THREADS FROM THE PAST: A GENETIC STUDY OF AFRICAN ETHNIC GROUPS AND HUMAN ORIGINS

A THESIS PRESENTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY IN BIOLOGICAL SCIENCE
BIOLOGICAL ANTHROPOLOGY

AT MASSEY UNIVERSITY

ELIZABETH EVELYN WATSON

1996

I am the vessel of fate
I sail between the reefs of the warp
that stand for life
from the right bank I pass to the left
unreeling my gut
to add to the fabric
then back from the left bank I pass to the right
unreeling my gut
life is perpetual coming and going
a permanent gift of the self

-a Fulbe weaver's invocation (Hampâté Bâ, 1990)

1. ABSTRACT

Mitochondrial DNA (mtDNA) provides an efficient tool for investigating human pre-history and evolution. In this thesis Hypervariable I sequences from 241 individuals representing nine diverse African ethnic groups are presented, and analysed together with published sequences. Each ethnic group studied represents a different combination of geography (East and West Africa), linguistic phylum (Niger-Kordofanian, Nilo-Saharan, and Afroasiatic), and subsistence mode (agricultural and pastoral). The African mtDNA dataset is expanded considerably by the addition of these samples. Questions concerning human evolution and human pre-history are readdressed.

These samples demonstrate a higher diversity in Africa than previously reported. An ethnic group such as the Turkana has a higher mitochondrial diversity than the rest of the world combined. There is some evidence of regionalism, in that some of the clusters are specific to West or East Africa; however, there is also sharing of lineages across Africa suggesting substantial migration and admixture both in the past and more recent times. In Africa, the !Kung being the exception, the mtDNA from one ethnic group does not show a common unique ancestor, rather the ethnic groups are composites of several mtDNA types, with distinct origins and population histories. Combined analysis with Median networks and pairwise distance methods, demonstrate that within Africa approximately 60,000 years ago, there was a phenomenal expansion of one cluster, which was associated with the people who subsequently moved out of Africa. Almost all non-African populations are derived from this cluster. The Out of Africa hypothesis is supported, but there may have been more than one expansion of modern humans out of Africa. Further sampling in Asia and Australia is crucial to conclude the Multiregional-Out of Africa debate.

ABBREVIATIONS

°C	degrees Celsius
μl	microlitre
A.L.F.	Automatic Laser Frequencer (Pharmacia)
AD	*anno Domini* years after the birth of Christ
APS	ammonium persulphate
BC	years before Christ
bp	base pairs
BP	years before present
c.	*circa*
ddNTP	de-deoxyribonucleotide triphosphate
DNA	deoxyribonucleic acid
DNA	deoxyribonucleotide acid
DTT	dithiothreitol
EDTA	ethylenediamine tetraacetic acid
g	grams
HVR-I	Hyper-variable Region I
HVR-II	Hyper-variable Region II
kb	kilobase
kya	thousand years before present
M	Molar
Mya	million years before present
mg	milligram
ml	millilitre

mtDNA	mitochondrial DNA
NaCl	Sodium Chloride
np	nucleotide position with reference to Anderson *et al.*, (1981)
^{o}C	degrees Celsius
PCR	polymerase chain reaction
RFLP	restriction fragment length polymorphism
RNA	ribonucleotide acid
SDS	sodium dodecyl sulphate
TEMED	N,N,N,N, tetramethylethylendiamine
tRNA	transfer ribonucleotide acid

NOTES

The term "race" is something which can not be visualised genetically; in humans there is usually more diversity within a population than between populations. Terms which have racial overtones have been replaced by the geographic origin of the peoples i.e. Asian, European, African, American.

- "African" replaces the term Negroid.
- "European/Middle Eastern" replaces the term Caucasoid and refers to people living in the geographical area of Europe and the Middle East, and who are not known to be recent immigrants.
- "Asian" replaces the term Mongoloid and refers to people living in Asia.
- "San" replaces the term Bushman and refers to hunter-gatherers who speak a Khoisan language. In this study the San population are known as the !Kung.
- "Dama" replaces the term Hottentot and refers to pastoralists who speak a Khoisan language.

I have consciously avoided describing the physical attributes of the people as most groups exhibit high variation, and physical attributes generally do not show genetic affinity. Stature and skin colour are, for instance, more an adaptation to the environment than determining any genetic relationship.

1. ABSTRACT .... **III**

2. INTRODUCTION **20**

2.1 THE OUT OF AFRICA HYPOTHESIS VERSUS THE MULTIREGIONAL HYPOTHESIS 20
2.2 AFRICAN ETHNIC GROUPS 22
2.2.1 Fulbe 26
2.2.2 Hausa 27
2.2.3 Kanuri 28
2.2.4 Kikuyu 29
2.2.5 !Kung 30
2.2.6 Mandenka 31
2.2.7 Pygmy (Biaka and Mbuti) 32
2.2.8 Somali 33
2.2.9 Songhai 34
2.2.10 Tuareg 35
2.2.11 Turkana 36
2.2.12 Yoruba 37
2.3 MITOCHONDRIAL DNA (MTDNA) 38
2.3.1 Characteristics of the Molecule 38
2.3.2 Mitochondrial DNA of Human Population Studies 39
2.4 NUCLEAR DNA STUDIES 40
2.5 STUDIES TO CLASSIFY AFRICAN GROUPS 42
2.5.1 Phenetic 42
2.5.2 Protein Immunology and Allozymes 43
2.5.3 Language 43
2.6 ETHNICITY 44
2.7 LANGUAGE 45
2.7.1 Correlation with Genetics 45
2.7.2 Language Replacement 46
2.7.3 Languages in this Study 46
2.7.4 History of Languages 50
2.8 PRE-HISTORY AND HISTORY OF AFRICA 51
2.8.1 Summary of African Epochs 51
2.8.2 Climate 52
2.8.3 Fossils and Tools 53
2.8.4 Culture 53
2.8.5 Domestication 53
2.8.6 Metal and the Iron Age 55

2.8.7 Population Expansions 55
2.9 Scope of Thesis 56

3. SAMPLE COLLECTION AND ETHICS 58

3.1 Introduction 58
3.2 Ethics 58
3.2.1 Informed Consent 59
3.2.2 Health, Exploitation and Use of Results 59
3.3 Sample Collection 60
3.3.1 Evolution of the Project and Selection Of Ethnic Groups 60
3.3.2 Fieldwork and Sampling Etiquette 65
3.3.3 Future of Samples 67
3.4 Summary 67

4. METHODS AND SEQUENCE DESCRIPTION 69

4.1 Introduction 69
4.2 Methods 69
4.2.1 A: Hypervariable Region I 70
4.2.2 B: HpaI 72
4.2.3 C: Nine base pair repeat 73
4.3 Hypervariable Region I 74
4.3.1 Substitution Rate within the Hypervariable Region I 74
4.3.2 Transitions and Transversions in the Hypervariable Region I 75
4.3.3 Variability within the C-Run Region 76
4.4 *Hpa* I (np 3594) 78
4.5 Nine Base Pair Sequence 80
4.6 African Ethnic Groups 81
*4.6.1 Pairwise distances ** 81
*4.6.2 Hudson Test ** 81
4.6.3 Distance Statistic 82
4.6.4 Ethnic Groups Studied 82
4.6.5 Number of Variable Positions in African populations 86
4.6.6 Simulations of Coalescent Times 87
4.6.7 Correlation between Language, Economy, Geography and Genetics 88
4.6.8 Africa in the World 93
4.6.9 Africa's diversity in comparison to the World's 95
4.7 Discussion 96

5. THE TURKANA PEOPLE OF EAST AFRICA HAVE THE HIGHEST MTDNA GENETIC DIVERSITY 98

5.1 ABSTRACT 98
5.2 INTRODUCTION 98
5.3 RESULTS 101
5.4 DISCUSSION 104

6. MITOCHONDRIAL FOOTPRINTS OF HUMAN EXPANSIONS WITHIN AFRICA 105

6.1 ABSTRACT 105
6.2 INTRODUCTION 105
6.3 MATERIALS AND METHODS 106
6.3.1 Subjects 106
6.3.2 Median Networks application to intraspecific studies 107
6.3.3 Network Analysis 109
6.4 RESULTS 111
6.4.1 Identification of Expansion events 111
6.4.2 Rooting of mtDNA Network 112
6.4.3 Age of Expansion events 115
6.4.4 Geographical Origin of Clusters 117
6.5 DISCUSSION 117
6.5.1 Expansion Events 117
6.5.2 Explanation of Expansions 118
6.5.3 Geographical Origin of Cluster-1 120
6.5.4 Summary 120

7. TREES FROM LANGUAGES AND GENES ARE VERY SIMILAR 121

7.1 ABSTRACT 121
7.2 INTRODUCTION 121
7.3 METHODS 122
7.4 RESULTS 123
7.5 DISCUSSION 123

8. DISCUSSION 125

8.1 PATTERNS WITHIN AFRICA 125
8.1.1 Economy, Language and Geography 125
8.1.2 Ethnicity 125
8.1.3 Migrations and Expansions within Africa 126
8.2 PATTERNS WITHIN THE WORLD 128
8.2.1 Diversity 128
8.2.2 One Mitochondrial Type Founded the Rest of the World 128
8.2.3 Bottlenecks and Expansions 129

8.2.4 MtDNA Survival and Selection *130*
8.3 HUMAN EVOLUTION 133
8.3.1 Fossil Record *133*
8.3.2 Europe *137*
8.3.3 Asia and Australia *138*
8.3.4 Dating/Calibration *139*
8.3.5 Phylogeny *140*
8.4 FUTURE DIRECTIONS 140
8.4.1 HVR-I *140*
8.4.2 Middle Eastern Contribution to the African genepool *140*
8.5 CONCLUSION 142
9. REFERENCES 145
10. APPENDICES 163
10.1 INFORMATION REGARDING THE USE OF SAMPLES COLLECTED IN THIS SURVEY 163
10.2 ETHNOLOGICAL INFORMATION FOR ALL INDIVIDUALS SAMPLED 164
10.2.1 Anuak *165*
10.2.2 Babur *165*
10.2.3 Baribar *165*
10.2.4 Bella *165*
10.2.5 Borana *165*
10.2.6 Bukusu *166*
10.2.7 Dinka *166*
10.2.8 Fulbe *166*
10.2.9 Gusii *170*
10.2.10 Hausa *170*
10.2.11 Higgi *173*
10.2.12 Ibibio *173*
10.2.13 Idoma *173*
10.2.14 Igbo *173*
10.2.15 Junkun *174*
10.2.16 Kalenjin *174*
10.2.17 Kanuri *174*
10.2.18 Kikuyu *175*
10.2.19 Luhyia *177*
10.2.20 Luo *178*
10.2.21 Maasai *181*
10.2.22 Mambila *182*

10.2.23 Margi *182*
10.2.24 Nuer *182*
10.2.25 Omotaboh *184*
10.2.26 Rendille *184*
10.2.27 Sabarumo *185*
10.2.28 Sakuye *185*
10.2.29 Samburu *185*
10.2.30 Shuwa *187*
10.2.31 Somali *187*
10.2.32 Songhai *189*
10.2.33 Tera *190*
10.2.34 Teso *190*
10.2.35 Tiv *190*
10.2.36 Tuareg *190*
10.2.37 Turkana *192*
10.2.38 Yoruba *194*
10.3 SEQUENCES USED IN THE ANALYSIS 196
10.4 SHARED LINEAGES 198
10.5 SEQUENCES GENERATED IN THIS STUDY 200
10.5.1 Fulbe Sequences n=61 .... *200*
10.5.2 Hausa Sequences n=20 .... *202*
10.5.3 Kanuri Sequences n=14 .... *203*
10.5.4 Kikuyu Sequences n=25 .... *204*
10.5.5 Somali Sequences n=27 .... *206*
10.5.6 Songhai Sequences n=10 .... *207*
10.5.7 Tuareg Sequences n=26 .... *208*
10.5.8 Turkana Sequences n=37 .... *210*
10.5.9 Yoruba Sequences n=21 .... *212*
10.5.10 Extra Sequences n=2 .... *213*
10.6 INFORMATION AND CONSENT SHEET 214
10.7 LACTOSE INTOLERANCE 215
10.8 INDIVIDUALS OF CROSS CULTURAL MARRIAGES 216
10.9 OTHER POSITIONS 217
10.10 POPULATION PORTRAITS 222

LIST OF FIGURES

FIGURE 1 REPRESENTATION OF *HPA*I DIGESTIONS 72

FIGURE 2 RESOLVING THE NINE BASE PAIR REPEAT 73

FIGURE 3 VARIABLE POSITIONS IN THE HVR-I REGION....75

FIGURE 4 DISTRIBUTION OF SUBSTITUTIONS ALONG THE HVR-I....76

FIGURE 5 PAIRWISE DISTANCE DISTRIBUTION OF AFRICAN ETHNIC GROUPS 85

FIGURE 6 NUMBER OF VARIABLE POSITIONS VERSUS SAMPLE SIZE FOR AFRICAN POPULATIONS....86

FIGURE 7 NUMBER OF VARIABLE POSITIONS VERSUS SAMPLE SIZE FOR A SELECTION OF AMERICAN AND CIRCUMARCTIC POPULATIONS 86

FIGURE 8 PAIRWISE DISTRIBUTION OF AFRICA VERSUS THE REST OF THE WORLD....95

FIGURE 9 PAIRWISE DISTANCE DISTRIBUTION OF THE TURKANA VERSUS THE WORLD....102

FIGURE 10 PAIRWISE DISTANCE DISTRIBUTION WITHIN ASIA AND AFRICA....103

FIGURE 11 KEY TO NETWORKS 108

FIGURE 12 INTERPOPULATION NETWORK WITHIN AFRICA 110

FIGURE 13 NETWORK OF THE WEST AFRICAN TWIN CLUSTER-II 111

FIGURE 14 THE RELATIONSHIP BETWEEN THE ISOLATED AFRICAN LINEAGES AND THE FIVE EXPANSION CLUSTERS....112

FIGURE 15 GRAPH REPRESENTING THE PERCENTAGE OF EXPANSION CLUSTERS IN EACH POPULATION (%)114

FIGURE 16 PAIRWISE DISTRIBUTION OF CLUSTERS I-VI 116

FIGURE 17 SCHEMATIC REPRESENTATION OF RELATIONSHIPS AND EXPANSIONS OF THE CLUSTERS. 119

FIGURE 18 THE TWO UNROOTED TREES FROM CAVALLI-SFORZA *ET AL.* (1988) 122

FIGURE 19 SCHEMATIC DIAGRAM OF THE RELATIONSHIPS OF THE TURKANA SEQUENCES....127

FIGURE 20 SCENARIOS FOR LOWER MTDNA DIVERSITY OBSERVED OUTSIDE OF AFRICA....129

FIGURE 21 POSSIBLE MODEL OF POPULATION CRASHES IN EURASIA 130

FIGURE 22 MITOCHONDRIAL EVE AND ANATOMICALLY MODERN HUMANS....131

FIGURE 23 MODEL OF DIFFERENT EXPANSIONS OUT OF AFRICA AND THEIR CONTRIBUTION TO TODAY'S MTDNA GENEPOOL 136

LIST OF TABLES

TABLE 1 HYPOTHESIS FOR MULTIREGIONAL AND OUT OF AFRICA SCENARIOS 21
TABLE 2 LANGUAGE CLASSIFICATION OF ETHNIC GROUPS 47
TABLE 3 SUMMARY OF GEOLOGICAL AND ARCHAEOLOGICAL PERIODS IN AFRICA. 51
TABLE 4 CLIMATIC SUMMARY OF THE LATE PLEISTOCENE AND HOLOCENE IN AFRICA 52
TABLE 5 ETHNIC GROUPS FROM WHICH BLOOD WAS SAMPLED. 61
TABLE 6 ETHNIC GROUPS FROM WHICH DNA WAS SEQUENCED 61
TABLE 7 PLACES WHERE SAMPLES WERE COLLECTED 62
TABLE 8 VARIABILITY WITHIN THE C-RUN REGION 77
TABLE 9 *HPA*I FREQUENCIES IN THE POPULATIONS IN THIS STUDY. 78
TABLE 10 NUMBER OF REPEATS IN THE POPULATIONS SURVEYED.. 80
TABLE 11 SAMPLED POPULATIONS 82
TABLE 12 MEAN PAIRWISE DISTANCES WITHIN AFRICAN ETHNIC GROUPS 83
TABLE 13 GENETIC DISTANCES AND RESULTS FROM THE HUDSON TEST 84
TABLE 14 COALESCENT TIMES WITH DIFFERENT POPULATION HISTORIES.. 87
TABLE 15 LINGUISTIC PAIRWISE DISTRIBUTION. 88
TABLE 16 GEOGRAPHIC PAIRWISE DISTRIBUTION. 89
TABLE 17 LIFESTYLE PAIRWISE DISTRIBUTIONS. 90
TABLE 18 HUNTER-GATHERER PAIRWISE DISTRIBUTIONS 91
TABLE 19 CONTINENTAL COMPARISON 93
TABLE 20 MEAN PAIRWISE DISTANCES BETWEEN AND WITHIN CONTINENTS 94
TABLE 21 AFRICA VERSUS NON AFRICA 94
TABLE 22 KEY TO MAP 6 100
TABLE 23 DIVERSITY OF THE TURKANA COMPARED TO AFRICA AND THE WORLD.. 101
TABLE 24 MEAN DISTANCE FROM OUTGROUPS 101
TABLE 25 POPULATIONS SAMPLED 107
TABLE 26 NUMBER OF INDIVIDUALS IN EACH CLUSTER 114
TABLE 27 DIVERSITY OF CLUSTERS 115
TABLE 28 EXPANSION TIMES OF CLUSTERS IN DIFFERENT GEOGRAPHICAL REGIONS 117

LIST OF MAPS

MAP 1 AFRICAN ETHNIC GROUPS FROM WHICH HVR-I SEQUENCES HAVE BEEN PUBLISHED 24
MAP 2 AFRICAN ETHNIC GROUPS FROM WHICH HVR-I SEQUENCES ARE STUDIED IN THIS THESIS..............25
MAP 3 POSSIBLE ORIGINAL LOCATIONS OF MAJOR LANGUAGE GROUPS...48
MAP 4 FRAGMENTATION BELT 51
MAP 5 PLACES WHERE SAMPLES WERE COLLECTED 64
MAP 6 PLACE OF BIRTH OF MATERNAL GRANDMOTHER'S OF THE 37 TURKANA INDIVIDUALS 100

ACKNOWLEDGEMENTS

I dedicate this thesis to the volunteers who took part in the survey. Their lives are often difficult but they approach life with pride and enthusiasm. We have more to learn from them than what their genes tell.

I also dedicate this to Anthonia Ekpa from Calabar and Musa Kadzai from Yola who helped teach me how to understand and love Africa.

I am fortunate to have worked with Prof. Rashid Aman, of the National Museums of Kenya, who supported my project and assisted my collection in Kenya. I thank James Maikweki and Meave Leakey also from the Museums, for enjoyable Safaris.

I thank the following clinics and hospitals for all their assistance:
Turkana District Hospital, Lodwar, Kenya
Maralal District Hospital, Maralal, Kenya
Isiolo District Hospital, Isiolo, Kenya
Provincial General Hospital Nakuru, Kenya
Provincial General Hospital, Nakuru, Kenya
Sinna Medical Centre, Nairobi, Kenya
South Nyanza District Hospital, Kisumu, Kenya
Hôpital National, Niamey, Niger
Akintunde Memorial Medical Centre, Saki, Nigeria
Baptist Medical Hospital, Saki, Nigeria
UDUTH, Sokoto, Nigeria
UMTH, Maiduguri, Nigeria

I thank the Hospital directors and staff including Dr. Abba Kyari Di, Prof. Idris Mohammed, Dr. Simon Obi, Prof. Chikwen, Dr.V. Brahasuriya, Dr. Jose P. Ambe, Dr. Simon Nwachuwa, Lawrence Ewesit, George Ochieng Otieno, Bashir Farah, Adirahman Abdullah Gacheru Muruthi, Obadish Nfunga, Michael Loter Kiyonga, Chaibou Tawaye, Dr. M. Akintunde, Dr. Amadou Sako, Dr. Hamma Soumana. I am especially grateful to Usman Aliu Wakili the Seriki Fulani of Saki Idedapo Local Government Area and Sister Connie Hall from Mekaneyessus for their help in collecting Fulbe and Nuer samples respectively.

Special thanks to field assistants who helped me with translation and struggling with cattle; the list is too long to mention here, but especially Mohammed Hada, Mohammed Ahnou Abbou, Efe Efiawore, David Laur, Joseph Owino Jung'a, and Jean-Romain da Silveira.

I am very grateful to ILCA (International Livestock Centre for Africa) for providing transport to some areas and providing a base to use libraries and faxes. I received a lot of assistance from many people, but I would especially like to thank Dr. E. Rege, Dr. Waudyallew Mulatu, Eshetu Zerihum, from Addis Ababa, Dr. Jimmy Smith, Jonathon Ekuwuruke, Alhaji Nata'ala Sambo, Dr. Doug Little, Dr. Kwaku Agyemang from Kaduna and Dr. Mark Powell and Dr. Peter Lawrence from Niamey.

Other agricultural research agencies provided assistance. Special thanks to Dr. Bruno Gerard of ICRISAT and Prof. E.O. Oyedipe and Dr. Voh of NAPRI, Dr. Christie Peacock of Farm Africa and Dr. Gérard Duvallet of CRTA in Bamako.

I am most grateful to the Betubizas, Bakers, Lawrences, Christine Schäfer, Brent Swallow and Urs Herren, who invited me into their homes and I am also grateful to all the people who offered me their generosity and friendship on the road.

Robert Hickson gave me a lot of support both practical and psychological in the earlier phases of the project for which I will always be grateful.

Thanks to Karin Bauer for the roller coaster rides we had while working together and also for sequencing the Kikuyu, Somali and Yoruba individuals. Thanks to Antti Sajantila and Tiiu Anttinen for their friendship and weekend adventures. Traudl Feldmaier-Fuchs for taking care of me. Walter Schartau and Frau Krella for invaluable assistance. Thanks to Barbara Petri for her friendship and compassion. Hendrik Poinar for demonstrating that science can have soul. Thanks to Konny Rassmann for many things but especially for her wonderful enthusiasm. Thanks to Arndt von Haeseler for writing programs for the SUN. I also am very grateful to all the people I had the pleasure of working with in the laboratory; there are too many to name. Søren Germer from Berkeley provided very useful references on the ethics of sampling and human research. I am very grateful to Hans-Jürgen Bandelt and Peter Forster from Hamburg and Martin Richards from Oxford for their enthusiasm and stimulating discussions and introduction to the world of Median Networks.

This thesis would not have been possible without David Penny for always being supportive of my ideas and me, and for being a wonderful supervisor. I thank Michael Hendy for his support and useful comments on my papers. I am grateful to Svante Pääbo for inviting me to Germany and providing the laboratory facilities for this study.

Sue Watson, Martin Richards, Robert Hickson and David Penny provided editorial advice for this thesis, smoothing out many rough edges.

I was very fortunate to receive the Vice Chancellor's Scholarship from Massey University. While in Germany I was funded by the Deutsches Akademisches Austausch Dienst (DAAD) and also the Leibniz Stiftung filled in some of the hiatuses while in Germany. The Fieldwork and travel was funded by my own finances. This research received approval from the Human Ethics Committee of Massey University.

I appreciate my family who have never provided barriers for my travels and science. Most of all thanks to Tom Södergren for being field assistant and friend.

CONTRIBUTION

The thesis is my own work. For some of the chapters, however, I have collaborated with other people. The following list summarises my contribution to the following chapters.

SAMPLE COLLECTION AND ETHICS

target:	**no journal targeted**
purpose:	**to introduce the fieldwork and design of project**
contribution:	**my own work**

SEQUENCE DESCRIPTION

target:	**no journal targeted**
purpose:	**to give a general overview of the sequences**
contribution:	**my own work**

THE TURKANA PEOPLE FROM KENYA HAVE THE WORLD'S DIVERSITY IN THEIR MTDNA GENEPOOL

target:	**Nature**
purpose:	**to emphasis diversity within Africa**
contribution:	**my own work with supervisoral advice from David Penny**
authors	**Watson, E.E. and Penny, D.**

HUMAN MITOCHONDRIAL FOOTPRINTS OF EXPANSIONS WITHIN AFRICA

target:	**to be decided**
purpose:	**to trace mitochondrial relationships within Africa and date expansion events**
contribution:	**written by myself with assistance from Hans-Jürgen Bandelt, Peter Forster and Martin Richards**
authors	**Watson, E.E., Forster, P., Richards, M. and Bandelt, H.-J.**

TREES FROM LANGUAGES AND GENES ARE VERY SIMILAR

target:	Systematic Biology 42(3):321-323
purpose:	**to compare trees generated globally from genetics and language.**
contribution:	**my original idea and tree construction**
authors	**Penny, D., Watson, E., and Steel, M.A.**

2. INTRODUCTION

The Out of Africa hypothesis was proposed on the basis mitochondrial DNA (mtDNA) data from human populations (Brown *et al.* 1981; Cann *et al.* 1987; Vigilant *et al.* 1991). Controversy ensued due to the methods used for analysis (Templeton, 1992; 1993; Maddison *et al.* 1992) but reanalysis with new methods concluded that the data did indeed support an out of Africa scenario (Penny *et al.* 1995). Some palaeontologists arrived independently at an African origin scenario for modern humans (Stringer, 1988; Stringer and Bräuer, 1994; Bräuer, 1989; 1992). Other palaeontologists had an opposing hypothesis termed the Multiregional hypothesis which proposed regional continuity of humans over a period of a million years (Wolpoff, 1989; Frayer and Wolpoff, 1993; Frayer, 1993; 1994).

Since Vigilant *et al.*'s (1991) study, much of Africa has remained unsurveyed in terms of mtDNA. This doctoral study has involved collection of a diverse set of African samples, sequencing and analysing them with established methods as well as with new methods such as median networks. With these new methods and the expanded dataset (including a large compilation of world-wide sequences published after Vigilant *et al.*'s (1991) study) the question of the origin of modern humans is readdressed. The Multiregional and out of Africa hypotheses are discussed in an attempt to bring together palaeontological and genetic data. In addition to questions of human evolution, the mtDNA data was used to determine the genetic basis of ethnicity within Africa.

2.1 The Out of Africa hypothesis versus the Multiregional hypothesis

Since the experiments of Wilson and Sarich (1969), molecular geneticists and palaeontologists have often disagreed. Wilson and Sarich (1969) proposed dates of human, chimpanzee and gorilla divergence from four to eight million years ago, much earlier than the excepted palaeontological dates. These dates have eventually reached general acceptance. The next significant controversy was due to the 'Out of Africa' Hypothesis proposed by members of the Wilson group (Brown *et al.* 1981; Cann *et al.* 1987; Vigilant *et al.* 1991). 'Out of Africa', proposes that modern humans share at least one common female ancestor who lived within the last 200,000 years; and that *Homo sapiens sapiens* then spread out from Africa around 100,000 years ago replacing earlier *Homo erectus* populations. After extensive debate (Templeton, 1992; 1993; Maddison *et al.* 1992; Hedges *et al.* 1992; Gibbons, 1992; Penny *et al.* 1995) the general agreement (although not unanimous) amongst geneticists is a model based on an out of Africa scenario. There are variations such as the 'Weak Garden of Eden' Hypothesis (Harpending *et al.* 1993) which proposes that the "racial"[sic] populations separated about 100,000 years ago and later expanded demographically. The Out of Africa hypothesis is also accepted by some palaeontologists (Stringer, 1988; Stringer and Bräuer, 1994; Bräuer, 1989; 1992). For the mtDNA to support the Out of Africa hypothesis the most recent common ancestor (MRCA) should predate the expansion of modern humans out of Africa and the highest diversity should be found in Africa.

Based on the fossil record, primarily in Asia, other palaeontologists (Wolpoff, 1989; Frayer and Wolpoff, 1993; Frayer *et al.* 1993; 1994) have proposed a model named the 'Multiregional Hypothesis'. The first proponent of this model was Weidenreich (1943; cited in Frayer *et al.* 1993) who proposed that human populations were an interconnected network, that retained regional continuity in some areas. This theory assumes significant geneflow between the world populations. The Multiregional hypothesis, proposes that *Homo sapiens* arose in independent regions derived from *Homo erectus* populations. This implicitly places the common ancestor of modern humans over one million years ago, when *Homo erectus* expanded out of Africa. For the mtDNA to support the Multiregional hypothesis the most recent common ancestor should predate the expansion of *Homo erectus* out of Africa and to support the high amount of geneflow there should be little evidence of regionalism.

Table 1 Hypothesis for Multiregional and Out of Africa scenarios

Hypotheses	Out of Africa	Multiregional
Hypothesis 1 **The MRCA of humans lived ......**	*c.* 200,000BP	*c.* 1,600,000BP
Hypothesis 2 ***Homo sapiens sapiens* evolved in ......**	a single region	several regions
Hypothesis 3 **the region was ......**	Africa	Africa and Eurasia
Hypothesis 4 ***Homo sapiens sapiens* ...... previous *Homo* groups**	replaced	interbred with

MRCA, is most recent common ancestor. (Later in the thesis modified versions of both these hypotheses will be discussed.)

These opposing views have been important for the scientists on both sides to consolidate their theories. No general consensus has been reached although some proponents of both theories propose that there may have been some degree of admixture between the modern and archaic humans (Stringer and Bräuer, 1994).

Using most calibrations the mtDNA data does not show divergences as far back as one million years, rather most divergence is seen in Africa dated to about 100,000-200,000 years ago. The high diversity in Africa (Cann *et al.* 1987; Vigilant *et al.* 1989; Bowcock *et al.* 1991a; Horai *et al.* 1993; Jorde *et al.* 1994; Rogers and Jorde, 1994; Horai *et al.* 1995), has been argued to support the Out of Africa hypothesis (Stoneking 1993). However, Aoki and Shida (1993) have argued that there are alternative explanations, such as population size and differential substitution rates. A recent paper by Jorde *et al.* (1995) demonstrates nuclear markers do not show such relatively deep branches in Africa as mtDNA. They propose the deep branches in the mtDNA are due to the mtDNA having an effective population one quarter of the size of nuclear DNA, and therefore being more susceptible to bottlenecks. It is conceivable that in Eurasia during the Pleistocene glacial advances the populations underwent severe bottlenecks. Africa being more tropical may have had less severe bottlenecks and therefore could have retained most of its diversity. The fossil record on which the Multiregional hypothesis is based, does indicate some conflict with the genetic data which needs to be addressed in the future. The conflict lies in the fossil evidence for the presence of early humans outside Africa at dates generally older than the mitochondrial Eve date and continuity of traits which are persistent in the Asia, European and Australia sequences (see Section 2.1).

Another explanation is that the mtDNA results are not reflecting human maternal population history but are complicated by selection on the mtDNA genome due to a favourable mutation. This phenomenon is known as selective sweeps and could account for the lower diversity observed in the mtDNA (see Section 9) .

2.2 African Ethnic Groups

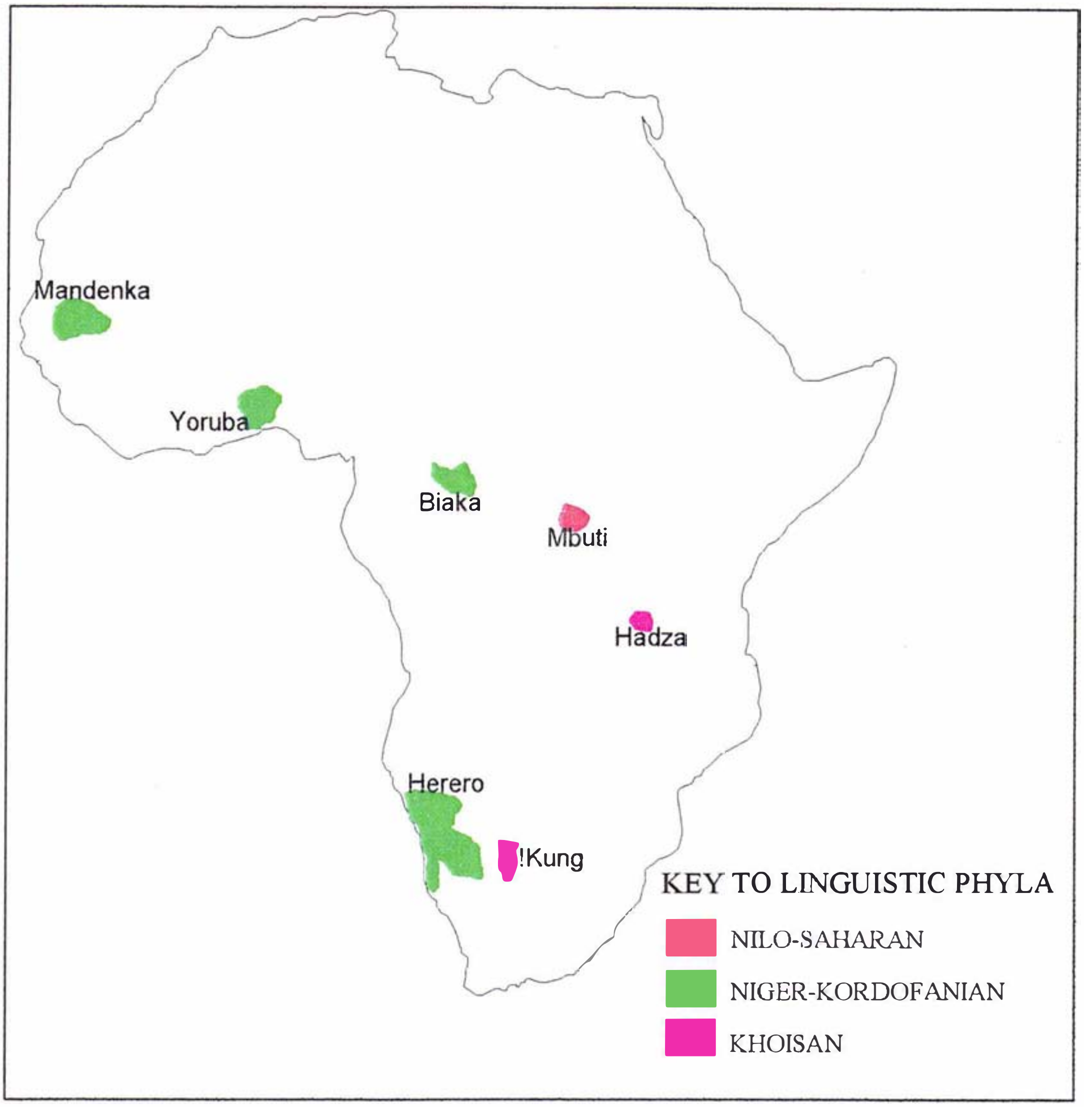
To address question concerning human origins and the dynamics of populations within Africa it was necessary to sample extensively. The groups were chosen from diverse linguistic phyla, economic modes and geographic areas. Generally demographically large populations were chosen to avoid problems of being limited to a small geographical area, with the possible complications of shared maternal ancestry.

In this study nine populations were chosen (see Figure 10). When combined with data from previous studies, it brings the number of populations to fifteen (see Table 11), although the Hadza and Herero are not included in the analysis due to ambiguous sequences.

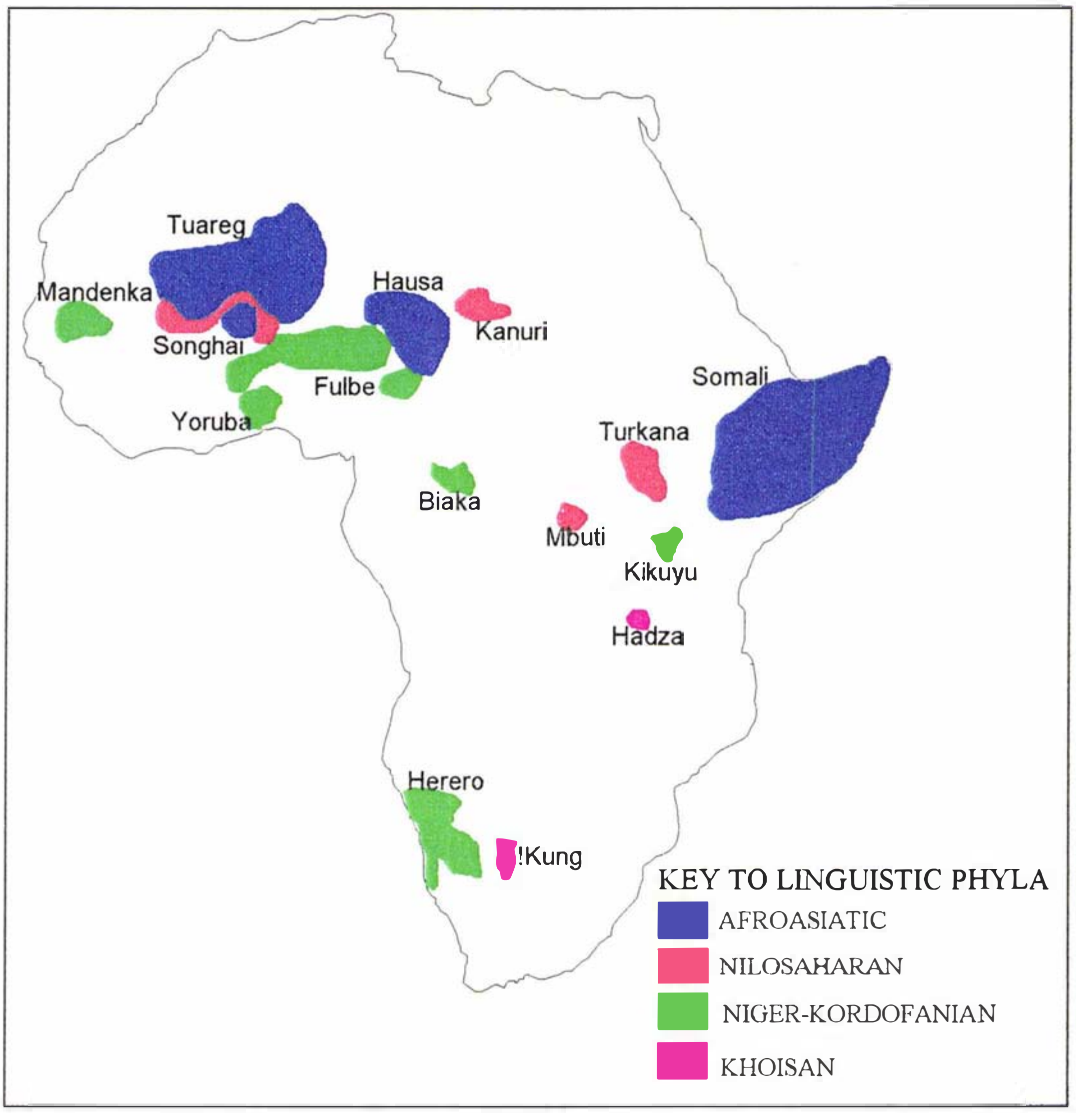
The following section (pp. 26-37) introduces the ethnic groups, their location and some details about their history and economy and is intended as a useful reference in relation to discussions in later sections of the thesis.

There is a strong tendency for people to assume a single origin and ancestry of an ethnic entity. The notion of ethnic identity reflecting single origins or a single common ancestor (Hiernaux, 1975) has been challenged as being bound up with the Hamite-Cushitic dichotomy assigned by European historians and anthropologists. The Hamites (a name associated with fairer skinned pastoralists) were associated with the origins of anything 'civilised' from iron working to the construction of Zimbabwe (Sobania, 1979). Many of the following origin theories (pp. 26-37) appear to fall into this category. Some African groups, however, have in their oral tradition a Middle Eastern origin. Since some are now predominantly Moslem, they may seek such an origin.

The terminology requires clarification. Descent is inheritance of name and identity from one's ancestors. "Patrilineal" is recognition of descent through the male line; "matrilineal is descent through the female line and "bilineal" is descent through both the male and female lines. "Residence" refers to the movement of one spouse after marriage; when the wife moves to her husband's in-laws it is termed "patrilocal", and when the husband moves to his wife's in-laws it is termed "matrilocal". When the couple move to neither of the in-laws it is termed "virilocal". Most of the African groups in this survey are primarily patrilineal and patrilocal.



Map 1 African Ethnic Groups from which HVR-I Sequences have been Published
HVR-I sequences have been published from the above ethnic groups. The sequences from the Herero and Hadza are not very informative due to ambiguous positions. Ethnic boundaries have been adapted from Murdock (1959). The ethnic groups are colour coded by their Linguistic phyla.



Map 2 African Ethnic Groups from which HVR-I sequences are studied in this thesis.

The ethnic groups from which the HVR-I sequences are studied in this thesis, including the published populations. Ethnic boundaries have been adapted from Murdock (1959). The ethnic groups are colour coded by their Linguistic phyla.

2.2.1 Fulbe

Other names	Fulani, Peul, Fula, Felaata, Pullo, Felani, Filani, Ful
Geographical Area	The Savannah belt from Senegal to Cameroun, and have been seen as far east as Ethiopia (H. Ghalib, *pers. comm.)*
Population	6 million (Stenning, 1965:p.363), 5 million (de Tressan, 1953 cited in Murdock, 1959:p.413)
Language and Affiliation	Fulfulde is a West Atlantic language of the Niger-Kordofanian linguistic phylum, affiliated closely with the Tukulor, Serer, and Wolof of Senegal.
Lifestyle	Primarily pastoral but there are semi-sedentary groups in Northern Nigeria, the delta region of Mali, and Senegal region.
Descent/ Residence	Patrilineal/ Patrilocal (Murdock, 1959:p416)
Proposed Origin	Jewish or Syrian origins (Guiraudon, 1888; Delafosse, 1912; Morel, 1902; cited in Stenning, 1959:p.18). Ethiopian affinities Barth (1857-8; cited in Stenning, 1959;p.19) North African Berbers (Passarge, 1895; Meyer, 1897; Crozals, 1883; cited in Stenning, 1959:p.19) Hindu Origins (Golberry, 1805; cited in Stenning, 1959:p.19) Malayo-Polynesian (Eichtal, 1841; cited in Stenning, 1959:p.19) The Muslim Fulbe believe in a Middle Eastern origin associated with the Islamic genesis (Stenning, 1959:p.19).
Recent History	Historical records favour a Senegalese origin and expansion beginning in the eleventh century (Delafosse, 1912; cited in Stenning, 1959). The expansions were linked to Islam, culminating in the taking of the seven Hausa states in Northern Nigeria early in the nineteenth century.

2.2.2 Hausa

Other names	Hausawa, Haussa, Haoussa.
Geographical Area	The Savannah region of Northern Nigeria and Southern Niger as well as in Ghana, Togo, Benin and Cameroun.
Population	9 million (Grimes, 1988:p.278); 12 million Hausa (Ruhlen, 1987:p.86)
Language and Affiliation	A Chadic Language of the Afroasiatic linguistic phylum.
Lifestyle	Agriculturists but some do have cattle.
Descent/ Residence	Patrilineal (Stride and Ifeka, 1969)/ Patrilocal
Proposed Origin	One theory is the Hausa were hunter-fisher people living by Lake Chad. When the lake dried they shifted West to present locations (Adamu, 1981). A visitor, Prince Bayagida from Baghdad, killed an evil snake and was rewarded by marrying Queen Daura. His legitimate descendants founded the seven Hausa States and his seven illegitimate sons founded the seven "non-pure" Hausa states (Stride and Ifeka, 1969).
Recent History	The Hausa states were founded during the sixteenth century.

2.2.3 Kanuri

Other names	Kanouri, Beriberi, Bornu, Bornouans, Kanoury, Kole, Sirata
Geographical Area	The shores of Lake Chad, in the countries of Niger, Nigeria, Central African Republic, and Cameroun
Population	1 million (Ruhlen, 1987:p.107), 3.5 million (Grimes, 1988:p.282).
Language and Affiliation	A language of the Saharan group of the Nilo-Saharan linguistic phylum. Related closely to Kanembu
Lifestyle	Agricultural and Pastoral
Descent/ Residence	Bilineal (Cohen, 1967)
Proposed Origin	Yemen (Cohen, 1967:p.12). Intermarriage between the Zaghawa nomads and the sedentary So (Stride and Ifeka, 1969) Central Sahara (Cohen, 1967:p.14).
Recent History	Islam arrived during the eleventh century (Cohen, 1967:p.14). The Kanuri founded the Kanem empire 9th to 14th centuries and the Bornu kingdom from the 15th to 17th centuries. These empires were important in the trans-Saharan trade.

2.2.4 Kikuyu

Other names	Gikuyu, Gigikuyu, Gekoyo, Akikuyu, Wakikuyu 13 clans
Geographical Area	Central highlands of Kenya
Population	4,356,000 (1987) (Grimes, 1988:p.244)
Language and Affiliation	Bantu language of the Niger-Kordofanian linguistic phylum. Affiliated closely with the Kamba, Embu, and Chuka (Grimes, 1988:p.244)
Lifestyle	Agriculturists with cattle, goats and sheep (Olivier and Mathew, 1963)
Descent	Patrilineal (Routledge and Routledge, 1910)/ Patrilocal (Murdock, 1959:p. 345)
Proposed Origin	Bantu expansion from West
Recent History	Moved and expanded from Mount Kilamanjaro (Olivier and Mathew, 1963), arrived in current location 1500's. Mixed with the Maasai.

2.2.5 !Kung

Other names	Zhu/twasi (Vigilant 1990)
Geographical Area	Kalahari area of Botswana and Namibia
Population	6,000 (Vigilant 1990), 40,000 (total San) (Lee, 1979)
Language and Affiliation	Northern San of the Khoisan linguistic phylum, related to Nama and Naron and more distantly to the Tanzania Khoisan speakers
Lifestyle	Hunter-Gatherers
Descent/ Residence	Bilineal, patrilocal (Murdock, 1959; p.56)
Proposed Origin	Khoisan speakers were previously covering large areas of Eastern and Southern Africa
Recent History	15th century Bantu expansions and the arrival of the Europeans have reduced the range of the Khoisan speakers.

2.2.6 Mandenka

Other names Manding, Mandingo, Mandingue, Mandique, Mande, Sose, Wangara

Geographical Area Senegal and Gambia

Population 766,600 (Grimes, 1988:p.299) 1.4 million (Cavalli-Sforza *et al.* 1994:p.182), 1 million (Murdock, 1959: p.72)

Language and Affiliation Mande, which gives its name to the Mande family of the Niger-Kordofanian linguistic phylum. Closely related to Malinke, although languages are distinct.

Lifestyle Agriculturists (Schaffer and Cooper, 1980:p.5)

Descent/ Residence Patrilineal (Murdock, 1959: p.75)/ Patrilocal (Schaffer and Cooper, 1980:p.45)

Proposed Origin Little is known but probably from the upper reaches of the Senegal and Niger Rivers (Ki-Zerbo, 1993).

Recent History Mandenka from Kangaba expanded to form the Empire of Mali which was dominant from the 13th to 16th centuries. Part of the success was due to the role of trade, especially from gold (Davidson, 1965). Mali went into decline, due to attacks on Djenne by the Songhai (Murdock, 1959:p 73).

2.2.7 Pygmy (Biaka and Mbuti)

Other names	Four groups, eastern (Mbuti, Efe), western (Biaka, Aka, Babinga, Bagombi, Baka), central and southern. The Eastern(Mbuti) and Western (Biaka) are in this study.
Geographical Area	South Western Central African Republic, Southern Cameroun, North Eastern Zaire (Ituri forest) (Vigilant *et al.* 1989).
Population	30,000 (Biaka) and 30,000 (Mbuti) (Vigilant, 1989)
Language and Affiliation	Original language has been replaced and they now speak Baka-Gundi of Adamawa, Niger-Kordofanian (Biaka (Western Pygmy)) and Nilo-Saharan (Mbuti (Eastern Pygmy)) (Cavalli-Sforza *et al.* 1994:p.179).
Lifestyle	Hunter-gatherers
Descent/ Residence	Patrilineal (Pederson and Woehle, 1985)
Proposed Origin	Biaka in the South West Sudan before the Arab invasions, believed to have travelled first to NW Zaire.
Recent History	Contact with Bantu farmers and adoption of languages; most of the geneflow is through marriage between the Agricultural male and pygmy female (Cavalli-Sforza *et al.* 1994:p.178).

2.2.8 Somali

Other names	6 clans the Dir, Isaaq, Daarod, Hawiye, Digil and Rahanwin
Geographical Area	Somalia, Northern Kenya and Eastern Ethiopia
Population	3.5 million (Lewis, 1965:p.321) 2 million (Ruhlen, 1987:p.86)
Language and Affiliation	A Cushitic language of the Afroasiatic linguistic phylum. They are related to the Dankil, Galla and Beja peoples
Lifestyle	Nomadic pastoralists with camels, cattle, sheep and goats (Elmi, 1991)
Descent/ Residence	Patrilineal (Lewis, 1967:p.1)/ Patrilocal (Murdock, 1959:p.322).
Proposed Origin	They claim Arabian ancestry (Lewis, 1967:p.11). The Somali were part of a Cushitic expansion from South Eastern Ethiopia which absorbed Bantu groups (Murdock, 1959:p.319).
Recent History	In the last millennium they have been in constant conflict with Christian Ethiopia (Lewis, 1967:p.17). During the sixteenth century the Somali expanded south into Southern Somalia and Northern Kenya (Lewis, 1967:p.23).

2.2.9 Songhai

Other names	Songai, Songoi, Songhay, Sonai, Sonrai, Sonrhai, Koroboro
Geographical Area	A wide belt along the Niger River in Mali and Niger
Population	600,000 (Davidson, 1965); 330,000 (Murdock, 1959:p.138); 528,000 (Grimes, 1988:p.258)
Language and Affiliation	Nilo-Saharan, Songhai. Closely related to Dendi and Djerma.
Lifestyle	Agricultural, fishing, two clans are pastoral (Davidson, 1965).
Descent/ Residence	Patrilineal
Proposed Origin	Interbreeding between small communities on the Niger River and desert nomads (probably Zaghawa) (Stride and Ifeka, 1969).
Recent History	The Songhai formed the great Songhai empire which ruled most of West Africa from the 7th to 14th centuries and controlled Saharan trade. Its two major cities were Gao and Timbuktu. In 1465 Timbuktu was occupied by the Tuareg and this marked the beginning of the demise of the kingdom.

2.2.10 Tuareg

Other names	Tamasheq, Tomacheck, Tamashekin, Tamajeq, Tourage. There are seven major clans (Ahaggeren, Antessar, Asben, Auliminden, and Azjer, Ifora, Udalan) (Murdock, 1959:p.114).
Geographical Area	Arid regions of Niger, Mali and Algeria.
Population	200,000 (Murdock, 1959:p. 114) 300,000 (Norris, 1984:p.315), 360,000 (Grimes, 1988:p.266)
Language and Affiliation	Tamashaq, a Berber dialect (Briggs, 1960:p.107; Norris, 1984:p.315) of the Afroasiatic linguistic phylum. Related closely to the Berber of Mauritania and Algeria
Lifestyle	Nomadic pastoralists with camels, sheep, goats and cattle (Briggs, 1960:p.140)
Descent/ Residence	Matrilineal descent (Murdock, 1959: p.117; Norris, 1975:p.6); the Auliminden clan are the exception
Proposed Origin	Descendants of the people who brought the camel to Africa (Norris, 1984:p.315) Forty visiting Tunisians killed an *ifrit* and as a reward were given forty wives. Later they abandoned the wives, the offspring of which were the founders of the Tuareg people (Norris, 1975:p.9)
Recent History	Moved from Libya to the centre of the Sahara to escape Arab attacks (Cavalli-Sforza *et al.* 1994:p.173; 1988). Established the Agadez sultanate in the fifteenth century (Norris, 1984:p.312).

2.2.11 Turkana

Other names	Bume, Buma, Elgume 19 clans
Geographical Area	The arid region in the north-west corner of Kenya, bounded by the Ugandan Sudanese and Ethiopian borders and Lake Turkana to the East.
Population	250,000 Turkana (McCabe, 1991; Grimes, 1988:p.248; Cavalli-Sforza *et al.* 1994:p. 182),
Language and Affiliation	A Nilotic language of the Nilo-Saharan linguistic Phylum. The Turkana are affiliated to the Iteso, Maasai, Kalenjin ethnic groups (Diets, 1987) and their language is mutually intelligible with the Toposa of Sudan, and Karamajong and Jie of Uganda and Nyangatom of Ethiopia and Sudan (Grimes, 1988:p.248).
Lifestyle	They are nomadic pastoralists (Gulliver, 1955:p.2), traditionally cattle, but due to dessertification there is a shift to camels.
Descent/ Residence	Patrilineal/ Patrilocal (Gulliver, 1955:p.124)
Proposed Origin	1000 BP Nilotic groups adopted pastoralism and expanded into Kenya from the West (Murdock, 1959:p.333).
Recent History	Late in the seventeenth century the Turkana separated from the Jie of Uganda and moved into the Rift Valley, displaced or absorbed the previous inhabitants the Rendille, Ngikor and Merille (Gulliver, 1955:p.5; Sobania, 1979).

2.2.12 Yoruba

Other names	Oyo
Geographical Area	South West Nigeria and Benin in the humid rainforest zone
Population	15 million (Eades, 1980:p.2; Cavalli-Sforza *et al.* 1994:p.182)
Language and Affiliation	Kwa language of the Niger-Kordofanian linguistic phylum.
Lifestyle	Agriculturists
Descent/ Residence	Patrilineal (Murdock, 1959:p.248)/ Virilocal (Eades, 1980:p.51)
Proposed Origin	Nubia, Arabia driven out by religious dispute (Talbot, 1926) Egypt (Talbot, 1926) Eastern Sudan (Lucas, 1948) Central Sudan (Davidson, 1965) Oral tradition from Ife-Ife (Davidson, 1965) Oral tradition from Mecca (Davidson, 1965) Possibly related to those who founded the Nok Culture (Davidson, 1965)
Recent History	Series of Kingdoms founded between the thirteenth and fifteenth centuries (Eades, 1980:p.17).

2.3 Mitochondrial DNA (mtDNA)

2.3.1 Characteristics of the Molecule

Mitochondria are DNA containing intracellular organelles associated with oxidative energy metabolism in eucaryotes. The DNA associated with this organelle, mitochondrial DNA (mtDNA), is a circular molecule of 16569bp in humans (Anderson *et al.* 1981). There are estimated to be between 1,000 to 10,000 molecules per cell (Bogenhagen *et al.* 1974; cited in Giles *et al.* 1980) and they almost universally follows a maternal inheritance (Giles *et al.* 1980). Paternal inheritance patterns have been suggested although this has not been confirmed in humans. Recently Kaneda *et al.* (1995) demonstrated that in mice paternal mtDNA transmission occurs with some interspecific crosses only. It can be presumed that paternal transmission is insignificant for the purpose of human population studies. Unless stated otherwise mtDNA in this thesis refers to humans, although most features are common to other mammals and vertebrates.

The mtDNA, due to its maternal inheritance pattern, has an effective population size which is one quarter of nuclear DNA (Kimura, 1983:p.40). The mtDNA molecular codes for 13 proteins associated with electron transport and oxidative phosphorylation, as well as 22 transfer RNAs (tRNAs) and 2 ribosomal RNAs (rRNAs). Most proteins required by mitochondria are encoded by the nuclear genome and imported from the cytoplasm. MtDNA is characterised by having few noncoding bases. The control region which has no coding function is the exception with a length of 1112bp. It contains the origin of transcription for both DNA strands, the origin of replication for the heavy strand, and the terminator associated sequence (TAS), as well as the binding sites for mtDNA transcription factors (Walberg and Clayton 1981; Chang and Clayton, 1984: Hixson and Clayton, 1985). Within this control region, there are two hypervariable regions (HVR-I and HVR-II from 15997-16401 and 00029-00408 respectively (using the numbering of Anderson *et al.* 1981)), which are generally thought to be selectively neutral. These regions have been used for intra-population studies particularly within human populations. The first region, being the most variable, was chosen to be sequenced in this study. The mtDNA has a substitution rate about five times faster than nuclear DNA (Brown *et al.* 1979; Wilson *et al.* 1985) and the Hypervariable region I is ten times faster than the average of the mtDNA genome. The reason for this is uncertain but it could be due to the mode of replication or template directed mutations which may account for the observed hotspots (regions of very high substitution rate). During replication the first part of HVR-II and all of

HVR-I is in a single stranded state. In the single stranded state the strand is subject to more deamination accounting for the excess of thymines on the light strand (Kornberg and Baker, 1992:p.682). Due to the high rate of substitution, the control region is useful for reconstructing the history of the molecule and implicitly the maternal history of the individual.

2.3.2 Mitochondrial DNA of Human Population Studies

Different types of mtDNA data have been used for human populations studied; this includes high and low resolution Restriction enzyme studies as well as DNA sequencing the control region (HVR-I and II) and also the entire genome. These methods have the advantage of detecting all variants in a population without prior knowledge. This has a major advantage over other methods which require the variants be known before the survey begins. Restriction Fragment Length Polymorphisms (RFLP) detect variation throughout the whole mtDNA molecule. Graven *et al.* (1995) found very high correlation between RFLP haplotype frequencies and associated control region sequences, which is what would be expected for a non-recombining genome.

Restriction Enzyme Studies

For a review of the methodology and applications of restriction enzymes on mtDNA see Avise (1994:pp. 60-68). Restriction enzymes recognise a sequence (usually an inverted repeat) and cut the DNA in or adjacent to the recognition sequence. For example *Hpa*I recognises the GTT/AAC sequence and cuts between the thymine and adenine, as indicated by the slash. The absence or presence of restriction sites produces various length variants which is a simple way to detect polymorphism either within or between populations or species. Studies of human populations have investigated restriction enzyme patterns using a six enzyme system: Johnson *et al.* (1983), Greenberg *et al.* (1983) and Cann *et al.* (1987) which have included Africans, and studies exclusive to Africa include Scozzari *et al.* (1988); Soodyall and Jenkins (1992); Scozzari *et al.* (1994); Graven *et al.* (1995). Recently a study by Chen *et al.* (1995) used high-resolution methodology which employs fourteen restriction enzymes to populations within Africa. High resolution methodology screens 15-20% of the mtDNA sequence, whereas the six enzyme method screens between 2-3% of the mtDNA. These studies have demonstrated a higher diversity of lineages within Africa than the rest of the world.

Control Region Sequencing Studies

For a review of the methodology of DNA sequencing see Avise (1994:pp. 82-87). Vigilant *et al.*'s initial study of the control region of the !Kung and the Pygmy (1989) was followed by the world-wide survey which included additional African groups the Yoruba, Hadza and Herero (1991). They reported both the high mitochondrial diversity within Africa and that the deepest branches on a phylogenetic tree are primarily African. These conclusions have been criticised on the basis of the tree building analysis which failed to find the most parsimonious trees (Templeton, 1992, 1993). However, closer detailed analysis still places the root of the phylogenetic tree within the African cluster and shows that the non-African sequences are a subset of the major African lineages (Penny *et al.* 1995). A recent study extended the African mtDNA database with the addition of control region sequences and restriction enzyme data from 119 Mandenka of West Africa (Graven *et al.* 1995). For distribution of these African populations see Map 1.

2.4 Nuclear DNA Studies

Most DNA is contained within the nucleus. It follows a diploid inheritance which is in one sense a more reliable picture of the genetic history of the individual, but analysis is complicated by recombination. The nuclear substitution rate is generally slow; even non-coding regions do not have the resolving power for intraspecies studies. Repetitive DNA such as microsatellites (repeated copies of a unit from two to five nucleotides in length) have fast gain or loss of repeat number, although the mechanisms of this are still poorly understood. The following studies are a summary of research in human nuclear DNA.

Nuclear DNA Studies in Human Population

Bowcock *et al.* (1994) summarised genetic variation of human groups by continents. Microsatellites like mtDNA, demonstrate Africa has significantly more genetic variation, whereas 110 blood and allozyme variants show higher diversity in Europe. This is probably an artefact of the technique where known classical markers are screened for polymorphisms that are already known, and the initial detection of polymorphisms was mostly carried out on European populations (Bowcock *et al.* 1994). Consequently, many possible polymorphisms in non-European populations will not be detected (Stoneking, 1994). Sequencing and microsatellites on the other hand find all variants regardless of whether they were known previously.

