina, Using Uniparentally and Biparentally Inherited Genetic Markers. Hum Biol 76: 543.557

Matthee H (2008) Muslim identities and political strategies: A case study of Muslims in the greater Cape Town of South Africa, 1994-2000. Unidruckerei, University of Kassel, Germany

Merriwether DA, Friedlaender JS, Mediavilla J, Mgone C, Gentz F, Ferrell RE (1999) Mitochondrial DNA Variation Is an Indicator of Austronesian Influence in Island Melanesia. Am J Phys Anthropol 110: 243-270

Mishmar D, Ruis-Pesini E, Golik P, Macualay V, Clark AG, Hosseini S, Brandon M, Easley K, Chen E, Brown MD, Sukernik RI, Olckers A, Wallace DC (2003) Natural selection shaped regional mtDNA variation in humans. PNAS 100:171-176

Montiel R, Bettencourt C, Santos C, Prata MJ, Lima M (2005) Analysis of Y-chromosome Variability and it's Comparison with mtDNA Variability Reveals Different Demographic Histories Between Islands in the Azores Archipelago (Portugal). Annals Hum Genet 69: 135-144

Navajas M and Fenton B (2000) The application of molecular markers in the study of diversity in acarology: a review. Exp Appl Acarol 21:685-696

Nelson TM, Just RS, Loreille O, Schanfield MS, Podini D (2007) Development of a Multiplex Single Base Extension Assay for Mitochondrial DNA Haplogroup Typing. Croat Med J 48: 460- 472

Onofri V, Alessandrini F, Turchi C, Pesaresi M, Buscemi L, Tagliabracci A (2006) Development of multiplex PCRs for evolutionary and forensic applications of 37 human Y chromosome SNPs. Forensic Sci Int 157: 23-35

Parkendorf B and Stoneking M (2005) Mitochondrial DNA and Human Evolution. Annu Rev Genomics 6:165-183

Parker PG, Snow AA, Schug MD, Booton GC, Fuerst PA (1998) What molecules can tell us about populations: Choosing and using a molecular marker. Ecol 79:361-382

Piechota L, Tońska K, Nowak M, Kabzińska D, Lorenc A, Bartnik E (2004) Comparison between the Polish population and European populations on the basis of mitochondrial morphs and haplogroups. Acta Biochimica Polonica 51:883-895

Quintáns B, Álvarez-Iglesias V, Salas A, Lareu M, Carracedo Á (2004) Typing mtDNA SNPs of forensic and population interest with snapshot. Int Congress Series 1261: 419-421.

Richards M and Macaulay V (2001) The Mitochondrial Gene Tree Comes of Age. Am J Hum Genet 68:1315-1320

Rootsi S (2004) Human Y Chromosomal Variation in European Populations. Masters Thesis, University of Tartu, Estonia

Rootsi S, Magri C, Kivisild T, Benuzzi G, Help H, Bermisheva M, Kutuev I, Barac´ L, Peric˘ic´M, Balanovsky O, Pshenichnov A, Dion D, Grobei M, Zhivotovsky LA, Battaglia V, Achilli A, Al-Zahery N, Parik J, King R, Cinniog˘lu C, Khusnutdinova E, Rudan P, Balanovska E, Scheffrahn W, Simonescu M, Brehm A, Goncalves R, Rosa A, Moisan JP, Chaventre A, Ferak V, Fu¨redi S, Oefner PJ, Shen P, Beckman L, Mikerezi I, Terzic´R, Primorac D, Cambon-Thomsen A, Krumina A, Torroni A, Underhill PA, Santachiara-Benerecetti AS, Villems R, Semino O (2004) Phylogeography of Y-Chromosome Haplogroup I Reveals Distinct Domains of Prehistoric Gene Flow in Europe. Am J Hum Genet 75:128-137

Roychoudhury S, Roy S, Badal D, Chakraborty M, Roy M, Roy B, Ramesh A, Prabhakaran, N, Usha Rani MV, Vishwanathan H, Mitra M, Sil SK, Majumder PP (2000) Fundamental genomic unity of ethnic India is revealed by analysis of mitochondrial DNA. Curr Sci 79: 1182-1192

Rubicz R, Melton PE, and Crawford MH (2007) Molecular Markers in Anthropological Studies. In MH Crawford (ed.): Anthropological Genetics: Theory, Method and Applications. Cambridge University Press, Cambridge

