

https://dspace.org/

Y-chromosome variation in the South African 'coloured' population

Motladiile, Thejane Wilson

2004
http://hdl.handle.net/10539/21454

Downloaded from a DSpace repository. https://dspace.org/

Y-CHROMOSOME VARIATION IN THE SOUTH AFRICAN 'COLOURED' POPULATION

THEJANE WILSON MOTLADIILE

A dissertation submitted to the faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Master of Science in Medicine in the Division of Human Genetics.

November 2004

DECLARATION

I declare that this dissertation is my own, unaided work. It is being submitted for the degree of Master of Science in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in any other University. Furthermore, this work has been approved by the Ethics Committee for Research on Human Subjects at the University of the Witwatersrand, and the certificate numeral is M02-05-27.



THEJANE WILSON MOTLADIILE

Signed at Johannesburg on this 17th day of November 2004.

ABSTRACT

Genetic polymorphisms within the non-recombining portion of Y-chromosome (NRY) preserve a record of human paternal genetic heritage that has persisted to the present, allowing human evolutionary inference, population affinity and demographic history, to be elucidated. To elucidate the geographic origins of the paternal ancestry of the present-day South African (SA) 'Coloured' population, a total sample of 167 individuals consisting of Cape Malay (N=54) and 'Coloured' groups from Cape Town (N=48) and Johannesburg (N=65) were analysed at 21 binary and eight short tandem repeat (STR) polymorphic loci within NRY. A SA White sample (present study, N=97) as well as other presumed parental populations were included for comparative analysis. Haplotypes constructed using both biallelic haplogroup and STR haplotype data assisted in resolving the geographic regions of origin of Y-chromosome in these groups. Altogether the proportions of African, European and Asian contributions were estimated to be 0%, 18.5% and 46.3% in the Cape Malay, 31.3%, 25% and 20.1% in the Cape 'Coloureds', and 24.6%, 40% and 16.9%, in the 'Coloured' group from Johannesburg. Those haplotypes that could not be unambiguously resolved to European or Asian origins were referred to as Eurasian lineages, and constituted 35.2%, 22.9% and 18.5% of Y-chromosomes in the Cape Malays, Cape 'Coloureds' and Johannesburg 'Coloureds', respectively. While the 'Coloured' groups currently residing in Cape Town and Johannesburg were not significantly different from each other, both groups were significantly different from the Cape Malay population. This was further supported from the association of these groups in population trees. For the most part, these data corroborate historical data concerning the history of 'Coloured' populations, but is the first study to show how males have contributed in shaping the gene pool of the 'Coloured' population from South Africa.

DEDICATION

This work is dedicated especially to my parents, brothers and sisters for their encouragement and support throughout my academic studies, and to my fiancé, Kelebogile, for her warm and everlasting love and our first baby-boy Thapelo.

ACKNOWLEDGMENTS

I am very thankful to the individuals who gladly donated blood samples for DNA extractions and use in this study. I am also grateful to Professors Himla Soodyall, Paul van Helden as well as Dr Eileen Hoal for providing the DNA samples. My special gratitude is extended to my supervisor Professor Himla Soodyall for her insight and helpful guidance throughout the study. Dr Almut Nebel is acknowledged for her helpfulness and willingness to take time and teach me how to design mismatch primers and to use computer software packages for statistical analyses. I'd like to thank Dr Manfred Kayser for providing us with the Southeast Asian comparative data.

Graciously, I would like to thank the Division of Human Genetics (Wits University), in particular Professor Himla Soodyall, for granting me the opportunity to further my academic studies. The Medical Research Council (MRC), National Health Laboratory Service (NHLS), and the University of the Witwatersrand, made the facilities available for the research reported in this dissertation, and have also supported me financially. I have also received a one-year additional funding from the Ernst and Ethel Eriksen Trust.

I wish to express my appreciation to colleagues in the Human Genomic Diversity and Disease Research Unit (HGDDRU) – Akashnie Maharaj, Vanessa Pillay and Heeran Makkan for their dependable companionship and be of assistance in proofreading all the results presented in this document. To the entire Departmental staff members as well as students whom have made my stay possible and enjoyable, I thank you. Finally, I would like to thank GOD for his mercy and all the strength and support his given me to accomplish the objectives of this research.

TABLE OF CONTENTS Page

Declaration..... ii

Abstract.....iii

Dedication.....iv

Acknowledgements.....v

Table of contents.....vi

List of figures.....xi

List of tables.....xvi

List of abbreviations.....xvii

CHAPTER 1

INTRODUCTION

1.1 South African History: Brief Overview.....1

1.2 "Coloured by history, shaped by place" – Some insights into the problem of nomenclature.....2

1.3 Tracing the male ancestors of the South African "Coloured" population – the historical archive.....5

1.4 Reconstructing ancestry: the Y-chromosome as a tool.....10

1.4.1 Structural features of the Y-chromosome.....10

1.4.2 Y-chromosome in population and evolutionary studies.....11

1.4.3 Y-chromosome polymorphisms.....12

1.4.4 Y-biallelic polymorphisms.....13

1.4.4.1 Y-chromosome phylogeography..... 14

1.4.4.2 Global distribution of Y-chromosome haplogroups.. 16

1.4.5 Y-short tandem repeats (Y-STRs).... 19

1.4.5.1 Genealogical applications of Y-STRs.....19

1.5 Objectives of this study.....21

CHAPTER 2

SUBJECTS AND METHODS

2.1 Population samples.....22

2.1.1 Comparative data sets 22

2.2 Molecular methods.....25

2.2.2 Mismatched base primer design.....25

2.2.3 Y-Biallelic polymorphisms typing.....27

2.2.3.1 PCR-RFLP analysis... 29

2.2.3.2 Typing strategy..... 31

2.2.4 Terminology and Nomenclature..... 32

2.2.5 Y-STR polymorphism typing 33

2.2.5.1 Multiplex PCR..... 33

2.2.5.2 Polyacrylamide gel electrophoresis... 35

2.2.6 DNA sequencing analysis..... 35

2.2.6.1 Cycle sequencing PCR.....36

2.2.7 Statistical analysis..... 36

2.2.7.1 Intrapopulation variation analysis..... 37

2.2.7.2 Interpopulation variation analysis..... 37

2.2.7.2.1 Network analysis.....38

CHAPTER 3

RESULTS

3.1 Assignment of alleles to haplogroups 39

3.2 Haplotyping Y-chromosome STRs.....40

3.3 Intrapopulation variation.....41

3.3.1 Y-biallelic haplogroup variation.....41

3.3.2 Y-STR haplotype variation.....41

3.3.3 Y-chromosome diversity in the 'Coloured' population.....45

3.3.3.1 Y-haplogroup and haplotype diversity.....45

3.3.3.2 Y-lineage diversity.....46

3.4 Interpopulation variation.....47

3.4.1 Population differentiation.....47

3.4.2 Tracing the ancestry of Y-lineages in the South African 'Coloureds'.....48

3.4.2.1 African origin of Y-chromosome lineages.....51

3.4.2.1.1 Haplogroup A*.....51

3.4.2.1.2 Haplogroup B-M150.....53

3.4.2.1.3 Haplogroup E-M2.....55

3.4.2.1.4 Haplogroup E-M75.....57

3.4.2.2 Summary.....58

3.4.2.3 European origin of Y-chromosome lineages.....59

3.4.2.3.1 Haplogroup E-M35.....59

3.4.2.3.2 Haplogroup I-M170.....61

3.4.2.3.3 Haplogroup R-M207..63

3.4.2.3.4 Haplogroup R-SRY10831.2...66

3.4.2.4 Summary.....67

3.4.2.5 Asian origin of Y-chromosome lineages.....68

3.4.2.5.1 Haplogroup C-M130..68

3.4.2.5.2 Haplogroup D-M174. 70

3.4.2.5.3 Haplogroup H-M52..70

3.4.2.5.4 Haplogroup KMN-M9.....71

3.4.2.5.5 Haplogroup L-M11...73

3.4.2.5.6 Haplogroup O-M175. 74

3.4.2.5.7 Haplogroup PQ-M74. 75

3.4.2.6 Summary.....76

3.4.2.7 Eurasian origin of Y-chromosome lineages...77

3.4.2.7.1 Haplogroup FG-M213.....77

3.4.2.7.2 Haplogroup J-p12f2...79

3.4.2.7.3 Haplogroup J-M172...81

3.4.2.8 Summary.....82

3.4.3 Population Affinities..84

CHAPTER 4

DISCUSSION AND CONCLUSION

4.1 Paternal ancestry of the South African 'Coloured' population.....88

4.2 Founding haplogroups for the South African 'Coloureds'... 91

4.3 Population affinities...95

4.4 Significance of population studies in disease management... 96

4.5 Significance of haploid markers in Forensic applications.... 98

4.6 Conclusion.... 99

REFERENCES.....100

Electronic Database Information..... 116

APPENDIX A.....117

Solutions..... 117

APPENDIX B.....122

Table 1: Y-chromosome PCR-based biallelic polymorphism reagents and product sizes... 122

Table 2: Y-chromosome PCR-based STR polymorphism reagents, product and repeat sizes... 123

APPENDIX C.....124

Table 1: Y-STR haplotypes within biallelic haplogroups and their frequencies in the SA 'Coloured' and SA White populations 124

APPENDIX D.....129

Ethics Approval..... 129

LIST OF FIGURES

CHAPTER 1

Figure 1.1: The Dutch East India Company's arrival and meeting with the Khoisan… 6

Figure 1.2: The slave trading routes to the Cape Peninsula. Different colour circles represent particular geographic locations from which slaves were drawn… 9

Figure 1.3: Schematic representation of the human Y-chromosome, including NRY and PARs.. 10

Figure 1.4: The distribution of geographic specific haplogroups in contemporary global populations, Underhill *et al.* (2001). Note: the author shall use the proposed YCC nomenclature to elucidate the global distribution of Y-chromosome haplogroups... 15

Figure 1.5: A phylogenetic tree for the 18 major NRY-specific biallelic haplogroups or paragroups identified in global populations, based on 20 markers (Hammer and Zegura 2002).. 17

CHAPTER 2

Figure 2.1: Sampling sites for the South African 'Coloured' and White populations. Note all the samples from Gauteng and Western Cape provinces, were collected in the region of Johannesburg and Cape Town, respectively..... 23

Figure 2.2: A schematic representation of mismatched oligonucleotide primer design. Note: the green typescripts denote the primer binding sites and the primers, whereas the mismatches and their positions are shown in red. Restriction sites are shown as blue boxes with the lines indicating where the enzyme cuts..... 26

Figure 2.3: Maximum parsimony phylogenetic tree showing all the haplogroup defining polymorphisms included for genotyping in the present study..... 32

CHAPTER 3

Figure 3.1a: Images of the agarose gels showing genotypes for some of the NRY specific polymorphisms included in this study..... 39

Figure 3.1b: An electrophoregram showing the non-overlapping ranges of the Y-chromosome STR alleles labelled with four different types of flourephores 40

Figure 3.2: Phylogenetic tree showing haplogroup composition in three South African 'Coloured' and the White populations. Dotted lines represent haplogroups not detected in a sample..... 43

Figure 3.3: Allelic variation at eight Y-STR loci in the three SA 'Coloured' groups and the SA White populations..... 44

Figure 3.4: Distributions of Y-biallelic haplogroups in the South African 'Coloured' (Cape Town and Johannesburg) and White populations..... 49

Figure 3.5: Frequency distributions of Y-chromosome biallelic haplogroups in some contemporary global human populations. The African populations include Khoisan (Nama and !Kung) and Bantu-speakers from South Africa (Soodyall, unpublished). Europeans include Dutch, French (Semino *et al.* 2000) and British (Wells *et al.* 2001). The Asian data used in the comparative analysis included Indians (Ramana *et al.* 2001), Iraqis (Al-zahery *et al.* 2002), Yakuts (Puzyrev *et al.* 2003), Chinese and Malaysians (Kayser *et al.* 2003).....50

Figure 3.6: Median-joining network showing the relationship of Y-STR haplotypes within paragroup A* found in the Khoisan, SA Bantu-speakers and SA 'Coloureds'. Nodes are represented as black circles. The lines joining haplotypes and nodes are

proportional to the number of mutational steps. The size of the circles is proportional to the frequencies of haplotypes 52

Figure 3.7: Median-joining network showing the relationship between Y-STR haplotypes within haplogroup B-M150 found in the SA Bantu-speaking (Soodyall, unpublished) and SA 'Coloured' populations.....54

Figure 3.8: Median-joining network showing the relationship between Y-STR haplotypes within haplogroup E-M2 found in the SA Bantu-speakers, Khoisan (Soodyall, unpublished), SA White and SA 'Coloured' populations.....56

Figure 3.9: Median-joining network showing the relationship between Y-STR haplotypes within haplogroup E-M75 found in the SA Bantu-speakers (Soodyall, unpublished) and SA 'Coloured' populations.....57

Figure 3.10: Median joining network showing the relationship between Y-STR haplotypes within haplogroup E-M35 found in the SA White, SA Bantu-speaking, Khoisan and SA 'Coloured' populations..... 60

Figure 3.11: Median-joining network showing the relationship between Y-STR haplotypes within haplogroup I-M170 found in the SA White, SA Bantu-speaking, Khoisan and SA 'Coloured' populations..... 62

Figure 3.12: Median-joining network showing the relationship between Y-STR haplotypes within haplogroup R-M207 found in the SA White, SA Bantu-speaking, Khoisan and SA 'Coloured' populations..... 65

Figure 3.13: Median-joining network showing the relationship of Y-STR haplotypes within haplogroup R-SRY10831.2 found in the Asians, SA Whites and SA 'Coloureds'. Note: DYS388 is excluded in this comparative analysis that includes SA Indian (i) (Soodyall, unpublished), Malaysian, Chinese, and Russian populations..... 67

Figure 3.14: Median-joining network showing the relationship of Y-STR haplotypes within haplogroup C-M130 found in the Asian – including SA Indian (Soodyall, unpublished), SA Bantu-speaking and SA 'Coloured' populations.. 69

Figure 3.15: Median-joining network showing the relationship of Y-STR haplotypes within haplogroup H-M52 found in the SA Indian (Soodyall, unpublished) and SA 'Coloured' populations. DYS388 is excluded in this comparative analysis that includes SA Indian, Malaysian, Chinese, and Russian populations.....71

Figure 3.16: Median-joining network showing the relationship of Y-STR haplotypes within haplogroup KMN-M9 found in the Asian, SA White, SA Bantu-speakers and SA 'Coloured' populations. DYS388 is excluded in this comparative analysis that includes SA Indian, Malaysian, Chinese, Russian and Anatolian populations..... 72

Figure 3.17: Median-joining network showing the relationship of Y-STR haplotypes within haplogroup L-M11 found in the Asian and SA 'Coloured' populations. DYS388 is excluded in this comparative analysis that includes SA Indian, Malaysian, Chinese, and Russian populations... 73

Figure 3.18: Median-joining network showing the relationship of Y-STR haplotypes within haplogroup O-M175 found in the Asian and SA 'Coloured' populations. DYS388 is excluded in this comparative analysis that includes SA Indian, Malaysian, Chinese, and Russian populations... 75

Figure 3.19: Median-joining network showing the relationship between Y-STR haplotypes within haplogroup FG-M213 found in the Asian, SA White, Khoisan and SA 'Coloured' populations. DYS388 is excluded in this comparative analysis that includes SA Indian, Malaysian, Chinese, Russian and Anatolian populations 78

Figure 3.20: Median-joining network showing the relationship between Y-STR haplotypes within haplogroup J-12f2 found in the Asian and SA White populations. DYS388 is excluded in this comparative analysis that includes SA Indian, Malaysian, Chinese, and Russian populations.....80

Figure 3.21: Median-joining network showing the relationship between Y-STR haplotypes within haplogroup J-M172 found in the Asian and SA White populations. DYS388 is excluded in this comparative analysis that includes SA Indian (Soodyall, unpublished), Malaysian, Chinese, and Russian populations. Note: Haplotypes H84/H88 and H90/H92 (Appendix C) are the same when the DYS388 is excluded.....82

Figure 3.22: Unweighted pair group method with arithmetic mean (UPGMA) tree illustrating the relationships among the SA 'Coloured' and SA White populations, based on the Slatkin linearised Fst distances.....85

Figure 3.23: Unrooted neighbour-joining (NJ) tree showing the relationships among the SA 'Coloured' and parental populations, based on the Slatkin linearised Fst distances.....86

CHAPTER 4

Figure 4.1: Proportion of Y-chromosome contributions from African, European, Asian and Eurasian ancestry in the Cape Malay, Cape 'Coloured' and Johannesburg 'Coloured' populations from South Africa.....91

LIST OF TABLES

CHAPTER 2

Table 2.1: Male parental population samples used in the comparative analysis with the South African 'Coloured' populations.....24

Table 2.2: NRY-specific unique event biallelic polymorphisms.....28

Table 2.3: The NRY locus specific biallelic polymorphisms and PCR-RFLP analysis conditions. Note: (*Forward and 'Reverse primers).....30

Table 2.4: Y-STR oligonucleotide primers and PCR amplification conditions.....34

CHAPTER 3

Table 3.1: Y-biallelic and STR polymorphism diversity in the SA 'Coloured' and SA White populations.....45

Table 3.2: Y-chromosome lineage diversity within the South African 'Coloured' population, based on the combined haplogroup-haplotype data.....47

Table 3.3: Differentiation of the 'Coloured' population pairs at 0.01 and 0.05 levels of significance.....48

Table 3.4: Total genetic contributions of parental population in the South African 'Coloured' population.....83

Table 3.5: Slatkin's population pair-wise linearised Fst distances (Below the Diagonal) and P-values (Above the Diagonal) between the SA 'Coloured' and SA White populations.....84

Table 3.6: Slatkin's population pair-wise linearised Fst distances (Below the Diagonal) and P-values (Above the Diagonal) between the SA 'Coloured' populations and their presumed parental populations.....87

LIST OF ABBREVIATIONS

A	Adenine nitrogenous base
ANA	Anatolia (Asia Minor)
BAC	Bacterial artificial chromosome
BMA	BioWhittaker Molecular Applications
BOR	Borneose
bp	Nucleotide base pair
BRI	British
BSA	Bovine serum albumin
C	Cytosine nitrogenous base
CAM	Cameroonian
CAR	Central African Republic
CC and CCO	Cape 'Coloureds'
CHI	Chinese
CM and CMA	Cape Malays
CMH	Cohen modal haplotype
CON	Congolese
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
ddH2O	double distilled water
dGTP	deoxyguanosine triphosphate
dNTP	deoxyribonucleotide triphosphate
dTTP	deoxythymidine triphosphate
DEIC	Dutch East India Company

DHPLC	Denaturing high-performance liquid chromatography
DNA	Deoxyribonucleic acid
DUT	Dutch
DWIC	Dutch West India Company
EDTA	Ethylenediamine tetra-acetic acid
FRE	French
Fst	Co-ancestry coefficient
G	Guanosine nitrogenous base
GER	German
HG	Haplogroup
ht	Haplotypes
HGDDRU	Human genomic diversity and disease research unit
IND	Indian
Indels	Insetion/deletion mutations
KUN	!Kung
l	Litre
JC and JCO	Johannesburg 'Coloured'
LD	linkage disequilibrium
M	mutations
MAL	Malaysian
MALD	Mapping by admixture linkage disequilibrium
mb	Mega base
$MgCl_2$	Magnesium chloride
MEGA	Molecular evolutionary genetic analysis
µg	Microgram

µl	Microlitre
µM	Micromolar
ml	Millilitre
mM	Millimolar
MOZ	Mozambican
MRC	Medical Research Council
mtDNA	mitochondrial DNA
NAM	Nama
NEB	New England Biolabs
ng	Nanogram
NHLS	National health laboratory service
NJ	Neighbour joining
nm	nanometer
NRY	Non-recombining portion of Y-chromosome
PAR	Pseudoautosomal region
PCR	Polymerase chain reaction
PE	Perkin-Elmer
pH	negative logarithmic measure of hydrogen ion concentration
PHI	Philippinos
RFLP	Restriction fragment length polymorphism
RW and SAW	South African Whites
SAB	South African Bantu-speakers
SD	Standard deviation / variance
SAI	South African Indians
SANAC	South African Native Affairs Commission

SNPs	Single nucleotide polymorphisms
STRs	Short tandem repeats
Taq	Thermus aquaticus
T	Thiamine nitrogenous base
TAN	Tanzanians
TB	Tuberculosis
TE	Tris-EDTA buffer
Tm	Melting temperature
U	Enzyme units
UEPs	Unique event polymorphisms
UPGMA	Unweighted pair group method with arithmetic mean
UV	Ultraviolet
v	Volume
VNTR	Variable number of tandem repeats
w	weight
Wits	University of the Witwatersrand
YAK	Yakuts (Russians)
YAP	Y chromosome *Alu* insertion polymorphism
YCC	Y Chromosome Consortium
Yp	Y chromosome short arm
Yq	Y chromosome long arm

CHAPTER 1

INTRODUCTION

1.1 *South African History: Brief Overview*

South Africa has an estimated population size of approximately 45 million people, consisting of 79.1% Africans, 9.4% Whites, 9% 'Coloureds' and 2.5% Indians/Asians (Census 2001). The present-day population is composed of various ethnic groups defined on the basis of a number of different criteria (e.g. culture, language, religion, "race", etc.) within the broader categories outlined above, each reflecting the influence of historical events dated to over 30,000 years ago[1]. While these events are outside the scope of this thesis, a few scenarios are important in understanding the main objective of this study (outlined in section 1.5).

The Khoisan – San and Khoikhoi, are the indigenous people of southern Africa. The San where until recently predominantly hunter-gatherers of the Later Stone Age, but some have resorted to pastoralism or have become assimilated with other communities. Evidence for hunter-gathering mode of subsistence can be traced from the archaeological record (art and tools) to 20,000 years ago (Deacon and Deacon 1999). The Khoikhoi, it is suggested (Elphick 1985; Deacon and Deacon 1999), have descended from the same common ancestry as the San who acquired cattle and became pastoralists as far back as ~2000 years ago. Both the San and the Khoikhoi have experienced varying degrees of admixture between them, and with other groups

[1] See website http://www.sahistory.org.za/pages/mainframe.htm for a chronological outline of some of these events.

- most notably with people speaking Bantu languages - who migrated to the southern African region within the past 1200 years (Huffman 1982).

The next major influence in shaping the history of South Africa involved the arrival of sea-borne immigrants from Europe since the 1650s. One of the consequences of European settlement in South Africa was the introduction of slaves from West Africa, East Africa, Madagascar and from different parts of the Far East: Mainland of India, Ceylon, Malay Archipelago, East Indies, Philippines, Southern Asia and Japan to the Cape (Marais 1957; Watson 1990). The bringing together of people with diverse geographic regions of origin within the past 2000 years, has vastly changed the social, political, cultural, linguistic and undoubtedly genetic structure of the present-day South African population. One of the consequences of such a complex history was the generation of a so-called "mixed ancestry race" referred to as 'Coloured' in South Africa.

1.2 *"Coloured by history, shaped by place" – Some insights into the problem of nomenclature*

A recent book edited by Zimitri Erasmus (2001) entitled "Coloured by history, shaped by place" resulted from discussions held at a conference in June 1998 in Cape Town on 'Coloured'[2] identities. 'Coloured' identity has been an intensely contested South African identity and issues on their nomenclature are unresolved. According to Hendricks (2001) the "Constructions of 'coloured' identity are intimately linked to the long-established language of 'race'-fixed identities and owe much to the beliefs that emanated from 'scientific' theories formulated in Europe and the New World during the nineteenth century". Hendricks (2001) argues that to "conceive of people as of

[2] The author shall use the word "Coloured" henceforth without any derogatory intentions

'mixed' descent, it is necessary to believe in the existence of separate 'races', 'racial purity' and concomitantly, that sexual intercourse between 'races' – 'miscegenation'- produces a hybrid group".

Historically, the word "Coloured" can be traced to the SANAC report (1905) based on the deliberations of the commission who convened in 1903. The commission made use of the category 'Coloured' in reference to generalized 'native' subjects and also to subjects with certain special and unique features. The word 'native' was used to "mean an aboriginal inhabitant of Africa, South of the Equator, and to include half castes and their descendants by Natives". The commission acknowledged "pure races" – Europeans, "Bantus" and "Asiatics", and "Coloureds" as "mixed groups" with "mixing of blood" (Reddy 2001). The enormous emphasis placed on 'pure blood' was based on the assumption that 'pure bloodiness' actually did exist in certain 'races'.

The Population Registration Act of 1950 allowed the state to classify people along the lines of colour, 'race' and ethnicity. A 'coloured person' was defined as 'a person who is not a white person or a native'; a 'native' means a person who in fact is or is generally accepted as a member of any aboriginal race or tribe of Africa', and a 'white person means a person who in appearance obviously is, or who is generally accepted as, a white person, but does not include a person who, although in appearance obviously a white person, is generally accepted as a coloured person' (Statutes 277). To perfect its classification system so that it corresponded more fittingly with 'reality', the state resorted to a series of amendments; many of which aimed to tighten the loopholes against people making appeals for 're-classification' (Reddy 2001). The government proclaimed in 1959 that the 'Coloured' category be further sub-divided

into 'Cape Coloured', Cape Malay, Griqua, Indian, Chinese, "other Asiatic" and "other Coloured" (Reddy 2001).

Jeppie (2001) addressed the question of identity of Muslims of the western Cape. According to his account, Muslims have been referred to as Malays, "*slamse*" (a corrupted form of 'Islam' and a wholly negative term) and also as coloured Moslems. Muslim slaves and free blacks in the Cape colony came mainly from South and Southeast Asia (Shell 1994, cited in Jeppie 2001). "The popular religious practices of slaves and exiles from the Indonesian archipelago emerged as the dominant theme in local Muslim culture....African slaves, and converts from among both the free and unfree population, brought to the Muslim community an ethnic variety that both confused and impressed visitors and locals outside the community" (Jeppe 2001). There was a great deal of confusion over how to classify the various groups in the colony. Soon the Muslim-as-Malay came to be constructed against the Christian-as-coloured in official and dominant discourses in the nineteenth century. The arrival of new immigrants from South Asia (and Mauritius) in South Africa added another element to Cape Town's population. Among these migrants were many Muslims. Soon distinction emerged between the newly arrived Indian Muslims and the long-settled Muslims of the city, many of whom themselves had distant South Asian backgrounds (Jeppie 2001). The first so-called Passenger Indians arrived in the Cape in the 1880s; compatriots from the Natal colony joined them after 1897. The majority of them were Muslims and became traders in the city, living among the then existing community of Muslims. Distinctions (e.g. language) and controversies developed between the incoming and resident communities and this contributed to the distinction of Malay Muslims and Indian or Asiatic Muslims (Jeppie 2001). However, threats of 'Asiatic repatriation' and other hardships placed on them by the government, made 'Malay' a preferable term of identification (Jeppie 2001).

Erasmus (2001) argues that 'coloured' identities need to be viewed as "cultural identities comprising detailed bodies of knowledge, specific cultural practices, memories, rituals and modes of being. 'Coloured' identities were formed in the colonial encounter between colonists (Dutch and British), slaves from South and East India and from East Africa, and conquered indigenous peoples, the Khoi and San.....The result has been a highly specific and instantly recognizable cultural formation – not just 'a mixture' but a very particular 'mixture' comprising elements of Dutch, British, Malaysian, Khoi and other forms of African culture appropriately, translated and articulated in complex and subtle ways".

This is just a glimpse of the sociological discourse with respect to the nomenclature of 'Coloured' populations, and the issue is far from being resolved.

1.3 *Tracing the male ancestors of the South African "Coloured" population – the historical archive*

Most 'Coloured' people reside in the Western Cape (5.5%) where two major groups: Cape 'Coloured' and Cape Malay are identified based on self-proclaimed identity. In other parts of the country, they are referred to collectively as "Coloured" and constitute <3% of the total population size of the various provinces [Eastern Cape (1.08%), Northern Cape (0.96%), Gauteng (0.76%), KwaZulu-Natal (0.32%), Free State (0.19%), Northwest (0.13%), Mpumalanga (0.05%) and Limpopo (0.02%)] (Census 2001).

Given the complex history of the "Coloured" population, this study attempts to elucidate the geographic origins of the male contribution in three samples of "Coloureds" by using Y-chromosome markers. "Coloureds" from two regions in South Africa – Cape Town and Johannesburg were examined and for the purposes of this study, the Cape Malay was distinguished from the Cape 'Coloured' and the group sampled in Johannesburg, the latter referred to in this thesis as Johannesburg 'Coloured'. In another complementary study being conducted in Prof Soodyall's Unit, mitochondrial DNA (mtDNA) is being used to trace the origins of the female contribution in these 'Coloured' groups.

The history of 'Coloured' populations in South Africa is related to the history following European colonization of South Africa. In 1652 the Dutch East India Company (DEIC) together with their employees from various parts of Europe, especially Germany, and some slaves moved to the Cape Peninsula at the southernmost-tip of Africa (Figure 1.1), in order to establish a refreshment station for company ships bound for the East Indies (Watson 1990). They confronted the first 'natives' of the country, the Khoikhoi (Figure 1.1), with whom welcoming relations were established (Marais 1957; Shell 1994). This was the year in which the DEIC took permanent possession of the Cape peninsula, even though the Europeans had made transient visits before (Shell 1994).

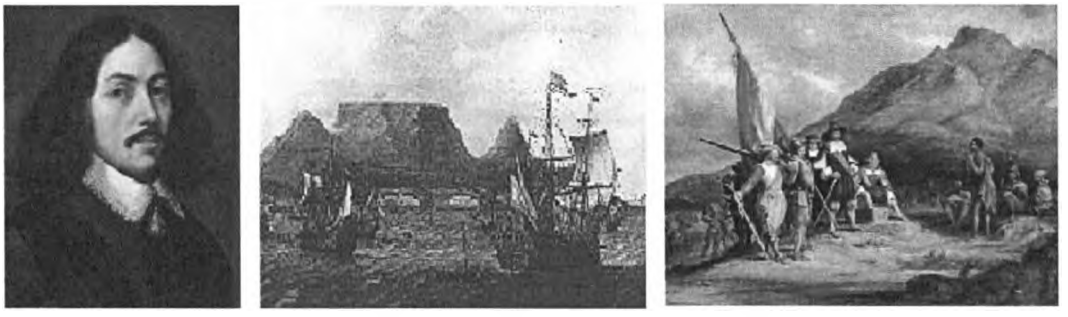


Figure 1.1: The Dutch East India Company's arrival and meeting with the Khoisan.