Blood groups and protein polymorphisms

Cavalli-Sforza *et al.* (1988) investigated 120 alleles in 42 world-wide populations. Africans/non-Africans were the first split in the phylogenetic tree based on genetic distances. Nei and Roychoudhury (1993) investigated 29 polymorphic loci in 26 world-wide populations and also observed Africans/non-Africans were the first split in a Neighbor Joining Tree. This African/non African split is also probably the result of an artefact of the forcing of genetic distances between populations into an evolutionary tree.

Results from β-globin haplotypes from 852 chromosomes from Asia and Pacific populations were compared to previously published data (from Africa, China, and Europe) showing some limited support for an African origin of modern humans (Chen *et al.* 1990) and apolipoprotein B haplotypes found Africans the most divergent group (Rapacz *et al.* 1991).

RFLP's (Restriction Fragment Length Polymorphisms)

Bowcock *et al.* (1991a) looked at one hundred DNA polymorphisms (RFLP's and polymorphisms detected with PCR) in five populations. Their conclusions were that Europe is an admixture of Asian and African elements.

Dinucleotide Repeat Polymorphisms

Di Rienzo *et al.* (1994) typed ten dinucleotide loci $(CA)_n$ in 46 Sardinians, 46 Egyptians and 25 Africans and found higher divergence between Africa and the other two populations. Bowcock *et al.* (1994) looked at 30 microsatellites mainly $(CA)_n$ in fourteen human populations and found the highest diversity within Africa. Deka *et al.*'s (1995) study investigates eight dinucleotide $(CA)_n$ repeats in eight human populations and chimpanzees. The predominantly Hausa group from northern Nigeria had the highest diversity and heterozygosity. Goldstein *et al.* (1995) investigated fifteen dinucleotide loci and found the deepest split was between Africans and non-Africans.

Y-specific polymorphisms

Y-specific polymorphisms are the male equivalent to the mtDNA. Inheritance is haploid and through the male line only. It is likely that the male side of the human story may be different to the female side. These differences, if detectable, could be informative about the differential movements of males and females. Torroni *et al.* (1990) observed polymorphism in six *Taq*I fragments, they found a relatively lower variability in Africa compared to Europeans. In a world-wide survey Dorit *et al.* (1995) found no variability in a 729 base pair intron of the *ZFY* locus, which they suggested could be due to a selective sweep, a recent origin for modern humans or a historically small effective male population size.

2.5 Studies to Classify African Groups

As with other continents there has been an interest in classifying the human populations of Africa. Earlier studies concentrated on physical traits, but with the advent of genetic studies, this question could be approached in a more objective way.

2.5.1 Phenetic

Phenetics is the study of observable physical characteristics of individuals. Phenetics has a long and at times contentious history (Gould, 1981). Hiernaux (1975) using anthropomorphic traits and a few genes divided Sub-Saharan Africa up into five categories: (1) Khoisan, (2) Pygmies, (3) Elongated Africans and Nilotes, (4) West Sudanese and Guinea Rainforest, and (5) Bantu. Winkler and Sokal (1987) investigated nineteen anthropometric and five pigmentation intensity variables for fifteen Kenyan tribes. They concluded "that linguistic and cultural characteristics of a population are strong but insufficient indicators of its biological relationships". Physical characteristics can be under environmental selection pressures, such as the trend for people living in rainforest to have lower stature. This does not imply shared recent history but shared ecological environment. Phenetic investigations seem antiquated when there are more sophisticated methods available, which can look at loci under no selective constraint.

2.5.2 Protein Immunology and Allozymes

Protein Immunology measures the antigenic properties of proteins. Allozyme studies measure enzyme polymorphisms by electrophoresing the enzymes on a starch or acrylamide gel. With staining techniques the enzymes can be visualised and the relative mobility determined.

Normally these studies were limited to local geographic regions. In Africa, however, Jacquard, (1974) investigated five blood groups from 27 populations from the Mediterranean to south of the Sahara. Excoffier *et al.* (1991) investigated GM and Rhesus RH haplotypes in sub-Saharan Africa. They found language family is the best predictor of genetic relationship, this may reflect historic migrations and expansions of ethnically different people. They claim, however, that most of the genetic variance remains unexplained by the models they tested. Their results indicate that sub-Saharan Africa does not consist of a homogenous genepool, and that migrations of peoples bound by common ethnic and linguistic affiliations have been important in fashioning the genetic landscape. Cavalli-Sforza *et al.*'s (1994) book attempted to bring results from many loci together. Data are lacking from some of the populations, which limits the ability to resolve some relationships. In the African context they concluded: this synthesis identifies the Bantu cluster by its general homogeneity and distinguishes the East and West Bantu movements; that there may have been a Nilo-Saharan influence in the Bantu genepool; that West Africa may have been the first part of continent to experience a population explosion from farming; that there were at least two independent expansions in West Africa, one in Senegal and the other in the Niger-Mali-Burkina Faso region; that the Ethiopians, Beja and perhaps Tuareg are an admixture of African and European/Middle Eastern genes; and that the Khoisan are intermediate between Africa and West Asia.

2.5.3 Language

Language classification, especially in Africa, has proved to be controversial. Greenberg's approach of comparing cognates over a selection of populations initially concerned many linguists, yet at the same time his findings have reached general acceptance (Ruhlen, 1987:p.120). For this study Ruhlen's classification (1987), which is largely based on

Greenberg's (1963) work, will be followed. This classifies Africa into four linguistic phyla, Afroasiatic, Nilo-Saharan, Niger-Kordofanian and Khoisan. These language phyla are discussed further in Section 2.7.3.

2.6 Ethnicity

Ethnic groups are defined as a biologically self perpetuating group, which shares cultural values, makes up a field of communication and interaction, has a membership which identifies itself, and is identified by others (Narroll, 1964; cited in Barth, 1969). Ethnicity is based on self identification, ancestry and language (Fisherman, 1977). In the time-scale of one hundred thousand years, it cannot be assumed that genetic affiliation would be identical to ethnic affiliation. In the past, population processes may be different to what is observed today.

The notion that each ethnic group has maintained its culture through ignorance of its neighbours is naive (Barth, 1969). Herren (1991) argues that the simplicity of ethnic or 'tribal' maps is a result of the colonial preconceptions, that being able to cross 'ethnic' bridges has been an important asset for the survival of people throughout African history. Herren (1991) presents the ethnic title as a fluid concept. When people are faced with conflict due to competition over resources, the smaller or weaker group has two options: to change niche or to assimilate. Ethnicity implies access to resources as well as a specific cultural and social way of exploiting the resources (Herren, 1991). Herren (1991) gives the example of 'the more powerful groups were.....interested in attracting their victims to join them, in order to boost both labour force and military power', and attributes the success of the Turkana to the strategy of assimilation. Herren (1991), provides examples of ethnic shifting in Kenya in the Samburu-Turkana interface at Isiolo (Hjort, 1981; cited in Herren, 1991) and in the Maasai-Samburu interface at Mukogodo. The Ariaal are a subsociety between the Rendille and Samburu (Spencer, 1973) and act as a stepping stone between the two cultures. "Even reference to common ancestors does not imply biological affinity, because the contribution of common ancestors to the gene pool, if not mythical, may vary largely between populations" (Hiernaux, 1975). Herren (1987) suggests the ethnic groups (of Kenya) are not older than the nineteenth century and are composites made up of predating smaller segments.

When discussing ethnicity authors usually refer back a few hundred years, but when questioning whether ethnic groups are genetic or cultural entities, it is important to consider both present time scale and long term dynamics. In the present, movements across ethnic boundaries are discernible, illustrated by the individuals involved in this study (see Section 10.8), 5% are children from marriages across ethnic boundaries (see p. 216). If this rate is representative of earlier times, after fifty generations only 8.11% would be expected to have no mixing in the previous fifty generations. It cannot be assumed, however, that the present is representative of the past. Prior to agricultural and pastoral production, population numbers and density would have been relatively low in Africa and populations may have had different dynamics to what is observed today. In recent times there has been a dramatic change in population distribution. From 1960 to 1980 the number of sub-Saharan urban centres over the size of 500,000 increased from 3 to 28, mainly as a result of massive rural to urban drift (Quinn, 1994). Urbanisation often breaks down the ethnic boundaries, complicating genetic studies such as this one, however, most of these samples were collected from rural centres. If the dynamics of ethnic groups are different now and were more stable in the past, even a very low rate of movement across ethnic boundaries would be quite significant when viewed over thousands of years.

2.7 Language

This section is not attempting to give a complete review, but a general background to help interpret the sequence data. The study of comparative linguistics has used syntax, morphology, phonology, and semantics to postulate relationships between languages. This process was extended to propose migration events and to reconstruct dates based on the now unfashionable field of glottochronology. Glottochronology was the "science" of estimating dates by the change of language, somewhat like the molecular clock but fraught with problems such as word borrowing and lack of a constant rate of change.

2.7.1 Correlation with Genetics

Cavalli-Sforza *et al.* (1988) and Olderogge (1981) represent two extremes on the continuum of correlation between genetic and linguistic groupings. Cavalli-Sforza *et al.* (1988) in a world-wide survey demonstrated "a considerable parallelism between genetic and linguistic history". This paper has attracted criticism for lack of data and inappropriate methods (Diamond, 1988; O'Grady *et al.* 1989). In Africa, Olderogge (1981) argues, there is no correlation between "racial" types and linguistics "as the early ethnic groups multiplied, migrated and crossbred over a long period". Obenga (in the Cairo Symposium, 1974: cited in Ki-Zerbo, 1981) stated that linguistics was the most obvious means of establishing whether peoples were culturally related. Barbujani, (1991) states that linguistic distances are better predictors of genetic distances than spatial distances in Africa. Excoffier *et al.* (1991) agrees but states that the greatest part of genetic variance remains unexplained. It is necessary to differentiate between strong correlation and weak correlation between language and genetics. Strong correlation does not hold as there are numerous examples world-wide of language replacement. The Pygmy groups are an extreme example, speaking three different languages from two linguistic phyla, all of which have been borrowed from neighbouring group. Weak correlation between language and genetics does hold at least at the world-wide scale as demonstrated in Section 7, p. 121.

2.7.2 Language Replacement

There are examples of language replacement in the different regions of the world. Renfrew (1987) proposes three models for language replacement, 1. **Demographic/subsistence** where there is a mass movement of people into the region usually with a new exploitative technology; 2. **Élite dominance**, whereby a small group of highly organised people dominate the local people and their language becomes dominant; 3. **System collapse** where an organised state collapses and the peoples of the periphery move towards the centre, some establishing their own organisation and language on the territory as a whole. Alternatives to these migration theories include "drift, the development of *linguae francae* used by traders, the emergence of new languages accepted over an area as 'standard intercommunity speech' and reflecting the preferences of the elite" (Hall, 1980; cited in Barker, 1988). Vansina (1980, cited in Barker, 1988) states that language did not necessarily spread with any of the more 'advanced' arts such as agriculture and metallurgy, let alone with new pottery styles. This will be further discussed in the Bantu section (p. 49).

Renfrew (1993) accounts for the present distribution of the world's language families, according to four main events. Firstly initial colonisation, prior to 15,000 BP (in Africa this includes the Khoisan and Nilo-Saharan phyla), and secondly, agricultural dispersal after 10,000 BP (in Africa the Niger-Kordofanian (especially the Bantu) and Afroasiatic) and the two other events: Élite dominance which has affected South Africa, Egypt and other areas to various degrees. The last event Northern climate adjustments did not affect Africa.

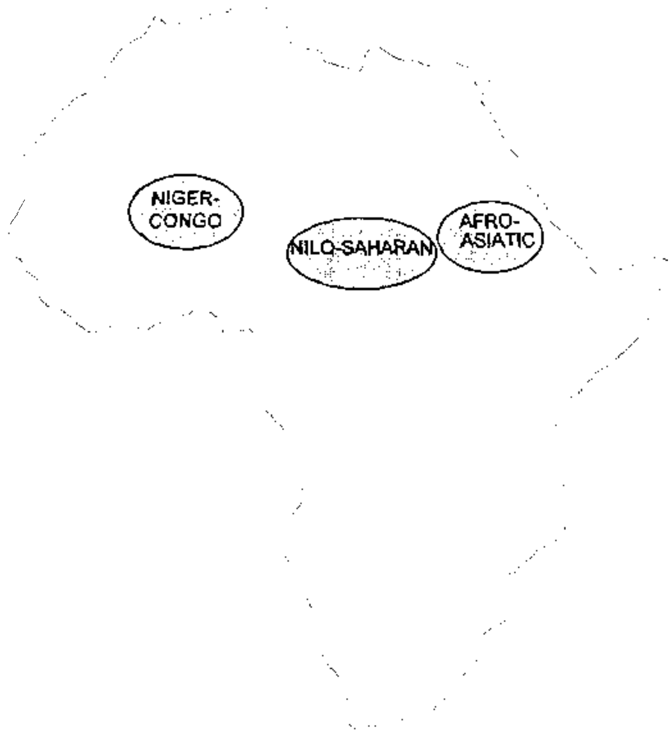
2.7.3 Languages in this Study

A summary of the languages of the groups mentioned in the text are presented in Table 2. Linguistic classification is a highly controversial field of study; instead of presenting all the arguments and possible contradictions, Ruhlen's (1987) comprehensive classification of languages will be presented. For Africa this is primarily based on Greenberg's work (1963). To locate the proposed origins of the families see Map 3.

Table 2 Language Classification of Ethnic Groups

Sub Phylum	Primary Division	Secondary Division	Language
I KHOISAN			
Hadza			*Hadza*
Sandawe			
Southern Africa	Northern		*Qxû (San/!Kung)*
II NIGER-KORDOFANIAN			
Kordofanian			
Mande	Mandenkan		*Mandinka*
Niger-Congo	West Atlantic	Northern	*Fula (Fulbe)*
	North Central	Adamawa-Ubanguan	*Baka-Gundi(Biaka)*
	South Central	Eastern	*Yoruba*
	Benue-Zambesi	Kikuyu-Kamba	*Kikuyu*
		Bantu C	*Biaka*
III NILO-SAHARAN			
Songhai			*Songhai*
Saharan	Western	Kanuri	*Kanuri*
Maban			
Fur			
East Sudanic	Nilotic	Teso-Turkana	*Turkana*
		?	*Mbuti*
IV AFRO-ASIATIC			
Ancient Egyptian			
Berber	Berber Proper	Tuareg	*Tamasheq (Tuareg)*
Chadic	West	Group A	*Hausa*
Omotic			
Cushitic	Eastern Cushitic	Omo-Tana	*Somali*
Semitic			

Language classification of languages of the ethnic groups mentioned in the text adapted from Ruhlen (1987). The four African linguistic phyla are illustrated in bold. The columns represent hierarchical classification, but are not comparable to each other in level of hierarchy. The populations studied in this thesis are indicated in italics.



Map 3 Possible original locations of major language groups.
From Ehret (1993), the locations are an approximation based on historical and present-day distributions. The origin of the Khoisan phylum cannot be placed as it is too old. The Niger-Congo is a subphylum of the Niger-Kordofanian phylum. The other subphylum of Niger-Kordofanian, Kordofanian, is spoken in the region between the proposed origin of Nilo-Saharan and Afroasiatic.

2.7.3.1 Khoisan

Khoisan has 31 extant languages and 120,000 speakers with a Southern Africa distribution (South Africa, Namibia, Southern Angola, Botswana and Northern Tanzania.). Khoisan is characterised by the click sounds. This phylum was originally spoken over most of the southern third of Africa (Ruhlen, 1987:p.114). Due to its high internal divergence, Khoisan is regarded as the oldest of the linguistic phylum in Africa. The most divergent Khoisan languages are in Tanzania. Some claim the Tanzanian languages Sandawe and Hadza are language isolates (Elderkin 1983; Fleming, 1983; cited in Blench, 1993), not classifiable within the four linguistic phyla of Africa. There are suggestions based on skeletal evidence that the Khoisan inhabited North Africa (Tobias, 1964, Cavalli-Sforza *et al.* 1994:p.176), although evidence is fragmentary.

2.7.3.2 Niger-Kordofanian

There are 180 million Niger-Kordofanian speakers who speak 1,064 extant languages throughout West, Central and Southern Africa (Ruhlen, 1987:p.95). Kordofanian is an isolated group of 30 languages in Sudan. In the case of this survey the Niger-Congo subgroup of the linguistic phylum has been surveyed. Proto Niger-Congo was spoken at least 8,000 years ago (Ehret, 1984). Fracturing of the Niger-Congo may be related to the expansion of peoples south from the Sahel with the desiccation of the Sahara (Posnansky, 1981).

Bantu

Bantu is a Niger-Kordofanian group which includes about 500 languages spoken by 100 million speakers. There is a great deal of controversy about the origins and expansions and migrations of the Bantu. There are many theories, based on evidence of linguistic proximity as well as archaeological evidence of iron production and pottery styles. Renfrew (1987; 1988) in the Indo-European context, cautions about over-interpreting conclusions based on assumptions which equate archaeology with a people and a language; this is also applicable for Africa. Previously African archaeological studies have focused on diffusionist theories, often with racist overtones, implying a superior group moves into a region displacing the previous inhabitants and bringing sophisticated technology, for example, iron and agriculture (see Sinclair *et al.* 1993).

Despite this caution about theories of Bantu origins, I will discuss the main findings of some of the more accepted hypotheses. Most agree the highest linguistic diversity of the Bantu languages is found in the area of the Cameroun-Nigerian border. The high linguistic diversity in the West has led some (Greenberg, 1963; Heine *et al.* 1977; Bastin *et al.* 1983; cited in Cavalli-Sforza *et al.* 1994:p.167) to propose a Western Bantu origin in the Cameroun area, although some propose a South Eastern origin (Guthrie, 1967; cited in Cavalli-Sforza *et al.* 1994:p.166). There is consensus that some words are borrowed from Central Sudanic groups. This is supported linguistically by Ehret (1984) and also by a genetic contribution by Cavalli-Sforza *et al.* (1994:p.183). This association with Central Sudanic groups has been interpreted as the Bantu streams passed through Central Sudanic areas, and borrowed technologies as well as words. Others have suggested the Central Sudanic groups have moved into Bantu regions. There is also agreement that there are two distinct branches of Bantu, Western and Eastern, who are distinguished by culture and language. Hypotheses of different migration paths have been proposed. The Eastern Bantu have iron technology , cereals and cattle (Huffman, 1989). Western Bantu have a different culture with stone tools and pottery but no metal (de Maret, 1982). Differentiation of Eastern Bantu and Western Bantu has been suggested based on genetic evidence (Excoffier *et al.* 1991).

2.7.3.3 Afroasiatic

Afroasiatic languages are spoken in Northern Africa and the Middle East. There are 240 Afro-Asiatic languages spoken by 175 million people of which 100 million speak Arabic. Proto-Afroasiatic may have been spoken more than 15,000 years ago (Munson, 1977; Ehret, 1979) in North Eastern Africa, where the most Afroasiatic linguistic diversity occurs today.

2.7.3.4 Nilo-Saharan

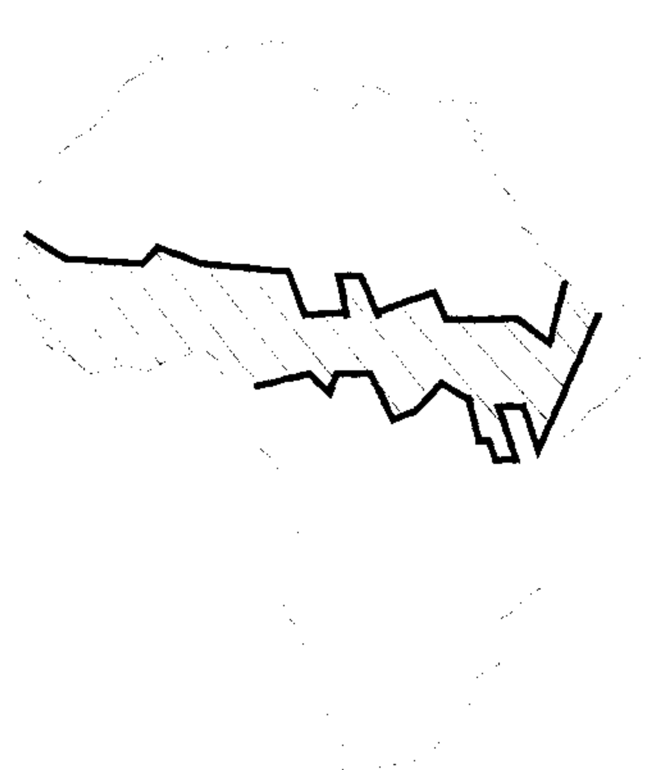
Nilo-Saharan has 138 extant languages, and eleven million speakers. Nilo-Saharan is the second most divergent phylum in Africa and is therefore seen as the second oldest after the Khoisan (Blench, 1993). Some theories claim the proto-Nilo-Saharans had an aquatic culture (based primarily on fishing), but Ehret (1993), with linguistic and archaeological evidence argues it was food producing as opposed to food collecting. Proto-Nilotes were pastoralists and Proto-Saharans agriculturists (Ehret, 1974). The arid period 8,000-6,000 B.P promoted migrations from the Sahara, perhaps leading to the scattered distribution of the Nilo-Saharan speakers. The Songhai isolate of West Africa, is the most controversial branch in this family.

2.7.4 History of Languages

Blench (1993) argues on the basis of lexical reconstructions that dispersal of phyla took place before agriculture. However he does state that there is evidence for early exploitation of domesticated animals in Nilo-Saharan and Afroasiatic communities. On the basis of this Blench (1993) argues that the speakers of these proto languages were hunting populations and, therefore, the principal stimulus to dispersal initially was pressure on the resources or sudden expansion of their availability.

Fragmentation Belt

The fragmentation belt, named by linguists (see Dalby, 1970) because of its high linguistic diversity, is wide belt running across Africa south of the Sahara. This region accounts for two thirds of African languages (Greenberg, 1981). The desiccation of the Sahara may have accounted for compressing diverse ethnic groups into a narrow region. A gradual desiccation process began 8,000 BP (Clark and Brandt, 1984) and the Sahara was completely desert 5,000 BP (Shaw, 1977).



Map 4 Fragmentation Belt
The area of high linguistic diversity is illustrated by hatching (this map is adapted from Dalby (1970). Most individuals sampled in this survey are from the fragmentation belt region.

2.8 Pre-History and History Of Africa

2.8.1 Summary of African Epochs

Table 3 provides a summary of the main geological and archaeological periods of African history.

Table 3 Summary of geological and archaeological periods in Africa.

European Terminology	African Terminology	Dates (BP)	Description
Lower Palaeolithic	Early Stone Age	2,500,000 - 200,000	Pebble culture
Middle Palaeolithic	Middle Stone Age	200,000-15,000	Blade industries
Upper Palaeolithic	Late Stone Age	15,000 - 500	Backed blades, burins, end scrapers
Neolithic		7,000	Pottery, ground stone axes, farming
Iron Age		2,000	Iron production and use

This table illustrates common terms for different eras within the African context. The Middle Pleistocene falls between 700,000-200,000 years before present; the Late Pleistocene 200,00-12,000 years before present and the Holocene from 12,000 years until the present. Information is from various sources but primarily from Clark (1994). The dates are in years before present.

2.8.2 Climate

There is some correlation in temperature changes between the different geographical regions as well as world-wide trends as suggested by the well documented Northern Hemisphere glaciations and the Ice cores from the poles and tropical glaciers (Thompson *et al.* 1995). The Intertropical convergence zone (ITCZ) is where the humid equatorial air meets the dry Saharan air. Presently this zone oscillates seasonally from 20-23^{o} North. As the polar air mass extends during glaciation episodes the ITCZ was pushed south leading to increased aridity (Said and Faure, 1990).

Table 4 Climatic summary of the Late Pleistocene and Holocene in Africa

kya		NORTH	WEST	EAST	CENTRAL	SOUTH
24				⇑	⇑	
	PLEISTOCENE			↑ ppt		
22	(Approx. 5°C↓)			°C↓	↓ ppt	
20				⇓		⇑
						°C↓
18			⇑		extension	S ↓ ppt
					of	N ↑ ppt
16			°C↓		savannah	
			↓ ppt			⇓
14	⇓		⇓			Northern
	⇑	↓ ppt	⇑		⇓	Kalahari
12				↓ ppt	⇑	Lakes
10		⇑	↑ ppt	⇑		
				↑ ppt		
8	HOLOCENE	↑ ppt		⇓	⇓	
			⇓		Lake	⇑
6			⇑	↓ ppt	regression	↑ ppt
		⇓	↓ ppt		⇑	⇓
4			⇓			⇑
			↑ ppt		↓ ppt	↓ ppt
2			↓ ppt			
0	⇓					

This is a simplified summary of climatic fluctuations. Information was collected from many sources but especially (Grove, 1993, Maley, 1993, Shaw, 1981, Clark, 1994). ⇓⇑ indicate the extent of an epoch or a climatic period. For some periods the information is incomplete. ↓ **ppt** indicates a lower amount of precipitation and ↑ **ppt** the converse. Generally lower temperatures (**°C**) are correlated with a decrease in precipitation. **kya** indicates thousand years before present.

2.8.3 Fossils and Tools

Controversy remains over the earliest dates of modern humans. The dating of fossils or artefacts is often doubted and the fossils themselves are questioned in terms of being archaic or modern. The Border Caves and Klasies River (South Africa) have been dated at 100,000 BP, and show archaic as well as modern types. Laetoli (Tanzania) has been dated at 125,000, Omo (Ethiopia) at 135,000 (Cavalli-Sforza *et al.* 1988). Smith (1992), however states that the chronological shift to modern humans may not be as far back as claimed. He states that there is no evidence exactly when anatomical modern humans first appear in southern Africa. Thermoluminescence (TL) and electron spin resonance (ESR) place the earliest definite evidence of early modern humans in Skhul and Qafzeh sites (at 92,000 BP) in Israel (Grün and Stringer, 1991; Stringer, 1988; cited in Smith, 1992). Smith, (1992) argues there is not sufficient evidence, in the African fossil record which identifies Africa as the source for all modern people.

Complex tools were first associated with European sites dated to around 30,000 years ago, but a recent find of complex small barbed bone and ivory points in Zaire dated to around 90,000 BP (Brooks *et al.* 1995), suggests that modern human behaviour is associated with anatomically modern humans. Brooks *et al.* (1995) argues these data support an African origin of behaviourally as well as biologically modern humans.

2.8.4 Culture

Two important mechanisms of cultural expansion are cultural diffusion and demic diffusion. Cultural diffusion is defined as when domestic plants or animals and technologies are passed from one group to another independent of population displacement (Edmonson, 1961). Demic diffusion is when populations grow and displace others (Ammerman and Cavalli-Sforza, 1984).

2.8.5 Domestication

Domestication is the process of withdrawing animals and plants from the natural selection process and modifying their characteristics by selective breeding to provide labour, material and/or food for humans. Associated with domestication is a change in lifestyle, from the hunter-gatherer system (food collecting) to agricultural or pastoral (food producing). Domestication increases the carrying capacity of the land and probably was associated with major population expansions.

The dates of domestication events are contentious (McIntosh and McIntosh, 1986). Cattle may be the earliest domesticate in Africa (Banks, 1984), the earliest date proposed being 9500 BP. However some claim that the earliest finds may have been wild not domestic.

Within Africa there have been several domestication centres known as cradles of agriculture. Harlan (1993) provides the following information. Ethiopia was the proposed centre for the ensete (banana) and teff (a grain crop). West Africa was a centre for the domestication of African rice, yams, and oil palm, proposed to have occurred 3500 BP although there is evidence that oil palm was domesticated by 5000 BP. Sahara was a domestication centre for sorghum, finger millet and pearl millet, cowpea, gourd and baobab proposed at around 6,000 BP. Middle East was a domestication centre for sheep, goats, cattle, emmer wheat and barley at about 10,000 BP. Apart from the coastal North Africa and the Nile Valley, only the sheep, goats and cattle where important in the rest of Africa. Egypt domesticated the cat and guinea fowl. Southern Africa domesticated Kefir/Kafir, a type of sorghum.

Domestication may have been responsible for population expansions and possibly migrations, although the transfer of technology and domesticates may have been by cultural diffusion.

Cattle

There is a theory that cattle were indigenous to the Sahara; however, most believe that cattle migrated from the Middle East area, or possibly from Europe via the Iberian peninsula. There were several migrations into Africa; the first occurred before 6500 BP when *Bos taurus* humpless longhorn cattle arrived in North Africa. The Kuri from Lake Chad area and the N'dama from West Africa are the descendants of this migration. The second migration was the *Bos taurus* humpless shorthorn which has representatives in the Sheku of Ethiopia and Muturu of West Africa. These *Bos taurus* breeds (with the possible exception of the Kuri) have adapted to the rainforest area, and are resistant to Trypannosomiasis. The third migration around 3500 BP was *Bos indicus* cervico-thoracic humped cattle via the Nile Valley; these interbred with the humpless long horn creating the Sanga breeds of Africa. The fourth migration was *Bos indicus* thoracic humped cattle associated with the Arab invasions of 680 AD (Ajayi and Crowder, 1985).

DNA results from Dan Bradley's group in Dublin have identified the *Bos taurus* and *Bos indicus* types. In Africa the pattern is curious. The DNA markers, mitochondrial and Y chromosomal, do not correlate with phenotype. All African cattle regardless of phenotype have mitochondrial markers of the *Bos taurus* type, and Y-chromosome markers of the *Bos indicus* type (Bradley *et al.* 1994; MacHugh *et al. in prep*). The autosomal markers however show an

East-West cline of *Bos indicus/Bos taurus* types. This demonstrates introgression of the *Bos indicus* genetic material through the male line.

2.8.6 Metal and the Iron Age

There are several theories about the origin of iron production and use in Africa, generally classified as the diffusionist and the indigenous (Okafor, 1993). The diffusionist school argue that iron smelting arose in West Asia and spread through Egypt and Meroe, or East Asia, and spread through Madagascar (Anfray, 1957; cited in Andah, 1979; Okafor, 1993). The indigenous school argue the people of Nok in Nigeria and Kagera in Tanzania initiated iron smelting independently possibly derived from copper smelting techniques. Nok has been dated at 2500-3000 BP (Shinnie, 1967 cited in Andah, 1979) other West African sites have been dated to 2000 BP in Chad and 2200 BP in Mali. The earliest East African sites are dated to the first millennium AD (Kiriama, 1993). The East and West African centres are supposed to have played a role in the expansion of the Bantu (Shinnie, 1967; cited in Moktar, 1981).

2.8.7 Population Expansions

Population expansions can occur when a population moves into an uninhabited area, or by replacing existing populations in an inhabited area. Humans can also expand by exploiting a new niche or making a cultural adaptation which increases the carrying capacity of the land. New innovations may have been the impetus for many expansion events, in the history and evolution of hominids and modern humans. Economic modes such as agriculture, fishing, and pastoralism could have opened up new environments. Techniques such as iron production allowed the environment to be modified further by facilitating the clearing of land for cultivation. Ammerman and Cavalli-Sforza (1994) modelled agriculture as generating a large increase in the effective population size.

Cavalli-Sforza *et al.* (1994:p.189) from genetic data has suggested several population expansions. These expansions were demonstrated with pockets of high or low allele frequencies. They suggest that West Africa underwent three population expansions, prior to the Bantu expansions. These had foci in (1), Senegal, (2), Nigeria and (3), the Mande, Gur, Ewe Volta and Kru people. Cavalli-Sforza *et al.* (1994:p.193) conclude that West Africa may have been the first part of the continent to experience population expansion from farming.

2.9 Scope of Thesis

The thesis describes the process of the genetic survey: choosing the populations; seeking permission; collecting blood samples; generating the sequence data; and analysing the data. A diverse selection of sub-Saharan groups from East and West Africa were chosen. Nine months were spent in East and West Africa, seeking permission through the local authorities (governmental and traditional) and collecting 472 blood samples. The hypervariable I region from the mtDNA was chosen as the locus to investigate, due to its high substitution rate and the large amount of existing sequences which could be used for comparison. Nine sub-Saharan ethnic groups were sequenced these groups were chosen to represent a different combination of the three variables: geography, language and economy as well as increasing the knowledge of African mtDNA sequences. The main data generated in this study was 360 bases from the mitochondrial DNA hypervariable region I from 243 individuals of the nine ethnic groups. The sequences are compared with each other and with published sequences from Africa and the rest of the World. Standard methods of analysis were used along with the new median network method to determine the genetic relationships between the ethnic groups and between the categories of linguistic phyla, economic patterns and spatial patterns and the relationship of Africa to the rest of the world.

Both regionalism and shared lineages over wide geographical areas were observed. Each ethnic group was found to be a composite of different genetic elements, obscuring linguistic, spatial and economic relationships. The mitochondrial DNA results support an African origin for all modern humans, these results are complemented by a linkage study of the nuclear CD4 locus. However, there are some still inconsistencies between the palaeontological record and the genetic information that need to be reconciled.

The thesis is arranged as three introductory sections containing a literature review, a description of sample collection and overview of mtDNA results; then follows a series of papers which are in various stages of publication.

The introduction (Section 2) provides orientation to the genetic background of this study, ethnic background of the populations chosen for the study, and a brief introduction to the palaeontological and archaeological history of Africa. Theories discussed in later papers are introduced but this is intended only as a useful reference for the reader, rather than a detailed analysis of topics. The information in the introduction needs to be placed alongside the evidence derived from the data collection and analysis.

A significant part of this doctoral study was devoted to the organisation and collection of blood samples in both East and West Africa. Section 3 is devoted to collection, and the ethics behind human diversity studies.

A significant amount of time was also spent on the laboratory work, which is summarised along with a description of the sequences in Section 4. Further results are in the Appendices (Section 11).

The concept of genetic diversity is addressed in Section 5, where Africa is again shown to have a comparatively high genetic diversity.

In Section 6, a new median network method is applied to the African sequences, showing that this method can yield informative results from mtDNA information. The median networks method extracts more information from the data than other distance based or tree building methods. The median networks method is suited to intraspecific questions as it illustrates the ancestral sequences which still exist in the species. Other tree building methods on the other hand are designed for interspecies studies. Interesting results include the dating of expansion events and the fact that almost all non-African sequences (and some African) appear to be derived from a single African lineage. This method reiterates the appropriateness of mtDNA for approaching questions concerning human ethnicity, pre-history and evolution.

Section 7 investigates previously published trees from Cavalli-Sforza *et al.* (1988) of world-wide human populations. One tree is generated from allele frequencies and the other from linguistic information. Tree comparison metrics are applied and show correlation between the two trees.

Section 8 is the discussion which synthesises the results from the previous sections, and draws conclusions and suggests which aspects need to be investigated further.

3. Sample Collection and Ethics

3.1 Introduction

Investigations into the biology and history of groups of people have the potential to conflict with deep-seated cultural and religious beliefs. In such situations there are important ethical issues to be considered, for example: whether or not the study is open ended (such as culturing cells or making a genomic library) or restricted to a particular locus; whether there is a possibility of commercial gain to researchers or volunteers; and whether or not the groups have been consulted and involved in the early planning stages.

This present project is on a much more modest scale than the Human Genome Diversity Project but faces similar questions and ethical decisions. The Human Genome Diversity Project has come under much international attention (Gillis, 1994) and criticism (Lock, 1994). The Human Genome Diversity Project is planning to establish cell lines from individuals from selected populations. This will essentially provide an infinite source of DNA for various studies in the future.

There are many contentious issues, from the scientific design such as, whether the sampling is based on ethnicity or geography (which caused the first major debates in the Human Genome Diversity Project planning), the selection of populations, and the number and storage of samples. Other ethical issues to be faced include those of the rights of indigenous people rights, fully informed consent, and feedback to the groups involved.

3.2 Ethics

A human study such as this raises many ethical concerns and issues. "Moral and ethical decisions occur at all stages of research, from selection of topic, area or population, sponsor and source of funding, to publication of findings and disposal of data" (Ellen, 1984).

In social sciences there was little formal concern about ethics among either researchers or their subjects before the mid-1950's (Ellen, 1984). Some early anthropologists, however, formulated an anthropological code of ethics, for example, Mead *et al.* 1949 (cited in Ellen, 1984). Today there is a "willingness to address these issues and processes publicly, explicitly, and more profoundly" (Emerson, 1981; cited in Ellen, 1984). Today laws exist such as an Anthropological Research Act, that was passed in Botswana in 1967 (Tebape, 1978; cited in Ellen, 1984), to protect indigenous people from potential exploitation from researchers.

Research projects dealing in any way with humans today, are required to go through approval by ethic committees. The existence of these committees can only help consolidate the research objectives and make the researcher aware of the obligations involved. It must be remembered that those who have studied indigenous groups have often been leaders in advocating respect of these groups.

3.2.1 Informed Consent

Informed consent is defined as "the knowing consent of an individual or his [/her] legally authorised representative, so situated as to be able to exercise free power of choice without undue inducement to any element of force, fraud, deceit, duress, or other forms of constraint or coercion" (Annas *et al.* 1977:291; cited in Thorne, 1980). Informed consent treats the volunteers as autonomous beings (Wax, 1980), with the power to make their own decisions. It has been claimed that informed consent is inappropriate for many situations and was originally designed to protect patients from potential abuses by medical researchers (Thorne, 1980). Wax (1980) argues that the standard consent form is irrelevant and a 'nuisance to all parties', especially when there is illiteracy and unfamiliarity to such forms, distrust of paper endorsements and lack of knowledge from which to make an appropriate response. Nevertheless, in this study informed consent was a principle adhered to.

3.2.2 Health, Exploitation and Use of Results

Responsibility to health was an issue that was highlighted by Neel's (1994) emphasis on the obligation to keep disease away from the Amerindians of the Amazonian basin. Most human genetic studies collect blood and with the actuality of the Human Immunodeficiency Virus (HIV), it is imperative in the drawing and processing of blood to protect the researcher, workers and volunteers from acquiring HIV and other diseases.

The Human Genome Diversity Project has been criticised for being exploitative; the average cost to sample each individual has been estimated to be US $2,399 (RAFI, 1993, cited in Lock, 1994) from people whose own existence is precarious. The results, even if not esoteric, will probably never affect the volunteers and their communities which took part in the survey. The potential of applications to patent the genetic material make the situation even more ironic.

The exponential growth of genetic information has opened up possibilities for abusing the data generated from diversity projects. When looking at neutral sites such as the mitochondrial control region the potential for abuse is slim; however other questions come into play including, who owns the results, who has access to the results, and who benefits from the results.

These issues were considered in the proposal made to the Massey University Ethics Committee. Approval was granted by the Committee prior to undertaking the data collection. In the next section, the procedures followed in the collection of blood samples is outlined.

3.3 Sample Collection

3.3.1 Evolution of the Project and Selection Of Ethnic Groups

The emphasis and strategy of the present study has evolved significantly since its first proposal. Initially, the emphasis was on the origins of the Fulbe of West Africa. There were several testable theories as to the origins of the Fulbe (see Introduction). Later literature research found other pastoral groups such as the Maasai of East Africa had similar theories as to their origins, so the project evolved to encompass pastoralists in general. This was the status of the project when sampling began. The ethnic groups chosen represent major pastoral groups from West and East Africa and their non-pastoral (in all cases, agricultural) neighbours for the purposes of providing outgroups for comparison (see Table 5). Pastoralists are defined as subsisting primarily on products from domesticated animals, especially, cattle, sheep, camels and goats. Non-pastoralists are defined as being agriculturists who subsist on cultivated crops.

To test for the relationships of economic mode, language and geography populations were chosen which represented a particular combination of lifestyle, language and geography (see Table 6). This also provided a diverse spectra of African peoples suitable to answer questions regarding African prehistory and human evolution in general.

Table 5 Ethnic Groups from which Blood was Sampled.

	WEST AFRICA			EAST AFRICA		
	PASTORAL⟷NONPASTORAL			PASTORAL⟷NONPASTORAL		
NIGER-KORDOFANIAN	Fulbe		Igbo Bozo Yoruba		Kikuyu	Luhyia Gusii
NILO-SAHARAN		Kanuri Songhai		Maasai Samburu Dinka Kalenjin Turkana		Luo
AFROASIATIC	Shewa Tuareg	Bella Hausa		Rendille Sakuye Gabbra Borana Somali		

The above table summarises the groups from which blood was sampled from. The groups are arranged according to attributes of language, geography and lifestyle. Pastoral and Non pastoral are illustrated as a continuum. Hausa, for example, are agriculturists, but some also own livestock. For detailed ethnological information see p. **Error! Bookmark not defined.**.

Table 6 Ethnic Groups from which DNA was Sequenced

	WEST			EAST		
	PASTORAL⟷NONPASTORAL			PASTORAL⟷NONPASTORAL		
NIGER-KORDOFANIAN	Fulbe		Yoruba		Kikuyu	
NILO-SAHARAN		Kanuri Songhai		Turkana		*
AFROASIATIC	Tuareg	Hausa		Somali		

Ethnic groups were chosen to represent a unique combination of the attributes, of geography, language and economy. Pastoral and Non-pastoral is illustrated as a continuum.

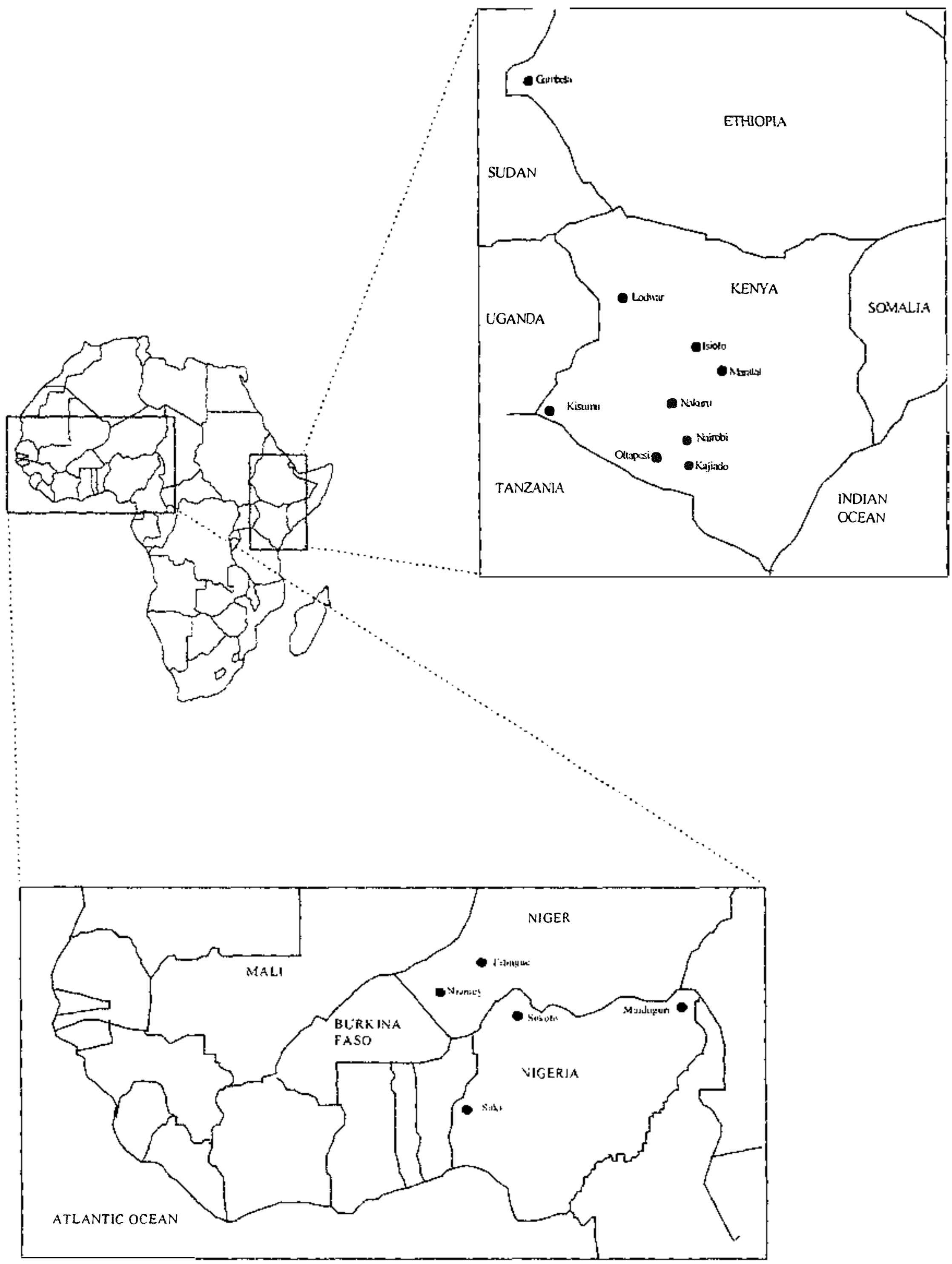
* The Luo were not sequenced as A. Di Rienzo planned to sequence the Luo from her independently collected samples.

Table 7 Places where Samples were Collected

DATE	PLACE	CLINIC	ASSISTANT	TRANSLATOR	POP.	*n*
16/6/92- 18/6/92	Oltapesi KENYA	Oltapesi Clinic		David Laur	Maasai	7
20/6/92- 23/6/92	Kisumu KENYA	South Nyanza District Hospital	Dr. Amollo Dr. Obota Dr. Okongó	Owino J'unga	Luo Maasai Luhyia Kikuyu	42 1 10 4
28/6/92- 3/7/92	Lodwar KENYA	Turkana District Hospital	George Otieno	Michael Loter	Turkana Kikuyu Luhyia Luo Borana	25 1 3 2 1
6/7/92- 8/7/92	Nakuru KENYA	Nakuru Provincial General Hospital	Dr. Nfugma	Hospital Staff	Luo Kalenjin Luhyia Kikuyu Gusii	6 5 6 10 5
28/7/92- 31/7/92	Maralal KENYA	Maralal Hospital	Dr. Okawa	Lawrence Ewesit	Samburu Maasai Rendille Kikuyu Turkana Luo	14 1 1 7 3 1
3/8/92	Maralal KENYA	Maralal Hospital	Dr. Okawa	Lawrence Ewesit	Samburu Somali Turkana Luhyia Gusii Kikuyu Kalenjin Luo	6 1 2 1 1 5 1 1
13/8/92- 28/8/92	Isiolo KENYA	Isiolo District Hospital	Dr. Gacheru	Hospital Staff	Borana Turkana Somali Sakuye Garre Gabbra Samburu Rendille	10 8 6 11 1 2 1 1

DATE	PLACE	CLINIC	ASSISTANT	TRANSLATOR	POP.	*n*
24/8/92	Kajiado KENYA	Kajiado Hospital	Bashir Farah	Hospital Staff	Maasai Rendille	7 1
31/8/92	Nairobi KENYA	Sinna Medical Clinic	Abdiraham Maalim Abdullah	Clinic Staff	Somali	21
6/10/92	Gambela, ETHIOPIA	Mekanessus		Connie Hall	Nuer Dinka Anuak	30 2 1
3/11/92	Sokoto NIGERIA	Uthman Danfodyio University Teaching Hospital	Dr. Simon Nwachuhwu	Hospital Staff	Fulbe Hausa	8 26
23/11/92	Maiduguri NIGERIA	University of Maiduguri Teaching Hospital	Prof. Idris Mohammed Dr. Jose Ambe Dr. Abba Kyani	Dr. V. Brahasuriya	Fulbe Hausa Igbo Shewa Kanuri Yoruba Others	5 5 3 4 14 1 11
30/11/92-3/12/92	Saki NIGERIA	Akintunde Memorial Hospital Baptist Medical Hospital	Dr. M Akintunde Dr. Fatula Seriki of Saki	Efe Efiawore	Fulbe Hausa Yoruba Others	36 3 22 3
17/12/92	Filingué NIGER	Centre Medical Filingué	Dr. Hamma Soumana	Hospital Staff	Songhai Hausa Fulbe Kanuri	2 6 1 1
	Niamey NIGER	Hôpital National	Dr. Amadou Sako	Mr. Chaibou Tawaye	Fulbe Tuareg Hausa Bella Songhai	10 27 4 1 10
						Σ 472

Illustrates in which hospitals and clinics the samples were collected and the dates of collection and who assisted in this collection. **N** indicates the number of samples collected for each population (**POP.**). From June 1992 to February 1993 samples were collected from clinics and hospitals in Kenya, Ethiopia, Nigeria and Niger; with the assistance of many people.



Map 5 Places where Samples were Collected

The locations where the samples were collected. Relevant regions in East (Top right) and West Africa (bottom) are enlarged.

3.3.2 Fieldwork and Sampling Etiquette

Nine months were spent collecting 472 human samples and over 1000 cattle samples. Most of the time was spent on administrative activities, i.e., talking to the local and hospital hierarchy, documenting records, labelling and filling tubes. Collecting both cattle and human samples was more efficient than collecting either one. While waiting for permission for one I could be collecting the other. Unfortunately due to time constraints, I was not able to investigate the cattle samples for this thesis.

(1) Informed Consent

Volunteers were informed of the study. Usually I employed a translator who could speak the language of the population. Most areas had a *lingua franca*, for example, Swahili in Kenya and Hausa in the Sahel region of West Africa, and this was sufficient to communicate with individuals from other ethnic groups. The aim of the project was described simply as looking at the blood to determine relationships between their ethnic group and others. Volunteers indicated consent by fingerprint, cross or signature, and the consent was confirmed by another person, usually the translator, or sometimes by the medical staff (see Information and Consent Sheet p. 214).

(2) Approaching Officials and Bribes

The initial approach to officials differed from region to region but usually the village chief and the government official chief were approached first and then the hospital director and depending on the size of the hospital all people in the hierarchy until finally the nurses. The officials and nurses from the first hospital visited, asked for substantial cash rewards. After this experience, I always stated, that as an appreciation of their assistance I would provide the hospital or clinic with gloves, lancelets and syringes (which were lacking in the Kenyan hospitals) and did not enter into any cash transactions. This was accepted, except refusing to give bribes resulted in being unable to sample in Burkina Faso due to a high official in the Department of Health. Volunteers were not given any remuneration.

(3) Volunteers

Generally volunteers were selected from people visiting Out Patient Clinics in rural clinics and hospitals. A high proportion of these visitors had malaria, and in this stressed physiological and psychological state, questions would have been obtrusive so these patients were not approached, nor were children under the age of eighteen. Most volunteers were in the clinic for routine blood tests, for HIV testing, prenatal tests, or venereal disease. Effectively the blood collected was the excess from the blood taken for the medical tests.

In most instances the nurses did not use gloves, and often needles were reused. I provided gloves and new sterile needles, syringes and lancets.

(4) Travelling with blood samples.

Letters from Massey University and the International Livestock Centre for Africa were sufficient to satisfy border officials and I experienced no problems taking samples over state or country borders. Due to the unreliability of the postal service, and other possible events, all samples and information sheets were in duplicate and sent back to Europe at different times.

(5) Selection of Subjects

I decided not to employ any selection criteria on phenetic type, due to the inherent ethical concerns and possibly biasing the sample. In West Africa for example, there are caste systems: the higher castes are referred to in the literature as being for example 'pure Tuareg', generally with fairer skin and Middle Eastern features. I based the ethnic criteria on self identification and mother tongue of the individual.

Sampling was conducted primarily in rural areas. There were two reasons for this; people living traditional lives were less likely to be recently mixed with other ethnic groups; in addition people living traditionally were more aware of their ethnic history than their urban counterparts.

(6) Blood Collection

Most towns have the government chiefs as well as the traditional chiefs. It was important to approach both and in an appropriate order, depending on cultural norms of the region.

After both groups of chiefs had given their approval, the Director/Superintendent of the hospital or clinic was approached. The research objectives and the procedure involved were explained. Permission to work within the hospital was requested. The research objectives and protocol were discussed with the technician in charge of the laboratory and then with the nurses.

The person (outpatient) was approached by myself with a translator and informed of the intentions of the project and asked if he/she would like to participate.

Five ml sterile syringes and needles were provided. Nurses withdrew approximately five mls of blood of which the standard tests required up to three mls. The volunteer was then directed to the questioner/translator, who noted the tube number and completed the information sheet (see p. 214) with questions concerning their ethnic background. The volunteer was required to sign the consent form. The translator/questioner also signed the consent form to show the volunteer gave consent freely.

Two aliquots of 0.75 ml of whole blood were suspended in 0.75 ml of buffer (100mM Tris, 100mM EDTA and 1% SDS (Rassmann *et al.* 1989) in 1.5ml screw-cap Nunc tubes. These were stored at ambient temperatures (4-50°C) before being sent to Germany where they were stored at 4°C.

3.3.3 Future of Samples

The samples were collected with the understanding that they would be used to investigate genetic relationships within African groups as well as with the rest of the world. Most officials and hospital directors requested that these samples should not be used to address medical questions. The samples will remain under my guardianship. The samples will be accessible for future research respecting the above conditions. As the material is limited, the applications will be judged on the merit of the experimental design (see Section 3).

3.4 Summary

All through this project ethical questions arose, from bribery to individuals' integrity. Although it is not realistic to assume that a list of rules can cover all ethical issues, the following is a summary of principles adhered to in this research and which I regard as important for future studies of the same nature as this one.

1. Samples should be collected with fully informed consent.
2. Appropriate officials should be approached, despite potential delays and complications. Governments' and officials' requests concerning the use of samples should be honoured.
3. No transactions of cash should be entered into.
4. Children under the age of eighteen and very ill patients should not be asked to take part in the survey.
5. Publications from the study should be sent to the government and hospitals officials involved.

The overall procedure for informed consent, and obtaining permission from central and local government, tribal elders, hospital officials and volunteers is time consuming, but important. It can also be rewarding and a great deal learned through the two-way process of consultation.

4. METHODS AND SEQUENCE DESCRIPTION

4.1 Introduction

This chapter describes the sequences generated in this study and compares them to previously published sequences, both within Africa (Vigilant *et al.* 1991; Graven *et al.* 1995) and from the rest of the world (Ward *et al.* 1991, 1993; Piercy *et al.* 1993; Shields *et al.* 1993; Torroni *et al.* 1993a, 1993b, 1994; Pult *et al.* 1994; Santos *et al.* 1994; Batista *et al.* 1995; Bertranpetit *et al.* 1995; Kolman *et al.* 1995; Mountain *et al.* 1995; Sajantila *et al.* 1995)*. The first section describes the patterns within Africa, in terms of ethnicity, linguistics, economic modes and geography. The second section places the sequences from Africa in context with the other continents. This is a descriptive chapter, without accompanying conclusions, more in depth analyses are in the following sections .

African mtDNA sequences have been studied in the Yoruba, Mbuti, Biaka, Herero, Hadza, !Kung and African-Americans (Vigilant *et al.* 1989); and Mandenka (Graven *et al.* 1995). This present study presents eight additional African populations and adds more sequences from the Yoruba population. These populations are diverse representing three linguistic phyla, are from both East and West Africa and with different economic modes (for summary see Table 6).

4.2 Methods

DNA extraction, amplification and sequencing followed standard laboratory procedures.

- **Blood Storage**

From the excess blood collected from each individual, two aliquots of 0.75ml of whole blood were suspended in 0.75ml of buffer (100mM Tris, 100mM EDTA and 1% SDS) (Rassman *et al.* 1989). These were stored at ambient temperatures in 1.5ml screw cap Nunc tubes.

- **Extraction**

To extract DNA, 400µl of the blood/buffer solution was added to 200µl of NaCl (100mM), DTT (80mg/ml), Proteinase K (10mg/ml) and digested at 37°C for 3 hours. After standard

* For overview see Section 11.3

phenol-chloroform extraction (two phenol-chloroform followed by one chloroform extraction), the DNA was concentrated by Centricon tubes (Amicon, Beverly, MA).

4.2.1 A: Hypervariable Region I

- **Amplification**

Polymerase Chain Reaction (PCR) amplification of the Hypervariable region I of the mitochondrial control region fragment was performed as a hotstart PCR using wax beads. The wax separated the upper solution containing Taq Polymerase and 10× Taq Buffer, ddH_20 and DNA, the lower solution contained Primers (H16498 5′-CCT GAA GTA GGA ACC AGA TG-3′ and L15926 5′-TCA AAG CTT ACA CCA GTC TTG TAA ACC-3′), BSA, 10× Taq Buffer and ddH_20. The 50µl PCR was amplified with step cycles of 1′ at 94°C, 1′at 58°C and 1′at 72°C for 35 cycles in a Perkin Elmer/Cetus thermal cycler. For each individual two PCR reactions were performed with the primers (H′ *normal* and L′ *biotinylated*) and (H′ *biotinylated* and L′ *normal*). Products were run on a 3% agarose gel to visualise the bands before continuing to the strand separation procedure.

- **Strand Separation and Purification**

The biotinylated product was attached to 30µl streptavidin coated Dynabeads (Dynal, Oslo, Norway) by incubating for 15 minutes in 40µl binding buffer (2M NaCl, 10mM Tris/HCl,7.5, 1mM EDTA,8.0). The beads and PCR product were washed in 40µl TE (10mM Tris,7.5, 1M EDTA,8.0). The biotinylated strand was separated by denaturing with 0.15M NaOH for 5 minutes. The supernatant containing the non-biotinylated strand was discarded. The biotinylated single strand was washed with 40µl Binding buffer followed by 40µl TE, then suspended in 15µl ddH_2O for the sequencing reaction.

- **Sequencing**

Both H (Heavy) and L (Light) strands were sequenced according to the Sequenase 2.0 protocol (United States Biochemicals Corp., Cleveland, Ohio, USA) with the nested fluorescent primers H16401 5′-TGA TTT CAC GGA GGA TGG TG-3′ and L15997 5′-CAC CAT TAG CAC CCA AAG CT-3′ (modified from Vigilant *et al.* 1989). Samples were run on a 0.35mm 6% polyacrylamide gel with an Automated Laser Frequencer (Pharmacia, Uppsala, Sweden) at 1500V, 38mA, 36W, with a 2 second sampling intervals for 400 minutes.

- **Alignment**

ASCII files were exported from the ALF Software (Pharmacia, Uppsala, Sweden), any ambiguities were resolved by consulting the printout, or by re-sequencing the region in question. The 360 bases (corresponding to positions 16024 to 16383 according to Anderson *et* *al.* 1981) were aligned with ESEE (Eyeball Sequence Editor) (Cabot 1988). The alignment was straight forward except for in the cytosine run (C-run) region from positions 16180-16193 (see Table 8). Deletions in the C-run region were optimally placed in np 16182 or 16183. When the C-run was not interrupted by another nucleotide such as a thymine, it was difficult to determine exactly how many C's were present, due to polymerase stuttering, therefore some insertion events may have been overlooked.

4.2.2 B: *HpaI*

The Restriction enzyme *Hpa*I recognises the sequence GTT/AAC, the slash indicates where the enzyme cuts. This is useful for determining a polymorphism found almost exclusively within Africa. Position 3594 (according to Anderson *et al.* 1981) is a cytosine in most human populations however in some Africans position 3594 is a thymine which the *Hpa*I enzyme recognises and cuts at position 3592. The state of the *Hpa*I site was determined by PCR amplifying the surrounding region with the primers L03526 5′-CAT-CAC-CCT-CAT-CAC-CG-3′ and H03706 5′-ATT-GTT-TGG-GCT-ACT-GCT-CG-3′. A control fragment (containing a *Hpa*I site at position 5691) was amplified by the primers L05269 5′-TTG-CCC-AAA-TGG-GCC-ATT-AT-3′ and H06011 5′-TGG-CCC-AGG-TCG-GCT-CGA-AT-3′. The individual's fragment was amplified in 25μl reactions with 5′ denaturation at 94^{o}C followed by step cycles of 1′ at 94^{o}C, 1′at 54^{o}C and 1′ at 72^{o}C for 25 cycles in a Perkin Elmer/Cetus thermal cycler. The control fragment was amplified under the same conditions except in 100μl reactions. 10μl of both the control fragment and the fragment amplified from the individual were digested together for three hours with *Hpa*I and run on a 4% Metaphor Gel (Figure 1). For further details and results see Section 4.4 p. 78.