Sahoo S, Singh A, Himabindu G, Banerjee J, Sitalaximi T, Gaikwad S, Trivedi R, Endicott P, Kivisild T, Metspalu M, Villems R, Kashyap VK (2006) A prehistory of Indian Y chromosomes:Evaluating demic diffusion scenarios. PNAS.103:843:848

Salas A, Richards M, De la Fe T, Lareu MV, Sobrino B, Sànchez-Diz P, Macaulay V, Carracedo Á (2002) The Making of the African mtDNA Landscape. Am J Hum Genet.71:1082-1111

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

Santos C, Montiel R, Angles N, Lima M, Francalacci P, Malgosa A, Abade A, Aluja MP (2004) Determination of Human Caucasian Mitochondrial DNA by Means of a Hierarchical Approach. Hum Biol 76: 431-453

Schlötterer C (2004) The evolution of molecular markers −just a matter of fashion? Nat Rev Genets 5: 63-69

Schurr TG (2000) Mitochondrial DNA and the Peopling of the New World. Proceedings of the National Academy of Sciences 91:1158-1162

Schurr TG and Wallace DC (2002) Mitochondrial DNA Diversity in Southeast Asian Populations. Hum Biol 74: 431-452

Shell RCH (2000) Islam in Southern Africa, 1658-1998.In: Nehemia Levtzon and Randall Pouwels (Eds). Islam in Africa. Ohio University Press, Athens

Semino O, Magri C, Benuzzi G, Lin AA, Al-Zackery N, Battaglia V, Maccioni L, Triantaphyllidis C, Shen P, Oefner PJ, Zhivotovsky LA, King R, Torroni A, Cavalli-Sforza LL, Underhill PA, Santachiara-Benerecetti (2004) Origin, Diffusion and Differentiation of Y-Chromoosme Haplogroup E and J: Inferences on the Neolithization of Europe and Later Migratory Events in the Mediterranean Area. Am J Hum Genet 74:1023-1034

Serk P (2004) Human mitochondrial DNA Haplogroup J in Europe and Near East. Masters Thesis, University of Tartu, Estonia

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

Soodyall H, Vigilant L, Hill AV, Stoneking M, Jenkins T (1996) mtDNA Control-Region Sequence Variation Suggests Multiple Independent Origins of an "Asian-Specific" 9-bp Deletion in Sub-Saharan Africans. Am J Hum Genet 58:595-608

Stone AC and Stoneking M (1998) MtDNA Analysis of a Prehistoric Oneata Population: Implications for the Peopling of the New World. Am J Hum Genet 62:1153-1170

Sunnucks P (2000) Efficient genetic markers for population biology. Tree 15:199-203

Skaletsky H, Kuroda-Kawaguchi T, Mix PJ, Cordum HS, Hillier L, Brown LG, Repping S et al (2003) The male specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nat 423:825-837

Tishkoff S and Kidd KK (2004) Implications of biogeography of human populations for 'race' and medicine. Nat Genet 36: S21-S27

Tolk HV, Barac L, Pericic M, Klaric IM, Janicijevic B, Campbell H, Rudan I, Kivisild T,Villems R, Rudan P (2001) The evidence of mtDNA haplogroup F in a European population and its ethnohistoric implications. Eur J Hum Gen 9:717-723

Torres JB, Kittles RA and Stone AC (2007) Mitochondrial and Y Chromosome Diversity in the English-Speaking Caribbean. Ann Hum Genetics (2007) 71:1–9

Torroni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, Obinu D, Savontaus M, Wallace DC (1996) Classification of European mtDNAs From an Analysis of Three European Populations. Genet 144: 1835-1850

Torroni A, Rengo C, Guida V, Cruciani F, Sellitto D, Coppa A. Calderon FL, Simionati B, Valle G, Richards M, Macualay V, Scozzari R (2001) Do the Four Clades of the mtDNA Haplogroup L2 Evolve at Different Rates?. Am J Hum Genet 69:1348-1356

Tripathy V, Nirmala A, Reddy BM (2008) Trends in Molecular Anthropological Studies in India. Int J Hum Genet 8:1-20

Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T,Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ (1997) Detection of Numerous Y Chromosome Biallelic Polymorphisms by Denaturing High- Performance Liquid Chromatography. Genome Res 7:996-1005