According to Nurse *et al.* (1985), other Europeans who visited the Cape came from Germany, Scotland, and Ireland. Furthermore, there were ancestors from Scandinavia, as well as the Baltic and Iberian regions. "Following the revocation of the edict of Nantes in 1685, there was a sudden influx of diverse French Huguenots in the colony" (Nurse *et al.* 1985). There were certainly Jewish people many of whom were Christianized, among the eighteen-century settlers. Some British settlers arrived during the later years of the Napoleonic wars, however the most significant influx took place in the 1820s. During the later half of the nineteenth century European immigrations became less organized and more sporadic, also more British, German and Dutch continued to arrive. The opening of the diamond mines at Kimberly, and subsequently the Gold mines of the Witwatersrand brought more Europe settlers and Caucasoids other than Afrikaner into the interior. Jews from Poland and Lithuania were the first occupants of Kimberly, and this immigration was enlarged by Pogroms from Russia. Middle Eastern Christians including the Lebanese, Syrians, Cypriots and a number of Copts and Armenians began to arrive at about the same time (Nurse *et al.* 1985).

When the DEIC colonists moved away from the Cape settlement to develop farms, their Dutch servants and Khoisan workers disappeared; as a result the DEIC experienced a labour shortage. Jan van Riebeeck recommended that slaves be introduced into the colony in order to condense labour shortages, nevertheless the Dutch government statutes forbade enslavement of the indigenous populations (Watson 1990). An urgent request for slaves was made to the Governor-General and Council of the Indies but the answer was not encouraging, as slaves were badly needed in Batavia (Böeseken 1977).

Prior to 1792 there was a great deal of rivalry between the Dutch West India Company (DWIC) and DEIC with respect to slave trade. The DWIC would not allow the DEIC to slave on the coasts of West Africa, arguing that Cape Town was technically on the west coast of Africa (Eldredge and Morton 1994; Shell 1994). Nevertheless, a small number of Bantu-speaking slaves were smuggled into the Cape from west and central West Africa (Figure 1.2). In 1658, 266 slaves from Dahomey and Angola were taken off two captured ships (Portuguese slaver bound for Brazil) and imported into the colony (Watson 1990; Shell 1994). Thereafter, West African slaves were clandestinely shipped from the DWIC territory by the DEIC from Cape Verde on outward-bound voyages (Eldredge and Morton 1994).

Subsequently the DEIC turned to territories east of the Cape for most of their slaves (Eldredge and Morton 1994; Shell 1994). Most imported African Cape slaves were drawn from areas including Madagascar, Mascarenes and Eastern Africa (Figure 1.2), especially Mozambique (Marais 1957). These were the prime sources of the Cape slave importation. As late as the 1870s on the periphery of the Transvaal, slave raids on the South African Bantu-speaking peoples were conducted (Eldredge and Morton 1994).

Other non-African slaves came from the different parts of the Far East: mainland of Indian, Ceylon, Malay Archipelago (Marais 1957), East Indies, Southeast Asia, Philippines, and Japan (Watson 1990). Moreover, some of the slaves were also drawn from the outer reaches of Borneo and the shores of China (Figure 1.2); these slaves included persons of Abyssinian, Arabian, Bengali, Borneose, Brazilian, Burmese, Chinese, Iranian, Japanese and Sri Lankan origin (Shell 1994).

The total number of slaves that were imported into the Cape from the Indonesian Archipelago, India, Madagascar, the Mascarenes, and Africa before the cessation of the oceanic slave trade, 1808, was estimated to 63,000. Slaves from Africa, the Indian subcontinent, Madagascar, and Indonesian archipelago constituted 26.4%, 25.9%, 25.1% and 22.7%, respectively (Eldredge and Morton 1994). By the year 1750, 79% of slave owners had ten slaves or fewer, and by the abolition of the slave trading, 1834, only 4.8% had more than twenty, and the largest slave holder owned eighty-four (Watson 1990). The African slaves brought to the Cape households did all the hard work in the vineyards, grain-lands, orchards and vegetable gardens of the colonists. The vast majority of slaves were held in the Cape Town and its agricultural hinterland. Many of the slaves from the East were Muslims who where for the most part highly skilled in various arts and crafts. Meat, course bread, salted fish, rice, and wine were supplied to the slaves (Marais 1957).

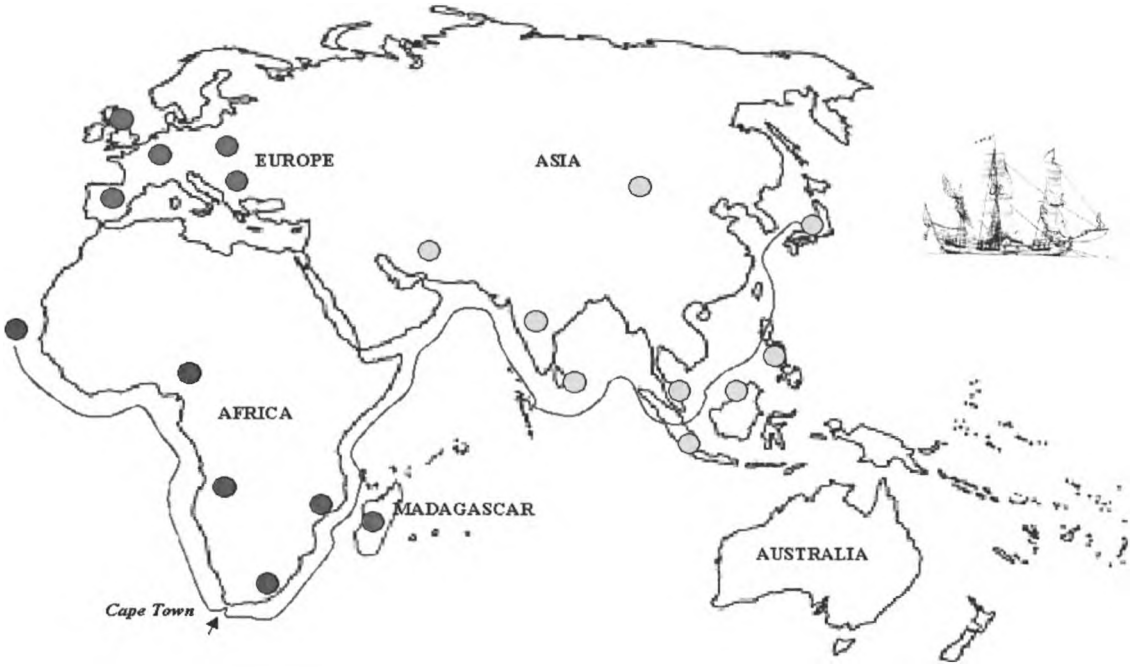


Figure 1.2: The slave trading routes to the Cape Peninsula. Different colour circles represent particular geographic locations from which slaves were drawn.

1.4 *Reconstructing ancestry: the Y-chromosome as a tool*

1.4.1 *Structural features of the Y-chromosome*

Genomic studies revealed that the Y-chromosome, the second smallest and a human sex chromosome, is ~60 megabases (Mb), represents about 2 - 3% of the haploid genome, and 95% of its entire length of Y-chromosome does not recombine. For that reason, the latter region came to be known as the non-recombining portion of Y-chromosome (NRY), also known as the male-specific region (Figure 1.3). This region consisting of euchromatic and heterochromatic regions, is flanked by the pseudoautosomal regions (PAR) at the terminal regions of both the short (Yp) and long (Yq) arms (Figure 1.3), were X-Y crossing over (recombination) event usually occurs frequently during male meiosis (Hammer and Zegura 1996; Skaletsky *et al.* 2003).

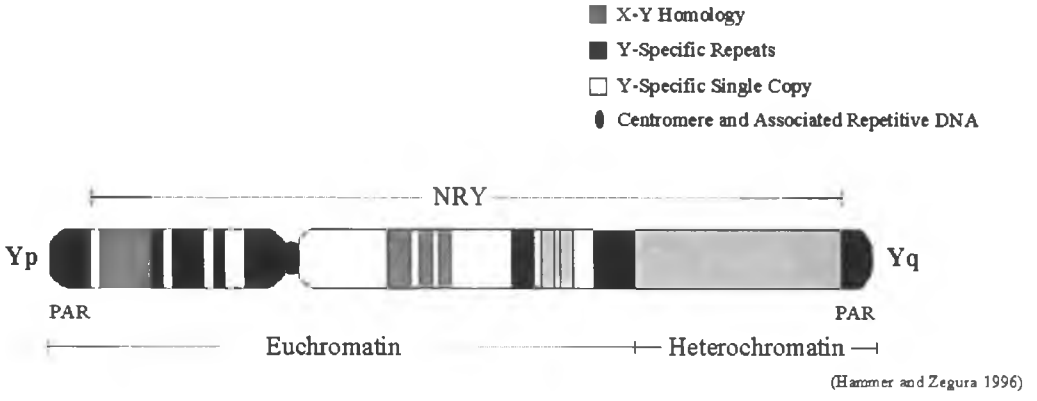


Figure 1.3: Schematic representation of the human Y-chromosome, including NRY and PARs.

In an effort to detect nucleotide sequence of the NRY from a single male individual, mapping and sequencing techniques were merged into a single analytic process. About 220 bacterial artificial chromosome (BAC) clones, each containing a portion of the NRY were used. A finished sequence containing 23 megabases (Mb) of euchromatic DNA (Figure 1.3) was obtained. Euchromatic sequences (X-transposed, X-degenerate and ampliconic) contained 156 transcriptional units, which include 78 protein-coding genes that collectively encode 27 distinct proteins. X-transposed is 99% identical to the X-chromosome, X-degenerate represents the remnants of ancient autosomes from which X and Y evolved, and ampliconic contains 30% MSY were sequence pairs exhibit >99.9% identity. Alternatively, complete sequences for 1Mb centromeric, 40Mb heterochromatic block on the distal long arm (Figure 1.3) as well as >3,000 tandem repeats of 125 base pairs (bp) in a third heterochromatic block were obtained (Skaletsky *et al.* 2003). These findings thus refute a postulated notion that the human Y-chromosome is a functional wasteland.

1.4.2 *Y-chromosome in population and evolutionary studies*

Several properties of the Y-chromosome make it a useful marker in population and evolutionary studies. First, its uniparental mode of inheritance through the patriline and haploid state make it possible to trace lineages found in present-day individuals to male ancestors who lived in the past. Second, in the absence of recombination, it is possible to construct the evolutionary histories of mutations found in contemporary divergent lineages to a common point in the past (Hammer 1994; Jobling *et al.* 1997; Scozzari *et al.* 1999; Underhill *et al.* 2001; Cruciani *et al.* 2002). Third, Y-chromosomes have a much smaller effective population size that is ¼ that of autosomal DNA, making it possible to discern subtle differences, for instance, founder effects in populations (Lell and Wallace 2000).

1.4.3 *Y-chromosome polymorphisms*

Genetic polymorphism has been defined as the existence of two or more alleles at significant frequencies in the population (Strachan and Read 1996). The well-known NRY-specific polymorphisms that are currently being employed to unravel the historical evolution of human populations as well as their interactive relationships include the most powerful genetic markers such as: a) the fast evolving hyper-variable minisatellites [(MSY1 variable number of tandem repeats (VNTR)], with an average mutation frequency of ~0.2% per generation, b) moderately evolving microsatellites or short tandem repeats (STRs), with a mutation frequency of ~6 - 11% per generation and c) the slow evolving biallelic or binary markers, with mutation rate in the order of 2-4 x 10^{-8} site per generation are unlikely to have arisen more than once in human evolution (Hammer and Zegura 1996; de Knijff 2000; Forster *et al.* 2000; Thomas *et al.* 2000; Hurles *et al.* 2002).

Very detailed descriptions of the Y-chromosome polymorphic markers employed in this particular study are outlined in the succeeding sections. In more recent years, a use has been found for these polymorphisms in DNA forensics (Jobling *et al.* 1997), genealogical reconstruction (Jobling 2001 as cited in the Y Chromosome Consortium, YCC 2002) medical genetics (Jobling and Tyler-Smith 2000) and human evolutionary studies (Hammer and Zegura 1996).

1.4.4 *Y-biallelic polymorphisms*

The slow evolving Y-chromosome biallelic polymorphisms include an *Alu* insertion polymorphism (YAP), single nucleotide polymorphisms (SNPs) as well as small insertions/deletions (indels). SNPs and indels occur at a frequency of ~1 in 1000bp throughout the genome are very stable and usually non-deleterious; hence they are been referred to as unique event polymorphisms (UEPs) that are able to partition Y-chromosomes into distinct geographic specific haplogroups (de Knijff 2000; Forster *et al.* 2000; Thomas *et al.* 2000; Shastry 2002). Before 1985 there were no known NRY-specific biallelic polymorphisms identified (Casanova *et al.* 1985). Until 1997, there were only 11 known polymerase chain reaction (PCR) based binary polymorphisms (Hammer 1994; Hammer and Horai 1995).

Underhill *et al.* (1997) discovered an additional 19 new PCR-based NRY specific biallelic polymorphisms by means of denaturing high-performance liquid chromatography (DHPLC) analysis. The use of DHPLC to detect more biallelic markers, further expanded the repertoire of these markers by introducing more than 200 additional SNPs and small indels (Rosser *et al.* 2000; Hammer *et al.* 2001; Underhill *et al.* 2000; 2001). Low mutation rate makes bialleic markers suitable for identifying stable paternal lineages that can be traced back in time over thousands of years, in an approach known as phylogeographic.

1.4.4.1 *Y-chromosome phylogeography*

Phylogeography is a systematic approach that is based on the geographic distribution of lineages to provide amazingly fascinating clues about the evolutionary history of human species (Avise *et al.* 1987, cited in Underhill *et al.* 2001). Such an approach has been previously used for mitochondrial DNA (mtDNA) networks (Richards *et al.* 1998; Kivisild *et al.*1999; Macaulay *et al.* 1999, cited in Underhill *et al.* 2001). Underhill *et al.* (2001) followed the phylogeographic approach to construct a detailed Y-chromosome phylogeography of modern global male individuals and past population movements and interactions (Figure 1.4) based on the information that for most human history, sons tended to live closer to their parents than daughters (Jobling and Tyler-Smith 2000). A sum of 218 biallelic polymorphisms were employed in the phylogeographic construction, in which ancestral allelic states were deduced from non-human primate sequences. The phylogeny is composed of 131 haplotypes (with mutations labelled with a prefix "M") that delineate the ten haplogroups (I-X) and exhibit more geographic specificity (Figure 1.4). Hence, the term 'haplogroup' was used to refer to NRY lineages defined by biallelic polymorphisms. This very detailed phylogenetic tree of Y-chromosome variation can be used to evaluate any particular population, since the distribution of its lineages is consistent with geographic origins.

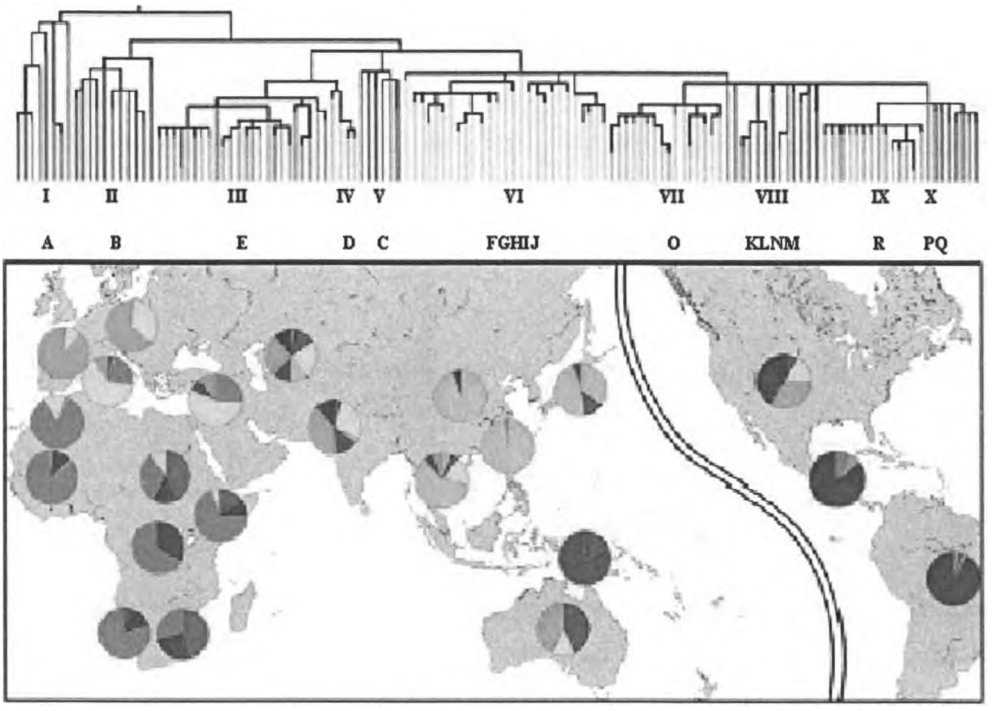


Figure 1.4: The distribution of geographic specific haplogroups in contemporary global populations, Underhill *et al.* (2001). Note: the author shall use the proposed YCC nomenclature to elucidate the global distribution of Y-chromosome haplogroups.

Due to the growing number of known binary polymorphisms there has been a corresponding increase in the number of different systems used to name these binary haplogroups, of which at least seven has been described. These different nomenclature systems made it difficult to compare results from one publication to the next. As a consequence, the YCC (2002) constructed a highly resolved phylogeographic tree of NRY biallelic haplogroups and described a new nomenclature system that utilizes the capital letters (A-R) to define major haplogroups (Figures 1.4). This principle is flexible enough to allow the anticipated changes that may result from the discovery of new polymorphisms and NRY lineages (Hammer and Zegura 2002; YCC 2002).

1.4.4.2 *Global distribution of Y-chromosome haplogroups*

In an analysis of 2007 males from 40 globally distributed populations, conducted by Hammer and Zegura (2002), the YCC-based phylogeny that comprised 237 NRY polymorphisms was modified into a truncated tree, composed of 245 polymorphisms and giving rise to 153 NRY haplogroups. The authors argued that not all the markers were variable in the YCC cell line panel and thus for a complete presentation of these markers additional samples were required to improve the resolution of the phylogeny. In their phylogeographic analysis, 20 out of 48 markers were genotyped to present the actual geographic distributions of Y-biallelic haplogroups (see Figure 1.5), since many were phylogenetically equivalent (define the same clade on the tree).

The distribution of Y-chromosome biallelic haplogroups were highly structured by geographic origins. The most divergent haplogroups A and B, placed on the deepest branch of the global phylogeny were found exclusively in African populations suggesting an African origin of human NRY diversity (Figures 1.4 and 1.5). Distinct patterns of haplogroup A lineages have been identified in the Khoisan, and among Ethiopians and Sudanese from southern and eastern Africa, respectively. Haplogroup B has a wide-ranging distribution in sub-Saharan Africans and shows some geographic structuring, B-M112 lineages are found mainly in the hunter-gatherer population and B-150 in the Bantu-speaking populations (Underhill *et al.* 2001; Semino *et al.* 2002; Cruciani *et al.* 2002; Hammer and Zegura 2002).

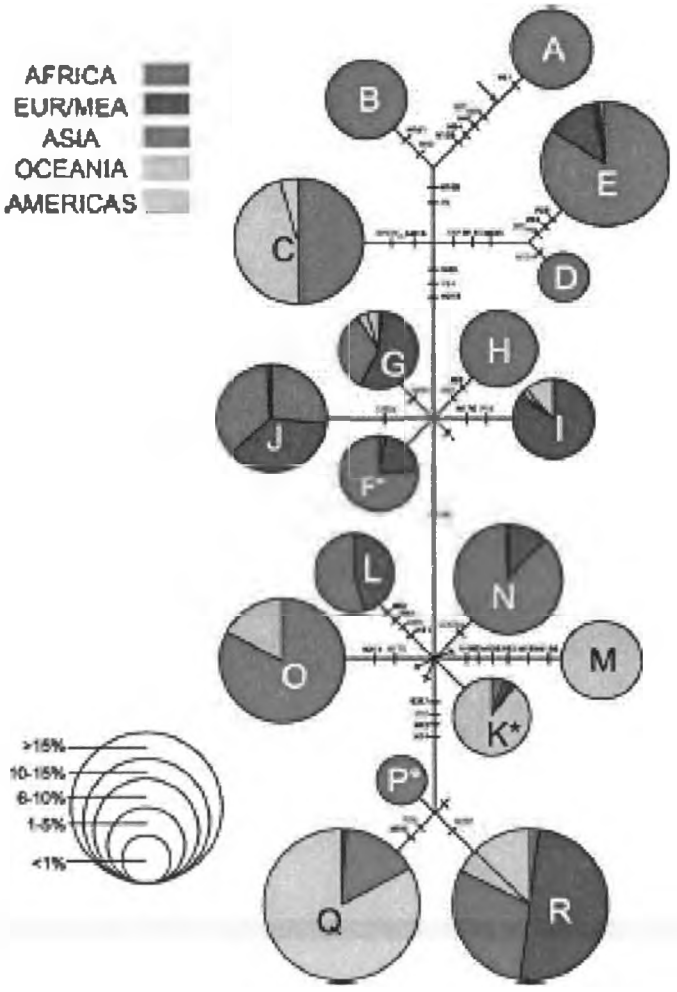


Figure 1.5: A phylogenetic tree for the 18 major NRY-specific biallelic haplogroups or paragroups identified in global populations, based on 20 markers (Hammer and Zegura 2002).

Haplogroup E is found at high frequency throughout Africa (Figures 1.4 and 1.5) and most likely it reflects the Bantu-agricultural expansion in the last 3000 years (Underhill *et al.* 2001). Haplogroup E is defined by the presence of a YAP, from which three main sub-lineages E-M2, E-M35 and E-M75 were derived (Spurdle *et al.* 1994; Scozzari *et al.* 1999; Underhill *et al.* 2001; Semino *et al.* 2002; Cruciani *et al.* 2002). E-M2 is found commonly in sub-Saharan and Central Africans (Underhill *et al.* 2001; Cruciani *et al.* 2002), whereas E-M35 is found mostly in north and east Africans, Mediterranean Islanders and Southern Europeans (Underhill *et al.* 2001;

Cruciani *et al.* 2002; Francalacci *et al.* 2003). E-M75 has been found mostly in the Bantu-speaking populations from East and South Africa (Underhill *et al.* 2001).

All of the non-African haplogroup distributions have been fully described in this section, also see Figure 1.5. Haplogroup C lineages were found at high frequencies in central Asians and Australo-Melanesia, but at a low frequency in the Northwest Americans, whereas haplogroup D lineages went as far as East Asia and hence have been confined to the Japanese and Tibetans, where they occur at high frequencies (Underhill *et al.* 2001; Capelli *et al.* 2001; Cavalli-Sforza and Feldmann 2003; Figure 1.4). Haplogroups G and J were commonly found in the Middle East and from there expanded to southern Europe probably with Neolithic farmers (Cavalli-Sforza and Feldmann 2003; Figure 1.5).

Haplogroups F, L and N were shared primarily between Asia and the Europe/Middle East region however, the highest frequency of haplogroup N has been observed in Russia (Puzyrev *et al.* 2003). The distribution of other haplogroups including H, K, M, O and P was more frequent among the Asian populations (Hammer and Zegura 2002; Figure 1.5). Together haplogroups H and L remained in India (Cavalli-Sforza and Feldmann 2003). Additionally, haplogroup H has been found predominantly in India, Pakistan and Sri Lanka (Qamar *et al.* 2002; Wells *et al.* 2001). Haplogroup O is exclusively Asian and has been found predominantly among Oceanic populations (Capelli *et al.* 2001; Kayser *et al.* 2003; Figure 1.5). Haplogroup Q has been found mostly in the Americas and with low frequencies among the Asian populations (Hammer and Zegura 2002). Finally, haplogroups I and R were commonly found in Europe (Figure 1.5; Semino *et al.* 2000). Haplogroup R-SRY10831.2 has been identified at higher frequency in Pakistan (Qamar *et al.* 2002; Cavalli-Sforza and Feldmann 2003).

1.4.5 *Y-short tandem repeats (Y-STRs)*

Y-STRs are the fairly evolving markers consisting of di-, tri-, or tetranucleotide repeat units. Y-STRs mutate through a step-wise mutation fashion (replication slippage) in which there is a change of one repeat unit in the array length (Hammer and Zegura 2002; YCC 2002). High mutation rates of the Y-STRs facilitate the reflection of more recent genealogical events (Nebel *et al.* 2000). Even though, Y-STRs tend to reveal higher levels of polymorphism, they cannot be treated by parsimony analysis instead, networks are used to resolve the variations within haplogroups. Thus, the term 'haplotype' is reserved for all sublineages of haplogroups that are defined by variation at STRs on the NRY. Y-STRS also allow separations of human populations to be more accurately measured and to give measure of time depth for the nodes of the biallelic phylogeny (Malaspina *et al.* 1998; Forster *et al.* 2000; Quintana-Murci *et al.* 2001). In this study Y-STRs have been utilized to resolve finer variations within haplogroups.

1.4.5.1 *Genealogical applications of Y-STRs*

Y-STRs are easier to find and cheaper to score than are biallelic markers, hence the analysis of Y-STR haplotypes with high degree of discrimination, has been extensively used for forensic applications, such as identification of male DNA in cases with male/female stain mixtures and paternity testing, especially for deficiency cases were alleged father is deceased (Quintana-Murci *et al.* 2001). Non-exclusion can only be problematic if a) all the male relatives of the suspect or alleged father are supposed to share the same haplotype and b) some haplotypes are over-represented depending on the particular demographic history of each population.

Another important genealogical application of the Y-STR is the association between closely related human populations (Hammer and Zegura 2002) or surnames and specific Y-STR haplotypes. Quintana-Murci *et al.* (2001) has given the Cohen modal haplotype (CMH), feature of the Cohanim (the hereditary priesthood), and the haplotype associated with the English surname 'Sykes' as two potential examples of identification of DNA samples. Therefore, forensic geneticists have joined together in a collaborative effort to produce standardised and highly reproducible protocols (http://www.stadnap.uni-mainz.de/s) and make available large-scale World Wide Web databases of allele and haplotype frequencies (Roewer *et al.* 2001, cited in Caglia *et al.* 2003).

1.5 *Objectives of this study*

The main objective of this study is to use Y-chromosome markers to elucidate the paternal ancestries of "Coloured" populations in South Africa. Both Y-biallelic and Y-STR markers were used to resolve Y-chromosomes into haplogroups and haplotypes, and then compared to other published and unpublished data. More specifically, Y- chromosome data were used to:

- Identify and estimate the various male parental contributions in the "Coloured" population,
- Examine the relationship between the Cape Malay and other "Coloured" groups sampled in Cape Town and Johannesburg to their presumed parental populations,
- Contribute in elucidating the genetic history of the "Coloured" population in South Africa.

CHAPTER 2

SUBJECTS AND METHODS

2.1 *Population samples*

The blood samples were collected from 264 healthy, paternally unrelated, random 'Coloured' and White male participants from South Africa, with the appropriate informed consent for use of their DNA in this study. This study was approved by the Ethics Committee for research on human subjects at the University of the Witwatersrand, Johannesburg (Ethical Protocol number: M020527; Appendix D). The 'Coloured' population consisted of the Cape 'Coloured' (N=48) and Cape Malay (N=54) groups sampled in the Western Cape (Figure 5), and were provided by Prof Paul van Helden and Dr Eileen Hoal (University of Stellenbosch). Both the 'Coloured' (N=65) and White (N=97) samples from Johannesburg in Gauteng (Figure 2.1) were provided by Prof Himla Soodyall (Ethical Protocol number M980553). Henceforth, the Cape and Johannesburg 'Coloured' groups would be referred to as Cape 'Coloured' and Johannesburg 'Coloured', respectively, to distinguish them from each other and from the Cape Malay. The author also wishes to emphasize that the name 'Coloured' is not meant to have any derogatory meaning and apologies to anyone who finds its use offensive.

2.1.1 *Comparative data sets*

The data collected from the present study were compared to data from other African, Asian and European parental populations (see Table 2.1).



Figure 2.1: Sampling sites for the South African 'Coloured' and White populations. Note all the samples from Gauteng and Western Cape provinces, were collected in the region of Johannesburg and Cape Town, respectively.

Table 2.1: Male parental population samples used in the comparative analysis with the South African 'Coloured' populations.

Population	Symbol	N	Ethnicity	Primary Reference
African samples:				
Nama	NAM	23	Khoikhoi	Soodyall (unpublished)
!Kung	KUN	35	San	Soodyall (unpublished)
African	SAB	127	Bantu-speaker	Soodyall (unpublished)
African	CAR	75	Bantu-speaker	Soodyall (unpublished)
Congolese	CON	67	Bantu-speaker	Soodyall (unpublished)
Cameroonian	CAM	48	Bantu-speaker	Cruciani *et al.* (2002)
Mozambican	MOZ	36	Bantu-speaker	Soodyall (unpublished)
Tanzanians	TAN	32	Bantu-speaker	Knight *et al.* (2003)
Indians	SAI	63	Indian	Soodyall (unpublished)
Whites	SAW	97	White	Present study
Asian samples:				
Chinese	CHI	36	Han Chinese	Kayser *et al.* (2003)
Malaysian	MAL	18	Malay	Kayser *et al.* (2003)
Philippines	PHI	39	Philippinos	Kayser *et al.* (2003)
Borneose	BOR	51	Kota Kinabalu	Capelli *et al.* (2001)
Indian	IND	44	Peruru Brahmins	Ramana *et al.* (2001)
Russian	YAK	47	Yakut	Puzyrev *et al.* (2003)
Asia Minor	ANA	81	Anatolian	Cinnioğlu *et al.* (2004)
European samples:				
Dutch	DUT	27	European	Semino *et al.* (2000)
French	FRE	23	European	Semino *et al.* (2000)
German	GER	16	European	Semino *et al.* (2000)
British	BRI	25	English	Wells *et al.* (2001)
Total		1012		

2.2 *Molecular methods*

All solutions used in this section of the thesis have been fully described in Appendix A.