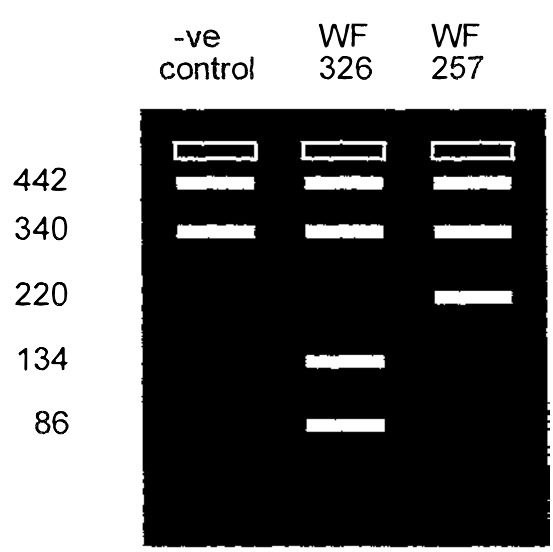


Figure 1 Representation of *Hpa*I Digestions
Representation of *Hpa*I digestion fragments on 4% Metaphor Gel. Negative (H_2O) control shows only the control fragment, which is completely digested. Individual WF326 is *Hpa*I positive showing the control fragment plus a previously amplified *Hpa*I positive fragment both of which have been digested and run together. Individual WF257 is *Hpa*I negative showing the fragment has not been cut at the HpaI site (np 3592). These were run with a 1kb ladder.

4.2.3 C: Nine base pair repeat

The Anderson reference sequence (Anderson *et al.* 1981) has two copies of a nine base pair sequence (CCCCCTCTA) in the COII/tRNA^{Lys} intergenic region. In some Asian, Native American, Melanesian and Polynesian individuals there is a deletion of one repeat. The deletion has also been reported in some Pygmy mtDNA (Vigilant, 1990). A triplication event (3 copies of the nine base pair segment) has been reported in the Tharu from Nepal (Passarino *et al.* 1993). The number of repeats was determined by amplifying the flanking regions with H8297 5′-ATG-CTA-AGT-TAG-CCT-TAC-AG-3′ and L8196 5′-ACA-GTT-TCA-TGC-CCA-TGG-TC-3′. The 25µl PCR reaction was amplified 5′ denaturation at 94°C followed by step cycles of 1′ at 94°C, 1′at 58°C and 1′ at 72°C for 30 cycles in a Perkin Elmer/Cetus thermal cycler. The products were run on a 4% Metaphor Gel with pre-amplified fragments of one and two repeats in length. For further details and results see Section 4.5 .

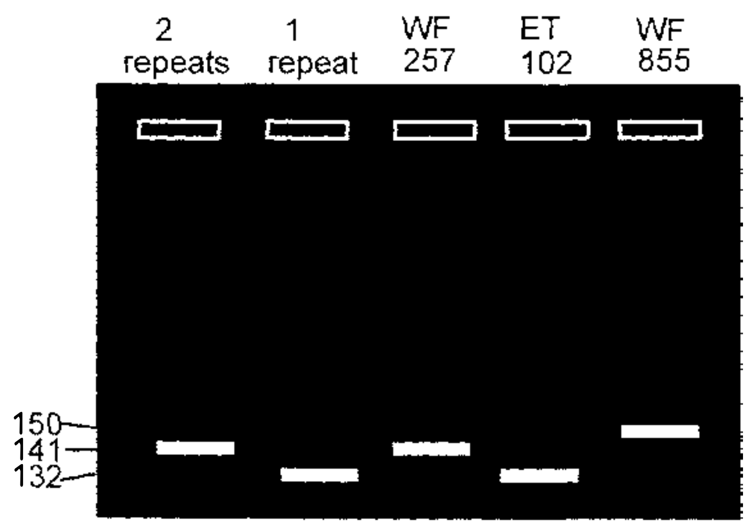


Figure 2 Resolving the Nine base pair repeat
Controls were run of individuals with previously determined nine base pair repeats. Individual WF257 shows two copies of the nine base pair repeat, individual ET102 shows a deletion event with only one copy of the nine base pair, and individual WF855 shows a triplication with three copies of the nine base pair repeat. These were run with a 1kb ladder.

4.3 Hypervariable Region I

This section reports 243 sequences of 360 unambiguous bases in the Hypervariable Region I (16024-16383) of the mtDNA from nine African populations (Kikuyu, Somali, Turkana, Kanuri, Songhai, Hausa, Tuareg, Yoruba and Fulbe) with two additional sequences from the Sakuye and Borana from East Africa. These sequences are analysed in addition to 168 previously published African sequences (for location see Map 2) and 779 non-African sequences. For each study in this thesis all sequences with ambiguous bases were excluded from analysis. The number of sequences analysed varies according to the region studied, for example, in Section 6 more sequences are used because the region studied stretches from 16090-16365, as opposed to other sections (such as Section 5) which study the region from 16030-16370.

4.3.1 Substitution Rate within the Hypervariable Region I

Hasegawa *et al.* (1993) collated $n_1 = 450$ sequences from different studies including Vigilant *et al.*'s (1989; 1991), Horai and Hayasaka (1990), Di Rienzo and Wilson (1991) and Ward *et al.* (1991); and counted the number of changes at each position on the most parsimonious tree known for this data. This was repeated for the data generated in this study by selecting the best parsimonious tree from fifty trees (DNA parsimony algorithm: PHYLIP 3.5 (Felsenstein, 1990)) and plotted together with Hasegawa *et al.*'s (1993) results (see Figure 3). The use of parsimony introduces errors into the analysis, but makes the method overall conservative (Wakeley 1993). The values above the x-axis are from Hasegawa *et al.* (1993), $n_1 = 450$. The values below the x-axis are from the data generated in this study ($n_2 = 243$). There is high correlation ($r=0.77$, $P<0.001$) between the independent data sets, confirming that the rate of change is highly variable between sites. Substitutions at some sites are frequent whereas other sites appear conserved. Wakeley (1993) also used a parsimonious method for rate comparison and arrived at a similar conclusion and concluded that the substitution rates are approximately gamma distributed among sites.

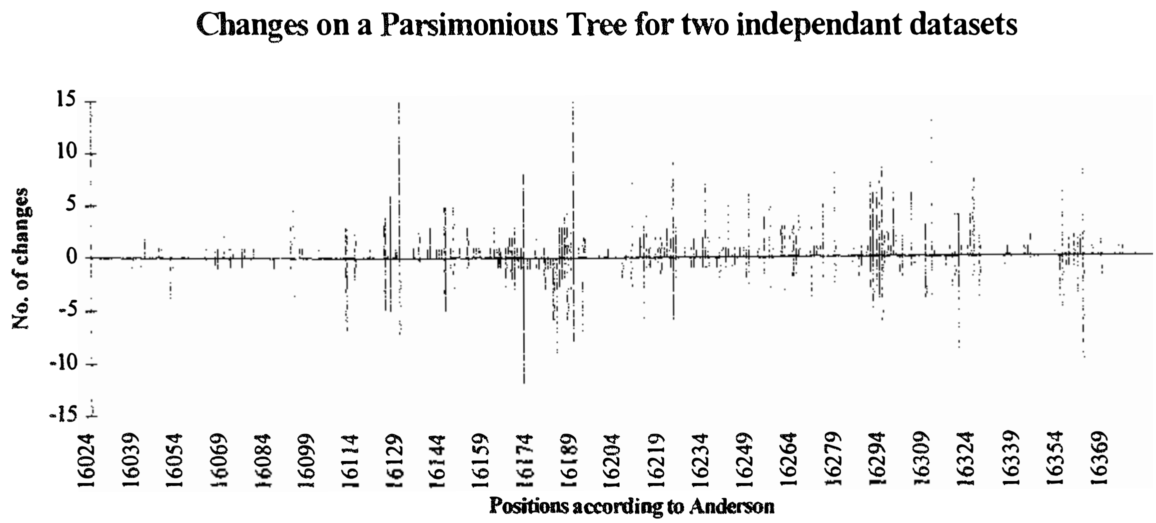


Figure 3 Variable Positions in the HVR-I region.
The positions (from 16024 to 16383; Anderson *et al.* (1981) are marked along the horizontal axis. The number of changes generated from the most parsimonious tree (DNA parsimony algorithm: PHYLIP 3.5 (Felsenstein, 1990)) are on the vertical axis. Hasegawa *et al.* (1993) results n_1 are indicated above the horizontal axis and the results from this study n_2 are plotted below the axis. The two samples are not normalised for sample size.

4.3.2 Transitions and Transversions in the Hypervariable Region I

Figure 4 illustrates which sites in the Hypervariable region I (HVR-I) have transition events, transversion events or both. The data set used is from a world-wide sample (n = 1036) combined with data generated in this study (n = 243), giving a total of 1279 sequences. Sites with transitions outnumber transversions 165 to 37 (approximately 4.5:1) in the 360 sites analysed. The transition/transversion ratio is obviously higher on a phylogenetic tree due to the same sites (usually transitions) flipping backwards and forwards. The transition/transversion ratio is characteristic of the HVR-I in particular (see Section 2.3.1).

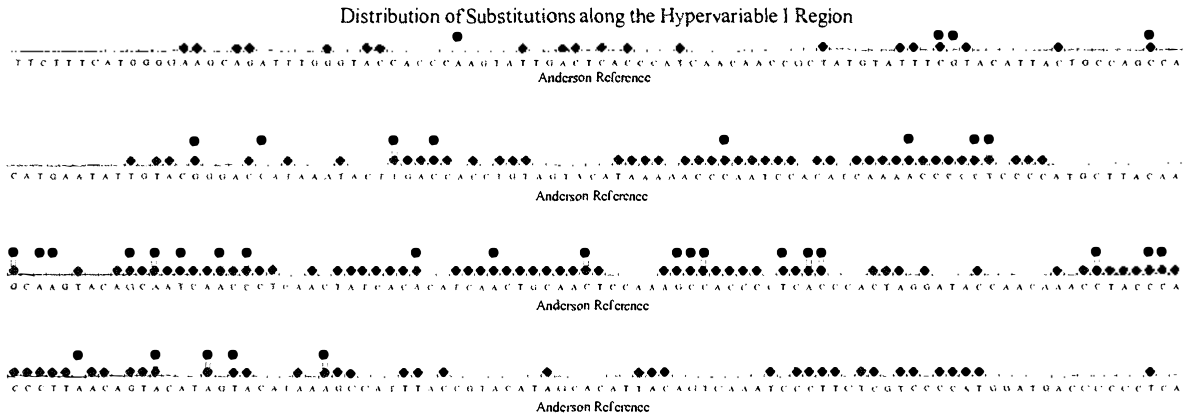


Figure 4 Distribution of Substitutions along the HVR-I
Circles represent transversions and the diamonds represent transitions. The vertical lines indicate that both transitions and transversions are observed at that site. The region illustrated is from np. 16024 to 16383.

4.3.3 Variability within the C-Run Region

Horai *et al.* (1993) illustrated deletion events associated with the C-run region which had not been reported earlier. This region np 16180 to 16193 is defined by a short run of A's (2-4 bases) followed by a run of cytosines (up to 12 bases) usually interrupted by one thymine although there are many possible variations on this (see Table 8). Different authors align this region differently, therefore potentially biasing results. Due to stuttering of the polymerase during sequencing it was not always possible to count the exact number of C's for the long C runs; in those cases either end of the run has been used as an anchor for alignment. The table below illustrates the frequencies of the corresponding sequence for 391 Africans (243 from this study and 148 from Vigilant *et al.* (1991) and Graven *et al.* (1995)). Horai *et al.'s* (1993) sample included 10 Africans, 20 Europeans, 91 Asians and 72 Native Americans. The number of possible combinations indicates how hypervariable this region is. Obviously in some cases the mechanism of change is slippage, therefore assigning the apparent transversions with the weight of transversions can add give misleading results.

Table 8 Variability within the C-Run Region

SEQUENCE	FREQUENCY 193 Samples Horai *et al.* (1993)	FREQUENCY 391 Africans
CAAAACCCCCTCCCCA	111	191
CAAAACCCTCCCCCCA	5	80
CAAAACCCCCCCCCCA	15	23
CAAAACCCTGCCCCCA	1	17
CAAACCCCCCCCCCCA	23	16
CAAAACCCCCCCCTCA		15
CAAAACCCTACCCCCA		6
CAACCCCCCCCCCCCA	15	6
CAAAATCCTCCCCCCA		4
CAAACCCCCCCCCTCA		4
CAAAACCCCCTCCTCA	3	4
CAAAACCTCCTCCCCA		3
CAA-CCCTCCCCCCCA		2
CAAAACCCCACCCCCA		2
CAAA-CCCTCCCCCCA		1
CAAA-CCTCCCCCCCA		1
CAAAACCCCGCCCCCA		1
CAAAACCCTCACCCCA		1
CAAAACCCTCCCCTCA		1
CAAAACCTTCCCCCCA		1
CAAAACTCCCTCCCCA	2	1
CAAAATCCCCCCCCCA		1
CAAAATCCCCTCCCCA		1
CAAAATCCTGCCCCCA		1
CAAACCCCCTCCCCCA		1
CAAAGCCCCCTCCCCA		1
CAGAACCCCCTCCCCA		1
CCAAACCCCCCCCCCA		1
TAAAACCCCCTCCTCA		1
TAAAACTCCCTCCCCA		1
TAAACCCCCCCCCCCA		1
TAACCCCCCCCCCCCA		1
NAAAACCCTCTCCCCN	14	
NAAAACCCCCTCCCTN	2	
NAAAATCCTACCCCCN	1	
NAAAACCTCCCCCCCN	1	

The first column illustrates the sequence surrounding the C-run region, from np 16179 to 16194 and the next columns indicate the frequency of occurrence in two different datasets.

4.4 *Hpa* I (np 3594)

*Hpa*I positive cuts at position 3592 if there is a thymine in position 3594. *Hpa*I has high frequencies in Africa (Denaro *et al.* 1981; Chen *et al.* 1995), and low in the Middle East and Sicily (Bonné-Tamir *et al.* 1986, Semino *et al.* 1989). This is probably due to African admixture. Thymine in position 3594 corresponds to the *Hpa*I-Morph 3 from Restriction Enzyme Analysis. This defines the group Chen *et al.* (1995) calls Haplogroup L, for which they estimate a divergence time of 98-129,000 years.

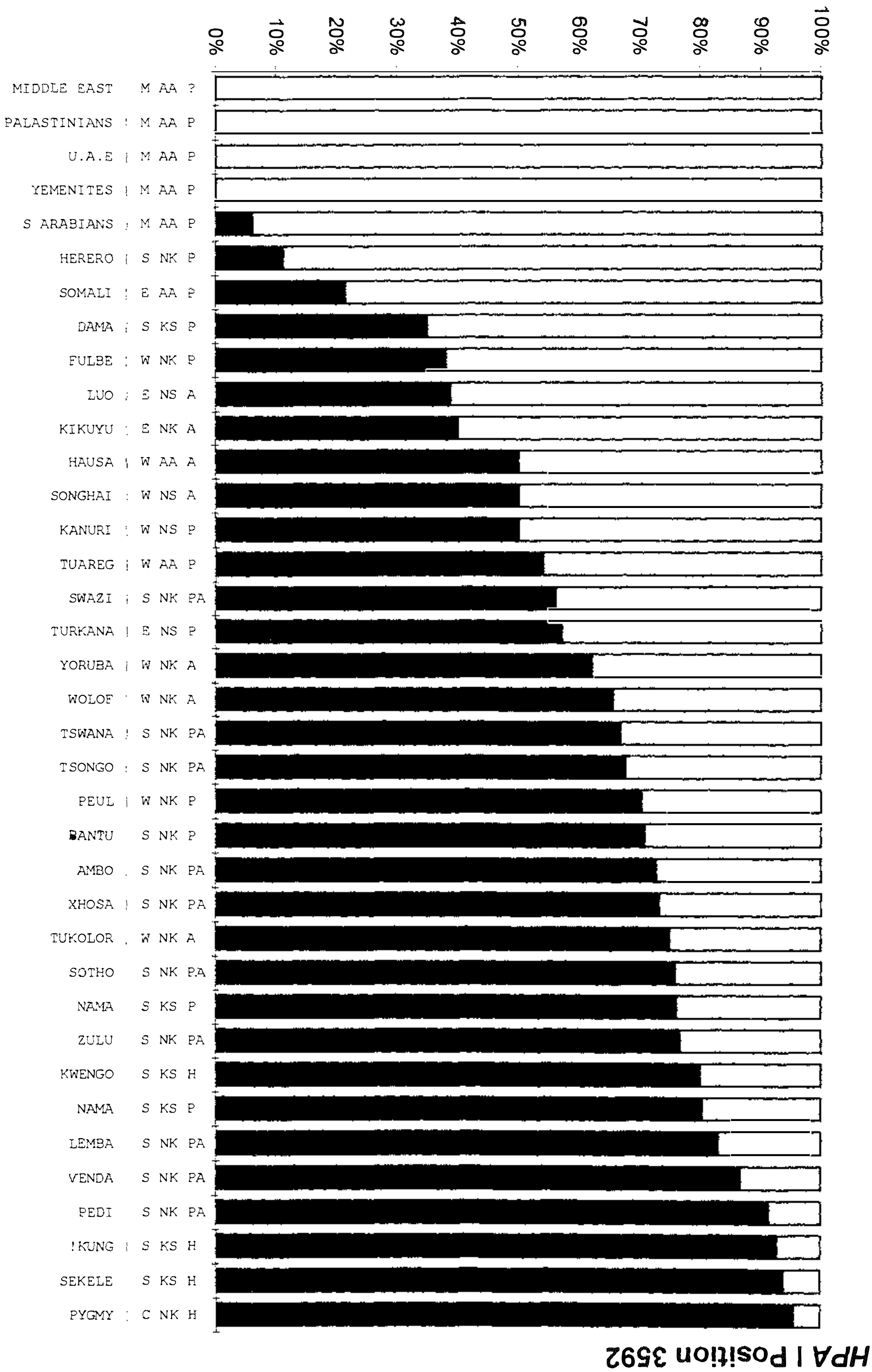
Table 9 *Hpa*I frequencies in the populations in this study.

Population	*Hpa* I -	*Hpa* I +	% *Hpa*I +
Borana	1	0	
Fulbe	38	23	38
Hausa	10	10	50
Kanuri	7	7	50
Kikuyu	15	10	40
Sakuye	1	0	
Somali	21	6	22
Songhai	5	5	50
Tuareg	12	14	54
Turkana	16	21	57
Yoruba	8	13	62

Key to following graph

The Populations are marked along the x axis with a code. The first category is geographical location: M = Middle East; W = Western; E = Eastern; C = Central; S = Southern Africa. The second category is linguistic phyla: AA = Afro-Asiatic; NS = Nilo-Saharan; NK = Niger-Kordofanian. The third category indicates lifestyle: P = Pastoralism, A = Agricultural; H = Hunter-Gatherer, PA = Pastoralism/Agricultural.

Sekele, Nama, !Kung, and Kwengo are from Soodyall and Jenkins (1992). Dama, Swazi, Tswana, Tsongo, Ambo, Xhosa, Sotho, and Lemba are from Soodyall (1993). Pygmy, Bantu and !Kung are from Denaro *et al.* (1981). Wolof, Tukolor and Peul are from Scozzari *et al.* (1988). N.B., for this graph the Pygmy groups are combined.


0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
MIDDLE EAST M AA ?
PALASTINIANS M AA P
U.A.E M AA P
YEMENITES M AA P
S ARABIANS M AA P
HERERO S NK P
SOMALI E AA P
DAMA S KS P
FULBE W NK P
LUO E NS A
KIKUYU E NK A
HAUSA W AA A
SONGHAI W NS A
KANURI W NS P
TUAREG W AA P
SWAZI S NK PA
TURKANA E NS P
YORUBA W NK A
WOLOF W NK A
TSWANA S NK PA
TSONGO S NK PA
PEUL W NK P
BANTU S NK P
AMBO S NK PA
XHOSA S NK PA
TUKOLOR W NK A
SOTHO S NK PA
NAMA S KS P
ZULU S NK PA
KWENGO S KS H
NAMA S KS P
LEMBA S NK PA
VENDA S NK PA
PEDI S NK PA
!KUNG S KS H
SEKELE S KS H
PYGMY C NK H
HPA I Position 3592

4.5 Nine Base Pair Sequence

From comparison with the HVR-I region, it can be determined that the triplication of the nine base pair sequence occurred as two independent events one occurring in Cluster-I (WH540) and the other in Cluster-III (WF855). The deletion occurred as three independent events one occurring in Cluster-III (WY720), another Cluster-V (ET129, EK68 & ET168) and another in Cluster-VI (ET102). The Pygmies with the deletion reported by Vigilant (1990) fall in Cluster-V. This was confirmed by a neighbor-joining tree (tree not shown). These results indicate that the loss or gain of the nine base pair sequence in the intergenic region is reasonably fast and is therefore by itself is not an informative marker within Africa. This polymorphic site should be used only in conjunction with Control Region sequences or RFLP information.

Table 10 Number of Repeats in the populations surveyed

Population	1	2	3	Total
Borana		1		1
Fulbe		60	1	61
Hausa		19	1	20
Kanuri		14		14
Kikuyu	1	24		25
Sakuye		1		1
Somali		27		27
Songhai		10		10
Tuareg		26		26
Turkana	3	34		37
Yoruba	1	20		21
Total	**5**	**236**	**2**	**243**

The number of the 9-base pair repeats in each population is recorded. The overall frequency is low, 2.1% of the sequences have only one repeat (i.e. deletion) and 0.8% have three repeats (i.e. triplication).

4.6 African Ethnic Groups

4.6.1 Pairwise distances *

Visualising population structure and history by pairwise distances (also known as mismatch distribution) was developed by Li (1977) and extended by Rogers and Harpending (1992). Pairwise distances are presented as histograms showing the relative frequencies of pairs of individuals who differ by i sites, where $i = 0,1,...$. "Waves" defined by one peak are seen in published pairwise distances (Rogers and Harpending, 1992). This has also been generated from simulation studies where a population has undergone an expansion event or gone through a bottleneck (Rogers and Harpending, 1992). Simulation studies of populations in equilibrium (i.e. maintaining a population size) show on the other hand very ragged curves (Rogers and Harpending, 1992). There will be more than one model that generates a particular distribution (Marjoram and Donnelly, 1994), for example, the fusion of two distinct populations each with relatively homogenous sequences would generate a bimodal distribution.

4.6.2 Hudson Test *

The Hudson Test (Hudson *et al.* 1992) aims to detect genetic subdivision between two or more sub-populations, based on genetic distances. The statistical significance of the test is generated with Monte Carlo Simulations.

Where, $K_{ST} = 1 - \frac{K_S}{K_T}$

K_S equals is a weighted average of K1 and K2,

where, $K_S = wK_1 + (1 - w)K_2$

K_T is the average number of differences between all sequences from the two subpopulations combined. When there is no subdivision K_{ST} is near 0. The Probability of K_{ST} gives the proportion of random partitions of the distance matrix that give the randomised K_{ST} > observed K_{ST}. In other words, the pairwise distance entries from two populations are randomly assigned

* Programs written by A. von Haeseler

between the subpopulations and the K_{ST} recalculated; this is repeated one thousand times. The K_{ST} value is compared to the randomised results to estimate the Probability of K_{ST} (P of K_{ST}). P of K_{ST} value (P>0.05) rejects the null hypothesis that the two subpopulations are not genetically differentiated, i.e. the populations can not be distinguished between each other by this test.

4.6.3 Distance Statistic

$D_S = \max\left\{0, D_{12} - \frac{D_1 + D_2}{2}\right\}$, where D_{12} is the mean pairwise distance between two populations 1 and 2 and D_1 and D_2 is the mean pairwise difference between population 1 and 2, respectively (Nei, 1987:p.276: formula 10.21). This statistic is based on the assumption that the two populations being compared represent two independently drifting samples from a single ancestral population.

4.6.4 Ethnic Groups Studied

Table 11 Sampled populations

Population	Code	Source	*n*	Language Phyla	Economy
Biaka	CB	2	17	NK	H
Mbuti	CM	2	11	NS	H
Kikuyu	EK	1	25	NK	A
Somali	ES	1	27	AA	P
Turkana	ET	1	37	NS	P
!Kung	SK	2	19	KS	H
Fulbe	WF	1	61	NK	P
Hausa	WH	1	20	AA	A
Kanuri	WK	1	14	NS	P
Mandenka	WM	3	87	NK	A
Songhai	WS	1	10	NS	A
Tuareg	WT	1	26	AA	P
Yoruba	WY	1+2	21+12	NK	A
			$\sum = 387$		

'Code' designates the population. The first letter represents geographical location ***C*** Central, ***E*** Eastern, ***S*** Southern, or ***W*** Western Africa. The second letter represents the population. This nomenclature for the ethnic groups will continue throughout the thesis. 'Source' of the samples are indicated by the numbers ***1*** this thesis, ***2*** Vigilant *et al.* (1991), and ***3*** Graven *et al.* (1995). *n* is the sample size of each population. The four language phyla present in Africa are designated: ***AA*** Afroasiatic, ***KH*** Khoisan, ***NK*** Niger-Kordofanian and ***NS*** Nilo-Saharan. The economic lifestyle of the population is ***H*** Hunter-Gatherer; ***P*** Pastoralist and ***A*** Agriculturist. Note that there are two additional sequences, not included in the population analysis, but are in the continental analysis. They are from one Sakuye individual from Kenya and one Borana individual from Ethiopia.

Table 12 Mean pairwise distances within African ethnic groups

	CB	CM	EK	ES	ET	SK	WF	WH	WK	WM	WS	WT	WY
CBiaka [FC]	**8.118** **5.294**	10.797 4.171	11.334 3.057	11.689 2.656	11.870 3.293	8.850 2.455	11.515 2.613	10.897 2.594	11.723 2.828	11.360 2.602	10.545 2.789	11.068 2.671	10.979 2.987
CMbuti [FC]		**6.364** **5.060**	9.404 2.645	9.983 2.416	10.133 2.976	8.809 1.421	9.955 2.224	9.636 2.565	9.714 2.486	10.100 2.302	10.545 2.788	9.738 2.598	9.807 2.438
EKikuyu			**8.670** **3.412**	8.388 3.216	9.664 3.533	9.387 1.686	8.306 3.205	7.842 3.093	8.226 3.373	8.292 3.211	9.304 3.291	8.118 3.398	8.342 3.086
ESomali				**7.687** **2.903**	9.799 3.327	9.747 1.692	7.704 2.749	7.174 2.604	7.667 2.929	7.853 2.795	8.881 2.967	7.486 2.937	7.693 3.026
ETurkana					**10.650** **3.657**	9.915 2.215	9.798 3.280	9.285 3.135	9.722 3.385	9.745 3.188	10.627 3.138	9.619 3.356	9.722 3.126
SKung [FC]						**2.866** **2.681**	10.058 1.467	9.536 1.433	10.083 1.845	10.136 1.514	10.627 3.138	9.626 1.699	9.405 1.565
WFulbe							**7.092** **3.196**	6.659 2.815	7.203 3.183	7.002 3.189	8.210 3.140	7.153 3.085	7.400 3.117
WHausa								**6.158** **2.372**	6.707 2.846	6.657 2.845	7.600 3.142	6.604 2.749	6.948 2.817
WKanuri									**7.451** **3.246**	7.169 3.186	8.407 3.363	7.129 3.258	7.481 3.153
WMandenka										**6.591** **3.547**	8.240 3.364	7.125 3.157	8.491 3.214
WSonghai											**8.407** **3.363**	8.146 3.406	8.490 3.214
WTuareg												**7.105** **3.214**	7.374 3.080
WYoruba													**7.693** **3.026**

The upper right triangle summarises the results from the inter-population mean pairwise distance analysis, the upper value in each cell is the mean and the lower value is the standard deviation. The cells on the diagonal (marked in bold) are the intra-population mean and standard deviation.

The Turkana have the highest mean pairwise distance and the !Kung the lowest. There is no obvious correlation between the diversity of an ethnic group (mean pairwise distance) and geography, language or economy. The food producing groups share lineages between them (see Interpopulation Network Figure 12 and Section 10.4), whereas the food collectors share no lineages between groups. The food producing groups show the mean pairwise distance is no higher between the groups than within them. However, in contrast the food collecting groups have higher mean pairwise distance between the groups than within.

[FC] Food Collectors

Table 13 Genetic Distances and results from the Hudson Test

	CB	CM	EK	ES	ET	SK	WF	WH	WK	WM	WS	WT	WY
CBiaka	x	3.56	2.94	3.79	2.79	3.36	3.91	3.76	3.94	4.01	2.93	3.46	3.07
CMbuti	0.191 0	x	1.89	2.96	1.63	4.19	3.23	3.28	2.81	3.62	2.74	3.00	2.78
EKikuyu	0.147 0	0.094 0	x	0.21	0.00	3.62	0.42	0.43	0.17	0.66	0.35	0.23	0.16
ESomali	0.190 0	0.146 0	0.013 0.048	x	0.63	4.47	0.32	0.25	0.10	0.71	0.42	0.09	0.25
ETurkana	0.010 0	0.057 0.001	**0.000 0.385**	0.003 0	x	3.16	0.93	0.88	0.67	1.12	0.68	0.74	0.55
SKung	0.244 0	0.327 0	0.228 0	0.280 0	**0.152 0.125**	x	5.08	5.03	4.93	5.41	4.66	4.64	4.12
WFulbe	0.156 0	**0.108 0.077**	0.029 0.001	0.018 0	0.050 0	0.234 0	x	0.03	0.00	0.16	0.04	0.06	0.01
WHausa	0.214 0	0.199 0	0.028 0.002	0.018 0.032	0.043 0	0.361 0	**0.002 0.269**	x	0.00	0.28	0.00	0.00	0.02
WKanuri	0.205 0	**0.171 0.132**	**0.009 0.142**	**0.000 0.164**	0.003 0.001	0.340 0	**0.003 0.658**	**0.007 0.774**	x	0.15	0.06	0.00	0.00
WMandenka	0.139 0	0.100 0	0.032 0	0.037 0	0.057 0	0.213 0	0.011 0.014	0.013 0.014	**0.000 0.132**	x	0.32	0.28	0.14
WSonghai	0.157 0.003	0.156 0.003	**0.016 0.105**	0.020 0.031	0.022 0.018	0.301 0	**0.001 0.322**	**0.003 0.536**	**0.000 0.329**	**0.008 0.077**	x	0.00	0.02
WTuareg	0.184 0	**0.158 0.115**	**0.015 0.060**	**0.003 0.144**	0.038 0	0.303 0	**0.003 0.191**	**0.003 0.612**	**0.001 0.896**	0.015 0.012	**0.002 0.471**	x	0.00
WYoruba	0.152 0	0.127 0	**0.010 0.089**	0.016 0.021	0.029 0	0.248 0	**0.000 0.224**	**0.002 0.334**	**0.000 0.699**	**0.004 0.384**	**0.001 0.384**	**0.002 0.534**	x

The upper right triangle summarises the Corrected D_S values (× 100). These were calculated from $D_S = \max\left\{0, D_{12} - \frac{D_1 + D_2}{2}\right\}$, where D_{12} is the mean pairwise distance between two populations 1 and 2 and D_1 and D_2 is the mean pairwise difference between population 1 and 2, respectively (Nei, 1987:p.276: formula 10.21). The lower left triangle summarises the results from the Hudson test, the top figure in each cell is the K_{ST} value $K_{ST} = 1 - \frac{K_S}{K_T}$ and the lower figure is the Probability of K_{ST} (zero is to 3 decimal places). The null hypothesis that the two subpopulations are not genetically differentiated can be rejected when the Probability of $K_{ST} < 0.05$. The comparisons which reject the null hypothesis are indicated with black squares and white text.

The Hudson test (see Table 13), suggests that most ethnic groups are distinct from each other, but that there is generally no subdivision between the West African populations. The Kikuyu and Turkana of East Africa also show no subdivision. The distance statistic (D_S) generally correlates to the Hudson test results (see Table 13). Ten other pairs of populations show no subdivision between them, including curiously the Turkana and the !Kung.

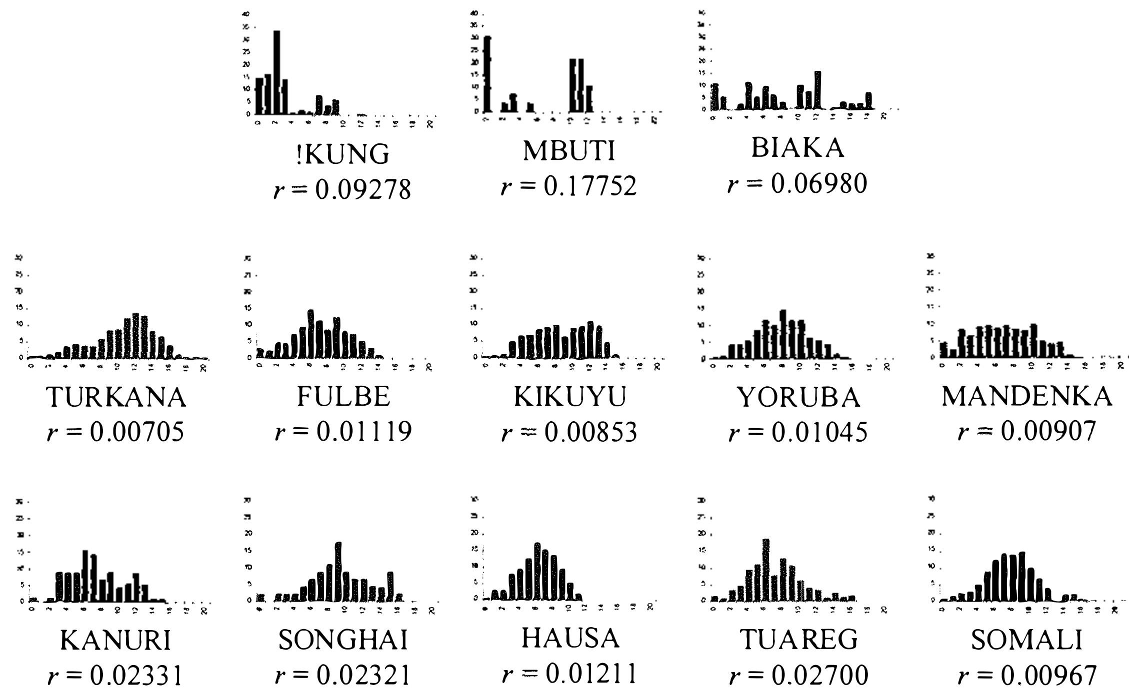


Figure 5 Pairwise Distance Distribution of African Ethnic Groups

The horizontal axis is the distribution of the number of substitutions between the pairs of sequences. The vertical axis is the percentage of pairs. Note the vertical axis for the food collectors (the top row) have a maximal value of 40%, whereas the vertical axis for the food producers have a maximal value of 30%. r represents the raggedness value described by Harpending *et al.* (1993) defined as $r = \sum_{i=1}^{d}(x_i - x_{i-1})^2$.

The pairwise distance distributions due to the small sample sizes should not be overinterpreted. There is a tendency, however, for the food producing groups to have a smooth distribution in contrast to the food collecting groups which have ragged curves, this has been confirmed with the raggedness value $r = \sum_{i=1}^{d}(x_i - x_{i-1})^2$ Harpending *et al.* (1993). Several measures of raggedness were calculated, but simulation studies demonstrated that these measures were not stable, for example small sample sizes tend to increase the raggedness values significantly (data not shown), and therefore caution should be applied to the interpretation of raggedness values. The raggedness statistic should be used only as a guide as it is an *ad hoc* measure of non-smooth distributions (Sherry *et al.* 1994) and is sensitive to sample size and diversity. The results (**Table 17** and **Table 18**) are consistent with simulation studies that have generated ragged curves from populations which remain a constant size (Rogers and Harpending, 1992). There are however, several population histories which could lead to this shape of distribution and the results should be interpreted as being consistent with a constant population size, rather than proof of it.

4.6.5 Number of Variable Positions in African populations

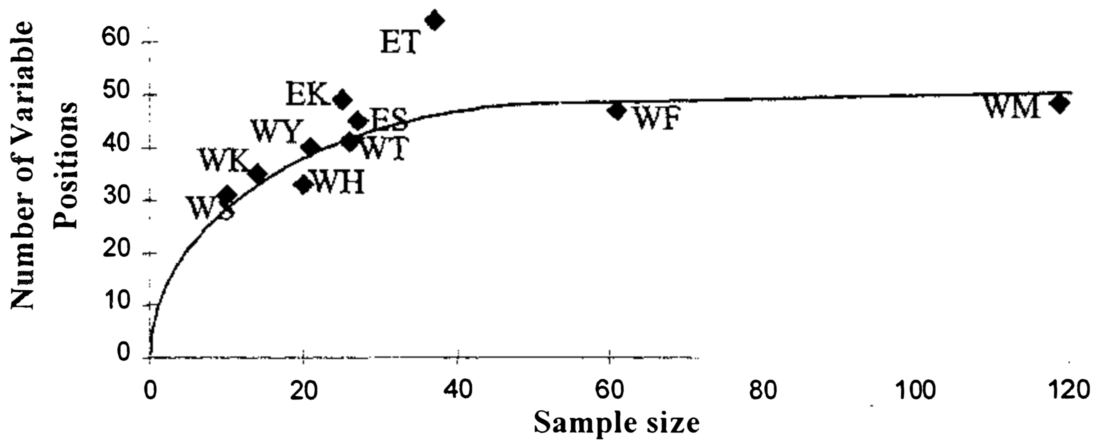


Figure 6 Number of variable positions versus sample size for African populations
The populations are labelled EK (Kikuyu), ES (Somali), ET (Turkana), WK (Kanuri), WF (Fulbe), WH (Hausa), WS (Songhai), WT (Tuareg), WY (Yoruba), and WM (Mandenka, Graven *et al.* 1995). The curved line is for comparison with Figure 7.

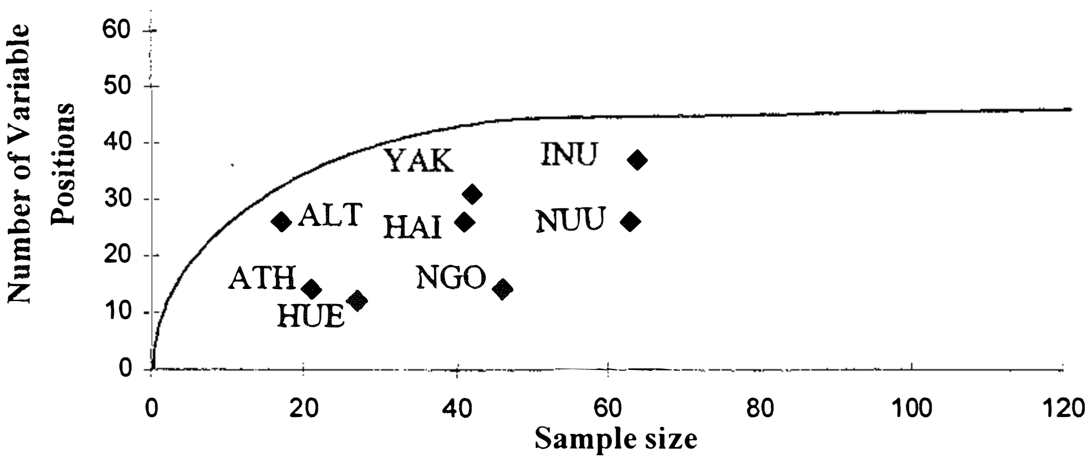


Figure 7 Number of variable positions versus sample size for a selection of American and Circumarctic populations
The populations are labelled NUU (Nuu-Chah-Nulth, Ward *et al.* 1991); ALT (Altai, Shields *et al.* 1993); HAI (Haida, Ward *et al.* 1993); HUE (Huetar, Santos *et al.* 1994); YAK (Yakima, Shields *et al.* 1993); INU (Inuit, Shields *et al.* 1993); and NGO (Ngöbé, Kolman *et al.* 1995). The curved line is for comparison with Figure 6.

The 389 Africans included in this survey had 127 variable positions in the region np 16024 to 16383. The figure above illustrates the relationship between sample size and number of variable positions. The Turkana show relatively more variable positions for their sample size which is reflected in their extremely high diversity. The African populations generally have more variable positions than other populations* as illustrated by a selection of Native American and Circumarctic populations (see Figure 7).

4.6.6 Simulations of Coalescent Times

(Analysis done with A. von Haeseler)

Table 14 Coalescent times with different population histories

Location in Africa	Population	constant size	exponentially growing	sudden expansion
South	**!Kung**	40,000	34,000	4,000
Central	**Biaka**	54,000	23,000	19,000
	Mbuti	40,000	23,000	10,000
East	**Kikuyu**	116,000	29,000	35,000
	Somali	103,000	27,000	34,000
	Turkana	140,000	33,000	45,000
West	**Fulbe**	88,000	11,000	26,000
	Hausa	81,000	21,000	n.a.
	Songhai	91,000	47,000	36,000
	Tuareg	94,000	25,000	26,000
	Yoruba	101,000	21,000	32,000
	Kanuri	94,000	33,000	29,000
	Mandenka	79,000	9,000	21,000

Estimates of coalescent times (years) back to the most recent common ancestor of the sample. Column two and three give the times for a constant and exponentially growing population, respectively. The method is described by Weiss and von Haeseler (*in prep.*) derived from Nee *et al.* (1995). The last column gives the estimate of the expansion time of the model of Pleistocene population explosion (Rogers, 1994).

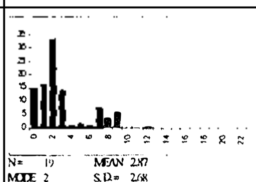
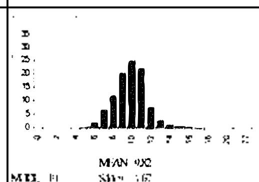
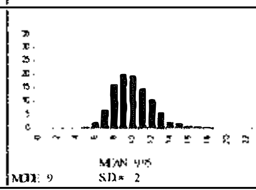
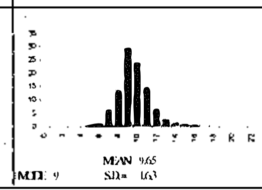
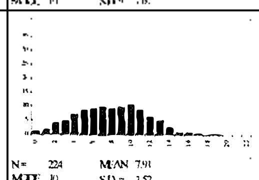
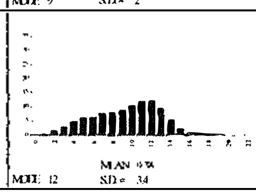
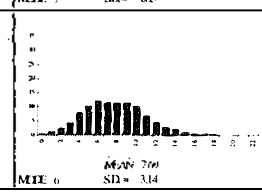
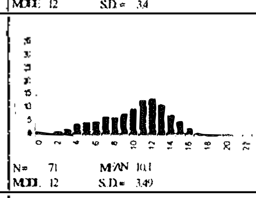
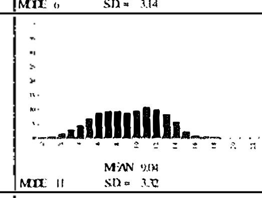
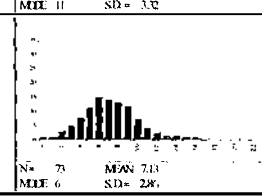
With African populations it is not appropriate to treat the ethnic group as an entity with a common history, as demonstrated later in Section 6. Rather ethnic groups are composites of different genetic elements, each with distinct histories in terms of origin and expansion events. This table is interesting, however, as the sudden expansion times in the West Africa ethnic groups, correlates roughly to the expansion events of clusters II (a and b), IV and V of West Africa. The sudden expansion times also in the East Africa groups correlate to the expansion of Clusters III and V (see Section 6). The constant population coalescence times probably approximate the more distant lineages coalescence times, for example the diverse lineages (predominantly Cluster-VI) of East Africa correlate with oldest coalescent dates. See Section 6 for reference to the Clusters mentioned above.

* It is important to note that due to the diversity within Africa, none of the African populations surveyed are representative of the population as a whole, that is, a significant proportion of the lineages within the population are not sequenced.

4.6.7 Correlation between Language, Economy, Geography and Genetics

4.6.7.1 Language

Table 15 Linguistic Pairwise Distribution.

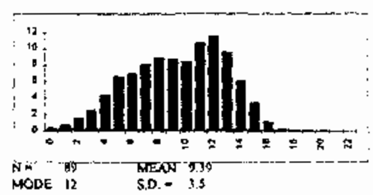
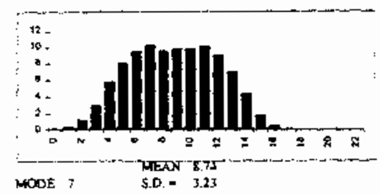
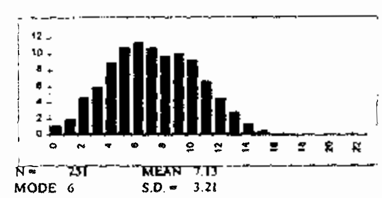
	KHOISAN	NIGER-KORDOFANIAN	NILO-SAHARAN	AFROASIATIC
KS				
NK	K_{ST} = 0.0786 P = 0.000 D_S = 4.43			
NS	K_{ST} = 0.1204 P = 0.000 D_S = 3.47	K_{ST} =0 .0164 P of K_{ST} = 0.000 D_S = 0.38		
AA	K_{ST} = 0.1977 P = 0.000 D_S = 4.65	K_{ST} =0.0079 P = 0.000 D_S = 0.17	K_{ST} = 0.0250 P = 0.000 D_S = 0.43	

The upper right triangle illustrates the pairwise distance distributions. The horizontal axis is the number of substitutions and the vertical axis the percentage of pairs. The upper figure in each cell of the lower left triangle is the K_{ST} value where $K_{ST} = 1 - \frac{K_S}{K_T}$, the middle value is the Probability that there is no difference between the populations and the lower value is the corrected distances, $D_S = \max\left\{0, D_{12} - \frac{D_1 + D_2}{2}\right\}$.

To test if pairwise distances could detect correlation with linguistic phyla, the sequences were pooled by phyla. The Biaka and Mbuti are suspected to have recently replaced their language (see p. 32). This may be the case for other populations too, therefore for the purposes of the study the current language was used to investigate relationships between the language phyla. The Nilo-Saharan group has the highest mean pairwise distance within the phylum, this is due in part to the diverse Turkana accounting for a high proportion of this phylum (67%). There is subdivision according to the Hudson test but the genetic distances(D_S) based on the mean pairwise distance (Nei 1987:p.276: formula 10.21) do not detect any obvious structure, except for the isolation of the Khoisan phylum (see Table 15). The other linguistic phyla have low genetic distances, suggesting that there is little to differentiate between them.

4.6.7.2 Geography

Table 16 Geographic Pairwise Distribution.

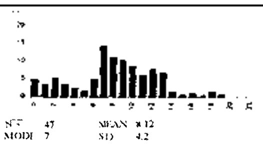
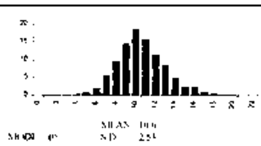
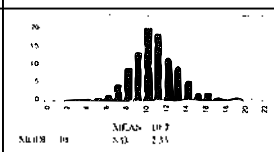
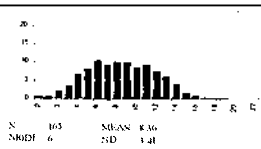
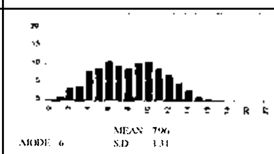
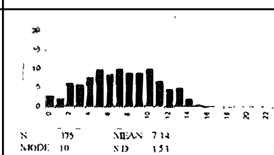
	EAST	WEST
EAST		
WEST	K_{ST} = 0.0234 P = 0.000 D_S = 0.48	

The upper right triangle illustrates the pairwise distance distributions. The horizontal axis is the number of substitutions and the vertical axis is the percentage of pairs. The lower left triangle are the results from the Hudson test (the K_{ST} value and the probability of K_{ST}) and the corrected distance $D_S = \max\left\{0, D_{12} - \frac{D_1 + D_2}{2}\right\}$.

To test if pairwise distances could detect correlation with geography, the sequences were grouped by geographic categories. 'West' is defined by individuals from Cameroun, Nigeria, Benin, Mali, Niger and Burkina Faso and 'East' by individuals from Kenya, Somalia and Ethiopia. There is a higher diversity in East Africa (mean±s.d. = 9.39±3.50), than in West (mean±s.d. = 7.13±3.21). The distributions show bimodal peaks which is consistent with two major expansion events (Rogers and Harpending, 1992). The corrected distance between the two regions is low. There is however, geographic subdivision as measured by the Hudson test.

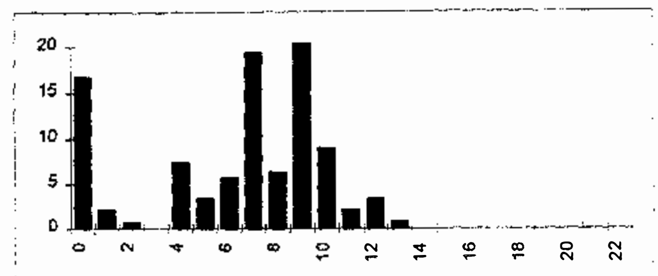
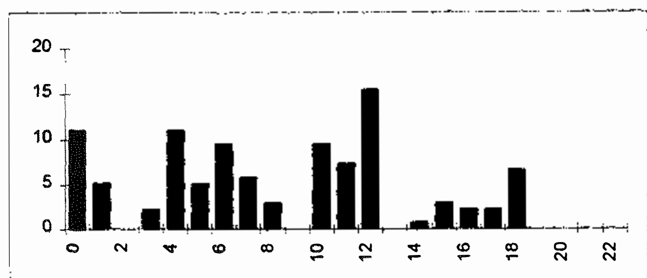
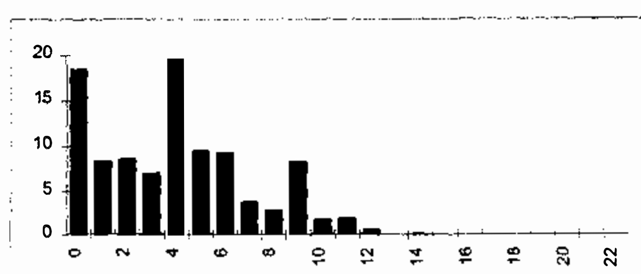
4.6.7.3 Economy

Table 17 Lifestyle pairwise distributions.

	HUNTER-GATHERERS	PASTORALISTS	AGRICULTURISTS
HUNTER-GATHERERS			
PASTORALISTS	$K_{ST} = 0.0892$ P = 0.000 $D_S = 2.36$		
AGRICULTURISTS	$K_{ST} = 0.1476$ P = 0.000 $D_S = 3.07$	$K_{ST} = 0.0127$ P = 0.000 $D_S = 0.21$	

The upper right triangle illustrates the pairwise distance distributions. The horizontal axis is the number of substitutions and the vertical axis the percentage of pairs. The lower left triangle are the results from the Hudson test (the K_{ST} value and the probability of K_{ST}) and the corrected distance $D_S = \max\left\{0, D_{12} - \frac{D_1 + D_2}{2}\right\}$. The raggedness values for the pastoralists, agriculturists and hunter-gatherers are 0.0043, 0.0058 and 0.0167 respectively where raggedness is calculated based on Harpending *et al.* (1993) defined as $r = \sum_{i=1}^{d}(x_i - x_{i-1})^2$.

Table 18 Hunter-Gatherer Pairwise Distributions

MUKRI (Asia) r = 0.1178	 N = 38 MEAN 6.41 MODE 9 S.D. = 3.6
BIAKA PYGMY (Africa) r = 0.0698	 N = 17 MEAN 8.12 MODE 12 S.D. = 5.29
SAAMI (Europe) r = 0.0466	 N = 115 MEAN 3.99 MODE 4 S.D. = 3.08

Three hunter-gatherer populations from different continents. Unlike the standard smooth unimodal curve observed in most world-wide populations, these curves are characterised by their raggedness. r represents the raggedness value described by Harpending *et al.* (1993) defined as $r = \sum_{i=1}^{d}(x_i - x_{i-1})^2$. Compare the raggedness values to food producing populations (10.9). 'Ragged curves are generated from simulations when the population size remains constant (Rogers and Harpending, 1992). Note the Mukri sample size is 38, which is lower than the 43 published by Mountain *et al.* (1995), because five sequences with ambiguous bases were removed.

To test if pairwise distances could detect correlation between the different economic modes (i.e. Pastoralism, Agricultural and Hunter-Gathering) (see Table 6), the sequences were pooled appropriately. Pastoralism is defined by a lifestyle dependant on herded animals and the groups are often nomadic or semi-nomadic, in Africa the herded animals include cattle, goats, sheep and camels. The groups which depend on a pastoral based economy include the Somali, Turkana, Fulbe, Kanuri and Tuareg. Agriculturalism is defined by a lifestyle based on growing and harvesting domesticate crops. The groups which depend on a agricultural based economy include the Yoruba, Mandenka, Songhai, Hausa and Kikuyu. Hunter-Gathering is a nomadic lifestyle based on collect plants and hunting animals. The groups which depend on a hunter-gatherer based economy include the !Kung, Biaka and Mbuti. The Hunter-Gatherer groups are genetically distinct from the other groups, in that they do not share any lineages with other populations. When pooling the three Hunter-Gatherer groups, the pairwise distribution is still ragged despite the larger sample size. This phenomenon is also observed in other populations (see Table 18): the Mukri of India (Mountain *et al.* 1995) and the Saami of Lapland (Sajantila *et al.* 1995). The Mukri are Hunter-Gatherers and the Saami were also, until they adopted reindeer herding sometime between the fourteenth to sixteenth centuries (A. Sajantila, *pers. comm.*). Here is evidence on three continents that the Hunter-Gatherer lifestyle correlates with ragged curves. The food collecting groups (!Kung, Mbuti and Biaka) all show ragged distributions which are consistent with a population that has experienced no growth (i.e. no increase in population density). The food producing groups on the other hand, show smoother curves, indicative of population growth. Presumably the food collecting groups outside Africa were part of a Palaeolithic expansion, therefore a smooth unimodal curve would be expected. Other food collecting groups such as the Nuu-Chah-Nulth (Ward *et al.* 1991) and Aboriginals (S. Easteal *pers. comm.*), demonstrate smooth unimodal curves. To explain this apparent anomaly, it may be the neighbours of the food producers who influence the population dynamics of the food collectors. Agricultural or pastoral neighbours may restrict the food collectors' territories or marginalise them to less fertile areas. Both these scenarios would reduce or at best maintain the population size.

4.6.8 Africa in the World

Compared to the other continents Africa has the highest diversity when observing the mean, the mode and the maximum pairwise distances (see Table 19). Sequences from Africa also have the most number of variable positions, with the much larger European sample just behind, Asia also has a relatively high number of variable positions for the sample size (Table 19). When calculating corrected distances the Americas had the highest value from the other continents (see Table 20). Steel *et al.* (1988) and Penny (1982) have cautioned that distance measures do not reflect all the information in the data. When observing pairwise distributions (Table 19), African diversity (mean±s.d. = 8.47±3.37) is as high as the diversity between Africa and the rest of the world (mean±s.d. = 8.31±3.18) suggesting the world's mitochondrial genepool is a subset of Africa. This contrasts with the rest of the world's mean of 5.86±2.60.

Table 19 Continental Comparison

CONTINENT	*n*	lineages	variable positions	max. pwd	mean pwd	s.d.	mode
AFRICA	389	232	126	21	8.47	3.37	10
MIDDLE EAST	41	38	68	18	6.89	2.65	6
EUROPE	612	254	119	16	4.29	2.38	4
ASIA	198	123	115	18	6.73	2.54	7
AMERICA	355	117	79	14	5.57	2.96	7

This is an overview of the HVR-I in five continents/sub-continents. **lineages** are the number of different sequences; **variable positions** are the number of variable positions from the 341 bases analysed (indel events were not counted); **max. pwd** is the maximum pairwise distance; **mean pwd** is the mean pairwise distance; **s.d.** is the standard deviation; and **mode** is the peak of the distribution. For this analysis, only 341bp (np 16030-16370)* were used, this was to maximise the sample size as many published sequences were not sequenced for the whole 360 bases or had ambiguous bases in the flanking regions.

* For overview of published studies see (Section 11.3)

Table 20 Mean pairwise distances between and within continents

	AFRICA	MIDDLE EAST	EUROPE	ASIA	AMERICA
AFRICA	**8.47**	8.78	7.81	8.60	9.01
MIDDLE EAST	*1.10*	**6.89**	5.84	7.39	7.80
EUROPE	*1.43*	*0.25*	**4.29**	6.22	6.76
ASIA	*1.00*	*0.58*	*0.71*	**6.73**	6.67
AMERICA	*1.99*	*1.57*	*1.83*	*0.52*	**5.57**

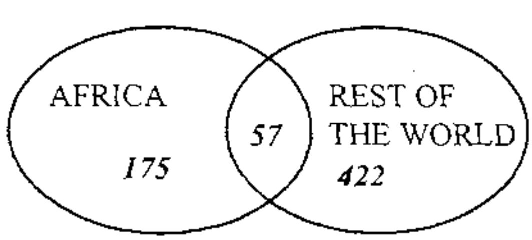
The upper right triangle represents the **inter**-continental mean pairwise distance. The cells on the diagonal marked in bold are **intra**-continental mean pairwise distance. The lower left triangle marked in italics represents the corrected distance $D_S = \max\left\{0, D_{12} - \frac{D_1 + D_2}{2}\right\}$. For this analysis only 341 bases were analysed (np 16030-16371).

Table 21 Africa versus non Africa

	n	lineages	max. pwd	mean pwd	s.d.	mode
AFRICA	389	232	21	8.47	3.37	10
BETWEEN AFRICA & NON AFRICA			23	8.31	3.18	7
NON AFRICA	1165	479	19	5.86	2.60	6
TOTAL (ALL CONTINENTS)	1554	654*	23	6.94	3.14	6

America, Asia and Europe were pooled to represent the non-African sequences. The results were unchanged when the Middle East sample was included (data not shown). **variable positions** are the number of variable positions from the 341bp analysed (indel events were not counted); **max. pwd** is the maximum pairwise distance; **mean pwd** is the mean pairwise distance; **s.d.** is the standard deviation; **mode** is the peak of the distribution.

* The number of lineages of all sequences combined can be represented by a Venn diagram.



4.6.9 Africa's diversity in comparison to the World's

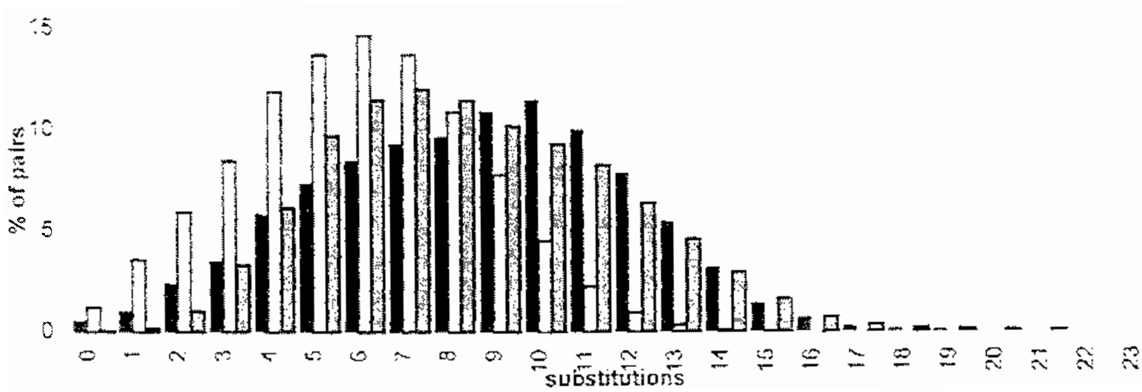


Figure 8 Pairwise Distribution of Africa versus the rest of the world
The histogram of the pairwise distribution of Africa is illustrated as black, the rest of the World is illustrated as white and between Africa and the rest of the world is illustrated as with hatched markings. Africa's diversity is higher than the other continents (Table 20), and is maintained even when the other continents are pooled (Table 21).

