

**Reconstruction of major male and female lineages of the
Strand 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*



Supervisor: Prof. Mongi Benjeddou (PhD)

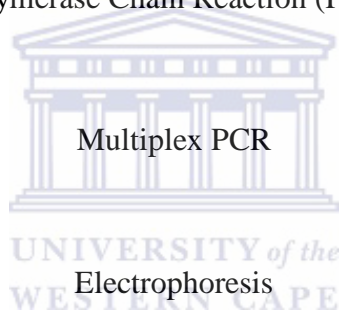
Keywords

Forensics

(SNP)

Haplogroup (Hg)

Polymerase Chain Reaction (PCR)



Electropherogram

Genotyping

Polymorphism

Abstract

Reconstruction of major male and female lineages of the Strand Muslim community

T. Geduld

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

There are currently two theories regarding the origins of the Strand Muslim community. Some historians believe that the community was founded by the descendents of the followers of Sheikh Yusuf. Other scholars, however, suggest that Imam Abdol Sammat and other free blacks of Indonesian origin might have founded this community. According to the latter theory, this community's founders might have moved from Cape Town to the interior to settle on farms in the Hottentots Holland basin before finally making home at the Strand (formally known as Mosterd Bay). This study was therefore carried out to provide new type of evidence regarding the origins and the development of the Strand Muslim community using molecular genetics techniques.

Initially, a pilot study was carried out in order to reconstruct the major paternal and maternal lineages of the Muslim population living in the Cape metropolitan area. The Study has shown the ability of molecular genetic tools to give insight into the origins and history of local communities. The study was also used as a point of reference for the Strand Muslim Community project. Genetic variations of the Y-chromosome and mitochondrial DNA for the pilot study were analyzed using the RFLP technique.

The SNaPshot mini-sequencing technique was used to genotype single nucleotide polymorphisms (SNP) on the Y-chromosome and mitochondrial DNA in 115 males from the Strand Muslim community.

Y-chromosome DNA variations were investigated in the Strand Muslim community using 14 binary markers located in the non-recombining region of Y-chromosome (NRY). Twelve haplogroups were found in the population. Asian haplogroups dominated the lineages found in the population's Y-pool with haplogroup NO-M214 occurring at the greatest frequency. African haplogroups were least observed. The African lineage with highest frequency was E-M96. European lineages were found at moderate frequencies with R1-M173 dominating the European category.



Mitochondrial DNA variations were examined by analysis of transition positions at sites located in the coding region. Seven haplogroups out of 14 possible haplogroups were detected. The mtDNA study of the Muslim Strand community has revealed a high matrilineal contribution of Africans and Asians. The African haplogroups, however, dominated the maternal ancestry with haplogroup L1/L2 occurring within 55 individuals. The presence of haplogroups L1/L2 and L3 clearly shows the importance of the African contribution in the maternal gene pool of the Strand Muslim community, whereas the presence of Asian haplogroups M, N, D and B in the population is likely due to a gene flow from across Asia.

Declaration

I declare that ‘Reconstruction of major male and female lineages of the Strand Muslim community’ is my own work, that it has not been submitted for any degree or examination in any other university and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Full Name: Tasneem Geduld

Signed:

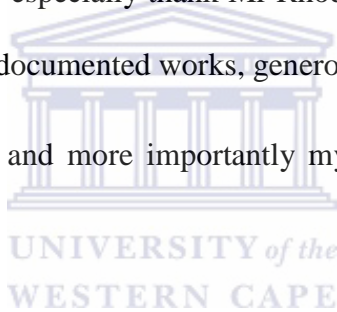


Date: September 2010

Acknowledgements

I would like to thank my research supervisor Professor Mongi Benjeddou for his guidance throughout this project. I would like to acknowledge and thank the National Research Foundation (NRF) for financial assistance. Many a thanks to all the men who generously donated biological material for the study. My gratitude also goes to all the students and staff who made working in the Department of Biotechnology such a pleasant experience. I would also like to especially thank Mr Rhoda a well known historian of the Strand community for his well documented works, generosity and assistance.

Thanks also go to my friends and more importantly my family and husband for their support and encouragement.



List of Abbreviations

DNA	Deoxyribonucleic Acid
SNP	Single Nucleotide Polymorphism
Y-STR	Y-chromosome STR
PAR	Pseudo-Autosomal Region
MtDNA	Mitochondrial DNA
NRY	Non-recombining Y-chromosome
YCC	Y-chromosome Consortium
CRS	Cambridge Reference Sequence
PCR	Polymerase Chain Reaction
HCl	Hydrochloric Acid
KCl	Potassium Chloride
MgCl ₂	Magnesium Chloride
dNTPs	deoxy Nucleotide TriPhosphates
dATP	deoxy Adenosine TriPhosphate
BSA	Bovine Serum Albumin
RFU	Relative Fluorescent Units
RFLP	Restriction Fragment Length Polymorphism

List of tables

Table 1.1. Historical documented data demonstrating the number and place of origin of slaves brought to the Cape in 1658- 1700

Table 1.2. South African population group percentages estimated for 2009 (adapted from www.statssa.gov).

Table 1.3. Western Cape population group percentages and percentage of Muslim obtained from 2001 census (adapted from www.statssa.gov).

Table 2.1. Y-chromosomal SNP amplification primer sequences, annealing temperatures and PCR fragment sizes. # indicates touchdown PCR

Table 2.2. Y-chromosomal SNP thermocycling conditions.

Table 2.3. List of Y-chromosomal haplogroups, defining markers, restriction enzymes with optimal temperature and the digestion product sizes. Lack of restriction site is defined as ancestral and presence of restriction site and SNP is defined as derived.

Table 2.4. Mitochondrial SNP amplification primer sequences, annealing temperatures and PCR fragment sizes.

Table 2.5. List of mitochondrial haplogroups, defining markers, restriction enzymes with optimal temperature and digestion product sizes. Lack of restriction site is defined as ancestral and presence of restriction site and SNP is defined as derived.

Table 3.1. Y-chromosomal multiplex panels depicting discriminatory SNPs defining specific haplogroups organized into four multiplexes named MY1, MY2, MY3 and MY4 and one singleplex named SY⁵.

Table 3.2. Y-chromosomal PCR conditions for MY1, MY2, MY3 and MY4

Table 3.3. Y-chromosomal PCR conditions for SY5

Table 3.4. Y-chromosomal extension primer sequences, final concentration of primers used in multiplex reactions MY1 and MY2 and primer length

Table 3.5. Y-chromosomal extension primer sequences, final concentration of primers used in multiplex reactions MY3 and MY4 and primer length

Table 3.6. Y-chromosomal extension primer sequences, final concentration of primers used in singleplex reaction SY⁵ and primer length

Table 4.1. Mitochondrial PCR amplification multiplex panels with organized specific markers used in the study.

Table 4.2. Mitochondrial PCR conditions for M1 and M2

Table 4.3. Mitochondrial extension primer sequences, primer length and final concentration of primers used in multiplex reactions M1 and M2.

Table 4.4. List of continent-specific mitochondrial DNA haplogroups observed within in the Strand Muslim community.

Table 5.1. Table depicting the number cross cultural relationship between males and females observed within in the Strand Muslim community,

List of Figures

Figure 1.1. Human mitochondrial DNA representing control region/D-Loop and HV regions I and II adapted from (http://www.nfstc.org/pdi/Subject09/pdi_s09_m02_01_a.htm).

Figure 1.2. Male specific region of the Y-chromosome (adapted from Skaletsky *et al.* 2003)
a: Schematic representation of the whole chromosome, including the pseudoautosomal and heterochromatic regions
b: Enlarged view of 24Mb portion of euchromatic region of the NRY, 1Mb bar indicates the scale of diagram

Figure 1.3. Single Nucleotide Polymorphisms are single-nucleotide mutations in the DNA sequence: shaded region indicates an A to G transition (adapted and altered from Hutchison *et al.* 2004).

Figure 1.4. Example of an allele SNP extension electropherogram depicted SNP in respective colour and peak size (adapted from Onofri *et al.*, 2006)

Figure 1.5. World mitochondrial haplogroups and mutations (adapted from mitomap: a human mitochondrial genome database, <http://www.mitomap.org>. 2008)

Figure 1.6. Y-chromosomal evolutionary tree indicating haplogroups and major haplogroups designated origin and migration Adapted from (<http://www.familytreedna.com/public/china>).

Figure 2.1. Paternal phylogenetic tree indicating haplogroups and examined markers indicated in red boxes for lineage assignment, marker denoted in the green box indicates marker not depicted but used for the assignment of haplogroup H. (adapted from <http://genographic.nationalgeographic.com/genographic/lan/en/population.html>).

Figure 2.2. Maternal phylogenetic tree indicating haplogroups and examined markers for lineage assignment. Haplogroup-defining polymorphisms, relative to CRS are preceded by (+) derived and (-) ancestral and followed by the Restriction Enzyme used.

Figure 2.3. PCR-RFLP test for Y-chromosome lineage K-M9. Lane designated M is a DNA ladder (marker) with 50bp increments. Lane 1 shows positive control undigested. Lanes 2-4 as

well as lanes 6 and 7 indicates *HinfI* digested PCR products from four unrelated males; fragment size of 164bp defines derived state of the marker. Lanes 4 and 5 indicated partially digested products. Lanes 8-9 indicate DNA fragments of 100 and 64 base pairs and is therefore indicative of the ancestral state of marker M9.

Figure 2.4. PCR electrophoresis of marker YAP indicating correct fragment size. Lane indicated by the letter M is a DNA sizing ladder (marker); Lane 1 shows negative PCR control. Lanes 1-11 shows expected PCR products of 10 unrelated males. Lanes 3 and 8 indicates derived state of marker YAP with expected size of 455bp.

Figure 2.5. PCR-RFLP test for mtDNA lineage M (10397). Lane designated M is a DNA ladder (marker) with 50bp increments. Lane 1 shows positive control undigested. Lanes 2-8 indicates a DNA fragment of 276bp is therefore indicative of the ancestral state of marker 10397. Lanes 9-11 indicated *AluI* digested PCR products from three unrelated males; fragment size of 147bp and 129bp indicates the derived state of the marker.

Figure 2.6. PCR-RFLP test for mtDNA lineage B (9bp Del region V). Lane designated M is a DNA ladder (marker) with 50bp increments. Lane 1 shows negative control. Lanes 2-5 indicates a DNA fragment of 120bp and is therefore indicative of the ancestral state of the marker. Lane 6 indicates a 100bp DNA ladder (*Fermentas*) and Lanes 7-11 indicates PCR products from five unrelated males of sizes 120bp; there were no positive results for this marker.

Figure 2.7. A pie chart representation of the percentage of paternal haplogroup frequencies observed in a 100 unrelated males from the Cape Muslim Community.

Figure 2.8. A pie chart representation of the percentage of maternal haplogroup frequencies observed in a 100 unrelated males from the Cape Muslim Community.

Figure 2.9. Cape Muslim donor survey information depicting the percentage of Muslim fathers, mothers, paternal grandfathers and maternal grandmothers.

Figure 2.10. Cape Muslim donor survey information depicting the percentage of fathers, mothers, paternal grandfathers and maternal grandmothers born in the Cape.

Figure 2.11. Cape Muslim donor survey information depicting the percentage of donors classified as Malay, Coloured, African, Asian, Muslim, Arab and not indicated.

Figure 2.12. Donor survey data obtained from Cape Muslim male donors indicating perceived ethnic groups of paternal and maternal lines.

Figure 2.13. A) Overall percentage of African, Asian and European Y-chromosomal haplogroups obtained in the sample population of the Cape Muslim community. B) Donor survey data obtained from males of the Cape Muslim Community depicting perceived ethnicity and true genetic ethnicity through Y-haplogroups assignment.

Figure 2.14. A) Overall percentage of African, Asian and European mtDNA haplogroups obtained in the sample population of the Cape Muslim community. B) Donor survey data obtained from males of the Cape Muslim Community depicting perceived ethnicity and true genetic ethnicity through mtDNA-haplogroups assignment.

Figure 3.1. Y-chromosomal tree depicting haplogroups A-R. Colour boxes indicate haplogroups investigated. Yellow: African haplogroups, Red: Asian haplogroups, Blue: European haplogroups and Black indicating Eurasian/European haplogroups (adapted from <http://www.genebase.com/image/ydnaSnpBackboneChart01.jpg>)

Figure 3.2. Electropherograms of MY1: Four-marker multiplex analysed with *Gene mapper 4.0 software*. X-axis represents the size (bp) of the primers with the incorporated nucleotide. The Y-axis corresponds to the RFU of the peak. Derived alleles are indicated with their respective colour for each peak where green represents nucleotide A, black: nucleotide C, blue: nucleotide G and red: nucleotide T. **A)** sample indicating derived state for marker R1-M173, **B)** sample indicating derived state for marker F-M89, **C)** sample indicating derive state for marker P-M45, **D)** sample indicating derive state for marker I-M170 and **E)** sample indicating no derive state for any marker within MY1.

Figure 3.3. Electropherograms of samples tested for MY2, a four marker multiplex analysed with *Gene Mapper 4.0 software*. X-axis represents the size (bp) of the primers with the incorporated nucleotide. The Y-axis corresponds to the RFU of the peak. Derived alleles are indicated with their respective colour for each peak where green represents nucleotide A, black: nucleotide C, blue: nucleotide G and red: nucleotide T. **a)** sample with derived state for marker E-M96, **b)** sample with derived state for marker B-M181, **c)** sample with derived state for marker A-M91 and **d)** sample that did not test positive for any of the markers tested within this multiplex.

Figure 3.4. Electropherograms of MY3 consisting of two markers. X-axis represents the size (bp) of the primers with the incorporated nucleotide. The Y-axis corresponds to the RFU of the peak. Derived alleles are indicated with their respective colour for each peak where green represents nucleotide A, black: nucleotide C, blue: nucleotide G and red: nucleotide T. **a)** sample indicating derived state for marker K-M9 and **b)** sample indicating derived state for marker J2-M172.

Figure 3.5. Electropherograms of three markers multiplex MY4 analysed with Gene Mapper 4.0. X-axis represents the size (bp) of the primers with the incorporated nucleotide. The Y-axis corresponds to the RFU of the peak. Derived alleles are indicated with the respected colours where green represents nucleotide A, black: nucleotide C, blue: nucleotide G and red: nucleotide T. **a)** sample indicating derive state for marker NO-M214, **b)** sample indicating derive state for marker C-M216 and **c)** sample that does not carry any of the derived state for MY4.

Figure 3.6. Electropherograms of SY5 analysed with Gene Mapper 4.0. X-axis represents the size (bp) of the primers with the incorporated nucleotide. The Y-axis corresponds to the RFU of the peak. Derived alleles are indicated with the respected colours where green represents nucleotide A, black: nucleotide C, blue: nucleotide G and red: nucleotide T. **a)** sample indicating ancestral state for marker D-M174, **b)** sample indicating derive state for marker D-M174

Figure 3.7. A pie chart representation of the percentage of paternal haplogroup frequencies observed in a 115 unrelated males from the Strand Muslim Community.

Figure 3.8. Representation of Y chromosomal haplogroup dominance observed within the Strand Muslim Community expressed in percentage. **A)** African haplogroups observed indicating E-M96 as the dominant African haplogroup found. **B)** Asian haplogroups observed indicating NO-M214 as the dominant Asian haplogroup found. **C)** European haplogroups observed indicating R1-M173 as the dominant European haplogroup found.

Figure 4.1. MtDNA haplogroups A-N. Colour circles indicating haplogroups prevalent on various continents. Peach: African Haplogroups, Blue: Asian Haplogroups and Green: European Haplogroups (adapted from <http://www.mitomap.org>)

Figure 4.2: Electropherograms of M1: Plots of fragment size (x-axis) relative to -120 Liz size standard (*Applied Biosystems*) and relative fluorescents units (RFUs, y-axis). Seven-marker multiplex analyzed with *Gene mapper 4.0 software*. Derived alleles are indicated with the respective colour for each peak where green represents nucleotide A, black: nucleotide C, blue: nucleotide G and red: nucleotide T. **A)** Sample indicating result typical for haplogroup B assignment, **B)** sample indicating result typical for haplogroup H assignment and **C)** sample indicating result typical for haplogroup L1\L2 assignment.

Figure 4.3: Electropherograms of M1 and M2: Plots of fragment size (x-axis) relative to -120liz size standard (*Applied Biosystems*) and relative fluorescents units (RFUs, y-axis).Seven-marker multiplex analyzed with *Gene mapper 4.0 software*. Derived alleles are indicated with the respective colour for each peak where green represents nucleotide A, black: nucleotide C, blue: nucleotide G and red: nucleotide T. **D)** Sample indicating result typical for haplogroup N assignment, **E)** sample indicating result typical for haplogroup M assignment and **F)** sample indicating result typical for haplogroup L3 assignment .**G)** Electropherogram associated with sample being typed in M2

Figure 4.4: A pie chart representation of the percentage of maternal haplogroup frequencies and geographic origin observed in a 115 unrelated males from the Strand Muslim Community.

Figure 5.1. Strand Muslim donor survey information depicting the percentage of Muslim fathers, mothers, paternal grandfathers and maternal grandmothers of Strand Muslim donors.

Figure 5.2. Strand Muslim donor survey depicting the percentage of fathers, mothers, paternal grandfathers and maternal grandmothers born in the Strand.

Figure 5.3. Strand Muslim donor survey information depicting the percentage of donors who defined their ethnic origin as Coloured, Malay, and Indian and not indicated.

Figure 5.4. Donor survey data obtained from Strand Muslim male donors indicating the perceived ethnic origin of paternal and maternal lines.

Figures 5.5. A): Donor survey data obtained from males of the Strand Muslim Community depicting perceived ethnicity and true genetic ethnicity through Y-haplogroups assignment. **B)** Donor survey data obtained from males of the Strand Muslim Community depicting perceived ethnicity and true genetic ethnicity through mtDNA haplogroup assignment.

Figure 5.6.Overall percentage of African, Asian and European Y-chromosomal haplogroups obtained in the sample population of the Strand Muslim community.

Figure 5.7. Overall Asian, African and European contribution to Strand mtDNA-pool haplogroups obtained in the sample population of the Strand Muslim community.

Table of Contents

Chapter 1 – Literature Review	1
1.1 Introduction	1
1.1.2 Pre-Cape Muslim history	1
1.1.3 Origin of Islam in the Cape	1
1.1.4 Present day Muslims in the Cape	3
1.1.5 Origin of the Strand Muslim community	5
1.2 Genetic evolutionary studies	6
1.3 Genetic inheritance models	7
1.3.1 Mitochondrial DNA as an evolutionary tool	7
1.3.2 Mitochondrial DNA Structure	7
1.3.2.1 Composition	7
1.3.2.2 Mitochondrial D-loop	9
1.3.2.3 Hypervariable regions	10
1.3.2.4 Mitochondrial DNA features for population analysis	11
1.4 Y-chromosome as an evolutionary tool	12
1.4.1 Y-chromosome structure	12
1.4.2 Y-chromosome features for population analysis	14
1.5 Genetic markers	16

1.5.1	Types of genetic markers used in evolutionary studies	17
1.5.1.1	Single Nucleotide Polymorphisms (SNP)	17
1.5.1.2	Alu insertion polymorphisms	19
1.5.1.3	Short tandem repeats	20
1.6	DNA polymorphism detection techniques	21
1.7	Continent specific haplogroups	24
1.7.1	Definition of haplogroup	24
1.7.2	Major mitochondrial haplogroups	24
1.7.2.1	Mitochondrial DNA variation in Africa	25
1.7.2.2	Mitochondrial DNA variation in Asia	26
1.7.2.3	Mitochondrial DNA variation in Europe	28
1.7.3	Major Y-chromosomal haplogroups	29
1.7.3.1	Y-chromosomal variation in Africa	30
1.7.3.2	Y-chromosomal variation in Asia	31
1.7.3.3	Y-chromosomal variation in Europe	34
1.8	Objectives of this Study	35



Chapter 2

Reconstruction of major paternal and maternal lineages of the Cape Muslim

Community: a pilot study.	37
2.1 Introduction	37
2.2 Materials and Methods	38
2.2.1 Haplogroup Nomenclature	38
2.2.2 Y-chromosomal SNP Selection	38
2.2.3 MtDNA SNP Selection	39
2.2.4 Biological samples and survey information	40
2.2.5 Ethical approval	41
2.2.6 DNA Extraction	41
2.2.7 DNA Quantification	41
2.2.8 Y-chromosomal DNA analysis	41
2.2.9 Y-chromosomal Restriction Digest	44
2.2.10 Mitochondrial DNA analysis	46
2.2.11 Mitochondrial Restriction analysis	47
2.3 Results and Discussion	48
2.3.1 Population sample and DNA amplification	48
2.3.1.1 Y- Chromosomal DNA PCR analysis	49
2.3.1.2 Y- Chromosomal SNP RFLP analysis	49
2.3.1.3 Mitochondrial DNA PCR analysis	50
2.3.1.4 Mitochondrial SNP-RFLP analysis	50
2.3.2 Y-chromosomal haplogroup variation	52



2.3.3 Mitochondrial haplogroup variation	55
2.4 Population survey	56
2.4.1 Y- chromosomal haplogroups and perceived donor ethnicity classification	63
2.4.2 Mitochondrial haplogroups and perceived donor ethnicity classification	65
2.5 The Cape Muslims genealogical history	67
2.6 Summary	71
Chapter 3	72
Reconstruction of major paternal lineages of the Strand Muslim community	
3.1 Introduction	72
3.2 Materials and Methods	72
3.2.1 SNP selection	72
3.2.2 Terminology	73
3.2.3 Biological samples	74
3.2.4 DNA quantification	75
3.2.5 PCR amplification	75
3.2.6 PCR product purification	78
3.2.6.1 Single base extension	79
3.2.6.2 Single base extension product purification	82
3.2.6.3 Capillary electrophoresis	82
3.2.6.4 Data analysis	83
3.3 Results and Discussion	83
3.3.1 Population samples and DNA amplification	83



3.3.2 Genotyping	84
3.3.2.1 Y-chromosomal multiplex I minisequencing and genotyping	84
3.3.2.2 Y-chromosomal multiplex II minisequencing and genotyping	86
3.3.2.3 Y-chromosomal multiplex III minisequencing and genotyping	88
3.3.2.4 Y-chromosomal multiplex IV minisequencing and genotyping	89
3.3.2.5 Y-chromosomal singleplex V	90
3.4 Y-chromosomal haplogroup observed	91
3.4.1 Y chromosomal haplogroup frequencies	91
3.4.2 Strand Y-chromosomal haplogroup dominance	94
3.5 Summary	96
Chapter 4	97
Reconstruction of major maternal lineages of the Strand Muslim Community	
4.1 Introduction	97
4.2 Materials and Methods	97
4.2.1 SNP Selection	97
4.2.2 Terminology	98
4.2.3 Biological Samples	100
4.2.4 DNA Quantification	100
4.2.5 PCR Amplification	100
4.2.6 PCR Product Purification	103
4.2.7 Single Base Extension	103
4.2.8 Single Base extension product purification	105



4.2.9 Capillary electrophoresis	105
4.2.10 Data analysis	105
4.3 Results and Discussion	106
4.3.1 Biological samples and DNA amplification	106
4.3.2 Genotyping	106
4.4 Mitochondrial haplogroups observed	110
4.4.1 Mitochondrial haplogroup frequencies	111
4.4.2 Strand mitochondrial haplogroup dominance	113
4.5 Summary	115
Chapter 5	117
Population survey and ancestry analysis of the Strand Muslim community	
5.1 Introduction	117
5.2 Materials and Methods	117
5.3 Results and Discussion	118
5.3.1 Population Survey	118
5.4 The Strand Muslims genealogical history	125
5.4.1 The Strand Muslim community's Y-chromosomal ancestral history	126
5.4.2 The African fraction of the Strand Y-chromosomal pool	126
5.4.3. The European fraction of the Strand Y-chromosomal pool	127
5.4.4 The Asian fraction of the Strand Y-chromosomal pool	129
5.5. The Strand Muslim community's maternal ancestral history	131
5.5.1. The African mitochondrial contribution	132



5.5.2 The Asian mitochondrial contribution	133
5.5.3. The European mitochondrial contribution	134
5.6 Combined Y-chromosomal and mitochondrial DNA ancestry	134
Chapter 6	137
6.1 Conclusion	137
References	140
Appendices	162



Chapter 1

Literature review

1.1 Introduction

Through the decades South African populations' structure was altered by various factors such as war, apartheid and post apartheid laws. One such area greatly influenced by the regulations of apartheid is the city of Cape Town or Cape of Good Hope (Da Costa, 1994; Cajee, 2003). This city is composed of various communities which are diverse in heritage and culture. One such a community is the community of Cape Muslims. This community reflects the same diversity that can be observed in the greater South African population (Benjeddou *et al.*, 2006).



1.1.2 Pre-Cape Muslim History

The history and civilization of Islam and Muslims in South Africa is directly linked to the arrival of the Dutch in the 17th century. Prior to this the country was solely inhabited by the indigenous Bantu, San and Khoi tribes (Cajee, 2003). During this period there was no evidence of Islam or Muslims inhabiting this area. Historical archives show that Muslim presence became evident only by early 17th century.

1.1.3 Origin of Islam in the Cape

In the 16th century the Cape of Good Hope was established as a port city to the Dutch East Indian Company. The Dutch had colonized and governed much of the East Indian trade route that included areas of Madagascar, Indonesian island and South, West and East India to name a few (Bloom, 1960; Da Costa, 1994; Cajee, 2003 and Shell, 2005).

Historical documentation demonstrated shows that for the period 1658 to 1700 more than 50% of total slaves brought to Cape were from India (Table 1.1).

Table 1.1: Historical documented data demonstrating the number and place of origin of slaves brought to the Cape in 1658- 1700

Country	Number	Origin %
Madagascar	397	30.63
Sri Lanka	20	1.54
India	653	50.34
Indonesia	189	14.58
Malaysia	4	0.32
Indo-china	1	0.08
Japan	1	0.08
Cape of Good Hope	10	0.77
Unidentified	21	1.62
Total		100

(Adapted from Cajee, 2003).

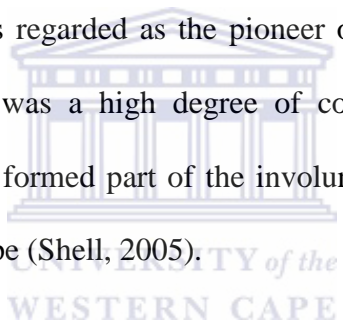


Areas such as Indonesian islands, South, West and East India were rich in Islamic culture and religion. In the beginning of the 17th Century slaves were brought by the Dutch East Indian Company (D.E.I.C) from these various eastern colonies as servants. The first historically recorded Muslim was Ibrahim of Batavia who alongside other Muslim slaves became known as the *Mardyckers* (Shell, 2005).

Apart from slaves, a larger group of prisoners and political exiles were brought by the Dutch as cheap labour. In 1743, a second fleet of convicts were brought from Indonesia to the Cape, many of these convicts remained in the Cape after completing their sentence; these convicts were often referred to as free blacks. Prisoners or free blacks with political

interest and objectivity were brought as political exile prisoners who had opposed the Dutch government in the East.

There are several well documented Muslim political exiles that have played a pivotal role in the spread and establishment of Islam in the Cape (Da Costa, 1994). Many of the political exiles were princes of Muslim descent dethroned in the East. One such a political exile was Sheikh Yusuf who was born in Macassar situated in Sulawesi modern day Indonesia. He played a pivotal role in the spread of Islam in the Cape often meeting secretly with slaves in lodges and his own home teaching them the Islamic faith. (Shell, 2005) Historically he is regarded as the pioneer of Islam in South Africa. It is during this period that there was a high degree of conversions to Islam. All these immigrants and prisoners thus formed part of the involuntary migrations that led to the introduction of Islam in the Cape (Shell, 2005).



1.1.4 Present day Muslims in the Cape

Stational population data in South Africa is estimated at 49.32 million (www.statssa.gov) (Table 1.2)

Table 1.2: South African population group percentages estimated for 2009

Population Group	Percentage (%)
African	79.3
Coloured	9.0
Indian/Asian	2.6
White	9.1

(Adapted from www.statssa.gov).

According to the mid year South African population estimates for 2009, African population group composed the highest percentage in the South African population. This was followed by the white and coloured population group that contributed 9.1 and 9.0 % respectively.

The population in the Western Cape was estimated in 2009 to contribute to 10.9% to the general South African population (Table 1.3).

Table 1.3: Western Cape population group percentages and percentage of Muslim obtained from 2001 census

Population Group	Percentage (%)	Percentage of Cape General Muslims %
African	3.4	0.7
Coloured	61.1	10.3
Indian\Asian	4.0	66.2
White	19.4	0.4

(Adapted from www.statssa.gov).

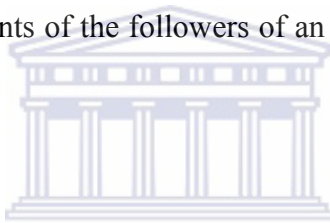
Population data obtained from the 2001 census indicated that of the Western Cape population the colored population group contributed the greatest percentage (www.statssa.gov). Islam was also the highest in the Indian/Asian population group and the coloured population group. Based on this it is clear that Islam is a minority group in the Western Cape as in total only 6.5% of the Western Cape was Muslim.

Population data in the Western Cape also revealed that Cape General Muslims classify themselves as Cape Indian Muslim or Cape Malay Muslim. Various social economics studies and historical studies on Cape Coloured Muslim have often based the question of

origin. It is thus of huge interest to Cape General Muslims to establish a link between history and genetic ancestral origin.

1.1.5 Origin of the Strand Muslim community

The community of Strand previously known as Mostered Bay is situated at the west coast of Cape Town (Rhoda, 2006). This local Muslim community is historically documented as the first cohesive rural Muslim community established in Cape Town. Oral history and historical data indicates two possible theories for the origin of the Strand Muslim Community. The first assumes the origin of the community to be linked to “Malay fishermen” who were descendents of the followers of an ex political exile prisoner Sheik Yusuf of Macassar.



The second theory is that of Mr. Rhoda, a founding historian of this community. Mr. Rhoda first showed interest in this community in the year 2006 when he conducted an in-depth historical report of the founding origins of the Muslim community of Strand, linking the community’s first inhabitants to slaves of Asian origin and followers of a Javanese imam named Abdol Sammat (Rhoda, 2006). Based on these historical records, data and oral history, no genetic evidence was available to validate these findings. The advancements in genetic studies can thus aid in answering and validating many questions concerning individual or population ancestral origin.

1.2 Genetic evolutionary studies

The basis of genetic evolution is focused on genetic variation in individuals and in populations. This makes it possible to deduce a population's ancestral origin, as human genetic diversity is the direct result of differential accumulation of mutations throughout evolution in individuals or populations (Novick *et al.*, 1995). Every living individual thus carries a combination of genetic signatures of their ancestors. This unique combination makes up an individual's unique genetic identity which is then passed on to become a constitute part of successive generations (Hutchison *et al.*, 2004).

The advent of Molecular science and the advances in genetic science has greatly increased the ability to interpret genetic variation in humans. Early genetic studies on human evolution dating back to the early 60s focused primarily on both proteins and matrilineal inherited mtDNA (Jobling and Tyler-Smith, 2003). However, thanks to the many studies conducted on world populations, two uni-parental inherited markers systems namely the Y-chromosome and mitochondrial DNA have served as the tool of choice for the past 20 years (Brión *et al.*, 2003; Jobling and Tyler-Smith, 2003; Rootsi, 2004). The Y-chromosome and mitochondrial DNA both escape recombination and thus provide paternal and maternal lineages respectively, which make a valuable contribution to the understanding of human fruition.

1.3 Genetic inheritance models

1.3.1 Mitochondrial DNA as an evolutionary tool

Mitochondrial DNA (mtDNA) has been an important source of phylogenetic information for studies of human settlement of different populations as well as for studying population migration (Richards and Macaulay, 2001; Ingman, 2003; Dubut *et al.*, 2004; Pakendorf and Stoneking, 2005). This is owed to the many advantages of the genetic molecule which have made it increasingly useful in comparison to nuclear DNA testing (Parsons, 2006).

In initial studies concerning human evolution, mtDNA analyses served as the first genetic model of choice for studying human evolution from a female perspective. One such study was that conducted in 1987 by Allan Wilson and colleagues who used a molecular approach to study Restriction Enzyme (RE) sites variations in human mitochondria. Their work, as described in (Shriver and Kittles, 2004; Umetsu and Yuasa, 2005), has thus provided the first genetic window into the past detailing the history of maternal lineages across human populations.

1.3.2 Mitochondrial DNA Structure

1.3.2.1 Composition

In (1981) Anderson and co-investigators described the first complete sequence of human mitochondrial DNA, which was later corrected by Andrew (1999). This sequence was coined the Cambridge Reference sequence (Budowle *et al.*, 2003; Nilsson, 2007). The mitochondrial DNA genome contains approximately 16.569bps of circular DNA

(Ingman, 2003; Vallone *et al.*, 2004). The genomic model encodes for 37 distinct transcribed genes located in the coding region, 13 of these are proteins, 22 are tRNAs and two are RNAs which are involved in the oxidation phosphorylation system that produces adenosine triphosphate ATP used in cellular energy (Budowle *et al.*, 2003; Ingman, 2003; Pakendorf and Stoneking, 2005; Nilsson, 2007). The number of mitochondrial DNA molecules\cells varies between different types of cells and tissues. Each cell has on average 107 mitochondria and a mitochondrion has between 1-5 mtDNA molecules (Nilsson, 2007).



1.3.2.2 Mitochondrial D-loop

The largest non-coding region within mitochondrial DNA is the D-loop or control region (Figure 1.1).

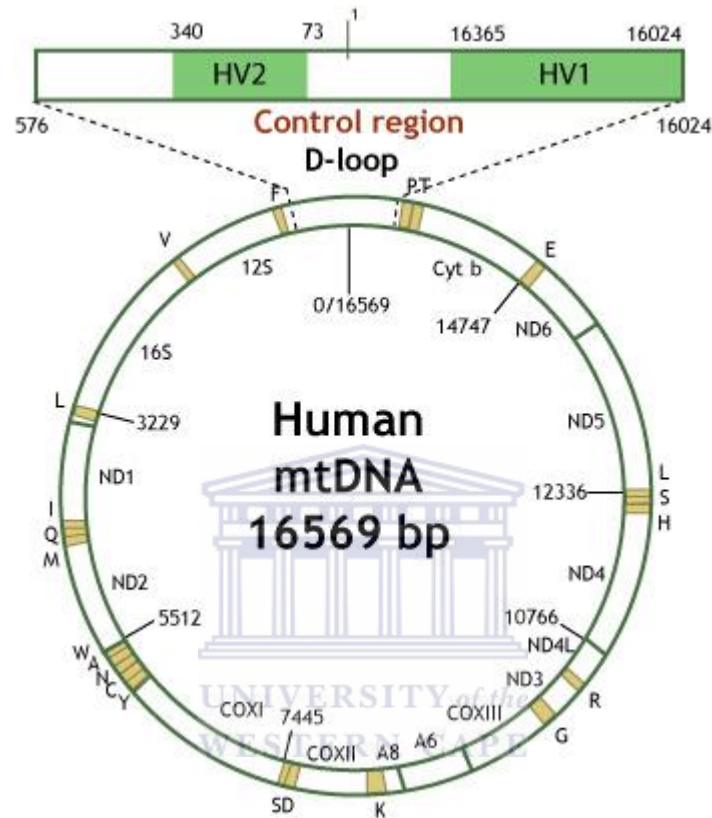


Figure 1.1: Human mitochondrial DNA representing control region/D-Loop and HV regions I and II adapted from (http://www.nfstc.org/pdi/Subject09/pdi_s09_m02_01_a.htm).

The entire region is 1.1kbp in size. The name D-loop refers to the looping out of the displaced heavy strand from the duplex DNA at the heavy strand origin of replication (Lunt *et al.*, 1998; Rowold, 2006). This region consist of two segments namely Hypervariable I (HVI) and Hypervariable II (HVII) in which mutations are frequent (Ingman, 2003; Shriver and Kittles, 2004; Nilsson, 2007).

Sequence variation in the HV regions has an evolutionary nucleotide substitution rate approximately ten times higher than in the coding region (Tambetsset *al*, 2004; Pakendorf

and Stoneking, 2005; Parsons, 2006). These mutations are highly heterogeneous and variable hence these regions are termed hypervariable regions. The fast evolving D-Loop mutation rate, which is estimated at ten times that of mtDNA average mutation, tends to discriminate within coding region haplogroup markers and provides more information about recent events in evolution than does a stretch of DNA of an equal length in the nuclear genome (Carracedo *et al.*, 2000; Parsons, 2006; Rowald, 2006).

1.3.2.3 Hypervariable regions

Hypervariable region I is described as the region between nucleotide position 16024 and 16383. Certain sites in this regions such as 16093, 16129, 16209 16311 and 1632 are mutational hotspots as they occur frequently in many different phylogenetic contexts.

Hypervariable region II lays between nucleotide position 50 and 73. This region like HVI contains mutational hotspots occurring at nucleotide position 146,150 and 152. Variation mutation rates are higher in HVII in comparison to HVI; this is mainly due to the region containing fewer sites and sequence variation. It is thus assumed that HV II is less informative than HV I. This makes it less favorable for use in phylogentic studies. Another disadvantage feature of HV II is that it contains polymorphism such as G/A at nucleotide position 263 that cannot be used in population analyses (Serk, 2004, Coble *et al.*, 2006). This transition along with transition A/G at nucleotide position 73 share ancestries in haplogroup U, K, T, J, I, W, X and Z and therefore cannot be consider defining polymorphism. This region also contains a tandem repeat poly cytosine trace

from 302- 309 that has a high mutation rate that changes repeat length. Although this region has many disadvantages, it is however used to define haplogroups.

1.3.2.4 Mitochondrial DNA features for population analysis

The human mitochondrial DNA has provided detailed estimation of human maternal genealogy. One of the various reasons for its value in studying human demographic history arises from its large copy number, strictly maternal inheritance and general homoplasmy (Pakendorf and Stoneking, 2005). There is no evidence for recombination in mitochondrial DNA although low frequency rearrangement of somatic mitochondrial DNA has been observed in heart muscle (Cavali- Sforza and Feldman, 2003).