2.2.2 *Mismatched base primer design*

The two polymorphic marker systems, M52 and M170, appeared to have no cut sites recognized by the available restriction endonucleases. Consequently, polymerase chain reaction and restriction fragment length polymorphism (PCR-RFLP) protocols were developed for each one of them (Tables 2.1 and 2.2). Using sequence information obtained from Underhill *et al.* (2000), reverse oligonucleotide primers (> 21-mer) complimentary to each strand were designed by introducing single mismatched bases, at least two bases away from the polymorphic site, to create the restriction endonuclease sites in the newly synthesized strands (Flores *et al.* 2003; Figure 2.2).

Melting temperatures [$\mathbf{T_m}$=4(G+C)+2(A+T)] of the mismatched reverse primers were employed in designing the usual forward primers, considering the total GC and AT contents that may lead to the formation of hairpins, the lengths of the preferred amplicons (usually < 200bp) and additional restriction sites within the primed sequence. An additional restriction site discovered within the M170 amplicon was abolished by introducing a mismatched base in the forward primer (Figure 2.2). Eppendorf Mastercycler gradient machine was used to work out the optimum annealing temperature for each system from a series of set PCRs. Then, resulting PCR products from the most optimum conditions were digested with the appropriate restriction endonucleases.



Figure 2.2: A schematic representation of mismatched oligonucleotide primer design. Note: the green typescripts denote the primer binding sites and the primers, whereas the mismatches and their positions are shown in red. Restriction sites are shown as blue boxes with the lines indicating where the enzyme cuts.

2.2.3 *Y-Biallelic polymorphisms typing*

All genomic DNA samples for the South African 'Coloured' and White populations were extracted from whole blood specimens by means of the standard salting out procedure (Miller *et al.* 1988). Subsequently all samples were quantified in a GeneQuant UV-visible spectrophotometer at a wavelength of 260 nm, and further diluted to a final concentration of 100 ng/µl suitable for any PCR amplification assay. Y-chromosome polymorphic markers that have previously been shown to delineate major haplogroups (A-R) found specifically in populations of African, Asian, European and American descent, were selected from the phylogeography of modern human populations (Underhill *et al.* 2001; YCC 2002; Hammer and Zegura 2002) based on the known published history of origins for the South African 'Coloured' populations. These PCR-based markers included a set of 21 NRY unique event biallelic polymorphisms consisting of an *Alu* insertion polymorphism (YAP), two insertion/deletion (*Indels*) as well as 18 single nucleotide polymorphisms (SNPs) (Table 2.2). Of the 18 SNP markers selected 11 were transition mutations, six were transversions and one was a reversion mutation (Table 2.2).

Table 2.2: NRY-specific unique event biallelic polymorphisms.

Marker	Nucleotide change	Polymorphism References
YAP (M1)	- →+Alu	Hammer (1994)
SRY10831	A → G, G → A	Whitfield *et al.* (1995)
M150	C → T	Underhill *et al.* (2000)
M40 (SRY4064)	G → A	Whitfield *et al.* (1995)
M2 (sY81)	A → G	Seielstad *et al.* (1994)
M35	G → C	Underhill *et al.* (2000)
M78	C → T	Underhill *et al.* (2000)
M75	G → A	Shen *et al.* (2000)
M168	C → T	Shen *et al.* (2000)
M130 (RPS4Y11)	C → T	Bergen *et al.* (1990)
M213	T → C	Shen *et al.* (2000)
M52	A → C	Underhill *et al.* (2000)
M170	A → C	Shen *et al.* (2000)
P12f2	88bpdel	Casanova *et al.* (1985)
M172	T → G	Shen *et al.* (2000)
M9	C → G	Underhill *et al.* (1997)
M175	TTCTCdel	Shen *et al.* (2000)
M11	A → G	Underhill *et al.* (1997)
M74	G → A	Shen *et al.* (2000)
M207	A → G	Shen *et al.* (2000)

Reversion, Transitions, and Transversions.

2.2.3.1 *PCR-RFLP analysis*

YAP and p12f2 biallelic markers (Table 2.2) were typed using PCR assays according to published protocols (Table 2.3). Thirteen other biallelic makers [M2, M9, M11, M40, M74, M78, M130, M168, M172, M175, M207, M213 and SRY10831 (Table 2.3)] were typed using PCR followed by restriction fragment length polymorphism (RFLP) analysis following published protocols given in Table 2.3. The remaining five-biallelic markers [M150, M35, M75, M52 and M170 (Table 2.2)], which did not have cut sites recognized by available restriction endonucleases were analyzed using standardized unpublished PCR-RFLP protocols (see section 2.2.2; Table 2.3).

PCR amplifications were carried out in either PE Biosystems GeneAmp® PCR System 9700 or Hybaid OmniGene thermal cyclers, in either 10 or 25µl reactions containing: autoclaved ultrafilted double distilled water (pH 7.0), 1X PCR Buffer (Roche), 1.5–3.5 mM $MgCl_2$ (Roche), 100-200µM of each dNTPs (dATP, dTTP, dGTP and dCTP nucleotides) (Invitrogen), 0.15-0.4 µM oligonucleotide primers (IDT® – Whitehead Scientific) see (Appendix B – Table 1), 0.02U/µl Taq DNA polymerase (Roche), and 100ng of genomic DNA template. Additional $MgCl_2$ and BSA (Sigma) (Appendix B – Table 1) were also used to improve the specificity of the PCR for some markers. For all the biallelic markers, cycling programs included an initial denaturation step at 95°C for 5-10minutes, the three-step standard thermal cycling profiles constituting one PCR cycle, and a final extension step at 72°C for 5-10minutes. Normally, the PCR conditions were 35 cycles of 30-45 seconds denaturing step at 94°C, annealing temperature (51 - 60°C) see (Table 2.3), and 72°C elongation step.

Table 2.3: The NRY locus specific biallelic polymorphisms and PCR-RFLP analysis conditions. Note: (*Forward and 'Reverse primers).

	PCR Condition				RFLP Condition		
Marker	Primer pairs (5' → 3')	Primer sizes (bp)	T (°C)	Primer References	Enzymes	Ancestral alleles	Derived alleles
YAP (M1)	*5'-caggggaagataaagaaata-3' '5'-actgctaaaaggggatggat -3'	20 20	51.0	Hammer and Horai (1995)	-	YAP- (150)	YAP+ (450)
SRY10831	*5'-tccttagcaaccattaatctgg-3' '5'-aaatagcaaaaaatgacacaaggc-3'	22 24	60.0	Santos *et al.* (1999)	Dra III (NEB)	A (167)	G (112+55)
M150	*5'-cccacacacacagatagacgt-3' '5'-cctactttccccctcttctg-3'	21 20	60.0	Nebel A (unpublished)	Aat II (NEB)	C (146+21)	T (167)
M40	*5'-ggtatgacaggggatgatgtga-3' '5'-ccacgcccagctaatttttg-3'	22 22	58.0	Thomas *et al.* (1999)	BsrB I (NEB)	G (135+90)	A (225)
M2	*5'-atgggagaagaacggaagga-3' '5'-tggaaaatacagctcccct-3'	20 20	58.0	Thomas *et al.* (1999)	Nla III (NEB)	A (105+43)	G (148)
M35	*5'-agggcatggtccctttctat-3' '5'-tgggttcaagtttccctgtc-3'	20 20	58.0	Nebel A (unpublished)	Bsr I (NEB)	G (122+64)	C (186)
M78	*5'-cttcaggcattatttttttggt-3' '5'-atagtgttccttcacctttcct-3'	23 22	63.0	Underhill *et al.* (2000)	Aci I (NEB)	C (196+105)	T (301)
M75	*5'-cacatcaagaaaacttgcaagac-3' '5'-tataaaggtatttctctgtatataca-3'	23 26	55.0	Nebel A (unpublished)	Nla III (NEB)	G (189)	A (165+24)
M168	*5'-agtttgaggtagaatactgtttgc t-3' '5'-aatctcataggtctctgactgttc-3'	25 24	56.0	Underhill *et al.* (2000)	Hinf I (NEB)	C (234+105+81+52)	T (234+186+52)
M130	*5'-ctgtacttacttttatctcctc-3' '5'-cagcaacagtaagtcgaatg-3'	22 20	50.0	Kayser *et al.* (2000)	Bsl I (NEB)	C (57+34)	T (91)
M213	*5'-tataatcaagttaccaattactggc-3' '5'-ttttgtaacattgaatggcaaa-3'	25 22	56.0	Underhill *et al.* (2001)	Nla III (NEB)	T (290+119)	C (409)
M52	*5'-gaaatggagtcatgaactcct-3' '5'-atatcaagaaacctatcaaacatac-3'	21 25	60.0	Motladiile TW (unpublished)	HpyCH4IV (NEB)	A (164)	C (138+26)
M170	*5'-aaataatttcacgtttgttcaaataa-3' '5'-acacaacccacactgaaaaaca-3'	26 22	59.0	Motladiile TW (unpublished)	Nla III (NEB)	A (109+20)	C (129)
p12f2	*5'-ctgactgatcaaaatgcttacagatc-3' '5'-ggatcccttccttacaccttatac-3'	26 24	58.0	Rosser *et al.* (2000)	-	no del (+88bp)	del (-88bp)
M172	*5'-atcccccaaacccatttttgatgcat-3' '5'-ggatccatctccatcttcactcaatgttg-3'	25 23	58.0	Nebel *et al.* (2001)	Nla III (NEB)	T (148)	G (122+26)
M9	*5'-gcagcatataaaactttcagg-3' '5'-aaaacctaactttgctcaagc-3'	21 21	54.0	Underhill *et al.* (1997)	Hinf I (NEB)	C (181+95+64)	G (245+95)
M175	*5'-ttgagcaagaaaaatagtaccca-3' '5'-ctccattcttaactatctcaggga-3'	23 24	60.0	Underhill *et al.* (2000)	Mbo II (Amersham)	no del (370+74)	del (444)
M11	*5'-ccctccctctctccttgtattctacc-3' '5'-ttcatcacaaggagcataaacaa-3'	26 23	58.0	Qamar *et al.* (2002)	Msp I (NEB)	A (215)	G (193+22)
M74	*5'- atgctataataactaggtgttgaag-3' '5'-aattcagcttttaccacttctgaa-3'	25 24	60.0	Underhill *et al.* (2000)	Rsa I (NEB)	G (385)	A (195+190)
M207	*5'-aggaaaaatcagaagtatccctg-3' '5'-caaaattcaccaagaatccttg-3'	23 22	56.0	Underhill *et al.* (2001)	Dra I (Roche)	A (356+77)	G (423)

Specific genotypes for each individual in a total sample were detected from digesting the whole volume of PCR products (either 10 or 25µl) with 0.35 – 0.5U of the appropriate restriction enzyme, according to the instructions of the New England Biolabs (NEB), Amersham and Roche manufacturers (Table 2.3). RFLP patterns were resolved by electrophoresis on either 2% or 3% agarose depending on the sizes (bp) of the PCR products. The negatively charged DNA fragments migrated towards the anode, against ethidium bromide that intercalated between stacked base pairs of the DNA, thus making it fluoresce when visualized for imaging under UV light on the SYNGENE-GeneSnap, transilluminator documentation system. For each marker typed, a molecular weight marker (Invitrogen), a negative control and two positive controls (cut and uncut) were loaded on each gel to permit the sizing of fragments, check for contamination in the PCR-RFLP typing, and verify the success of the PCR-RFLP assay, respectively. Finally, the resulting individual fingerprints (RFLPs) were scored of which the derived allelic states were used to define respective haplogroups for individuals within particular populations.

2.2.3.2 Typing strategy

A hierarchical typing strategy was adopted in characterizing all the samples in the present study. In this top-down approach (Bosch *et al.* 2001), YAP was typed on all samples first to resolve them into either YAP- (absence) or YAP+ (presence). This was followed by typing for M9 on all the YAP- chromosomes, while the marker M40 was typed on the YAP+ chromosomes (Figure 2.3). Then derived allelic forms for other biallelic markers were resolved in a hierarchical approach that is based on the known Y-chromosome phylogeny (Figure 2.3; Underhill *et al.* 2001; YCC 2002), to define respective haplogroups for each of the samples.

Opportunity for doing multiplexing of biallelic markers was not possible, as the PCR conditions (Table 2.3) were variable for each marker. In the case of similar screening conditions, some of the samples would have been needlessly screened for other markers and thus interfering with the proposed hierarchical typing strategy.

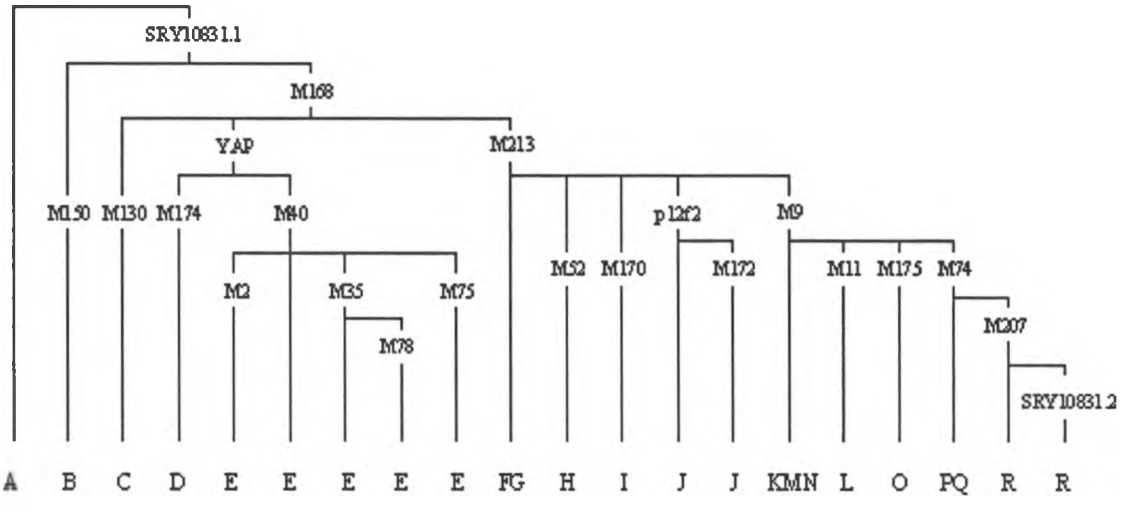


Figure 2.3: Maximum parsimony phylogenetic tree showing all the haplogroup defining polymorphisms included for genotyping in the present study. Each polymorphism defines corresponding haplogroup lineage below it.

2.2.4 *Terminology and Nomenclature*

We used the terms 'haplogroups' and 'haplotypes' according to de Knijff (2000). Haplogroups/Clades are NRY lineages defined by biallelic polymorphisms, and haplotypes are sublineages of haplogroups that are defined by Y-STR variation. We adopted the conventional nomenclature system proposed by the YCC (2002) for naming biallelic haplogroups, A-R. This alphabetical designation identifies the 18 major Y-chromosome haplogroups.

2.2.5 *Y-STR polymorphism typing*

2.2.5.1 *Multiplex PCR*

All DNA samples were screened for six tetranucleotide STRs: DYS19, DYS389I, DYS389II, DYS390, DYS391 and DYS393, and two trinucleotide STRs: DYS388 and DYS392 (Table 2). The eight STR loci were PCR amplified with a pair of primers that included one fluorescently labelled and unlabelled primer (Table 2.4), in two separate multiplex PCR reactions. Initially, The four loci DYS19, DYS390, DYS391 and DYS393 were co-amplified in the first multiplex according to the method suggested by Lane *et al.* (2002), and subsequently the other loci DYS388, DYS389I, DYS389II and DYSS392 were co-amplified in the second multiplex with a slightly different method (Soodyall *et al.* 2003).

The two multiplex PCR reactions were performed in 25µl final volume containing: autoclaved ultrafilted double distilled water (pH 7.0), 1X PCR Buffer (Roche), 250µM of each dNTPs (dATP, dTTP, dGTP and dCTP nucleotides), 0.08 - 0.16 µM primers (IDT® - Integrated DNA technologies, inc.), 1mM $MgCl_2$ (Roche), and 1mg/ml BSA (Sigma) (Appendix B – Table 2), 0.5U Roche-Taq DNA polymerase (5U/µl), 50ng genomic DNA template, and, by means of a Hybaid OmniGene thermal cycler. The first multiplex PCR was amplified by means of 30 seconds denaturing steps at 95°C, annealing temperature of 54°C (Table 2) and elongation steps of 72°C in 28 cycles. For the second multiplex PCR, annealing temperature and the number of cycles were all increased by a factor of 2. An initial denaturing step at 95°C for 2 minutes was included in both protocols. To distinguish between PCR fragments with the same label and size, products of each multiplex reaction were run on separate gels.

Table 2.4: Y-STR oligonucleotide primers and PCR amplification conditions.

Marker (Labeling)	Repeat motif	Primer pairs	Primer size (bp)	~T (°C)	Primer References
DYS19 (Blue)	(GATA)n	5'- FAM - ctactgagtttctgttatagt -3' 5' - atggcatgtagtgaggaca -3'	21 19	54	Roewer *et al.* (1992)
DYS388 (Blue)	(ATT)n	5'- FAM – gtgagttagccgtttagcga -3' 5'- cagatcgcaaccactgcg -3'	20 18	56	Jobling and Tyler-Smith (1995)
DYS389 (Blue)	(GATA)n (GACA)n	5'- FAM - ccaactctcatctgtattatctat -3' 5'- tcttatctccacccaccaga -3'	24 20	56	Jobling and Tyler-Smith (1995)
DYS390 (Green)	(GATA)n (GACA)n	5'- HEX – tatattttacacatttttgggcc -3' 5'- tgacagtaaaatgaacacattgc -3'	23 23	54	Jobling and Tyler-Smith (1995)
DYS391 (Green)	(GATA)n	5'- HEX - ctattcattcaatcatacaccca -3' 5'- gattctttgtggtgggtctg -3'	23 20	54	Jobling and Tyler-Smith (1995)
DYS392 (Yellow)	(ATT)n	5'- TAMRA - tcattaatctagcttttaaaaacaa -3' 5' - agacccagttgatgaaatgt -3'	25 20	56	Jobling and Tyler-Smith (1995)
DYS393 (Green)	(GATA)n	5'- HEX - gtggtcttctacttgtgtcaatac -3' 5'- aactcaagtccaaaaaatgagg-3'	24 22	54	Jobling and Tyler-Smith (1995)

Note: Only forward primers are fluorescently labelled and ~T denotes primer-annealing temperature.

2.2.5.2 *Polyacrylamide gel electrophoresis*

The fluorescently labeled PCR products including a negative control and two positive controls were mixed with GeneScan™ - 500 ROX™ size standard (Applied Biosystems) and resolved on 4% polyacrylamide vertical gel electrophoresis in an ABI 377® Prism automated DNA sequencer. The fragment sizes were estimated using previously sequenced positive controls of known repeat numbers with the aid of GeneScan version 3.1.2 and Genotyper version 2.5 software packages (PE Applied Biosystems).

Allele sizes for DYS389IIc were obtained by subtraction of the DYS389I allele length from DYS389II. The repeats were named according to the nomenclature proposed by Kayser *et al.* (1997). Y-STR haplotypes were constructed using the number of repeats in the order: DYS19-DYS388-DYS389I-DYS389II-DYS390-DYS391- DYS392-DYS393.

2.2.6 *DNA sequencing analysis*

At the DYS19 and DYS388 loci, two individuals presenting with unusual repeat sizes of 10 and 9 repeats, respectively were sequenced to ascertain these number of repeats using an ABI 377® Prism automated DNA sequencer. The two samples were PCR amplified following the abovementioned Y-chromosome STR protocols. PCR products were run on a 2% NuSieve® CTG® agarose (BMA). The bands were then cut out purified according to the manufacturer's recommendations using nucleospin extraction column kit (Macherey Nagel). The quality and the yield of the purified DNA were estimated on 2% agarose gel.

2.2.6.1 *Cycle sequencing PCR*

The purified DNA together with 4µl of Big-Dye terminator kit (Applied Biosystems), 3.3 µM of untagged reverse primer were cycle sequenced. Cycle sequencing involves, 25 cycles of denaturing at 96°C for 30sec, primer annealing at 50°C for 15sec and chain elongation at 60C° for 4min. unincorporated dye terminators were removed from the sequencing products using spin™ purification columns (Sigma) according to the manufacturers' instructions. Subsequently, the two samples were dried in a vacuum centrifuge. The dried products were then mixed with dextran-formamide loading buffer and denaturing of the double stranded DNA was performed at 100°C for 2min on a heating block prior to loading. Samples were then loaded on a 0.2mm thick 4% denaturing polyacrilamide gel and run by means of electrophoresis for 7 hours in an ABI 377® Prism automated DNA sequencer. All sequences were analyzed using ABI 377® Prism software.

2.2.7 *Statistical analysis*

Direct counting method was employed to determine the frequencies of Y-chromosome biallelic haplogroups (HG) and STR haplotypes (ht) presented in Figure 3.1 as well as Appendix C, respectively. Y-biallelic phylogeny of HG relationships was constructed. Haplogroup and haplotype frequency data sets were used as described in the following sub-sections to evaluate the state of genetic variation within and between populations.

2.2.7.1 *Intrapopulation variation analysis*

Y-chromosome variation within each population was inferred from biallelic-HG and STR-ht frequency data, by estimating genetic diversity *(h)* and its sampling variance V*(h)* were calculated from unbiased formulae given in Nei (1987), using ARLEQUIN version 1.1 software package (Schneider *et al.* 1997). Diversity within each haplogroup and the mean number of pair-wise differences that could not be estimated at once using ARLEQUIN 1.1, were estimated using the combined biallelic-HG and STR-ht data on the advanced ARLEQUIN version 2.0 (Schneider *et al.* 2000).

2.2.7.2 *Interpopulation variation analysis*

For the combined UEP-HG and STR-ht data, significant population differentiation estimated from 10, 000 permutations of Markov Chain steps, was inferred from an exact test of Rousset and Raymond (1995) included in ARLEQUIN version 1.1. Population pair-wise F_{ST} genetic distances and associated probability (*P*) values based on HG frequency data were computed using ARLEQUIN version 1.1. Then the relationships of HGs were inferred from constructing population trees using relevant programs in MEGA version 2 (Kumar *et al.* 2001). Unrooted neighbour-joining (NJ) tree for 21 populations including the 'Coloureds' and the presumed parental populations was constructed from the matrix of Slatkin linearised pair-wise F_{ST} distances. On the other hand, unweighted pair group method with arithmetic mean (UPGMA) tree for the three South African 'Coloured' and Whites populations was constructed from the matrix of Slatkin linearised pair-wise F_{ST} distances.

2.2.7.2.1 *Network analysis*

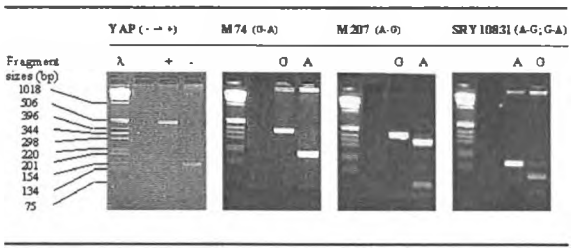
The relationships of STR-haplotypes within each UEP-haplogroups were also deduced from constructing networks using NETWORK versions 3.0 and 4.0. The reduced median algorithm (r = 1) and median joining network (ϵ = 0) were applied sequentially to generate an input file for the network (Bandelt *et al.* 1999).

CHAPTER 3

RESULTS

3.1 *Assignment of alleles to haplogroups*

The allelic states (derived or ancestral) for each biallelic marker were determined for each individual sampled. These allelic states were determined following agarose gel electrophoresis of either digested or undigested PCR fragments (Figure 3.1a) using the size standard (1Kb ladder) and known controls from whom the product sizes were confirmed by DNA sequencing. Only the derived allelic forms determined the path towards the eventual haplogroups. For instance, (YAP-) individuals whom were derived when typed for the polymorphisms M74, M207 and SRY10831 (Figures 3.1; Table 2.3), where not all assigned to haplogroup R-SRY10831.2, since some of the individuals whom were for ancestral with SRY10831 turned out to be falling in haplogroup R-M207. Moreover, those whom were ancestral for R-M207, by default belonged to haplogroup PQ-M74.

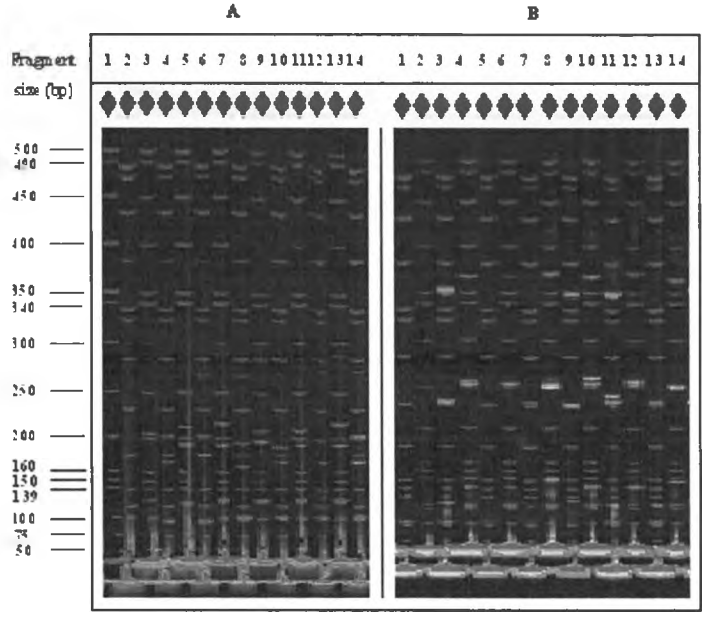


Note: Only four of the 20 polymorphisms included in this study have been depicted in the figure. Information regarding the fragment sizes as well as allelic states is given in Table 2.3.

Figure 3.1a: Images of the agarose gels showing genotypes for some of the NRY specific polymorphisms included in this study.

3.2 *Haplotyping Y-chromosome STRs*

The two multiplexed Y-STR PCR products were resolved separately on a 36 well polyacrylamide gel through the process of electrophoresis on an ABI 377® Prism automated DNA sequencer with the aid of the GeneScan filter set A matrix (Figure 3.1b). The matrix selected helped in the analysis of the resulting non-overlapping and fluorescently labelled (Table 2.4) PCR fragments, which were scored in the presence of known sequenced controls and ROX marker and subsequently converted to repeats sizes given in Appendix B. For the most part, the scoring of the alleles was consistently unambiguous, except for the few runs that showed some slight variations. Hence, a binning was constructed in order to maintain the scoring.



Note: A comprise of the first multiplex and B the second multiplex including the polymorphic markers DYS19, DYS390, DYS391 and DYS393, and DYS388, DYS389I, DYS389II and DYSS392, respectively. The flourephores labelling has been described in Table 2.4. Lane 1 = ROX 500 marker, lane 2 = negative control, lanes 3 and 4 = positive controls, and lanes 5-14 = some of the samples examined in this study.

Figure 3.1b: An electrophoregram showing the non-overlapping ranges of the Y-chromosome STR alleles labelled with four different types of flourephores.

3.3 *Intrapopulation variation*

3.3.1 *Y-biallelic haplogroup variation*

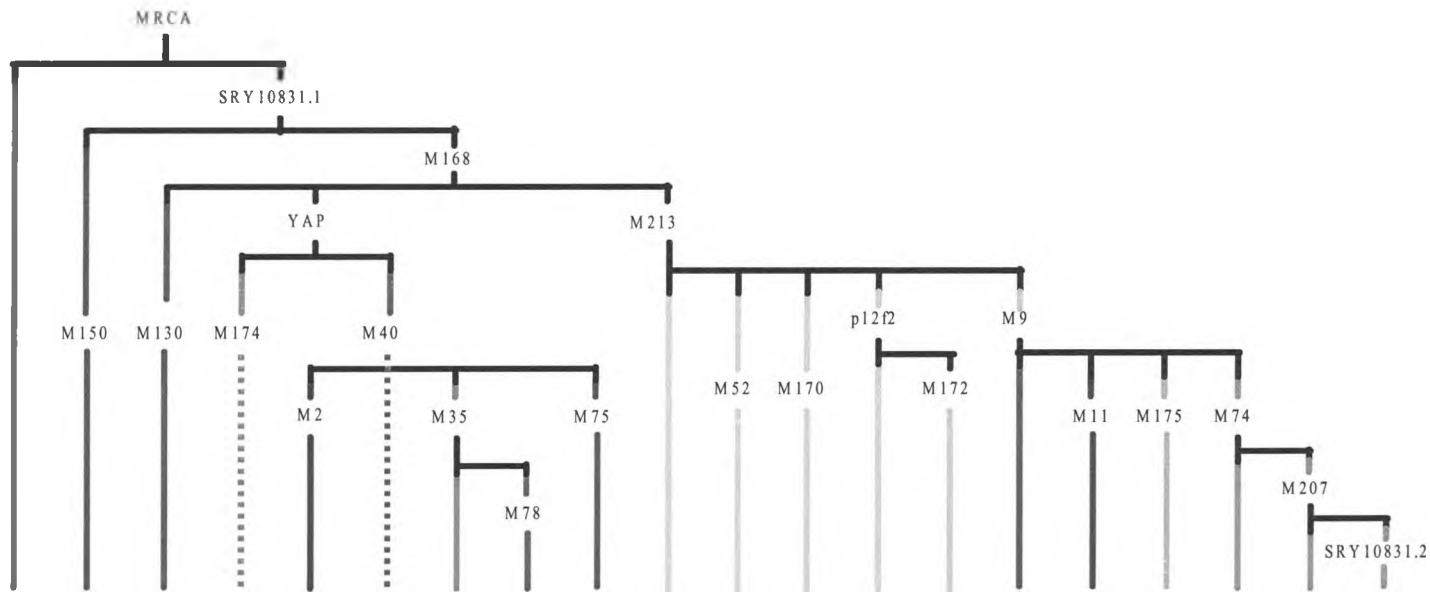
It was possible to resolve the Y-chromosome lineages in the South African (SA) 'Coloured' (N = 167) and SA White (N = 97) populations using 20 of the 21 NRY biallelic polymorphisms used in this study (Figure 3.2). Using the nomenclature suggested by the YCC (2002) for haplogroup assignment, 18 unique haplogroups were resolved in the total sample of 264 individuals. The distribution and frequencies of haplogroups in 'Coloureds' from Johannesburg and the Cape as well as the Cape Malay and SA White populations are shown in Figure 3.2. Haplogroups E-M40 and D-M174, shown as broken lines in Figure 3.2, were not found in our sample.