4.7 Discussion

The hypervariable region has many advantages for human population studies due to its high substitution rate. There is, however, rate inconsistency along the region (Wakeley, 1993; Hasegawa *et al.* 1993). This was supported in the present study with an independent data set of African sequences. Figure 3 illustrates that some sites are indeed very fast while others tend to be conserved. There is also a bias of transitions to transversion, as illustrated by Figure 4. However, what appears to be transversions in the C-run region (np 16180-16193) may in fact be slippage events (see Section 4.3.3). Variable rates and transition bias should be taken into account especially with phylogenetic reconstruction methods.

Using pairwise distance distributions it is possible to get an overview of the data, for example, the higher mtDNA diversity in Africa as compared to other continents. Analysis in the later sections, however, demonstrates that it is more informative to have a sequence based approach which reveal the different genetic components (defined in this thesis as clusters). Pairwise distances used in parallel with other methods can provide an estimate of diversity and expansion dates of the clusters (see Section 6).

Pult *et al.* (1994) reported that the African populations are genetically isolated from each other in comparison to European populations. The populations used in the Pult *et al.* (1994) analysis were originally from Vigilant *et al.* (1989; 1991), the populations (especially the !Kung and Pygmy groups) were probably chosen for this study because they are phenotypically distinct from other African groups. Results from additional nine African populations reported here, show evidence of sequences shared across ethnic boundaries (see Section 6) and large geographical distances, for example, a sequence is shared between Kikuyu, Mandenka, Somali, Tuareg, and Fulbe individuals (Cluster-V, **Figure 12**), the maximum distance between these individuals is almost 6,000km.

Despite the evidence of geneflow over large areas, the Hudson test and Nei's distance statistic does detect subdivision between most Africa populations. The West African populations show no subdivision between them with the Hudson and Nei's distance test. The East African populations of the Turkana and Kikuyu also show no subdivision between them, suggesting a degree of regional differentiation.

Subdivision is detected with the Hudson test according to the categories of language, economic mode and geography. Further investigations, particularly simulation studies, need to be conducted into what the Hudson test measures and if it is an appropriate test for population studies such as this.

All categories (language, economy and geography) probably contribute to some degree to the genetic structure of each ethnic group and will be discussed in more depth in the Section 6. Section 6 describes various elements (cluster) in ethnic groups which have different histories, some reflecting the geographical region, some may reflect cultural diffusion which correlates to language and/ or to economy.

Observing the raggedness values of the pairwise distance distributions suggests the Hunter-Gatherers (Food Collectors) have more ragged curves, suggesting that their populations may have been constant. These results should not be overinterpreted due to sample size limitations. This is consistent with some food collecting populations which live outside Africa, such as the Mukri and Saami, however other food collecting populations do not show this. It is not the food collecting lifestyle *per se*, but the food producing neighbours which may influence the population dynamics of the food collectors. Even in the last millennium the population territories of food collectors are known to have been severely reduced, for example the !Kung in Southern Africa and the Saami in Scandinavia.

This section described the sequences using standard methods. This is presented to orientate the reader and provide an overview. The following sections will analyse the data more extensively.

5. THE TURKANA PEOPLE OF EAST AFRICA HAVE THE HIGHEST MTDNA GENETIC DIVERSITY

5.1 Abstract

The mitochondrial DNA hypervariable region I was sequenced from 241 individuals from nine African ethnic groups. When comparing these and other published sequences, the Turkana people of Kenya have the highest diversity of any ethnic group reported to date. Indeed the 37 Turkana individuals have a higher diversity than the Vigilant *et al.* (1991) world wide sample of 189 individuals. Diversity can be indicative of the centre origin of species. An alternative explanation of lower mutation rate outside Africa is eliminated because the same rate of mutation was found from a nuclear insert copy (Zischler *et al.* 1995) to 232 African lineages and 475 non-African lineages. The data is most compatible with larger effective population size due to age rather than population size. This is compatible with the non-African mitochondrial DNA being a subset of African DNA.

5.2 Introduction

The origin of modern humans (*Homo sapiens sapiens*) has caused debate for several centuries. Currently the two main competing hypotheses concerning the origin of modern humans accept that *Homo erectus* arose in Africa and spread out of Africa more than one million years ago. However, the Out of Africa hypothesis maintains that *Homo sapiens* evolved within Africa, and then expanded to the rest of the world within the last 100,000 years ago, replacing *Homo erectus* (Cann *et al.* 1987; Vigilant *et al.* 1989; Bowcock *et al.* 1991a; Horai *et al.* 1993; Jorde *et al.* 1994; Horai *et al.* 1995) it predicts the highest genetic diversity would be found in Africa. In contrast the Multiregional hypothesis maintains that *Homo erectus* evolved into *Homo sapiens sapiens* in several regions particularly Africa, Europe, Asia and Australia (Wolpoff, 1989, 1992; Thorne and Wolpoff, 1992; and consequently full diversity would be found in most regions of the world.

Most genetic studies favour the Out of Africa hypothesis (Cann *et al.* 1987; Vigilant *et al.* 1991; Horai *et al.* 1993; Jorde *et al.* 1994; Horai *et al.* 1995; Penny *et al.* 1995), but the palaeontological record, particularly in Asia, is difficult to interpret. Palaeontologists have different ways of interpreting the fossil record and some view it as supporting Multiregional hypothesis (Wolpoff, 1989, 1992; Thorne and Wolpoff 1992; Frayer and Wolpoff, 1993; Frayer *et al.* 1994) but others arrived at an Out of Africa scenario independently from the geneticists (Bräuer, 1984a; 1984b; 1989; 1990; Stringer, 1992). These regions of high genetic diversity are interesting in that they can indicate the origin of a species; however, alternative explanations could be differential mutation rates (either an acceleration within Africa or deceleration outside Africa) or some other more complex model involving population dynamics. These alternatives can be tested against the simple model of centre of diversity being the region of origin for a species.

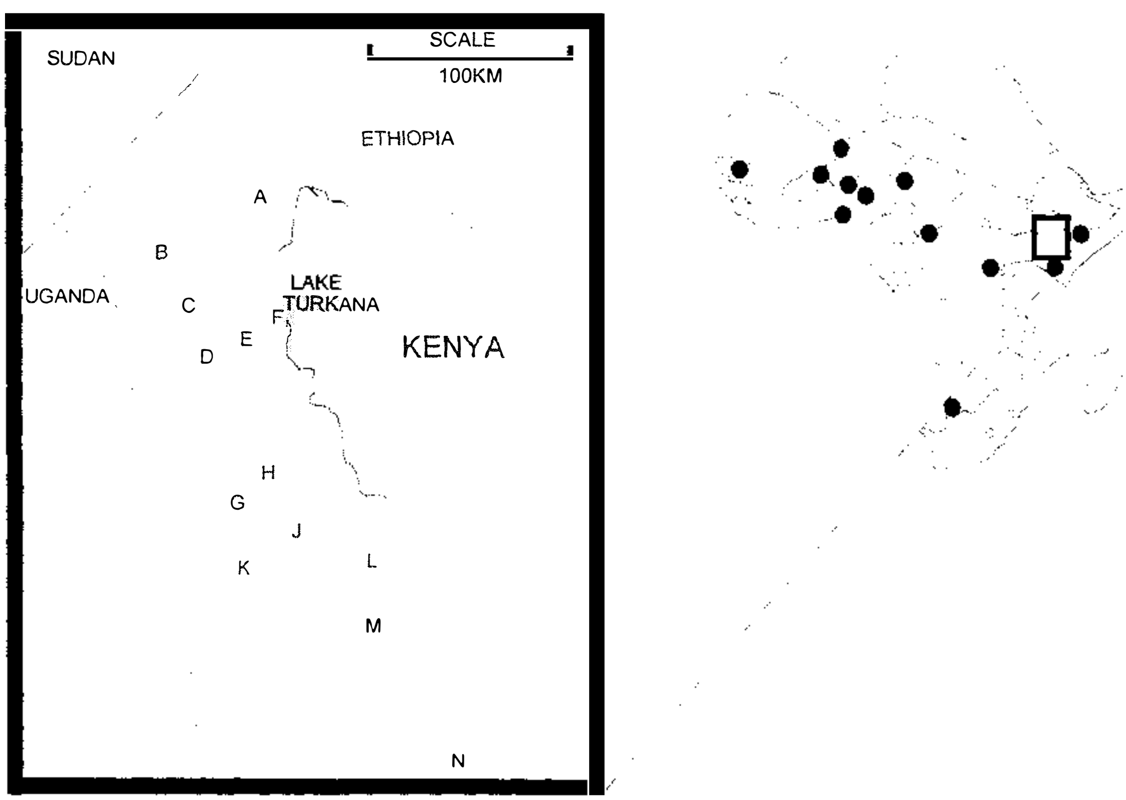
The hypervariable region I of the control region, np 16030-16370* , was sequenced from 241 individuals from nine ethnic groups (as presented in this thesis). The groups are the Kanuri, Songhai, Hausa, Tuareg, Yoruba and Fulbe from West Africa, and the Somali, Kikuyu and Turkana from East Africa (see Map 2). These groups are linguistically and culturally diverse. In addition, the world-wide dataset of 189 individuals with 135 lineages (Vigilant *et al.* 1991), which included Africans, was used for comparison.

A useful marker in African populations is the presence of a thymine at position 3594 (Denaro *et al.* 1981; Chen *et al.* 1995) is detected by a restriction enzyme *Hpa*I which recognises the sequence GTT/AAC, *Hpa*I positive has been found to be specific to Africa (see p. 78) with frequencies as high as one hundred per cent in the !Kung and it does occurs in low frequencies in the Middle East and Sicily (Bonné-Tamir *et al.* 1986, Semino *et al.* 1989), probably due to African admixture. *Hpa*I positive has not yet been reported elsewhere in the world.

* numbering according to Anderson *et al.* 1981

	Place of Birth	frequency of individuals
A	Lokitaung	2
B	Kakuma	2
C	Loima	1
D	Lorugumi	7
E	Lodwar	4
F	Fergusson's Gulf	4
G	Kapitir	1
H	Lokichar	1
J	Lokori	3
K	Kerio	1
L	Baragoi	6
M	Merti	1
N	Isiolo	4

Table 22 Key to Map 6
Place of Birth refers to the Maternal Grandmother's place of birth (or the individual's place of birth when the information is not available).



Map 6 Place of Birth of Maternal Grandmother's of the 37 Turkana individuals
The Turkana are a pastoral group (dependant on herding cattle, goats and camels) living on the Western side of Lake Turkana. Their population has been estimated at 250,000 (Grimes, 1988:p.248). Their language is of the Nilotic group of the Nilo-Saharan Language Phylum (Ruhlen, 1987:p.318), which is closely related to Toposa in Sudan (Grimes, 1988:p.248). The Turkana are presumed to have moved into Northern Kenya at the end of the 17th century, and may have absorbed the people who were living there including the Ngikor and Rendille peoples (Sobania, 1979). The Map of Africa illustrates the locations of other African groups included in the analysis, including the Mandenka from Graven *et al.* (1995) and Pygmy (Biaka and Mbuti), !Kung and Yoruba from Vigilant *et al.* (1991).

5.3 Results

Table 23 Diversity of the Turkana compared to Africa and the World.

population	diversity	*n*
Turkana	3.12×10^{-2}	37
East Africa	2.75×10^{-2}	89
Africa (all)	2.48×10^{-2}	389
West Africa	2.09×10^{-2}	251
Asia	1.97×10^{-2}	198
America	1.63×10^{-2}	355
Europe	1.26×10^{-2}	612

In addition to the 241 African sequences, other published sequences were used (Vigilant *et al.* 1991; Ward *et al.* 1991, 1993; Piercy *et al.* 1993; Shields *et al.* 1993; Torroni *et al.* 1993a, 1993b, 1994; Pult *et al.* 1994; Santos *et al.* 1994; Batista *et al.* 1995; Bertranpetit *et al.* 1995; Kolman *et al.* 1995; Graven *et al.* 1995; Mountain *et al.* 1995; Sajantila *et al.* 1995). Genetic diversity (substitutions per site) is calculated from the mean pairwise distance divided by the sequence length.

The simplest measure of diversity is the average distance within groups. Table 23 illustrates that the highest diversity observed in the HVR-I region is found within Africa and that the diversity is especially high within the Turkana of East Africa. The simplistic model would predict on the basis of diversity that the centre of origin of modern humans would be Africa and more specifically East Africa. More complex models need to be addressed. It is possible that the mutation rate was higher in Africa or slower outside Africa. Thus we compared all African and non-African lineages a fragment of mitochondrial Hypervariable I region inserted in the nuclear genome (Zischler *et al.* 1995) and to a second outgroup of 64 chimpanzees *Pan troglodytes* (Morin *et al.* 1994). The mean distance for the 232 African and 475 non-African lineages from the nuclear insert was 21.29±1.50 and 21.32±1.81 respectively. Similarly the mean distance from the 64 chimpanzees was 57.51±1.65 and 56.69±1.27 respectively (see Table 24). These results indicate no acceleration in the African lineages or deceleration in the non African lineages and therefore exclude differential rates of mitochondrial DNA substitution as an explanation of the high diversity observed in Africa.

Table 24 Mean distance from Outgroups

	lineages *n*	nuclear copy of HVR-I as outgroup	64 chimpanzees as outgroup
Africa	232	21.29	57.51
Non-Africa	475	21.32	56.69

Mean pairwise distance of the datasets to the outgroups. The numbers are not significantly different, indicating that there has been no acceleration on the African lineages, or deceleration on the non-African lineages. Thus the higher diversity in Africa is not a result of a higher substitution rate.

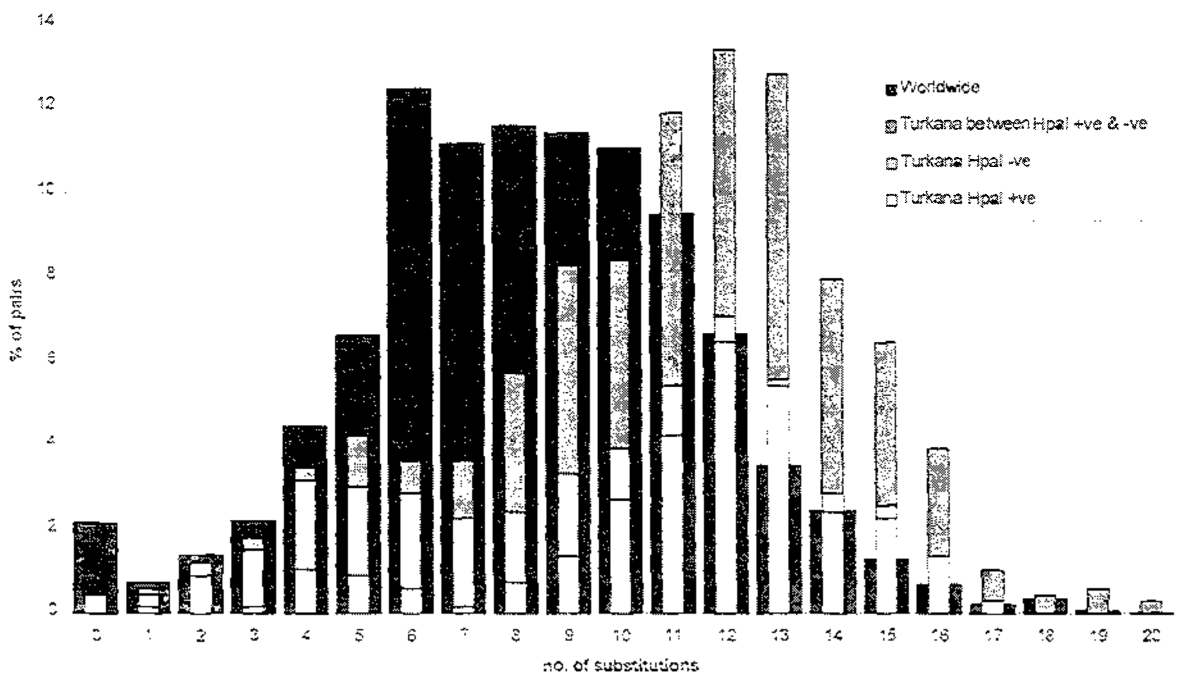


Figure 9 Pairwise Distance Distribution of the Turkana versus the World
Mean pairwise distances illustrate the average distance of each sequence to every other sequence within a population. The diversity within the Turkana is higher (mean = 10.650±3.657) as opposed to the Vigilant *et al.* (1991) world-wide sample of 189 (mean = 8.262±3.269). The difference is highly significant statistically by randomly sampling 37 sequences from the full Vigilant *et al.* (1991) set. Subsampling from the 135 lineages (Vigilant *et al.* 1991) show only 12 (with replacement) and 86 (without replacement) subsamples from 100 000 subsamples have a higher mean than the Turkana group of 37. Subsampling from the 189 individuals show 0 (with replacement) and 2 (without replacement) subsamples from 100 000 subsamples have a higher mean than the Turkana. To place this in context with the 37 Turkana samples, 3 Turkana individuals have identical sequences.

If the Turkana are divided by the *Hpa*I restriction site at np 3592 (see p. 78), quite different pairwise distributions can be observed. *Hpa*I negative individuals are relatively closely related (mean = 7.108±3.089). The *Hpa*I positive group are diverse (mean = 11.071±3.564) and equivalent in diversity to the distribution of pairwise distances between the *Hpa*I positive and negative group (mean = 11.652±3.103), both which account for the highest peak. The *Hpa*I positive sequences are found almost exclusively within Africa, the negative sequences are found world-wide including within Africa. This figure demonstrates that the Turkana have a higher diversity than the rest of the world. The *Hpa*I positive haplotypes are responsible for high diversity both from within the *Hpa*I positive group and their distance from the *Hpa*I negative individuals. This is illustrated in section 10.9 where the Clusters III, V and VI have the *Hpa*I site. Generally individual sequences with the *Hpa*I site are more diverse.

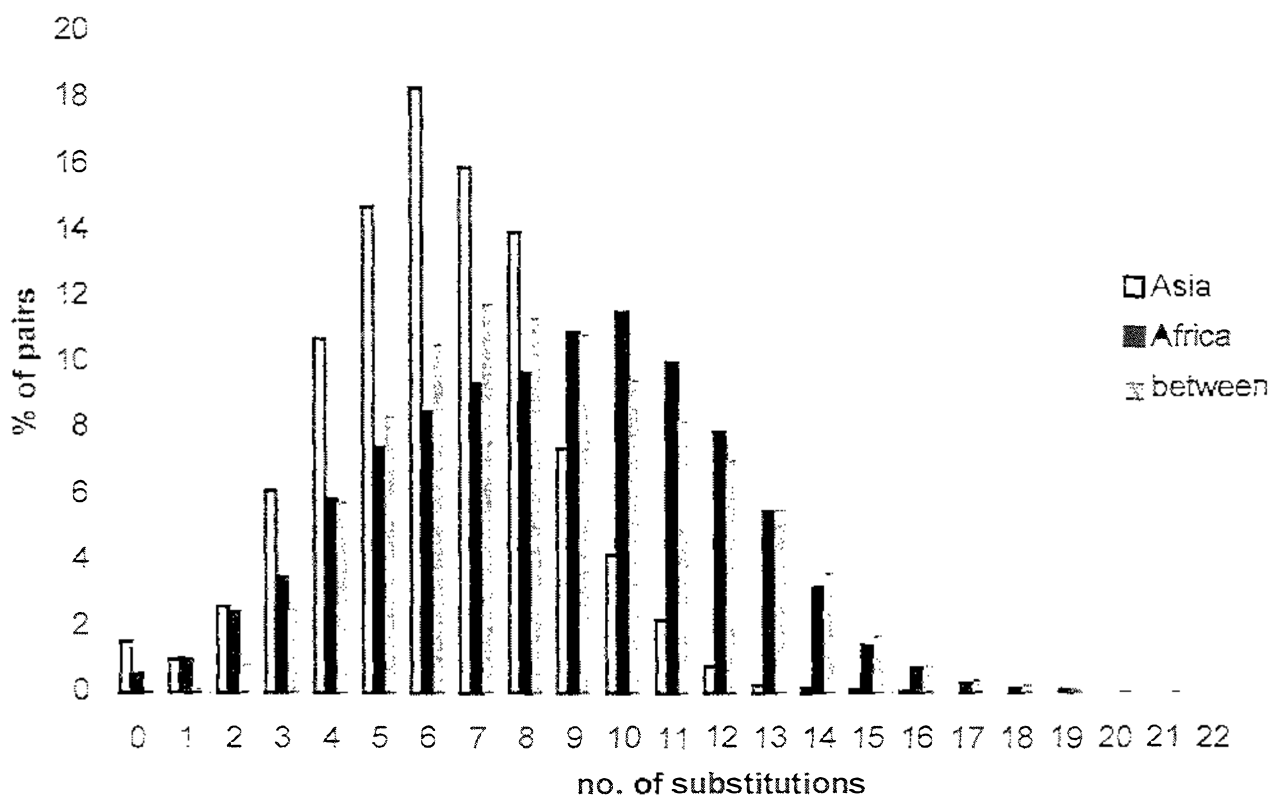


Figure 10 Pairwise Distance Distribution within Asia and Africa
Asian sequences n = 198 (Vigilant *et al.*'s 1991; Shields *et al.* 1993; Torroni *et al.* 1993a, 1993b, 1994; Mountain *et al.* 1995). African sequences n = 389 (Vigilant *et al.* 1991; Graven *et al.* 1995). Between indicates the pairwise distance between African and Asian sequences. Individuals containing ambiguous bases were excluded from the dataset. Asia and Africa are selected because they have the highest diversity of the continents (see Table 23) and the main alternative to Africa origin is an Asian centre. The mean pairwise distance of Africa was 8.466±3.368; Asia 6.734±2.542 and between Asia and Africa 8.604±3.215 Not only does Africa have the higher diversity, but the intercontinental mean pairwise distance being almost the same as the distance within Africa is consistent with the hypothesis that the Asian mitochondrial genepool is a subset of the African. This interpretation is limited on the representativeness of the mtDNA data set. The Asian sequence set is limited in that a high proportion of sequences are from the Circumarctic region, which show low diversity as compared with Amerind sequences. East India may also not be representative of the Asia genepool as a whole.

5.4 Discussion

The diversity observed in the Turkana is a reflection of the high diversity in East Africa. The diversity of the Turkana does not indicate that the Turkana as an ethnic entity have an extraordinary ancient history, rather ethnic groups are composites of different elements, which may differ in their history. The idea of ethnic groups as stable units through time has been challenged in the literature (Barth, 1969; Herren, 1991) and in the present study 5% of volunteers are from interethnic marriages. The high diversity is accounted for by the *Hpa*I positive elements which are as diverse from each other as to the *Hpa*I negative elements. The two other East African groups (the Somali and the Kikuyu) also have high diversity in relationship to other African groups.

The fact that the non-African sequences appear to be a subset of the African sequences (as illustrated in Figure 10) argues against the diversity being explained by higher effective population size although this does require further research. These results are consistent with a general Out of Africa model; however this may be modified with data from Australia, or a significantly different calibration of the mtDNA molecular clock. It is necessary to compare these conclusions with a similar study of a nuclear locus, or preferably loci to determine if the Turkana also show high diversity at the nuclear level.

The genetic record suggests that Africa is a 'Hominid epicentre', from where waves of expansion occurred out of Africa. With higher diversity there may have been more opportunity to generate new hominid types: first *Homo erectus* then two distinctive *Homo sapiens* (early and modern). These hominids have successively migrated out of Africa replacing the earlier hominids, who have left, as yet, no detectable trace in the genetic record (see Map 2). There is a possibility which requires further investigation, that there were in fact two modern human *Homo sapiens sapiens* expansions out of Africa. The first which may have remnants in Australia as suggested by sequence number 49 from Vigilant *et al.*'s (1991) dataset, and possibly other parts of Asia.

6. Mitochondrial Footprints of Human Expansions within Africa

6.1 Abstract

Mitochondrial DNA studies support an African origin for modern humans, but the expansion events within Africa have not previously been investigated. 237 HVR-I sequences from a diverse selection of nine African ethnic groups were studied in addition to previously published sequences. The African sequences fell into five major clusters, whose star-like structures affirm demographic expansions. A number of sequences (14%) fell outside these clusters, these were divergent from each other and accounted for the high mtDNA diversity seen within Africa. The oldest of these expansions includes the major founding sequence of Eurasian mtDNA. Other younger expansions remained within Africa. Coalescence dates were estimated for each cluster and the cause of the expansions is discussed.

6.2 Introduction

Most mtDNA studies support an Out of Africa scenario for the evolution of modern humans (Cann *et al.* 1987; Cavalli-Sforza *et al.* 1988; Vigilant *et al.* 1991; Chen *et al.* 1990; Rapacz *et al.* 1991; Bowcock *et al.* 1991a; Nei and Roychoudhury, 1993; Di Rienzo *et al.* 1994; Bowcock *et al.* 1994; Deka *et al.* 1995; Goldstein *et al.* 1995; Penny *et al.* 1995). These studies demonstrate the highest diversity in the world is found within Africa, and predict human's last common ancestor lived between 100 kya and 300 kya. With the advent of large datasets and intraspecific phylogenetic methods, it is now becoming possible to study the geographic origin of mtDNA diversity in Africa, as well as the expansion events within Africa that led to the colonisation of Eurasia.

Different nuclear and mitochondrial loci have been studied in African populations and compared with other African populations and with populations in other parts of the world. MtDNA has been studied in African populations with a six restriction enzyme system (Cann *et al.* 1987, Soodyall 1993), and control region sequences (Vigilant *et al.* 1991). Recently Graven *et al.* (1995) generated information from both restriction enzyme data and control region sequences from 119 Mandenka individuals; and Chen *et al.* (1995) employed fourteen restriction enzyme system on populations from Central and West Africa.

6.3 Materials and Methods

6.3.1 Subjects

African mtDNA sequences from Hypervariable Region I and II (HVR I and II) have been published from the Yoruba, Mbuti (East Pygmy), Biaka (West Pygmy), Hadza, Herero, !Kung and African-Americans (Vigilant *et al.* 1991), and from Mandenka (Graven *et al.* 1995). The Hypervariable Region I was analysed from these published sequences[1] together with eight additional African populations (Turkana, Kikuyu, Somali, Hausa, Tuareg, Kanuri, Songhai and Fulbe) and additional sequences from the Yoruba population (see Table 6). Networks were constructed from information from np 16090-16365[2], in order to include as many sequences as possible without losing too much phylogenetic information. The *Hpa*I Morph-3 cuts at np 3592 when np 3594 has a thymine. *Hpa*I at this position was determined in all individuals and added to the HVR I data as an additional character. *Hpa*I Morph-3 has been shown to be specific to Africa, and informative within Africa (Denaro *et al.* 1981; Chen *et al.* 1995). *Hpa*I Morph-3 has a low frequency in the Middle East and Sicily (Bonné-Tamir *et al.* 1986, Semino *et al.* 1989), which is probably due to African admixture, and has been reported elsewhere in the world. The *Ava*II cuts at np 16390 when np 16390 has a guanine. *Ava*II Morph 3 has an adenine at np 16390 which defines the 2-2 RFLP Haplotype group and Chen *et al.*'s (1995) L_2 group. This position was determined in approximately half of the individuals see Section 11.8.

---

[1] All sequences with ambiguities were excluded from the analysis.

[2] All nucleotide positions (np) are numbered according to Anderson *et al.* (1981).

Table 25 Populations Sampled

Population	Code	Source	*n*	Language Phylum
Biaka	CB	2	17	NK
Mbuti	CM	2	13	NS
Kikuyu	EK	1	24	NK
Somali	ES	1	27	AA
Turkana	ET	1	37	NS
!Kung	SK	2	19	KS
Fulbe	WF	1	60	NK
Hausa	WH	1	20	AA
Kanuri	WK	1	14	NS
Mandenka	WM	3	110	NK
Songhai	WS	1	10	NS
Tuareg	WT	1	23	AA
Yoruba	WY	1+2	21+12	NK
			$\sum = 407$	

'Code' designates the population. The first letter represents geographical location: ***C*** Central, ***E*** Eastern, ***S*** Southern, or ***W*** Western Africa. The second letter represents the population. The four language phyla present in Africa are designated: ***AA*** Afroasiatic, ***KH*** Khoisan, ***NK*** Niger-Kordofanian and ***NS*** Nilo-Saharan (Ruhlen, 1991). 'Source' of the samples are indicated by numbers: ***1*** This thesis, ***2*** Vigilant *et al.* (1991) and ***3*** Graven *et al.* (1995). For geographical location of the ethnic groups see **Map 2**. Note the sample sizes are different from **Table 11** due to a different region being investigated.

6.3.2 Median Networks application to intraspecific studies

Network Methods, are suited to *intra*-species mtDNA studies, in contrast to other phylogenetic methods which were developed to represent *inter*-species relationships. Intraspecific mtDNA data has both ancestral and derived sequences still present in the population. The reduced median network method employed here (Bandelt *et al.* 1995), resolves character conflicts by frequency guided compatibility arguments. Unresolved character conflicts (primarily caused by np. 16189) are maintained in the form of reticulations, so that the resulting Interpopulation Network (Figure 12) displays alternative phylogenies. It is this ability to retain many alternatives simultaneously that makes the networks more useful and less arbitrary, than forcing the data into just one tree.

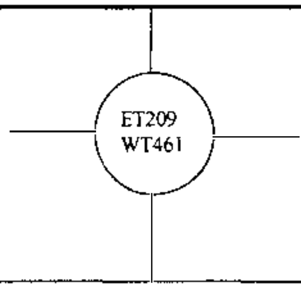
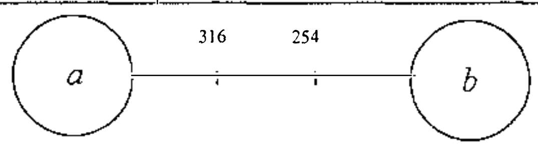
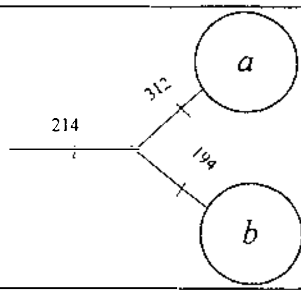
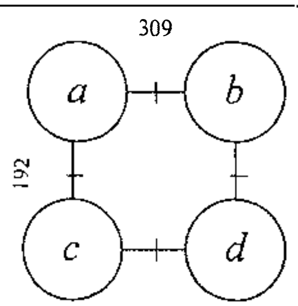
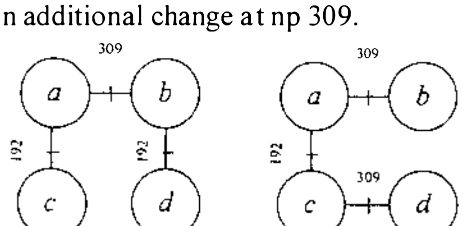
	Sequences from individuals ET209 and WT461 are identical and therefore share the same node.
	Two individuals *a* and *b* are separated by two nucleotide changes at position 316 and 254, the order of these events can not be determined without additional information, nor can the direction of change.
	When *a* and *b* share a substitution at position 214, the most parsimonious pathway is drawn representing the 214 as a single change.
	The parallel edges have the same nucleotide changes. This diagram represents four alternative pathways, two requiring an additional change at np 192 and two requiring an additional change at np 309.  or etc.

Figure 11 Key to Networks

A guide to interpreting the network diagrams.

Within the network it is useful to identify distinct groups of closely related sequences, or clusters. Clusters from recent expansions have central lineages with high frequency and numerous neighbouring lineages. These clusters are compared to established RFLP haplogroups (Graven *et al.* 1995; Soodyall, 1993), and to results from a high resolution restriction enzyme method (Chen *et al.* 1995). This method has no relationship to cluster analysis based on distance measures.

6.3.3 Network Analysis

As we are interested in the origins and causes of human expansions, we first defined a simple but effective criterion to distinguish between sequences that may reflect recent expansion events. Any major demographic expansion event would be expected to lead to a concomitant geographic expansion. Therefore we identified all sequences that were shared between at least two different populations, submitted these matches to a network analysis (see Figure 12) to distinguish derived from ancestral character states, and then identified in the remaining dataset sequences sharing the sequence motif of one these matches. 86% of the sequences fell into the five major clusters. This network was named the Interpopulation network (see Figure 12). Some reticulations are resolved by employing a reduction strategy based on compatibility and frequency arguments (Bandelt *et al.* 1990). Position 182 (np 16182) and 183 (np 16183) were excluded from the network, due to their high rate of substitution or slippage in the C-run region which is noted for being hypervariable (Horai *et al.* 1995; Bendall and Sykes, 1995). The resulting network (Figure 12) is relatively free of ambiguities. The reticulations can be resolved into a binary tree in several ways, and there is no need to distinguish between these alternatives. The reticulations in the Interpopulation network are mainly associated with np 189 (16189), which is also within the C-run. Position 189 has one of the highest substitution rates in Hypervariable Region I (Hasegawa *et al.* 1993; Wakeley, 1993).

Clusters are defined as a set of closely related lineages, radiating from a central lineage. The assignment of clusters is to some extent arbitrary and were chosen based on two types of situations. (1) Natural clusters, based on isolation from the other sequences, for example, Clusters IV and V. (2) Congruence with RFLP data. Data is available with both HVR-I sequences and RFLP types (Soodyall, 1993; Graven *et al.* 1995), and information from the unpublished sequences includes the informative *Hpa*I and *Ava*II sites. The assignments of Cluster-I, II and III are congruent with RFLP haplogroups 1-2, 7-2 and 2-2 respectively.

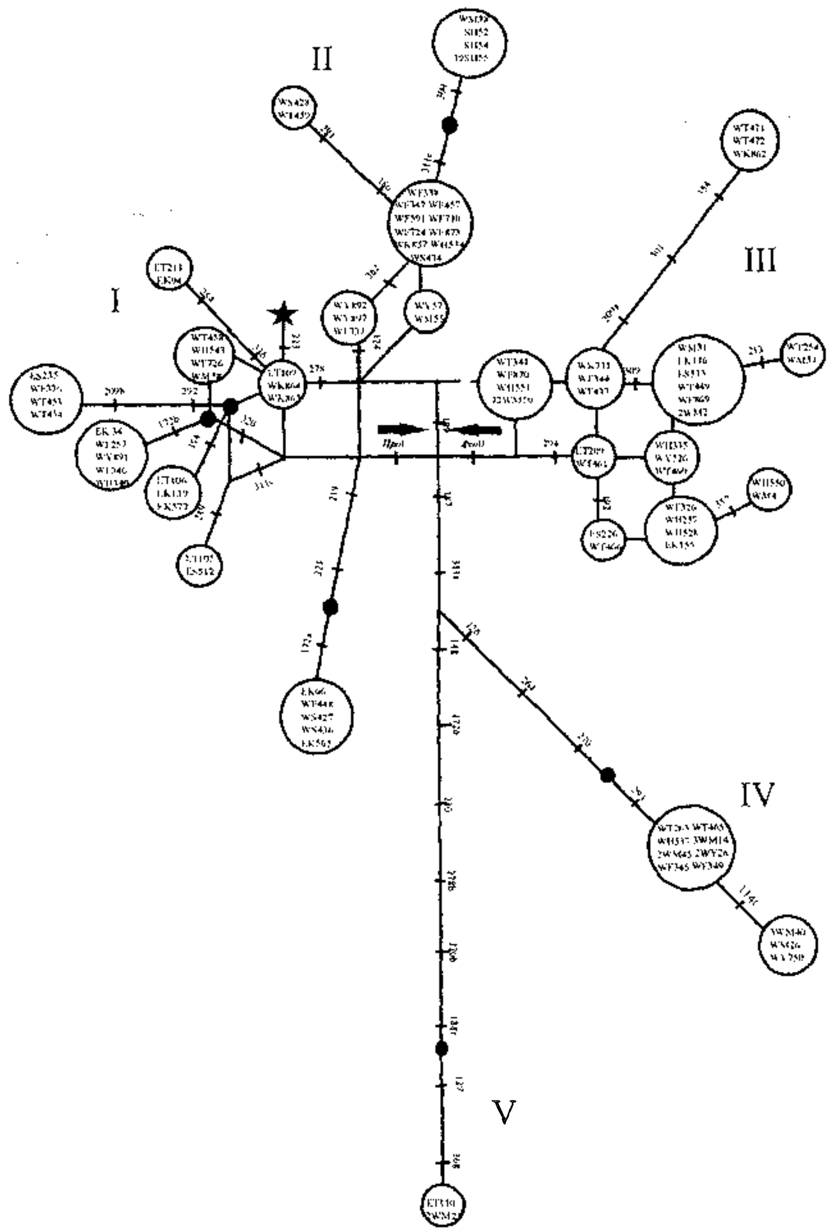


Figure 12 Interpopulation Network within Africa
This is a median network, with the different possible pathways illustrated. Shared lineages between populations are illustrated. Solid nodes indicate that the lineage is existent within Africa. The shared lineages are classified into five expansion clusters (I to V) the rationale behind these assignments are discussed in the text. Labels along the edges correspond to the numbering of Anderson *et al.* 1981). Gain of a restriction enzyme site is indicated with →. Two Tuareg individuals are at the node marked ★ which is the Cambridge reference sequence (Anderson *et al.* 1981).

6.4 Results

6.4.1 Identification of Expansion events

The Interpopulation Network (see Figure 12) could be divided into several clusters (named with Roman numerals), based on sequence similarity. These clusters are identified with a high frequency central node with many radiating edges. Interpopulation lineages are part of star-like expansions. An example of a cluster is the twin Cluster-II are illustrated in detail in Figure 13. The twin cluster is named due to two foci (central lineages with many radiating lineages).

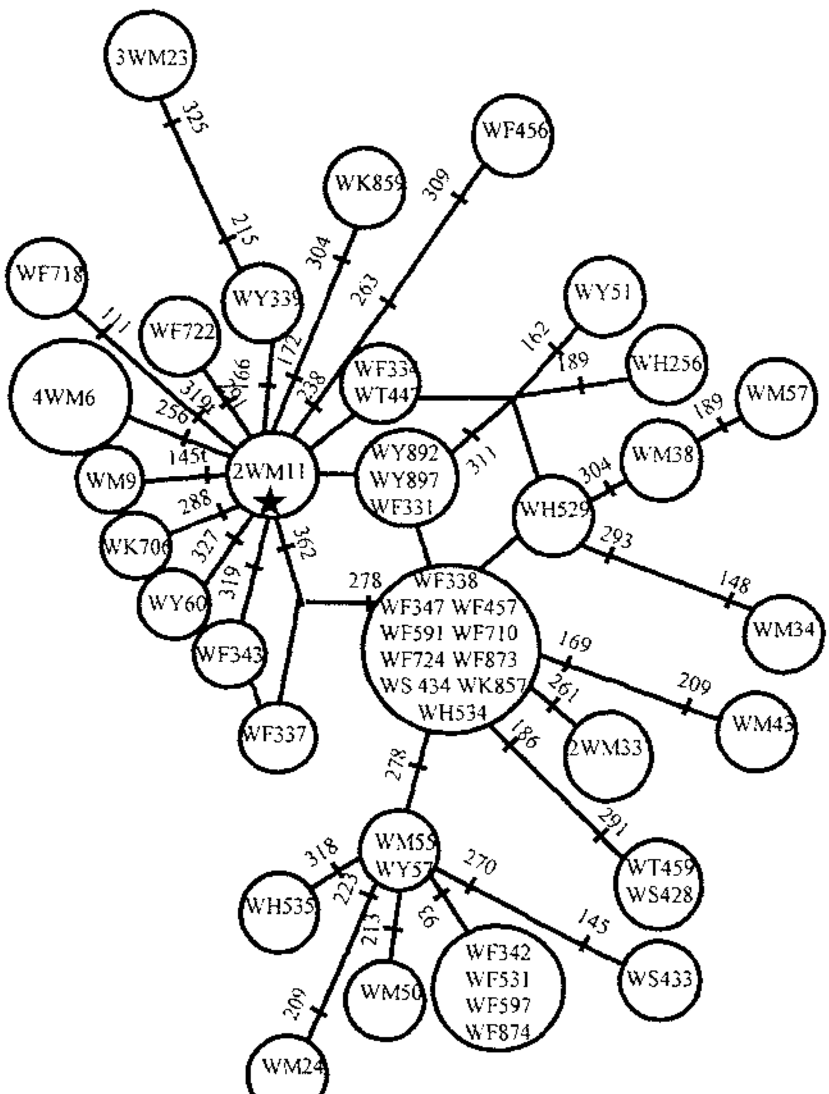


Figure 13 Network of the West African twin Cluster-II

The sequence at the node indicated with a ★ is 124, 223. The expansion dates for the two clusters based around the central nodes 124, 223 and 124, 223, 278, 362, are both dated at approximately 29,000 years ago.

14% of lineages do not fit naturally into any cluster, are not shared between populations and do not form any star-like clusters, are not unimodal in pairwise comparisons and are therefore termed 'Isolated lineages'. Figure 14 illustrates the relationships of the isolated lineages to the five expansion clusters (illustrated in Figure 12).

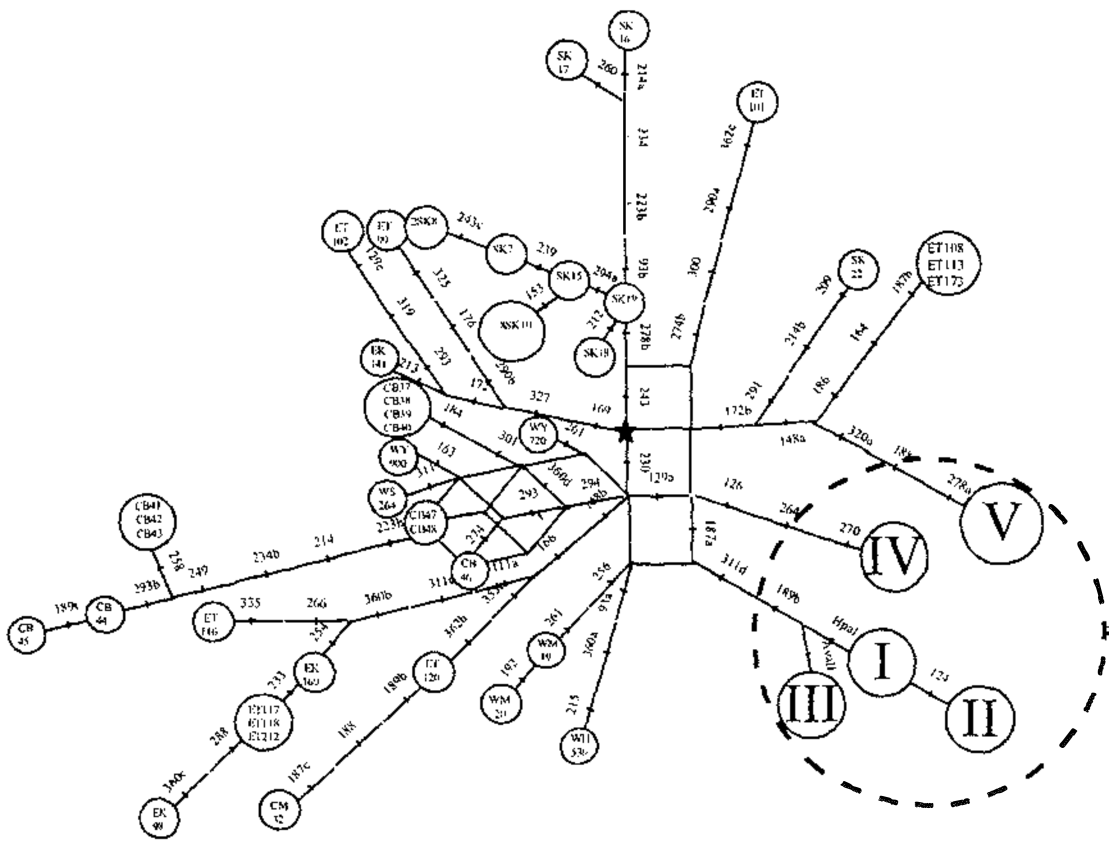


Figure 14 The relationship between the Isolated African lineages and the five expansion clusters.
The root for the mitochondrial insert in the nuclear genome (Zischler *et al.* 1995) is marked with a ★ 129, 187, 189, 223, 230, 278, 311. The nuclear insert has an additional number of substitutions from this point, including 187, 189, 218, 223, 249, 259t*, 263, 264, 274, 284, 288, 290, 293t, 301, 355,356. The clusters within the dotted ellipse are the expansion clusters illustrated in Figure 12.

6.4.2 Rooting of mtDNA Network

Human mtDNA coalescent outgroup rooting of the network using the recently published nuclear fragment of mitochondrial DNA, which is believed to have integrated chromosomally before the most recent common ancestor of human mtDNA existed (Zischler et al 1995) supports that the node marked by a star in Figure 13 is the sequence of the human mitochondrial coalescent. This is supported by the fact that this is the only node that population-specific networks of isolated lineages have in common. The age estimate in Figure 16 of 100000 years for the coalescent is a minimum estimate since some of the parallel mutations, indicated by reticulations (Figure 12), cannot be resolved on the basis of the present data. East Africa appears to represent the origin of the human mtDNA coalescent: 66% of all East African isolated lineages coalesce on the coalescent, compared to 10% for the !Kung, 0% for East and West Pygmies, and 0% for West Africans.

The three largest African clusters (comprising 70% of the samples) are closely related relative to the complete African phylogeny (Figure 13) and coalesce on a peripheral sequence characterised by transitions at np16223, np16278 and a *Hpa*I restriction site gain at np3592 relative to the Cambridge reference sequence (Anderson *et al.* 1981). The coalescence age of this sequence is about 60000 years, making this the oldest detectable expansion event in Africa (Figure 16). The expansion of this sequence affected all sampled African populations with the exception of the !Kung, and created a clear frequency gradient from East to West (Figure 15), suggesting an East African origin for this expansion. A single sequence (np16223, loss of HpaI np3592) of Cluster-I within this expansion is ancestral to virtually all Eurasian sequences (Excoffier *et al.* 1991). Taken together, these findings suggest that one or a small set of closely related lineages (which would be indistinguishable in today's clusters) expanded in East Africa and spread into both Africa and Eurasia approximately 60000 years ago. This scenario is strongly supported by results from nuclear DNA variation (Tishkoff *et al.* 1996).

---

* t represents a transversion event.

Table 26 Number of individuals in each cluster

	SK	CM	CB	WM	WS	WT	WY	WH	WF	WK	ET	EK	ES	Africa
I	0	0	0	6	2	8	10	6	20	4	15	15	19	**105**
II	0	0	0	19	3	2	6	4	18	3	0	0	0	**55**
III	0	6	0	59	2	10	9	8	11	6	4	3	7	**125**
IV	0	0	0	22	1	2	1	1	11	1	0	0	0	**39**
V	0	6	4	2	1	1	6	0	0	0	7	3	1	**31**
VI	19	1	13	2	1	0	1	1	0	0	11	3	0	**52**
Total	**19**	**13**	**17**	**110**	**10**	**23**	**33**	**20**	**60**	**14**	**37**	**24**	**27**	**407**

The Interpopulation Network (see Figure 12) could be divided into clusters based on sequence similarity Cluster-I (RFLP nomenclature 1-2) is called the universal haplotype (Excoffier *et al.* 1992) and is present in world populations (Tharu, Oriental, Wolof, Peul, Pima, Maya, Finnish, Sicilian, Israeli Jews and Israeli Arabs). 'Isolated VI' is not an expansion cluster in that it has not undergone an expansion event, but represents the lineages that fall outside the five expansion clusters, this cluster is not monophyletic. The 'isolated' lineages have *Hpa*I Morph 3 (i.e. are positioned on the right of the Interpopulation Network, Figure 12). The column labelled 'Africa' is the number of individuals in each of the clusters in Africa.

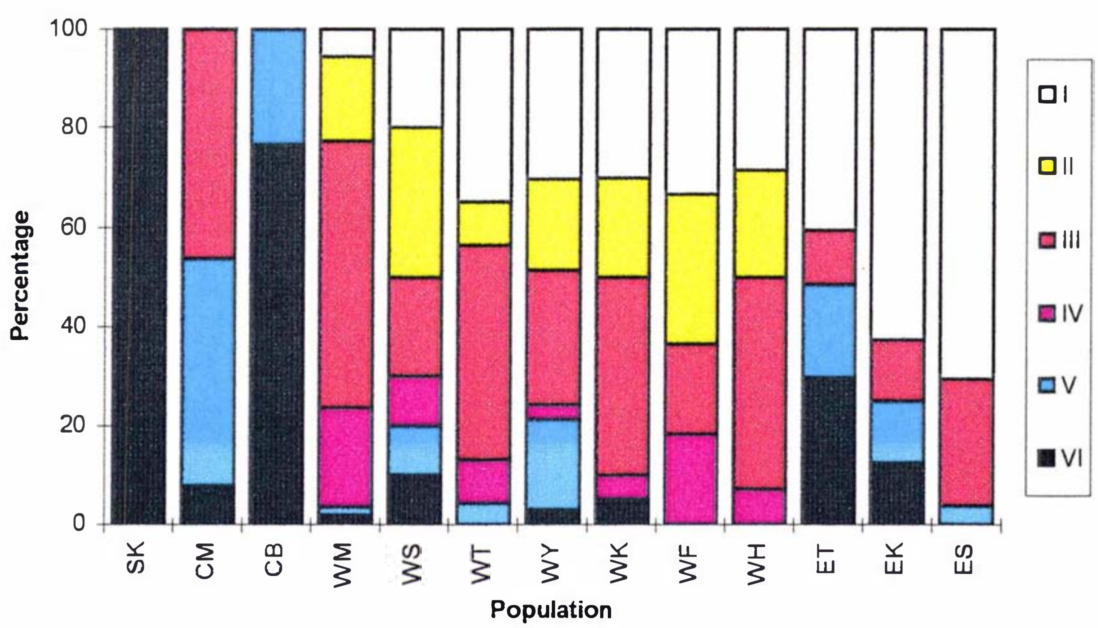


Figure 15 Graph representing the Percentage of Expansion Clusters in each population (%)

6.4.3 Age of Expansion events

Mean Pairwise Distance was calculated for all sequences within a cluster. Transversions were excluded from age estimations, although they are illustrated in the Interpopulation and Isolated lineage networks. Rogers (1995) applied the τ estimator which is based on method of moments. However, for most of the datasets this was not applicable as the mean exceeded the variance. Therefore the uncorrected mean was used as the estimate of the age of an expansion event of a cluster. A transition rate of 1.61×10^{-7} site/year was employed. This rate was calculated from expansion events of the A2 haplotype in Eskimos and Na-Dene. This assumed the diversity within these ethnic groups arose after the last glaciation 11300±500 years (Forster *et al. in prep.*). The rate is calculated assuming ρ is the measure for time depth. ρ is the average number of sites differing between a set of sequences and their last common ancestor. This transition rate corresponds to other calibrations (Ward *et al.* 1991; Hasegawa and Horai 1991), which were based on human-ape divergence. The rate was estimated on the equivalent region (np 16090-16365) as used in this study.

Table 27 Diversity of Clusters

Cluster	mpwd±s.d	Estimated age (expanding)	Geographical Range
I	5.255±2.078	59,000±2,000	World-wide
IIa	2.571±1.362	29,000±3,000	West Africa
IIb	2.585±1.569	29,000±3,000	West Africa
III	4.457±2.199	50,000±2,000	Africa
IV	1.801±1.391	20,000±2,000	West Africa
V	3.188±1.570	36,000±4,000	Africa
Isolated (VI)	9.134±3.493	103,000±5,000	Africa

Estimated expansion dates (assuming expanding populations), to the nearest 1000 years, based on the archaeological dating of the recolonization of Alaska, at the end of the Younger Dryas, with 1 transition/22,500 years. Cluster-VI has not undergone an expansion event.

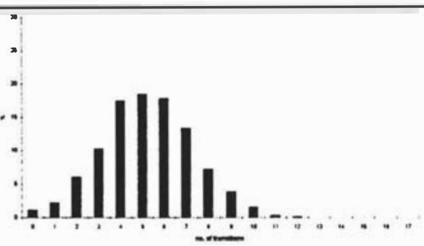
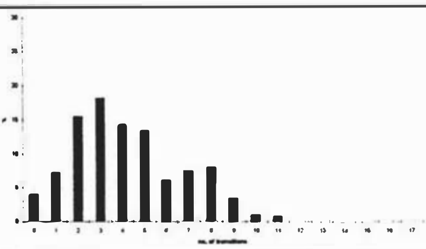
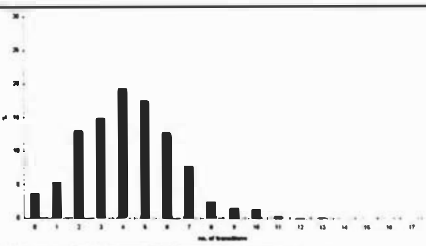
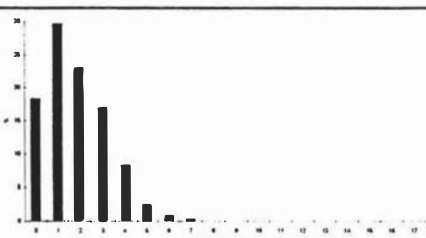
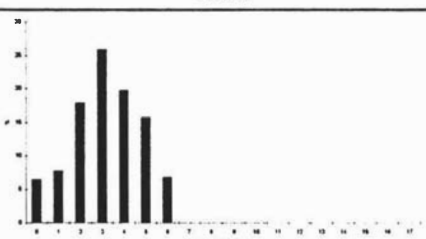
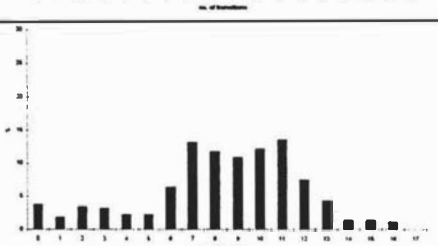
Cluster	distribution	mean and s.d.	τ
I		5.215 2.067	
II		4.236 2.463	2.883
III		4.242 2.137	3.672
IV		1.801 1.391	1.435
V		3.188 1.570	
VI		9.134 3.493	6.343

Figure 16 Pairwise distribution of Clusters I-VI

The pairwise distance distributions, with the horizontal axis representing the number of transitions between pairs and the vertical axis represents the percentage (%) of pairs. Note Cluster-II shows bimodality indicating dual clusters as illustrated in Figure 13. The mean, standard deviation and τ when applicable, are included for each cluster. τ is calculated from Rogers (1995) $\tau = m - \left(\sqrt{v - m}\right)$.

6.4.4 Geographical Origin of Clusters

Table 28 Expansion times of clusters in different geographical regions

	I	II	III	IV	V	VI
East Africa	5.378±0.290 61±3		3.527±0.453 40±5		3.382±0.379 38±4	8.835±1.087 100±12
Central Africa					1.267±0.401 14±5	5.637±0.985 63±11
West Africa	4.868±0.291 55±3	4.236±0.232 48±3	4.190±0.212 47±2	1.801±0.210 20±2	1.800±0.521 20±6	6.000±1.040 68±12
South Africa						3.099±0.612 35±7

The upper value in each cell is the mean pairwise distance ± the standard error of the mean. The lower value is the estimated date of the expansion (kya) ± the standard error of the mean. Not all sources of error are known. The standard error is not an appropriate statistic to apply to non-independent categories but is used as a guide to the variability in the distributions. The dates are in thousands of years and are rounded to the nearest thousand years. The shaded cells either have no representatives from that cluster or a small sample size (<5).

6.5 Discussion

Network methods prove very applicable to intraspecific data from HVR-I sequences. Network identify clusters which are shown to have distinct histories. The Africa ethnic groups (with the exception of the !Kung) are composite of these different clusters. The results indicate that for genetic studies, especially mitochondrial, that a traditional population based approach is not sufficient. Most African ethnic groups do not share a common history within the group but are rather composites of different clusters (Table 26). These groups share clusters with other ethnic groups, often over a wide geographical and cultural distances.

6.5.1 Expansion Events

Most calibrations are based on the 13 million year old separation of Orang-utans from apes and humans. For this paper the internal transition based calibration of Forster *et al. (in prep.)* was used. In Africa there is evidence of several expansion events. The early expansion event in Cluster-I occurred in East Africa or the Middle East approximately 60 kya and expanded westwards. In West Africa there was an expansion of Cluster-III at 50kya which later expanded to East Africa and may have been associated with the expansion of Cluster-V 40kya. In West Africa 30kya, there was an expansion of the twin clusters IIa and IIb, and 20 kya there were further expansions of Cluster-IV and V.

Using RFLP and sequence information in mtDNA, Harpending *et al.* (1993) observed expansions in Africa 80 kya and in Europe 40 kya, with a world-wide average of 50 kya. With pairwise distributions, Sherry *et al.* (1994) proposed a world-wide expansion approximately 40 kya, associated with the Middle Palaeolithic/Upper Palaeolithic boundary. This interpretation is not phylogenetically based, and as is seen in Africa, a lot of valuable information is lost when reducing the analysis to pairwise distances only. Assuming the population or group investigated shared a common history, this method would be appropriate. Instead however, each group/population is a composite of mitochondrial types with very different origins and histories.

6.5.2 Explanation of Expansions

The explanation of expansion events is difficult to determine, but may have been multifactorial. There are two types of hypothesis, climate change and biological change, including cultural change. Climate was proposed by (Rogers and Jorde, 1995) based on the evidence that Eastern Chimpanzees (*Pan troglodytes schweinfurthii*) underwent an expansion at the same time as humans (50 kya). Climate was probably not involved in the expansion of Cluster-I, as other African lineages did not undergo an expansion at the same time. Possibly there was a hitchhiker effect, where a mitochondrial type was associated with a behavioural, biological, or technological advance, which gave a group a selective advantage. A tightly linked positive mutation undergoing a selective sweep is unlikely, while although most non-African populations have a Cluster-I type, in Africa Cluster-I constitutes about a quarter of the mtDNA sequences involved in this study. The later expansions may have been associated with a ripple effect, whereby an expanding group absorbed or transferred technology to a neighbouring group, causing the previously constant sized population to also expand. There are examples of geographic neighbours simultaneously expanding; Clusters III and V in East Africa *c.* 40kya; Clusters IIa and IIb in West Africa *c.* 30kya; and clusters IV and V in West Africa *c.* 20kya.