During fertilization, the sperm mitochondria are imported into the oocyte but are selectively eliminated in early stages of embryogenesis through the ubiquitine pathway (Ingman, 2003; Hutchison *et al.*, 2004; Torpey, 2007). This maternal mode of inheritance and lack of recombination thus provides a record of the maternal lineage of our species (Di Rienzo and Wilson, 1999). Due to its genetically haploid state, the effective size of a population of mitochondrial DNA is a quarter of that of the corresponding autosomes. The high copy number makes mtDNA a molecule of choice in certain forensics DNA applications especially when dealing with biological samples that may be degraded (Pakendorf and Stoneking, 2005). Mitochondrial DNA mutates at a rate of 2-4% per site per million years that is on the average 10 to 100 fold faster than nuclear genome (Leat *et al.*, 2004). This high mutation rate has been attributed to the lack of proof-reading activity of mitochondrial DNA polymerase and to the high concentration of oxidation radicals

inside the mitochondria. The high mutation rate of the mitochondrial DNA thus allows distinct mtDNA haplotypes to be assembled into a phylogenetic tree using the maximum parsimony method (Ingman, 2003). This molecule can therefore be useful tool to construct the maternal lineage of a population.

1.4 Y-chromosome as an evolutionary tool

The Y-chromosome is a superb tool for investigating human evolution from a male perspective. This chromosome is associated with sexual development in males and is inherited clonally from father to son (Jobling, 2001; Jobling and Tyler-Smith, 2003). The haploid nature of the Y-chromosome is due to no material being exchanged with the homologue X chromosome through recombination (Stoneking, 1998). (Hutchison *et al.*, 2004) summarized that it is useful to genealogy studies as Y-chromosome follows the same inheritance pattern as that of surnames in many western and non-western societies.

However, owing to adoption and non-paternities it does not render a 100% correspondence except that it provides means to identify common biological ancestors on the paternal line where there was an unknown biological relationship.

1.4.1 Y-chromosome structure

In initial studies, the human Y-chromosome was always viewed as a non-polymorphic chromosome riddled with non coding DNA. The first report on Y-chromosome size was that by the Human Genome Project that generated a publicly available human Y-chromosome sequence of approximately 51 million base pairs (Mb)

(www.ornl.gov/hgmis). Later (Rootsi, 2004) and (Fletcher, 2004) described Y-chromosome to be encompassed to contain 50 to 60 million base pairs of DNA. The study conducted by (Tilford *et al.*, 2001) described in (Rootsi, 2004), revealed a detailed physical map of the human Y-chromosome. The genome of Y-chromosome observed was found to contain very few genes and was divided into three distinct sections namely, the pseudoautosomal region, the euchromatin and heterochromatin region (Jobling *et al.*, 1998; Charlesworth, 2003; Fletcher, 2004) (Figure 1.2.)

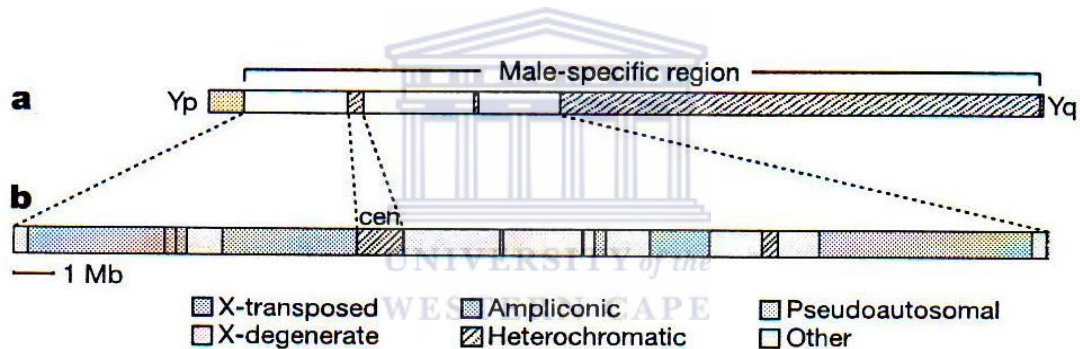


Figure 1.2: Male specific region of the Y-chromosome

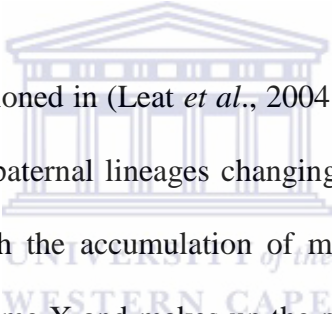
a: Schematic representation of the whole chromosome, including the pseudoautosomal and heterochromatic regions

b: Enlarged view of 24Mb portion of euchromatic region of the NRY, 1Mb bar indicates the scale of diagram

(Adapted from Skaletsky *et al.*, 2003)

Majority of Y-chromosome does not undergo recombination in male meiosis and in the order of 95% of Y-chromosome is thus termed the non recombining region NRY or male specific region MSY which is demonstrated in figure 1.2 above (Rootsi, 2004). The NRY region of one male was fully sequenced and was found to contain 156 transcription units that are located in the euchromatic region. Half of the transcription units encode for 27

distinct proteins, 12 of these are expressed freely in most tissues and 11 are testis specific (Rootsi, 2004). The Euchromatin region, which is 24Mb long, was found to be diverse in composition and size. Most of Y-specific genes or functional genes dealing with male sexuality and fertility could be found within euchromatin regions (Bachtrog and Charlesworth, 2001; Chakravati, 2001; Jobling *et al.*, 2001; Fletcher, 2004). The areas that are composed of repetitive satellites are commonly found in the heterochromatin region (Jobling *et al.*, 1998). This region is 30Mb in length and contains areas of condensed inert genetic material (Chakravati, 2001; Callafell *et al.*, 2002; Fletcher, 2004).



The 95% NRY region, as mentioned in (Leat *et al.*, 2004 and Jones, 2004), was stated as being inherited intact through paternal lineages changing only by a mutation at a given locus and thus evolves through the accumulation of mutations. The remaining 5% is genetically similar to chromosome X and makes up the pseudoautosomal region or PAR which is located at the distal ends of each arm of Y-chromosome (Jobling *et al.*, 1998; Fletcher, 2004). This is the only region, which has 99% homology with the X chromosome and is responsible for the Y-chromosome correct pairing with the X chromosome during meiosis (Fletcher, 2004).

1.4.2 Y - chromosome features useful for population analysis

The basis for Y-chromosome being such an effective instrument in evolutionary studies or population studies has been due to its uniqueness when compared to other chromosomes. The Y-chromosome is distinctive in nature when compared to other

chromosomes; Y-chromosome is in a permanent hemizygous state and is passed only through the male germ line (Jobling and Tyler-Smith, 2003). The Y-chromosome also has singularity feature assuming a 1:1 sex ratio. Thus the effective population size of the chromosome in a whole population is expected to be one – quarter of that of any autosome, one third of X chromosome and similar to that of mitochondrial DNA (Rootsi, 2004).

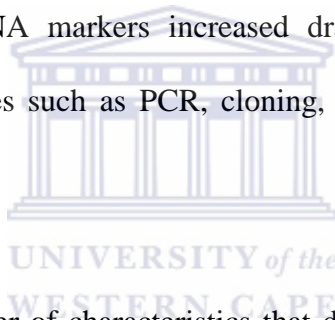
The haploid nature of Y-chromosome renders the informative NRY region which escapes meiosis. As the chromosome is passed intact through generations changing only by mutations, it provides a simple preserved record of history compared to other autosomes (Blanco *et al.*, 2000; Carvalho- Silva and Tyler- Smith, 2008). These mutations are found on all chromosomes however, the haploid Y-chromosome provides a model for studying them without the complicating factors of interallelic events and allelic diversity. The non-recombining region also houses many polymorphisms making the Y-chromosome the most informative haplotype system with applications in evolutionary studies and genealogical reconstructs (Jobling and Tyler-Smith, 2003).

In the NRY, binary polymorphism can be combined into haplogroups with a monophyletic origin that allows phylogenetic relationships to be depicted in a single most parsimonious tree.

1.5 Genetic markers

Genetic markers and their uses have evolved from focusing on phenotypes via immunological parameters and proteins to genotypes (Rowold, 2006; Väli *et al.*, 2008). Initial DNA studies focused on classical blood and protein polymorphism systems however, due to the undeveloped technology of molecular genetics it resulted in a scarcity of sequence polymorphisms which would estimate times of evolutionary events such as major migrations and ages of haplotypes, these systems were in time replaced (Thomson *et al.*, 2000).

The use of molecular or DNA markers increased dramatically with the advent of molecular biological techniques such as PCR, cloning, DNA sequencing and southern hybridization (Rowold, 2006).

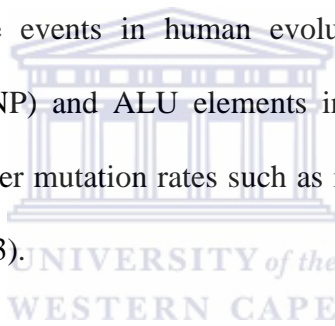


DNA markers vary in a number of characteristics that determine their levels of genetic resolution. (Rowold, 2006) summarized the most important features of DNA markers and listed them as dependant on mutation, evolutionary rate, allelic diversity, number of loci as well as mode of transmission and effective population size. Polymorphisms of particular interest are point mutations.

These polymorphisms have progressively been used in numerous applications such as medical genetics, diagnostics, studies of pharmacological response, as well as human evolutionary genetics (Quintans *et al.*, 2004). The search for polymorphisms in the human genome by both conventional and novel techniques has produced large numbers

of new markers (Bosch *et al.*, 2001). Insertions, deletions and SNPs are all biallelic markers, which represent an important class of markers referred to as unique event polymorphisms (Butler, 2003). Rowold (2006) described genetic markers to behave as Mendelian traits, which follow the laws of segregation. These binary DNA variants are associated and found within the NRY region of the Y-chromosome and the non-coding region of mitochondrial DNA. These mutations are thus ideal since they are evolutionary stable and are sequentially accumulative across generations (Underhill, 2003).

Genetic variations fall into two broad categories, biallelic markers with low mutation rates, which represent unique events in human evolution; these are namely single nucleotide polymorphisms (SNP) and ALU elements insertions. The second group is multi-allelic markers with higher mutation rates such as microsatellites and minisatellites (Jobling and Tyler-Smith, 2003).



1.5.1 Types of genetic markers used in evolutionary studies

1.5.1.1 Single Nucleotide Polymorphisms (SNP)

Single nucleotide polymorphism or SNPs, are Single base sequence variations represented in figure 1.3 below. These single base mutations occur in a DNA sequence where one base changes to another (Hutchison *et al.*, 2004). These mutations tend to be rare events or unique events in the history of the human race with mutations rates estimated at 175 total SNP mutations per individual per generation. A vast majority of SNPs are termed bi-allelic meaning that they have two possible alleles.

Reference Sequence: TAATCTGCCTTTACTTTTTTCGGTACTGGAGAGC

Comparison Sequence: TAATCTGCCTTTACTTTTTTCGGTGCTGGAGAGC

Figure 1.3: Single nucleotide Polymorphisms are single-nucleotide mutations in the DNA sequence: shaded region indicates an A to G transition (adapted and altered from Hutchison *et al.*, 2004).

SNPs are the most common type of polymorphism in the genome and are generally easier to type in comparison to other DNA markers such as for example STRs. Their availability have served as a promising tool in a wide range of human genetic applications including pharmacogenetics, population evolution studies, analysis of forensic samples and the identification of susceptibility genes involved in complex diseases (Hellard *et al.*, 2002; Alessandrini *et al.*, 2004; Endicott *et al.*, 2006).

There are 1042 million SNPs found at a density of one per 1000 bases and occurring every 100 to 300 bases along the 3 billion base human genome thus in 90% of sequences, 20-kilo bases will certainly contain one or more SNPs, this frequency however increases in gene containing regions (Chakaravarti, 2001; Wang and Moul, 2001)

(http://www.ornl.gov/sci/techresources/Human_Genome/faq/snps.shtml).

The abundance of SNPs means that they are more beneficial for differentiating between individuals. The vast majority of SNPs allows these mutations to be identified and analysed from very low quantities of DNA in small fragments (Brión *et al.*, 2004). These markers have a low mutation rate by comparison to microsatellite loci; as a result they have a modest capacity to discriminate between individuals unless a substantial number are analyzed simultaneous (Giordano *et al.*, 1999; Leat *et al.*, 2004). The low mutation

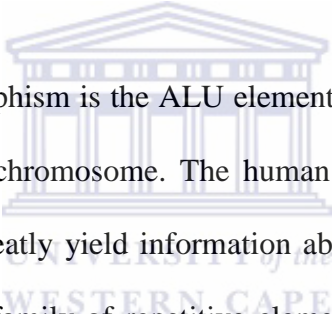
rate also renders these polymorphisms population specific and relatively stable (Naito *et al.*, 2001). SNPs in both the Y-chromosome and mitochondrial DNA are particularly valuable since they possess uni-parental inheritance and do not undergo recombination. Quintàns *et al.* (2004) goes on to describe these polymorphisms as inherited haplotype blocks since these mutations are recorded in the molecule and transferred from generation to generation as a genetic fingerprint of the evolutionary history of the locus (Quintàns *et al.*, 2004, Amorim and Pereira, 2005).

Thus for evolutionary studies SNPs are proving to be a promising tool, they are increasingly and consequently used and there is a large amount of population data available (Thacker *et al.*, 2004). However, SNP diversity should be interpreted with great caution since most previously published studies type variants have led to preconceived results and in order for successful usage and selection, SNPs should thus not be artifacts and should be polymorphic (Wang *et al.*, 1998; Ramana *et al.*, 2001). Butler (2003) stated that SNPs would gradually come to replace the use of microsatellites in molecular ecology and population genetics research.

1.5.1.2 ALU insertion polymorphisms

Among the different polymorphisms, short interspersed nuclear elements are another form of repeated DNA that has been investigated for population variation studies. Indels has been recognized as an abundant source of genetic markers that are found widely spread across the genome though not as common as SNPs. In the study of (Mills *et al.*, 2006) re-sequencing, data info was used in identifying 415-436 indels segregating in

human populations. It was found that among the total number of >10 million polymorphism known in humans, 1.5 million represented indels. Thus these markers can form a very basic common class of genetic markers. These polymorphisms consist of short identifiable sequences that are inserted at a location in the genome. These indels or deletions can range from one nucleotide to hundreds of nucleotides. Analysis of these indels show that no two indels will have the same length at the same genomic position, meaning that shared indels can confidently be seen as representing identity by descent (Väli *et al.*, 2008). They are also readily easily genotyped by the simple process based on size separation.

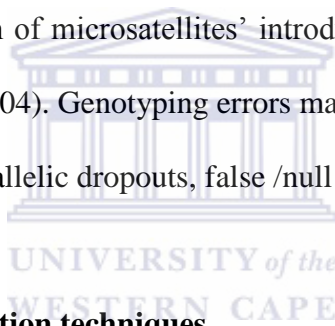


An example of such a polymorphism is the ALU element, which is often considered as a unique event found on the Y-chromosome. The human specific ALU repeat has been shown in various studies to greatly yield information about geographic or ethnic origin (Väli *et al.*, 2008). The ALU family of repetitive elements was originally defined as a fraction of repetitive DNA that was distinctly cleaved with restriction enzyme *Alu I*. The family is present in an excess of 500 000 copies per haploid genome and constitute 5% of total mass of the genome (Novick *et al.*, 1995).

1.5.1.3 Short tandem repeats

Short tandem repeats (STR) are stretches of DNA containing motifs of 1-6 bp that are tandemly repeated. Approximately 3% of the human genome is occupied by STR sequences, which are found on all autosomal, as well as the sex chromosomes (Jorde *et al.*, 1998). STRs made an introduction about 15 years ago and have been the genotype

based marker approach for many applications concerning relatedness between individuals or populations (Jorde *et al.*, 1998). STRs are more prone to mutation due to replication slippage that arise by intramolecular slipped strand mispairing repeats of 1-4bp with the exception of some trinucleotide repeat expansions causing inherited diseases (Pastinen, 2000; Naito *et al.*, 2001). The highly polymorphic nature of these microsatellites' allows for a good discriminatory capacity in forensic application, which can be used to investigate demographic events that occurred in more recent time scale (Jobling, 2001; Benjeddou *et al.*, 2006). The 2×10^{-3} per generation mutation rate also allows that a limited number of markers are needed for applications however; the complex and heterogeneous mutation pattern of microsatellites' introduces ambiguities to further data analysis (Zhivotovsky *et al.*, 2004). Genotyping errors may occur because of stutter bands and technical artefacts such as allelic dropouts, false /null alleles and size homoplasy.

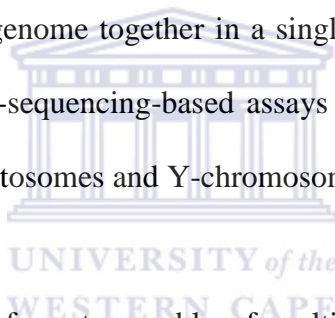


1.6 DNA polymorphism detection techniques

Classical methods for studying point mutations in mitochondrial DNA and Y-chromosome were based on PCR amplification followed by restriction fragment analysis using polyacrilimide gel electrophoresis coupled with silver staining or using agarose gels stained with ethidium bromide (Fletcher, 2004). These methods have proved to be tedious, time consuming, costly and required large amounts of DNA. Fletcher (2004) summarized that it is crucial for any DNA based technology to work with minimal amounts of starting DNA as DNA evidence are usually found in limited quantities. SNP identification has shown to be a promising application in forensic human identification (Brandstätter *et al.*, 2006).

The past years has seen the development of various new and improved high throughput approaches for examining SNP markers (Butler, 2003; Quintans *et al.*, 2004). These include DNA micro arrays, FRET analysis by light cycler, Malditof spectrometry and various others (Lareu *et al.*, 2001; Quintans *et al.*, 2004).

Since SNPs often require forensic genetists to analyze large number of loci simultaneously, multiplexed PCR systems are used in which two or more target sequences are amplified in the same reaction (Leat *et al.*, 2004). Many phylogenetical errors can be overcome by multiplexed PCR reactions, as multiple primer pairs will amplify different parts of the genome together in a single reaction tube (Endicott *et al.*, 2006). Various multiplex mini-sequencing-based assays have been used successfully to validate analysis of mtDNA, autosomes and Y-chromosome (Bouakaze *et al.*, 2007).



There are two SNP detection formats capable of multiplex analysis; these are allele-specific primer extension and allele-specific hybridization (Vallone *et al.*, 2004). These methods are all based on allele specific extension or mini-sequencing (Hellard *et al.*, 2002). Minisequencing or single base extension is a rapid and robust assay to simultaneously genotype SNPs. It involves allele-specific primer extension with fluorescent dye labeled dideoxynucleotide triphosphates ddNTPs to help visualize the results in electropherograms that are represented as peaks (Figure 1.4).

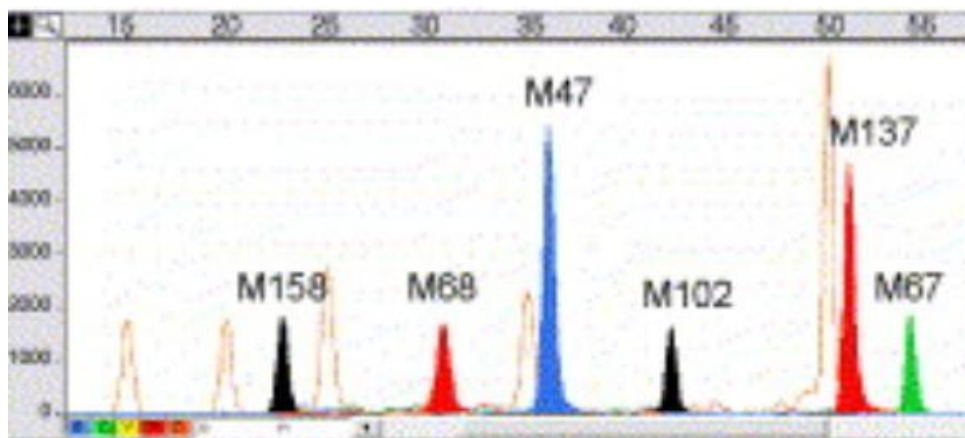


Figure 1.4: Example of an allele SNP extension electropherogram depicted SNP in respective colour and peak size (adapted from Onofri *et al.*, 2006)

The process involves three primary steps: amplification, primer extension and analysis. First, the region around each SNP locus is amplified using PCR. Amplicons can be pooled following singleplex PCR or simultaneously generated using multiplex PCR (Fletcher, 2004). Hereafter a primer binds to the template DNA one nucleotide away from the SNP of interest, where DNA polymerase extends the primer. Once extended multiple analysis techniques are used to determine which of the four dNTPs is incorporated.

The ABI Prism SNaPshot ddNTP is a common single base extension technique used. The analysis is dependent on the availability of 5-dye detection with electrophoretic platforms. Extended products are run through an ABI Prism 310 Genetic analyzer (Fletcher, 2004). The base that was incorporated can be determined by the fluorescent signal observed since each of the fluorescently ddNTPs has a specific colour assigned to indicate each of the four nucleotides. For example, a green fluorescent signal corresponds to the elution time when the primer and one sized nucleotide existed in the column. This would indicate that an oligonucleotide primer was extended using an Adenosine

nucleotide, indicated that the template DNA had a thymine nucleotide 3' to the primer binding site (Fletcher, 2004).

1.7 Continent specific haplogroups

1.7.1 Definition of haplogroup

Genetic markers used to measure variation within populations are found in the Y-chromosome and mitochondrial DNA. Variations in these two types of DNA can be grouped into continent-specific haplogroups or lineages (Benjeddou *et al.*, 2006).

1.7.2 Major mitochondrial haplogroups

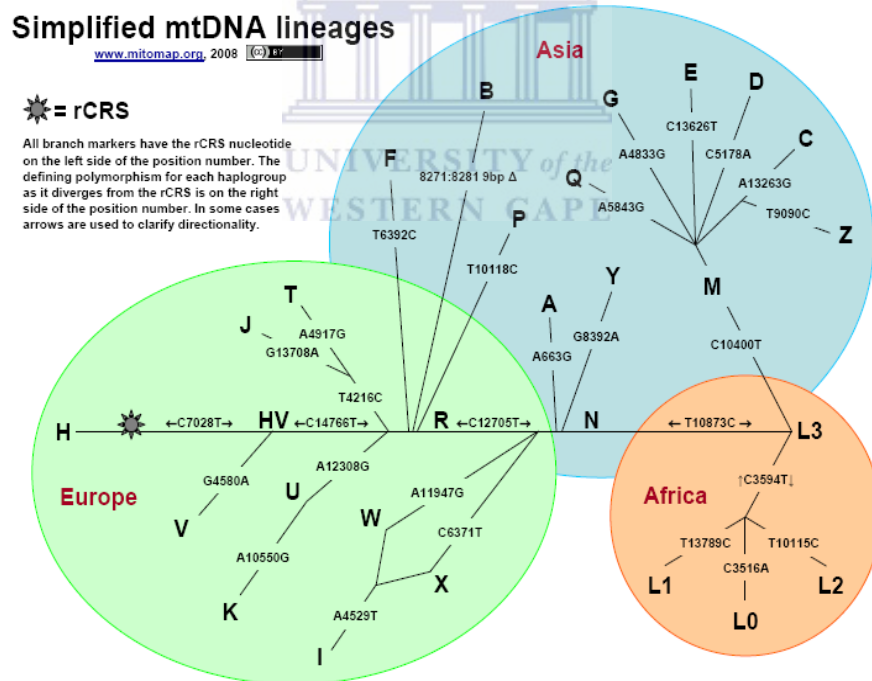


Figure 1.5: World mitochondrial haplogroups and mutations (adapted from mitomap: a human mitochondrial genome database, <http://www.mitomap.org>, 2008)

1.7.2.1 Mitochondrial haplogroups in Africa

In 1987, Cann and colleagues as described in (Batini, 2008) discovered a deep split between Africans and non Africans. The deepest branches of the mtDNA tree lead exclusively to African haplotypes (Herrnstadt and Howell, 2004; Underhill and Kivisild, 2007). Modern mtDNA variation was thus found to be characterised as comprising L lineages exclusively in Africa and M and N outside Africa (Hurles *et al.*, 2005; Underhill and Kivisild, 2007). Three of these most ancient and major lineages L1, L2, L3 are common groups for sub-Saharan African lineages and make up the macrohaplogroup L (Maca-Meyer *et al.*, 2001; Mishmar *et al.*, 2003). Mitochondrial eve L0 represents the earliest female root of the human family tree (Herrnstadt and Howell, 2004; Wells, 2007).

This haplogroup is the oldest and most divergent genetic sequence and is found in 20-25% of people in central, east and south east Africa. L0 includes four sub-haplogroups LOa, LOd, LOf and LOk. These subgroups are found predominantly in southern Khoisan and east African. The second haplogroup descendent directly from mtDNA eve is L1 (Pakendorf and Stoneking, 2005). Descendants of L1 can trace their female ancestry to some of the first modern humans to leave Africa. L1 descendants thus left Africa and populated rest of the world while LO remained in Africa (Wells, 2007; Batini, 2008). L2 is descendant from LO/L1 and also populate central Africa to as far as South Africa.

Compared to predecessor LO and L1 individuals that remained predominately in Eastern and Southern Africa, L2 ancestors went off to populate the rest of the world. L2 is the most frequent and widespread mtDNA haplogroup in Africa. The haplogroup can be

separated into four unique subsets, L2a, L2b, L2c, and L2d (Wells, 2007). Haplogroup L3 is the sister group of haplogroup L1 and L2 and is associated with the first humans to leave Africa and inhabit Europe and the new world. L3 is composed of numerous sub haplogroups which further sub-divided into subclades, some of these are L3b, L3e, L3f, L3h and L3g. This haplogroup with its subclades are found throughout Africa at various frequencies.

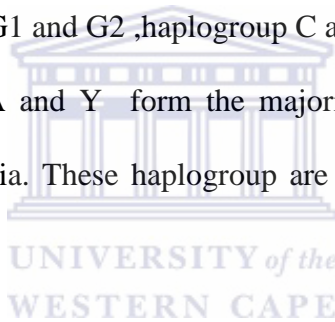
1.7.2.2 Mitochondrial haplogroups in Asia

Eastern and Western Eurasia specific mitochondrial haplogroups meet in central Asia which as stated by Tambets (2004), is the contracted zone between these regionally differentiated groups. The mitochondrial haplogroups stemming from this composite are more or less equal in contribution. An example is the frequency distribution of Western Eurasia mtDNA lineages found in India, namely H, I, J, T, X and sub cluster group of U (Kivisild *et al.*, 1999). These haplogroups were initially considered Western Eurasian but have now been observed to be the second most dominant haplogroup observed in India.

Two hypothetical routes for the Asian colonization were proposed; the first involves a route through central Asia and second one through south Asia. Haplogroups associated with the south Asian route have all been found to be associated with haplogroup M and it's the derivatives which are most prominent in India and Eastern Asia. The Asian phylogenetic mitochondrial tree thus exhibits the ancestral lineage known as M with descendant branches E, F and G as well as A through D, which are also founded in the Americas (Maca- Meyer *et al.*, 2001; Herrnstadt and Howell, 2004, Underhill and

Kivisild, 2007; Wells, 2007). Many branches of M have been found throughout Asia, especially India where it constitutes 80 percent of Indian mtDNA lineages (AraÚjo Silva *et al.*, 2006; Friedlaender *et al.*, 2007). Haplogroup M spread initially to the South and East Eurasia through to Oceania and today is virtually absent in Europe, however it is present in Africa as M1. M is thus a broader genetic marker for an early Southern migration route of humans from the Indian sub continent East ward (Endicott *et al.*, 2006).

Derived lineage groups of macro haplogroup M such as D with subgroups D4 and D5, haplogroup G with subgroups G1 and G2 ,haplogroup C and Z and those of haplogroup N with derived lineage groups A and Y form the majority of mtDNA haplogroups in northern and north Eastern Asia. These haplogroup are also the most common in East Asia (Tambets, 2004).

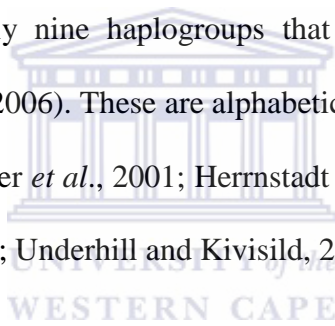


Haplogroup D is considered an East Eurasian haplogroup and today accounts for 20 percent of mitochondrial DNA pool in East Asia (AraÚjo Silva, 2006; Wells, 2007). Haplogroup D exhibits a gradual decrease in frequency moving West across Eurasia. Haplogroup C today accounts for five to ten percent of the people in central Asia and around 3 percent of people living in East Asia. The spread of haplogroup C and D shows a decreasing gradient towards South and South East Asia and are more frequent in Northern Asia populations. Haplogroup Z, the sister clade of haplogroup C accounts for 3 percent of entire mtDNA genomic pool found in Central Asia (Tambets, 2004).

1.7.2.3 Mitochondrial haplogroups in Europe

Throughout most mtDNA studies such as that of Wallace and colleagues in 1989, European mtDNA variants were often classified according to data obtained from analysis of coding region and the control region in the mitochondrial DNA.

When examining the MtDNA haplogroup tree, European mtDNA haplogroups radiates from L3 to haplogroup N. In 1989 as described in (Serk, 2004), Douglas Wallace produced a group of European specific haplogroups namely H, I, J, and K. Torroni *et al.* later discovered five more European specific haplogroups namely T, U, V, X and W. Today there are thus currently nine haplogroups that encompass the 98 percent of European populations (Isaacs, 2006). These are alphabetically arranged and are H, I, J, K, T, U, V, W and X (Maca- Meyer *et al.*, 2001; Herrnstadt and Howell, 2004; Houshmand, 2004; AraÚjo Silva *et al.*, 2006; Underhill and Kivisild, 2007).



1.7.3 Major Y-chromosome haplogroups

The theory of Y-chromosomal phylogeny has through numerous studies revealed that it is rooted in Africa (Cruciani *et al.*, 2004). This has been made evident through linguistics, archaeological and more recent genetic studies which all suggest that the Y-chromosomal lineages have an African origin (Batini, 2008) (Figure 1.6).

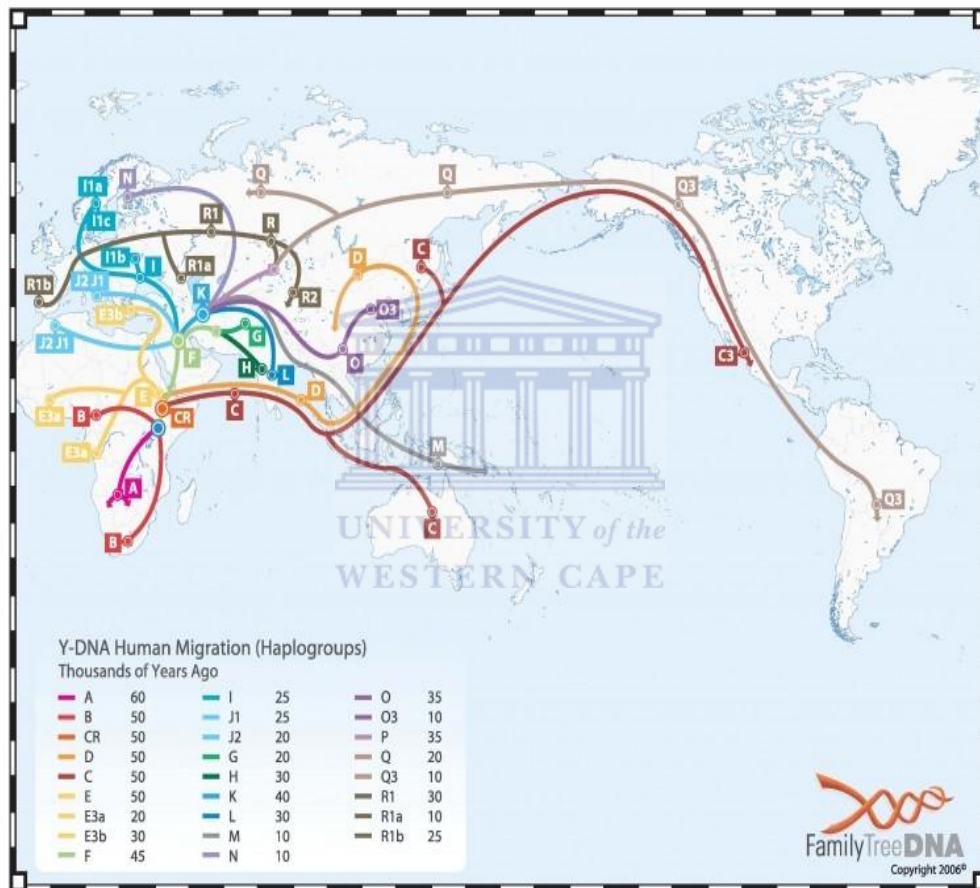


Figure 1.6: Y-chromosomal evolutionary tree indicating haplogroups and major haplogroups designated origin and migration Adapted from (<http://www.familytreedna.com/public/china>).

1.7.3.1 Y-chromosome haplogroups in Africa

When examining the Y-chromosome tree it can be observed that the two deepest branches of the Y tree namely Haplogroup A and haplogroup B are rooted and widely distributed in sub-Saharan Africa (Rootsi, 2004; Batini, 2008). Haplogroup A defined by marker M91 possesses deep genealogical heritage and is found across Africa at moderate frequencies (Underhill, 2003).

Haplogroup A is found in several Khoisan speaking populations, South Africans, West and central Africans, Sudanese and Ethiopians (Cruciani *et al.*, 2002; Luis *et al.*, 2004). However Sub-Saharan African populations are characterized by the presence of four main haplogroups A, B, E and R (Batini, 2008).

Haplogroup B is defined by several polymorphisms and this Y-chromosomal lineage is characterized by Marker M60. This haplogroup has a broad dispersal and has similar distribution pattern as haplogroup A but are more frequent in Biaka, Mbuti pygmies and East Africa (Trovoada *et al.*, 2007; Batini, 2008).

A third haplogroup, haplogroup E designated by marker M96 encompasses about 80% of the present day African Y-chromosome and its high frequency is most likely associated with bantu speaking farmers from West Africa (Hammer, 1994; Weale *et al.*, 2003). This haplogroup along with haplogroup D share three phylogenetically equivalent markers. This is indicative of shared heritage originating most probably in Africa. Descendants with these mutations that remained in Africa gave rise to haplogroup E which is most

frequent and divergent in Africa (Weale *et al.*, 2003; Rootsi, 2004). Haplogroup E has three subgroups found moderately throughout Africa. Subgroup E1 and E2 are rare and mainly found in East African populations. Subgroup E3 is the widest spread within Africa and is found at very high frequencies. E3 is subdivided into two namely E3a and E3b. E3a is predominantly sub Saharan Africa and has 65% frequency in Western Africa (Al-Zahery *et al.*, 2003; Weale *et al.*, 2003; Arredi *et al.*, 2004; Wetton *et al.*, 2005; Trovada *et al.*, 2007).

Haplogroup R which is predominantly a European and Asian haplogroup has been observed in north Cameroon population and proposes a back migration from Asia to sub-Saharan Africa (Underhill and Kivisild, 2007; Batini, 2008).

1.7.3.2 Y-chromosome haplogroups in Asia

There have been 15 of 18 haplogroups reported in Asia but their distribution is highly region specific (Rootsi, 2004). This vast number of haplogroups observed is thus indicative of various migratory routes that influenced these haplogroups distribution and frequency (Cruciani *et al.*, 2002).

There are two major migratory routes considered to play a role in the initial peopling of East Asia. The first model suggests that humans to colonize Southeast Asia took a South route followed by a Northward migration pattern. Evidence for this migratory pattern is mentioned in Rootsi (2004) when he describes the presence of haplogroups O as being far wide spread in the south then the north. Haplogroups O today accounts for 80%-90% of

lineages in East and Southeast Asia (Underhill, 2004). Two of its subgroups O1 and O2 are found in Taiwan, Indonesia and Malaysia and southern China respectively. The presence of Haplogroup C and D supports the Northern route wide overall Asian distribution. The mutation M130 observed for Haplogroups C has not been found anywhere in Africa and is it likely proposed that it arose somewhere in Asia prior to the arrival of modern humans to Sahul in South East Asia (Underhill, 2004). The presence of Haplogroups C extends widely into Central and Northern Asia (Jobling and Tyler –Smith, 2003).

Haplogroup D however is very specific and restricted and has been found in high frequencies in Tibetan and Japanese populations (Hammer, 1994; Weale *et al.*, 2003). The Near East is commonly referred to as the Levant and includes regions such as the Middle East. The most observed haplogroups associated with the Near East are that of J and E3b (Flores *et al.*, 2003).

Haplogroup J has a high frequency and is believed to originate in Near East. This lineage has also been observed in India (Underhill, 2003). The mutation M168 is the root of all out of Africa Y haplogroups and this mutation designates haplogroup F which is the third major haplogroup (Cox, 2006). An expansion of Haplogroups F towards Central Asia gave rise to the M9 mutation, which designates the clade- K and ultimately L and M.

Haplogroup K is predominantly found in central Asia where as South East Asia was found to have the greatest frequency of L. The population carrying the M9 expanded

widely into Northern Asia which is characterized by haplogroup P, this haplogroup then expanded as haplogroup Q which is now found in Siberia and ultimately West ward as haplogroup R. These diversifications suggest fragmentation of population carrying F related lineage through most of Asia, displacing the earlier haplogroup C and D lineages (Underhill, 2003). Mutation M9 thus lies at the root of a spectrum of lineages present throughout Eurasia (Underhill, 2004).

Central Asia is located at the crossroad of West and East Eurasia. Well *et al.*, (2001) mentioned in Rootsi (2004) states that central Asia is the starting point of migration to Europe, India as well as the Americas (Cruciani *et al.*, 2002). Influences of near East Asia, East Asia and Southeast Asia can all be observed in the Y chromosomal pool of central Asia. A few of numerous haplogroups observed in central Asia at various frequencies are haplogroups R with subgroups R1 and R1a, C3 and J2 represents the most frequent haplogroups.

The lesser observed haplogroups are F, E, K, O, L and R2. The fact that these haplogroups are all present in central Asia at different frequencies shows that the origin of Y-chromosome diversity in central Asia thus reflects a recent admixture region of paternal lineages arriving from different directions rather than the regions where they originated from (Rootsi, 2004).

1.7.3.3 Y-chromosome haplogroups in Europe

European populations have been extensively studied making this continent the most thoroughly studied region worldwide. European Y variation is thus best understood compared to Asian and African. In the study conducted by Semino *et al.*, (2000), cited in Rootsi, (2004), it was found that 95% of studied European Y- chromosome can be grouped into ten phylogenetically defined haplogroups. However the majority of Y-European lineages belong to haplogroup R1a, R1b, I and N3. These haplogroups covers about 70-80% of the total European Y pool. The remaining 20% belongs to haplogroups J2, E3b or G.