Underhill PA, Shen P, Lin AA, Jin L, Passarino G,Yang WH, Kauffman E, Bonne-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ (2000) Y chromosome sequence variation and the history of human populations. Nat Genet 26:358–361

Underhill PA, Passarino G, Lin AA, Shen P, Lahr MM, Foley RA, Oefner PJ, Cavalli-Sforza LL (2001) The phyleogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann Hum Genet 65:43-62

Underhill PA (2003) Inferring Human History: Clues of Y Chromosome Haplotypes Cold Spring Habour. Symposia on Quantitative Biology, Vol LXVIII, Cold Spring Harbor Laboratory Press: 487-493

Underhill PA (2005) A synopsis of extant Y chromosome diversity in East Asia and Oceania: The peopling of East Asia: putting together archaeology, linguistics and genetics. Routledge Curzon, London

Underhill PA and Kivisild T (2007) Use of Y Chromosome and Mitochondrial DNA Population Structure in Tracing Human Migrations. Annu Rev Genet 41:539–64

Vallone PM and Butler JM (2004) Y-SNP Typing of U.S. African American and Caucasian Samples Using Allele-Specific Hybridization and Primer Extension. J Forensic Sci 49:723-732

Vallone PM, Just RS, Coble MD, Butler JM, Parsons TJ (2004) A multiplex allele-specific primer extension assay for forensically informative SNPs distributed throughout the mitochondrial genome. Int J Legal Med 118:147:157

Van der Ross RE (2005) Up from Slavery Slaves at the Cape their origins, treatment and contribution. Ampersand Press, Cape Town

Van Oven M and Kayser M (2008) Updated Comprehensive Phylogenetic Tree of Global Human Mitochondrial DNA Variation. Mutuation in Brief 29:E386-E394

Wallace DC, Brown MD, Lott MT (1999) Mitochondrial DNA variation in human evolution and disease. Gene 238:211-230

Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, Jin L, Su B, Pitchappan R, Shanmugalakshmi S, Balakrishan K, Read M, Pearson NM, Zerjal T, WebsterMT, Zholoshvili I, Jamariashvili E, Gambarov S, Nikbin B, Dostiev A, Aknazarov O, Zalloua P, Tsoy I, Kitaev M, Mirrakhimov M, Chariev A, Bodmer WF (2001) The Eurasian Heartland: A continental perspective on Y-chromosome diversity. PNAS 98:10244-10249

Wen B, Li H, Gao S, Mao X, Gao Y, Li F, Zhang F, He Y, Dong Y, Zhang Y, Huang W, Jin J, Xiao C, Lu D, Chakraborty R, Su B, Deka R, Jin L (2005) Genetic Structure of Hmong-Mien Speaking Populations in East Asia as Revealed by mtDNA Lineages. Mol Biol Evol 22:725-734

Wood ET, Stover DA, Ehret C, Destro-Bisol G, Spedini G, Mcleod H, Louie L, Bamshad M, Strassman BI, Soodyall H, Hammer MF (2005) Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes. Eur J Hum Genet 13:867-876

Yonggang Y and Yaping Z (2003) Pitfalls in the analysis of ancient human mtDNA. Chinese Sci Bulletin 48:826-830

Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L Santos FR, Schiefenhovel W, Fretwell N, Jobling MA, Harihara S, Shimizu K, Semjidmaa D, Sajantila A, Salo P, Crawford MH, Ginter EK, Evgrafov OV, Tyler-Smith C (1997) Genetic Relationships of Asians and Northern Europeans, Revealed by Y-Chromosomal DNA Analysis. Am J Hum Genet 60:1174-1183

Electronic Supplementary Resources

Mitomap (2002) Human MtDNA Migrations [online]. Available from:www.mitomap.org/mitomap/WorldMigrations.pdf [11 Oct 2009].

National Geographic (2009) Population Genetics [online]. Available from: genographic.nationalgeographic.com/genographic/index.html [11 Oct 2009].

Parsons TJ (2006) Mitochondrial DNA Genome Sequencing and SNP Assay Development for Increased Power of Discrimination [online]. Available from:www.ncjrs.gov/pdffiles1/nij/grants/213502.pdf [21 Sept 2009].

Appendix I: Protocols

DNA Extraction from buccal swabs using Epicenter protocol

1. Place the buccal swap tip into 250ul of Epicenter extraction buffer after removing it aseptically with a sterile scalpel blade.