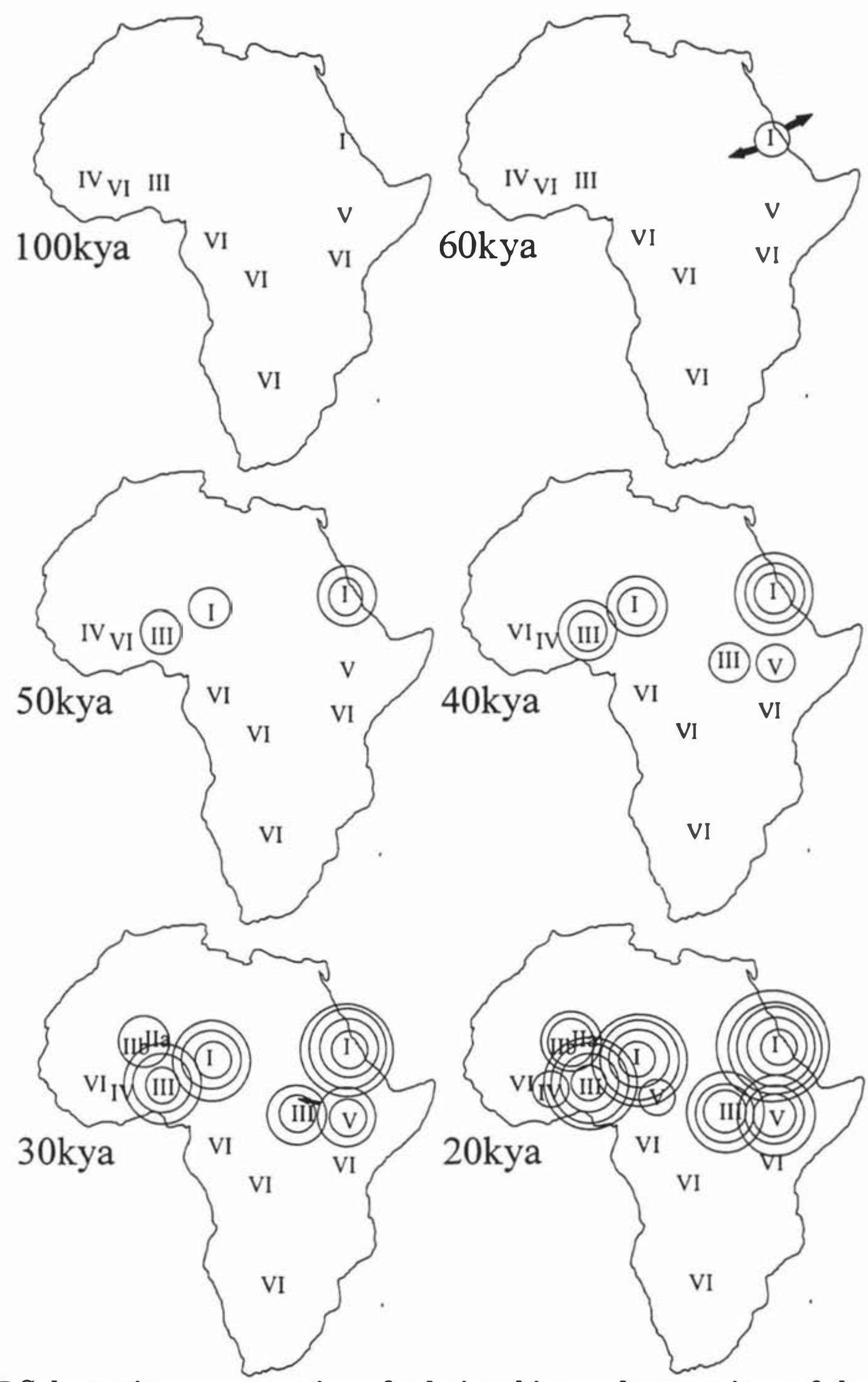


Figure 17 Schematic representation of relationships and expansions of the clusters. Circles indicate a demographic expansion event and concentric circles indicate a continuation of the expansion. Expansion in this context means demographic expansion, although in some cases was associated with a geographic expansion. 100 kya there were isolated types including some of the progenitors of the clusters which later underwent expansion. 60kya Cluster-I underwent an expansion event in the East Africa or the Middle East. 50kya Cluster-III underwent an expansion in West Africa and expanded eastwards. 40kya Cluster-III arrived in East Africa and during this time Cluster-V underwent an expansion event and expanded westwards. 30kya Cluster-IIa and IIb derivatives from Cluster-I underwent an expansion event. 20kya there were further expansion events in West Africa affecting Clusters IV and V. The Isolated (VI) types remained unaffected by expansion events.

6.5.3 Geographical Origin of Cluster-I

Outside Africa, Cluster-I is the expansion cluster from which other non-African lineages are derived (Excoffier *et al.* 1992). Within Africa, Cluster-I exists at high frequency in the groups associated with this survey (30-74%), and at a lower frequency (2-33%) in the Southern African populations (Soodyall, 1993).

To determine whether Cluster-I expanded in Africa or in the Middle East, the diversity of Cluster-I in the various regions can be determined. Cluster-I in Africa has a mean pairwise distance of 5.215 (East Africa 5.378±0.2900 and West Africa 4.868±0.2913) and in the Middle East (Di Rienzo and Wilson, 1991) a mean pairwise distance of 5.685±0.2800. These mean distances correspond to an expansion date of around 60kya. The order of expansion can not be established as the errors associated with the calculation are not known. However, the Middle East is a possibility for geographic origin. Geneflow may have been substantial between the Middle East and Africa. In historical times there may have been a significant amount of immigration from Middle Eastern groups; many African ethnic groups claim some Middle Eastern origin. However, Cluster-I is derived from African lineages which supports an African origin. Therefore Cluster-I probably originated in Africa, but may have expanded in the Middle East.

6.5.4 Summary

Detailed network analysis identifies five clusters which have undergone expansion events. Fourteen percent of the African sequences do not fit into these clusters. Generally these sequences are highly divergent from each other, accounting for the high mitochondrial diversity observed in Africa.

Cluster-I expanded first around 60kya in the Middle East or Africa. The cause of this expansion is not known. Although it occurred in only one lineage, it is unlikely to have been associated with a selective sweep, because it achieved only partial success in Africa. The initial expansion requires more information as to the cause. The other clusters expanded later and often simultaneously, suggesting a cultural innovation was causal to their expansions.

7. TREES FROM LANGUAGES AND GENES ARE VERY SIMILAR

7.1 Abstract

Tree diagrams have been used to represent relationships between both human populations and languages but only recently have the trees been compared (Figure 18) and the claim made "of considerable parallelism between genetic and linguistic evolution" (Cavalli-Sforza *et al.* 1988:6002). However in rejecting this claim it was pointed out (O'Grady *et al.* 1989) that the original study did not have an objective measure to support their conclusion of high similarity. But the lack of an objective measure invalidates the denial (O'Grady *et al.* 1989) just as much as it invalidates the claim that the trees are similar (Cavalli-Sforza *et al.* 1988; 1989). The original work has recently been supported by computer simulation to estimate the probability of finding such agreement between the gene and language trees. Tree comparison metrics are perhaps a more natural way of comparing trees quantitatively and we have recently extended knowledge of such metrics (Steel and Penny, 1993). We report here that an objective tree comparison measure can resolve the issue of the similarity of the gene and language trees from Cavalli-Sforza *et al.* (1988) showing that they are indeed far more similar than expected by chance.

7.2 Introduction

Several tree comparison metrics are available (Williams and Clifford, 1971; Robinson and Foulds, 1981; Penny and Hendy, 1985; Day, 1986; Swofford, 1991) though a limitation has been that their properties and distributions are poorly known. Recent progress (Steel and Penny, 1993) allows the question to be reconsidered. We have used the path difference metric to measure the similarity between the gene and language trees and can now report that the similarity between the trees is very highly significant.

7.3 Methods

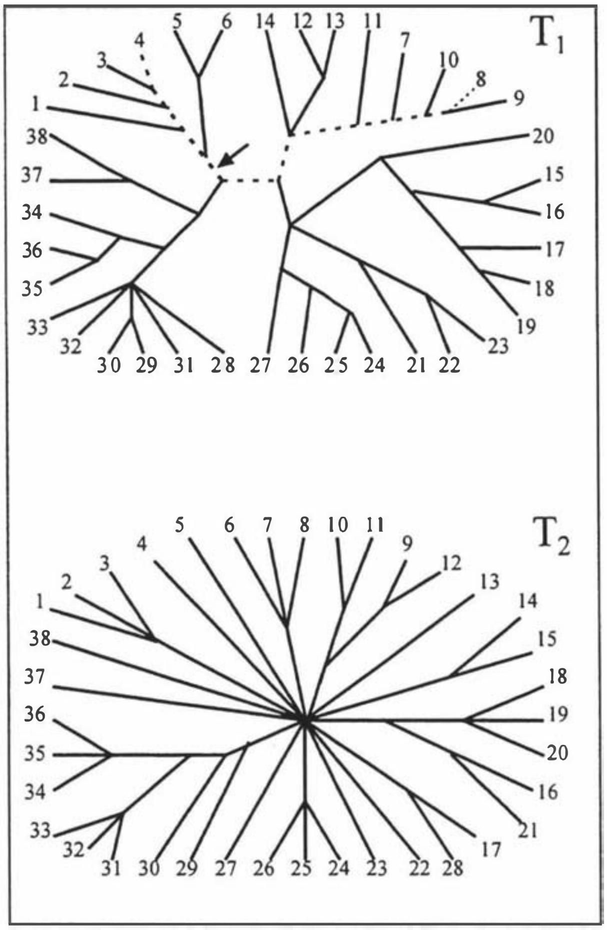


Figure 18 The two unrooted trees from Cavalli-Sforza *et al.* (1988) with T_I the genetics tree and T_2 the language tree. T2 has some slight differences from that in Cavalli-Sforza *et al.* (1988) to make it follow more closely the data in Ruhlen (1987) thus allowing the trees to be considered more independent. The changes are to link; Sardinian and European (Ruhlen, 1987:294); and Iranian and Indian (Ruhlen, 1987:293); rather than have a four-way split. In addition, Mbuti pygmies are included as a Bantu language which may well be a language displacement (Ruhlen, 1987:292). These changes have little effect on the path difference metric because they are local (Steel and Penny, 1993). If the European groups are separated that were combined in the original (Cavalli-Sforza *et al.* 1988) then the trees would be even more similar. The path length metric is illustrated by dotted lines showing the paths between one pair of (taxa 4 and 8) which have the largest difference in lengths. Labels: (1) Mbuti, (2) W. Africa, (3) Bantu, (4) Nilo-Saharan, (5) San, (6) Ethiopian, (7) Berber, (8) S. W. Asia, (9) Iranian, (10) European, (11) Sardinian, (12) Indian, (13) Dravidian, (14) Lapp, (15) Uralic, (16) Mongol, (17) Tibetan, (18) Korean, (19) Japanese, (20) Ainu, (21) N. Turkic, (22) Eskimo, (23) Chukchi, (24) S. Amerind, (25) C. Amerind, (26) N. Amerind, (27) Na-Dene, (28) S. Chinese, (29) Mon Khmer, (30) Thai, (31) Indonesian, (32) Malaysian, (33) Filipino, (34) Polynesian, (35) Micronesian, (36) Melanesian, (37) New Guinean, (38) Australian. (Additional details are in Cavalli-Sforza *et al.* 1988).

The following reasons were considered in selecting a path difference metric. The trees are compared as unrooted trees (Penny *et al.* 1992) and each edge had the same weighting. Problems considered in selecting the tree comparison metric were (1) that the underlying topologies (unlabeled trees (Penny *et al.* 1992)) differ, with the genetics tree being almost fully resolved and the linguistic tree much less resolved, and (2) as pointed out by the original authors (Cavalli-Sforza *et al.* 1988, 1992), some groups show "language displacement" where some groups adopt another language and so appear at a different place on the language tree. The partition metric (Robinson and Foulds, 1981) was therefore not used as it is particularly sensitive to a taxon being shifted on the tree (Penny and Hendy, 1985) and its properties are not so useful for nonbinary trees (Steel, 1988). Nevertheless the partition metric found three internal edges (six similarities) between the two trees, and this alone would be highly significant (Penny and Hendy, 1985).

The path difference metric is the square root of $\Sigma(P_{i,j}-Q_{i,j})^2$, where $P_{i,j}$ and $Q_{i,j}$ are the path lengths between pairs of taxa i and j on the trees T_1 and T_2 respectively (Penny and Hendy, 1985). The sum is over all pairs i and j. The metric measures the Euclidean distance between the path lengths for all pairs of taxa (populations or languages) on each tree. The dotted lines in Fig. 1 show the paths between the pair of taxa 4 and 8, which have the largest difference in lengths.

7.4 Results

Evaluating the path length difference metric for the language and genetic trees gives the value 126.02 (a zero value would mean the trees are identical). For this metric the distribution is only known when all binary trees are equally likely and so Monte Carlo simulation was used to estimate how frequently two trees with 38 taxa and with the same underlying topologies (unlabeled trees, Penny *et al.* 1992) would be at least as similar by chance. All taxa were randomly repositioned on each tree 107 times and we found only 126 cases where the metric was this value or less. Our conclusion is unambiguous - the two trees are indeed far more similar than expected by chance. There is about one chance in 100,000 that such similar trees would be found, given that the trees differ in resolution (number of internal branches).

7.5 Discussion

The present results do not explain why the trees are similar; this has been discussed by Cavalli-Sforza *et al.* (1988, 1992). We are interested here only in the factual question of whether two trees are more similar than expected by chance. With both data sets the trees are approximations to a more complex set of relationships. With languages there can be transfer of words as well as language displacement. With populations there is interbreeding and genetic amalgamation (Ward *et al.* 1991). However with both sets of data this largely occurs between contiguous groups and a tree is still a useful representation for both data sets.

The high similarity of the language and genetic trees appears to place some constraints on the timing of human evolution and therefore may help illuminate current controversies (for example, Hedges *et al.* 1992). The results are consistent with early *Homo sapiens* having language capabilities (Foley, 1991; Corballis, 1992), rather than language having developed only very recently (Noble and Davidson, 1991). It seems unlikely that the trees would be so similar if language developed after the separation of early human groups. Nor would the results support a continuity between all *Homo erectus* and *Homo sapiens* populations without languages having developed very early.

Tree comparison metrics have the potential to be useful in phylogenetic studies in that they readily allow quantitative answers to specified questions. We have reported some of their applications in testing evolutionary hypotheses and measuring rates of convergence to a single tree as longer sequences are used (Penny *et al.* 1991). A current limitation is their restricted availability in standard computer packages. But then because they are not used very often, there is little incentive to provide them in packages!

8. DISCUSSION

This thesis presents a range of subjects focusing on human genetic history in Africa as well as addressing issues of African ethnicity and human evolution. This discussion will synthesise results and conclusions from the previous sections. Areas which need further investigation will be discussed and questions will be proposed which need to be addressed in the future.

The Multiregional and Out of Africa hypotheses do have testable hypotheses. The mtDNA data supports an Out of Africa scenario, however, the data and analysis have limitations. It is critical that the calibration of the clock is correct and that the world has been sampled extensively, both of which are not certainties. It is also important to acknowledge that human origins and evolution are probably a result of multifactorial processes, superimposed on one another and that one hypothesis may best summarise the evolution of humans, but other processes may be acting and be expressed in different data sets such as the palaeontological record.

8.1 Patterns within Africa

8.1.1 Economy, Language and Geography

Hypotheses regarding African ethnicity and the relationship of genetics, economy, language, and geography were tested. Ethnic groups appear to be a composite of mtDNA elements, relating to the geographic area they live, the language they speak and the economic mode that sustains them.

Two phenomena are observed with the African HVR-I sequences. Firstly, mtDNA lineages are shared across the categories of economy, language phyla and geographic regions. Secondly, the Hudson test does detect structure based on these boundaries, i.e. there is more sequence identity within the categories of economy, language and geography; than would be expected by chance.

8.1.2 Ethnicity

Pult *et al.* (1994) noted the high proportion of shared lineages between European populations compared with the lack of shared lineages in African populations (the Yoruba, !Kung and the Pygmy). The !Kung and Pygmy populations, however, are not representative of Africa as a whole. The African populations studied in this thesis share lineages with each other despite often large geographical distances (<5,700km).

The composite nature of mtDNA types found within African ethnic groups is consistent with the concept that groups are a product of fission-fusion population dynamics. Sequences within an ethnic group do not share a common mtDNA ancestor unique to the group. The sequences instead represent different aspects of the groups history, including the geographical location of the group, which language they speak and the lifestyle they lead, each factor has contributed culturally to the group and genetically to the mtDNA genepool of the group. Implicitly with the mosaic distribution of both economy and language in Africa (see **Map 1**), strong regionalism would not be expected. The linguistic phyla are approximately 12,500 years old (Blench, 1993), and economic patterns are even younger, and both these factors appear to be superimposed on the mitochondrial types specific to the geographic region. The pattern is further complicated by recent admixture as shown in this survey, of which 5% of the individuals had parents with different ethnic backgrounds (see p. 216). All individuals identified with the paternal ethnic group.

8.1.3 Migrations and Expansions within Africa

Based on mitochondrial lineage frequencies and diversity it is possible to determine a putative geographical origin and an expansion date for each mitochondrial cluster. The median Networks method (Section 6) identifies different elements within populations (clusters), and by determining a gradient of the frequency and diversity of the elements, it is possible to hypothese the geographic origin of each cluster.

Most African groups (except the !Kung and Pygmy populations (see Section 6) have Cluster-I which is the central lineage for all world populations (Excoffier *et al.* 1992). The low level of diversity in humans as compared to other species has been interpreted as representing a severe population reduction, but this lack of diversity can be accounted for by a phenomenal expansion a population with one cluster. This result is supported by the nuclear study (Tishkoff *et al.* 1996) whereby non-African individuals (and North East Africans) have the Alu deletion associated with a single STRP allele, whereas Africans have the Alu deletion associated with many STRP alleles, supporting a recent origin of non-Africans.

In Africa several major population expansions occurred: the first at approximately 60 kya which was associated with the expansion of Cluster-I, this expansion had a East African/ Middle Eastern focus. Several other expansions followed occurring approximately 50, 40, 30 and 20 kya, each with a East or West African focus. The explanation of Cluster-I's expansion is not known, a selective sweep is unlikely as only 25% of African mtDNA are of that type. It may have been associated with a climatic change although no other mitochondrial type underwent expansion. Possible explanations include the Cluster-I type hitchhiking with a nuclear mutation or cultural innovation. Hitchhiking is defined as the process whereby an allele at one locus has a selective advantage, an allele associated with it with little or no selection may increase in frequency as a result (Kojima and Schaffer 1967; cited in Hedrik 1985:p.354). In this thesis the term hitchhiking is expanded to include indirect linkage with a nuclear locus (epistatic) or with a cultural innovation. The later expansions of the other cluster appear to be culturally induced as they had a geographic focus and usually included more than one cluster.

Some African groups have as much diversity as the rest of the world combined for instance the Turkana of East Africa. The common ancestor of one ethnic group such as the Turkana may be over 100,000 years old (see 4.6.6). This due to the high diversity of Cluster VI lineages as well as each ethnic group consisting of a composite of different genetic elements each with their own, origin and expansion event (see **Table 26**)

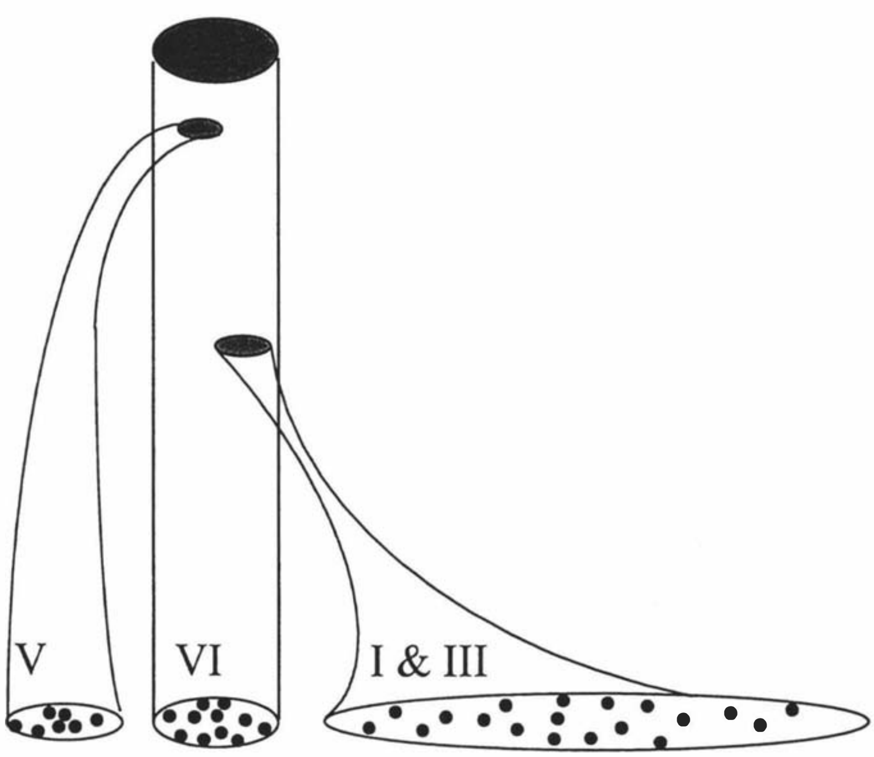


Figure 19 Schematic diagram of the relationships of the Turkana sequences
The 37 Turkana sequences are illustrated as dots. Cluster-VI is most divergent group and V and I & III have undergone expansion events. With the addition of the two West African clusters (II and IV) this diagram would be representative of the World. The most recent common ancestor of the Turkana may in fact be the most recent common ancestor for *Homo sapiens sapiens*.

8.2 Patterns within the World

8.2.1 Diversity

Africa has the highest diversity of all the continents in most genetic systems studied. The only apparent exception are the results from allozyme studies, which are not designed to detect all polymorphisms (Stoneking, 1994). Selection for polymorphism in these studies was based primarily on European populations. Hence results from classical enzymes will be disregarded and the focus will be on explanations for the high diversity in Africa.

Many authors have interpreted the high African genetic diversity as indicating the geographical origin of modern humans. This is the simplest explanation but other explanations should be considered. Diversity could also be indicative of high effective population size or accelerated substitution rate and these alternatives were addressed in Section 5. Substitution rates were tested by independently comparing African and non-African sequences with different outgroups, the mitochondrial DNA nuclear insert (Zischler *et al.* 1996) and a chimpanzee *Pan troglodytes* (Morin *et al.* 1994). No variation in rate was observed. Therefore differential rates were eliminated leaving the older age of the African population (i.e. centre of origin of species) or the higher population size of the African population, as the explanation for high diversity.

The definition of effective population size includes the age and size of the population. Eurasia may have had a lower population size due to population crashes associated with glaciation events. Even during the full extent of glaciation there would have been large areas which would have been able to sustain relatively large populations. The dominance of Cluster-I and its derivatives outside Africa, suggests that population crashes is not a feasible explanation on its own. Eurasia was presumably a diverse widespread population. If it underwent a population crash there would have been several refugia. Stochastically each refugia would comprise of different sequences and therefore it would be expected that most variability would be maintained in Eurasia as a whole (see Figure 21). By a process of elimination an African origin for modern humans is the best supported interpretation as to the high diversity within Africa.

8.2.2 One Mitochondrial Type Founded the Rest of the World

Cluster-I (corresponds to the RFLP group 1-2) has been identified as the central haplotype of different world populations (Excoffier *et al.* 1992). Almost all lineages outside Africa are from Cluster-I or its derivatives, while within Africa Cluster-I accounts for only 25% of the sequences studied (see Section 6). Linkage disequilibrium analysis of the CD4 locus also indicates that the populations outside Africa experienced a founder effect (Tishkoff *et al.* 1996), whereby there was an expansion of one genetic element.

8.2.3 Bottlenecks and Expansions

The relatively low level of mtDNA diversity observed in human populations can be interpreted in several ways. Jorde *et al.*'s (1995) comparison of nuclear and mtDNA markers indicated different ratios of the depths of the African lineages to the rest of the world's (with the mtDNA lineages being relatively deeper). One explanation is that there may have been a bottleneck which would effect the mtDNA diversity (due to its lower effective population size) more severely than the nuclear genepool. These results may be consistent with Africa maintaining most of its diversity due to stable or expanding populations, while Eurasia underwent one or more bottlenecks. This would explain a significant loss of mitochondrial diversity outside Africa. However even if there was a severe bottleneck (due to a population crash) in Eurasia one would expect that the population was geographically widespread and therefore most of the diversity would have been maintained as discussed in 8.2.1. Australia and some parts of Asia are not extensively sampled; therefore some of these conclusions may change with more sequence information. Extensive modelling needs to be done to test effects of a bottleneck on the diversity of a large geographical range such as Eurasia.

FOUNDER EFFECT **expansion of one type due to hitchhiking, selection, or colonising a new area**	**BOTTLENECK** **collapse of a population possibly due to climatic severity**

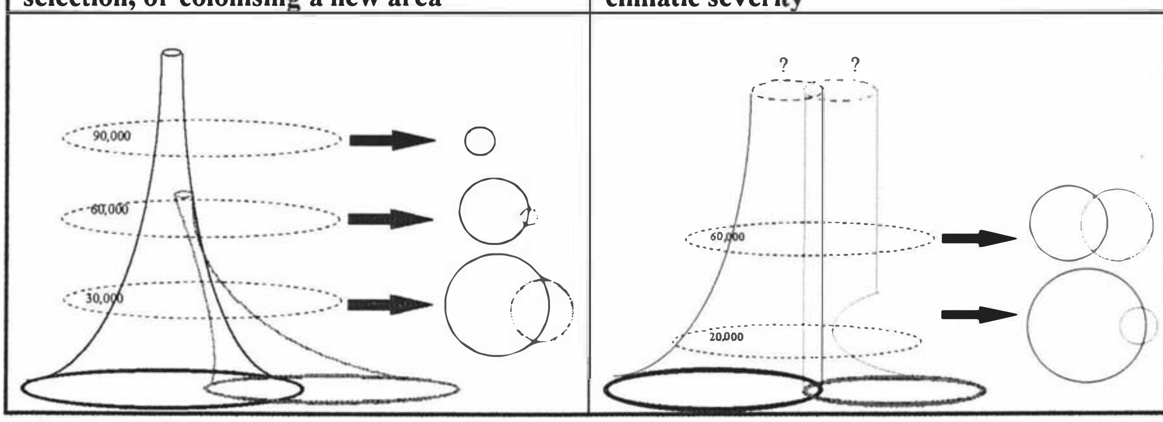


Figure 20 Scenarios for lower mtDNA diversity observed outside of Africa
Black illustrates sequences from the African continent and red from sequences outside Africa. The diameter of these structures represents diversity. At different time intervals a cross section is shown, illustrated as Venn diagrams. The Bottleneck scenario is presented as being consistent with the Multiregional Hypothesis.

Various scenarios can explain the patterns of diversity and phylogenetic relationships between the continents. Although a population crash, perhaps associated with successive glaciation events could lower the diversity outside Africa (see Figure 21), while Africa maintained its population size and therefore its diversity. Results from the Network analysis (see Section 6.5.1) indicate that a population crash outside Africa is implausible, due to the demographic success of one mtDNA lineage outside Africa, is implausible for presumably there would be several refugia during the glaciations and the refugia would be unlikely to all have the same lineage. This does acknowledge that there may have been a population crash outside of Africa, but not that the population crash is the only explanation for low diversity.

Early distribution of *Homo sapiens*	**Population crash in the temperate zones, possibly due to a glaciation event**	**Population growth and expansion.**

Figure 21 Possible Model of Population crashes in Eurasia
Loss in mtDNA diversity outside Africa may be accounted for by severe population crashes, possibly caused by glaciation events. The refugia populations could expand and recolonise in the interstadials. At the local level, it may be possible to have reduced diversity, but when observing Eurasia as a whole, most diversity should be retained, thus contradicting this model.

8.2.4 MtDNA Survival and Selection

Avise *et al.* (1984) demonstrated that by stochastic mechanisms in a population of constant size lineage extinction would occur. If n is the number of females, a population of constant size would take on average $4n$ generations until all descendants would trace back to a single founder female. The date of mitochondrial Eve does not necessarily equate with the origin of *Homo sapiens sapiens*. The date of mitochondrial Eve could be earlier or later than the date of the first palaeontological modern humans, depending on population dynamics (see Figure 22). It must be emphasised that the history of the mtDNA molecule may not necessarily equate with the history of humans as a species.

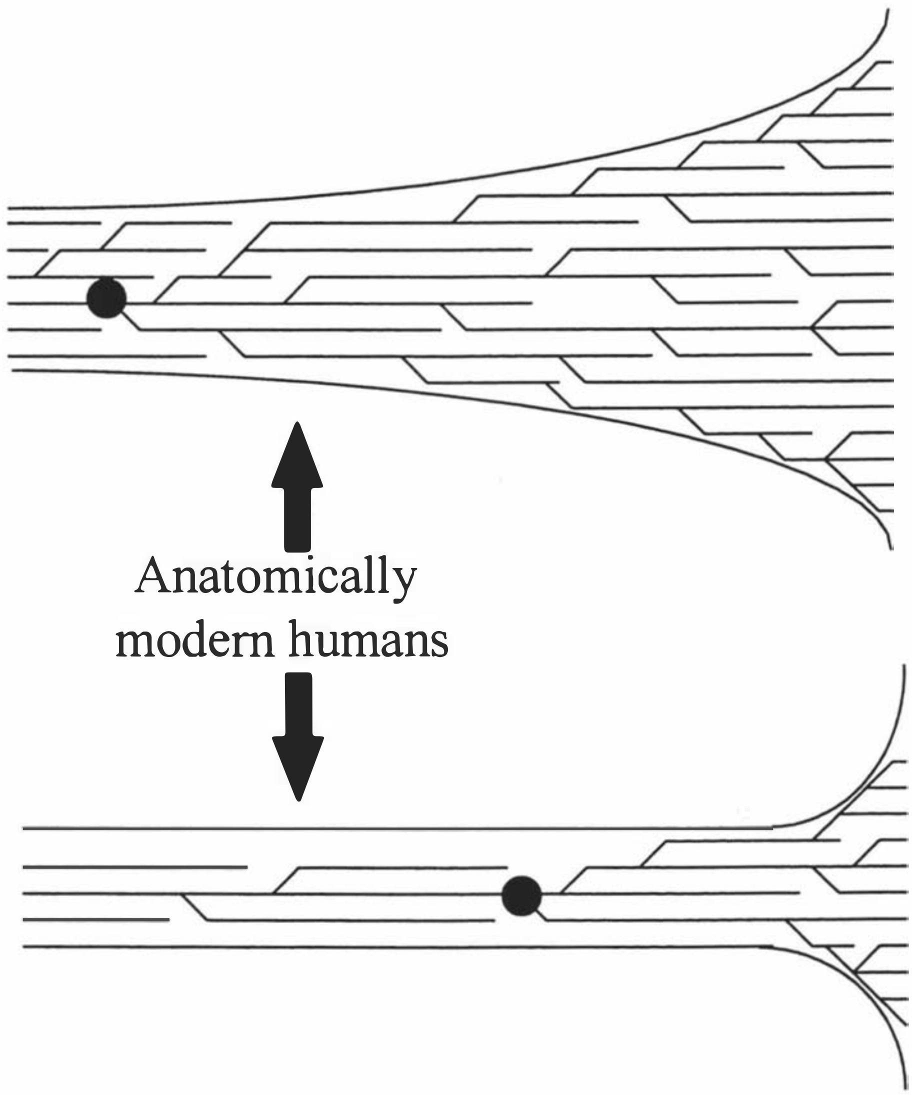


Figure 22 Mitochondrial Eve and Anatomically Modern Humans
Mitochondrial Eve (indicated with a dot) may be earlier or later than the date of anatomically modern humans. In a population which is gradually growing (top illustration), or a large constant population (i.e. not having experienced bottlenecks) the most recent common ancestor may precede the emergence of the species. In a population which maintains a small constant size over a long period of time (bottom illustration) or a population which has undergone a population reduction (i.e. having experienced bottlenecks) the most recent common ancestor may be younger than the emergence of a species. Recent selective sweeps would also to make the date of the mitochondrial Eve more recent (not illustrated).

Selection is a process whereby a allele has an advantageous mutation which improves its chance of survival, and therefore becomes more frequent in a population. Sometimes other alleles and neutral alleles which are associated with the gene under selection, also become more frequent in a population, this process is termed genetic hitchhiking. This process can be due to linkage or some other factor. The process where the allele which is under positive selection pressure and all linked alleles becomes frequent in a population is termed selective sweep. For this discussion **directly linked hitchhiking** will refer to the tight linkage of a neutral allele to a mitochondrial gene under selection, and **indirectly linked hitchhiking** will refer to a more flexible association of the neutral allele to a nuclear gene under positive selection (this is also defined as epistatic interaction) or association of a neutral allele to a favourable cultural trait such as more sophisticated tools. Success of a mitochondrial type may also be due to a chance event such as the dominant mitochondrial type in a founder group.

Linked hitchhiking is associated with non-recombining genomes, or parts of genomes, for example, mitochondrial DNA, the non-recombining part of the Y-chromosome, or genes that are very close to the gene under positive selection pressure. There are a wide spectrum of degenerative diseases (including blindness, movement disorder, dementia, stroke, heart disease, renal failure, and diabetes) associated with mtDNA (Wallace 1995). These diseases are caused by missense mutations (affecting the polypeptide sequence) and protein synthesis mutations (affecting the tRNAs and rRNAs) (Wallace 1995). As mtDNA is involved with metabolism it is feasible to expect a favourable mutation could convey a selective advantage to the individual and therefore any neutral substitutions associated with it. Directly linked hitchhiking, in this case a favourable mutation anywhere in the mtDNA genome, could favour neutral mitochondrial types linked with it. There is no evidence, however, of metabolic superiority of individuals with the Cluster-I lineage as Cluster-I which accounts for almost all mtDNA sequences outside Africa, accounts for only one quarter within Africa. Another possibility is an early selective sweep which accounts for all mitochondrial types observed today in modern humans. This theory has been proposed by S. Easteal (*pers. comm.*). The early selective sweep is proposed on the apparent anomaly of the mitochondrial and nuclear genetic record (Jorde *et al.* 1995). The selective sweep began in Africa gradually spreading throughout Africa and then expanded into the rest of the world, suggesting the success of the Cluster-I may be due to its geographical location.

Unlinked hitchhiking seems more plausible. Indirectly linked hitchhiking is similar to a directly linked hitchhiking in the sense that neutral substitutions are linked with a positive mutation, although in this case not directly linked. A mitochondrial type may be associated with a favourable mutation in the nuclear genome. The favourable mutation would not be linked to a specific mitochondrial type, due to maternal only inheritance and recombination events, but this may explain the initial success of one lineage.

Another alternative is if a mitochondrial type was frequent in a population which innovated a technological advance. A new technology would allow the population to expand and may give them a selective advantage over other populations. Some variation of the hitchhiker effect is the most realistic way to explain the world-wide success of the lineage associated with Cluster-I (RFLP type 1-2), without completely replacing other types within Africa. Another possibility for the success of Cluster-I may have been due to the geographic location enabling that type to be in the position to colonise the rest of the world.

The early selective sweep of mtDNA does not influence the results of Section 6, however, it would change the interpretation of the origins of humans, in that it could support a Multiregional hypothesis. This theory requires further investigation.

8.3 Human Evolution

8.3.1 Fossil Record

Genetically there is evidence supporting the Out of Africa hypothesis (Gibbons, 1995), however, at the palaeontological level there appears to be conflicting interpretations. Stringer and Bräuer (1994) state that there is agreement with the Out of Africa hypothesis and the fossil record. Whereas the opposing Multiregional hypothesis is supported by other palaeontologists (Wolpoff, 1989; Frayer and Wolpoff, 1993; Frayer *et al.* 1994; Frayer, 1994).

Homo sapiens

There is evidence for early humans or archaic humans in Asia, between 200,000 and 300,000 BP (Wu and Bräuer 1993; Tiemei *et al.* 1994). Bräuer argues the archaic Chinese (dated between 300,000-400,000 BP) are closer to modern Africans. According to almost all calibrations these dates are older than the out of Africa expansion proposed by geneticists. Discussions regarding the out of Africa hypothesis generally do not acknowledge this information. This information suggests; according to the standard interpretation of out of Africa; not only did modern humans replace *Homo erectus* but also archaic modern humans *Homo sapiens*. The calibration of the molecular clock may be conservative, in which case, the above dates may fit in with the expansion out of Africa. Therefore the out of Africa expansion would have been archaic humans *Homo sapiens* rather than modern *Homo sapiens sapiens*.

Stringer (1992) compared archaic forms with modern forms from Africa and Eurasia and found the African archaic was "narrowly" favoured over the East Asian archaic as being closer to modern humans. This information suggests that there was an early expansion out of Africa of *Homo sapiens*, although there may had been some degree of regional continuity between archaic and modern humans.

Homo erectus

A *Homo erectus* mandible found in East Georgia (Caucasus) has been reliably dated from 1.6 to 1.8 Mya (Gabunia and Vekua, 1995). Some of the Asian *Homo erectus* dates have been suggested to be 1.1 Mya for the Lantian Gongwangling crania from China (An and Ho, 1989; cited in Gabunia and Vekua, 1995) and 1.8 Mya and 1.7 Mya for *Homo erectus* from Mojokerto and Sangiran sites respectively (Swisher *et al.* 1994). These dates places the age of the last common ancestor of modern humans, as proposed by the Multiregional hypothesis, as greater than one million years

Multiple waves Reconcile the Multiregional with the Out of Africa Hypotheses

The genetic and palaeontological record may support both the Multiregional and Out of Africa hypotheses. There may have been several expansions out of Africa, with each wave almost completely replacing the earlier hominid. This could explain the persistence of some phenotypic features, and the existence of a few mtDNA outliers.

The early dates of archaic human outside Africa suggest that there were at least three and possibly four, hominid expansions out of Africa, all which replaced or almost completely replaced the previous inhabitants. The first expansion would be *Homo erectus* (1.6-1.8 million years ago) followed by archaic *Homo sapiens* (200,000-300,000 years ago) followed later by an expansion of modern human *Homo sapiens sapiens* (within the last 100,000 years). In addition mtDNA sequences suggest there was a fourth expansion out of Africa. Evidence for this is the presence of a few individuals outside of Africa with sequences from Cluster-VI, (7 Europeans[1] and 1 Australian (Vigilant *et al.* 1991)); Cluster-IV (1 European), Cluster-III (2 Sardinians and 3 Middle Easterners (Di Rienzo and Wilson, 1991)). This could represent an early expansion out of Africa of *Homo sapiens sapiens.* The estimated dates of the separations of these clusters (see Section 6) preceded the *Homo sapiens* fossils. If the calibration was incorrect then some of these outlying sequences could be from a *Homo sapiens* expansion.

Africa may have been a hominid epicentre, generating more technologically or biologically advanced hominids. Such an interpretation leads to further questions which need to be answered. Was there geneflow between these different hominids? Were these hominids out-competed by humans with advanced technologies or superior mental capacities, or were they victims of more active aggression? Were the hominids susceptible to diseases the new humans brought with them? Would further genetic surveying yield further outliers (i.e. belonging to Cluster-VI) in Eurasia, or perhaps sequences distinct enough to be *Homo erectus*?

Some palaeontologists see degrees of regional continuity in human populations (Thorne and Wolpoff, 1992) in Asia, Australia and Europe but others see no evidence for this (Stringer and Bräuer, 1994). Nei (1995) discusses the regional characteristics proposed to support the Multiregional hypothesis and discusses them in context with the genetic results. Proponents of the Multiregional hypothesis claim features such as shovel shaped incisors in East Asians, prominent brow ridges in Australians, large noses in Europeans, and other traits have demonstrated regional continuity through the fossil record (Thorne and Wolpoff, 1992). Most genetic evidence suggests that gene flow (if it was not complete replacement) must have been quite extensive. The Multiregional hypothesis predicts there was substantial multidirectional geneflow, whereas Out of Africa predicts substantial geneflow in one direction. With the postulated high levels of geneflow, it would have been unlikely, although not impossible, to maintain some regional features. These features illustrating regional continuity could be an adaption to the environment or merely due to chance. It should be noted that there is very little mtDNA information from Australia; and East Asia is not extensively surveyed. More detailed studies of these populations are required to fully evaluate the genetic support for the Out of Africa and Multiregional hypotheses.

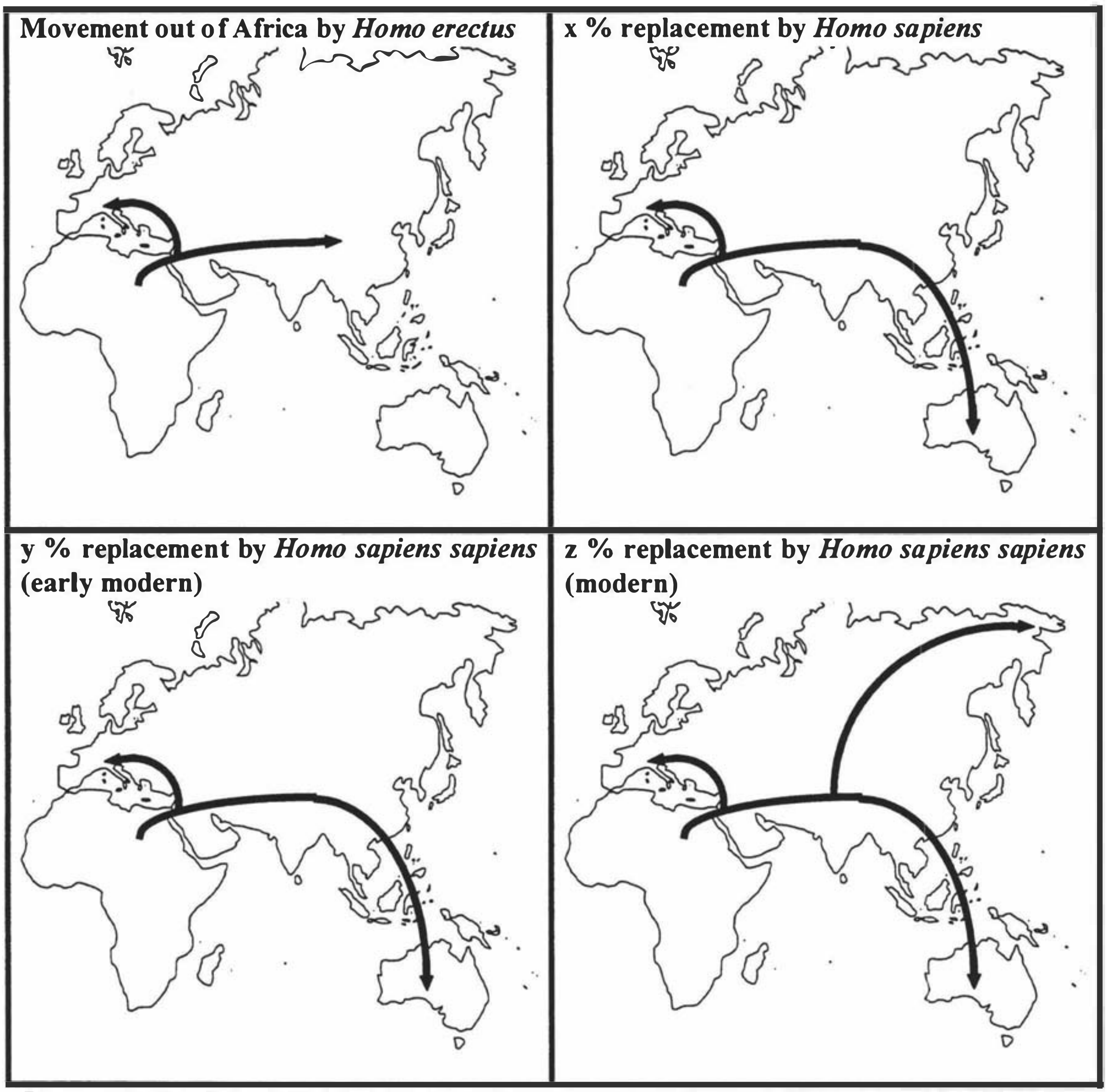


Figure 23 Model of different expansions out of Africa and their contribution to today's mtDNA genepool

This model assumes interbreeding between the various hominid species, therefore allowing for some continuity of traits as presented by the proponents of the Multiregional hypothesis. Assuming that the mtDNA types can be distinguished between the various hominid groups this could be tested with more mitochondrial data. The various variables may be specific to different regions.

	Variables	
Percentage of modern genepool	x = 95% y = 95% z = 95%	x = 0% y = 90% z = 95%
Homo erectus	0.0125	0.5000
Homo sapiens	0.2375	0.0000
Homo sapiens sapiens I	4.7500	4.5000
Homo sapiens sapiens II	95.0000	95.0000

To reconcile palaeontological information there are several explanations: the archaic humans along with *Homo erectus* were outcompeted; or their contribution to the genepool is not detectable due to a rapid influx of modern genes; or the calibration of the molecular clock is incorrect (see 8.3.4); or Asia, Australia and New Guinea have not been sampled extensively enough .

8.3.2 Europe

The debate of the origin of modern humans cannot be pursued without addressing the question of the European/Middle Eastern Neanderthal *Homo sapiens neanderthalensis*, or *Homo erectus neanderthalensis*. There is palaeontological evidence that they lived in Europe and the Middle East from 300,000-30,000 BP. There are three popular hypotheses regarding the origins of the Neanderthals and their relationship to the modern humans (Wenke, 1990), representing three points on a continuum.

1. The Neanderthals were the direct ancestor to *Homo sapiens sapiens*. This is supported by the fact that there is a gap in the fossil record between 45,000-25,000 BP, in which period proponents claim that Neanderthals evolved into *Homo sapiens sapiens*. Thorne and Wolpoff (1992) state that characteristics such as a big nose and bony bridge on mandibular nerve canal, have persisted through the Upper Palaeolithic to recent times.

2. The Neanderthals were cold specialists, most probably from *Homo sapiens* stock. The arrival of the cognitively and/or technologically superior *Homo sapiens sapiens* in Europe placed the Neanderthal under severe selection pressure, where they were either absorbed or displaced. Zubrow (1989; cited in Bräuer, 1989) demonstrated with demographic modelling that a small demographic advantage by modern humans would have been sufficient to result in a extinction of the Neanderthals in the course of only a thousand years. This is contradicted by the evidence that Neanderthals had a lower mortality rate than humans during the Pleistocene (Wenke, 1990).

3. The Neanderthals were a distinct *Homo erectus* lineage which is now extinct. When Stringer (1992) compared archaic forms with modern forms from Africa and Eurasia, the Neanderthal group were distinctive. This theory implies there was no interbreeding between Neanderthals and modern humans.

There is evidence that Neanderthals and modern humans overlapped spatially and temporally in the Middle East. The Qafzeh site in Israel shows early evidence of modern human occupation, dated by thermoluminescence at 92,000 years BP. Neanderthals found in Kebara (Israel), have been dated at 60,000 BP (Valladas *et al.* 1988; cited in Stringer, 1988). The Neanderthals and modern humans may not have been living as neighbours, rather alternating habitat according to climatic fluctuations. There would have been opportunity to interbreed, and some palaeontologists claim there is evidence of hybrids (Bräuer, 1982, 1984, Vandermeesch, 1985:p.195, cited in Bräuer and Rimbach, 1990). This information argues against regional continuity as humans preceded Neanderthal occupation in the Middle East. Questions arise from this. Why did the humans not move into Europe? Were the Neanderthals so established, or did it take a technological advance in modern humans to colonise Europe (Stringer, 1988).

Early Neanderthals were in Europe at least 300,000 years ago and therefore one would expect the Neanderthal mtDNA to reflect this early separation from the human line. Assuming that the molecule clock is correctly calibrated there is no evidence of mtDNA sequences in Europe which could have separated so early. The large survey of European HVR-I sequences presented by M. Richards *(pers. comm.)* show only eight outliers from a very monomorphic genetic pool of over 600 individuals. These European outliers have sequences close to or identical to Africans. The distance from these outliers to the Anderson reference sequence (the most common sequence in Europe), is about six substitutions. These sequences probably represent more recent migrations.

8.3.3 Asia and Australia

Although mtDNA has been surveyed in some parts of Asia, there are crucial regions and populations which have yet to be sequenced. Populations from the Andaman and Nicobar Islands, as well as the Semang of Malaysia and the Aeta of the Philippines, are termed "Oceanic Negritos", exhibiting African phenotypic traits (Cavalli-Sforza *et al.* 1994: p.349). These groups are not shown to be genetically distinct based on allozyme information (Cavalli-Sforza *et al.* 1994:p.234), however, should be investigated mitochondrially. Affiliation with Africans could be tested with the analysis of the *Hpa*I position 3592. Presence of the *Hpa*I site (thymine in position 3594) would warrant further detailed investigation.

Anthropologists have suggested three (Birdsell 1977) migrations into Australia although Thorne (1981) suggests that two populations were already present in the Pleistocene. The earliest dates for human fossils in Australia are estimated at 60,000BP (Roberts *et al.* 1990; cited Cavalli-Sforza *et al.* 1994: p.344), although the Multiregional proponents Thorne and Wolpoff (1992); cited Cavalli-Sforza *et al.* 1994: p.349) have suggested the robust remains from Kow swamp, and two other skulls are descendants of *Homo erectus*.

MtDNA RFLP analysis showed Australian and New Guinean groups have lower heterozygosities (Engels, 1981 cited Stoneking and Wilson, 1989) and a higher frequencies of private polymorphism (Whittam *et al.* 1986 cited Stoneking and Wilson 1989) than Europeans, Africans and Asians respectively. The RFLP analysis on Papuans with populations from the rest of the world showed 14 clusters of Papuans scattered throughout the tree, although the topology of the tree showed a group of exclusively Africans one side of the root (Stoneking *et* *al.* 1986). Despite these initial results further investigation should continue. Cluster-VI is present in at least one Australian Aboriginal individual number 49 from Vigilant *et al.* (1991). Cluster-VI is the most diverse group of all the clusters and its presence in Australia suggests an early migration out of Africa of *Homo sapiens sapiens*. Initial results from the HVR-I show the Australian Aboriginals to have relatively divergent DNA (D. Betty *pers. comm.*).

8.3.4 Dating/Calibration

Precise dating is critical to resolve many issues of human origin and evolution. Although some attempts to determine the coalescence time of human mtDNA are based on intraspecific calibration (Stoneking *et al.* 1992; Forster *et al. in prep.*), most studies that have attempted to calibrate the molecular clock, are based on the date of 13 million years ago for the orang-utan split from the other great apes. This date has its own error-value that is usually not taken into account when calibrating.

The dates for most recent common ancestor for modern humans are surprisingly consistent from 117,000-280,000 years ago (without a 95% confidence interval) (Gibbons, 1995). These dates are generated from diverse datasets including: RFLP, control region, COII and the entire genome from the mtDNA (Horai *et al.* 1995); protein polymorphisms (Nei and Roychoudhury, 1993); and microsatellites (Bowcock *et al.* 1991a).

If the calibrations are slower by an order of magnitude, instead of observing modern human genetic variation, one may be observing *Homo erectus* variation. The out of Africa hypothesis would be referring to expansion of *Homo erectus*, rather than *Homo sapiens sapiens* out of Africa. The wide geographic spread of *Homo erectus* would ensure (based on the assumptions presented in Figure 21) the most recent common ancestor for modern humans would predate the movement of *Homo erectus* out of Africa. If the dating supported a common mitochondrial ancestor one million years old, then the Multiregional hypothesis would be consistent with the mitochondrial record. This is unlikely, if the calibration are slower by an order of magnitude, then the divergence dates of the great apes would become unrealistic. Based on these assumptions there is no evidence in the mtDNA genetic record for *Homo erectus*.

If the calibration was slower by a half, then the out of Africa expansion may have been *Homo sapiens*, rather than *Homo sapiens sapiens*. This would be consistent with the palaeontological information regarding the presence of *Homo sapiens* out of Africa 300,000 years ago and would not require another replacement event. According to this calibration there would have been two expansions out of Africa *Homo erectus* and *Homo sapiens* in which *Homo sapiens* transformed into modern humans in a manner similar to the Multiregionalists hypothesise.

The possibility must be entertained that the calibration was fast and therefore the dates are earlier than those hypothesised by the proponents of Out of Africa. Under this scenario the expansion event may correlate dates of the modern tools.

More data will be needed to establish weights for positions within the control region and a better understanding for the mechanisms behind the substitutions within the HVR-I, adding to the accuracy of calibration.

8.3.5 Phylogeny

Phylogenetic arguments can be used to approach the origins of modern humans. Penny *et al.* (1995) demonstrated that a group of forty-nine exclusively Africans (from Vigilant *et al.*'s (1991) survey) lay closest to the root of the modern human tree. The results from the median network method indicate that Cluster-I is a derivative of African specific clusters. As Cluster-I is the central cluster (Excoffier *et al.* 1992) of almost all non-Africans, this indicates that Africa was the geographical origin for modern human mtDNA.

8.4 Future Directions

8.4.1 HVR-I

There are still many questions to be addressed with these data. The published HVR-I sequences now exceed 2,000 and with the discovery of HVR-I fragments inserted in the nuclear genome (Zischler *et al.* 1995) as well as great ape HVR-I sequences being available (e.g. Morin *et al.* 1994; Horai *et al.* 1995), one can begin to build up a stringent model of the evolutionary processes involved in this region. A better understanding of the process acting in this region can refine the calibration. Also sequences from unsampled or poorly sampled regions such as Australia may provide data that will settle the Multiregional versus Out of Africa debate.

8.4.2 Middle Eastern Contribution to the African genepool

The original question of this thesis concerning the origins of the pastoralists, could not be resolved using mtDNA data. There is some genetic evidence supporting a Middle Eastern contribution (see 6.5.3) in African groups. Cluster-I exists in higher frequencies in the pastoralists (41%) as opposed to the agriculturists (20%) in West and East Africa, although geography may be influencing these figures. Also phenotypically, linguistically and culturally there are affiliations between the pastoralists of Africa and the Middle East. In their oral tradition, almost all pastoral groups surveyed suggest a Middle Eastern origin, though some of these groups are Moslem and may seek such an origin. It is possible that there would be more Middle Eastern contribution to the genepool from males. This is an interesting question to address, with reference populations from the Middle East and Africa. To approach the question of male contribution, microsatellite loci on the non recombining part of the Y chromosome should be examined.

The practice of relying on cattle as a basis of economy is a relatively recent phenomenon in African History (see Section 2.8.5). Dan Bradley's group have investigated African cattle in detail and have seen that, mitochondrially, they are of the relatively homogeneous *Bos taurus* type, irrespective of phenotype (*Bos indicus* or *Bos taurus* (MacHugh *et al. in prep.).* In terms of nuclear markers there is a strong cline from east to west of Zebu types *Bos indicus*, also irrespective of the phenotype. In terms of Y markers all cattle irrespective of phenotype have the *Bos indicus* type (Bradley *et al.* 1994, MacHugh *et al. in prep.*). This supports introgression of the *Bos indicus* type through the male line.

In some African pastoral groups, the young men travel with the male animals large distances in search of grazing. The Fulbe of Mali, for instance, travel 1000 km to Mauritania and back each year. The large distances travelled by male animals may account for the pattern of introgression of the Zebu type of Y chromosome in Africa. The milk cows and calves stay close to the villages with the women, children and older people. Simplistically, one may predict a similar pattern of Y introgression for humans. The introgression in humans (as in the cattle) would be a relatively recent phenomenon. It would be interesting to look at the groups involved in this mtDNA study and with reference Middle Eastern groups to determine the amount of introgression of Middle Eastern Y types. Preliminary work showed high diversity in the Y-chromosome of the African samples (data not shown).

To investigate further relationships within Africa and Africa's relationship to the rest of the world, it would be useful to look at several microsatellite loci for linkage analysis, such as demonstrated by Bowcock *et al.* (1994), or to undertake a linkage study similar to Tishkoff *et al.* (1996). However, while there are many microsatellite loci available for study, the mutation mechanisms are poorly understood, so dating based on microsatellite data are still not reliable (Nei, 1995).

8.5 Conclusion

This thesis primarily examined HVR-I data with respect to the hypotheses concerning African diversity and human origins. The African mtDNA dataset was enhanced by the addition of East and West African sequences, providing a diverse sample geographically, linguistically and culturally. MtDNA has proved to be highly informative in the study of the evolution and dispersal of *Homo sapiens sapiens.* The combination of pairwise distances comparisons, phylogenetic trees, and median networks, allows much more information to be extracted from the sequences than any single technique by itself.

The median networks method identified clusters within Africa. The expansion event of the clusters could be dated and the geographic origin of the cluster could be determined by the frequency and diversity of these clusters in different geographical areas.

The mtDNA sequences within an Africa ethnic group are comprised of a composite of clusters. The composite nature illustrates, together with the observation of recent admixture, that each ethnic group does not have a unique common ancestor to the group. Rather the data is consistent with a fission fusion process which has given rise to today's ethnic groups. Clusters specific to geographic origin and language group contribute to the diverse mitochondrial nature of an ethnic group. Treating ethnic groups as a single entity, such as a taxa on a phylogenetic tree, is not an appropriate way to represent the history of a group. In addition to this the regional differentiation which is observed within Africa argues against treating Africa as a genetically homogenous unit.

Africa has a higher mitochondrial diversity than has been previously reported, which is exemplified by the East African group the Turkana. The Turkana have a higher diversity than Vigilant *et al.* (1981) dataset, which includes Africans.

Both the mitochondrial sequence information and CD4 locus support an African origin for modern humans. There is no evidence that *Homo erectus* has contributed to the genepool outside of Africa, assuming that the calibrations used are of the right magnitude. However, earlier *Homo sapiens* expansions out of Africa may have contributed to today's genepool, as suggested by the presence of a Cluster-VI group in Australia. What we may be observing if the calibrations are two times faster, is evidence of the *Homo sapiens* (known as archaic humans) as contributing to the modern day genepool. The regional continuity as observed by some palaeontologists, may be consistent with the mtDNA data, although the degree of replacement appears high (>95%). With this amount of replacement one would predict the earlier phenetic traits to be too diluted to be evident.

Future work on existing HVR-I sequences should allow more detailed understanding of the substitution patterns in this region, refining the calibration. Also additional sequences from specific Asian, as well as a detailed survey of Australian populations may provide further insight into the early migrations of humans.

Complex models need to be developed that test various scenarios extending the Out of Africa and Multiregional scenarios. These could allow for selective sweeps, multiple expansion events and population declines. It is important to acknowledge that not one event accounted for the distribution of the world's mitochondrial types, but that it was probably a multifactorial process. The Out of Africa hypothesis may best summarise the evolution of humans, but other processes may be acting and be expressed in different data sets such as the palaeontological record.

-

To follow this project from the inspiration, research, design, travel, sample collection, laboratory work and analysis - and to discover that the key to humanness lies in Africa - is a journey home in itself.

Somewhere the Sky touches the Earth,

and the name of that place is the End

-a Kamba saying (Lindblom, 1920)

9. References

Adamu, M.,
The Hausa and their neighbours. In *General History of Africa II* Moktar, G. (Ed). California: UNESCO: Heinemann pp. 226-230, 1981.

Ajayi, J.F.A. and Crowder, M.,
Historical Atlas of Africa. London: Longman , 1985.

Ammerman, A.J. and Cavalli-Sforza, L.L.,
Farming in Europe. *Man* **6**:674-688, 1984.

Andah, B.W.,
Iron Age beginnings in West Africa: reflections and suggestions. *West Afr J Arch* **9**:125-141, 1979.

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H. and Young, I.G.
Sequence and organisation of the human mitochondrial genome. *Nature* **290**:457-465, 1981.

Aoki, K. and Shida, M.,
A Monte Carlo simulation study of coalescence times in a successive colonisation model with migration. In *Prehistoric Mongoloid dispersals*, T Akazawa *et al.* (Ed) Oxford: Oxford University Press , 1993.

Avise, J.C., Neigel, J.E. and Arnold, J.,
Demographic influences on Mitochondrial DNA Lineage Survivorship in Animal Populations. *Journal of Molecular Evolution* **20**:99-105, 1984.

Avise, J.C.,
Molecular Markers, Natural History and Evolution. New York: Chapman & Hall, 1994.

Bandelt, H-J., Forster, P., Sykes, B.C. and Richards, M.B.,
Mitochondrial Portraits of Human Populations Using Median Networks. *Genetics* **141**:743-753, 1995.

Banks, K.M.,
Climates, cultures and cattle: the Holocene archaeology of the eastern Sahara. Dallas: Southern Methodist University , 1984.

Barbujani, G.,
What do Languages tell us about Human Microevolution? *Trends Ecol Evol* **6**(5):151-156, 1991.

Barker, G.,
Review of Archaeology and Language by Renfrew, C. *Current Anthropology* **29**(3):448-449, 1988.

Barth, F.,
Ethnic groups and boundaries. Boston: Little Brown , 1969.

Batista, O., Kolman, C.J. and Bermingham, E.,
Mitochondrial DNA diversity in the Kuna Amerindians. *Hum Mol Genet* **4**(5):921-929, 1995.

Bendall, K.E. and Sykes,B.C.,
Length heteroplasmy in the first hypervariable segment of the human mtDNA control region. *Am J Hum Genet* **57**:248-256, 1995.

Betranpetit, J., Sala, J., Calafell, F., Underhill, P.A., Moral, P and Comas, D.,
Human mitochondrial DNA variation and the origin of the Basques. *Ann Hum Genet* **59**:63-81, 1995.

Binford, L.R.,
Faunal Remains from Klasies River Mouth. New York: Academic Press , 1984.

Birdsell, J.B.,
The recalibration of a paradigm for the first peopling of Greater Australia. In *Sundra and Sahul, Prehistoric Studies in Southeast Asia, Melanesian and Australia* (Eds) Allen,J., Golson, J. and Jones, R. London: Academic Press, pp. 113-167, 1977.