3.3.2 *Y-STR haplotype variation*

The inclusion of eight Y-chromosome specific STR markers resolved Y-chromosomes in the three SA 'Coloured' and the SA White populations defined at the haplogroup level into 185 unique haplotypes of which 150 were singletons (Appendix C). The number of haplotypes found in the SA Whites, Johannesburg 'Coloureds', Cape Malays and Cape 'Coloureds' were 66, 58, 50 and 42, respectively. Only 22 of these haplotypes were randomly shared among the SA 'Coloured' and SA White populations (Appendix C).

The frequencies of alleles (number of repeats) found at each Y-STR locus are shown in Figure 3.3. With the exception of DYS392, the distributions of number of repeats at all the loci are indicative of unimodal distributions, with less frequent alleles differing from the most frequent alleles by single repeat units. The highest frequent alleles were commonest in the SA White population, except at the DYS391 locus, and were found at varying frequencies in the three SA 'Coloured' groups. The highest peak around 14, 16, and 13 repeat alleles at DYS19, DYS389II and DYS392 including DYS393 loci, respectively, have been also identified at high frequencies in Europeans compared to the Indians and Sri Lankan populations as reported by Jobling *et al.* (1997). The most frequent allele around 10 repeat at DYS391 found in the three SA 'Coloured' populations, was also found at high frequency in the Central Africans. Furthermore, alleles with the highest frequencies around 13 and 24 repeats at the loci DYS389I and DYS390, respectively, have been found to be more frequent in Central Africans and Asians (Papua New Guineans), in that order (Kayser *et al.* 2001; Figure 3.3). DYS392 has a bimodal distribution of alleles around 11 and 13 repeats. The 13 repeats were found to be associated with the European populations (Jobling *et al.* 1997), while 11 repeats were associated mostly with the Asian population - Tolai New Britain (Kayser *et al.* 2001).

One allele, consistent with 10 repeats at DYS19 locus, was found in one individual with a Cape Malay ancestry. In addition, one allele consistent with 9 repeats at DYS388 locus, was found in one individual from the Johannesburg 'Coloured' group (Figure 3.3). These allelic changes occurred as a result of stepwise mutations or replication slippage favouring very small changes in the total array length (Hammer and Zegura 2002).



	Number of individuals within haplogroups																			
Haplogroup:	A*	B	C	D	E	E	E	E	E	FG	H	I	J	J	KMN	L	O	PQ	R	R
POPULATION																				
CC (N = 48)	2	1	1	...	11	...	...	1	1	6	2	2	...	3	1	...	6	...	9	2
CM (N = 54)	...	...	5	...	...	...	...	2	...	9	1	1	...	6	8	4	7	...	7	4
JC (N = 65)	...	4	3	...	9	...	4	2	2	4	2	3	...	5	1	1	3	1	18	3
RW (97)	...	...	...	...	1	...	1	2	...	4	...	13	1	2	1	...	...	1	64	7

The population names are shortened into two letter codes: CC = Cape 'Coloured', CM = Cape Malay, JC = Johannesburg 'Coloured', and RW = Random Whites.
N = number of individuals.

Figure 3.2: Phylogenetic tree showing haplogroup composition in three South African 'Coloured' and the White populations. Dotted lines represent haplogroups not detected in a sample.*

* = Y-chromosomes within haplogroup/clade A that have not been resolved further using other polymorphisms into sub-lineages.

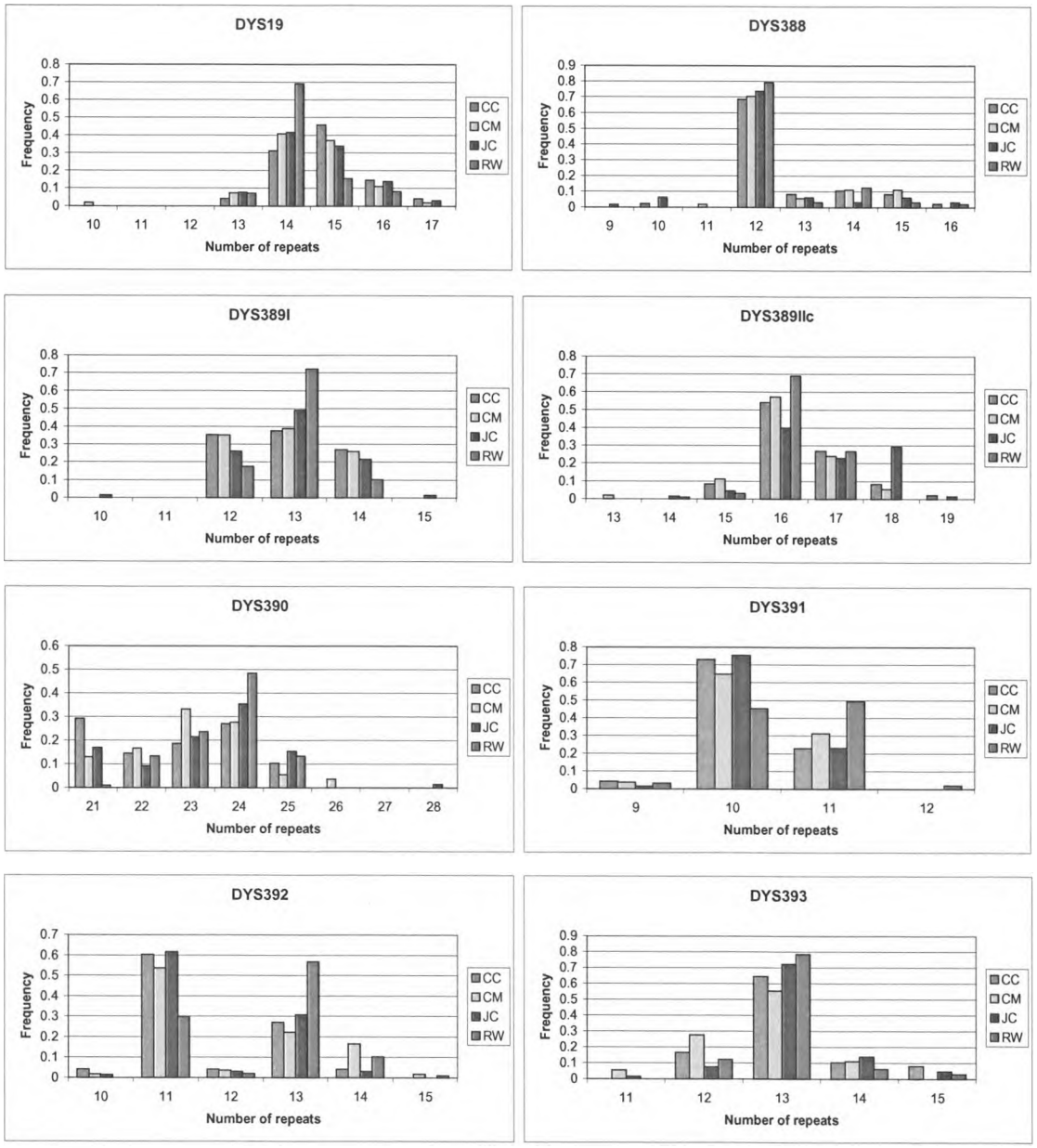


Figure 3.3: Allelic variation at eight Y-STR loci in the three SA 'Coloured' groups and the SA White populations. Note: population names are also abbreviated in two letter codes as in Figure 3.2.

3.3.3 *Y-chromosome diversity in the 'Coloured' population*

3.3.3.1 *Y-haplogroup and haplotype diversity*

The standard indices of Y-chromosome diversity are summarized in Table 3.1. Biallelic haplogroup diversity was high (h>0.887) in the three SA 'Coloured' populations, but was lower in the SA Whites (h=0.544; Table 3.1). The same trend was observed using Y-STR haplotype diversity; h>0.993 in the three SA 'Coloured' populations and (h=0.981) in the SA Whites (Table 3.1). In general, a high degree of genetic variation was observed in all three SA 'Coloured' groups.

Table 3.1: Y-biallelic and STR polymorphism diversity in the SA 'Coloured' and SA White populations.

POPULATION	N	$^{a}\eta$	Haplogroup Diversity ± SD	$^{b}\kappa$	Haplotype Diversity ± SD
Cape 'Coloured'	48	11	0.887 ± 0.024	42	0.993 ± 0.006
Cape Malay	54	14	0.899 ± 0.014	50	0.997 ± 0.005
Johannesburg 'Coloured'	65	16	0.889 ± 0.026	58	0.996 ± 0.004
Whites	97	11	0.544 ± 0.057	66	0.981 ± 0.007

N = number of individuals, $^{a}\eta$ = number of haplogroups, $^{b}\kappa$ = number of haplotypes, and SD = variance.

3.3.3.2 *Y-lineage diversity*

Due to similar diversity within each SA 'Coloured' population (Table 3.1), all three groups were pooled and Y-chromosome lineage (combination of haplogroup and haplotype data) diversity and the mean number of pair-wise differences in the SA 'Coloured' and SA White populations were estimated and the results are summarized in Table 3.2. In general, higher lineage diversity values and mean number of pair-wise differences were observed in the SA 'Coloured' populations compared with the SA White population. However, the diversity within FG-M213 lineage was higher in the SA Whites compared with SA 'Coloureds', but no diversity was detected at the PQ-M174 locus in both groups since they occurred as singletons in each group (Table 3.2).

Table 3.2: Y-chromosome lineage diversity within the South African 'Coloured' population, based on the combined haplogroup-haplotype data.

	'Coloured'			White		
Haplogroup	κ	Haplotype Diversity ± SD	π ± SD	κ	Haplotype Diversity ± SD	π ± SD
A*	2	0.375 ± 0.433	3.000 ± 2.449	...	...	...
B-M150	4	0.138 ± 0.125	1.100 ± 0.853	...	...	...
C-M130	9	0.579 ± 0.356	4.639 ± 2.510	...	...	...
E-M2	13	0.347 ± 0.214	2.779 ± 1.534	1	0.000 ± 0.000	0.000 ± 0.000
E-M35	4	0.542 ± 0.403	4.333 ± 2.701	1	0.000 ± 0.000	0.000 ± 0.000
E-M78	5	0.288 ± 0.219	2.300 ± 1.504	2	0.250 ± 0.306	2.000 ± 1.732
E-M75	2	0.083 ± 0.104	0.667 ± 0.667	...	...	...
FG-M213	15	0.499 ± 0.291	3.988 ± 2.086	4	0.583 ± 0.431	4.667 ± 2.885
H-M52	5	0.325 ± 0.243	2.600 ± 1.663	...	...	...
I-M170	6	0.550 ± 0.365	4.400 ± 2.526	11	0.436 ± 0.267	3.487 ± 1.899
J-p12f2	...	...	...	1	0.000 ± 0.000	0.000 ± 0.000
J-M172	13	0.468 ± 0.282	3.747 ± 2.010	2	0.250 ± 0.306	2.000 ± 1.732
KMN-M9	9	0.650 ± 0.389	5.200 ± 2.749	1	0.000 ± 0.000	0.000 ± 0.000
L-M11	4	0.250 ± 0.196	2.000 ± 1.343	...	...	...
O-M175	13	0.510 ± 0.301	4.083 ± 2.148	...	...	...
PQ-M74	1	0.000 ± 0.000	0.000 ± 0.000	1	0.000 ± 0.000	0.000 ± 0.000
R-M207	22	0.348 ± 0.209	2.786 ± 1.510	37	0.297 ± 0.182	2.375 ± 1.311
R-SRY10831.b	8	0.365 ± 0.239	2.917 ± 1.686	5	0.298 ± 0.210	2.381 ± 1.468
TOTAL	135			66		

κ = number of haplotypes within haplogroups, π = mean number of pair-wise difference and SD = variance.

3.4 *Interpopulation variation*

3.4.1 *Population differentiation*

Fisher's exact test of population pair-wise differentiation based on the combined haplogroup-haplotype frequencies was used to assess the level of difference between the SA 'Coloured' populations (Table 3.3). The test showed no significant difference between the Cape and Johannesburg 'Coloured' groups ($0.01<P>0.05$), but a significant difference was observed between the two 'Coloured' groups and the Cape Malays ($0.01>P<0.05$) (see Table 3.3).

Table 3.3: Differentiation of the 'Coloured' population pairs at 0.01 and 0.05 levels of significance.

Population	Cape Coloured	Cape Malays
Cape Malays	0.000 ± 0.000 (+)	...
Johannesburg 'Coloured'	0.589 ± 0.036 (-)	0.000 ± 0.000 (+)

Note: Only significant values at the 0.01 level are shown in the table.

(+) = Significant difference, (-) = No significant difference.

Population	Cape Coloured	Cape Malays
Cape Malays	0.000 ± 0.000 (+)	...
Johannesburg 'Coloured'	0.567 ± 0.026 (-)	0.000 ± 0.000 (+)

Note: Only significant values at the 0.05 level are shown in the table.

(+) = Significant difference, (-) = No significant difference.

3.4.2 *Tracing the ancestry of Y-lineages in the South African "Coloureds"*

Both Y-chromosome biallelic haplogroup and STR haplotype data were used in conjunction with other published and unpublished data to trace the geographic origins of Y-chromosomes in the three 'Coloured' populations. The distributions of the biallelic haplogroups in the SA 'Coloured' and SA White populations are shown in Figure 3.4. In addition, haplogroup distributions in the presumed parental populations of the SA 'Coloureds' are shown in Figure 3.5. Median-joining networks (Figures 3.6–3.21) were constructed using haplotype information within haplogroups to examine the relationship of haplotypes found in the 'Coloureds' with their presumed parental populations.

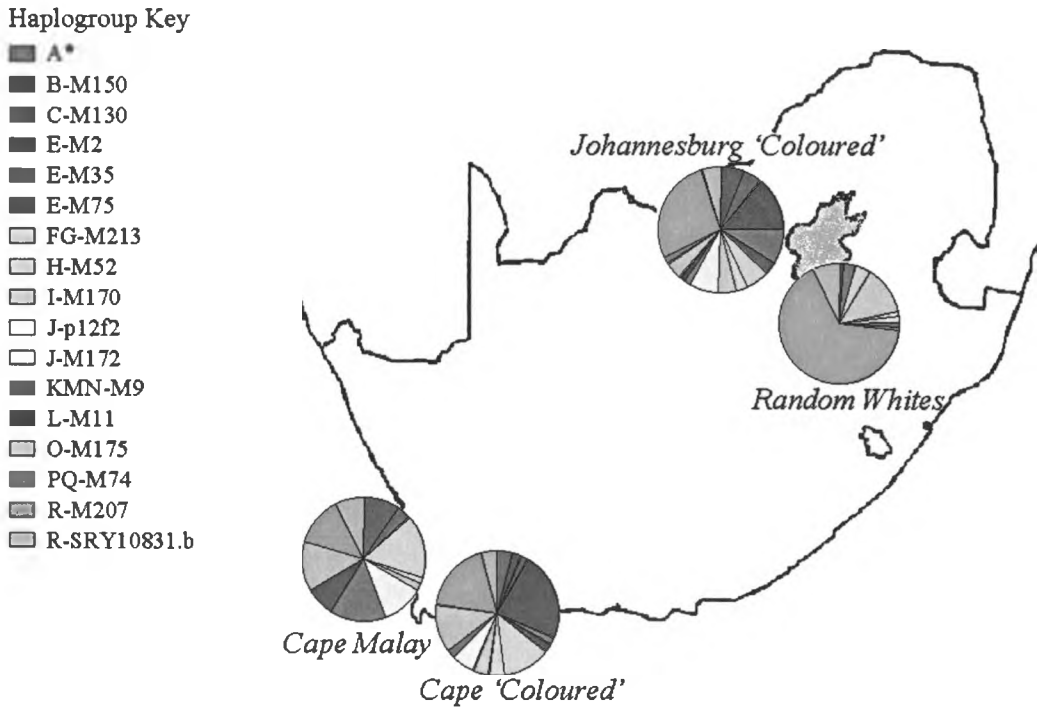


Figure 3.4: Distributions of Y-biallelic haplogroups in the South African 'Coloured' (Cape Town and Johannesburg), Cape Malay, and White populations.

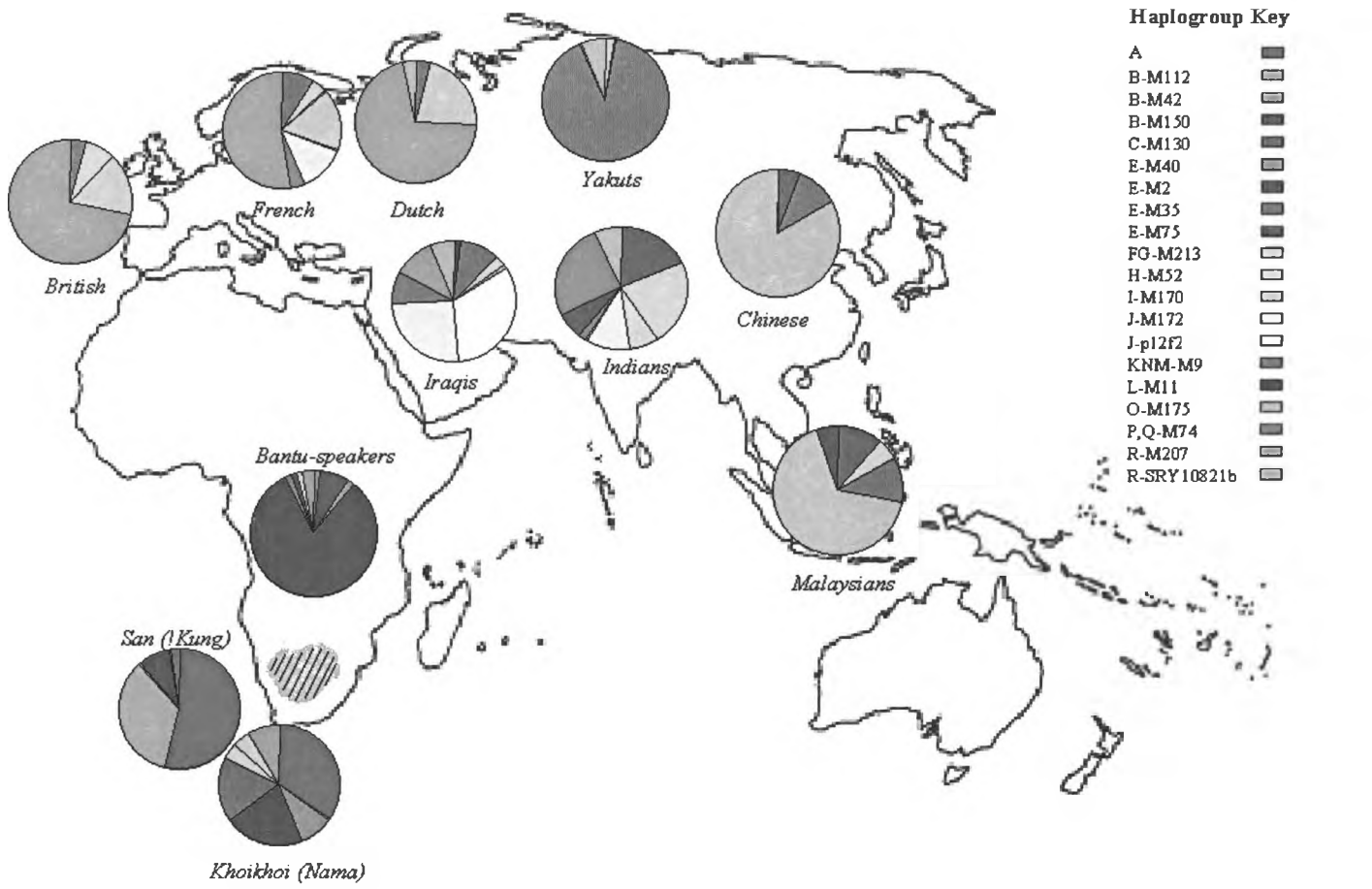


Figure 3.5: Frequency distributions of Y-chromosome biallelic haplogroups in some contemporary global human populations. The African populations include Khoisan (Nama and !Kung) and Bantu-speakers from South Africa (Soodyall, unpublished). Europeans include Dutch, French (Semino *et al.* 2000) and British (Wells *et al.* 2001). The Asian data used in the comparative analysis included Indians (Ramana *et al.* 2001), Iraqis (Al-zahery *et al.* 2002), Yakuts (Puzyrev *et al.* 2003), Chinese and Malaysians (Kayser *et al.* 2003).

3.4.2.1 *African origin of Y-chromosome lineages*

*3.4.2.1.1 Haplogroup A**

Haplogroups A, B and E found most commonly in African populations, were found at varying frequencies in the different groups examined (Figures 3.2 and 3.4). A paragroup A* [Y-chromosomes within a clade that have not been resolved further using other polymorphisms into sub-lineages, YCC (2002)] was only found in the Cape 'Coloured' group at a frequency of 4.2% (Figures 3.2 and 3.4). This is the oldest haplogroup lineage found in modern humans (dated to ~90 000 years old) and is found exclusively in Africa at frequencies of 46.3% in the Khoisan from southern Africa (Figure 3.5), 45% in Sudanese and 13.6% in Ethiopians from East Africa, 4.6% in South Africans and 2.3% in Malians from Northwest Africa (Underhill *et al.* 2000, 2001).

The network in Figure 3.6 shows the relationship of haplotypes within the A* paragroup found in the Cape 'Coloured', Khoisan and SA Bantu-speaking populations (Soodyall, unpublished). Haplotypes H1 and H2 were the only two haplotypes found in the Cape 'Coloureds' (Appendix C). H2 was also found in the Khoisan, while H1 was only found in one Cape 'Coloured' individual and differed by two mutational steps from a haplotype found in a Khoisan individual. These findings suggest a Khoisan ancestry of these haplotypes in the Cape 'Coloured' group. The other haplotypes in the network were found in the SA Bantu-speakers and Khoisan individuals and were not detected in the sample examined in this study. While the network shows a complex relationship of all haplotypes, with the various branches being separated by nodes (either representing missing haplotypes due to sampling biases or additional structure within haplogroup), the

presence of these haplotypes found in the Cape 'Coloureds' suggest a recent admixture from the Khoisan.

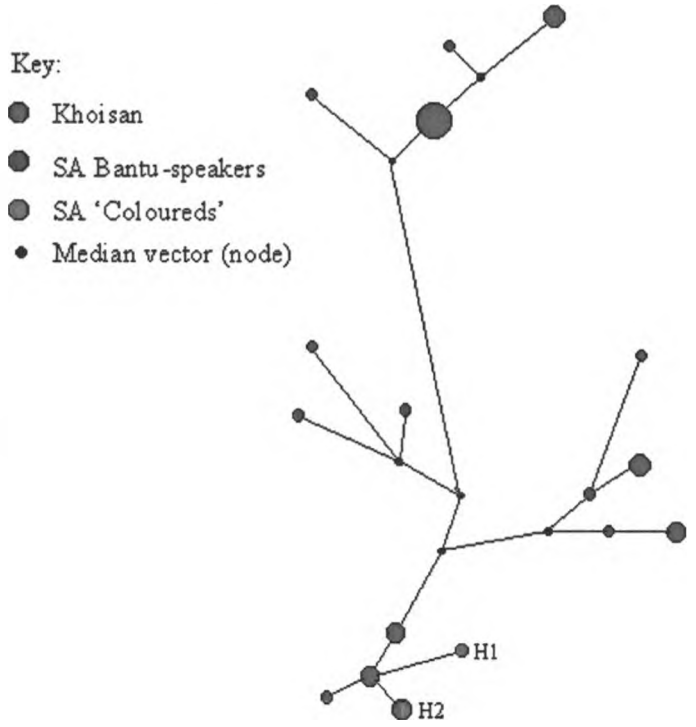


Figure 3.6: Median-joining network showing the relationship of Y-STR haplotypes within paragroup A* found in the Khoisan, SA Bantu-speakers and SA 'Coloureds'. Nodes are represented as black circles. The lines joining haplotypes and nodes are proportional to the number of mutational steps. The size of the circles is proportional to the frequencies of haplotypes.

3.4.2.1.2 Haplogroup B-M150

Haplogroup B-M150 was found at frequencies of 6.2% in the Johannesburg 'Coloured' and 2.1% in the Cape 'Coloured' populations (Figures 3.2 and 3.4). This haplogroup has been observed mostly in sub-Saharan Africans (Figure 3.5) and has been found at frequencies of 18% in central West Africans (Fali), 13.2% in South Africans, 10% in Cameroonians (Ewondo), 9% in East Africans (Ethiopians), and 8% in Central Africans (Mbuti pygmies) (Underhill *et al.* 2000; Cruciani *et al.* 2002), but it is not found in the Khoisan (Soodyall, unpublished).

The B-M150 modal haplotype H5 (Appendix C) found at a frequency of 6.3% in the Bantu-speakers and not in the Khoisan (Soodyall, unpublished), was found in 3.1% of the Johannesburg 'Coloureds' (Figure 3.7). H4 and H6 were found in both the Cape 'Coloured' and Johannesburg 'Coloured' groups and have also been found in Bantu-speaking individuals from South Africa (Soodyall, unpublished). Haplotype H3, found solely in the Johannesburg 'Coloureds', differed from the modal haplotype by one mutational step. All haplotypes within haplogroup B-M150 found in the Johannesburg and Cape 'Coloured' groups were most likely introduced by admixture from Bantu-speaking people.

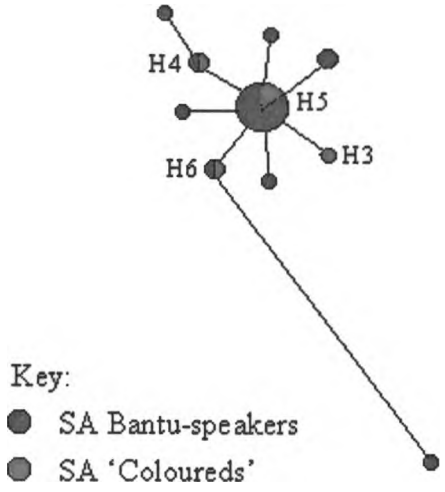


Figure 3.7: Median-joining network showing the relationship between Y-STR haplotypes within haplogroup B-M150 found in the SA Bantu-speaking (Soodyall, unpublished) and SA 'Coloured' populations.

3.4.2.1.3 Haplogroup E-M2

Haplogroup E-M2 was found at frequencies of 22.9% in the Cape 'Coloureds', 13.8% in the Johannesburg 'Coloureds' and 1% in the SA Whites, but it was not found in the Cape Malay (Figures 3.2 and 3.4). E-M2 lineages have been observed in the Bantu-speaking populations (Figure 3.5) at considerable frequencies of 96% in central West Africans (Bamileke), 91% in West Africans (Mossi), 65% in central Africans (Biaka Pygmies), 64.2% in South Africans, 18% in Khoisan from southern Africa, 7.8% in North Africans (Berbers), and 3.4% in East Africans (Ethiopians) (Underhill *et al.* 2000; Cruciani *et al.* 2002).

Haplotypes (H16 - H28) with the exception of H20 and H21 (Appendix C) found in the SA 'Coloureds' are derived from haplotypes found in the SA Bantu-speaking populations (Figure 3.8). H20 and H21 were derived from a nodal position, which is connected to two commonly found haplotypes among Bantu-speakers. H17 was also found in 1% of the total SA White population. H26 and H25 were also found in the Khoisan. It is possible that these Khoisan individuals have a male ancestry derived from the Bantu-speakers. These findings suggest a Bantu-speaking ancestry of E-M2 lineages in the Cape and Johannesburg 'Coloured' groups.

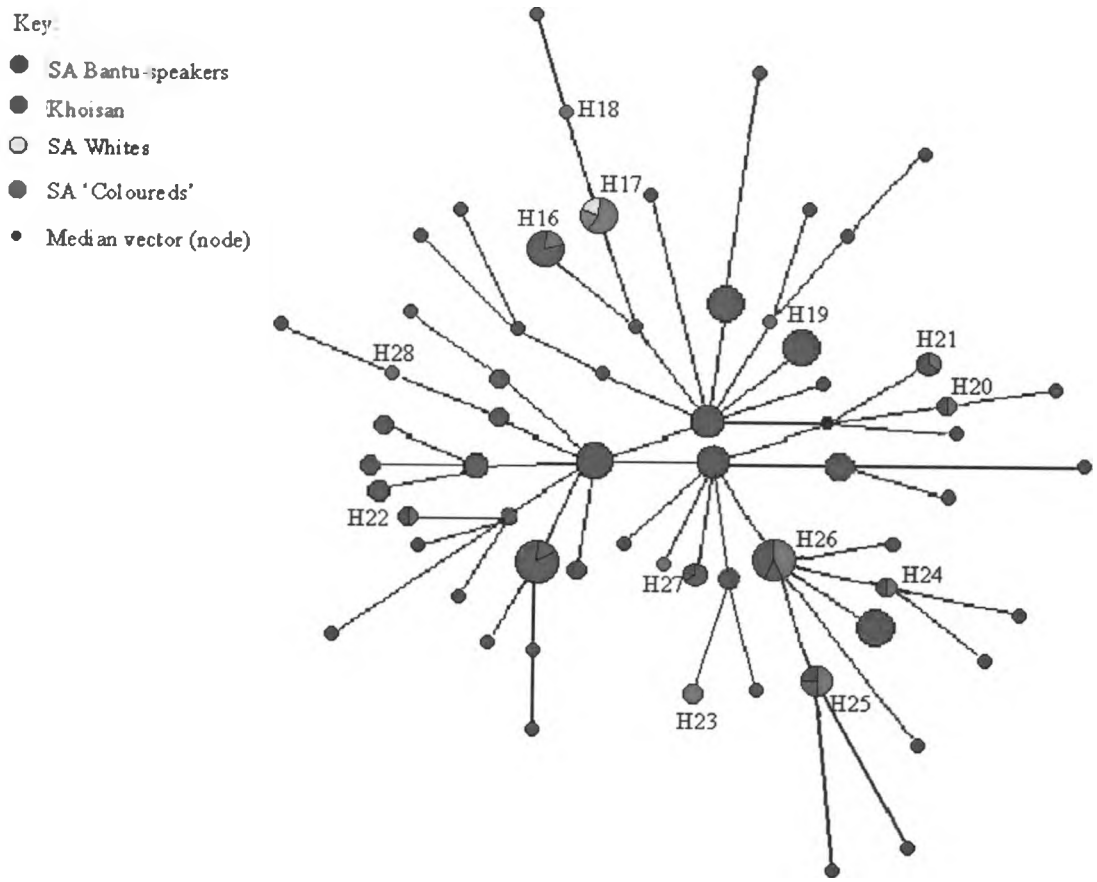


Figure 3.8: Median-joining network showing the relationship between Y-STR haplotypes within haplogroup E-M2 found in the SA Bantu-speakers, Khoisan (Soodyall, unpublished), SA White and SA 'Coloured' populations.