The most dominant European haplogroup is haplogroup R1 defined by marker M173 is an ancient Eurasian marker that was believed to originate in descendants of men who, arrived from the near east 25 000 years ago who initially entered Europe (Kayser *et al.*, 2005). This marker is predominantly observed in Western Europe and about 50% of European Y-chromosomes share this haplogroup. The haplogroup can be subdivided into groups R1a and R1b (Underhill, 2003).

Haplogroup R1a shows a high frequency in areas of Eastern Europe as oppose to haplogroup R1b that has low frequency in Eastern Europe (Wetton *et al.*, 2005). The second most dominate European haplogroup is I' defined by marker M170, this haplogroup is found largely in central and northern Europe (Wetton *et al.*, 2005; Whit Athey and Nordtvedt, 2005; Wells, 2007).

Haplogroup I-M170 has more likely arisen in Europe as it is well dispersed over the continent and includes about a quarter of all north west European men. Two of its largest subgroups are I1a and I1b. Subgroup I1a is common in Scandinavia, Germany and occurs in Britain at a frequency of about 15%. Haplogroup I1b is more frequent in Eastern Europe (Whit Athey and Nordtvedt, 2005).

Haplogroup N defined by LLY22 is predominantly found in north eastern Eurasia to Eastern Europe. Haplogroup N3 is the most common subclade of haplogroup N. This haplogroup is prevalent in northern and eastern European populations and is relatively absent in southern and western Europe. Apart from the above mentioned haplogroups various other haplogroups have been reported in European populations namely E3b, J2, F and G (Flores *et al.*, 2003; Rootsi, 2004). Haplogroup E3b Haplogroup G is related to groups H, I and J and its subclade G2 expanded northward and eastward to repopulate Europe (Wells, 2007).

1.8 Objectives of this study

The community of Strand has quite an extensive record of historical origins of the earliest Muslim inhabitants; however, there are still gaps to be filled which leaves room for genetics to complete the phylogeny of this community and trace its paternal and maternal lineages.

Initially, a pilot study was conducted on Cape Muslims living in the Cape metropolitan area, where Y-chromosome mitochondrial DNA variations were investigated to give an insight on the genetic diversity of the Cape Muslim community. Genetic evolutionary tools can indeed help in studying past demographic events in present day populations.

The data obtained from the pilot study was also used to design the genetic study of the smaller Muslim local community, the Strand Muslim community, and also serve as a point of reference for future studies involving local communities living in the Cape metropolitan area and in the broader Western Cape.

In Chapter 2, major paternal and maternal lineages were reconstructed for the Cape Muslim community.

Chapter 3 describes the reconstruction of the major paternal lineages of the Strand Muslim Community.

Chapter 4 describes the reconstruction of the major maternal lineages of the Strand Muslim Community.

Chapter 5 discusses the results of the population survey and ancestry analysis of the Strand Muslim community.

Chapter 2

Reconstruction of major paternal and maternal lineages of the Cape Muslim

Community: a pilot study.

2.1 Introduction

A pilot study was conducted in order to reconstruct the major paternal and maternal lineages of the Cape Muslim Community. The aim of the study was to use and show the ability of Restriction Fragment Length Polymorphisms (RFLP) based molecular techniques as a tool for investigating the origins of Muslims living in the Cape Metropolitan area. Based on previously published literature of the Asian, African, European and mixed populations as mentioned in (Chapter one), Y-SNP and mtDNA (RFLPs) that were prevalent to these regions were chosen for this study. The frequency of these major Asian, African and European haplogroups were examined within the Cape Muslim Community in order to provide some insight into the genetic ancestry of this community and to establish the nature of Y-chromosomal and mitochondrial DNA variation of general Cape Muslims.

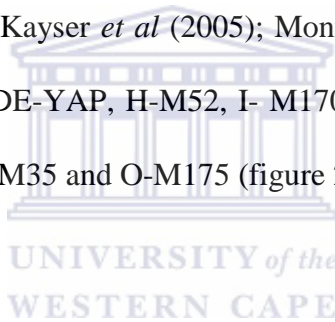
2.2 Materials and Methods

2.2.1 Haplogroup Nomenclature

Y-chromosome haplogroup nomenclature follows that of the Y-Chromosome Consortium using designation of a haplogroup linked to a SNP example K-M9. Mitochondrial nomenclature follows that of the Cambridge Reference Sequence.

2.2.2 Y-chromosomal SNP Selection

Eight Y-chromosomal SNP markers which represent some of the major internal branches of the Y-tree were selected from the literature of Hammer and Horai, (1995); Underhill *et al* (2000); Flores *et al* (2003); Kayser *et al* (2005); Montiel *et al* (2005). These markers define the major haplogroups DE-YAP, H-M52, I- M170, R1-M173 and K-M9, and the sub haplogroups E3a-M2, E3b-M35 and O-M175 (figure 2.1).



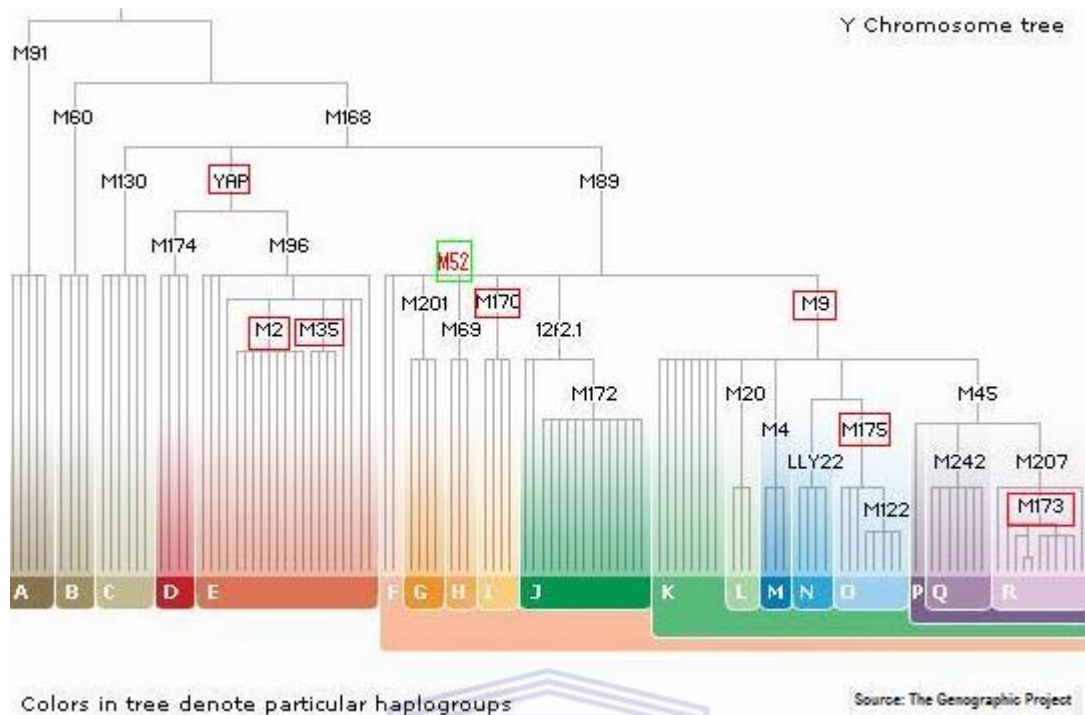


Figure 2.1: Paternal phylogenetic tree indicating haplogroups and examined markers indicated in red boxes for lineage assignment, marker denoted in the green box indicates marker not depicted but used for the assignment of haplogroup H. (adapted from <http://genographic.nationalgeographic.com/genographic/lan/en/population.html>).

2.2.3 Mitochondrial DNA SNP Selection

Six defining haplogroup-specific markers for mtDNA haplogroup determination were selected from the literature of Martinez-Cruzado *et al* (2001) and Santos *et al* (2004). The mtDNA markers define selected sites at (np) positions 7025, 13704, 3592, 10397, 10394 and 9bp deletion within region V, defining haplogroups (B) 9bp deletion, (H) np 7025 *Alu I*, (J) np 13704 *Bst OI*, and (L) np 3592 *Hpa I* and (M) np 10397 *Alu I* (Figure 2.2).

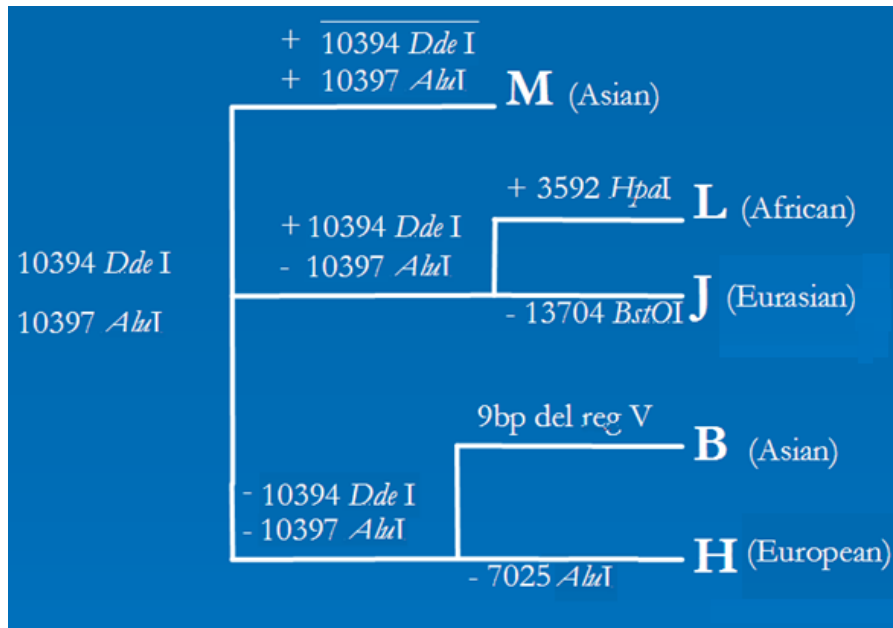


Figure 2.2: Maternal phylogenetic tree indicating haplogroups and examined markers for lineage assignment. Haplogroup-defining polymorphisms, relative to CRS are preceded by (+) derived and (-) ancestral and followed by the Restriction Enzyme used.

2.2.4 Biological Samples and Survey Information

Buccal Samples were collected from a 100 Muslim males born and residing within the Cape Metropolitan area. All donors were unrelated and had signed informed consented forms. Donors place of birth; religion and other vital information were ascertained through a donor survey.

The survey was in the form of a unique bar-coded questionnaire that was also able to exclude patrilineal and matrilineal relatives and individuals with origins outside of the Cape Metropolitan Area. Each sample was accompanied by this unique barcode envelope and questionnaire and all subsequent procedures that followed utilized the unique codes. Later the information gathered from donors' survey was compiled in a single database. A copy of the survey can be found in the Appendix section.

2.2.5 Ethical Approval

Ethical clearance for the study was obtained from the Senate Research Committee of the University of the Western Cape.

2.2.6 DNA Extraction

Genomic DNA was extracted from buffy coats of buccal swabs using the Buccal Amp™ Quick Extract DNA extraction kit (*Epicentre Technologies*) as per the manufacturers' recommendations. Hereafter samples were further purified using Phenol-Chloroform-Isoamyl alcohol purification method. The step by step protocol for this method can be found in Appendix III.



2.2.7 DNA Quantification

DNA samples were quantified using a Nanodrop *ND 1000 UV-Vis* spectrophotometer (*Applied Biosystems*) as per the manufacturers' recommendations.

2.2.8 Y-chromosomal DNA Analysis

Polymerase chain reaction amplifications were carried out in a total volume of 25µl, containing 1X PCR Buffer (*Bioline*), 1U of *Taq* DNA polymerase (*New England Biolabs*), 0.2µM of each dNTP (*Roche*), 1.5-2.5mM MgCl₂ (*Bioline*), 0.4µM-0.8µM of each primer, 1.8mg/ml of BSA and 20ng-50ng of template DNA.

Primers were synthesized by (*Whitehead Scientific*) using previously reported sequences cited in Hammer & Horai (1995); Underhill *et al* (2000); Flores *et al* (2003); Kayser *et al*

(2005) and Montiel *et al* (2005) (Table 2.1). Amplification was carried out in a GeneAmp 9600 thermo cycler (*Applied Biosystems*) (Table 2.2). The standard operating procedure for preparing the primer mixtures can be found in Appendix III.



Table 2.1: Y-chromosomal SNP amplification primer sequences, annealing temperatures and PCR fragment sizes.

Polymorphs	Primers	Annealing temp (°C)	Size (bp)	Reference
K-M9 (C-G) (Asian)	F5 ' GCAGCATATAAACTTTCAGG 3' R5 ' GAAATGCATAATGAAGTAAGCG3'	54	164	Kayser et al., 2005
I-M170 (A-C) (European)	F5 ' TATTTACTTAAAAATCATGGTTC3' R5 ' CCAATTACTTTCAACATTTAAGACC3'	49	99	Kayser et al., 2005
R1-M173 (A-C) (European)	F5 ' TTTCTGAATATTAACAGATGACAACG3' R5 ' CAGTACTCACTTTAGGTTTGCCA3'	63/56#	128	Kayser et al., 2005
E3b-M35 (G-C) (African)	F5 ' TAAGCCTAAAGAGCAGTCAGAG3' R5 ' AGAGGGAGCAATGAGGACA3'	63/59#	513	Kayser et al., 2005
DE-Yap(Alu insertion) (African)	F5 ' CAGGGGAAGATAAAGAAATA3' R5 ' ACTGCTAAAAGGGGATGGAT3	55	See table below	Hammer & Horai, 1995
H-M52 (A-C) (Asian)	F5 ' ATACCTATAAGAATATTGCCTCCA 3' R5 ' GACGAAGCAAACATTTCAAG3'	56	88	Flores et al., 2003
E3a-M2 (A-G) (African)	F5 ' AGGCACTGGTCAGAATGAAG 3' R5 ' AATGGAAAATACAGAGCTCCCC3'	62	209	Underhill et al., 2000; Montiel et al., 2005
O-M175 (deletion) (Asian)	F5 ' TTGAGCAAGAAAAATAGTACCCA 3' R5 ' CTCCATTCTTAACTATCTCAGGGA3'	56	439	Underhill et al., 2000; Montiel et al., 2005

indicates touchdown PCR

Table 2.2: Y-Chromosomal SNP thermocycling conditions.

Marker	Thermocycling conditions
K-M9	Initial denaturation step at 95°C for 4min followed by 40 cycles at 94°C for 30s, 30s at annealing temperature 54°C, 45s at 72°C. Final extension at 72°C for 10mins.
I-M170	Initial denaturation step at 95°C for 4min followed by 40 cycles at 94°C for 30s, 30s at annealing temperature 49°C, 45s at 72°C. Final extension at 72°C for 10mins.
R1-M173	Touch-down conditions: decreasing annealing temperature 63/56°C by 0.5°C for 14 cycles, 40 cycles at constant temperature.
E3b-M35	Touch-down conditions: decreasing annealing temperature 63/59°C by 0.5°C for 14 cycles, 40 cycles at constant temperature.
DE-YAP	Initial denaturation step of 94°C for 2min, 40 cycles starting at 94°C for 1min, 55°C for 30s, and 72°C for 1min. Final extension at 72°C for 7mins.
H-M52	Initial denaturation step of 94°C for 5min, followed by 40 cycles starting at 94°C for 30s, 54°C for 30s, 72°C for 30s. Final extension step of 72°C for 5mins.
E3a-M2	Initial denaturation step of 95°C for 4min, followed by 40 cycles starting at 94°C for 30s, 62°C for 30s, 72°C for 45s. Final extension step of 72°C for 10 mins.
O-M175	Initial denaturation step of 94°C for 5min, followed by 40 cycles starting at 94°C for 30s, 56°C for 30s; 72°C for 30s. Final extension step of 72°C for 5mins.

Agarose gel electrophoresis was used to verify PCR products. An aliquot of 5µl of each product was run on a 2% agarose gel (*Whitehead Scientific Grade low EEO*) stained with 0.1µg/ml Ethidium Bromide.

2.2.9 Y-chromosomal Restriction Digests

The restriction digests were performed according to the manufacturer conditions. Aliquots of 10µl of PCR product were used in the restriction digest reaction with the appropriate enzyme (Table 2.3). Reactions were carried out in a total volume of 20µl.

The results of the restriction analysis were resolved on 3% agarose gels (*Whitehead Scientific Grade low EEO*) and the observed fragment size was sized by comparison with a 100bp DNA ladder (*New England Biotechnologies*). The gels were then stained with ethidium bromide and photo-documented under UV light using *Alpha Digi Doc* cam.

Table 2.3: List of Y-chromosomal haplogroups, defining markers, restriction enzymes with optimal temperature and the digestion product sizes. Lack of restriction site is defined as ancestral and presence of restriction site and SNP is defined as derived.

Markers	Enzyme	Temperature (°C)	Fragments (bp)
M9	<i>Hinfl</i>	37	164 derived
			100+64 ancestral
M170	<i>NlaIV</i>	37	23+76 derived
			99 ancestral
M173	<i>HpyCH4IV</i>	37	128 derived
			102+26 ancestral
M35	<i>BtsI</i>	55	351+162 derived
			513 ancestral
Yap	N/A	N/A	455 derived
			150 ancestral
M52	<i>Sty I</i>	37	82 derived
			63+25 ancestral
M2	<i>NlaIII</i>	37	68+141 derived
			102+68+39 ancestral
M175	<i>Ear I</i>	37	439 derived
			87+357 ancestral

2.2.10 Mitochondrial DNA (mtDNA) Analysis

The standard amplification reactions were carried out in a final volume of 25µl containing, 1X PCR Buffer (*New England Biolabs*), 1U of *Taq* DNA polymerase (*New England Biolabs*), 0.2mM of each dNTP(*Roche*), 2.5mM MgCl₂ (*Roche*), 0.2µM of each primer and 2.5µl of DNA.

Primers were synthesized by (*Whitehead Scientific*). Primer sequence for transition position containing 9bp deletion within region V was cited in (*Martinez-Cruzado et al.*, 2001). All subsequent motif primer sequences were obtained from (*Santos et al.*, 2004) (Table 2.4).

Thermocycling conditions were conducted in a GeneAmp 9600 thermo cycler (*Applied Biosystems*). Amplification condition for 9bp deletion within region V involved: 94°C for 5mins following 30 cycles for 1min at 94°C, 1min at 55°C and 1.5min at 72°C. The final extension was at 72°C for 10mins.

All other motifs obtained from (*Santos et al.*, 2004) involved an initial denaturation step at 94°C for 5mins before being subjected to 40 cycles of 50s at 94°C, 1 min at annealing temperature and 1min at 72°C and final extension cycle of 5min at 72°C.

Table 2.4: Mitochondrial SNP amplification primer sequences, annealing temperatures and PCR fragment sizes.

Polymorphs	Primers	Annealing Temp(°C)	Size (bp)	Reference
9bp del within region V B (Asian)	L8216: 5'- CAGTTTCATGCCCATCGTC -3'	55	121 without deletion	(Martinez-Cruzado <i>et al.</i> , 2001)
	H8296: 5'- TGCTAAGTTAGCTTTACAG-3'		111 deletion present	
7025 <i>AluI</i> H(European)	L6949: 5'-CCGTAGGTGGCCTGACTGGC-3'	56	123	(Santos <i>et al.</i> , 2004)
	H7052: 5'- TGATGGCAAATACAGCTCCT-3'			
13704 <i>BstOI</i> J (European)	L13627: 5'- TCGAATAATTCTTCTCACCC-3'	47	137	(Santos <i>et al.</i> , 2004)
	H13725: 5'- TAGTAATGAGAAATCCTGCG-3'			
3592 <i>HpaI</i> L (African)	L3388: 5'-CTAGGCTATATACAACACTACGC-3'	50.9	330	(Santos <i>et al.</i> , 2004)
	H4153:5'-GGCTACTGCTCGCAGTG -3'			
10397 <i>AluI</i> M (Asian)	L10252:5'- TTGATCTAGAAATTGCCCTC-3'	48.2	276	(Santos <i>et al.</i> , 2004)
	H10527: 5'- GTATTCCTAGAAGTGAGATG-3'			
10394 <i>Dde I</i>	L10252:5'- TTGATCTAGAAATTGCCCTC-3'	48.2	276	(Santos <i>et al.</i> , 2004)
	H10527: 5'- GTATTCCTAGAAGTGAGATG-3'			

Agarose gel electrophoresis was used to verify PCR products. An aliquot of 5µl of each product was run on a 2% agarose gel (*Whitehead Scientific Grade low EEO*) stained with 0.1µg/ml Ethidium Bromide.

2.2.11 Mitochondrial DNA Restriction Digests

The restriction digest were performed according to the manufacturer conditions using 10µl PCR template and the appropriate enzyme (Table 2.5). Reactions were made up to a final volume of 25µl. The results of the restriction analysis was then resolved on 3% agarose gels (*Whitehead Scientific Grade low EEO*). Identification of 9bp deletion within

region V defining haplogroup B in a (-/-) context was resolved on a 2% standard agarose gels. Hereafter the gels were stained with ethidium bromide and photo-documented under UV light using *Alpha Digi Doc* cam.

Table 2.5: List of mitochondrial haplogroups, defining markers, restriction enzymes with optimal temperature and digestion product sizes. Lack of restriction site is defined as ancestral and presence of restriction site and SNP is defined as derived.

Marker	Enzyme	Temperature (°C)	Fragments (bp)
9bp del region V	N/A	N/A	121 ancestral
			111 derived
7025 <i>AluI</i>	<i>AluI</i>	37	108,15 ancestral
			78,30,15 derived
13704 <i>BstOI</i>	<i>BstOI</i>	60	137 ancestral
			98,39 derived
3592 <i>HpaI</i>	<i>HpaI</i>	37	330 ancestral
			207,123 derived
10397 <i>AluI</i>	<i>AluI</i>	37	276 ancestral
			147,129 derived
10394 <i>Dde I</i>	<i>Dde I</i>	37	171,105 ancestral
			133,105,38 derived

UNIVERSITY of the
WESTERN CAPE

2.3 Results and Discussion

2.3.1 Population sample and DNA amplification

Hundred buccal samples were collected from healthy male donors from the Cape Muslim Community, and DNA was extracted using a silica based extraction kit namely the Buccal Amp extraction kit (*Eppicentre Technologies*). The DNA extraction technique proved efficient with regard to DNA stocks having concentrations of more than 100ng/μl; however amplifications of these samples were problematic. The DNA was found to have low purity level, and this was evident from the failure of the PCR amplifications especially during Y-chromosomal analysis. All DNA samples were then subjected to undergo Phenol-chloroform-Isoamyl precipitation.

2.3.1.1 Y- Chromosomal DNA PCR Analysis

Analysis of Y-chromosomal SNPs proved to be complicated using the RFLP method of analysis. Amplification was successful only in samples that exhibited a high degree of purity. Continuous use of fresh DNA stocks was required for most samples as well as increased DNA concentration. However, successful amplified products were clearly visible during agarose electrophoresis.

2.3.1.2 Y- Chromosomal SNP RFLP Analysis

The polymorphic states of all Y-chromosome SNPs could be determined by examining restriction fragments with highest fragment size of 455bps, this range of fragments sizes could readily allow distinguishing using conventional agarose gel electrophoresis.



Figure 2.3: Representative of PCR-RFLP analysis for Y-chromosome lineage K-M9 of 8 unrelated males. Lane designated M is a DNA ladder (marker) with 50bp increments. Lane 1 shows positive control undigested. Lanes 2-4 as well as lanes 6 and 7 indicates *Hinfl* digested PCR products from five unrelated males; fragment size of 164bp defines derived state of the marker. Lanes 4 and 5 indicated partially digested products. Lanes 8-9 indicate DNA fragments of 100 and 64 base pairs and is therefore indicative of the ancestral state of marker M9.

RFLP analysis with use of a restriction enzyme was applicable to all binary markers except the YAP element. The PCR product size of this marker would determine haplogroup assignment.

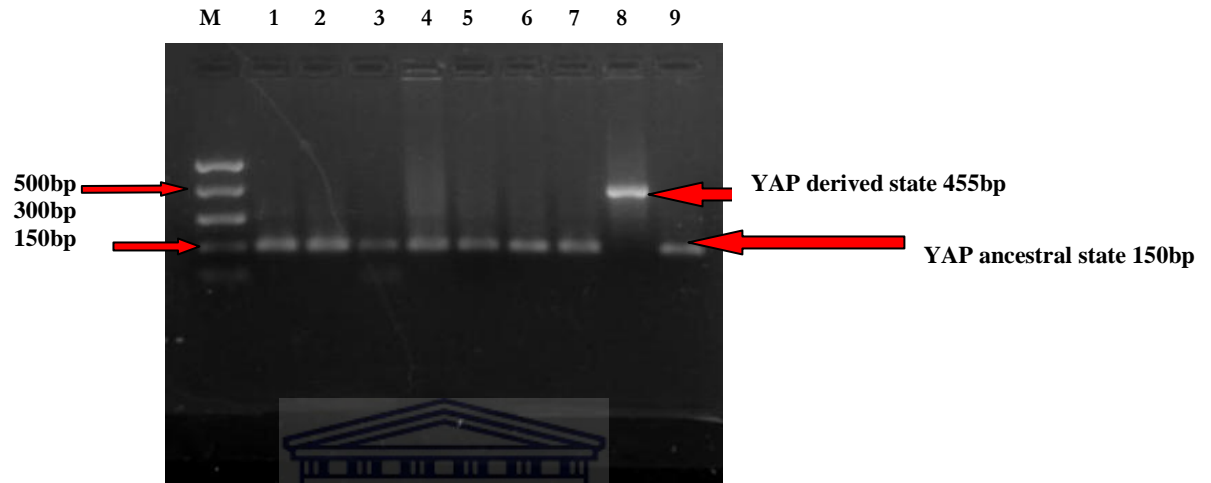


Figure 2.4: Representative of PCR electrophoresis analysis of lineage DE-YAP indicating correct fragment size for 9 unrelated males. Lane indicated by the letter M is a PCR marker (*New England BioLabs*); Lanes 1-9 shows expected PCR products of 10 unrelated males. Lanes 8 indicates derived state of marker YAP with expected size of 455bp.

2.3.1.3 Mitochondrial DNA PCR Analysis

Mitochondrial analysis was achieved far more easily compared to Y-chromosomal analysis. Amplification was successful with low amounts of DNA that had a relatively high purification result.

2.3.1.4 Mitochondrial SNP-RFLP Analysis

The polymorphic states of mtDNA SNPs were determined by examining restriction fragments with a highest fragment size of 207bps, this range of fragments sizes could readily allowed distinction using conventional 3% agarose gel electrophoresis.

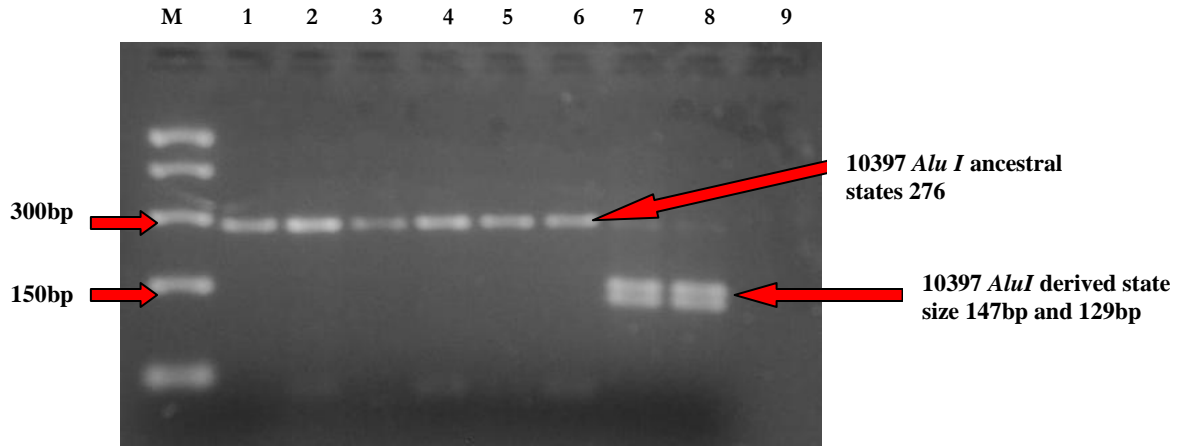


Figure 2.5: Representative of PCR-RFLP analysis for mtDNA lineage M (10397) of 8 unrelated males. Lane designated M is a PCR marker (*New England BioLabs*). Lane 1 shows positive control undigested. Lanes 2-6 indicates a DNA fragment of 276bp is therefore indicative of the ancestral state of marker 10397. Lanes 7 and 8 indicated *AluI* digested PCR products from two unrelated males; fragment size of 147bp and 129bp indicates the derived state of the marker.

Haplogroup designation for lineage B (defined by 9bp deletion within region V) was not based on RE incisions but on PCR product fragment size.

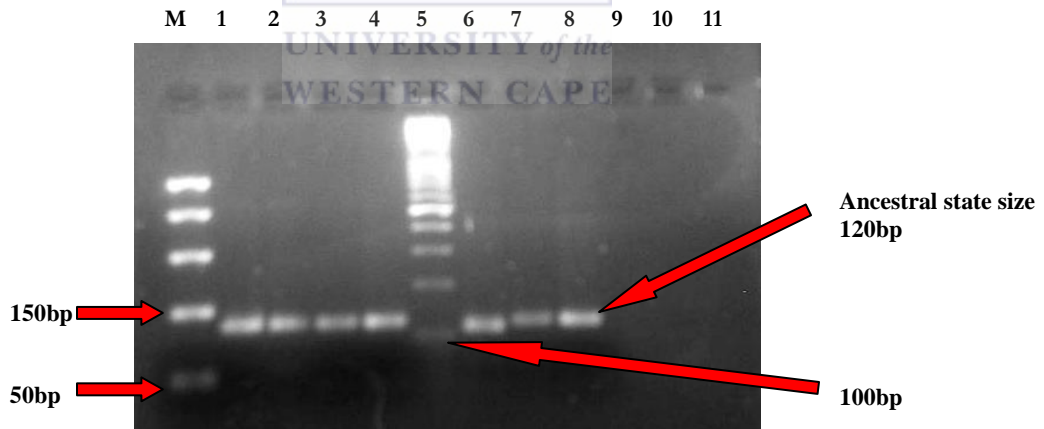


Figure 2.6: Representative of PCR electrophoresis analysis of mtDNA lineage B (9bp deletion within region V). Lane designated M is a PCR marker (*New England BioLabs*). Lanes 1-4 and 6-8 indicates PCR products from 7 unrelated males indicating a DNA fragment of 120bp and is therefore indicative of the ancestral state of the marker. Lane 5 indicates a 100bp DNA ladder (*Fermentas*); there were no positive results for this marker.

2.3.2 Y-chromosomal Haplogroup Variation

Eight Y-chromosome binary polymorphisms were examined in 100 DNA samples collected from Cape Muslim males. The analysis of these markers allowed the assignment of unknown male samples to the major Asian, European and African Y-chromosomal haplogroup in 93 samples. The remaining seven samples being were not assigned to particular haplogroups. The collective proportion of haplogroups observed is shown (Figure 2.7).

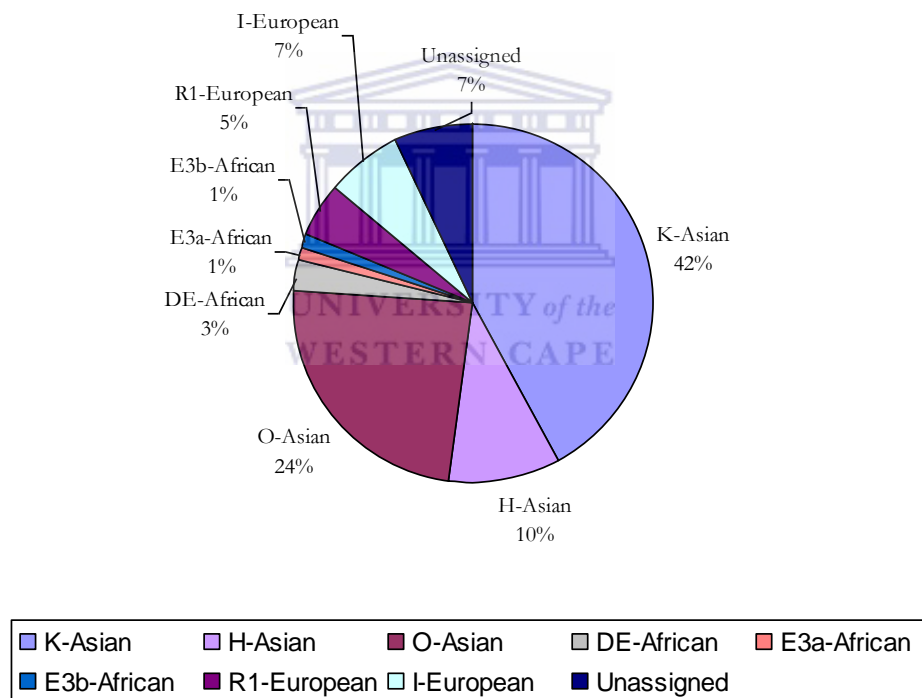


Figure 2.7: A pie chart representation of the percentage of paternal haplogroup frequencies observed in a 100 unrelated males from the Cape Muslim Community.

The markers defining major Asian, African and European haplogroups of the Y-chromosomal tree were placed into panels and named Asian, African or European.

The Asian panel consisting of haplogroups K-M9, O-M175 and H-M52 displayed the greatest frequency amongst the Cape Muslim Y-pool. Haplogroup K defines lineages

found in Asian populations such as Melanesia, Indonesia and populations' residing within Central Asia (Underhill, 2004). It occupies the major internal node within F-M89 clade of the Y chromosomal tree and occurs at considerable frequencies in all non-African population. K-M9 in this study was considered as a single group (Underhill, 2004; Cox, 2006). This haplogroup showed the highest frequency of the three Asian haplogroups mentioned constituting a 42% fraction of the Cape Muslim Community Y chromosomal pool.

Haplogroup O-M175 accounts for 80-90% of Y-chromosomes in Southeast Asia and East Asia with high frequencies documented in a number of population studies concerning the Indian subcontinent (Underhill, 2004; Wells, 2007). In the Cape Muslim Community, this haplogroup appeared at nearly half the frequency of Haplogroup K-M9, accounting for 24% of the population Y chromosomal pool. The least frequent of the Asian haplogroups was observed for haplogroup H-M52, which is normally associated with Indian populations, and accounted for 10% of the Cape Muslim Y-pool.

Haplogroup I-M170 and R1-M173 represent two of the most dominant haplogroups in European populations (Kayser *et al.*, 2005). Haplogroup I-M170, predominantly found in North West European, accounted for 7% of the Cape General Muslim Y-pool. Haplogroup R1-M173 is found in more than 50% of the Y-chromosomal pool in Europe and is prevalent in Western European populations (Wetton *et al.*, 2005; Whit Athey and Nordtvedt, 2005; Wells, 2007). However, amongst the Cape General Muslim males this haplogroup accounted for only 5% of the sample population.

African Haplogroups E3a-M2, E3b-M35 and DE-YAP were found in low frequency within the samples. Haplogroup E3a-M2 and E3b-M35 both only appeared in one individual; however, haplogroup DE-YAP were most frequent among all the examined African lineages accounting for 3% of the Cape General Muslim Y-pool. This haplogroup is a lineage often found in Sub-Saharan Africa however (Hammer and Horia, 1995) stated that traces of this lineage have also been found in Japan but at low frequencies.



2.3.3 Mitochondrial Haplogroup Variation

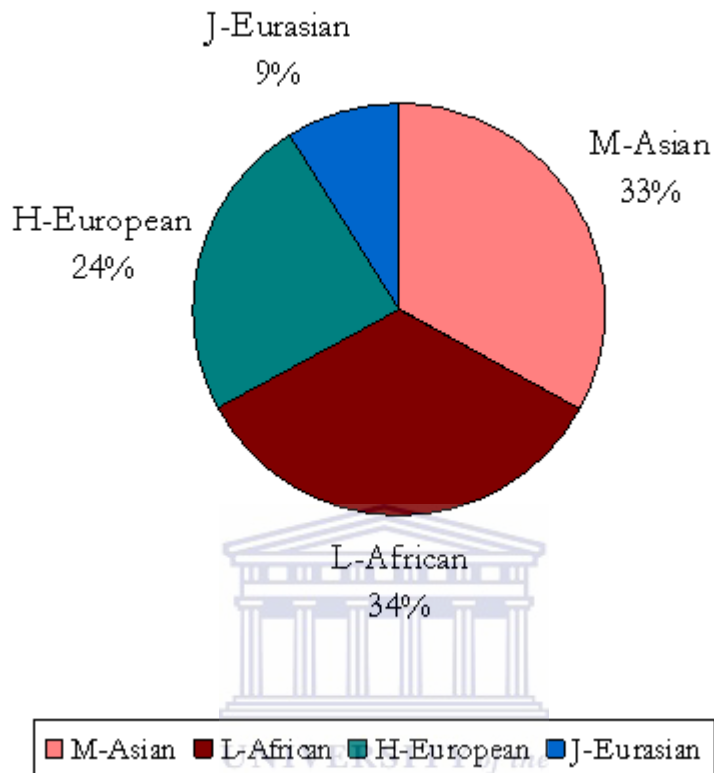


Figure 2.8: A pie chart representation of the percentage of maternal haplogroup frequencies observed in a 100 unrelated males from the Cape Muslim Community.

Screening of the six diagnostic mitochondrial RFLP markers revealed that the Cape Muslim Community's mtDNA pool encompassed haplogroups B, M, L, H and J.

Haplogroup B defined by 9bp deletion within region V was not detected within the population. Haplogroup M a super haplogroup found at high to intermediary frequencies across Central and South Asia and defined by gain site at np 10394 *Dde I* and np 10397 *Alu I* was detected in 33% of Cape Muslim community's mtDNA pool (AraÚjo Silva *et al.*, 2006). This concludes an Asian origin for individuals typed positive for this haplogroup (Bermisheva *et al.*, 2002; Malhi *et al.*, 2007; Wells, 2007).

Haplogroup L defined by transition position at 3592 *Hpa I* was detected in 34% of the population. This haplogroup identifies mtDNA of African origin as it is found across Africa with increased frequency in Sub-Saharan African populations (Maca-Meyer *et al.*, 2001; Wells, 2007).

Haplogroup H defined by cut site at *Dde I* and 7025 *Alu I* is a dominant European mitochondrial haplogroup and is dispersed across Northern and western European populations (Houshmand *et al.*, 2004; AraÚjo Silva *et al.*, 2006). In the Cape Muslim Community this haplogroup was interestingly detected within 24% of the Cape Muslim mtDNA.