2. Then vortex for 15 seconds and heat for 5 minutes at 65°C

3. Thereafter vortex for 15 seconds and heat for 2 minutes at 98°C

4. Remove the swab and dispose into a medical waste container

5. DNA extract can then be quantified and then stored at -20 C

DNA Purification using the Phenol Chloroform method

1. An equal volume of Phenol/Chloroform/Isoamyl Alcohol (25:24:1) was added to the DNA solution to be purified (100µl to 400µl) in a 1.5ml eppendorf tube. A volume of 40µl of DNA solution was used which was made-up to a final volume of 100µl by adding TE buffer to reduce sample loss during the extraction procedure.

2. The appropriately labelled tube was then shaken for 5 minutes.

3. The tube was then placed in a bench-top centrifuge (Eppendorf, 5415D) and the mixture was centrifuged at 13200rpm for 2-5minutes at room temperature.

4. The top aqueous phase was then carefully transferred to a clean appropriately labeled tube using a 200µl pipette.

5. To this, an equal volume of chloroform/isoamyl alcohol (24:1) was added.

6. The tube was then shaken for 2 minutes and centrifuged for 5minutes at 1000rpm.

7. The top aqueous phase was then transferred to a clean-labelled eppendorf tube.

8. $1/10^{th}$ volume of 3M Sodium acetate pH 5.5 was added to the mixture and mixed briefly

9. Thereafter, 1.25 volume of isopropanol was added ,mixed and left at -70 °C for 30 minutes alternatively it could be left overnight at -20 °C

10. The tube was then spun in a centrifuge at 8000-10000rpm for 15 minutes.

11. The supernatant was then carefully removed as not to disturb the DNA pellet

12. 200µl of cold 70% Ethanol was added to the DNA pellet which was then centrifuged for 5 minutes.

13. After centrifugation the 200µl of supernatant was removed carefully from the tube without disturbing the DNA pellet. The remaining ethanol was removed by allowing it to evaporate in a 37°C water bath or by leaving it on a bench desk for 10 minutes.

14. Once the ethanol was evaporated the DNA pellet was resuspended in the desired volume of TE buffer.

Electrophoresis on the ABI PRISM 3130xl Genetic Analyzer

Preparation of 96 well Plate

1. Use a clean 96 well plate when preparing purified minisequencing products for electrophoresis on the ABI PRISM 3130xl Genetic Analyzer. These preparations are always performed on ice to minimise degradation of samples and reagents.

2. To each well aliquot the desired volume of minisequencing product undergoing analysis. A sample sheet is used to record the arrangement and description of samples.

3. To this loading mixture is added containing 0.4µl of size standard (Liz 120) and HiDi ™ Formamide (HiDi ™ Formamide volume may vary depending on the minisequencing product used for analysis).

4. Sample arrangement is always co-ordinated in rows of two as the Genetic Analyzer is a 16 capillary system allowing 16 samples to be analyzed simultaneously. Therefore when an uneven number of samples (e.g. eight samples) needs to be analyzed fill the remaining wells with equal volumes of HiDi ™ Formamide.

5. Once the plate has been setup, place a plate septa or covering over.

6. Plates may be made-up on the run date or alternatively may be setup a few days prior to run date and stored in the -4°C freezer.

Setting up the Genetic Analyzer

1. Press restart all on the service console allocated on the desktop of ABI computer (lights will change to amber then green)

2. Once this has been achieved press the button allocated on the ABI machine which moves the buffer and water dispenser forward

3. Buffer and water containers should be changed daily, however may be changed every alternate day.

4. Discard the water and 1X buffer then rinse containers with distilled water. This is followed by wiping the inside surface with paper towel thereafter replenishing these containers with the specified liquid. In addition, ensure that the Genetic Analyzer is setup with a 36cm capillary array and POP -7 polymer

5. Prepare the plate run on the Genetic Analyzer by clicking on the ga 3130xl icon then proceed to Plate manager.

6. At plate manager select new, a screen will appear with the following questions

this needs to be completed in the following manner: **Name** S.Isaacs, **Description**: MtDNA Multiplex I, Application: GeneMapper-UWC-1UU01VLCILX, Plate Type: 96-well, **Owners Name** S.Isaacs, **Operator Name**: S.Isaacs. Bold letters indicate fields that will vary according to user.