3.4.2.1.4 Haplogroup E-M75

Haplogroup E-M75 was found at low frequencies in the Johannesburg 'Coloured' (3.1%) and 'Cape Coloured' (2.1%) populations (Figures 3.2 and 3.4). E-M75 lineages have been found at varying frequencies of 17.1% in Ethiopians, 15.1% in South African Bantu-speakers, 8.1% in Central Africans, 5% in Sudanese (Underhill *et al.* 2000) and 1.5% in Congolese (Barkhan and Soodyall, unpublished).

The commonest haplotype H40 was found in both the Bantu-speaking and the two 'Coloured' groups from Cape Town and Johannesburg (Figure 3.9; Appendix C). H41, which was found exclusively in the Johannesburg 'Coloured' group, differed by one mutational step from the commonest haplotype. Thus, the introgression of this lineage into the SA 'Coloured' groups was from Bantu-speaking ancestry.

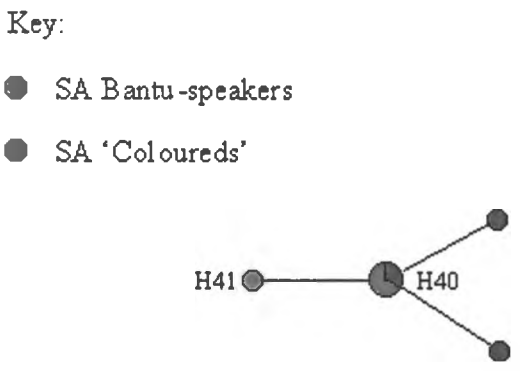


Figure 3.9: Median-joining network showing the relationship between Y-STR haplotypes within haplogroup E-M75 found in the SA Bantu-speakers (Soodyall, unpublished) and SA 'Coloured' populations.

3.4.2.2 *Summary*

In summary, using haplogroups A, B and E we estimated the total African contribution of Y-chromosomes in the Cape Malays, Johannesburg 'Coloureds' and Cape 'Coloureds' to be of 0%, 24.6% and 31.3%, respectively. Using the distribution of Y-chromosome haplogroups and accompanying haplotypes within Africa, we estimate that 4.2% of Y-chromosomes in the 'Cape Coloureds' is derived from Khoisan ancestry, while 18.9% is derived from the Bantu-speakers. We did not detect any Khoisan ancestry in the 'Coloured' group from Johannesburg; thus the African contribution in them is derived exclusively from the Bantu-speakers.

3.4.2.3 *European origin of Y-chromosome lineages*

3.4.2.3.1 Haplogroup E-M35

E-M35 has been found at a frequency of 9.2% in the Johannesburg 'Coloureds', 3.7% in the Cape Malays, 2.1% in the Cape 'Coloureds' and 3.1% in the SA White population (Figures 3.2 and 3.4). This haplogroup ranges in frequency from 69% in North Africans (Berbers), 14% in Ethiopians Jews, 11.8% in Corsicans from Mediterranean, 8.7% in French, 4% in British and 3.7% in Dutch (Figure 3.5) (Semino *et al.* 2000, Cruciani *et al.* 2002; Francalacci *et al.* 2003).

Using the comparative data at hand, the associations of haplotypes in E-M35 found in the SA 'Coloured' populations suggest two possible contributions (Figure 3.10; Appendix C). H33 is closely associated with haplotypes found in Africans (Khoisan and Bantu-speakers) and consequently is most likely to have South African ancestry (cluster 1). Some of the other haplotypes (H34, H35, H36, H38 and H39) are found in close association with lineages in the SA Whites (cluster 2). In fact, H34 is also found in a SA White individual. The other haplotypes H29 and H30 are connected to nodal positions, presumably due to inadequate sampling. We suggest that these haplotypes have a European ancestry. H32 differed by two mutational steps from a haplotype found in a Khoisan individual that differs to H37 found in the SA White by one mutational step. The haplotype found in the Khoisan individual most likely has a European ancestry. Therefore, we have interpreted the ancestry of H32 to a European source.

Based on the above interpretation, we estimated the proportion of African derived E-M35 lineages to be 1.5% in the Johannesburg 'Coloured' and European derived lineages to be 3.7% in the Cape Malays, 2.1% in the Cape 'Coloureds' and 9.2% in the Johannesburg 'Coloureds'.

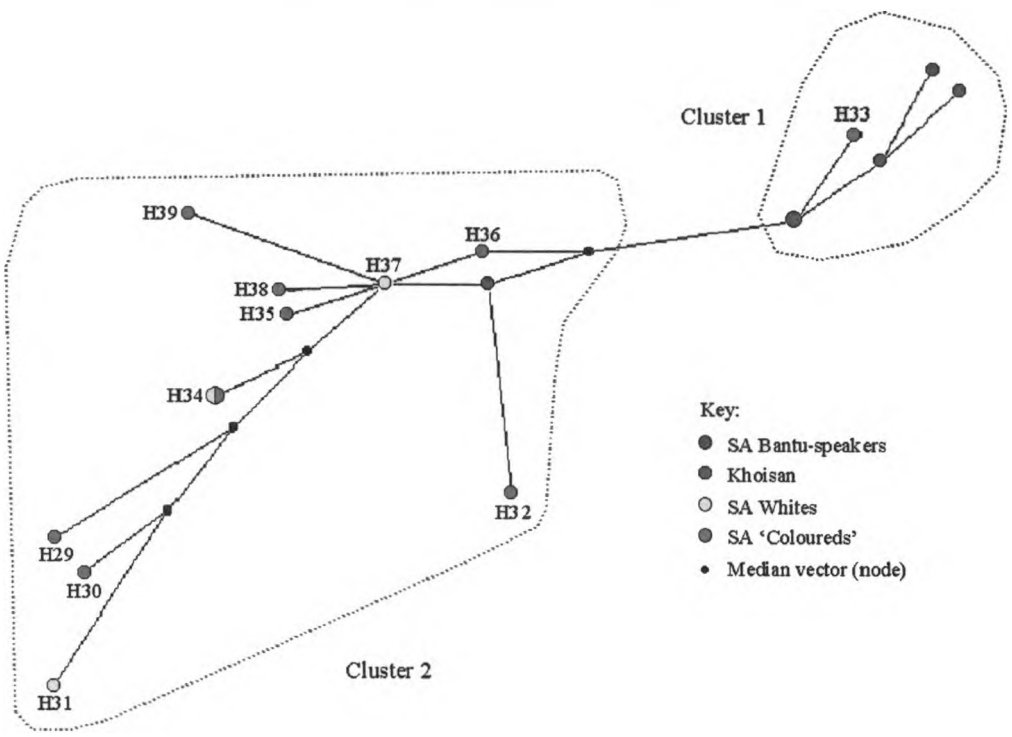


Figure 3.10: Median joining network showing the relationship between Y-STR haplotypes within haplogroup E-M35 found in the SA White, SA Bantu-speaking, Khoisan and SA 'Coloured' populations.

3.4.2.3.2 Haplogroup I-M170

Haplogroup I-M170 was found at frequencies of 1.9% in Cape Malays, 4.2% in 'Cape Coloureds', 4.6% in Johannesburg 'Coloureds' and 13.4% in the SA White population (Figures 3.2 and 3.4). The main cluster in the network is associated with H79 with the majority of other haplotypes found in the SA Whites and SA 'Coloureds' being connected to it by one or two step mutational changes within this network (Figure 3.11; Appendix C). Given the distribution of I-M170 in Europe [44.8% in Croatians, 41.7% in Saami, 37.7% in Sardinians, 37.5% in Germans, 23.6% in Polish, 22.2% in Dutch, 19.6% in Albanians, 17.4% in French, 8% in Italians (Semino *et al.* 2000; Figure 3.5)], we have inferred this haplogroup and its associated haplotypes to having a European origin. The two haplotypes – one in a Khoisan individual and one in a SA Bantu-speaker must also have Y-chromosome that trace to a European ancestry. H67 differed from the commonest haplotype H79 by eight mutational steps, including a nodal position. This could be due to: a) extinction of some lineages within a population and b) substructure within the haplogroup attributable to step-wise mutations. In particular, this haplotype was associated with a lower sized repeat at the DYS388 locus. Most individuals harbor alleles that range between (10 -16); this individual was found to have 9 repeats, which was confirmed by sequencing (see Figure 3.3; Appendix C). Thus, the presence of this haplogroup in the SA 'Coloureds' can be traced to European ancestry.

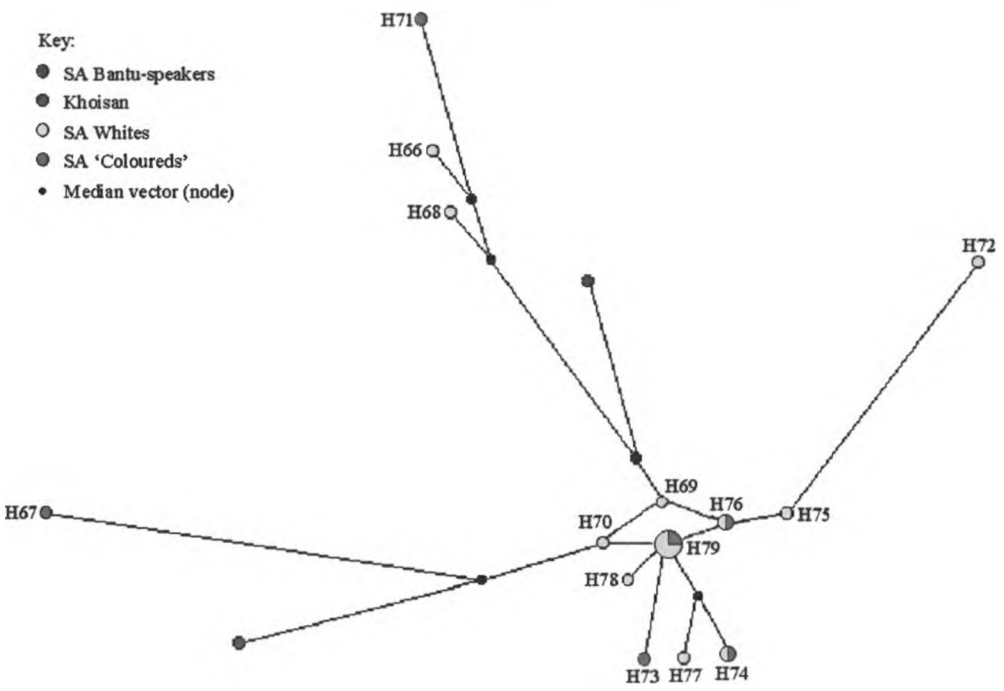


Figure 3.11: Median-joining network showing the relationship between Y-STR haplotypes within haplogroup I-M170 found in the SA White, SA Bantu-speaking, Khoisan and SA 'Coloured' populations.

3.4.2.3.3 Haplogroup R-M207

Haplogroup R-M207 was found at appreciable frequencies in Cape Malays (13%), Cape 'Coloureds' (18.8%), Johannesburg 'Coloureds' (27.7%) and SA Whites (66%) (Figures 3.2 and 3.4). R-M207 lineages have been identified at high frequencies in European populations including: Spanish Basque (88.9%), French Basque (86.4%), Catalan (79.2%), Dutch (70.5%), Italian (62%), French (52.2%), German (50%), Slovakian (35.6%), Greek (27.6%), Sardinian (22.1%) and Albanian (17.6%) (Semino *et al.* 2000). Moreover, this lineage was also found in Asian populations, albeit at lower frequencies [Central Asia (32%), Caucasus (17%), Russians (9%), and East Asia (2%)], but it was not found in India (Wells *et al.* 2001; Ramana *et al.* 2001; Figure 3.5)

Among the SA Whites the modal haplotype was H148 (Figure 3.12; Appendix C). This haplotype was found at a frequency of 6.3% in the Cape 'Coloureds', 3.1% in the Johannesburg 'Coloureds' and 1.9% in the Cape Malays. Five haplotypes (H128, H133, H147, H150 and H152) differed from the modal haplotype by a single mutational step and were identified at varying frequencies in both the SA White and SA 'Coloured' populations. Moreover, H144 and H164 were also single step neighbours, which were found exclusively in the SA 'Coloured' population. Four haplotypes (H132, H134, H135, H145 and H161) were found only in the SA 'Coloured' population and differ from the modal haplotype by two mutational steps. H136 was found in the SA 'Coloureds' and differed from H148 by three mutational steps. Two haplotypes H126 and H131 found in the SA 'Coloureds' differed from the modal haplotype by four mutational steps.

Haplotypes H141 and H155 differed from the commonest haplotype, H148 by two mutational steps and were found at varying frequencies in both the SA 'Coloureds' and SA Whites. H173, H125 and H166 found in the SA 'Coloured' population differed from H248 by two, three and five mutational steps, respectively, incorporating single step median vectors. From these associations it can be concluded that these haplotypes found in the SA 'Coloured' population have a European ancestry (Figure 3.12).

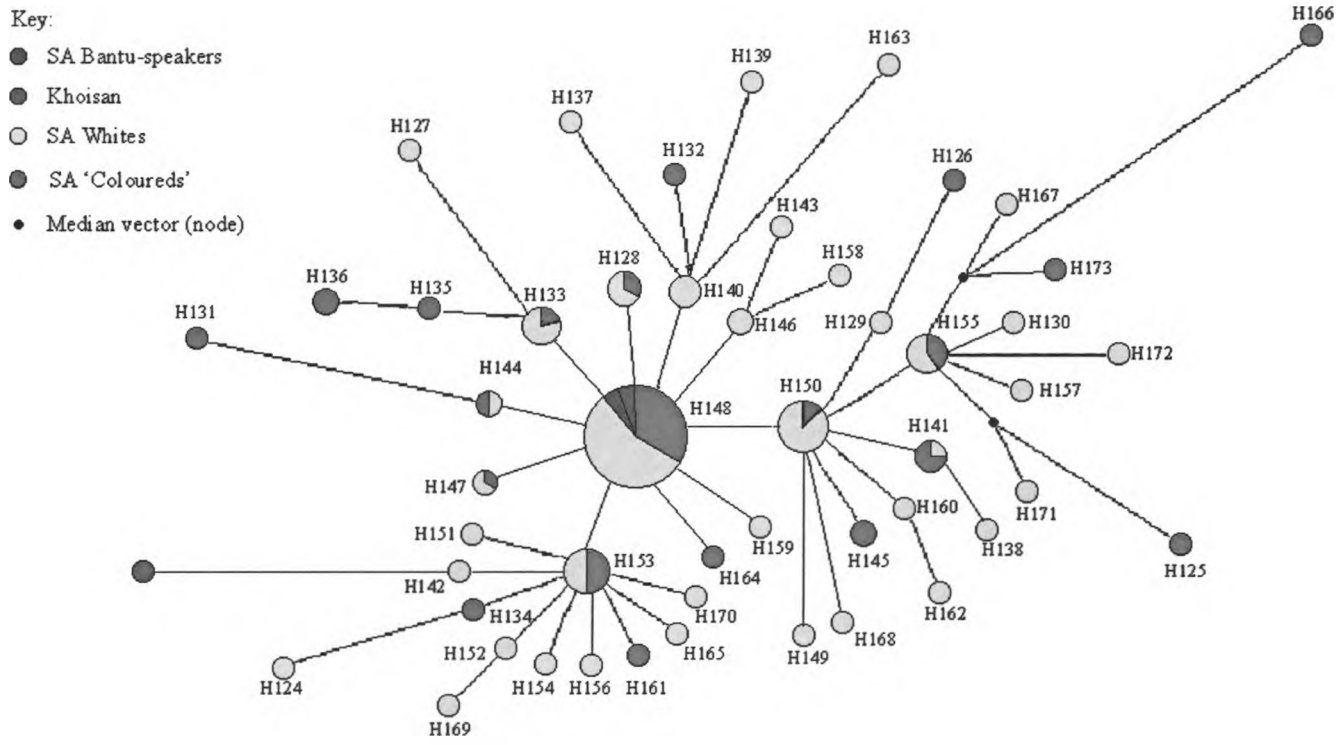


Figure 3.12: Median-joining network showing the relationship between Y-STR haplotypes within haplogroup R-M207 found in the SA White, SA Bantu-speaking, Khoisan and SA 'Coloured' populations.

3.4.2.3.4 Haplogroup R-SRY10831.2

Haplogroup R-SRY10831.2 was found at varying frequencies of 7.4% in Cape Malays, 4.2% in 'Cape Coloureds', 4.6% in Johannesburg 'Coloureds', and 7.2% in the White population (Figures 3.2 and 3.4). This lineage has previously been found at frequencies of 68% in Ishkashimi from Central Asia (Wells *et al.* 2001), 60% in Hungarian from Europe (Semino *et al.* 2000) and 47% in Punjab from India (Kivisild *et al.* 2003).

Data from Malaysians, Han Chinese (Kayser *et al.* 2003), Yakuts (Puzyrev *et al.* 2003) and SA Indians (Soodyall, unpublished), excluding locus DYS388, were used in the comparative analysis to compute the network presented in Figure 3.13. Haplotypes H174 and H182 are the commonest R-SRY10831.2 haplotypes found in the SA Whites (Appendix C), although H182 was also found at a frequency of 25% of in the SA Indians. Haplotype H184 was found in both the Cape 'Coloured' and SA White populations. Many of the haplotypes found in the SA 'Coloured' populations were either one-step or two-step neighbours (except H185) to haplotypes found both in the SA White and SA Indian populations. Haplotype H185 found in Cape Malay individual differed from H182 by five mutational steps, either due to sample size bias, lineage extinction or substructure within the haplogroup resulting from step-wise mutations. In particular, this haplotype was associated with a lower sized repeat at the DYS319. Most individuals harbor alleles that range between (13-17), however this individual was found to have 10 repeats, which was confirmed by sequencing (see Figure 3.3; Appendix C).

This suggests a recent introduction of these lineages in the SA 'Coloureds' from Indian and European ancestors. However, we cannot unambiguously resolve the origin of this haplogroup and will therefore refer to it as having Eurasian ancestry.

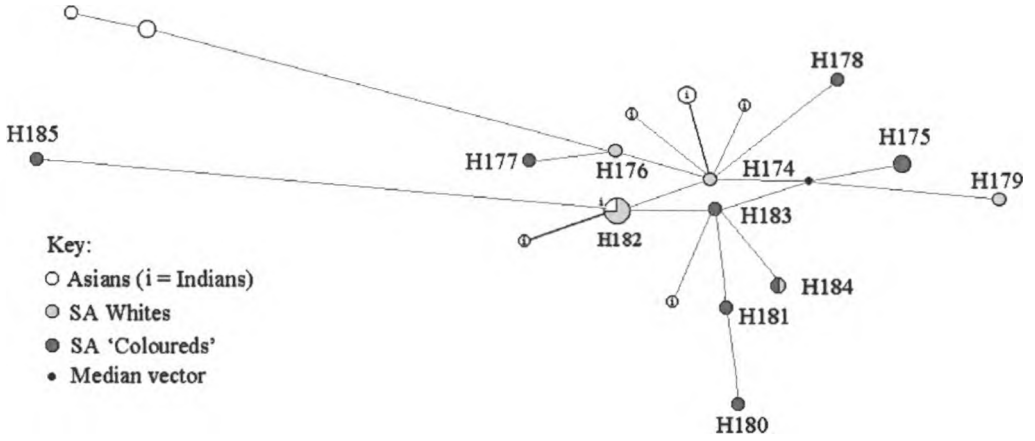


Figure 3.13: Median-joining network showing the relationship of Y-STR haplotypes within haplogroup R-SRY10831.2 found in the Asians, SA Whites and SA 'Coloureds'. Note: DYS388 is excluded in this comparative analysis that includes SA Indian (i) (Soodyall, unpublished), Malaysian, Chinese, and Russian populations.

3.4.2.4 *Summary*

While haplogroups E-M35, I-M170, R-M207 and R-SRY10831.2 were intended to be used as indicators of European ancestry, haplotype analysis revealed that 1.5% of E-M35 in the Johannesburg 'Coloured' was derived from African ancestry, while R-SRY10831.2 could not be unambiguously resolved and therefore referred to as Eurasian in the present study. Overall, the lineages which could be traced to European ancestry was found at frequencies of 18.5% in Cape Malays, 25.0% in Cape 'Coloureds' and 40.0% in Johannesburg 'Coloureds'.

3.4.2.5 *Asian origin of Y-chromosome lineages*

3.4.2.5.1 Haplogroup C-M130

C-M130 lineages have been found in Central Asia [Kazaks (73.7%), Mongolians (53.8%)], India (Gujarat and Lambadi Indians each at 17.2%), Australians (43%), New Guineans (17.4%) and Japanese (13.0%), 11.1% Malaysians (Figure 3.4), 5.6% Chinese and 1.3% in Anatolia (Underhill *et al.* 2000; Zerjal *et al.* 2002; Kivisild *et al.* 2003, Kayser *et al.* 2003; Cinnioğlu *et al.* 2004). However, it is also found at low frequencies in some European populations, [Lebanese (3.2%), and Sardinians and Greeks each at 1.3%] (Semino *et al.* 2000). The most likely origin of haplogroup C-M130 found at frequencies of 9.3% in the Cape Malay, 4.6% in the Johannesburg 'Coloured' and 2.1% in the Cape 'Coloured' populations is Asia since it was not found in SA Whites (Figures 3.2 and 3.4).

Data from Malaysians, Han Chinese (Kayser *et al.* 2003), Yakuts (Puzyrev *et al.* 2003), SA Indians (Soodyall, unpublished) and Anatolians (Cinnioğlu *et al.* 2004), excluding locus DYS388, were used in the comparative analysis to compute the network presented in Figure 3.14. It is not easy to resolve the haplotypes found in this haplogroup network. There is evidence of a founder effect haplotype from which haplotypes H10, H12, H13 and H14 (Appendix C) are derived since they are found predominantly in the Cape Malay (except H12 found in one Cape 'Coloured' individual). The other haplotypes found in SA 'Coloureds' are derived from multiple founders (at least four others), presumably of Asian origin from populations from whom there is no comparative data available. The closest neighbours to haplotypes H11, which differed by two-step mutation from H7, is three mutational steps to a haplotype found in a SA Indian [haplotypes from Indian populations marked with an (i)]. Also, haplotype H9 is derived from a lineage that is

found in a SA Indian, albeit several mutation step differences apart. The most likely source of at least two of the founders could be Indian, given the prevalence of this haplogroup in India (Kivisild *et al.* 2003).

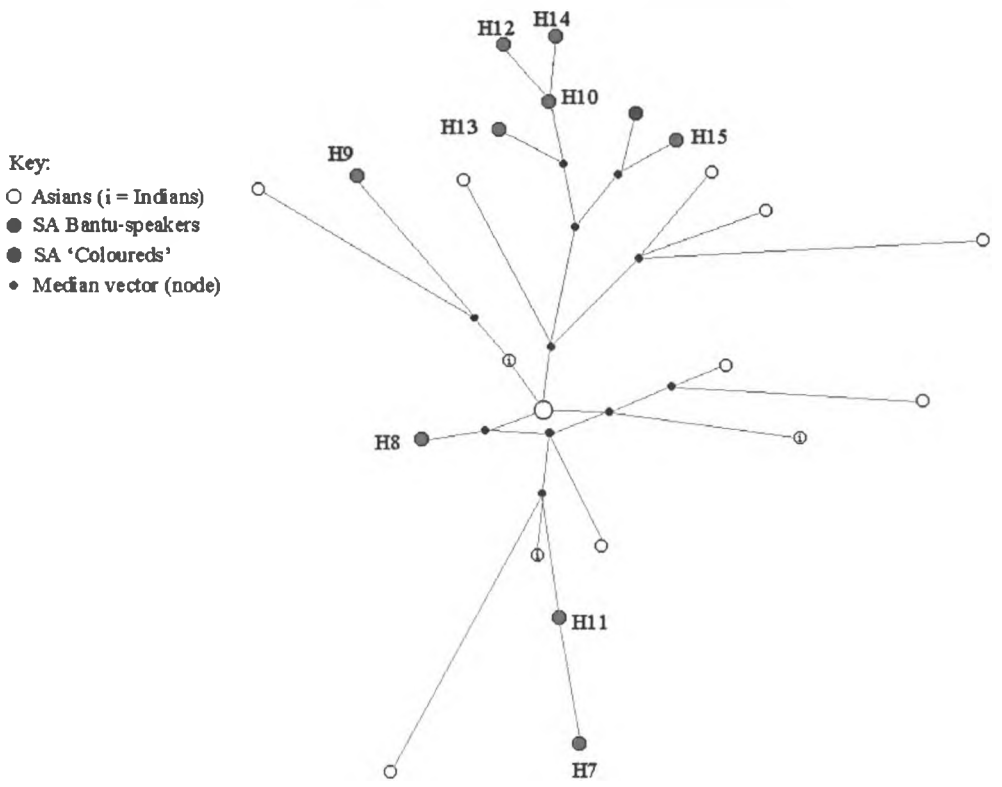


Figure 3.14: Median-joining network showing the relationship of Y-STR haplotypes within haplogroup C-M130 found in the Asian – including SA Indian (Soodyall, unpublished), SA Bantu-speaking and SA 'Coloured' populations.

3.4.2.5.2 Haplogroup D-M174

Haplogroup D-M174 has been found at reasonably high frequencies in Southeast Asia (30% in Yao Jinxiu), Northeast Asia (25% in Tibetans), but does not occur in Oceanic, European and American Indian populations (Su *et al.* 1999). Interestingly, D-M174 was **not** found in the South African 'Coloured' population groups examined (Figure 3.2).

3.4.2.5.3 Haplogroup H-M52

Haplogroup H-M52 was found at frequencies of 4.2% in the Cape 'Coloured', 3.1% in the Johannesburg 'Coloured' and 1.9% in the Cape Malay populations (Figures 3.2 and 3.4). H-M52 lineages were found at varying frequencies of 60.9% in the Koya from India (Kivisild *et al.* 2003), 13% in Dushanbe from Central Asia (Wells *et al.* 2001), 6.8% in Peruru Brahmins from India (Ramana *et al.* 2001; Figure 3.5), and 5% in Syrians (Semino *et al.* 2000). This haplogroup is not found in Western Europe (Wells *et al.* 2001). Haplotypes H62 and H63 occur in both the SA Indian and SA 'Coloured' populations and differed from the H-M52 founding haplotype in the SA Indians by one mutational step (Figure 3.15; Appendix C). All haplotypes found in the SA 'Coloured' populations are closely associated with haplotypes found in the SA Indian (except H61, which is connected to a node). The presence of these haplotypes in the SA 'Coloured' populations are suggested to have an Indian ancestry.

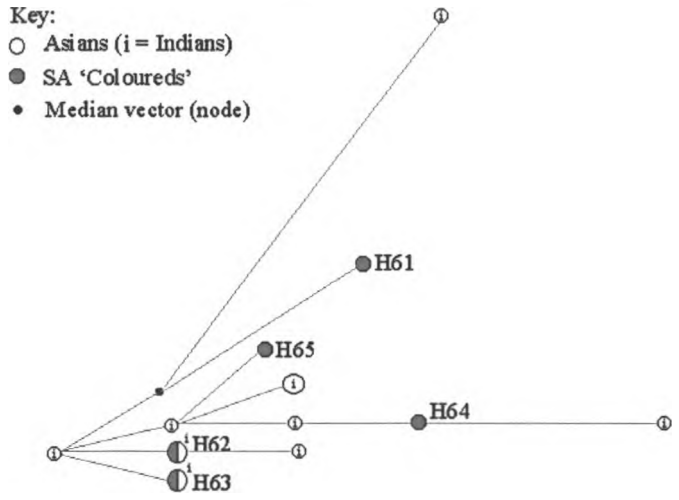


Figure 3.15: Median-joining network showing the relationship of Y-STR haplotypes within haplogroup H-M52 found in the SA Indian (Soodyall, unpublished) and SA 'Coloured' populations. DYS388 is excluded in this comparative analysis that includes SA Indian, Malaysian, Chinese, and Russian populations.

3.4.2.5.4 Haplogroup KMN-M9

Haplogroup KMN-M9 was detected at frequencies of 14.8% in the Cape Malay, 2.1% in the Cape 'Coloured', 1.5% in the Johannesburg 'Coloured' and 1% in the SA White populations (Figure 3.2). This haplogroup has been found in Melanesians (83%), Southeast Asians (34%), Polynesians (8%) and Taiwanese (2%) (Capelli *et al.* 2001) as well as in Yakuts from Russia (91%) (Puzyrev *et al.* 2003; Figure 3.5), East Asians (25%), Central Asians (12%) (Wells *et al.* 2001), Malaysians and Han Chinese each at (11.1%) (Kayser *et al.* 2003) and Anatolians (6.5%) (Cinnioğlu *et al.* 2004). Also, Semino *et al.* (2000) found that this haplogroup was found at frequencies <6.5% in European populations and attributed its presence in Europeans to Asian ancestry. Given the above, this haplogroup can also be traced to Asian ancestry.

Haplotype H104 found in a Cape Malay individual is a closest neighbour to a haplotype found in a Russian individual and differed from it by one mutation (Figure 3.16; Appendix C). Eight other haplotypes found in the SA 'Coloureds', predominantly in the Cape Malay were closely related in the network and differed from an Anatolian founding haplotype by more than four mutational steps including nodal positions. These findings may suggest additional structure or origins of haplotypes from Asian parental sources excluding those used in this comparative analysis.