Haplogroup J defined by np 13704 *Bst OI* is a Eurasian haplogroup that has greatest frequency in the Near East accounting for 25% of Yemini and Bedouin populations (Wells, 2007). In the sample population this haplogroup was found at a low frequency of 9%.

2.4 Population Survey

The question of identity in the Cape Muslim community has always been a point of interest as demonstrated in a study conducted in Da Costa (1994). Dr Da Costa researched and considered the problem of identity amongst Muslims, post Apartheid. His research suggested that although many Muslim chose a religious identity other forms identity such as of national origin or ethnicity group do exist. Data collected revealed that donors' age groups ranged from 20- 39 and 50-63 years of age. This was ideal as ethnic

perspective could be gained in both younger and older generations. All donors shared a common language of English and Afrikaans. Older individuals however preferred Afrikaans as opposed to younger individuals who preferred both English and Afrikaans. The survey also contained vital information concerning religion, place of birth and ethnic classification that will be described in more detail below.

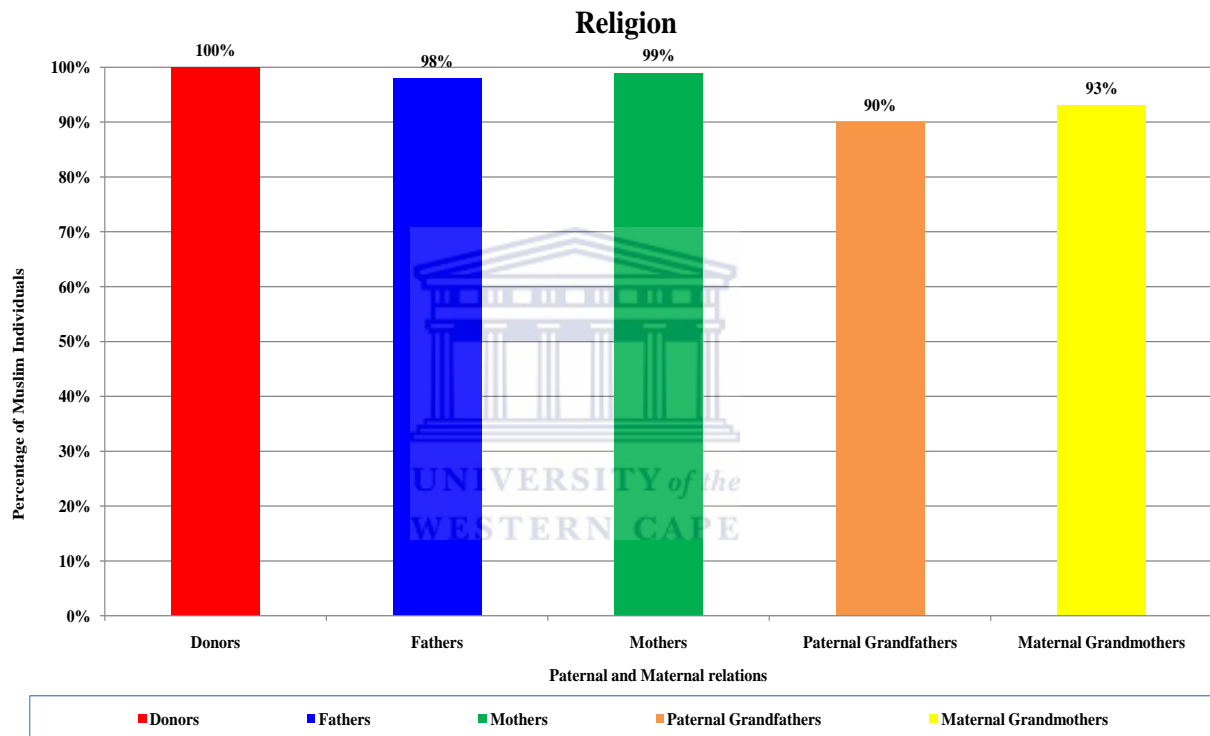


Figure 2.9: Cape General Muslim donor survey information depicting the percentage of Muslim fathers, mothers, paternal grandfathers and maternal grandmothers.

The study was aimed at general Cape General Muslims thus all donors were of Islamic faith. The category Religion in the survey however showed different values for parents and grandparents who were of Islamic faith. Examining the information stipulated in the above figure 2.9 it was noted that majority of donors were second generation Muslims with the exception of two individual's fathers and one individual's mother being of

Christian faith. It was also found that whilst a high number of parents and donors shared the Islamic faith, only 90% of individuals' paternal grandfathers' and 93% of individuals' maternal grandmothers were Muslim. This is in fact conclusive evidence that those individuals who had Muslim parents but not Muslim grandparents had undergone conversion to Islam. The high number of grandparents that were Muslim is also indicative of these families exhibiting a long and rich Islamic background in their ancestral history.

A second general requirement for participation in the study was that all donors had to be born within Cape Town. The survey however showed different results for donors' paternal and maternal lines (Figure 2.10).

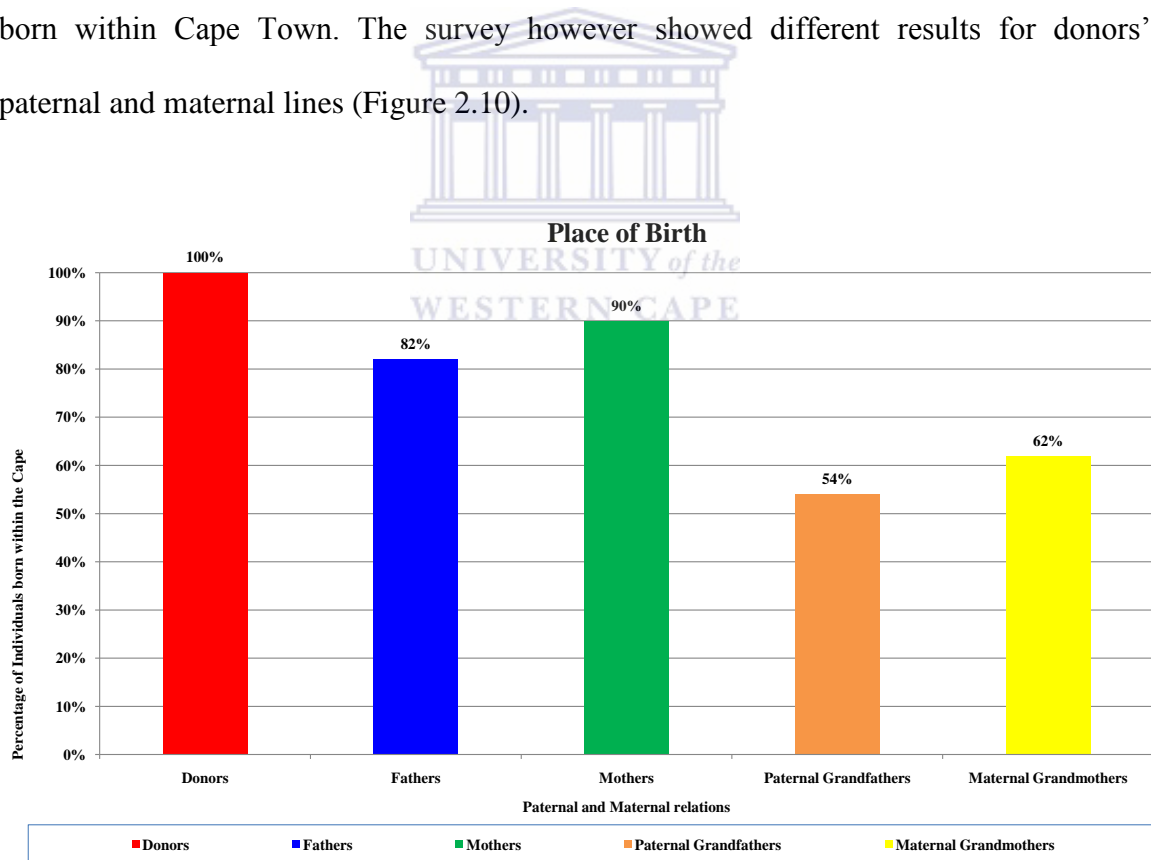
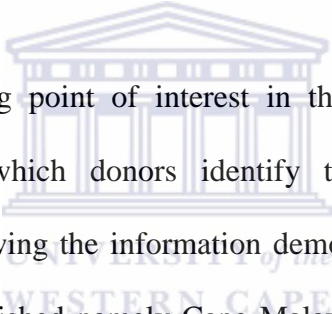


Figure 2.10: Cape General Muslim donor survey information depicting the percentage of fathers, mothers, paternal grandfathers and maternal grandmothers born in the Cape.

Eighty two percent of individuals stipulated that their paternal parent was Cape born and 90% of individuals stated their maternal parent was Cape born. This value decreased when the grandparents' data was examined. Interestingly more individuals stated that their maternal grandmothers were born within Cape Town as compared to their paternal grandfathers. This thus showed that more mothers and grandmothers were Cape born. In contrast less fathers and paternal grandfathers were born within the Cape, many of whom were stated as being born with in India. This concluded the theory that donors share the paternal ancestry of the influx of males who migrated into the Cape during slavery and post emancipation (Cajee, 2003).



The third and most interesting point of interest in the survey was the category of perceived ethnic group to which donors identify themselves, their parents and grandparents to belong. Reviewing the information demonstrated three major groups of ethnic classification was established namely Cape Malay Muslim, Cape Indian Muslim and Cape Coloured Muslim (Figure 2.11).

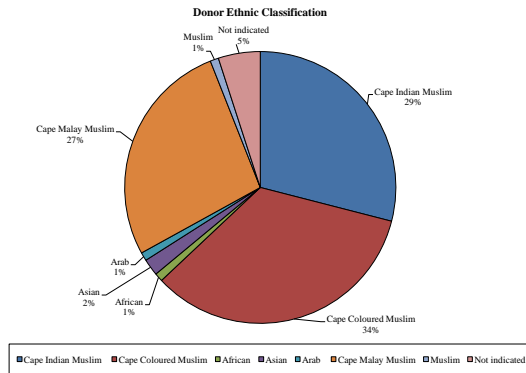


Figure 2.11: Cape Muslim donor survey information depicting the percentage of donors classified as Malay, Coloured, African, Asian, Muslim, Arab and not indicated.

Ethnic classification revealed no standardized ethnic group but diverse groups to which the donors identified them as. A large proportion of donors 34% were described as Cape Coloured Muslims, 29% of individuals considered they were Cape Indian Muslims and 27% individuals described themselves as being Cape Malay Muslims. It was also established that only one individual was considered African, two individuals chose to be defined as Asian, one individual as Arab and only one individual chose to define him as being Muslim. Five donors chose not to be affiliated to an ethnic group designated Cape Other Muslims, and this was observed more within older individuals.

Examining the ethnic groups more strictly it was evident that individuals would assign different ethnic groups between their mother and father. Represented below in figure 2.12 are the results of donors' fathers and paternal grandfathers sharing similar ethnicity and mothers and maternal grandmothers sharing similar ethnicity.

Paternal Ethnic Classification

Maternal Ethnic Classification

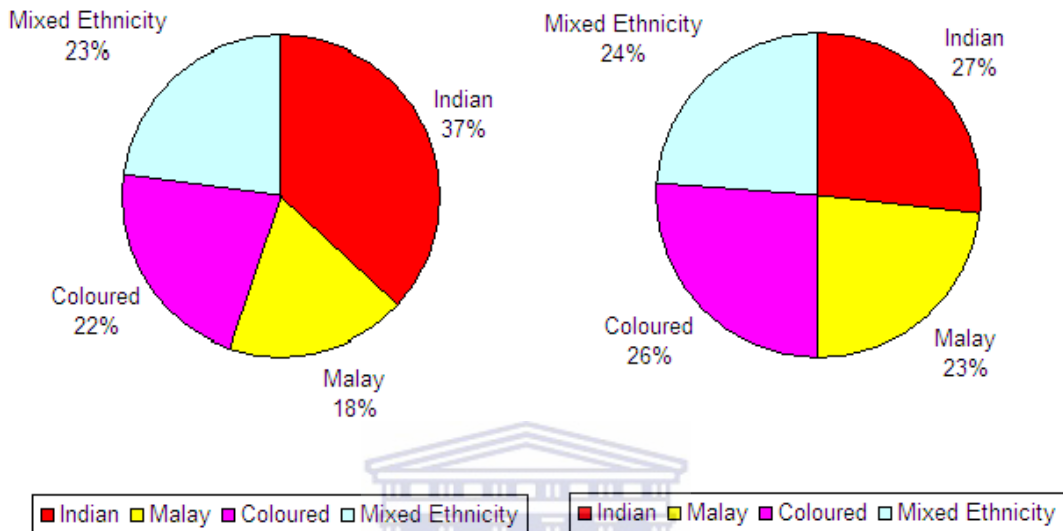


Figure 2.12: Donor survey data obtained from Cape Muslim male donors indicating perceived ethnic groups of paternal and maternal lines.

When viewing the paternal classification results, 36% of donors defined their fathers and paternal grandfathers' to be Cape Indian Muslims. Eighteen percent of donors defined their fathers and paternal grandfathers' as Cape Malay Muslims and 22% defined their fathers and grandfathers as Cape Coloured Muslims.

When comparing the maternal ethnicity results to the paternal results less (27%) mothers and maternal grandmother was described as Cape Indian Muslims. More mothers and maternal grandmothers were Cape Malay Muslims (23%) and Cape Coloured Muslims (26%). A fourth category was established in both paternal and maternal results and was named mixed ethnicity as donor would assign one ethnicity to the parent but a different one to the grandparent this results covered 23% of paternal result and 24% maternal

results. Results of the ethnicity groups' in relation to perceived ethnicity and genetic ancestry could hereafter be viewed (Figure 2.13a and b).



2.4.1 Y- chromosomal Haplogroups and Perceived Donor Ethnicity Classification

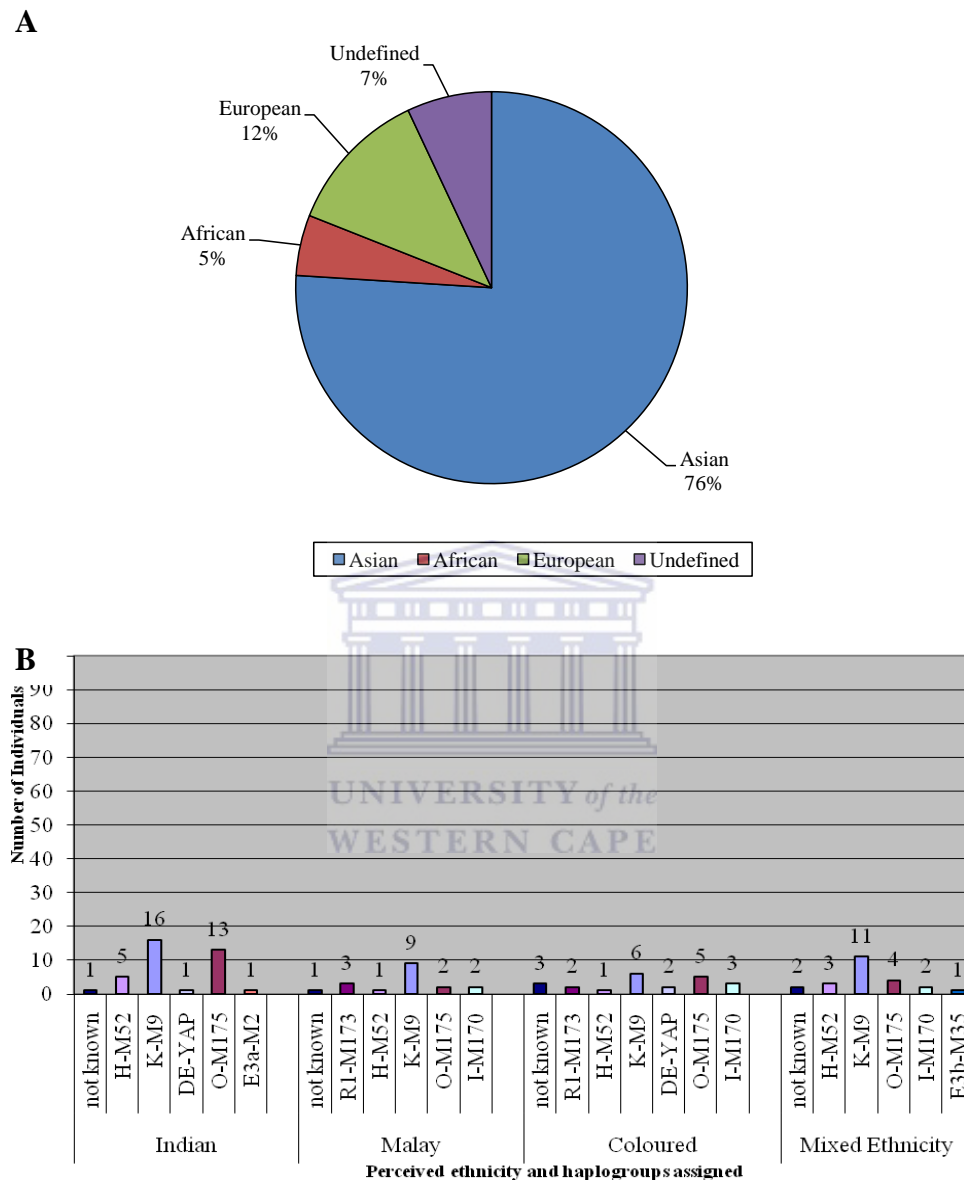


Figure 2.13: A) Overall percentage of African, Asian and European Y-chromosomal haplogroups obtained in the sample population of the Cape Muslim community. B) Donor survey data obtained from males of the Cape Muslim Community depicting perceived ethnicity and true genetic ethnicity through Y-haplogroups assignment.

During initial sampling stages (29\100) donors classified themselves as Cape Indian Muslim as depicted in figure 2.11. From the above figures 2.13, it is observed that the majority of individuals who considered themselves, their paternal parent and paternal

grandparent as Indian did indeed carry one of the Asian markers. The perceived Cape Indian Muslims ethnicity group thus exhibited the greatest frequency for the derived state of haplogroup K-M9, haplogroup O-M175 marker and haplogroup H-M52, with two individuals being positive for African haplogroup markers DE-YAP and E3a-M2. Increased numbers for haplogroup K-M9 was also observed in mixed ethnicity, Cape Malay Muslims and Cape Coloured Muslims perceived ethnicity groups.

The Cape Malay Muslims perceived ethnicity group was shown to exhibit all Asian markers and European markers at low to intermediary frequencies. The presence of I-M170 and R-M173 (European defining markers) within the group is indicative of mixed Eurasian populations. The Cape Coloured Muslim population group was most diverse in YSNP designation. This is evident as Asian, African and European haplogroups appeared in this ethnicity group. All markers except West African M35 and M2 were prevalent within this defined population group. In addition to this the few individuals that belonged to African haplogroups also opted to define themselves as Cape Coloured Muslims.

2.4.2 Mitochondrial haplogroups and perceived donor ethnicity classification

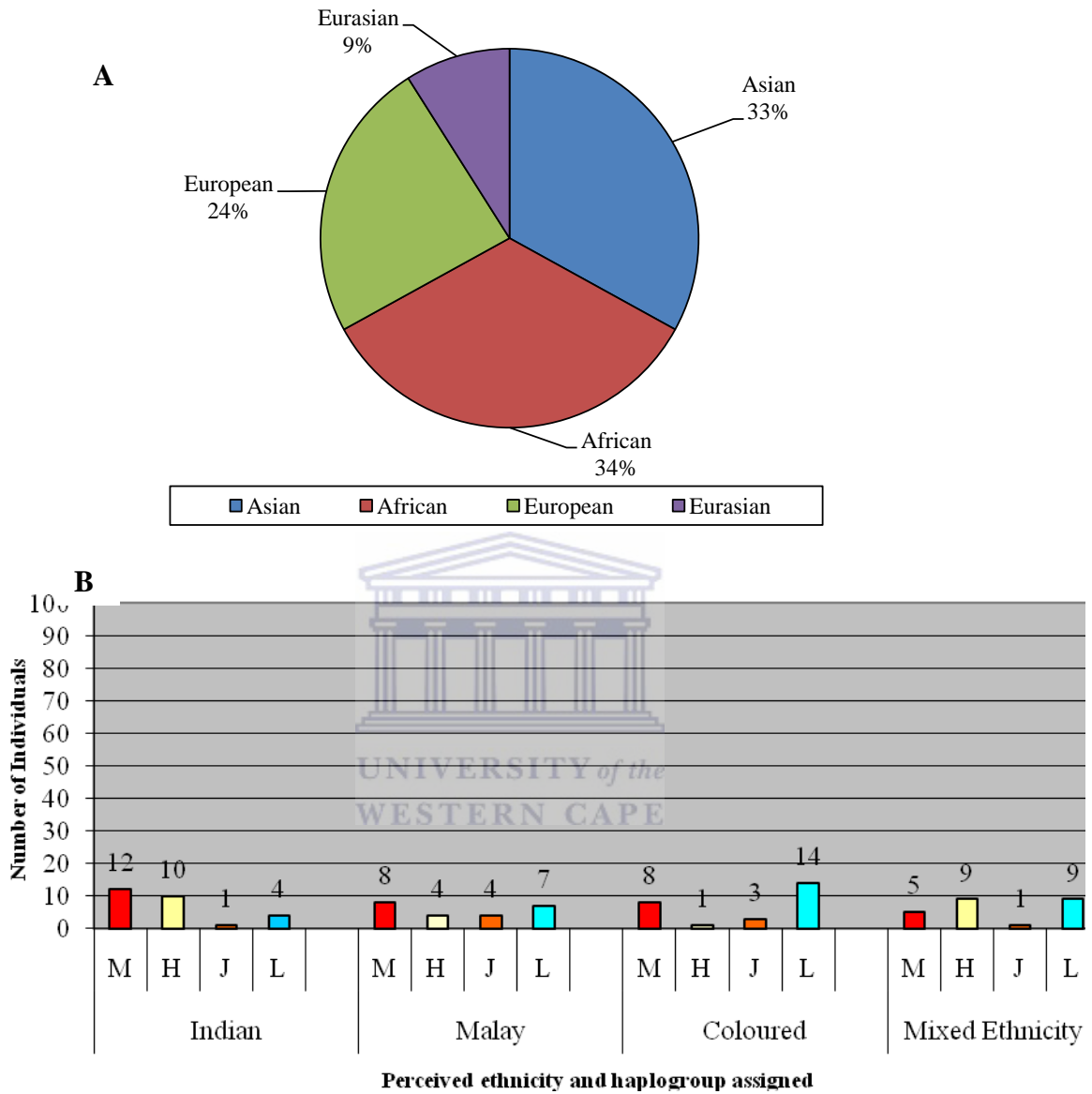
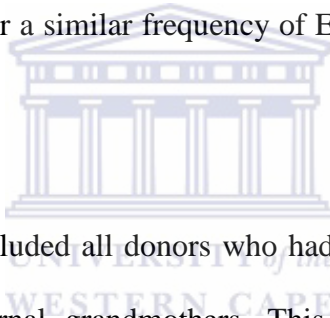


Figure 2.14: A) Overall percentage of African, Asian and European mtDNA haplogroups obtained in the sample population of the Cape Muslim community. B) Donor survey data obtained from males of the Cape Muslim Community depicting perceived ethnicity and true genetic ethnicity through mtDNA-haplogroups assignment.

Donors who defined their mothers and maternal grandmothers as Cape Indian Muslims exhibited the highest number of positives for mtDNA marker defining haplogroup M see

above figure 2.14b. These individuals maternal lines could thus be linked to an Indian ethnicity based on the observance of the Asian haplogroup M. Interestingly European haplogroup H was also found more frequent in Indian defined mothers and maternal grandmothers correlating to the ethnic result.

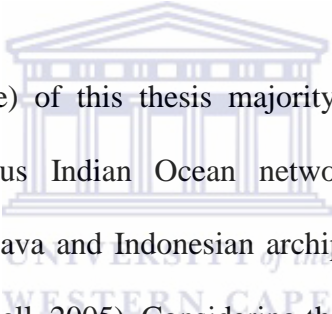
The Cape Malay Muslim category and Cape Coloured Muslims category showed similar frequency for haplogroup M. African haplogroup L however dominated mothers and maternal grandmothers' defined as Cape Coloured Muslims. Haplogroup J a Eurasian haplogroup was observed more frequent in Cape Malay Muslims defined mother and maternal grandmothers however a similar frequency of European haplogroup H was also found in this group.



The mixed ancestry groups included all donors who had defined a different ethnicity to their mothers and their maternal grandmothers. This group termed as Cape Other Muslims contained many donors who chose to define their mothers based on their maternal grandfathers' ethnicity which resulted in two different ethnicity groups for donor, mother and maternal grandmother. A high number of donors were positive for haplogroup H within the Cape other Muslims. This haplogroup was indeed evident in three donors whom defined their grandmothers as white and had indeed tested positive for the European haplogroup H.

2.5 The Cape Muslims Genealogical History

In the 19th century, there was an influx of many voluntary immigrants from the Indian subcontinent to the Cape (Cajee, 2003; Dangor, 2003; Shell, 2005). The Donor survey results showed that many donors considered their paternal lines to be of Indian descent. The haplogroup data generated from the Y-chromosomal RFLP analysis also indicated an increased frequency and prevalence of Asian haplogroups K-M9, O-M175 and H-M52. These haplogroups represented the greatest proportion of Y- lineages in the population with K-M9 representing 43% of the sample male population suggesting a correlation with historical documentation and survey results.

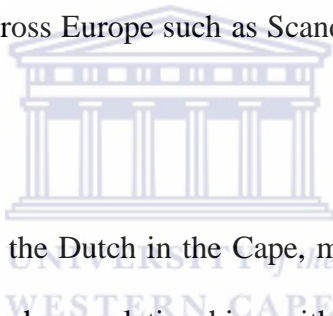


As mentioned in (Chapter one) of this thesis majority of voluntary and involuntary immigrants came from various Indian Ocean networks of slave trades such as Madagascar, Malaysia, India, Java and Indonesian archipelago between 1680 and 1731 (Cajee, 2003; Dangor, 2003; Shell, 2005). Considering the origin and prevalence of these haplogroups in present day Asian population, the presence of these lineages within this sample population concludes that Males being positive for these haplogroups can thus trace their origin to these forced and voluntary immigrants of Asian origin.

Supporting evidence of the presence of these Asian lineages within these populations can also be observed in the many donors who defined themselves and their paternal and maternal lines as Cape Malay Muslims. In the survey results 18% of donors defined themselves as Cape Malay Muslims and strongly links this ethnic group to being of Islamic faith. Historically the introduction of Islam to South Africa and Cape Town is

directly owned to slavery brought on by the Dutch East Indian Company. Majority of slaves and later political exiles and convicts came from the Indonesian archipelago and Java (Dangor, 2003; Shell, 2005). Today these Javanese and Malay customs and traditions can still be seen in many of Cape General Muslim homes (Rochlin, 1939).

Apart from the increased frequency of Asian haplogroups European haplogroups I-M170 and R1-M173 appeared at intermediary frequencies and at nearly equal contribution within this community. The presence of European lineages proposes definite admixture within this population. Historically the arrival of the Dutch to the Cape ultimately led to an influx of migrations from across Europe such as Scandinavia, Germany and France to name a few (Byrnes, 1996).



During the initial settlement of the Dutch in the Cape, many of the European men were single and were encouraged to have relationships with indigenous females or female Asian slaves (Byrnes, 1996; Shell, 2005). Later in the year 1688 with the arrival of the French, more men that are Dutch opted for relationships with French women (Byrnes, 1996). The initial admixture however between Asian, African women and European men is the most likely reason for the presence of these dominant European haplogroups within Cape General Muslim males.

The admixed history of this community thus revealed a high degree of diversity within the y-chromosomal pool. One particular ethnic group which majority of donors chose was the Cape Coloured Muslim ethnic group. Coloured is a term coined by the Population

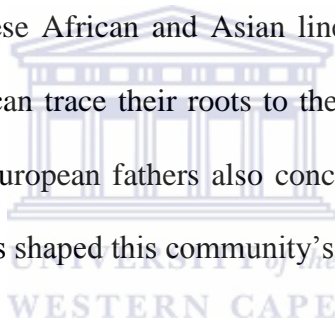
Registration Act of 1950, that defined an individual who was not white or native (Da Costa, 1994; Shell, 2005). This population group has for decades been the middle ground of much racial segregation (Bloom, 1960). Individuals of mixed ancestry were often pulled between being white or African. However over the years coloured people gradually became to except their mixed ancestry and formed a distinct population group with in South Africa and more importantly within the Cape population. (Da Costa, 1994) suggested that the Cape Coloured Muslim group consist of Afro-Asian national origin.

The RFLP analyses also revealed the presence of a small percentage of African lineages DE-YAP E3a-M2 and E3b-M35. The presence of these predominantly African haplogroups most likely accounts for the slaves brought from across west and east Africa that converted to Islam as most conversions to Islam occurred in the 16th century among slaves, confined in slave lodges and indigenous Khoi San farmers (Shell, 2005).

When comparing the mtDNA RFLP analysis to the sample population's Y-chromosomal analysis dominant Asian, African and European lineage were also found. However in contrast to the increased frequency of Asian Y lineages, the mtDNA lineages were dominated by African haplogroups namely haplogroup L. This exclusively African lineage found at great frequency in Southern and sub-Saharan African population accounting for 34% of the mtDNA pool.

The presence of this lineage is thus directly associated to indigenous African population such as the Khoi Khoi and San. Although the VOC was barred from enslaving these

indigenous people, many of these individuals became farm labourers and worked in the Cape (Byrnes, 1996). Khoi women were often involved in domestic chores and working within households of settlers (Byrnes, 1996; Shell, 2005) The second dominant lineage found was haplogroup M a dominant Asian lineage found across Asia in various population at various frequencies. The presence of this haplogroup can be associated to the female Asian slaves enslaved as labourers by the Dutch. As the ratio of men to women increased over these period relations between these women often considered free black and settlers amplified (Shell, 2005). The ratio of females to males amongst the settlers could also be seen amongst the slaves and thus women were often married off at early age. The presence of these African and Asian lineages thus concludes that these positive L and M individuals can trace their roots to these slaved women. And the mix result of Asian, African and European fathers also concludes the fact that relationships between both settlers and slaves shaped this community's mixed ancestral heritage.



Interestingly European contributions of 24% were detected amongst the mtDNA pool of this population. Historically many European settlers who later became farmers known as *free burghers* settled and spread out throughout the Cape (Byrnes, 1996). The European women was often involved in trades such as in- keeping and brewing and elite European women could be found managing households. With the ration of women to men in the Cape colony many European women were pressured to marry at a young age and with the even greater imbalance of females slaves to male slaves, it is plausible that many slave men had secret relationships with European women (Shell, 2005)

Individuals typed positives for haplogroup J can trace their origin to western Eurasia. This haplogroup appearing in both Asian and European population can be linked to either contribution.

2.6 Summary

The ancestral origin of the Cape Muslim Community was found to exhibit Asian, African and European origins. Asian Haplogroups were most dominant in the paternal line with K-M9 exhibiting the greatest frequency amongst the Cape Muslim males. European haplogroups were moderately observed as both European markers M170 and M173 were nearly equal in distribution. Mitochondrial haplogroup L was most dominant in the maternal pool. Based on these findings, population genetics can be applied on a broader scale to a sub –Cape Muslim community. The method of RFLP analysis however was found to be time consuming and laborious. This method of SNP detecting is ideal for a laboratory starting out dealing with few SNPs however when dealing with perhaps more markers a high throughput SNP genotyping technique may be more suitable.

Chapter 3

Reconstruction of major paternal lineages of the Strand Muslim community

3.1 Introduction

The recent demonstration of the Y-chromosome as an effective tool for studying and tracing paternal lineages prompted investigation into the origins of the paternal lineages of the Strand Muslim community. The use of SNP SNaPshot genotyping system would allow for more information to be gathered which would complement previous historical studies of this community that have already taken advantage of the availability of the huge amount of historical data surrounding the Strand Muslim community.

Four multiplex reactions and 1 singleplex were used based on the literature of Onofri *et al* (2006). These were as follows. Multiplex 1: M89, M173, M170, M45, Multiplex 2: M52, M 96, M91, M181, Multiplex 3: M9, M172, Multiplex 4: M201, M216, M214 and Singleplex 5: M174. The results of this investigation are described in this chapter.

3.2 Materials and Methods

3.2.1 SNP Selection

SNPs selection was based on variations of paternal lineages found in African, Asian and European populations. Fourteen Y chromosomal markers and their respective primers were selected from the literature of Onofri *et al.* (2006).

3.2.2 Terminology

Haplogroup nomenclature followed the conventions recommended by the (Y-chromosome Consortium, 2002) for naming Y-chromosome lineages. Capital letters A-R identify the 18 major Y-chromosome clades and haplogroups. In this study the shorthand mutation- based naming system was employed. This system is based on retaining the major haplogroup information (i.e. 19 capital letters) followed by the name of the terminal mutation that defines a given haplogroup example F-M89, I-M170 and R-M173 used in text and figures (Figure 3.1.).

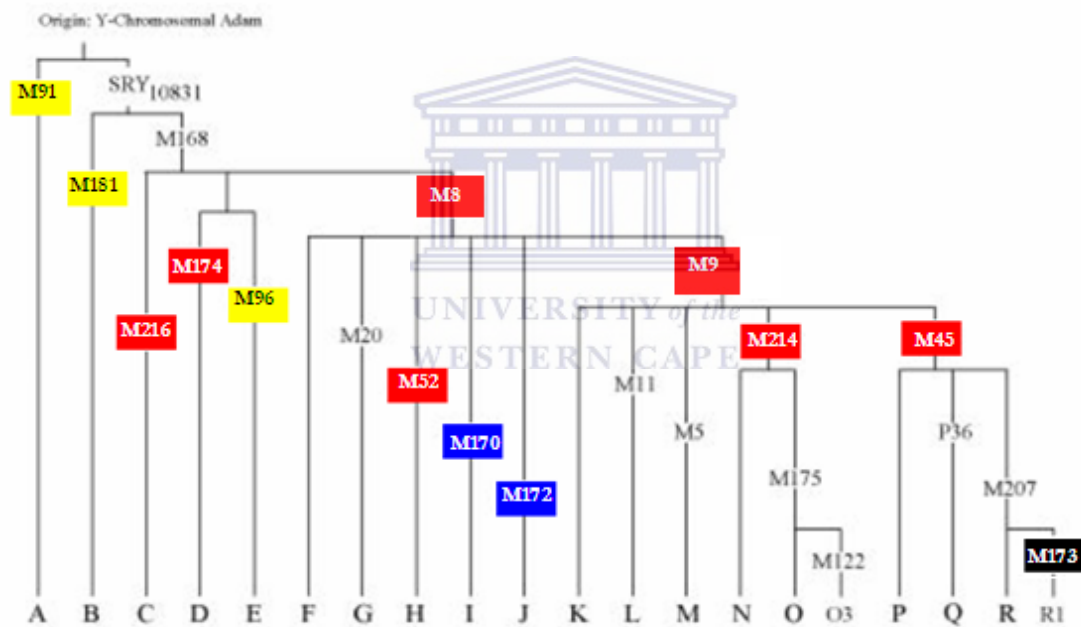


Figure 3.1: Y-chromosomal tree depicting haplogroups A-R. Colour boxes indicate haplogroups investigated. Yellow: African haplogroups, Red: Asian haplogroups, Blue: European haplogroups and Black indicating Eurasian/European haplogroups (adapted from <http://www.genebase.com/image/ydnaSnpBackboneChart01.jpg>)

Binary Markers were placed into four multiplex panels and a singleplex of an additional marker obtained later in the study. Onofri *et al* (2006) stated that markers be pooled to work in two multiplexes however initial amplification attempts at this failed and

selected SNPs were organized into multiplex panels namely MY1, MY2, MY3, MY4 and singleplex SY5 (Table 3.1).

Table 3.1: Y-chromosomal multiplex panels depicting discriminatory SNPs defining specific haplogroups organized into four multiplexes and one singleplex.

Multiplex	Marker	Haplogroup
MY1	M89	F
	M173	R1
	M170	I
	M45	P
MY2	M52	H
	M96	E
	M91	A
	M181	B
MY3	M9	K
	M172	J2
MY4	M201	G
	M216	C
	M214	NO
SY5	M174	D

3.2.3 Biological Samples

Samples were obtained in the form of buccal swabs from a 115 unrelated Muslim males born in the Muslim community of the Strand. Total DNA from buccal cells was extracted using the Buccal AmpTM Quick Extract DNA extraction kit (*Epicentre Technologies*) as per the manufacturers' recommendations. Hereafter samples were further purified using Phenol-Chloroform-Isoamyl alcohol purification method. The step by step protocol for this method can be found in Appendix III.

3.2.4 DNA Quantification

DNA samples were quantified using a Nanodrop *ND 1000 UV-Vis* spectrophotometer (*Applied Biosystems*) as per the manufacturers' recommendations.

3.2.5 PCR Amplification

The standard reaction components for the multiplex and singleplex amplifications were performed in a final volume of 17µl containing 20-40ng genomic DNA, 8µl QIAGEN Multiplex reaction mix (*Qiagen*), 1.7µl of Qiagen Q Solution (*Qiagen*), 4.6µl of RNase free water (*Qiagen*) and the appropriate primers (Table 3.2 and Table 3.3).

Primers were synthesized by (*Whitehead Scientific*) using previously reported sequences cited in Onofri *et al* (2006). The reported primer sequences were aligned against the human genome reference sequence. This was a precautionary measure to avoid any unspecific amplification and extension products. The standard operating procedure for preparing the primer mixtures can be found in Appendix III.