7. A run sheet will then automatically appear list all the samples being analyzed and complete the run sheet as indicated: Size Standard: GS120LIZ, Panel: None, Analysis: SNaPshot Default, Result Group Forensics_SNPs and Instrument Protocol: Snapshot1_JM. Save the run sheet as this creates a plate for analysis. During this time samples in the 96 well plate may also be denatured at 95˚C for 5 min and thereafter be kept on hold at 4˚C or placed on ice.

8. Proceed to Run Scheduler, select Plate View and locate the plate recently saved once the plate is located, select it then assemble the standard 96-well plate for electrophoresis. Place the 96-well plate on a plate adapter ensure the notch at A1 matches to the plate adapter. Press down on all four corners of the adapter to ensure a proper seating. Then use 96 well plate septa to seal samples and load plate onto the platform designated on the Genetic Analyzer.

9. Link the plate to the Genetic Analyzer by selecting either bay A or B (if only one plate is processed select A) if this is done successfully the bay will appear green. Thereafter, return to **Plate View** (check if the correct lanes were chosen for analyses) then click the green arrow located on the top left corner of toolbar as this will initiate the run

10. Once the run is completed retrieve the run folder

Processing and Genotyping of Samples from the 3130xl Genetic Analyzer

1. Open the Genemapper software select new project and the SNaPshot analysis format.

2. Select file and import the relevant run folder then click the analyze tab to analyze samples.

3. Once analysis has been completed examine each individual sample to ensure that all the SNP markers have been detected. During this stage samples may be genotyped as well.

4. Genotyping is done manually by comparing the observed SNP to the two allelic states known to occur for a particular SNP marker.

MtDNA PCR Reactions

Preparing primer mixtures from 100µM stocks:

From the 100µM primer stocks 10µl of the forward and reverse primer is aliquoted into a labelled tube, which subsequently is used to setup 10X primer mix

10x primer mix for Multiplex I

8272-8280	5µM	(from the 50 uM F& R mix take 5 µl)
12406	5µM	
7600	5µM	
10400/10398	5µM	
3594	5µM	
7028	5µM	

Dilute in 20µl H_20

10x primer mix for Multiplex II

13263	5µM	(from the 50 uM F& R mix take 5 µl)
1719	5µM	
5178	5µM	
4833	5µM	
663	5µM	

Dilute in 25µl H_20

Y-Chromosome PCR Reactions

Preparing primer mixtures from 100µM stocks:

From the 100µM, primer stocks 10 µl of the forward and reverse primer is aliquoted into a labelled tube, which subsequently is used to setup 10X primer mix

10x primer mix for Multiplex I

M170	5µM	(from the 50 uM F& R mix take 5 µl)
M173	5µM	
M45	5µM	
M89	5µM	
Dilute in	30µl H_20	

10x primer mix for Multiplex II

M52	5µM	(from the 50 uM F& R mix take 5 µl)
M96	5µM	
M181	5µM	
M91	5µM	
Dilute in	30µl H_20	

10x primer mix for Multiplex III

M172	5µM	(from the 50 uM F& R mix take 5 µl)
M9	5µM	
Dilute in	40µl H_20	

10x primer mix for Multiplex IV

M216	5µM	(from the 50 uM F& R mix take 5 µl)
M201	5µM	
M214	5µM	
Dilute in	35µl H_20	

Appendix II. Composition of Buffers and Solutions

Commercial Stock

10x 3730 Running Buffer with EDTA

Phenol /Chloroform/Isoamyl alcohol

Chloroform/ Isoamyl alcohol

Isopropanol

10X TBE Buffer (1L)

108g Tris Base EDTA
~55g Boric Acid

Add Tris Base, EDTA and Boric Acid to 700mL dH_2O and mix to dissolve. Thereafter makeup to the final volume of 1L

1X TBE Buffer (1L)

100mL 10X TBE Buffer

add TBE to 900mL dH_2O to makeup the final volume to 1L.

70% Ethanol (500mL)

350 mL 100% Ethanol

Add 150 ml dH_2O to makeup the final volume of 500mL

TE Buffer

Dissolve 1ml of Tris HCl (1M) in 200µl of EDTA (ph 8.0). Add 100ml of dH2O. The ph of the solution should be 8.0

3M Sodium Acetate (500mL)

3M Sodium Acetate

123.052g Sodium Acetate (anhydrous)

400 mL dH20

pH to 5.5 then adjust volume to 500mL with distilled thereafter autoclave solution