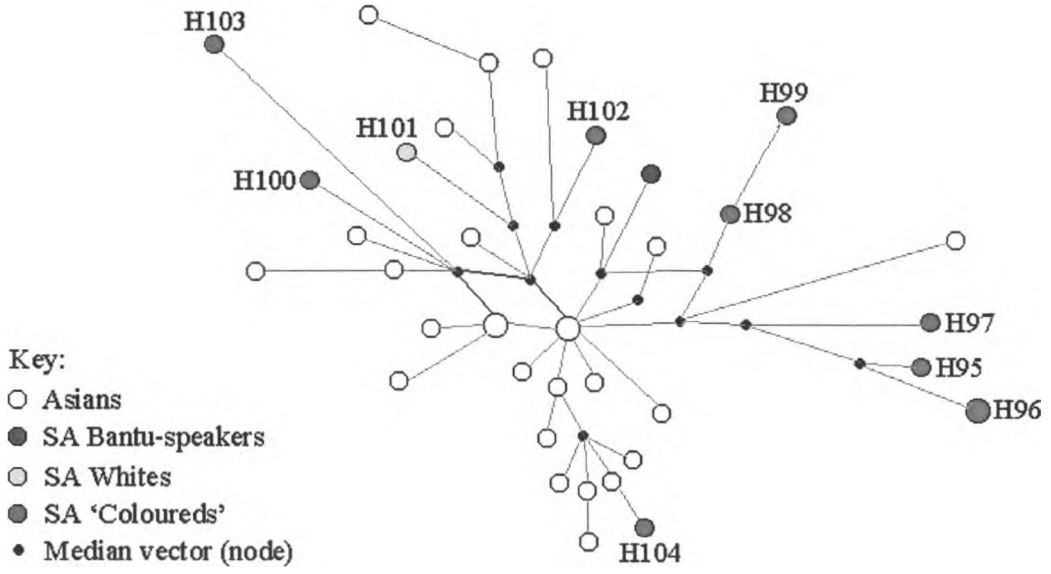


Figure 3.16: Median-joining network showing the relationship of Y-STR haplotypes within haplogroup KMN-M9 found in the Asian, SA White, SA Bantu-speakers and SA 'Coloured' populations. DYS388 is excluded in this comparative analysis that includes SA Indian, Malaysian, Chinese, Russian and Anatolian populations.

3.4.2.5.5 Haplogroup L-M11

Haplogroup L-M11 has been found in 48% Kalar from Indian, 16% Bartanei from Central Asia, 8% Sairaz from Middle East (Wells *et al.* 2001), 28.8% Baluch from Pakistan (Qamar *et al.* 2002), 6.8% Peruru Brahmin from India (Ramana *et al.* 2001) 5.4% Calabrian (Semino *et al.* 2000), and 4.8% Armenia from Europe (Zerjal *et al.* 2002). It was only found in the Cape Malay and Johannesburg 'Coloured' groups at frequencies of 7.4% and 1.5%, respectively (Figures 3.2 and 3.4). H108 is the SA Indian founding haplotype defined at haplogroup L-M11, which was also found in the Cape Malay and Johannesburg 'Coloured' individuals (Figure 3.17; Appendix C). Haplotypes H106 and H107 were closely associated with the founding haplotype and differed from it by one mutation. H105 is a two-neighbour to a haplotype found in a SA Indian. Therefore, the presence of L-M11 lineages in the SA 'Coloured' population can also be traced to an Indian ancestry.



Figure 3.17: Median-joining network showing the relationship of Y-STR haplotypes within haplogroup L-M11 found in the Asian and SA 'Coloured' populations. DYS388 is excluded in this comparative analysis that includes SA Indian, Malaysian, Chinese, and Russian populations.

3.4.2.5.6 Haplogroup O-M175

Haplogroup O-M175 was also found at frequencies of 13.0% in the Cape Malay, 12.5% in the Cape 'Coloured' and 4.6% in the Johannesburg 'Coloured' populations. It was not found in the SA White population (Figures 3.2 and 3.4), but has been identified at considerable frequencies among South East Asian populations [88.6% in Malaysians (Kayser *et al.* 2003; Figure 3.5), 83% Borneose from Southeast Asia (Capelli *et al.* 2001) 84% in Southern Chinese from Taiwan, 97% Philippinos from Polynesia, 15% Fiji from Melanesia (Capelli *et al.* 2001), 3.2% Indians from west Bengal (Kivisild *et al.* 2003)]. This is another useful Y chromosome marker for tracing Asian ancestry in the SA 'Coloured' populations (Figure 3.5).

The network shown in Figure 3.18 shows a single cluster relating all haplotypes found in Asian and SA 'Coloured' populations. The SA 'Coloured' haplotypes H109, H110, H111, H114, H118, H119, H120 and H121 derived from a common node, presumably of an Asian descent (Figure 3.18; Appendix C) were found in the network to be in close association with haplotypes found in the SA Indian as well as Chinese populations. Furthermore, haplotypes H112, H113, H115 and H116 that were derived from a different node with the Chinese and Malaysian populations appeared to be in close association with the two groups. H117 found in a Cape Malay individual differed from its closest Indian neighbour by one mutational step. From these haplotype associations we suggest a Southeast Asian contribution of the O-M175 lineages in the SA 'Coloured' population.

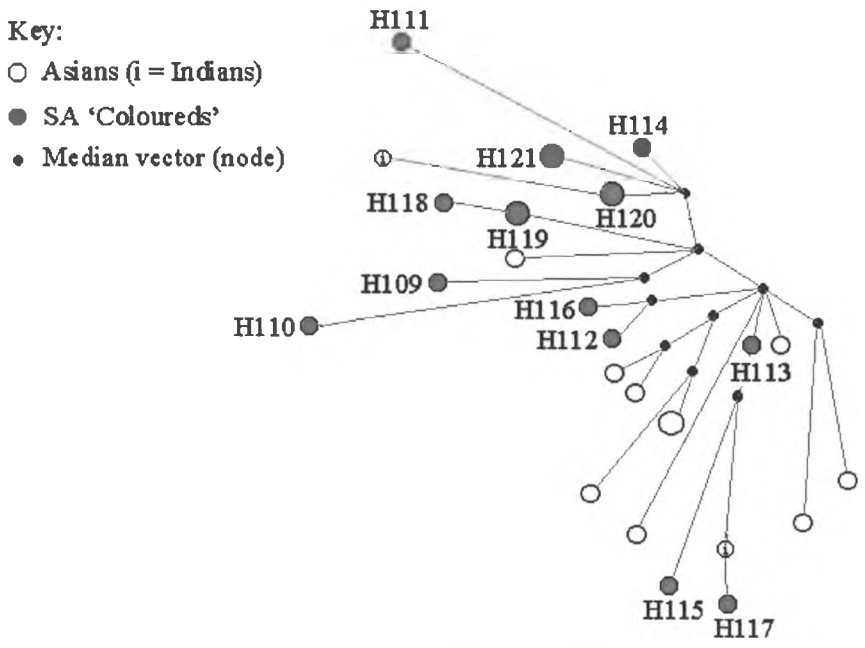


Figure 3.18: Median-joining network showing the relationship of Y-STR haplotypes within haplogroup O-M175 found in the Asian and SA 'Coloured' populations. DYS388 is excluded in this comparative analysis that includes SA Indian, Malaysian, Chinese, and Russian populations.

3.4.2.5.7 Haplogroup PQ-M74

Haplogroup PQ-M74 was found at a frequency of 1.5% in the Johannesburg 'Coloured' and 1% in the SA White populations (Figures 3.2 and 3.4). This haplogroup has been found in 30% Turkmen from Central Asia, 25% Peruru Brahmins from South India (Figure 3.5) (Ramana *et al.* 2001) 17% Tuvlotan from East Asia (Wells *et al.* 2001), 14% Hvar from Croatian Islands (Barac *et al.* 2003), and <4.3% European populations, excluding the ones on the West (Semino *et al.* 2000), 4% Koreans, 3.1% Papua New Guinean, 2.5% Southern Borneo (Kayser *et al.* 2003). The introduction of this lineage in the Johannesburg 'Coloureds' is thus from the Asian ancestry (Figure 3.5).

3.4.2.6 *Summary*

Using haplogroups C-M130, H-M52, KMN-M9, L-M11, O-M175 and PQ-M74, the Asian contribution to the Y-chromosome pool of Johannesburg 'Coloured', Cape 'Coloured', and Cape Malay populations were estimated to be 16.9%, 20.8% and 46.3%, respectively. The inclusion of haplotype data helped resolve the geographic ancestry of some of the haplogroups to Indian ancestry (H-M52 and L-M11). Using these haplogroups only, we estimated that the Indian contribution in the Cape 'Coloureds', Johannesburg 'Coloureds', and Cape Malays is 4.2%, 4.6% and 9.3%, respectively.

3.4.2.7 *Eurasian origin of Y-chromosome lineages*

Given the distribution of some Y-chromosome haplogroups in both Europe and Asia, they could not be unambiguously resolved to either Asian or European ancestry. These haplogroups have been referred to as having Eurasian ancestry. In the present study haplogroups FG-M213, J-p12f2, and J-M172 are examples of Y-chromosomes that are referred to as Eurasian Y-chromosomes.

3.4.2.7.1 Haplogroup FG-M213

Haplogroup FG-M213 was observed at a frequency of 12.5% in the Cape 'Coloured', 16.7% in the Cape Malay, 6.2% in the Johannesburg 'Coloured' and 4.1% in the SA White populations (Figures 3.2 and 3.4). This haplogroup has been identified mostly in European and Asian populations (Figure 3.5). The highest frequencies were observed in Europeans [33.3% in Georgians, 30% in Syrians, 18.8% in Calabrians, 18.3% in Catalans, 19.3% in Lebanese and 10% in Italians] (Semino *et al.* 2000). Among Asians, it has been found at frequencies of 22.7%, 15.4%, 12.2% and 8.3% in the Peruru Brahmins (Figure 3.5), Siddis, Vizag Brahmins, and Valmiki of South India, respectively (Ramana *et al.* 2001).

Haplotype H51 was found in both the SA 'Coloured' and Asian populations, and was derived from an Asian haplotype from which it differed by a single mutational step (Figure 3.19; Appendix C). There were no haplotypes that were found in both the SA 'Coloured' and the SA white populations, though some haplotypes found in the SA 'Coloured' populations were closely associated with haplotypes found in Asians, but to a lesser extent to those found in the SA Whites. From these haplotype associations one may

suggest a substructure within the haplogroup or origins from other sources than those used in this comparative analysis. Thus, the introduction of this lineage in the SA 'Coloured' population could not be resolved either into an Asian or European contribution, because of a lack of other Asian Y-STR comparative data.

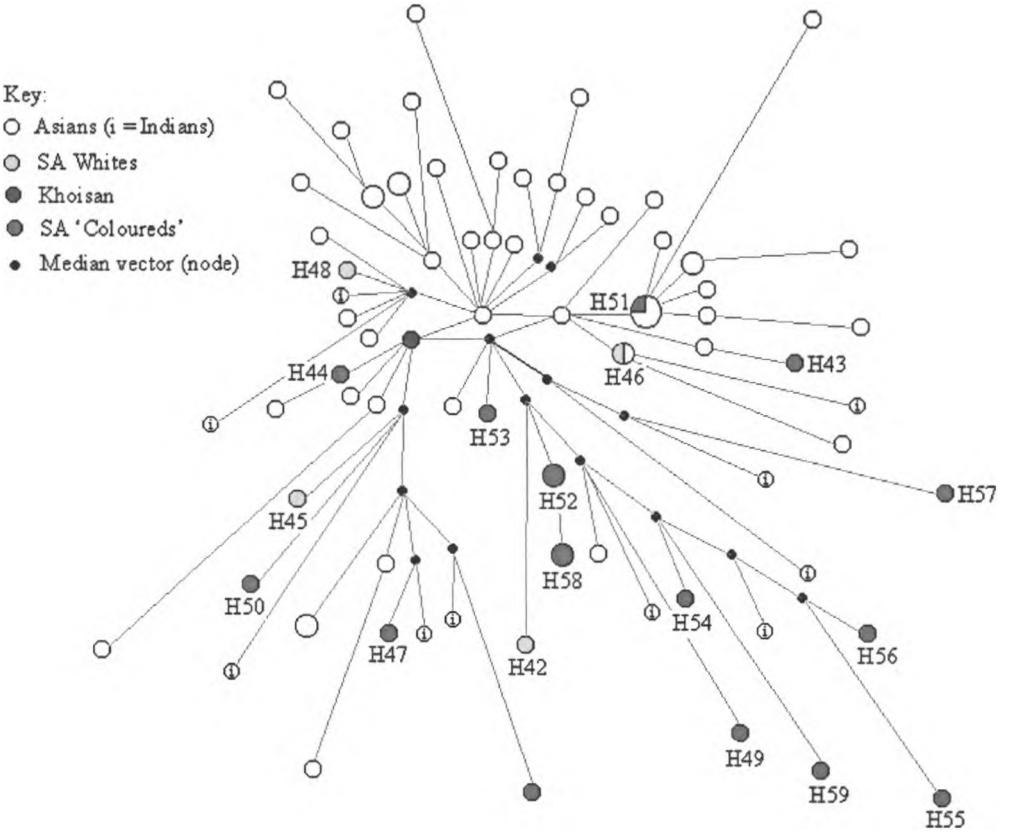


Figure 3.19: Median-joining network showing the relationship between Y-STR haplotypes within haplogroup FG-M213 found in the Asian, SA White, Khoisan and SA 'Coloured' populations. DYS388 is excluded in this comparative analysis that includes SA Indian, Malaysian, Chinese, Russian and Anatolian populations.

3.4.2.7.2 Haplogroup J-p12f2

Haplogroup J-p12f2 was not found in any of the three SA 'Coloured' populations examined, except in the SA Whites at a frequency of 1% (Figures 3.2 and 3.4). Hence, neither the Europeans nor the Asians have introduced this lineage in the SA 'Coloureds'. This haplogroup occurs at high frequencies of 33.1% in the Iraqis (Figure 3.5), 30% in the Syrians, 13% in the Lebanese and 3.3% in the Turks (Semino *et al.* 2000; Al-Zahery *et al.* 2003). Haplogroup J-p12f2 network shows the relationship between the haplotypes found in both the SA Indian and SA White populations (Figure 3.20). H80 found in the SA White individual is the commonest haplotype from which a haplotype found in a SA Indian was derived (Appendix C). Other haplotypes found in this haplogroup were of Asian/Indian origin and derived from the nodes. A haplotype from a SA Bantu-speaking individual was also derived from a node, which could be of either Asian or European origin. From, these haplotype associations, we therefore suggest a Eurasian origin of this haplotype in 1% of the SA Whites.

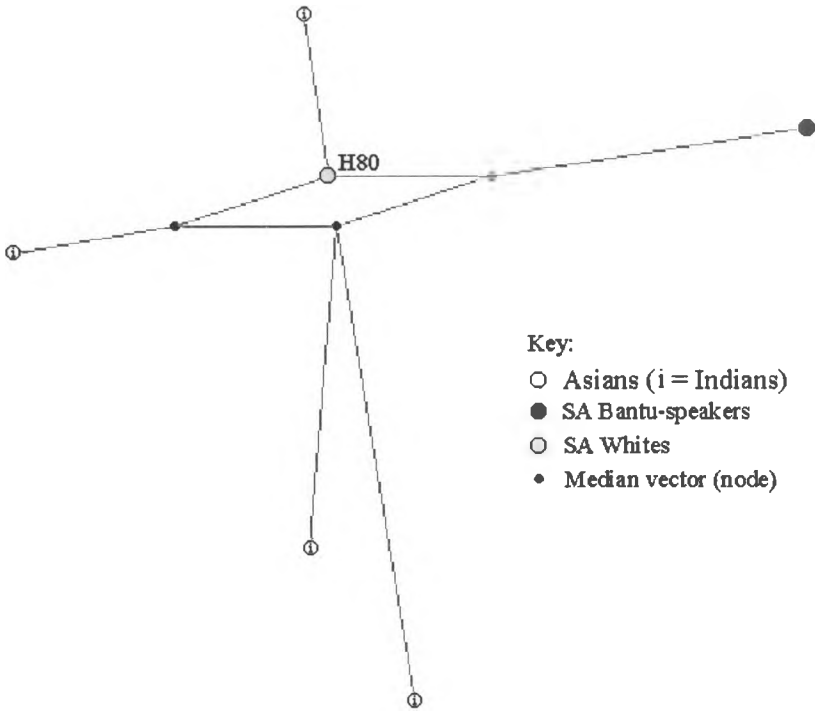


Figure 3.20: Median-joining network showing the relationship between Y-STR haplotypes within haplogroup J-12f2 found in the Asian and SA White populations. DYS388 is excluded in this comparative analysis that includes SA Indian, Malaysian, Chinese, and Russian populations.

3.4.2.7.3 Haplogroup J-M172

Lineage J-M172 which is derived from J-p12f2 was found at frequency of 11.1% in the Cape Malay, 7.7% Johannesburg 'Coloured', 6.3% 'Cape Coloured' and 2.1% SA White Populations (Figures 3.2 and 3.4). Haplogroup J-M172 lineages have been identified quiet considerably in both Asian and European populations with the highest frequency of 40% in the Turkish, 33% Georgian, 29% Lebanese, 23.5% Albanians and 21.6% Calabrian populations from Europe (Semino *et al.* 2000). These lineages have also been found in Indians ranging in frequency from 21.2% in the Punjab, 20.7% in Gujarat, and 14% in Konka Brahmin (Kivisild *et al.* 2003). Regarding network analysis, haplotype H93 was found in both the SA 'Coloured' and SA Indian populations (Figure 3.21; Appendix C). H94 found in the SA 'Coloureds' differed from H93 by two mutations. H90, H91 and 92 also found in the SA 'Coloureds' were derived from a common node with a SA Indian. On the other hand, haplotypes H84 and H88 were found in the SA White, SA Indian and SA 'Coloured' populations. H87 found in one SA 'Coloured' individual differed from these closest neighbour haplotypes (H84 and H88) by one mutational step. Moreover, H82 was found in both the SA White and SA 'Coloured' groups. Other haplotypes found in the SA 'Coloureds' were derived from the nodes that could either be of Asian or European origins. Thus, we suggest a Eurasian contribution of these J-M172 Y-chromosome lineages in the SA 'Coloured' populations.

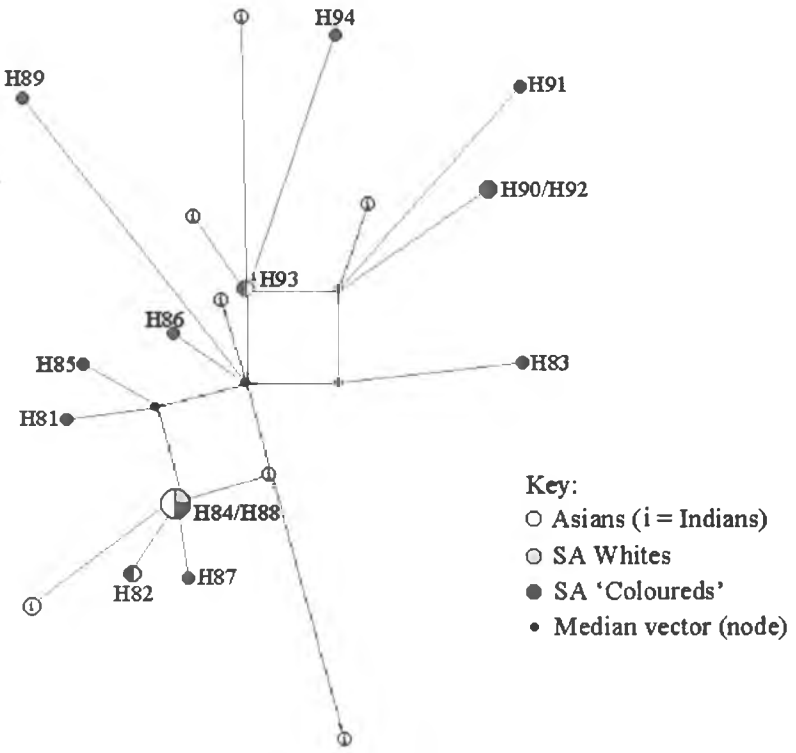


Figure 3.21: Median-joining network showing the relationship between Y-STR haplotypes within haplogroup J-M172 found in the Asian and SA White populations. DYS388 is excluded in this comparative analysis that includes SA Indian (Soodyall, unpublished), Malaysian, Chinese, and Russian populations. Note: Haplotypes H84/H88 and H90/H92 (Appendix C) are the same when the DYS388 is excluded.

3.4.2.8 *Summary*

Therefore, Eurasian derived Y chromosome haplogroups in the Johannesburg 'Coloured', Cape 'Coloured', and Cape Malay populations were estimated to be of 18.5%, 22.9% and 35.2%, respectively.

The overall contributions of Y-chromosome haplogroup and haplotype lineages in the present-day 'Coloured' population from the African, Asian, European and Eurasian ancestry are summarized in Table 3.4. The contributions were estimated through a direct counting method of the overall frequencies based on the combined biallelic and STR data. Due to effective population size, our data is thus sensitive to potential sampling bias.

Table 3.4: Total genetic contributions of parental population in the South African 'Coloured' population.

		Haplotype contributions (%)					
		African		**Asian**		**European**	**Eurasian**
POPULATION	N	Bantu-speakers	Khoisan	Mixed	Indian	Mixed	Mixed
Cape 'Coloureds'	48	27.1	4.2	16.6	4.2	25.0	22.9
Cape Malays	54	...	...	37.0	9.3	18.5	35.2
Johannesburg 'Coloureds'	65	24.6	...	12.3	4.6	40.0	18.5
Random Whites	97	1.0	...	2.1	...	82.5	14.4
Total	264						

N = number of individuals.
Note: mixed refers to different population groups represented.

3.4.3 *Population Affinities*

F_{ST} genetic distances based on the use of 18 haplogroups were used to examine the genetic affinities of the four populations surveyed in the present study. The pair-wise F_{ST} distance values were statistically significant ($P<0.05$) for a number of population pairs, except for the 'Cape Coloured' and 'Johannesburg Coloured' populations (Table 3.5), suggesting that these two groups have similar Y-chromosome composition. The association of these groups in the UPGMA tree presented in Figure 3.22 reinforced this observation. A NJ tree constructed before was not consistent with the actual Fst distances; hence a UPGMA tree was instead constructed.

Table 3.5: Slatkin's population pair-wise linearised Fst distances (Below the Diagonal) and P-values (Above the Diagonal) between the SA 'Coloured' and SA White populations.

POPULATION	1	2	3	4
1 Cape 'Coloured'	...	*	*0.188*	*
2 Cape Malay	0.045	...	*	*
3 Johannesburg 'Coloured'	0.005	0.046	...	*
4 White	0.291	0.326	0.154	...

Note: * = Significantly different values ($P<0.05$), and non-significantly different values ($P>0.05$) are italicized (see the above diagonal).

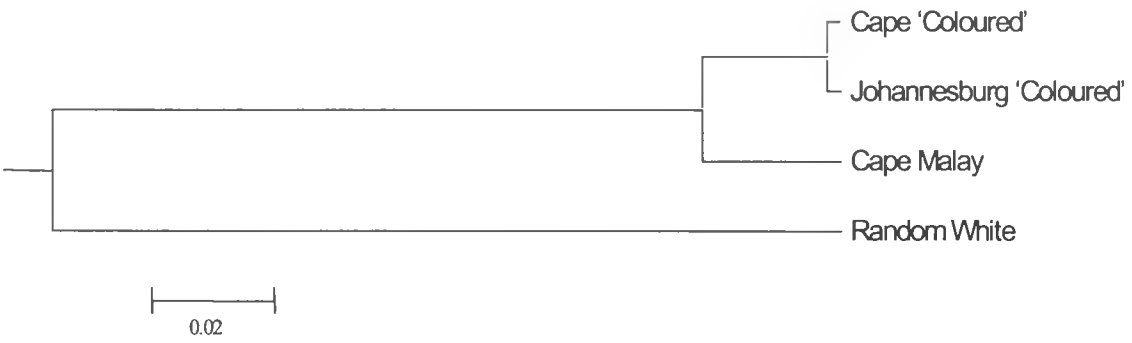


Figure 3.22: Unweighted pair group method with arithmetic mean (UPGMA) tree illustrating the relationships among the SA 'Coloured' and SA White populations, based on the Slatkin linearised Fst distances.

To assess the affinities of the South African 'Coloured' population with their presumed parental populations, a neighbour joining (NJ) tree based on reduced haplogroup data (14; due to lack of data on some of the biallelic markers in the comparative data) was constructed from Slatkin pair-wise F_{ST} values (Table 3.6; Figure 3.23). The pair-wise F_{ST} distance values were statistically significant ($P<0.05$) for many population pairs, except for the Johannesburg 'Coloured' and (Cape 'Coloured', Dutch and French) populations (Table 3.6). On the other hand, the Cape Malays appeared to be not statistically different from the SA Indians (Table 3.6). Populations cluster in three major branches of the tree. Cluster 1 includes the Bantu-speakers from South Africa as well as other Bantu-speakers from other regions in sub-Saharan Africa. Both the Nama and !Kung groups cluster together (Cluster 2). Cluster 3 includes the European, Asian, Cape Malays and Johannesburg 'Coloureds', with the Cape Malay showing closer affinity with Asian populations and the Johannesburg 'Coloured' lying intermediate between the Asian and European populations. The Cape

'Coloured' population is placed at a branch that lies intermediate between the Khoisan (Cluster 2) and Eurasian (Cluster 3) populations.

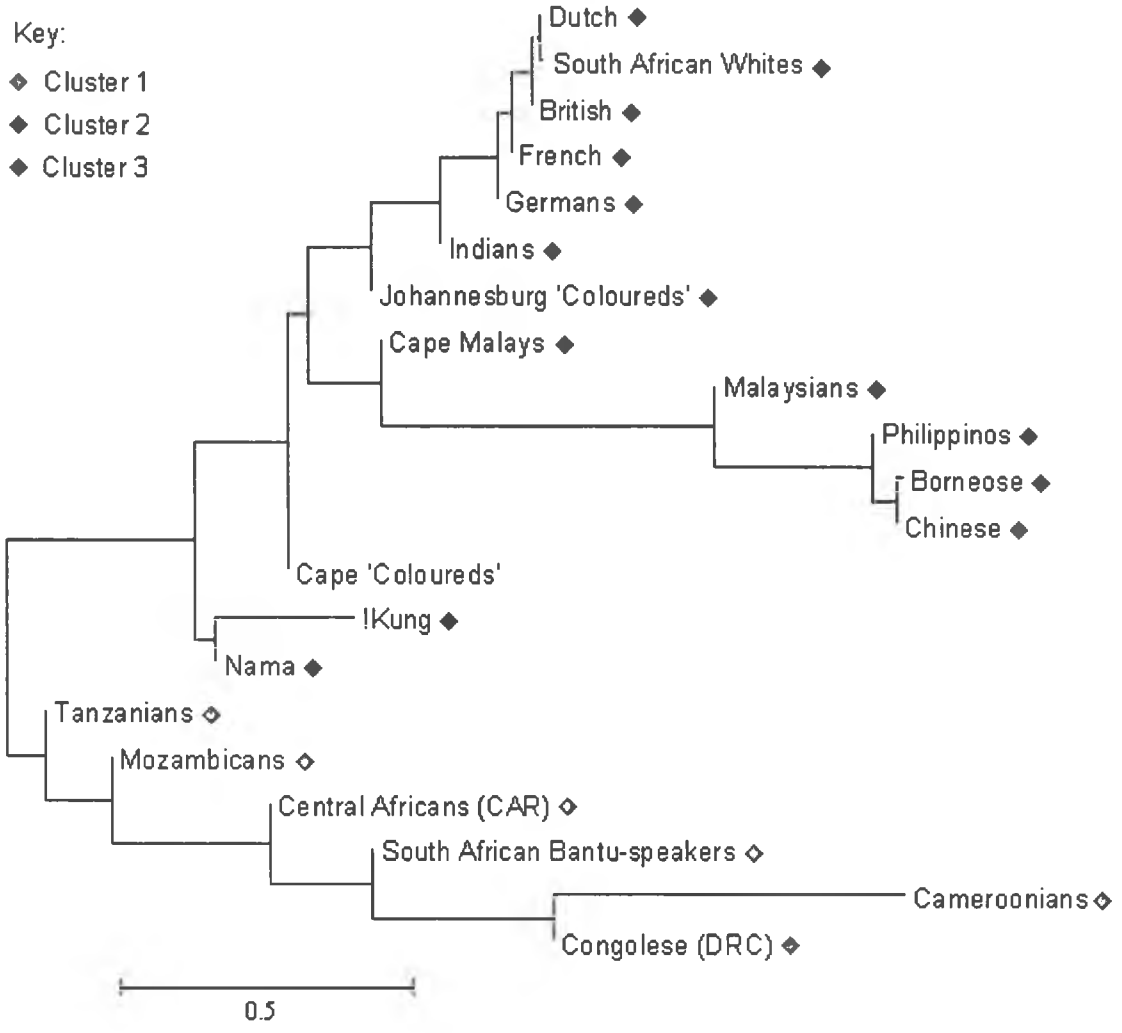


Figure 3.23: Unrooted neighbour-joining (NJ) tree showing the relationships among the SA 'Coloured' and parental populations, based on the Slatkin linearised Fst distances.

Table 3.6: Slatkin's population pair-wise linearised Fst distances (Below the Diagonal) and the corresponding P-values (Above the Diagonal) between South African 'Coloured' groups and their presumed parental populations.