Table 3.2: Y-chromosomal PCR conditions for MY1, MY2, MY3 and MY4

Multiplex set		Primer Sequence ('5—3')	Amplicon size (bp)	Primer concentration (µM)
MY1	MY2			
M170		F:GTTTTCATATTCTGTGCATTATACAAATTACTAT	96	0.23
		R:CATTTTACAGTGAGACACAACCCAC		
M173		F:AAAATTTTCTTACAATTCAAGGGCATT	104	0.23
		R:GCTGCAGTTTCCAGATCCT		
M45		F:GGTGTGGACTTTACGAACCAACCT	109	0.23
		R:CCTGGACCTCAGAAGGAGCTTT		
M89		F:CTGCTCAGCTTCCTGGATTCA	110	0.23
		R:CACTTTGGGTCCAGGATCACC		
	M52	F:CTCCACCTCAACTTCCCAGAG	153	0.29
		R:AGCAAACATTTCAAGAGAGAATGAAA		
	M96	F:TTCTCCATATCTGTGTAAAGGCAAGT	165	0.29
		R:CCATAGGTTTTTAATATTATACCTGAGTG		
	M181	F:GCTAGCAAAGTTGGCTTGGG	166	0.29
		R:GCACACTAGCTATAAGCAAAAAGAAAT		
	M91	F:ATTGCGATGTTTTATTTCAAACAAGATG	170	0.29
		R:GCGTATTTTTCAAAAATATATGGAGAA		



Multiplex set		Primer Sequence ('5—3')	Amplicon size (bp)	Primer concentration (μM)
MY3	MY4			
M9		F: AGAACTGCAAAGAAACGGCCTA	127	0.35
		R:TGCATATAATGAAGTAAGCGCTACCT		
M172		F:TTTTATCCCCAAACCCATTT	136	0.35
		R:CATGTTGGTTTGGAACAGTTTATCC		
	M216	F:AAGCCACTTAAATTCCAATGGA	171	0.29
		R:CACTGCTAGTTATGTATACCTGTTGAA		
	M214	F:CAATTGTACAGCACAAATATATGCCTGTAAA	207	0.29
		R:GAGGTCAAGGGTGTGGTGAG		
	M201	F:TATGCATTTGTTGAGTATATGTCAAAT	163	0.29
		R:TCCAACACTAAGTACCTATTACGAAAA		

Table 3.3: Y-chromosomal PCR conditions for SY5.

Singleplex	Primer Sequence ('5--3')	Amplicon size (bp)	Primer concentration. (μ M)
M174	F:AAATGTACGTTTTTGGTTTACTCATAATG R:TGCAAAAGGAGAAGGACAAGA	169	0.29

PCR thermocycling was conducted using a GeneAmp 9600 thermo cycler (*Applied Biosystems*) according to the PCR conditions stipulated in the QIAGEN Multiplex kit (*QIAGEN*) protocol manufacturers' manual. Conditions were as follow: 1 cycle at 95°C for 15 minutes, 40 cycles of 94°C for 30 seconds, 60°C for 90 seconds, 72°C for 90 seconds, followed by 1 final cycle at 72°C for 10 minutes.

Agarose gel electrophoresis was used to verify PCR products. An aliquot of 5 μ l of each product was run on a 2% agarose gel stained with 0.1 μ g/ml Ethidium Bromide.

3.2.6 PCR Product Purification

Following verification, the PCR products were purified using *SAP* and *EXO I* (*USB Corporation*) according to the manufacturer's instructions. The *SAP* (shrimp alkaline phosphatase) digests unincorporated dNTPs while the exonuclease digests residual primers. Aliquots of 5.5 μ l of *SAP* and 0.55 μ l of *EXO I* were mixed with 15 μ l of each product and incubated at 37°C for an hour followed by 15min of denaturation at 75°C.

3.2.6.1 Single Base Extension

Sequencing was performed using the ABI PRISM® SNaPshot™ Multiplex kit (*Applied Biosystems*) according to the manufacturer's instructions. The SNaPshot protocol uses fluorescently-labelled dideoxy terminator incorporation at the site of interest followed by detection on a capillary electrophoresis instrument. Multiplexing was achieved by varying the length of extension primers which provide spatial separation in detection.

Extension reactions for MY1 and MY2 were performed in a final reaction volume of 13.2µl. Extension reactions for MY3, MY4 and SY5 were performed in a final reaction volume of 14µl. All reactions contained 2µl SNaPshot Ready Reaction mix, (*Applied Biosystems*), 2µl template DNA (*Exo* and *SAP* treated), appropriate volume of Sabax water and appropriate primer synthesised by (*Applied Biosystems*) at the specific concentration (Tables 3.4, 3.5 and 3.6).

Extension cycling conditions was performed in a GeneAmp 9600 thermo cycler (*Applied Biosystems*) as follows: 25 cycles of 95°C for 10 seconds, 55°C for 5 seconds and 60°C for 30 seconds.

Table 3.4: Y-chromosomal extension primer sequences, final concentration of primers used in multiplex reactions MY1 and MY2 and primer length

Marker Multiplex set		Primer Sequence ('5--3')	Primer length (nt)	Poly-morphism	Primer concentration (µM)
MY1	MY2				
M170		F:AAATTACTATTTTATTTACTTAAAAATCATTGTTC	35	A/C	0.15
M173		F: CAATCAAGGGCATTAGAAC	21	A/C	0.15
M45		F:(6-POLY-GACT) AAATTGGCAGTGAAAAATTATAGATA	50	G/A	0.15
M89		R:(34-POLY-T) CAACTCAGGCAAAGTGAGAGAT	56	C/T	0.15
	M52	F:ATACCTATAAGAATATTGCCTGCA	24	A/C	0.35
	M96	F:(6-POLY-GACT) GTAACCTGGAAAAACAGGTCTCTCATA ATA	53	G/C	0.35
	M181	F:(4-POLY-GACT) GGACAACTTGATCATCTTTTGA	39	T/C	0.35
	M91	R:(4-POLY-GACT) GATACTACAGTAGTGAAGTATTAATAAAAAA	47	9T/8T	0.35

Table 3.5: Y-chromosomal extension primer sequences, final concentration of primers used in multiplex reactions MY3 and MY4 and primer length

Marker Multiplex set		Primer Sequence (5'-3')	Primer length (nt)	Poly-morphism	Primer concentration (μ M)
MY3	MY4				
	M9	F:(6-POLY-GACT)AAATTGGCAGTGAAAAATTATATAGATA	44	C/G	0.53
	M172	R:AAGAAAATAATAATTGAAGACCTTTTAAGT	30	C/T	0.35
	M216	R:CACTGCTAGTTATGTATACCTGTTGAAT	28	T/C	0.21
	M214	R:GA(8-POLY-GACT)AGTGTGAGACACTGTCTGAAAACAAC	60	G/T	0.21
	M201	F:(4-POLY-GACT)AGATCTAATAATCCAGTATCAACTGAGG	44	G/T	0.28



Table 3.6: Y-chromosomal extension primer sequences, final concentration of primers used in singleplex reaction SY5 and primer length

SY5 Marker	Primer Sequence (5' → 3')	Primer length (nt)	Poly-morphism	Primer concentration (μ M)
M174	R:(16-POLY-T TAIL) CACCCCTCACTTCTGCACT	35	T/C	0.35

3.2.6.2 Single Base Extension Product Purification

Following single base extension reactions, the sequencing products were further purified to remove unincorporated Fluorescent (F) ddNTPs using *SAP (USB Corporations)* as described by the manufacturers. Aliquots of 1 μ l *SAP (USB Corporations)* was used to treat extended products. Samples were incubated at 37⁰C for an hour and the enzyme was inactivated at 75⁰C for 15 minutes.

3.2.6.3 Capillary Electrophoresis

Amplicons were analysed using an automated ABI 310 *XL excelsior* colour sequencer (*Applied Biosystems*). In preparation for running, 0.35 μ l product were resuspended in 0.5 μ l internal size standard Liz 120 (*Applied Biosystems*) and 10.65 μ l of Hi- Di Formamide (*Applied Biosystems*) in a final volume of 11.5 μ l. Samples were then loaded onto a MicroAmp Optical 96-Well Reaction Plate and denatured for 5 minutes at 95⁰C in a GeneAmp PCR System 9600 thermal cycler (*Applied Biosystems*)

3.2.6.4 Data Analysis

Electrophoresis was conducted and the data was analysed according to manufacturer instructions. Electrophoresis was performed using a 36 cm X 16 µM capillary filled with denaturing performance polymer POP 7 (*Applied Biosystems*). Run voltage was 15kV, Injection voltage was 12 kV, injection duration was 23 seconds and temperature was 60°C. A five dye of dR110, dR6G, dTAMRA and dROX was used with E5 filter according to manufacturer's instructions.

Data was collected with the ABI 310 *XL* collection software and analyzed with Gene Mapper software v4.0 (*Applied Biosystems*). Electropherograms were reviewed to detect any potential problematic results that may include the presence of any peak other than the expected peak, unexpected migration of expected peaks, peak overlap and low peak relative fluorescence unit (RFU). Genotypes were compared both manually and using published nomenclature of YCC. Genotyping was based on peak size where primer sizes were compared to the internal size standard, (RFU) and peak colour.

3.3 Results and Discussion

3.3.1 DNA Samples and Amplification

All 115 samples collected from unrelated Muslim males born within the Strand were successfully extracted, genotyped and analyzed. The relatively small SNP panels and the implementation of the Qiagen multiplexing Kit for amplification resulted in intense bands in all multiplexes that required minimum amounts of DNA reducing the complexity of optimizing the multiplex PCR reactions for the samples used in the study.

3.3.2 Genotyping

Electrophoresis using the ABI 310 XL excelsior DNA sequencer generated electropherograms that could be easily interpreted. However 9 samples were reanalyzed to achieve better profiles.

3.3.2.1 Y-chromosomal multiplex I mini-sequencing and genotyping

Four markers were amplified within this multiplex therefore when analysing the SNaPshot products four fluorescently detectable nucleotide extension peaks were obtained. Genotyping was clear as each peak was visible with no room for misinterpretation. However it was observed in three subjects that Marker M173 and M89 exhibited double peaks this made typing for these samples difficult and required that the samples be reamplified in PCR amplification. Possible reasons for double peaks were attributed to insufficient purification before and after extension. Extension products for this multiplex ranged from 21-56 nucleotides. Primer concentrations allowed for sufficient RFU peak heights of above 400 that can be observed in the example electropherograms (Figure 3.2).

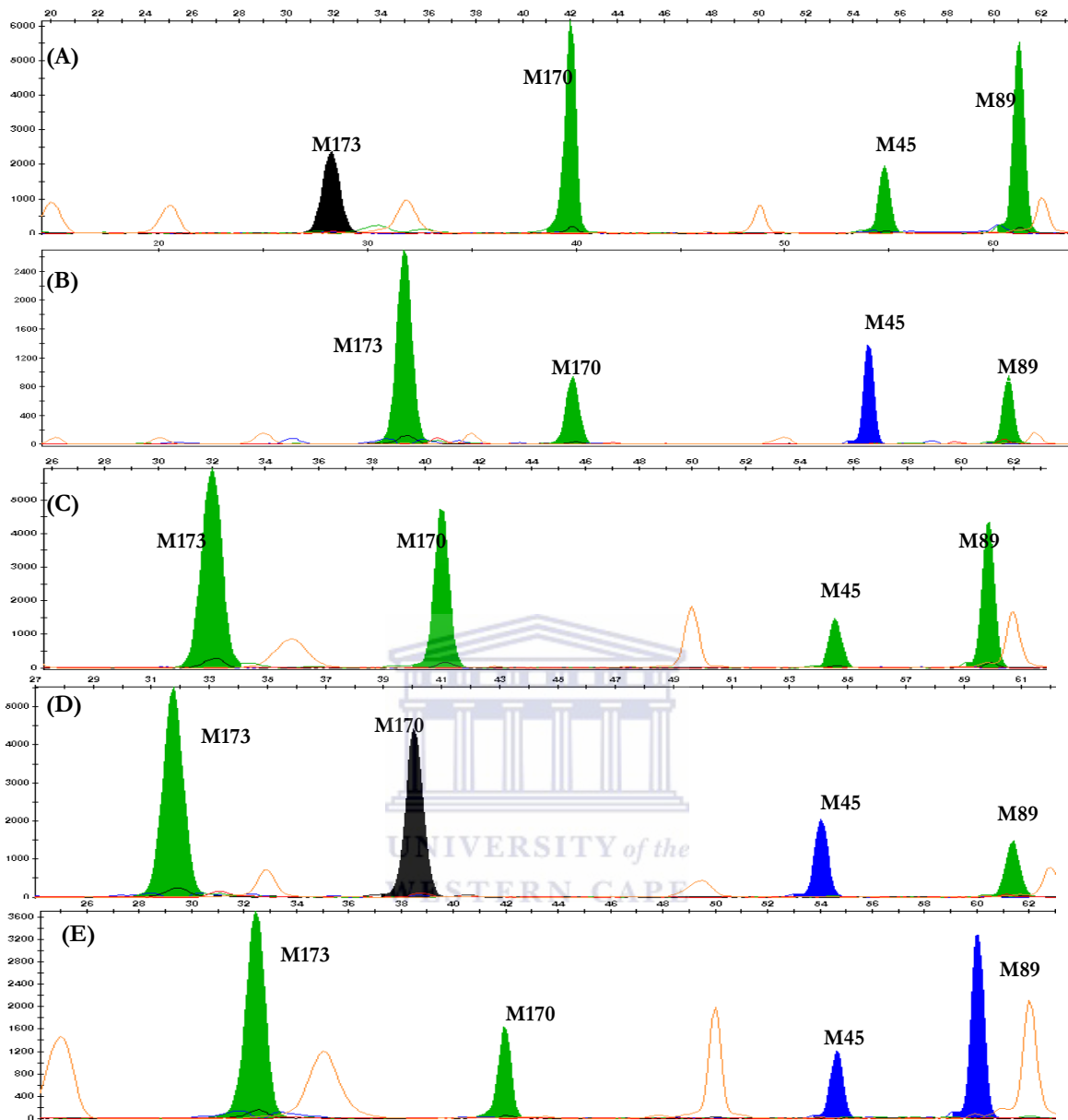


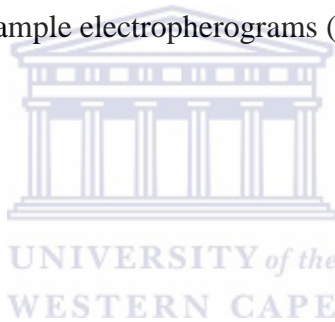
Figure 3.2: Electropherograms of MY1: Four-marker multiplex analysed with *Gene mapper 4.0 software*. X-axis represents the size (bp) of the primers with the incorporated nucleotide. The Y-axis corresponds to the RFU of the peak. Derived alleles are indicated with their respective colour for each peak where green represents nucleotide A, black: nucleotide C, blue: nucleotide G and red: nucleotide T. **A)** sample indicating derived state for marker R1-M173, **B)** sample indicating derived state for marker F-M89, **C)** sample indicating derive state for marker P-M45, **D)** sample indicating derive state for marker I-M170 and **E)** sample indicating no derive state for any marker within MY1.

Figure 3.2 (D) shows an individual that was classified as having the C allele (black peak) for M170. Subjects that tested positive for I-M170, R1-M173 and P-M45 in MY1 were

further typed in MY3 and samples that tested positive for F- M89 were carried over to both MY3 and MY4. Figure 3.2 (E) shows a sample not having any of the derived states for MY1 and is an example of a sample that would be subjected to MY2.

3.3.2.2 Y-chromosomal multiplex II mini-sequencing and genotyping

Genotyping was clear as each peak was visible with no room for misinterpretation. However initial results of MYII rendered low RFUs in nearly all samples typed. Increasing the amount of DNA for these samples in the wells resulted in peak heights of above 400RFUs. Extension products for this multiplexed ranged from 24-53 nucleotides these can be observed in the example electropherograms (Figure 3.3.).



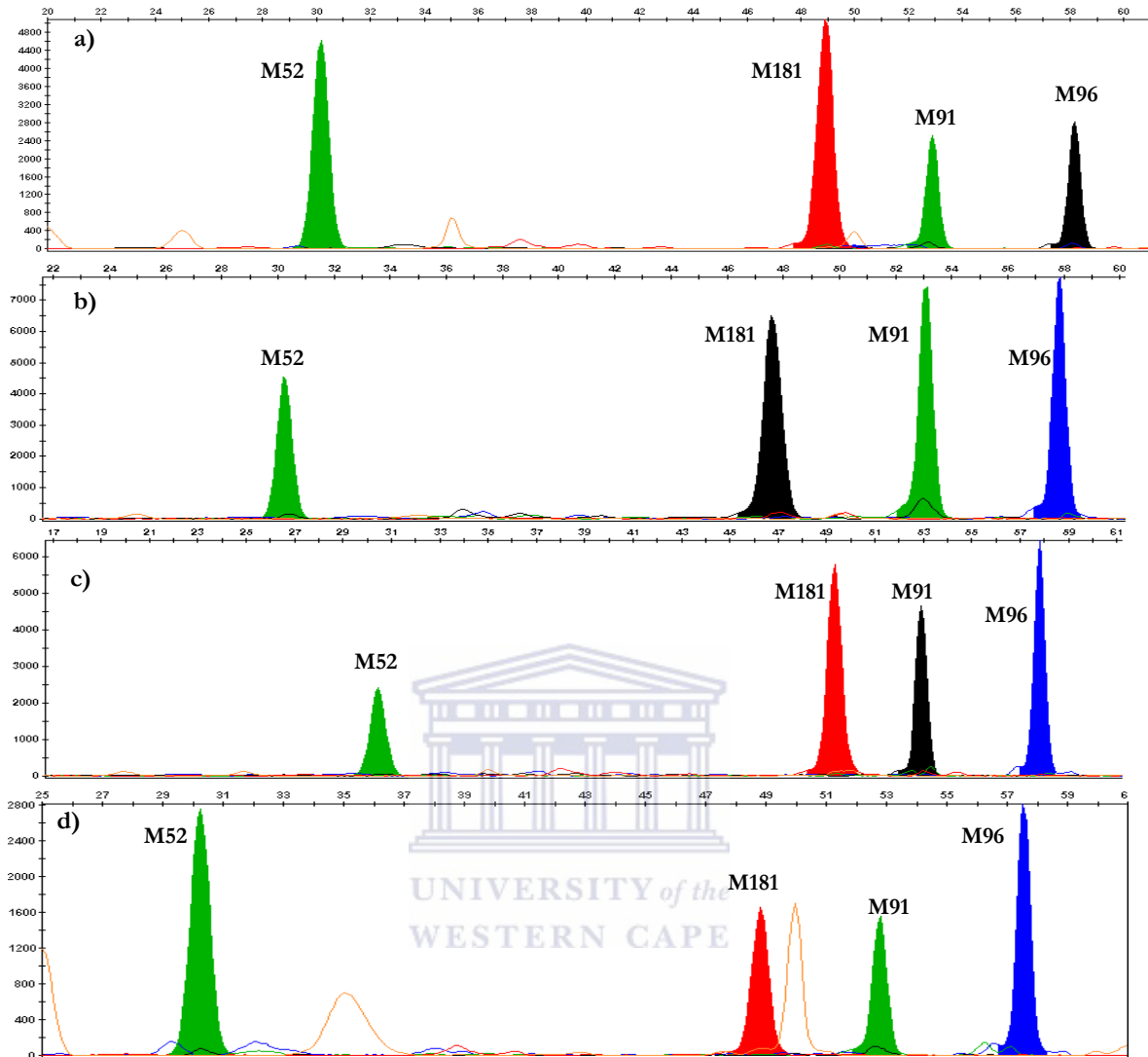


Figure 3.3: Electropherograms of samples tested for MY2, a four marker multiplex analysed with *Gene Mapper 4.0* software. X-axis represents the size (bp) of the primers with the incorporated nucleotide. The Y-axis corresponds to the RFU of the peak. Derived alleles are indicated with their respective colour for each peak where green represents nucleotide A, black: nucleotide C, blue: nucleotide G and red: nucleotide T. **a)** sample with derived state for marker E-M96, **b)** sample with derived state for marker B-M181, **c)** sample with derived state for marker A-M91 and **d)** sample that did not test positive for any of the markers tested within this multiplex.

As demonstrated in figure 3.3 (a), (b) and (c), subjects that tested positive for a MY2 marker could be successfully genotyped and placed within a haplogroup. In the case of the male individual indicated in the figure 3.3 in the electropherogram marked (d) the individual had not tested positive for any of the markers in MY2 and was thus subjected to further genotyping in SY5.

3.3.2.3 Y-chromosomal multiplex III mini-sequencing and genotyping

MY3 multiplex consisted of two markers M172 and M9. These markers followed sequential typing to further verify haplogroup assignment in all subjects that had tested positive for the derived state of markers R-M173, P-M45 and F-M89 typed in MY1. In this multiplex a male individual would be classified as having either a G allele (blue peak) for K-M9 locus or a C allele (black peak) for J2-M172 locus. A positive result obtained for samples that were assigned to haplogroup J were not further genotyped however subjects that had tested positive for K-M9 were further typed in MY4 (Figure 3.4).

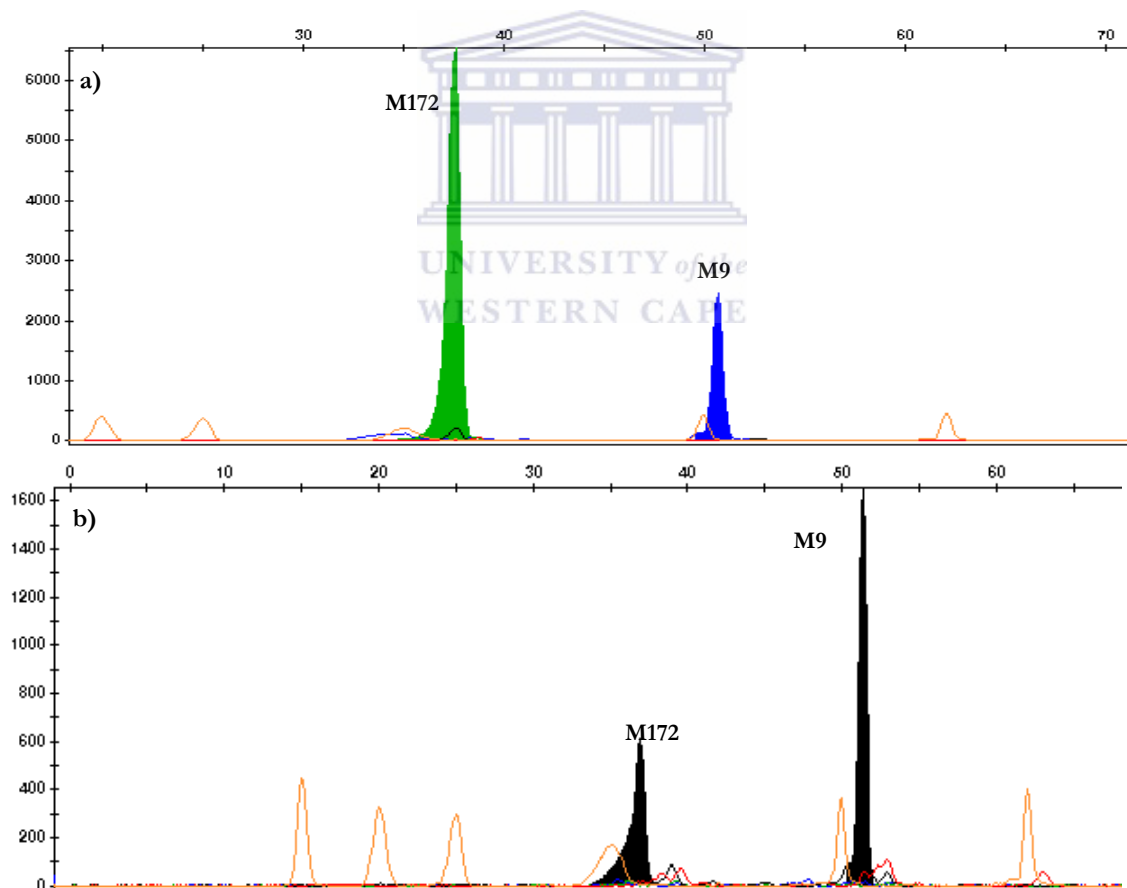


Figure 3.4: Electropherograms of MY3 consisting of two markers. X-axis represents the size (bp) of the primers with the incorporated nucleotide. The Y-axis corresponds to the RFU of the peak. Derived alleles are indicated with their respective colour for each peak where green represents nucleotide A, black: nucleotide C, blue: nucleotide G and red: nucleotide T. **a)** sample indicating derived state for marker K-M9 and **b)** sample indicating derived state for marker J2-M172.

3.3.2.4 Y-chromosomal multiplex IV mini-sequencing and genotyping

MY4 multiplex was the last step in the hierarchical approach of genotyping samples that had tested positive for K-M9 in multiplex MY3. Based on results obtained male individuals could be classified as belonging to haplogroup K, R1, F or NO (Figure 3.5).

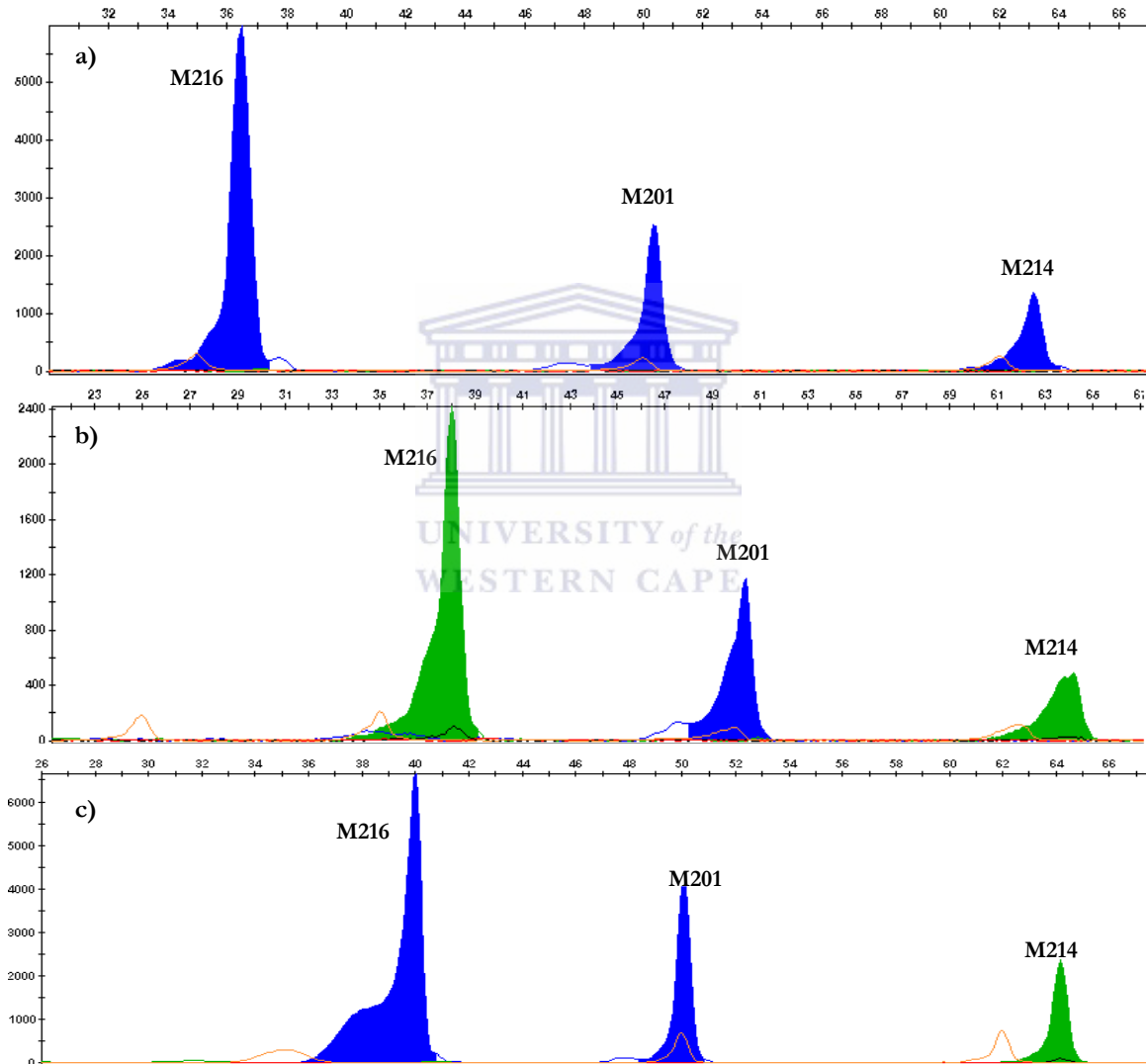


Figure 3.5: Electropherograms of three markers multiplex MY4 analysed with Gene Mapper 4.0. X-axis represents the size (bp) of the primers with the incorporated nucleotide. The Y-axis corresponds to the RFU of the peak. Derived alleles are indicated with the respected colours where green represents nucleotide A, black: nucleotide C, blue: nucleotide G and red: nucleotide T. **a)** sample indicating derive state for marker NO-M214, **b)** sample indicating derive state for marker C-M216 and **c)** sample that does not carry any of the derived state for MY4.

In the above electropherogram marked (e) the individual that did not carry the derive allele for any of the markers were either classified as R1, P, K or F. These haplogroups were assigned based on derived states observed in MY1 and MY3 for these defining markers. This multiplex was the last step to test for haplogroup assignment. However two samples did not fall into any multiplex and an additional singleplex was implemented namely SY5 (figure 3.6).

3.3.2.5 Y-chromosomal Singleplex V

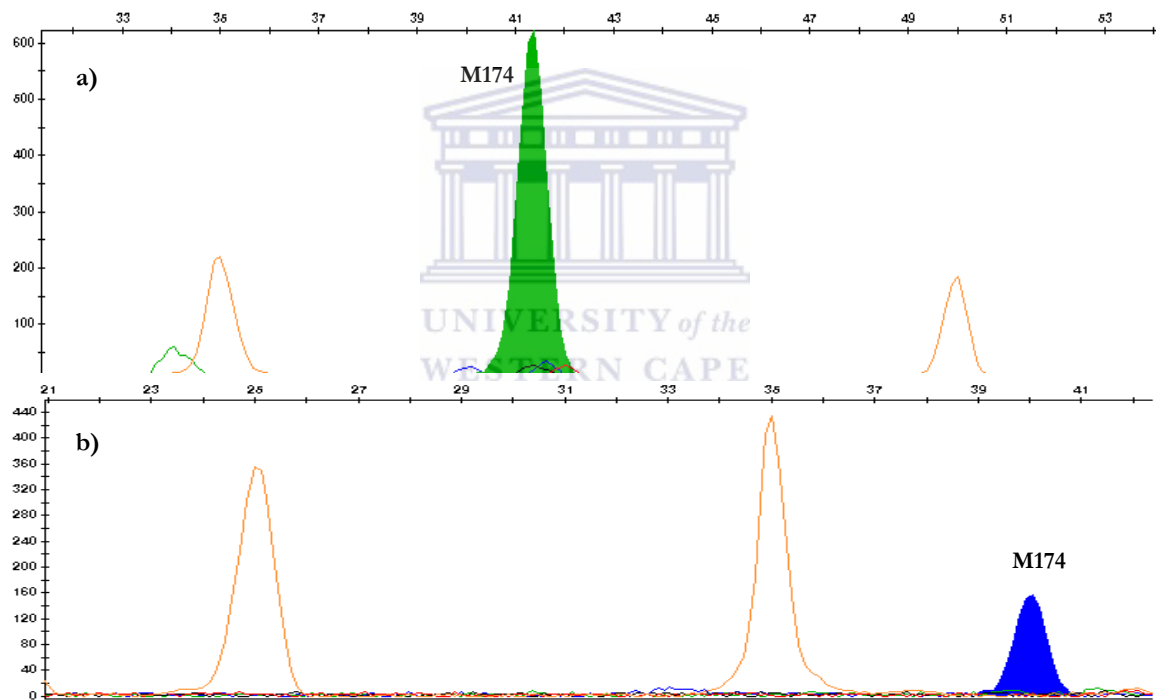


Figure 3.6: Electropherograms of SY5 analysed with Gene Mapper 4.0. X-axis represents the size (bp) of the primers with the incorporated nucleotide. The Y-axis corresponds to the RFU of the peak. Derived alleles are indicated with the respected colours where green represents nucleotide A, black: nucleotide C, blue: nucleotide G and red: nucleotide T. **a)** sample indicating ancestral state for marker D-M174, **b)** sample indicating derive state for marker D-M174

Singleplex SY5 consisted of a single marker D-M174. Genotyping this marker in singleplex was relatively easy. Samples that did not have any of the derived states for

markers listed in MY1, MY2, MY3 and MY4 was subjected this singleplex. Sample in figure 3.6, sample (b) was typed positive for this marker and could be successfully assigned to a haplogroup.

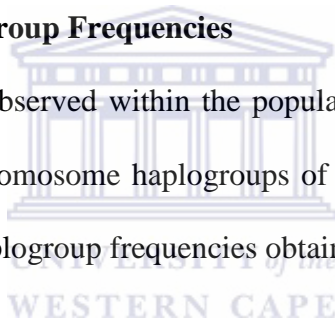
3.4 Y-chromosomal Haplogroup Observed

Following successful multiplex amplification and genotyping of SNPs, major Y-chromosome haplogroups were determined. A total of fourteen haplogroups were examined and only twelve Y-chromosome haplogroups were found in the population

.

3.4.1 Y-chromosomal Haplogroup Frequencies

The twelve biallelic markers observed within the population were used to construct the major continent-specific Y-chromosome haplogroups of the Strand Muslim community. Figure 3.7 below shows the haplogroup frequencies obtained within this population set.



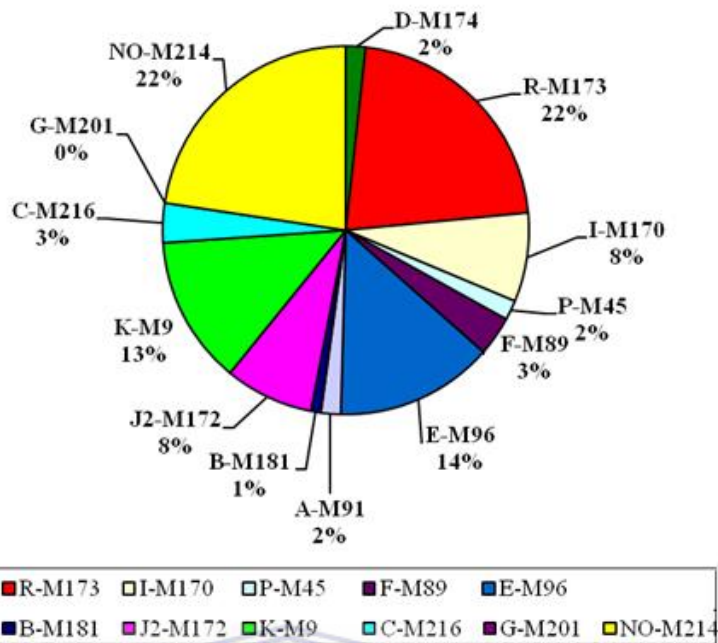
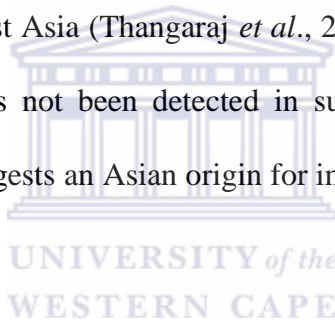


Figure 3.7: A pie chart representation of the percentage of paternal haplogroup frequencies observed in a 115 unrelated males from the Strand Muslim Community.

The Y-chromosome haplogroups observed in the population were B, A, D, P, C, F, I, J, K, E, R1 and NO. Haplogroup B accounted for the lowest percentage frequency of 1%. This haplogroups is one of the three African haplogroups on the Y-chromosome tree that shares a broad dispersal across the African continent. However it is predominantly found within east sub Saharan Africa with greatest frequency among the Pygmy population (Jobling and Tyler-Smith, 2003). Haplogroup A, D and P shared similar frequencies of 2% each within the population. As haplogroup A and haplogroup B represents the deepest branch of the Y-chromosome phylogeny it is believe to provide the first genetic link to the earliest common ancestor of humans (Wells, 2007). These haplogroup are also scattered widely at low frequency across the African continent with high frequencies only being observed in hunter gathers populations (ancient lineages).

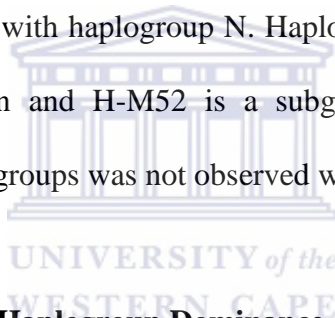
Haplogroup D have not been found anywhere outside of Asia where it most likely originated. This lineage is therefore commonly found in central Asia, Japan and Tibet and at low frequencies in Southeast Asia and Andaman Islanders (Karafet *et al.*, 2001; Thangaraj *et al.*, 2002; Shi *et al.*, 2008). Another Asian haplogroup, haplogroup P was detected at similar frequencies as haplogroups A and D. Haplogroup P is found predominantly across Asia and East Asia. Haplogroup F is a lineage that confirms an out of Africa migration. It is found primarily on the Indian subcontinent at low to moderate frequency. This haplogroup appeared in the Strand Muslim at a low frequency of 3%, and shared a similar frequency as haplogroup C. Haplogroup C is a common Asian lineage found in central; South and East Asia (Thangaraj *et al.*, 2002; Zerjal *et al.*, 2002; Karafet *et al.*, 2005). This lineage has not been detected in sub-Saharan African populations (Karafet *et al.*, 2008). This suggests an Asian origin for individuals typed positive for this marker.



The proportion of haplogroup I and J2 within the population was equal as they shared a similar frequency of 8%. Haplogroup I is widespread within Europe more so Western Europe and is virtually absent anywhere else in the world (Cinnioglu *et al.*, 2004). Haplogroup J2 lineage is found at high frequencies in Middle Eastern, Balkan, Hungarian and southern Italian populations (Wells, 2007; Battaglia *et al.*, 2008). Haplogroup K and haplogroup E were found at moderate frequencies of 13 and 14 % respectively. Haplogroup K is widely distributed across North and Southeast Asia with prevalent frequencies found in Philippine and Indonesian populations (Tajima *et al.*, 2002; Karafet *et al.*, 2008). Low frequencies have also been found in Pakistani and East Indian

population. Haplogroup E is found at high frequencies in north African populations and at low frequencies within Central and south Asian populations suggesting a back migration of individuals who left Africa (Weale *et al.*, 2003; Rootsi, 2004).

Haplogroup R1 and NO had the highest frequency of 22% each in the Strand population. Haplogroup R1 is common in all non-African men and is found at high frequencies in western Eurasia (Kayser *et al.*, 2005, Battaglia *et al.*, 2008). Haplogroup NO combines the N and O lineage into one super clade. Haplogroup O accounts for 83.7% of South East Asians (Wise *et al.*, 2005; Krithika *et al.*, 2007). It is also observed at moderate frequencies in East Asia along with haplogroup N. Haplogroup G-M201 is a haplogroup confined to the Mediterranean and H-M52 is a subgroup of haplogroup H that is restricted to India. These haplogroups was not observed within the population.