Blench, R.,
Recent developments in African language classification of African languages. In *The Archaeology of Africa: Food, metals and towns.* Shaw, T., Sinclair, P., Andah, B and Okpoko, A. (Eds) London: Routledge pp. 126-138, 1993.

Bonné-Tamir, B., Johnson, M.J., Natali, A., Wallace, D.C., Cavalli-Sforza, L.L.,
Mitochondrial types in two Israeli populations: a comparative study at the DNA level. *Am J Hum Genet* **38**:341-351, 1986.

Bowcock, A.M., Herbert, J.M., Mountain, J.L., Kidd, J.R., Rogers, J., Kidd, K.K. and Cavalli-Sforza, L.L.,
Study of an additional 58 DNA markers in five human populations from four continents. *Gene Geography* **5**:151-173, 1991a.

Bowcock, A.M., Kidd, J.R., Mountain, J.L., Herbert, J.M., Carotenuto, L., Kidd, K.K. and Cavalli-Sforza, L.L.,
Drift, admixture, and selection in human evolution: A study with DNA polymorphisms. *Proc Natl Acad Sci USA* **88**:839-843, 1991b.

Bowcock, A.M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J.R. and Cavalli-Sforza, L.L.,
High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368**:455-457, 1994.

Bradley, D.G., MacHugh, D.E., Loftus, R.T., Sow, R.S., Hoste, C.H. and Cunningham, E.P.,
Zebu-taurine variation in Y chromosomal DNA: a sensitive assay for genetic introgression in West African trypanotolerant cattle populations. *Animal Genetics* **25**:7-12, 1994.

Bräuer, G.,
Evolution of Modern Humans: A Comparison of the African and non-African Evidence. In *The Human Revolution* (Eds) Mellars, P. and Stringer, C. Edinburgh: Edinburgh University Press pp123-154, 1989.

Bräuer, G.,
A craniological approach to the origin of anatomically modern *Homo sapiens* in Africa and implications for the appearance of modern Europeans. In *The Origins of Modern Humans: a World Survey of the Fossil Evidence.*(Eds) Smith,F.H. and Spencer,F. New York: Alan R Liss pp.327 410, 1984a.

Bräuer, G.,
The 'Afro-European sapiens hypothesis' and hominid evolution in East Africa during the late Middle and Upper Pleistocene. *Courier Forschungsinstitut Senckenberg* **69**:145-165, 1984b

Bräuer, G. and Rimbach, K.W.,
Late archaic and modern *Homo sapiens* from Europe, Africa, and Southwest Asia: Craniometric comparisons and phylogenetic implications. *J Hum Evol* **19**:789-807, 1990.

Bräuer, G.,
Africa's place in the evolution of *Homo sapiens*. In *Continuity or Replacement: Controversies in Homo sapiens evolution.* (Eds) Bräuer, G and Smith, F.H., Rotterdam: Balkema 83-98, 1992.

Bräuer, G., Deacon, H.J. and Zipfel, F.,
Comment on the new maxillary finds from Klasies River, South Africa. *J Hum Evol* **23**:419-422, 1992.

Brooks, A.S., Helgren, D.M., Cramer, J.S., Franklin, A., Hornyak, W., Keating, J.M., Klein, R.G., Rink, W.J., Schwarcz, H., Smith, J.M.L., Stewart, K., Todd, N.E., Verniers, J. and Yellen, J.E.,
Dating and Context of Three Middle Stone Age Sites with Bone Points in the Upper Semliki Valley, Zaire. *Science* **268**:548-553, 1995.

Brown, W.M., George, M. and Wilson, A.C.,
Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci USA* **76**:1967-1971, 1979.

Brown, W.M., Prager, E.M., Wang, A. and Wilson, A.C.,
Mitochondrial DNA sequences of primates: tempo and mode of evolution. *J Mol Evol* **18**:225-239., 1982.

Cabot, J.,
ESEE (Eyeball Sequence Editor) Version 104, 1993.

Cann, R.L. and Wilson, A.C.,
Length mutations in human mitochondrial DNA. *Genetics* **104**:699-711, 1983.

Cann, R.L., Stoneking, M. and Wilson, A.C.,
Mitochondrial DNA and human evolution. *Nature* **325**:31-36, 1987.

Cavalli-Sforza, L.L., Piazza, A., Menozzi, P. and Mountain, J.,
Reconstruction of human evolution: Bringing together genetic, archaeological and linguistic data. *Proc Natl Acad Sci USA* **85**:6002-6006, 1988.

Cavalli-Sforza, L.L.,
Genetic and linguistic evolution. *Science* **244**:1128-1129, 1989.

Cavalli-Sforza, L.L., Minch, E. and Mountain, J.L.,
Coevolution of Genes and languages revisited. *Proc Natl Acad Sci USA* **89**:5602-5624, 1992.

Cavalli-Sforza, L.L., Menozzi, P. and Piazza, A.,
The History and Geography of Human Genes. Princeton: Princeton University Press , 1994.

Chang, D.D. and Clayton, D.A.,
Precise identification of individual promoters for transcription of each strand of human mitochondrial DNA. *Cell* **36**:635-643, 1984.

Chen, L.Z., Easteal, S., Board, P.G., Kirk, R.L.,
Evolution of beta-globin haplotypes in human populations. *Mol Biol Evol* 7(5):423-437, 1990.

Chen, T.M., Tiemei, C., Zhang, Y.Y. and Yinyun,
Palaeolithic chronology and possibly coexistence of *Homo erectus* and *Homo sapiens* in China. *World Archaeol* **23**:147-154, 1991.

Chen, Y-S., Torroni, A., Excoffier, L., Santachiara-Benerecetti, A.S. and Wallace, D.C.,
Analysis of mtDNA Variation in African Populations Reveals the Most Ancient of All Human Continent-Specific Haplogroups. *Am J Hum Genet* **57**:133-149, 1995.

Clark, J.D and Brandt, S.A., (Eds)
From Hunters to Farmers: The Causes and Consequences of Food Production in Africa. Berkeley: University of California Press , 1984.

Clark, J.D.,
Africa from the appearance of *Homo sapiens* to the beginnings of food production. In *History of Humanity. Volume I: Prehistory and the beginnings of civilisation.* de Laet, S.J., (Ed) London: Routledge pp. 191-206, 1994.

Cohen, R.,
The Kanuri of Bornu. New York: Holt, Rinehart & Winston , 1967.

Coles, C. and Mack, B.,
Hausa Women in the Twentieth Century. University of Wisconsin Press , 1991.

Corballis, M.C.,
On the evolution of language and generativity. *Cognition* **44**:197-226, 1992.

Dalby, D.,
Language and History in Africa. New York: Africana, 1970

Davidson, B.,
The Growth of African Civilisation: A History of West Africa 1000-1800. London: Longman , 1965.

Day, W.H.E.,
Analysis of quartet dissimilarity measures between undirected phylogenetics. *Syst Zool* **35**:325-333, 1986.

De Maret, P.,
The Iron Age in the West and South. In *The Archaeology of Central Africa*, F von Noten(Ed) Graz: Akademisches Druk pp. 77-96, 1982.

Deka, R., Jin, L., Shriver, M.D., Yu, L.M., DeCroo, S., Hundrieser, J., Bunker, C.H., Ferrell, R.E. and Chakraborty, R.,
Population Genetics of Dinucleotide $(dC\text{-}dA)_n$:$((dG\text{-}dT)_n$ Polymorphisms in World Populations. *Am J Hum Genet* **56**:461-474, 1995.

Denaro, M., Johnson, M.J., Chen, K.H., Wilmsen, E., Cavalli-Sforza, L.L. and Wallace, D.C.,
Ethnic variation in *Hpa*I endonuclease cleavage patterns of human mitochondrial DNA. *Proc Natl Acad Sci USA* **78**(9):5768-5772, 1981.

Di Rienzo, A. and Wilson, A.C.,
Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proc Natl Acad Sci USA* **88**:1597-1601, 1991.

Di Rienzo, A., Peterson, A.C., Garza, J.C., Valdes, A.M., Slatkin, M. and Freimer, N.B.,
Mutational processes of simple-sequence repeat loci in human populations. *Proc Natl Acad Sci USA* **91**:3166-3170, 1994.

Diamond, J.M.,
Genes and the tower of Babel. *Nature* **336**:, 1988.

Dietz, T.,
Pastoralists in dire straits. *Nederlandse Geografishe Studies* **49**:, 1987.

Dorit, R.L., Akashi, W. and Gilbert, W.,
Absence of Polymorphism at the *ZFY* Locus on the Human Y Chromosome. *Science* **268**:1183, 1995.

Eades, J.S.,
The Yoruba Today. Cambridge: Cambridge University Press, 1980.

Edmonson, M.S.,
Neolithic Diffusion Rates. *Current Anthropology* **2**:71-102, 1961.

Ehret, C.,
Ethiopians and East Africans: the problem of contact. Nairobi: East African Publishing House, 1974.

Ehret, C.,
On the antiquity of agriculture in Ethiopia. *Journal of African History* **20**:161-177, 1979.

Ehret, C.,
Historical/Linguistic Evidence for Early African Food Production In *From Hunters to Farmers: The Cause and Consequence of Food Production in Africa* (Eds) Clark, J.D. and Brandt, S. Berkeley: University of California Press , 1984.

Ehret, C.,
Nilo-Saharans and the Saharan-Sudanese Neolithic. In *The Archaeology of Africa: Food, metals and towns*. Shaw, T., Sinclalr, P., Andah, B and Okpoko, A. (Eds) London: Routledge pp. 104-125, 1993.

Ellen, R.F.,
Ethics in relation to informants, the profession and governments. In *Ethnographic research: A guide to general conduct* R,F Ellen (Ed) London: Academic Press , 1984.

Elmi, A.A.,
Livestock production in Somalia with special emphasis on camels. *Nomadic Peoples* **29**:87-103, 1991.

Excoffier, l., Harding, R.M., Sokal, R.R., Pellegrini, B and Sanchaz-Mazas, A.,
Spatial Differentiation of RH and GM Haplotype Frequencies in Sub-Saharan Africa and Its Relation to Linguistic Affinities. *Hum Biol* **63**(3):273-307, 1991.

Excoffier, L., Smouse, P.E. and Quattro, J.M.,
Analysis of Molecular Variance Inferred From Metric Distances Among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data. *Genetics* **131**:479-491, 1992.

Felsenstein, J.,
PHYLIP, Version 3.5. University of Seattle, Washington, 1990.

Ferris, S.D., Wilson, A.C. and Brown, W.M.,
Evolutionary tree for apes and humans based on cleavage maps of mitochondrial DNA. *Proc Natl Acad Sci USA* **78**:2432-2436, 1981.

Fisherman, J.A.,
Language and Ethnicity. In *Language, Ethnicity and Intergroup Relations.* Giles, M (Ed) London , 1977.

Foley, R.A.,
The silence of the past. *Nature* **353**:114-115, 1991.

Forster,P., Harding,R., Torroni, R and Bandelt, H.-J.,
Native American mtDNA. *in prep.*

Frayer, D.W., Wolpoff, M.H.,
Theories of modern human origins: The palaeontological test. *Am Anthro* **95**:14-50, 1993.

Frayer, D.W., Wolpoff, M.H., Thorne, A.G., Smith, F.H. and Pope, G.G.,
Getting it Straight. *Am Anthro* **96**:424-437, 1994.

Gabunia, L. and Vekua, A.,
A Plio-Pleistocene hominid from Dmanisi, East Georgia, Caucasus. *Nature* **373**:509-512, 1995.

Gibbons, A.,
Mitochondrial Eve: Wounded, but not dead. *Science* **257**:873-875, 1992.

Gibbons, A.,
Pleistocene Population Explosions. *Science* **262**:27-28, 1993.

Gibbons, A.,
Out of Africa-at Last? *Science* **267**:1272-1273, 1995.

Giles, R.E., Blanc, H., Cann, H.M. and Wallace, D.C.,
Maternal inheritance of human mitochondrial DNA. *Proc Natl Acad Sci USA* **77**:6715-6719, 1980.

Gillis, A.M.,
Getting a picture of human diversity. *BioScience* **44**(1):8-10, 1994.

Goldstein, D.B., Ruiz-Linares, A., Cavalli-Sforza, L.L. and Feldman, M.W.,
Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc Natl Acad Sci USA* **92**:6723-6727, 1995.

Gould, S.J.,
Mismeasure of Man. New York: Norton, 1981.

Graven, L., Passarino, G., Semino, O., Boursot, P., Santachiara-Benerecetti, S., Langaney, A. and Excoffier, L.,
Evolutionary Correlation between Control Region Sequence and Restriction Polymorphisms in the Mitochondrial Genome of a Large Senegalese Mandenka Sample. *Mol Biol Evol* **12**(2):334-345, 1995.

Greenberg, B.D., Newbold, J.E. and Sugino, A.,
Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. *Gene* **21**:33-49, 1983.

Greenberg, J.H.,
The Languages of Africa. Bloomington: Indiana University Press , 1963.

Greenberg, J.H.,
African Linguistic Classification. In *Methodology and African Prehistory* Ki-Zerbo, J. (Ed*),* Berkeley: University of California Press , 1981.

Grimes, B.F.,
Ethnologue: Languages of the World. Dallas: Summer Institute of Linguistics , 1988.

Grove, A.T.,
Africa's climate in the Holocene. In *The Archaeology of Africa: Food, metals and towns.* Shaw, T., Sinclair, P., Andah, B and Okpoko, A. (Eds) London: Routledge pp. 32-42, 1993.

Gulliver, P.H.,
The Family Herds: A study of two pastoral tribes in East Africa the Jie and Turkana. London: Routledge & Kegan, 1955.

Hampâté Bâ, A.,
The living tradition. In *Methodology and African Prehistory* Ki Zerbo, J. (Ed), UNESCO California: James Currey, 1990.

Harlan, R.R.,
Tropical African cereals. In *The Archaeology of Africa: Food, metals and towns.* Shaw, T., Sinclair, P., Andah, B and Okpoko, A. (Eds) London: Routledge pp. 53-60, 1993.

Harpending, H.C., Sherry, S.T., Rogers, A.R. and Stoneking, M.,
The genetic structure of ancient human populations. *Current Anthropology* **34**:483-496, 1993.

Harpending, H.C.,
Signature of Ancient Population Growth in a Low-Resolution Mitochondrial DNA Mismatch Distribution. *Hum Biol* **66**(4):591-600, 1994.

Hasegawa, M., Di Rienzo, A. and Wilson, A.C.,
Toward a more accurate time scale for the human mitochondrial DNA tree. *J Mol Evol* **37**:347-354, 1993.

Hedges, S.B., Kumar, S., Tamura, K. and Stoneking, M.,
Human Origins and the analysis of mitochondrial DNA sequences. *Science* **255**:737-739, 1992.

Hendrik, P.W.,
Genetics of Populations. Boston: Jones and Bartlett, 1985.

Hendy, M.D., Foulds, L.R. and Penny, D.,
Proving phylogenetic trees minimal with 1-clustering and set partitioning. *Math BioScience* **51**:71-88, 1980

Herren, U.,
The people of Mukogodo Division Laikipia District. Bern: Laikipia Reports 9:1-39, 1987.

Herren, U.,
Socioeconomic Strategies of Pastoral Maasai Households in Mukogodo, Kenya. Ph.D. Thesis University of Bern , 1991.

Hiernaux, J.,
The People of Africa. London: Weidenfeld & Nicholson , 1975.

Hixson, J.E. and Clayton, D.A.,
Initiation of transcription form each of the two human mitochondrial promoters requires unique nucleotides at transcriptional start sites. *Proc Natl Acad Sci USA* **82**:2660-2664, 1985.

Horai, S. and Hayasaki, K.,
Intraspecific Nucleotide Sequence Differences in the Major Noncoding region of Human Mitochondrial DNA. *Am J Hum Genet* **46**:828-842, 1990.

Horai, S., Kondo, R., Nakagawa-Hattori, Y., Hayashi, S., Sonoda, S. and Tajima, K.,
Peopling of Americas, Founded by Four Major Lineages of Mitochondrial DNA. *Mol Biol Evol* **10**(1):23-47, 1993.

Horai, S., Hayasaki, K., Kondo, R., Tsugane, K and Takahata, N.,
Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proc Natl Acad Sci USA* **92**:532-536, 1995.

Hudson, R.R., Boos, D.D. and Kaplan, N.L.,
A Statistical Test for Detecting Geographic Subdivision. *Mol Biol Evol* **9**:138-151, 1992.

Hudson,R.R.,
Gene genealogies and the coalenscence process. *Oxford Surveys in Evol Biol* **7**:1-44, 1990.

Huffman, T.N.,
Ceramics, settlements and Late Iron Age migrations. *African Archaeological Reviews* **7**:155-182, 1989.

Jacquard, A.,
Distances généalogiques et distances génétiques. *Cahiers d'Anthropologie et d'Ecologie Humaine:Paris* , 1974.

Johnson, M.J., Wallace, D.C., Ferris, S.D., Rattazzi, M.C. and Cavalli-Sforza, L.L.,
Radiation of human mitochondrial DNA types analyzed by restriction endonuclease cleavage patterns. *J Mol Evol* **19**:255-271, 1983.

Jorde, L., Watkins, M., Bamshad *et al.*
Worldwide Genetic variation in tetranucleotide repeat polymorphisms, restriction site polymorphisms, and mitochondrial DNA sequence. *Am J Hum Genet* **55**:A154, 1994.

Jorde, L.B., Bamshad, M.J., Watkins, W.S., Zenger, R., Fraley, A.E., Krakowiak, P.A., Carpenter, K.D., Soodyall, H., Jenkins, T. and Rogers, A.R.,
Origins and Affinities of Modern Humans: A Comparison of Mitochondrial and Nuclear Genetic Data. *Am J Hum Genet* **57**:523-538, 1995.

Kaneda, H., Hayashi, J.I., Takahama, S., Taya, C., Lindahl, K.F., Yonehawa, H.,
Elimination of paternal mitochondrial DNA in intraspecific crosses during early mouse embryogenesis. *Proc Natl Acad Sci USA* **92**(10):4542-4546, 1995.

Ki-Zerbo, J.,
Editorial note: theories on the 'races' and history of Africa. In *Methodology and African Prehistory* Ki-Zerbo, J (Ed), Berkeley: University of California Press , 1981.

Ki-Zerbo, J.,
Die Geschichte Schwarz-Afrikas. *Fischer* , 1993.

Kimura, M.,
The neutral theory of molecular evolution. Cambridge: Cambridge University Press , 1983.

Kiriama, H.O.,
The iron-using communities in Kenya. In *The Archaeology of Africa: Food, metals and towns.* Shaw, T., Sinclair, P., Andah, B and Okpoko, A. (Eds) London: Routledge pp.484-498, 1993.

Kolman, C.J., Bermingham, E., Cooke, R., Ward, R.H., Arias, T.D. and Guionneau-Sinclair, F.,
Reduced mtDNA diversity in the Ngöbe Amerindians. *Genetics* **140**:275-283, 1995.

Kornberg, A. and Baker, T.A.,
DNA replication (2nd ed) New York: Freeman and Company , 1992.

Lee, R.B.,
The !Kung San: Men Women and Work in a Foraging Society. Cambridge: Cambridge University Press , 1979.

Lewis, I.M.,
The Northern Pastoral Somali of the Horn in Peoples of Africa. Gibbs. J.R.,(Ed) New York: Holt, Rinehart and Winston pp321-326, 1965.

Li, W.H. and Sadler, L.A.,
Low nucleotide diversity in man. *Genetics* **129**:513-523, 1991.

Li, W.H.,
Distribution of nucleotide differences between two randomly chosen cistrons in a finite population. *Genetics* **85**:331-337, 1977.

Lindblom, G.,
The Akamba in British East Africa. Uppsala , 1920.

Lock, M.,
Interrogating the Human Diversity Genome Project. *Soc Sci Med* **39**(5):603-606, 1994.

Lucas, J.O.,
The Religion of the Yoruba. Lagos , 1948.

Lynch, M.,
Phylogenetic hypotheses under the assumption of neutral quantitative-genetic variation. *Evolution* **43**:1-17, 1989.

MacHugh, D.E., Bradley, D.G., Loftus, R., Shriver, M.D. and Cunningham, P.,
Cattle Evolution, Domestication and Gene Flow Revealed with Polymorphic Microsatellites. *in prep.*

Maddison, D.R., Ruvolo, M. and Swofford, D.L.,
Geographic origins of human mitochondrial DNA: Phylogenetic evidence from control region sequences. *Syst Biol* **41**:111-124, 1992.

Maley, J.,
The climatic and vegetational history of the equatorial regions of Africa during the upper Quaternary. In *The Archaeology of Africa: Food, metals and towns.* Shaw, T., Sinclair, P., Andah, B and Okpoko, A. (Eds) London: Routledge pp. 43-52, 1993.

McCabe, J.T.,
The Turkana in Kenya. In *Nomads* (Ed*) P.J. Carmichael.* London: Collins and Brown , 1991.

McIntosh, S.K. and McIntosh, R.J.,
Recent archaeological research and dates from West Africa. *Journal of African History* **27**:413-442, 1986.

Moktar, G.,
Ancient Civilizations of Africa. Berkeley: University of California Press , 1981.

Morin, P.A., Moore, J.J., Chakraborty, R.,
Kin Selection, social structure, gene flow, and the evolution of chimpanzees. *Science* **265**:1193-1201, 1994.

Mountain, J., Hebert, J.M., Bhattacharyya, S., Underhill, P.A., Ottolenghi, C., Gadgil, M., Cavalli-Sforza, L.L.,
Demographic History of India and mtDNA-Sequence Diversity. *Am J Hum Genet* **56**:979-992, 1995.

Munson, P.J.,
Africa's Prehistoric Past. In *Africa* P. O'Meara (Ed) Bloomington: Indiana University Press pp. 62-82, 1977.

Murdock, G.P.,
Africa: Its Peoples and their Culture History. New York: McGraw-Hill , 1959.

Murdock, G.P.,
Ethnographic Atlas. Pittsburgh: University of Pittsburgh Press , 1967.

Nee, S., Holmes ,E.C., Rambaut, A. and Harvey, P.,
Inferring population history from molecular phylogenies. *Phil Trans R Soc Lond* **349**:25-31, 1995.

Neel, J.V.,
Physician to the Gene Pool. Genetic Lessons and Other Stories. New York: John Wiley , 1994.

Nei, M. and Rodchoudhury, A.K.,
Evolutionary Relations of human populations on a global scale. *Mol Biol Evol* **10**:927-943, 1993.

Nei, M.,
Genetic support for the out-of-Africa theory of human evolution. *Proc Natl Acad Sci USA* **92**:6720-6722, 1995.

Nei, M.,
Molecular Evolutionary Genetics. New York, Columbia University Press. 1987

Noble, W. and Davidson, I.,
The evolutionary emergence of modern human behavior: language and its archaeology. *Man* **26**:223-254, 1991.

Norris, H.T.,
Indigenous Peoples of the Sahara. In Cloudsley-Thompson, J.L.(Ed) Oxford: Pergamon Press , 1984.

Norris, H.T.,
The Tuaregs: Their Islamic Legacy and Its Diffusion in the Sahel. Wilts: Aris & Phyllips, 1975.

O'Grady, R.T., Goddard, I., Bateman, R.T., Di Michelle, W.D., Funk, V.A., Kress, W.J., Mooi, R. and Cannell, P.F.,
Genes and Tongues. *Science* **243**:1651, 1989.

Okafor, E.E.,
New Evidence on early iron-smelting from southeastern Nigeria. In *The Archaeology of Africa: Food, metals and towns.* Shaw, T., Sinclair, P., Andah, B and Okpoko, A. (Eds) London: Routledge pp. 432-448, 1993.

Olderogge, D.,
Migrations and ethnic and linguistic differentiations. In *Methodology and African Prehistory.* Ki-Zerbo, J (Ed), Berkeley: University of California Press , 1981.

Olivier, R. and Mathew, G.,
History of East Africa. Oxford: Clarendon Press , 1963.

Passarino, G., Semino, O., Modiano, G. and Santachiara-Benerecetti, A.S.,
COII/tRNALys Intergenic 9-bp Deletion and Other mtDNA Markers Clearly Reveal That the Tharus (Southern Nepal) Have Oriental Affinities. *Am J Hum Genet* **53**:609-618, 1993.

Pederson, J. and Woehle, E.,
Bangombi (Baka) In *Hunters and Gatherers (Vol I)* Imgold, T., Riches, D., Woodburn, J (Eds) Oxford: Berg , 1985.

Penny, D.,
Towards a basis for classification: the incompleteness of distance measures, incompatibility analysis and phenetic classification. *J Theoretical Biol* **96**:129-142, 1982

Penny, D. and Hendy, M.D.,
The use of tree comparison metrics. *Syst Zool* **34**:75-82, 1985.

Penny, D., Hendy, M.D. and Steel, M.A.,
Testing the theory of descent. In *Phylogenetic analysis of DNA sequences.* (Eds) M.M Miyamoto and J. Cracraft. New York: Oxford University Press pp.155-183, 1991.

Penny, D., Hendy, M.D. and Steel, M.A.,
Progress with methods for constructing evolutionary trees. *Trends Ecol Evol* **7**:73-79, 1992.

Penny, D., Steel, M., Waddell, P.J. and Hendy, M.D.,
Improved Analyses of Human mtDNA Sequences Support a Recent African Origin for *Homo sapiens. Mol Biol Evol* **12**(5):863-882, 1995.

Piercy, R., Sullivan, K.M., Benson, N. and Gill, P.,
The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. *Int J Leg Med* **106**:85-90, 1993.

Posnansky, M.,
Introduction to the later prehistory of sub-Saharan Africa. In *Methodology and African Prehistory.* Ki Zerbo, J (Ed) Berkeley: University of California Press , 1981.

Pult, I., Sajantila, A., Simanainenen, O., Georgiev, O., Schaffner, W. and Pääbo, S.,
Mitochondrial DNA sequences from Switzerland reveal striking homogeneity of European populations. *Biol Chem Hoppe-Seyler* **375**:837-840, 1994.

Quinn, T.C.,
Population migration and the spread of types 1 and 2 human immunodeficiency viruses. *Proc Natl Acad Sci USA* **91**:2407-2414, 1994.

Rapacz, J., Chen, L., Bulter-Brunner, E., Wu, M-J., Hasler-Rapacz, J.O., Bulter, R. and Schumaker, V.N.
Identification of the ancestral haplotype for apolipoprotein B suggests an African origin of *Homo sapiens sapiens* and traces their subsequent migration to Europe and the Pacific. *Proc Natl Acad Sci USA* **88**:1403-1406, 1991.

Rassman, K., Ebenhöh, P. and Tautz, D.,
Easy Blood. *Fingerprint News* **4**:7, 1989.

Renfrew, C.,
Archaeology and Language: The puzzle of Indo-European Origins. London: Jonathan Cape , 1987.

Renfrew, C.,
Archaeology and Language: The puzzle of Indo-European Origins (Book Review). *Current Anthropology* **29**(3):437-441, 1988.

Renfrew, C.,
Archaeology, Genetics and Linguistic Diversity. *Man* **27**:445-478, 1993.

Robinson, D.F. and Foulds, L.R.,
Comparison of phylogenetic trees. *Math Biosci* **53**:131-147, 1981.

Rogers, A.R. and Harpending, H.,
Population Growth Makes Waves in the Distribution of Pairwise Genetic Differences. *Mol Biol Evol* **9**(3):552-569, 1992.

Rogers, A.R. and Jorde, L.B.,
Genetic Evidence on Modern Human Origins. *Hum Biol* **67**:1-36, 1995.

Routledge, W.S. and Routledge, K.,
With a Prehistoric People: The Akikuyu of British East Africa. London: Edward Arnold , 1910.

Ruhlen, M.,
A Guide to the World's Languages: Volume 1:Classification. London: Edward Arnold 76-124, 1987.

Said, R. and Faure, H.,
Chronological framework: African pluvial and glacial epochs. In *Methodology and African Prehistory* Ki Zerbo, J. (Ed), UNESCO California: James Currey, 1990.

Sajantila, A., Lahermo, P., Anttinen, T., Lukka, M., Sistonen, P., Savontaus, M-L., Aula, P., Beckman, L., Tjanebaerg, L., Gadde-Dahl, T., Issel-Tarver, L., Di Rienzo, A. and Pääbo, S.,
Genes and Languages in Europe: and analysis of mitochondrial lineages. *Genome Research* **1**:, 1995.

Santos, M., Ward, R.H. and Barrantes, R.,
mtDNA Variation in the Chibcha Amerindian Huetar from Costa Rica. *Hum Biol* **66**(6):963-977, 1994.

Scozzari, R., Torroni, A., Semino, O., Sirugo, G., Brega, A. and Santachiara-Benerecetti, A.S.,
Genetic Studies on the Senegalese Population I: Mitochondrial DNA Polymorphisms. *Am J Hum Genet* **43**(4):534-544, 1988.

Scozzari, R., Torroni, A., Semino, O., Cruciani, F., Spedini, G. and Santachiara-Benerecetti, S.A.,
Genetic studies in Cameroon: mitochondrial DNA polymorphisms in Bamileke. *Hum Biol* **66**:1-12, 1994.

Schaffer, M. and Cooper, C.,
Mandinko: The Ethnography of a West African Holy Land. New York: Holt, Rinehart & Winston, 1980.

Semino, O., Torroni, A., Scozzari, R., Brega, A., De Benedictis, G. and Santachiara-Benerecetti, A.S.,
mtDNA polymorphisms in Italy. III. Population data from Sicily: a possible quantitation of maternal African ancestry. *Annals of Hum Genet* **53**(2):193-202, 1989.

Shaw, T.,
Hunters, gatherers and first farmers in West Africa. In *Hunters, gatherers and first farmers beyond Europe: An archaeological survey.* Megaw, J.V.S. (Ed) Leicester: Leicester Press , 1977.

Shaw, C.T.,
The Prehistory of West Africa In *Methodology and African Prehistory* Ki-Zerbo, J (Ed) Berkeley: University of California Press , 1981.

Sherry, S.T., Rogers, A.R., Harpending, H., Soodyall, H., Jenkins, T. and Stoneking, M.,
Mismatch Distributions of mtDNA Reveal Recent Human Population Expansions. *Hum Biol* **66**(5):761-776, 1994.

Shields, G.F., Schmiechen, A.M., Frazier, B.L., Redd, A., Voevoda, M.I., Reed, J.K. and Ward, R.H.,
mtDNA Sequences Suggest a Recent Evolutionary Divergence of Beringian and Northern North American Populations. *Am J Hum Genet* **53**:549-562, 1993.

Sinclair, P.J.J., Shaw, T. and Andah, B.,
Introduction. In *The Archaeology of Africa: Food, metals and towns.* Shaw, T., Sinclair, P., Andah, B and Okpoko, A. (Eds) London: Routledge pp. 1-31, 1993.

Smith, F.H.,
Models and realities in modern human origins: the African fossil evidence. *Phil Trans R Soc Lond* **337**:243-250, 1992.

Sobania, N.W.,
Background History of the Mt Kulal Region of Kenya: UNEP-MAB Integrated Project in Arid Lands. *IPAL Technical Report* **A-2**:221-238, 1979.

Soodyall, H. and Jenkins, T.,
Mitochondrial DNA polymorphisms in Negroid populations from Namibia: new light on the origins of the Dama, Herero and Ambo. *Annals of Hum Biol* **20**(5):477-485, 1992.

Soodyall, H.,
Mitochondrial DNA polymorphisms in Southern African Populations. Ph.D. Thesis, University of Witwatersrand, Johannesburg , 1993.

Spencer, P.,
Nomads in Alliance: Symbiosis and Growth among the Rendille and Samburu of Kenya. London, 1973.

Steel, M.A.,
Distributions of the symmetric difference metric on phylogenetic trees. *J Discr Math* **1**:541-551, 1988.

Steel, M.A. and Penny, D.,
Distribution of tree comparison metrics- Some new results. *Syst Biol* **43**:, 1993.

Stenning, D.J.,
The Pastoral Fulani of Northern Nigeria in Peoples of Africa. New York: Holt, Rinehart and Winston pp. 363-385, 1965.

Stoneking, M., Bhatia, K. and Wilson, A.C.,
Rate of Sequence Divergence Estimated from Restriction Maps of Mitochondrial DNAs from Papua New Guinea. *CSHS on Quant Biol* **LI**:433-439, 1986

Stoneking, M. and Wilson, A.C.,
Mitochondrial DNA. In *The Colonization of the Pacific: A Genetic Trail.* (Eds) Hill, A.V.S. and Serjeantson,S.W. Oxford: Clarendon, 1989.

Stoneking, M., Jorde, L.B., Bhatia, K. and Wilson, A.C.,
Geographic Variation in Human Mitochondrial DNA from Papua New Guinea. *Genetics* **124**:717-733, 1990.

Stoneking, M., Sherry, S.T., Redd, A.J. and Vigilant, L.,
New Approaches to dating suggest a recent age for the human mtDNA ancestor. *Phil Trans R Soc Lond* **337**:167-175, 1992.

Stoneking, M.,
In defence of "Eve"- a response to Templeton's critique. *Am Anthropol* **96**:131-141, 1994.

Stoneking, M.,
DNA and recent human evolution. *Evolutionary Anthropology* **2**:60-73, 1993.

Stride, G.T. and Ifeka, C.,
Peoples and Empires of West Africa: West Africa in History 1000-1800. *Nelson* , 1969.

Stringer, C.,
The Dates of Eden. *Nature* **331**:565-566, 1988.

Stringer, C.B.,
Reconstructing recent human evolution. *Phil Trans R Soc Lond* **337**:217-224, 1992.

Stringer, C. and Bräuer, G.,
Methods, Misreading and Bias. *Am Anthro* **96**:416-424, 1994.

Swisher, C.C., Curtis, G.H., Jacob, T., Getty, A.G., Suprijo, A. and Widiasmoro
Age of the Earliest Known Hominids in Java, Indonesia. *Science* **263**:1118-1121, 1994.

Swofford, D.L.,
When are phylogeny estimates from molecular and morphological data incongruent? In *Phylogenetic analysis of DNA sequences.* (Eds) M.M Miyamoto and J. Cracraft, New York: Oxford University Press pp. 295-333, 1991.

Talbot, P.A.,
The Peoples of Southern Nigeria. Volume I. London , 1926.

Templeton, A.R.,
Human Origins an analysis of mitochondrial DNA sequences. *Science* **255**:737, 1992.

Templeton, A.R.,
The "Eve" Hypothesis: a genetic reanalysis. *American Anthropologist* **95**:51-72, 1993.

Thompson, L.G., Mosley-Thompson, E., Davis, M.E., Davis, M.E., Lin, P.-N., Henderson, K.A., Cole-Dai., Bolzen, J.F., Liu, K.B.,
Late Glacial Stage and Holocene Ice Core Records from Huascarán, Peru. *Science* **269**:46-50, 1995.

Thorne, A.G. and Wolpoff, M.H.,
The Multiregional Evolution of Humans. *Scientific American* **266**:28-33, 1992.

Thorne, A.G.,
The arrival and adaptation of Australian aboriginals. In *Ecological Biogeography of Australia* (Ed) Illies, J. The Hague: Junk, 1981.

Thorne, B.,
"You Still Takin' Notes?" Fieldwork and Problems of Informed Consent. *Social Problems* **27**(3):284-297, 1980.

Tiemei, C., Quang, Y. and En, W.,
Antiquity of *Homo sapiens* in China. *Nature* **368**:55-56, 1994.

Tishkoff, S.A., Dietzsch, E., Speed, W., Pakstis, A.J., Kidd, J.R., Cheung, K., Bonné-Tamir, B., Santachiara-Benerecetti, A.S., Moral, P., Krings, M., Pääbo, S., Watson, E., Risch, N., Jenkins, T. and Kidd, K.K.,
Global Patterns of Linkage Disequilibrium at the CD4 locus and Modern Human Origins. *Science* **271**:1380-1387, 1996.

Tobias, P.V.,
Bushman hunter-gatherers: a study in human ecology. In *Ecological Studies in Southern Africa.* D.H.S. Davies (Ed) The Hague: Junk pp. 67-86, 1964.

Torroni, A., Semino, O., Scozzari, R., Sirugo, G., Spedini, G., Abbas, N., Fellous, M., Santachiara-Benerecetti, A.S.,
Y chromosome DNA polymorphisms in human populations: differences between Caucasoids and Africans detected by 49 and 49f probes. *Ann Hum Genet* **54**(4):287-296, 1990.

Torroni, A., Schurr, T.G., Cabell, M.F., Brown, M.D., Neel, J.V., Larsen, M., Smith, D., Vullo, C.M. and Wallace, D.,
Asian Affinities and Continental Radiation of the Four Founding Native American mtDNAs. *Am J Hum Genet* **53**:563-590, 1993a.

Torroni, A., Sikernik, R.I., Schurr, T.G., Starkovskaya, Y.B., Cabell, M.F., Crawford, M.H., Comuzzie, A.G. and Wallace, D.C.,
mtDNA Variation of Aboriginal Siberians Reveals Distinct Genetic Affinities with Native Americans. *Am J Hum Genet* **53**:591-608, 1993b.

Torroni, A., Lott, M.T., Cabell, M.F., Chen, Y-S., Lavergne, L. and Wallace, D.,
mtDNA and the Origin of Caucasians: Identification of Ancient Caucasian-specific Haplogroups, One of Which Is Prone to a Recurrent Somatic Duplication in the D-Loop Region. *Am J Hum Genet* **55**:760-776, 1994.

Vigilant, L., Pennington, R., Harpending, H., Kocher, T.D. and Wilson, A.C.,
Mitochondrial DNA sequences in single hairs from a southern African population. *Proc Natl Acad Sci USA* **86**:9350-9354, 1989.

Vigilant, L.,
Control region sequences from African populations and the evolution of human mitochondrial DNA. Ph.D. Thesis, University of California, Berkeley , 1990.

Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. and Wilson, A.C.,
African Populations and the Evolution of Human Mitochondrial DNA. *Science* **253**:1503-1507, 1991.

Wakeley, J.,
Substitution Rate Variation Among Sites in Hypervariable Region 1 of Human Mitochondrial DNA. *J Mol Evol* **37**:613-623, 1993.

Walberg, M.W. and Clayton, D.A.,
Sequences and properties of the human KB cell and mouse L cell D-loop regions of mitochondrial DNA. *Nucleic Acids Research* **9**:5411-5421, 1981.

Wallace, D.C.,
Mitochondrial DNA Variation in Human Evolution, Degenerative Disease and Aging. *Am J Hum Genet* **57**:201-228, 1995.

Ward, R.H., Frazier, B.S., Dew, K. and Pääbo, S.,
Extensive Mitochondrial Diversity within a Single Amerindian Tribe. *Proc Natl Acad Sci USA* **88**:8720-8724, 1991.

Ward, R.H., Redd, A., Valencia, D., Frazier, B. and Pääbo, S.,
Genetic and linguistic differentiation in the Americas. *Proc Natl Acad Sci USA* **90**:10663-10667, 1993.

Wax, M.L.,
Paradoxes of "Consent" to the Practice of Fieldwork. *Social Problems* **27**(3):272-283, 1980.

Wendorf, F. and Schild, R.,
The Emergence of Food Production in the Egyptian Sahara. In *From Hunters to Farmers: The causes and consequences of Food Production in Africa.* Clark, J.D. and Brandt, S. (Eds) California: University of California Press , 1984.

Wenke, R.J.,
Patterns in Prehistory : Humankind's first three million years (3rd ed). Oxford: Oxford University Press , 1990.

Williams, W.T. and Clifford, H.T.,
On the comparison of two classifications of the same set of elements. *Taxon* **20**:519-522, 1971.

Wilson, A.C. and Sarich, V.M.,
A molecular time scale for human evolution. *Proc Natl Acad Sci USA* **63**:1088-1093., 1969.

Wilson, A.C., Cann, R.L., Carr, S.M., George, M., Gyllensten, U.B., Helm-Bychowski, K.M., Higuchi, R.G., Palumbi, S.R., Prager, E.M., Sage, R.D. and Stoneking, M.
Mitochondrial DNA and two perspectives on evolutionary genetics. *Biological Journal of the Linnean Society* **26**:375-400, 1985.

Winkler, E-M. and Sokal, R.R.,
A Phenetic Classification of Kenyan Tribes and Subtribes. *Hum Biol* **59**(1):121-145, 1987.

Wolpoff, M.H.,
Multiregional Evolution: The fossil alternative to Eden. In *The Human Revolution : Behavioural and Biological Perspectives on the Origins of Modern Humans.* Mellars, P. and Stringer, C., (Ed), Princeton: Princeton University Press pp. 62-108, 1989.

Wolpoff, M.H.,
In *Continuity or Replacement: controversies Homo sapiens evolution.* G. Bräuer and F.H. Smith (Eds) Rotterdam: Balkema pp. 25-63, 1992.

Wu, X. and Bräuer, G.,
Morphological comparison of archaic *Homo sapiens* crania from China and Africa. *Z Morph Anthrop* **79**(3):241-259, 1993.

Yellen, J.E., Brooks, A.S., Cornelissen, E., Mehlman, M.J. and Stewart, K.,
A Middle Stone Age Worked Bone Industry from Katanda, Upper Semliki Valley, Zaire. *Science* **268**:553-556, 1995.

Zischler, H., Gelsert, H., von Haeseler, A. and Pääbo, S.,
A nuclear 'fossil' of the mitochondrial D-loop and the origin of modern humans. *Nature* **378**:489-492, 1995.

10. Appendices

10.1 Information regarding the Use of Samples Collected in this Survey

Samples remain indefinitely under the guardianship of myself (Elizabeth Watson) and written permission must be sought for any studies which wish to use them. Requests can be directed to **Elizabeth Watson c/- Massey University Ethics Committee, Massey University, Palmerston North, New Zealand**. The samples were collected with the understanding that they will not be used for medical studies.

10.2 Ethnological Information for all Individuals Sampled

This is the information collected for the individuals who participated in the survey. Empty spaces are where information was not available. The country of collection as well as country of origin of the individual is recorded. As this study concentrates on mtDNA, the individuals are coded according to their maternal line.

Key to following pages

The first row of each individual's entry records the ethnic affiliation of the individual and parents and grandparents. The second row records the **Town of Birth.** The third row indicates the **Province of Birth.**

The columns are titled at the of each page. Maternal is abbreviated **M.** and Paternal **P.**. **Country** indicates in which country the sample was collected.

Code corresponds to the number of the tube in which blood was collected. The first letter represents the **Country of Birth** on the maternal line, and the second letter the ethnic group within that country (see Table below).

Code	Country	Population	Code	Country	Population
BF	Burkina Faso	Fulbe	MS	Mali	Songhai
CF	Cameroun	Fulbe	MT	Mali	Tuareg
EA	Ethiopia	Anuak	N?	Nigeria	Misc.
EB	Ethiopia	Borana	NF	Nigeria	Fulbe
ED	Ethiopia	Dinka	NH	Nigeria	Hausa
EN	Ethiopia	Nuer	NI	Nigeria	Igbo
EZ	Ethiopia	Somali	NK	Nigeria	Kanuri
GH	Ghana	Hausa	NS	Nigeria	Shuwa
KB	Kenya	Borana	NT	Nigeria	Yoruba
KC	Kenya	Sakuye	NY	Nigeria	Yoruba
KG	Kenya	Gabbra	RB	Niger	Bozo
KJ	Kenya	Kalenjin	RD	Niger	Djerma
KK	Kenya	Kikuyu	RF	Niger	Fulbe
KL	Kenya	Luo	RH	Niger	Hausa
KM	Kenya	Maasai	RK	Niger	Kanuri
KR	Kenya	Rendille	RL	Niger	Bella
KS	Kenya	Samburu	RS	Niger	Songhai
KT	Kenya	Turkana	RT	Niger	Tuareg
KU	Kenya	Gusii	SN	Sudan	Nuer
KY	Kenya	Luhyia	YF	Benin	Fulbe
KZ	Kenya	Somali	ZB	Somalia	Borana
MF	Mali	Fulbe	ZZ	Somalia	Somali

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country

10.2.1 Anuak

361 EA	Nuer	Nuer	Anuak	Nuer	Nuer	Ethiopia
Town of Birth	Jikao	Jikao	Jikao	Jikao	Jikao	
Prov. of	Ilubabor	Ilubabor	Ilubabor	Ilubabor	Ilubabor	

10.2.2 Babur

714 NU	Babur	Babur	Babur	Babur	Babur	Nigeria
Town of Birth	Maiduguri	Biu	Biu	Biu	Biu	
Province	Borno	Bomo	Bomo	Bomo	Borno	
727 NU	Babur	Babur	Babur	Babur	Babur	Nigeria
Town of Birth	Bui	Wajaffa	Wajaffa	Wajaffa	Wajaffa	
Province	Bomo	Borno	Borno	Borno	Bomo	
735 NU	Babur	Babur	Babur	Babur	Babur	Nigeria
Town of Birth	Biu	Biu	Bui	Beteri	Bui	
Province	Borno	Borno	Bomo	Bauchi	Borno	

10.2.3 Baribar

871 NB	Bariba	Bariba	Bariba	Bariba	Bariba	Nigeria
Town of Birth	Ilesa	Ilesa	Ilesa	Ilesa	Ilesa	
Province	Kwara	Kwara	Kwara	Kwara	Kwara	
875 NB	Baribar	Baribar	Baribar	Baribar	Baribar	Nigeria
Town of Birth	Saki	Gbane	Gbane			
Province	Oyo	Kwara	Kwara	Benin	Benin	

10.2.4 Bella

266 RL	Bella	Bella	Bella	Bella	Bella	Niger
Town of Birth	Balleyara	Balleyara	Balleyara	Balleyara	Akama	
Province	Filingué	Filingué	Filingué	Filingué	Filingué	

10.2.5 Borana

96 EB	Borana	Borana	Borana	Borana	Borana	Kenya
Town of Birth	Damballa	Nagelle				
Province	Marsabit	Ethiopia	Ethiopia	Ethiopia	Ethiopia	
126 ZB	Borana	Borana	Borana	Borana	Borana	Kenya
Town of Birth	Kitale		Islo	Nairobi	Islo	
Province		Somalia	Somalia		Somalia	
174 KB	Borana	Borana	Borana	Borana	Borana	Kenya
Town of Birth	Garba Tulla	Sericho	Moyale	Moyale	Wajir	
Province	Isiolo	Isiolo	Marsabit	Marsabit	Wajir	
176 KB	Borana	Borana	Borana	Borana	Borana	Kenya
Town of Birth	Merti	Wajir	Moyale	Sololo		
Province	Isiolo	Wajir	Moyale		Ethiopia	
210 KB	Borana	Borana	Borana	Borana	Borana	Kenya
Town of Birth	Malkagala	Sericho	Gutatu	Siribde	Marti	
Province	Isiolo	Isiolo	Isiolo	Isiolo	Isiolo	
217 KB	Borana	Borana	Borana	Borana	Borana	Kenya
Town of Birth	Garfasa	Moyore	Merti	Marsabit		
Province	Isiolo	Isiolo	Isiolo	Marsabit		

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
224 KB	Borana	Borana	Borana	Borana	Borana	Kenya
Town of Birth	Mega	Moyale	Moyale	Bubisa	Bubisa	
Province	Ethiopia	Marsabit	Marsabit	Marsabit	Marsabit	
227 KB	Borana	Borana	Borana	Borana	Borana	Kenya
Town of Birth	Isiolo		Sericho			
Province	Isiolo		Isiolo			
230 KB	Borana	Borana	Borana	Borana	Borana	Kenya
Town of Birth	Kina	Sercho	Sericho	Moyale	Moyale	
Province	Isiolo	Isiolo	Isiolo	Moyale	Moyale	
239 KB	Somali	Borana	Borana	Somali	Somali	Kenya
Town of Birth	Isiolo	Merti	Merti	Wajir	Wajir	
Province	Isiolo	Isiolo	Isiolo	Wajir	Wajir	
240 EB	Borana	Borana	Borana	Borana	Borana	Kenya
Town of Birth	Bulesa					
Province	Isiolo	Ethiopia	Ethiopia	Ethiopia	Ethiopia	

10.2.6 Bukusu

32 K?	Bukusu	Bukusu	Bukusu	Bukusu	Luhyia	Kenya
Town of Birth	Ugenya					
Province	Siaya	Busia	Busia	Busia	Busia	

10.2.7 Dinka

374 ED	Nuer	Nuer	Dinka	Nuer	Nuer	Ethiopia
Town of Birth	Jikao	Jikao	Jikao	Jikao	Jikao	
Province	Ilubabor	Ilubabor	Ilubabor	Ilubabor	Ilubabor	
382 ED	Dinka	Dinka	Dinka	Dinka	Dinka	Ethiopia
Town of Birth	Jikao	Jikao	Jikao	Jikao	Jikao	
Province	Ilubabor	Ilubabor	Ilubabor	Ilubabor	Ilubabor	

10.2.8 Fulbe

257 RF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Niger
Town of Birth	Filingué	Abala	Abala	Saudiné	Saudiné	
Province	Filingué	Filingué	Filingué	Filingué	Filingué	
326 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth	Okutu	Okutu		Okutu		
Province	Kwara	Kwara		Kwara		
327 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth	Kontagora	Sokoto	Sokoto	Sokoto	Sokoto	
Province		Sokoto	Sokoto	Sokoto	Sokoto	
328 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth	Gbabe	Sokoto	Sokoto	Sokoto	Sokoto	
Province	Kwara	Sokoto	Sokoto	Sokoto	Sokoto	
329 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth	Bode Sadu					
Province	Kwara	Sokoto	Sokoto	Sokoto	Sokoto	
330 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth	Kaduna					
Province	Kaduna	Sokoto	Sokoto	Sokoto	Sokoto	
331 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth						

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
Province	Kwara	Kwara	Kwara	Kwara	Kwara	
332 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth						
Province	Kwara	Kwara	Kwara	Kwara	Kwara	
333 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth						
Province	Kwara	Kwara	Kwara	Kwara	Kwara	
334 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth						
Province	Oyo	Oyo	Oyo	Oyo	Oyo	
336 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth						
Province	Sokoto	Sokoto	Sokoto	Sokoto	Sokoto	
337 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth						
Province	Sokoto	Sokoto	Sokoto	Sokoto	Sokoto	
338 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth	Saki	Saki	Saki	Saki	Saki	
Province	Oyo	Oyo	Oyo	Oyo	Oyo	
341 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth						
Province	Kano	Kano	Kano	Kano	Kano	
342 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth	Saki	Saki	Saki	Saki	Saki	
Province	Oyo	Oyo	Oyo	Oyo	Oyo	
343 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth						
Province	Kano	Kano	Kano	Kano	Kano	
344 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth						
Province	Katsina	Katsina	Katsina	Katsina	Katsina	
345 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth						
Province	Kano	Kano	Kano	Kano	Kano	
346 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth						
Province	Sokoto	Sokoto	Sokoto	Sokoto	Sokoto	
347 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth						
Province	Sokoto	Sokoto	Sokoto	Sokoto	Sokoto	
348 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth						
Province	Sokoto	Sokoto	Sokoto	Sokoto	Sokoto	
349 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth						
Province	Sokoto	Sokoto	Sokoto	Sokoto	Sokoto	
435 RF	Songhai	Fulbe	Fulbe	Songhai	Songhai	Niger

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
Town of Birth	Niamey	Tillaberi	Tillaberi	Sicasso	Sicasso	
Province						
437 BF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Niger
Town of Birth	Niamey	Dou	Samai	Dou	Dou	
Province						
441 RF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Niger
Town of Birth	Assani	Assani	Assani	Assani	Assani	
Province						
442 RF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Niger
Town of Birth	Boboye	Boboye	Boboye	Bibiye	Boboye	
Province						
444 RF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Niger
Town of Birth	Saguia	Saguia	Saguia	Saguia	Saguia	
Province						
445 RF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Niger
Town of Birth	Saguia	Saguia	Saguia	Saguia	Saguia	
Province						
446 RF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Niger
Town of Birth	Kirkissoye	Doleo	Kirkissoye	Kirkissoye	Karague	
Province						
448 RF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Niger
Town of Birth	Dessa	Dessa	Dessa	Dessa	Dessa	
Province						
456 RF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Niger
Town of Birth	Baribé	Badawal	Badawa;	Barbé	Barbé	
Province						
457 RF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Niger
Town of Birth	Jangeré	Baouledé	Baouledé	Jangeré	Jangeré	
Province						
527 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth	Mabera					
Province	Sokoto	Sokoto	Sokoto	Sokoto	Sokoto	
531 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth	Gwarayo	Gwarayo	Gwarayo	Lugu	Lugu	
Province						
544 NF	Hausa	Hausa	Fulbe	Hausa	Hausa	Nigeria
Town of Birth	Sokoto	Kamba	Kamba	Sokoto	Gusau	
Province	Sokoto	Kebbi	Kebbi	Sokoto	Sokoto	
553 BF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth	Sakka	Sakka	Seba	Seba	Seba	
Province		Burkina Faso	Burkina Faso	Burkina Faso	Burkina Faso	
591 NF	Hausa	Fulbe	Fulbe	Hausa	Hausa	Nigeria
Town of Birth	Bunza	Masama	Masama	Dono Vabo	Dono Vabo	
Province						
597 NF	Hausa	Fulbe	Fulbe	Hausa	Hausa	Nigeria
Town of Birth	Bakura	Kano	Kano	Bakura	Bakura	
Province		Kano	Kano			

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
598 NF	Hausa	Fulbe	Fulbe	Hausa	Hausa	Nigeria
Town of Birth	Tambuwal	Tambuwal	Tambuwal	Kamba	Kamba	
Province	Sokoto	Sokoto	Sokoto	Sokoto	Sokoto	
710 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth						
Province	Adamawa	Cameroun	Cameroun	Adamawa	Gongola	
718 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth						
Province	Kwara	Kwara	Kwara	Kwara	Kwara	
719 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulani	Nigeria
Town of Birth						
Province	Sokoto	Sokoto	Sokoto	Sokoto	Sokoto	
722 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth	Bargodo	Bargodo	Bargodo	Bargodo	Bargodo	
Province	Sokoto	Sokoto	Sokoto	Sokoto	Sokoto	
723 YF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth	Ilorin	Nikki	Nikki	Nikki	Nikki	
Province	Kwara	Benin	Benin	Benin	Benin	
724 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth						
Province	Katsina	Katsina	Katsina	Katsina	Katsina	
725 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth						
Province	Katsina	Katsina	Katsina	Katsina	Katsina	
726 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth	Bauchi					
Province	Bauchi					
728 CF	Fulbe	Fulbe		Fulbe	Fulbe	Nigeria
Town of Birth	Garwa	Garwa		Garwa		
Province	Cameroun	Cameroun		Cameroun		
737 NF	Kanuri	Fulbe	Fulbe	Kanuri	Kanuri	Nigeria
Town of Birth	Maiduguri	Geidan	Geidan	Maiduguri	Maiduguri	
Province	Borno			Borno	Borno	
738 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth	Bauchi					
Province	Baushi	Bauchi	Bauchi	Bauchi	Bauchi	
746 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth	Anka					
Province	Sokoto	Sokoto	Sokoto	Sokoto	Sokoto	
747 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth						
Province	Sokoto	Sokoto	Sokoto	Sokoto	Sokoto	
854 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth	Madigala	Madigala	Madigala	Madigala	Madigala	
Province	Adamawa	Adamawa	Adamawa	Adamawa	Adamawa	
855 NF	Fulbe	Fulbe	Fulbe	Kanuri	Kanuri	Nigeria
Town of Birth	Madagali	Madagali	Madagali	Madagali	Madagali	
Province	Adamawa	Adamawa	Adamawa	Adamawa	Adamawa	

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
856 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth	Kunini	Kunini	Kunini	Kunini	Kunini	
Province	Adamawa	Adamawa	Adamawa	Adamawa	Adamawa	
866 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth						
Province	Sokoto	Sokoto	Sokoto	Sokoto	Sokoto	
869 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth	Saki	Saki		Ilesa		
Province	Oyo	Oyo		Kwara		
870 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth						
Province	Sokoto	Sokoto	Sokoto	Sokoto	Sokoto	
872 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth	Kiyoma	Barkay	Barkay	Kaima	Wawa	
Province	Kwara	Kwara	Kwara	Kwara	Kwara	
873 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth		Hasat	Ladi	Sumano	Ladi	
Province	Sokoto	Sokoto	Sokoto	Sokoto	Sokoto	
874 NF	Fulbe	Fulbe	Fulbe	Fulbe	Fulbe	Nigeria
Town of Birth	Yaweri	Bargudu	Bargudu	Bargudu	Bargudu	
Province	Sokoto	Sokoto	Sokoto	Sokoto	Sokoto	

10.2.9 Gusii

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
148 KU	Gusii	Gusii	Gusii	Gusii	Gusii	Kenya
Town of Birth						
Province	Kisii	Kisii	Kisii	Kisii	Kisii	
149 KU	Gusii	Gusii	Gusii	Gusii	Gusii	Kenya
Town of Birth						
Province	Kisii	Kisii	Kisii	Kisii	Kisii	
153 KU	Gusii	Gusii	Gusii	Gusii	Gisii	Kenya
Town of Birth						
Province	Kisii	Kisii	Kisii	Kisii	Kisii	
157 KU	Gusii	Gusii	Gusii	Gusii	Gusii	Kenya
Town of Birth						
Province	Kisii	Kisii	Kisii	Kisii	Kisii	
162 KU	Gusii	Gusii	Gusii	Gusii	Gusii	Kenya
Town of Birth						
Province	Kisii	Kisii	Kisii	Kisii	Kisii	
166 KU	Gusii	Gusii	Gusii	Gusii	Gusii	Kenya
Town of Birth						
Province	Kisii	Kisii	Kisii	Kisii	Kisii	

10.2.10 Hausa

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
255 RH	Hausa	Hausa	Hausa	Hausa	Arab	Niger
Town of Birth	Keita	Tchédia	Tchédia	Baguay	Egypt	
Province	Tahoua	Tahoua	Tahoua	Tahoua		
256 RH	Hausa	Hausa	Hausa	Hausa	Hausa	Niger
Town of Birth	Tounfalis	Chikal	Chikal	Tounfalis	Tounfalis	
Province	Filingué	Filingué	Filingué	Filingué	Filingué	

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
258 RH	Songhai	Hausa	Hausa	Songhai	Sonhai	Niger
Town of Birth	Tabla	Tabla	Chikal	Bangario	Bangario	
Province	Filingué	Filingué	Filingué	Filingué	Filingué	
259 RH	Hausa	Hausa	Hausa	Hausa	Hausa	Niger
Town of Birth	Gaffati	Magaria	Magaria	Zinder	Zinder	
Province	Mirriah	Mirriah	Mirriah	Bimi	Birni	
260 NH	Hausa	Hausa	Hausa	Hausa	Hausa	Niger
Town of Birth	Toudou	Filingué	Tounfalis	Filingué	Garka	
Province	Filingué	Filingué	Filingué	Filingué	Filingué	
261 RH	Hausa	Hausa	Hausa	Hausa	Hausa	Niger
Town of Birth	Kwakaré	Kany	Filingué	Filingué	Filingué	
Province	Filingué	Filingué	Filingué	Filingué	Filingué	
335 NH	Hausa	Hausa	Hausa	Hausa	Hausa	Nigeria
Town of Birth						
Province	Jos	Kano	Kano	Kano	Kano	
340 NH	Hausa	Hausa	Hausa	Hausa	Hausa	Nigeria
Town of Birth	Saki	Kano	Kano	Sokoto	Sokoto	
Province	Oyo	Kano	Kano	Sokoto	Sokoto	
425 RH	Hausa	Hausa	Hausa	Hausa	Hausa	Niger
Town of Birth	Natameye	Natameye	Natameye	Natameye	Natameye	
Province						
429 RH	Hausa	Hausa	Hausa	Hausa	Hausa	Niger
Town of Birth	Niamey	Tibini	Natankau	Natankau	Natankau	
Province						
439 RH	Hausa	Hausa	Hausa	Hausa	Hausa	Niger
Town of Birth	Kelémi	Ayaware	Ayaware	Kelémi	Kelémi	
Province						
440 RH	Hausa	Hausa	Hausa	Hausa	Hausa	Niger
Town of Birth	Kelémi	Kelémi	Laway	Kelémi	Kelémi	
Province						
528 NH	Hausa	Hausa	Hausa	Hausa	Hausa	Nigeria
Town of Birth	Aljima	Aljima	Aljima	Aljima	Aljima	
Province						
529 NH	Hausa	Hausa	Hausa	Hausa	Hausa	Nigeria
Town of Birth				Namoda	Namoda	
Province	Sokoto	Sokoto	Sokoto			
530 NH	Hausa	Hausa	Hausa	Hausa	Hausa	Nigeria
Town of Birth	Birnin Kebb	Birnin Kebbi	Birnin Kebbi	Birnin Kebbi	Birnin Kebbi	
Province	Sokoto	Sokoto	Sokoto	Sokoto	Sokoto	
532 NH	Hausa	Hausa	Hausa	Hausa	Hausa	Nigeria
Town of Birth	Sakaba	Zuru		Zuru		
Province						
533 NH	Hausa	Hausa	Hausa	Hausa	Hausa	Nigeria
Town of Birth		Koko	Koko	Koko	Koko	
Province	Kaduna					
534 NH	Hausa	Hausa	Hausa	Hausa	Hausa	Nigeria
Town of Birth		Tangaza	Tangaza	Tangaza	Tangaza	