POPULATION	BOR	BRI	CAR	CAM	CCO	CMA	CHI	DRC	DUT	FRE	GER	IND	JCO	KUN	MAL	MOZ	NAM	PHI	SAB	SAW	TAN
BOR		*	*	*	*	*	*0.416*	*	*	*	*	*	*	*	*0.267*	*	*	*0.921*	*	*	*
BRI	1.774		*	*	*	*	*	*	*0.99*	*0.327*	*0.386*	*	*	*	*	*	*	*	*	*0.802*	*
CAR	1.889	1.389		*	*	*	*	*0.178*	*	*	*	*	*	*	*	*	*	*	*	*	*
CAM	4.057	3.664	0.057		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
CCO	0.585	0.198	0.377	0.733		*	*	*	*	*	*	*	*0.188*	*	*	*	*	*	*	*	*
CMA	0.565	0.226	0.726	1.213	0.045		*	*	*	*	*	*0.05*	*	*	*	*	*	*	*	*	*
CHI	0.002	1.774	1.889	4.656	0.556	0.519		*	*	*	*	*	*	*	*0.416*	*	*	*0.426*	*	*	*
DRC	2.158	1.676	0.006	0.045	0.433	0.803	2.189		*	*	*	*	*	*	*	*	*	*	*0.406*	*	*
DUT	1.839	—	1.435	3.758	0.221	0.251	1.846	1.732		*0.287*	*0.277*	*	*	*	*	*	*	*	*	*0.99*	*
FRE	1.419	—	1.152	2.823	0.102	0.116	1.39	1.37	0.009		*0.99*	*0.05*	*0.05*	*	*	*	*	*	*	*0.079*	*
GER	1.506	—	1.197	3.395	0.089	0.101	1.489	1.442	0.009	—		*0.158*	*0.119*	*	*	*	*	*	*	*0.129*	*
IND	0.919	0.096	0.862	1.566	0.091	0.032	0.844	0.983	0.11	0.051	0.045		*	*	*	*	*	*	*	*	*
JCO	0.651	0.096	0.465	0.821	0.005	0.046	0.622	0.523	0.11	0.036	0.029	0.041		*	*	*	*	*	*	*	*
KUN	1.292	0.902	0.909	1.902	0.343	0.419	1.241	1.132	0.949	0.697	0.698	0.528	0.375		*	*	*	*	*	*	*
MAL	0.014	0.922	1.307	3.325	0.257	0.228	0.007	1.517	0.981	0.674	0.678	0.42	0.312	0.726		*	*	*0.327*	*	*	*
MOZ	1.388	0.097	0.082	0.259	0.214	0.435	1.352	0.079	1.021	0.743	0.747	0.559	0.289	0.648	0.797		*	*	*	*	*0.644*
NAM	0.958	0.457	0.44	1.117	0.092	0.186	0.9	0.565	0.495	0.302	0.294	0.248	0.116	0.123	0.433	0.239		*	*	*	*
PHI	—	1.617	1.815	4.219	0.519	0.506	—	2.092	1.683	1.272	1.35	0.823	0.584	1.196	0.031	1.295	0.858		*	*	*
SAB	1.777	1.455	0.024	0.076	0.42	0.785	1.775	—	1.498	1.225	1.256	0.944	0.518	0.999	1.296	0.059	0.504	1.722	—	*	*0.079*
SAW	1.696	—	1.39	2.458	0.291	0.326	1.693	1.599	—	0.039	0.038	0.145	0.154	1.084	1.081	1.123	0.635	1.579	1.468	—	*
TAN	1.336	0.925	0.055	0.237	0.207	0.425	1.286	0.058	0.975	0.708	0.712	0.537	0.269	0.56	0.744	—	0.206	1.238	0.036	1.096	—

Note. * = Significantly different values ($P<0.05$), and non-significantly different values ($P>0.05$) are italicized (see the above diagonal)

Key for abbreviated population names is given in Table 2.1

CHAPTER 4

DISCUSSION AND CONCLUSION

4.1 *Paternal ancestry of the South African 'Coloured' population*

The usefulness of genetic markers such as autosomal, Y-chromosome, and mitochondrial DNA (mtDNA) polymorphisms in assessing genetic structure of admixed populations has been well documented (Parra *et al.* 1998; Sans *et al.* 2002; Vieira *et al.* 2002, Soodyall *et al.* 2003). As a consequence, it has prompted the recent interest of elucidating geographic origins of the paternal ancestry as well as their contributions in the admixed populations of South Africa. This study employed a set of the most informative NRY-specific biallelic and STR polymorphisms at the different loci to delineate the geographic specific haplogroups and resolve haplotype variations within haplogroups, respectively, in the modern SA 'Coloured' population. The populations under study included the Cape Malays and the 'Coloured' groups from Cape Town and Johannesburg.

Biallelic polymorphisms have revealed heterogeneity of Y-chromosome haplogroup lineages (A-R) in the total gene pool of the SA 'Coloured' population, which are indicative of genetic admixture from the African, Asian, European and Eurasian paternal ancestral sources (Figures 3.2, 3.4, 3.5, 4.1). These findings corroborated historic data from the past 350 years when reconstructing the geographic origins of the male founders. The arrival of Europeans has been considered as one of the events that has contributed in

shaping the history of South Africa. First they conquered and resulted in displacing the indigenous people of southern Africa, viz. the Khoisan (whom where initially been displaced from the interior of the country to the coastal lines by the Bantu-speakers) from their original habitats. Secondly, the introduction of slaves from different parts of the world, have resulted in a mixture of the gene pool of the present-day South African population. Offspring from unions between people of different geographic origins resulted in the generation of the South African 'Coloured' population (Marais 1957; Nurse *et al.* 1985; Watson 1990; Shell 1994; Eldredge and Morton 1994; Erusmus 2001; Jeppie 2001).

Earlier genetic studies based on the use of blood groups and serogenetic markers have also shown support for the African, Asian and European ancestry in the two major 'Coloured' groups including the present-day Cape Malays and Cape 'Coloureds'. These studies revealed that the total gene pool of the Cape 'Coloured' population was composed of 46% African, 32% Western Europeans and 22% Asian genes whereas in the Cape Malay, Africans, Europeans and Asians had contributed 25%, 33% and 42%, respectively (Botha 1972, cited in Jenkins 1972). However, these studies have not demonstrated by degree of difference the extent at which males and females from parental populations have contributed in the total gene pool of these groups. In the present study, it was possible to assess how males have contributed in shaping the gene pool of the South African 'Coloured' population. Other ongoing studies in the HGDDRU using mtDNA are examining how females have contributed in shaping the gene pool of these populations.

NRY-specific biallelic haplogroup and STR haplotype frequency data were separately employed to deduce the level of genetic diversity **within** the SA 'Coloured' and SA White populations. These data revealed a high degree of genetic diversity; $h > 0.887$ at the haplogroup level and $h > 0.993$at the haplotype level in all the three SA 'Coloured' groups as compared with the SA Whites, since these values were higher than those obtained in the SA Whites (Table 3.1). High diversities were as a result of genetic admixture from African, Asian and European ancestries (Figures 3.2). Moreover, the most frequent alleles around particular repeats at the different STR loci were not only detected in the SA Whites (see Figure 3.3), other commonest alleles found at some loci, for example 10 repeats at DYS391 locus and 11 repeats at DYS392 locus identified in the SA 'Coloureds', were from African and Asian ancestral contributions, respectively. The same trend of a higher degree of genetic diversity was also observed, when the two data sets were used jointly to estimate the overall lineage diversity and the mean number of pair-wise differences within the SA 'Coloured' and SA White populations (Table 3.2).

4.2 *Founding haplogroups for the South African 'Coloureds'*

Haplogroups A, B and E (including E-M35) were used to resolve the African ancestral contributions in the present-day 'Coloured' groups from Cape Town and Johannesburg, with the exception of the Cape Malays. The lack of Y-chromosomes traced to African ancestry in the Cape Malays (see Figure 4.1; Table 3.4) tends to support the historical data. We estimated the total African contribution in the Cape 'Coloureds' and Johannesburg 'Coloureds' to be 31.3% and 24.6%, respectively (Figure 4.1), of which 27.1% and 23.1% were traced to contributions from Bantu-speakers (Figures 3.2 and 3.4). The Khoisan ancestral contribution inferred from the distribution of haplogroup A lineages was only 4.2% in the Cape 'Coloured' group. The combination of biallelic and STR data in each of the populations enabled further resolution of Y-chromosomes, particularly that of E-M35, from which it was estimated that 1.5% of E-M35 lineages in the Johannesburg 'Coloured' group were derived from Khoisan paternal ancestry. The latter was based on the association of E-M35 haplotypes in the network presented in Figure 3.10.

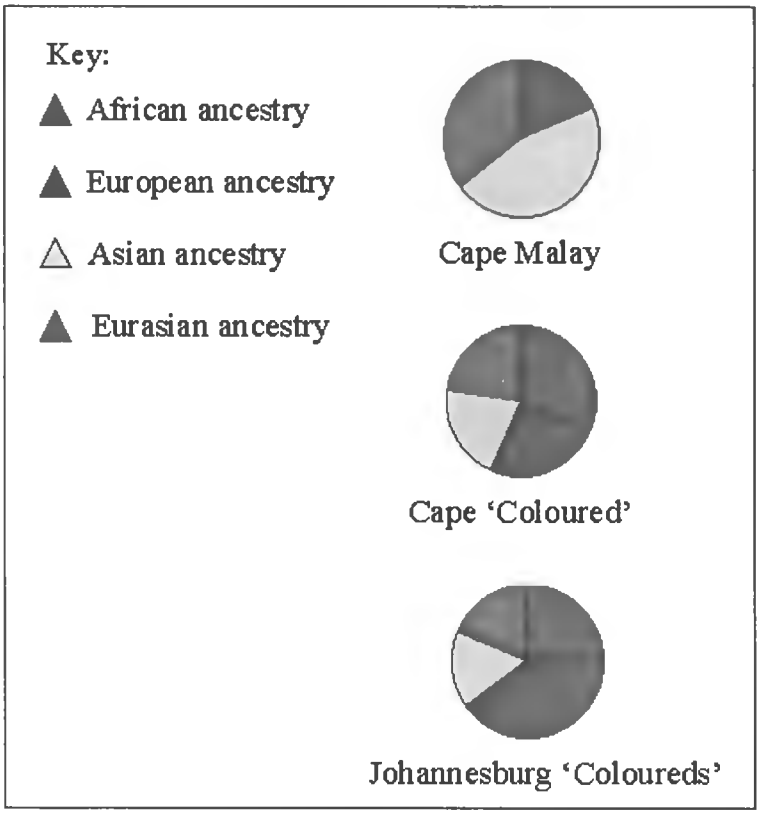


Figure 4.1: Proportion of Y-chromosome contributions from African, European, Asian and Eurasian ancestry in the Cape Malay, Cape 'Coloured' and Johannesburg 'Coloured' populations from South Africa.

Collectively, haplogroup E-M35, I-M170 and R-M207 lineages (Figures 3.1 and 3.3) that are European related (Semino *et al.* 2001), have unambiguously traced the European paternal ancestral contributions in the Cape Malays, Cape 'Coloureds' and Johannesburg 'Coloureds' to be of 18.5%, 25% and 40%, respectively (Figure 4.1; Table 3.4). Haplogroup R-M207 lineages were among the other European specific lineages, the most contributed lineage in the South African 'Coloured' populations (Figures 3.2 and 3.4). Historically, it has been postulated that the introduction of the European lineages in the total gene pool of the SA 'Coloured' population was promoted by a singular shortage of European females at the beginning of the Cape colony, which prompted unions between European males and indigenous African women (Nurse *et al.* 1985).

Y-chromosome modal haplotype H148 defined by haplogroup R-M207 (Appendix C), was found at a frequency of 10.3% in the SA White population, and was also found in the three SA 'Coloured' groups at varying frequencies of 6.3%, 3.1% and 1.9% in the Cape 'Coloureds', Johannesburg 'Coloureds' and Cape Malays, respectively (Figure 3.12). It is possible that this lineage was introduced by a single male founder in the SA 'Coloured' population, but since H148 is found widely elsewhere in South Africa and the rest of the world, the possibility of multiple introductions at various times in the history of South Africa cannot be excluded.

Haplogroup C-M130, H-M52, KMN-M9, L-M11, O-M175, and PQ-M74 lineages, were useful in resolving the Asian ancestral contributions in the Cape Malay, Cape 'Coloured' and Johannesburg 'Coloured' populations. These contributions were estimated to be 46.3%, 20.8% and 16.9%, in that order (Figure 4.1; Table 3.4). It was possible to dissect out the Indian paternal contributions from these data. Approximately 9.3%, 4.2% and 4.6% of the lineages found in the Cape Malay, Cape 'Coloured' and Johannesburg 'Coloured' groups, respectively, were traced to Indian founders carrying the H-M52 and H-M11 haplogroups (Figures 3.15 and 3.17).

It was not surprising to find that the highest proportion of the Asian ancestral contribution was observed in the Cape Malays. These findings are compatible with historical data that suggest that Cape Malay ancestors were the descendents of the Caucasoid Muslims of Bengal who were brought to the Cape from the islands of Indonesian Archipelago in Southeast Asia. Moreover, it is also suggested that the marked Indian contribution in Cape Malay ancestors could have been from the admixture process between them and the Gujarat (Nurse *et al.* 1985).

Haplogroup O-M175, which is informative for tracing Southeast Asian lineages (Kayser *et al.* 2003), was found at a high frequency of 12.5% in the Cape 'Coloureds' and 13% in the Cape Malays, suggesting a gene flow between the two groups (Figure 3.2). The Asian YAP derived haplogroup D lineages that are exclusive to the Japanese and Tibetan populations (Su *et al.* 1999) were not found in our sample. This finding is inconsistent with the historic data, and it could be as a result of artifacts of sampling.

Unfortunately, even with the use of our informative collection of Y-chromosome biallelic and STR markers, it was not possible to resolve the geographic origins of the paternal ancestors who contributed haplogroup lineages FG-M213 and J-M172. Therefore, we referred to the ancestry of these lineages as Eurasian. Also in network analysis, many of the haplotypes found in SA 'Coloured' population tended to be associated with either European or Asian haplotypes. A major problem was the paucity of sufficient comparative data to resolve the ancestries of the haplotypes further. Thus the Eurasian ancestry were estimated to be 18.5%, in the Johannesburg 'Coloureds', 22.9% in the Cape 'Coloureds' and 35.2% in the Cape Malays.

4.3 *Population affinities*

With the use of the Fisher's exact test (extended version) of population pair-wise differentiation that determines the measure of genetic variation **within** each population, this study found no significance difference ($P>0.01$ and $P>0.05$) between the Cape 'Coloured' and Johannesburg 'Coloured' populations (Table 3.3). The similar trend was revealed by population pair-wise Fst P-values ($P>0.05$), which determines the measure of genetic variation **between** populations (Tables 3.5). Therefore, the Cape 'Coloureds' were more closely related to the Johannesburg 'Coloureds' than they are to the Cape Malays, since they tended to cluster together on the UPGMA tree (Figure 3.23).

The three SA 'Coloured' groups were also related to the presumed parental population groups. There was no significant difference (P >0.05) between the 'Coloured' groups from Cape Town and Johannesburg. Furthermore, no significant difference (P >0.05) was found between the Johannesburg 'Coloureds' and the French as well as German populations. This may be due to the high frequency of European derived Y-chromosome lineages in Johannesburg 'Coloured' group. Even though, there was no significant difference between the Johannesburg 'Coloureds' and Cape 'Coloureds', the latter was placed at a branch lying intermediate between the Khoisan and Eurasian populations. The Johannesburg 'Coloureds' were more closely related to the Indians and Europeans, and the Cape Malays showed a closer affinity with the Asian populations, particularly the Malaysians.

4.4 *Significance of population studies in disease management*

Understanding of the population history, for example, genetic structure of populations and processes that have contributed in producing the spectrum of variations in populations, is of paramount importance in understanding the aetiology of disease. A new method for mapping disease genes has been described using admixed populations (Parra *et al.* 1998). The linkage disequilibrium (LD) created when two ethnically distinct populations reproduce is useful in mapping disease genes showing high prevalence among the parental populations (Chakraborty and Weiss 1988). For example, non-insulin dependent diabetes mellitus and obesity, hypertension, lung and prostate cancer, and other anthropological traits have been studied by this method.

Also, several studies making use of linkage disequilibrium (LD), and studies of admixed populations have demonstrated the power of mapping disease genes by admixture linkage disequilibrium (MALD) (Stephens *et al.* 1994; Parra *et al.* 1998). Thus, admixed populations have a significant contribution to make when trying to map disease loci.

Other studies have made use of haploid genetic markers including Y-chromosome and mitochodrial DNA to clarify the genetic origins and establish a method for genetic mapping (Vieira e*t al.* 2002). Given that there is high prevalence of tuberculosis (TB) in the Western Cape, particularly that of the Cape 'Coloured' population, knowledge of the ancestries that have contributed in shaping the gene pool of the populations would contribute to studies focusing on the genetic susceptibility to TB. Also, the sound effects of founder effects on some diseases, for example, lipoid proteinosis in the SA 'Coloured' populations from Namaqualand, could be traced from a joint analysis of haploid markers and other haplotype data. "For recently admixed populations, linkage disequilibrium can be found over relatively large distances of up to 10 or 20 cM, proffering the scenario of performing whole genome linkage disequilibrium with only a few hundred markers and using a case/control or family triad-based approach, rather than the families needed for linkage" (Vieira e*t al.* 2002). Haploid marker haplotypes may be ideal for using to perform a whole genome linkage disequilibrium screen in conjunction with autosomal markers, because theoretically they tend to express the same paternal or maternal founder ethnic groups, and might share the same genetic causes of autosomal conditions.

4.5 *Significance of haploid markers in Forensic applications*

Several people are interested in understanding their ancestry; as a result researchers have used genetic ancestry tracing involving haploid genetic markers such as Y-chromosome and mtDNA to uncover the ancestry of a number of different populations including the Lemba (southern African tribe whose history and customs have long suggested a Jewish history) and Melungeons, a 'mixed ancestry' group in eastern Tennessee and Virginia. Currently, plans have been proposed to offer genetic ancestry tracing to Africa-Americans who want to trace their genetic lineages back to Africa (Elliot and Brodwin 2002). The implementation of these valuable genetic applications in the South African context could be of a great help in resolving issues of social discourse relating to the 'Coloured' identities and nomenclature, as described in section 1.2 of this thesis. However, identities including cultural practices, memories, rituals and modes of being would remain unresolved. In the case of Medical Forensics, validation of Y-chromosome haploid markers was assessed, and it has been found that the NRY-specific STR loci that exhibit moderate diversity, chance of exclusion, and discrimination power are useful in forensic and paternity testing applications, see section 1.4.5.1 of this thesis.

4.6 *Conclusion*

Genetic polymorphisms within the haploid, paternally transmitted non-recombining portion of Y-chromosome have been able to resolve and substantiate the complexity of historical human admixture processes that were to a large extent associated with the European colonization of South Africa, 351 years ago. The present-day study has undoubtedly traced back the geographic origins of the paternal ancestry of the two 'Coloured' groups from Cape Town and Johannesburg to Africa, Asia and Europe, and that of the Cape Malays to Asia and Europe. Varying degrees of genetic admixture were discerned in the total gene pool of the three 'Coloured' groups, of which the highest proportion of African, Asian and European ancestry were observed in the Cape 'Coloured', Cape Malay and Johannesburg 'Coloured' populations, respectively. The study has also shown that the 'Coloureds' living in Cape Town and Johannesburg are not significantly different from each other even though they are geographically separated. Moreover, two groups were found to significantly different from the Cape Malays, suggesting that historical events were more significant in shaping their gene pool.

REFERENCES

Al-Zahery N, Semino O, Benuzzi G, Magri C, Passarino G, Torroni A, Santachiara-Benerecetti AS. (2003) Y-chromosome and mtDNA polymorphisms in Iraq, a crossroad of the early human dispersal and of post-Neolithic migrations. *Mol. Phylogenet. Evol.* **28**:458-472.

Avise J, Arnold J, Ball RM, Birmingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC. (1987) Intraspecific phylogeography: The molecular bridge between populations genetics and systematics. *Ann. Rev. Ecol. Syst.* **18**:489-522.

Bandelt HJ, Forster P, Rohl A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**:37-48.

Barać L, Pericic M, Klaric IM, Rootsi S, Janicijevic B, Kivisild T, Parik J, Rudan I, Villems R, Rudan P. (2003) Y chromosomal heritage of Croatian population and its island isolates. *Eur. J. Hum. Genet.* **11**:535-542.

Bergen AW, Wang CY, Tsai J, Jefferson K, Dey C, Smith KD, Park SC, Tsai SJ, Goldman D. (1999) An Asian-Native American paternal lineage identified by RPS4Y resequencing and by microsatellite haplotyping. *Ann. Hum. Genet.* **63**:63-80.

Böeseken AJ. (1977) Slaves and Free Blacks at the Cape, 1658-1700. *Tafelberg Uitgewers*, Cape Town.

Bosch E, Calafell F, Comas D, Oefner PJ, Underhill PA, Bertranpetit J. (2001) High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between northwestern Africa and the Iberian Peninsula. *Am. J. Hum. Genet.* **68**:1019-1029.

Botha MC. (1972) Blood group gene frequencies: An indication of the genetic constitution of population samples in Cape Town. *Am. J. Roentgenol Radium Ther. Nucl. Med.* **115**: 1- 27.

Caglia A, Tofanelli S, Coia V, Boschi I, Pescarmona M, Spedini G, Pascali V, Paoli G, Destro-Bisol G. (2003) A study of Y-chromosome microsatellite variation in sub-Saharan Africa: a comparison between FST and RST genetic distances. *Hum. Biol.* **75**:313-330.

Capelli C, Wilson JF, Richards M, Stumpf MP, Gratrix F, Oppenheimer S, Underhill P, Pascali VL, Ko TM, Goldstein DB. (2001) A predominantly indigenous paternal heritage for the Austronesian-speaking peoples of insular Southeast Asia and Oceania. *Am. J. Hum. Genet.* **68**:432-443.

Casanova M, Leroy P, Boucekkine C, Weissenbach J, Bishop C, Fellous M, Purrello M, Fiori G, Siniscalco M. (1985) A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance. *Science* **230**:1403-1406.

Cavalli-Sforza LL and Feldman MW. (2003) The applications of molecular genetic approaches to the study of human evolution. *Nat. Genet.* **33**:266-275.

Chakraborty R, Weiss KM. (1988) Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. *Proc. Natl. Acad. Sci.* **85**:9119-9123.

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

Cruciani F, Santolamazza P, Shen P, Macaulay V, Moral P, Olckers A, Modiano D, Holmes S, Destro-Bisol G, Coia V, Wallace DC, Oefner PJ, Torroni A, Cavalli-Sfroza LL, Scozzari R, and Underhill PA. (2002) A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. *Am. J. Hum. Genet.* **70**:1197-1214.

de Knijff P. (2000) Messages through bottlenecks: On the combined use of slow and fast evolving polymorphic markers on the human Y-Chromosome. *Am. J. Hum. Genet.* **67**:1055-1061.

Deacon HJ, Deacon J. (1999) Human beginnings in South Africa: Uncovering the secrets of the Stone Age. *David Philip Publishers*, Cape Town.

Eldredge EA and Morton F. (1994) Slavery in South Africa: Captive labour on the Dutch frontier. *University of Natal Press*, Pietermaritzburg.

Elliott C, Brodwin P. (2002) Identity and genetic ancestry tracing. *BMJ.* **325**:1469-1471.

Elphick R. (1985) Khoikhoi and the founding of White South Africa. *Ravan Press,* Johannesburg.

Erasmus Z. (2001) Coloured by history, shaped by place: New perspectives on Coloured identities in Cape Town edited by Zimitri Erasmus. *Kwela Books and South African History Online*, Cape Town.

Flores C, Maca-Meyer N, Pérez JA, Gonzalez AM, Larruga JM, and Cabrera VM. (2003) A predominant European ancestry of paternal lineages from Canary Islanders. *Ann. Hum. Genet.* **67**:138-152.

Forster P, Röhl A, Lünnemann P, Brinkmann C, Zerjal T, Tyler-Smith C, Brinkmann B. (2000) A short tandem repeat-based phylogeny for the human Y chromosome. *Am. J. Hum. Genet.* **67**:182-196.

Francalacci P, Morelli L, Underhill PA, Lillie AS, Passarino G, Useli A, Madeddu R, Paoli G, Tofanelli S, Calo CM, Ghiani ME, Varesi L, Memmi M, Vona G, Lin AA, Oefner P, Cavalli-Sforza LL. (2003) Peopling of three Mediterranean islands (Corsica, Sardinia, and Sicily) inferred by Y-chromosome biallelic variability. *Am. J. Phys. Anthropol.* **121**:270-279.

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

Hammer MF and Horai S. (1995) Y chromosomal DNA variation and the peopling of Japan. *Am. J. Hum. Genet.* **56**: 951-962.

Hammer MF, Zegura SL. (1996) The role of the Y chromosome in human evolutionary studies. *Evol. Anthr.* **5**:116-134.

Hammer MF, Karafet TM, Redd AJ, Jarjanazi H, Santachiara-Benerecetti S, Soodyall H, Zegura SL. (2001) Hierarchical patterns of global human Y-chromosome diversity. *Mol. Biol. Evol.* **18**:1189-1203.

Hammer MF, Zegura SL. (2002) The human Y chromosome haplogroup tree: Nomenclature and phylogeography of its major divisions. *Annu. Rev. Anthropol.* **31**:303-321.

Hendricks C. (2001) 'Ominous' Liaisons: Tracing the interface between 'race' and sex at the Cape, pp. 29-44. In: Coloured by history, shaped by place: New perspectives on Coloured identities in Cape Town, *Zimitri Erasmus*. Kwela Books and South African History Online, Cape Town.

Huffman TN. (1982) Archaeology and the ethnohistory of the African Iron Age. *Annu. Rev. Anthropol.* **11**:133-150.

Hurles ME, Nicholson J, Bosch E, Renfrew C, Sykes BC, Jobling MA. (2002) Y chromosomal evidence for the origins of Oceanic-speaking peoples. *Genetics* **160**:289-303.

Jenkins T. (1972) Genetic polymorphisms of man in southern Africa. MD Thesis, *University of London*, London.

Jeppie S. (2001) Reclassifications: Coloured, Malay, Muslim, pp. 80-96. In: Coloured by history, shaped by place: New perspectives on Coloured identities in Cape Town, *Zimitri Erasmus*. Kwela Books and South African History Online, Cape Town.

Jobling MA. (2001) In the name of the father: Surnames and genetics. *Trends. Genet.* **17**:353-357.

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

Jobling MA, Tyler-Smith C. (1995) Fathers and sons: The Y-chromosome and human evolution. *Trends. Genet.* **11**:449-456.

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

Kayser M, Brauer S, Weiss G, Schiefenhovel W, Underhill P, Shen P, Oefner P, Tommaseo-Ponzetta M, Stoneking M. (2003) Reduced Y-chromosome, but not

mitochondrial DNA, diversity in human populations from West New Guinea. *Am. J. Hum. Genet.* **72**:281-302.

Kayser M, Brauer S, Weiss G, Underhill PA, Roewer L, Schiefenhovel W, Stoneking M. (2000) Melanesian origin of Polynesian Y chromosomes. *Curr. Biol.* **10**:1237-1246.

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

Kayser M, Krawczak M, Excoffier L, Dieltjes P, Corach D, Pascali V, Gehrig C, Bernini LF, Jespersen J, Bakker E, Roewer L, de Knijff P. (2001) An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. Am J *Hum. Genet.* **68**:990-1018.

Kivisild T, Bamshad MJ, Kaldma K, Metspalu M, Metspalu E, Reidla M, Laos S, Parik J, Watkins WS, Dixon ME, Papiha SS, Mastana SS, Mir MR, Ferak V, Villems R. (1999) Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. *Curr. Biol.* **18**:1331-1334.

Kivisild T, Rootsi S, Metspalu M, Mastana S, Kaldma K, Parik J, Metspalu E, Adojaan M, Tolk HV, Stepanov V, Golge M, Usanga E, Papiha SS, Cinnioğlu C, King R, Cavalli-

Sforza L, Underhill PA, Villems R. (2003) The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am. J. Hum. Genet.* **72**:313-332.

Knight A, Underhill PA, Mortensen HM, Zhivotovsky LA, Lin AA, Henn BM, Louis D, Ruhlen M, Mountain JL. (2003) African Y chromosome and mtDNA divergence provides insight into the history of click languages. *Curr. Biol.* **13**:464-473.

Kumar S, Tamura K, Jakobsen I, Nei M. (2001) MEGA2: Molecular evolutionary genetics analysis. *Arizona State University*, Tempe.

Lane AB, Soodyall H, Arndt S, Ratshikhopha ME, Jonker E, Freeman C, Young L, Morar B, Toffie L. (2002) Genetic substructure in South African Bantu-speakers: Evidence from autosomal DNA and Y-chromosome studies. *Am. J. Phys. Anthropol.* **119**:175-185.

Lell JT, Wallace DC. (2000) The peopling of Europe from maternal and paternal perspectives. *Am. J. Hum. Genet.* **67**:1376-1381.

Macaulay V, Richards M, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonne-Tamir B, Sykes B, Torroni A. (1999) The emerging tree of West Eurasian mtDNAs: A synthesis of control-region sequences and RFLPs. *Am. J. Hum. Genet.* **64**:232-249.

Malaspina P, Cruciani F, Ciminelli BM, Terrenato L, Santolamazza P, Alonso A, Banyko J, Brdicka R, Garcia O, Gaudiano C, Guanti G, Kidd KK, Lavinha J, Avila M, Mandich P, Moral P, Qamar R, Mehdi SQ, Ragusa A, Stefanescu G, Caraghin M, Tyler-Smith C,

Scozzari R, Novelletto A. (1998) Network analyses of Y-chromosomal types in Europe, northern Africa, and western Asia reveal specific patterns of geographic distribution. *Am. J. Hum. Genet.* **63**:847-860.

Marais JS. (1957) The Cape Coloured people. *Witwatersrand University Press*, Johannesburg.

Miller SA, Dykes DD, Polesky HF. (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16**:1215.

Nebel A, Filon D, Brinkmann B, Majumder PP, Faerman M, Oppenheim A. (2001) The Y chromosome pool of Jews as part of the genetic landscape of the Middle East. *Am. J. Hum. Genet.* **69**:1095-1112.

Nebel A, Filon D, Weiss DA, Weale M, Faerman M, Oppenheim A, and Thomas MG. (2000) High-resolution Y chromosome haplotypes of Israeli and Palestinian Arabs reveal geographic substructure and substantial overlap with haplotypes of Jews. *Hum. Genet.* **107**:630-641.

Nei M. (1987) Molecular evolutionary genetics. *Columbia University Press*, New York.

Nurse GT, Weiner JS, Jenkins T. (1985) The peoples of Southern Africa and their affinities. *Clarendon Press*, London.