3.4.2 Strand Y-chromosomal Haplogroup Dominance

Polymorphisms that define major Asian, African and European male lineages were selected for the study. The fourteen biallelic markers examined were grouped into three categories (Figure 3.8). African consisted of the three major African lineages A-M91, B-M181 and E-M96. Asian consisted of the markers F-M89, K-M9, P-M45, C-M216, D-M174, H-M52, G-M201 and NO-M214 and European that included the three markers I-M170, R1-M173 and J2-M172. Haplogroups having the same continent of origin were grouped together and the overall frequencies in the population were also determined (Figure 3.8.).

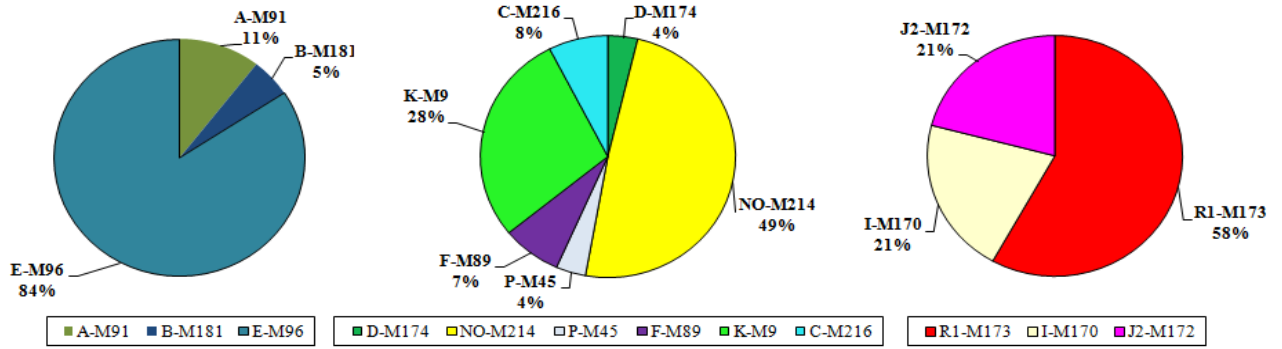
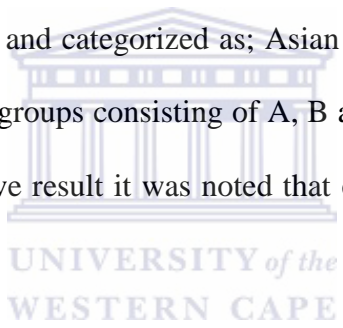


Figure 3.8: Representation of Y chromosomal haplogroup dominance observed within the Strand Muslim Community expressed in percentage. **A)** African haplogroups observed indicating E-M96 as the dominant African haplogroup found. **B)** Asian haplogroups observed indicating NO-M214 as the dominant Asian haplogroup found. **C)** European haplogroups observed indicating R1-M173 as the dominant European haplogroup found.

The resulting pie charts above indicates the continent specificity of the twelve out of fourteen haplogroups observed and categorized as; Asian haplogroup consisting of F, K, C, NO, D and P; African haplogroups consisting of A, B and E; European haplogroups I, R1 and J2. Examining the above result it was noted that each population category had a defining dominant haplogroup.



Haplogroup E dominates the African category as 84% of individuals assigned to an African haplogroup tested positive for this particular haplogroup. Haplogroup NO dominated the Asian category being present in 49% of all Asian defined individuals. Haplogroup R1 was found amongst 58% of those individuals classified as having European descent and dominated the European category. These three haplogroups E, NO and R1 as described by the YCC are representative of the most dominant lineages found to date with in present African, Asian and European population.

Analysing the haplogroup data it was noted that there was haplogroup prevalence for Asian lineages as six of the major Asian haplogroups out of the eight was found in this population. This suggests that Asian lineages are more prominent within this population; in fact the most frequent European haplogroup R1-M173 was unfortunately not further typed with sub markers to differentiate between R1-Asian and R1-European. This marker that is defined as European has subgroups of Asian descent making it in some studies a Eurasian marker. However in this analysis the M173 positive R1 individuals were classified as European although based on the Asian haplogroups prevalence it can be assumed that some unknown proportion could be R1-Asian.

3.5 Summary

The analysis of the Y-chromosome composition of the sample Muslim population allowed the detection of particular characteristics that is likely related to specific migratory events that have occurred in this area since the time of settlement. Asian haplogroups were dominated by the NO-M214 lineage found at 49% frequency in the population's Y-pool. African haplogroups were least observed however The African lineage with highest frequency was E-M96 occurring at 84% frequency. European lineages were found at moderate frequencies with R1-M173 dominating the European category with 58% frequency.

Chapter 4

Reconstruction of major maternal lineages of the Strand Muslim Community

4.1 Introduction

Mitochondrial polymorphisms have been extensively used to study the maternal composition and relationships of human populations and past migration events. The combination of control and coding region data allows grouping of mtDNA variants into haplogroups defined by one or more mtDNA coding region polymorphisms and particular control region sequences (Peričić *et al.*, 2005). In this chapter, the origin of the maternal lineages of the Strand Muslim community was investigated. As discussed in chapter 3, the efficient single base extension platform namely the SNP SNaPshot for analysis was used to determine major continent-specific mitochondrial haplogroups.

Two multiplex reactions were used each containing 7 and 5 mtDNA coding region SNPs respectively. These were as follow: Multiplex 1: 8272-8280 del, 10398, 10400, 3594, 7028, 12406, 7600 and Multiplex 2: 13263, 1719, 5178,663 and 4833.

4.2 Materials and Methods

4.2.1 SNP Selection

Twelve mtDNA SNPs were selected from previous published literature Nelson *et al* (2007). These SNPs are found throughout the coding region of the mtDNA and define haplogroups A, B, C, D, E, F, G, H, L1/L2, L3, M and N. The SNPs were selected for

their ability to discriminate among major African, Asian and European lineages (Figure 4.1).

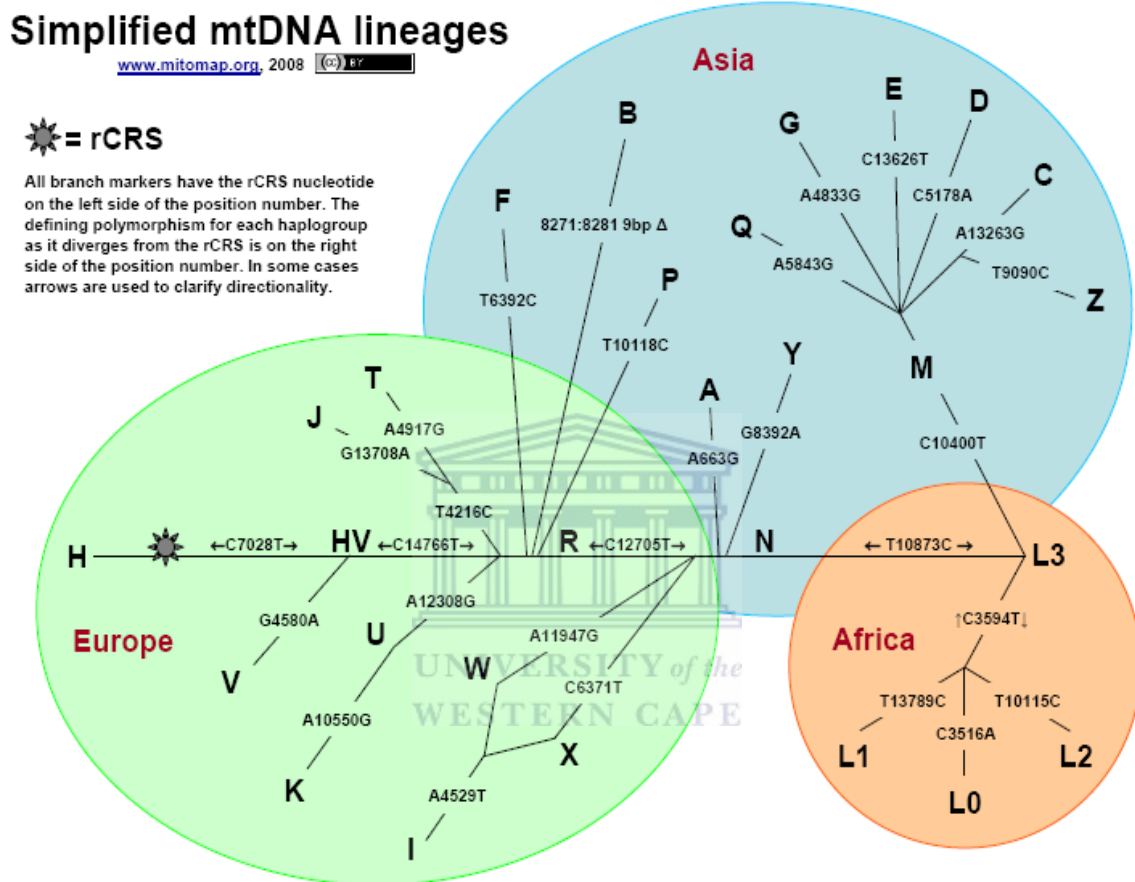


Figure 4.1: MtDNA haplogroups A-N. Colour circles indicating haplogroups prevalent on various continents. Peach: African Haplogroups, Blue: Asian Haplogroups and Green: European Haplogroups.

(adapted from <http://www.mitomap.org>)

4.2.2 Terminology

Nomenclature of mtDNA types follows that of the Cambridge Reference Sequence (CRS) (Carracedo *et al.*, 2000; Budowle *et al.*, 2003). The sequence of the sample will be identical to CRS except at that particular site (Carracedo *et al.*, 2000). The nomenclature system was established based on the difference detected between an individual's

sequence and CRS where only the site designated by number and nucleotide differing are recorded. Mitochondrial DNA SNPs were typed in two multiplexes namely multiplex one named M1 and multiplex two described as M2 (Table 4.1)..

Table 4.1 Mitochondrial PCR amplification multiplex panels with organized specific markers used in the study.

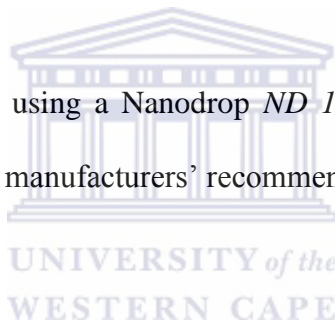
Multiplex one :M1	Multiplex two :M2
8272-8280 del	13263
10398	1719
10400	5178
3594	663
7028	4833
12406	
7600	

4.2.3 DNA Samples

Samples previously used in the analysis of the Y-chromosome in chapter 3 were used again in this part of the study. Buccal swabs from a 115 unrelated Muslim males born in the Muslim community of Strand were obtained. Total DNA from buccal cells was extracted using the Buccal Amp™ Quick Extract DNA extraction kit (*Epicentre Technologies*) as per the manufacturers' recommendations. Hereafter Samples were further purified using Phenol-Chloroform-Isoamyl alcohol purification method. The step by step protocol for this method can be found in Appendix III.

4.2.4 DNA Quantification

DNA samples were quantified using a Nanodrop *ND 1000 UV-Vis* Spectrophotometer (*Applied Biosystems*) as per the manufacturers' recommendations.



4.2.5 PCR Amplification

The standard reaction components for the multiplex amplifications were performed in a final volume of 12µl containing 20ng genomic DNA, 6µl QIAGEN Multiplex reaction mix (*Qiagen*), 0.6µl of Qiagen Q Solution (*Qiagen*), 2.4µl of *RNAse* free water (*Qiagen*) and the appropriate primers (Table 4.2).

Primers were synthesized by (*Whitehead Scientific*) using previously reported sequences cited in Nelson *et al* (2005). The reported primer sequences were aligned against the human genome reference sequence. This was a precautionary measure to avoid any unspecific amplification and extension products. The standard operating procedure for

preparing the primer mixtures can be found in Appendix III. A negative control reaction containing all reagents except DNA was also included with each PCR reaction.

PCR thermocycling was conducted using a GeneAmp 9600 thermo cycler (*Applied Biosystems*) according to the PCR conditions stipulated in the *QIAGEN Multiplex kit* (*QIAGEN*) protocol manufacturers' manual. Conditions were as follow: 1 cycle at 95°C for 15 minutes, 40 cycles of 94°C for 30 seconds, 60°C for 90 seconds, 72°C for 90 seconds, followed by 1 final cycle at 72°C for 10 minutes.

Agarose gel electrophoresis was used to verify PCR products. An aliquot of 5µl of each product was run on a 1.2% agarose gel stained with 0.1µg/ml Ethidium Bromide.

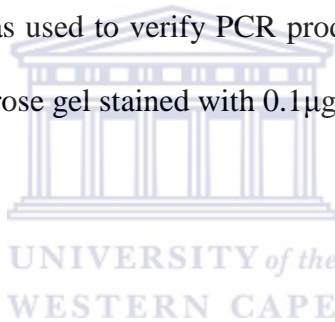


Table 4.2 Mitochondrial PCR conditions for M1 and M2

Multiplex set	Primer Sequence ('5--3')	Amplicon size (bp)	Primer concentration. (μM)
8272-8280	F:TAAAAATCTTTGAAATAGGGCCC R:GTTAATGCTAAGTTAGCTTTACAGTGG	89(del)80	0.35
10398/10400	F:AGTCTGGCCTATGAGTGACTAC R:AATGAGTCGAAATCATTTCGTTT	86	0.35
3594	F:CTTAGCTCACCATCGCTCT R:AGAATAAATAGGAGGCCTAGGTTG	90	0.35
7028	F:TATTAGCAAACATCACTAGACATCGT R:TGGCAAATACAGCTCCTATTGA	96	0.35
12406	F:AATTCCCCCATCCTTACC R:GCGACAATGGATTTACATAATG	78	0.35
7600	F:GGCTAAATCCTATATATCTTAATGGCA R:GGGAAGTAGCGTCTTGTAGACC	64	0.35
13263	F:CAAAAAAATCGTAGCCTTCTCC R:GTTGATGCCGATTGTAAC TATTATG	67	0.35
1719	F:CCC ACTCCACCTTACTACCAGA R:TGCGCCAGGTTTCAATTT	84	0.35
5178	F:TAAACTCCAGCACCACGACC R:GTGGATGGAATTAAGGGTGTAG	79	0.35
4833	F:AATAGCCCCCTTTCACCTCCTG R:AGAAGAAGCAGGCCGGA	72	0.35
663	F:ACATCACCCATAAAACAAATAGG R:TGGTGATTTAGAGGGTGA ACTCA	108	0.35

4.2.6 PCR Product Purification

Following verification, the PCR products were purified using *SAP* and *EXO I* (*USB Corporation*) according to the manufacturer's instructions. The *SAP* (shrimp alkaline phosphatase) digests unincorporated dNTPs while the exonuclease digests residual primers. Aliquots of 2µl of *SAP* and 2µl of *EXO I* were mixed with 5µl of each product and incubated at 37°C for an hour followed by 15min of denaturation at 75°C.

4.2.7 Single Base Extension

Sequencing was performed using the ABI PRISM® SNaPshot™ Multiplex kit (*Applied Biosystems*) according to the manufacturer's instructions. The SNaPshot protocol uses fluorescently-labelled dideoxy terminator incorporation at the site of interest followed by detection on a capillary electrophoresis instrument. Multiplexing was achieved by varying the length of extension primers which provide spatial separation in detection.

Extension reactions for M1 and M2 were performed in a final reaction volume of 14µl. Reactions contained 2µl SNaPshot Ready Reaction mix (*Applied Biosystems*), 2µl template DNA (*Exo* and *SAP* treated), appropriate volume of Sabax water and appropriate primer synthesised by (*Applied Biosystems*) at the specific concentration (Tables 4.3). Extension cycling conditions was performed in a GeneAmp 9600 thermo cycler (*Applied Biosystems*) as follows: 25 cycles of 95°C for 10 seconds, 55°C for 5 seconds and 60°C for 30 seconds.



UNIVERSITY *of the*
WESTERN CAPE

Table 4.3 Mitochondrial extension primer sequences, Primer length and final concentration of primers used in multiplex reactions M1 and M2.

Marker Multiplex set	Primer Sequence ('5--3')	Primer length (nt)	Polymorphism	Primer concentration (μ M)
8272-8280	F:CCCTATAGCACCCCCTCTA	19	C/G	0.37
10398	F:(21-POLY-T) GAGTGA CTACAAAAAGGATTAGACTGA	48	A/G	0.25
10400	R:(24-POLY-T) TTCGTTTTGTTAAACTATATACCAATTC	53	C/T	0.29
3594	R:(29-POLY-T) TAGGAGGCCTAGGTTGAGGTT	58	C/T	0.25
7028	R:(33-POLY-T) CCTATTGATAGGACATAGTGGAAGTG	63	C/T	0.27
12406	F:(50-POLY-T) CCCATCCTTACCACCCTC	68	G/A	0.25
7600	F:(51-POLY-T) TATCTTAATGGCACATGCAGC	78	G/A	0.25
13263	F:(3-POLY-T) TAGCCTTCTCCACTTCAAGTCA	25	A/G	0.13
1719	F:(7-POLY-T) CTCCACCTTACTACCAGACAACCTTA	33	G/A	0.25
5178	F:(13-POLY-T-tail) CTACTATCTCGCACCTGAAACAAG	37	C/A	0.14
4833	F: (54-POLY-T-tail) CCAGAGGTTACCCAAGGC	73	A/G	0.25
663	F:(19-POLY-T-tail) CCATAAACAAATAGGTTTGGTCCT	43	A/G	0.15

4.2.8 Single Base extension product purification

Following single base extension reactions, the sequencing products were further purified to remove unincorporated Fluorescent (F) ddNTPs using *SAP (USB Corporations)* as described by the manufacturers. Aliquots of 2µl *SAP (USB Corporations)* was used to treat extended products. Samples were incubated at 37⁰C for an hour and the enzyme was inactivated at 75⁰C for 15 minutes.

4.2.9 Capillary Electrophoresis

Amplicons were analysed using an automated ABI 310 *XL excelsior* colour sequencer (*Applied Biosystems*). In preparation for running, 0.35µl product were resuspended in 0.5µl internal size standard Liz 120 (*Applied Biosystems*) and 10.65µl of Hi- Di Formamide (*Applied Biosystems*) in a final volume of 11.5µl. Samples was then loaded onto a MicroAmp Optical 96-Well Reaction Plate and denatured for 5 minutes at 95⁰C in a GeneAmp PCR System 9600 thermal cycler (*Applied Biosystems*)

4.2.10 Data Analysis

Electrophoresis was conducted and the data was analysed according to manufacturer instructions. Electrophoresis was performed using a 36 cm X 16 µM capillary filled with denaturing performance polymer POP 7 (*Applied Biosystems*). Run voltage was 15kV, Injection voltage was 12 kV, injection duration was 23 seconds and temperature was 60⁰C. A five dye of dR110, dR6G, dTAMRA and dROX was used with E5 filter according to manufacturer's instructions.

Data was collected with the ABI 310 *XL* collection software and analyzed with Gene Mapper software v4.0 (*Applied Biosystems*). Electropherograms were reviewed to detect any potential problematic results that may include the presence of any peak other than the expected peak, unexpected migration of expected peaks, peak overlap, and low peak relative fluorescence unit (RFU). Genotypes were compared both manually and according to the Cambridge Reference Sequence. Genotyping was based on peak size where primer sizes were compared to the internal size standard, (RFU) and peak colour.

4.3 Results and Discussion

4.3.1 Biological sample and DNA amplification

All 115 samples collected from unrelated Muslim males born within the Strand were successfully extracted, genotyped and analyzed. The implementation of the Qiagen Kit for amplification resulted in intense bands in all multiplexes that required minimum amounts of DNA reducing the complexity of optimizing the multiplex PCR reactions for the samples used in the study.

4.3.2 Genotyping

Upon the completion of amplification and SNP SNaPshot mini-sequencing haplogroups were successfully assigned. Fairly sufficient separations were obtained during electrophoresis for both multiplexes. However, overlaps in peaks were obtained for some of the samples. Reasons for this were associated with the use of polymer 7 and not polymer 4 as the difference in polymer was found to induce peak migration. Peak overlaps were also thought to be the result of difference in electrophoretic mobility of

extension primers that is determined by the length, sequence and dye. The overlap in peaks did not however interfere with analysis as genotyping was based on both peak colour and peak size. Extension primers of multiplex panel M2 did not produce any of these problems although an error in the journal resulted in re-synthesising of marker 1719 as it exhibited the wrong peak colour in electropherogram. Corresponding with the authors of the paper identified a typo in the extension primer sequence and analysis was repeated with correct extension primer. Following troubleshooting haplogroup assignment was made possible through analysing the electropherograms (Figure 4.2 and 4.3).



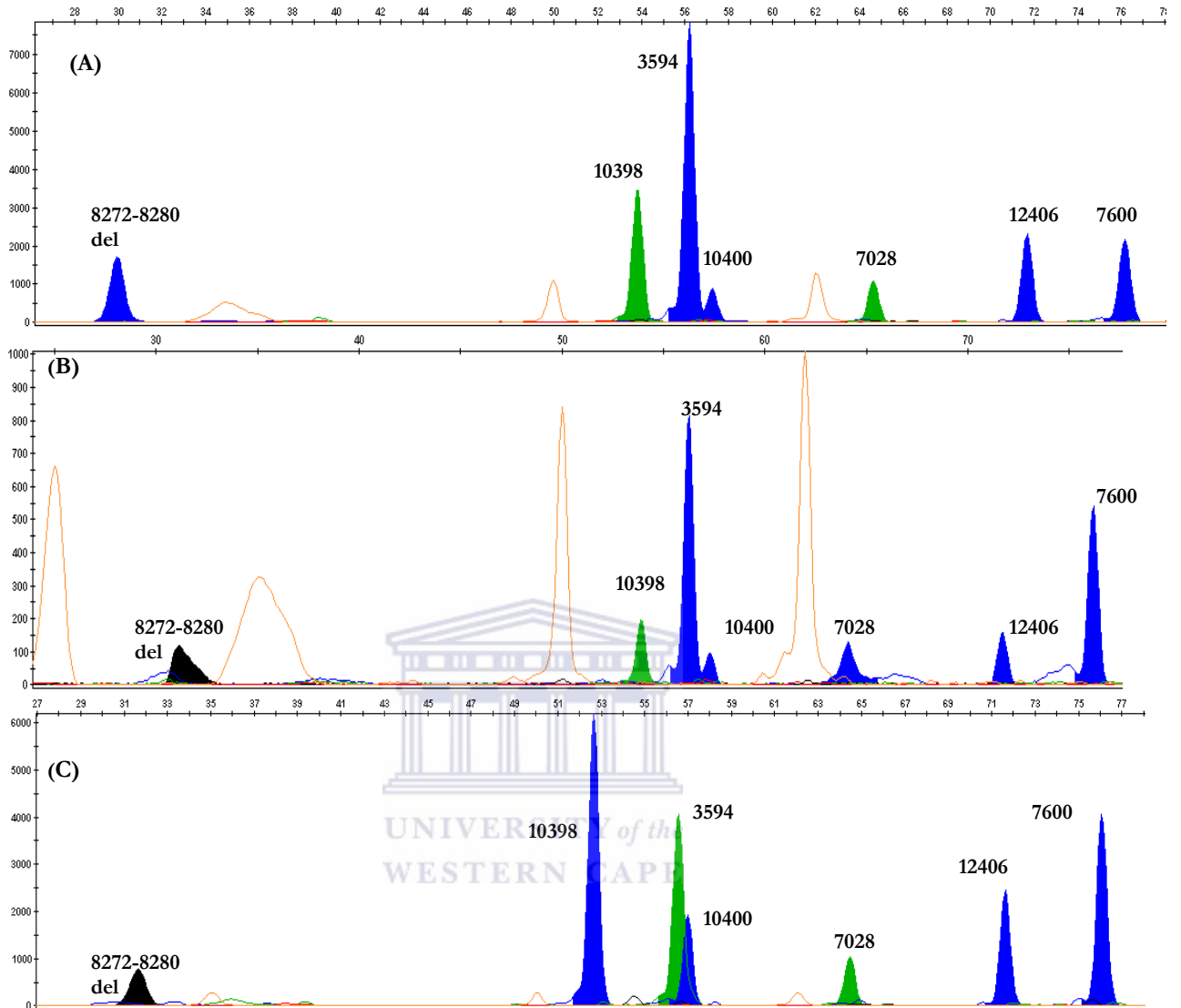


Figure 4.2: Electropherograms of M1: Plots of fragment size (x-axis) relative to -120 Liz size standard (*Applied Biosystems*) and relative fluorescent units (RFUs, y-axis). Seven-marker multiplex analyzed with *Gene mapper 4.0* software. Derived alleles are indicated with the respective colour for each peak where green represents nucleotide A, black: nucleotide C, blue: nucleotide G and red: nucleotide T. **A)** Sample indicating result typical for haplogroup B assignment, **B)** sample indicating result typical for haplogroup H assignment and **C)** sample indicating result typical for haplogroup L1\L2 assignment.

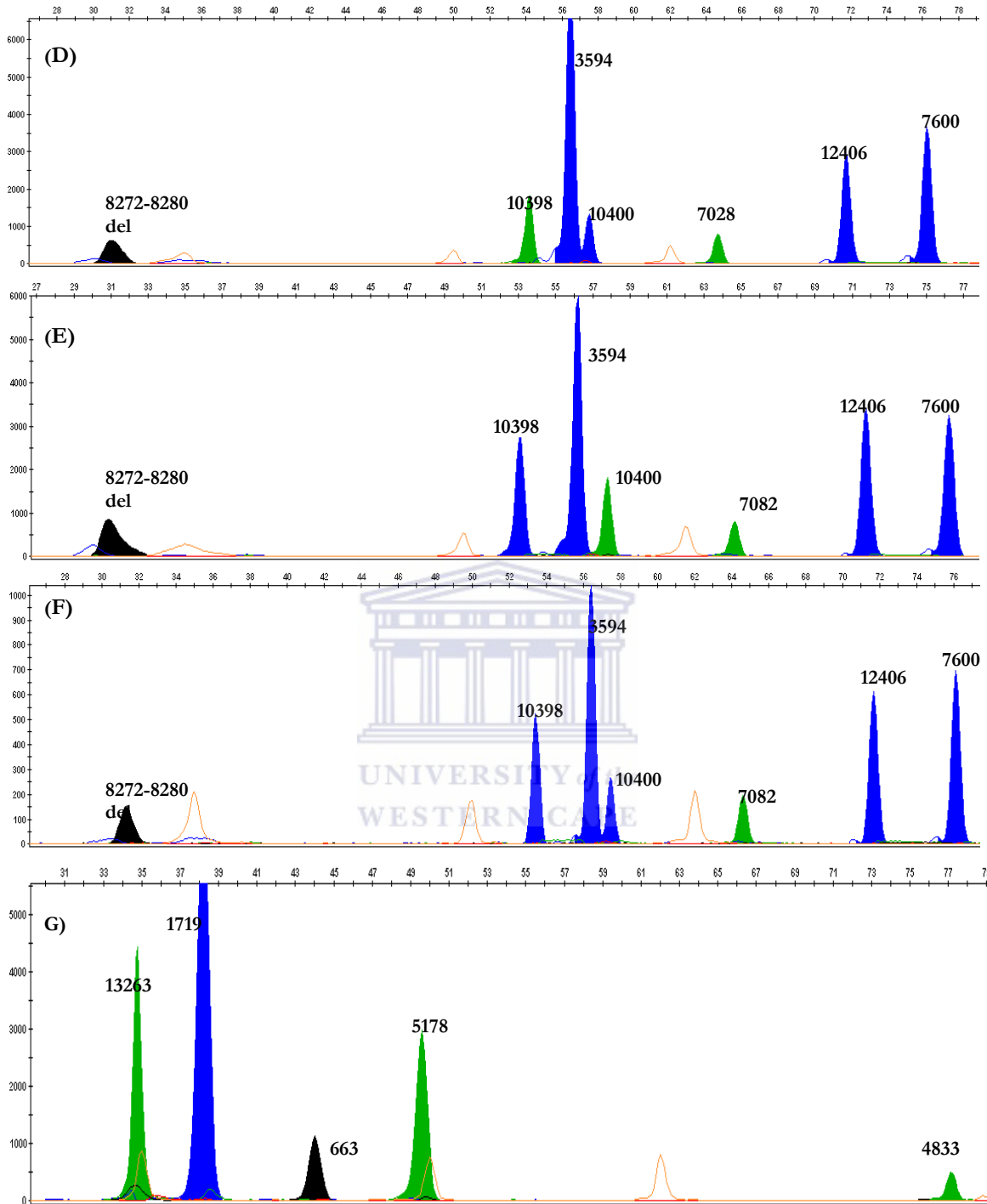
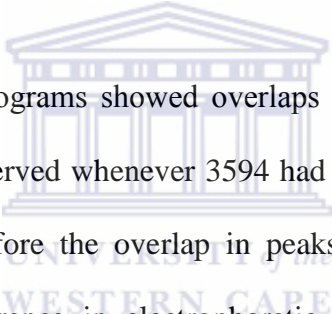


Figure 4.3: Electropherograms of M1 and M2: Plots of fragment size (x-axis) relative to -1201iz size standard (*Applied Biosystems*) and relative fluorescent units (RFUs, y-axis). Seven-marker multiplex analyzed with *Gene mapper 4.0* software. Derived alleles are indicated with the respective colour for each peak where green represents nucleotide A, black: nucleotide C, blue: nucleotide G and red: nucleotide T. **D)** Sample indicating result typical for haplogroup N assignment, **E)** sample indicating result typical for haplogroup M assignment and **F)** sample indicating result typical for haplogroup L3 assignment. **G)** Electropherogram associated with sample being typed in M2

The resulting electropherograms indicated above in figure 4.2 A, B and C indicated samples that could be assigned a haplogroup with only typing in M1. The electropherograms in figure 4.3 D, E and F indicated samples that were retyped in M2 to confirm haplogroup assigned. The electropherogram results for figure 4.3 was straightforward within the multiplex as extension primers were well spaced. However it was found that when a sample was typed for the ancestral state at the transition position 3594 and 10400 the blue peaks did not migrate far from one another and close peaks were observed for these markers as shown above in A, B, D and F.



L1/L2 individuals' electropherograms showed overlaps in peaks at transition positions 3594 and 10400. This was observed whenever 3594 had the derived state indicated by a green peak. As mentioned before the overlap in peaks or close peak migration was considered the result of difference in electrophoretic mobility of extension primers mainly attributed to length, dye and the short oligonucleotide sequence. All electropherograms frequently gave good consistent results with no extraneous peaks; weak signal strength for 8272-8280 was also improved in all cases except in assignment of haplogroup H. Future work would require redesign of primer for transition positions as well as 3594 and 10400 so that better peak RFU and peak separation can be obtained.

4.4 Mitochondrial Haplogroups Observed

Subsequent analysis of the genotyping results allowed individuals to be placed within particular haplogroups. The two multiplex panels together characterize twelve mtDNA

coding region polymorphisms which defined 14 haplogroups, however only 7 haplogroups was observed in the population. It was found that M1 discriminated well among nearly all samples resulting in six haplogroups whereas M2 only characterised one individual to haplogroup D.

4.4.1 Mitochondrial Haplogroup Frequencies

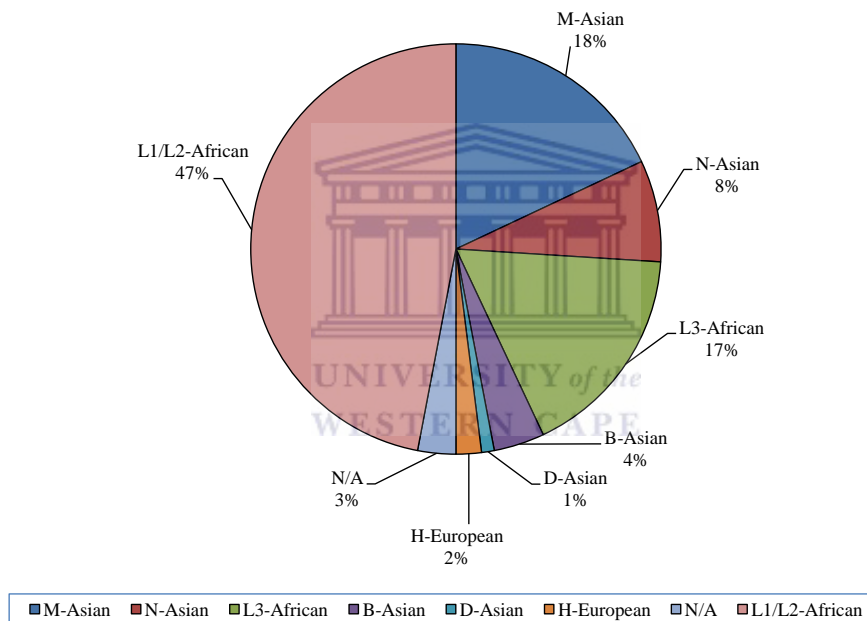


Figure 4.4: A pie chart representation of the percentage of maternal haplogroup frequencies and geographic origin observed in a 115 unrelated males from the Strand Muslim Community.

The maternal lineages of the Strand Muslim community were successfully determined for 97% of the samples. The major MtDNA haplogroups observed within the population were L1/L2, M, N and L3, and subgroups of M and N, B, H, and D.

Reviewing the single haplogroups revealed that haplogroup L1/L2 was well represented in the population occurring at a 47% frequency. This haplogroup had the highest

frequency of all mtDNA lineages observed. According to published data L1/L2 likely arose in East Africa. It has the greatest frequency within central Africa and Khoisan of Southern African populations (Wells, 2007). These two lineages also constitute the most ancient lineages of the mtDNA phylogenetic tree (Maca-Meyer *et al.*, 2001; Hurles *et al.*, 2005; Underhill and Kivisild, 2007).

Haplogroup M and L3 were found at 18 and 17% frequency in the population respectively. Haplogroup M, which is a derived macrohaplogroup that split from L3. It is found throughout Asia and constitutes 30-50% of East Eurasian lineages occurring in Pakistani, Indian and South Asian populations (Friedlaender *et al.*, 2007; Malhi *et al.*, 2007). Haplogroup L3 was observed at a 17% frequency. This haplogroup represents the first individual to leave Africa and is found in high frequencies in populations across North Africa. Haplogroup N appeared at an intermediary frequency of 8%. This haplogroup, which as haplogroup M split from L3 is found across western Eurasia with maximum frequencies in western Asia and central Asia.

Haplogroups B, H and D were found to constitute the lowest frequencies of all mtDNA haplogroups. Haplogroup B accounted for 4%, Haplogroup H for 2% and haplogroup D for 1%. Haplogroup B is a derivative of haplogroup N and haplogroup D is a derivative of haplogroup M. Both lineages are found across central Asia and East Asia where haplogroup B has been observed at a 48% frequency of mongoloid populations, 17% frequency of Southeast Asian and 20% of Chinese populations (Wells, 2007). Haplogroup D accounts for 20% of East Asian populations.

Haplogroup H appeared at a low frequency of 2% in the Strand Muslim population. This haplogroup is a dominant lineage found across Europe. It has been found at a 20-40% frequency in western and northern European populations.

As haplogroup assignment was based on shared constellations of polymorphisms in mtDNA coding regions that have been previously associated with named haplogroups 3% of the population could not be assigned to a particular haplogroup. This was believed to be the result of reverse mutations present with the mtDNA that obstructed the process of haplogroup assignment.

4.4.2 Strand Mitochondrial Haplogroup Dominance

The phylogeny of mtDNA variation present in modern humans can be crudely characterized and classified into major clades discriminating between African, Asian and European. The seven haplogroups observed and found at various frequencies within this population were clustered into categories African, Asian and European that are presented below.

Table 4.4: List of continent-specific mitochondrial DNA haplogroups observed within in the Strand Muslim community.

African	Asian	European
L1\L2	B	H
L3	D	
	M	
	N	

African lineages were comprised of L1/L2 and L3; Asian lineages consisted of haplogroups B, D, M and N. European lineages only consisted of haplogroup H. This European lineage was only observed within two individuals.

Based on haplogroups found, it was deduced that Asian haplogroups were more prevalent in this population occurring at low to intermediate frequencies (Figure 4.5).

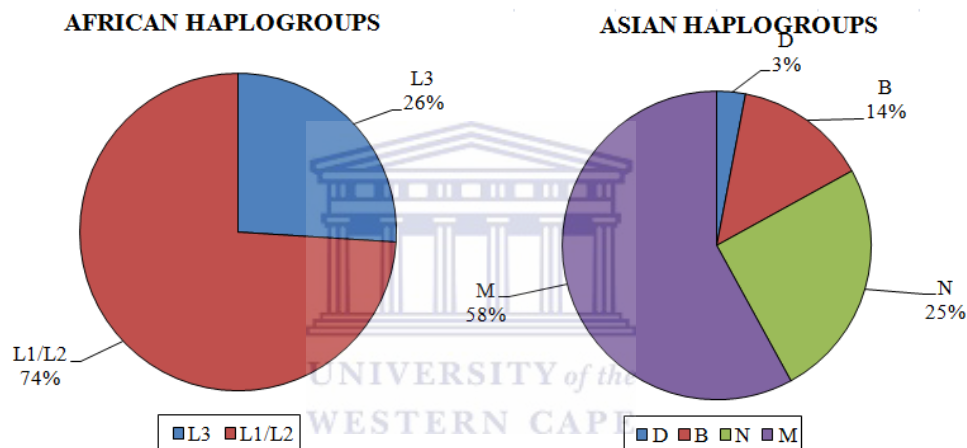


Figure 4.5: Representation of Mitochondrial haplogroup dominance observed within the Strand Muslim Community expressed in percentage format. **A)** African haplogroups observed indicating dominant African haplogroup found. **B)** Asian haplogroups observed indicating dominant Asian haplogroup found.

The African category was dominated by the lineage L1/L2 which is defined by transition sites at positions 10398G, 3594T and 7028T. This lineage was most frequent occurring in 74% of individuals. L3 defined by transition positions at 10398G and 7028T was found in 26% of individuals only. African haplogroup L1/L2 was thus the dominating African lineage. The Asian category showed that only 3% of the community was assigned to haplogroup D defined by transition positions at 5178A, 10398G, 10400T and 7028T. Haplogroup B defined by transition sites at position 8272-8280G del and 7028T were

found in 14% of individuals. Haplogroup N which is defined by transition sites at position 3594G, 10398A and 10400C was typed positive in 25% of individuals. 58% of individuals typed positive for transition positions at 10398G 10400T and 7028T which define haplogroup M making this the dominant Asian mitochondrial lineage within the Strand Muslim community.

Overall haplogroup results revealed that within the Strand population of 112 males, more Asian lineages appeared in the population. However African lineages dominated as more individuals was typed for African maternal ancestry.

4.5 Summary

The mtDNA study of the Muslim Strand community has revealed a high matrilineal contribution of Africans and Asians. MtDNA pool was diverse in composition but had dominant African ancestry. The presence of haplogroups L1/L2 and L3 identifies the mtDNA of African origin whereas Asian haplogroups M, N, D and B are likely due to gene flow from across Asia.

Chapter 5

Population survey and ancestry analysis of the Strand Muslim community

5.1 Introduction

The Strand Muslim community stems from Indonesian, Javanese and Asian origins that had strong links to Islam with some African influence Rhoda, (2006). The many Indonesian Malay Muslim customs still practiced in the area to date supports this hypothesis. The present genetic ancestry study clearly shows a significant contribution of the Asians in the genetic pool of the Strand Muslim community. However, there is a clear difference in the contribution the Asian, African and European haplogroups into the paternal and maternal lineages of this community. This chapter aims therefore to provide an analysis of the ancestry of this community using both the results of the genetic study and the population survey obtained through a detailed questionnaire completed by the donors.

5.2 Materials and Methods

All donors included in this study completed a questionnaire covering their paternal and maternal ancestry. Information obtained from these questionnaires included language, religion, birthplace and perceived ethnicity group. The questionnaire was in the form of a unique bar-coded questionnaire that was also able to exclude patrilineal relatives and individuals with origins outside of the Strand. A copy of this questionnaire is included in the appendix section.

5.3 Results and Discussion

5.3.1 Population Survey

The information collected was as detailed as possible. It included geographical information of paternal and maternal lines' birthplace, ethnicity and religion. This permitted the construction of a database, which now forms part of the Forensic Laboratory at University of the Western Cape (UWC) main Populations database.

The 115 Muslim males surveyed ranged from 18 to 84 years of age. These age groups made it possible to gain insight into the community's ancestral history over generations. The home language spoken by many donors was English or Afrikaans and in many cases, donors considered themselves bilingual. Creolized Afrikaans was observed in many donors. Three donors reported the home language of their paternal and maternal lines as Urdu and two donors reported their paternal grandfather's home language as Chinese.

As the study was aimed at the Muslim community, all donors were of Islamic faith. Religion within the sample population was of great interest as to determine if Islam was the most prominent religion within the donor's paternal and maternal ancestry and ultimately if Islam shaped this community. The establishment of Islam in Strand during the year 1822 was undertaken with the purpose of creating a core Muslim community Rhoda, (2006). The Javanese Imam Abdol Sammat and his many free black followers were responsible for the strong influence and presence of Islam in the community. Rhoda, (2006) inferred this information stating that the establishment of Islam within Strand was the direct and deliberate desire of Abdol Sammat to propel Islam as to oppose the rapid

expansion of Christian missionaries in the Strand and the greater Cape Town area. The presence of Islam in this community as a core religion is seen in (Figure 5.1.).

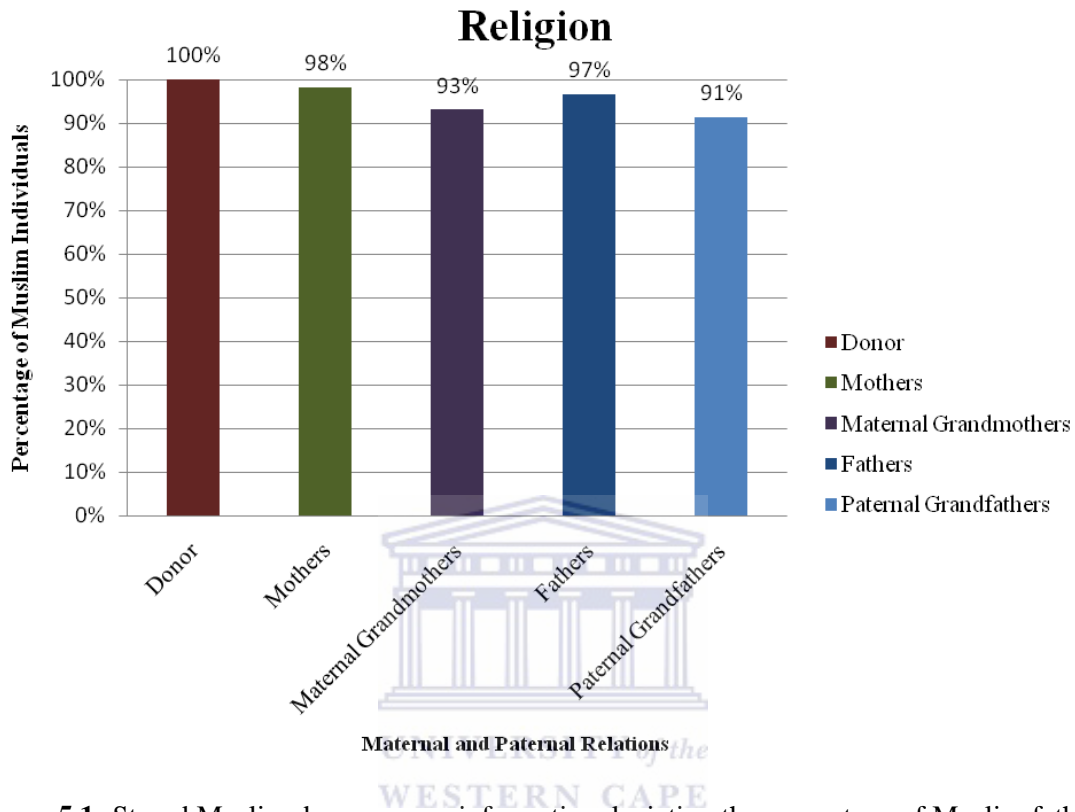


Figure 5.1: Strand Muslim donor survey information depicting the percentage of Muslim fathers, mothers, paternal grandfathers and maternal grandmothers of Strand Muslim donors.

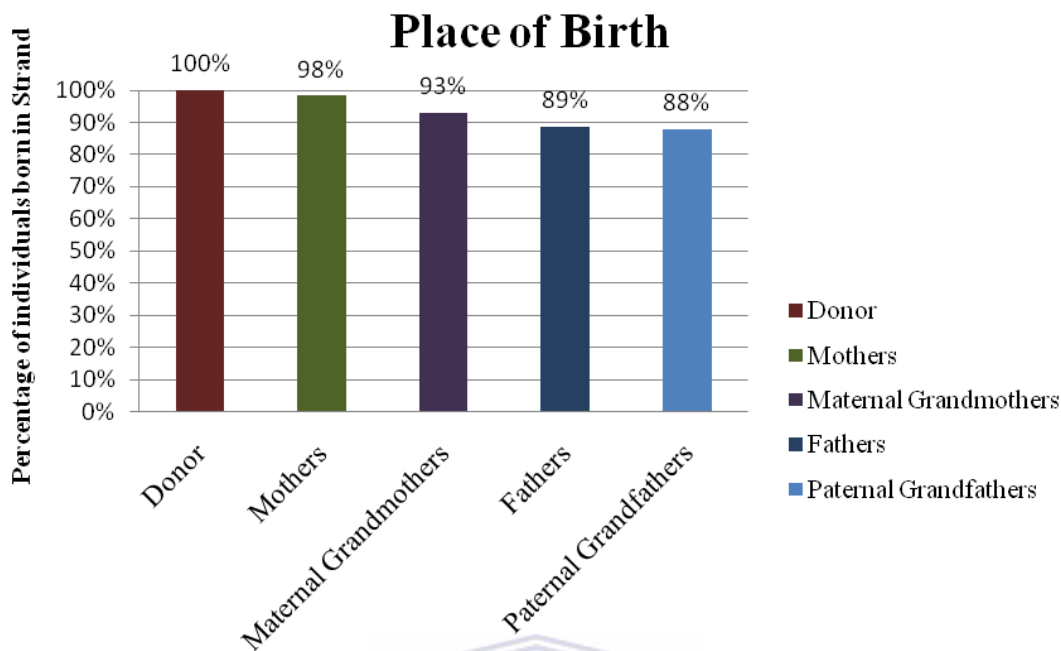
The graph above shows that Islam was a prominent religion within this community, and had been so for some time. The survey data revealed a high number of donors’ parents and grandparents were of Islamic faith. This indicates that a large proportion of donors are third generation Muslims and that Islam may have even dated back to further generations. The present day Imam in Strand is the great grandson of a Javanese Imam that settled in Strand Rhoda, (2006). This indicates that Islam persisted throughout the years within this community. It was also found that four donors had Christian fathers and Christian paternal grandfathers and two donors shared Christian mothers and Christian maternal grandmothers; an indication that conversion within the community was present

and in some cases did occur over the generations. More mothers and fathers were found to be of the Islamic faith in comparison to grandfathers and grandmothers.

Interfaith ancestry information surveyed was found in 18 out of 115 (16%) donors, this implies that conversion to Islam for these donors was the result of interfaith marriages or voluntary conversion.

The survey results proved that this community has a strong Islamic background. This may be attributed to the community's strong Islamic leadership, which can be dated back to the year 1822 suggesting this community as having a long preserved Islamic identity Rhoda, (2006). Many of the free blacks that settled in the Strand area were of Indonesian origin and brought with them many of the Islamic socio-religious practices that can still be found in this community to date. This can also be seen in the many Islamic Masjids erected in the area during the 18th century that still stands today.

Fundamental information that was covered by the survey was the donors' place of birth. All donors had to be born within the area of Strand. Donors were asked to supply as detailed as possible geographical information of their paternal and maternal lines' birth place (Figure 5.2).



Maternal and Paternal Relations

Figure 5.2: Strand Muslim donor survey depicting the percentage of fathers, mothers, paternal grandfathers and maternal grandmothers born in the Strand.

A large proportion of donors had paternal and maternal ancestry originating in Strand. Based on the data the present day Strand population can be traced back three generations to have been born within Strand. This implies that these individuals' grandparents and great grandparents might have been part of the first inhabitants to settle within Strand. Rhoda, (2006) also confirmed this, stating that present day Muslim families who are seventh generation descendant of early inhabitants could still be found in Strand to date.

The question on ethnicity was based on self- identification with respondents stating what their ethnicity was. Ethnicity data, which, is often seen as affiliation to nationality or birthplace, reported three major ethnic classification groups (figure 5.3.). This was Coloured, Malay and Indian.

Donor Ethnic Classification

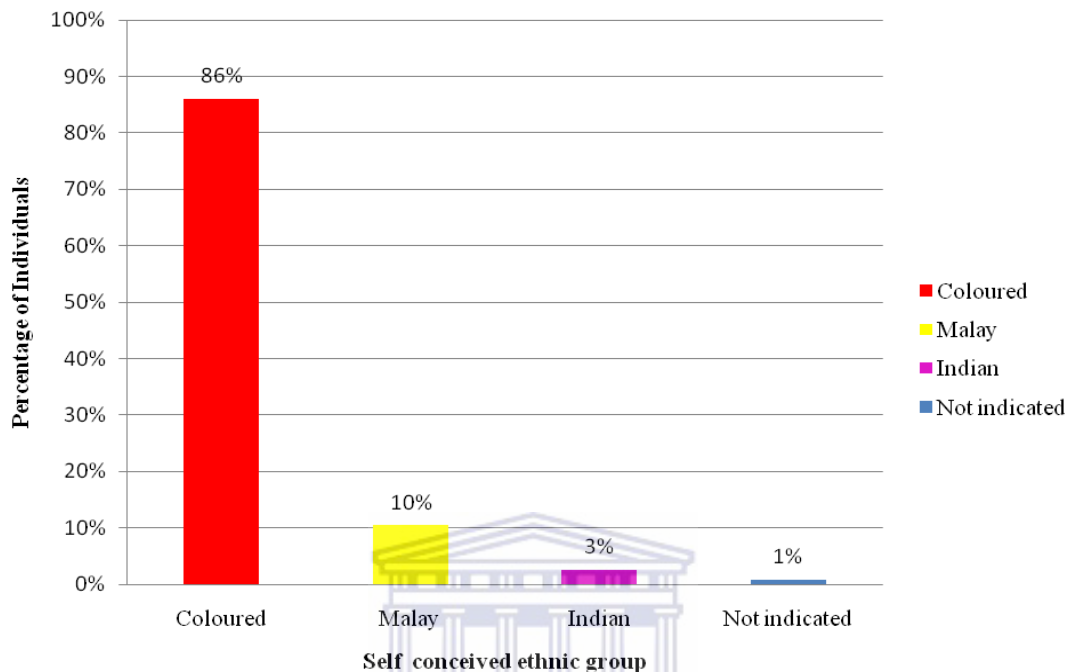


Figure 5.3: Strand Muslim donor survey information depicting the percentage of donors who defined their ethnic origin as Coloured, Malay, and Indian and not indicated.

Eighty six percent of donors reported their ethnicity as Coloured. Coloured is political term used to define an individual of European and African or Asian origin (Da Costa, 1994; Bloom, 1960). Coloured was also the largest population group reported in 1842 during a Hottentots Holland Opgraavol (census) in the area. This community was one of the many communities in the Cape that was destroyed by apartheid's Slum Clearance ACT.

The second largest group reported in the survey was Malay in which only 10% of donors were placed. The low number of Malay defined individuals was surprising as historically this community makes preference to a strong Malay or Indonesian background. The third

group was Indian and this was the smallest group with only three individuals. One donor chose not indicate any ethnic group. This pattern of categorization was also observed in the paternal and maternal lines (Figure 5.4).

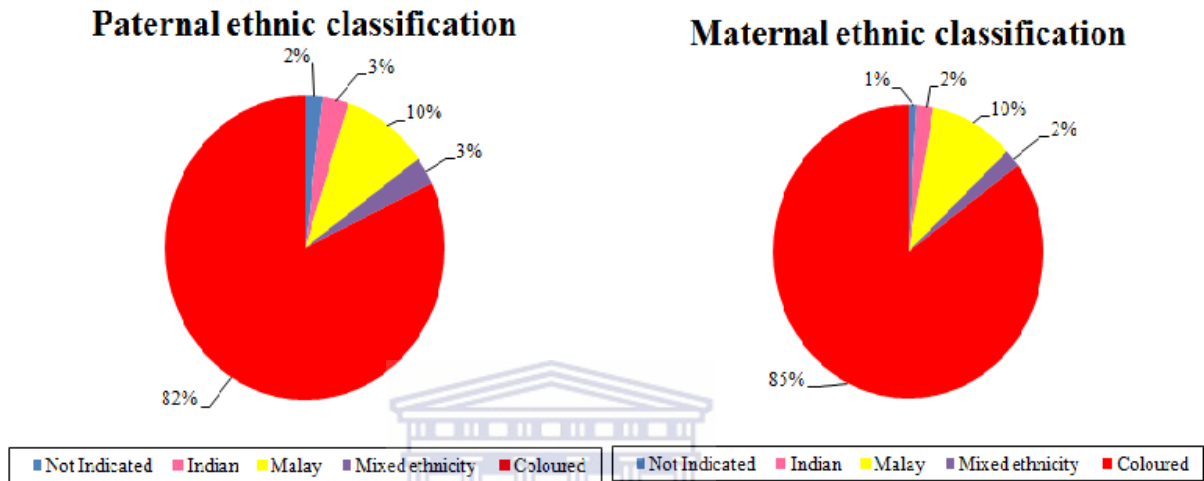
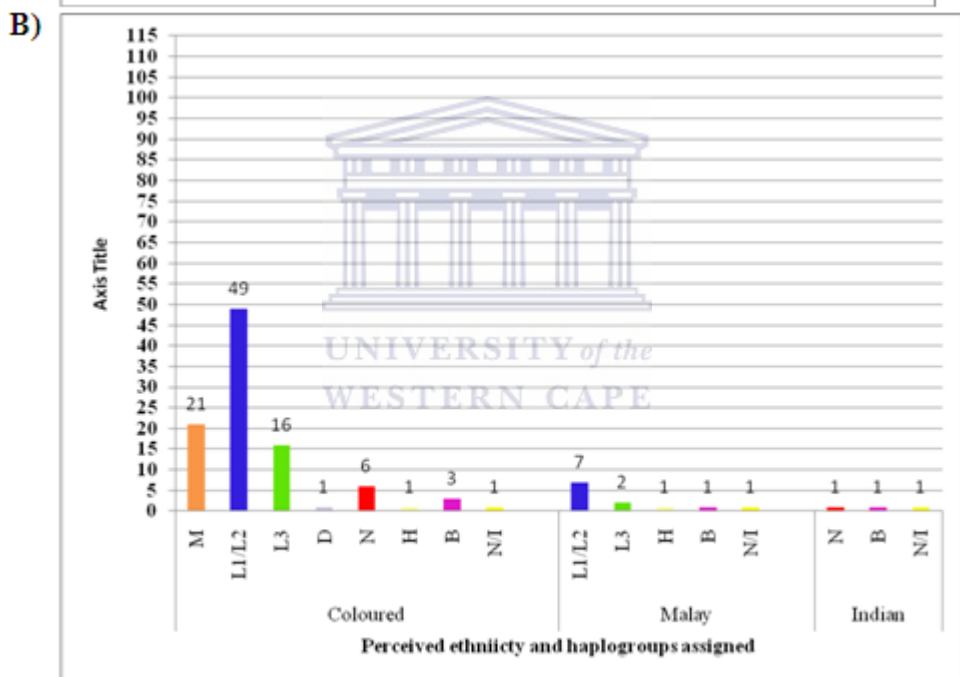
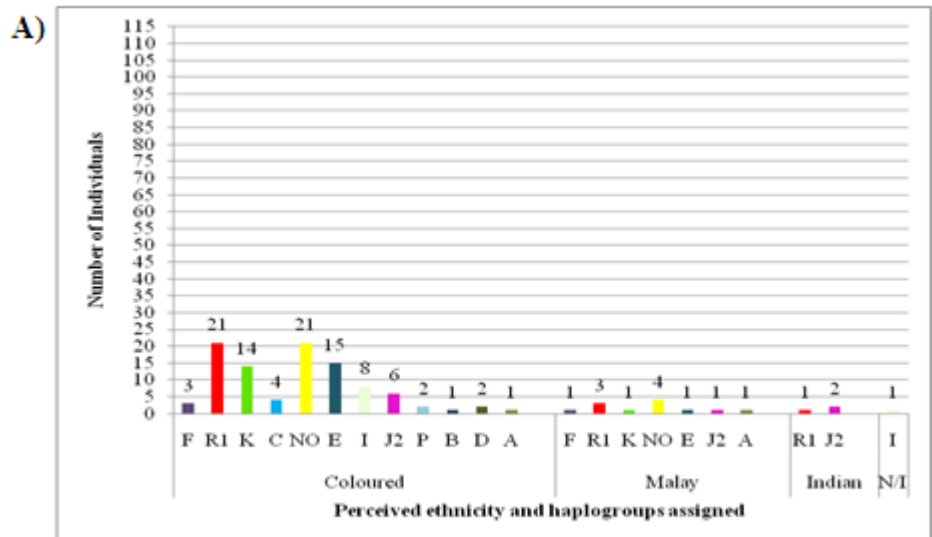


Figure 5.4: Donor survey data obtained from Strand Muslim male donors indicating the perceived ethnic origin of paternal and maternal lines.

A large number of respondents classified their maternal and paternal lines as coloured. A fourth group termed mixed ethnicity was also used. This group consisted of donors who reported two or three different groups in their ancestral line. .



Figures 5.5: **A):** Donor survey data obtained from males of the Strand Muslim Community depicting perceived ethnicity and true genetic ethnicity through Y-haplogroups assignment. **B)** Donor survey data obtained from males of the Strand Muslim Community depicting perceived ethnicity and true genetic ethnicity through mtDNA haplogroup assignment.

As depicted above in (figures 5.5a and b), the coloured ethnic group contained the greatest haplogroup diversity in both the Y chromosomal pool and mitochondrial pool. Donors

who reported this group as their ethnicity represent the diversity of this politically coined term.

5.4 The Strand Muslims genealogical history

5.4.1 The Strand Muslim community's Y-chromosomal ancestral history

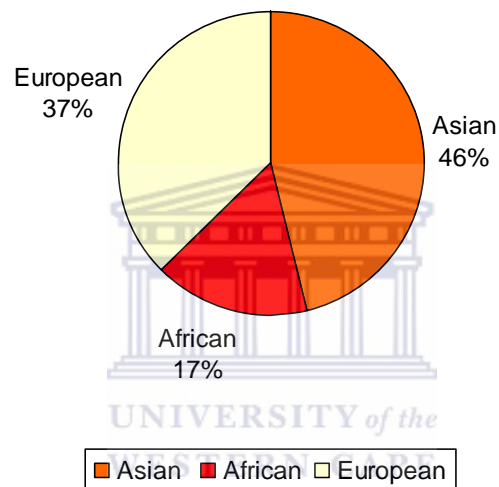


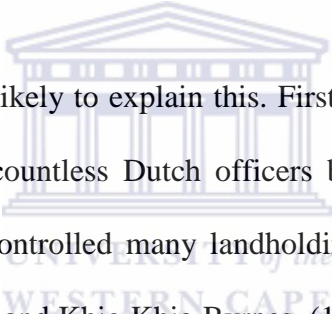
Figure 5.6: Overall percentage of African, Asian and European Y-chromosomal haplogroups obtained in the sample population of the Strand Muslim community.

The total sample results depicted in figure 5.6 reveal as much as 17% African, 37% European and an astounding 46% Asian contribution to the total Y chromosomal pool of the Strand Muslim community.

5.4.2 The African fraction of the Strand Y-chromosomal pool

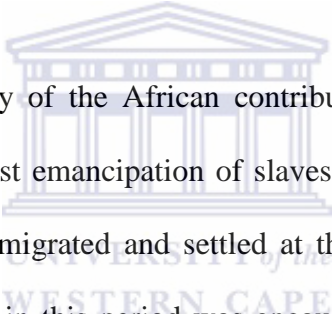
The Y chromosomal ancestral study revealed that 17% (19 out of 115) of the Strand sample population had African ancestry. Haplogroups A-M91, B-M181 and E-M96 were the major African haplogroups detected.

Outside of Africa, carriers of Y-chromosome haplogroup A-M91 and B-M181 are rare. Among the Strand samples, two individuals were typed positive for haplogroup A-M91 and a single haplogroup B-M181 individual was detected. The fact that these indigenous African haplogroups appeared at a low frequency is viewed as surprising given the populations geographic vicinity. However, a historical census described in (Rhoda, 2006) showed that in the year 1825 no indigenous males of African origin were found in Strand. The low incidence of haplogroup A-M91 and B-M181 thus suggests that indigenous males did not form part of the initial inhabitants of this community.



Three processes in history are likely to explain this. Firstly, the increase in settlement at the Cape was encouraged as countless Dutch officers become landowners called *free burghers*. The *free burghers* controlled many landholdings that resulted in a series of wars between European settlers and Khoi Khoi Byrnes, (1996). Countless Khoi Khoi thus lost their lives to the Dutch's superior weapons. Secondly, as *free burghers* moved further into the land the Khoi Khoi population began to disintegrate Byrnes, (1996). Some Khoi Khoi became labourers on European farms following their religion and many moved out of the Cape seeking better opportunities. The third event was the small pox epidemic of 1713, which eradicated 90% of the Khoi Khoi population Byrnes, (1996). Thus, it is plausible that as the Strand was a growing Muslim population being composed of mostly Asian free blacks of Islamic faith not many indigenous male individuals joined this enclave.

The third African haplogroup E-M96 appeared in 16% of individuals. This haplogroup occurs in high frequency in North African and East African populations. The first slaves enslaved by the D.E.I.C had ancestry in these regions and across Madagascar and Mozambique Shell, (2005). This suggests that individuals typed positive for this marker are descendent from male migrants or free blacks who joined the enclave at Strand in later years. The area of Strand was an ideal place for countless free blacks to settle. The ocean was a source of free food and the Crown lands on which the community was established provided land and homes for many free blacks who could not afford to buy land Rhoda, (2006).



This indicates that the majority of the African contribution to the Strand Muslim Y-chromosomal pool occurred post emancipation of slaves. During this time, an influx of free blacks of various origins migrated and settled at the then Mosterd bay. It is also noted that conversion to Islam in this period was encouraged and prevalent within this historically rich Islamic community.

5.4.3. The European fraction of the Strand Y-chromosomal pool

European ancestry was found in 37% (43 out of 115) of the sample population. The European fraction of the Y-chromosomal pool revealed the presence of European haplogroups I-M170, J2-M172 and R1-M173. Haplogroups I-M170 and J2-M172 are dominant lineages found in European populations. In the Strand sample population it was moderately distributed in nine individuals. Haplogroup I-M170 is a common lineage found in Western Europe and is most dominant in French populations. Haplogroup J2-

M172 is found in both the Middle East and Southern Europe, where it is most commonly found in Italian and Spanish populations. The increased frequency of haplogroup R1-M173 that is considered a Eurasian marker occurred in 25 individuals. This haplogroup is commonly found in Western European and Middle Eastern populations.

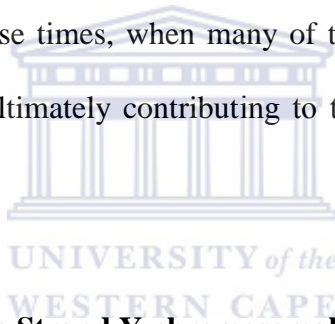
Individuals typed positive for markers I-M170, R1-M173 and J2-M172 can thus trace their ancestry to immigrants who were of European origin coming from countries such as Spain, Italy, France, England and Germany (Byrnes, 1996). European presence in Strand was officially documented in a census cited in (Rhoda, 2006). The census of the year 1842 however only enumerated five males of European descent at that particular time.

However, examining the historical timeline of Cape Town and Strand, it is seen that prior to the arrival of Abdol Sammat European influence was already present, supporting the notion that European males might have been present in that area before the community was established. It is plausible that many of the first European men were *free burghers* who were given land by the V.O.C. or Trek Boers farmers of primarily Dutch, German and French ancestry (Byrnes, 1996). Evidence of this is seen in the following

- In the year 1690, the Dutch East Indian Company sends Sheikh Yusuf to an area close by Mosterd Bay.
- During the period of 1700-1706, the son of a well-known Dutch officer Willem Adriaan had fishing rights between Gordon bay and Lourens River.

- In the year 1822, the land on which the community was illegally founded on was Crown Lands.
- In the year 1839, the advent of missionaries sent to Mosterd Bay concludes that Europeans were moving in and out around this area.
- In the census of 1842, it was found that 10% of the population residing in Strand was of European descent.

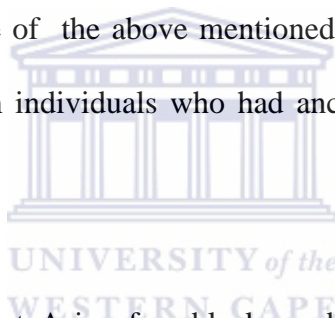
The presence of these European haplogroups in the present day Muslim of Strand today confirms the history of Cape Town and Strand and characterizes the colonial based systems established during those times, when many of the European men had relations with free black women thus ultimately contributing to this community's Y pool Shell, (2005).



5.4.4 The Asian fraction of the Strand Y-chromosomal pool

The genetic analysis results showed that 46% (53 of 115) of the Strand paternal pool had Asian Ancestry. Asian haplogroups observed were C-M216, D-M174, NO-M214, F-M89, K-M9 and P-M45. Haplogroups C-M216, D-M174, P-M45 and F-M89 appeared in low frequency. Haplogroup C-M216 and P-M45 were typed positive in four and two individuals respectively. Haplogroup D-M174 was typed positive in two Strand individuals. Haplogroup F-M89 was typed positive in four individuals and haplogroup K-M9 that was found at a moderate frequency occurred in fifteen individuals. Haplogroup NO-M214 occurred with the highest frequency, accounting for 26% of the Asian contribution. The haplogroup that combines haplogroups N-M214 and O-M175 is widely

distributed in present day Asian populations. (Karfet *et al.*, 2005) describes samples belonging to this haplogroup being positive for O-M175. In the study conducted by (Karefet *et al.*, 2005) and colleagues, South East Asian populations including Balinese, Eastern Indonesian, Western Indonesian, Malaysian, Taiwanese, Philippines and Vietnamese samples were found to be polymorphic for haplogroup C-M216, F-M89, K-M9 and N with predominate presence of Haplogroup O-M175 subgroup haplogroups O-M95 and O-M165. This was also found in Southern Chinese and South Asian populations (Thangaraj *et al.* 2002; Tajima *et al.*, 2002; Zerjal *et al.*, 2002; Montiel *et al.*, 2005; Xue *et al.*, 2006; Krithika *et al.* 2007). As the presence and frequency of these y-chromosomal haplogroups is similar to those of the above mentioned populations it suggest a strong male migration to Strand from individuals who had ancestry from these diverse Asian population.



Rhoda (2006) hypothesized that Asian free blacks and free blacks of various origins founded the Strand Muslim community. In 1822 five free black slaves of Indonesian origins settled in Strand originally from Semarang Java accompanied a Javanese Imam. These free blacks formed the core group of the initial inhabitants that resulted in the first cohesive Muslim community in the Cape. Evidence of his theory was found in two documents cited in Rhoda (2006). These are presented in the Appendices of this thesis (Appendices). Appendix IIb shows that prior to the arrival of the Imam in the year 1822, he resided in the greater Cape Town area with many other free blacks of Indonesian origin who later settled with him at Strand.

Appendix IIa indicated that between 1822 and 1840 there was an influx of free blacks of Indonesian origin settling at Strand. This is also confirmed in numerous censuses that are described by Rhoda (2006). He stipulated that in the first census of 1825, 14 out of 19 free blacks were of Indonesian origin. Considering this historical evidence and examining the genetic ancestry, the male population of Strand does indeed have predominant South Eastern Asian origins.

5.5. The Strand Muslim community's maternal ancestral history

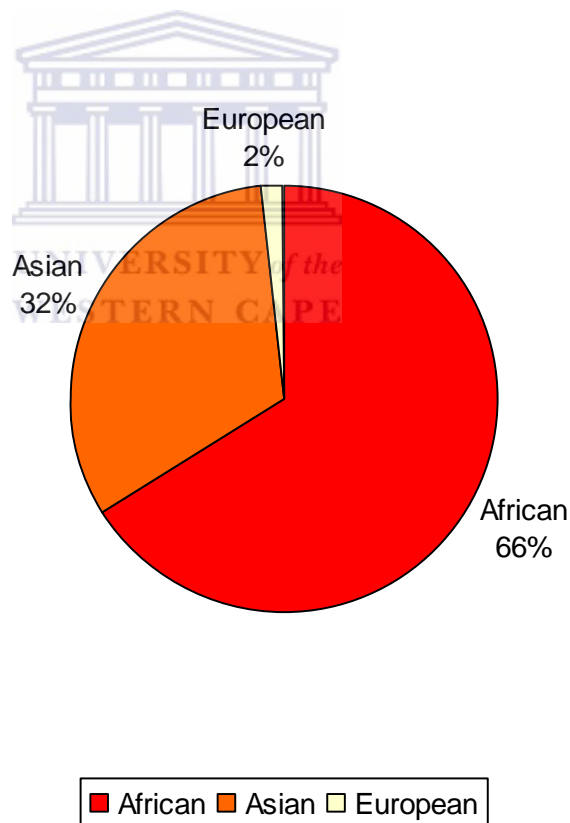


Figure 5.7: Overall Asian, African and European contribution to Strand mtDNA-pool haplogroups obtained in the sample population of the Strand Muslim community.

Major continent-specific mitochondrial haplogroups were successfully assigned to 112 samples. A total of seven mtDNA lineages were found in the population, and they included major African, Asian and European haplogroups. The Strand mitochondrial pool consisted of 66% African ancestry, 32% Asian ancestry and 2% European ancestry.

5.5.1. The African mitochondrial contribution

Genetic data analysis showed that 66% (74 out of 112) of individuals of African origins. Two major African haplogroups were detected in the population; haplogroup L1/L2 and haplogroup L3. Haplogroup L1/L2 was assigned to 56 individuals and Haplogroup L3 was found in 19 individuals. Haplogroup L1/L2 (as described in chapter one and four of this thesis) is considered the most ancient of all mitochondrial lineages. It is commonly found at high frequency in Southern African and Sub-Saharan African populations Brandstätter *et al* (2004). Haplogroup L3 that reflects similar distribution as L1/L2 in the world populations, but predominantly found in East and West African populations.

The large proportion of African mtDNA ancestry in the Strand Muslim population suggests that initial female presence within this population was in fact linked to predominantly indigenous African females with some western African females. Historically this is supported in a census of the year 1825 Rhoda (2006). The free black settlement at then Mosterd bay was found to consist of 41 individuals, of which eight were female. These women were also listed as being indigenous Hottentots. A small proportion of Khoi Khoi presence was also observed later in the year 1832. This suggests that the original mtDNA must have persisted throughout the generations. The increased

frequency of L1/L2 lineage could be the result of the influx of free black males either that settled at Strand with their female companions or who seek female companions to marry which stimulated the spread of Islam in the area.

5.5.2 The Asian mitochondrial contribution

Genetic analysis revealed that Asian ancestry encompassed 36 of the 112 (32%) of the mtDNA pool of present day Strand Muslims. The Asian haplogroups detected were M, N, B and D. These haplogroups were found at various frequencies within the sample population. Haplogroup M was detected in 29 individuals constituting the highest result for the Asian ancestry. This super haplogroup is found across Asia namely Southeast Asian populations at varying frequencies. Individuals carrying the definitive substitution can thus trace their maternal ancestry to population such as Indian, Pakistani, Malaysian and Vietnamese to name just a few Schurr and Wallace (2002).

Haplogroup N was positively typed in nine mtDNA samples. Haplogroup B, a subgroup of haplogroup N, was found in the population at a low frequency (five individuals). This haplogroup is found in both East Asia and Southeast Asia and is believed to have an Indonesian source (Schurr and Wallace, 2002; Alvarez-Iglesias *et al.*, 2007).

The final mtDNA Asian haplogroup detected was haplogroup D that was only observed in one individual. This haplogroup is common in East Asia, Thailand and Siberia but is rare among Malay and Vietnamese populations. The Asian contribution of present day Strand mtDNA can thus be linked to populations across Asia. The historical data cited in

(Rhoda, 2006) and presented as appendix IIa in this thesis shows that during the periods of 1823, 1826 and 1827 many Asian females companions were listed to have settled with their husbands in Strand.

5.5.3. The European mitochondrial contribution

European ancestry within the Strand mitochondrial pool was only observed in two individuals. The haplogroup observed is haplogroup H which is a dominate mtDNA European lineage found in western European populations. This result was somewhat expected as it is documented that in year 1842 four white females were enumerated in a census at Strand. Overall only 10% of the then population was listed as European. From this it is conclusive that mtDNA of European origin was not prevalent in the area. The low incidence of Europeans lineages in the maternal pool was plausibly affiliated to more relations occurring between European men and free black women than between free black men and European women.

5.6 Combined Y-chromosomal and mitochondrial DNA ancestry

As demonstrated above in figures 5.6 and 5.7 the Muslim community of Strand reflects the heterogeneous nature of South African populations. The genetic analysis confirms that this community has a strong Asian male line and surprisingly a strong African maternal line. This does not however mean that the community was shaped by only these two lineages as the presence of African, Asian and European lineages was found in both the Y- chromosomal and mitochondrial results.

Examining donor's paternal and maternal ancestry results revealed some interesting facts. The diversity of Y- chromosomal haplogroups found that suggests that there was a higher rate of male migration to the community than female migration. This is also indicated by the fact that indigenous African mitochondrial haplogroups dominated the maternal ancestry. Thus, fewer mtDNA haplogroups were introduced into the community. Rhoda (2006) demonstrated this in his data when he stipulated that in 1825 the ratio of men to women was 3:1. This resulted in many cross-cultural relationships being formed as many men either had relations with African women or sought partners elsewhere.

The moderate frequency of Asian mitochondrial lineages might have thus been encouraged by the ratio of men to women, as many of the Asian male free blacks would have likely sought Asian female partners who shared similar belief systems. This system of relationships proved effective as is documented by the year 1842, when the presence of 47 coloured females and 43 coloured men was enumerated thus increasing the ratio of women to men. It can thus be concluded that many of the men with European, Asian and African ancestry had relationship with African women (Table 5.1).

Table 5.1: table depicting the number cross cultural relationship between males and females observed within in the Strand Muslim community.

	African maternal ancestry	Asian maternal ancestry	European maternal ancestry
African paternal ancestry	14	4	1
Asian paternal ancestry	31	19	1
European paternal ancestry	27	15	0

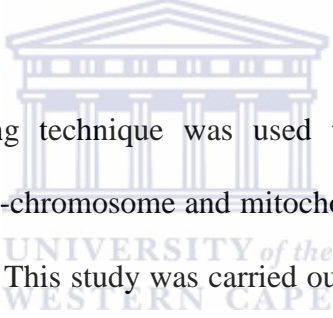
Demonstrated above it is seen that Majority of individuals who had African, Asian or European paternal ancestry also had African maternal ancestry. Nineteen individuals with Asian paternal ancestry also shared Asian maternal ancestry and no individuals shared European paternal ancestry and European maternal ancestry.



Chapter 6

Conclusion

A pilot study was carried out in order to reconstruct the major paternal and maternal lineages of the Muslim population living in the Cape metropolitan area. The Study has shown the ability of molecular genetic tools to give insight into the origins and history of local communities. The study was also used as a point of reference for the Strand Muslim Community project. Genetic variations of the Y-chromosome and mitochondrial DNA for the pilot study were analyzed using the RFLP technique.



The SNaPshot mini-sequencing technique was used to genotype single nucleotide polymorphisms (SNP) on the Y-chromosome and mitochondrial DNA in 115 males from the Strand Muslim community. This study was carried out to give additional evidence of the origins and the development of the Strand Muslim community using molecular genetics techniques. Over all, the mini-sequencing method provided a time and cost effective approach to analyze populations' ancestries to that of the RFLP technique.

Y-chromosomal haplogroups were successfully assigned to all 115 individuals. The Y-chromosomal pool of the Strand Muslims was found to exhibit a greater frequency of Asian lineages and a low to moderate frequencies of African and European lineages and. The dominant Asian haplogroup observed was NO-M214. The Asian haplogroups found reflects a Javanese, Indonesian and Malaysian origins.

Mitochondrial haplogroups were successfully assigned to all individuals, except for 2 samples where a back mutation occurred. Matrilineal contribution revealed a high frequency of indigenous African and moderate frequency of Asian lineages. The dominant L1/L2 haplogroup was suggestive of an indigenous African origin.

Genetic data thus support the view that the Muslim community of Strand belongs to lineages found in Asia most likely Indonesia, Southeast Asia and central Asia. The data correlates with historical studies as many Malay practices and customs are still practised in the area to this day.

There is a strong evidence of the influence of Asian haplogroups on the male genetic pool of both Cape and Strand Muslim communities. For the female genetic pool, however, there is a greater influence of the African haplogroups in both cases. None the less, there is a significant contribution of the Asian female lineages in the female gene pool of both communities although with much lower influence than that of their male counterpart in the paternal lineages. These results suggest that core founding members of the communities were mainly male slaves from Asia who later settled with indigenous African female partners.

As a results of the Group Areas Act of 1950 (Act No. 41 of 1950), communities were split according to their racial groups and put into different residential areas of any given town or city. These areas have become real enclaves where populations had different origins and grown separately influenced by clear different emigrational events. The difference in genetic makeup of the Cape General Muslims and the Strand General

Muslims is a clear evidence of this segregation. There is a greater influence of Asian male haplogroups in the Cape Muslim population than in the Strand Muslim population. The influence of the African female haplogroups is twice higher in the Strand Muslim Community than in the Cape Muslim Community.



References

Alessandrini F, Pesaresi M, Onofri V *et al.* (2004). Multiplex PCR development of Y-chromosomal biallelic polymorphism for forensic applications. *International Congress Series*.1261: 363-365.

Al- Zahery N, Semino O, Benuzzi G, Magri, C *et al.* (2003). Y-chromosome and mtDNA polymorphisms in Iraq, a crossroad of the early human dispersal and of post-Neolithic migrations. *Molecular Phylogenetics and Evolution*.1-15.

Amorim A, Pereira L (2005). Pros and Cons in the use of SNPs in forensic kinship investigation: a comparative analysis with STRs. *Forensic Science international*. 150:17-21.

Anderson S, Bankier AT, Barrell BG *et al.* (1981). Sequence and organization of the human mitochondrial genome. *Nature* . 290: 457-465.

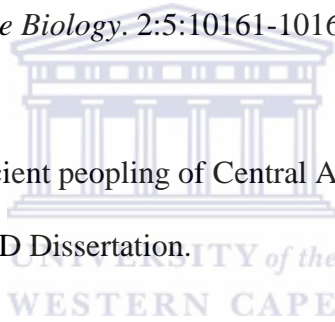
Andrews RM, Kubacka I, Chinnery PF *et al.* (1999).Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA, *Nature genetics*. 23: 147.

Araújo Silva W, Bortolini MC, Cruz Schneider MP *et al.* (2006). mtDNA Haplogroup analysis of Black Brazilian and Sub-Saharan Populations: Implications for the Atlantic Slave Trade. *Human Biology*. 78:1:29-41.

Arredi B, Poloni ES, Paracchini S *et al.* (2004). A predominantly Neolithic origin for Y-chromosomal DNA variation in North Africa. *Am. J. Hum. Genet.* 75:338-345.

Bachtrog D, Charlesworth B (2001). Towards a complete sequence of the human Y-chromosome. *Genome Biology*. 2:5:10161-10165.

Batini C (2008). The ancient peopling of Central Africa: a genetic perspective. *University of Roma*. PhD Dissertation.



Battaglia V, Fornarino S, Al-Zahery N *et al.* (2008). Y-chromosomal evidence of the cultural diffusion of agriculture in Southeast Europe. *European Journal of Human Genetics*. 1-11.

Benjeddou M, Leat N, Davidson S (2006). Use of molecular genetics and historical records to reconstruct the history of local communities. *African Journal of Biotechnology*. 5: 25: 2516-2519.

Bermisheva MA, Tambets K, Villems R *et al.* (2002). Diversity of Mitochondrial DNA haplogroups in Ethnic populations of the Volga-Ural Region. *Molecular Biology*. 36:6: 802-812.

Blanco P, Shlumukova M, Sargent CA *et al.* (2000). Divergent outcomes of intra-chromosomal recombination on the human Y-chromosome: male infertility and recurrent polymorphism. *J. Med. Genet.* 37: 752-758.

Bloom L (1960). The coloured People of South Africa. *The British Journal of Sociology*; 28: 2:139-150.

Bosch E, Calafell F, Comas D *et al.* (2001). High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between North-western Africa and Iberian Peninsula. *Am.J. Hum. Genet.* 68:1019-1029.

Bouakaze C, Keyser C, Amory S *et al.* (2007). First successful assay of Y-SNP typing by SNaPshot minisequencing on ancient DNA. *Int. J. Legal Med.*

Brandstätter A, Peterson CT, Irwin JA *et al.* (2004). Mitochondrial DNA control region sequences from Nairobi (Kenya): inferring phylogenetic parameters for the establishment of a forensic database. *Int J Legal Med.* 118: 294-306.

Brandstätter A, Salas A, Gassner C *et al.* (2006). Dissection of mitochondrial haplogroup H using coding region SNPs. *International Congress Series.* 1288: 7-9.

Brión M, Blanco- Vereá A, Lareu V *et al.* (2004). 29 Y-chromosome SNP analysis in European populations. *International Congress Series.* 1261:73-75.

Brión M, Quintans B, Gonzales-Neira A *et al.* (2003). Microgeographic patterns of highly informative Y-chromosome haplotypes (using biallelic markers and STRs) in Galicia (NW Spain): forensic and anthropological implications. *International Congress Series.* 1239:61-66.

Brown WM (1980). Polymorphisms in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. *P. Natl Acad. Sci USA* 77:360 5-9.

Budowle B, Allard MW, Wilson MR *et al.* (2003). Forensics and Mitochondrial DNA: Applications, debates, and Foundations. *Annu. Rev.Genomics Hum. Genet.* 4: 119-141.

Butler JM (2003). Recent Developments in Y-Short Tandem Repeat and Y-Single Nucleotide Polymorphism Analysis. *Forensic Science Review.* 15: 2: 92-111.

Byrnes RM (1996). South Africa: a country study. Washington: GPO for the Library of Congress.

Cajee ZA (2003). Islamic History & Civilisation in South Africa: The Impact of Colonialism, Apartheid, and Democracy (1652- 2004). *Symposium on "Islamic civilisation in Eastern Africa"*, Islamic university of Uganda, Kampala, Uganda.

Calafell F, Comas D, Bertranpetit J (2002). Why names. *Genome research*. 12: 219-22.

Carracedo A, Bär W, Lincoln P *et al.* (2000). DNA Commission of the International Society for Forensic Genetics: Guidelines for mitochondrial DNA typing. *Forensic Science International*. 110: 79-85.

Carvalho- Silva DR, Tyler-Smith C (2008). The Grandest Genetic Experiment Ever performed on Man? A Y-Chromosomal Perspective on Genetic Variation in India. *Int J Hum Genet*. 8:1-2: 21-29.

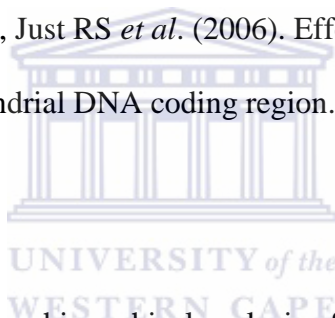
Cavalli-Sforza LL, Feldman MW (2003). The application of molecular genetics approaches to the study of human evolution. *Nature Genetics supplement* . 33: 266-275.

Chakaravarti A (2001). Single nucleotide polymorphism...to a future of genetic medicine. *Nature*. 409: 822-823.

Charlesworth B (2003). The organization and evolution of the human Y-chromosome. *Genome Biology*. 4:1: 226.

Cinnioglu C, King R, Kivisild T *et al.* (2004). Excavating Y-Chromosome haplotype strata in Anatolia. *Hum Genet*. 114: 127-148.

Coble MD, Vallone PM, Just RS *et al.* (2006). Effective strategies for forensics analysis in the Mitochondrial DNA coding region. *Int J Legal Medicine*. 120: 27-32.



Cox PM (2006). Minimum hierarchical analysis of global Y-chromosome SNP diversity by PCR-RFLP. *Anthropological Science*. 114:69-74.

Cruciani F, Fratta RL, Santolamazza P *et al.* (2004). Phylogeographic analysis of haplogroup e3b (E-M215) Y-chromosome reveals multiple migratory events within and out of Africa. *Am.J. Hum. Genet*. 74: 1014-1022.

Cruciani F, Santolamazza P, Shen P *et al.* (2002). A back migration from Asia to sub-Saharan Africa is supported by high Resolution analysis of human Y-chromosome haplotypes. *Am.J. Humn. Genet*. 70:1197-1214.

Da Costa Y (1994). Muslims in Greater Cape Town: a Problem of Identity. *The British Journal of Sociology*. 45:2: 235-246.

Dangor SE (2003). The establishment and consolidation of Islam in South Africa: from the Dutch colonisation of the Cape to the present. *Historia*. 48:1: 203-220.

Dangor SE (2006). The origins of Muslims. *Journal Magazine Awqaf Insight*. 1: 10-12.

Di Rienzo A, Wilson AC (1999). Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proc. Natl. Acad. Sci. USA*. 88: 1597- 1601.

Dubut V, Chollet L, Murail P *et al.* (2004). MtDNA polymorphism in five French groups: importance of regional sampling. *European Journal of Human Genetics*. 12: 293-300.

Endicott P, Metspalu M, Stringer C *et al.* (2006). Multiplexed SNP typing of Ancient DNA clarifies the origin of Andaman mtDNA Haplogroups amongst South Asian Tribal Populations. *Plos one* (1):e81.doi:10.1371/journal.pone.000008.

Fletcher JC (2004). The use of Pyrosequencing for the analysis of Y-chromosome single nucleotide polymorphisms. *Department of Chemistry, University of Central Florida .MSc dissertation submitted.*

Flores C, Maca- Meyer N, Pèrez JA *et al.* (2003). A Predominant European ancestry of paternal lineages from Canary Islanders. *Annals of Human Genetics.* 67:138-152.

Friedlaender JS, Friedlaender FR, Hodgson JA *et al.* (2007). Melanesia mtDNA Complexity. *PLOS one.* 2:2:e248.

Giordano M, Oefner PJ, Underhill PA *et al.* (1999). Identification by denaturing high performance liquid chromatography of numerous polymorphisms in a candidate region for multiple sclerosis susceptibility. *Genomics.* 56: 247-25.

Goncalves R, Rosa A, Freitas A *et al.* (2003). Y-chromosome lineages in Cabo Verde Islands witness the diverse geographic origin of its first male settler. *Human Genet.* 113: 467-472.

Hammer M.F(1994). A recent insertion of an *Alu* element on the Y-chromosome is a useful marker for human population studies. *Mol. Boil. Evol .* 11:5: 749-761.

Hammer M.F, and Horai Satoshi (1995). Y chromosomal DNA variation and the peopling of Japan *Am, J. Humn. Genet.* 56: 951-962.

Hammer MF, Zegura SL (2002). The human Y-chromosome Haplogroup tree: nomenclature and Phylogeography of Its Major Divisions. *Annu. Rev. Anthropology.* 31: 303-321.

Hellard SL, Ballereau SJ, Visscher PM *et al.* (2002). SNP genotyping on pooled DNAs: comparison of genotyping technologies and a semi automated method for data storage and analysis. *Nucleic Acids Research.* 30:15:1-10.

Herrnstadt C, Howell N (2004). An evolutionary perspective on pathogenic mtDNA mutations: haplogroup associations of clinical disorders. *Mitochondrion.* 4: 791-798.

Houshmand M, Sanati MH, Vakilian M *et al.* (2004). Investigation of the mitochondrial haplogroups M, BM, N, J, K and their frequencies in five regions in Iran. *Iranian Journal of Biotechnology.* 2:1:44-48.

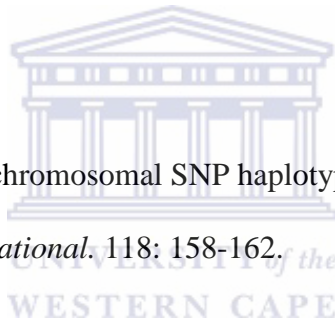
Hurles ME, Sykes BC, Jobling MA *et al.* (2005). The Dual origin of the Malagasy in island Southeast Asia and East Africa: Evidence from Maternal and Paternal lineages. *Am. J. Hum. Genetics.* 76: 000-000.

Hutchison LAD, Myres NM, Woodward SR (2004). Growing the family tree: the power of DNA in reconstructing family relationships. *Proceedings of the first symposium on bioinformatics and biotechnology*. 42-49.

Ingman M (2003). Mitochondria and Human evolution. *University of Uppsala, Comprehensive summaries of Uppsala Dissertations from the Faculty of Science and Technology 1288, PhD dissertation*.

Jobling MA (2001).: In the name of the father: surnames and Genetics. *Trends in Genetics*. 17: 6.

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



Jobling MA, Tyler-Smith C (2003). The Human Y-chromosome: an evolutionary marker comes of age. *Nature*. 4: 598-612.

Jobling MA, Williams G, Schiebel K *et al.* (1998). A selective difference between human Y chromosomal DNA haplotypes. *Current Biology*. 8: 25: 1391-1394.

Jones PN (2004). American Indian mtDNA and Y-chromosome Genetic Data: A comprehensive report of their use in migration and other anthropological studies. 1-135.

Jorde LB, Bamshad M, and Rogers AR (1998). Using mitochondrial and nuclear DNA markers to reconstruct human evolution, *Bioessays*. 20.2:126-36.

Karafet TM, Lansig JS, Redd AJ *et al.* (2005). Balinese y-chromosome perspective on the peopling of Indonesia: genetic contributions from pre-Neolithic hunter-gatherers, Austronesian farmers and Indian traders. *Human Biology*. 77:1: 93-114.

Karafet TM, Mendez FL, Meilerman MB *et al.* (2008). New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Research*. doi:10.1101/gr.7172008.

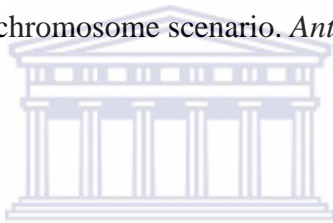
Karafet T, Xu L, Du R *et al.* (2001). Paternal population history of East Asia: sources, patterns and microevolutionary processes. *Am. J. Hum. Genet* . 69:615-628.

Kayser M, Brauer S, Cordaux R, Casto A *et al.* (2006). Melanesian and Asian origins of Polynesians: mtDNA and Y-chromosome gradient across the pacific. *Mol. Biol. Evol.* 23:11: 2234-2244.

Kayser M, Lao O, Anslinger K *et al.* (2005). Significant genetic differentiation between Poland and Germany follows present day political borders as revealed by Y-chromosome analysis. *Hum Genet.* 117: 428-443.

Kivisild T, Bamshad MJ, Kaldma K *et al.* (1999). Deep common ancestry of Indian and Western Eurasian mitochondrial DNA lineages. *Current Biology.* 9:1331-1334.

Krithika S, Maji S, Vasulu TS *et al.* (2007). Molecular genetic perspective of Indian populations: a y-chromosome scenario. *Anthropologist Special Chapter 30.* 3:385-392.



Lareu M, Puente J, Sobrino B (2001). The use of the Lightcycler for the detection of Y-chromosome SNPs. *Forensic Science International.* 118:163-168.

Leat N, Ehrenreich L, Benjeddou M *et al.* (2004). Developments in the use of Y-chromosome markers in forensic genetics. *African Journal of Biotechnology.* 3:12: 637-342.

Luis JR, Rowold DJ, Regueiro M *et al.* (2004). The Levant versus the horn of Africa: Evidence for Bidirectional corridors of human Migration. *Am, J. Humn. Genet.* 74, 000-000.

Lunt DH, Whipple LE, Hyman BC (1998). Mitochondrial DNA variable number tandem repeats (VNTRs): utility and problems in molecular ecology. *Molecular Ecology*. 7: 1441-1455.

Maca-Meyer N, González AM, Larruga JM *et al.* (2001). Major genomic mitochondrial lineages delineate early human expansions. *BMC Genetics*. 2:13:8.

Malhi RS, Kemp BM, Eshleman JA *et al.* (2007). Mitochondrial haplogroup M discovered in prehistoric North Americans. *Journal of Archaeological Science*. 34: 642-648.

Martínez-Cruzado JC, Toro-Labrador G, Ho-Fung V *et al.* (2001). Mitochondrial DNA analysis reveals substantial Native American Ancestry in Puerto Rico. *Human Biology*. 73:4:491-511.

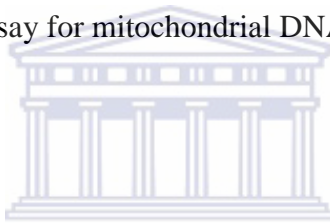
Mills RE, Bennett EA, Iskow RC *et al.* (2006). Recently mobilized transposons in the Human and chimpanzee genomes. *The American Journal of human genetics*. 78: 671-679.

Mishmar D, Ruiz-Pensini E, Golik P *et al.* (2003). Natural selection shaped regional mtDNA variation in humans. *PNAS*. 100:1: 171-176.

Montiel R, Bettencourt C, Silva C *et al.* (2005). Analysis of Y-chromosome variability and its comparison with mtDNA variability Reveals different demographic histories between Islands in the Azores Archipelago (Portugal). *Annals of Human Genetics.* 69: 135-144.

Naito E, Umetsu K, Yuasa I *et al.* (2001). A novel dimorphism in the human SRY gene: usefulness in human migration studies. *Int. J. Legal Med.* 114:274-277.

Nelson TM, Just RS, Schanfield MS *et al.*(2007). Development of a multiplex single base extension assay for mitochondrial DNA haplogroup typing. *Croat Med J.* 48: 460-472.



Nilsson M(2007). Mitochondrial DNA in sensitive forensic analysis. *Digital comprehensive summaries of Uppsala Dissertations from the Faculty of Medicine, Department of Genetics and pathology.* 220: 78.

Novick GE, Novick CC, Yunis J *et al.*(1995). Polymorphic human specific *Alu* insertions as markers for human identification. *Electrophoresis.* 16: 1596- 1601.

Onfori V, Alessandrini F, Turchi C *et al.* (2006). Development of Multiplex PCRs for evolutionary and forensic applications of 37 human Y-chromosome SNPs. *Forensic Science International.* 157: 23-35.

Pakendorf B, Stoneking M (2005). Mitochondrial DNA and Human evolution. *Annu. Rev. Genomics Hum. Genet.* 6:165-183.

Parsons TJ (2006). Mitochondrial DNA genome sequencing and SNP Assay Development for increased Power of discrimination. *USA Department of Justice.* 213502: CX-KO10.

Pastinen T (2000). Scoring human genomic SNPs and mutations: Multiplexed primer extension with manifold and microarrays as solid –support. *University of Helsinki, Department of Human molecular genetics, National Public Health Institute and Department of Medical genetics, Academic dissertation.*

Perić M, Lauc LB, Klarić IM *et al.* (2005). Review of Croatian Genetic Heritage as Revealed by mitochondrial DNA and Y chromosomal lineages. *Croat. Med. J.* 46:4: 502-513.

Quintàns B, Álvarez –Iglesias V, Salas S *et al.* (2004). Typing of mitochondrial DNA coding region SNPs of forensic and anthropological inserts using SNaPShot mini-sequencing. *Forensic Science International.* 140: 251-257.

Ramana GV, Bing S, Jin L *et al.* (2001). Y-chromosome SNP haplotypes suggest evidence of gene flow among caste tribe, and the migrant Siddi populations of Andhra Pradesh, South India. *European Journal of Human Genetics*. 9:695-700.

Rhoda E (2006). The founding and development of the Strand Muslim community, 1822-1928. *Department of History, University of the Western Cape. Msc. Thesis* submitted.

Richards M, 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. *University of Tartu, Department of Evolutionary Biology, PhD Dissertation.*

Rowald DJ (2006). Molecular markers and Human History: A tale of two haplogroups systems. *Florida International University.*

Santos C, Montiel R, Anglés N *et al.* (2004).Determination of Human Caucasian Mitochondrial DNA haplogroups by means of a hierarchical approach. *Human Biology*. 76:3: 431-453.

Schurr TG, Wallace DC (2002).Mitochondrial DNA diversity in Southeast Asian populations. *Human Biology*.74:3: 431-452.

Serk P (2004). Human mitochondrial DNA haplogroup J in Europe and near east. *University of Tartu, Faculty of biology and Geography, Institute of Molecular and Cell Biology, Department of Evolutionary biology, MSC Dissertation.*

Shell RCH (2005). Islam in South Africa, 1653- 2001. *Van Leer Institute, Jerusalem.*

Shi H, Zhong H, Peng Y *et al.* (2008). Y-chromosome evidence of earliest modern human settlement in Eastasia and multiple origins of Tibetan and Japanese populations. *BMC Biology*. 6:45.

Shriver MD, Kittles RA (2004). Genetic ancestry and the search for personalized genetic histories. *Nature Reviews Genetics*. 5:611-618.

Skaletsky H, Kuroda-Kawaguchi T, Minx PJ *et al.* (2003). The male specific region of the human Y-chromosome is a mosaic of discrete sequence classes, *Nature* . 423: 825-837.

Stoneking M. (1998). Women on the move. *Nature genetics*. 20: 219-220.

Tajima A, Pan HI, Fuchqroen G *et al.*(2002). Three major lineages of Asian Y-chromosomes: implications for the peopling of East and Southeast Asia. *Hum. Genet.*110:80-88.

Tambets K, Rootsi S, Kivisild T *et al.* (2004).The western and eastern Roots of the Saami- the story of genetics outliers told buy mitochondrial DNA and Y-chromosome. *Am. J. Hum. Genet.* 74: 661-682.

Thacker C, Harrison C, Brion M *et al.* (2004). Population studies using single nucleotide polymorphisms-how important is detailed sample origin information? *International Congress Series.* 1261: 30-32.

Thangaraj K, Singh L, Reddy A.G *et al.* (2002).Genetic Affinities of the Andaman Islanders, a vanishing human population. *Current Biology.* DOI: 10.1016/S0960982202013362.

Thomson R. Pritchard JK, Shen P *et al.* (2000).Recent common ancestry of human Y-chromosomes: Evidence from DNA sequence data. *PNAS.* 97:13:7360-7365.

Torpey JL (2007).Developing a DNA Microarray for Human Ancestry and Migration. *University of Tufts, Department of Chemistry, Hons. Dissertation.*

Trovoada MJ, Tavares L, Guzman L *et al.* (2007). Dissecting the Genetic History of Sao Tome e Principe, a new window from Y-chromosome biallelic markers. *Annals of Human Genetics*.71: 77-85.

Umetsu K, Yuasa I (2005). Recent progress in mitochondrial DNA analysis. *Legal Medicine*. 7: 259-262.

Underhill PA (2003). Inferring human history: clues from Y-chromosome haplotypes. *Cold Spring Harbor symposia on Quantitative Biology*. LXVIII: 487-493.

Underhill PA (2004). A synopsis of extant Y-chromosome diversity in East Asian and Oceania. 301-315.



Underhill PA, Kivisild T (2007). Use of Y-chromosome and mitochondrial DNA population Structure in Tracing Human Migration. *Annu Rev. Genet.* 41:539-564.

Underhill PA, Shen P, Lin A.A *et al.* (2000) Y-chromosome sequence variation and the history of human populations. *Nature Genetics*. 26: 358-361.

Vahed G (2006). Muslims in Post- Apartheid South Africa, Department of Anthropology, Gender and Historical Studies, University KwaZulu- Natal, article *Awaqaf Insights Journal Magazine*.1: 44-48.

Väli U, Brandström M, Johansson M *et al.* (2008). Insertion- deletion polymorphisms (indels) as genetic markers in natural populations. *BMG Genetics*. 9: 8: 1-8.

Vallone PM, Just RS, Coble MD *et al.* (2004). A multiplex allele-specific primer extension assay for forensically informative SNPs distributed throughout the mitochondrial genome. *Int Journal Legal medicine*. 118:147-157.

Wang DG, Fan JB, Siao CJ *et al.* (1998). Large-scale identification, mapping and genotyping of single nucleotide polymorphism in the human genome. *Science*. 280: 1077-1082.

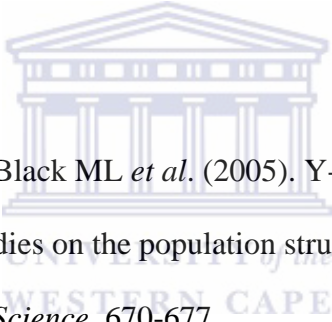
Wang Z, Moulton J (2001). SNPs, proteins structure, and disease. *Human Mutation*. 17:263-270.

Weale ME, Shah T, Jones AL *et al.* (2003). Rare deep rooting Y-chromosome lineages in humans; lessons for phylogeography. *Genetics*. 165:229-234.

Wells S (2007). Deep Ancestry: Inside the Genographic project. National Geographic. 1-247.

Wetton JH, Tsang KW, Khan H (2005). Inferring the population of origin of DNA evidence within the UK by allele-specific hybridization of Y-SNPs. *Forensic Science International*. xxx-xxx.

Whit Athey T, Nordtvedt K (2005). Resolving the placement of haplogroup I-M223 in the Y-chromosome Phylogenetic Tree. *Journal of Genetic Genealogy*. 1: 54-55.



Wise CA, Sullivan SG, Black ML *et al.* (2005). Y-Chromosome and Mitochondrial DNA studies on the population structure of the Christmas Island community. *Wiley InterScience*. 670-677.

Zerjal T, Well RS, Yuldasheva N *et al.* (2002). A genetic landscape reshaped by recent events: Y-chromosomal insights into Central Asia. *Am. J. Hum. Genet* .71:466-482.

Zhivotovsky LA, Underhill PA, Cinnioglu C *et al.* (2004). The effective mutation rate at Y-chromosome short tandem repeats, with application to human populations-divergence time. *Am J Hum Genet*. 74:1: 50-61.

Electronic supplementary resources

Stats South Africa: (<http://www.statssa.gov.za>)

Human

Mitochondrial

DNA:

http://www.nfstc.org/pdi/Subject09/pdi_s09_m02_01_a.htm)

Y-chromosomal evolutionary tree: (<http://www.familyreedna.com/public/china>)

World mitochondrial haplogroups: (<http://www.mitomap.org.2008>)

Paternal phylogenetic tree:

<http://genographic.nationalgeographic.com/genographic/lan/en/population.html>).

SNPs: http://www.ornl.gov/sci/techresources/Human_Genome/faq/snps.shtml

www.ornl.gov/hgmis



Appendices

Appendix I: Donor Questionnaire

Donor year of birth	<input type="text"/>	Date of Sampling	<input type="text"/>	Lab code	<input type="text"/>	
PLACE OF BIRTH						
Donor	Mother	Mother's mother	Mother's father	Father	Father's mother	Father's father
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
RELIGION						
Donor	Mother	Mother's mother	Mother's father	Father	Father's mother	Father's father
Islam	<input type="checkbox"/> Islam	<input type="checkbox"/> Islam	<input type="checkbox"/> Islam	<input type="checkbox"/> Islam	<input type="checkbox"/> Islam	<input type="checkbox"/> Islam
Christianity	<input type="checkbox"/> Christianity	<input type="checkbox"/> Christianity	<input type="checkbox"/> Christianity	<input type="checkbox"/> Christianity	<input type="checkbox"/> Christianity	<input type="checkbox"/> Christianity
Hinduism	<input type="checkbox"/> Hinduism	<input type="checkbox"/> Hinduism	<input type="checkbox"/> Hinduism	<input type="checkbox"/> Hinduism	<input type="checkbox"/> Hinduism	<input type="checkbox"/> Hinduism
Judaism	<input type="checkbox"/> Judaism	<input type="checkbox"/> Judaism	<input type="checkbox"/> Judaism	<input type="checkbox"/> Judaism	<input type="checkbox"/> Judaism	<input type="checkbox"/> Judaism
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

UNIVERSITY of the
WESTERN CAPE

HOME LANGUAGE						
Donor	Mother	Mother's mother	Mother's father	Father	Father's mother	Father's father
English	<input type="checkbox"/> English	<input type="checkbox"/> English	<input type="checkbox"/> English	<input type="checkbox"/> English	<input type="checkbox"/> English	<input type="checkbox"/> English
Afrikaans	<input type="checkbox"/> Afrikaans	<input type="checkbox"/> Afrikaans	<input type="checkbox"/> Afrikaans	<input type="checkbox"/> Afrikaans	<input type="checkbox"/> Afrikaans	<input type="checkbox"/> Afrikaans
Urdu	<input type="checkbox"/> Urdu	<input type="checkbox"/> Urdu	<input type="checkbox"/> Urdu	<input type="checkbox"/> Urdu	<input type="checkbox"/> Urdu	<input type="checkbox"/> Urdu
Xhosa	<input type="checkbox"/> Xhosa	<input type="checkbox"/> Xhosa	<input type="checkbox"/> Xhosa	<input type="checkbox"/> Xhosa	<input type="checkbox"/> Xhosa	<input type="checkbox"/> Xhosa
Zulu	<input type="checkbox"/> Zulu	<input type="checkbox"/> Zulu	<input type="checkbox"/> Zulu	<input type="checkbox"/> Zulu	<input type="checkbox"/> Zulu	<input type="checkbox"/> Zulu
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
POPULATION GROUP						
Donor	Mother	Mother's mother	Mother's father	Father	Father's mother	Father's father
Coloured	<input type="checkbox"/> Coloured	<input type="checkbox"/> Coloured	<input type="checkbox"/> Coloured	<input type="checkbox"/> Coloured	<input type="checkbox"/> Coloured	<input type="checkbox"/> Coloured
African	<input type="checkbox"/> African	<input type="checkbox"/> African	<input type="checkbox"/> African	<input type="checkbox"/> African	<input type="checkbox"/> African	<input type="checkbox"/> African
White	<input type="checkbox"/> White	<input type="checkbox"/> White	<input type="checkbox"/> White	<input type="checkbox"/> White	<input type="checkbox"/> White	<input type="checkbox"/> White
Indian	<input type="checkbox"/> Indian	<input type="checkbox"/> Indian	<input type="checkbox"/> Indian	<input type="checkbox"/> Indian	<input type="checkbox"/> Indian	<input type="checkbox"/> Indian
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

Appendix IIa: Data (adapted from Rhoda 2006)

Movement of free blacks of Indonesian & other origins to Mosterd Bay between 1822 and 1840																
	1822	1823	1824	1825	1826	1827	1828	1829	1830	1832	1833	1836	1837	1838	1839	1840
Opgaafrol reference	J268	J268	J274	J276	J283	J286	J288	1/STB 16/25	J294	J302	J302	J304	J307	J309	J309	J310
Salomon van Jafa	*#	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Abdol Sammat	*#	*	*	*	*	*	*	*	*	*	*	*	*			
Borno van Jafa	*#		*		*	*		*	*	*						
Seding van Jafa	*#	*					*		*	*	*	*	*	*	*	*
Cadier (van Timor)	*#			*	*	*	*		*	*	*	*				
Rachiem	*#				*	*	*	*	*							
Wiro van Jafa		#*	*	*	*	*	*	*	*	*	*					
Ongo van Jafa		#*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Marjam van Jafa		#*	*	*	*	*	*									
Bonzo van Jafa		#*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Azor van Jafa		#*		*					*							
Zingo van Jafa			#*	*				*			*	*		*	*	*
Abdol van de Kaap				#*	*	*	*	*	*							
Vryswart van Jafa				#*	*	*	*	*	*							
Nollo van Jafa				#*	*	*	*	*	*							
Azam van Jafa				#*	*	*	*	*								
Kwassa van Jafa				#*	*		*	*								
Ronnie van Jafa				#*	*	*										
Isakiem van Jafa				#*												
Singo van Jafa				#*			*	*								
Allie van Jafa					#*											
Darius van de Kaap						#*	*	*	*	*	*	*	*	*	*	*
Minerva van Batavia						#*	*	*	*	*	*	*	*	*	*	*
Podro van Jafa							#*		*	*	*	*	*	*	*	*
Seding van Jafa Jnr.								#*	*	*	*	*	*	*	*	*
Jamier van Timor								#*	*		*					*
Sakier van Timor									#*	*	*	*	*	*		*
Simon van Jafa										#*	*	*	*	*	*	*
Kameding van de Kaap										#*	*	*	*	*	*	*
# = Year of arrival at M/Bay.																
Annual arrivals at M/Bay	6	5	1	8	1	2	1	2	1	2						
Of the 29 free blacks of Indonesian and Cape-born origin who settled at Mosterd Bay between 1822 & 1832, seventeen were still at Mosterd Bay in 1832. Thus almost 60% of those free blacks who settled at Mosterd Bay in this period chose to remain there. Imam Abdus Sammat had his will drawn up in 1832.											17	In 1840 almost 45% of the 29 free blacks were still at Mosterd Bay. Imam Abdus Sammat died on 7 March 1838.	13			

Appendix Iib: Data (adapted from Rhoda 2006)

List of Javanese free blacks who stayed in the same area with Imam Abdus Sammat in Cape Town, circa 1810.				
No.	Name	Place of birth	Age	Address
1	Nolla (Nollo)	Samarang	40	32 Loop Street
2	Singo	Java	50	4 Draay Steeg
3	Seding (Sen.)	Java	50	<i>In de kelder van de predikant Fleck</i>
4	Seding (Jun.)	Java	40	12 Riebeeck Street
5	Azam	Java	50	3 Draay Steeg
6	Abdol Sammat	Samarang	50	1 Zee Street
7	Ongo	Java	40	46 Water[kant] Street
8	Salomon	Batavia	43	12 Visch Steeg

Appendix III: Protocols

Buccal Amp DNA extraction protocol for buccal swabs.

1. Place swab fragment in 250ul of Epicenter extraction buffer
2. Vortex for 15 seconds
3. Heat at 65C for five minutes
4. Vortex for 15 seconds
5. Heat at 98C for 2 minutes
6. Vortex for 15 seconds.

Purification and concentration of DNA samples

1. Add an equal volume of Phenol/Chloroform/Isoamyl Alcohol (25:24:1) to DNA solution to be purified (50µl to 100µl) in a eppendorf tube. Add TE buffer to reduce sample loss.
2. Vortex for 5 minutes
3. Centrifuge at 13200rpm for 10 minutes at room temperature.
4. Transfer the top ,aqueous phase to clean tube using 200µl pipettor
5. Add an equal volume of chloroform/isoamyl alcohol (24:1)
6. Vortex 2 minutes and centrifuge for 5minutes
7. Transfer top phase to clean eppendorf tube
8. Add 1/10th volume 3M Sodium acetate ph 5.5, mix briefly
9. Add 1.25 volume isopropanol, mix and leave for 30 minutes at -70C
10. Spin in centrifuge at 8000-10000 rpm for 20 minutes

11. Remove supernatant , do not disturb pellet
12. Add 200µl of cold 70% ethanol, centrifuge for 10minutes
13. Remove 200µl supernatant and allow ethanol to evaporate
14. Resuspend pellet in desired volume

DNA Quantitation

Nanodrop[®] ND-1000 Spectrophotometer V3.3.0

The NanoDrop[®] ND-1000 is a full-spectrum (220-750nm) spectrophotometer that measures 2µl samples with high accuracy and reproducibility. It utilizes a patented sample retention technology that uses surface tension alone to hold the sample in place. This eliminates the need for cuvettes and other sample containment devices and allows for clean up in seconds. In addition, the ND-1000 has the capability to measure highly concentrated samples without dilution (75X higher concentration than the samples measured by a standard cuvette spectrophotometer) (*Nanodrop[®] ND-1000 V3.3.0 User's manual*). The Nucleic Acid Measurement Module was used to check the concentration and quality of the DNA samples.

Primer mix of Multiplex I (Onofri *et al.* 2006) M170, M173, M45 and M89

Primers were prepared as follows

10µl of 100µM of revers and forward primer for each marker was pooled together.

This was done to obtain 20µl mix of primers at a concentration of 50uM each.

Hereafter a working Stock was prepared

Primer	Final	{ 10x }
M170	0.5 μ M	5 μ l
M89	0.5 μ M	5 μ l
M173	0.5 μ M	5 μ l
M45	0.5 μ M	5 μ l
H20		30 μ l

Final volume 50 μ l

Primer mix of Multiplex II (Onofri *et al.* 2006) M181, M96, M91 and M52



Primers were prepared as follows

10 μ l of 100 μ M of reverse and forward primer for each marker was pooled together.

This was done to obtain 20 μ l mix of primers at a concentration of 50 μ M each.

Hereafter a working Stock was prepared

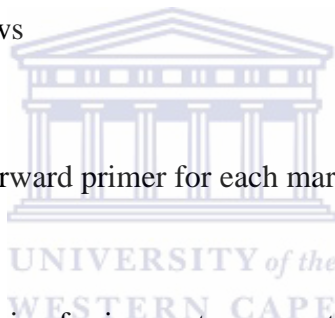
Primer	Final	{ 10x }
M181	0.5 μ M	5 μ l

M96	0.5µM	5µl
M91	0.5µM	5µl
M52	0.5µM	5µl
H2O		30µl

Final volume 50µl

Primer mix of Multiplex III (Onofri *et al.* 2006) M172 and M9

Primers were prepared as follows



10µl of 100µM of reverse and forward primer for each marker was pooled together.

This was done to obtain 20 µl mix of primers at a concentration of 50µM each.

Hereafter a working Stock was prepared

Primer	Final	{ 10x }
M172	0.5µM	5µl
M9	0.5µM	5µl
H2O		40µl

Final volume 50µl

Primer mix of Multiplex IV (Onofri *et al.* 2006) M216, M201 and M214


Primers were prepared as follows

10µl of 100µM of reverse and forward primer for each marker was pooled together.

This was done to obtain 20 µl mix of primers at a concentration of 50µM each.

Hereafter a working Stock was prepared

Primer	Final	{ 10x }
M216	0.5µM	5µl
M214	0.5µM	5µl
M201	0.5µM	5µl
H2O		35µl



Final volume 50µl

Primer mix of Multiplex M1 (Nelson *et al.*)

Primers were prepared as follows

10µl of 100µM of reverse and forward primer for each marker was pooled together.

This was done to obtain 20µl mix of primers at a concentration of 50µM each.

Hereafter a working Stock was prepared

Primer	Final	{ 10x }
--------	-------	---------

8272-8280	0.5µM	5µl
-----------	-------	-----

10398/10400	0.5µM	5µl
-------------	-------	-----

3594	0.5µM	5µl
------	-------	-----

7028	0.5µM	5µl
------	-------	-----

12406	0.5µM	5µl
-------	-------	-----

7600	0.5µM	5µl
------	-------	-----

H2O		20µl
-----	--	------

Final volume 50µl



Primer mix of Multiplex M2 (Nelson *et al.*)

Primers were prepared as follows

10µl of 100µM of reverse and forward primer for each marker was pooled together.

This was done to obtain 20µl mix of primers at a concentration of 50µM each.

Hereafter a working Stock was prepared

Primer	Final	{ 10x }
13263	0.5µM	5µl
1719	0.5µM	5µl
5178	0.5µM	5µl
4833	0.5µM	5µl
633	0.5µM	5µl
H2O		25µl

Final volume 50µl





UNIVERSITY *of the*
WESTERN CAPE