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
Province	Sokoto					
535 NH	Hausa	Hausa	Hausa	Hausa	Hausa	Nigeria
Town of Birth	Arkilla	Kutar Atiku	Marina	Marina	Gobirawa	
Province	Sokoto	Sokoto	Sokoto	Sokoto	Sokoto	
536 NH	Hausa	Hausa	Hausa	Hausa	Hausa	Nigeria
Town of Birth	Illo Baguda	Illo Baguda	Illo Baguda	Illo Baguda	Illo Baguda	
Province						
537 NH	Hausa	Hausa	Hausa	Hausa	Hausa	Nigeria
Town of Birth	Bungudu	Bungudu	Bungudu	Bungudu	Bungudu	
Province						
538 NH	Hausa	Hausa	Hausa	Hausa	Hausa	Nigeria
Town of Birth	Achida	Kwabare	Kwabare	Achida	Achida	
Province						
539 NH	Hausa	Hausa	Hausa	Hausa	Hausa	Nigeria
Town of Birth	Anka	Anka	Anka	Anka	Anka	
Province	Sokoto	Sokoto	Sokoto	Sokoto	Sokoto	
540 NH	Hausa	Hausa	Hausa	Hausa	Hausa	Nigeria
Town of Birth	Jega	Jega	Jega	Jega	Jega	
Province						
543 NH	Hausa	Hausa	Hausa	Hausa	Hausa	Nigeria
Town of Birth	Kontogora	Kontogora	Kontogora	Kontogora	Kontogora	
Province	Niger	Niger	niger	Niger	Niger	
550 NH	Hausa	Hausa	Hausa	Hausa	Hausa	Nigeria
Town of Birth	Sokoto					
Province	Sokoto	Kano	Kano	Kano	Kano	
551 NH	Hausa	Hausa	Hausa	Hausa	Hausa	Nigeria
Town of Birth						
Province	Katsina	Katsina	Katsina	Katsina	Katsina	
552 NH	Hausa	Hausa	Hausa	Hausa	Hausa	Nigeria
Town of Birth	Zaria	Zaria	Zaria	Zaria	Zaria	
Province	Kaduna	Kaduna	Kaduna	Kaduna	Kaduna	
554 NH	Hausa	Hausa	Hausa	Hausa	Hausa	Nigeria
Town of Birth	Bagudo	Bagudo	Bagudo	Mungadi	Mungadi	
Province	Kebbi	Kebbi	Kebbi			
555 NH	Hausa	Hausa	Hausa	Hausa	Hausa	Nigeria
Town of Birth	Gando Anka	Gando Anka	Gando Anka	Gando Anka	Gando Anka	
Province						
592 NH	Hausa	Hausa	Hausa	Hausa	Hausa	Nigeria
Town of Birth	Minna	Minna	Minna	Minna	Minna	
Province	Niger	Niger	Niger	Niger	Niger	
593 NH	Hausa	Hausa	Hausa	Hausa	Hausa	Nigeria
Town of Birth						
Province	Kaduna	Jos	Jos	Kaduna	Kaduna	
594 NH	Hausa	Hausa	Hausa	Hausa	Hausa	Nigeria
Town of Birth	Sokoto	Denge	Denge	Durbawa	Durbawa	
Province	Sokoto	Sokoto	Sokoto			
595 NH	Hausa	Hausa	Hausa	Hausa	Hausa	Nigeria

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
Town of Birth	Sokoto	Gumi	Gumi	Sokoto	Sokoto	
Province	Sokoto			Sokoto	Sokoto	
596 NH	Hausa	Hausa	Hausa	Hausa	Hausa	Nigeria
Town of Birth	Sakobe	Shuri				
Province	Sokoto	Sokoto	Sokoto	Sokoto	Sokoto	
599 NH	Hausa	Hausa	Hausa	Hausa	Hausa	Nigeria
Town of Birth	Lagos	Rabah	Rabah	Azara	Azara	
Province		Sokoto	Sokoto	Kano	Kano	
651 NH	Hausa	Hausa	Hausa	Hausa	Hausa	Nigeria
Town of Birth	Lagos	Lagos	Wurno	Wurno	Wurno	
Province			Sokoto	Sokoto	Sokoto	
652 NH	Hausa	Hausa	Hausa	Hausa	Hausa	Nigeria
Town of Birth	Sabon Birni	Marina	Kofar Atiku	Kofar Atiku	Kofar Atiku	
Province	Sokoto	Sokoto	Sokoto	Sokoto	Sokoto	
713 NH	Hausa	Hausa	Hausa	Hausa	Hausa	Nigeria
Town of Birth	Birnin Kebb	Birnin Kebbi	Bimin Kebbi	Bimin Kebbi	Birnin Kebbi	
Province	Sokoto	Sokoto	Sokoto	Sokoto	Sokoto	
721 GH	Hausa	Hausa	Hausa	Hausa	Hausa	Nigeria
Town of Birth	Sokode	Sokode	Sokode	Sokode	Sokode	
Province	Ghana	Ghana	Ghana	Ghana	Ghana	
851 NH	Hausa	Hausa	Hausa	Hausa	Hausa	Nigeria
Town of Birth	Maiduguri	Maiduguri	Maiduguri	Maiduguri	Maiduguri	
Province	Borno	Borno	Borno	Borno	Bomo	
853 NF	Hausa	Hausa	Hausa	Hausa	Hausa	Nigeria
Town of Birth	Maiduguri	Babaniya	Babaniya	Kali		
Province	Borno					
858 NH	Hausa	Hausa	Hausa	Hausa	Hausa	Nigeria
Town of Birth						
Province	Kano	Kano	Kano	Kano	Kano	

10.2.11 Higgi

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
703 NF	Higgi	Higgi	Higgi	Higgi	Higgi	Nigeria
Town of Birth	Michika	Michika	Michika	Michika	Michika	
Province						

10.2.12 Ibibio

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
705 NF	Ibibio	Ibibio	Ibibio	Ibibio	Ibibio	Nigeria
Town of Birth						
Province	Akwa Ibom	Akwa Ibom	Akwa Ibom	Akwa Ibom	Akwa Ibom	

10.2.13 Idoma

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
708 NF	Idoma	Idoma	Idoma	Idoma	Idoma	Nigeria
Town of Birth						
Province	Benue	Benue	Benue	Benue	Benue	

10.2.14 Igbo

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
702 NI	Igbo	Igbo	Igbo	Igbo	Igbo	Nigeria
Town of Birth	Abia	Owerri	Owerri	Owerri	Owerri	
Province	Abia	Abia	Abia	Abia	Abia	
704 NI	Igbo	Igbo	Igbo	Igbo	Igbo	Nigeria
Town of Birth						

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
Province	Ananbra	Anambra	Anambra	Anambra	Anambra	
729 NI	Igbo	Igbo	Igbo	Igbo	Igbo	Nigeria
Town of Birth	Achina	Achina		Achina		
Province	Anambra	Anambra		Anambra		

10.2.15 Junkun

715 NF	Hausa	Junkun	Junkun	Hausa	Hausa	Nigeria
Town of Birth	Ara	Ara	Ara	Ara	Ara	
Province	Katsina	Katsina	Katsina	Katsina	Katsina	

10.2.16 Kalenjin

133 KJ	Kalenjin	Kalenjin	Kalenjin	Kalenjin	Kalenjin	Kenya
Town of Birth	Kartonjo	Kartijin	Baringo	Kartonjo		
Province	Baringo	Baringo	Baringo	Baringo	Baringo	
144 KJ	Kalenjin	Kalenjin	Kalenjin	Kalenjin	Kalenjin	Kenya
Town of Birth						
Province	Kericho	Kericho	Kericho	Kericho	Kericho	
151 KJ	Kalenjin	Kalenjin	Kalenjin	Kalenjin	Kalenjin	Kenya
Town of Birth						
Province	Kericho	Kericho	Kericho	Kericho	Kericho	
154 KJ	Kalenjin	Kalenjin	Kalenjin	Kalenjin	Kalenjin	Kenya
Town of Birth						
Province	Kericho	Kericho	Kericho	Kericho	Kericho	
161 KJ	Kalenjin	Kalenjin	Kalenjin	Kalenjin	Kalenjin	Kenya
Town of Birth						
Province	Baringo	Baringo	Baringo	Baringo	Baringo	
167 KJ	Kalenjin	Kalenjin	Kalenjin	Kalenjin	Kalenjin	Kenya
Town of Birth						
Province	Baringo	Baringo	Baringo	Baringo	Baringo	

10.2.17 Kanuri

262 RK	Kanuri	Kanuri	Kanuri	Kanuri	Kanuri	Niger
Town of Birth	Bimi Kazoé	Gadori	Birni Kazoé	Zinder	Zinder	
Province	Goué	Goué	Goué	Birni	Birni	
706 NK	Kanuri	Kanuri	Kanuri	Kanuri	Kanuri	Nigeria
Town of Birth						
Province	Borno	Borno	Borno	Borno	Borno	
707 NK	Kanuri	Kanuri	Kanuri	Kanuri	Kanuri	Nigeria
Town of Birth	Damboa	Damboa	Damboa	Damboa	Damboa	
Province						
712 NK	Shuwa	Kanuri	Kanuri	Shuwa	Shuwa	Nigeria
Town of Birth	Maiduguri	Maiduguri	Maiduguri	Maiduguri	Maiduguri	
Province	Borno	Borno	Borno	Borno	Borno	
731 NK	Kanuri	Kanuri	Kanuri	Kanuri	Kanuri	Nigeria
Town of Birth	Maiduguri	Kanduga	Kanduga		Kanduga	
Province	Borno	Borno	Borno	Yobe	Borno	
852 NK	Kanuri	Kanuri	Kanuri	Kanuri	Kanuri	Nigeria
Town of Birth	Maiduguri	Kodaga	Maiduguri	Damaturu	Maiduguri	
Province	Borno		Borno	Yobe	Borno	

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
857 NK	Kanuri	Kanuri	Kanuri	Kanuri	Kanuri	Nigeria
Town of Birth	Maiduguri	Maiduguri	Maiduguri	Maiduguri	Maiduguri	
Province	Bomo	Bomo	Bomo	Bomo	Bomo	
859 NK	Kanuri	Kanuri	Kanuri	Kanuri	Kanuri	Nigeria
Town of Birth	Maiduguri	Dabuwa	Dabuwa	Dabuwa	Dabuwa	
Province	Bomo	Bomo	Bomo	Bomo	Bomo	
861 NK	Kanuri	Kanuri	Kanuri	Kanuri	Kanuri	Nigeria
Town of Birth	Maiduguri	Maiduguri	Maiduguri	Maiduguri	Maiduguri	
Province	Bomo	Bomo	Bomo	Bomo	Borno	
862 NK	Kanuri	Kanuri	Kanuri	Kanuri	Kanuri	Nigeria
Town of Birth	Damboa	Damboa	Damboa	Damboa	Damboa	
Province	Bomo	Bomo	Borno	Borno	Borno	
863 NK	Kanuri	Kanuri	Kanuri	Kanuri	Kanuri	Nigeria
Town of Birth	Maiduguri	Damboa	Kukawa	Kukawa	Kukawa	
Province	Bomo	Bomo	Borno	Bomo	Borno	
864 NK	Kanuri	Kanuri	Kanuri	Kanuri	Kanuri	Nigeria
Town of Birth	Maiduguri	Maiduguri	Maiduguri	Maiduguri	Maiduguri	
Province	Bomo	Borno	Bomo	Bomo	Borno	
865 NK	Kanuri	Kanuri	Kanuri	Kanuri	Kanuri	Nigeria
Town of Birth	Maiduguri	Maiduguri	Maiduguri	Maiduguri	Maiduguri	
Province	Bomo	Bomo	Borno	Borno	Borno	
867 NK	Kanuri	Kanuri	Kanuri	Kanuri	Kanuri	Nigeria
Town of Birth	Manguno	Manguno	Manguno	Manguno	Manguno	
Province	Bomo	Bomo	Bomo	Borno	Borno	
868 NK	Kanuri	Kanuri	Kanuri	Kanuri	Kanuri	Nigeria
Town of Birth	Maiduguri	Maiduguri	Maiduguri	Maiduguri	Maiduguri	
Province	Borno	Borno	Borno	Borno	Borno	

10.2.18 Kikuyu

27 KK	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kenya
Town of Birth						
Province	Nakuru	Kiambu	Kiambu	Murang'a	Murang'a	
34 KK	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kenya
Town of Birth						
Province	Kiambu	Kiambu	Kiambu	Kiambu	Kiambu	
66 KK	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kenya
Town of Birth						
Province	Murang'a	Murang'a	Murang'a	Murang'a	Murang'a	
68 KK	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kenya
Town of Birth						
Province	Kiambu	Kiambu	Kiambu	Kiambu	Kiambu	
94 KK	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kenya
Town of Birth	Uasingishu	Uasingishu	Uasingishu	Kiambu	Kiambu	
Province	Eldoret	Eldoret	Kiambu	Kiambu	Kiambu	
97 KK	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kenya
Town of Birth						
Province	Murang'a	Murang'a	Murang'a	Murang'a	Murang'a	
98 KK	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kenya

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
Town of Birth	Meru	Nanyuki	Meru	Nyeri	Nyeri	
Province	Meru	Lakipia	Meru	Nyeri	Nyeri	
115 KK	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kenya
Town of Birth						
Province	Kiambu	Kiambu	Kiambu	Kiambu	Kiambu	
136 KK	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kenya
Town of Birth						
Province	Kiambu	Nakuru				
138 KK	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kenya
Town of Birth						
Province	Nakuru	Kiambu	Kiambu			
139 KK	Kikuyu	Kikuyu	Kikuyu	Kikuyu		Kenya
Town of Birth						
Province	Nakuru					
140 KK	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kenya
Town of Birth	Surai	Surai	Surai	Surai	Surai	
Province						
141 KK	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kenya
Town of Birth						
Province	Nakuru	Nakuru	Kiambu	Nakuru	Kiambu	
142 KK	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kenya
Town of Birth						
Province	Nakuru	Kiambu				
150 KK	Kikuyu	Kikuyu		Kikuyu		Kenya
Town of Birth						
Province						
155 KK	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kenya
Town of Birth						
Province	Nakuru	Nakuru	Nakuru	Nakuru	Nakuru	
156 KK	Swahili	Kikuyu	Kikuyu	Swahili	Swahili	Kenya
Town of Birth						
Province	Mombasa	Nyeri				
165 KK	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kenya
Town of Birth						
Province	Kiambu	Kiambu	Kiambu	Kiambu	Kiambu	
169 KK	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kenya
Town of Birth	Machakos	Embu	Embu			
Province		Embu	Embu	Murang'a	Murang'a	
170 KK	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kenya
Town of Birth	Kerugoya	Kerugoya	Kerugoya	Kerugoya	Kerugoya	
Province	Embu	Embu	Embu	Embu	Embu	
172 KK	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kenya
Town of Birth						
Province	Murang'a	Murang'a	Murang'a	Murang'a	Murang'a	
565 KK	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kenya
Town of Birth						
Province	Nyeri	Nyeri	Nyeri	Nyeri	Nyeri	

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
566 KK	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kenya
Town of Birth						
Province	Nyeri	Nyeri	Nyeri	Nyeri	Nyeri	
568 KK	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kenya
Town of Birth						
Province	Nakuru	Kitale	Kitale	Kiambu	Kiambu	
572 KK	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kikuyu	Kenya
Town of Birth	Molo	Molo	Molo	Molo	Kiambu	
Province	Nakuru	Nakuru	Nakuru	Nakuru	Kiambu	

10.2.19 Luhyia

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
24 KY	Luhyia	Luhyia	Luhyia	Luhyia	Luhyia	Kenya
Town of Birth	Kakamega	Lubao	Lubao	Lukusi	Lukusi	
Province	Kakamega	Kakamega	Kakamega	Kakamega	Kakamega	
28 KY	Luhyia	Luhyia	Luhyia	Luhyia	Luhyia	Kenya
Town of Birth	Kakmega	Isuka	Isuka	Ikolomani	Ikolomani	
Province	Kakamega	Kakamega	Kakamega	Kakamega	Kakamega	
29 KY	Luhyia	Luhyia	Luhyia	Luhyia	Luhyia	Kenya
Town of Birth	Emuhaya	Emuhaya	Emuhaya	Emuhaya	Emuhaya	
Province	Kakamega	Kakamega	Kakamega	Kakamega	Kakamega	
30 KY	Luhyia	Luhyia	Luhyia	Luhyia	Luhyia	Kenya
Town of Birth	Butere					
Province	Kakamega	Kakamega	Kakamega	Kakamega	Kakamega	
35 KY	Luhyia	Luhyia	Luhyia	Luhyia	Luhyia	Kenya
Town of Birth						
Province	Kakamega	Kakamega	Kakamega	Kakamega	Kakamega	
42 KY	Luhyia	Luhyia	Luhyia	Luhyia	Luhyia	Kenya
Town of Birth	Kakamega	Kakamega	Maragoli	Kakamega		
Province	Kakamega	Kakamega	Kakamega	Kakamega		
67 KY	Luhyia	Luhyia	Luhyia	Luhyia	Luhyia	Kenya
Town of Birth						
Province	Kakamega	Kakamega	Kakamega	Kakamega	Kakamega	
69 KY	Luhyia	Luhyia	Luhyia	Luhyia	Luhyia	Kenya
Town of Birth						
Province	Kakamega	Kakamega	Kakamega	Kakamega	Kakamega	
70 KY	Luhyia	Luhyia	Luhyia	Luhyia	Luhyia	Kenya
Town of Birth						
Province	Kakamega	Kakamega	Kakamega	Kakamega	Kakamega	
72 KY	Luo	Luhyia	Luhyia	Luhyia	Luhyia	Kenya
Town of Birth	Seme	Emuhaya				
Province	Kisumu	Kakamega	Kakamega	Kakamega	Kakamega	
119 KY	Luhyia	Luhyia	Luhyia	Luhyia	Luhyia	Kenya
Town of Birth	Kitale	Kitale	Kitale	Kitale	Kitale	
Province	Trans Nzoia	Trans Nzoia	Trans Nzoia	Trans Nzoia	Trans Nzoia	
121 KY	Luhyia	Luhyia	Luhyia	Luhyia	Luhyia	Kenya
Town of Birth	Nairobi					
Province		Siaya	Siaya	Kakamega	Kakamega	
125 KY	Luhyia	Luhyia	Luhyia	Luhyia	Luhyia	Kenya

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
Town of Birth						
Province	Bungoma	Bungoma	Bungoma	Bungoma	Bungoma	
134 KY	Luhyia	Luhyia	Luhyia	Luhyia		Kenya
Town of Birth						
Province	Bungoma	Busia	Busia	Bungoma		
137 KY	Luhyia	Luhyia	Luhyia	Luhyia	Luhyia	Kenya
Town of Birth						
Province	Kakamega	Kakamega	Kakamega	Kakamega	Kakamega	
143 KY	Luhyia	Luhyia	Luhyia	Luhyia	Luhyia	Kenya
Town of Birth						
Province	Kakamega	Kakamega	Kisumu	Kakamega	Kakamega	
158 KY	Luhyia	Luhyia	Luhyia	Luhyia	Luhyia	Kenya
Town of Birth	Samia					
Province	Busia	Busia	Busia	Busia	Busia	
159 KY	Luhyia	Luhyia	Luhyia	Luhyia	Luhyia	Kenya
Town of Birth						
Province	Nandi	Kakamega	Kakamega	Kakamega	Kakamega	
160 KY	Luhyia	Luhyia	Luhyia	Luhyia	Luhyia	Kenya
Town of Birth						
Province	Busia	Busia	Busia	Busia	Busia	
163 KY	Luhyia	Luhyia	Luhyia	Luhyia	Luhyia	Kenya
Town of Birth						
Province	Kakamega	Kakamega	Kakamega	Kisumu	Kakamega	

10.2.20 Luo

20 KL	Luo	Luo	Luo	Luo		Kenya
Town of Birth	Kisumu	Seme				
Province	Kisumu	Kisumu	Kisumu	Kisumu		
21 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth	Seme	Seme		Seme		
Province	Kisumu	Kisumu	Kisumu	Kisumu	Kisumu	
22 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth	Kano	Nyahera	Nyahera	Kano	Kakmir	
Province	Kisumu	Kisumu	Kisumu	Turkana	Kisumu	
23 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth	Ugenya	Ugenya	Ugenya	Ugenya		
Province	Siaya	Siaya	Siaya	Siaya		
25 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth	Kisumu		Kajulu	Kajulu		
Province	Kisumu	South Nyanza	Kisumu	Kisumu	Tanzania	
26 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth	Gembe	Jula		Kamrere		
Province	South Nyanza	South Nyanza	South Nyanza	South Nyanza	South Nyanza	
31 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth	Homa bay	Homa Bay	Homa Bay	Homa Bay	Homa Bay	
Province	South Nyanza	South Nyanza	South Nyanza	South Nyanza	South Nyanza	
33 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth						

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
Province	Kisumu	Kisumu	Kisumu	Kisumu	Kisumu	
36 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth		Seme	Seme	Seme	Seme	
Province	Kisumu	Kisumu	Kisumu	Kisumu	Kisumu	
37 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth	Osiri	Chiga	Chiga	Osiri	Osiri	
Province	Kisumu	Kisumu	Kisumu	Kisumu	Kisumu	
38 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth	Bondo	Jula	Alego	Sakwa	Sakwa	
Province	Siaya	South Nyanza	Siaya	Siaya	Siaya	
39 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth	Kajulu			Kisumu		
Province	Kisumu	South Nyanza	South Nyanza	Kisumu	Kisumu	
40 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth		Alego	Alego	Alego	Alego	
Province	Siaya	Siaya	Siaya	Siaya	Siaya	
41 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth	Yala	Ugenya	Ugenya	Yala	Gem	
Province	Siaya	Siaya	Siaya	Siaya	Siaya	
43 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth	Karachuonyo	Kochia	Mbita	Karachuonyo	Karachuomo	
Province	South Nyanza	South Nyanza	South Nyanza	South Nyanza	South Nyanza	
44 KL	Luo	Luo		Luo		Kenya
Town of Birth	Kajulu			Kisumu		
Province	Kisumu	Kisumu		Kisumu		
45 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth	Seme	Seme	Seme	Seme	Seme	
Province	Kisumu	Kisumu	Kisumu	Kisumu	Kisumu	
46 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth	Kisumu			Katito	Katito	
Province	Kisumu	South Nyanza	South Nyanza	Kisumu	Kisumu	
47 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth						
Province	Siaya	Busia	Busia	Busia	Busia	
48 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth						
Province	Kisumu	Kisumu	Kisumu	Kisumu	Kisumu	
49 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth	Ahero	East Korwa				
Province	Kisumu	Kisumu	Kisumu	Kisumu	Kisumu	
50 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth						
Province	Siaya	Siaya	Siaya	Siaya	Siaya	
52 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth						
Province	Kisumu	Kisumu	Kisumu	Kisumu	Kisumu	
53 KL	Luo	Luo	Luo	Luo	Luo	Kenya

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
Town of Birth						
Province	Kisumu	Kisumu	Kisumu	Kisumu	Kisumu	
54 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth	Ugenya					
Province	Siaya	Siaya	Siaya	Siaya	Siaya	
55 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth	Kisumu	Kendo Bay	Kendo Bay	Kawabwaya	Kawabwaya	
Province	Kisumu	South Nyanza	South Nyanza	South Nyanza	South Nyanza	
57 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth						
Province	South Nyanza	South Nyanza	South Nyanza	South Nyanza	South Nyanza	
58 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth						
Province	Kisumu	Kisumu	Kisumu	Kisumu	Kisumu	
59 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth						
Province	Siaya	Siaya	Siaya	Siaya	Siaya	
60 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth						
Province	Siaya	Siaya	Siaya	Siaya	Siaya	
61 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth		Kano				
Province	Kisumu	Kisumu	Kisumu	Kisumu	Kisumu	
62 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth						
Province	Kisumu	Siaya	Siaya	Siaya	Siaya	
63 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth	Nyakach					
Province	Kisumu	Kisumu	Kisumu	Kisumu	Kisumu	
64 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth	Nyakach	Kabondo	Kabondo	Nyakach	Nyakach	
Province	Kisumu	South Nyanza	South Nyanza	Kisumu	Kisumu	
65 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth	Kano	Chemeli	Chemeli	Nyakach	Nyakach	
Province	Kisumu	Kisumu	Kisumu	Kisumu	Kisumu	
71 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth						
Province	Kisumu	Kisumu	Kisumu	Kisumu	Kisumu	
73 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth		Ugenya				
Province	Kisumu	Siaya	Kisumu	Kisumu	Kisumu	
74 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth		Seme			Seme	
Province	South Nyanza	Kisumu	South Nyanza	South Nyanza	Kisumu	
75 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth			Alego			
Province	Siaya	Siaya	Siaya	Siaya	Siaya	

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
76 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth	Nyakach	Kano	Kano	Nyakach	Nyakach	
Province	Kisumu	Kisumu	Kisumu	Kisumu	Kisumu	
77 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth	Kaksingiri	Gem	Kaksingeri	Gem	Kaksingeri	
Province	South Nyanza	Siaya	South Nyanza	Siaya	South Nyanza	
130 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth				Kabondo		
Province	South Nyanza	Kisumu	Kisumu	Kisii	Kisii	
132 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth						
Province	Kisumu	Siaya	Siaya	Kisumu	Kisumu	
135 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth						
Province	Siaya	Siaya	Siaya	Siaya	Siaya	
145 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth						
Province	South Nyanza	Kakamega	South Nyanza	South Nyanza	South Nyanza	
146 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth						
Province	South Nyanza	South Nyanza	South Nyanza	South Nyanza	South Nyanza	
147 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth						
Province	Kisumu	Kisumu	Siaya	Kisumu	Kisumu	
152 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth						
Province	South Nyanza	South Nyanza	South Nyanza	South Nyanza	South Nyanza	
171 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth						
Province	Kisumu	Kisumu	Kisumu	Kisumu	Kisumu	
575 KL	Luo	Luo	Luo	Luo	Luo	Kenya
Town of Birth						
Province	Siaya	Siaya	Siaya	Siaya	Siaya	

10.2.21 Maasai

1 KM	Maasai	Maasai	Maasai	Maasai	Masaai	Kenya
Town of Birth	Oltapesi	Oltapesi	Oltapesi	Oltapesi	Oltapesi	
Province	Kajiado	Kajiado	Kajiado	Kajiado	Kajiado	
2 KM	Maasai	Maasai	Maasai	Maasai	Masaai	Kenya
Town of Birth	Oltapesi	Oltapesi	Oltapesi	Oltapesi	Oltapesi	
Province	Kajiado	Kajiado	Kaijiado	Kajiado	Kajiado	
3 KM	Maasai	Maasai	Maasai	Maasai	Masaai	Kenya
Town of Birth	Oltapesi	Oltapesi	Oltapesi	Oltapesi	Oltapesi	
Province	Kajiado	Kajiado	Kajiado	Kajiado	Kajiado	
4 KM	Maasai	Maasai	Maasai	Maasai	Masaai	Kenya
Town of Birth	Oltapesi	Oltapesi	Oltapesi	Oltapesi	Oltapesi	
Province	Kajiado	Kajiado	Kajiado	Kajiado	Kajiado	
5 KM	Maasai	Maasai	Maasai	Maasai	Masaai	Kenya

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
Town of Birth	Oltapesi	Oltapesi	Oltapesi	Oltapesi	Oltapesi	
Province	Kajiado	Kajiado	Kajiado	Kajiado	Kajiado	
6 KM	Maasai	Maasai	Maasai	Maasai	Masaai	Kenya
Town of Birth	Oltapesi	Oltapesi	Oltapesi	Oltapesi	Oltapesi	
Province	Kajiado	Kajiado	Kajiado	Kajiado	Kajiado	
7 KM	Maasai	Maasai	Maasai	Maasai	Masaai	Kenya
Town of Birth	Oltapesi	Oltapesi	Oltapesi	Oltapesi	Oltapesi	
Province	Kajiado	Kajiado	Kajiado	Kajiado	Kajiado	
12 KM	Maasai	Maasai	Maasai	Maasai	Masaai	Kenya
Town of Birth	Oltapesi	Oltapesi	Oltapesi	Oltapesi	Oltapesi	
Province	Kajiado	Kajiado	Kajiado	Kajiado	Kajiado	
13 KM	Maasai	Maasai	Maasai	Maasai	Masaai	Kenya
Town of Birth	Ngoire	Ngoire	Ngoire	Ngoire	Ngoire	
Province	Kajiado	Kajiado	Kajiado	Kajiado	Kajiado	
18 KM	Maasai	Maasai	Maasai	Maasai	Masaai	Kenya
Town of Birth	Kiloliti	Kiloliti	Kiloliti	Kiloliti	Kiloliti	
Province	Kajiado	Kajiado	Kajiado	Kajiado	Kajiado	
51 KM	Maasai	Maasai	Maasai	Maasai	Masaai	Kenya
Town of Birth						
Province	Kajiado	Kajiado	Kajiado	Kajiado	Kajiado	
80 KM	Maasai	Maasai	Maasai	Maasai	Masaai	Kenya
Town of Birth	Mabartni	Mabatini	Mabatini	Mabatini	Mabatini	
Province	Kajiado	Kajiado	Kajiado	Kajiado	Kajiado	
84 KM	Maasai	Maasai	Maasai	Maasai	Masaai	Kenya
Town of Birth	Oloiyangala	Oloiyangalan	Oloiyangalan	Oloiyangalani	Oloiyangalani	
Province	Kajiado	Kajiado	Kajiado	Kajiado	Kajiado	
89 KM	Maasai	Maasai	Maasai	Maasai	Masaai	Kenya
Town of Birth	Mabatini	Mabatini	Mabatini	Mabatini	Mabatini	
Province	Kajiado	Kajiado	Kajiado	Kajiado	Kajiado	
570 KM	Maasai	Maasai	Maasai	Maasai	Maasai	Kenya
Town of Birth						
Province	Narok	Narok	Narok	Narok	Narok	

10.2.22 Mambila

732 NK	Mambila	Mambila	Mambila	Mambila	Mambila	Nigeria
Town of Birth	Warwar					
Province						

10.2.23 Margi

730 NF	Margi	Margi	Margi	Margi	Margi	Nigeria
Town of Birth	Uba					
Province	Borno					

10.2.24 Nuer

350 EN	Nuer	Nuer	Nuer	Nuer	Nuer	Ethiopia
Town of Birth	Jikao	Jikao	Jikao	Jikao	Jikao	
Province	Ilubabor	Ilubabor	Ilubabor	Ilubabor	Ilubabor	
351 EN	Nuer	Nuer	Nuer	Nuer	Nuer	Ethiopia
Town of Birth	Jikao	Jikao	Jikao	Jikao	Jikao	
Province	Ilubabor	Ilubabor	Ilubabor	Ilubabor	Ilubabor	

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
352 EN	Nuer	Nuer	Nuer	Nuer	Nuer	Ethiopia
Town of Birth	Jikao	Jikao	Jikao	Jikao	Jikao	
Province	Ilubabor	Ilubabor	Ilubabor	Ilubabor	Ilubabor	
353 SN	Nuer	Nuer	Nuer	Nuer	Nuer	Ethiopia
Town of Birth	jikao	Jikao	Jikao	Jikao	Jikao	
Province	Ilubabor	Ilubabor	Ilubabor	Ilubabor	Ilubabor	
354 SN	Nuer	Nuer	Nuer	Nuer	Nuer	Ethiopia
Town of Birth	Jikao	Jikao	Jikao	Jikao	Jikao	
Province	Sudan	Sudan	Sudan	Sudan	Sudan	
355 SN	Nuer	Nuer	Nuer	Nuer	Nuer	Ethiopia
Town of Birth	Jikao	Jikao	Jikao	Jikao	Jikao	
Province	Sudan	Sudan	Sudan	Sudan	Sudan	
356 SN	Nuer	Nuer	Nuer	Nuer	Nuer	Ethiopia
Town of Birth	Jikao	Jikao	Jikao	Jikao	Jikao	
Province	Ilubabor	Sudan	Sudan	Sudan	Sudan	
357 EN	Nuer	Nuer	Nuer	Nuer	Nuer	Ethiopia
Town of Birth	Jikao	Jikao	Jikao	Jikao	Jikao	
Province	Ilubabor	Ilubabor	Ilubabor	Ilubabor	Ilubabor	
358 SN	Nuer	Nuer	Nuer	Nuer	Nuer	Ethiopia
Town of Birth	Jikao	Jikao	Jikao	Jikao	Jikao	
Province	Ilubabor	Sudan	Sudan	Ethiopia	Ilubabor	
359 SN	Nuer	Nuer	Nuer	Nuer	Nuer	Ethiopia
Town of Birth	Jikao	Jikao	Jikao	Jikao	Jikao	
Province	Sudan	Sudan	Sudan	Sudan	Sudan	
360 SN	Nuer	Nuer	Nuer	Nuer	Nuer	Ethiopia
Town of Birth	Malakal	Malakal	Malakal	Malakal	Makalal	
Province	Sudan	Sudan	Sudan	Sudan	Sudan	
362 EN	Nuer	Nuer	Nuer	Nuer	Nuer	Ethiopia
Town of Birth	Jikao	Jikao	Jikao	Jikao	Jikao	
Province	Ilubabor	Ilubabor	Ilubabor	Ilubabor	Ilubabor	
363 SN	Nuer	Nuer	Nuer	Nuer	Nuer	Ethiopia
Town of Birth	Jikao	Nasir	Nasir	Jikao	Jikao	
Province	Sudan	Sudan	Sudan	Sudan	Sudan	
364 SN	Nuer	Nuer	Nuer	Nuer	Nuer	Ethiopia
Town of Birth	Yom	Yom	Yom	Yom	Yom	
Province	Sudan	Sudan	Sudan	Sudan	Sudan	
365 EN	Nuer	Nuer	Nuer	Nuer	Nuer	Ethiopia
Town of Birth	Akobo	Jikao	Jikao	Nasir	Nasir	
Province	Ilubabor	Ilubabor	Ilubabor	Sudan	Sudan	
366 EN	Nuer	Nuer	Nuer	Nuer	Nuer	Ethiopia
Town of Birth	Jikao	Jikao	Jikao	Jikao	Jikao	
Province	Ilubabor	Ilubabor	Ilubabor	Ilubabor	Ilubabor	
367 EN	Nuer	Nuer	Nuer	Nuer	Nuer	Ethiopia
Town of Birth	Jikao	Jikao	Jikao	Jikao	Jikao	
Province	Ilubabor	Ilubabor	Ilubabor	Ilubabor	Ilubabor	
368 SN	Nuer	Nuer	Nuer	Nuer	Nuer	Ethiopia
Town of Birth	Jikao	Jikao	Jikao	Jikao	Akobo	

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
Province	Sudan	Sudan	Ilubabor	Ilubabor	Ilubabor	
369 SN	Nuer	Nuer	Nuer	Nuer	Nuer	Ethiopia
Town of Birth	Jikao	Jikao	Jikao	Jikao	Jikao	
Province	Sudan	Sudan	Sudan	Sudan	Sudan	
370 SN	Nuer	Nuer	Nuer	Nuer	Nuer	Ethiopia
Town of Birth	Nasir	Nasir	Nasir	Jikao	Jikao	
Province	Sudan	Sudan	Sudan	Sudan	Sudan	
371 EN	Nuer	Nuer	Nuer	Nuer	Nuer	Ethiopia
Town of Birth	Jikao	Jikao	Jikao	Jikao	Jikao	
Province	Ilubabor	Ilubabor	Ilubabor	Ilubabor	Ilubabor	
372 SN	Nuer	Nuer	Nuer	Nuer	Nuer	Ethiopia
Town of Birth	Jikao	Jikao	Jikao	Jikao	Jikao	
Province	Sudan	Sudan	Sudan	Sudan	Sudan	
373 SN	Nuer	Nuer	Nuer	Nuer	Nuer	Ethiopia
Town of Birth	Jikao	Jikao	Jikao	Jikao	Jikao	
Province	Sudan	Sudan	Sudan	Sudan	Sudan	
375 SN	Nuer	Nuer	Nuer	Nuer	Nuer	Ethiopia
Town of Birth	Jikao	Jikao	Jikao	Jikao	Jikao	
Province	Sudan	Sudan	Sudan	Sudan	Sudan	
376 EN	Nuer	Nuer	Nuer	Nuer	Nuer	Ethiopia
Town of Birth	Jikao	Jikao	Jikao	Jikao	Jikao	
Province	Ilubabor	Ilubabor	Ilubabor	Ilubabor	Ilubabor	
377 EN	Nuer	Nuer	Nuer	Nuer	Nuer	Ethiopia
Town of Birth	Jikao	Jikao	Jikao	Jikao	Jikao	
Province	Ilubabor	Ilubabor	Ilubabor	Ilubabor	Ilubabor	
378 SN	Nuer	Nuer	Nuer	Nuer	Nuer	Ethiopia
Town of Birth	Jikao	Jikao	Jikao	Jikao	Jikao	
Province	Sudan	Sudan	Sudan	Sudan	Sudan	
379 EN	Nuer	Nuer	Nuer	Nuer	Nuer	Ethiopia
Town of Birth	Jikao	Jikao	Jikao	Jikao	Jikao	
Province	Ilubabor	Ilubabor	Ilubabor	Ilubabor	Ilubabor	
380 EN	Nuer	Nuer	Nuer	Nuer	Nuer	Ethiopia
Town of Birth	Jikao	Jikao	JIkao	Jikao	Jikao	
Province	Ilubabor	Ilubabor	Ilubabor	Ilubabor	Ilubabor	
381 SN	Nuer	Nuer	Nuer	Nuer	Nuer	Ethiopia
Town of Birth	Jickao	Jickao	Jickao	Jickao	Jickao	
Province	Sudan	Sudan	Sudan	Sudan	Sudan	

10.2.25 Omotaboh

124 K?	Omotaboh	Omotaboh	Omotaboh	Omotaboh	Omotaboh	Kenya
Town of Birth	Kisii	Kisii	Kisii	Kisii	Kisii	
Province	Kisii	Kisii	Kisii	Kisii	Kisii	

10.2.26 Rendille

16 KR	Rendille	Rendille	Rendille	Rendille	Rendille	Kenya
Town of Birth	Marsabit	Marsabit	Marsabit	Marsabit	Marsabit	
Province	Marsabit	Marsabit	Marsabit	Marsabit	Kajiado	
234 KR	Somali	Rendille	Rendille	Somali	Somali	Kenya
Town of Birth	Bulapesa	Bulapesa				

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
Province	Isiolo	Isiolo				
569 KR	Samburu	Rendille	Rendille	Samburu	Samburu	Kenya
Town of Birth	Ilaut	Ilaut	Ilaut	Ilaut	Ilaut	
Province						

10.2.27 Sabarumo

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
716 NF	Sabarumo	Sabarumo	Sabarumo	Sabarumo	Sabarumo	Nigeria
Town of Birth	Kanguia	Salaratu				
Province	Kebbi	Kebbi	Kebbi	Kebbi	Kebbi	

10.2.28 Sakuye

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
178 K?	Sakuye	Sakuye		Sakuye	Sakuye	Kenya
Town of Birth	Isiolo	Garba Tulla	Isiolo	Isiolo	Isiolo	
Province	Isiolo	Isiolo	Isiolo	Isiolo	Isiolo	
208 K?	Sakuye	Sakuye	Sakuye	Sakuye	Sakuye	Kenya
Town of Birth	Isiolo	Garba Tulla	Garba Tulla	Garba Tulla	Isiolo	
Province	Isiolo	Isiolo	Isiolo	Isiolo	Isiolo	
215 K?	Sakuye	Sakuye	Sakuye	Sakuye	Sakuye	Kenya
Town of Birth	Garba Tulla	Isiolo	Isiolo	Isiolo	Isiolo	
Province	Isiolo	Isiolo	Isiolo	Isiolo	Isiolo	
221 K?	Sakuye	Sakuye	Sakuye	Sakuye	Sakuye	Kenya
Town of Birth	Merti					
Province	Isiolo			Isiolo	Isiolo	
228 K?	Sakuye	Sakuye	Sakuye	Sakuye	Sakuye	Kenya
Town of Birth	Garisa	Marsabit	Moyale			
Province	Garisa	Marsabit	Moyale			
232 K?	Borana	Borana	Sakuye	Borana	Borana	Kenya
Town of Birth	Merti	Merti	Merti	Merti	Merti	
Province	Isiolo	Isiolo	Isiolo	Isiolo	Isiolo	
233 K?	Borana	Sakuye	Sakuye	Borana	Boran	Kenya
Town of Birth	Merti					
Province	Isiolo					
236 K?	Sakuye	Sakuye	Sakuye	Sakuye	Sakuye	Kenya
Town of Birth	Malkadaka	Garba Tulla				
Province	Isiolo	Isiolo				
237 KC	Borana	Borana	Sakuye	Borana	Boran	Kenya
Town of Birth	Idilola	Merti	Isiolo	Gufu	Gufu	
Province	Ethiopia	Isiolo	Isiolo	Ethiopia	Ethiopia	
242 K?	Borana	Sakuye	Sakuye	Borana	Borana	Kenya
Town of Birth	Garba Tulla	Malkadaka	Garba Tulla	Merti	Garba Tulla	
Province	Isiolo	Isiolo	Isiolo	Isiolo	Isiolo	
244 K?	Sakuye	Borana	Sakuye	Sakuye	Sakuye	Kenya
Town of Birth	Garba Tulla	Kulamane	Garfarsa			
Province	Isiolo	Isiolo	Isiolo	Somalia	Somalia	

10.2.29 Samburu

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
95 KS	Rendille	Samburu	Samburu	Rendille	Rendille	Kenya
Town of Birth	Kisima	Poro	Ndoto	Korr	Korr	
Province	Samburu	Samburu	Samburu	Samburu	Samburu	

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
164 KS	Samburu	Samburu	Samburu	Samburu	Samburu	Kenya
Town of Birth	Baragoi	Ndoto		Nyero	Nyero	
Province	Marsabit	Samburu		Samburu	Samburu	
200 KS	Samburu	Samburu	Samburu	Samburu	Samburu	Kenya
Town of Birth	Bahawa	Bahawa	Bahawa	Marmar	Marmar	
Province						
201 KS	Samburu	Samburu	Samburu	Samburu	Samburu	Kenya
Town of Birth	Arsim	Nguronit	Latakweny	Arsim	Lodosit	
Province	Samburu	Samburu	Samburu	Samburu	Samburu	
203 KS	Samburu	Samburu	Samburu	Samburu	Samburu	Kenya
Town of Birth	Lodokogen	Koriyen		Korigen		
Province		Samburu		Samburu		
204 KS	Samburu	Samburu	Samburu	Samburu	Samburu	Kenya
Town of Birth	Ledera	Ngiro	Ngirerit	Ngirerit		
Province	Samburu	Samburu	Samburu	Samburu		
206 KS	Samburu	Samburu	Samburu	Samburu	Samburu	Kenya
Town of Birth	Lopungokwe	Bahawa	Bahawa	Sugutamarmar		
Province	Samburu	Samburu	Samburu	Samburu		
231 KS	Turkana	Samburu	Samburu	Turkana	Turkana	Kenya
Town of Birth	Baragoi	Baragoi				
Province	Maralal	Maralal	Maralal			
296 KS	Maasai	Samburu	Samburu	Maasai	Maasai	Kenya
Town of Birth	Rumuruti	Maralal	Maralal	Maralal	Marsabit	
Province	Laikipia	Samburu	Samburu	Samburu	Marsabit	
560 KS	Samburu	Samburu	Samburu	Samburu	Samburu	Kenya
Town of Birth	Logutane	Logutane		Logutane	Logutane	
Province	Samburu	Samburu		Samburu	Samburu	
561 KS	Samburu	Samburu	Samburu	Samburu	Samburu	Kenya
Town of Birth	Ngati	Kisuma		Poro	Poro	
Province						
562 KS	Samburu	Samburu	Samburu	Samburu	Samburu	Kenya
Town of Birth	Lolkunono	Marti	Marti	Lokunono	Lokunono	
Province						
563 KS	Samburu	Samburu	Samburu	Samburu	Samburu	Kenya
Town of Birth	Tamiyoi	Tamiyoi	Suguta	Lodokejek	Lodokejek	
Province	Maralal	Maralal				
564 KS	Samburu	Samburu	Samburu	Samburu	Samburu	Kenya
Town of Birth	Suguta	Lkerei	Wamba	Suguta	Suguta	
Province						
567 KS	Samburu	Samburu	Samburu	Samburu	Samburu	Kenya
Town of Birth	Power	Power	Power	Power	Power	
Province	Samburu	Samburu	Samburu	Samburu	Samburu	
571 KS	Samburu	Samburu	Samburu	Samburu	Samburu	Kenya
Town of Birth	Poro	Poro	Siambu	Lesinde	Poro	
Province	Samburu	Samburu	Samburu	Samburu	Samburu	
573 KS	Samburu	Samburu	Samburu	Samburu	Samburu	Kenya
Town of Birth	Siambu	Morijo	Ngorika	Ndoto	Ndoto	
Province						

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
574 KM	Samburu	Samburu	Samburu			Kenya
Town of Birth	Simisi	Latakweng	Leraka	Lekerae	Ilaut	
Province						
576 KS	Samburu	Samburu	Samburu	Samburu	Samburu	Kenya
Town of Birth	Baragoi	Lesikan	Loroki	Baragoi	Ndoto	
Province	Samburu	Samburu	Samburu	Samburu	Samburu	
577 KS	Samburu	Samburu	Samburu	Samburu	Samburu	Kenya
Town of Birth	Opiriol	Waso		Opiriol	Opiriol	
Province	Samburu	Samburu	Lakipia	Samburu	Samburu	

10.2.30 Shuwa

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
709 NS	Kanuri	Shuwa	Shuwa	Kanuri	Kanuri	Nigeria
Town of Birth	Maiduguri	Monguno	Monguno	Mongono	Mongono	
Province	Borno	Borno	Borno	Borno	Borno	
711 NS	Shuwa	Shuwa	Shuwa	Shuwa	Shuwa	Nigeria
Town of Birth	Maiduguri	Maiduguri	Monguno	Maiduguri	Monguno	
Province	Borno	Borno	Borno	Borno	Borno	
734 NS	Shuwa	Shuwa	Shuwa	Shuwa	Shuwa	Nigeria
Town of Birth	Gwange	Gwange	Gwange	Gwange	Gwange	
Province	Borno	Borno	Borno	Borno	Borno	
860 NS	Shuwa	Shuwa	Shuwa	Shuwa	Shuwa	Nigeria
Town of Birth	Maka	Maiduguri	'Maiduguri	Maiduguri	Maiduguri	
Province		Borno	Borno	Borno	Borno	

10.2.31 Somali

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
202 KZ	Somali	Somali	Somali	Somali	Somali	Kenya
Town of Birth	Isiolo	Isiolo	Isiolo			
Province	Isiolo	Isiolo	Isiolo	Ethiopia	Ethiopia	
213 KZ	Somali	Somali	Somali	Somali	Somali	Kenya
Town of Birth	Isiolo	Wajir	Wajir	Wajir	Wajir	
Province	Isiolo	Wajir	Wajir	Wajir	Wajir	
222 KZ	Somali	Somali	Somali	Somali	Somali	Kenya
Town of Birth	Bulapesa	Mandera	Mandera			
Province	Isiolo	Mandera	Mandera			
225 KZ	Somali	Somali	Somali	Somali	Somali	Kenya
Town of Birth	Wajir		Buna	Buna	Buna	
Province	Wajir		Wajir	Wajir	Wajir	
226 KZ	Somali	Somali	Somali	Somali	Somali	Kenya
Town of Birth	Mado	Mado	Mado			
Province	Isiolo	Isiolo	Isiolo			
229 KZ	Somali	Somali	Somali	Somali	Somali	Kenya
Town of Birth	Bulapesa	Isiolo				
Province	Isiolo	Isiolo				
235 KZ	Somali	Somali	Somali	Somali	Somali	Kenya
Town of Birth	Kilimani	Wajir	Wajir	Girifta	Girifta	
Province	Isiolo	Wajir	Wajir	Wajir	Wajir	
238 KZ	Somali	Somali	Somali	Somali	Somali	Kenya
Town of Birth	Giriftu	Giriftu	Giriftu	Giriftu	Giriftu	
Province	Wajir	Wajir	Wajir	Wajir	Wajir	

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
241 KZ	Somali	Somali	Somali	Somali	Somali	Kenya
Town of Birth	Wajir	Elwak	Mandera	Elwak	Simprifade	
Province	Wajir	Mandera	Mandera	Mandera		
500 EZ	Somali	Somali	Somali	Somali	Somali	Kenya
Town of Birth	Warbiweye	Warbiweye	Warbiweye	Dolow	Amino	
Province	Ethiopia	Ethiopia	Ethiopia	Ethiopia	Ethiopia	
501 EZ	Somali	Somali	Somali	Somali	Somali	Kenya
Town of Birth	Mandora	Nakele		Moyale		
Province	Ethiopia	Ethiopia	Ethiopia	Moyale		
502 ZZ	Somali	Somali	Somali	Somali	Somali	Kenya
Town of Birth	Elwak	Elwak		Mandera	Elwak	
Province			Somalia	Mandera		
503 EZ	Somali	Somali	Somali	Somali	Somali	Kenya
Town of Birth	Aminou	Aminou				
Province	Ethiopia	Ethiopia	Ethiopia	Somalia	Somalia	
505 EZ	Somali	Somali	Somali	Somali	Somali	Kenya
Town of Birth	Warir	Wajir		Wajir		
Province	Wajir	Wajir	Ethiopia	Wajir	Ethiopia	
506 EZ	Somali	Somali	Somali	Somali	Somali	Kenya
Town of Birth	Moyale	Moyale	Nageyle	Moyale	Nageyle	
Province	Moyale	Moyale	Ethiopia	Moyale	Ethiopia	
507 EZ	Somali	Somali	Somali	Somali	Somali	Kenya
Town of Birth	Mandera					
Province	Mandera	Somalia	Ethiopia	Ethiopia	Somalia	
508 EZ	Somali	Somali	Somali	Somali	Somali	Kenya
Town of Birth	Mandera	Mandera		Mandera		
Province	Mandera	Mandera	Ethiopia	Mandera	Somalia	
509 EZ	Somali	Somali	Somali	Somali	Somali	Kenya
Town of Birth	Garissa	Modogashe				
Province			Ethiopia	Somalia	Somalia	
510 EZ	Somali	Somali	Somali	Somali	Somali	Kenya
Town of Birth	Wajir	Wajir		Wajir		
Province	Wajir	Wajir	Ethiopia	Wajir	Somalia	
511 EZ	Somali	Somali	Somali	Somali	Somali	Kenya
Town of Birth	Mandera	Mandera		Mandera		
Province	Mandera	Mandera	Ethiopia	Mandera	Ethiopia	
512 KZ	Somali	Somali	Somali	Somali	Somali	Kenya
Town of Birth	Wajir	Wajir	Wajir	Mandera	Mandera	
Province	Wajir	Wajir	Wajir	Mandera	Mandera	
513 KZ	Somali	Somali	Somali	Somali	Somali	Kenya
Town of Birth	Mandera	Rhamu	Rhamu	Rhamu		
Province	Mandera				Ethiopia	
514 EZ	Somali	Somali	Somali	Somali	Somali	Kenya
Town of Birth	Mandera			Mandera		
Province	Mandera	Ethiopia	Ethiopia	Mandera	Ethiopia	
515 EZ	Somali	Somali	Somali	Somali	Somali	Kenya
Town of Birth	Nakele	Mandera	Sidamo			

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
Province	Ethiopia	Mandera	Ethiopia	Somalia	Somalia	
516 ZZ	Somali	Somali	Somali	Somali	Somali	Kenya
Town of Birth	Bulahaya	Luugh	Luugh	Garbahare	Luugh	
Province	Somalia	Somalia	Somalia	Somalia	Somalia	
517 KZ	Somali	Somali	Somali	Somali	Somali	Kenya
Town of Birth	Mandera	Elwak	Elwak	Mandera	Mandera	
Province	Mandera			Mandera	Mandera	
518 ZZ	Somali	Somali	Somali	Somali	Somali	Kenya
Town of Birth	Luugu	Bokol	Bokol	Luugh	Konai	
Province	Somalia	Somalia	Somalia	Somalia	Somalia	
519 ZZ	Somali	Somali	Somali	Somali	Somali	Kenya
Town of Birth	Namanga	Moshi	Hargesa	Hargesa	Hargesa	
Province		Tanzania	Somalia	Somalia	Somalia	

10.2.32 Songhai

264 MS	Songhai	Songhai	Songhai	Songhai	Songhai	Niger
Town of Birth	Niamey	Tahoua				
Province		Tahoua	Mali	Tillaberi	Tillaberi	
265 RS	Hausa	Songhai	Songhai	Hausa	Hausa	Niger
Town of Birth	Tera	Famale	Famale	Tahoua	Tahoua	
Province		Tillaberi	Tillaberi	Tahoua	Tahoua	
426 RS	Songhai	Songhai	Songhai	Songhai	Songhai	Niger
Town of Birth	Hamdallaye	Hamdallaye	Hamdallaye	Hamdallaye	Hamdallaye	
Province						
427 RS	Songhai	Songhai	Songhai	Songhai	Songhai	Niger
Town of Birth	Niamey	Niamey	Dosso	Niamey	Dosso	
Province						
428 RS	Songhai	Songhai	Songhai	Songhai	Songhai	Niger
Town of Birth	Dosso	Dosso	Dosso	Dosso	Dosso	
Province						
430 RS	Songhai	Songhai	Songhai	Songhai	Songhai	Niger
Town of Birth	Koygolo	Koygolo	Koygolo	Koygolo	Koygolo	
Province						
431 RS	Songhai	Songhai	Songhai	Songhai	Songhai	Niger
Town of Birth	Gardi	Gardi	Koygolo	Koygolo	Koygolo	
Province						
432 RS	Songhai	Songhai	Songhai	Songhai	Songhai	Niger
Town of Birth	Niamey	Niamey	Niamey	Niamey	Niamey	
Province						
433 RS	Songhai	Songhai	Songhai	Songhai	Songhai	Niger
Town of Birth	Niamey	Niamey	Dantchandou	Dantchandou	Dantchandou	
Province						
434 RS	Songhai	Songhai	Songhai	Songhai	Songhai	Niger
Town of Birth	Agadez	Niamey	Dambon	Niamey	Tillberi	
Province						
436 RS	Songhai	Songhai	Songhai	Songhai	Songhai	Niger
Town of Birth	Bawabou	Loga	Loga	Koygolo	Koygolo	
Province						

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country

10.2.33 Tera

736 NF	Tera	Tera	Tera	Tera	Tera	Nigeria
Town of Birth	Kaduna	Jos	Bui	Bauchi	Gombe	
Province	Kaduna	Jos	Borno	Bauchi		

10.2.34 Teso

56 K?	Teso	Teso	Teso	Teso	Teso	Kenya
Town of Birth						
Province	Busia	Busia	Busia	Busia	Busia	

10.2.35 Tiv

701 NF	Tiv	Tiv	Tiv	Tiv	Tiv	Nigeria
Town of Birth						
Province	Benue	Benue	Benue	Benue	Benue	

10.2.36 Tuareg

254 MT	Tuareg	Tuareg	Tuareg	Tuareg	Tuareg	Niger
Town of Birth	Adaraubouko	Anékou	Anékou	Adaraubokou	Adaraubokou	
Province						
263 RT	Tuareg	Tuareg	Tuareg	Tuareg	Tuareg	Niger
Town of Birth	Talcho					
Province	Filingué	Mali	Azaouak	Azaouak	Azaouak	
438 RT	Tuareg	Tuareg	Tuareg	Tuareg	Tuareg	Niger
Town of Birth	Bonkoukou	Bonkoukou	Bonkoukou	Bonkoukou	Bonkoukou	
Province						
443 RT	Tuareg	Tuareg	Tuareg	Tuareg	Tuareg	Niger
Town of Birth	Niamey	Indikitan	Balleyara	Indikitan	Indikitan	
Province						
447 RT	Tuareg	Tuareg	Tuareg	Tuareg	Tuareg	Niger
Town of Birth	Baleyara	Baleyara	Imbama	Balleyara	Imbama	
Province						
449 RT	Tuareg	Tuareg	Tuareg	Tuareg	Tuareg	Niger
Town of Birth	Balleyara	Imbama	Imbama	Sassabadé	Sassabadé	
Province						
450 RT	Tuareg	Tuareg	Tuareg	Tuareg	Tuareg	Niger
Town of Birth	Balleyara	Balleyara	Balleyara	Balleyara	Balleyara	
Province						
451 RT	Tuareg	Tuareg	Tuareg	Tuareg	Tuareg	Niger
Town of Birth	Balleyara	Balleyara	Balleyara	Balleyara	Balleyara	
Province						
452 RT	Tuareg	Tuareg	Tuareg	Tuareg	Tuareg	Niger
Town of Birth	Agadez	Elmiki	Elmiki	Tarwadji	Tarwadji	
Province						
453 RT	Tuareg	Tuareg	Tuareg	Tuareg	Tuareg	Niger
Town of Birth	Zinder	Tabla	Tabla	Tabla	Tabla	
Province						
454 RT	Tuareg	Tuareg	Tuareg	Tuareg	Tuareg	Niger
Town of Birth	Kochilan	Tilloa	Tilloa	Kochilan	Kochilan	
Province						

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
455 RT	Fulbe	Tuareg	Tuareg	Fulbe	Fulbe	Niger
Town of Birth	Agadez	Tahoua	Tahoua	Magazia	Sokoto	
Province						
458 RT	Tuareg	Tuareg	Tuareg	Tuareg	Tuareg	Niger
Town of Birth	Ourihimaza	Ourihinaza	Kohehé	Ourihimaza	Ourihimaza	
Province						
459 RT	Tuareg	Tuareg	Tuareg	Tuareg	Tuareg	Niger
Town of Birth	Diguina	Diguina	Diguina	Diguina	Diguina	
Province						
460 RT	Tuareg	Tuareg	Tuareg	Tuareg	Tuareg	Niger
Town of Birth	Bonkoukou	Bonkoukou	Bonkoukou	Bonkoukou	Bonkoukou	
Province						
461 RT	Tuareg	Tuareg	Tuareg	Tuareg	Tuareg	Niger
Town of Birth	Agadez	Aderbissamat	Aderbissamat	Aderbissamat	Aderbissamat	
Province						
462 RT	Tuareg	Tuareg	Tuareg	Tuareg	Tuareg	Niger
Town of Birth	Aderbissnet	Aderbissnet	Aderbissnet	Aderbissnet	Aderbissnet	
Province						
463 RT	Tuareg	Tuareg	Tuareg	Tuareg	Tuareg	Niger
Town of Birth	Naredet	Naredet	Nareget	Naredet	Naredet	
Province						
464 RT	Tuareg	Tuareg	Tuareg	Tuareg	Tuareg	Niger
Town of Birth	Amataltal	Agadez	Agadez	Teguirwet	Agadez	
Province						
465 RT	Tuareg	Tuareg	Tuareg	Tuareg	Tuareg	Niger
Town of Birth	Tiguidan Ta	Agadez	Agadez	Agadez	Agadez	
Province						
466 RT	Tuareg	Tuareg	Tuareg	Tuareg	Tuareg	Niger
Town of Birth	Toumbalaga	Adubisnet	Adubisnet	Adubisnet	Adubisnet	
Province						
467 RT	Tuareg	Tuareg	Tuareg	Tuareg	Tuareg	Niger
Town of Birth	Dakori	Wourhamidja	Wourhamidja	Wourhamidja	Ally	
Province						
468 RT	Tuareg	Tuareg	Tuareg	Tuareg	Tuareg	Niger
Town of Birth	Niamey	Abala	Abala	Abala	Abala	
Province						
469 RT	Tuareg	Tuareg	Tuareg	Tuareg	Tuareg	Niger
Town of Birth	Agadez	Adubissnet	Adubissnet	Goffat	Tcheroseme	
Province						
471 RT	Tuareg	Tuareg	Tuareg	Tuareg	Tuareg	Niger
Town of Birth	Agadez	Aderbissnet	Aderbissnet	Aderbissnet	Aderbissnet	
Province						
472 RT	Tuareg	Tuareg	Tuareg	Tuareg	Tuareg	Niger
Town of Birth	Agadez	Aderbissnet	Aderbissnet	Aderbissnet	Aderbissnet	
Province						
473 RT	Tuareg	Tuareg	Tuareg	Tuareg	Tuareg	Niger
Town of Birth	Arlit	Agadez	Tacrisat	Tacrisat	Tacrisat	
Province						

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
474 RT	Tuareg	Tuareg	Tuareg	Tuareg	Tuareg	Niger
Town of Birth	Balleyara	Balleyara	Fafa	Balleyara	Toumbora	
Province						

10.2.37 Turkana

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
19 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Lokitaung	Lokitaung	Lokitaung	Lowaregaka	Lowaregaka	
Province	Turkana	Turkana	Turkana	Turkana	Turkana	
93 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Marala	Baragoi	Lomutono	Kerio	Kerio	
Province	Samburu	Samburu	Turkana	Turkana	Turkana	
99 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Baragoi	Baragoi	Baragoi	Baragoi	Baragoi	
Province	Marsabit	Marsabit	Marsabit	Marsabit	Marsabit	
101 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Lorrukumi	Lorukumi	Lorukumi	Lorukumi	Lorukomi	
Province	Turkana	Turkana	Turkana	Turkana	Siaya	
102 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Loima	Loima	Loima	Loima	Loima	
Province	Turkana	Turkana	Turkana	Turkana	Turkana	
103 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Lokichar	Lokichar	Lokichar	Lokichar	Lokichar	
Province	Turkana	Turkana	Turkana	Turkana	Turkana	
104 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Lokori	Lokori	Lokori	Lokori	Lokori	
Province	Turkana	Turkana	Turkana	Turkana	Turkana	
105 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Kalokol	Kalokol	Kalokol	Kalokol	Kalokol	
Province	Turkana	Turkana	Turkana	Turkana	Turkana	
106 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Lokitaung	Lokitaung	Lokitaung	Lokitaung	Lokitaung	
Province	Turkana	Turkana	Turkana	Turkana	Turkana	
107 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Kaputir	kaputir	Kaputir	Kaputir	Kaputir	
Province	Turkana	Turkana	Turkana	Turkana	Turkana	
108 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Kakuma	Kakuma	Kakuma	Kakuma	Kakuma	
Province	Turkana	Turkana	Turkana	Turkana	Turkana	
109 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Lokitaung	Lokitaung	Lokitaung	Lokitaung	Lokitaung	
Province	Turkana	Turkana	Turkana	Turkana	Turkana	
110 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Lurugum	Lorugum	Lorugum	Lorugum	Lorugum	
Province	Turkana	Turkana	Turkana	Turkana	Turkana	
111 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Lorugum	Lorugum	Lorugum	Lorugum	Lorugum	
Province	Turkana	Turkana	Turkana	Turkana	Turkana	

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
112 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Lokori	Lokori	Lokori	Lokori	Lokori	
Province	Turkana	Turkana	Turkana	Turkana	Turkana	
113 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Lokogum	Lokogum	Lokogum	Lokogum	Lokogum	
Province	Turkana	Turkana	Turkana	Turkana	Turkana	
114 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Kerio	Lorugum	Lorugum	Kalokol	Kalokol	
Province	Turkana	Turkana	Turkana	Turkana	Turkana	
116 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Lorugum	Lorugum	Lorugum	Lorugum	Lorugum	
Province	Turkana	Turkana	Turkana	Turkana	Turkana	
117 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Kalokol	Kalokol	Kalokol	Kalokol	Kalokol	
Province	Turkana	Turkana	Turkana	Turkana	Turkana	
118 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Lodwar	Lodwar	Lodwar	Lodwar	Lodwar	
Province	Turkana	Turkana	Turkana	Turkana	Turkana	
120 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Lokori	Lokori	Lokori	Lokori	Lokori	
Province	Turkana	Turkana	Turkana	Turkana	Turkana	
122 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Lodwar	Lodwar	Lodwar	Lodwar	Lodwar	
Province	Turkana	Turkana	Turkana	Turkana	Turkana	
123 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Kerio	Kerio	Kerio	Kerio	Kerio	
Province	Turkana	Turkana	Turkana	Turkana	Turkana	
127 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Elike Sprin	Lorugum	Lorugum	Lokitaung	Lokitaung	
Province	Turkana	Turkana	Turkana	Turkana	Turkana	
128 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Fergusson G	Fergusson Gu	Fergusson Gu	Fergusson Gul	Fergussin Gulf	
Province	Turkana	Turkana	Turkana	Turkana	Turkana	
129 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Kalekol	Kalekol	Kalekol	Kalekol	Kalekol	
Province	Turkana	Turkana	Turkana	Turkana	Turkana	
131 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Kalekol	Kakuma	Kakuma	Kakuma	Kakuma	
Province	Turkana	Turkana	Turkana	Turkana	Turkana	
168 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Marti	Marti	Marti	Marti	Marti	
Province	Samburu	Samburu	Samburu	Samburu	Samburu	
173 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Moyale	Isiolo		Baragoi		
Province	Marsabit	Isiolo		Maralal		
175 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Maralal	Baragoi		Baragoi	Maralal	
Province	Maralal	Maralal		Maralal	Maralal	

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
179 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth		Isiolo		Baragoi	Baragoi	
Province	Isiolo	Isiolo		Maralal	Maralal	
205 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Maralal	Lodwar	Baragoi			
Province	Maralal	Turkana	Turkana			
209 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Chumui	Isiolo	Isiolo	Isiolo	Isiolo	
Province	Isiolo	Isiolo	Isiolo	Isiolo	Isiolo	
211 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Nakuru	Lenrumt	Isiolo	Lenrurmt	Lenrurmt	
Province	Nakuru	Kitale	Isiolo	Kitale	Kitale	
212 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Isiolo	Lodwar	Lodwar	Lodwar	Lodwar	
Province	Isiolo	Turkana	Turkana	Turkana	Turkana	
214 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Isiolo	Lodwar	Lodwar	Kapecho		
Province	Isiolo	Lodwar	Lodwar	Lodwar		
218 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Isiolo	Baragoi	Baragoi	Baragoi	Baragoi	
Province	Isiolo	Samburu	Samburu	Samburu	Samburu	
243 KT	Turkana	Turkana	Turkana	Turkana	Turkana	Kenya
Town of Birth	Baragoi	Nachola	Nachola	Sokota	Sokota	
Province	Maralal	Maralal	Maralal	Maralal	Maralal	

10.2.38 Yoruba

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
339 NY	Yoruba	Yoruba	Yoruba	Yoruba	Yoruba	Nigeria
Town of Birth	Saki	Saki	Saki	Saki	Saki	
Province	Oyo	Oyo	Oyo	Oyo	Oyo	
526 NY	Hausa	Yoruba	Yoruba	Hausa		Nigeria
Town of Birth	Sokoto	Ilorin	Ilorin	Sokoto	Sokoto	
Province	Sokoto	Kwara	Kwara	Sokoto	Sokoto	
717 NY	Yoruba	Yoruba	Yoruba	Yoruba	Yoruba	Nigeria
Town of Birth	Ede	Ede	Ede	Ede	Ede	
Province	Osun	Osun	Osun	Osun	Osun	
720 NY	Yoruba	Yoruba	Yoruba	Yoruba	Yoruba	Nigeria
Town of Birth	Ede	Ede	Ede	Ede	Ede	
Province	Osun	Osun	Osun	Osun	Osun	
748 NY	Yoruba	Yoruba	Yoruba	Yoruba	Yoruba	Nigeria
Town of Birth	Togo	Saki	Saki	Saki	Saki	
Province		Oyo	Oyo	Oyo	Oyo	
749 NY	Yoruba	Yoruba	Yoruba	Yoruba	Yoruba	Nigeria
Town of Birth	Kangi	kisi	Kisi	Kisi	Kisi	
Province	Oyo	Oyo	Oyo	Oyo	Oyo	
750 NY	Yoruba	Yoruba	Yoruba	Yoruba	Yoruba	Nigeria
Town of Birth	Iseyin	Iseyin	Iseyin	Iseyin	Iseyin	
Province	Oyo	Oyo	Oyo	Oyo	Oyo	

Code	individual	Mother	M. Grandmother	Father	P. Grandfather	Country
881 NY	Yoruba	Yoruba	Yoruba	Yoruba	Yoruba	Nigeria
Town of Birth	Sepeteri	Sabe	Sabe	Ago Amodu	Ago Amodu	
Province						
886 NY	Yoruba	Yoruba	Yoruba	Yoruba	Yoruba	Nigeria
Town of Birth						
Province						
887 NY	Yoruba	Yoruba	Yoruba	Yoruba	Yoruba	Nigeria
Town of Birth						
Province						
888 NY	Yoruba	Yoruba	Yoruba	Yoruba	Yoruba	Nigeria
Town of Birth	Okuta	Gbararo	Gbararo	Gbure	Gbure	
Province						
889 NY	Yoruba	Yoruba	Yoruba	Yoruba	Yoruba	Nigeria
Town of Birth	Sabe	Basii	Basii	Basii	Basii	
Province						
890 NY	Yoruba	Yoruba	Yoruba	Yoruba	Yoruba	Nigeria
Town of Birth	Agurege	Ago Are	Agurere	Ago Are	Agurere	
Province		Oyo		Oyo		
891 NY	Yoruba	Yoruba	Yoruba	Yoruba	Yoruba	Nigeria
Town of Birth	Okeho	Okeho	Okeho	Okeho	Okeho	
Province	Oyo	Oyo	Oyo	Oyo	Oyo	
892 NY	Yoruba	Yoruba	Yoruba	Yoruba	Yoruba	Nigeria
Town of Birth	Ago Are	Ago Are	Ago Are	Ago Are	Ago Are	
Province	Oyo	Oyo	Oyo	Oyo	Oyo	
893 NY	Yoruba	Yoruba	Yoruba	Yoruba	Yoruba	Nigeria
Town of Birth	Isemi	Isemi	Isemi	Isemi	Isemi	
Province						
894 NY	Yoruba	Yoruba	Yoruba	Yoruba	Yoruba	Nigeria
Town of Birth						
Province						
895 NY	Yoruba	Yoruba	Yoruba	Yoruba	Yoruba	Nigeria
Town of Birth	Ibadan	Ibadan				
Province	Oyo	Oyo				
896 NY	Yoruba	Yoruba	Yoruba	Yoruba	Yoruba	Nigeria
Town of Birth						
Province						
897 NY	Yoruba	Yoruba	Yoruba	Yoruba	Yoruba	Nigeria
Town of Birth	Saki	Igboho	Igboho	Saki	Saki	
Province	Oyo	Oyo	Oyo	Oyo	Oyo	
898 NY	Yoruba	Yoruba	Yoruba	Yoruba	Yoruba	Nigeria
Town of Birth	Komu	Komu	Komu	Komu	Komu	
Province						
899 NY	Yoruba	Yoruba	Yoruba	Yoruba	Yoruba	Nigeria
Town of Birth	Ago Amodu	Ago Amodu	Ago Amodo	Ago Amodu	Ago Amodu	
Province	Oyo	Oyo	Oyo	Oyo	Oyo	
900 NY	Yoruba	Yoruba	Yoruba	Yoruba	Yoruba	Nigeria
Town of Birth	Iseyin	Iseyin	Iseyin	Iseyin	Iseyin	
Province						

10.3 Sequences used in the Analysis

The following sequences were used in the analysis. These sequences were free from ambiguities between nucleotide positions 16030 and 16370 (numbered according to Anderson *et al.* 1981), the total sample was n = 1595. The sequences are arranged according to continent **AF** (Africa); **AM** (America); **AS** (Asia); **EU** (Europe); **ME** (Middle East).

Population	AF	AM	AS	EU	ME	REFERENCE
Aboriginal			1			Vigilant *et al.* 1991
Altai			17			Shields *et al.* 1993
Apache		1				Torroni *et al.* 1993a
Arab					12	Di Rienzo and Wilson 1991
Asian			21			Vigilant *et al.* 1991
Athapaskan		21				Torroni *et al.* 1993a
Basque				45		Bertranpetit *et al.*,1994
Bedouin					29	Di Rienzo and Wilson 1991
Bella Coola		44				Ward *et al.* 1993; Torroni *et al.* 1993
Biaka Pygmy	18					Vigilant *et al.* 1991
Borana	1					this thesis
Boruca		2				Torroni *et al.* 1993a
British				95		Piercy *et al.* 1993
Chukchi			7			Shields *et al.* 1993
Dogrib		2				Torroni *et al.* 1993a
Estonian				28		Sajantila *et al.* 1995
European				15		Vigilant *et al.* 1991
Evenk			11			Torroni *et al.* 1993b
Finns				48		Sajantila *et al.* 1995
Fulbe	61					this thesis
Guaymi		1				Torroni *et al.* 1993a
Haida		42				Ward *et al.* 1993; Torroni *et al.* 1993
Han			7			Torroni *et al.* 1993a
Hausa	20					this thesis
Havik			42			Mountain *et al.*,1995
Huetar			27			Santos *et al.* 1994
Icelander				40		Sajantila *et al.* 1995
Inuit		17	6			Shields *et al.* 1993
Inupiaq		5				Shields *et al.* 1993
Kadar			7			Mountain *et al.* 1995
Kanuri	14					this thesis
Kikuyu	25					this thesis
Karelian				83		Sajantila *et al.* 1995

Population	AF	AM	AS	EU	ME	REFERENCE
Korean			4			Torroni *et al.* 1993a
Kraho		3				Torroni *et al.* 1993a
Kuna		3				Batista *et al.* 1995
Kung	19					Vigilant *et al.* 1991
Makiritare		1				Torroni *et al.* 1993a
Mandenka	87					Graven *et al.* 1995
Mapuche		39				Torroni *et al.* 1993a
Mataco		3				Torroni *et al.* 1993a
Maya		4				Torroni *et al.* 1993a
Mbuti Pygmy	10					Vigilant *et al.* 1991
Mukri			38			Mountain *et al.*,1995
Navajo		2				Torroni *et al.* 1993a
New Guinean			5			Vigilant *et al.* 1991
Ngöbé		46				Kolman *et al.* 1995
Nivkh			2			Torroni *et al.* 1993b
Nuu-Chah-Nulth		63				Ward *et al.* 1991
Ojibwa		3				Torroni *et al.* 1993a
Pima		3				Torroni *et al.* 1993a
Saami				115		Sajantila *et al.* 1995
Sakuye	1					this thesis
Sardinian				69		Di Rienzo and Wilson 1991
Somali	27					this thesis
Songhai	10					this thesis
Swiss				74		Pult *et al.* 1994
Ticuna		3				Torroni *et al.* 1993a
Tlingit		2				Torroni *et al.* 1993a
Tuareg	26					this thesis
Turkana	37					this thesis
Udegey			3			Torroni *et al.* 1993b
Yakima		42				Shields *et al.* 1993
Yanomama		3				Torroni *et al.* 1993a
Yoruba	33					this thesis; Vigilant *et al.* 1991
TOTALS	**389**	**355**	**198**	**612**	**41**	

10.4 Shared Lineages

- 2CMa 2CMb 2CMc
- 37CM 38CM 39CM 40CM
- 41CM 42CM 43CM
- 47CM 48CM 48CMb
- 68CBb 68CB
- 96EB 123ET
- 257WF 34EK 89WY
- 326WF 348WF 259WH 528WH 155EK
- 329WF 330WF 738WF 747WF
- 331WF 892WY 897WY
- 332WF 333WF
- 336WF 235ES 453WT* 454WT*
- 338WF 347WF 457WF 591WF 710WF 724WF 873WF 534WH 857WK 434WS
- 341WF 870WF 551WH 10aWM 10bWM 10cWM 10dWM 10eWM 10fWM 10gWM 16aWM 6bWM 6cWM
- 342WF 531WF 597WF 874WF
- 344WF 437WF 731WK
- 345WF 349WF 537WH 14WM 45aWM 45bWM 263WT 465WT 26HWYa 26HWYb 346WF 340WH
- 435WF 441WF 598WF 723YF 856WF
- 444WF* 445WF *
- 726WF 543WH 58WM 458WT
- 869WF 136EK 513ES 449WT
- 335WH 526WY
- 540WH 106WY
- 862WK 471WT* 472WT*
- 864WK 865WK 109ET
- 66EK 565EK 427WS 436WS
- 68EKΨ 168ET Ψ
- 94EK 211ET
- 139EK 572EK 106ET
- 7aSK 7bSK 7cSK 7dSK 7eSK 7fSK
- 10SK 11SK 12SK 13SK
- 19SK 20SKa 20SKb
- 8SK 9SK
- 18aWM 18bWM 18cWM
- 25aWM 25bWM 25cWM
- 26WM 750WY
- 52aWM 52bWM 52cWM
- 40aWM 40bWM 40cWM
- 8aWM 8bWM
- 21aWM 21bWM 110ET
- 3aWM 3bWM 3cWM 3dWM 3eWM 3fWM 3gWM 3hWM 3iWM 3jWM 17WM
- 37aWM 37bWM
- 12aWM 12bWM 12cWM 12dWM 12eWM 12fWM 12gWM
- 1aWM 1bWM 1cWM 1dWM 1eWM 1fWM
- 2aWM 2bWM
- 27aWM 27bWM
- 23aWM 23bWM 23cWM
- 6aWM 6bWM 6cWM 6dWM
- 463WT* 464WT*
- 48aWM 48bWM
- 11aWM 11bWM
- 33aWM 33bWM
- 6CB 4CBa 4CBb 5CBa 5CBb 5CBc

---

* Indicates the samples collected in Niamey which may be a duplication.

Ψ May be a case of mislabelling.

- 69CB 73CB
- 213ES 510ES
- 226ES 466WT
- 503ES 505ES
- 506ES 508ES
- 512ES 105ET
- 428WS 459WT
- 452WT 468WT
- 117ET 118ET 212ET

10.5 Sequences generated in this study

10.5.1 Fulbe Sequences n=61

```
              1111111 1111111122 2222222222 2222222222 2333333
           5891222447 7788888901 1122333346 6667778999 9011267
           1631469582 3923789293 4923458940 3450486123 4919028
Anderson   ATTCTTGGCT CCAACCTCTG CACCCATCGC TCACGCCCCA CATGCTC
  257RF    .........C ...C..C... ...T...... .......... ....T..
  326NF    .......... ......CT.. ...T...... .....T.... TG.....
  327NF    .....C.... ....T.C..A ...T.....T .T..AT...G ..C....
  328NF    .......... ......CT.. .......... ...T...... ....T..
  329NF    .......A.. .......... ..T....... .......... .......
  330NF    .......A.. .......... ..T....... .......... .......
  331NF    ....C..... .......... ...T...... .....T.... .......
  332NF    .......... ........C. ...T.G.... ........T. ..C....
  333NF    .......... ........C. ...T...... ........T. ..C....
  334NF    .C..C..... ...G...... ...T...... .......... ..C....
  336NF    .......... ........C. ...T...... ........T. ..C....
  337NF    ....C..... .......... ...T...... .......... ...A.C.
  338NF    ....C..... .......... ...T...... .....T.... .....C.
  341NF    .......... .......... ...T...... .....T.... .......
  342NF    ..C....... .......... ...T...... .....T.... .....C.
  343NF    ....C..... .......... ...T...... .......... ...A...
  344NF    .......... .......... ...T...... .....T.... T......
  345NF    .....C.... ....T.C... ...T...... .T.T.T...G ..C....
  346NF    .........C ......C... ...T...... .......... ....T..
  347NF    ....C..... .......... ...T...... .....T.... .....C.
  348NF    .......... ......CT.. ...T...... .....T.... TG.....
  349NF    .....C.... ....T.C... ...T...... .T.T.T...G ..C....
  435RF    ..C..C.... ....T.C... ...T...... .T.T.T...G ..C....
  437BF    .......... .......... ...T...... .....T.... T......
  441RF    ..C..C.... ....T.C... ...T...... .T.T.T...G ..C....
  442RF    ........T. .......... ...T....A. .....TT... TG.....
  444RF    ..C....... ......CT.. ...T...... .....T.... TG.....
  445RF    ..C....... ......CT.. ...T...... .....T.... TG.....
  446RF    .......... .TCC..C... ...T..T... .......... ..C.TC.
  448RF    .........C ......C... .G........ .....T.... .......
  456RF    ....C..... .......... ...T..C... C......... .G.....
  457RF    ....C..... .......... ...T...... .....T.... .....C.
  527NF    .....C.... ....T.C..A ...T.....T .T.T.T...G ..C....
  531NF    ..C....... .......... ...T...... .....T.... .....C.
  544NF    .......... ....T.CT.. ...T...... .....T.... TG.....
  553BF    .....C.... ....T.C... T..T...... .TGT.T...G ..C....
  591NF    ....C..... .......... ...T...... .....T.... .....C.
  597NF    ..C....... .......... ...T...... .....T.... .....C.
  598NF    ..C..C.... ....T.C... ...T...... .T.T.T...G ..C....
  710NF    ....C..... .......... ...T...... .....T.... .....C.
  718NF    ...TC..... .......... ...T...... .......... ...C...
  719NF    .......A.C ...C.TC... ...T...... .......... ....T..
  722NF    G...C..... .......... ...T...... .......T.. .......
```

```
             1111111 1111111122 2222222222 2222222222 2333333
          5891222447 7788888901 1122333346 6667778999 9011267
          1631469582 3923789293 4923458940 3450486123 4919028
Anderson  ATTCTTGGCT CCAACCTCTG CACCCATCGC TCACGCCCCA CATGCTC
 723YF    ..C..C.... ....T.C... ...T...... .T.T.T...G ..C....
 724NF    ....C..... .......... ...T...... .....T.... .....C.
 725NF    .......... ........CA ...TT..... .......... .......
 726NF    .......... .......... ...T...... .......... ....T..
 728CF    .......... .......... ...T...... .......... ..C.T..
 737NF    .......C.. .......... ..T....... .......... .......
 738NF    .......A.. .......... ..T....... .......... .......
 746NF    .........C ..CC..C... ...T...... .....T.... ....T..
 747NF    .......A.. .......... ..T....... .......... .......
 854NF    .......... ........C. ...T...... .......... ..C....
 855NF    .......... .........A .......... .....T.... TG.....
 856NF    ..C..C.... ....T.C... ...T...... .T.T.T...G ..C....
 866NF    .........C T......... .G........ .......... ..C.T..
 869NF    .......... .......... ...T...... .....T.... TG....T
 870NF    .......... .......... ...T...... .....T.... .......
 872NF    .....C...C ....T.C... ...T...... ...T.T.... ..C....
 873NF    ....C..... .......... ...T...... .....T.... .....C.
 874NF    ..C....... .......... ...T...... .....T.... .....C.
```

10.5.2 Hausa Sequences n=20

```
             11111111 1222222222 2223333333 333
           5922278888 9012666778 9990001125 566
           1346922379 2953145086 2340191894 702
ANDERSON   ATTTGTAACT CTACCCACCC CACACATACC TCT
  256RH    ..C...CC.C ...T....T. ......C... ...
  259RH    .........C T..T....T. ..T..G.... ...
  260NH    .......... .C.TT..... T.....C... ...
  261RH    ....A....C ........T. ...G.....T ..C
  335NH    .........C ...T....T. ..T..G.... ...
  340NH    .....C...C ...T...... ........T. ...
  528NH    .........C T..T....T. ..T..G.... ...
  529NH    ..C....... ...T....T. ......C... ..C
  530NH    .C..A..C.C ..GT....T. ......C... .T.
  532NH    .........C T.......T. .......... ...
  534NH    ..C....... ...T....T. .......... ..C
  535NH    G......... ...T....T. .......G.. ..C
  537NH    ...C....TC ...T.T.TT. .G....C... ...
  538NH    .........C ...T..T... .......... ...
  539NH    .......... ...T....TT ..T.TG.... ...
  540NH    .C........ ...T..T... .......... ...
  543NH    .......... ...T...... ........T. ...
  550NH    .........C T..T....T. ..T..G.... C..
  551NH    .......... ...T....T. .......... ...
  552NH    .C........ ...T....T. .......... ...
```

10.5.3 Kanuri Sequences n=14

```
            11111111 1111122222 2222223333 33333
          4912244677 8888802367 7889990001 12556
          1214658612 1347993940 8682341491 80452
Anderson  ATCTTGCAAT AACCTTCCCC CCTCACCTAT ACCCT
  262RK   .C....T... ....C.T... TT......G. .....
  706NK   ...C...... ......T... ..C....... .....
  707NK   .......... ....CCT... ...T.....C .T...
  731NK   .......... ......T... T....T.... .....
  852NK   ..A..A..G. ..T.C.TT.. T..T.....C ...T.
  857NK   ...C...... ......T... T......... ....C
  859NK   ...C.....C ......T... .......C.. .....
  861NK   .........C .C..C.T... .........C .T...
  862NK   .......... .....CT... T....TT... ..T..
  863NK   ....C..... ....C..... T......... .....
  864NK   G......... ......T... .......... .....
  865NK   G......... ......T... .......... .....
  867NK   .......... G.....T... T....T..G. .....
  868NK   ....C..-.C ...TC.T.TT T...G....C G....
```

10.5.4 Kikuyu Sequences n=25

```
              11111111 1111111122 2222222222 2222222333
            6912456667 7888888901 1112335566 6788999001
            7319836892 9356789293 7893034604 5878034391
Anderson    CTCGCGACCT CACCCCTCTG TCACAAACCC ACCTCACGAT
  027KK     .......... ......C... ...T...... ..G..T...C
  034KK     .........C .C....C... ...T...... ..........
  066KK     .........C .C....C... ..G....... .T........
  068KK     ....T....C ....TGC... ...TG..... .........C
  094KK     .......... .......... ...T..G... ..........
  097KK     ...AT..T.C ....TGC... ...TG..... .T...G...C
  098KK     ..TAT.G... ....T.C... ...T.GG... .T.C......
  115KK     .......... ..T....... ...T...... .........C
  136KK     .......... .......... ...T...... .T....T.G.
  138KK     .......... T.T....... ...T....T. .........C
  139KK     .......... .......... ...T...... .........C
  140KK     .......... .......... .T.T...... .......A..
  141KK     ...A....TC ....T.C..A ...TG..... .T.......C
  142KK     .......... ...T....C. ...T...... .........C
  150KK     .......... .......... ...T...... .........C
  155KK     .......... ......CT.. ...T...... .T....T.G.
  156KK     ...AT....C ....TGC... C..TG..... .........C
  165KK     .C........ .......... ...T....T. .T.......C
  169KK     ..TAT.G... ....T.C... ...T..G... .T........
  170KK     .....A.... .......... ...T...... T.........
  172KK     T......... .......... ...T...... ..........
  565KK     .........C .C....C... ..G....... .T........
  566KK     .......... .......... ...T...T.T .T....T...
  568KK     ...AT..T.C ....TGC... ...TG..... ....T....C
  572KK     .......... .......... ...T...... .........C
```

```
          333333333
          122255666
          605745028
Anderson  ACTCCCCTT
  027KK   .....T.C.
  034KK   .T.......
  066KK   .........
  068KK   .T.......
  094KK   G........
  097KK   .T.......
  098KK   .........
  115KK   ...T.....
  136KK   .........
  138KK   .........
  139KK   ....T....
  140KK   ......T..
  141KK   ...T....C
  142KK   ...T.....
  150KK   .........
  155KK   .........
  156KK   .T.......
  165KK   .........
  169KK   ......T..
  170KK   .........
  172KK   ..-T.....
  565KK   .........
  566KK   .........
  568KK   .T.......
  572KK   ....T....
```

10.5.5 Somali Sequences n=27

```
               111111 1111111111 2222222222 2222223333 33333
           5699122344 6678888889 0112234467 7899990001 15566
           1623169678 8922347892 9383408904 8412343591 95902
Anderson   AATTCTGTCC CCTAACCCTC TGCCTACTCG CACCACGAAT GCTCT
  213KZ    ....A.A... ..C....... ...T..T... ..T......C .....
  222KZ    ...C...... ....C...C. ...T...... T....T.... .....
  225KZ    .......... ...CC...C. ..TT...... ......A... ...T.
  226KZ    .......... ....C...CT ...T...... T....T.... .....
  229KZ    .......... ..C..T.... ...T....T. .........C .....
  235KZ    .......... .......... C..T...... ...T.....C .....
  238KZ    G.....A... ...CC...C. ...T...C.. .........C .....
  241KZ    ........G. ..C....... .A.T..T... .......... .T...
  500KZ    .......... ....C...CT ...T...... T.T..T.... .....
  501KZ    ...C..A... .......... ...T...C.. .........C ..C..
  502ZZ    .......... ....C...C. ...T...... T.T..T.... .....
  503EZ    .......... .......... ...T....T. .........C .....
  505EZ    .......... .......... ...T....T. .........C .....
  506EZ    .....C.... .......... .......... .......T.. ....C
  507EZ    ......A..T T.C...TGC. ...T.G.... .G..G....C .....
  508KZ    .....C.... .......... .......... .......T.. ....C
  509EZ    .......... .T.......T ...T...... T......... .....
  510EZ    ....A.A... ..C....... ...T..T... ..T......C .....
  511EZ    .G.....C.. .T........ ...T...... T......... .....
  512KZ    .......... ...CC...C. ...T...C.. .........C .....
  513KZ    .......... .......... ...T...... T....T..G. .....
  514EZ    ..C....... ....C...CT ...T...... T.T..T.... .....
  515EZ    .....C.... ..C....... C..T...... .......... .....
  516ZZ    .......... .......... ....C..... .........C .....
  517KZ    .....C.... .......... .......... .......... ....C
  518ZZ    .......... ........CT ...T.....A T....T..G. .....
  519KZ    .......... .......... ...T...... .......... A....
```

10.5.6 Songhai Sequences n=10

```
             11111111 1111112222 2222223333 3
           3812224467 8888891236 7789990126 6
           8614695882 3678929304 0861349100 2
Anderson   ATCTTGGCCT ACCCTCACAC CCCCACATCC T
  264MS    GC...A.... ..T.C..T.. .T..GT...T C
  426RS    .......... ....CT.... .T...TG... .
  427RS    .........C C...C.G... .T........ .
  428RS    ...C...... .T.....T.. .T.T...... C
  430RS    ....C..... ..T.C....T TT..G..C.. .
  431RS    .......... .......T.. .TT..TG... .
  432RS    ..T..A.TTC ..TGC..TG. .......CT. .
  433RS    ......A... .......T.. TT........ C
  434RS    ...C...... .......T.. .T........ C
  436RS    .........C C...C.G... .T........ .
```

10.5.7 Tuareg Sequences n=26

```
           111111111 1111112222 2222222222 2233333333 3
          8122244677 8888890111 2345677799 9900112255 6
          6146958824 3678929389 3086404812 3419180746 2
Anderson  TCTTGGCCTC ACCCTCTGCA CACCCCGCCC ACCATACCCT T
  254MT   .......... .......A.. T......T.. .T.G...... .
  263RT   ...C...... ..T.C..... T...TT.T.. G...C..... .
  438RT   .......... .......... T......T.. .T.G.....C .
  443RT   .......... .......... T.T....... .......T.. .
  447RT   C.C....... .......... T......... ....C..... .
  449RT   .......... .......... T......T.. .T.G...... .
  450RT   .......... .......A.. T.....AT.. .T........ .
  451RT   .....A.... .......A.. T.....AT.. .T........ .
  452RT   .......... .......... .......... .......... .
  453RT   .......... ......C... T........T ....C..... .
  454RT   .......... ......C... T........T ....C..... .
  458RT   .......... .......... T......... ......T... .
  459RT   ..C....... .T........ T......TT. .......... C
  460RT   .......... C...C..... T......T.. .T.G...... .
  461RT   .......... ....C..... T......T.. .T........ .
  462RT   ....A....T .....T..T. T..A...... ....C..... C
  463RT   ........C. C...C....G .......T.. ....C..... .
  464RT   ........C. C...C....G .......T.. ....C..... .
  465RT   ...C...... ..T.C..... T...TT.T.. G...C..... .
  466RT   .......... C...CT.... T......T.. .T........ .
  467RT   .T..A.TTC. ..TGC..... TG.....T.. G...C.T... .
  468RT   .......... .......... .......... .......... .
  469RT   ....A..... ....C..... .......... .......... .
  471RT   .......... ......C... T......T.. .TT.....T. .
  472RT   .......... ......C... T......T.. .TT.....T. .
  474RT   .......... .......... T......T.. .....G.... .
```

10.5.8 Turkana Sequences n=37

```
                 111 1111111111 1111111122 2222222222
          4556779122 4666677777 8888888901 2223344455
          1123153169 8468902569 2346789298 3490313946
Anderson  AACTCTTCTG CAACCATACC AACCCCTCTC CTTAAATTAC
   093KT  .........A .......... .......... T.........
   099KT  ..T......A ....T...T. ....T.C... T..G......
   101KT  .......... .......... ....T.C... T..G..C...
   102KT  .......... ....T.C... ....T.C... T..G......
   103KT  .......... .......... .......... T......C..
   104KT  .........A T..T..C... .....GC... T..G......
   105KT  .......... .......... CC....C... T......C..
   106KT  .......... .......... .......... T.........
   107KT  .......... .......... .......... T.........
   108KT  .G........ TG....C... -C.T..C... T..G......
   109KT  G......... .......... .......... T.........
   110KT  .........A T..T..C... ....TGC... T..G......
   111KT  .........A .......... ........C. T.........
   112KT  .......... .......... ......CT.. T.........
   113KT  .G.C...... TG....C... -C.T..C... T..G......
   114KT  .......... .......... ......CT.. T........T
   116KT  ....T..T.A T.G....... ....T.C... T.........
   117KT  .......T.A T.G....... ....T.C... T...G...G.
   118KT  .......T.A T.G....... ....T.C... T...G...G.
   120KT  .........A T.G....... .-..T.C... T.........
   122KT  .......... .....G.... ......C... T.C.......
   123KT  .......... .......... .......... T....C....
   127KT  ......C... .......... ........CT T.........
   128KT  .........A T.....C... .....AC... T..G......
   129KT  .......... T.....C... ..T.TGC... T..G......
   131KT  .......... .......... .......... TC........
   168KT  .......... T.....C... ....TGC... T..G......
   173KT  .G.C...... TG....C... -C.T..C... T..G......
   175KT  .....C.... .........T ........T. T.........
   179KT  .......... .......... ........C. T.........
   205KT  .....C.... .......... .......... T.........
   209KT  .......... .......... .C....C... T.........
   211KT  .......... .......... .......... T.......G.
   212KT  .......T.A T.G....... ....T.C... T...G...G.
   214KT  ........C. ....T..... .......... T.........
   218KT  .........A T..T..C... ....TGC... T..G......
   243KT  .........A T..T..C... .....AC... T..G......
```

```
          2222222222 3333333333 33333
          6677899999 0011122234 55666
          4648902345 0916905754 45028
Anderson  CCGCACCACC AATAGCTCAC CCCTT
   093KT  .......... .......... .....
   099KT  ...T.T.... ..C...CT.. ....C
   101KT  ..AT.T.... G.C....... ...C.
   102KT  ...T...... ..C.A..T.. T...C
   103KT  ...T...TT. ..C....... .T.C.
   104KT  .......... ..C....... .....
   105KT  .......... ..C....... .....
   106KT  .......... ..C....... T....
   107KT  .......... ..C....... ...C.
   108KT  ...T...... ..C....... .....
   109KT  .......... .......... .....
   110KT  .......... ..C..T.... .....
   111KT  ......T..T ..C....... .....
   112KT  .......TT. ..C......T .T.C.
   113KT  ...T...... ..C....... .....
   114KT  ...T....T. .G........ .....
   116KT  .T.T...... ........G. ..T..
   117KT  ...T...... .......... ..T..
   118KT  ...T...... .......... ..T..
   120KT  ...T...... ..C....... .T.C.
   122KT  T..T....T. ..C....... .....
   123KT  .......... ..C....... ...C.
   127KT  ......T... ..C....... .....
   128KT  ....G..... ..C..T.... .....
   129KT  .......... ..C..T.... ....C
   131KT  ...T...... .G........ .....
   168KT  .......... ..C..T.... .....
   173KT  ...T...... ..C....... .....
   175KT  ..A....T.. ..C..T.... .T.C.
   179KT  .......... ..C....... T....
   205KT  ..A....T.. ..C....... .T.C.
   209KT  ...T....T. .......... .....
   211KT  .......... ...G...... .....
   212KT  ...T...... .......... ..T..
   214KT  .......T.. ..C....... .T.C.
   218KT  .......G.. ..C....... .....
   243KT  ....G..... ..C..T.... .....
```

10.5.9 Yoruba Sequences n=21

```
             11111111 1111111112 2222222222 2223333333
           4811222466 6778888891 2336666778 9990122566
           8614469836 8290378923 3090145086 2349107502
Anderson   GTCCTTGCAA CTCAACCTCG CACCCCACCC CACATCCCCT
  339NY    ....C....G .......... T......... ..........
  526NY    .......... .......C.. T.......T. ..TG......
  717NY    .......... .......... T..T....T. ..T.......
  720NY    ......A... .....T.C.. T...T...T. ....C...T.
  748NY    .......... ....C..C.. T......... .....T....
  749NY    .......... .......... T.......TT ..T.......
  750NY    ...A.C.... .....T.C.. T....T.TT. .G..C.....
  881NY    .......... .......CT. T.....C.T. ..TG......
  886NY    ......A... .......... T.......T. ..TG......
  888NY    ......AT.. TC...TGC.. TG........ ....CT...C
  889NY    .......... .C.C...C.. T......... .....T....
  890NY    .......... .......... T.......T. T.TG......
  891NY    .......... .C..C..C.. T......... .....T....
  892NY    ....C..... .......... T.......T. ..........
  893NY    .......... .......... T...T...T. ..TG......
  894NY    ..T....... .......... T......... ....C.T...
  895NY    A......... ..T.C..C.. T.T....... ....CT...C
  897NY    ....C..... .......... T.......T. ..........
  898NY    .......... .....T.C.. T.....T... ..........
  899NY    ...A..A... .........A T.......T. .......T.C
  900NY    .C....A.G. .....T.C.. T.......T. .GT.C...T.
```

10.5.10 Extra Sequences n=2

```
           112222233
          9790245616
          3727317012
Anderson  TACACACCTT
  237KC   CGTTT.ATCC
  096EB   ....TC..CC
```

KC= Sakuye from Kenya

EB= Borana from Ethiopia

INFORMATION SHEET & CONSENT FORM

The Origins of the Pastoral Peoples of Africa

Investigators

E.E. Watson	Massey University, New Zealand
Associate Professor D. Penny	Massey University, New Zealand
Professor S. Pääbo	University of Munich, Germany.

Aim of Investigation:

1. To investigate the origins of the Pastoral people of Africa.
2. To observe genetic diversity within different ethnic groups.

Confidentiality

The people who donate samples will remain anonymous. The results of this study will be available from the researchers in the Molecular Genetics Unit, Massey University, and a copy will be forwarded to the Institutions which assisted in this study. Participants are able to withdraw their consent at any time.

BHNY

Tube Number 748

Your Place of Birth TOGO

Your Ethnic Affiliation YORUBA

Do you drink fresh milk ? Yes __ No ✓

Your mother's Place of Birth SAKI, OYO STATE

Her Ethnic Affiliation YORUBA

Your father's Place of Birth SAKI OYO STATE

His Ethnic Affiliation YORUBA

Your mother's mother's Place of Birth AHA, OYO STATE

Her Ethnic Affiliation YORUBA

Your father's father's Place of Birth SAKI OYO STATE

His Ethnic Affiliation YORUBA

B.P 110/70 mmHg

CONSENT

PARTICIPANT

I agree to take part in this study.

Signature .... Date 3/12/92

TRANSLATOR

I have discussed this consent form with the person giving the sample and I am satisfied that he/she fully understands it, and that his/her consent is freely given.

Signature .... Date 3/12/92

INVESTIGATOR

Signature Elizabeth Watson Date 3/12/92

10.7 Lactose Intolerance

Lactose is hydrolysed to the absorbable monosaccharides, glucose and lactose. This is done by lactase, an enzyme that is found in the brush borders of the epithelial cells of the small intestine. Typical responses for lactose intolerant people (Malabsorbers) is borborygmus, abdominal pain, loose stools and diarrhoea. The volunteers were asked whether they had any of these symptoms after drinking fresh milk. It was not a reliable way of detecting intolerance, but gave an indication of which ethnic groups are tolerant. The table below illustrates the populations sampled (based on maternal identity) and the number of individuals with symptoms of lactose intolerance. The final column 'Economy' indicates whether the population are pastoralists (P) or non pastoralists (NP). NP/P are populations which are agricultural which sometimes keep livestock and utilise their milk products. Pastoralists show a higher lactose tolerance (84%) than non pastoralists (67%).

POPULATION	INTOLERANT	TOLERANT	ECONOMY
Anuak		1	P
Bella		1	NP/P
Borana	4	7	P
Dinka		2	P
Fulbe	11	59	P
Gusii		6	NP
Hausa	3	39	NP/P
Igbo	1	2	NP
Kalenjin	2	4	P
Kanuri	2	14	P
Kikuyu	2	23	NP/P
Luhyia	7	13	NP
Luo	11	39	NP
Maasai		16	P
Misc.	5	13	
Nuer	2	28	P
Rendille		3	P
Sakuye		1	P
Samburu	1	18	P
Shuwa		4	P
Somali	5	23	P
Songhai	4	8	NP
Tuareg	6	21	P
Turkana	11	27	P
Yoruba	15	8	NP

10.8 Individuals of Cross Cultural Marriages

From the 472 individuals sampled, 23 (4.9%) came from mixed marriages. That is 6.4% of the West African sample and 3.7% from the East African sample. All individuals claimed their father's ethnic identity.

MOTHER	FATHER
Borana	Sakuye
Borana	Somali
Fulbe	Kanuri
Fulbe	Songhai
Fulbe	Hausa
Fulbe	Hausa
Fulbe	Hausa
Fulbe	Kanuri
Hausa	Songhai
Junkun	Hausa
Kanuri	Shuwa
Kikuyu	Swahili
Rendille	Samburu
Rendille	Somali
Sakuye	Borana
Sakuye	Borana
Samburu	Maasai
Samburu	Turkana
Samburu	Rendille
Shuwa	Kanuri
Songhai	Hausa
Tuareg	Fulbe
Yoruba	Hausa

10.9 Other Positions

This table represents the positions which were determined outside the 360 base region (np 16024-16383).

HpaI, cuts at position 3592 (when there is a T in position 3594), has high frequencies in Africa (Denaro *et al.* 1981; Chen *et al.* 1995), and low in the Middle East and Sicily (Bonné-Tamir *et al.* 1986, Semino *et al.* 1989). This is probably due to African admixture and is non existent elsewhere in the world. Position 3594 corresponds to the *HpaI*-Morph 3 from Restriction Enzyme Analysis. This defines the group Chen *et al.* (1995) calls Haplogroup L, for which they estimate a divergence time of 98-129,000 years. *HpaI* site is determined by amplifying its surrounding region with the primers L03526 5′-CAT-CAC-CCT-CAT-CAC-CG-3′ and H03706 5′-ATT-GTT-TGG-GCT-ACT-GCT-CG-3′. The amplified fragment is added to a control fragment (with the presence of *HpaI* site) amplified by the primers L05269 5′-TTG-CCC-AAA-TGG-GCC-ATT-AT-3′ and H06011 5′-TGG-CCC-AGG-TCG-GCT-CGA-AT-3′ . These fragments are digested together with *HpaI* and run on a 4% Metaphor Gel.

AvaII, cuts at position 16390 (when there is a G in position 16390). This polymorphism is also detectable with *HinfI*. The state of the polymorphism was determined in half of the individuals in this study. The polymorphism was determined from the sequencing gel of the HVR-I, rather than with a restriction enzyme digest. *AvaII* negative corresponds to Chen *et al.* (1995) haplogroup L_2, which they estimate has a divergence time of 59-78,000 years. Semino *et al.* (1989) and Chen *et al.* (1995) observed 2 substitution events, in this study there are *AvaII* negative sequences which fall outside Cluster-III (in WM34, 2WM11 in Cluster-II, and in WF333 in Cluster-I).

Nine Base Pair Sequence (CCCCCTCTA) occurs in the intergenic COII/tRNA^{Lys} between positions 8272-8289. In most human populations there is a duplication of the nine basepair unit, however, in some Asian, American, and Oceania populations one repeat is deleted (Cann and Wilson, 1983; Stoneking *et al.*1990; Torroni *et al.*1993a). One copy (i.e. deletion) has been reported in Pygmy mtDNA's (Cann and Wilson, 1983;Vigilant, 1990). A triplication has also been reported in one Tharu from Nepal (Passarino *et al.* 1993).

The number of repeats was determined by amplifying the flanking regions with H8297 5′-ATG-CTA-AGT-TAG-CCT-TAC-AG-3′ and L8196 5′-ACA-GTT-TCA-TGC-CCA-TGG-TC-3′. The products were run on a 4% Metaphor Gel with control fragments of 1 and 2 repeats. Presence or absence of restriction sites are indicated by '1', '0' respectively. The nine basepair repeat is scored according to the number of repeats.

Population	ID	*Hpa*I	*Ava*II	9-BP
Borana	96	0		2
Fulbe	257	0		2
Fulbe	326	1	0	2
Fulbe	327	1	1	2
Fulbe	328	0	1	2
Fulbe	329	0	1	2
Fulbe	330	0	1	2
Fulbe	331	0	1	2
Fulbe	332	0	1	2
Fulbe	333	0	0	2
Fulbe	334	0	1	2
Fulbe	336	0		2
Fulbe	337	0	1	2
Fulbe	338	0	1	2
Fulbe	341	1		2
Fulbe	342	0		2
Fulbe	343	0		2
Fulbe	344	1	0	2
Fulbe	345	1	1	2
Fulbe	346	0	1	2
Fulbe	347	0		2
Fulbe	348	1		2
Fulbe	349	1		2
Fulbe	435	1		2
Fulbe	437	1	0	2
Fulbe	441	1		2
Fulbe	442	1		2
Fulbe	444	1		2
Fulbe	445	1		2
Fulbe	446	0		2
Fulbe	448	0		2

Population	ID	*Hpa*I	*Ava*II	9-BP
Fulbe	456	0	1	2
Fulbe	457	0	1	2
Fulbe	527	1	1	2
Fulbe	531	0	1	2
Fulbe	544	1	0	2
Fulbe	553	1	1	2
Fulbe	591	0	1	2
Fulbe	597	0		2
Fulbe	598	1	1	2
Fulbe	710	0	1	2
Fulbe	718	0	1	2
Fulbe	719	0		2
Fulbe	722	0	1	2
Fulbe	723	1		2
Fulbe	724	0	1	2
Fulbe	725	0		2
Fulbe	726	0	1	2
Fulbe	728	0		2
Fulbe	737	0	1	2
Fulbe	738	0	1	2
Fulbe	746	0		2
Fulbe	747	0		2
Fulbe	854	0	1	2
Fulbe	855	1	0	3
Fulbe	856	1		2
Fulbe	866	0		2
Fulbe	869	1		2
Fulbe	870	1		2
Fulbe	872	1		2
Fulbe	873	0	1	2
Fulbe	874	0		2

* A triplication event.

Population	ID	*HpaI*	*AvaII*	9-BP
Hausa	256	0		2
Hausa	259	1	0	2
Hausa	260	0	1	2
Hausa	261	1	0	2
Hausa	335	1	0	2
Hausa	340	0		2
Hausa	528	1		2
Hausa	529	0	1	2
Hausa	530	1	1	2
Hausa	532	1	0	2
Hausa	534	0	1	2
Hausa	535	0		2
Hausa	537	1	1	2
Hausa	538	0	1	2
Hausa	539	1	0	2
Hausa	540	0	1	3*
Hausa	543	0	1	2
Hausa	550	1	0	2
Hausa	551	1	0	2
Hausa	552	0	1	2
Kanuri	262	1		2
Kanuri	706	0	1	2
Kanuri	707	0	1	2
Kanuri	731	1		2
Kanuri	852	1	0	2
Kanuri	857	0	1	2
Kanuri	859	0	1	2
Kanuri	861	0		2
Kanuri	862	1	0	2
Kanuri	863	1		2
Kanuri	864	0	1	2
Kanuri	865	0		2
Kanuri	867	1		2
Kanuri	868	1		2
Kikuyu	27	0		2
Kikuyu	34			2
Kikuyu	66	0	1	2
Kikuyu	68	1		1
Kikuyu	94	0		2
Kikuyu	97	1		2
Kikuyu	98	1		2
Kikuyu	115	0		2
Kikuyu	136	1		2
Kikuyu	138	0	1	2
Kikuyu	139	0		2
Kikuyu	140	0		2
Kikuyu	141	1		2
Kikuyu	142	0		2
Kikuyu	150	0		2

Population	ID	*HpaI*	*AvaII*	9-BP
Kikuyu	155	1	0	2
Kikuyu	156	1	1	2
Kikuyu	165	0	1	2
Kikuyu	169	1		2
Kikuyu	170	0		2
Kikuyu	172	0	1	2
Kikuyu	565	0	1	2
Kikuyu	566	1		2
Kikuyu	568	1	1	2
Kikuyu	572	0		2
Sakuye	237	0		2
Somali	213	0	1	2
Somali	222	1		2
Somali	225	0		2
Somali	226	1	0	2
Somali	229	0	1	2
Somali	235	0	1	2
Somali	238	0		2
Somali	241	0	1	2
Somali	500	1		2
Somali	501	0		2
Somali	502	1		2
Somali	503	0		2
Somali	505	0	1	2
Somali	506	0	1	2
Somali	507	1	1	2
Somali	508	0	1	2
Somali	509	0	1	2
Somali	510	0	1	2
Somali	511	0	1	2
Somali	512	0		2
Somali	513	0		2
Somali	514	0		2
Somali	515	0	1	2
Somali	516	0	1	2
Somali	517	0	1	2
Somali	518	1	0	2
Somali	519	0		2
Songhai	264	1	1	2
Songhai	426	1		2
Songhai	427	0		2
Songhai	428	0		2
Songhai	430	1	1	2
Songhai	431	1		2
Songhai	432	1		2
Songhai	433	0		2
Songhai	434	0	1	2
Songhai	436	0	1	2
Tuareg	254	1	0	2

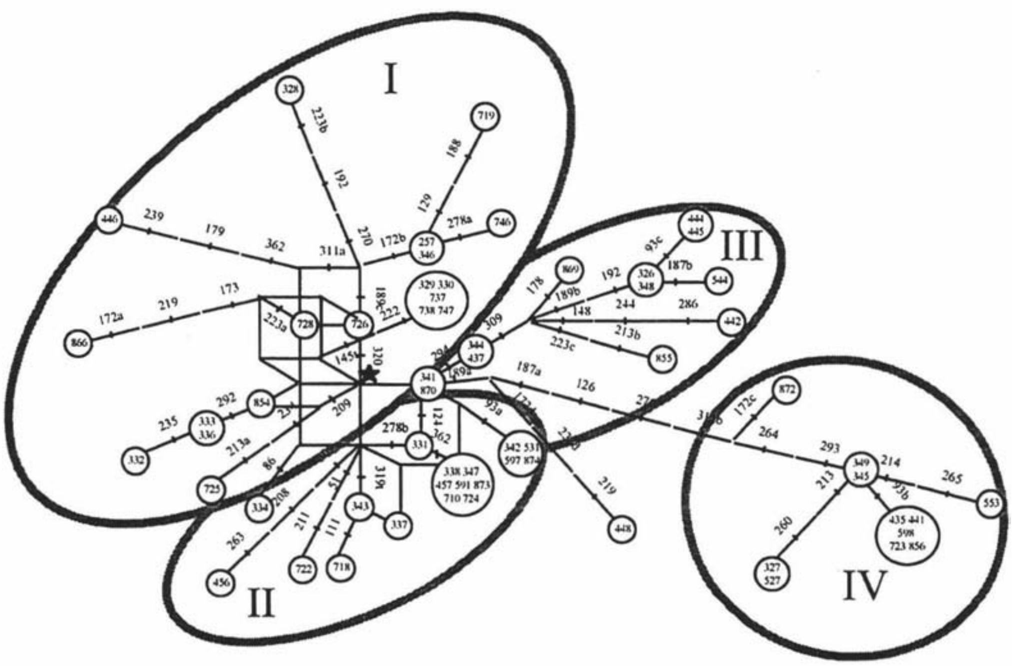
Population	ID	*Hpa*I	*Ava*II	9-BP
Tuareg	263	1		2
Tuareg	438	1		2
Tuareg	443	0		2
Tuareg	447	0	1	2
Tuareg	449	1	0	2
Tuareg	450	1	0	2
Tuareg	451	1	0	2
Tuareg	452	0	1	2
Tuareg	453	0		2
Tuareg	454	0		2
Tuareg	458	0	1	2
Tuareg	459	0	1	2
Tuareg	460	1		2
Tuareg	461	1		2
Tuareg	462	0		2
Tuareg	463	0		2
Tuareg	464	0	1	2
Tuareg	465	1		2
Tuareg	466	1	0	2
Tuareg	467	1	1	2
Tuareg	468	0	1	2
Tuareg	469	0	1	2
Tuareg	471	1	0	2
Tuareg	472	1		2
Tuareg	474	1		2
Turkana	93	0	1	2
Turkana	99	1	1	2
Turkana	101	1		2
Turkana	102	1	1	1
Turkana	103	0		2
Turkana	104	1		2
Turkana	105	0	1	2
Turkana	106	0	1	2
Turkana	107	0		2
Turkana	108	1		2
Turkana	109	0		2
Turkana	110	1	1	2
Turkana	111	0	1	2
Turkana	112	0		2
Turkana	113	1	1	2
Turkana	114	1	0	2
Turkana	116	1	1	2
Turkana	117	1		2

Population	ID	*Hpa*I	*Ava*II	9-BP
Turkana	118	1	1	2
Turkana	120	1	1	2
Turkana	122	1	0	2
Turkana	123	0		2
Turkana	127	0	1	2
Turkana	128	1		2
Turkana	129	1	1	1
Turkana	131	1	0	2
Turkana	168	1	1	1
Turkana	173	1	1	2
Turkana	175	0	1	2
Turkana	179	0	1	2
Turkana	205	0	1	2
Turkana	209	1	0	2
Turkana	211	0	1	2
Turkana	212	1	1	2
Turkana	214	0		2
Turkana	218	1	1	2
Turkana	243	0	1	2
Yoruba	339	0		2
Yoruba	526	1		2
Yoruba	717	1		2
Yoruba	720	1		1
Yoruba	748	0	1	2
Yoruba	749	1		2
Yoruba	750	1	1	2
Yoruba	881	1		2
Yoruba	886	1		2
Yoruba	888	1	1	2
Yoruba	889	1		2
Yoruba	890	1		2
Yoruba	891	0	1	2
Yoruba	892	0	0	2
Yoruba	893	1	0	2
Yoruba	894	0	1	2
Yoruba	895	0		2
Yoruba	897	0		2
Yoruba	898	0		2
Yoruba	899	1		2
Yoruba	900	1	1	2

10.10 Population Portraits

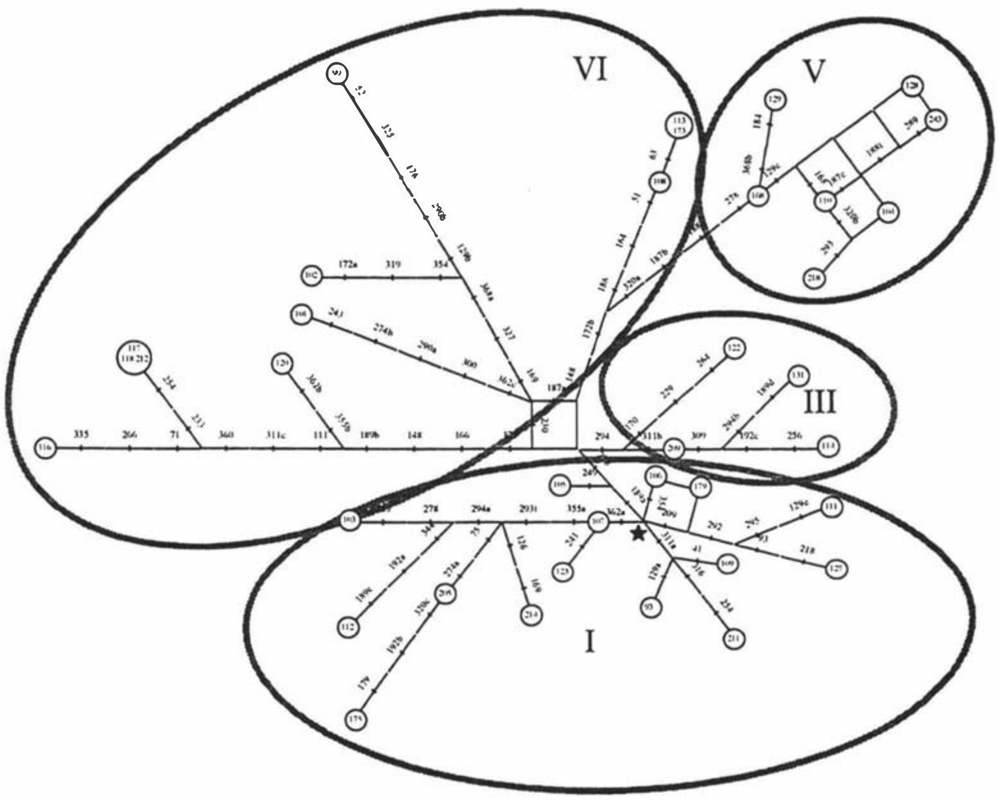
The Networks method is described in Section 6.3.2. Illustrated below and on the following page are Network Portraits of two African populations. Nucleotide position 182, and 183 (16182, 16183) are removed because they are very fast sites. According to the clusters assigned in Section 6, the Fulbe below are a composite of Cluster-I (33.3%), Cluster-II (30.0%), Cluster-III (18.3%), and Cluster-IV (18.3%) clusters. For orientation see the Interpopulation Network Figure 12.

The Turkana network (see next page) contrasts with the Fulbe, by exhibiting very long edges. The Turkana are composed of Cluster-I (40.5%), Cluster-III (10.8%), Cluster-V (18.9%) and the isolated lineages (Cluster-VI) (29.7%).



Fulbe Network Portrait ★ 223

The Fulbe individuals are assigned to four of the six African clusters. The Clusters are circled and numbered with Roman Numerals. Note Cluster I and II are *Hpa*I negative, all other clusters are *Hpa*I positive.



Turkana Network Portrait ★ 223, 311

The Turkana individuals are assigned to four of the six African clusters. The Clusters are circled and numbered with Roman Numerals. Note Cluster I is *Hpa*I negative, all other clusters are *Hpa*I positive.