Parra EJ, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R, Forrester T, Allison DB, Deka R, Ferrell RE, Shriver MD. (1998) Estimating African American admixture proportions by use of population-specific alleles. *Am. J. Hum. Genet.* **63**:1839-1851.

Puzyrev VP, Stepanov VA, Golubenko MV, Puzyrev KV, Maksimova NR, Khar'kov VN, Spiridonova MG, Nogovitsyna AN. (2003) MtDNA and Y-chromosome lineages in the Yakut population. *Russian Journal of Genetics* **39**:816-822.

Qamar R, Ayub Q, Mohyuddin A, Helgason A, Mazhar K, Mansoor A, Zerjal T, Tyler-Smith C, Mehdi SQ. (2002) Y-chromosomal DNA variation in Pakistan. *Am. J. Hum. Genet.* **70**:1107-1124.

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

Ramana GV, Su B, Jin L, Singh L, Wang N, Underhill P, Chakraborty R. (2001) Y-chromosome SNP haplotypes suggest evidence of gene flow among caste, tribe, and the migrant Siddi populations of Andhra Pradesh, South India. *Eur. J. Hum. Genet.* **9**:695-700.

Reddy T. (2001) The Constitution of Coloured Subjects in South Africa, pp. 64-79. In: Coloured by history, shaped by place: New perspectives on Coloured identities in Cape Town, *Zimitri Erasmus*. Kwela Books and South African History Online, Cape Town.

Richards MB, Macaulay VA, Bandelt HJ, Sykes BC. (1998) Phylogeography of mitochondrial DNA in western Europe. *Ann. Hum. Genet.* **62**:241-260.

Roewer L, Arnemann J, Spurr NK, Grzeschik KH, Epplen JT. (1992) Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts. *Hum. Genet.* **89**:389-394.

Roewer L, Krawczak M, Willuweit S, Nagy M, Alves C, Amorim A, Anslinger K, Augustin C, Betz A, Bosch E, Caglia A, Carracedo A, Corach D, Dekairelle AF, Dobosz T, Dupuy BM, Furedi S, Gehrig C, Gusmao L, Henke J, Henke L, Hidding M, Hohoff C, Hoste B, Jobling MA, Kargel HJ, de Knijff P, Lessig R, Liebeherr E, Lorente M, Martinez-Jarreta B, Nievas P, Nowak M, Parson W, Pascali VL, Penacino G, Ploski R, Rolf B, Sala A, Schmidt U, Schmitt C, Schneider PM, Szibor R, Teifel-Greding J, Kayser M. (2001) Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes. *Forensic Science International* **118**:106-113.

Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, Amorim A, Amos W, Armenteros M, Arroyo E, Barbujani G, Beckman G, Beckman L, Bertranpetit J, Bosch E, Bradley DG, Brede G, Cooper G, Corte-Real HB, de Knijff P, Decorte R, Dubrova YE, Evgrafov O, Gilissen A, Glisic S, Golge M, Hill EW, Jeziorowska A, Kalaydjieva L, Kayser M, Kivisild T, Kravchenko SA, Krumina A, Kucinskas V, Lavinha J, Livshits LA, Malaspina P, Maria S, McElreavey K, Meitinger TA, Mikelsaar AV, Mitchell RJ, Nafa K, Nicholson J, Norby S, Pandya A, Parik J, Patsalis PC, Pereira L, Peterlin B, Pielberg G, Prata MJ, Previdere C, Roewer L, Rootsi S, Rubinsztein DC, Saillard J, Santos FR, Stefanescu G, Sykes BC, Tolun A, Villems R, Tyler-Smith C, Jobling MA.

(2000) Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am. J. Hum. Genet.* **67**:1526-1543.

Rousset F, Raymond M. (1985) Testing heterozygote excess and deficiency. *Genetics* **140**:1413-1419.

Sans M, Weimer TA, Franco MH, Salzano FM, Bentancor N, Alvarez I, Bianchi NO, Chakraborty R. (2002) Unequal contributions of male and female gene pools from parental populations in the African descendants of the city of Melo, Uruguay. *Am. J. Phys. Anthropol.* **118**:33-44.

Santos FR, Pandya A, Tyler-Smith C, Pena SD, Schanfield M, Leonard WR, Osipova L, Crawford MH, Mitchell RJ. (1999) The central Siberian origin for native American Y chromosomes. *Am. J. Hum. Genet.* **64**:619-628.

Scozzari R, Cruciani F, Santolamazza P, Malaspina P, Torroni A, Sellito D, Arredi B, Destro-Bisol G, De Stefano G, Rickards O, Martinez-Labarga C, Modiano D, Biondi G, Moral P, Olckers A, Wallace DC, Novelletto A. (1999) Combined use of biallelic and microsatellite Y-chromosome polymorphisms to infer affinities among African populations. *Am. J. Hum. Genet.* **65**:829-846.

Schneider S, Kueffer JM, Roessli D, Excoffier L. (1997) Arlequin ver 1.1: A software for population genetic data analysis. *Genetics and Biometry laboratory - University of Geneva*, Switzerland.

Schneider S, Roessli D, Excoffier L. (2000) Arlequin: A software for population genetic data. *Genetics and Biometry Laboratory - University of Geneva*, Switzerland.

Seielstad MT, Hebert JM, Lin AA, Underhill PA, Ibrahim M, Vollrath D, Cavalli-Sforza LL. (1994) Construction of human Y-chromosomal haplotypes using a new polymorphic A to G transition. *Hum. Mol. Genet.* **3**:2159-2161.

Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill PA. (2000) The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: A Y chromosome perspective. *Science* **10**:1155-1159.

Semino O, Santachiara-Benerecetti AS, Falaschi F, Cavalli-Sforza LL, Underhill PA. (2002) Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny. *Am. J. Hum. Genet.* **70**:265-268.

Shastry BS. (2002) SNP alleles in human disease and evolution. *J. Hum. Genet.* **47**:561-566.

Shell RC. (1994) Children of Bondage: A social history of the slave society at the Cape of goodhope, 1652-1834. *Witwatersrand University Press*, Johannesburg.

Shen P, Wang F, Underhill PA, Franco C, Yang WH, Roxas A, Sung R, Lin AA, Hyman RW, Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ. (2000) Population genetic

implications from sequence variation in four Y chromosome genes. *Proc. Natl. Acad. Sci.* **97**:7354-7359.

Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, Pyntikova T, Ali J, Bieri T, Chinwalla A, Delehaunty A, Delehaunty K, Du H, Fewell G, Fulton L, Fulton R, Graves T, Hou SF, Latrielle P, Leonard S, Mardis E, Maupin R, McPherson J, Miner T, Nash W, Nguyen C, Ozersky P, Pepin K, Rock S, Rohlfing T, Scott K, Schultz B, Strong C, Tin-Wollam A, Yang SP, Waterston RH, Wilson RK, Rozen S, Page DC. (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **19**:825-837.

Soodyall H, Nebel A, Morar B, Jenkins T. (2003) Genealogy and genes: Tracing the founding fathers of Tristan da Cunha. *Eur. J. Hum. Genet.* **11**:705-709.

Spurdle AB, Hammer MF, Jenkins T. (1994) The Y Alu polymorphism in southern African populations and its relationship to other Y-specific polymorphisms. *Am. J. Hum. Genet.* **54**:319-330.

Stephens JC, Briscoe D, O'Brien SJ. (1994) Mapping by admixture linkage disequilibrium in human populations: Limits and guidelines. *Am. J. Hum. Genet.* **55**:809-824.

Strachan T, Read AP. (1996) Human molecular genetics. *BIOS Scientific Publishers Ltd. John Wiley & Sons Inc.*, New York.

Su B, Xiao J, Underhill P, Deka R, Zhang W, Akey J, Huang W, Shen D, Lu D, Luo J, Chu J, Tan J, Shen P, Davis R, Cavalli-Sforza L, Chakraborty R, Xiong M, Du R, Oefner P, Chen Z, Jin L. (1999) Y-Chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age. *Am. J. Hum. Genet.* **65**:1718-1724.

Thomas MG, Bradman N, Flinn HM. (1999) High throughput analysis of 10 microsatellite and 11 diallelic polymorphisms on the human Y-chromosome. *Hum. Genet.* **105**:577-581.

Thomas MG, Parfitt T, Weiss DA, Skorecki K, Wilson JF, le Roux M, Bradman, Goldstein DB. (2000) Y chromosome travelling south: The Cohen modal haplotype and the origins of the Lemba-the "Black Jews of Southern Africa". *Am. J. Hum. Genet.* **66**:674-686.

Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ. (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res.* **7**:996-1005.

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

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

Vieira AR, Karras JC, Orioli IM, Castilla EE, Murray JC. (2002) Genetic origins in a South American clefting population. *Clin. Genet.* **62**:458-463.

Watson RL. (1990) The slave question: Liberty and property in South Africa. *Witwatersrand University Press*, Johannesburg.

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

Whitfield LS, Sulston JE, Goodfellow PN. (1995) Sequence variation of the human Y chromosome. *Nature* **378**:379-380.

Y Chromosome Consortium. (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res.* **12**:339-348.

Zerjal T, Wells RS, Yuldasheva N, Ruzibakiev R, Tyler-Smith C. (2002) A genetic landscape reshaped by recent events: Y-chromosomal insights into central Asia. *Am. J. Hum. Genet.* **71**:466-482.

Electronic-Database Information

The URLs for data in this thesis are as follows:

ARLEQUIN version 1.1: http://anthropologie.unige.ch/arlequin

ARLEQUIN ver 2.0: http://lgb.unige.ch/arlequin/software/2.000/manual/Arlequin.pdf

HISTORICAL SITE: http://www.sahistory.org.za/pages/mainframe.htm

MEGA ver 2.0: http://megasoftware.net/

NETWORK versions 3.0 and 4.0: http: www.fluxus.engineering.com/

Y-STRs: http://ystr.charite.de/

APPENDIX A

SOLUTIONS

DNA Markers

Commercial preparations of molecular weight markers were used.

ROX Marker

Fluorescently labelled GeneScan™ - 500 ROX™ (prepared by Applied Biosystems) was used as an internal size standard for Y-STR fragments.

UV-light visible DNA molecular weight marker (1kb ladder)

100µl of λ-DNA (Gibco BRL) was mixed with 50µl Ficoll dye and 850µl TE and stored at 4°C. Use 1µl/5µl PCR-RFLP product per well in agarose to size DNA fragments.

Oligonucleotide primers

10µM primers

Concentrations (mM) of oligonucleotide primer stocks (stored at –70°C) were determined spectrophotometrically and then converted to µM. Primers were diluted with ddH_20 to concentrations of 10µM and stored at –20°C.

3.3µM primers

Add 66µl of 10µM primer to 134µl ddH_20 to produce 200µl of 3.3µM primer.

Agarose gels

2% Agarose gel (100ml)

Dissolve 2g molecular grade agarose in 100ml of 1X TBE and boil to dissolve. Add 3µl of 10mg/ml Ethidium bromide just before pouring.

3% Agarose gel (100ml)

Dissolve 3g molecular grade agarose in 100ml of 1X TBE and boil to dissolve. Add 3µl of 10mg/ml Ethidium bromide just before pouring.

2% NuSieve agarose gel (100ml)

Dissolve 2g NuSieve agarose in 100ml of 1X TAE and boil to dissolve. Add 3µl of 10mg/ml Ethidium bromide just before pouring.

Polyacrylamide gel

4.0% denaturing polyacrylamide gel (PAG)

For a 100ml gel mix: add together 10.6ml de-ionized 40% polyacrylamide [19:2 (w/v) acrylamide: bisacrylamide], 10ml of 10X TBE, 36g of Urea and ~50ml of ddH_20. Stir the mix until urea has dissolved, filter any visible particles and store at 4°C in the dark. To 10ml of the gel mix, add 50µl of 10% APS and 7µl TEMED just before pouring on 12cm gel plates.

Buffers

50X Tris-acetic acid acetic acid EDTA (TAE) buffer

Dissolve 121g of Tris in 28.6ml of glacial acetic acid and 50ml of 0.5M EDTA (pH 8.0). Adjust the volume to 500 ml with ddH_20.

1X Tris-acetic acid acetic acid EDTA (TAE) buffer

Mix 20ml of 50X TAE with 980ml of ddH_20 to make 1l of 1X TAE.

10X Tris boric acid EDTA (TBE) buffer

Dissolve 108g of hydroxymethylaminomethane (Tris base) and 55g Boric acid in 900ml ddH_20. Add 40ml of 0.5M EDTA (pH 8.0). Adjust to pH 8.0 with hydrochloric acid (HCl), make up to 1l and autoclave.

Tris-EDTA (TE) buffer

Dissolve 121.1mg Tris and 37.2mg Na_2EDTA in 90ml of ddH_20. Adjust the pH to 8.0 and make up to 100ml with ddH_20.

1M Tris-HCl buffer

Dissolve 121g hydroxymethylaminomethane (Tris base) to 800ml ddH_20. While stirring, adjust the pH to 8.0 with 1M HCl and make up to 1l with ddH_20 and autoclave.

Dyes

Dextran-formamide dye

Dissolve 100mg of blue Dextran (Sigma) in 5ml formamide, aliquot and store at 20°C.

Ficoll dye

Dissolve 50g sucrose, 0.1g bromophenol and 10g Ficoll in 90ml ddH_20. Add 10ml of 0.5M EDTA and allow to dissolve overnight. Filter through Whatmann 1MM filter paper and make up to 100ml with ddH_20.

Other reagents

10% Ammonium persulphate (APS)

Dissolve 1g of APS in 10ml ddH_20 and store the mix at 4°C in the dark.

Bovine serum albumin (BSA)

Dissolve 1g of BSA in 10ml of ddH_20, aliquot into 1ml amounts and store at –20°C.

0.5M ethylenediamine tetra-acetic acid (EDTA)

Dissolve 9.3g Na_2EDTA in 40ml water while adjusting pH to 8.0 with 10N sodium hydroxide (NaOH), when totally dissolved make up to 50ml and autoclave.

Formamide

Stir together 0.5g amberlite MB-6 resin and 100ml formamide at room temperature for 30min. Filter twice through Whatmann No 1 filter paper and store at –20°C.

Magnesium chloride ($MgCl_2$)

Dissolve 20.34g $MgCl_2$ in 90ml of ddH_20, make up to 100ml, and dispense into 25ml aliquots and autoclave.

APPENDIX B

Table 1

Y-chromosome PCR-based biallelic polymorphism reagents and product sizes

Polymorphism	SRY10831	M168	M150	M130	YAP	M40	M2	M35	M78	M75	M213	M170	M52	p12f2	M172	M9	M175	M11	M74	M207
PCR reagents	Final concentrations (units are the same as those for the stock soloutions)																			
$MgCl_2$ (25 mM)	1 5	2 5	1.5	1.5	1.5	1	1.5	2	2 5	2	2.5	2 5	2 5	1 5	2	1.5	1 5	3 5	1.5	1.5
Primer F[a] (10 μM)	0.4	0 4	0 4	0.4	0.4	1 5	0 2	0 4	0 2	0 3	0 4	0.4	0 4	0.3	0.4	0.2	0 2	0.4	0.3	0 4
Primer R[b] (10 μM)	0.4	0 4	0.4	0.4	0.4	1.5	0 2	0 4	0 2	0 3	0 4	0 4	0 4	0.3	0 4	0.2	0.2	0.4	0 3	0 4
BSA (10 mg/ml)	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PCR product size(bp)	167	473	167	91	150 (YAP-) 450 (YAP+)	225	148	186	301	189	409	129	164	88 (p12f2) 148 (M2)	148	340	444	215	385	423

Note F[a] = forward primer and R[b] = reverse primer

APPENDIX B

Table 2

Y-chromosome PCR-based STR polymorphism reagents, product and repeat sizes.

Polymorphism	DYS19	DYS388	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393
PCR reagents	Final concentrations (units are the same as those for the stock soloutions)							
$MgCl_2$ (25 mM)	1.5	1	1	1	1.5	1.5	1	1.5
Primer F[a] (4 μM)	1.6	0.08	0.08	0.08	1.6	1.6	1.6	1.6
Primer R[b] (4 μM)	1.6	0.08	0.08	0.08	1.6	1.6	1.6	1.6
BSA (10 mg/ml)	1	1	1	1	1	1	1	1
Allele size ranges (bp)	174–211	128–143	237–265	355–383	191–227	271–299	233–263	108–136
Number of repeats	10–19	12–17	9–16	26–33	18–27	7–14	6–16	9–16

Note: F[a] = forward primer and R[b] = reverse primer.The number of repeats are consistent with those on the http://ystr.charite.de/

APPENDIX C

Table 1

Y-STR haplotypes within biallelic haplogroups and their frequencies in the SA 'Coloured' and SA White populations.

									Frequency in population[a]				
Haplotypes	**DYS19**	**DYS388**	**DYS389I**	**DYS389II**	**DYS390**	**DYS391**	**DYS392**	**DYS393**	**CCO**	**CMA**	**JCO**	**SAW**	**TOTAL**
A-SRY10831.1													
H1	17	12	13	16	22	10	10	12	1				**1**
H2	17	12	12	17	22	10	10	13	1				**1**
B-M150													
H3	15	10	15	18	24	10	11	13			1		**1**
H4	15	10	14	19	24	10	11	13	1				**1**
H5	15	10	14	18	24	10	11	13			2		**2**
H6	15	10	13	18	24	10	11	13			1		**1**
C-M130													
H7	16	13	14	18	25	11	11	12		1			**1**
H8	16	13	13	17	24	10	11	13			1		**1**
H9	15	14	12	18	24	10	10	14		1			**1**
H10	15	14	12	16	21	10	11	12		1			**1**
H11	15	13	14	18	25	11	11	13			1		**1**
H12	15	13	12	17	21	10	11	12	1				**1**
H13	15	13	12	16	21	11	11	13		1			**1**
H14	15	13	12	16	21	11	11	12		1			**1**
H15	14	16	13	17	21	10	11	13			1		**1**
E-M2													
H16	17	12	14	17	21	10	11	14			1		**1**
H17	16	12	14	17	21	10	11	15	2		1	1	**2**
H18	16	12	14	16	21	10	11	15	1				**1**
H19	16	12	13	18	21	10	11	14			1		**1**
H20	16	12	13	18	21	10	11	13	1				**1**
H21	16	12	12	17	21	10	11	13			1		**1**
H22	15	12	14	17	21	10	11	15	1				**1**
H23	15	12	14	16	21	10	11	13	3				**3**
H24	15	12	13	19	21	10	11	13			1		**1**
H25	15	12	13	18	21	11	11	13	1		1		**2**
H26	15	12	13	18	21	10	11	13			3		**3**
H27	15	12	13	17	21	10	11	12	1				**1**
H28	15	12	12	17	21	11	11	14	1				**1**
E-M35													
H29	15	12	14	18	24	10	11	13			1		**1**
H30	14	12	14	17	24	11	11	14			1		**1**
H31	14	12	14	16	23	11	11	13				1	**1**
H32	13	12	12	18	23	10	11	13			1		**1**
H33	13	12	10	17	24	9	11	14			1		**1**

Table 1 (Continued)

Haplotypes	DYS19	DYS388	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	Frequency in population[a] CCO	CMA	JCO	SAW	TOTAL
H34	14	12	13	17	25	10	11	13	1			1	**2**
H35	13	12	13	18	24	10	11	13			1		**1**
H36	13	12	13	17	24	10	11	14			1		**1**
H37	13	12	13	17	24	10	11	13				1	**1**
H38	13	12	13	17	23	10	11	13		1			**1**
H39	13	12	13	16	25	10	11	13		1			**1**
E-M75													
H40	14	12	12	16	25	10	11	13	1		1		**2**
H41	14	12	12	16	24	10	11	13			1		**1**
FG-M213													
H42	16	14	12	16	23	9	11	12				1	**1**
H43	16	12	13	16	21	10	11	14	1				**1**
H44	16	12	12	17	22	10	11	13			1		**1**
H45	15	15	12	17	24	10	11	12				1	**1**
H46	15	13	12	16	23	10	11	14				1	**1**
H47	15	12	14	17	24	10	11	13		1			**1**
H48	15	12	14	17	22	10	11	14				1	**1**
H49	15	12	13	15	23	10	14	13		1			**1**
H50	15	12	12	17	23	10	14	13		1			**1**
H51	15	12	12	16	21	10	11	14			1		**1**
H52	14	14	12	16	23	10	11	13	2		1		**3**
H53	14	14	12	16	22	10	11	13	1				**1**
H54	14	13	13	16	24	10	11	13			1		**1**
H55	14	12	13	16	24	11	13	12		1			**1**
H56	14	12	13	16	24	10	13	13		1			**1**
H57	14	12	13	16	21	11	13	13		1			**1**
H58	14	12	12	16	24	10	11	13		3			**3**
H59	13	12	14	15	23	9	11	13	1				**1**
H60	13	12	13	17	24	10	13	13	1				**1**
H-M52													
H61	17	12	14	14	22	10	11	12			1		**1**
H62	16	12	14	17	22	10	11	12	1				**1**
H63	16	12	13	16	22	10	11	12			1		**1**
H64	15	12	14	17	22	10	11	13		1			**1**
H65	15	12	13	16	22	10	11	12	1				**1**
I-M170													
H66	16	13	13	17	23	10	12	15				1	**1**
H67	16	9	12	15	23	11	11	13			1		**1**
H68	15	14	13	17	24	10	12	15				1	**1**
H69	15	14	12	17	22	10	11	13				1	**1**
H70	15	14	12	16	22	10	11	13				1	**1**

Table 1 (Continued)

Haplotypes	DYS19	DYS388	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	Frequency in population[a] CCO	CMA	JCO	SAW	TOTAL
H71	15	13	14	18	22	10	12	15			1		1
H72	14	15	13	17	24	9	11	12				1	1
H73	14	15	12	15	22	10	11	13	1				1
H74	14	14	13	16	23	10	11	13		1		1	2
H75	14	14	12	17	23	10	11	13				1	1
H76	14	14	12	17	22	10	11	13	1			1	2
H77	14	14	12	16	23	10	11	12				1	1
H78	14	14	12	16	22	10	11	14				1	1
H79	14	14	12	16	22	10	11	13			1	3	4
J-p12f2													
H80	14	16	13	17	23	10	11	12				1	1
J-M172													
H81	16	15	13	16	24	10	11	12			1		1
H82	16	15	12	16	24	10	11	12	1			1	2
H83	15	16	13	18	23	10	11	13	1				1
H84	15	16	12	16	24	10	11	12				1	1
H85	15	15	14	16	24	10	11	12		1			1
H86	15	15	13	16	22	10	11	12		1			1
H87	15	15	12	17	24	10	11	12			1		1
H88	15	15	12	16	24	10	11	12	1				1
H89	15	14	14	18	23	11	11	12		1			1
H90	14	16	13	18	23	10	11	13			1		1
H91	14	15	14	15	23	9	11	13		1			1
H92	14	15	13	18	23	10	11	13			2		2
H93	14	15	13	16	23	10	11	12		1			1
H94	14	15	13	13	23	10	11	12		1			1
KMN-M9													
H95	15	12	14	16	22	10	11	12		1			1
H96	15	12	13	15	21	10	11	12		2			2
H97	14	15	14	16	23	10	11	14		1			1
H98	14	14	12	16	23	10	11	12		1			1
H99	14	14	12	16	22	10	11	13		1			1
H100	14	12	14	17	25	11	14	13	1				1
H101	14	12	14	16	24	11	14	14				1	1
H102	14	12	13	16	25	11	13	13			1		1
H103	13	12	14	17	24	11	11	14		1			1
H104	13	11	14	16	22	10	13	13		1			1
L-M11													
H105	15	12	13	16	22	10	14	12		1			1
H106	14	12	12	16	22	11	14	11		1			1
H107	14	12	12	16	22	10	15	11		1			1
H108	14	12	12	16	22	10	14	11		1	1		2

Table 1 (Continued)

Haplotypes	DYS19	DYS388	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	Frequency in population[a] CCO	CMA	JCO	SAW	TOTAL
O-M175													
H109	17	12	12	16	23	10	12	14		1			1
H110	16	12	12	17	21	9	12	14		1			1
H111	15	15	12	15	21	9	12	13	1				1
H112	15	14	12	17	24	10	13	14	1				1
H113	15	13	12	16	24	10	13	13	1				1
H114	15	13	12	16	23	11	13	13	1				1
H115	15	12	14	15	24	11	13	14			1		1
H116	15	12	13	17	24	10	13	13		1			1
H117	15	12	13	16	26	11	13	14		1			1
H118	15	12	12	18	23	10	14	13			1		1
H119	15	12	12	17	23	10	14	13		2			2
H120	15	12	12	16	23	10	14	13	1	1			2
H121	15	12	12	16	23	10	12	14	1		1		2
PQ-M74													
H122	15	12	14	16	28	10	13	13			1		1
H123	13	12	13	16	22	10	15	13				1	1
R-M207													
H124	16	12	14	16	23	11	13	13				1	1
H125	16	12	12	16	23	10	14	13		1			1
H126	15	13	13	16	24	10	13	14	1				1
H127	15	12	14	16	25	11	13	13				1	1
H128	15	12	13	16	24	11	13	13	1			2	3
H129	15	12	13	16	24	10	13	13				1	1
H130	15	12	13	16	23	10	13	13			1		1
H131	15	12	12	16	25	11	13	12			1		1
H132	14	12	14	17	24	11	13	13			1		1
H133	14	12	14	16	24	11	13	13	1			4	5
H134	14	12	14	16	23	11	13	13		1			1
H135	14	12	14	15	24	11	13	13		1			1
H136	14	12	14	15	24	11	13	12	1	1			2
H137	14	12	13	17	25	11	13	12				1	1
H138	14	12	13	17	25	10	13	13				1	1
H139	14	12	13	17	24	12	13	12				1	1
H140	14	12	13	17	24	11	13	13				3	3
H141	14	12	13	17	24	10	13	13		1	2	1	4
H142	14	12	13	17	23	11	13	13				1	1
H143	14	12	13	16	25	11	14	13				1	1
H144	14	12	13	16	25	11	13	13			1	1	2
H145	14	12	13	16	25	10	13	13			2		2
H146	14	12	13	16	24	11	14	13				2	2
H147	14	12	13	16	24	11	13	14			1	2	3
H148	14	12	13	16	24	11	13	13	3	1	2	10	16
H149	14	12	13	16	24	10	14	12				1	1
H150	14	12	13	16	24	10	13	13			1	7	8

Table 1 (Continued)

Haplotypes	DYS19	DYS388	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	Frequency in population[a] CCO	CMA	JCO	SAW	TOTAL
H151	14	12	13	16	23	12	13	13				1	**1**
H152	14	12	13	16	23	11	14	13				1	**1**
H153	14	12	13	16	23	11	13	13	1	1	1	3	**6**
H154	14	12	13	16	23	11	13	12				1	**1**
H155	14	12	13	16	23	10	13	13	1		1	3	**5**
H156	14	12	13	16	22	11	13	13				1	**1**
H157	14	12	13	16	22	10	13	13				1	**1**
H158	14	12	13	15	24	11	14	13				1	**1**
H159	14	12	13	15	24	11	13	13				1	**1**
H160	14	12	13	15	24	10	13	13				1	**1**
H161	14	12	13	15	23	11	13	13			1		**1**
H162	14	12	13	14	24	10	13	13				1	**1**
H163	14	12	12	17	25	11	14	13				1	**1**
H164	14	12	12	16	24	11	13	13			1		**1**
H165	14	12	12	16	23	11	13	13				1	**1**
H166	14	12	12	16	23	10	10	15			1		**1**
H167	14	12	12	16	22	10	13	13				1	**1**
H168	13	12	13	16	24	9	13	13				1	**1**
H169	13	12	13	16	23	11	14	13				1	**1**
H170	13	12	13	16	23	11	13	13				1	**1**
H171	13	12	13	16	23	10	14	13				1	**1**
H172	13	12	13	16	22	10	13	13				1	**1**
H173	13	12	12	16	23	10	13	13			1		**1**
R-SRY10831.2													
H174	16	13	13	17	25	11	11	13				1	**1**
H175	16	12	14	17	25	10	11	13		1	1		**2**
H176	16	12	13	17	24	11	11	13				1	**1**
H177	16	12	13	17	23	11	11	13		1			**1**
H178	16	12	13	16	26	11	11	13		1			**1**
H179	16	12	13	16	25	10	11	12				1	**1**
H180	15	12	14	18	25	10	11	13			1		**1**
H181	15	12	13	18	25	10	11	13	1				**1**
H182	15	12	13	17	25	11	11	13				3	**3**
H183	15	12	13	17	25	10	11	13			1		**1**
H184	15	12	13	16	25	10	11	13	1			1	**2**
H185	10	12	13	17	24	11	11	13		1			**1**
TOTAL									**48**	**54**	**65**	**97**	**768**

[a]Populations are denoted by codes: CCO=Cape 'Coloured', CMA=Cape Malay, JCO=Johannesburg 'Coloured', SAW=SA Whites.
Note:DYS389II = DYS389IIc.

ETHICS APPROVAL

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

Division of the Deputy Registrar (Research)

COMMITTEE FOR RESEARCH ON HUMAN SUBJECTS (MEDICAL)
Ref: R14/49 Motladiile

CLEARANCE CERTIFICATE PROTOCOL NUMBER M02-05-27

PROJECT Y-Chromosome Variation In The South African Coloured Population

INVESTIGATORS Mr TW Motladiile

DEPARTMENT School of Pathology, NHLS/SAIMR

DATE CONSIDERED 02-05-31

DECISION OF THE COMMITTEE *

Approved unconditionally

DATE 02-07-22 CHAIRMAN .....(Professor P E Cleaton-Jones)

* Guidelines for written "informed consent" attached where applicable.

c c Supervisor: Dr H Soodyall
Dept of School of Pathology, NHLS/SAIMR

Works2\Iain0015\HumEth97.wdb\M 02-05-27

DECLARATION OF INVESTIGATOR(S)

To be completed in duplicate and **ONE COPY** returned to the Secretary at Room 10001, 10th Floor, Senate House, University.

I/we fully understand the conditions under which I am/we are authorized to carry out the abovementioned research and I/we guarantee to ensure compliance with these conditions. Should any departure to be contemplated from the research procedure as approved I/we undertake to resubmit the protocol to the Committee.

DATE 01/07/2002 SIGNATURE .....

PROTOCOL NO.: M 02-05-27

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES