# HAPLOID GENETIC VARIATION IN POPULATIONS FROM UGANDA, ZAMBIA AND THE CENTRAL AFRICAN REPUBLIC

Debra Barkhan

A thesis submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Doctor of Philosophy.

Johannesburg, 2004

# DECLARATION

I declare that this thesis is my own unaided work. It is being submitted for the Degree of Doctor of Philosophy in Human Genetics at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.

I declare that this work has been approved by the Ethics Committee of the University of the Witwatersrand for Research on Human Subjects, and the certificate numbers are M980601 and and 24/4/87.

Debra Barkhan

9 NOV 2004

Date

## ABSTRACT

Y chromosome DNA and mitochondrial DNA (mtDNA) variation were examined in Ugandans, Zambians, Biaka Pygmies and non-Pygmies from the Central African Republic. Y chromosome DNA variation was also examined in populations from the Democratic Republic of Congo. Data generated in this study were analysed together with published data to (1) clarify the understanding of the overall patterns of haploid genetic variation in Africa; (2) examine genetic affinities among central African and other African populations; (3) assess the concordance of haploid markers with different mutation rates in assessing population affinities; (4) compare male and female migration rates in African populations; and (5) refine theories regarding the prehistory of central Africa populations based on linguistics and archaeology.

Sixteen biallelic and eight microsatellite Y-specific markers were examined in 369 central African individuals. Eleven Y chromosome haplogroups (HGs A, B*, B-M150, B-M112, B-M211, E-M191, E-M2, E-M35, E-M40, FJ and R) and 174 compound haplotypes were identified. The mtDNA 9-bp deletion, 3592 *Hpa*I and 10397 *Alu*I restriction polymorphisms, and two hypervariable regions (HVRs) were examined in 397 individuals. A total of 246 mtDNA types were identified and classified into 19 mtDNA subhaplogroups.

Using Y chromosome data, central African populations shared close genetic affinities with each other and with populations from west and southern Africa. Extensive unidirectional Y chromosome gene flow from non-Pygmy populations to Biaka Pygmies was observed. Using mtDNA data, central African non-Pygmy populations shared close genetic affinities with each other and with populations from west, east and southern Africa. MtDNA studies indicated almost complete maternal genetic isolation of Biaka. Overall, using both mtDNA and Y chromosome data, pan-African populations were best grouped by geographic rather than by linguistic criteria.

Different mtDNA and Y chromosome data types revealed similar genetic relationships among African populations. Female migration rates appear to have exceeded male migration rates in non-Pygmy central African populations in this study, whilst the opposite was found in Biaka Pygmies. Data types at different levels of resolution suggested that male and female migration rates in Africa may have differed over time, and may not have been significantly different. This research has provided new insights into the complex demographic history that shaped the present-day genetic landscape of central African populations.

# ACKNOWLEDGEMENTS

I would like to thank the individuals who contributed blood samples for DNA extraction and analysis and who therefore made this study possible. This study was also facilitated by those who collected the samples, including Professor Trefor Jenkins, Dr Thomy de Ravel, Professor Wilson Byarugaba, Professor Kensese Mossanda, and Professor Himla Soodyall.

I am extremely grateful to my supervisor, Professor Himla Soodyall, for her patience and guidance throughout this study, and for reading several drafts of this thesis. Dr Bharti Morar and Dr Almut Nebel were both invaluable during the time that I was privileged to work with them, and provided advice, inspiration and support. I also wish to express my appreciation to my colleagues in the PEG lab/ HGDDRU and Molecular Genetics Lab for their friendship, and to Akashnie Maharaj and Thejane Motladile for the use of their unpublished data for comparative purposes.

A special thank-you goes to Professor Michele Ramsay for her on-going support. Dr Luisa Pereira kindly provided much of the published mtDNA data in electronic format, and Dr Tony Lane and Caren Ginsburg gave me helpful statistical advice. Thank you also to Dr Adrian Puren and Dr Clive Gray of the NICD for allowing me time to complete the write-up, and for organising a laptop computer. Many thanks to the NHLS Graphics Unit for assistance in printing this thesis. Most of all, my appreciation goes to my husband Alain, my parents Les and Rhoda, and my sister Leila and her family, who have all been an amazing source of help, love, encouragement, and patience through the years.

During my studies I was supported by a National Research Foundation Prestigious Doctoral Scholarship, and by generous scholarships and bursaries from the University of the Witwatersrand. Travel grants from The Paul and Stella Loewenstein Trust and from the University of the Witwatersrand allowed me to present some of this work at international conferences. This research was supported by grants awarded to Professor Himla Soodyall by the NHLS (formerly SAIMR), University of the Witwatersrand, the NRF and the South African MRC.

## TABLE OF CONTENTS Page

## CHAPTER ONE – INTRODUCTION.....1

# LIST OF FIGURES

## LIST OF TABLES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| AMOVA | analysis of molecular variance |
| ASD | average square distance |
| BMH | Bantu Modal Haplotype |
| bp | base pairs |
| BSA | bovine serum albumin |
| $^{o}$C | degrees Centigrade |
| C.A.R. | Central African Republic |
| CI | confidence interval |
| COII | cytochrome oxidase II gene |
| comm. | communication |
| CRS | Cambridge reference sequence |
| d | mean sequence divergence |
| DNA | deoxyribonucleic acid |
| DHPLC | denaturing high performance liquid chromatography |
| $ddH_2O$ | deionised distilled water |
| dNTP | deoxyribonucleotide triphosphate |
| ddNTP | dideoxyribonucleotide triphosphate |
| D.R.C. | Democratic Republic of Congo |
| EDTA | ethylene-diamine-tetra-acetic acid |
| fig. | figure |
| g | gram |
| *h* | genetic diversity |
| HG | haplogroup |
| HGDDRU | Human Genomic Disease and Diversity Research Unit |
| HG-STR | combined haplogroup – short tandem repeat data |
| ht | haplotype |
| HVR | hypervariable region |
| ins | insertion |
| kb | kilobase |

| | |
|---|---|
| kb | kilobase |
| km | kilometre |
| kya | thousand years ago |
| LGM | Last Glacial Maximum |
| $\mu$ | mutation rate |
| μg | microgram |
| μl | microlitre |
| μM | micromolar |
| M | molar |
| Mb | megabase |
| mg | milligrams |
| $MgCl_2$ | magnesium chloride |
| min | minutes |
| MJ | median joining |
| mM | millimolar |
| ml | millilitre |
| mpd | mean number of pairwise differences |
| MRC | Medical Research Council |
| MRCA | most recent common ancestor |
| MSY | male-specific region of the Y chromosome |
| mtDNA | mitochondrial DNA |
| N | sample size |
| NHLS | National Health Laboratory Service |
| NICD | National Institute of Communicable Diseases |
| NJ | neighbour joining |
| np | nucleotide position |
| NRF | National Research Foundation |
| NRY | non-recombining portion of the Y chromosome |
| $P$ | probability |
| PC (plot) | principle components |
| PEG | Population and Evolutionary Genetics |

| | |
|---|---|
| PCR | polymerase chain reaction |
| r | regression co-efficient |
| RE | restriction enzyme |
| RFLP | restriction fragment length polymorphism |
| RNA | ribonucleic acid |
| s | seconds |
| SAIMR | South African Institute of Medical Research |
| SNP | single nucleotide polymorphism |
| STR | short tandem repeat |
| subHG | subhaplogroup |
| TAE | tris acetate-EDTA |
| *Taq* | *Thermus aquaticus* |
| TBE | tris borate-EDTA |
| TMRCA | time to the most recent common ancestor |
| $\text{tRNA}^{\text{Lys}}$ | transfer RNA for Lysine |
| U | units |
| V | volts |
| v | version |
| YAP | Y *Alu* polymorphism |
| YCC | Y Chromosome Consortium |

# 1. INTRODUCTION

Haploid genetic data, in conjunction with anthropological, archaeological and linguistic evidence, have contributed significantly to the current debate about the origins, migrations and relationships of modern human populations. Haploid areas of the genome include both the male-specific region of the Y chromosome (MSY), and mitochondrial DNA (mtDNA).

In recent years the use of highly informative mtDNA (Soodyall et al. 1996; Watson et al. 1997; Bandelt et al. 2001; Salas et al. 2002), and Y chromosome genetic markers (Scozzari et al. 1999; Bosch et al. 1999, 2000, 2001; Cruciani et al. 2002; Semino et al. 2002), together with the phylogeographic approach (Avise 1987), has greatly improved the understanding of the prehistory of African populations. However the geographic distribution of the sampled populations has been rather uneven; many regions and populations within the African continent are under-represented either due to small sample sizes or lack of sampling. In particular there is a lack of data from populations living within the central African region (fig. 1.1; Tishkoff and Williams 2002). Studies consistently indicate that Africa is the most genetically diverse continent in the world (Batzer et al. 1994; Bowcock et al. 1994; Armour et al. 1996; Jorde et al. 1997; Stoneking et al. 1997; Jorde et al. 2000; Tishkoff et al. 2000), but haploid genetic variation in African populations has not been as thoroughly studied as it has in non-African populations (e.g. Macaulay et al. 1999; Forster et al. 2000; Richards et al. 2000; Kayser et al. 2001). African populations show high levels of subdivision (Tishkoff and Williams 2002), with extensive genetic variation among even geographically close African populations (e.g. Tofanelli et al. 2003). For these reasons, the further study of [central] African populations is of fundamental importance for understanding both global and regional genetic variation.

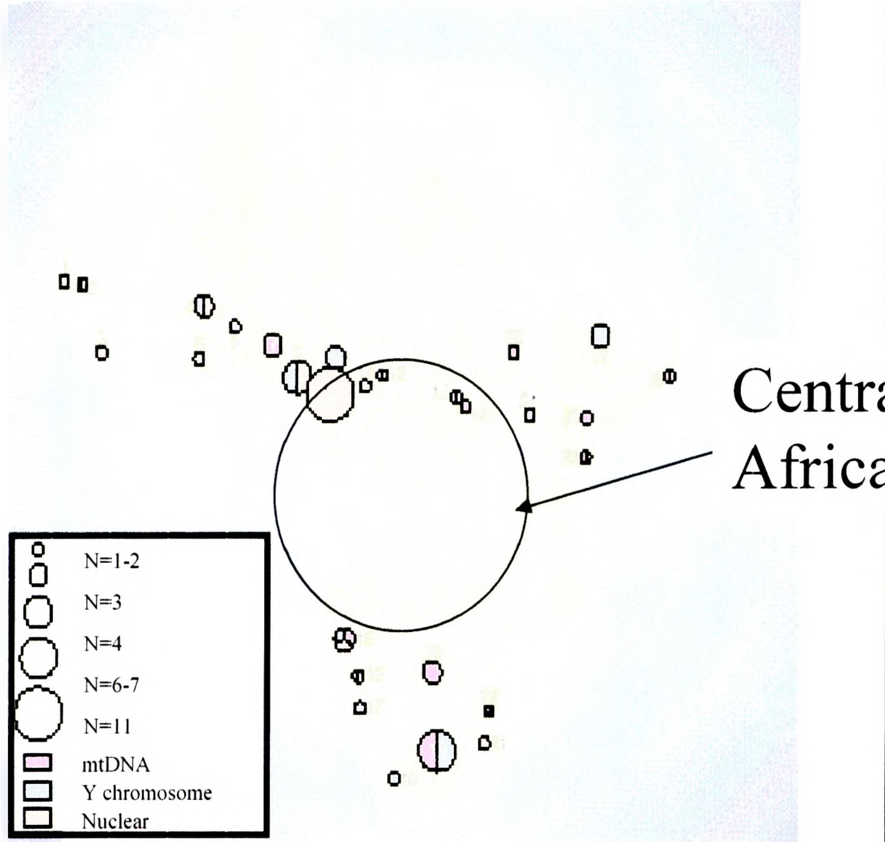


**Figure 1.1** The geographical distribution of non-biomedical studies in sub-Saharan African populations (Tishkoff and Williams 2002). N= number of ethnic groups studied with more than 20 individuals examined.

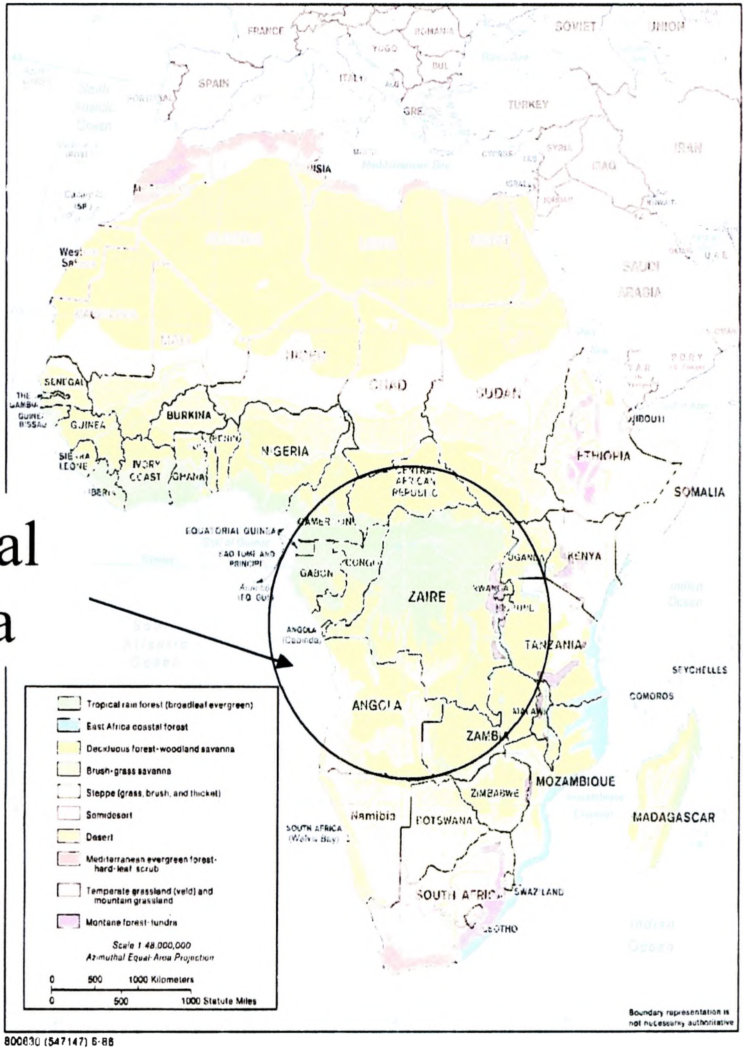


**Figure 1.2** Map showing African environmental zones

Insights into the patterns of haploid genetic variation in African populations have previously been hampered by the use of different polymorphic markers in different studies, making comparative investigations difficult. Additionally, very few studies (Passarino et al. 1998; Pereira et al. 2001; Pereira et al. 2002; Knight et al. 2003) have examined both mtDNA and MSY-DNA variation in the same African populations, which can reveal interesting differences in the patterns of male and female migrations (Seielstad et al. 1998).

In order to address the paucity of haploid genetic data from central African populations, both mtDNA and MSY-DNA variation in populations from the Central African Republic, Uganda and Zambia were investigated in this study. Data from these populations have been analysed within the context of the recently published MSY (Y Chromosome Consortium [YCC] 2002) and mtDNA (Salas et al. 2002) phylogenies, and compared to published MSY and mtDNA data (Tables 2.2 and 2.3) to examine how males and females have shaped the gene pools of African populations. In addition, genetic data in this study have been collated with the information gained from non-genetic sources, in order to gain new insights into the prehistory of central African populations.

## 1.1 *The prehistory of central Africa using non-genetic studies*

The central African geographic region extends on both sides of the equator, from the Atlantic Ocean in the west, to the Great Rift Valley and Great Lakes in the east. The area is largely characterised by tropical rainforest, bounded to the north and south by woodland savannah environments (fig. 1.2). Knowledge of the settlement and prehistory of central Africa has been partially reconstructed from archaeological and linguistic studies, although the understanding of the central African past by archaeological means is somewhat limited due to a scarcity of data (van Noten 1982; Phillipson 1985; Shaw et al. 1995; Vansina 1990), as well as by the lack of hominid fossils (van Noten 1982; Phillipson 1985).

### 1.1.1 *Stone Age archaeology of central Africa.*

Early Stone Age (modes 1 and 2; Clark 1968) Acheulian and pre-Acheulian tools have been discovered in peripheral parts of central Africa (Cahen 1982; Vansina 1990), indicating that these regions were populated since at least ~1.5 million years ago. However there is a lack of Acheulian artefacts from the densely forested Congo basin (Cahen 1982; Phillipson 1985), suggesting that the core central African regions were only occupied at later stages.

A Middle Stone Age industry (mode 3; Clark 1968) called the Sangoan provides archaeological evidence for a first penetration of the true central African forest, and for the intensive occupation of the surrounding woodlands (Oliver 1999). Entry into the true forest may be associated with the evolution of modern *Homo sapiens* (Vansina 1990). This is substantiated by the technological innovations seen in Sangoan tools (Oliver 1999); such innovations may indicate modern behaviour, consistent with the evolution and expansion of early modern humans at this time (Lahr and Foley 1998). In central Africa, the Sangoan Industry has been dated to at least 80 thousand years ago [kya] (Clark 1982). By ~40 kya, the central African Sangoan Industry was succeeded by more specialised Middle Stone Age industries, such as the Lupemban and the Bambatan (Phillipson 1985; Oliver 1999). Such diversification and regionalization is thought to indicate population subdivision (Lahr and Foley 1998).

Late Stone Age technologies (mode 5, Clark 1968) such as the Tshitolian Industry appeared in central Africa ~18-14 kya (Phillipson 1977; Ambrose 1982; Phillipson 1985). During this time period, extremely arid conditions associated with the Last Glacial Maximum may have facilitated further occupation of the central African region, since rainforest regions changed to savannah environments and were thus more easily penetrable (Oliver 1999, Vansina 1990).

About 10 kya, very warm climatic conditions known as the Holocene Wet Phase caused a re-expansion of the equatorial forest (Maley 1995; Vansina 1990). Wet conditions stimulated population growth in the Sahel region to the north of today's central rainforest region, which in turn may have stimulated the development of settled life and food production in Africa (Phillipson 1985). A common fishing and hunting culture is thought to have developed 9 000 -

7 000 years ago at the latitude of present-day Khartoum in Sudan, extending as far south as Uganda and Kenya (Oliver 1999). The advent of agriculture occurred ~8 kya in regions to the north, west and east of the central African rainforest, in the Sahel belt, west Africa and Ethiopia highlands respectively (Phillipson 1985; Diamond 1997). Indigenous domesticated African crops include yams, guinea rice, sorghum, millet, ensete (a plant in the banana family), teff (a grain similar to millet) and noog (an oilseed) (Phillipson 1985; Diamond 1997). Domesticated indigenous sub-Saharan animals include only guinea fowl (Diamond 1997) and possibly cattle (Bradley et al. 1996).

Food production is thought to confer several advantages to farming populations in comparison with hunter-gatherers, including the ability to support higher population densities and the ability to store food surpluses. These developments lead to population growth, the development of complex technology, social stratification, and resistance to infectious disease acquired from social domesticated animals (Diamond 1997; Diamond and Bellwood 2003). Food production is considered to have been an important trigger for major expansions and dispersals of [farming] populations all over the world within the last ~10 000 years (Diamond 1997; Diamond and Bellwood 2003). Within central Africa, the expansions and dispersals of agriculturist populations had begun by ~5 kya; the impact of these demographic shifts is best understood by analyses of African linguistics in conjunction with Iron Age archaeological findings.

### 1.1.2 *African linguistic studies*

Linguistic studies have proven to be a useful adjunct to archaeological studies, offering insights into the more recent past. Language usually indicates group membership since you learn a language from the society into which you are born (Phillipson 1985) although linguistic change can occur extremely fast. By studying present-day languages and their distributions, historical linguists can reconstruct certain features of ancestral languages. Such studies can suggest the areas in which the ancestral languages may have been spoken. Reconstructed vocabulary can indicate lifestyle and technology, and can reveal population interactions (e.g. Ehret 1998).

Languages spoken by the peoples of Africa are classified into four phyla (Greenberg 1963): Afro-Asiatic (including the Semitic, Berber, Cushitic and Chadic language families), Nilo-Saharan (east and central Sudanic, Saharan and Songhai), Niger-Congo (Atlantic, Mande, Voltaic, Kwa, Adamawa and Bantu), and Khoisan (San and Khoikhoi). The distribution of these language phyla in Africa today is shown in fig. 1.3.

The great majority of subequatorial Africans (>200 million people) today speak one of ~500 very closely related languages, even though they are distributed over an area of ~500 000 km$^2$. Wilhelm Bleek (1862) collectively referred to these languages as "Bantu" languages, based on the word meaning "people" in many of them. The current distribution of the Bantu languages is likely to have been achieved as a consequence of the movement of people (demic diffusion) rather than merely by a diffusion of language (Ehret and Posnansky 1982; Huffman 1982). This expansion is commonly referred to in the literature as the "Bantu Expansion" (Greenberg 1963). Linguists have suggested that the "Bantu Expansion" began ~3-5 kya (Ehret 1982; Vansina 1990) and probably originated in the Cross River Valley, in the region of present-day eastern Nigeria and western Cameroon (Johnston 1913; Greenberg 1972; Huffman 1982; Vansina 1990; Vogel 1994).

Bantu languages have been divided into three major groups, including North-Western Bantu (subgroups A, B, C), Eastern Bantu (subgroups E, F, G, J, N, P and S) and Western Bantu (subgroups H, K, L, R, D and M) (fig. 1.4; Guthrie 1948; Vansina 1984, 1990; Nurse and Tucker 2002; Holden 2002). Speakers of Eastern and Western Bantu languages are thought to have followed different migratory routes until they reached their present-day distributions. The ancestors of Eastern Bantu-speakers are thought to have migrated eastwards out of the Cross River Valley, reaching the Great Lakes region by ~3 kya (Ehret 1998). After establishing their presence in the interlacustrine area, speakers of Eastern Bantu languages expanded southwards, reaching roughly their full contemporary distribution across most of eastern and southern Africa by the early centuries of our era (fig.1.4).

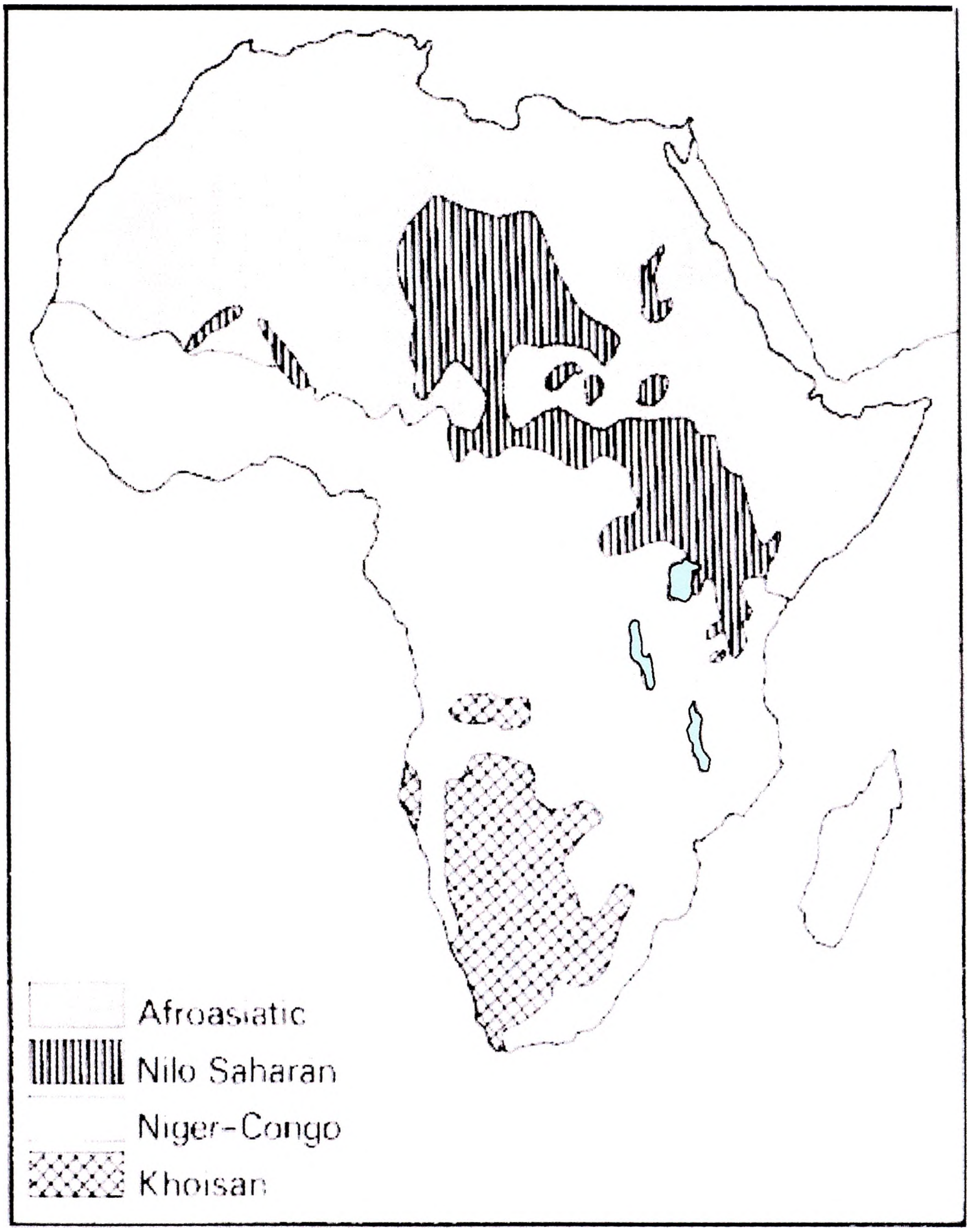
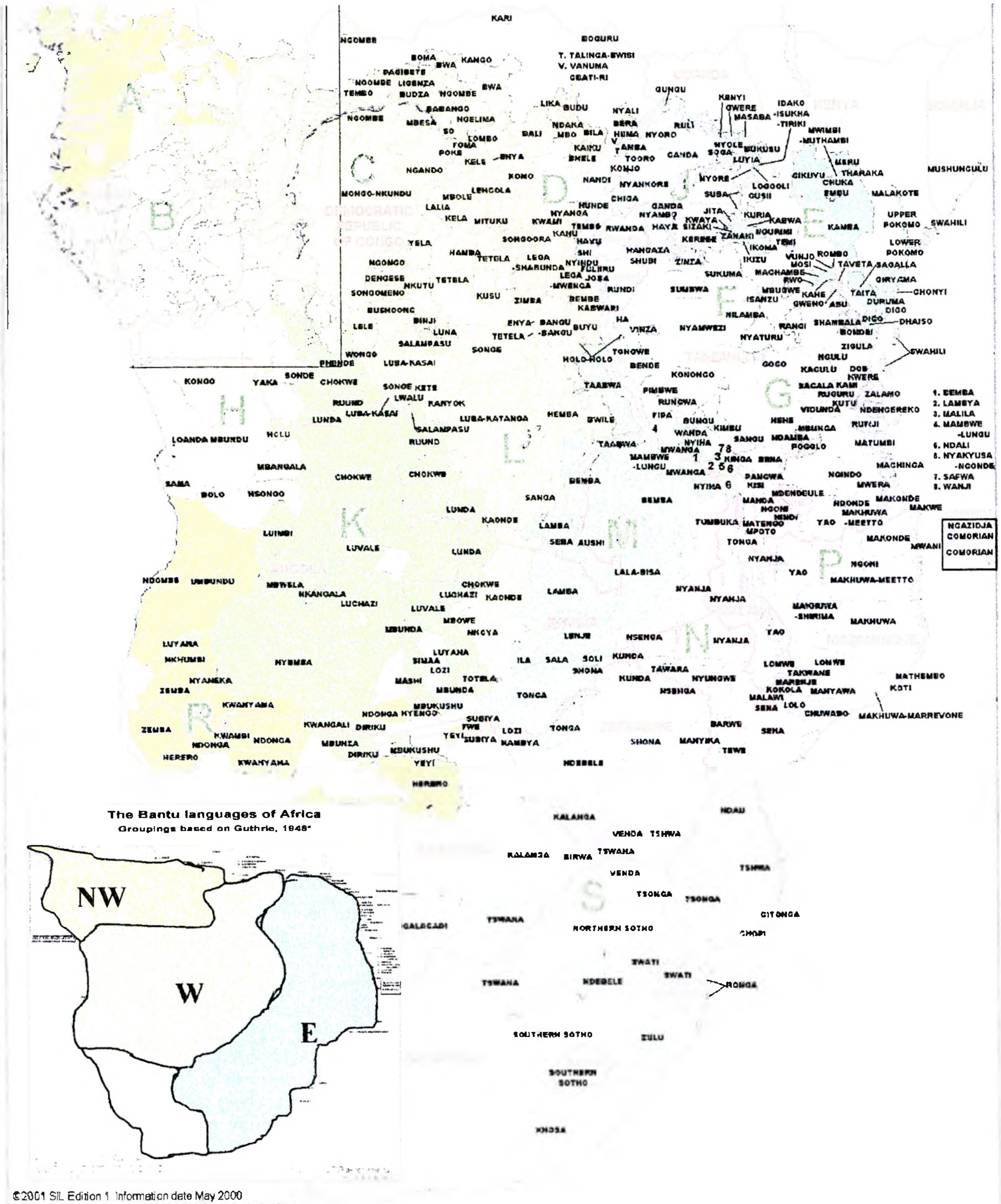


**Figure 1.3** The geographical distribution of the four major African language phyla



©2001 SIL Edition 1 Information date May 2000

**Figure 1.4.** The distribution of Eastern (E) and Western (W) Bantu languages (inset) and their subgroups (letters in main figure) (Guthrie 1948 and Grimes 2001).

Western Bantu-speakers followed a second major route of migration, moving from the Cameroon homeland directly south through the rainforests, possibly following the Atlantic coast. Vansina (1990) has used linguistic data to map their movements in some detail; they eventually settled the whole of the rainforest region, extending as far south-east as the present-day Zambia-Malawi border (fig. 1.4). Western Bantu-speakers are believed to have started their expansion well before Eastern Bantu-speakers (Ehret 1982; Phillipson 1985).

### 1.1.3 *Iron Age archaeology of central Africa*

There is no linguistic or archaeological evidence for knowledge of metallurgy by proto-Bantu-speakers (Ehret 1982; Vansina 1984, 1990). Nevertheless, iron technology is considered to be one of the technological innovations that drove the "Bantu Expansion", and certainly characterized most of the later stages of the expansion (Huffman 1982). The earliest evidence of metallurgy in central Africa comes from the Great Lakes region, and dates to ~ 2 300 years ago (van Noten 1982). There is evidence of iron-working at sites in Cameroon and the Central African Republic since ~ 2 100 years ago (Lanfranchi et al. 1998) and possibly earlier in other parts of west-central Africa (Vansina 1990). Metallurgy spread later to the southern parts of central Africa e.g. to Zambia by ~1 600 years ago (van Noten 1982; Vansina 1984).

The pottery of iron-using communities from areas in Africa where Eastern Bantu languages are spoken is very homogenous, and has been described as belonging to a single Early Iron Age stylistic tradition called the Chifumbaze Complex (Phillipson 1977, 1985; Huffman 1982). The Chifumbaze complex is associated with evidence of food-production, settled village life, metallurgy, and pottery manufacture (Phillipson 1985). Different ceramic traditions within the Chifumbaze are suggestive of branching migratory routes. The earliest manifestation of the Chifumbaze is a type of pottery called Urewe Ware (Soper 1971), which first appeared in the Lake Victoria region ~2.5 kya. Phillipson (1985) suggested that Urewe Ware gave rise to two different Chifumbaze pottery types in the regions to its immediate south, called the eastern and western facies. Huffman (1989) redescribed Phillipson's "western facies" pottery as the Kalundu ceramic tradition. The eastern facies has been subdivided into two further branches, called highland and lowland, inland and coastal, or Nkope and Kwale respectively (Huffman

1982; Phillipson 1985; Huffman 1989). The distribution of the Chifumbaze Complex components is depicted in fig. 1.5. The pottery found in regions where Western Bantu languages are spoken differs from the Chifumbaze Complex.

The archaeological record also provides evidence of cultural differences between Eastern and Western Bantu-speakers. Eastern-Bantu-speakers gained pastoralism and cereal agriculture in the region of the Great Lakes, most probably adopted from Nilo-Saharan-speakers in that area (Vogel 1994). A strong focus on cattle as currency and source of status evolved among the Eastern Bantu-speaking populations; this is reflected in the distinctive arrangement of their settlements in the "Central Cattle Pattern" (Huffman 1982; 1986; 1989). However due to the humidity of the environment and the tsetse fly barrier, the east African cereal and cattle package was not viable in the rainforest regions occupied by Western Bantu-speakers (Ehret 1982; Vansina 1984). Instead, the economy in the west incorporated a combination of fishing, hunting, vegeculture, banana and oil palm horticulture (Huffman 1989). In Western Bantu-speaking societies, metal currency took the place of cattle as a source of prestige (Vansina 1984; Huffman 1989), and their villages were typically arranged in the "Forest Pattern" instead of the "Central Cattle Pattern" (Huffman 1989). Eastern and Western Bantu-speakers also differed in their views of identity inheritance: Western Bantu were 'matrilineal", believing that "a person was created through his mother's blood" whilst the Eastern Bantu were "patrilineal" (Hammond-Tooke 1974).

Eastern and Western Bantu-speaking peoples are thought to have met and intermingled in south-central Africa during the first millennium of the common era (Vansina 1984; Phillipson 1985). In some cases, for example in the region of present-day Zambia, archaeological evidence shows that areas previously populated by speakers of eastern Bantu languages were repopulated by later expansions of Western Bantu speakers (Huffman 1989). This dual influence can be detected in the mixture of Central Cattle and Forest settlement patterns used in Zambian villages (Huffman 1989). In other cases, eastern Bantu-speakers are thought to have moved into areas initially occupied by Western-Bantu speakers (Vansina 1984).

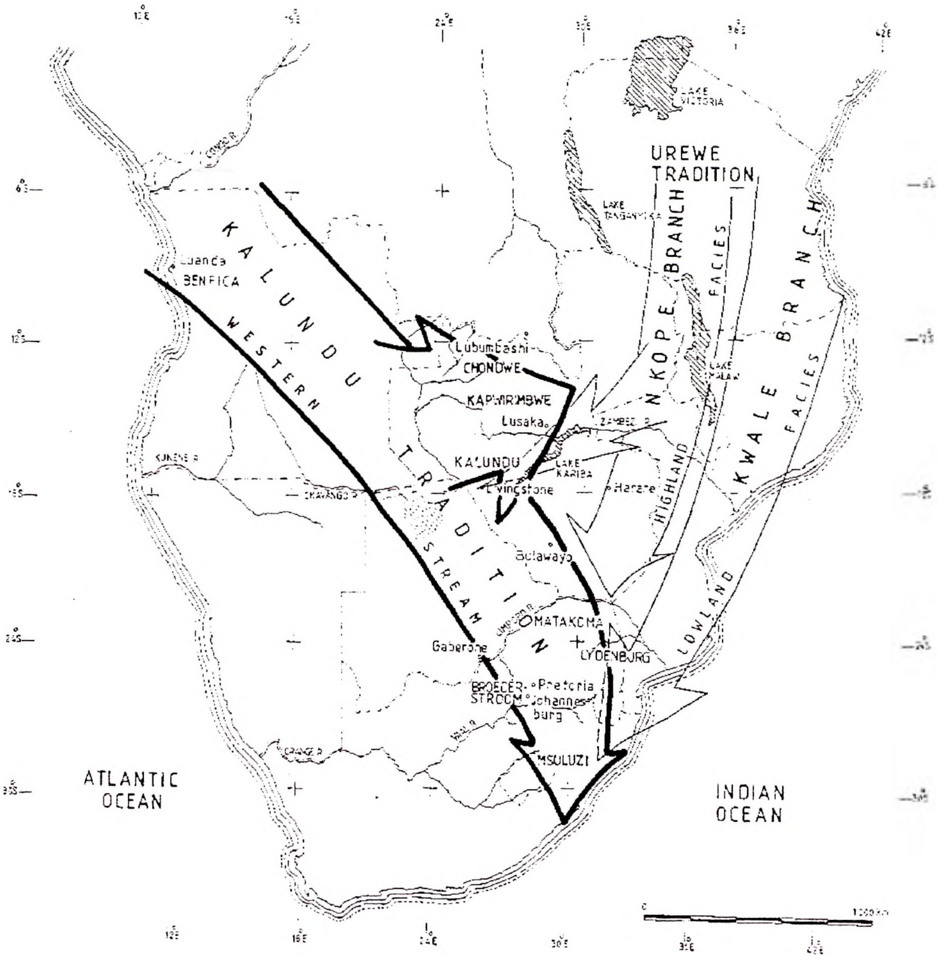


**Figure 1.5** Distribution of the archaeological Iron Age Chifumbaze complex and its components in sub-Saharan Africa (Huffman 1989)

### 1.1.4 *Non-Bantu-speaking populations of central Africa*

The demographics of central Africa were greatly influenced by the influx of Bantu-speakers described above, but were also shaped by pre-existing populations as well as other migratory (non-Bantu-speaking) populations. The former includes the indigenous hunter-gatherer populations of central Africa, whilst the latter includes speakers of Ubangian languages within the northern part of the rainforest. The Great Lakes region has a particularly complex demographic history since it has been repopulated by a succession of groups speaking Afro-Asiatic (southern and eastern Cushitic), Nilo-Saharan (southern and eastern Nilotic) as well as Niger-Congo (Bantu) languages (Ambrose 1982).

Very little is known about the autochthonous populations of central Africa; their descendants are believed to be the current populations known collectively as 'Pygmies". The term comes from a Greek word meaning "fist", denoting their short stature (Cavalli-Sforza 1986) but is considered somewhat derogatory. In the absence of alternatives, I use the collective term with apologies, but wherever possible, refer to specific populations by name. At present the two major Pygmy populations in central Africa are the Biaka, who live in the forests spanning Cameroon and the Central African Republic (C.A.R.), and the Mbuti, who are found in the Ituri forest of north-eastern Democratic Republic of Congo (D.R.C.). Pockets of other Pygmy populations exist or existed in present-day Gabon, Congo, Rwanda, Burundi and Zambia (Cavalli-Sforza 1986), suggesting that Pygmy populations were once much more widespread. The Biaka and Mbuti speak the languages of neighbouring farming populations (Grimes et al. 2001). It is thought that their original language has been lost (Letouzey 1976; Cavalli-Sforza 1986), therefore linguistic studies are not concordant with their biological history. Archaeological studies also do not contribute to the reconstruction of Pygmy history since very few archaeological finds are preserved in the rainforest environment. Pygmies still partially retain their hunter-gatherer lifestyles (Cavalli-Sforza 1986) and their microlithic technology survived long after the appearance of metallurgy in central Africa (Ehret and Posnansky 1982; Phillipson 1985).

People speaking Ubangian languages (classified as Adamawa-Ubangian within Niger-Congo) expanded from their homeland in the northwest of the Ubangi River in an easterly direction along the northern edge of the rainforest, at approximately the same time period as the first stages of the "Bantu Expansion" (Saxon 1982). The Ubangian dispersal has been described as a "roughly contemporaneous northern counterpart" of the "Bantu Expansion" and Ubangian-speakers came to occupy the northern third of the equatorial rainforest over the course of the last ~5 000 years (Ehret and Posnansky 1982). Linguistic evidence suggests that Ubangian-speakers were the first food-producers in their early areas of settlement (Ehret 1974; David 1982; Saxon 1982) and banana cultivation is thought to be an important factor in their expansion (Vansina 1984). Like the Western Bantu, the areas settled by Ubangian-speakers were generally not suitable for pastoralism, although at the eastern fringes of their distribution knowledge of cows and milking can be linguistically reconstructed (Saxon 1982). Contact between Ubangian-speakers and Bantu-speakers had already occurred by ~4 kya, exemplified by the Ubangian borrowing of a Bantu word for oil palm (David 1982), but Ubangian-speakers appear to have gained knowledge of iron-working independently of the Bantu-speakers, probably from the Nigerian Nok complex (David 1982).

Within the last thousand years, peoples of European, Middle Eastern and Asian origins reached central Africa, and French, Portuguese, Belgian and British colonies were established. Of the countries examined in this study, the C.A.R. was colonised by France, whilst Uganda and Zambia were colonised by England. Most central African countries established their independence only within the past 50 years.

### 1.2 *Haploid genetic markers*

MtDNA and MSY-DNA constitute the haploid areas of the human genome. They do not pair with any other chromosomes during meiosis and do not experience genetic recombination. These loci are particularly powerful tools for population genetic studies because their genetic variation stems only from the accumulation of mutations over time. Genetic variation at these loci found in living individuals can be used to reconstruct ancestral genetic variation. In addition, each haploid molecule is in complete linkage disequilibrium. Therefore polymorphisms detected at different loci in the molecule may be used in combination (haplotypes), which allows more powerful analyses than if the polymorphisms were each considered separately.

#### 1.2.1 *The male-specific region of the Y chromosome*

The human Y chromosome is a small acrocentric chromosome, ~59Mb in length, which contributes ~ 2% of the total genome length (Skaletsky et al. 2003). Unlike other human nuclear chromosomes, the Y chromosome is found only in males and is critical for male sexual determination and differentiation. Normal males carry one Y chromosome and one X chromosome per cell. The distal tips of the Y chromosome pair with the X chromosome during meiosis (Burgoyne et al. 1982), but most of the Y chromosome remains unpaired during meiosis. This region has been called the male-specific region of the Y chromosome (MSY; Skaletsky et al. 2003) or the non-recombining region of the Y chromosome (NRY). The MSY is inherited from fathers to sons only, termed holandric or paternal inheritance. The study of MSY variation has emerged as a powerful tool for the investigation into variation and evolution of male lineages.

Studies in the 1980s and early 1990s failed to discover much variation on the Y chromosome (Casanova et al 1985; Lucotte and Ngo 1985; Jakubicza et al. 1989; Malaspina et al. 1990; Seielstad et al. 1994; Dorit et al. 1995; Whitfield et al. 1995). These studies led to the Y chromosome being described as being "gene-poor" and "nearly genetically blank" (Spurdle and Jenkins 1992; Mitchell and Hammer 1996). More recently, increasing numbers of polymorphisms on the MSY have been discovered, partially due to the advent of new technologies such as denaturing high performance liquid chromatography (DHPLC; Underhill et al. 1997). There are now several hundred informative MSY markers of different types that can be typed by convenient PCR-based techniques (e.g. Hammer 1994; Jobling and Tyler-Smith 1995; Underhill et al. 1997; Bergen et al. 1999; Semino et al. 2000; Shen et al. 2000; Underhill et al. 2000, 2001; Hammer et al. 2001). Polymorphism types include single nucleotide polymorphisms (SNPs), microsatellites (also called short tandem repeats or STRs), minisatellites, insertion-deletion polymorphisms and translocations. No doubt the recent completion of the sequencing of the euchromatic portion of the Y chromosome (Skaletsky et al. 2003) will facilitate the discovery of further polymorphisms.

The different types of polymorphisms present on the MSY have different mutation rates, and are caused by different mutational mechanisms. Repetitive DNA polymorphisms such as STRs have high mutation rates, with estimates ranging from 1.2 x $10^{-3}$ (Bianchi et al. 1998) to 2.8 x $10^{-3}$ (Kayser et al. 2000). This property is advantageous for population studies, because STRs are polymorphic in potentially all humans (Kayser et al 2001). However recurrent mutations can cause identical alleles in unrelated individuals (also referred to as homoplasy), so that STR alleles may or may not be identical by descent; microsatellites are not the markers of choice for constructing unique and reliable phylogenies (Jobling and Tyler-Smith 2000). SNPs and insertion-deletion polymorphisms have a very low rate of occurrence and are generally assumed to have occurred only once in human history, therefore most of these markers tend to be biallelic or binary (Jobling and Tyler-Smith 1995). One estimate for the mutation rate of binary markers is 2 x $10^{-8}$ per generation (Jobling 2001). Slowly-evolving markers unambiguously define groups of chromosomes that are descendants of single common ancestors (Thomas et al 1998; Weale et al. 2001) and identify stable paternal lineages (haplogroups) that can be traced back in time over thousands of years (YCC 2002).

The combined use of MSY polymorphisms with different mutation rates is advantageous for anthropological genetic studies. Potentially homoplasic STR data are resolved by their combined use with biallelic data, whilst the use of STRs allows the high-resolution exploration of diversity within haplogroups and estimations of time depths (Hurles et al. 1999; Helgason et al. 2000; Kayser et al. 2001).

The MSY is also particularly well-suited for human evolutionary studies because of its small effective population size. For each Y chromosome in a population, there are four of each autosome and three X chromosomes. The MSY is therefore particularly sensitive to founder effects, bottlenecks and genetic drift, and may show finer geographic structure than autosomal variation (Maynard Smith 1990; Hammer 1995; Perez-Lezaun et al. 1999; Weale et al. 2001). The lower rate of male migration (Seielstad et al. 1998) and/or unequal reproductive success amongst males are likely to lower the MSY effective population size even below that of mtDNA, and thus MSY variation may show even finer geographic structure than mtDNA variation.

The MSY shows a deficit of genetic diversity in comparison with the autosomes and the X chromosome, despite the fact that about twice as many mutations occur in male than in female meiosis due to the larger number of cell divisions during spermatogenesis (Shen et al. 2000). The lack of Y chromosome variation may be due to the smaller effective MSY population size relative to that of autosomes or the X chromosome (see above). A combination of natural selection plus lack of recombination could also lower diversity (Jobling and Tyler-Smith 1995, Paabo 1995; Mitchell and Hammer 1996). This theory posits that a selectively advantageous mutation on the MSY would spread through the population. Because all MSY genes are completely linked, the rest of the chromosome would simultaneously be spread through the population in an effect known as "genetic hitch-hiking" (Maynard-Smith and Haigh 1974). The resulting "selective sweep" would remove neutral genetic variation from the populations. However there is no genetic evidence to suggest that natural selection has had a significant influence on human Y chromosomes (Dorit et al. 1995; Hammer 1995; Goldstein et al. 1996; Underhill et al. 1997; Nachman 1998; Seielstad et al. 1999). Human male behavioural patterns may affect MSY diversity. For example, in polygyny, a small number of males may be reproductively more successful than others, causing a further reduction in the effective population size of the Y chromosome and resulting in lowered diversity (Jobling and Tyler-Smith 1995; Mitchell and Hammer 1996). Another example with similar effects is the levirate, where widows are remarried to a brother-in-law (Mitchell and Hammer 1996). Male population bottlenecks could also be caused by male involvement in wars and migrations.

Several MSY phylogenies and nomenclature systems (Su et al. 1999; Jobling and Tyler-Smith 2000; Semino et al. 2000; Underhill et al. 2000; Capelli et al. 2001; Hammer et al. 2001; Kaladjieva et al. 2001; Underhill et al. 2001) have recently been incorporated into one cohesive global MSY phylogeny (YCC 2002). This phylogeny was based largely on that proposed by Underhill et al. (2000, 2001), and used 245 Y biallelic markers to define 153 MSY lineages. These lineages are grouped into 18 major monophyletic Y chromosome haplogroups, named HGs A to R (YCC 2002). Most of these HGs have clear population-specific and or continent-specific distributions; for example, HGs A and B are found exclusively in African populations (Underhill et al. 2000; 2001). The root of the tree has been placed between HG A and the rest of the tree, and time to the most recent common ancestor (MRCA) has been estimated to be ~59 000 years (95% CI 40 000-140 000 years; Thomson et al. 2000). This date is in good agreement with a previous estimate of 46 000-91 000 years, based on eight microsatellites (Pritchard et al. 1999).

### 1.2.2 *Mitochondrial DNA*

The mitochondrial genome consists of a circular molecule of double-stranded DNA 16 569 bp in length (Anderson et al. 1981). MtDNA is found within the mitochondria, the cytoplasmic organelles responsible for oxidative phosphorylation in eukaryote cells. Each mtDNA codes for 13 polypeptides, 22 transfer RNAs and two ribosomal RNAs (Anderson et al. 1981; Wallace 1995). Almost all the non-coding DNA of the entire mtDNA molecule is concentrated in one 1.122kb region known as the control region or D-loop (Anderson et al. 1981). The control region has an extremely high mutation rate and includes two hypervariable regions, commonly called hypervariable regions I-II or HVRI-HVRII between positions 16024-16383 and 57-372 (Stoneking and Soodyall 1996; Gurven 2000; Stoneking 2000). MtDNA has other unique properties that have made it a popular marker in population genetics studies, including:

- strict maternal inheritance – mtDNA is passed from mothers only to their children (Giles et al. 1980). Studying mtDNA polymorphisms specifically allows maternal lineages to be traced;
- high copy number, with up to 9 000 copies of mtDNA in a single cell. This property was especially useful in the early days of population genetics when DNA extraction and amplification techniques were not yet perfected. Today this property allows the extraction and analysis of mtDNA from ancient sources;
- the relatively early date of knowledge of the complete mtDNA genome sequence (Anderson et al. 1981), allowing comparative analyses to be performed;
- fast overall mutation rate, estimated to be between 5 to 20 times faster than nuclear genes (Brown et al. 1979; Wallace 1995), therefore mtDNA is highly polymorphic;
- non-uniform mutation rate across the molecule. This is useful because polymorphisms with slower mutation rates have been used to define monophyletic lineages (haplogroups) whilst polymorphisms with faster mutation rates reveal patterns of diversity at higher resolution.

MtDNA has been used extensively in the field of population genetics to deal with diverse topics, including the issue of the origins and subsequent expansions of modern humans (Cann et al. 1987; Vigilant et al. 1991; Jorde et al. 1995; Quintana-Murci et al. 1999), the peopling of different parts of the world (Sykes et al. 1995; Forster et al. 1996; Richards et al. 2000), diversity at the continental level (Chen et al. 1995; Horai et al. 1996; Redd and Stoneking 1999; Simoni et al. 2000), population admixture (Comas et al. 1998) and the study of ancient DNA (Krings et al. 1997, 2000; Ovchinnikov et al. 2000; Adcock et al. 2001).

The first evidence that mtDNA variation was correlated with the ethnic and geographic origin of the individual came from studies of mtDNA restriction fragment length polymorphisms (RFLPs) in wide-spread populations (Brown 1980; Denaro et al. 1981; Johnson et al. 1983; Cann et al. 1987; Merriwether et al. 1991; Torroni et al. 1993, 1996). The use of a core set of six restriction enzymes (REs) to produce RFLPs was later replaced by the use of a "high-resolution" set of twelve REs, and then by an "extended set" of 14 REs (e.g. Chen et al. 1995). The existence of population- and continent-specific mtDNA lineages was also confirmed by studies of variation in one or both hypervariable regions of the control region (e.g. Horai and Hayasaka 1990; Vigilant et al. 1991; Watson et al. 1996; Bandelt and Forster 1997). The deletion of one of two copies of a 9-bp repeat motif between the COII and tRNA$^{Lys}$ genes (commonly referred to as a 9-bp deletion) has also proven to be an informative phylogeographic marker (Wrischnik et al. 1987; Hertzberg et al. 1989; Soodyall et al. 1996). Only recently were attempts made to combine marker types (Graven et al. 1995; Soodyall et al. 1996; Watson et al. 1997; Passarino et al. 1998; Alves-Silva et al. 2000; Bandelt et al. 2001; Pereira et al. 2001; Salas et al. 2002), and in the past three years complete mtDNA genome sequencing has become popular (Horai et al. 1995; Ingman et al. 2000; Finnila et al. 2001; Maca-Meyer et al. 2001; Parsons and Coble 2001; Torroni et al. 2001; Herrnstadt et al. 2002).

The different types of mtDNA data have now been synthesised into a holistic global mtDNA phylogeny (Chen et al. 1995; Graven et al. 1995; Macaulay et al. 1999; Finnila et al. 2001; Torroni et al. 2001; Salas et al. 2002; Herrnstadt et al. 2002). MtDNA types from African populations are divided into three major haplogroups (HGs) called L1, L2 and L3. HG L1 (defined by +3592 *Hpa*I and -10806 *Hinf*I) and HG L2 (defined by +3592 *Hpa*I and +16389 *Hinf*I or -16390 *Ava*II) are found exclusively in African populations, and collectively are sometimes referred to as macrohaplogroup L* (Chen et al. 2000). L1 has also been called a 'paragroup" (Brehm et al. 2002) since it is paraphyletic (Salas et al. 2002). Seven clades have been identified within L1, and are named alphabetically L1a to L1f, and L1k (Chen et al. 1995; Pereira et al. 2001; Salas et al. 2002), and four clades within L2 have been identified and are named L2a-L2d (Torroni et al. 2001). The root of the global mtDNA phylogeny, representing the MRCA, is based within HG L1 and is at least 150 000 – 170 000 years old (Horai et al. 1995; Ingman et al. 2000).

MtDNA lineages within HG L3 (defined by –3592 *Hpa*I) are named differently depending on whether they are found in African or non-African populations. Eurasian L3 mtDNA types are classified into the two macrohaplogroups M and N; HGs within these macrohaplogroups are named A, B, C etc. MtDNA types from HG L3 in African populations are classified into at least five subHGs, called L3b and L3d-g (Watson et al. 1997; Rando et al. 1998; Bandelt et al. 2001; Salas et al. 2002). The nomenclature L3a and L3c (Watson et al. 1997) has been discontinued, because mtDNA types in L3c have been reclassified as belonging to HG U6 (Rando et al. 1998), and mtDNA types in L3a need further resolution. A significant proportion of mtDNA types from African populations that belong to L3 have not yet been classified into subHGs. Following the nomenclature system of Macaulay et al. (1999), these mtDNA types are generally referred to collectively as belonging to L3* (the as yet further undefined default L3 lineage). This has also been referred to as L3A* i.e. the African-specific portion of L3* (Rando et al. 1998; Salas et al. 2002).

### 1.2.3 *Limitations of haploid markers*

Genetic studies have generally affirmed the utility of both mtDNA and MSY DNA in the reconstruction of human evolution and for tracing population relationships (Stoneking and Soodyall 1996). However since these haploid markers each represent single genetic loci, interpretations of mtDNA and MSY data in studies of population history should be made with caution. It is possible that in studying these loci we are only learning about the histories of these two genes and not the history of the population in which they are found, since only one out of $2^n$ ancestral lineages from n generations is represented by each mtDNA or MSY lineage. Moreover their peculiarities in transmission may limit their value as indicators of human evolutionary events. The genetic "hitch-hiking" effect described for the MSY (above) is equally applicable to mtDNA, so that their patterns of variation, even in non-coding regions, may reflect natural selection rather than population history (Jorde et al. 1995). There is some evidence that natural selection has affected certain mtDNA haplogroups (Torroni et al. 2001; Mishmar et al. 2003; Ruiz-Pesini et al. 2004); this is supported by the observations that certain mtDNA HGs appear to be associated with specific mitochondrial diseases (Brown et al. 1997; Brown et al. 2002; Sudoyo et al. 2002; Howell et al. 2003). In summary, patterns of variation at haploid loci may exist due to drift or selection and may differ drastically from patterns at other loci. Wherever possible haploid data should be compared to data from other independent (nuclear) loci (Stoneking and Soodyall 1996).

## 1.3 *Studies of haploid genetic variation in African populations*

Studies of mtDNA and MSY-DNA variation in African populations have increased in frequency in the last few years (Soodyall et al. 1996; Watson et al. 1997; Passarino et al. 1998; Rando et al. 1998; Scozzari et al. 1999; Bosch et al. 2000, 2001; Bandelt et al. 2001; Salas et al. 2002; Cruciani et al. 2002; Pereira et al. 2002; Semino et al. 2002; Knight et al. 2003). These studies have revealed a great amount of genetic diversity and interesting lineage structure, as well as allowing the affinities of some African populations to be examined.

### 1.3.1 *Y chromosome studies in African populations*

Both biallelic markers and short tandem repeat (STR) loci within the MSY have previously been used to examine genetic variation in Africa. Of the 18 major Y chromosome HGs recently defined by biallelic markers (Underhill et al. 2000, 2001; YCC 2002), only seven have been reported in African populations, including HGs A, B, E, G, J, K and R (Underhill et al. 2000; Semino et al. 2002; Cruciani et al. 2002). In contrast to mtDNA and autosomal loci, Y binary markers in African populations do not show the highest diversity in worldwide comparisons (Underhill et al. 2001). This may be due to ascertainment bias in the discovery of SNP markers. Alternatively, the failure to detect intermediates for deeply divergent African Y chromosome lineages suggests that repeated episodes of population contractions have eliminated Y chromosome binary diversity that accumulated during periods of expansion (Underhill et al. 2001).

MSY HGs A and B have been found exclusively in Africa and are suggested to have deeper genealogical heritage than the other HGs (Underhill et al. 2000, 2001). The most frequent and widespread MSY HG in Africa is HG E (Spurdle et al. 1994; Hammer et al. 1998; Scozzari et al. 1999; Underhill et al. 2001; Cruciani et al. 2002). Two sublineages within HG E, called E-M2 and E-M35, have different geographic distributions within Africa. HG E-M2 is common in sub-Saharan Africa, whereas HG E-M35 is found in populations from north and east Africa, as well as in the Mediterranean Basin and Europe (Underhill et al. 2001; Cruciani et al. 2002). Y chromosomes belonging to HGs G, J, K and R have also been observed in African populations, albeit at varying and relatively low frequencies (Bosch et al. 2001; Underhill et al. 2001; Cruciani et al. 2002; Semino et al. 2002).

MSY STR variation in African populations has been analysed in a global context (Pritchard et al. 1999; Seielstad et al. 1999; Forster et al. 2000; Kayser et al. 2001), and to focus on African-specific diversity (see below). Global studies have usually used the whole or partial forensic panel of STRs comprising DYS19-DYS388-DYS389-DYS390-DYS391-DYS392-DYS393. These studies have not consistently shown higher genetic diversity in African populations than other global populations as seen using mtDNA and nuclear data (see Seielstad et al. 1999 vs. Pritchard et al. 1999). The same set of STRs have been used in studies focussing on north African populations (Bosch et al. 1999, 2000, 2001), southern African Bantu-speaking populations (Thomas et al. 2000; Lane et al. 2002; Pereira et al. 2002), populations from Sao Tome and Principe (Corte-Real et al. 2000; Trovoada et al. 2001), and other previous Portuguese colonies, including Guinea Bissau and Angola (Corte-Real et al. 2000). Other authors, notably Scozzari et al. (1999) and Cruciani et al. (2002) utilised different Y STRs, including YCAIIa , YCAIIb, DYS413a, DYS413b, A7.2 and DYS439, to study the genetic affinities of African populations.

In some studies (e.g. Seielstad et al. 1999, Trovoada et al. 2001; Caglia et al. 2003), Y-STR data only have been used to examine the affinities of genetic African populations even though homoplasy may affect the ability of these markers to reflect the true genetic and historical relationships among populations. In the present study it was questioned whether Y chromosome markers with different mutation rates were capable of revealing similar population affinities. In other studies, Y STR loci have been typed in conjunction with biallelic loci (Bosch et al. 1999, 2000, 2001; Scozzari et al. 1999; Thomas et al. 2000; Cruciani et al. 2002; Pereira et al. 2002). This approach showed that Y chromosome STR variation is highly structured by biallelic background (Bosch et al. 1999); allowed the expansion times of particular lineages to be estimated (e.g. Thomas et al. 2000); allowed the prediction of monophyletic lineages in African populations before the defining biallelic markers had been described (Scozzari et al. 1999); and showed geographic structuring of STR haplotype variation within particular Y chromosome HGs (Scozzari et al. 1999; Cruciani et al. 2002). Combined use of biallelic and STR markers has also allowed founder effects in African populations to be detected. For example, Thomas et al. (2000) and Pereira et al (2002) described a strong founder effect for Y STRs in South African Bantu-speakers; they suggested that one specific STR haplotype within HG E-M2 was a signature of the 'Bantu Expansion", and named this haplotype the "Bantu Modal haplotype" (BMH). Using the STR markers DYS19-DYS388-DYS390-DYS391-DYS392-DYS393, this "BMH" was defined as 15-12-21-10-11-13.

### 1.3.2 *MtDNA studies in African populations*

African populations were among the first human groups to be analyzed for mtDNA variation, although early studies focused on African variation within a global context only. Some of the first mtDNA RFLP analyses showed that most African mtDNAs had different *HpaI* restriction patterns to those in Eurasian populations (Denaro et al. 1981), and suggested that the greatest diversity of mtDNA occurred in African populations (Johnson et al. 1983). African-American populations were represented in the work of Cann et al. (1987) in their pioneering use of mtDNA RFLPs to support the hypothesis that modern *Homo sapiens* evolved in Africa, and Africans were also represented in early studies of mtDNA control region sequence variation in global populations (Vigilant et al. 1989; Horai and Hayasaka 1990; Vigilant et al. 1991).

Attention has subsequently turned towards the analysis of African mtDNA variation at an intra-continental level. One approach has been the use of mtDNA data to describe the genetic relationships among African populations. The genetic affinities of African hunter-gatherer populations (Khoisan-speaking populations, and Biaka and Mbuti Pygmy populations) have been a particularly popular topic in the literature (fig. 1.1; Soodyall and Jenkins 1992; Chen et al. 1995; Watson et al. 1996; Bandelt and Forster 1997; Chen et al. 2000; Knight et al. 2003). Several American populations with African ancestry and populations from specific geographic areas within Africa have also been studied, including populations from north-west, west, north-east, east and southern Africa (see Table 2.3). There is a lack of mtDNA data from the populations of central Africa, which are represented in the literature only by populations from Sudan and Equatorial Guinea, the island populations of Sao Tome e Principe, and very small samples of Biaka and Mbuti Pygmies (Table 2.3).

A second approach to the analysis of African mtDNA has been the identification of mtDNA signatures of African population expansions and migrations, versus geographically isolated or population-specific variants (Soodyall et al. 1996; Watson et al. 1997; Rando et al. 1998; Quintana-Murci et al. 1999; Chen et al. 2000; Bandelt et al. 2001; Salas et al. 2002). The geographic origins and directions of dispersals of many of the more widespread African lineages have been deduced by the use of phylogenetic network analysis and from patterns of lineage frequency and diversity (Watson et al. 1997; Salas et al. 2002). Several mtDNA markers have been proposed as signals of the "Bantu Expansion", including L1a2 (the subset of L1a associated with a 9-bp deletion), L2a, L3b and L3e1 (Bandelt et al. 1995; Chen et al. 1995; Soodyall et al. 1996; Watson et al. 1997; Alves-Silva et al. 2000; Bandelt et al. 2001; Pereira et al. 2001; Salas et al. 2002).

## 1.4 *Aims of this study*

In order to address the paucity of mtDNA and MSY data from central African populations, genetic variation at these loci was examined in nearly 400 individuals from populations from Uganda, Zambia and the Central African Republic (C.A.R.). Individuals from the Democratic Republic of Congo (D.R.C) were also included in the Y chromosome studies. More specifically, this study was directed towards:

1) exploring the patterns of mtDNA and Y chromosome variation in each central African population sampled, in order to contribute to the understanding of the overall patterns of haploid genetic variation in Africa;
2) examining the genetic affinities of central African (Pygmy and non-Pygmy) populations with each other and with other African populations;
3) assessing the concordance of haploid markers with different mutation rates in assessing population affinities;
4) comparing male and female migration rates in African populations;
5) refining theories regarding the prehistory of central Africa populations based on traditional linguistic and archaeological methods.

# 2. SUBJECTS AND METHODS

## 2.1 *Subjects*

Over 800 unrelated individuals from the Central African Republic, Zambia and Uganda volunteered to participate in this study. A total of 820 individuals were screened for the mitochondrial 9-bp deletion, whilst 397 samples were utilised for studies of other mtDNA markers (fig. 2.1, Table 2.1). Y chromosome variation was examined in a total of 369 individuals, including samples from individuals from the Democratic Republic of Congo (fig. 2.1, Table 2.1). All samples were collected with the subjects' informed consent and with the necessary Government consent from regions where sampling was conducted. This study was approved by the Committee for Research on Human Subjects at the University of Witwatersrand, Johannesburg, South Africa (Trefor Jenkins, Protocol Number 24/4/87 and Debra Barkhan, Protocol Number M980601; Appendix 8.2).

Ten to twenty ml of blood from volunteers were collected into EDTA venipuncture tubes and were air freighted to Johannesburg from the point of collection. At the time of sampling, subjects were asked to provide information on their place of birth and the language spoken by them and by their parents. This information was used to group individuals into ethnic groups defined by their mother's ancestry for mtDNA studies, or ethnic groups defined by their father's ancestry for Y chromosome studies. However in the Zambian samples, details of the individual's ethnic identity only (and not maternal and paternal ethnicity) were provided.

The Zambian samples were collected in Lusaka, and the Ugandan samples were obtained from the districts of Kissoro, Kabale and Rukungiri in south-western Uganda (fig. 2.1) Sampling in the D.R.C. was conducted at Luozi and Kinshasha (fig. 2.1). All of the ethnic groups sampled from Zambia, Uganda and the D.R.C. speak languages classified as Bantu within the Niger-Kordofanian language family (Greenberg 1963; Grimes et al. 2001; Table 2.1).

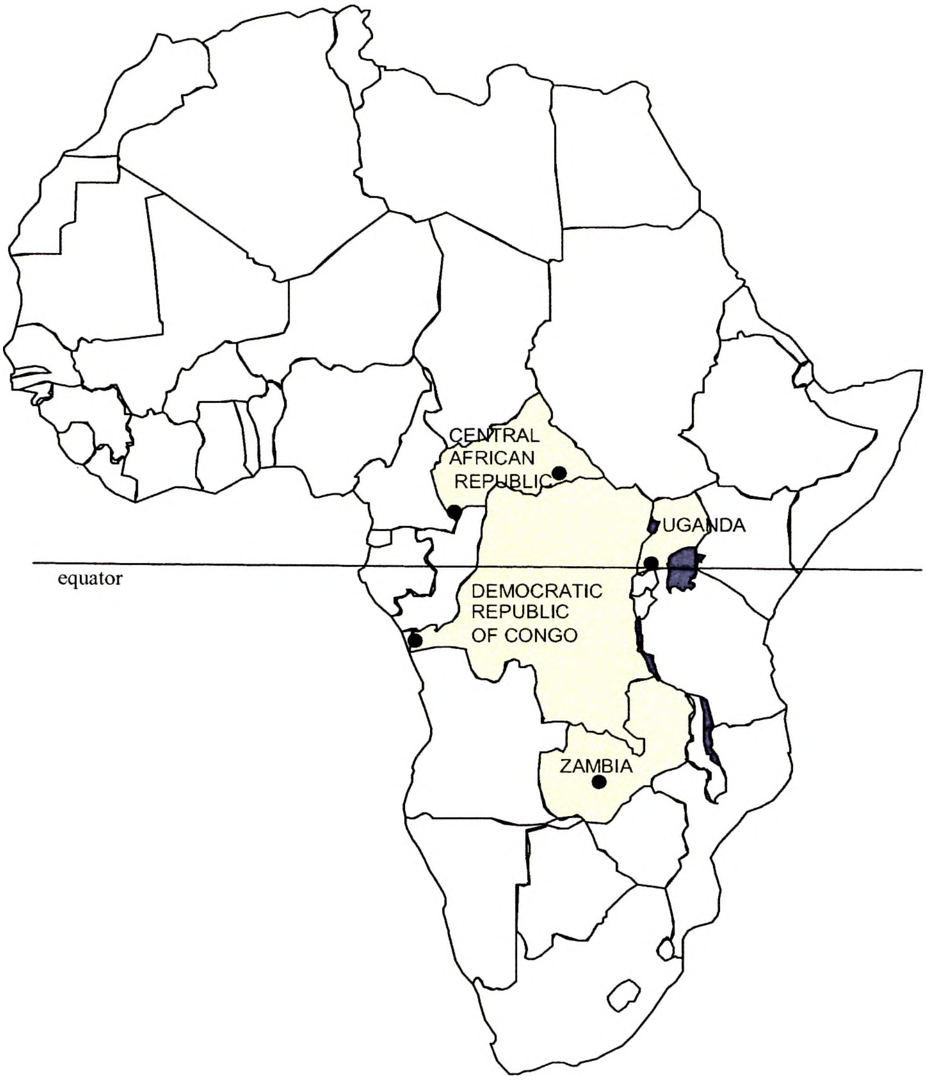


**Figure 2.1** Map of Africa showing countries (highlighted in green) and towns/villages (black dots) from which samples were collected for use in the present study. Blue areas indicate lakes.

**Table 2.1** Number of central African individuals in which mtDNA and MSY variation were examined, and their ethnic and linguistic affiliations, and country of origin.

| Country | Ethnic group | N (mtDNA) | N (Y chromosome) | Language family: subgroup[1] |
|---|---|---|---|---|
| **Uganda** | Bakiga | 28 | 30 | NK: Bantu,Narrow Bantu, J |
| | Bafumbira | 17 | 9 | NK: Bantu,Narrow Bantu, J |
| | Banyankole | 22 | 39 | NK: Bantu,Narrow Bantu, J |
| | Bahororo | 19 | 20 | NK: Bantu,Narrow Bantu, J |
| | Other | 6 | - | |
| **total** | | **92** | **98** | |
| **C.A.R. (Pygmy)** | Biaka | 71 | 20 | NK: Bantu, Narrow Bantu, C |
| | Babenzele | 43 | 20 | NK: Bantu, Narrow Bantu, C |
| **total** | | **114** | **40** | |
| **C.A.R. (non-Pygmy)** | Sangha-Sangha | 16 | 18 | NK: Adamawa-Ubangian |
| | Nzakara | 28 | 22 | NK: Adamawa-Ubangian |
| | Mbimou | 17 | 4 | NK: Adamawa-Ubangian |
| | Gbaya | 34 | 31 | NK: Adamawa-Ubangian |
| **total** | | **95** | **75** | |
| **Zambia** | Nyanja | 17 | 22 | NK: Bantu,Narrow Bantu, N |
| | Lozi | 24 | 22 | NK: Bantu, Narrow Bantu, S |
| | Bemba | 23 | 16 | NK: Bantu, Narrow Bantu, M |
| | Tonga | 13 | 12 | NK: Bantu, Narrow Bantu, M |
| | Other | 19 | 15 | NK: Bantu |
| **total** | | **96** | **87** | |
| **D.R.C.** | Manyanga | - | 61 | NK: Bantu, Narrow Bantu, D |
| | Other | - | 8 | |
| **total** | | - | **69** | |
| CENTRAL AFRICA TOTAL | | 397 | 369 | |

[1] Grimes et al. (2001); for Bantu language subgroups, see fig 1.4

Both Pygmy and non-Pygmy groups from the C.A.R. were sampled. The sample of Pygmies consisted of Biaka and Babenzele, both of whom belong to the Aka subgroup (Cavalli-Sforza 1986). Using both Y chromosome data and autosomal loci, Coia et al. (2004) have shown that the Biaka and Babenzele are closely related to each other. These two populations were pooled in this study and are collectively referred to as "Biaka". The Biaka have lost their original language and now speak the Bantu languages of the farming communities among whom they live (Grimes et al. 2001). The sampled non-Pygmy individuals speak languages classified as Adamawa-Ubangian, also within the Niger-Kordofanian language family (Grimes et al. 2001).

The breakdown of the samples from each country by ethnic group, their linguistic affiliations (Ruhlen 1987; Grimes et al. 2001), and the numbers of individuals from each group analysed for mtDNA and Y chromosome variation in this study, are shown in Table 2.1. Because sample sizes of many of the sampled individual ethnic groups were small (N<25), groups sampled in the same geographic locations were pooled. Linguistic classifications also support the groupings used. For instance, all ethnic groups sampled from the Uganda region speak languages classified as belonging to zone J within central Bantu (Guthrie 1967; Grimes et al. 2001; fig. 1.4), and all non-Pygmy groups from the C.A.R. speak languages classified within the Adamawa-Ubangian linguistic family.

Published mtDNA and NRY from other African populations (Tables 2.2, 2.3) were used for comparative analyses. Populations were classified according to geographic origins and by their linguistic affiliations (Grimes et al. 2001).

**Table 2.2** Geographic origins and linguistic affiliations of populations used for comparative Y chromosome studies

| GEOGRAPHIC REGION | COUNTRY | ABBREVIATION | POPULATION | LANGUAGE FAMILY[a] | N[b] | N[c] | N[d] | REFERENCE[e] |
|---|---|---|---|---|---|---|---|---|
| north-east | Ethiopia | ORO | Oromo | AA: Semitic | 78 | | | 4 |
| north-east | Ethiopia | AMH | Amhara | AA Semitic | 48 | | | 4 |
| north-east | Ethiopia | ETH | Bench, Berta, Dasenech, Dizi, Dogon, Hamar, Konso, Majangir, Nyangatom Ongota,Surma, Tsamako | AA Semitic | | 91 | 91 | 3 |
| north-west | Morocco | ARA | Arabs | AA Semitic | 49 | 44 | 44 | 1, 2 |
| north-west | Morocco | BER | Berbers | AA Semitic | 64 | 60 | 60 | 1, 2 |
| north-west | Morocco | SAH | Saharawis | AA Semitic | | 29 | 29 | 1 |
| west | Mali | MAL | Fulbe, Bozo, Songhai | NK and NS | | 15 | 15 | 3 |
| west | Cape Verde | CVD | | IE Portuguese | | | 47 | 7 |
| west | Guinea-Bissau | GNB | | NK Northern Atlantic, IE: Portuguese | | | 33 | 7 |
| west | B Faso | MOS | Mossi | NK Voltaic | 49 | | | 2 |
| west | B Faso | RIM | Rimaibe | NK West Atlantic | 37 | | | 2 |
| west | Senegal | SEN | Senegal | NK. Northern Atlantic | 139 | | | 4 |
| west | B Faso and Cameroon | FLB | Fulbe | NK West Atlantic | 37 | | | 2 |
| central | Cameroon | FAL | Fali | NK Adamawa | 39 | | | 2 |
| central | Cameroon | BAM | Bamileke | NK Bantoid | 48 | | | 2 |
| central | Cameroon | EWO | Ewondo | NK Bantu | 29 | | | 2 |
| central | Sao Tome e Principe | STP | Angolares, Forros, Tongas | IE Portuguese | | | 104 + 34 | 6, 7 |
| central | Angola | ANG | | NK Bantu, IE Portuguese | | | 50 | 7 |
| **central** | **CAR** | **CAR** | **Sangha, Gbaya, Nzakara, Mbimou** | **NK: Adamawa** | **75** | **75** | **75** | **this study** |
| **central** | **Uganda** | **UGA** | **Bakiga, Banyankole, Bahororo, Bufumbira** | **NK: Bantu** | **98** | **98** | **98** | **this study** |
| **central** | **DRC** | **DRC** | **Manyanga** | **NK: Bantu** | **69** | **69** | **69** | **this study** |
| **central** | **Zambia** | **ZAM** | **Lozi, Nyanja, Tonga, Bemba, others** | **NK: Bantu** | **87** | **87** | **87** | **this study** |
| **central - Pygmy** | **CAR** | **CAP** | **Biaka and Babenzele** | **NK: Bantu** | **40** | **40** | **40** | **this study** |
| central - Pygmy | CAR | BIA | Biaka | NK: Bantu | 20 | 20 | 31+20 | 2, 3, 5 |
| central - Pygmy | DRC | MBT | Mbuti | NS. Central Sudanic | 12 | | | 2 |
| south | Mozambique | MOZ | Changana, Ronga, Chope, others | NK Bantu | | | 37 +66 | 7, 8 |
| south | South Africa | SAF | Sotho, Swazi, Tswana, Xhosa, Zulu | NK Bantu | | 56 | 56 + 77 | 3, 9 |
| south - Khoisan | Namibia | KNG | !Kung | KS: northern | 64 | | | 2 |
| south - Khoisan | Namibia | KHW | Khwe | KS. central | 26 | | | 2 |
| south - Khoisan | Namibia | SAN | Sekele, "Omega San" | KS northern | | 29 | 29 | 3 |
| TOTAL AFRICA | | | | | 1108 | 713 | 1192 | |

[a] African language families AA = Afro-Asiatic, NK = Niger-Kordofanian, NS = Nilo-Saharan, KS = Khoisan, IE= Indo-European

[b] sample size using haplogroups

[c] sample size using 8 STRs (DYS19, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393)

[d] sample size using 5 STRs (DYS19, DYS390, DYS391, DYS392, DYS393)

[e] 1 = Bosch et al. 2001, 2 = Cruciani et al. 2002, 3 = Seielstad et al 1999, 4 = Semino et al. 2002
5= Kayser et al 2001, 6= Trovoada et al 2001, 7=Corte-Real et al 2000, 8=Pereira et al 2002, 9=Thomas et al 2000

**Table 2.3** Geographic origins and linguistic affiliations of populations used for comparative mtDNA studies

| GEOGRAPHIC REGION | COUNTRY | NUMBER CODE | POPULATION | LANGUAGE FAMILY[a] | N[b] | N[c] | REFERENCE |
|---|---|---|---|---|---|---|---|
| non-African | South America | | Brazil | n/a | | 247 | Alves-Silva et al 2000 |
| non-African | mixed | | L3e types | n/a | | 66 | Bandelt et al 2001 |
| non-African | South Africa, Malawi, C A R , D R C | | types with 9-bp deletion | n/a | | 41 | Soodyall et al 1996 |
| non-African | USA | | African-American | n/a | | 8 | Vigilant et al 1991 |
| north-west | Canary Islands | 27 | | IE | 54 | | Pinto et al 1996 |
| north-west | Mauritania | 24 | | AA Semitic, NK: Northern Atlantic | 30 | | Rando et al 1998 |
| north-west | Algeria | 28 | | AA Semitic, AA: Berber | 85 | | Corte-Real et al 1996 |
| north-west | Morocco | 25 | non-Berber | AA Semitic, IE | 32 | | Rando et al 1998 |
| north-west | Morocco | 29 | Berbers | AA Berber | 60 | | Rando et al 1998 |
| north-west | West Sahara | 26 | | AA Semitic | 25 | | Rando et al 1998 |
| west | Nigeria, Niger Mali | 21 | Songhai | NS Songhai | 10 | | Watson et al 1997 |
| west | Nigeria, Niger, Mali | 17 | Tuareg | AA Berber | 23 | | Watson et al 1997 |
| west | Nigeria, Niger, Mali | 20 | Yoruba | NK Voltaic | 33 | 13 | Watson et al 1997, Vigilant et al 1991 |
| west | Nigeria, Niger | 22 | Hausa | AA Chadic | 20 | | Watson et al 1997 |
| west | Nigeria, Niger, Benon, Cameroon, B Faso | 14 | Fulbe | NK Northern Atlantic | 60 | | Watson et al 1997 |
| west | Nigeria, Niger | 13 | Kanuri | NS Saharan | 14 | | Watson et al 1997 |
| west | Senegal | 18 | Mandenka | NK Northern Atlantic | 110 | 60 | Graven et al 1995 |
| west | Senegal | 15 | | NK Northern Atlantic | 50 | | Rando et al 1998 |
| west | Senegal | 16 | Wolof | NK Northern Atlantic | 48 | | Rando et al 1998 |
| west | Senegal | 19 | Serer | NK Northern Atlantic | 23 | | Rando et al 1998 |
| north-east | Ethiopia | 10 | | AA Semitic | 74 | | Thomas et al 2002 |
| north-east | Egypt | 12 | | AA Semitic | 68 | | Krings et al 1999 |
| north-east | Sudan, Egypt | 9 | Nubian | AA Semitic | 80 | | Krings et al. 1999 |
| north-east | Somalia, Kenya, Ethiopia | 11 | Somali | AA Cushitic | 27 | | Watson et al 1997 |
| east | Kenya | 7 | Turkana | NS. Eastern Sudanic | 37 | | Watson et al 1997 |
| east | Kenya | 6 | Kikuyu | NK Bantu | 25 | | Watson et al 1997 |
| east | Tanzania | 23 | Hadza | KS Hadza | 49 | 66 | Knight et al 2003, Vigilant et al 1991 |
| east | Tanzania | 5 | Datoga | NS Eastern Sudanic | 18 | 18 | Knight et al 2003 |
| east | Tanzania | 4 | Iraqw | AA Cushitic | 12 | 12 | Knight et al 2003 |
| east | Tanzania | 8 | Sukuma | NK Bantu | 21 | 21 | Knight et al 2003 |
| central | Sudan | 36 | Nuba, Shilio, Duba, Nuer | NS Eastern Sudanic | 76 | | Krings et al 1999 |
| central | Sao Tome e Principe | 38 | | IE Portuguese | 50 | | Mateu et al 1997 |
| central | Equatorial Guinea | 34 | Fang | NK Bantu | 11 | | Pinto et al. 1996 |
| central | Equatorial Guinea | 39 | Bubi | NK Bantu | 45 | | Mateu et al. 1997 |
| **central** | **CAR** | **37** | **Sangha, Gbaya, Nzakara, Mbimou** | **NK: Adamawa** | **95** | **95** | **this study** |
| **central** | **Uganda** | **32** | **Bakiga, Banyankole, Bahororo, Bufumbira** | **NK: Bantu** | **92** | **92** | **this study** |
| **central** | **Zambia** | **33** | **Lozi, Nyanja, Tonga, Bemba, others** | **NK: Bantu** | **96** | **96** | **this study** |
| **central - Pygmy** | **CAR** | **41** | **Biaka and Bahenzele** | **NK: Bantu** | **114** | **114** | **this study** |
| central - Pygmy | CAR | 40 | Biaka | NK Bantu | 17 | 17 | Watson et al 1997, Vigilant et al 1991 |
| central - Pygmy | DRC | 30 | Mbuti | NS Central Sudanic | 13 | 20 | Watson et al 1997, Vigilant et al 1991 |
| south | Mozambique | 31 | mixed | NK Bantu | 109 | 109 | Pereira et al 2001 |
| south | Mozambique | 35 | mixed | NK Bantu | 307 | | Salas et al. 2002 |
| south | Namibia | | Herero | NK Bantu | | 27 | Vigilant et al 1991 |
| south - Khoisan | Botswana, Namibia, Angola | 1 | !Kung | KS Northern | 43 | 43 | Chen et al 2000 |
| south - Khoisan | Botswana, Namibia, Angola | 3 | !Kung | KS Northern | 25 | 25 | Watson et al 1997, Vigilant et al 1991 |
| south - Khoisan | Angola, Namibia | 2 | Khwe | KS Central | 31 | 31 | Chen et al 2000 |
| TOTAL AFRICA | | | | | 2212 | 859 | |

[a] African language families. AA = Afro-Asiatic, NK = Niger-Kordofanian, NS = Nilo-Saharan, KS = Khoisan, IE= Indo-European

[b] sample size using HVRI sequences only

[c] sample size using HVRI and HVRII sequences

## 2.2 *DNA extraction*

DNA was extracted from stored buffy coats using the salting out method (Miller et al. 1988) and samples were stored at $-20^{0}$ C. Aliquots of extracted DNA were quantified by gel electrophoresis using standards of known concentration.

## 2.3 *Y chromosome molecular methods and analyses*

A total of 369 samples were typed for 16 Y-specific biallelic markers and eight Y-specific STRs.

### 2.3.1 *Biallelic marker typing*

The recent nomenclature suggested by the Y Chromosome Consortium (2002) was adopted in describing HGs defined by biallelic markers (fig. 3.1). Ten biallelic markers (M42, $SRY_{10831}$, M168, RPS4Y, M1, M40, M213, M9, M74 and M207) were used to define the major Y chromosome HGs. Three markers (M2, M35 and M191) were used to resolve sublineages within HG E, and another three (M150, M112 and M211) were used to resolve sublineages within HG B (fig. 3.1). Not all markers were typed in all individuals since a hierarchical typing system was followed according to the phylogeny defined by Underhill et al. (2000, 2001) and the YCC (2002). Biallelic markers in the sample of individuals from the D.R.C. were typed by Akashnie Maharaj as part of another study.

The M1 polymorphism or Y *Alu* insertion polymorphism (YAP) was detected using primers and PCR conditions suggested by Hammer and Horai (1995). PCR/ RFLP assays for $SRY_{10831}$, RPS4Y, M2, M40 and M9 were adapted from published protocols (Santos et al. 1999; Bergen et al. 1999; Thomas et al. 1999; Underhill et al. 1997, respectively). The primer sequences for detecting the M42, M168, M213, M74 and M207 polymorphisms were published by Underhill et al. (2001), while the M35, M191, M150, M112 and M211 polymorphisms were amplified using primers designed by A. Nebel (personal communication). Primer sequences for all 16 biallelic markers are presented in Table 2.4. PCR reactions mixes and cycling conditions used to amplify each biallelic marker are shown in Table 2.5. Table 2.5 also shows PCR amplicon size, the restriction enzyme used to detect each of the SNPs, and the product sizes after digestion.

### 2.3.2 *STR marker typing*

Six tetranucleotide Y-specific STRs (DYS19, DYS389I, DYS389II, DYS390, DYS391 and DYS393) and two trinucleotide STRs (DYS388 and DYS392) were amplified by PCR with fluorescently-labelled primers, and resolved by electrophoresis on polyacrylamide gels in an automated ABI 377 DNA sequencer. Primer sequences for all STR loci are shown in Table 2.4. DYS19 is also sometimes referred to as DYS394. Note that the DYS389 forward primer binds in two places on the Y chromosome and when used in conjunction with DYS389-R, amplifies two PCR fragments DYS389I and DYS389II.

The eight STR loci were amplified in two separate multiplex PCR reactions. In the first multiplex, loci DYS19, DYS390, DYS391 and DYS393 were amplified according to the method suggested by Lane et al. (2002; Table 2.6). The DYS388, DYS389I, DYS389II, and DYS392 loci were amplified in a second multiplex using similar PCR conditions but an annealing temperature of $56^{\circ}$ C (Table 2.6).

The number of repeats at each locus was determined using sequenced controls with known numbers of repeats and the aid of GeneScan v3.1.2 and Genotyper v2.5 Software (Applied Biosystems/PE). The correspondence between allele size in bp and number of repeats at each locus is shown in Appendix 8.3. We adopted the allele nomenclature proposed by Kayser et al. (1997). The DYS389II allele was scored in the standard manner by subtracting the size of the corresponding DYS389I allele. STR haplotypes (hts) were constructed using number of repeats in the order DYS19-DYS388-DYS389I-DYS389II-DYS390-DYS391-DYS392-DYS393.

**Table 2.4** Sequences of primers used to amplify sixteen Y-chromosome biallelic markers and eight Y-chromosome STRs.

| Y polymorphism | Primer | Primer sequence | Reference |
|---|---|---|---|
| M42 | M42-F | 5'-AAA GCG AGA GAT TCA ATC CAG-3' | Underhill et al. 2000 |
| | M42-R | 5'- TTT TAG CAA GTT AAG TCA CCA GC -3' | Underhill et al. 2000 |
| SRY$_{10831}$ | SRY$_{10831}$- F | 5'- TCC TTA GCA ACC ATT AAT CTG G-3' | Santos et al. 1996 |
| | SRY$_{10831}$ -R | 5'- AAA TAG CAA AAA ATG ACA CAA GGC -3' | Santos et al. 1996 |
| M168 | M168-F | 5'- AGT TTG AGG TAG AAT ACT GTT TGC T-3' | Underhill et al. 2000 |
| | M168-R | 5'- AAT CTC ATA GGT CTC TGA CTG TTC -3' | Underhill et al. 2000 |
| M150 | M150-F | 5' - CCC ACA CAC ACA GAT AGA CGT-3' | A. Nebel (unpublished) |
| | M150-R | 5' -CCT ACT TTC CCC CTC TTC TG-3' | A. Nebel (unpublished) |
| M112 | M112-F | 5' - GAA AAG CAA AAG AGA ACT GCC-3' | A. Nebel (unpublished) |
| | M112-R | 5' - TTT CTT GAT GAT GAG ACC AAT ATT TC-3' | A. Nebel (unpublished) |
| M211 | M211-F | 5' - CAT TTA CCC ACC CAA TCC AC-3' | A. Nebel (unpublished) |
| | M211-R | 5' - TGT GGA GAA ATT TGC TGG AA-3' | A. Nebel (unpublished) |
| RPS4Y711 | RPS4Y-F | 5'- CTG TAC TTA CTT TTA TCT CCT C -3' | Bergen et al. 1999 |
| | RPS4Y-R | 5'- CAG CAA CAG TAA GTC GAA TG -3' | Bergen et al. 1999 |
| M1 (YAP) | M1 -F | 5'- CAG GGG AAG ATA AAG AAA TA -3' | Hammer and Horai 1995 |
| | M1 -R | 5'- ACT GCT AAA AGG GGA TGG AT -3' | Hammer and Horai 1995 |
| M40 (SRY$_{4064}$) | M40 - F | 5'- GGT ATG ACA GGG GAT GAT GTG A -3' | Thomas et al. 1999 |
| | M40 -R | 5'- CCA CGC CCA GCT AAT TTT TTG -3' | Thomas et al. 1999 |
| M2 (sY81) | M2 -F | 5'- ATG GGA GAA GAA CGG AAG GA -3' | Thomas et al. 1999 |
| | M2- R | 5'- TGG AAA ATA CAG CTC CCC CT -3' | Thomas et al. 1999 |
| M191[2] | M191-F | 5' - GGA GGA GCA AGT ACA GCG AG-3' | A. Nebel (unpublished) |
| | M191-R | 5' - CAC ACC AAA ATA TCT CAT ATT TT**G** AT | A. Nebel (unpublished) |
| M35 | M35 -F | 5' - AGG GCA TGG TCC CTT TCT AT -3' | A. Nebel (unpublished) |
| | M35 -R | 5'- TGG GTT CAA GTT TCC CTG TC -3' | A. Nebel (unpublished) |
| M213 | M213 -F | 5'- TAT AAT CAA GTT ACC AAT TAC TGG C -3' | Underhill et al. 2000 |
| | M213 -R | 5'- TTT TGT AAC ATT GAA TGG CAA A -3' | Underhill et al. 2000 |
| M9 | M9 -F | 5'- GCA GCA TAT AAA ACT TTC AGG -3' | Underhill et al. 1997 |
| | M9 -R | 5'- AAA ACC TAA CTT TGC TCA AGC -3' | Underhill et al. 1997 |
| M74 | M74 -F | 5'- ATG CTA TAA TAA CTA GGT GTT GAA G -3' | Underhill et al. 2000 |
| | M74 -R | 5'- AAT TCA GCT TTT ACC ACT TCT GAA -3' | Underhill et al. 2000 |
| M207 | M207 -F | 5'- AGG AAA AAT CAG AAG TAT CCC TG -3' | Underhill et al 2000 |
| | M207-R | 5'- CAA AAT TCA CCA AGA ATC CTT G -3' | Underhill et al. 2000 |
| DYS19[1] | DYS19-A | 5'-**FAM**-CTA CTG AGT TTC TGT TAT AGT-3' | Roewer et al. 1992 |
| | DYS19-B | 5'-ATG GCA TGT AGT GAG GAC A-3' | Roewer et al. 1992 |
| DYS388[1] | DYS388-A | 5'-**FAM**-GTG AGT TAG CCG TTT AGC GA-3' | Jobling and Tyler-Smith 1995 |
| | DYS388-B | 5'-CAG ATC GCA ACC ACT GCG-3' | Jobling and Tyler-Smith 1995 |
| DYS389[1] | DYS389–A | 5'- **FAM** - CCA ACT CTC ATC TGT ATT ATC TAT -3' | Jobling and Tyler-Smith 1995 |
| | DYS389-B | 5'- TCT TAT CTC CAC CCA CCA GA -3' | Jobling and Tyler-Smith 1995 |
| DYS390[1] | DYS390-A | 5'-**HEX**-TAT ATT TTA CAC ATT TTT GGG CC-3' | Jobling and Tyler-Smith 1995 |
| | DYS390-B | 5'-TGA CAG TAA AAT GAA CAC ATT GC-3' | Jobling and Tyler-Smith 1995 |
| DYS391[1] | DYS391-A | 5'-**HEX**-CTA TTC ATT CAA TCA TAC ACC CA-3' | Jobling and Tyler-Smith 1995 |
| | DYS391-B | 5'-GAT TCT TTG TGG TGG GTC TG-3' | Jobling and Tyler-Smith 1995 |
| DYS392[1] | DYS392-A | 5'-**TAMRA**-TCA TTA ATC TAG CTT TTA AAA ACA A-3' | Jobling and Tyler-Smith 1995 |
| | DYS392-B | 5'-AGA CCC AGT TGA TGC AAT GT-3' | Jobling and Tyler-Smith 1995 |
| DYS393[1] | DYS393-A | 5'-**HEX**-GTG GTC TTC TAC TTG TGT CAA TAC-3' | Jobling and Tyler-Smith 1995 |
| | DYS393 -B | 5'-AAC TCA AGT CCA AAA AAT GAG G-3' | Jobling and Tyler-Smith 1995 |

[1] FAM, HEX and TAMRA are fluorescent dyes which fluoresce as blue, green and yellow respectively using Filter set A on the ABI377 Prism DNA Sequencer

[2] The underlined G in primer M191R represents the site of a sequence mismatch.

**Table 2.5** PCR ingredients and cycling conditions used for amplification of Y chromosome biallelic markers, and RFLP detection of polymorphisms Final concentrations of ingredients are shown.

| **Y chromosome polymorphism** | **M42** | **SRY$_{10831}$** | **M168** | **M150** | **M112** | **M211** | **M130 (RPS4Y)** | **M1 (YAP)** |
|---|---|---|---|---|---|---|---|---|
| **Mutation (ancestral-derived)** | A-T | A-G | C-T | C-T | G-A | C-T | C-T | absence - presence |
| **PCR ingredients** | | | | | | | | |
| buffer (includes 1 5mM $MgCl_2$) | 1X | 1X | 1X | 1X | 1X | 1X | 1X | 1X |
| $MgCl_2$ | 2 mM | 1.5 mM | 2.5 mM | 1.5 mM | 2 mM | 2 mM | 1.5 mM | 1 5 mM |
| forward primer | 0.4 µM | 0.4 µM | 0.4 µM | 0.4 µM | 0.4 µM | 0.4 µM | 0.4 µM | 0 4 µM |
| reverse primer | 0.4 µM | 0.4 µM | 0.4 µM | 0.4 µM | 0.4 µM | 0 4 µM | 0.4 µM | 0 4 µM |
| BSA | - | - | - | - | - | - | - | 1 µg/µl |
| dNTPs | 0.1mM | 0.1mM | 0.1mM | 0.1mM | 0 1mM | 0.1mM | 0.1mM | 0 1mM |
| *Taq* DNA polymerase | 0.5U | 0.5U | 0.5U | 0.5U | 0 5U | 0.5U | 0.5U | 0 5U |
| volume used | 25ul | 10ul | 25ul | 25ul | 50ul | 25ul | 25ul | 25ul |
| **PCR cycling conditions** | | | | | | | | |
| initial denaturation temperature (°C), time | 95, 10 min | 94, 5 min | 95, 10 min | 94, 5 min | 94, 5 min | 94, 5 min | 95, 10 min | 94, 2 min |
| denaturation temperature (°C), time | 94, 30s | 94, 45s | 94, 30s | 94, 45s | 94, 45s | 94, 45s | 94, 30s | 94, 1min |
| annealing temperature (°C), time | 58, 1min | 60, 45s | 56, 1 min | 60, 45s | 61, 45s | 58, 45s | 50, 30s | 51, 1 min |
| extension temperature (°C), time | 72, 1 min | 72, 45s | 72, 1 min | 72, 45s | 72, 45s | 72, 45s | 72, 45s | 72, 1 min |
| number of cycles | 35 | 35 | 35 | 35 | 35 | 35 | 30 | 30 |
| final extension temperature (°C), time | 72, 10 min | 72, 5 min | 72, 10 min | 72, 5 min | 72, 5 min | 72, 5 min | 72, 7 min | |
| **RFLP analysis** | | | | | | | | |
| PCR product size (bp) | 340 | 167 | 473 | 167 | 227 | 208 | 91 | 150 **(YAP-)** / 450 **(YAP+)** |
| restriction enzyme | *Alu* I | *Dra* III + BSA | *Hinf* I | *Aat* II | *TspR* I + BSA | *Rsa* I | *Bsl* I | - |
| restriction enzyme manufacturer | New England Biolabs | New England Biolabs | New England Biolabs | New England Biolabs | New England Biolabs | Roche | New England Biolabs | - |
| digestion conditions (°C) overnight | 37 | 37 | 37 | 37 | 65 | 37 | 55 | - |
| gel electrophoresis | 3% NuSieve | 3% agarose | 3% NuSieve | 3% NuSieve | 2% agarose | 2% agarose | 3% agarose | 2% agarose |
| ancestral allele - product sizes (bp) | 208 + 110 +22 **(A)** | 167 **(A)** | 234 + 105 + 81 + 52 **(C )** | 146 + 21 **(C)** | 155 + 72 **(G)** | 208 **(C)** | 57 + 34 **(C)** | 150 **(YAP-)** |
| derived allele - product sizes (bp) | 208 + 87 + 23 + 22 **(T)** | 112 + 55 **(G)** | 234 + 186 + 52 **(T)** | 167 **(T)** | 227 **(A)** | 137 + 71 **(T)** | 91 **(T)** | 450 **(YAP+)** |
| **Reference for assay** | | | | | | | | |
| | unpublished | Santos et al 1999 | unpublished | unpublished<br>A. Nebel, personal comm | unpublished<br>A Nebel, personal comm. | unpublished<br>A Nebel, personal comm. | Kayser et al 2000 | Hammer and Horai 1995 |

**Table 2.5 continued**

| **Marker** | **M40 (SRY$_{4064}$)** | **M2 (sY81)** | **M191** | **M35** | **M213** | **M9** | **M74** | **M207** |
|---|---|---|---|---|---|---|---|---|
| **Mutation** | G-A | A-G | T-G | G-C | T-C | C-G | G-A | A-G |
| **PCR ingredients** | | | | | | | | |
| buffer (includes 1 5mM $MgCl_2$) | 1X | 1X | 1X | 1X | 1X | 1X | 1X | 1X |
| $MgCl_2$ | 1 mM | 1.5 mM | 2 5 mM | 2 mM | 2 5 mM | 1.5 mM | 1 5 mM | 1.5 mM |
| forward primer | 0.15 uM | 0.2 uM | 0 3 uM | 0.4 uM | 0.4 uM | 0.2 uM | 0 3 uM | 0 4 uM |
| reverse primer | 0.15 uM | 0 2 uM | 0 3 uM | 0.4 uM | 0 4 uM | 0.2 uM | 0 3 uM | 0 4 uM |
| BSA | - | - | - | - | - | - | - | - |
| dNTPs | 0 1mM | 0 1mM | 0 1mM | 0.1mM | 0 1mM | 0.1mM | 0.1mM | 0 1mM |
| *Taq* DNA polymerase | 1U | 1U | 1U | 1U | 1U | 1U | 1U | 1U |
| volume used | 25ul | 25ul | 25ul | 25ul | 25ul | 25ul | 10ul | 25ul |
| **PCR cycling conditions** | | | | | | | | |
| initial denaturation temperature (°C), time | 94, 5 min | 94, 5 min | 94, 5 min | 95, 10 min | 95, 10 min | 94, 2 min | 94, 5 min | 95, 10 min |
| denaturation temperature (°C), time | 94, 45s | 94, 45s | 94, 45s | 94, 30s | 94, 30s | 94, 45s | 94, 45s | 94, 30s |
| annealing temperature (°C), time | 58, 45s | 58, 45s | 60, 45s | 58, 1 min | 56, 1 min | 54, 45s | 60, 45s | 56, 1 min |
| extension temperature (°C), time | 72, 45s | 72, 45s | 72, 45s | 72, 1 min | 72, 1 min | 72, 45s | 72, 45s | 72, 1 min |
| number of cycles | 35 | 35 | 35 | 35 | 35 | 30 | 35 | 35 |
| final extension temperature (°C), time | 72, 5 min | 72, 5 min | 72, 5 min | 72, 10 min | 72, 10 min | | 72, 5 min | 72, 10 min |
| **RFLP analysis** | | | | | | | | |
| PCR product size (bp) | 225 | 148 | 156 | 186 | 409 | 340 | 385 | 423 |
| restriction enzyme | *BsrB* I | *Nla* III + BSA | *Mbo* I | *Bsr* I | *Nla* III + BSA | *Hinf*I | *Rsa* I | *Dra* I |
| restriction enzyme manufacturer | New England Biolabs | New England Biolabs | New England Biolabs | New England Biolabs | New England Biolabs | New England Biolabs | Roche | Roche |
| digestion conditions (°C) overnight | 37 | 37 | 37 | 65 | 37 | 37 | 37 | 37 |
| gel electrophoresis | 3% agarose | 3% agarose | 3% agarose | 2% agarose | 2% agarose | 3% agarose | 3% agarose | 2% agarose |
| ancestral allele - product sizes (bp) | 135 + 90 **(G)** | 105 + 43 **(A)** | 156 **(T)** | 122 + 64 **(G)** | 290 + 119 **(T)** | 181 + 95 + 64 **(C )** | 385 **(G)** | 356 + 77 **(A)** |
| derived allele - product sizes (bp) | 225 **(A)** | 148 **(G)** | 129 + 27 **(G)** | 186 **(C)** | 409 **(C )** | 245 + 95 **(G)** | 195 + 190 **(A)** | 423 **(G)** |
| **Reference for assay** | | | | | | | | |
| | Thomas et al 1999 | Thomas et al 1999 | unpublished<br>A. Nebel, personal comm | unpublished | unpublished | unpublished | unpublished | unpublished |

**Table 2.6** PCR ingredients and cycling conditions used for amplification of Y chromosome STR multiplexes. Final concentrations of ingredients are shown.

| **multiplex** | **Y-STR multiplex 1** | **Y-STR multiplex 2** |
|---|---|---|
| **markers** | DYS19, DYS390, DYS391, DYS393 | DYS388, DYS389 (I and II), DYS392 |
| **PCR ingredients** | | |
| buffer | 1X (no $MgCl_2$ included) | 1X (1.5mM $MgCl_2$ included) |
| $MgCl_2$ | 1.5mM | 1mM |
| each forward primer | 0.16μM | 0.16μM |
| each reverse primer | 0.16μM | 0.16μM |
| Spermadine | 0.25mM | 0.25mM |
| dNTPs | 0.2mM | 0.25mM |
| *Taq* DNA polymerase | 1U | 1U |
| volume used | 25μl | 25μl |
| **PCR cycling conditions** | | |
| initial denaturation temperature (°C), time | 95, 2 min | 95, 2 min |
| denaturation temperature (°C), time | 95, 30s | 95, 30s |
| annealing temperature (°C), time | 54, 30s | 56, 30s |
| extension temperature (°C), time | 72, 30s | 72, 30s |
| number of cycles | 28 | 28 |

### 2.3.3 *Y chromosome statistical analyses*

Intra-population Y chromosome variation was inferred from HG-STR data using the programme ARLEQUIN v2.0 (Schneider et al. 2000) to estimate genetic diversity (*h*) and its sampling variance (*v*) described by Nei (1987), and to estimate the mean number of pairwise differences (*mpd*) within each population.

The relationships among five central African populations were analysed using (a) Y chromosome HG data only, (b) Y chromosome STR ht data only, and (c) combined Y chromosome HG-STR ht data. $F_{ST}$ genetic distances between pairs of populations were calculated using (a) HG frequency data and (c) HG-STR ht frequency data using the programme ARLEQUIN v2.0, whilst for (b) STR ht data, $R_{ST}$ values (Slatkin 1995) were calculated instead of $F_{ST}$ values. $R_{ST}$ incorporates information regarding relationships among STR alleles as well as their frequencies. For all data types, population differentiation was examined using an exact test (Raymond and Rousset, 1995) implemented in the programme ARLEQUIN v2.0. The correlation among the resulting three distance matrices was tested using a Mantel test implemented in ARLEQUIN v2.0. The $F_{ST}$ distance matrix calculated from HG-STR data was used in the programme MEGA v2.1 (Kumar et al. 2001) to construct a population neighbour-joining (NJ) tree, which reflected the genetic affinities among the five central African populations in this study.

The genetic relationships among STR hts within HGs were assessed by means of networks constructed using NETWORK v2.0e (Bandelt et al. 1995; 1999). The reduced median (RM, r=1) and median-joining (MJ, ε=0) algorithms were applied sequentially to resolve patterns of variation generated by the fast-evolving STRs (Forster et al. 2000). The relationships among STR hts within HGs B and E were also assessed using NJ gene trees which were constructed using the Average Square Distance (ASD) genetic distance (Goldstein et al. 1995) and the MICROSAT programme.

Time to the most recent common ancestor (TMRCA) for HG E-M2 was estimated using the formula $S$=$u$t (Slatkin 1995; Thomas et al. 2000), where $S$ is mean STR repeat variance, $u$ is mutation rate and t is time in generations (assumed to be 20 years). Two different mutation rates ($u$) were used: $u$= 0.0028 (95% CI 0.0017 – 0.0043, Kayser et al. 2000), and $u$= 0.0018 (95% CI 0.00098 – 0.0031, Quintana-Murci et al. 2001).

HG data from the five populations in this study were compared to compatible published HG data from 15 other African populations (Table 2.2), and eight-locus STR ht data from this study were compared to published data from eight other African populations (Table 2.2). However because relatively few data utilising the same eight STR loci used in the present study were available, haplotypes using 5 STR loci only (excluding DYS388, DYS389I and DYS389II) from 13 other African populations were also examined (total N=1192, Table 2.2). The geographic distribution of Y chromosome HGs, and of the ten most frequent 5-locus STR hts in Africa were examined.

Using HG data from a total of 20 populations, $F_{ST}$ distances (Nei 1987) were calculated between pairs of populations using ARLEQUIN v2.0 (Schneider et al. 2000). Using 5-locus or 8-locus STR haplotype data from 13 and 18 populations respectively, $R_{ST}$ distances (Slatkin 1995) were calculated between pairs of populations using ARLEQUIN v2.0 (Schneider et al. 2000). Each of the three distance matrices was used to construct a PC plot using the programme NT-SYS PC (Rohlf 1997).

Inter-population $F_{ST}$ or $R_{ST}$ distances were used in analyses of molecular variance (AMOVA), implemented in ARLEQUIN v2.0. Populations were grouped according to (a) geographic origin, (b) a combination of ethnicity and geographic origin, (c) language affiliation, or (d) a combination of ethnicity and language affiliation. The distribution of variance among three hierarchical levels was tested in order to assess relationships among groups of populations.

A matrix of geographic distances among all populations was constructed using either the known point of sample collection or the capital of the country from which each population originated. Mantel tests of the correlation between population pairwise $F_{ST}$ or $R_{ST}$ distances and geographic distances were performed using ARLEQUIN v2.0 (Schneider et al. 2000).

## 2.4 *mtDNA molecular methods and analyses*

The mtDNA COII/tRNA$^{Lys}$ intergenic 9-bp deletion polymorphism was examined in 820 individuals from Uganda, Zambia and the C.A.R., whilst the mtDNA 3592 *Hpa*I restriction polymorphism was examined in 341 individuals. Both hypervariable regions of the mtDNA control region were sequenced in 397 samples.

### 2.4.1 *Screening for the mitochondrial 9-bp deletion*

The presence or absence of the mtDNA COII/tRNA$^{Lys}$ intergenic 9-bp deletion was detected using PCR primers H8297 and L8215 (Hertzberg et al. 1989; Table 2.7). The products were resolved by gel electrophoresis using 2% Metaphor agarose gels (run at ~5 V/cm for 2 hours in 1XTBE buffer) and sized with a 1kb ladder (Gibco BRL) as described by Redd et al. (1995, Table 2.8). Positive and negative controls were included in each gel run. Fragment sizes of 112 bp and 121 bp corresponded to samples with and without the deletion, respectively. A total of 820 individuals from four central African populations were screened for the 9-bp deletion (Table 4.1).

### 2.4.2 *Screening for the 3592 Hpa*I *polymorphism*

The 3592 *Hpa*I polymorphism was typed by PCR / RFLP analysis using primers 5F and 5R, and PCR conditions modified from Rieder et al. (1998) (Tables 2.7 and 2.8). The 831 bp PCR fragment was digested with *Hpa*I at 37°C overnight. The PCR products were then resolved by gel electrophoresis using 2% agarose gels (at ~5V/cm for 2 hours in 1XTBE buffer) and sized with a 1 kb ladder (Gibco BRL). Positive and negative controls were included in each gel run. If the *Hpa*I recognition site was present, the PCR product was digested into two fragments of 445 bp and 386 bp (Table 2.8). The presence of the 3592 *Hpa*I recognition site represented a T allele at np 3594. The loss of the 3592 *Hpa*I recognition site could have been due to any change at the 3594 (or surrounding sites), but was assumed to represent a C allele at np 3594. Only 341 of 398 samples were screened successfully for the 3592 *Hpa*I RFLP due to poor sample quality. The 3592 *Hpa*I polymorphism results were used together with sequence information from the 16390 site in HVRI to classify samples into the mtDNA HGs L1, L2 and L3.

2.4.3 *Sequencing of the two mtDNA hypervariable regions*

HVRI and HVRII sequencing began with an initial PCR-amplification of the entire 1.1 kb mtDNA control region using primers L15996 and H408 (Vigilant et al. 1989, Table 2.7) and the method initially described by Vigilant (1990, Table 2.8). The amplified 1.1 kb PCR products were resolved by electrophoresis on 2% NuSieve agarose gels in 1XTAE buffer (run at ~5V/cm for 1.5 hours). The 1.1kb bands were excised from the gel, and purified using either Wizard columns (Promega) or Nucleospin columns (Macherey-Nagel) according to the manufacturer's recommendations.

Purified PCR products were then used in cycle sequencing PCR amplification reactions (Table 2.8). HVRI and II were cycle- sequenced separately using Perkin Elmer / Applied Biosystems FS AmpliTaq or BigDye DNA Sequencing Kits. The L-strand was sequenced using primer L15996 for HVR1 and L29 for HVRII (Table 2.7). In some cases, the H-strand was also sequenced, using primer H16401 for HVR1 (in 63 samples) and primer H408 for HVRII (in 32 samples; Table 2.7).

The COII/tRNA$^{Lys}$ intergenic region was also sequenced in certain samples to confirm results of typing for the 9-bp deletion polymorphism. The sequencing protocols was as described above for sequencing of the control region, except that the initial amplification used primers H8297 and L8215, and primer L8215 was used for cycle sequencing (Table 2.7).

The fluorescent PCR products were purified using Centriseps Columns (Princeton Separations) or DyeEx Spin Columns (Quiagen) and were dried in a vacuum centrifuge for two hours. The dried PCR products were resuspended in 4μl dextran-formamide dye, and electrophoresed on 4.3% polyacrylamide gels in an ABI 377 DNA sequencer for seven hours. Sequences were analysed using the ABI Prism 377 Collection Software v2.1.

**Table 2.7** Sequences of primers used to amplify four regions of the mtDNA genome

| MtDNA polymorphism | Primer | Primer sequence | Reference |
|---|---|---|---|
| **PCR primers** | | | |
| 9-bp deletion | L8215 | 5'-ACA GTT TCA TGC CCA TCG TC-3' | Hertzberg et al. 1989 |
| | H8297 | 5'-ATG CTA AGT TAG CTT TAC AG-3' | Hertzberg et al. 1989 |
| 3592 *Hpa* I | 5F | 5'-TAC TTC ACA AAG CGC CTT CC-3' | Rieder et al. 1998 |
| | 5R | 5'-ATG AAG AAT AGG GCG AAG GG-3' | Rieder et al. 1998 |
| 10397 *Alu* I | 15F | 5'-TCT CCA TCT ATT GAT GAG GGT CT-3' | Rieder et al. 1998 |
| | 15R | 5'-AAT TAG GCT GTG GGT GGT TG-3' | Rieder et al. 1998 |
| Control region | L15996 | 5'-CTC CAC CAT TAG CAC CCA AGC-3' | Vigilant et al. 1989 |
| | H408 | 5'-CTG TTA AAA GTG CAT ACC GCC A-3' | Vigilant et al. 1989 |
| **Cycle sequencing primers** | | | |
| HVRI - forward | L15996 | 5'-CTC CAC CAT TAG CAC CCA AGC-3' | Vigilant et al. 1989 |
| HVRI - reverse | H16401 | 5'-TGA TTT CAC GGA GGA TGG TG-3' | Vigilant et al. 1989 |
| HVRII - forward | L29 | 5'-GGT CTA TCA CCC TCT TAA CCA C-3' | Vigilant et al. 1989 |
| HVRII - reverse | H408 | 5'-CTG TTA AAA GTG CAT ACC GCC A-3' | Vigilant et al. 1989 |

**Table 2.8** PCR ingredients and cycling conditions for amplification of mtDNA markers. Final concentrations of ingredients are shown.

| **mtDNA polymorphism** | **9-bp deletion** | **3592 *Hpa* I** | **10 397 *Alu* I** | **1.1 kb Control region** | **Cycle sequencing** |
|---|---|---|---|---|---|
| **Mutation (ancestral - derived)** | **2 copies - 1 copy** | **3594 T-C** | **10 400 C-T** | - | - |
| **PCR ingredients** | | | | | |
| buffer (incl 1.5mM $MgCl_2$) | 1X | 1X | 1X | 1X | |
| dNTPs | 0.25mM | 0.2mM | 0.2mM | 0.25mM | |
| forward primer | 0.4 μM | 0.4 μM | 0.4 μM | 0.4 μM | 0.33μM |
| reverse primer | 0.4 μM | 0.4 μM | 0.4 μM | 0.4 μM | 0.33μM |
| BSA | 1 mg/ml | | | 1 mg/ml | |
| gelatin | | 0.0001% | 0.0001% | | |
| *Taq* DNA polymerase | 1U | 1U | 1U | 1U | |
| Applied Biosystems FS or BigDye cycle sequencing kit | | | | | 4μl |
| volume used | 50μl | 25μl | 25μl | 50μl | 10μl |
| **PCR cycling condiitons** | | | | | |
| initial denaturation temperature (°C), time | | 94, 1 min | 94, 1 min | | |
| denaturation temperature (°C), time | 94, 1 min | 94, 30s | 94, 30s | 94, 1 min | 96, 30s |
| annealing temperature (°C), time | 56, 1 min | 61, 45s | 61, 45s | 56, 1 min | 50, 15s |
| extension temperature (°C), time | 74, 1 min | 72, 2 min | 72, 2 min | 74, 1 min | 60, 4 min |
| number of cycles | 30 | 35 | 35 | 30 | 25 |
| final extension temperature (°C), time | | 72, 5 min | 72, 5 min | | |
| **RFLP analysis** | | | | | |
| PCR product size (bp) | - | 831 | 892 | - | - |
| restriction enzyme | - | *Hpa* I | *Alu* I | - | - |
| restriction enzyme manufacturer | - | New England Biolabs | New England Biolabs | - | - |
| digestion conditions (°C) overnight | - | 37 | 37 | - | - |
| gel electrophoresis | 2% Metaphor | 2% agarose | 3% agarose | 2% Nusieve | - |
| ancestral allele - product sizes (bp) | 121 **(2 copies)** | 445 + 386 **(T)** | 267 + 366 + 259 **(C)** | 1.1kb | - |
| derived allele - product sizes (bp) | 112 **(1 copy)** | 831 **(C)** | 267 + 164 + 202 + 259 **(T)** | - | - |

Control region sequence data included 780 bases of sequence data from positions 15997 – 16400 and 31 – 407. These data incorporate and extend slightly beyond hypervariable region I (HVRI; positions 16024-16383) and hypervariable region II (HVRII; positions 57 – 372). HVRI and HVRII sequences were compared with the published reference sequence (Anderson et al. 1981, Andrews 1999) using the GELIN software (Sherry 1991). They were then aligned manually for comparison with each other and with published sequence data. Following suggestions made by Handt et al. (1998) and Burckhardt et al. (1999), alignment gaps were introduced at the following positions in HVRI: 16104.1, 16139.1, 16169.1, 16174.1-16174.2, 16183.1-16183.4, 16227.1, 16259.1, 16296.1, 16366.1 and 16386.1, and in HVRII at 56.1-56.2, 174.1, 190.1, 291.1-291.2, 294.1, 302.1-302.4 and 315.1 – 315.2.

### 2.4.4 *Screening for the 10397 Alu*I *polymorphism*

The 10397 *Alu*I polymorphism was typed in those samples which could not be classified into mtDNA subHGs on the basis of their control region sequence variation. This polymorphism was typed by PCR / RFLP analysis using primers 15F and 15R, and PCR conditions modified from Rieder et al. (1998) (Tables 2.7 and 2.8). The 892 bp PCR fragment was digested with *Alu*I at 37$^{O}$C overnight. The PCR products were then resolved by gel electrophoresis using 3% agarose gels (at ~5V/cm for 2 hours in 1XTBE buffer) and sized with a 1 kb ladder (Gibco BRL). Positive and negative controls were included in each gel run. If the *Alu*I recognition site was present, the PCR product was digested into four fragments of 267, 164, 202 and 259 bp, whilst in the absence of the *Alu*I recognition site, the PCR product was digested into three fragments of 267, 366 and 259 bp (Table 2.8). The presence of the 10397 *Alu*I recognition site represented a T allele at np 10 400.

### 2.4.5 *MtDNA statistical analyses*

The frequencies of the mtDNA 9-bp deletion polymorphism and of the 3592 *Hpa*I polymorphism were calculated by direct counting. The 3592 *Hpa*I polymorphism was used together with sequence information from the HVRI 16390 site to classify each sample within one of the three major African mtDNA HGs, L1, L2 and L3. Frequencies of these HGs were then calculated in the four central African populations. In some cases where the 3592 *Hpa*I RFLP was not typed, HG classifications were performed on the basis of HVRI and II sequence variation only.

Control region sequence variation and presence/absence of the mtDNA 9-bp deletion were used to identify unique mtDNA types and classify them into subHGs following the definitions suggested by Salas et al. (2002). SubHG definitions suggested by Salas et al. (2002) on the basis of HVRI variation only were extended to include HVRII variation using data from this study, as well as sequence data from other publications (Table 2.3). Associations between the mtDNA 9-bp deletion polymorphism and mtDNA subHGs were also examined.

Intra-population mitochondrial variation was inferred from HVRI and HVRII sequence data by using the programme ARLEQUIN v2.0 (Schneider et al. 2000) to estimate genetic diversity (*h*) and its sampling variance (*v*) described by Nei (1987).

Using (a) HVRI and HVRII sequence data, and (b) mtDNA subHG frequency data, ARLEQUIN v2.0 (Schneider et al. 2000) was used to calculate $F_{ST}$ genetic distances between pairs of central African populations. The correlation between the resulting two distance matrices was tested by means of a Mantel test implemented in ARLEQUIN v2.0. These distance matrices were used in the programme MEGA v2.1 (Kumar et al. 2001) to construct population neighbour-joining (NJ) trees in order to examine the genetic affinities among the four central African populations in this study. ARLEQUIN v2.0 was also used to examine population differentiation by means of exact tests of differentiation (Raymond and Rousset, 1995).

The genetic relationships among mtDNA types within HGs L1, L2 and L3 were assessed by means of networks constructed using NETWORK v2.0e (Bandelt et al. 1995; 1999). The reduced median (RM, r=2) and median-joining (MJ, ε=0) algorithms were applied sequentially to resolve patterns of variation generated in the fast-evolving mtDNA control region. The relationships among mtDNA types within HGs L1, L2 and L3 were also assessed using neighbour-joining (NJ) gene trees (Saitou and Nei 1987), which were constructed using the p-distance and the programme MEGA v2.1 (Kumar et al. 2001). Bootstrapping, based on 1000 replications, was applied to estimate the confidence of the branching patterns in NJ trees (Felsenstein 1975). Phylogenetic trees and networks were rooted using Neanderthal HVR1 and HVRII sequences (Krings et al. 1997; 1999).

Salas et al. (2002) estimated TMRCA for all mtDNA subHGs in African populations using HVRI data only. These dates were recalculated in this study using combined HVRI and HVRII sequence data from this study, and wherever possible, from published data (Table 2.3). Dates were estimated using the formula $t=d/2\mu$, where d is mean sequence divergence within subHGs and $\mu$ is the mutation rate. The program SENDBS (N. Takezaki) was used to calculate d as described in Nei and Jin (1989), and a mutation rate of 16.6%/million years (calibrated by Soodyall et al. (1996) for both hypervariable regions) was used.

MtDNA HVRI sequence data from this study were compared to published HVRI sequence data from 37 other African populations (Table 2.3). The geographic distribution of mtDNA subHGs in African populations was examined. $F_{ST}$ distances (Nei 1987) were calculated between pairs of populations using ARLEQUIN v2.0 (Schneider et al. 2000). A second set of $F_{ST}$ distances were calculated using frequencies of mtDNA subHGs in each population. Each distance matrix was used to construct a PC plot using the programme NT-SYS PC (Rohlf 1997).

Inter-population $F_{ST}$ distances based on HVRI sequence data, or based on mtDNA subHG frequency data, were used in analyses of molecular variance (AMOVA), implemented in ARLEQUIN v2.0. Populations were grouped according to (a) geographic origin, (b) a combination of ethnicity and geographic origin, (c) language affiliation, or (d) a combination of ethnicity and language affiliation. The distribution of variance among three hierarchical levels was tested in order to assess relationships among groups of populations.

A matrix of geographic distances among all populations was constructed using either the known point of sample collection, or the capital of the country from which each population originated (Appendices 8.4 and 8.7). Mantel tests of the correlation between population pairwise $F_{ST}$ distances and geographic distance were performed using ARLEQUIN v2.0 (Schneider et al. 2000).

## 2.5 *Estimation of migration rates using Y chromosome and mtDNA data*

Firstly, Y chromosome and mtDNA data were used to estimate migration rates in populations from Zambia, Uganda and the C.A.R. The D.R.C. population was excluded in this set of comparisons, to ensure that the same populations were used in each analysis. Secondly, the total number of populations available for each mtDNA or Y chromosome data type (Tables 2.2, 2.3) was used to estimate migration rates. Total $F_{ST}$ and $N\upsilon$ estimates, the apportionment of molecular diversity estimated in AMOVA analyses and regressions of genetic and geographic distances were compared between mtDNA and Y chromosome data types as follows.

For each data type, populations were treated as a single group, and total $F_{ST}$ was estimated using ARLEQUIN v2.0. Following the method of Seielstad et al. (1998), $F_{ST}$ values were then used to estimate migration rates using the formula $F_{ST}=1/1+N\upsilon$. In this equation, $\upsilon$ is actually the sum of the migration and mutation rates (Cavalli-Sforza and Bodmer 1971); however since the mutation rate of mtDNA and Y chromosome DNA are substantially lower than estimates of the migration rate, $\upsilon$ may be assumed to be equal to the migration rate (Seielstad et al. 1998). N is the effective population size.

AMOVA analyses were also used to calculate the percentages of variance within and among populations, and these figures were compared using the different data types. High estimates of the fraction of total variance *among* (between) populations suggests that populations are highly differentiated and therefore suggests lower migration rate, whilst low estimates of the fraction of total variance between populations suggests that populations are not very differentiated and therefore suggests higher migration rate. Alternatively, estimates of the fraction of total variance *within* populations may be considered: the higher the fraction of total variance within populations, the higher the migration rate.

For each data type, matrices of inter-population $F_{ST}$ values were compared to pairwise geographic distances as described above from correlation analyses. Regression lines were fitted to these data sets using Microsoft Excel. Slopes of the regression lines indicate how quickly genetic distance increases with geographic distance, another indicator of migration rate. The slopes of each pair of lines (from different data sets) were compared using paired t-tests.

# 3. Y CHROMOSOME STUDIES

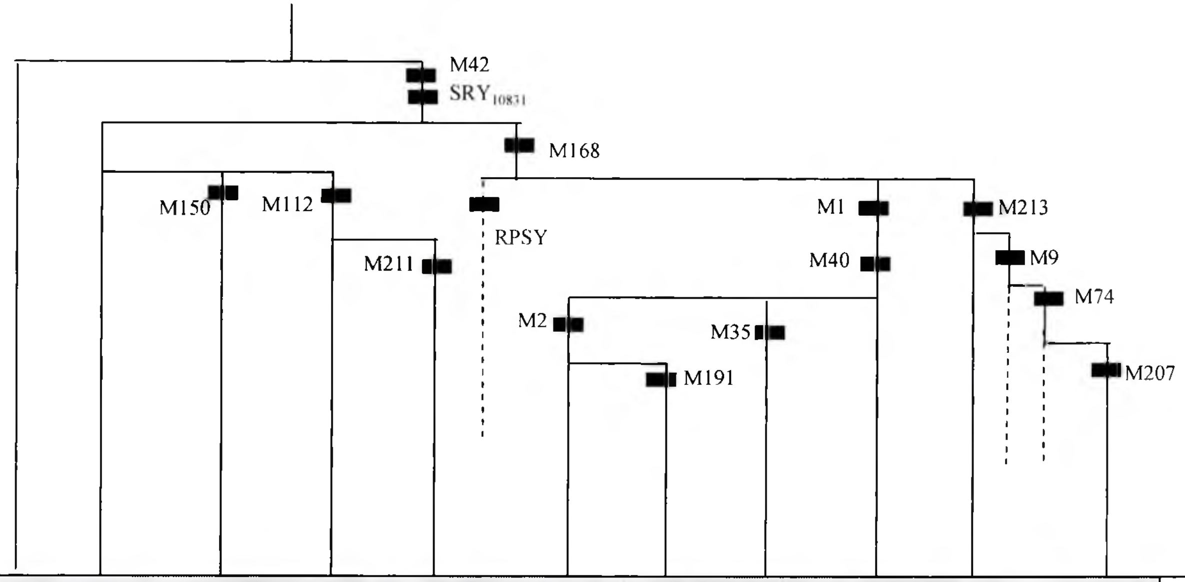
## 3.1 RESULTS

Sixteen biallelic and eight STR polymorphisms on the Y chromosome were used to assess the variation in 98 Ugandans, 87 Zambians, 40 Biaka Pygmies and 75 non-Pygmies from the Central African Republic, and 69 individuals from the D.R.C. These data were also used to examine the genetic affinities among the five central African populations, and among them and other African populations.

### 3.1.1 *Patterns of Y chromosome variation in central African populations*

#### 3.1.1.1 *HG data*

Using biallelic markers, eleven Y chromosome HGs were observed, including, for the first time in sub-Saharan Africa, HG R (fig. 3.1). HG E-M191 was the most frequent HG in all populations analysed (fig. 3.1), and was significantly more frequent in Uganda than in the pooled other four populations (Fisher's exact test, $P<0.005$). HG E-M2 was also frequent in all populations except the Biaka (fig. 3.1). HG B was significantly more frequent in the Biaka than in the other four populations (Fisher's exact test, $P<0.001$) and sublineages within HG B had distinct distributions: B-M150 was not found in the Biaka or C.A.R. samples, whilst B-M112 was frequent (0.300) in the Biaka (fig. 3.1). HG B-M211 was found in only 0.050 of our Biaka sample, compared to 0.200 of the sample of 20 Biaka examined by Cruciani et al. (2002). The remaining HGs A, FJ and R were present at frequencies of less than 0.10 in each population (fig. 3.1).



| population | N[3] | A | B* | B-M150 | B-M112 | B-M211 | E-M2 | E-M191 | E-M35 | E-M40 | FJ | R | *h*[1] | +/- |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| BIAKA | 40 |  |  |  | 0.300 | 0.050 | 0.075 | 0.500 | 0.025 | 0.050 |  |  | 0.960 | 0.014 |
| CAR | 75 | 0.053 | 0.013 |  | 0.040 |  | 0.347 | 0.440 | 0.040 |  |  | 0.067 | 0.971 | 0.009 |
| UGANDA | 98 |  |  | 0.010 |  |  | 0.265 | 0.612 | 0.082 | 0.020 |  | 0.010 | 0.971 | 0.009 |
| ZAMBIA | 87 | 0.011 |  | 0.057 |  |  | 0.414 | 0.425 |  | 0.069 | 0.011 | 0.011 | 0.983 | 0.007 |
| DRC | 69 | 0.014 | 0.014 | 0.072 |  |  | 0.333 | 0.464 | 0.014 | 0.043 | 0.014 | 0.029 | 0.980 | 0.008 |
| TOTAL | 369 | 0.016 | 0.005 | 0.030 | 0.041 | 0.005 | 0.309 | 0.493 | 0.035 | 0.035 | 0.005 | 0.024 | 0.984 | 0.002 |
| no. hts within HG | 174 | 5 | 2 | 8 | 10 | 1 | 41 | 75 | 13 | 11 | 2 | 6 |  |  |
| *h*[1] |  | 0.933 | n/a | 0.927 | 0.933 | n/a | 0.925 | 0.965 | 1 | 0.974 | n/a | 0.889 |  |  |
| *mpd*[2] |  | 8.000 | 7.000 | 2.255 | 6.933 | 0.000 | 2.512 | 3.158 | 7.795 | 3.205 | 6.000 | 2.667 |  |  |

[1] genetic diversity (Nei 1987), calculated from combined HG-STR data

[2] mean pairwise differences within HGs

[3] number of individuals

**Figure 3.1**. Phylogeny of 16 biallelic markers used (black bars), and the distribution of Y chromosome haplogroups (HGs) in five central African populations. Dotted lines indicate HGs that were not observed in the present study.

### 3.1.1.2 *HG-STR data*

Using biallelic data in conjunction with STR data resulted in the derivation of 174 hts (fig. 3.1, Table 3.1). Genetic diversity (*h*) calculated from these data was high (>0.96) and similar in all central African populations (fig. 3.1). Fourteen STR hts were observed in more than one HG i.e. were homoplasic (Table 3.1). All homoplasy occurred within HG E, with ten of the fourteen homoplasic types shared between HGs E-M2 and E-M191.

Of 174 hts found in the total sample, 142 hts (0.816) were not shared among the sampled populations (Table 3.1). Most of these "population-specific" hts were not very frequent and together accounted for only 174 of 369 Y chromosomes (0.472). The remaining 32 hts (0.184) were shared between two or more populations, and several were very frequent (Table 3.1); together these hts accounted for 195 of 369 Y chromosomes (0.528). Two hts within HG E-M2 (ht 61:15-12-13-17-21-10-11-13 and ht 67: 15-12-13-18-21-10-11-13), and two hts within HG E-M191 (ht 147: 17-12-13-17-21-10-11-14 and ht 157: 17-12-14-17-21-10-11-15) were observed at particularly high frequency, each representing approximately five percent of the total sample (Table 3.1). Together these four hts accounted for 0.214 of the total sample (N=369). Each of these hts was modal in at least one of the populations sampled: ht 61 was modal in the C.A.R. (non-Pygmy) sample, ht 67 was modal in the D.R.C. sample, ht 147 was modal in Zambians, and ht 157 was modal in both Ugandans and the Biaka (Table 3.1). The "Bantu Modal Haplotype" (BMH) defined by Thomas et al. (2000) on the basis of six STR loci (excluding DYS389I and II), corresponded to two different hts in this study, hts 61 and 67, which differed by a single step mutation at DYS389II.

Table 3.1 STR haplotypes (hts) within each Y chromosome haplogroup (HG), and their distribution among five central African populations
HG nomenclature is according to the YCC (2002)

| ht number | HG | DYS19 | DYS388 | DYS389I | DYS389II | DYS390 | DYS391 | DYS392 | DYS393 | total | CAP | CAR | ZAM | UGA | DRC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *1* | A | 14 | 12 | 13 | 17 | 20 | 10 | 14 | 14 | 1 |  | 1 |  |  |  |
| *2* | A | 15 | 10 | 13 | 19 | 24 | 10 | 11 | 13 | 1 |  |  | 1 |  |  |
| *3* | A | 15 | 11 | 12 | 16 | 21 | 10 | 12 | 13 | 1 |  |  |  |  | 1 |
| *4* | A | 15 | 12 | 14 | 16 | 21 | 10 | 13 | 11 | 2 |  | 2 |  |  |  |
| *5* | A | 16 | 11 | 14 | 17 | 21 | 10 | 11 | 13 | 1 |  | 1 |  |  |  |
| *6* | B-M112 | 14 | 9 | 13 | 15 | 23 | 9 | 11 | 12 | 2 |  | 2 |  |  |  |
| *7* | B-M112 | 15 | 12 | 11 | 17 | 24 | 11 | 11 | 14 | 3 | 3 |  |  |  |  |
| *8* | B-M112 | 15 | 12 | 11 | 17 | 25 | 11 | 11 | 14 | 1 | 1 |  |  |  |  |
| *9* | B-M112 | 15 | 12 | 11 | 17 | 25 | 11 | 11 | 15 | 1 | 1 |  |  |  |  |
| *10* | B-M112 | 15 | 12 | 11 | 17 | 25 | 12 | 11 | 14 | 1 | 1 |  |  |  |  |
| *11* | B-M112 | 15 | 12 | 12 | 17 | 24 | 11 | 11 | 14 | 1 | 1 |  |  |  |  |
| *12* | B-M112 | 15 | 12 | 13 | 16 | 25 | 8 | 11 | 12 | 1 | 1 |  |  |  |  |
| *13* | B-M112 | 15 | 12 | 14 | 14 | 24 | 10 | 11 | 12 | 1 |  | 1 |  |  |  |
| *14* | B-M112 | 15 | 12 | 14 | 16 | 24 | 10 | 11 | 13 | 1 | 1 |  |  |  |  |
| *15* | B-M112 | 17 | 12 | 12 | 16 | 23 | 11 | 11 | 14 | 3 | 3 |  |  |  |  |
| *16* | B-M211 | 15 | 12 | 11 | 16 | 24 | 9 | 11 | 12 | 2 | 2 |  |  |  |  |
| *17* | B-M150 | 15 | 10 | 13 | 19 | 23 | 10 | 11 | 13 | 1 |  |  |  |  | 1 |
| *18* | B-M150 | 15 | 10 | 14 | 18 | 23 | 10 | 11 | 13 | 1 |  |  | 1 |  |  |
| *19* | B-M150 | 15 | 10 | 14 | 18 | 24 | 10 | 11 | 13 | 3 |  |  | 2 | 1 |  |
| *20* | B-M150 | 15 | 10 | 14 | 19 | 24 | 11 | 11 | 13 | 1 |  |  |  |  | 1 |
| *21* | B-M150 | 16 | 10 | 14 | 18 | 23 | 10 | 11 | 13 | 2 |  |  | 1 |  | 1 |
| *22* | B-M150 | 16 | 10 | 14 | 18 | 24 | 10 | 11 | 12 | 1 |  |  |  |  | 1 |
| *23* | B-M150 | 16 | 10 | 14 | 18 | 24 | 10 | 11 | 13 | 1 |  |  |  |  | 1 |
| *24* | B-M150 | 17 | 10 | 14 | 18 | 24 | 11 | 11 | 13 | 1 |  |  | 1 |  |  |
| *25* | B* (default) | 16 | 12 | 12 | 18 | 23 | 11 | 11 | 13 | 1 |  |  |  |  | 1 |
| *26* | B* (default) | 16 | 13 | 12 | 15 | 22 | 10 | 11 | 14 | 1 |  | 1 |  |  |  |
| *27* | E-M35 | 11 | 12 | 13 | 17 | 24 | 10 | 12 | 13 | 1 |  | 1 |  |  |  |
| *28* | E-M35 | 11 | 12 | 13 | 19 | 24 | 10 | 12 | 13 | 1 |  | 1 |  |  |  |
| *29* | E-M35 | 11 | 12 | 14 | 17 | 23 | 11 | 11 | 14 | 1 |  |  |  | 1 |  |
| *30* | E-M35 | 13 | 12 | 12 | 18 | 25 | 11 | 11 | 13 | 1 |  |  |  | 1 |  |
| *31* | E-M35 | 13 | 12 | 13 | 17 | 23 | 11 | 11 | 13 | 1 |  |  |  | 1 |  |
| *32* | E-M35 | 13 | 12 | 13 | 17 | 24 | 11 | 11 | 14 | 1 |  |  |  | 1 |  |
| *33* | E-M35 | 13 | 12 | 14 | 17 | 24 | 10 | 11 | 13 | 1 |  |  |  | 1 |  |
| *34* | E-M35 | 13 | 13 | 14 | 17 | 24 | 11 | 11 | 14 | 1 |  |  |  | 1 |  |
| *35* | E-M35 | 15 | 12 | 13 | 18 | 21 | 11 | 11 | 13 | 1 | 1 |  |  |  |  |
| *36* | E-M35 | 15 | 13 | 12 | 18 | 24 | 12 | 11 | 12 | 1 |  |  |  | 1 |  |
| *37* | E-M35 | 15 | 14 | 13 | 19 | 25 | 10 | 11 | 13 | 1 |  |  |  | 1 |  |
| *38* | E-M35 | 16 | 12 | 14 | 17 | 21 | 10 | 11 | 15 | 1 |  | 1 |  |  |  |
| *39* | E-M35 | 17 | 12 | 13 | 17 | 21 | 10 | 11 | 15 | 1 |  |  |  |  | 1 |
| *40* | E-M40 | 14 | 12 | 12 | 16 | 24 | 10 | 11 | 13 | 2 |  |  | 1 |  | 1 |
| 41 | E-M40 | 14 | 12 | 12 | 16 | 24 | 11 | 11 | 13 | 1 |  |  | 1 |  |  |
| *42* | E-M40 | 14 | 12 | 12 | 16 | 25 | 10 | 11 | 13 | 1 |  |  |  | 1 |  |
| *43* | E-M40 | 14 | 12 | 12 | 16 | 25 | 11 | 11 | 13 | 1 |  |  | 1 |  |  |
| 44 | E-M40 | 14 | 12 | 12 | 17 | 24 | 11 | 11 | 13 | 1 |  |  | 1 |  |  |
| *45* | E-M40 | 14 | 12 | 12 | 17 | 25 | 10 | 11 | 13 | 1 |  |  | 1 |  |  |
| *46* | E-M40 | 14 | 12 | 12 | 18 | 25 | 10 | 11 | 13 | 1 |  |  | 1 |  |  |
| 47 | E-M40 | 14 | 12 | 13 | 16 | 24 | 10 | 11 | 13 | 1 | 1 |  |  |  |  |
| *48* | E-M40 | 15 | 12 | 12 | 16 | 25 | 11 | 11 | 13 | 2 | 1 |  |  |  | 1 |
| *49* | E-M40 | 15 | 12 | 13 | 18 | 21 | 10 | 11 | 13 | 1 |  |  |  |  | 1 |
| 50 | E-M40 | 15 | 13 | 12 | 17 | 24 | 10 | 11 | 13 | 1 |  |  |  | 1 |  |
| *51* | E-M2 | 14 | 12 | 12 | 16 | 25 | 10 | 11 | 13 | 1 |  |  |  |  | 1 |
| *52* | E-M2 | 14 | 12 | 12 | 18 | 21 | 10 | 11 | 13 | 1 |  | 1 |  |  |  |
| *53* | E-M2 | 14 | 12 | 13 | 17 | 21 | 10 | 11 | 12 | 1 |  |  | 1 |  |  |
| *54* | E-M2 | 14 | 12 | 13 | 18 | 21 | 10 | 11 | 13 | 1 |  |  |  | 1 |  |
| *55* | E-M2 | 15 | 12 | 12 | 16 | 21 | 10 | 11 | 13 | 1 |  |  |  | 1 |  |
| *56* | E-M2 | 15 | 12 | 12 | 17 | 21 | 10 | 11 | 13 | 2 |  |  | 1 | 1 |  |
| *57* | E-M2 | 15 | 12 | 12 | 17 | 21 | 10 | 11 | 14 | 1 |  | 1 |  |  |  |
| *58* | E-M2 | 15 | 12 | 12 | 17 | 21 | 11 | 11 | 14 | 1 |  | 1 |  |  |  |
| *59* | E-M2 | 15 | 12 | 13 | 16 | 21 | 11 | 11 | 14 | 1 |  |  |  | 1 |  |
| *60* | E-M2 | 15 | 12 | 13 | 17 | 19 | 10 | 11 | 13 | 1 |  |  |  | 1 |  |
| *61* | E-M2 | 15 | 12 | 13 | 17 | 21 | 10 | 11 | 13 | 20 | 1 | 9 | 3 | 5 | 2 |
| *62* | E-M2 | 15 | 12 | 13 | 17 | 21 | 10 | 11 | 14 | 7 |  | 1 | 2 | 3 | 1 |
| *63* | E-M2 | 15 | 12 | 13 | 17 | 21 | 11 | 10 | 14 | 1 |  |  |  |  | 1 |
| *64* | E-M2 | 15 | 12 | 13 | 17 | 21 | 11 | 11 | 13 | 5 |  |  | 3 |  | 2 |
| *65* | E-M2 | 15 | 12 | 13 | 17 | 21 | 11 | 11 | 14 | 1 |  |  |  | 1 |  |
| *66* | E-M2 | 15 | 12 | 13 | 17 | 21 | 11 | 11 | 15 | 1 |  |  | 1 |  |  |
| *67* | E-M2 | 15 | 12 | 13 | 18 | 21 | 10 | 11 | 13 | 20 |  | 6 | 4 | 3 | 7 |
| *68* | E-M2 | 15 | 12 | 13 | 18 | 21 | 11 | 11 | 13 | 10 | 2 |  | 3 |  | 5 |
| *69* | E-M2 | 15 | 12 | 13 | 18 | 21 | 11 | 11 | 14 | 1 |  |  | 1 |  |  |
| *70* | E-M2 | 15 | 12 | 13 | 18 | 22 | 9 | 11 | 13 | 1 |  |  | 1 |  |  |
| *71* | E-M2 | 15 | 12 | 13 | 18 | 22 | 10 | 11 | 13 | 2 |  | 2 |  |  |  |
| *72* | E-M2 | 15 | 12 | 13 | 19 | 21 | 10 | 11 | 13 | 3 |  |  | 3 |  |  |
| *73* | E-M2 | 15 | 12 | 13 | 19 | 21 | 11 | 11 | 14 | 1 |  |  | 1 |  |  |
| *74* | E-M2 | 15 | 12 | 14 | 17 | 21 | 10 | 11 | 13 | 2 |  |  | 1 |  | 1 |
| *75* | E-M2 | 15 | 12 | 14 | 17 | 21 | 10 | 11 | 14 | 1 |  |  | 1 |  |  |
| *76* | E-M2 | 15 | 12 | 14 | 17 | 21 | 10 | 11 | 15 | 1 |  |  |  | 1 |  |
| *77* | E-M2 | 15 | 12 | 14 | 18 | 21 | 10 | 11 | 13 | 3 |  | 1 | 1 | 1 |  |
| *78* | E-M2 | 15 | 12 | 14 | 18 | 22 | 10 | 11 | 13 | 1 |  | 1 |  |  |  |
| *79* | E-M2 | 15 | 13 | 13 | 18 | 21 | 9 | 11 | 13 | 1 |  |  | 1 |  |  |
| *80* | E-M2 | 16 | 12 | 12 | 16 | 21 | 10 | 11 | 13 | 1 |  |  | 1 |  |  |
| *81* | E-M2 | 16 | 12 | 12 | 19 | 21 | 10 | 11 | 14 | 1 |  |  |  |  | 1 |
| *82* | E-M2 | 16 | 12 | 13 | 17 | 21 | 10 | 11 | 13 | 2 |  |  | 2 |  |  |
| *83* | E-M2 | 16 | 12 | 13 | 17 | 21 | 10 | 11 | 14 | 5 |  | 1 | 2 | 1 | 1 |
| *84* | E-M2 | 16 | 12 | 13 | 18 | 21 | 10 | 11 | 13 | 2 |  |  | 1 | 1 |  |
| *85* | E-M2 | 16 | 12 | 13 | 18 | 21 | 10 | 11 | 15 | 1 |  |  | 1 |  |  |
| *86* | E-M2 | 16 | 12 | 13 | 18 | 21 | 11 | 11 | 14 | 2 |  | 2 |  |  |  |
| *87* | E-M2 | 16 | 12 | 13 | 19 | 21 | 10 | 11 | 13 | 3 |  |  |  | 3 |  |
| *88* | E-M2 | 16 | 12 | 13 | 19 | 21 | 10 | 11 | 14 | 1 |  |  |  | 1 |  |
| *89* | E-M2 | 16 | 12 | 14 | 16 | 21 | 10 | 12 | 14 | 1 |  |  |  |  | 1 |
| *90* | E-M2 | 16 | 13 | 14 | 18 | 21 | 10 | 11 | 13 | 1 |  |  | 1 |  |  |
| *91* | E-M2 | 17 | 12 | 14 | 18 | 21 | 10 | 11 | 14 | 1 |  |  |  | 1 |  |

**Table 3.1 continued**

| ht number | HG | DYS19 | DYS388 | DYS389I | DYS389II | DYS390 | DYS391 | DYS392 | DYS393 | total | CAP | CAR | ZAM | UGA | DRC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *92* | E-M191 | 12 | 12 | 13 | 17 | 21 | 10 | 11 | 15 | 1 | | 1 | | | |
| *93* | E-M191 | 14 | 12 | 13 | 17 | 22 | 10 | 11 | 13 | 1 | | | | | 1 |
| *94* | E-M191 | 14 | 12 | 14 | 18 | 21 | 10 | 11 | 16 | 1 | | | | | 1 |
| *95* | E-M191 | 15 | 12 | 12 | 17 | 21 | 10 | 11 | 14 | 1 | | | | | 1 |
| *96* | E-M191 | 15 | 12 | 12 | 17 | 21 | 10 | 11 | 15 | 2 | | | | | 2 |
| *97* | E-M191 | 15 | 12 | 12 | 17 | 22 | 10 | 11 | 14 | 1 | | 1 | | | |
| *98* | E-M191 | 15 | 12 | 13 | 16 | 21 | 11 | 11 | 14 | 1 | | | 1 | | |
| *99* | E-M191 | 15 | 12 | 13 | 17 | 21 | 10 | 11 | 14 | 2 | | | 1 | | 1 |
| *100* | E-M191 | 15 | 12 | 13 | 17 | 21 | 10 | 11 | 15 | 6 | 4 | 1 | 1 | | |
| *101* | E-M191 | 15 | 12 | 13 | 17 | 22 | 10 | 11 | 14 | 2 | | 2 | | | |
| *102* | E-M191 | 15 | 12 | 13 | 17 | 25 | 10 | 11 | 14 | 1 | 1 | | | | |
| *103* | E-M191 | 15 | 12 | 13 | 18 | 21 | 10 | 11 | 13 | 1 | | 1 | | | |
| *104* | E-M191 | 15 | 12 | 13 | 18 | 22 | 10 | 11 | 14 | 1 | | 1 | | | |
| *105* | E-M191 | 15 | 12 | 13 | 18 | 22 | 10 | 11 | 15 | 1 | | | | | 1 |
| *106* | E-M191 | 15 | 12 | 14 | 16 | 21 | 10 | 11 | 14 | 1 | | 1 | | | |
| *107* | E-M191 | 15 | 12 | 14 | 17 | 21 | 10 | 11 | 14 | 1 | | | | 1 | |
| *108* | E-M191 | 15 | 12 | 14 | 17 | 21 | 10 | 11 | 15 | 1 | | 1 | | | |
| *109* | E-M191 | 15 | 12 | 14 | 17 | 21 | 10 | 12 | 15 | 1 | | | 1 | | |
| *110* | E-M191 | 15 | 12 | 14 | 18 | 21 | 10 | 11 | 15 | 1 | | | | 1 | |
| *111* | E-M191 | 15 | 14 | 13 | 17 | 22 | 10 | 11 | 15 | 1 | | | | 1 | |
| *112* | E-M191 | 16 | 11 | 12 | 18 | 21 | 10 | 11 | 15 | 2 | | 2 | | | |
| *113* | E-M191 | 16 | 11 | 13 | 17 | 21 | 10 | 11 | 15 | 1 | | 1 | | | |
| *114* | E-M191 | 16 | 12 | 12 | 17 | 20 | 10 | 11 | 15 | 4 | 4 | | | | |
| *115* | E-M191 | 16 | 12 | 13 | 16 | 21 | 10 | 11 | 14 | 1 | | 1 | | | |
| *116* | E-M191 | 16 | 12 | 13 | 16 | 21 | 10 | 11 | 15 | 2 | 1 | | 1 | | |
| *117* | E-M191 | 16 | 12 | 13 | 16 | 21 | 11 | 11 | 15 | 1 | | | | 1 | |
| *118* | E-M191 | 16 | 12 | 13 | 17 | 21 | 10 | 10 | 14 | 2 | | | 1 | 1 | |
| *119* | E-M191 | 16 | 12 | 13 | 17 | 21 | 10 | 11 | 13 | 1 | | | | 1 | |
| *120* | E-M191 | 16 | 12 | 13 | 17 | 21 | 10 | 11 | 14 | 9 | | 3 | 3 | 1 | 2 |
| *121* | E-M191 | 16 | 12 | 13 | 17 | 21 | 10 | 11 | 15 | 8 | | 1 | | 4 | 3 |
| *122* | E-M191 | 16 | 12 | 13 | 17 | 21 | 10 | 11 | 16 | 1 | | | 1 | | |
| *123* | E-M191 | 16 | 12 | 13 | 17 | 21 | 11 | 11 | 14 | 2 | | | 2 | | |
| *124* | E-M191 | 16 | 12 | 13 | 17 | 21 | 11 | 11 | 15 | 4 | | | 2 | 1 | 1 |
| *125* | E-M191 | 16 | 12 | 13 | 17 | 22 | 10 | 11 | 15 | 3 | | 3 | | | |
| *126* | E-M191 | 16 | 12 | 13 | 18 | 21 | 10 | 11 | 14 | 2 | | | | 1 | 1 |
| *127* | E-M191 | 16 | 12 | 13 | 18 | 21 | 10 | 11 | 15 | 3 | | | 1 | | 2 |
| *128* | E-M191 | 16 | 12 | 13 | 18 | 21 | 10 | 12 | 14 | 1 | | | 1 | | |
| *129* | E-M191 | 16 | 12 | 13 | 18 | 21 | 11 | 11 | 15 | 1 | | | | 1 | |
| *130* | E-M191 | 16 | 12 | 13 | 18 | 22 | 10 | 11 | 15 | 6 | | 6 | | | |
| *131* | E-M191 | 16 | 12 | 13 | 19 | 21 | 10 | 11 | 14 | 2 | | | | 2 | |
| *132* | E-M191 | 16 | 12 | 14 | 15 | 21 | 10 | 11 | 14 | 1 | | | 1 | | |
| *133* | E-M191 | 16 | 12 | 14 | 16 | 21 | 10 | 11 | 14 | 1 | | | | | 1 |
| *134* | E-M191 | 16 | 12 | 14 | 16 | 21 | 10 | 11 | 15 | 1 | | | | | 1 |
| *135* | E-M191 | 16 | 12 | 14 | 17 | 21 | 10 | 11 | 14 | 6 | | 1 | 1 | 4 | |
| *136* | E-M191 | 16 | 12 | 14 | 17 | 21 | 10 | 11 | 15 | 6 | 2 | | 2 | 2 | |
| *137* | E-M191 | 16 | 12 | 14 | 17 | 21 | 11 | 11 | 14 | 1 | | | | 1 | |
| *138* | E-M191 | 16 | 12 | 14 | 18 | 21 | 10 | 11 | 14 | 1 | | | | | 1 |
| *139* | E-M191 | 17 | 10 | 14 | 18 | 21 | 10 | 11 | 14 | 1 | | | 1 | | |
| *140* | E-M191 | 17 | 12 | 12 | 16 | 21 | 10 | 11 | 15 | 1 | | 1 | | | |
| *141* | E-M191 | 17 | 12 | 12 | 17 | 21 | 10 | 11 | 14 | 1 | | 1 | | | |
| *142* | E-M191 | 17 | 12 | 12 | 17 | 21 | 10 | 11 | 15 | 1 | | | | 1 | |
| *143* | E-M191 | 17 | 12 | 13 | 16 | 21 | 10 | 11 | 15 | 3 | | | 1 | 1 | 1 |
| *144* | E-M191 | 17 | 12 | 13 | 16 | 21 | 10 | 11 | 16 | 1 | | | | | 1 |
| *145* | E-M191 | 17 | 12 | 13 | 17 | 21 | 10 | 8 | 14 | 1 | | | | 1 | |
| *146* | E-M191 | 17 | 12 | 13 | 17 | 21 | 10 | 10 | 15 | 1 | | | 1 | | |
| *147* | E-M191 | 17 | 12 | 13 | 17 | 21 | 10 | 11 | 14 | 17 | | | 9 | 4 | 4 |
| *148* | E-M191 | 17 | 12 | 13 | 17 | 21 | 10 | 11 | 15 | 4 | | 1 | | 2 | 1 |
| *149* | E-M191 | 17 | 12 | 13 | 17 | 21 | 10 | 11 | 16 | 1 | 1 | | | | |
| *150* | E-M191 | 17 | 12 | 13 | 17 | 21 | 11 | 11 | 15 | 2 | | 1 | | | 1 |
| *151* | E-M191 | 17 | 12 | 13 | 17 | 22 | 10 | 11 | 14 | 1 | | | 1 | | |
| *152* | E-M191 | 17 | 12 | 13 | 18 | 20 | 10 | 11 | 14 | 1 | 1 | | | | |
| *153* | E-M191 | 17 | 12 | 13 | 18 | 21 | 10 | 8 | 13 | 1 | | | | 1 | |
| *154* | E-M191 | 17 | 12 | 13 | 18 | 21 | 10 | 11 | 14 | 1 | 1 | | | | |
| *155* | E-M191 | 17 | 12 | 14 | 17 | 21 | 10 | 10 | 15 | 1 | | | | 1 | |
| *156* | E-M191 | 17 | 12 | 14 | 17 | 21 | 10 | 11 | 14 | 10 | | | 1 | 7 | 2 |
| *157* | E-M191 | 17 | 12 | 14 | 17 | 21 | 10 | 11 | 15 | 22 | 5 | 2 | | 13 | 2 |
| *158* | E-M191 | 17 | 12 | 14 | 17 | 21 | 10 | 12 | 14 | 1 | | | 1 | | |
| *159* | E-M191 | 17 | 12 | 14 | 17 | 21 | 11 | 11 | 14 | 2 | | | | 2 | |
| *160* | E-M191 | 17 | 12 | 14 | 17 | 22 | 10 | 11 | 14 | 1 | | | | | 1 |
| *161* | E-M191 | 17 | 12 | 14 | 18 | 21 | 10 | 11 | 15 | 1 | | | | 1 | |
| *162* | E-M191 | 17 | 12 | 14 | 18 | 21 | 11 | 11 | 14 | 1 | | | 1 | | |
| *163* | E-M191 | 17 | 12 | 14 | 20 | 21 | 10 | 11 | 14 | 1 | | | | 1 | |
| *164* | E-M191 | 17 | 12 | 15 | 17 | 21 | 10 | 11 | 14 | 1 | | | | 1 | |
| *165* | E-M191 | 17 | 12 | 15 | 17 | 21 | 10 | 11 | 15 | 1 | | | | 1 | |
| *166* | E-M191 | 18 | 12 | 13 | 16 | 21 | 10 | 11 | 14 | 1 | | | 1 | | |
| *167* | FJ | 15 | 12 | 14 | 16 | 24 | 10 | 13 | 13 | 1 | | | | | 1 |
| *168* | FJ | 15 | 13 | 14 | 18 | 23 | 10 | 12 | 14 | 1 | | | 1 | | |
| *169* | R | 15 | 12 | 13 | 15 | 24 | 11 | 13 | 12 | 1 | | | 1 | | |
| *170* | R | 15 | 12 | 13 | 17 | 24 | 11 | 13 | 13 | 1 | | 1 | | | |
| *171* | R | 15 | 12 | 13 | 18 | 24 | 10 | 13 | 13 | 1 | | 1 | | | |
| *172* | R | 15 | 12 | 14 | 14 | 24 | 10 | 13 | 12 | 1 | | | | | 1 |
| *173* | R | 15 | 12 | 14 | 16 | 24 | 11 | 13 | 13 | 3 | | 1 | | 1 | 1 |
| *174* | R | 15 | 12 | 14 | 17 | 24 | 11 | 13 | 13 | 2 | | 2 | | | |
| | | | | | | | | **total** | | **369** | **40** | **75** | **87** | **98** | **69** |

Abbreviations:
CAP: Pygmies from Central African Republic
CAR: Central African Republic
DRC: Democratic Republic of Congo
UGA: Uganda
ZAM: Zambia

### 3.1.1.3 *Haplogroup structure*

The genetic relationships among STR hts within HGs B, E and R were assessed by means of phylogenetic networks and NJ trees.

*HG B*

In a network of hts within HG B, hts from lineage B-M150 were clearly differentiated from hts from lineage B-M112 (fig. 3.2). This differentiation was also observed in a NJ tree of HG B hts (fig. 3.3). The single M211 haplotype (ht 16, found in two individuals, fig. 3.2) clustered with M112 hts as expected since M211 is derived from M112. HG B-M150 was characterised by alleles of 9 and 10 repeats at the DYS388 locus, whilst HG B-M112 was partially characterised by the DYS389I-11 allele (Table 3.1). TMRCA of HG B-M112 was estimated to be six to ten times older than TMRCA of HG B-M150. The estimated TMRCA of HG B-M112 was between 5 935 years ago (using *u*=0.0028, 95% CI 3 865- 9 775 years) and 9 232 years (using *u*=0.0018, 95% CI 5 361 – 16 957 years), whilst the TMRCA of HG B-M150 was estimated to be between 1 047 (using *u*=0.0028, 95% CI 682 – 1 725 years) and 1 629 years ago (using *u*=0.0018, 95% CI 946 – 2 992 years). Taking both B-M112 and B-M150 into account, the TMRCA of HG B was estimated to have existed between 6 395 (using *u*=0.0028, 95% CI 4 164 – 10 534 years) and 9 948 years ago (using *u*=0.0018, 95% CI 5 777 - 18 273 years).

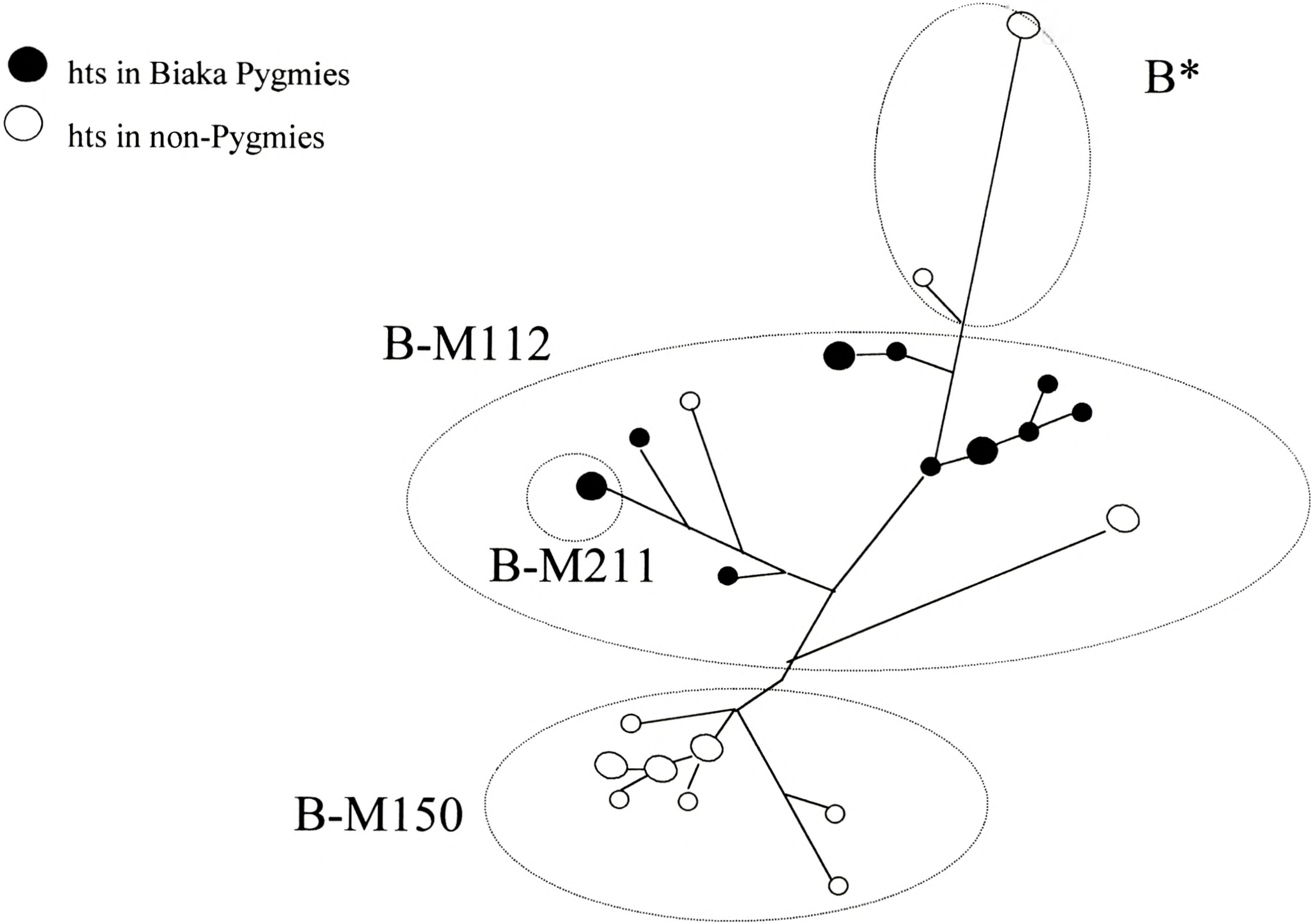


**Figure 3.2**. MJ network depicting the relationships among haplotypes within HG B. Black circles indicate haplotypes found in Pygmies from the C.A.R., and white circles indicate haplotypes found in non-Pygmies; circle size is proportional to haplotype frequency. Dotted lines indicate HG affiliation.

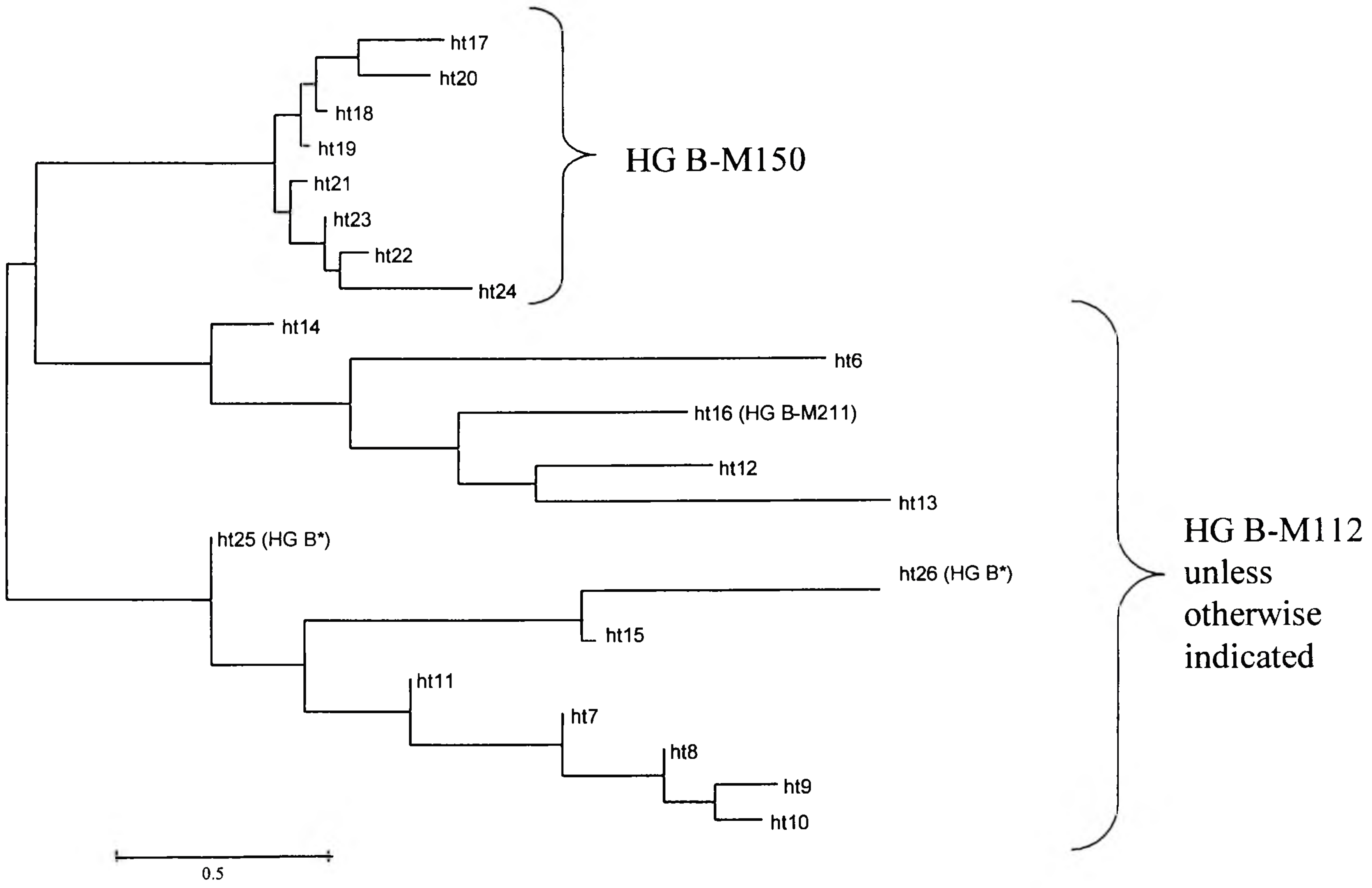


**Figure 3.3**. NJ tree constructed using ASD distance, depicting the relationships among haplotypes within HG B.

*HG E*

Four lineages within HG E were delineated in this study: HG E-M2, HG E-M191, HG E-M35, and HG E-M40 (fig. 3.1). HGs E-M40 and E-M35 were not common in this study (frequencies <0.07, Table 3.1). Both of these HGs incorporate several sublineages (YCC 2002), and hts within HGs E-M40 and E-M35 did not appear to be closely related to each other in phylogenetic networks (not shown). HG E-M35 also had a high mean number of pairwise differences (mpd), indicating that hts within it were not very closely related (fig. 3.1).

HG E-M35 was found at significantly higher frequency in Ugandans than in other sampled populations (Fisher's exact test, *P*<0.05). The E-M35 hts from this study (N=13) were compared to 34 E-M35 hts from north Africans (Bosch et al. 2001). No hts were shared between central African and north African populations. A MJ network (fig. 3.4) revealed that hts from central African populations either formed their own clusters, or were placed with the few hts from north Africa characterised by M78; they were not located within the major north African cluster characterised by M81.

HGs E-M2 and E-M191 were observed at high frequencies in central African populations in this study (fig. 3.1). Many of the hts from HGs E-M2 and E-M191 were not easily distinguished from each other; these two HGs exhibited considerable homoplasy and overlap of their haplotype distributions (Table 3.1). In a network of 106 hts (N=296) from HGs E-M2 and E-M191 (fig. 3.5), several branches included hts from both HG E-M2 and E-M191. Similarly, in a NJ tree (fig. 3.6), many of the observed clusters included hts from both HGs. It appears that much of the evolution of the two HGs has been along parallel or convergent lines (fig. 3.5, fig 3.6). An examination of the allele distributions for HGs E-M2 and E-M191 revealed that there was no significant difference in allele distribution between the two HGs at three STR loci: DYS388, DYS390 and DYS392 (exact test of population differentiation, *P*>0.001).

A very recent shared common ancestry of haplogroups E-M2 and E-M191 was confirmed by the similar estimated times to the most recent common ancestor (TMRCA) for each HG:

TMRCA of HG E-M2 was estimated to between 1 431 years ago (using *u*=0.0028, 95% CI 932 - 2 357 years) and 2 226 years (using *u*=0.0018, 95% CI 1 292 – 4 088 years), whilst TMRCA of HG E-M191 was estimated to be between 1 953 (using *u*=0.0028, 95% CI 1 272 – 3 216 years) and 3 038 years ago (using *u*=0.0018, 95% CI 1 764 – 5 579 years). Note that whilst estimated dates for HG E-M191 are slightly older than those estimated for HG E-M2, it is known that HG E-M191 is derived from E-M2 (Underhill et al. 2001). The TMRCA of HG E-M2* (incorporating both HG E-M2 and E-M191) was estimated to be between 2 407 (using *u*=0.0028, 95% CI 1 567 – 3 965 years) and 3 744 years ago (using *u*=0.0018, 95% CI 2 174 – 6 877 years).

Despite the close relationship between HGs E-M2 and E-M191, some differences between haplotypes from the two HGs were apparent (figs. 3.5, 3.6). Hts from HG E-M2 radiating from types 67 and 61 in the network (fig. 3.5) also clustered together near the base of the NJ tree (fig. 3.6), and were distinct from clusters of hts from HG E-M191 in the NJ tree (fig. 3.6) and network (fig. 3.5). Hts from HG E-M2 had significantly higher frequencies of the DYS19-15 and DYS393-13 alleles (Fisher's exact tests, *P*<0.0001), whilst hts from HG E-M191 had significantly higher frequencies of the DYS19-16, DYS19-17, and DYS393-15 alleles (Fisher's exact tests, *P*<0.0001).

Hts from all five central African populations were dispersed throughout the HG E-M2* network and no population-specific clustering was observed. The star-like shape of the network and the central placement of high-frequency hts (fig. 3.5) both provide strong evidence of founder effects within HGs E-M2 and E-M191. Ht157 (modal in Ugandans and Biaka) was placed in a peripheral position in the network (fig. 3.5), which is unusual for high-frequency hts. This ht may be associated with a very recent demographic expansion; this hypothesis is further supported by the limited number of hts derived from ht157 (fig. 3.5). The four very frequent hts (61, 67, 147, 157) could not be linked by single step mutations to each other, and there were at least six mutational steps between ht67 and ht157, the two most divergent pair of hts of the four (fig. 3.5).

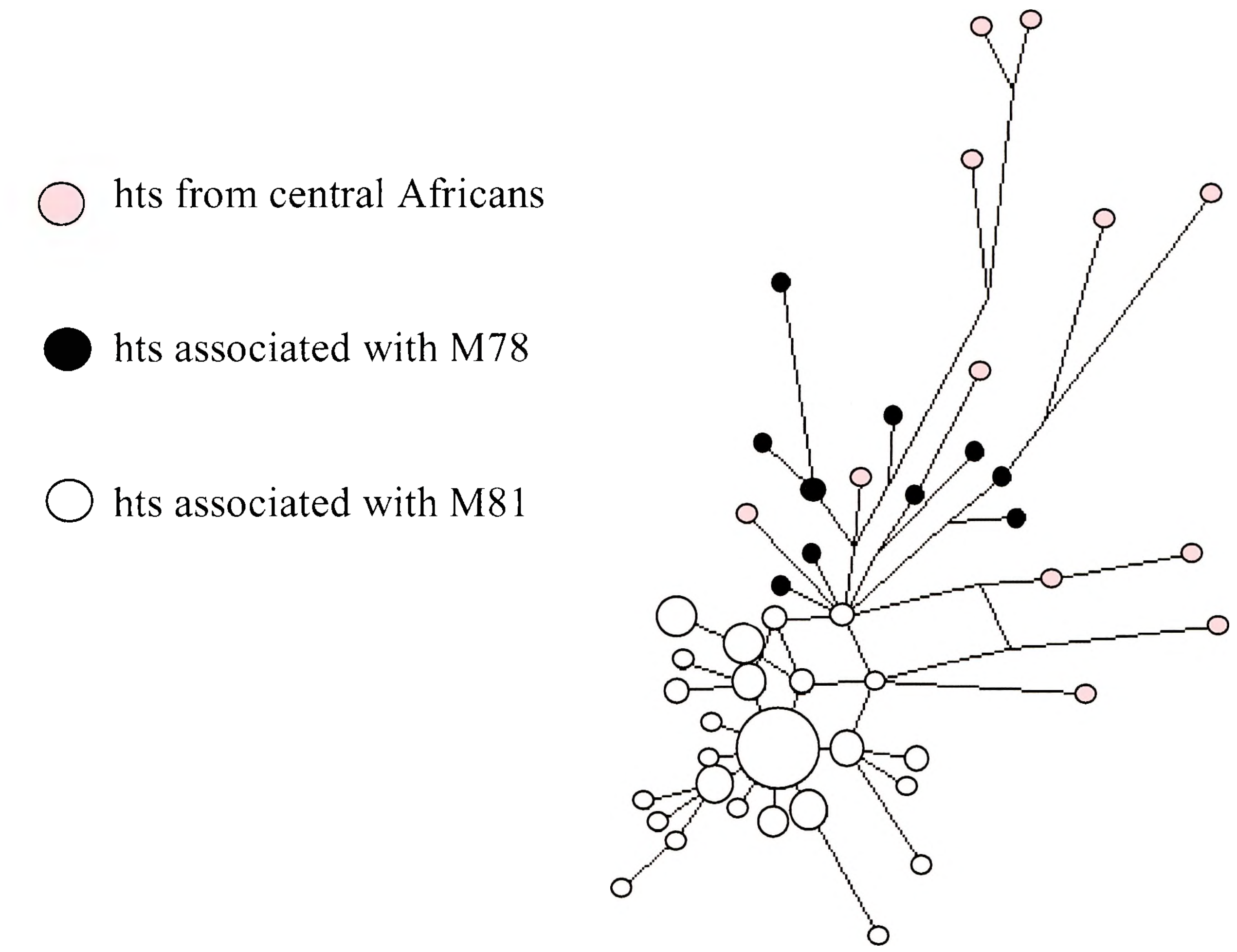


**Figure 3.4** MJ network depicting the relationships among haplotypes within HG E-M35. Red circles indicate haplotypes from central African populations, white circles indicate haplotypes from north Africa associated with M81 (Bosch et al. 2001) and black circles indicate haplotypes from north Africa associated with M78 (Bosch et al. 2001). Circle size is proportional to haplotype frequency.

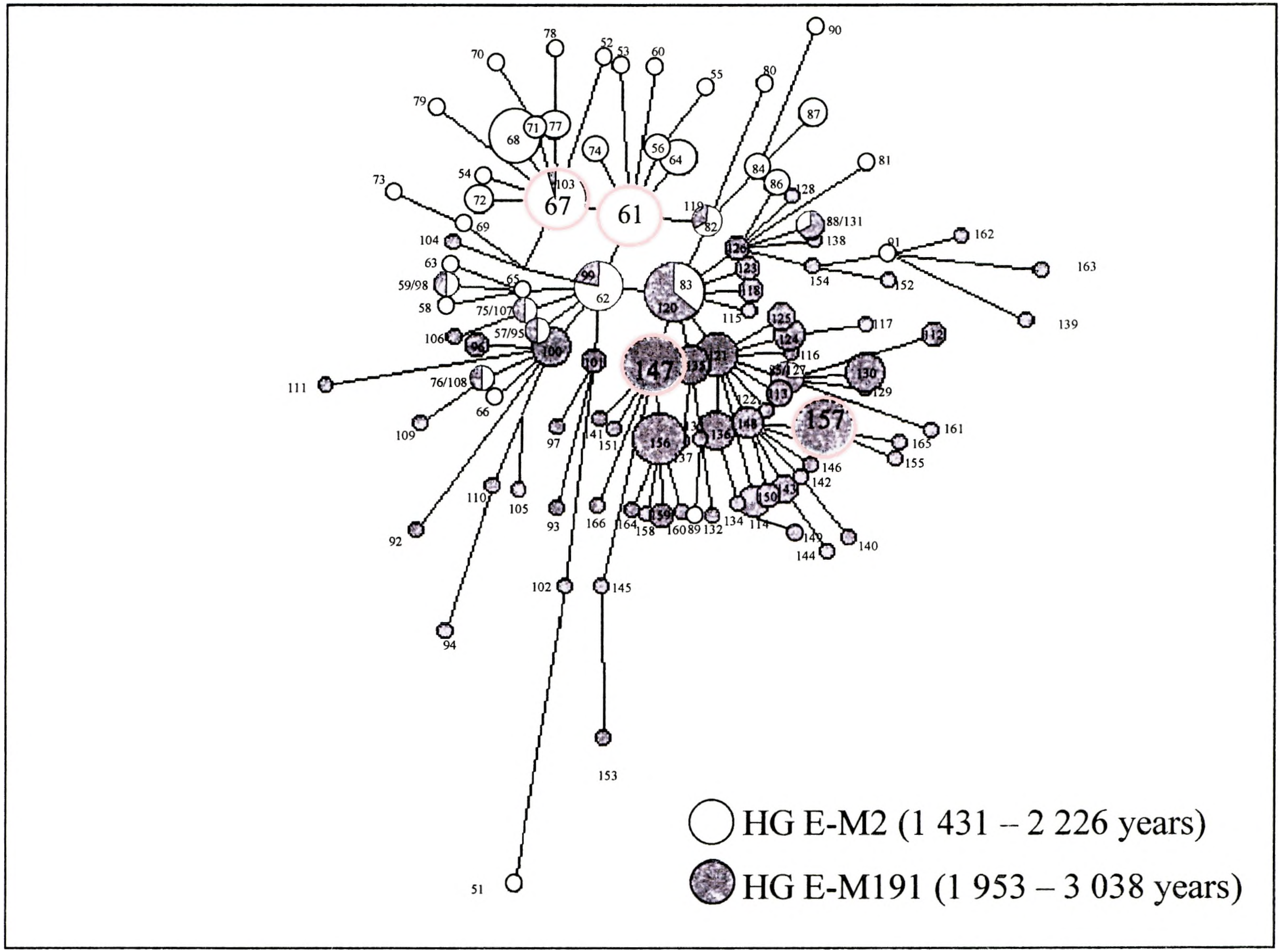


**Figure 3.5**. MJ network depicting the relationships among hts within HG E-M2*. White circles indicate haplotypes from HG E-M2 and gray circles indicate haplotypes from HG E-M191; circle size is proportional to ht frequency. Numbers represent ht numbers as listed in Table 3.1. The four very frequent hts are ringed in red. Estimated dates of TMRCA of each HG are based on mutation rates of 2.8 mutations per 1000 years (Kayser et al. 2000) and 1.8 mutations per 1000 years (Quintana-Murci et al. 2001).

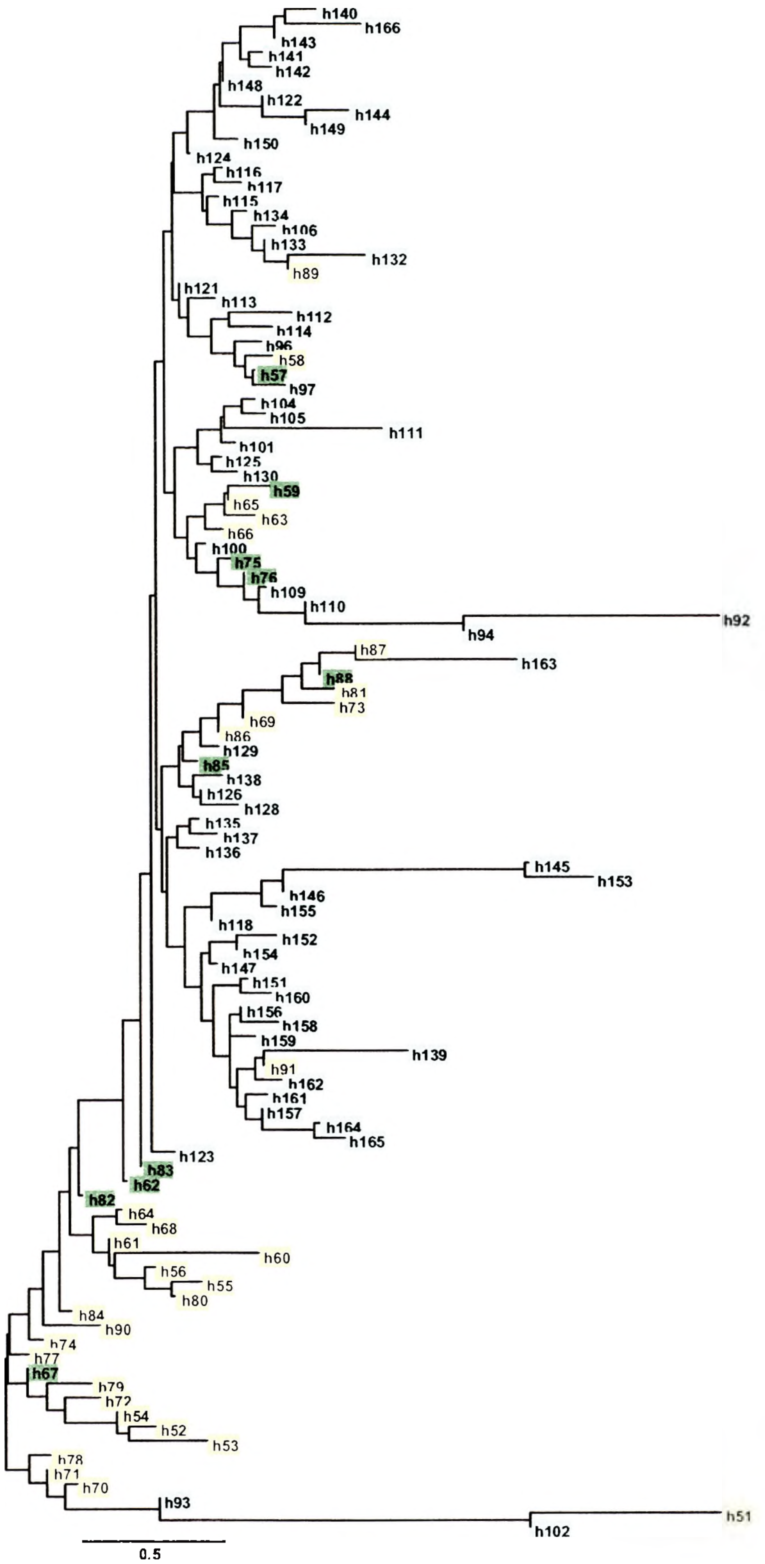


**Figure 3.6** NJ tree constructed using ASD distance, depicting the relationships among hts within HG E-M2*. Ht numbers are as listed in Table 3.1.Blue labels indicate hts associated with HG E-M191; yellow labels indicate hts associated with HG E-M2; green labels indicate hts associated with both HG E-M2 and HG E-M191 (homoplasic types).

*HG R*

The six Y chromosome hts in this study belonging to HG R were identical at four of the eight STR loci examined (Table 3.1), and this HG had relatively low genetic diversity (*h*=0.889, fig. 3.1). The HG R hts from the present study were compared to hts from HG R found in 71 South African whites (Thejane Motladile, unpublished data), five north Africans and 73 individuals from Iberia (Bosch et al. 2001). The most frequent HG R ht found in central Africans (ht173, N=3, Table 3.1) was also observed in one Basque individual, in whom it was associated with M65 within HG R (Bosch et al. 2001). No other hts were shared between central Africans and other populations used for comparative purposes (fig. 3.7). In a network of 79 hts from HG R, the six central African hts did not cluster with each other or with hts from north Africans (fig. 3.7). Thus despite the apparent similarity of the hts from HG R in central Africans, phylogenetic analysis showed that they were not closely related to each other.

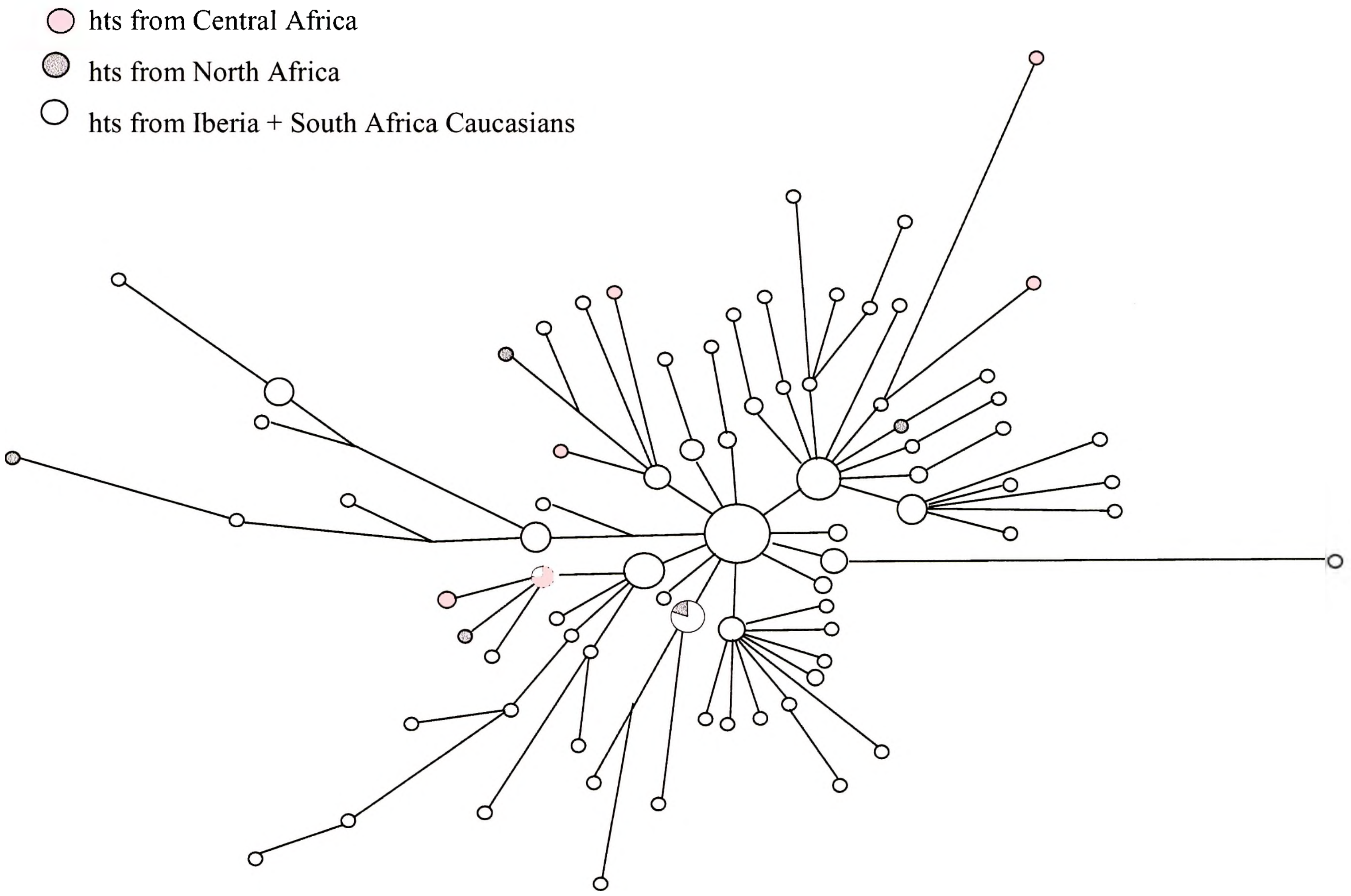

○ hts from Central Africa
● hts from North Africa
○ hts from Iberia + South Africa Caucasians

**Figure 3.7** MJ network depicting the relationships among haplotypes within HG R. White circles indicate haplotypes from Iberians and South African Caucasians (Bosch et al. 2001; TW Motladile, unpublished data); gray circles indicate haplotypes from north Africa (Bosch et al. 2001), and red circles indicate haplotypes from central Africa from the present study. Circle size is proportional to haplotype frequency.

### 3.1.2 *Genetic affinities among central African populations using Y chromosome data*

The genetic relationships among the five central African populations were assessed using exact tests of differentiation performed on HG data, combined HG-STR data, and STR ht data. When HG data from pairs of populations were analysed (Table 3.2a), the Biaka were significantly different from the other central African populations ($P<0.001$), and Ugandans were significantly different from populations from the C.A.R. and Zambia ($P<0.05$). Results were similar when STR ht data were used, with both Biaka and Ugandan populations being significantly different from other central African populations (Table 3.2c). However when HG-STR data were used (Table 3.2b), all pairs of populations, with the exception of Zambians vs. the D.R.C. population, were significantly different from each other ($P<0.01$). These results reflect the increased power of resolution of HG-STR data over HG data or STR data alone. In a NJ-tree constructed from HG-STR data and Slatkin's linearised $F_{ST}$ values (fig. 3.8) the difference between the Biaka and other central African populations was reflected by their placement on different branches. The distance between Ugandans and other non-Pygmy populations, and the similarity between populations from the D.R.C. and Zambia, were also reflected in the NJ tree (fig. 3.8).

### 3.1.3 *Correlation of Y chromosome data types with each other*

Pairwise $F_{ST}$ distances among the five central African populations calculated from HG frequencies were well-correlated ($r=0.89$, $P<0.05$) with pairwise $R_{ST}$ distances among them based on STR hts only (Table 3.2), and with pairwise $F_{ST}$ distances calculated from frequencies of HG-STR hts ($r=0.82$, $P<0.05$). This suggested that different Y chromosome data types are capable of revealing similar population relationships. The correlation of pairwise $F_{ST}$ distances calculated from frequencies of HG-STR hts with $R_{ST}$ distances calculated from hts constructed from STRs only was not statistically significant ($r=0.75$, $P=0.07$). It is likely that the high degree of homoplasy seen in HG-STR hts affected estimates of genetic distances among populations, and that the use of STR haplotypes alone (i.e. without HG data) may cause a slight decrease in the resolution of genetic distances among populations.

**Table 3.2** Pairwise genetic distances among five central African populations calculated from Y chromosome data

**a) Matrix of Fst genetic distances calculated from HG frequency data**

| | CAP | CAR | DRC | UGA |
|---|---|---|---|---|
| CAR | 0.08853** | | | |
| DRC | 0.09498*** | -0.00417 | | |
| UGA | 0.09648*** | 0.02555* | 0.01765 | |
| ZAM | 0.12923*** | 0.00344 | -0.00577 | 0.04445** |

**b) Matrix of Fst genetic distances calculated from HG-STR frequency data**

| | CAP | CAR | DRC | UGA |
|---|---|---|---|---|
| CAR | 0.2672*** | | | |
| DRC | 0.02158*** | 0.00951** | | |
| UGA | 0.01591*** | 0.01467*** | 0.0085** | |
| ZAM | 0.02312*** | 0.01282*** | -0.00009 | 0.01194*** |

**c) Matrix of Rst genetic distances calculated from STR haplotype data**

| | CAP | CAR | DRC | UGA |
|---|---|---|---|---|
| CAR | 0.07778*** | | | |
| DRC | 0.06957** | 0.00319 | | |
| UGA | 0.11782*** | 0.05952*** | 0.01879** | |
| ZAM | 0.0904** | 0.00978 | -0.01061 | 0.02376** |

Abbreviations:
CAP: Pygmies from Central African Republic
CAR: Central African Republic
DRC: Democratic Republic of Congo
UGA: Uganda
ZAM: Zambia
*significant difference, P<0.05
**significant difference, P<0.01
***significant difference, P<0.001

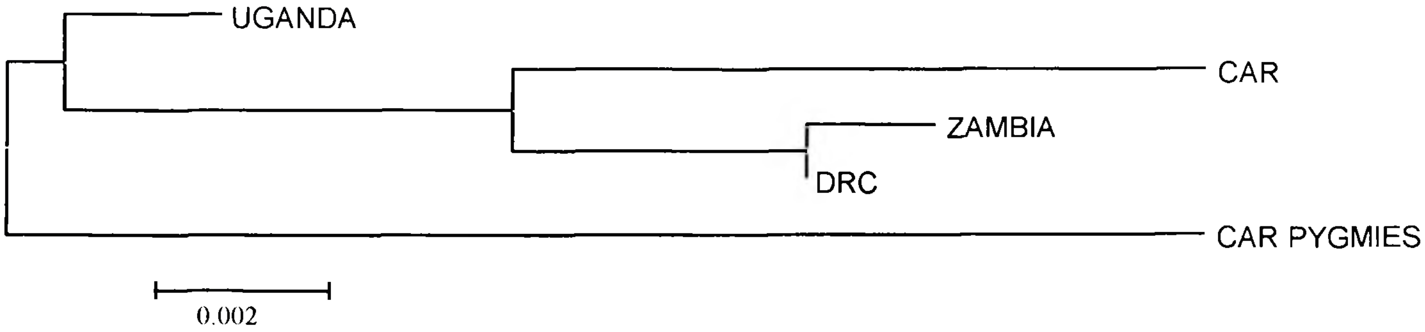


**Figure 3.8** NJ tree constructed using Y chromosome HG-STR data and $F_{ST}$ genetic distances, showing the genetic affinities among five central African populations

### 3.1.4 *Comparison to Y chromosome data from other African populations*

Y chromosome data from the present study were compared to HG and STR data from other African populations (Table 2.2) in order to examine the geographic distribution of Y chromosome variation in Africa, and to understand the genetic affinities of central African populations with other African populations.

#### 3.1.4.1 *Distribution of Y chromosome HGs in Africa*

The geographic distribution of Y chromosome HGs found in African populations is depicted in fig. 3.9. HGs E-M191 and E-M2 were found at high frequencies in sub-Saharan African populations. HG E-M191 was the most common HG in central African populations, while HG E-M2 was the most common HG in the populations of west Africa (fig. 3.9). HG E-M40 was frequent in west and north-east African populations, but was rare in the populations of central Africa. As previously reported by Underhill et al. (2001) and Semino et al. (2002), HGs E-M35 and FJ were particularly frequent in the populations of north-west and north-east Africa, whilst HG A was common in Khoisan populations and populations from Ethiopia. Two west African populations (Fulbe and Fali) had relatively high proportions of Y chromosomes belonging to typically Eurasian HGS (HGs FJ, KO and R; fig. 3.9).

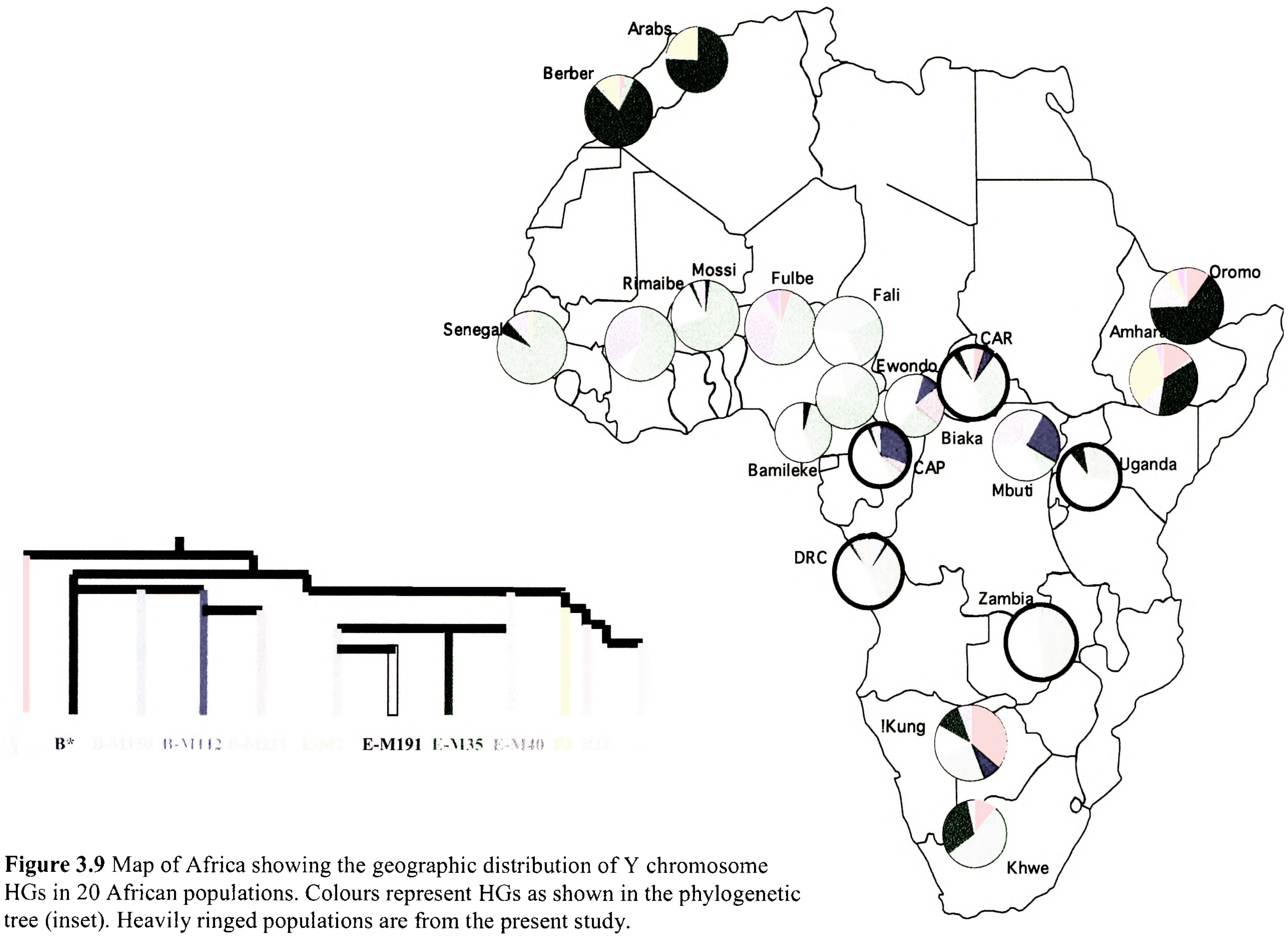


**Figure 3.9** Map of Africa showing the geographic distribution of Y chromosome HGs in 20 African populations. Colours represent HGs as shown in the phylogenetic tree (inset). Heavily ringed populations are from the present study.

### 3.1.4.2 *Distribution of 5-locus Y- STR haplotypes in Africa*

The distribution of haplotypes constructed from 5 STR loci only was examined in 17 African populations (N=1192, Table 2.2). Ten haplotypes (Table 3.3, fig. 3.10) together accounted for between 0.230 and 0.720 of the Y chromosomes in each of the 17 populations [with the exception of Ethiopians, in whom they only accounted for 0.060 of the sample (N=91)]. Altogether, these ten haplotypes accounted for 546 / 1192 (0.458) of Y chromosomes from 17 African populations.

Using the distribution of hts in HGs in the present study, it was possible to tentatively assign the ten frequent hts to HGs (Table 3.3, fig 3.10). Three haplotypes could be placed in HG E-M2, and four in HG E-M191; one could have belonged to either HG E-M2 or E-M191. One haplotype was assigned to HG E-M35, and one tentatively to HG B-M150 (although this haplotype was particularly homoplasic, Table 3.3). Eight of the ten haplotypes (all those suggested to be from HG E-M2 and M191) could be very closely linked in a hand-drawn phylogenetic network (fig. 3.10). Ht 2 corresponded to the 5-locus "Bantu Modal Haplotype" described by Pereira et al. (2002).

Hts 1-3 (potentially associated with HG E-M2, shown in yellow in fig. 3.10) were particularly frequent in central and south Africa, with very high frequencies of the "BMH" (ht 2) along the west coast (Angola and D.R.C) and in the C.A.R. (fig. 3.10). Hts 5-8 (potentially associated with HG E-M191, shown in blue in fig. 3.10) were observed frequently in central African populations, particularly in Uganda; these populations had correspondingly lower frequencies of hts 1-3. Hts 5-8 were reduced in frequency or nearly absent along the west coast (e.g. Angola, Sao Tome e Principe). Both the San and Biaka populations had low frequencies of both M2 and M191-associated hts (fig. 3.10). Ht10 was restricted in its distribution to populations of north Africa.

**Table 3.3** The ten most frequent 5-STR Y chromosome hts in 17 African populations (N=1192), and their likely HG of origin

| | DYS19-390-391-392-393 | N | probable HG [1] | homoplasy in other HG[1] |
|---|---|---|---|---|
| **ht1** | 15-21-10-11-13 | **124** | E-M2 (13.8%) | E-M191 (0.27%), E-M40 (0.27%) |
| **ht2** | 15-21-11-11-13 | **54** | E-M2 (4.1%) | E-M35 (0.27%) |
| **ht3** | 16-21-10-11-13 | **35** | E-M2 (2.44%) | E-M191 (0.27%), A (0.27%) |
| **ht4** | 15-21-10-11-14 | **59** | E-M2 (2.44%), E-M191 (1.36%) | |
| **ht5** | 16-21-10-11-14 | **55** | E-M191 (6.23%) | E-M2 (1.90%) |
| **ht6** | 16-21-10-11-15 | **51** | E-M191 (6.23%) | E-M2 (0.27%), E-M35 (0.27%) |
| **ht7** | 17-21-10-11-14 | **55** | E-M191 (8.67%) | E-M2 (.027%) |
| **ht8** | 17-21-10-11-15 | **43** | E-M191 (8.94%) | E-M35 (0.27%) |
| **ht9** | 15-24-10-11-13 | **26** | B-M150 (0.81%) | A (0.27%), B-M112 (0.27%), E-M40 (0.27%) |
| **ht10** | 13-24-9-11-13 | **44** | E-M35 [2] | |
| total | | 546 | | |

[1] HG assignment was based on data from the present study. Numbers in brackets are the frequency of the ht in each HG observed in the present study Colours represent assigned HGs and are consistent with the key in Figure 3.10.

[2] ht10 was not observed in the present study; HG assignment was from Bosch et al. (2001)

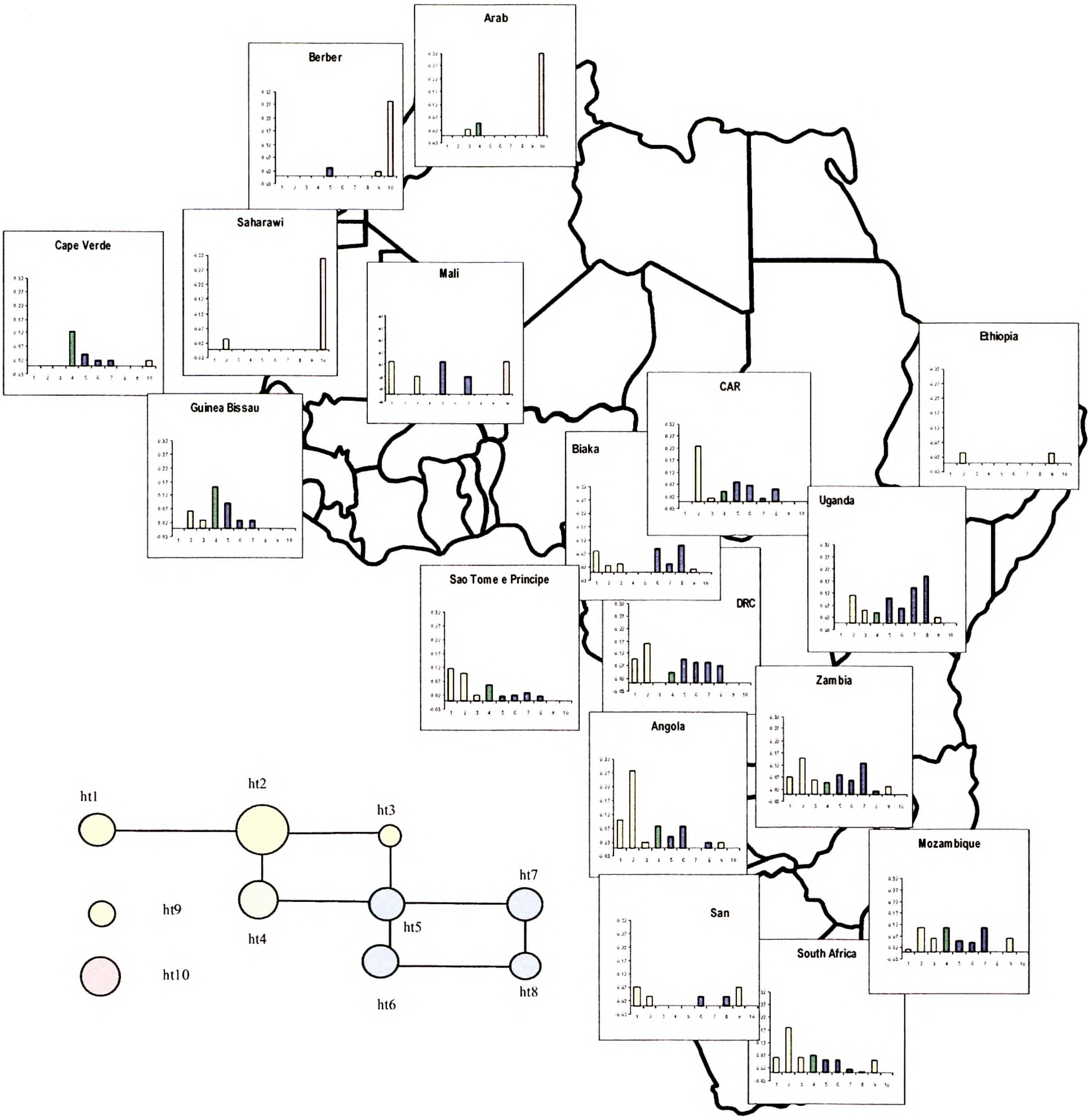


**Figure 3.10** Map of Africa showing the geographic distribution of the ten most frequent haplotypes constructed from five Y-specific STR loci (DYS19-DYS390-DYS391-DYS392-DYS393) in 17 African populations. Hts are listed in Table 3.3 and colours indicate suggested associated HGs (yellow=HG E-M2; green=HG E-M2*; blue=HG E-M191; orange=HG B-M150; pink=HG E-M35). Inset: phylogenetic network showing the close relationships among hts 1-8.

### 3.1.4.3 *Genetic affinities of African populations using Y chromosome data*

Genetic affinities of African populations were assessed using PC plots and AMOVA analyses. Three PC plots and three sets of AMOVA were performed using (a) HG frequencies from 20 African populations (N=1108, Table 2.2); (b) STR hts constructed from eight STR loci from 13 African populations (N=713, Table 2.2); (c) STR hts constructed from five STR loci from 17 African populations (N=1192, Table 2.2). The different data types were utilised to compare their ability to reveal population relationships.

*PC plots*

Using (a) HG data, three major clusters of populations were observed in the PC plot featured in fig. 3.11. The first cluster consisted of populations from west-central Africa, the second cluster included southern African (Khoisan) populations, and the third cluster comprised populations from north Africa. These clusters are also consistent with a linguistic subdivision of African populations, with populations speaking Niger-Congo, Khoisan, and Afro-Asiatic languages respectively. Within the west-central cluster, populations from central Africa speaking Bantu, Bantoid or Adamawa-Ubangian languages were separated from populations from west Africa speaking Atlantic or Voltaic languages (shown in red and orange respectively in fig. 3.11; Table 2.2). All five populations examined in the present study were placed within the Niger-Congo-speaking central African cluster (fig. 3.11). The same three-way split between Afro-Asiatic-speaking populations from north Africa, Khoisan populations, and populations from sub-Saharan Africa speaking Niger-Congo languages, was observed in PC plots constructed using (b) 8-locus STR hts and (c) 5-locus STR haplotypes (figs. 3.12, 3.13). The differentiation between Niger-Congo and Khoisan-speaking populations, and between populations from south-central and west Africa was less clear when 5-locus STR data were used (fig. 3.13).

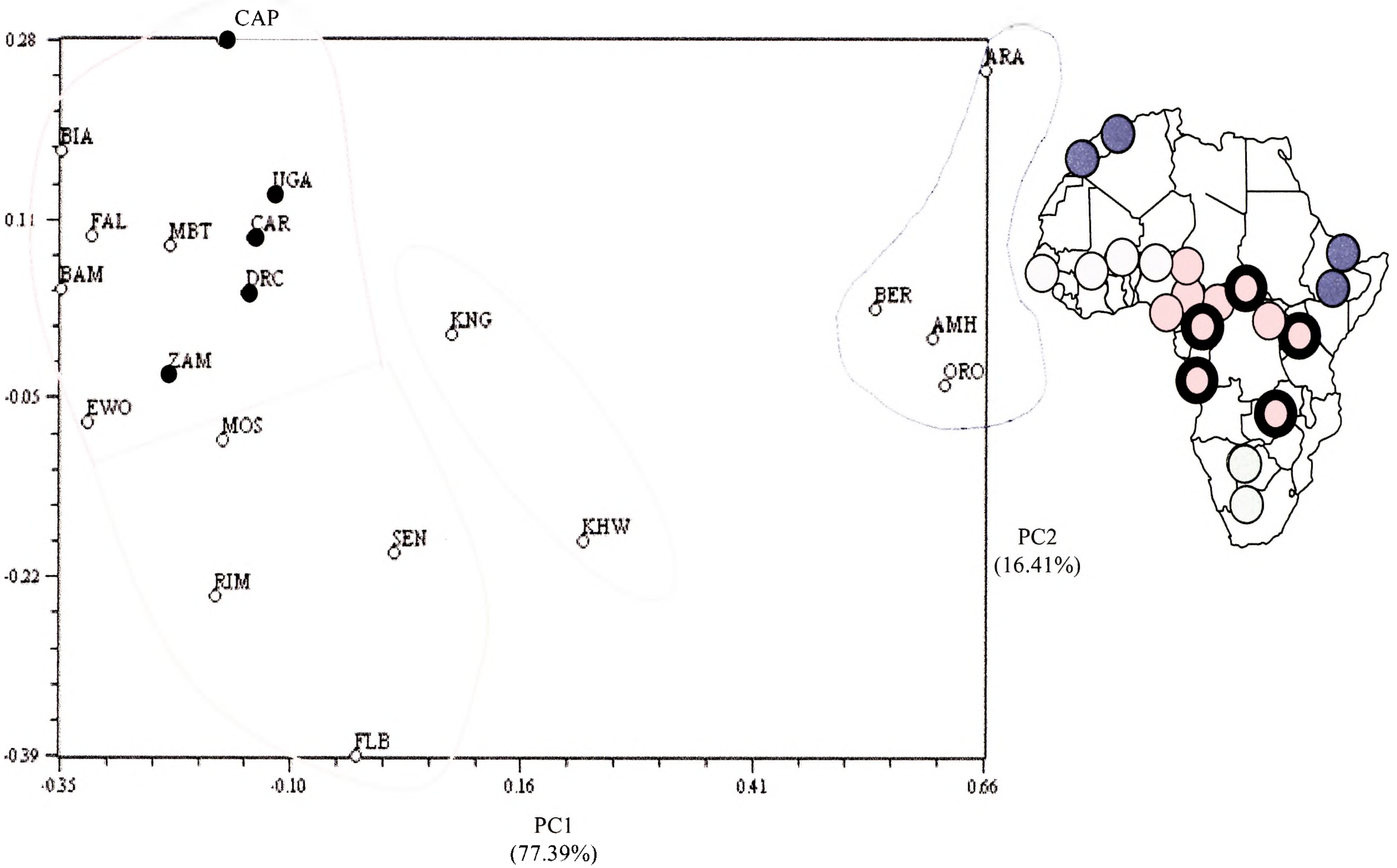


**Figure 3.11** PC plot constructed using HG data and $F_{ST}$ genetic distances among 20 African populations. The inset map shows the geographic origin of the sampled populations, and circles in black or ringed in black indicate central African populations from the present study. Population abbreviations: refer to Table 2.2.

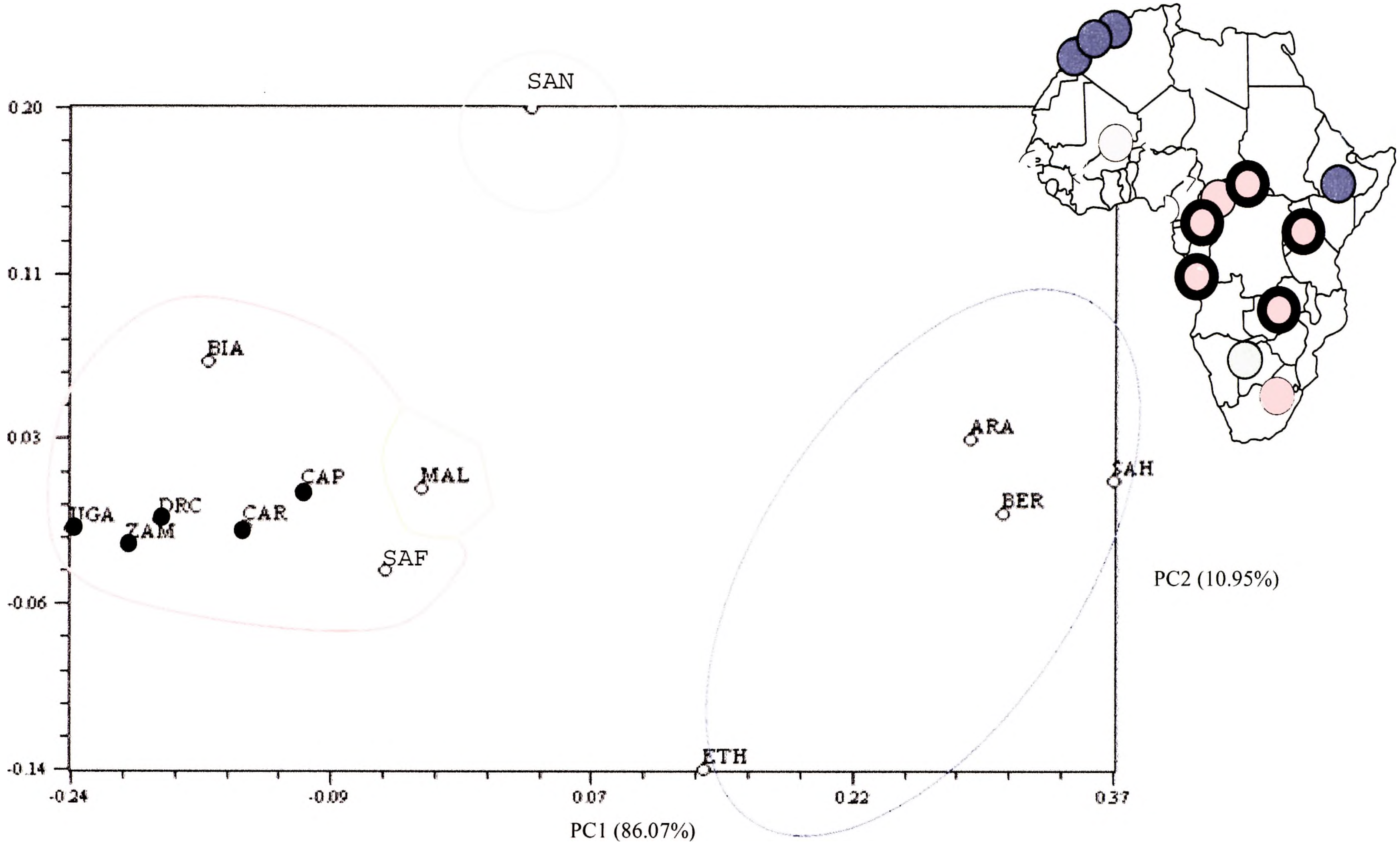


**Figure 3.12** PC plot constructed using 8-locus STR ht data and $R_{ST}$ genetic distances among 13 African populations. The inset map shows the geographic origin of the sampled populations, and circles in black or ringed in black indicate central African populations from the present study. Population abbreviations: refer to Table 2.2.

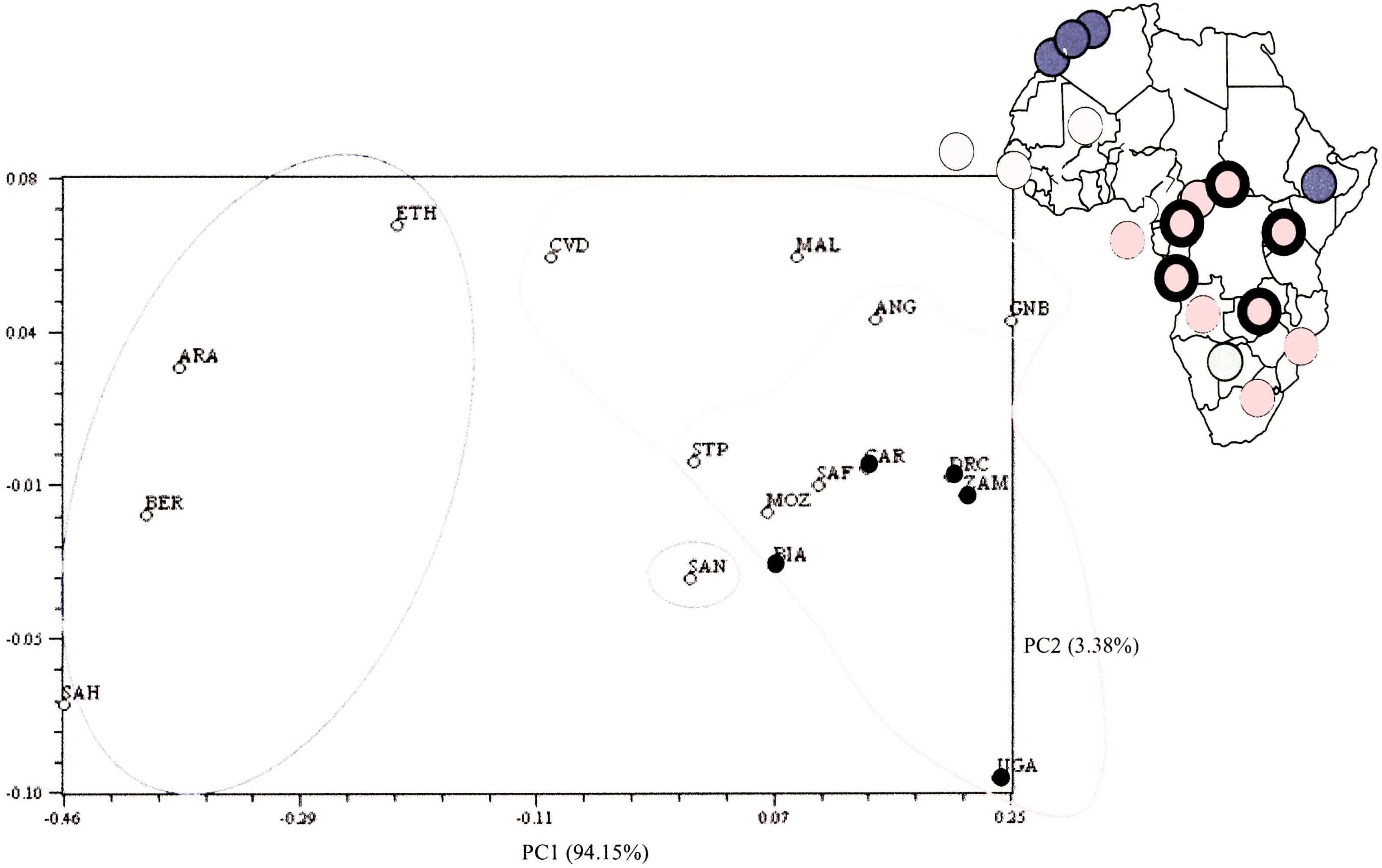


**Figure 3.13** PC plot constructed using 5-locus STR ht data and $R_{ST}$ genetic distances among 17 African populations. The inset map shows the geographic origin of the sampled populations, and circles in black or ringed in black indicate central African populations from the present study. Population abbreviations: refer to Table 2.2.

### *AMOVA analyses*

Genetic affinities among African populations were also assessed using AMOVA. Y chromosome HG, HG-STR or STR ht data from African populations were grouped according to a series of linguistic or geographic criteria (Table 3.4 and Table 2.2) and analysed by AMOVA. "Ethnic criteria" i.e. the placement of Pygmy populations and of Khoisan populations into separate groups, were also used in combination with geographic or linguistic criteria (Table 3.4). All groups classified according to geographic criteria were well-supported by AMOVA results (higher inter-group $F_{CT}$ than intra-group variance $F_{SC,}$ Table 3.4A). The best results (lowest $F_{SC}$ values i.e. lowest intra-group variance) were obtained when populations were grouped using a combination of four geographic regions (east, north, west and combined south-central) and ethnic criteria (Table 3.4B). This was true both when HG data were used, and when 8-locus or 5-locus STR ht data were used. Groups classified according to linguistic criteria (including or excluding ethnic criteria) were also supported by AMOVA analyses (Table 3.4C, 3.4D) but these groups produced higher $F_{SC}$ values than groups categorized according to geographic criteria i.e. Y chromosome HG, HG-STR or STR ht data from African populations were better grouped according to geographic origin than according to language group. Data from Nilo-Saharan-speaking populations are at present not well represented in the literature; such data may further help to resolve patterns of Y chromosome structuring in Africa.

**Table 3.4** Results of AMOVA tests using Y chromosome data from African populations

| Criteria | No.of groups | Groups[a] | | | | | | | A. HG data from 20 populations | | | | B. 8-locus STR data from 13 populations | | | | C 5-locus STR data from 17 populations | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | b | c | d | P | b | c | d | P | b | c | d | P |
| A) Geography[e] (2 regions) | 2 | NE+NW | C+W+S | | | | | | 31.14 | 11.23 | 57.63 | all P< 0.001 | 14.67 | 6.08 | 79.26 | all P< 0.001 | 17.31 | 4.9 | 77.79 | all P < 0.005 |
| A) Geography[e] (3 regions) | 3 | NE+NW | C+W | S | | | | | 26.2 | 11.33 | 62.05 | all P < 0.005 | 11.4 | 6.41 | 82.19 | all P < 0.005 | 9.44 | 5.78 | 84.78 | all P < 0.005 |
| A) Geography[e] (4 regions) | 4 | NE | NW | C+W | S | | | | 25.37 | 11.67 | 62.96 | all P < 0.005 | 14.64 | 3.55 | 81.81 | all P< 0.001 | 11.01 | 4.56 | 84.42 | all P < 0.005 |
| A) Geography[e] (4 regions) | 4 | NE | NW | W | C+S | | | | 27.83 | 6.98 | 65.19 | all P< 0.001 | 17.09 | 3.52 | 79.39 | all P< 0.001 | 15.6 | 3.62 | 80.78 | all P< 0.001 |
| A) Geography[e] (all 5 regions separate) | 5 | NE | NW | W | C | S | | | 28.11 | 5.13 | 66.76 | all P< 0.001 | 14.37 | 3.44 | 82.19 | all P< 0.001 | 10.46 | 4.13 | 85.41 | all P< 0.001 |
| **B) Geography[f] and ethnicity** | **6** | **NE** | **NW** | **W** | **C+S** | **Pygmy** | **Khoisan** | | **27.67** | **4.38** | **67.95** | **all P< 0.001** | **15.63** | **2.33** | **82.05** | **all P< 0.001** | **11.95** | **3.39** | **84.66** | **all P< 0.001** |
| C) Language[f] | 5 | AA | NS | NC | KS | IE | | | 26.15 | 11.33 | 62.52 | all P< 0.001 | 13.8 | 5.87 | 80.33 | all P < 0.005 | 12.21 | 4.27 | 83.52 | all P< 0.001 |
| D) Language[f] and ethnicity | 6 | AA | [NS][g] | NC | KS | IE | Pygmy | | 25.35 | 10.52 | 64.13 | all P< 0.001 | 12.38 | 5.61 | 82.01 | all P< 0.001 | 11.29 | 3.95 | 84.77 | all P< 0.001 |

[a] Populations were placed into groups according to geographic or linguistic criteria; refer to Table 2.3

[b] percent variance among groups

[c] percent variance among populations within groups

[d] percent variance within populations

[e] Geographic regions NE=north-east, NW=north-west, W=west, C=central. S=south

[f] language families AA = Afro-Asiatic, NC = Niger-Congo, NS = Nilo-Saharan, KS = Khoisan , IE=Indo-European

[g] square brackets indicate lack of data for that category

[h] populations from Angola and Guinea Bissau were classified as speaking NK languages

### 3.1.5 *Correlations between Y chromosome data and geography*

The pairwise geographic distances among African populations were significantly correlated with the pairwise $F_{ST}$ genetic distances among them, whether HG data from 20 populations (r=0.538, *P*<0.001), 8-locus STR ht data from 13 populations (r=0.526, *P*<0.001) or 5-locus STR ht data from 17 populations were used (r=0.434, *P*<0.001). Graphs showing these correlations are shown in Appendix 8.6A. The correlation between pairwise genetic $F_{ST}$ distances and pairwise geographic distances was then tested among populations from south and central (sub-Saharan) Africa only using HG or STR ht data (Table 2.2). Poor correlations were obtained using all three data types (using HG data, r=0.229, *P*=0.061; using 8-locus STR ht data, r=0.160 and *P*=0.180; using 5-locus STR ht data, r=0.006 and *P*=0.495). Graphs showing these correlations are shown in Appendix 8.6B. These results suggest that geographically separated populations in sub-Saharan Africa are genetically closely related and /or that populations living in close proximity are genetically different. These scenarios could have been caused by migrations of populations across sub-Saharan Africa, or by population admixture.

Next the correlation between genetic and geographic distances was tested among African populations from the south-central sub-Saharan region speaking languages classified within the Niger-Congo language group only (Table 2.2). Again, poor correlations were obtained using all three data types (with HG data, r= -0.231, *P*= 0.774; with 8-locus STR ht data, r= 0.096 and *P*= 0.264; with 5-locus STR ht data, r= 0.112 and *P*= 0.266). Graphs showing these correlations are shown in Appendix 8.6C. These results suggest that the lack of correlation between genetic and geographic distances seen in sub-Saharan African populations above, is largely due to a similar lack of correlation in populations from this region speaking Niger-Congo languages. Populations speaking Niger-Congo languages are known to have experienced recent and rapid migrations within sub-Saharan Africa (Phillipson 1977, Ehret 1998).

Similar poor correlations of genetic distance with geography were obtained in an analysis which included only the five Niger-Congo speaking central African populations from this study ($-0.259< r < -0.161$, $0.221<=P<=0.307$, using genetic distances measured by $F_{ST}$ from HG data or HG-STR data, or by $R_{ST}$ from STR ht data). The exclusion of the Biaka from the analysis did not improve the correlation ($- 0.402< r < 0.028$, $0.167<=P<=0.436$).

## 3.2 DISCUSSION

The present-day patterns of genetic variation in African populations are complex and harbour information that can be traced to both ancient and recent evolutionary events. Data collected in this study have allowed the genetic affinities of five central African populations to be examined, and have facilitated an understanding of the structure of Y chromosome variation in Africa.

### 3.2.1 *Ancient Y chromosome lineages in central African populations*

HGs A and B are the most ancient Y chromosome lineages described globally and may have been associated with population expansions that occurred in Africa between 130 000 and 70 000 years ago (Underhill et al. 2001). Presently these HGs are restricted to African populations.

HG A has previously been reported in populations from southern Africa (Kung 36%, Khwe 12%, South Africans 5%), east Africa (Sudan 45%, Ethiopia 14%) and west Africa (Mali 2%) and distinct clades were found in Khoisan and east African populations (Scozzari et al. 1999; Underhill et al. 2000; Cruciani et al. 2002; Semino et al. 2002). These findings have been interpreted as showing that HG A was part of the proto-African Y chromosome gene pool, and that Khoisan and east African populations show ancient divergence from a common ancestral population (Semino et al. 2002). HG A was found at very low frequencies in three of five central African populations in the present study (Zambia 1.1%, D.R.C. 1.4%, C.A.R. 5.3%) and was not found in Ugandans or the Biaka (figs. 3.1, 3.9). The absence, or near-absence, of HG A in central African populations in the present study, as well as in 20 Biaka and 17 Mbuti Pygmies examined by Underhill et al. (2000), indicates either that populations carrying this early lineage did not occupy the central African region, or that HG A did not persist there. Central African populations are therefore likely to be derived from more recent ancestral populations than are Khoisan and east African populations. This relates well with archaeological evidence (Cahen 1982; Phillipson 1985) which suggests that the central African region was populated at a somewhat later date (Middle Stone Age) than the surrounding eastern and southern African regions (populated since the Early Stone Age).

HG B has previously been reported at varying frequencies in populations from south, east, west and central Africa (Underhill et al. 2000; Cruciani et al. 2002; Semino et al. 2002). In the present study, HG B was present at significantly higher frequency ($P<0.001$) in the Biaka (35%) than in the other four central African populations studied (D.R.C. 9%; C.A.R. 5%, Zambia 6%, Uganda 1%, figs. 3.1, 3.9). This frequency is similar to those previously observed in samples of 20 Biaka (35%) and 12 Mbuti Pygmies (33%; Underhill et al. 2001; Cruciani et al. 2002). The observation of HG B at high frequency in Pygmy populations suggests that HG B may be a signature of the original central African Y chromosome gene pool, since the Pygmy populations are thought to be descended from the autochthonous inhabitants of central Africa. When data from this study were analysed together with published data, HG B was significantly more frequent in populations of central Africa than in other African populations ($P<0.05$).

HG B sublineages have different distributions: B-M112 is found mostly in hunter-gatherer populations from southern and central Africa, whilst B-M150 is found in other sub-Saharan African populations (Underhill et al. 2000; Cruciani et al. 2002; Semino et al. 2002). This difference was confirmed in the present study: HG B-M112 was found mostly in the Biaka, whilst HG B-M150 was found mostly in non-Pygmy populations (figs 3.1, 3.2, 3.3). The Biaka may share some ancestry with proto-Khoisan populations because Biaka and Khoisan share the HG B-M112 default lineage (Underhill et al. 2000, Cruciani et al. 2002). However the absence of HG A in the Biaka suggests that this relationship post-dates the earliest Khoisan Y chromosome diversification.

The overall pattern of distribution of HGs B-M150 and B-M112 in Africa (fig. 3.9) and their estimated TMRCAs, suggests that HG B-M150 has been dispersed widely quite recently (possibly only within the last thousand years). HG B-M112, on the other hand, is more ancient and has either undergone a more limited dispersal, or once had a wider geographic range but has only been retained in present-day hunter-gatherer populations. The typical existence of ancient hunter-gatherer populations in small, relatively isolated communities could account for the retention of B-M112 in these populations. The observation of several population-specific HG B sublineages [e.g. HG B-M211 in Biaka (present study), and other lineages in Mbuti, Mossi, Mali, Lissongo in Cruciani et al. (2002)], further suggests that population fragmentation and subdivision affected the evolution of HG B.

### 3.2.2 *More recent contributions to the central African Y chromosome gene pool*

Y HGs defined by M168 represent relatively more recent African (and non-African) lineages (Underhill et al. 2001) and include HGs E, FJ and R in the present study.

#### 3.2.2.1 *The "Bantu Expansion"*

Most of the existing Y chromosome variation in central African populations appeared to be due to a massive expansion of HG E-M2*, including both the M2 and M191 sublineages (figs. 3.1, 3.5). The star-like shape of a MJ network drawn with HG E-M2* ht data from the central African populations (fig. 3.5) and high frequencies of particular hts within this HG (Table 3.1) confirmed a rapid growth in numbers of Y chromosomes associated with this HG. In addition, geographic range expansion of HG E-M2* was inferred from the close relationship among hts associated with E-M2* from different central African populations studied (fig. 3.5), and the presence of many of these hts in more than one population (Table 3.1). Geographic range expansion of HG E-M2* was also implied by the wide geographic distribution in other African populations of a common core of frequent and closely-related 5-locus STR hts inferred to be associated with E-M2* (fig. 3.10).

HG E-M2* has been suggested to be associated with the "Bantu Expansion" (Passarino et al. 1998; Scozzari et al. 1999; Underhill et al. 2001; Cruciani et al. 2002). The lack of correlation between genetic and geographic distances among Niger-Congo-speaking populations from sub-Saharan Africa, among whom HG E-M2* accounted for the majority of their Y chromosome gene pool, attested to the role of this linguistic group in an expansion/s. The dates obtained in this study for TMRCA of HG E-M2* (2 407 -3 744 years ago) were similar to those obtained in other studies (Thomas et al. 2000; Underhill et al. 2001) and were also in accordance with dates suggested for the start of the "Bantu Expansion" on the basis of linguistic and archaeological data.

However this explanation may be overly simplistic, since HG E-M2* was also observed at high frequencies in non-Bantu-speaking populations of west Africa, as well as in an Adamawa-Ubangian-speaking population from the C.A.R. (fig. 3.9). HG E-M2* could therefore also have been dispersed via other migrations, such as the expansion of Adamawa-Ubangian-speakers that occurred at roughly the same time as the expansion of Bantu-speakers, or the expansion of speakers of Atlantic or Voltaic languages to west Africa. It seems likely that HG E-M2* is a signature not just of Bantu Expansion, but of the dispersal of several different language families within the Niger-Congo phylum.

The sublineages E-M2 and E-M191 within HG E-M2* both appeared have been involved in the expansions described above, but their distributions and therefore probable routes of dispersal were not identical. HG E-M2 was very frequent in west Africa (fig. 3.9) and haplotypes know or suggested to be associated with HG E-M2 were most frequent in populations along the west coast of Africa (fig. 3.10). Pereira et al. (2002) also noted a "stronger reduction of diversity" associated with the "Bantu Modal Haplotype" (ht 2 in fig. 3.10 in the present study) along the African west coast. On the other hand, HG E-M191 reached a maximum frequency in Uganda, was frequent in central Africa and rare in west and south Africa (figs. 3.1, 3.9). Five-locus STR hts suggested to be associated with E-M191 were also most frequent in Uganda and decreased in frequency in a radiating pattern to the west and south (fig. 3.10).

It is tempting to interpret the patterns of HGs E-M2 and E-M191 distribution as signatures of the expansion of Western and Eastern Bantu-speakers, respectively. In this scenario, HG E-M2, the more ancestral lineage, would have spread from the Bantu source population in the region of Cameroon/Nigeria along the west coast, and into regions associated with the expansion of Western Bantu languages (fig. 1.4). The more derived lineage, HG E-M191, may have arisen in the eastern Great Lakes region, and become frequent in the secondary expansion of speakers of Eastern Bantu languages ~3 kya (Ehret 1998). This lineage may have then dispersed towards the west and south when Eastern Bantu-speakers later moved into areas initially occupied by Western Bantu-speakers (Vansina 1984). Zambia, for example, showed nearly equal proportions of HGs E-M2 and E-M191, as would be expected in this model since Zambia is thought to have experienced input from both Eastern and Western Bantu-speakers (Huffman 1989).

However, the exclusive association of specific Y chromosome HGs with specific migrations must be cautioned against, since it over-simplifies the interpretation of Y chromosome variation in African populations. Firstly, as described above, the distribution of lineages within HG E-M2* are likely to be associated not just with the Bantu Expansion, but with other population movements and migrations as well. Secondly, an unknown proportion of HG E-M2* lineages may previously have been present in sub-Saharan Africa prior to the Bantu Expansion. For instance, Underhill et al. (2001) suggested that HG E evolved in east Africa, and its sublineages may have dispersed into central or south Africa before the Bantu Expansion. Thirdly, the patterns of Y chromosome variation in [central] Africa can often be explained in more than one way. For example, the M191 mutation may have occurred in west Africa and later spread to the Great Lakes region where it was driven to high frequency by a further population expansion and /or by drift. This scenario is supported by the position of ht 157 in fig. 3.5. Thus HG E-M191 may have achieved its present distribution by a migration from the west towards the east and south, rather than from the east, towards the west as described above.

A further note of caution regarding generalised statements concerning the Bantu Expansion: although this migration undoubtedly had great impact on Y chromosome variation in [central] African populations, each of the five populations examined in the present study had a unique pattern of Y chromosome variation. All pairs of populations could be statistically distinguished from each other, with the exception of Zambian vs. D.R.C. populations. Four of the five populations examined in this study had different modal hts, and all populations had a large percentage of hts that were population-specific (Table 3.1). Thus although the Y chromosomes of Bantu-speaking populations may show great similarities, these populations are still genetically distinct from each other. Studies of autosomal variation have also shown that Bantu-speaking populations are not necessarily genetically homogenous (Lane et al. 2002; Tofanelli et al. 2003).

Furthermore, by including the DYS389I and II loci in the present study, the "Bantu Modal Haplotype" described by Thomas et al. (2000) and Pereira et al. (2002) was split into two different haplotypes (hts 61 and 67, Table 3.1). Ht 67 was modal in only the D.R.C. (Bantu-speaking) population, whilst ht 61 was modal in the non-Pygmy C.A.R. population. Neither of these hts was modal in Ugandans, who are also Bantu-speaking. Thus the putative founder ht described by Thomas et al. (2000) and Pereira et al. (2002) was NOT the most common ht in all Bantu-speaking populations in the present study, and was modal in a population that are not Bantu-speaking. The use of the term "Bantu Modal Haplotype" by Thomas et al. (2000) and Pereira et al. (2002) is premature because the term was coined after studying a limited number of African populations. We suggest that the term "modal haplotype" should be confined to populations (e.g. "Ugandan Modal Haplotype"), or abandoned altogether.

### 3.2.2.2 *Eurasian lineages*

Of the total 369 Y chromosomes from central African populations, 11 chromosomes (0.030) were from typically Eurasian HGs. The two chromosomes (0.005) from HG FJ probably represent a recent European contribution to the central African gene pool, likely to have occurred since the colonization of the region by Europeans. The two countries in which the FJ chromosomes were found, Zambia and the D.R.C., were colonized by Britain and Belgium, respectively.

HG R was recently described in Africa at high frequency in populations from Cameroon, and was suggested to have been introduced to Africa at least 4 100 years ago in a back-migration from Asia to sub-Saharan Africa (Cruciani et al. 2002). This HG is also present in north Africa (Bosch et al. 2001) and probably in Egypt (Scozzari et al. 1999). The presence of HG R in central African populations from this study (0.024) shows that this HG has a more wide-spread distribution in Africa than previously described, but occurs at low frequencies (fig. 3.1, Table 3.1).

The six HG R haplotypes observed in this study were not closely related to each other (fig. 3.7), suggesting that HG R in central Africa is not derived from a single founding male. Rather, multiple introductions of HG R into central Africa seem likely, possibly in the form of contributions from several different males from the same contributing population. Europeans may be one possible source of HG R in Africa. However most hts from central Africans were not closely related to the high-frequency HG R hts which appear to have experienced a founder effect in Iberians and South African whites (fig. 3.7). A single HG R ht (ht 173, Table 3.1) found in three of five central African populations examined [as well as in one Basque individual] may represent a recent dispersal of one particular HG R haplotype in central Africa from a recent European founder. Salas et al. (2002) suggested that HG R may have spread from Eurasia to north Africa, and from there to sub-Saharan Africa. Although HG R hts from north and central Africa were not closely related in a phylogenetic network (fig. 3.7), this analysis was hampered by the small number of HG R Y chromosomes from north African populations. Thus north Africa may still be the primary source of HG R in sub-Saharan Africa as suggested by Salas et al. (2002).

In the present study, HG R was most frequent in the C.A.R. (4/6 hts, 5/9 chromosomes, Table 3.1) and all of these C.A.R. individuals were from the Gbaya ethnic group. The C.A.R. borders Cameroon, the region in which HG R has previously been reported at high frequency (Cruciani et al. 2002). A comparison of STR ht data from this study to very recently published

STR data from Cameroon populations (Caglia et al. 2003) showed no matching STR hts, but two of the STR hts from Cameroon could be one- or two-step neighbours of hts 173 and 174 from HG R in this study. Unfortunately no SNPs were typed in the Caglia et al. (2003) study to confirm the membership of these STR hts in HG R. It seems plausible that HG R Y chromosomes from Cameroon and the C.A.R. have a common north African origin, and that HG R was then distributed from this new core area to the rest of sub-Saharan Africa. If HG R was contributed by north African males to central Africa prior to or during the "Bantu Expansion" as suggested by Cruciani et al. (2002), then HG R Y chromosomes may have been carried by males participating in the "Bantu Expansion" to reach the distribution observed in the present study. More specifically, based on its observed distribution (figs. 3.1, 3.9), HG R may have been carried by the western stream of the "Bantu Expansion".

### 3.2.3 *The genetic affinities of central African populations using Y chromosome data*

Each central African population examined had a unique set of Y chromosomes (fig. 3.1, Table 3.1) and all pairs of populations except for the D.R.C. and Zambia were significantly different when high-resolution HG-STR data were used (fig. 3.8, Table 3.2). The close relationship between populations from the D.R.C. and Zambia corresponded well to archaeological data suggesting that Zambia was partially populated by populations migrating from the D.R.C. since ~500 AD (Phillipson 1985). The genetic data may also represent more ancient shared male ancestry between populations from the neighbouring regions of present-day Zambia and the D.R.C. The Adamawa-Ubangian-speaking (non-Pygmy) population from the C.A.R., rather than the Ugandan populations, was the next most closely related to the Bantu-speaking populations from Zambia and the D.R.C. (fig. 3.8). This indicated that the use of linguistic subfamilies was not a reliable method to predict the genetic similarity of [central] African populations. The Ugandan population appeared to have a somewhat different Y chromosome history from those of the other central African populations examined, whilst Y chromosomes from Biaka Pygmies from the C.A.R. differed greatly and significantly from the other central African populations (fig. 3.8, Table 3.2). These two populations are discussed further below.

### 3.2.3.1 *Genetic affinities of Ugandans*

The differences between Ugandans and other central African populations were largely due to the higher frequencies of HGs E-M35 and E-M191 and a very low frequency of HG B in Ugandans (fig. 3.1, 3.9). Particular hts within HG E-M191 were also much more common in Uganda than in other African populations (Table 3.1, fig. 3.10).

The presence of HG E-M35 in Ugandans suggests gene flow from populations where this HG is frequent, such as in neighbouring east Africa. For example HG E-M35 has been found in 31.8% of 88 Ethiopians (Underhill et al. 2000). STR hts from central Africans may be associated with M78 (fig. 3.4), a sublineage of HG E-M35 that is common in east Africa (Underhill et al. 2000), or with other M35 sublineages (but probably not M35-M81). Linguistic evidence suggests a great deal of interaction of Bantu-speaking peoples in the Great Lakes region with Cushitic (Afro-Asiatic) and Nilo-Saharan-speaking populations originating in regions north of Uganda since ~ 3 kya (Ehret 1998). Such interactions may have resulted in the introduction of Y chromosomes from these populations into the gene-pool of present-day Ugandans.

The high frequency of HG E-M191 and particular hts within it may have occurred due to a further expansion of Bantu-speaking peoples in ~3 kya in the Great Lakes region (Ehret 1998), or merely due to genetic drift. The increase in frequency of HG E-M191 in Ugandan populations may have replaced pre-existing HG B chromosomes in this population. The possibility that HG E-M191 was introduced to Uganda from east Africa was considered; however this is unlikely since HG E-M2* was absent in Sudan (N=40), and present in only 3% of 88 Ethiopians (Underhill et al. 2000).

### 3.2.3.2 *Genetic affinities of the Biaka*

The Biaka Pygmies from the C.A.R. differed significantly from the other central African populations in this study (Table 3.2) using HG, HG-STR or STR ht data. They had significantly lower frequencies of HGs E-M2, E-M191 and E-M35 and did not have Y chromosomes from HGs A, B-M150 or R, whilst HG B-M112 and its derivative M211 were found nearly exclusively in the Biaka (fig. 3.1, Table 3.1).

It has been suggested that the Pygmies' forest habitat has afforded them a degree of isolation from surrounding non-Pygmy populations (Cavalli-Sforza 1986). Data from this study both supported and challenged this hypothesis. The significant difference between Y chromosomes of the Biaka and other central African populations supported genetic isolation of the Biaka from the other populations, and the extremely low frequency of HG B-M112 in non-Pygmy populations (fig. 3.1, Table 3.1) indicated that Y chromosome gene flow from Biaka Pygmies to non-Pygmies has been rare. This is similar to findings from mtDNA data that most mtDNA types from Biaka Pygmies differ significantly from those found in other African populations (Chen et al. 1995; Chen et al. 2000). On the other hand, chromosomes from HG E accounted for 0.650 of the Y chromosomes found in the Biaka (fig. 3.1) and can be assumed to represent substantial male gene flow from non-Pygmy populations into the Biaka. Studies of autosomal loci also inferred extensive gene flow from non-Pygmy populations (Cavalli-Sforza 1986; Wijsman 1986; Cavalli-Sforza et al. 1994), and analyses of the distribution of subtypes of human polyomavirus JC in central Africa also supported a high extent of contact between Biaka and neighbouring populations (Chima et al. 1998). It appears that male gene flow between the Biaka and neighbouring populations has occurred mostly in one direction, from non-Pygmy populations to the Biaka. This is consistent with local social constraints which allow marriages only between Pygmy women and non-Pygmy men, but not vice versa (Cavalli-Sforza 1986), but implies that male offspring of such unions (or of extra-marital liaisons) have returned to Biaka society instead of staying with their non-Pygmy father's society (as otherwise expected).

### 3.2.4 *Structure of Y chromosome variation in Africa*

Comparisons of data at different levels of resolution from populations spanning the major geographic regions of Africa showed that Y chromosome variation is highly structured in Africa. AMOVA testing of HG, HG-STR or STR data supported the grouping of African populations by geography rather than by language family (Table 3.4). A strong contrast was observed between Afro-Asiatic-speaking populations from North Africa, with high frequencies of HGs E-M35 and FJ, and Niger-Congo-speaking populations from sub-Saharan Africa, with high frequencies of HG E-M2* (fig. 3.9, 3.11-3.13). Within sub-Saharan Africa, central and south African populations were very similar to each other. These populations were also similar to, but distinct from west African populations.

Despite the relative geographic proximity of central African populations to all other geographic regions of Africa, Y chromosome exchange seems to have only occurred substantially between central and south populations, and to a lesser degree between central and west populations. The extent of Y chromosome gene flow between populations in central and west Africa is unclear, since their gene pools are similar due to shared ancestry. The Sahara Desert and the rainforest may both have acted as barriers to Y chromosome flow between populations in north and central Africa, whilst the change in environment in the Great Lakes / Rift Valley region seems to have acted as a partial genetic barrier between populations in east and central Africa. Only a small amount of Y chromosome gene flow from east African populations into central African populations has potentially occurred, as indicated by HG E-M35, especially into Uganda (discussed above). Further Y chromosome STR data from the other regions of Africa may help to resolve the relationships among African populations more clearly.

### 3.2.5 *Utility of different Y chromosome data types*

Analyses were performed using HG data, STR ht data, and combined HG-STR data in order to evaluate the usefulness of these Y chromosome data types in depicting the affinities of African populations. Using five central African populations, the genetic distances between them based on HG frequencies were extremely well-correlated with corresponding distances based on STR hts only, and with corresponding distances based on HG-STR data. The correlation between genetic distances based on HG-STR hts and genetic distances based on STR hts only was only slightly affected by the high degree of homoplasy observed in the HG-STR data set. Overall, these observations provide additional support for the conclusion drawn by Bosch et al. (1999) that "STR variation is deeply structured by genetic background on the human Y chromosome". They also suggest that datasets in which STRs only were typed (e.g. Seielstad et al. 1999, Trovoada et al. 2001) may still be useful for resolving African population affinities. Note however that the correlation may only hold for fairly "low-resolution" HGs such as those studied here. As more SNPs are used and more specific lineages are examined, the specificity of STR variation in those lineages decreases, as seen in this study in sub-lineages of HG E.

Y chromosome HG data in this study proved useful for mapping the evolution of ancestral male lineages, whilst STR data provided finer levels of resolution capable of distinguishing individual populations from each other, and allowed founder effects and episodes of gene flow to be dissected. Both data types allowed the tracing of population affinities. Ultimately the combined use of HG-STR hts is the most informative data type for understanding Y chromosome evolution in Africa and it is hoped that in the future similar data will become available for many more African populations.

# 4. MTDNA STUDIES

## 4.1 RESULTS

The mitochondrial COII/tRNA$^{Lys}$ intergenic 9-bp deletion, the 3592 *Hpa*I RFLP, and the two hypervariable regions of the mtDNA control region were used to assess variation in 92 Ugandans, 96 Zambians, 114 Biaka Pygmies and 95 non-Pygmies from the Central African Republic. These data were also used to examine the genetic affinities among the four central African populations, and among them and other African populations.

### 4.1.1 *Patterns of mtDNA variation in central African populations*

### 4.1.1.1 *The mtDNA 9-bp deletion polymorphism*

The mtDNA COII/tRNA$^{Lys}$ intergenic 9-bp deletion was observed in a total of 49 of 820 individuals (0.059) from four central African populations (Table 4.1). Frequencies in individual populations were within the range previously reported in other sub-Saharan African populations (0.000 – 0.300, Soodyall et al. 1996). The deletion was significantly more frequent in Zambians than in the other three populations (Fisher's exact test, $P$<0.001). The frequency of the deletion in 124 Biaka Pygmies (0.032) in this study was significantly lower (Fisher's exact test, $P$<0.05) than the frequency of 0.235 reported previously in a sample of 17 Biaka Pygmies (Vigilant 1990; Soodyall et al. 1996). Differences could be due to the difference in sample sizes and/or due to sampling bias at the time of collection during the present or previous studies.

Whilst screening for the mtDNA 9-bp deletion, an insertion in the COII/tRNA$^{Lys}$ intergenic region was observed in six of the 820 samples. All six of these samples were from non-Pygmies from the C.A.R. (Table 4.1). Sequencing of the COII/tRNA$^{Lys}$ intergenic region in these samples showed that the insertions were due to the presence of three complete copies of the 9-bp repeat motif. This is the first time that the triplication of the COII/tRNA$^{Lys}$ intergenic 9-bp motif has been reported in African populations.

**Table 4.1** The frequencies of the mtDNA 9-bp deletion and 9-bp triplication in four Central African populations.

| **population** | N | **Number of samples with 9-bp deletion (frequency)** | **Number of samples with 9-bp triplication (frequency)** |
|---|---|---|---|
| Biaka Pygmies | 124 | 4 (0.032) | 0 |
| C.A.R. non-Pygmies | 267 | 1 (0.004) | 6 (0.023) |
| Uganda | 259 | 18 (0.070) | 0 |
| Zambia | 170 | 26 (0.153) | 0 |
| **TOTAL** | **820** | **49 (0.059)** | **6 (0.007)** |

### 4.1.1.2 *The mtDNA 3592 HpaI polymorphism*

Both alleles of the mtDNA 3592 *Hpa*I restriction polymorphism were observed in this study (Table 4.2). The presence of the 3592 *Hpa*I recognition site (caused by 3594 C) was more frequent than its absence (Table 4.2), and frequencies of this allele were within the range (0.110-1.00) reported for other sub-Saharan African populations (Fox 1997). The 3594 C allele defines macrohaplogroup L* (L1 plus L2), found only in African populations.

Data regarding the 3592 *Hpa*I polymorphism were used together with HVR sequence information to classify samples as belonging to mtDNA HGs L1, L2 or L3 (Table 4.2). HG L1 was significantly more frequent in the Biaka than in the other three populations (Fisher's exact test, $P$<0.001). HG L3 was the most frequent HG in the three non-Pygmy central African populations, and HG L2 was the least frequent mtDNA HG observed in all four central African populations (Table 4.2).

### 4.1.1.3 *Control region sequence variation*

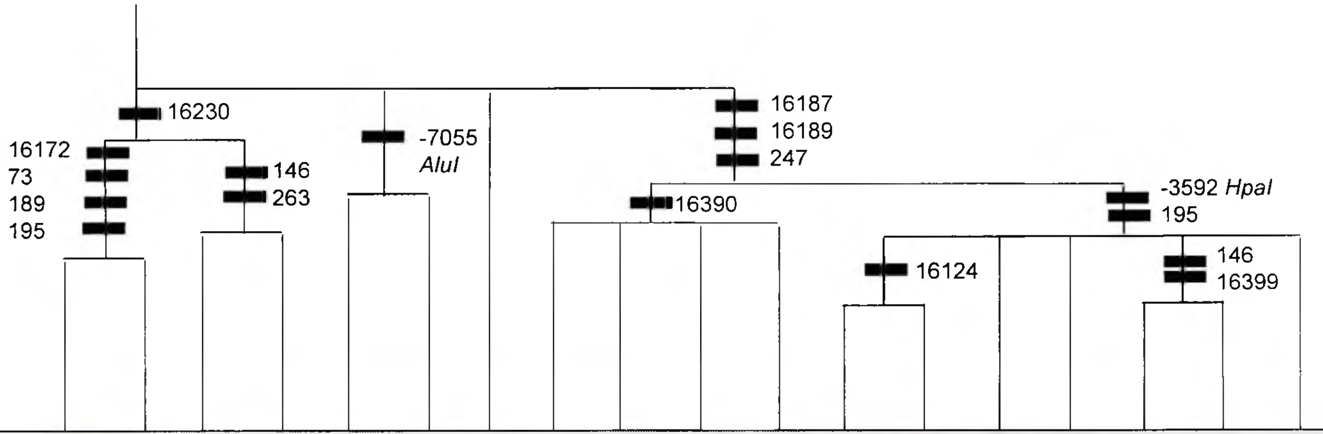
Sequence variation at 403 sites in HVRI and at 377 sites in HVRII was examined in 397 individuals. Altogether, 185 of the total 780 positions (0.237) showed variation (Tables 4.3-4.5). There were more variable sites in HVRI (114 positions or 0.283) than in HVRII (71 positions or 0.188). Transversions were observed at eight sites (16018, 16111, 16221, 16241, 16287, 16291, 85, 186); three different alleles were seen at another eight sites (16093, 16188, 16189, 16265, 16286, 16293, 95, 189); and all four possible alleles were observed at three sites (16114, 16166, 16197). Insertions were noted at positions 16018, 16183, 16192, 291, 302, and 315.1, and deletions were seen at positions 16182, 16183, 16263, 16325, 242, 243, 244, 247 and 291. All insertions and deletions were point mutations, with the exception of a 15bp insertion at 16018 observed in one Zambian individual (Table 4.3, type 63 in HG L1). This 15-base insertion consisted of a duplication of nps 16018-16032, and was confirmed by repeated sequencing of this sample in both directions.

**Table 4.2** Frequency of the 3592 *HpaI* RFLP, and of mtDNA HGs L1, L2 and L3 in four Central African populations.

| **population** | N[1] | ***HpaI* + (frequency)** | ***Hpa* I - (frequency)** | N[2] | **L1** | **L2** | **L3** |
|---|---|---|---|---|---|---|---|
| Biaka Pygmies | 80 | 80 (1.000) | 0 | 114 | 112 (0.982) | 0 | 2 (0.018) |
| C.A.R. non-Pygmies | 94 | 54 (0.574) | 40 (0.426) | 95 | 26 (0.274) | 27 (0.284) | 42 (0.442) |
| Uganda | 76 | 40 (0.526) | 36 (0.474) | 92 | 35 (0.380) | 14 (0.152) | 43 (0.467) |
| Zambia | 90 | 46 (0.511) | 44 (0.489) | 96 | 37 (0.385) | 14 (0.146) | 45 (0.469) |
| **TOTAL** | **340** | **220 (0.647)** | **120 (0.353)** | **397** | **210 (0.529)** | **55 (0.139)** | **132 (0.332)** |

[1] number of individuals typed for the 3592 *HpaI* RFLP

[2] number of mtDNA HVR sequences classified into HGs



| | subHG | L1a | L1f | L1d | L1k | L1b | L1c | L1e | L2a | L2b | L2c | L2d | L3b | L3d | L3e | L3f | L3g | L3h | L3A* | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| population | N[2] | | | | | | | | | | | | | | | | | | | *h*[1] | +/- |
| BIAKA | 114 | 0.035 | | | | | 0.947 | | | | | | 0.018 | | | | | | | 0.897 | 0.016 |
| CAR | 95 | 0.074 | | | | 0.105 | 0.084 | 0.011 | 0.189 | 0.021 | 0.042 | 0.032 | 0.053 | 0.021 | 0.211 | 0.063 | 0.021 | | 0.074 | 0.994 | 0.003 |
| UGANDA | 92 | 0.185 | 0.109 | | | | 0.076 | 0.011 | 0.109 | 0.033 | | 0.011 | 0.033 | 0.076 | 0.163 | 0.022 | 0.076 | 0.065 | 0.033 | 0.995 | 0.003 |
| ZAMBIA | 96 | 0.146 | 0.031 | 0.010 | 0.010 | 0.021 | 0.156 | 0.010 | 0.125 | 0.021 | | | 0.083 | 0.042 | 0.250 | 0.073 | 0.010 | | 0.010 | 0.994 | 0.003 |
| total | 397 | 0.106 | 0.033 | 0.003 | 0.003 | 0.030 | 0.348 | 0.008 | 0.101 | 0.018 | 0.010 | 0.010 | 0.045 | 0.033 | 0.149 | 0.038 | 0.025 | 0.015 | 0.028 | 0.990 | 0.002 |
| no. hts within HG | 246 | 31 | 10 | 1 | 1 | 11 | 47 | 3 | 33 | 5 | 2 | 4 | 12 | 7 | 46 | 13 | 8 | 4 | 8 | 246 | |
| *h*[1] | 0.990 | 0.988 | 0.949 | - | - | 0.985 | 0.929 | 1.000 | 0.996 | 0.905 | 0.500 | 1.000 | 0.974 | 0.923 | 0.989 | 0.981 | 0.956 | 0.800 | 0.891 | | |

[1] genetic diversity (Nei 1987), calculated from HVRI and HVRII sequence data

[2] number of individuals

**Figure 4.1** Distribution of mtDNA subHGs in four central African populations

Using the 185 variable sites from both hypervariable regions together with 3592 *Hpa*I data, 246 unique mtDNA types were identified, and were classified into subHGs (Tables 4.3 - 4.5). The 104 mtDNA types within HG L1 are shown in Table 4.3; the 44 types within HG L2 are shown in Table 4.4; and the 98 types from HG L3 are shown in Table 4.5. The distribution of mtDNA subHGs from L1, L2 and L3 in the four central African populations is shown in fig. 4.1.

### 4.1.1.4 *Haplogroup structure*

The genetic relationships among mtDNA types within HGs L1, L2 and L3 were assessed by means of phylogenetic networks and NJ trees (figs. 4.2 - 4.7). The phylogenetic relationships among HGs L1, L2 and L3 was elucidated through the construction of a skeleton phylogenetic network using mtDNA types defined by HVRI and HVRII variation from each subHG (fig. 4.8).

### *HG L1*

In both a NJ tree (fig. 4.2) and a MJ network (fig. 4.3), the 104 mtDNA types from HG L1 clustered clearly into the seven subHGs L1a, L1b, L1c, L1d, L1e, L1f and L1k previously described by Salas et al. (2002). There was strong bootstrap support in the NJ tree for each cluster (fig. 4.2). SubHGs L1d, L1e and L1k were represented by mtDNA types belonging to their subclades only (L1d2, L1e2, and L1k1, respectively). All mtDNA types within subHG L1f in this study are suggested to belong to a new clade, L1f1, defined by an additional mutation at position 16368 (Table 4.3). The polymorphisms defining each subHG are highlighted in Table 4.3 and are also depicted in fig. 4.8.

SubHGs L1a, L1d, L1f and L1k are previously thought to have been derived from a common ancestral node defined by the 16230 mutation, from which the root of the global network is also derived (Salas et al. 2002). Data from this study incorporating HVRII variation showed that subHGs L1a and L1f could be paired on the basis of mutations at sites 189, 195, 73 and 16172, whilst subHGs L1d and L1k were paired due to mutations at sites 146 and 263 (Table

4.3, fig. 4.8). The root of HG L1, represented by a Neanderthal sequence (Krings et al. 1997, 1999), was placed in subHG L1a2 in the L1 network (fig. 4.3), and between L1a and other L1 subHGs in the NJ tree (fig. 4.2). In a network of all African mtDNA subHGs, the root was placed along the branch leading to subHGs L1a and L1f (fig. 4.8).

SubHGs L1c and L1a were particularly common in the central African populations examined, accounting for 0.657 and 0.200 of the 210 L1 samples respectively (Table 4.3, fig. 4.1). The Khoisan-specific subHGs L1d and L1k were observed only in one Zambian individual each (Table 4.3, fig. 4.1). SubHG L1f was common only in Ugandans, L1b was common in the C.A.R. non-Pygmy population, and L1e was found at very low frequency in Uganda, Zambia and the C.A.R. (Table 4.3, fig. 4.1).

Structuring of mtDNA types within subHG L1a into subclusters was apparent in the NJ tree and MJ network (figs. 4.2, 4.3). In the present study, most L1a mtDNA types belonged to L1a1a and L1a2; only one type (type 2, Table 4.3) from the more ancestral L1a* clade (Salas et al. 2002) was observed. All mtDNA types in clade L1a2, as well as the single type from L1a*, were associated with the mtDNA 9-bp deletion, whilst mtDNA types from L1a1a were not (Table 4.3). Clade L1a2 was only slightly more common and more diverse in central African populations in this study than L1a1a (L1a2 frequency = 0.060 and *h* = 0.9800; L1a1a frequency = 0.040 and *h* = 0.9667). Estimates of TMRCAs of L1a1a and L1a2 suggested that the clades are of similar ages (45 058 and 39 376 years respectively). The date calculated for L1a2 here is much earlier than the ~ 10 000 years proposed by Salas et al. (2002), although this might be affected by the different dating methods and mutation rates used.

Structuring of mtDNA types within subHGs L1c into subclusters was also apparent in the NJ tree and MJ network (figs. 4.2, 4.3). The three clades within subHG L1c (L1c1, L1c2 and L1c3) described by Salas et al. (2002) were all observed in the present study. However types from L1c1 (excluding those from L1c1a1) did not cluster together in the tree or network, and in some cases clustered with types defined as L1c* (figs. 4.2, 4.3). This suggests that the mutation at 16293 which defines L1c has occurred repeatedly, and that the definition of a clade on the basis of this mutation alone is not supported. Two new clades within L1c were also observed, and referred to as L1c4 and L1c5 (figs. 4.2, 4.3, Table 4.3). A search of the published literature showed that subHGs L1c4 and L1c5 were also present in the Biaka examined by Vigilant et al. (1991). Using HVRI and HVRII data from this study and from Vigilant et al. (1991), TMRCA of subHG L1c4 was estimated to be 7 191 years old, whilst TMRCA of subHG L1c5 was estimated to be 16 308 years old. In the present study, subHGs L1c1a1, L1c4 and L1c5 were nearly exclusively observed in the Biaka and were extremely frequent in this population, accounting for 0.465, 0.333 and 0.140 of 114 Biaka samples, respectively. SubHGs L1c2 and L1c3 were observed mostly in the non-Pygmy populations of central Africa (Table 4.3).

Table 4.3 List of 104 mtDNA types within HG L1, grouped into subHGs according to their HVR sequence variation.

| type no | subHG | 9bp del | 9bp seq | AND | 3594<br>C | 16070<br>A | 16086<br>T | 16093<br>T | 16111<br>C | 16126<br>T | 16129<br>G | 16148<br>C | 16153<br>G | 16166<br>A | 16168<br>C | 16169<br>C | 16171<br>A | 16172<br>T | 16173<br>C | 16182<br>A | 16183<br>A | 16183<br>ins | 16183<br>ins | 16184<br>C | 16187<br>C | 16188<br>C | 16189<br>T | 16190<br>C | 16209<br>T | 16213<br>G | 16214<br>C | 16218<br>C | 16223<br>C | 16230<br>A | 16234<br>C | 16239<br>C | 16241<br>A | 16243<br>T | 16255<br>G | 16256<br>C | 16264<br>C | 16265<br>A | 16270<br>C | 16274<br>G | 16278<br>C | 16287<br>C | 16290<br>C | 16291<br>C | 16293<br>A | 16294<br>C | 16311<br>T | 16319<br>G | 16320<br>C | 16325<br>T | 16327<br>C | 16354<br>C | 16355<br>C | 16362<br>T | 16368<br>T | 16399<br>A | 64<br>C | 73<br>A | 85<br>G | 89<br>T | 93<br>A | 95<br>A | 103<br>G | 114<br>C | 143<br>G | 146<br>T | 150<br>C | 151<br>C | 152<br>T | 173<br>T | 182<br>C | 183<br>A | 184<br>G | 185<br>G | 189<br>A | 194<br>C | 195<br>T | 198<br>C | 199<br>T | 200<br>A | 203<br>G | 204<br>T | 207<br>G | 225<br>G | 226<br>T | 228<br>G | 236<br>T | 247<br>G | 257<br>A | 263<br>A | 297<br>A | 302<br>ins | 302<br>ins | 315<br>ins | 357<br>A | ZAM | UGA | CAP | CAR | total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | L1a2 | yes | | | ? | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | T | A | C | . | . | . | . | . | T | G | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C | . | T | C | . | . | . | . | . | . | T | . | . | . | G | . | . | . | . | C | . | . | C | . | . | . | . | . | . | . | C | . | . | . | . | . | . | . | . | . | C | A | . | G | . | . | . | C | . | | 1 | | | 1 |
| 2 | L1a' | yes | | | T | . | . | C | . | . | A | T | . | . | . | . | . | . | . | . | G | . | . | . | T | G | C | . | . | . | . | . | T | G | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C | . | T | . | . | . | . | . | . | . | . | G | . | C | G | . | . | . | . | C | . | . | C | . | . | . | . | . | G | . | . | . | . | . | . | . | . | . | . | . | C | A | . | G | . | . | . | C | . | | 1 | | | 1 |
| 3 | L1a2 | yes | | | T | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | T | A | C | . | . | . | . | . | T | G | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C | . | T | . | . | . | . | . | . | . | T | . | . | . | G | . | . | . | . | . | . | . | C | . | . | . | . | . | G | . | . | . | . | . | . | . | . | . | . | . | C | A | . | G | . | . | . | C | . | 1 | | | | 1 |
| 4 | L1a2 | yes | | | ? | . | C | . | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | T | T | G | C | . | . | . | . | . | T | G | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C | . | T | . | . | . | . | . | . | . | T | . | . | . | G | C | . | . | . | . | . | . | C | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | . | C | A | . | G | . | . | . | C | . | | 1 | | | 1 |
| 5 | L1a2 | yes | | | T | . | C | . | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | T | T | G | C | . | . | . | . | . | T | G | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C | . | T | . | . | . | . | . | . | . | T | . | . | . | G | C | . | . | . | . | . | . | C | . | . | . | . | A | G | . | . | . | . | . | . | . | . | . | . | . | C | A | . | G | . | . | . | C | . | | 1 | | | 1 |
| 6 | L1a2 | yes | | | T | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | T | G | C | . | . | . | . | . | T | G | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C | . | T | . | . | . | . | . | . | . | T | . | . | . | G | . | . | . | . | . | . | . | C | . | . | . | . | . | G | . | . | . | . | . | . | . | . | . | . | . | C | A | . | G | . | . | C | C | . | | 1 | | | 1 |
| 7 | L1a2 | yes | | | T | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | T | G | C | . | . | . | . | . | T | G | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C | . | T | . | . | . | . | . | . | . | T | . | . | . | G | . | . | . | . | . | . | . | C | . | . | . | . | . | G | . | . | . | . | . | . | . | . | . | . | . | C | A | . | G | . | . | C | C | . | 2 | | | | 2 |
| 8 | L1a2 | yes | | | T | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | T | G | C | . | . | . | . | . | T | G | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C | . | T | . | . | . | . | . | . | . | . | . | . | . | G | . | . | . | . | . | . | . | C | . | . | . | . | . | G | . | . | . | . | . | . | . | . | . | . | . | C | A | . | G | . | . | C | C | . | | | 2 | | 2 |
| 9 | L1a2 | yes | | | T | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | T | G | C | . | . | . | . | . | T | G | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | C | . | T | . | . | . | T | . | . | . | T | . | . | . | G | . | . | . | . | . | . | . | C | . | . | . | . | . | G | . | . | . | . | . | . | . | . | . | . | . | C | A | . | G | . | . | C | C | . | 1 | | | | 1 |
| 10 | L1a2 | yes | | | T | . | . | C | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | T | G | C | . | . | . | . | . | T | G | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C | . | T | . | . | . | . | . | . | . | . | . | . | . | G | . | . | . | . | . | . | . | C | . | . | . | . | . | G | . | . | . | . | . | . | . | . | . | . | . | C | A | . | G | . | . | C | C | . | | | 2 | | 2 |
| 11 | L1a2 | yes | | | T | . | . | C | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | T | G | C | . | . | . | . | . | T | G | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C | . | T | . | . | . | . | . | . | . | T | . | . | . | G | . | . | . | . | . | . | . | C | . | . | . | . | . | G | . | . | . | . | . | . | . | . | . | . | . | C | A | . | G | . | . | C | C | . | | 1 | | | 1 |
| 12 | L1a2 | yes | | | T | . | . | C | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | T | G | C | . | . | . | . | . | T | G | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C | . | T | . | . | . | . | . | . | . | T | . | . | . | G | . | . | . | . | . | . | . | C | . | . | . | . | . | G | . | . | . | . | . | . | . | . | . | . | . | C | A | . | G | . | . | . | C | . | | | | 1 | 1 |
| 13 | L1a2 | yes | | | T | . | . | C | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | T | G | C | . | . | . | . | T | T | G | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C | . | T | . | . | . | . | . | . | . | T | . | . | . | G | T | . | . | . | . | . | . | C | . | . | . | . | . | G | . | . | . | . | . | . | . | . | . | . | . | C | A | . | G | . | . | . | C | . | | 1 | | | 1 |
| 14 | L1a2 | yes | | | T | . | . | C | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | T | G | C | . | . | . | . | . | T | G | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C | . | T | . | . | . | . | . | . | . | . | . | . | . | G | C | . | . | . | . | . | . | C | . | . | . | . | . | G | . | . | . | . | . | . | . | . | . | . | . | C | A | . | G | . | . | . | C | . | 1 | | | | 1 |
| 15 | L1a2 | yes | | | T | . | . | C | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | T | G | C | . | . | . | . | . | T | G | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C | . | T | . | . | . | . | . | . | . | T | . | . | . | G | . | . | . | . | . | . | . | C | . | . | . | . | . | G | . | . | . | . | . | . | . | A | . | . | . | C | A | . | G | . | . | . | C | . | 1 | | | | 1 |
| 16 | L1a2 | yes | | | T | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | T | G | C | . | . | . | . | . | T | G | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C | . | T | . | . | . | . | . | . | . | T | . | . | . | G | . | . | . | . | . | . | . | C | . | . | . | . | . | G | . | . | . | . | . | . | . | A | . | . | . | C | A | . | G | . | . | . | C | . | 1 | | | | 1 |
| 17 | L1a2 | yes | 1 | | T | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | T | G | C | . | . | . | . | . | T | G | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C | . | T | . | . | . | . | . | . | . | T | . | . | . | G | C | . | . | . | . | T | . | C | . | . | . | . | . | G | . | . | . | . | . | . | C | A | . | . | . | C | A | . | G | . | . | C | C | . | | 1 | | | 1 |
| 18 | L1a2 | yes | | | ? | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | T | G | C | . | . | . | . | . | T | G | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C | . | T | . | . | . | . | . | . | . | T | . | . | . | G | . | . | . | . | . | T | . | C | . | . | . | . | . | G | . | . | . | . | . | . | C | A | . | . | A | C | A | . | G | . | . | C | C | . | | 1 | | | 1 |
| 19 | L1a2 | yes | | | T | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | T | G | C | . | . | . | . | . | T | G | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C | . | T | . | . | . | . | . | . | . | T | . | . | . | G | . | . | . | . | . | . | . | C | . | . | . | . | . | G | . | . | . | . | . | . | C | A | . | . | . | C | A | . | G | . | . | . | C | . | 3 | 1 | | | 4 |
| 20 | L1a2 | yes | | | T | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | T | G | C | T | . | . | . | . | T | G | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C | . | T | . | . | . | . | . | . | . | T | . | . | . | G | . | . | . | . | . | . | . | C | . | . | . | . | . | G | . | . | . | . | . | . | C | A | . | . | . | C | A | . | G | . | . | . | C | . | 1 | | | | 1 |
| 21 | L1a1a | no | | | T | . | . | . | . | . | A | T | . | . | T | . | . | . | . | . | . | . | . | . | T | G | C | . | C | . | . | . | T | G | . | . | . | . | . | . | . | . | . | . | T | . | . | . | G | . | C | . | T | . | . | . | . | . | . | . | T | . | . | . | G | . | . | . | . | . | . | . | C | . | . | . | . | A | G | . | . | . | . | . | . | . | . | . | . | . | C | A | . | G | . | . | C | C | . | | | | 1 | 1 |
| 22 | L1a1a | no | | | T | . | . | . | . | . | A | T | . | . | T | . | . | . | . | . | . | . | . | . | T | G | C | . | C | . | . | . | T | G | . | . | . | . | . | . | . | . | . | . | T | . | . | . | G | . | C | . | T | . | . | . | . | . | . | . | . | . | . | . | G | C | . | . | . | . | . | . | C | . | . | . | . | A | G | . | . | . | . | . | . | . | . | . | . | . | C | A | . | G | . | . | . | C | . | | | | 1 | 1 |
| 23 | L1a1a | no | 2 | | T | . | . | C | . | . | A | T | . | . | T | . | . | . | . | . | . | . | . | . | T | G | C | . | C | . | . | . | T | G | . | . | . | . | . | . | . | . | . | . | T | . | . | . | G | . | C | . | T | . | . | . | . | . | . | . | . | . | . | . | G | C | . | . | . | . | . | . | C | . | . | . | . | A | G | . | . | . | . | . | . | . | . | . | . | . | C | A | . | G | . | . | . | C | . | | | | 2 | 2 |
| 24 | L1a1a | no | | | T | . | . | C | . | . | A | T | . | . | T | . | . | . | . | . | . | . | . | . | T | G | C | . | C | . | . | . | T | G | . | . | . | . | . | . | . | . | . | . | T | . | . | . | G | . | C | . | T | . | . | . | . | . | . | . | . | . | . | . | G | C | . | . | . | . | . | . | C | . | . | . | . | A | G | . | . | . | . | . | . | . | . | . | . | . | C | del | . | G | . | . | . | C | . | | | | 1 | 1 |
| 25 | L1a1a | no | 2 | | T | . | . | . | . | . | A | T | . | . | T | . | . | . | . | . | . | . | . | . | T | G | C | . | . | . | . | . | T | G | . | . | . | . | . | . | . | . | . | . | T | . | . | . | G | . | C | . | T | . | . | . | . | . | . | . | . | . | . | . | G | C | . | . | . | . | . | . | C | . | . | . | . | A | G | . | . | . | . | . | . | . | . | . | . | . | C | A | . | G | . | . | . | C | . | | 1 | | | 1 |
| 26 | L1a1a | no | | | T | . | . | C | . | . | A | T | . | . | T | . | . | . | . | . | . | . | . | . | T | G | C | . | . | . | . | . | T | G | . | . | . | . | . | . | . | . | . | . | T | . | . | . | G | . | C | . | T | . | . | . | . | . | . | . | . | . | . | . | G | C | . | . | . | . | . | T | . | . | . | . | . | A | G | . | . | . | . | . | . | . | . | . | . | . | C | A | . | G | . | . | . | C | . | 1 | | | | 1 |
| 27 | L1a1a | no | 2 | | T | . | . | . | . | . | A | T | . | . | T | . | . | . | . | . | . | . | . | . | T | G | C | . | . | . | . | . | T | G | . | . | . | . | . | . | . | . | . | . | T | . | . | . | G | . | C | . | T | . | . | . | . | . | . | . | . | . | . | . | G | C | A | . | . | . | . | . | . | . | . | . | . | A | G | . | . | . | . | . | . | . | . | . | . | . | C | A | . | G | . | . | C | C | . | | 1 | | | 1 |
| 28 | L1a1a | no | | | T | . | . | . | . | . | A | T | . | . | T | . | . | . | . | . | . | . | . | . | T | G | C | . | . | . | . | . | T | G | . | . | . | . | . | . | . | . | . | . | T | . | . | . | G | . | C | . | T | . | . | . | . | . | . | . | . | . | . | . | G | C | . | . | . | . | . | . | . | . | . | . | . | A | G | . | . | . | . | . | . | . | . | . | . | . | C | A | . | G | . | . | C | C | . | 2 | 2 | | | 4 |
| 29 | L1a1a | no | | | T | . | . | . | . | . | A | T | . | . | T | . | . | . | . | . | . | . | . | . | T | G | C | . | . | . | . | . | T | G | . | . | . | . | . | . | . | . | . | A | T | . | . | . | . | . | C | . | T | . | . | . | . | . | . | . | . | . | . | . | G | C | . | . | . | . | . | . | . | . | . | . | . | A | G | . | . | . | . | . | . | . | . | . | . | . | C | A | . | G | . | . | . | C | . | | | | 1 | 1 |
| 30 | L1a1a | no | | | T | . | . | C | . | . | . | T | . | . | T | . | . | . | . | . | . | . | . | . | T | G | C | . | . | . | . | . | T | G | . | . | . | . | . | . | . | . | . | . | . | T | . | . | G | . | C | . | T | . | . | . | . | . | . | . | . | . | . | . | G | . | . | . | . | . | . | . | . | . | . | . | . | A | G | . | . | . | . | . | . | . | . | . | . | . | C | A | . | G | . | . | . | C | . | | 1 | | | 1 |
| 31 | L1a1a | no | 2 | | T | . | . | C | . | . | . | T | . | . | T | . | . | . | . | . | . | . | . | . | T | G | C | . | . | . | . | . | T | G | . | . | . | . | . | . | . | . | . | . | . | T | . | . | G | . | C | . | T | . | . | . | . | . | . | . | . | . | . | . | G | . | . | . | . | . | . | . | C | . | . | . | . | A | G | . | . | . | . | . | . | . | . | . | . | . | C | A | . | G | . | . | C | C | . | | 2 | | | 2 |
| 32 | L1b | no | | | T | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | C | . | . | . | . | . | T | . | . | . | . | . | . | . | T | . | T | . | T | . | . | . | . | . | C | . | . | . | . | . | . | C | . | . | . | G | . | . | . | . | . | . | . | . | . | . | C | . | T | . | . | T | G | . | C | T | C | . | . | . | . | . | . | . | . | A | . | G | . | . | . | C | G | | | | 1 | 1 |
| 33 | L1b | no | | | T | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | C | . | . | . | . | . | T | . | . | . | . | . | . | . | T | . | T | . | T | . | T | . | . | . | C | . | . | . | . | . | . | . | . | . | . | G | . | . | . | . | . | . | . | . | . | . | C | . | T | . | . | T | G | . | C | T | C | . | . | . | . | . | . | . | . | A | . | G | . | . | . | C | G | | | | 1 | 1 |
| 34 | L1b | no | | | T | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | C | . | . | . | . | . | T | . | . | . | . | . | . | . | T | . | T | . | T | . | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | G | . | . | . | . | . | . | . | . | T | . | C | . | T | . | . | T | G | . | C | T | C | . | . | . | . | . | . | . | . | A | . | G | . | . | . | C | G | | | | 1 | 1 |
| 35 | L1b | no | | | T | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | C | . | . | . | . | . | T | . | . | . | . | . | . | . | T | . | T | . | T | . | . | . | . | . | C | . | . | . | . | . | . | C | . | . | . | G | . | . | . | . | . | T | . | . | . | . | C | . | T | . | . | T | G | . | C | . | . | . | . | . | . | . | . | . | . | A | . | G | . | . | C | C | G | | | | 1 | 1 |
| 36 | L1b1 | no | | | T | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | C | . | . | . | . | . | T | . | . | . | . | . | . | . | T | . | T | . | T | . | . | . | G | . | C | . | . | . | . | . | . | . | . | . | . | G | . | . | . | . | . | . | . | . | . | . | C | . | T | . | . | T | G | . | C | . | . | . | . | . | . | . | . | . | . | A | . | G | G | . | . | C | G | 1 | | | | 1 |
| 37 | L1b | no | | | T | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | C | . | . | . | . | . | T | . | . | . | . | . | . | T | T | . | T | . | T | . | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | G | . | . | . | . | . | . | . | . | . | T | C | . | T | . | . | T | G | . | C | . | . | . | . | . | . | . | . | . | . | A | . | G | . | . | . | C | G | | | | 1 | 1 |
| 38 | L1b | no | | | T | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | C | . | . | . | . | . | T | . | . | . | . | . | . | . | T | . | T | . | T | . | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | G | . | . | . | . | . | . | . | . | . | T | C | . | T | . | . | T | . | . | C | . | . | . | . | . | . | . | . | . | . | A | . | G | . | . | . | C | G | 1 | | | | 1 |
| 39 | L1b1 | no | | | T | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | C | . | . | . | . | . | T | . | . | . | . | . | . | . | T | . | T | . | T | . | . | . | G | . | C | . | . | . | . | . | . | . | . | . | . | G | . | . | . | . | . | . | . | . | . | . | C | . | . | . | . | T | . | . | C | . | . | . | . | . | . | . | . | . | . | A | . | G | . | . | . | C | G | | | | 1 | 1 |
| 40 | L1b1 | no | | | T | . | . | . | T | C | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | C | . | . | . | . | . | T | . | . | T | . | . | . | . | . | . | T | . | T | . | . | . | G | . | C | . | . | . | . | . | . | . | . | . | . | G | . | . | . | . | . | . | . | C | . | . | C | . | T | . | . | T | G | . | . | . | . | . | . | . | . | . | . | . | . | A | . | G | . | . | . | C | G | | | | 2 | 2 |
| 41 | L1b1 | no | | | T | . | . | . | . | C | . | . | A | . | . | . | . | . | . | . | . | . | . | . | T | . | C | . | . | . | . | . | T | . | . | . | . | . | . | . | T | . | T | . | T | . | . | . | G | . | C | . | . | . | . | . | . | . | . | . | . | G | . | . | . | . | . | . | . | . | . | . | C | . | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | G | G | . | . | . | C | . | | | | 1 | 1 |
| 42 | L1b | no | | | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | del | del | . | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | C | . | . | T | . | T | . | T | T | . | . | . | . | C | . | . | . | . | . | . | . | . | ? | . | G | . | . | . | . | . | . | . | C | . | . | C | . | T | . | . | . | . | . | C | . | . | . | A | . | . | A | C | . | . | del | . | G | . | . | . | C | G | | | | 1 | 1 |
| 43 | L1d2 | no | | | ? | . | . | C | . | C | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | C | . | . | . | . | . | T | G | T | . | . | C | . | . | . | . | . | . | . | . | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | G | . | . | . | . | . | . | . | C | . | . | C | . | . | . | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | . | G | . | . | . | C | . | 1 | | | | 1 |
| 44 | L1e2 | no | | | T | . | . | . | . | . | A | T | . | G | . | . | . | . | . | . | C | . | . | . | T | . | C | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | G | . | C | . | . | . | . | . | T | C | . | G | . | G | . | . | . | . | . | . | . | C | . | . | C | . | T | . | . | . | . | . | . | . | . | G | . | . | . | . | . | . | . | A | . | G | . | . | . | C | . | 1 | | | | 1 |
| 45 | L1e2 | no | | | T | . | . | . | . | . | A | T | . | G | . | . | . | . | . | . | C | . | . | . | T | . | C | . | . | . | . | . | T | . | . | . | . | . | A | . | . | . | . | . | T | . | . | . | . | . | C | . | . | . | . | . | T | C | . | . | . | G | . | . | . | . | . | . | . | . | . | . | C | . | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . | G | . | . | . | C | . | | | | 1 | 1 |
| 46 | L1e2 | no | | | T | . | . | . | . | . | A | T | . | G | . | . | . | . | . | . | C | . | . | . | T | . | C | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | C | . | . | . | . | . | T | C | . | . | . | G | . | . | . | . | . | . | . | . | . | . | C | . | T | . | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | A | . | G | . | . | . | C | . | | 1 | | | 1 |
| 47 | L1k1 | no | | | T | . | . | C | . | . | . | . | . | . | . | . | G | . | . | . | . | . | . | . | T | . | C | . | . | . | T | . | . | . | . | . | C | . | . | . | . | . | . | . | T | . | . | A | . | . | C | . | . | . | . | . | . | . | . | . | . | G | . | . | . | . | . | . | . | . | . | . | C | . | . | . | . | . | G | . | C | . | C | G | . | . | . | . | . | . | . | A | . | . | . | . | . | C | . | 1 | | | | 1 |
| 48 | L1f1 | no | | | T | . | . | . | . | . | A | . | . | G | . | T | . | C | T | . | . | . | . | . | T | . | C | . | . | . | . | . | T | G | . | T | . | . | . | . | . | . | . | . | T | . | . | . | . | . | C | . | . | . | T | . | . | . | C | . | . | . | . | . | . | . | . | . | A | C | . | . | C | . | . | . | . | A | G | . | C | . | . | . | . | . | . | . | . | . | . | A | . | G | . | . | . | C | . | | 1 | | | 1 |
| 49 | L1f1 | no | | | T | . | . | . | . | . | A | . | . | . | . | T | . | C | T | . | . | . | . | . | T | . | C | . | . | . | . | . | T | G | . | T | . | . | . | . | . | . | . | . | T | . | . | . | . | . | C | . | . | . | T | . | . | . | C | . | . | . | . | . | . | . | . | . | A | C | . | . | C | . | . | . | . | A | G | . | C | . | . | . | . | . | . | . | . | . | . | A | . | G | . | C | C | C | . | | 3 | | | 3 |
| 50 | L1f1 | no | | | T | . | . | . | . | . | A | . | . | . | . | T | . | C | T | . | . | . | . | . | T | . | C | . | . | A | . | . | T | G | . | T | . | . | . | . | . | . | . | . | T | . | . | . | . | . | C | . | . | . | T | . | . | C | C | . | . | . | . | . | . | . | . | . | A | C | . | . | C | . | . | . | . | A | G | . | C | . | . | . | . | . | . | . | . | . | . | A | . | G | . | . | . | C | . | | 1 | | | 1 |
| 51 | L1f1 | no | | | T | . | . | . | . | . | A | . | . | . | . | T | . | C | T | . | . | . | . | . | T | . | C | . | . | . | . | . | T | G | . | T | . | . | . | . | . | . | . | . | T | . | . | . | . | . | C | . | . | . | T | . | . | . | C | . | . | . | . | . | . | . | . | . | A | C | . | . | C | . | . | . | . | A | G | . | . | . | . | . | . | . | . | . | . | . | . | A | . | G | . | . | . | C | . | | 1 | | | 1 |
| 52 | L1f1 | no | | | . | . | . | . | . | . | A | . | . | . | . | T | . | C | . | . | . | . | . | . | T | . | C | . | . | . | . | T | T | G | . | . | . | . | . | . | . | . | . | . | T | . | . | T | . | . | C | . | . | . | T | T | . | . | C | . | . | . | . | . | . | . | . | . | . | C | . | . | C | . | . | G | A | A | G | . | C | . | . | . | . | . | A | . | . | . | . | A | . | G | . | . | C | C | . | | 1 | | | 1 |
| 53 | L1f1 | no | | | T | . | . | . | . | . | A | . | . | . | . | T | . | C | . | . | . | . | . | . | T | . | C | . | . | . | . | . | T | G | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | C | . | . | . | T | . | . | . | C | . | . | . | . | . | . | . | . | . | . | C | . | . | C | . | . | . | . | A | G | . | . | . | . | . | . | C | A | . | . | . | . | A | . | G | . | . | C | C | . | | 1 | | | 1 |
| 54 | L1f1 | yes | 1 | | T | . | . | . | . | . | . | . | . | . | . | T | . | C | . | . | . | . | . | . | T | . | C | . | . | . | . | . | T | G | . | . | . | . | . | . | . | G | . | . | T | . | . | . | . | . | C | A | . | . | T | T | . | . | C | . | . | . | . | . | G | C | . | . | . | . | . | T | C | C | . | . | . | A | G | . | . | . | . | . | . | . | . | . | . | . | . | A | . | G | . | . | . | C | . | | 2 | | | 2 |
| 55 | L1f1 | no | | | T | . | . | . | . | . | A | . | . | . | . | T | . | C | . | . | del | C | . | . | . | . | C | . | . | . | . | . | T | G | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | C | . | . | . | T | T | . | . | . | ? | . | . | . | . | . | . | . | . | . | C | . | . | C | . | . | . | . | A | G | . | . | . | . | . | . | . | . | . | . | . | . | A | . | G | . | C | C | C | . | 1 | | | | 1 |
| 56 | L1f1 | no | | | T | . | . | . | . | . | A | . | . | . | . | T | . | C | . | . | del | C | . | . | . | . | C | . | . | . | . | . | T | G | . | . | . | . | . | . | . | . | T | . | T | . | . | . | . | . | C | . | . | . | T | T | . | . | C | . | . | G | T | . | . | . | . | . | . | C | . | . | C | . | . | . | . | A | G | . | . | . | . | . | . | . | . | . | . | . | . | A | . | G | . | C | C | C | . | 1 | | | | 1 |
| 57 | L1f1 | no | | | ? | G | . | . | . | . | A | . | . | . | . | T | . | C | . | del | del | C | C | . | . | . | . | . | . | . | . | . | T | G | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | T | C | . | . | . | T | . | . | . | C | G | . | . | . | . | . | . | . | . | . | C | . | . | C | . | . | . | . | A | G | T | . | T | . | . | . | C | . | . | . | . | . | A | . | G | . | C | C | C | . | 1 | | | | 1 |

**Table 4.3 continued**: List of 104 mtDNA types within HG L1, grouped into subHGs according to their HVR sequence variation.

|  |  | 9bp | 9bp |  | 3594 | 16017 | 16018 | 16051 | 16071 | 16078 | 16086 | 16093 | 16114 | 16129 | 16140 | 16145 | 16163 | 16169 | 16181 | 16183 | 16183 | 16184 | 16187 | 16189 | 16213 | 16214 | 16215 | 16223 | 16234 | 16241 | 16249 | 16258 | 16260 | 16261 | 16265 | 16274 | 16278 | 16286 | 16291 | 16293 | 16294 | 16301 | 16311 | 16320 | 16360 | 16364 | 16368 | 16390 | 16399 | 73 | 81 | 89 | 93 | 95 | 124 | 146 | 151 | 152 | 182 | 183 | 186 | 189 | 194 | 195 | 198 | 200 | 204 | 236 | 247 | 248 | 256 | 263 | 264 | 280 | 291 | 291 | 297 | 302 | 302 | 315 | 316 | 366 | 372 | 390 | 395 |  |  |  |  |  |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| type no | subHG | del | seq | AND | C | T | T | A | C | A | T | T | C | G | T | G | A | C | A | A | ins | C | C | T | G | C | A | C | C | A | T | A | C | C | A | G | C | C | C | A | C | C | T | C | C | C | T | G | A | A | G | T | A | A | G | T | C | T | C | A | C | A | C | T | C | A | T | T | G | A | C | A | C | C | A | ins | A | ins | ins | ins | G | G | T | A | A | ZAM | UGA | CAP | CAR | total |
| 58 | L1c2 | no |  |  | T | . | . | . | . | . | . | . | . | A | . | . | . | . | G | . | . | . | T | C | . | . | . | T | . | . | . | . | . | . | C | . | T | G | . | . | T | . | C | . | T | . | . | . | . | G | . | . | . | . | . | . | T | C | T | . | A | C | . | C | T | . | . | . | A | . | . | G | . | . | . | . | G | . | . | C | A | . | . | . | . |  | 1 |  |  | 1 |
| 59 | L1c2 | no |  |  | T | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | T | C | . | . | . | T | . | . | . | . | . | . | C | . | T | G | . | . | T | . | C | . | T | . | . | . | . | G | . | . | . | . | . | . | T | C | T | . | A | C | . | C | T | . | . | . | A | . | . | G | T | . | . | . | G | . | . | C | A | . | . | . | . | 1 |  |  |  | 1 |
| 60 | L1c2 | no |  |  | T | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | T | C | . | . | . | T | . | . | . | . | . | . | C | . | T | G | . | . | T | . | C | . | T | . | . | . | . | G | . | . | . | . | . | . | T | C | T | . | A | C | . | C | T | . | C | . | A | . | . | G | . | . | . | . | G | . | C | C | A | . | . | . | . | 1 |  |  |  | 1 |
| 61 | L1c2 | no |  |  | T | . | . | . | T | . | . | . | . | A | . | A | . | . | . | . | . | . | T | C | A | . | . | T | T | . | . | . | . | . | C | . | T | G | . | . | T | . | C | . | T | . | . | . | . | G | . | . | . | . | . | . | T | C | T | . | A | C | . | C | T | . | . | . | A | . | . | G | . | . | . | . | G | . | . | C | A | . | . | . | . | 3 |  |  |  | 3 |
| 62 | L1c2 | no |  |  | ? | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | T | C | . | T | . | T | . | . | . | . | . | . | C | . | T | A | T | . | T | . | C | . | T | . | . | . | . | G | . | . | . | . | . | . | T | C | T | . | A | C | . | C | T | . | . | . | A | . | . | G | . | . | . | . | G | . | . | C | A | . | . | . | . |  | 1 |  |  | 1 |
| 63 | L1c2 | no |  |  | T | . | in | . | . | . | C | . | . | A | . | . | . | . | . | . | . | . | T | C | . | . | . | T | . | . | . | . | . | . | C | . | . | G | . | . | T | . | C | . | T | . | . | . | . | G | . | . | . | . | . | . | T | C | T | . | A | C | . | C | T | G | . | . | A | . | . | G | . | . | . | . | G | . | . | C | A | . | . | . | . | 1 |  |  |  | 1 |
| 64 | L1c2 | no |  |  | T | . | . | . | . | G | . | . | . | A | . | . | . | . | . | . | . | . | T | C | . | . | . | T | . | . | . | . | . | . | C | . | . | A | . | . | T | . | C | T | T | . | . | . | . | G | . | . | . | . | . | . | T | C |  | . | A | C | . | C | T | . | . | . | A | . | . | G | . | . | . | . | G | . | . | C | A | . | . | . | . |  | 1 |  |  | 1 |
| 65 | L1c2 | no |  |  | T | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | T | C | . | . | . | T | . | . | . | . | . | . | C | . | T | G | . | . | T | . | C | T |  | . | . | . | . | G | . | . | . | . | . | . |  | C | T | . | A | C | . | C | T | . | . | . | A | . | . | G | . | . | . | . | G | . | . | C | A | . | . | . | . |  |  |  | 3 | 3 |
| 66 | L1c2 | no |  |  | T | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | T | C | . | . | . | T | . | . | . | . | . | . | C | . | T | G | . | . | T | . | C | T |  | . | . | . | . | G | . | . | . | . | . | . | T | C | T | . | A | C | . | C | T | . | . | . | A | . | . | G | . | . | . | . | G | . | C | C | A | . | . | G | . |  |  |  | 1 | 1 |
| 67 | L1c2 | no |  |  | T | . | . | . | . | . | . | . | . | A | . | . | . | T | . | . | . | . | T | C | . | . | . | T | . | . | . | . | . | . | C | . | T | G | . | . | T | . | C | . |  | . | . | . | . | G | . | . | . | . | . | . | T | C | T | . | A | C | . | C | T | . | . | . | A | . | . | G | . | . | . | . | G | . | C | C | A | . | . | . | . | 1 |  |  |  | 1 |
| 91 | L1c1 | no |  |  | T | C | . | . | . | . | . | . | . | A | . | . | G | . | . | . | . | . | T | C | . | . | . | T | . | . | . | . | T | . | . | . | T | . | . | G | T | . | C | . | T | . | . | . | . | G | . | . | . | . | . | . | T | C | T | . | A | C | . | . |  | . | . | . | A | . | . | G | . | . | . | . |  | . | . | C | A | . | . | . | . |  | 1 |  |  | 1 |
| 92 | L1c1 | no |  |  | T | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | T | C | . | . | . | T | . | . | . | . | . | . | . | . | T | . | . | G | T | . | C | . | T | . | . | . | . | G | . | . | . | . | . | . |  | C | T | . | A | C | . | C | . | . | . | . | A | G | . | G | . | . | . | . |  | . | . | C | A | . | . | . | . |  | 1 |  |  | 1 |
| 94 | L1c1 | no |  |  | T | . | . | . | . | . | C | . | . | A | . | . | . | . | . | . | . | . | T | C | . | . | . | T | . | G | . | . | . | . | . | . | T | . | . | G | T | . | C | . | T | . | . | . | . | G | . | . | . | . | . | . | T | C | T | . | A | C | . | C | T | . | . | . | A | . | . | G | . | . | . | . | G | . | . | C | A | . | . | . | . | 2 |  |  |  | 2 |
| 95 | L1c1 | no |  |  | ? | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | T | C | . | . | . | T | . | . | . | . | . | . | . | . | T | . | . | G | T | . | C | . | T | . | . | . | . | G | . | . | . | . | . | . | T | C | T | . | A | C | . | C | T | G | . | . | A | . | . | G | . | . | . | . | G | . | . | C | A | . | . | . | . |  | 1 |  |  | 1 |
| 96 | L1c1 | no |  |  | T | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | T | C | . | . | . | T | . | . | . | . | . | . | . | . | T | . | . | G | T | . | C | . | T | . | . | . | . | G | . | . | . | . | . | . | T | C | T | . | A | T | . | C | T | . | . | . | A | . | . | G | . | . | . | . | G | . | . | C | A | . | . | . | . | 1 |  |  |  | 1 |
| 68 | L1c1a | no |  |  | ? | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | T | C | . | . | . | T | . | . | . | . | . | . | . | A | T | . | . | G | T | . | C | . | T | . | . | . | . | G | . | . | . | . | . | . | T | C |  | . | A | C | . | C |  | . | C | . | A | . | . | G | . | . | . | . | G | . | . | C | A | . | . | . | . |  |  | 1 |  | 1 |
| 69 | L1c1a1 | no |  |  | T | . | . | . | . | . | . | . | . |  | . | . | . | . | . | . | . | . | T | C | . | T | . | . | T | . | C | . | . | . | . | A | T | . | . | . | T | . | C | . | T | . | . | . | . | G | . | . | . | . | . | . | T | C |  | . | A | C | . | C |  | . | C | . | A | . | . | G | . | . | . | . | G | . | . | C | A | . | . | . | . |  |  | 3 | 1 | 4 |
| 70 | L1c1a1 | no |  |  | ? | . | . | . | . | . | . | . | . |  | . | . | . | . | . | . | . | . | T | C | . | T | . | . | T | . | C | . | . | . | . | A | T | . | . | . | T | . | C | . | T | . | . | . | . | G | . | . | . | . | . | . | T | C |  | . | A | C | . | C |  | . | C | . | A | . | . | G | . | . | . | . | G | . | . | C | A | . | . | . | G |  |  | 1 |  | 1 |
| 71 | L1c1a1 | no |  |  | T | . | . | . | . | . | . | . | . |  | . | . | . | . | . | . | . | . | T | C | . | T | . | . | T | . | C | . | . | . | . | A | T | . | . | . | T | . | C | . | T | . | . | . | . | G | . | . | . | . | . | . | T | C |  | . | A | C | . | C |  | . | C | . | A | . | . | G | . | . | . | . | G | . | . | C | A | . | . | . | . |  |  | 1 |  | 1 |
| 72 | L1c1a1 | no |  |  | T | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | T | C | . | T | . | . | T | . | C | . | . | . | . | A | T | . | . | . | T | . | C | . | T | . | . | . | . | G | . | . | . | . | . | . | T | C |  | . | A | C | . | C |  | . | C | . | A | . | . | G | . | . | . | . | G | . | . | C | A | . | . | . | . |  |  | 5 |  | 5 |
| 73 | L1c1a1 | no |  |  | T | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | T | C | . | T | . | . | T | . | C | . | . | . | . | A | T | . | . | . | T | . | C | . | T | . | . | . | . | G | . | . | . | . | . | . | T | C |  | G | A | C | . | C |  | . | C | . | A | . | . | G | . | . | . | . | G | . | . | C | A | . | . | . | . |  |  | 3 |  | 3 |
| 74 | L1c1a1 | no |  |  | T | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | T | C | . | T | . | . | T | . | C | . | . | . | . | A | T | . | . | . | T | . | C | . | T | . | . | . | . | G | . | . | . | . | . | . | T | C |  | . | A | C | . | C | . | . | . | . | A | . | . | G | . | . | . | . | G | . | . | C | A | . | . | . | . |  |  | 1 |  | 1 |
| 75 | L1c1a1 | no |  |  | ? | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | T | C | . | T | . |  | T | . | C | . | . | . | . | A | T | . | . | . | T | . | C | . | T | . | . | . | . | G | . | . | . | . | . | . | T | C |  | . | A | C | . | C | . | . | . | . | A | . | . | G | . | . | . | . | G | . | C | C | A | . | . | . | G |  |  | 1 |  | 1 |
| 76 | L1c1a1 | no |  |  | T | . | . | G | . | . | . | . | . | A | . | . | . | . | . | . | . | . | T | C | . | T | . | . | T | . | C | G | . | . | . | A | T | . | . | . | T | . | C | . | T | . | . | . | . | G | . | . | . | . | . | . | T | C |  | . | A | C | . | C | . | . | . | . | A | . | . | G | . | . | . | . | G | . | . | C | A | . | . | . | . |  |  | 5 |  | 5 |
| 77 | L1c1a1 | no |  |  | T | . | . | G | . | . | . | . | . | A | . | . | . | . | . | . | . | . | T | C | . | T | . | . | T | . | C | G | . | . | . | A | T | . | . | G | T | . | C | . |  | . | . | . | . | G | . | . | . | . | . | . | T | C |  | . | A | C | . | C | . | . | . | . | A | . | . | G | . | . | . | . | G | . | . | C | A | . | . | . | . |  |  | 1 |  | 1 |
| 78 | L1c1a1 | no |  |  | T | . | . | G | . | . | . | . | . | A | . | . | . | . | . | . | . | . | T | C | . | T | . |  | T | . | C | G | . | . | . | A | T | . | . | G | T | . | C | . | T | . | . | . | . | G | . | . | . | . | . | . | T | C |  | . | A | C | . | C | T | . | . | . | A | . | . | G | . | . | . | . | G | . | . | C | A | . | . | . | . |  |  | 3 |  | 3 |
| 79 | L1c1a1 | no |  |  | T | . | . | G | . | . | . | . | . | A | . | . | . | . | . | . | . | . | T | C | . | T | . | . | T | . | C | G | . | . | . | A | T | . | . | G | T | . | C | . | T | . | . | . | . | G | . | . | . | . | . | . | T | C |  | . | A | C | . | C | . | . | . | . | A | . | . | G | . | . | . | . | G | . | . | C | A | . | . | . | . |  |  | 25 |  | 25 |
| 80 | L1c1a1 | no |  |  | T | . | . | G | . | . | . | . | . | A | . | . | . | . | . | . | . | . | T | C | . | T | . | . | T | . | C | G | . | . | . | A | T | . | . | G | T | . | C | . | T | T | . | . | . | G | . | . | . | . | . | . | T | C |  | . | A | C | . | C | . | . | . | . | A | . | . | G | . | . | . | . | G | . | C | C | A | . | . | . | . |  |  | 1 |  | 1 |
| 81 | L1c1a1 | no |  |  | T | . | . | G | . | . | . | . | . | A | . | . | . | . | . | . | . | . | T | C | . | T | . | . | T | . | C | G | . | . | . | A | T | . | . | G | T | . | C | . | T | . | . | . | . | G | A | . | . | . | . | . | T | C |  | . | A | C | . | C | . | . | . | . | A | . | . | G | . | . | . | . | G | . | . | C | A | . | . | . | . |  |  | 1 |  | 1 |
| 82 | L1c1a1 | no |  |  | T | . | . | G | . | . | . | . | . | A | . | . | . | . | . | . | . | . | T | C | . | T | . |  | T | . | C | G | . | . | . | A | T | . | . | G | T | . | C | . | T | . | . | . | . | G | . | . | . | . | A | . | T | C |  | . | A | C | . | C | . | . | . | . | A | . | . | G | . | . | . | . | G | . | . | C | A | . | . | . | . |  |  | 1 |  | 1 |
| 83 | L1c3 | no |  |  | T | . | . | . | . | . | . | . | . | A | . | . | . | . | . | del | C | . |  | C | . | . | G | T | . | . | . | . | . | . | . | . | T | . | . | . | T | . | C | . | T | . | C | . | . | G | . | . | . | . | . | . | T | C | T | . | A | C | . | . | . | . | . | . | A | . | . | G | . | . | . | . |  | . | C | C | A | . | . | . | . | 1 |  |  |  | 1 |
| 84 | L1c3 | no |  |  | T | . | . | . | . | . | . | C | T | A | . | . | . | . | . | del | C | . |  | C | . | . | G | T | . | . | . | . | . | . | . | . | T | . | . | . | T | . | C | . | T | . | . | A | . | G | . | . | . | . | . | . | T | C | T | . | A | C | . | . | . | . | . | . | A | . | . | G | . | . | . | . |  | . | C | C | A | . | . | . | . | 1 |  |  |  | 1 |
| 85 | L1c3 | no |  |  | T | . | . | . | . | . | . | . | . | A | . | . | . | . | . | del | . | . |  | A | . | . | G | T | . | . | . | . | . | . | . | . | T | . | . | . | T | . | C | . | T | . | . | . | . | G | . | . | . | . | . | . | T | C | T | . | A | C | . | . | . | . | . | . | A | . | . | G | . | . | . | . |  | . | . | C | A | . | . | . | . |  |  |  | 1 | 1 |
| 86 | L1c3 | no |  |  | T | . | . | . | . | . | . | . | . |  | . | . | . | . | . | del | C | . |  | C | . | . | G | T | . | . | . | . | . | . | . | . | T | . | . | . | T | . | C | . | T | . | . | . | ? | G | . | . | . | . | . | . | T | C | T | . | A | C | . | . | . | . | . | . | A | . | . | G | . | . | . | . |  | C | C | C | A | . | . | . | . | 1 |  |  |  | 1 |
| 87 | L1c3 | no |  |  | T | . | . | . | . | . | . | . | . | A | . | . | . | . | . | del | C | . |  | C | . | . |  | T | . | . | . | . | . | . | . | . | T | . | . | . | T | . | C | . | T | . | . | . | ? | G | . | . | . | . | . | . | T | C | T | . | A | C | . | . | . | . | . | . | A | . | . | G | . | . | . | . |  | . | C | C | A | . | . | . | . | 1 |  |  |  | 1 |
| 88 | L1c3 | no |  |  | T | . | . | . | . | . | . | . | . | A | . | . | . | . | . | del | C | . |  | C | . | . |  | T | . | . | . | . | . | . | . | . | T | . | . | . | T | . | C | . | T | . | . | . | ? | G | . | . | . | . | . | . |  | C | T | . | A | C | . | . | . | . | . | . | A | . | . | G | . | . | . | . |  | . | C | C | A | . | . | . | . |  |  |  | 1 | 1 |
| 89 | L1c* | yes |  |  | T | . | . | . | . | . | . | . | G | A | . | . | . | . | . | . | . | . | T | C | . | . | . | T | . | . | . | . | . | T | . | . | T | . | . | . |  | . | C | . | T | . | . | . | . | G | . | . | . | . | . | . | T | C | T | . | A | C | . | . | . | . | . | . | A | . | . | G | . | . | del | T | G | . | C | C | A | . | . | . | . |  | 1 |  |  | 1 |
| 90 | L1c* | yes | 1 |  | T | . | . | . | . | . | . | . | G | A | . | . | . | . | . | . | . | . | T | C | . | . | . | T | . | . | . | . | . | T | . | . | T | . | . | . |  | . | C | . | T | . | . | . | . | G | . | . | . | . | . | . |  | C | T | . | A | C | . | . | . | . | . | . | A | . | . | G | . | . | del | T | G | . | . | C | A | . | . | . | . | 1 |  |  |  | 1 |
| 93 | L1c* | no |  |  | T | . | . | . | . | . | C | . | . | A | . | . | . | . | . | . | . | . | T | C | . | . | . | T | . | G | . | . | . | . | . | A | T | . | T | . | T | . | C | . | T | . | . | . | . | G | . | . | . | . | . | . | T | C | T | . | A | C | . | C | T | . | . | . | A | . | . | G | . | . | . | . | G | . | . | C | A | . | . | . | . |  |  | 1 |  | 1 |
| 97 | L1c4 | no |  |  | T | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | T | C | . | . | . | T | . | . | . | . | . | . | . | A | T | . | . | G | T | . | C | . | T | . | . | . | . | G | . | . | G | C | . | . |  | C | T | . | A | C | . | C | . | . | . | C | A | . | . | G | . | . | . | . | G | . | . | C | A | . | . | . | . |  |  | 9 |  | 9 |
| 98 | L1c4 | no |  |  | T | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | T | C | . | . | . | T | . | . | . | . | . | . | . | A | T | . | . | G | T | . | C | . | T | . | . | . | . | G | . | . | G | C | . | . |  | C | T | . | A | C | . | C | . | . | . | C | A | . | . | G | . | . | . | . | G | . | . | C | A | . | C | . | . |  |  | 6 |  | 6 |
| 99 | L1c4 | no |  |  | T | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | T | C | . | . | . | T | . | . | . | . | . | . | . | A | T | . | . | G | T | . | C | . | T | . | . | . | . | G | . | . | G | C | . | . |  | C | T | . | A | C | . | C | . | . | . | C | A | . | . | G | . | . | . | . | G | . | . | C | A | A | C | . | . |  |  | 1 |  | 1 |
| 100 | L1c4 | no |  |  | T | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | T | C | . | . | . | T | . | . | . | . | . | . | . | A | T | . | . | G | T | . | C | . | T | . | . | . | . | G | . | C | G | C | . | . |  | . | T | . | A | C | . | C | . | . | . | C | A | . | . | G | . | . | . | . | G | . | . | C | A | . | . | . | . |  |  |  | 1 | 1 |
| 101 | L1c4 | no |  |  | T | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | T | C | . | . | . | T | . | . | . | . | . | . | . | A | T | . | . | G | T | . | C | . | T | . | . | . | . | G | . | . | G | C | . | . |  | . | T | . | A | C | . | C | . | . | . | C | A | . | . | G | . | . | . | . | G | . | C | C | A | . | . | . | . |  |  | 22 |  | 22 |
| 102 | L1c5 | no |  |  | T | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | T | T | C | . | . | . | T | . | . | . | . | . | . | . | . | T | . | . | . | T | T | C | . | T | . | . | . | . | G | . | . | . | . | . | C |  | C | T | . | A | C | . | C | . | . | . | . | A | . | T | G | . | . | del | . | G | . | . | C | A | . | . | . | . |  |  | 7 |  | 7 |
| 103 | L1c5 | no |  |  | T | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | T | T | C | . | . | . | T | . | . | . | . | . | . | . | . | T | . | . | . | T | T | C | . | T | . | . | . | . | G | . | . | . | . | . | C |  | C | T | . | A | C | T | C | . | . | . | . | A | . | T | G | . | . | del | . | G | . | . | C | A | . | . | . | . |  |  | 8 |  | 8 |
| 104 | L1c5 | no |  |  | T | . | . | . | . | . | . | C | . | A | C | . | . | . | . | . | . | T | T | C | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | T | T | C | . | T | . | . | A | . | G | . | . | . | . | . |  |  | C | T | . | A | C | . | C | . | . | . | . | A | . | . | G | . | T | . | . | G | . | . | C | A | . | . | . | . |  |  | 1 |  | 1 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | total | 37 | 35 | 112 | 26 | 210 |

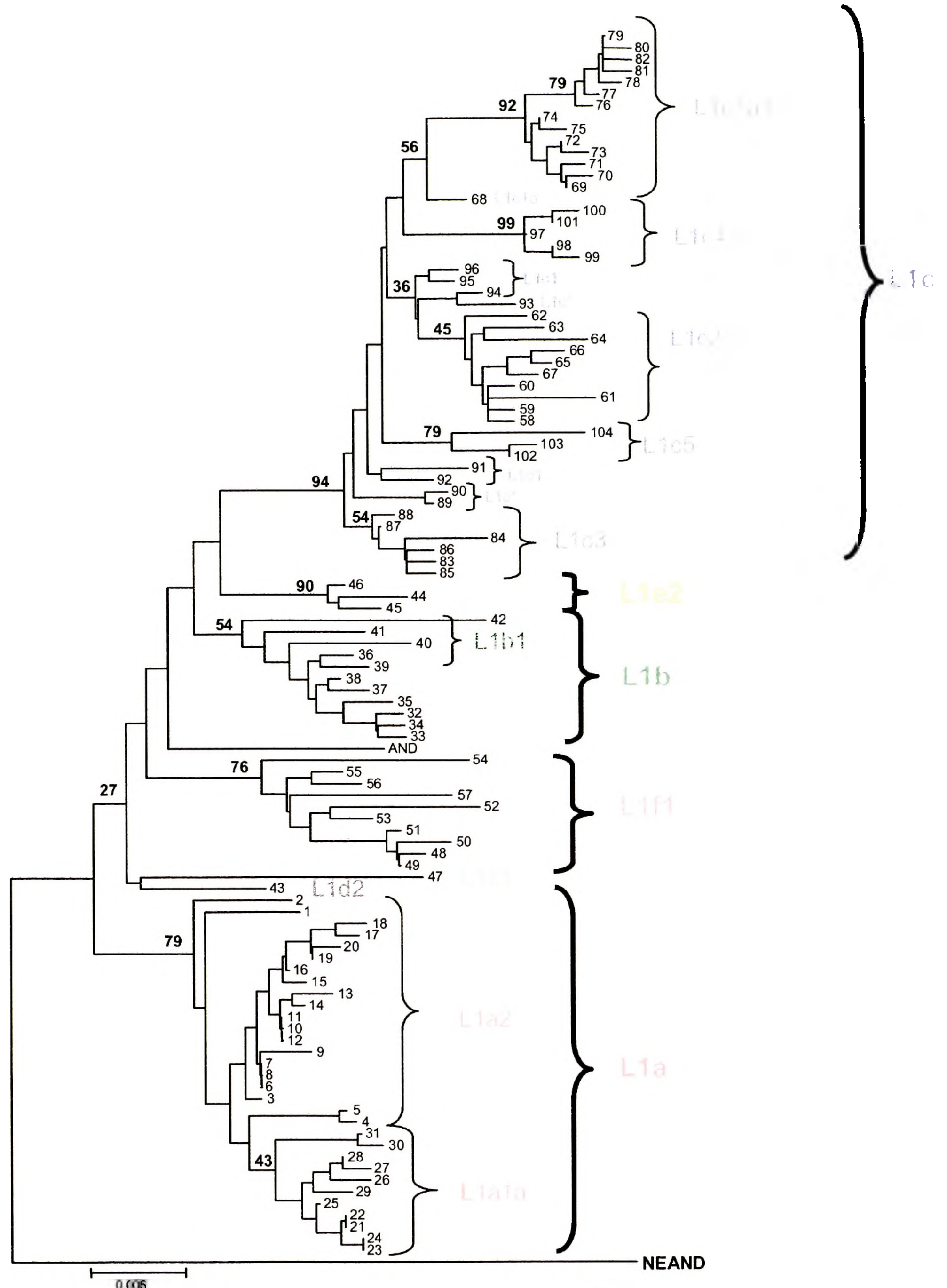


**Figure 4.2** NJ tree showing the relationships among mtDNA types from HG L1. Bootstrap values are shown in bold. AND is the reference sequence (Anderson et al. 1981; Andrews et al. 1999), NEAND is the root, represented by a Neanderthal sequence (Krings et al. 1997, 1999). Numbers represent mtDNA types as shown in Table 4.3.

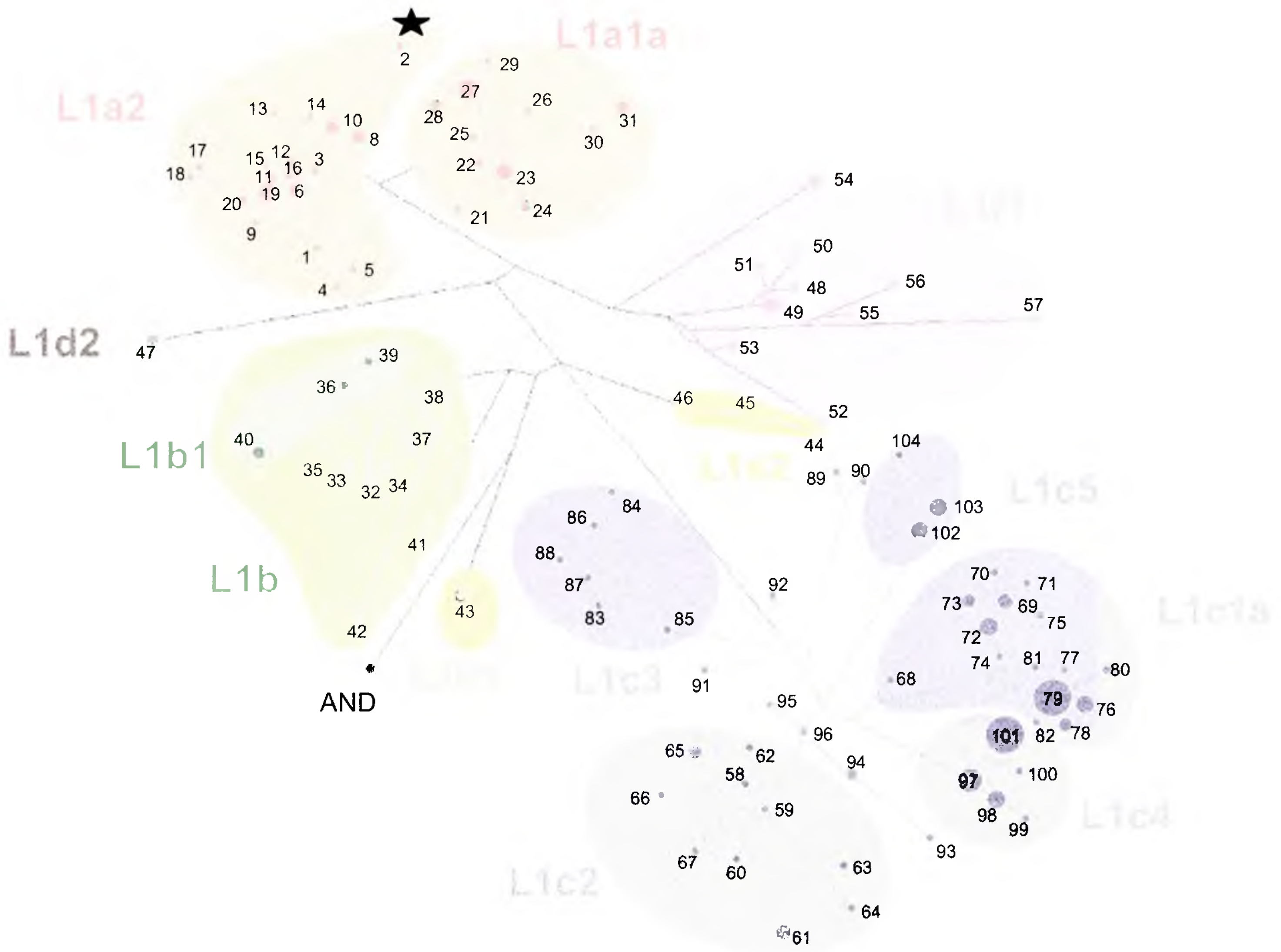


**Figure 4.3** MJ network showing the relationships among 104 L1 mtDNA types from four central African populations. Shaded areas represent mtDNA subHGs within HG L1. AND is the reference sequence (Anderson et al. 1981; Andrews et al. 1999). The star indicates the position of the root / Neanderthal sequence (Krings et al. 1997, 1999). Numbers are mtDNA types as shown in Table 4.3.

*HG L2*

The 44 L2 mtDNA types (Table 4.4) were classified into the four L2 subHGs, L2a, L2b, L2c and L2d previously described by Torroni et al. (2001) and Salas et al. (2002). The polymorphisms defining each subHG are highlighted in Table 4.4 and in fig. 4.8. It has previously been noted that subHG L2c is defined only by the 325 mutation in HVRII or by coding-region mutations (Torroni et al. 2001; Salas et al. 2002). From this study, and using data from other studies with HVRII data (Table 4.3) it is suggested that L2c is further distinguished from other L2 subHGs by mutations at HVRII sites 93 and 182 (Table 4.4, fig. 4.8).

In a NJ tree (fig. 4.4) and a MJ network (fig. 4.5), clusters representing subHGs L2a, L2b and L2c were observed. However the three L2d1 types and single L2d2 type from the present study did not form monophyletic clusters in either the NJ tree or the MJ network (figs. 4.4, 4.5). L2d1 and L2d2 have been suggested to be descended from the same ancestral mtDNA type since they share a mutation at 16 399 and a -*Mbo*I 3 693 polymorphism (Torroni et al. 2001; Salas et al. 2002). L2d1 types from this study lacked the 16 399 mutation, causing their separation in both the NJ tree and the MJ network. Based on types observed by Torroni et al. (2001) and Salas et al. (2002), it is likely that L2d1 types from this study have experienced a reversion at 16399.

SubHG L2a was the most frequent of the L2 subHGs in this study (Table 4.4, fig. 4.1), accounting for 40 of the 55 L2 mtDNAs (0.727). Salas et al. (2002) described extensive substructuring within this subHG, and clades L2a-α1, L2a1-β1, L2a1-β2 and L2a1-β3 were observed in this study, as well as a single instance of L2a1b-γ2 (figs. 4.4, 4.5). These clades were not well-resolved in the NJ tree or MJ network (fig. 4.4 - note low bootstrap values; fig. 4.5), as expected due to the homoplasic nature of the sites defining these clades (Salas et al. 2002).

Table 4.4 List of 44 mtDNA types within HG L2, grouped into subHGs according to their HVR sequence variation

| type no | subHG | 9bp del | AND<br>3594<br>C | 16037<br>A | 16086<br>T | 16092<br>T | 16093<br>T | 16111<br>C | 16114<br>C | 16124<br>T | 16129<br>G | 16131<br>T | 16145<br>G | 16153<br>G | 16172<br>T | 16182<br>A | 16183<br>A | 16183<br>Ins | 16183<br>ins | 16184<br>C | 16186<br>C | 16189<br>T | 16192<br>C | 16192<br>i | 16193<br>C | 16197<br>C | 16209<br>T | 16212<br>A | 16213<br>G | 16217<br>T | 16223<br>C | 16225<br>C | 16229<br>T | 16234<br>C | 16239<br>C | 16249<br>T | 16263<br>T | 16264<br>C | 16265<br>A | 16278<br>C | 16286<br>C | 16290<br>C | 16291<br>C | 16292<br>C | 16294<br>C | 16295<br>C | 16300<br>A | 16301<br>C | 16303<br>G | 16309<br>A | 16311<br>T | 16318<br>A | 16354<br>C | 16355<br>C | 16362<br>T | 16368<br>T | 16390<br>G | 16399<br>A | 16400<br>C | 73<br>A | 93<br>A | 143<br>G | 146<br>T | 150<br>C | 152<br>T | 182<br>C | 183<br>A | 185<br>G | 189<br>A | 195<br>T | 198<br>C | 199<br>T | 200<br>A | 204<br>T | 207<br>G | 235<br>A | 242<br>C | 243<br>A | 244<br>A | 247<br>G | 251<br>G | 263<br>A | 302<br>Ins | 302<br>ins | 315<br>Ins | 324<br>C | 325<br>C | 357<br>A | 385<br>A | ZAM | UGA | CAP | CAR | TOTAL |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | L2a-α1 | no | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C | . | . | . | T | . | . | . | . | . | . | . | . | T | . | . | . | . | T | . | . | T | . | . | . | . | T | . | . | . | A | . | . | G | . | A | C | . | C | T | . | . | G | C | . | . | . | . | . | . | . | . | . | . | . | G | . | . | C | . | . | . | . |  |  |  | 1 | 1 |
| 2 | L2a-α1 | no | T | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | T | . | . | . | . | T | . | . | . | . | . | C | . | . | . | . | . | A | . | . | G | . | A | C | . | C | T | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | G | . | . | C | . | . | . | . | 1 |  |  |  | 1 |
| 3 | L2a-α1 | no | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | T | . | C | . | . | . | T | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | . | A | . | . | G | . | . | C | . | C | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | G | . | . | C | . | . | . | . | 1 |  |  |  | 1 |
| 4 | L2a-α2 | no | T | . | . | . | . | . | . | . | . | . | . | . | . | . | del | . | . | . | . | C | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | T | . | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | A | . | . | G | . | A | C | . | C | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | A | G | C | . | C | . | . | . | . |  |  |  | 1 | 1 |
| 5 | L2a1-β1 | no | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | T | . | . | . | . | T | . | . | . | . | G | . | . | . | . | . | . | A | . | . | G | . | A | C | . | C | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | G | . | . | C | . | . | . | . |  | 1 |  | 1 | 2 |
| 6 | L2a1-β1 | no | T | . | G | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | T | . | . | . | . | T | . | . | . | . | G | . | . | . | . | . | . | A | . | . | G | . | A | C | . | C | . | . | . | . | C | T | . | . | . | . | . | . | . | . | . | . | G | . | . | C | . | . | . | . | 1 |  |  |  | 1 |
| 7 | L2a1-β1 | no | T | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | T | . | . | . | . | T | . | . | . | . | G | . | . | . | . | . | . | A | . | . | G | . | A | C | . | C | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | G | . | . | C | . | . | . | . |  |  |  | 1 | 1 |
| 8 | L2a1-β1 | no | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | T | . | . | . | . | . | . | . | . | T | . | . | . | . | T | . | . | . | . | G | . | . | . | . | . | . | A | . | . | G | . | A | C | . | C | . | . | . | G | C | . | . | . | . | . | . | . | . | . | . | . | G | . | C | C | . | . | . | . |  | 1 |  |  | 1 |
| 9 | L2a1-β1 | no | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | T | . | . | . | . | T | . | . | . | . | G | . | . | . | . | . | . | A | . | . | G | . | . | C | . | C | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | G | . | C | C | . | . | . | . | 1 |  |  | 1 | 2 |
| 10 | L2a1-β1 | no | ? | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | T | . | . | . | . | T | . | . | . | . | G | . | . | . | . | . | C | A | . | . | G | . | . | C | . | C | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | G | . | C | C | . | . | . | . | 1 |  |  |  | 1 |
| 11 | L2a1a | no | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | T | T | . | . | . | T | . | . | . | . | G | . | . | . | . | . | . | A | . | . | G | . | . | C | . | C | . | . | . | G | C | . | C | G | C | . | . | . | . | . | A | . | . | . | C | C | . | . | . | . | 1 |  |  |  | 1 |
| 12 | L2a1a | no | T | . | . | C | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | T | T | . | . | . | T | . | . | . | . | G | . | . | . | . | . | . | A | . | . | G | . | . | C | . | C | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | G | . | . | C | . | . | . | . |  |  |  | 1 | 1 |
| 13 | L2a1a | no | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | T | . | . | . | . | T | . | . | . | . | G | . | . | . | . | . | . | A | . | . | G | . | . | C | . | C | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | G | . | C | C | . | . | . | . | 1 |  |  |  | 1 |
| 14 | L2a1-β2 | no | T | . | . | . | . | . | . | . | . | . | . | . | . | del | del | . | . | . | . | . | . | . | . | C | . | . | . | . | T | . | . | . | . | . | . | . | . | T | . | . | . | . | T | . | . | . | . | G | . | . | . | . | . | . | A | . | . | G | . | . | C | . | C | . | . | . | . | C | T | . | . | . | . | . | . | . | . | . | . | G | . | C | C | . | . | . | . |  |  |  | 1 | 1 |
| 15 | L2a1-β2 | no | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | T | . | . | . | . | T | . | . | . | . | G | . | . | . | . | . | . | A | . | . | G | . | . | C | . | C | . | . | . | . | C | T | . | . | . | . | . | . | . | . | . | . | G | . | . | C | . | . | . | . | 1 |  |  |  | 1 |
| 16 | L2a1-β2 | no | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | T | . | . | . | . | T | . | . | . | . | G | . | . | . | . | . | . | ? | . | . | G | . | . | C | . | C | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | G | . | . | C | . | . | . | . |  | 1 |  |  | 1 |
| 17 | L2a1-β2 | no | T | G | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | T | . | . | . | . | T | . | . | . | . | G | . | . | . | . | . | . | A | . | . | G | . | A | C | . | C | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | G | . | . | C | . | . | . | . |  |  |  | 1 | 1 |
| 18 | L2a1-β2 | no | ? | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | T | . | . | . | . | T | . | . | . | . | G | . | . | . | . | . | . | A | . | . | G | . | A | C | . | C | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | G | C | C | C | . | . | . | . | 1 |  |  |  | 1 |
| 19 | L2a1-β2 | no | T | . | . | . | . | . | . | . | . | . | . | . | . | . | del | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | del | . | . | T | . | . | . | . | T | . | . | . | . | G | . | . | . | . | . | . | A | . | . | G | . | A | C | . | C | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | G | . | . | C | . | . | . | . |  |  |  | 1 | 1 |
| 20 | L2a1-β2 | no | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | T | . | . | . | T | T | . | . | . | . | G | . | . | . | . | . | . | A | . | . | G | . | A | C | . | C | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | G | . | . | C | . | . | . | . |  | 1 |  |  | 1 |
| 21 | L2a1b-γ2 | no | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | T | . | . | . | . | T | . | . | . | . | G | . | . | . | . | . | . | A | . | . | G | . | . | C | . | C | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | G | . | . | C | . | . | . | . | 2 |  |  |  | 2 |
| 22 | L2a1-β3 | no | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | T | . | . | . | . | T | . | . | . | . | G | . | . | . | . | . | . | A | . | . | G | . | A | C | . | C | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | G | . | . | C | . | . | . | . |  | 1 |  |  | 1 |
| 23 | L2a1-β3 | no | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | T | . | . | T | . | . | . | . | . | T | . | . | . | . | T | . | . | . | . | G | . | . | . | . | . | . | A | . | . | G | . | . | C | . | C | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | G | . | C | C | . | . | . | . |  |  |  | 1 | 1 |
| 24 | L2a1-β3 | no | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | C | T | . | . | . | . | . | . | . | . | T | . | . | . | . | T | . | . | . | . | G | . | . | . | . | . | . | A | . | . | G | . | . | C | . | C | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | G | . | C | C | . | . | . | . |  | 1 |  |  | 1 |
| 25 | L2a1-β3 | no | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | T | . | . | . | . | T | . | . | . | . | G | . | . | . | . | . | . | A | . | . | G | . | . | C | . | C | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | G | . | C | C | . | . | . | . | 1 | 2 |  | 1 | 4 |
| 26 | L2a2 | no | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | G | . | . | . | . | T | . | C | . | . | . | . | . | . | T | . | . | T | . | T | . | . | . | . | . | C | . | . | . | . | . | A | . | . | G | . | . | . | . | C | T | . | . | . | C | . | . | G | . | . | . | . | . | . | . | . | G | . | C | C | . | . | . | . |  |  |  | 1 | 1 |
| 27 | L2a2 | no | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | G | . | . | . | . | T | . | C | . | . | . | . | . | . | T | . | . | T | . | T | . | . | . | . | . | C | . | . | . | . | . | A | . | . | G | . | . | . | . | C | T | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | G | . | C | C | . | . | . | . |  |  |  | 1 | 1 |
| 28 | L2a2 | no | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | C | . | . | . | . | . | . | T | . | . | T | . | T | . | . | . | . | . | C | . | . | . | . | . | A | . | . | G | . | . | . | . | C | T | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | G | . | . | G | . | . | . | . |  |  |  | 2 | 2 |
| 29 | L2a2 | no | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | C | . | . | . | . | . | . | T | . | . | T | . | T | . | . | . | A | . | C | . | . | . | . | . | A | . | . | G | . | . | . | . | C | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | G | . | C | C | . | . | . | . |  |  |  | 1 | 1 |
| 30 | L2a2 | no | T | . | . | . | . | . | . | . | . | . | . | . | . | . | del | C | . | . | . | . | . | . | . | . | . | . | . | . | T | . | C | . | . | . | . | . | C | T | . | . | T | . | T | . | . | . | . | . | C | . | . | . | . | . | A | . | . | G | . | . | . | . | C | T | . | . | . | . | . | . | . | C | A | G | . | . | . | . | . | G | . | . | C | . | . | . | . |  | 1 |  |  | 1 |
| 31 | L2a2 | no | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | T | . | C | . | . | . | . | . | . | T | . | . | T | . | T | . | . | . | . | . | C | . | . | . | . | . | A | . | . | G | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | G | . | . | C | . | . | . | . |  |  |  | 1 | 1 |
| 32 | L2a2 | no | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | T | . | C | . | . | . | . | . | . | T | . | . | T | . | T | . | . | . | . | . | C | . | . | . | . | . | A | . | . | G | . | . | . | . | C | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | G | . | . | C | . | . | . | . |  |  |  | 1 | 1 |
| 33 | L2a2 | no | ? | . | . | . | . | . | . | . | . | . | . | . | C | del | del | C | C | . | . | . | . | . | . | . | . | . | . | . | T | . | C | . | . | . | . | . | . | T | . | . | T | . | T | . | . | . | . | . | C | . | . | . | . | . | ? | . | . | G | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | G | C | C | C | . | . | . | . |  | 1 |  |  | 1 |
| 34 | L2b | no | T | . | . | . | . | . | A | . | A | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . | T | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | A | . | . | G | . | . | C | T | C | T | . | . | . | C | T | . | . | C | . | . | . | . | . | . | . | G | . | C | C | . | . | . | . |  |  |  | 1 | 1 |
| 35 | L2b | no | ? | . | . | . | . | . | A | . | A | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | G | A | . | T | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . | . | G | . | . | C | T | C | T | . | . | . | C | T | . | . | C | A | . | . | . | . | . | . | G | . | . | C | . | . | . | . |  | 1 |  |  | 1 |
| 36 | L2b1 | no | T | . | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . | T | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C | . | A | . | . | G | . | . | C | T | C | T | . | . | . | C | T | . | . | C | A | . | . | . | . | . | . | G | . | . | C | . | . | . | . |  |  |  | 1 | 1 |
| 37 | L2b1 | no | T | . | . | . | . | . | A | . | A | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . | T | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | . | T | C | . | A | . | . | G | . | . | . | T | C | T | . | . | . | C | T | . | . | C | . | . | . | . | . | . | . | G | . | C | C | . | . | . | . |  | 2 |  |  | 2 |
| 38 | L2b1 | no | T | . | . | . | . | . | A | . | A | . | . | A | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . | T | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | C | . | . | . | . | C | . | A | . | . | G | . | . | C | T | C | T | G | . | . | C | T | . | . | C | . | . | . | . | . | . | . | G | . | . | C | . | . | . | G | 2 |  |  |  | 2 |
| 39 | L2c2 | no | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | T | . | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . | . | G | G | . | C | T | C | T | G | . | . | C | . | . | . | . | . | . | . | . | . | . | . | G | . | . | C | . | T | . | . |  |  |  | 1 | 1 |
| 40 | L2c1 | no | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | G | . | . | . | . | A | . | . | G | G | . | C | T | C | T | . | . | . | C | T | . | . | . | . | . | . | . | . | . | . | G | . | . | C | . | T | . | . |  |  |  | 3 | 3 |
| 41 | L2d1 | no | T | . | . | . | C | . | . | . | A | . | . | . | . | . | del | C | . | . | . | . | . | . | . | . | . | . | . | . |  | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | . | C | . | T | . | . | . | A | ? | . | G | . | . | C | T | . | . | . | . | . | C | . | . | . | C | . | . | . | . | . | . | . | G | . | . | C | . | . | . | . |  | 1 |  |  | 1 |
| 42 | L2d1 | no | T | . | . | . | . | . | . | . | A | . | . | . | . | . | del | . | . | . | . | . | . | . | . | . | . | . | . | . |  | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | A | . | . | G | . | . | C | T | . | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | G | . | C | C | T | . | . | . |  |  |  | 1 | 1 |
| 43 | L2d1 | no | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . |  | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | A | . | . | G | . | . | C | T | . | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | G | . | C | C | . | . | . | . |  |  |  | 1 | 1 |
| 44 | L2d2 | no | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . | T | G | . | . | C | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | G | . | . | C | . | . | . | . |  |  |  | 1 | 1 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | total | 14 | 14 | 0 | 27 | 55 |

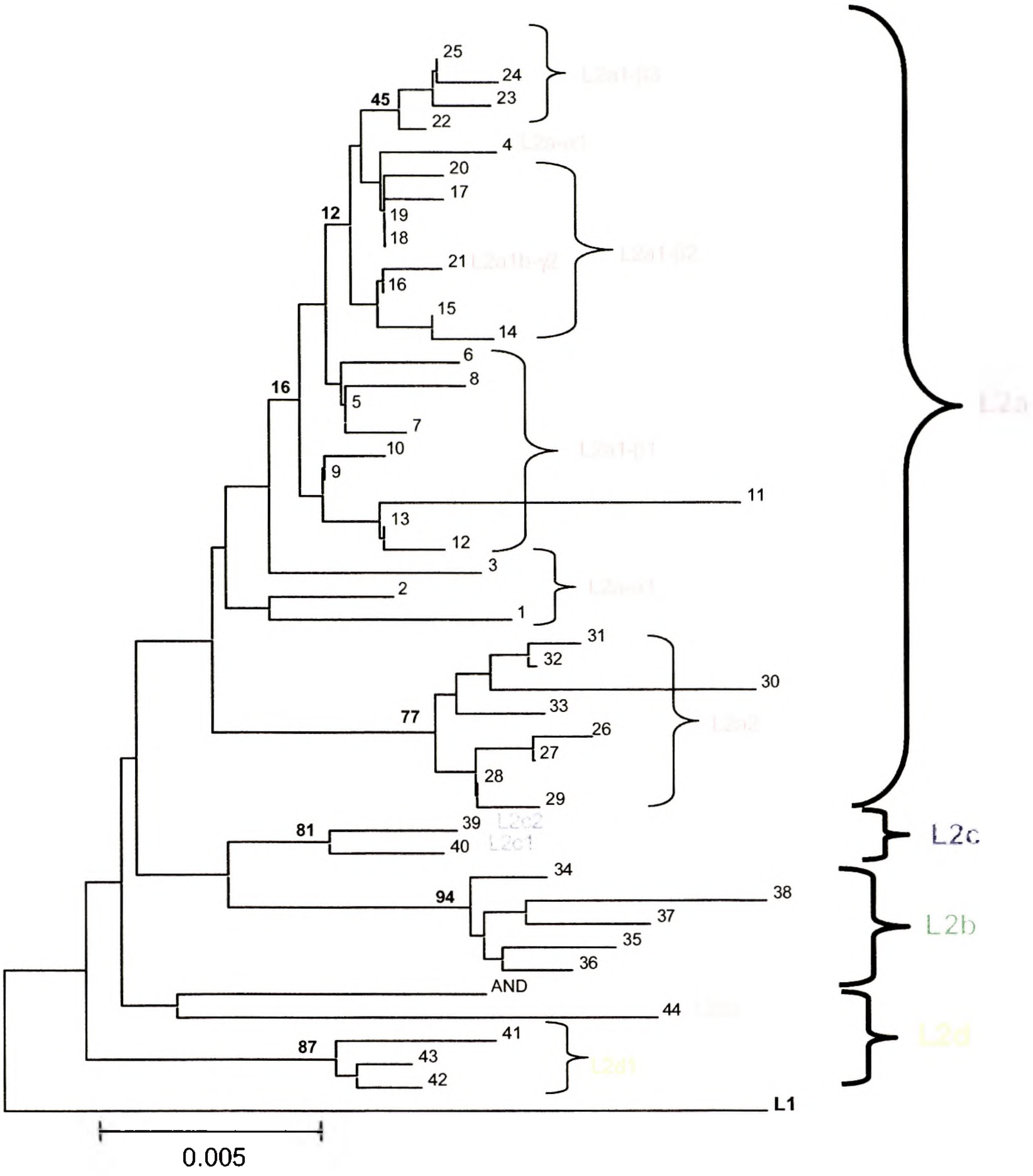


**Figure 4.4** NJ tree showing the relationships among 44 mtDNA types from HG L2. Bootstrap values are shown in bold. AND is the reference sequence (Anderson et al. 1981; Andrews et al. 1999). The tree is rooted with a L1 sequence. Numbers represent mtDNA types as shown in Table 4.4.

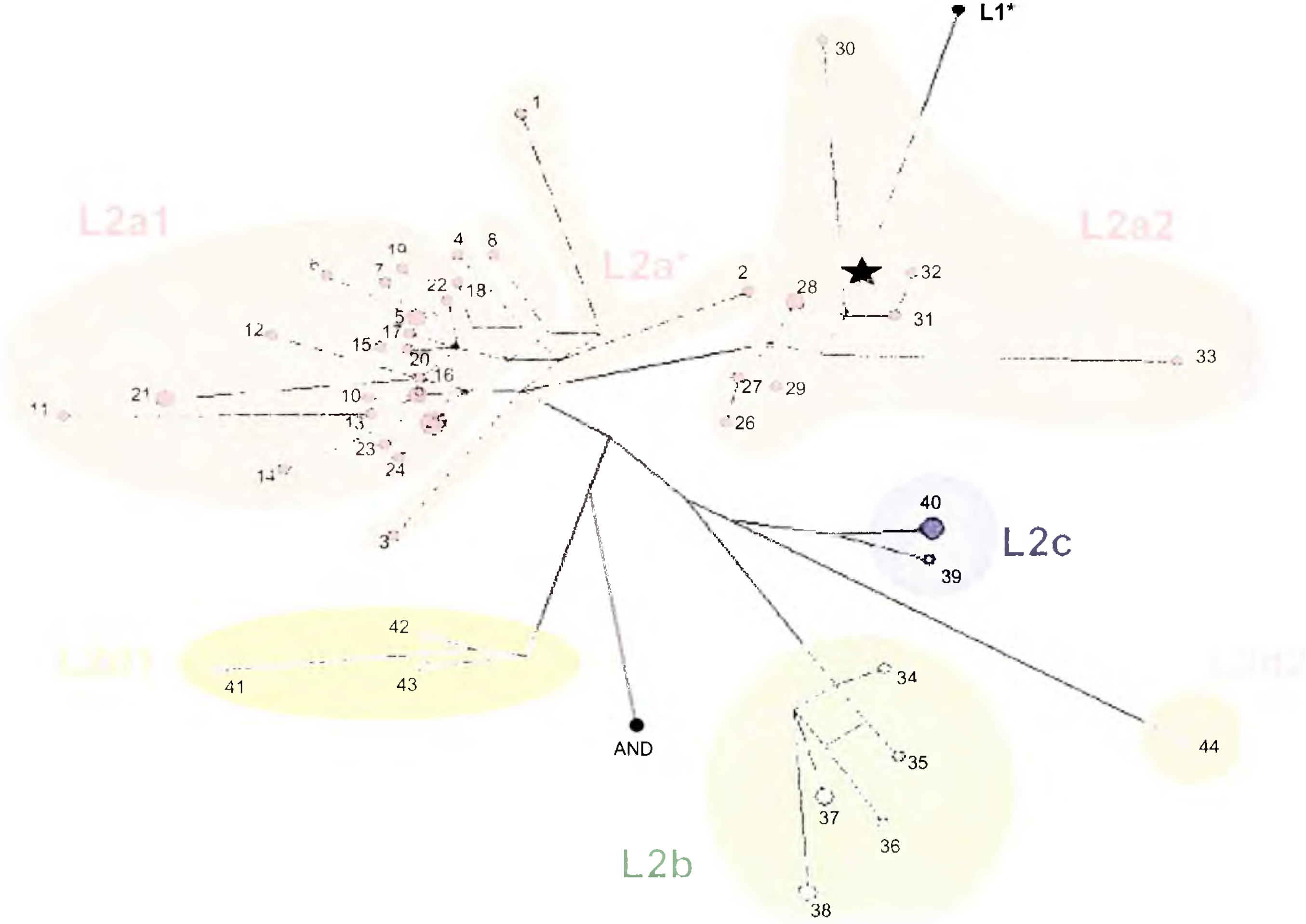


**Figure 4.5** MJ network showing the relationships among 55 L2 mtDNA types from four central African populations. Shaded areas represent mtDNA subHGs within HG L2. AND is the reference sequence (Anderson et al. 1981; Andrews et al. 1999), the star indicates the position of the root. Numbers represent mtDNA types as shown in Table 4.4.

One new clade of L2a mtDNA types was identified and called L2a2. This clade was characterised by mutations at sites 16189-16223-16229-16278-16291-16294-16311-16390-73-152-182-195-263-315ins, and was observed in 7/95 (0.074) individuals from the C.A.R. and 2/92 (0.022) individuals from Uganda (Table 4.4). The root of the L2 network (represented by an L1 sequence) was placed within the L2a2 cluster in the network (fig. 4.5), and L2a2 also appeared to be an early offshoot of HG L2 in a network constructed using all subHGs found in African populations (fig. 4.8). This placement is probably due to the retention in L2a2 of the 16189 and 16311 mutations, and the lack of the 146 mutation which defines all other L2 subHGs. A search of the published literature revealed sublineage L2a2 in six of 13 Mbuti Pygmies (46.2%, Watson et al. 1997), 9.2% of 76 individuals from Sudan (Krings et al. 1999), 2.7% of 37 Turkana from Kenya (Watson et al. 1997), and 1.5% of 68 Egyptians (Krings et al. 1999).

*HG L3*

Specific variant sites within the control region were used to sort the 98 unique mtDNA types within HG L3 (Table 4.5) into the subHGs L3b, L3d, L3e, L3f and L3g defined by Salas et al. (2002). The polymorphisms defining each subHG are highlighted in Table 4.5 and in fig. 4.8. MtDNA types from L3g in this study all had a mutation at 16399 which has not been previously described, but this is probably because HVRI typing of L3g samples in other studies did not reach this point. One new subHG, called L3h, was identified. This subHG was characterised by mutations at sites 16223-16311-16354-16399-73-146-153-263-315ins, and appeared to be related to subHG L3g due to shared mutations at 16399 and 146 (figs. 4.7a, 4.8). L3h was observed only in six of 92 Ugandans in this study (0.065, fig. 4.1); a search of the published literature revealed L3h in 1/37 Turkana (0.027), 2/25 Kikuyu (0.080) from Kenya (Watson et al. 1997) and 1/12 Iraqw (0.083) from Tanzania (Knight et al. 2003).

The relationships among mtDNA types from HG L3 were most easily resolved when L3e mtDNA types were placed in a separate network to types from other L3 subHGs (figs. 4.7a, 4.7b). A single NJ tree drawn with all 98 types did manage to resolve most of the L3 subHGs, but the extremely low bootstrap values reflected the difficulty observed in resolving L3 types based on sequence data alone (fig. 4.6). Clusters of mtDNA types in the NJ tree (fig. 4.6) and MJ networks (figs. 4.7, 4.8) generally represented L3 clades as described by Salas et al. (2002). However subHGs L3b and L3d were placed separately in a network drawn with all African mtDNA subHGs (fig. 4.8), suggesting that the 16124 mutation may have evolved twice. The 16185 mutation defining subHG L3e1a (Torroni et al. 2001; Salas et al. 2002) also appeared to be homoplasic; types with this mutation did not cluster together in the NJ tree (fig. 4.6) or MJ network (fig. 4.7b). All three mtDNA types with a triplication of the COII/tRNA$^{Lys}$ intergenic 9-bp repeat motif (types 53, 65 and 74, Table 4.5) were classified as belonging to subHG L3e; although they did not cluster together in the NJ tree, their placement in a phylogenetic network (fig. 4.7b) confirmed that the triplication has a single evolutionary origin.

SubHG L3e was the most common L3 subHG in the four central African populations (Table 4.5, fig. 4.1). The other L3 subHGs were not common (less than 0.05 of each population) but, with the exception of L3h, were widespread in central Africa (fig. 4.1).

Eight mtDNA types within HG L3 could not be classified into subHGs (types 1-6, 8, Table 4.5). These types were included in networks including and excluding subHG L3e (black circles, figs 4.7a, 4.7b), but most types did not cluster into defined subHGs. Type 7 might belong to subHG L3f, but lacked the defining HVRI mutations at 16209 and 16311; also, unusually for HG L3 types, it had a 9-bp deletion (Table 4.5). Based on the presence of the 16223 mutation in these samples, they do not belong to HG R, and RFLP typing of the 10397 *Alu*I polymorphism showed that they do not belong to the Asian superhaplogroup M. These types therefore either belong to the African HG L3A*, or to the Eurasian HG N. One mtDNA type (type 8, Table 4.5), putatively belongs to mtDNA HG X2c, since it has control region sequence variation similar to that described in Reidla et al. (2003).

Table 4.5 List of 98 mtDNA types within HG L3, grouped into subHGs according to their HVR sequence variation

| new no. | subHG | 9bp del | 9bp seq | ZAM | UGA | CAP | CAR | total |
|---|---|---|---|---|---|---|---|---|
| 1 | ? | no | | | | | 1 | 1 |
| 2 | ? | no | | | 1 | | | 1 |
| 3 | ? | no | | | | | 4 | 4 |
| 4 | ? | no | | | | | 1 | 1 |
| 5 | ? | no | | | | | 1 | 1 |
| 6 | ? | no | | | 1 | | | 1 |
| 7 | ? | yes | T | 1 | | | | 1 |
| 8 | K2c | no | | | 1 | | | 1 |
| 9 | L3b1 | no | | | | | 1 | 1 |
| 10 | L3b1 | no | | 2 | | | | 2 |
| 11 | L3b | no | | | | | 1 | 1 |
| 12 | L3b | no | | | 1 | | | 1 |
| 13 | L3b2 | no | | | | 2 | | 2 |
| 14 | L3b2 | no | | 2 | | | | 2 |
| 15 | L3b2 | no | | | | | 1 | 1 |
| 16 | L3b | no | | 1 | | | 2 | 3 |
| 17 | L3b | no | | 1 | | | | 1 |
| 18 | L3b | no | | 1 | | | | 1 |
| 19 | L3b | no | | | 1 | | | 1 |
| 20 | L3b | no | | 1 | 1 | | | 2 |
| 21 | L3d3 | no | | | 1 | | | 1 |
| 22 | L3d | no | | 1 | | | | 1 |
| 23 | L3d1 | no | | | 1 | | | 1 |
| 24 | L3d1 | no | | 1 | | | | 1 |
| 25 | L3d1 | no | | 2 | 3 | | | 5 |
| 26 | L3d | no | | | | | 2 | 2 |
| 27 | L3d | no | | | 2 | | | 2 |
| 28 | L3f | no | | | | | 2 | 2 |
| 29 | L3f | no | | | | | 1 | 1 |
| 30 | L3f | no | | 1 | | | | 1 |
| 31 | L3f | no | | | | | 1 | 1 |
| 32 | L3f | no | | 2 | | | | 2 |
| 33 | L3f1 | no | | 1 | | | | 1 |
| 34 | L3f1 | no | | | 1 | | | 1 |
| 35 | L3f1 | no | | | | | 1 | 1 |
| 36 | L3f1 | ? | | 1 | | | | 1 |
| 37 | L3f1 | no | | | | | 1 | 1 |
| 38 | L3f1 | no | | | 1 | | | 1 |
| 39 | L3f1 | no | | 1 | | | | 1 |
| 40 | L3f1 | no | | 1 | | | | 1 |
| 41 | L3g | ? | | 1 | | | | 1 |
| 42 | L3g | no | | | 2 | | | 2 |
| 43 | L3g | no | | | 2 | | | 2 |
| 44 | L3g | no | | | | | 1 | 1 |
| 45 | L3g | no | | | | | 1 | 1 |
| 46 | L3g | no | | | 1 | | | 1 |
| 47 | L3g | no | | | 1 | | | 1 |
| 48 | L3g | no | | | 1 | | | 1 |
| 49 | L3h | no | | | 1 | | | 1 |
| 50 | L3h | no | | | 1 | | | 1 |
| 51 | L3h | no | | | 3 | | | 3 |
| 52 | L3h | no | | | 1 | | | 1 |
| 53 | L3e1 | trip | T | | | | 4 | 4 |
| 54 | L3e1a | no | | | 1 | | | 1 |
| 55 | L3e1a | no | | 5 | | | | 5 |
| 56 | L3e1a | no | | 1 | | | | 1 |
| 57 | L3e1 | no | | 2 | | | | 2 |
| 58 | L3e1 | no | | | | | 1 | 1 |
| 59 | L3e1b | no | | | 3 | | | 3 |
| 60 | L3e1 | no | | 2 | | | | 2 |
| 61 | L3e1a | no | | | 1 | | | 1 |
| 62 | L3e1a | no | | | 1 | | | 1 |
| 63 | L3e1a | no | | | | | 1 | 1 |
| 64 | L3e1a | no | | 1 | | | | 1 |
| 65 | L3e1 | trip | T | | | | 1 | 1 |
| 66 | L3e1 | no | | | 1 | | | 1 |
| 67 | L3e1a | no | | | 1 | | | 1 |
| 68 | L3e1a | no | | 1 | | | | 1 |
| 69 | L3e1 | no | | 1 | | | | 1 |
| 70 | L3e1 | no | | 1 | | | | 1 |
| 71 | L3e1 | no | | 1 | | | | 1 |
| 72 | L3e1 | no | | 1 | | | | 1 |
| 73 | L3e1 | no | | 1 | | | | 1 |
| 74 | L3e1 | trip | | | | | 1 | 1 |
| 75 | L3e1 | no | | | | | 2 | 2 |
| 76 | L3e1 | no | | 1 | | | | 1 |
| 77 | L3e2 | no | | | | | 1 | 1 |
| 78 | L3e2a | no | | | | | 2 | 2 |
| 79 | L3e2a | no | | | | | 1 | 1 |
| 80 | L3e2b | no | | 1 | | | | 1 |
| 81 | L3e2b | no | | 1 | | | | 1 |
| 82 | L3e2b | no | | 1 | | | | 1 |
| 83 | L3e2b | no | | 1 | | | | 1 |
| 84 | L3e2b | no | | | 1 | | | 1 |
| 85 | L3e2b | no | | | | | 1 | 1 |
| 86 | L3e2b | no | | | | | 1 | 1 |
| 87 | L3e2b | no | | 1 | | | | 1 |
| 88 | L3e2a | no | | | | | 1 | 1 |
| 89 | L3e2b | no | | | 1 | | | 1 |
| 90 | L3e2b | no | | | 1 | | | 1 |
| 91 | L3e3 | no | | | | | 1 | 1 |
| 92 | L3e3 | no | | | 1 | | | 1 |
| 93 | L3e3 | no | | | 1 | | | 1 |
| 94 | L3e3 | no | | | 1 | | | 1 |
| 95 | L3e3 | no | | 1 | | | | 1 |
| 96 | L3e3 | no | | | | | 1 | 1 |
| 97 | L3e3 | no | | | 1 | | | 1 |
| 98 | L3e4 | no | | | | | 1 | 1 |
| | | | total | 45 | 43 | 2 | 42 | 132 |

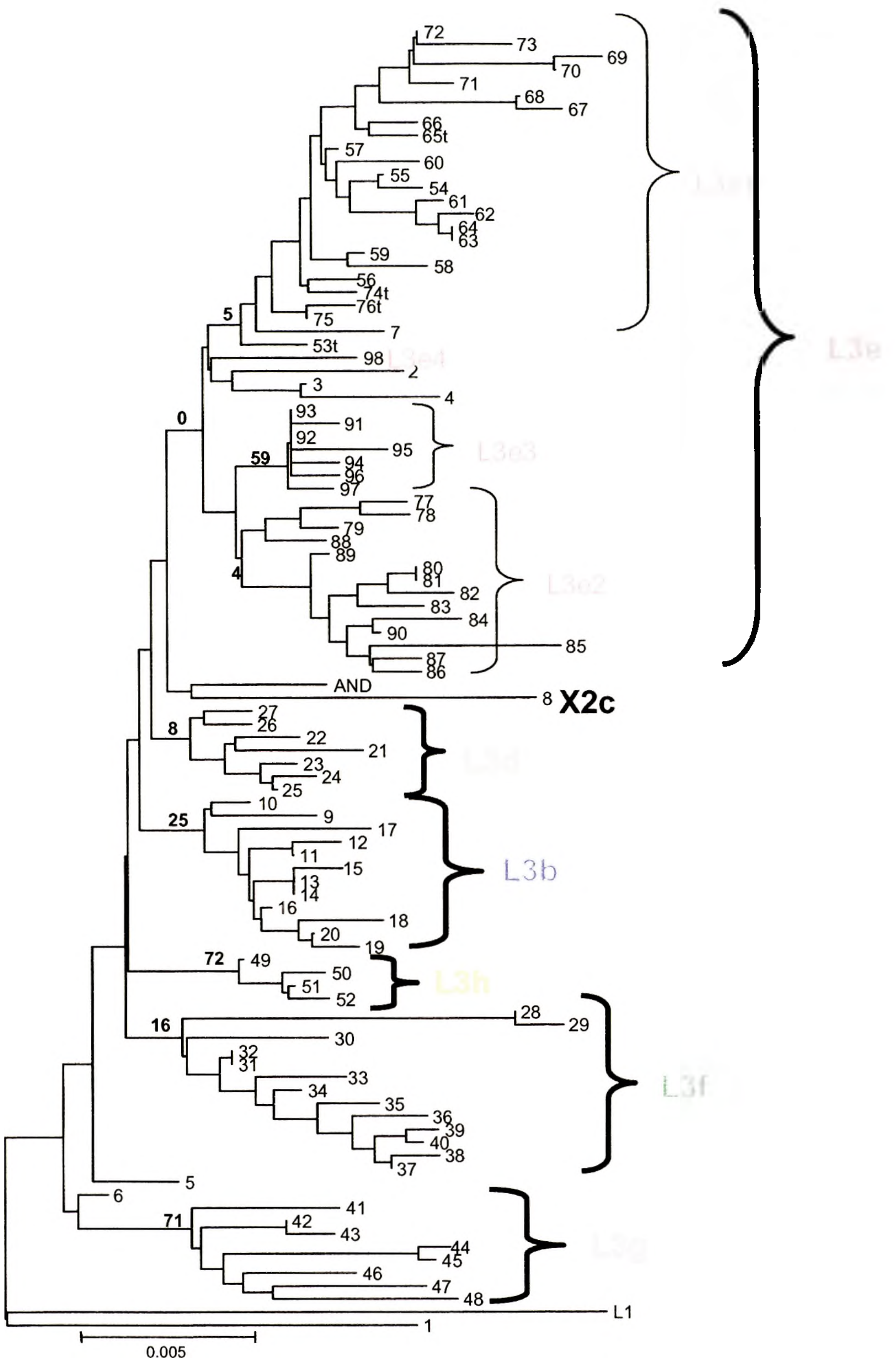


**Figure 4.6** NJ tree showing the relationships among 98 mtDNA types from HG L3. Numbers refer to mtDNA types described in Table 4.5. Bootstrap values are shown in bold. AND is the reference sequence (Anderson et al. 1981; Andrews et al. 1999). The tree is rooted with a L1 sequence.

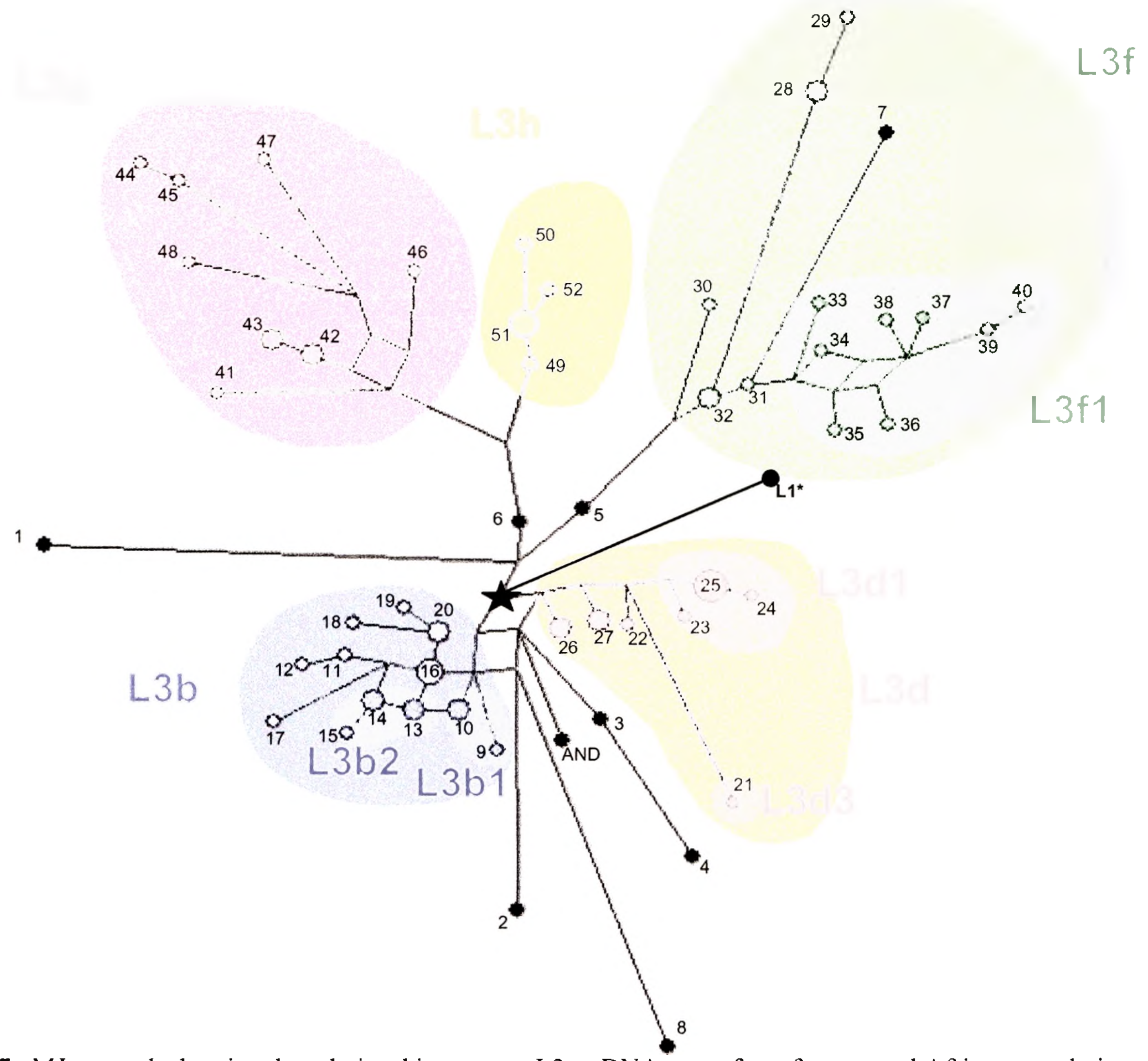


**Figure 4.7a** MJ network showing the relationships among L3 mtDNA types from four central African populations. Numbers refer to mtDNA types described in Table 5. Types from L3e are excluded. Shaded areas represent mtDNA subHGs within HG L3. Black circles indicate mtDNA types which have not been classified into subHGs. AND is the reference sequence (Anderson et al. 1981; Andrews et al. 1999), the star is the L1 root.

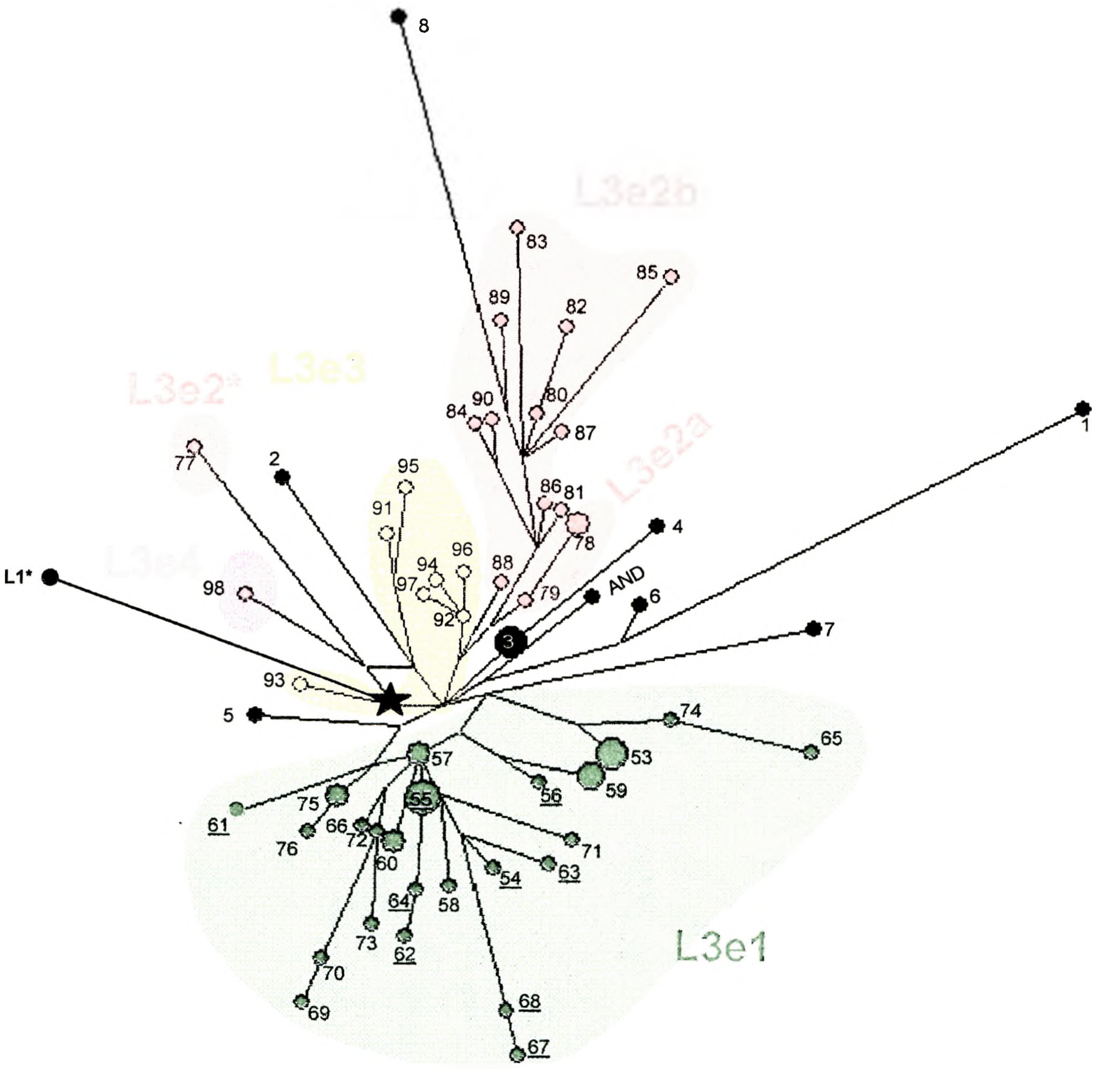


**Figure 4.7b**. MJ network showing the relationships among L3e and L3* mtDNA types from four central African populations. Numbers refer to mtDNA types described in Table 5. Shaded areas represent mtDNA subHGs within HG L3e. Black circles indicate mtDNA types which have not been classified into subHGs. Underlined types are classified as L3e1a according to Salas et al. (2002). AND is the reference sequence (Anderson et al. 1981; Andrews et al. 1999), the star is the root.

**Table 4.6** Estimated TMRCA of each mtDNA subHG (shown in years)

| | | | | 95% CI | |
|---|---|---|---|---|---|
| **HG** | **n** | **d** | **TMRCA** | **upper** | **lower** |
| L1 | 406 | 0.0277 | **167101** | 478255 | 84569 |
| L1a | 114 | 0.0110 | **66305** | 189769 | 33557 |
| L1a* | 5 | 0.0115 | **69129** | 197852 | 34986 |
| L1a1a | 43 | 0.0075 | **45058** | 128959 | 22804 |
| L1a2 | 66 | 0.0065 | **39376** | 112696 | 19928 |
| L1f | 16 | 0.0141 | **85177** | 243782 | 43108 |
| L1d | 65 | 0.0096 | **57764** | 165324 | 29234 |
| L1k | 20 | 0.0038 | **22870** | 65455 | 11574 |
| L1e | 8 | 0.0056 | **33947** | 97159 | 17181 |
| L1b | 20 | 0.0102 | **61203** | 175166 | 30974 |
| L1c | 163 | 0.0158 | **95139** | 272295 | 48150 |
| L1c* | 6 | 0.0107 | **64191** | 183720 | 32487 |
| L1c1 | 7 | 0.0066 | **39973** | 114404 | 20230 |
| L1c1a | 59 | 0.0037 | **22409** | 64136 | 11341 |
| L1c2 | 19 | 0.0091 | **55009** | 157439 | 27840 |
| L1c3 | 9 | 0.0042 | **25375** | 72624 | 12842 |
| L1c4 | 43 | 0.0012 | **7191** | 20580 | 3639 |
| L1c5 | 20 | 0.0027 | **16308** | 46674 | 8253 |
| L2 | 151 | 0.0111 | **66573** | 190536 | 33692 |
| L2a | 129 | 0.0079 | **47851** | 136953 | 24217 |
| L2a* | 6 | 0.0104 | **62545** | 179009 | 31654 |
| L2a1 | 100 | 0.0038 | **22791** | 65230 | 11535 |
| L2a2 | 23 | 0.0070 | **42108** | 120515 | 21311 |
| L2b | 13 | 0.0067 | **40515** | 115957 | 20505 |
| L2c | 5 | 0.0041 | **24689** | 70661 | 12495 |
| L2d | 4 | 0.0123 | **74067** | 211984 | 37485 |
| L3 | 250 | 0.0151 | **90714** | 259630 | 45910 |
| L3b | 31 | 0.0055 | **33237** | 95127 | 16821 |
| L3d | 49 | 0.0064 | **38314** | 109657 | 19391 |
| L3f | 18 | 0.0087 | **52444** | 150098 | 26542 |
| L3g | 49 | 0.0065 | **39049** | 111760 | 19763 |
| L3h | 7 | 0.0018 | **10581** | 30283 | 5355 |
| L3e | 96 | 0.0099 | **59930** | 171524 | 30331 |
| L3e1 | 59 | 0.0075 | **44991** | 128768 | 22770 |
| L3e2 | 24 | 0.0054 | **32516** | 93063 | 16456 |
| L3e3 | 11 | 0.0019 | **11222** | 32119 | 5680 |
| L3e4 | 2 | 0.0020 | **12344** | 35331 | 6248 |

4.1.1.5 *Phylogenetic relationships among HGs L1, L2 and L3*

The relationships among mtDNA HGs and subHGs found in African populations have recently been well-defined on the basis of HVRI sequence variation, incorporating some coding regions mutations (Bandelt et al. 2001; Pereira et al. 2001; Torroni et al. 2001; Salas et al. 2002). Data from this study, combined with HVRII sequence data from various other studies (Table 2.3) have allowed HVRII sequence variation, and information regarding the mtDNA 9-bp deletion, to be incorporated into the definitions of HGs and subHGs observed in African populations (fig. 4.8).

Overall, HG L1 can now be defined by mutations at sites 16187, 16189, 16223, 16278, 16311, 73, 152, 195, 247, 263 and 315 ins C; HG L2 mtDNA types is defined by mutations at sites 16223, 16278, 16390, 73, 146, 150, 152, 195, 263, 315 ins C; and L3 is defined by 16223, 73, 263, 315 ins C (fig. 4.8). The placement of some of the sites defining major HGs, such as 16311 and 16278, changed from the suggestions made by Salas et al. (2002). For example, subHGs L3f, L3g and L3h were united by their retention of the 16311 mutation, and 16311 was also present in subHG L2a2, so that this mutation was no longer considered to be one of the polymorphisms defining HG L1 (fig. 4.8). [The change in placement to of this polymorphism was also one of the causes of the uncoupling of subHGs L3b and L3d (fig. 4.8)]. The 16278 C allele was not placed on the branch defining L3 as in Salas et al. (2002), but rather within the HG L3 cluster (fig. 4.8). This change was accompanied by the placement of subHGs L3b and L3d as early offshoots of L3, and created a new node from which all other L3 subHGs (L3e, L3f, L3g and L3h) were derived (fig. 4.8). The validity of these changes is uncertain, since the 16311 and 16278 polymorphisms are known to be "hot spots" for recurrent mutations (Malyarchuk et al. 2002).

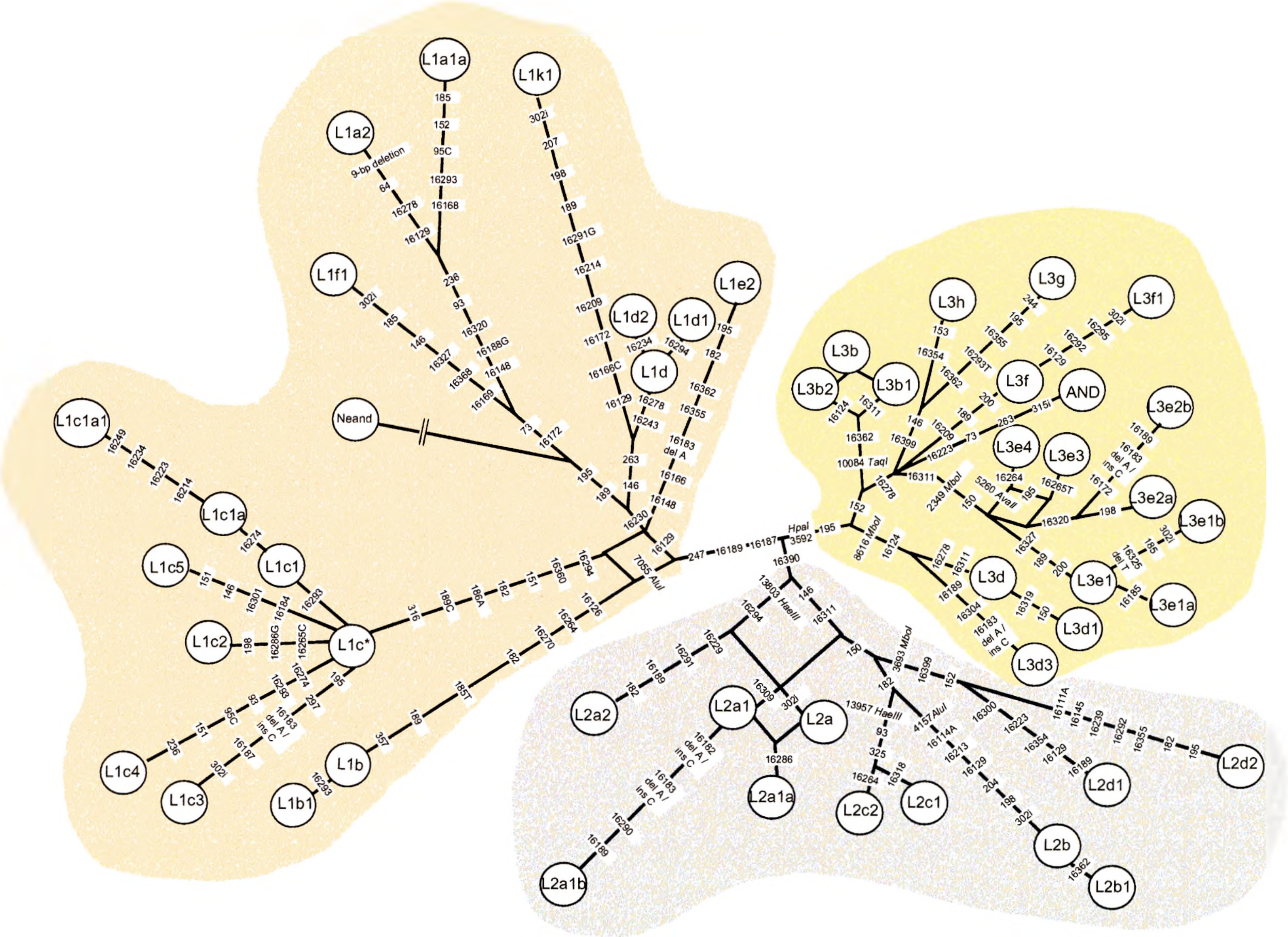


**Figure 4.8** Skeleton phylogenetic network showing the relationships among mtDNA HGs L1, L2, L3 and their subHGs. All types also have mutations at 16623 and 263 relative to the CRS unless otherwise indicated.

HVRII variation remains to be described for several of the more basal clades in the phylogeny, including L1a*, L1e*, L1f*, L1k*, L2d*, L2c* and L3e*. However because HVRII variation is known for the more derived clades in most cases, it is unlikely that the overall phylogeny suggested in fig. 4.8 will change substantially. HVRII variation also remains to be described for L1e1, L3d2 and L3e2. Whole mtDNA genome sequencing would be the best way of determining the true phylogenetic relationship among HGs L1, L2 and L3.

### 4.1.1.6 *Associations between mtDNA subHGs and the 9-bp deletion*

As previously reported, only subHG L1a2 is consistently associated with the mtDNA 9-bp deletion (Soodyall et al. 1996; Watson et al. 1997; Salas et al. 2002).However sporadic re-occurrences of the 9-bp deletion have been observed in subHGs L1c* (types 89 and 90, Table 4.3), L1d (Soodyall et al. 1996), L1f1 (type 54, Table 4.3), and in an unclassified L3 type (type 7, Table 4.5). Therefore, of at least five independent occurrences of the 9-bp deletion in African populations, four of them have been in mtDNA types from HG L1, suggesting that mtDNA types from HG L1 may be more predisposed to the occurrence of the 9-bp deletion than mtDNA types from other HGs. Re-occurrence of the 9-bp deletion in HG L1 may not be unexpected, given the age of HG L1.

### 4.1.2 *Genetic affinities among central African populations using mtDNA data*

The genetic relationships among the four central African populations were assessed by exact tests of differentiation using HVRI and II sequence data, and using subHG frequency data. Both when HVRI and II sequence data were used (Table 4.7a), and when subHG frequency data were used (Table 4.7b), all populations were significantly different to each other ($P$<0.001). The $F_{ST}$ genetic distances between the Biaka and the other central African populations were approximately twenty to forty times greater than the distances among the non-Pygmy populations (Table 4.7). In a NJ-tree constructed from HVRI and II sequence data and $F_{ST}$ values (fig. 4.9) the difference between the Biaka and other central African populations was reflected by their placement on different branches.

**Table 4.7** Pairwise genetic distances among four central African populations calculated from mtDNA data.

**a) Matrix of $F_{ST}$ genetic distances calculated from HVRI and II sequence data**

| | CAP | CAR | UGA |
|---|---|---|---|
| CAR | 0.41511*** | | |
| UGA | 0.40176*** | 0.02747*** | |
| ZAM | 0.38445*** | 0.01629*** | 0.01058*** |

**b) Matrix of $F_{ST}$ genetic distances calculated from mtDNA subHG frequency data**

| | CAP | CAR | UGA |
|---|---|---|---|
| CAR | 0.47875*** | | |
| UGA | 0.47894*** | 0.02262*** | |
| ZAM | 0.44372*** | 0.00811** | 0.01079*** |

Abbreviations:

CAP: Pygmies from Central African Republic

CAR: Central African Republic

UGA: Uganda

ZAM: Zambia

*significant difference, P<0.05

**significant difference, P<0.01

***significant difference, P<0.001

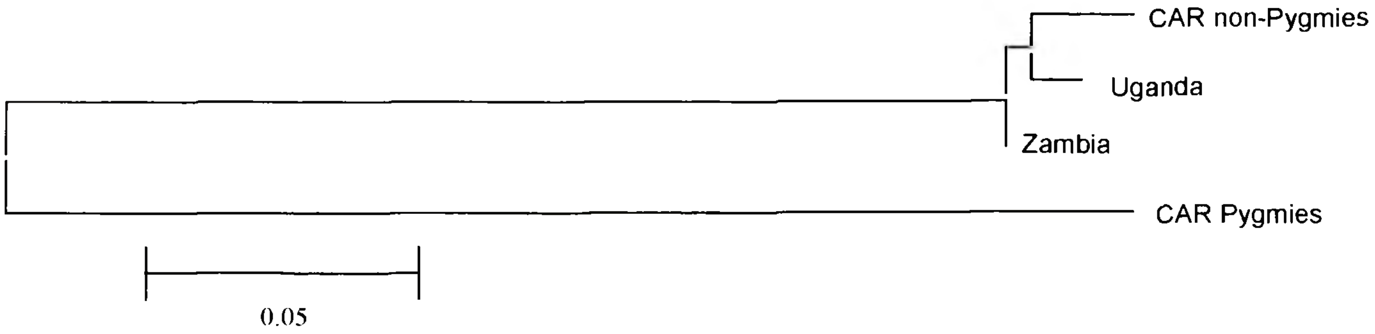


**Figure 4.9** NJ tree constructed using mtDNA HVRI data and $F_{ST}$ genetic distances, showing the genetic affinities among four central African populations.

### 4.1.3 *Comparison to mtDNA data from other African populations*

Data from the present study were compared to HVRI sequence data from 37 other African populations (Table 2.3) to examine the geographic distribution of mtDNA variation in Africa, and to understand the genetic affinities of central African populations with other African populations.

#### 4.1.3.1 *Distribution of mtDNA subHGs in Africa*

The geographic distribution of mtDNA subHGs from HG L1 within Africa is depicted in fig. 4.10. The Biaka from the C.A.R. and the Khoisan populations were notable for their very high frequencies of L1 lineages. L1 mtDNA types accounted for between 20% and 66% of the gene pools of most sub-Saharan African populations, but were less frequent in west African populations and rare in north-east and north-west African populations. Most of the L1 mtDNA subHGs had fairly restricted geographic distributions: L1b was largely found in west Africa, L1c was concentrated in central Africa, L1d and L1k were restricted to Khoisan populations from southern Africa, and L1e and L1f were present mostly in east Africa. However subHG L1a had a wide geographic distribution, with its highest frequency in east Africa.

The geographic distribution of mtDNA subHGs from HG L2 within Africa is depicted in fig. 4.11. SubHG L2a had an extremely wide distribution in Africa, and was frequent in many populations from different geographic regions. SubHGs L2b, L2c and L2d were largely confined to west African populations, although traces of L2b extended to east and southern Africa.

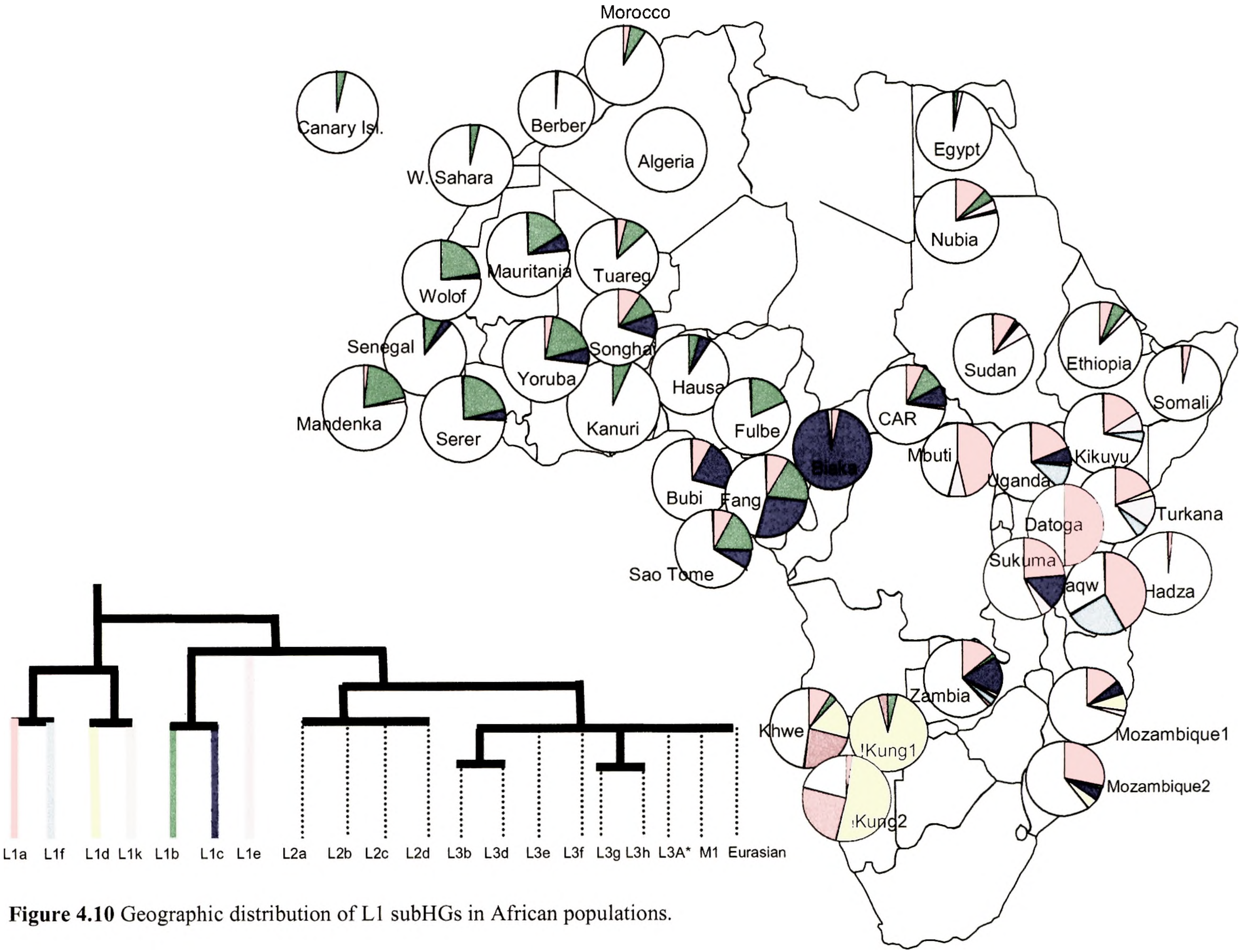


**Figure 4.10** Geographic distribution of L1 subHGs in African populations.

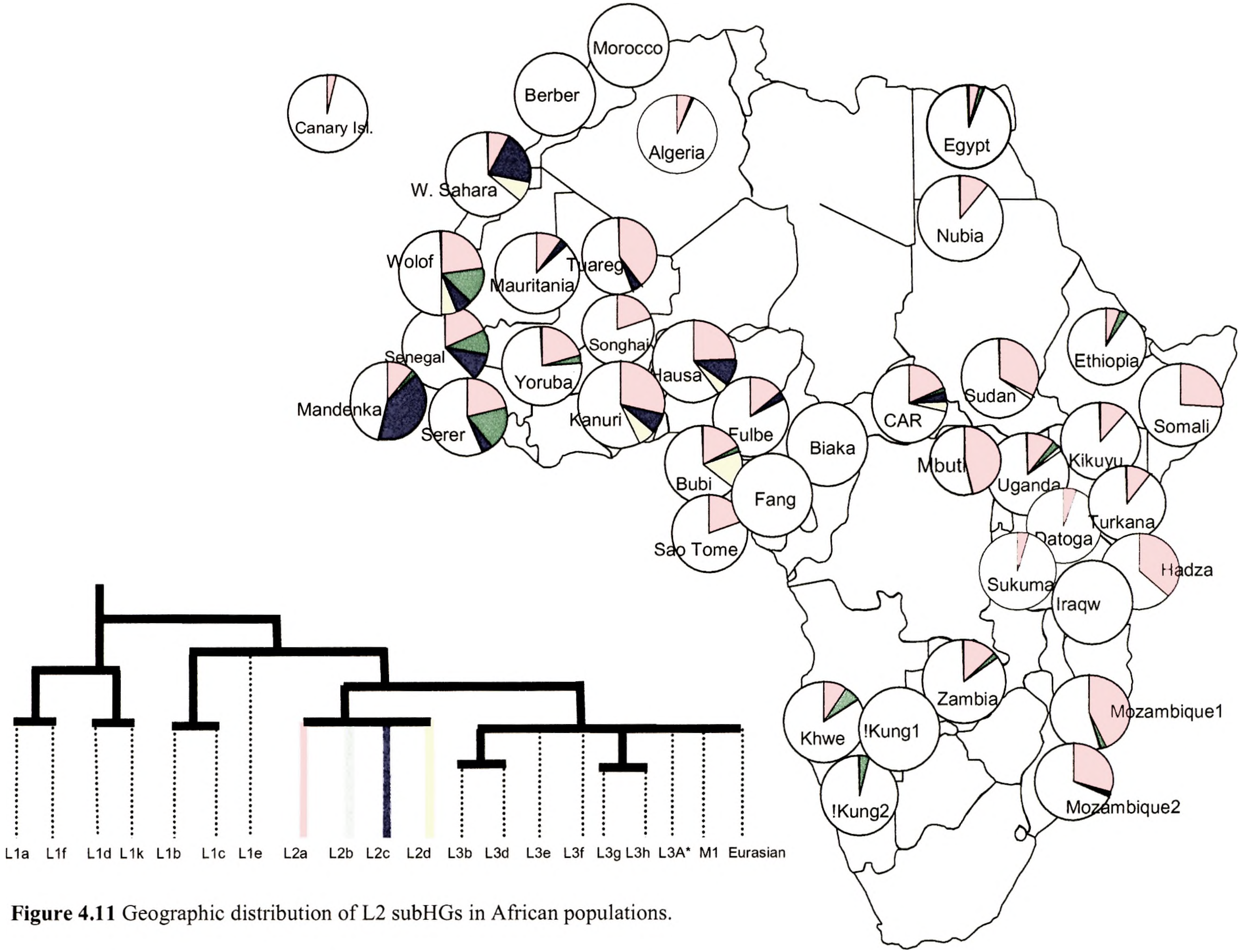


**Figure 4.11** Geographic distribution of L2 subHGs in African populations.

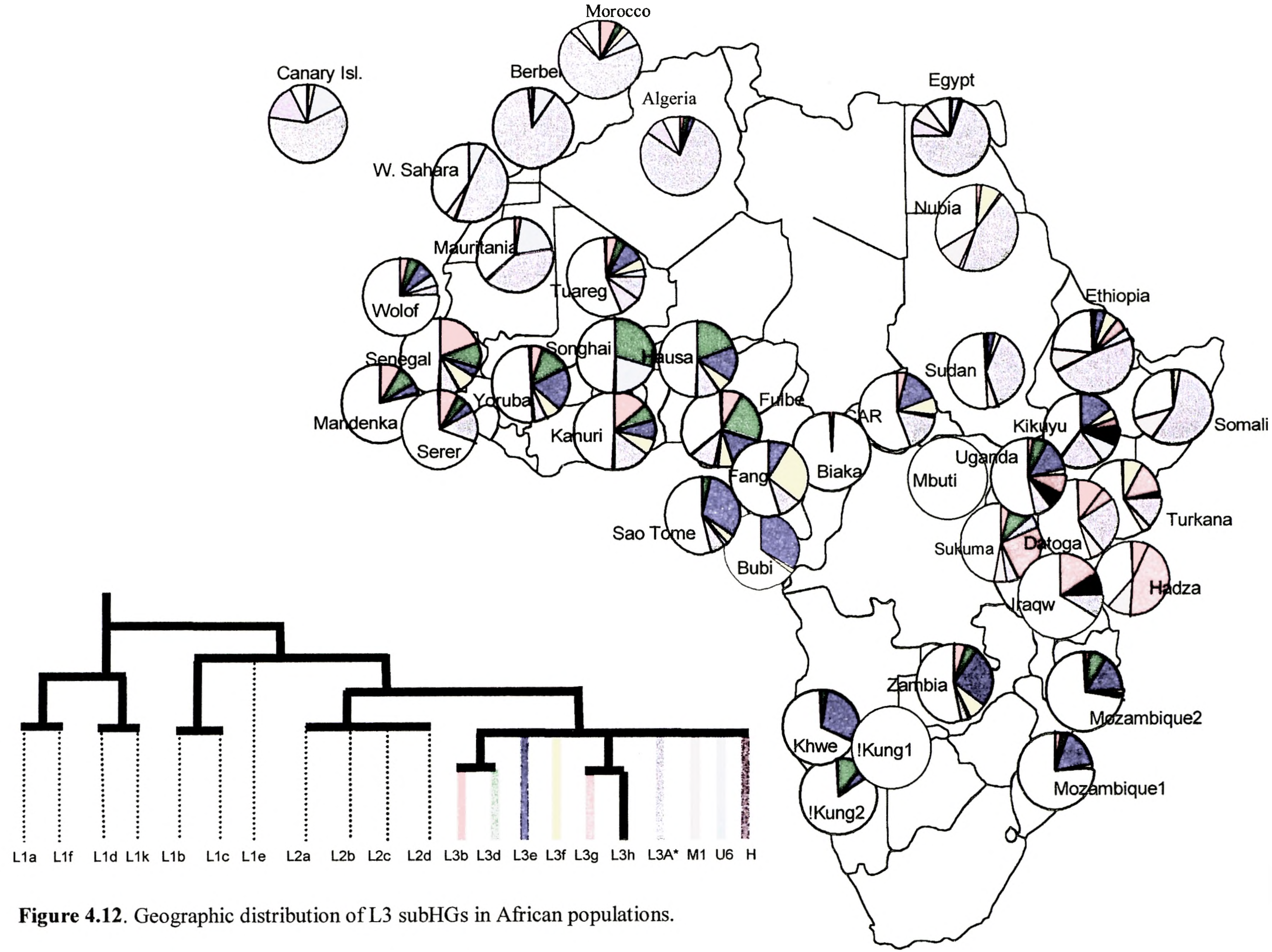


**Figure 4.12**. Geographic distribution of L3 subHGs in African populations.

The geographic distribution of mtDNA subHGs from HG L3 within Africa is depicted in fig. 4.12. SubHG L3e was generally the most frequent and wide-spread L3 subHG in sub-Saharan Africa, although it was absent from the non-Bantu speaking populations of east Africa. L3e was present but not very frequent in north-east and north-west African populations. L3f also had a fairly wide distribution, although it was not as frequent as L3e; its seemingly high frequency in the Fang is mitigated by the small sample size of this population (N=11, Pinto et al. 1996). L3g and L3h were specific to east Africa, whilst L3b and L3d were most frequent in west Africa and extended into east and southern African populations. Relatively large proportions of the L3 mtDNA types in north-east, east and north-west African populations were as yet further uncharacterised (L3A*), and several mtDNA lineages belonging to the Eurasian HGs M and N were observed these populations, including M1, U6 and H. SubHG M1 was mostly found in north-east and east Africa, and also occurred at low frequency in some north-west African populations. U6 was observed largely in north-west and west Africa, and sequences identical to the reference sequence (Anderson et al. 1981, Andrews et al. 1999; HG H) were present in the Canary Islands, Egyptians and the Tuareg (fig. 4.12). It is also likely that some of the mtDNA types within the L3A* category belong to other Eurasian HGs.

### 4.1.3.2 *Genetic affinities of African populations using mtDNA data*

Genetic affinities of African populations were assessed using PC plots and AMOVA analyses. Two PC plots were drawn, one using high-resolution HVR1 sequence data from 41 African populations, and the other using lower-resolution subHG frequency data from 41 African populations (N=2212, Table 2.3). AMOVA analyses were also performed using the two different data types.

*PC plots*

In PC plots drawn both with HVRI sequence data and with mtDNA subHG frequency data from 41 populations, most of the populations clustered closely together, with only a few outliers (figs. 4.13 and 4.14). In both plots, the two Biaka samples (populations 40 and 41) and one of the !Kung samples (3) were outliers, whilst Berbers (29) were also outliers in the PC plot drawn from subHG frequency data. The placements of these populations in the plots can be attributed to the extremely high frequencies of L1c mtDNA types in the Biaka, L1d mtDNA types in the !Kung, and L3A* types in the Berber (see figs. 4.10 and 4.12).

In each plot (figs. 4.13, 4.14), populations from individual geographic regions were placed closely to each other (each colour represents a different geographic region), indicating that mtDNA variation in Africa is strongly structured by geography. Only a few populations did not cluster with other populations from their geographic region of origin. These included the Mbuti Pygmies (population 30) from central Africa, who were similar to the Iraqw (4) and Datoga (5) populations from east Africa; the Sudanese (population 36), who were more similar to Nubians (9) and Ethiopians (10) than to other central African populations when subHG frequency data were used; and the Hadza (population 23) from east Africa.

Populations from central Africa (as well as from Mozambique, shown in red) clustered particularly closely together in both plots, indicating extremely close genetic affinities (figs. 4.13, 4.14). Three of the four populations examined in the present study (Zambians, Ugandans and non-Pygmies from the C.A.R.) were placed within the "central African" cluster, whilst the Biaka were outliers. In the plot drawn using HVRI sequence data (fig. 4.13), Ugandans (population 32) were placed close to the Turkana (7) and Sukuma (8) from east Africa, Zambians (33) were most similar to Mozambiquans (31, 35) and the Fang (34) from Equatorial Guinea, and the non-Pygmies from the C.A.R showed close affinities with the Fang (34) and Bubi (39) from Equatorial Guinea, Sudanese (36), Sao Tome e Principe (38), and Mozambiquans (35). In the plot drawn using subHG frequency data (fig. 4.14), the three non-Pygmy populations from this study were most closely related to each other, and to Mozambiquans.

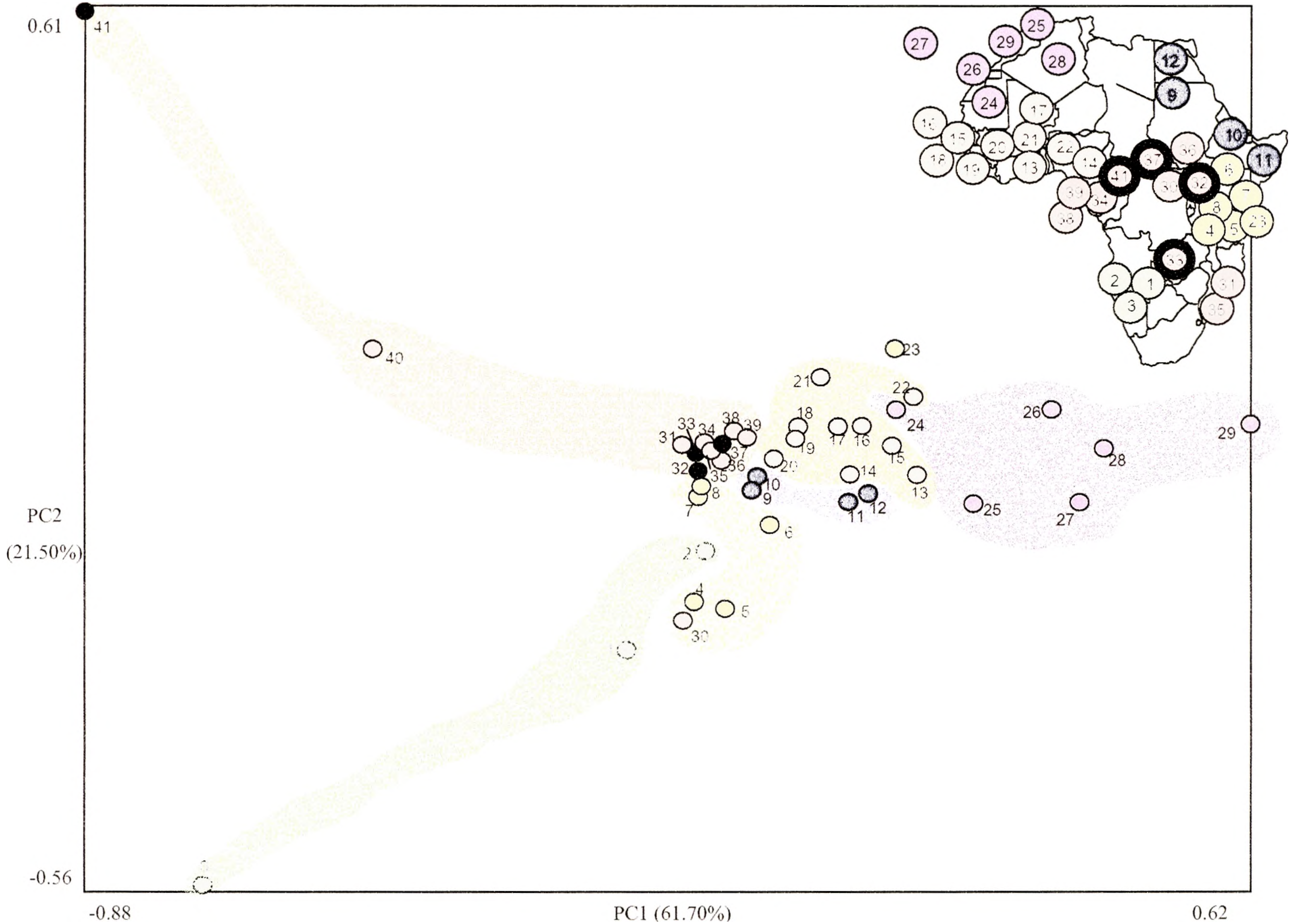


**Figure 4.13** PC plot constructed using mtDNA HVRI sequence data, showing the relationships among 41 African populations. Numbers represent populations as in Table 2.3. Circles in black or ringed in black indicate populations examined in this study.

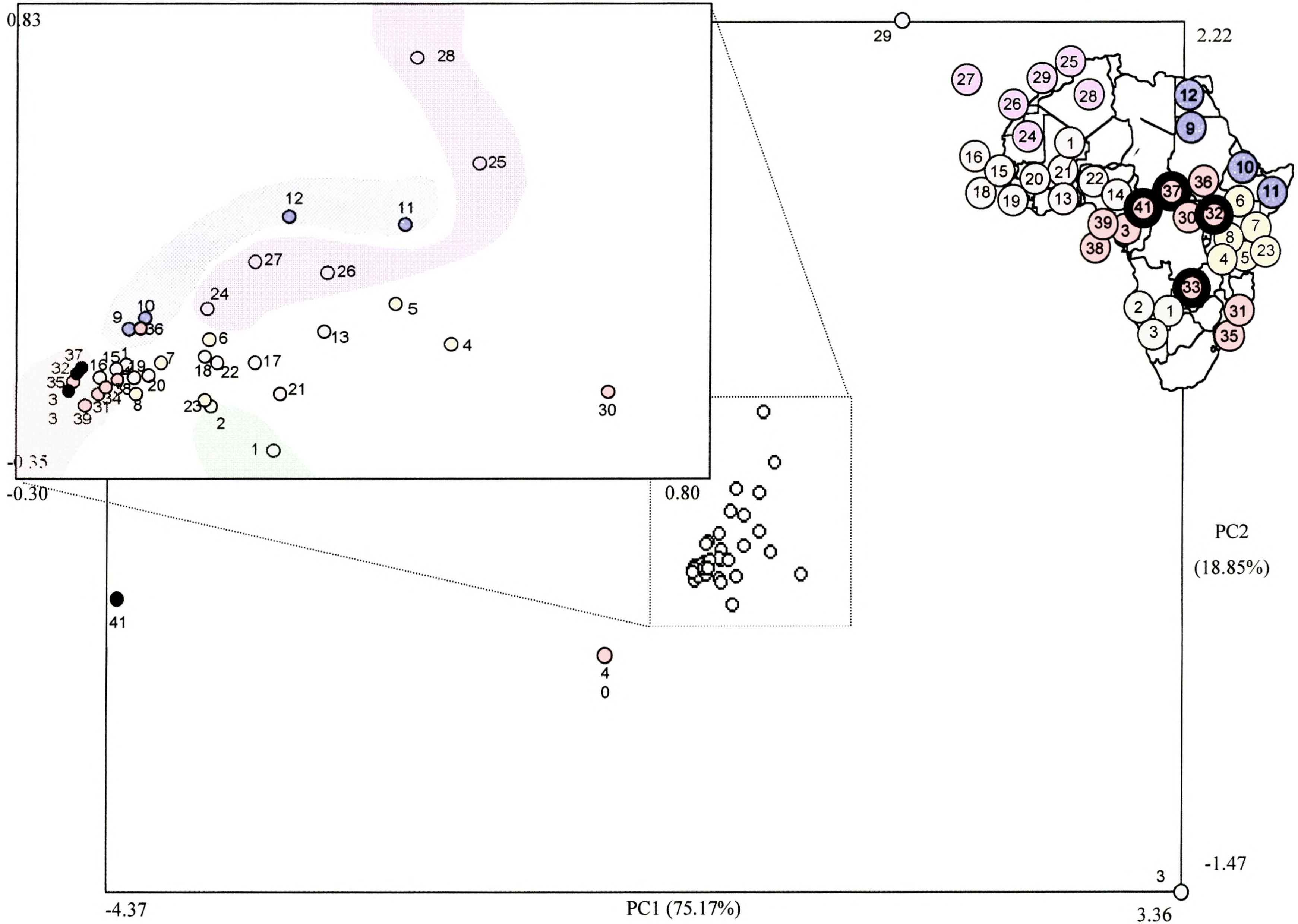


**Figure 4.14** PC plot constructed using mtDNA subHG frequency data, showing the relationships among 41 African populations. Numbers represent populations as in Table 2.3. Circles in black or ringed in black indicate populations examined in this study.

A comparison of the two plots showed that all populations appeared more closely related to each other when subHG data were used than when HVRI sequence data were used. Also, central, west and east African populations overlapped more in the PC plot drawn from subHG frequency data (fig. 4.14) than in the PC plot constructed using HVRI sequence data (fig. 4.13). These results are probably due to the effects of high-frequency subHGs such as L1a, L2a, L3e and L3A*, which make populations appear more similar to each other than when higher-resolution HVRI sequences are used.

*AMOVA analyses*

Genetic affinities among 41 African populations were also assessed using AMOVA analyses. MtDNA HVRI sequence data or subHG frequency data from 41 African populations were grouped according to a series of linguistic or geographic criteria (Table 4.8A, 4.8C, and Table 2.3) and analysed by AMOVA. "Ethnic criteria" i.e. the placement of Pygmy populations and of Khoisan populations into separate groups, were also used in combination with geographic or linguistic criteria (Table 4.8B, 4.8D).

The grouping of HVRI sequence data according to either geographic regions or linguistic groups was poorly supported by AMOVA analyses, since there was much more intra-group ($F_{SC}$) variance than inter-group ($F_{CT}$) variance (Table 4.8A, 4.8C). Groups were much more strongly supported (lowest $F_{SC}$ values and highest $F_{CT}$ values) when a combination of geographic and ethnic criteria, or linguistic and ethnic criteria, were used to define groups (Table 4.8B, 4.8D). The best results (lowest $F_{SC}$ values and highest $F_{CT}$ values) were obtained when populations were grouped using a combination of five geographic regions (north-east, north-west, east, west and combined south-central) and ethnic criteria (Table 4.8B). The same results were achieved with subHG frequency data (Table 4.8). Thus mtDNA HVRI or subHG frequency ht data from 41 African populations were better grouped according to a combination of geographic and ethnic criteria, than according to language group.

**Table 4.8** Results of AMOVA tests using HVRI sequence data and mtDNA subHG frequency data from 41 African populations

| | | | | | | | | | | A HVRI sequence data from 41 populations | | | | B mtDNA subHG frequency data from 41 populations | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Criteria | No.of groups | Groups[a] | | | | | | | | b | c | d | P | b | c | d | P |
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | | | | | | | | |
| A) Geography[e] (2 regions) (North vs sub-Saharan) | 2 | NE + NW | C+W+E+S | | | | | | | 7.25 | 12.46 | 80.29 | P<0.001 | 17.46 | 12.37 | 70.16 | P<0.001 |
| A) Geography[e] (4 regions) | 4 | NW+W | NE+E | C | S | | | | | 3.97 | 12.73 | 83.29 | P<0.005 | 5.19 | 17.1 | 77.71 | P<0.001 |
| A) Geography[e] (all 6 regions separate) | 6 | NE | NW | W | E | C | S | | | 6.04 | 10.72 | 83.24 | P<0.001 | 11.3 | 11.49 | 77.22 | P<0.001 |
| B) Geography[e] and ethnicity | 8 | NE | NW | W | E | C | Pygmy | S | Khoisan | 12.76 | 4.40 | 82.84 | P<0.001 | 17.05 | 5.82 | 77.12 | P<0.001 |
| **B) Geography[e] and ethnicity** | **7** | **NE** | **NW** | **W** | **E** | **S+C** | **Pygmy** | **Khoisan** | | **13.72** | **4.48** | **81.80** | **P<0.001** | **18.27** | **5.9** | **75.83** | **P<0.001** |
| B) Geography[e] and ethnicity (central group split) | 7 | NE | NW | E | W | S | Pygmy | Khoisan | | 12.91 | 4.62 | 82.48 | P<0.001 | 17.07 | 6.25 | 76.68 | P<0.001 |
| C) Language[f] | 5 | AA | NS | NK | KS | IE | | | | 3.93 | 13.23 | 82.83 | P<0.05 | 10.75 | 13.71 | 75.53 | P<0.001 |
| D) Language[f] and ethnicity | 6 | AA | NS | NK | KS | IE | Pygmy | | | 11.04 | 7.72 | 81.24 | P<0.001 | 15.78 | 9.34 | 74.88 | P<0.001 |
| D) Language[f] and ethnicity | 6 | AA | NS incl Mbuti | NK | KS excl Hadza | IE | Biaka | | | 13.19 | 6.05 | 80.76 | P<0.001 | 18.05 | 7.62 | 74.33 | P<0.001 |

[a] Populations were placed into groups according to geographic or linguistic criteria; refer to Table 2.3

[b] percent variance among groups

[c] percent variance among populations within groups

[d] percent variance within populations

[e] Geographic regions NE=north-east, NW=north-west, W=west, E=east, C=central, S=south

[f] language families AA = Afro-Asiatic, NK = Niger-Kordofanian, NS = Nilo-Saharan, KS = Khoisan , IE=Indo-European

### 4.1.4 *Correlations between mtDNA data and geography*

The results of correlation analyses using pairwise geographic distances and pairwise $F_{ST}$ genetic distances measured from mtDNA data, were very similar to results obtained using Y chromosome data. Statistically significant correlations were obtained only when data from populations spanning the entire African continent were analysed (using mtDNA HVRI sequence data from 41 populations, r=0.177, $P<0.005$; using mtDNA subHG frequency data from 41 populations, r=0.164, $P<0.05$). Graphs showing these correlations are shown in Appendix 8.9A.

Pairwise $F_{ST}$ distances among 31 populations from sub-Saharan Africa (i.e. excluding populations from north-west and north-east Africa, Table 2.3) were not statistically correlated with pairwise geographic distances among them (using HVRI sequence data, r=0.083, $P=0.176$; using subHG frequency data, r=-0.004 and $P=0.456$). Graphs showing these correlations are shown in Appendix 8.9B. Also, correlations between pairwise genetic $F_{ST}$ distances and pairwise geographic distances among 15 populations from central and southern Africa only were not statistically significant (using HVRI sequence data, r=-0.099, $P=0.746$; using subHG frequency data, r=-0.024 and $P=0.531$). Graphs showing these correlations are shown in Appendix 8.9C. Similarly to the Y chromosome data, these results suggested that geographically separated populations in sub-Saharan Africa are genetically closely related though their female lineages and/or that populations living in close proximity have different mtDNAs. These scenarios could have been caused by migrations of populations across sub-Saharan Africa, by population admixture, or by genetic barriers between neighbouring populations.

Again similarly to Y chromosome analyses, genetic distances among 17 African populations speaking languages classified within the Niger-Congo language group (Table 2.3) were poorly correlated with geographic distances among them (using HVRI sequence data, r=-0.232, $P=0.538$; using subHG frequency data, r= -0.061 and $P= 0.605$). Graphs showing these correlations are shown in Appendix 8.9D. These results suggest that the lack of correlation between genetic and geographic distances seen in sub-Saharan African populations above is largely due to a similar lack of correlation in populations from this region speaking Niger-Congo languages. Results also further strengthen the hypothesis that this lack of correlation is due to the effects of migration, since populations speaking Niger-Congo languages are known to have experienced recent and rapid migrations within sub-Saharan Africa (Phillipson 1977, Ehret 1998).

## 4.2 DISCUSSION

The present-day patterns of genetic variation in African populations are complex and harbour information that can be traced to both ancient and recent evolutionary events. Data collected in this study have facilitated an understanding of the structure of mtDNA variation in central Africa, and have allowed the genetic affinities of four central African populations to be examined.

### 4.2.1 *Ancient mtDNA lineages in central African populations*

The subHGs within mtDNA HG L1 are the most ancient mtDNA lineages described globally, and includes the MRCA of human mtDNAs, which is at least 150-170 000 years old (Horai et al. 1995; Ingman et al. 2000). In the present study, subHGs L1a and L1f are suggested on the basis of phylogenetic analysis to represent the most ancestral mtDNA types known in human populations, followed closely by subHGs L1d and L1k, and later by subHGs L1b, L1c and L1e. Note that calculated dates of TMRCA of these subHGs (Table 4.6) do not necessarily match their order of evolution since estimates are highly influenced by sample sizes available for each subHG.

SubHG L1a is thought to have originated in east Africa and then to have dispersed widely across sub-Saharan Africa (Salas et al. 2002). The cline of L1a frequencies observed in central African populations in the present study is consistent with dispersal from an east African source. In addition, only one mtDNA type from the ancestral east African clade L1a* was observed in central Africans in the present study; all other mtDNA types were from the more derived clades L1a1a or L1a2, supporting the suggestion by Salas et al. (2002) that subHG L1a did not evolve in central Africa.

Soodyall et al. (1996) suggested that the L1a2 clade (associated with the 9-bp deletion) might have evolved in central Africa. However the frequencies of L1a2 in the populations from central Africa in the present study (excluding Zambia) were much lower than the frequencies of L1a2 in south-eastern Africa e.g. Biaka 3%, non-Pygmies from the C.A.R. 0.4%, Uganda

7% (Table 4.1), compared to Mozambique 15% (Salas et al. 2002), Zimbabwe 18% (H. Soodyall, unpublished data), Zambia 15% (present study), Malawi 26.7% and South African non-Khoisan populations 16.3% (Soodyall et al. 1996, inferred from the frequency of the 9-bp deletion). The super-imposition of mtDNA types from this study and from Soodyall et al. (1996) onto fig. 4a in Salas et al. (2002, p. 1095), also supported this conclusion (not shown). This suggests that the mtDNA 9-bp deletion evolved not in central Africa but in south-eastern Africa. Therefore all mtDNA types from subHG L1a can be assumed to have an east (or south-east) African origin.

SubHG L1f is the sister clade of L1a, as seen in this study. SubHG L1f was relatively rare in the central African populations in the present study (overall frequency of 0.033), but was common (0.109) in Uganda. In a phylogenetic network constructed with L1f mtDNA types, mtDNA types from Ugandan individuals were the most ancestral. This was also true when a network defined by HVRI data only was constructed (including 13 individuals from Uganda and Zambia in this study, 2 Turkana and 1 Kikuyu from Watson et al. (1997), 3 Iraqw from Knight et al. (2003) and 1 Nubian from Krings et al. (1999); network not shown). This suggests that L1f evolved in the Lake Victoria area, and spread to a limited extent to populations in neighbouring areas. The absence of L1f in Bantu-speakers from Mozambique (Pereira et al. 2001; Salas et al. 2002) suggests that this subHG was not involved in the recent "Bantu Expansion".

The next pair of subHGs in the African mtDNA phylogeny consists of subHGs L1d and L1k. Both of these subHGs are frequent only in southern African Khoisan-speaking populations. In this study, these subHGs were each represented by single individuals from Zambia. This suggests a very small degree of gene flow from Khoisan populations which may be explained by the geographic proximity of Zambians to areas presently and anciently populated by Khoisan-speakers.

Overall, the most ancestral mtDNA subHGs, L1a, L1f, and L1dk, appear to have evolved in east Africa, in the Great Lakes region, and in southern Africa respectively, and not in central Africa. MtDNA data therefore support the hypothesis based on archaeological evidence

(Cahen 1982; Phillipson 1985) that the central African region (excluding the Great Lakes area) was populated at a somewhat later date (Middle Stone Age) than eastern or southern Africa (populated since the Early Stone Age).

SubHG L1c is thought to have evolved in central Africa (Salas et al. 2002), and was present in all central African populations examined in this study. The frequency of L1c was significantly higher (P<0.001) in the Biaka from the C.A.R. (0.947) than in the other three central African populations studied (non-Pygmy C.A.R. 0.084, Zambia 0.156, Uganda 0.076, figs. 4.1, 4.10). The observation of subHG L1c at high frequency in the Biaka confirms that this subHG may be a signature of the original central African mtDNA gene pool, since Pygmy populations are thought to be descended from the autochthonous inhabitants of central Africa (Cavalli-Sforza et al. 1986). The absence of L1c in Mbuti Pygmies (fig. 4.10) should not be regarded as contradictory to this hypothesis since generally the Mbuti Pygmies' mtDNA seems much more closely related to that of east African populations than to other central African populations (figs 5.13 and 5.14).

HG L1c sublineages had different distributions: L1c1a1, L1c4 and L1c5 were found nearly exclusively in the Biaka Pygmy populations, whilst L1c1 / L1c*, L1c2 and L1c3 were found in other central African populations. The evolution of the Biaka-specific clades from L1c* types suggests that the Biaka mtDNA gene pool evolved from that of the proto-central African population by isolation and drift. The oldest Biaka-specific clade, L1c1a1, was dated to 22 409 years ago, whilst the other clades were younger (L1c5: 16 308 years, L1c4: 7 191 years). These dates represent the oldest possible ages for the start of Biaka divergence from other populations since sequence divergence precedes population divergence. Thus the Biaka may have separated from other central African populations no earlier than ~20 000 years ago. This time period coincides with the onset of the Last Glacial Maximum ~18 kya in Africa, in which retreat of the tropical forest occurred due to dryness (Oliver 1990, Vansina 1990). If the Biaka ancestors preferred forest habitat as the Biaka do today (Cavalli-Sforza 1986), they may have become secluded in the forest habitat refuges that remained during the LGM. The very high frequencies of specific L1c mtDNA types in the Biaka population are also suggestive of a maternal founder effect in this population. The presence of single L1c1a and L1c4 mtDNA types in the non-Pygmy populations of the C.A.R. (Table 4.3) suggests a very low level and recent occurrence of maternal gene flow from Biaka Pygmies to their non-Pygmy neighbours. Bandelt et al. (2001) suggested a central African origin for the L3 subHG L3e ~45 000 years ago. This subHG was extremely common and diverse in the three non-Pygmy populations from central Africa in this study, and especially in Zambians (Table 4.1, fig. 4.12). When L3e mtDNA types from the present study were superimposed onto the networks of L3e mtDNA types drawn by Salas et al. (2002, fig 9b, p. 1104; not shown here), mtDNA types from central African populations matched the L3e1, L3e2 and L3e3 founder types, and also matched many of the other centrally placed and common L3e1 and L3e2 types. These observations supported the possibility of a central African origin of L3e in central Africa, and possibly in south-central Africa in the region of present-day Zambia.

### 4.2.2 *Other contributions to the central African mtDNA gene pool*

As discussed above, mtDNA types from subHGs L1a, L1d and L1k represent rather recent contributions to the central African mtDNA gene pool from east and southern Africa. Other contributions are represented by subHGs L1b and L1e, HG L2 and most of the lineages within HG L3.

#### 4.2.2.1 *Gene flow from west African populations*

SubHGs L1b and L1c are united as a pair in the mtDNA phylogeny by the 7055 *Alu*I RFLP, but have different geographic distributions centring in west and central Africa respectively. A central African origin of L1b, followed by a bottleneck and re-expansion in west Africa has been suggested by Salas et al. (2002) to explain this pattern, and seems plausible. Several of the central African L1b types in this study were classified as belonging to the more ancestral L1b (not L1b1) clade, which is also consistent with a central African origin of L1b. Dating in this study using HVRI and II sequence data confirmed a more recent TMRCA of L1b than of L1c. In the present study L1b was found at its highest frequency in the non-Pygmies of the C.A.R. (the most western of the population sampled), and was absent in Uganda, the most eastern population sampled, suggesting gene flow from a later west African source. This could also support an origin of L1b in west-central Africa. The presence of L1b in Zambians may support the idea that this region was partially populated by Bantu-speakers coming from the western stream of the Bantu migration.

SubHGs L2b, L2c, L2d, L3b and L3d are also assumed to have originated in west Africa (Salas et al. 2002). All together the subHGs of clearly west African origin (L1b, L2b, L2c, L2d, L3b and L3d) accounted for 0.018 of Biaka Pygmies, 0.274 of non-Pygmies from the C.A.R., 0.152 of Ugandans, and 0.167 of Zambians (Table 4.1).

4.2.2.2 *Gene flow from east African populations*

In the present study, the early branching clade of L2a called L2a2 was found mostly in the non-Pygmy population of the C.A.R. (fig. 4.1), which might indicate a west African origin of this clade and subHG. However when these mtDNA types were superimposed on the network of L2a lineages in fig. 6 of Salas et al. (2002; p. 1100; not shown here), they were found to be derived from east African L2a-α2 types. Therefore it is suggested that L2a2 mtDNA types in central African populations in this study could have an east African origin. Most of the other L2a mtDNA types found in the present study belonged to the L2a1-β clades, which are extremely widely distributed in sub-Saharan Africa; it is not possible to decipher the origins of these mtDNA types in the central African populations studied here.

SubHGs L1e and L3g are found almost exclusively in east Africa and therefore are thought to have evolved there (Salas et al. 2002). MtDNA types from L3f also probably originated in east Africa, with types from the L3f1 clade dispersing to west Africa (Salas et al. 2002). All three of these subHGs were found at low frequencies in all three non-Pygmy central African populations examined in this study (fig. 4.1). Their distribution indicates that the spread of L3f1 types from east to west Africa described by Salas et al. (2002) may have been accompanied by the spread of L1e and L3g types. Interestingly, subHG L3h, the sister clade of L3g, is nearly exclusively found in Uganda and east African populations (fig. 4.12; Watson et al. 1997; Knight et al. 2003), and did not participate in this east-to-west dispersal. Based on its frequency and diversity, L3h can be assumed to have originated in the Uganda region and spread to other east African populations.

Together the subHGs of clearly east African origin (L1a, L1e, L2a2, L3f and L3g) accounted for 0.035 of Biaka Pygmies, 0.242 of non-Pygmies from the C.A.R., 0.315 of Ugandans, and 0.240 of Zambians (fig. 4.1). Note that the east African lineage M1 (Quintana-Murci et al. 1999; fig. 4.12), was not found in central African populations in the present study.

### 4.2.2.3 *The "Bantu Expansion"*

The "Bantu Expansion" both introduced new mtDNA variation into central Africa, and dispersed lineages that evolved in central Africa to other parts of the continent.

Lineages that were introduced to central Africa via the "Bantu Expansion" should be expected to have come mostly from west Africa, since both the western and eastern streams of the migration originated in the region of Cameroon/Nigeria (see Introduction). MtDNA lineages of west African origin that have previously been suggested to be associated with the "Bantu Expansion" include subHGs L3b, L3d and L2a (Watson et al. 1997; Pereira et al. 2001; Salas et al. 2002).

None of the L3d mtDNA types found in central African populations in the present study matched the founder type described by Salas et al. (2002, fig. 9a. p.1104), but were further derived. Two of the seven L3d types from this study (types 23 and 25, table 4.5) matched the common L3d1 type found in Bantu-speaking Mozambiquans (Salas et al. 2002), and the most common L3d type in this study was found in Zambians and Ugandans, both of whom are Bantu-speaking populations. Therefore data from this study support the dispersal of L3d from west Africa via the "Bantu Expansion". The presence of L3d in Khoisan populations (Watson et al. 1997; Chen et al. 2000) suggests more specifically that L3d is a marker of the western stream of the "Bantu Expansion". MtDNA data from the D.R.C. and Angola could confirm this possibility.

SubHG L3b, the sister clade of L3d, was also present in the central African populations of this study (fig 4.1). Unlike L3d, L3b's distribution also extends into non-Bantu-speaking east African populations, and is absent from Khoisan populations (fig. 4.12; Watson et al. 1997; Chen et al. 2000; Knight et al. 2003). This pattern confirms that this subHG is likely to be a signal of the "Bantu Expansion", but may have been transmitted by the eastern stream of the "Bantu Expansion", rather than the western stream.

Starbursts of specific L2a mtDNA types (clades L2a1a, and L2a1b) in Mozambiquans led to the suggestion that these clades were markers of the "Bantu Expansion" (Pereira et al. 2001; Salas et al. 2002). However the rarity of these clades in west Africans, central Africans in this study, and east Africans, suggests instead that these clades are signals of a very recent founder effect in Bantu-speaking populations of Mozambique, and were not gained along the suggested routes of the "Bantu Expansion".

SubHG L2b may be an additional west African linage that was re-distributed via the "Bantu Expansion". This subHG is rare outside of west Africa (fig. 4.11). However in the instances where L2b is found outside of west Africa, it usually occurs in populations that speak Bantu languages or who have been in contact with Bantu-speakers, including the Bubi, Khoisan populations, the three non-Pygmy populations of this study and Mozambiquans (data from Pinto et al. 1996; Watson et al. 1997; Pereira et al. 2001; Salas et al. 2002). This potentially indicates that L2b played a minor role in the "Bantu Expansion".

SubHG L1a is a lineage of east African origin that has previously been suggested to be associated with the "Bantu Expansion" (Soodyall et al. 1996; Watson et al. 1997). Salas et al. (2002) suggested that L1a was transmitted from east Africans to Bantu-speaking populations in the Lake Victoria / Uganda region. From the Great Lakes region, speakers of eastern Bantu languages may have carried L1a southwards, as well as westwards in their later "reverse' expansions (Vansina 1984). This hypothesis is supported by data based on the geographic distribution of L1a mtDNA types from the present study and published data (fig. 4.10).

However the involvement of the above-motioned subHGs *only* in the "Bantu Expansion" must be regarded with caution. As noted by Pereira et al. (2001), mtDNA in Bantu-speaking African populations does not show a massive reduction in diversity parallel to that seen with the Y chromosome, suggesting that the mtDNA founder effects associated with the "Bantu Expansion" were not on a scale equal to those involving the Y chromosome. Some of the dispersals of mtDNA subHGs in Africa may date to much earlier times than the "Bantu Expansion". This is especially true for mtDNA subHGs such as L2b and L1a, whose distribution extends to north Africa, where their presence cannot be explained by the 'Bantu Expansion".

The "Bantu Expansion" would also have dispersed lineages that evolved in central Africa to other parts of the continent. Bandelt et al. (2001) previously suggested subHG L3e1 (and L3e2 and L3e3 to a lesser extent) in this context. Many of the L3e mtDNA types found in Bantu-speaking Mozambiquans match those found in central Africans in the present study, supporting this hypothesis. L1c2 and L1c3 may also have dispersed from a central African origin via the "Bantu Expansion" since these mtDNA types were found in Bantu-speaking populations of central, east and south-east Africa (present study; Salas et al. 2002; Pereira et al. 2001).

### 4.2.2.4 *Eurasian lineages*

Seven mtDNA types in the present study either belonged to the African HG L3A* or to the Eurasian HG N. Their control region variation did not match known motifs of HGs within superHG N (Macaulay et al. 1999), and these types can therefore be considered to be of non-Eurasian descent.

Only one mtDNA in the present study could definitely be assigned to a Eurasian mtDNA haplogroup. This single mtDNA type from a single Ugandan individual putatively belonged to mtDNA HG X2c (Table 4.5). This mtDNA type may have been gained from north Africans such as Egyptians: HG X2 occurs in Egyptians at a frequency of 0.5% (Reidla et al. 2003), and parts of Uganda belonged to the Egyptian province of Equatoria established in 1871 (Briggs

1996). Arab slave traders, whose presence in Uganda dates since the mid-19th century (Briggs 1996), may also have contributed this mtDNA type; X2c occurs in the Near East at a frequency of 2.9% (Reidla et al. 2003). Alternatively the X2 mtDNA type could represent recent gene flow from the British, who colonised Uganda in 1894, and in whom X2 occurs at a frequency of 0.9% (Reidla et al. 2003).

### 4.2.3 *The genetic affinities of central African populations using mtDNA data*

Each central African population examined had a unique mtDNA gene pool (fig. 4.1) and all pairs of populations were significantly different to each other, both when HVR sequence data were used, and when mtDNA subHG frequency data were used (*P*<0.001, Table 4.6). Despite these differences, the Adamawa-Ubangian-speaking (non-Pygmy) population from the C.A.R., and the Bantu-speaking Ugandan and Zambian populations were fairly closely related to each other in phylogenetic analysis, whilst the Bantu-speaking Biaka Pygmies from the C.A.R. differed greatly from the other central African populations (figs. 4.9, 4.13, 4.14). This indicated that the use of linguistic subfamilies was not a reliable method to predict the genetic similarity of [central] African populations.

This study supported others who have shown that the mtDNA of Biaka Pygmies differs greatly from that of other African populations (Chen et al. 1995; Chen et al. 2000). The majority (0.947) of mtDNA types in Biaka from the present study belonged to the Biaka-specific subHGs L1c1a, L1c4 and L1c5, while only a few of their mtDNA types belonged to subHGs found in other African populations (subHGs L1a and L3b, total frequency of 0.053; fig. 4.1). Only two non-Pygmy individuals from the C.A.R. (and none from other populations examined) had mtDNA types from Biaka-specific subHGs (Table 4.3). These data support extreme genetic isolation of Biaka maternal lineages, and mtDNA gene flow between Biaka and neighbouring populations has been very limited in either direction. Therefore any gene flow between Biaka and non-Pygmy populations inferred from autosomal loci (Cavalli-Sforza 1986; Wijsman 1986; Cavalli-Sforza et al. 1994) can be assumed to be due to male-sponsored gene flow between these populations. This was confirmed by studies of Y chromosome variation in the Biaka in the present study.

### 4.2.4 *Structure of mtDNA variation in Africa*

Comparisons of mtDNA from populations spanning the major geographic regions of Africa showed that mtDNA variation is highly structured in Africa. AMOVA testing supported the grouping of mtDNA data from African populations by geography rather than by language family (table 4.7). The strong structuring of mtDNA types from African populations by geography was also supported by PC plots (figs. 4.13 and 4.14). In most cases, high-resolution mtDNA HVR sequence data from populations from the geographic regions of north-west Africa, west Africa, north-east Africa and east Africa were easily distinguished, although there was a slight degree of overlap between adjoining regions (fig. 4.13). MtDNA from Khoisan-speaking populations from south-western Africa could also easily be identified (fig. 4.13, 4.14). This was consistent with previous phylogenetic analyses of mtDNA data from African populations which have shown consistent clustering of mtDNA gene pools from the broad geographical regions of eastern, western, southern and northern Africa (Watson et al. 1996; Pereira et al. 2001; Salas et al. 2002). These observations indicate that populations from these geographic regions have a large proportion of shared common mtDNA ancestry, and have experienced similar gene flow patterns.

The mtDNA gene pools of central African populations, with the exception of Biaka and Mbuti Pygmies, were very similar to each other and to those of south-east African Bantu-speaking populations (figs. 4.13, 4.14). As previously discussed, Biaka Pygmies' mtDNA was quite divergent to mtDNAs in other African populations (present study; Chen et al. 2000), whilst mtDNA from Mbuti Pygmies was quite closely related to mtDNA from non-Pygmy east African populations (present study; Salas et al. 2002). Within the central African group of populations, Ugandans, Zambians and Mozambiquans showed slightly closer affinities with east African populations (fig. 4.13), whilst the Fang, Bubi and the Sao Tome population showed slightly closer affinities with west African populations (fig. 4.13; Salas et al. 2002).

Overall, from a mtDNA perspective, the greatest amount of gene flow between central Africa and other geographic regions in Africa has been between central and south-east Africa. This was similar to observations made using Y chromosome data in this study, and may be partially attributed to the effects of the recent “Bantu Expansion”. There appears to have been comparatively little mtDNA gene flow among central African populations and those living in the north-west or north-east, which may be attributed to the Sahara Desert barrier between north and central Africa. MtDNA gene flow between east, central and west Africa has occurred at a considerable extent: it was estimated that mtDNA of combined east and west African populations have contributed to ~0.406 of the Zambian mtDNA gene pool, 0.467 of the Uganda mtDNA gene pool, 0.516 of the C.A.R. non-Pygmy mtDNA gene pool, and 0.053 of the Biaka mtDNA gene pool.

### 4.2.5 *Utility of HVRII sequence data*

The use of HVRII variation has allowed further resolution of mtDNA variation in African populations. Certain clades which were not recognisable from HVRI data alone, can be characterised by HVRII variation (e.g. L2c, L3e2a). In addition, the use of HVRII data has allowed phylogenetic relationships among subHGs to be clarified, especially within HG L1, where subHGs L1a and L1f could be paired on the basis of their shared HVRII variation, as could L1d and L1k (fig. 4.8).

There are some single HVRII sites which appear to be unique and stable, allowing subHGs to be defined, including 247 (defines HG L1), 151, 186A, 189C and 316 (define subHG L1c), 357 and 185T (L1b), 64 (L1a2), 143 (in a subset of L2a1β), 325 (L2c), 244 (subHG L3g) and 153 (subHG L3h). However the extent of HVRII sequence variation is far less than that of HVRI, and it appears that the same sites (e.g. 146, 150, 152, 189, 195) have evolved repeatedly in different lineages. This is consistent with the high mutation rate observed at such sites by Malyarchuk et al. (2002). Thus single HVRII sites are usually not diagnostic of subHG membership, and generally the use of the whole HVRII haplotype is necessary to allow subHG identity to be established.

# 5. COMPARISON OF MALE AND FEMALE MIGRATION RATES IN AFRICAN POPULATIONS

Y chromosome and mtDNA data were used to compare male and female migration rates in the four central African populations for which both data types were collected in this study. Migration rates were also compared in the total number of comparative African populations for each data type (Tables 2.2, 2.3). Analyses included comparisons of total $F_{ST}$ and Nυ estimates, comparisons of the apportionment of diversity estimated in AMOVA analyses, and comparisons of regressions of genetic and geographic distances.

## 5.1 *Migration rates in central African populations*

First, mtDNA and Y chromosome data from four central African populations (Zambians, Ugandans, Biaka and non-Pygmies from the C.A.R.) were compared (Table 5.1a). Estimates of Nυ in the four central African populations were approximately seven times higher when calculated from any type of Y chromosome data than when calculated from any type of mtDNA data (Table 5.1a). In addition, there were higher proportions of variance within populations using Y chromosome data, than the proportions of variance within populations calculated from mtDNA data (Table 5.1a). Together these observations suggested higher male than female migration rates in central Africa. However because haploid variation in the Biaka was consistently significantly different to variation in the non-Pygmy populations (Tables 3.2 and 4.7), the analysis was repeated excluding the Biaka (Table 5.1b).

When mtDNA and Y chromosome data from the three non-Pygmy populations were analysed (Table 5.1b), estimates of Nυ calculated from any type of mtDNA data were about twice that calculated from any type of Y chromosome data (Table 5.1b). Assuming equal effective population sizes, these observations suggested that female migration rate has been about double the male migration rate in central Africa. The slightly higher proportions of variance within populations observed using mtDNA data than using Y chromosome data. also supported a higher female than male migration rate in central Africa (Table 5.1b).

**Table 5.1** Comparison of $F_{ST}$ and Nυ values, and apportionment of diversity, using mtDNA and Y chromosome data from African populations. The highest values in each category are highlighted in bold.

|  | No of populations | $F_{ST}$ | Nυ | % variance within populations | % variance among populations |
|---|---|---|---|---|---|
| **a. four central African populations[1]** |  |  |  |  |  |
| mtDNA subHG frequency | 4 | 0.254 | 2.935 | 74.59 | 25.41 |
| mtDNA HVRI sequences | 4 | 0.260 | 2.852 | 74.04 | 25.96 |
| mtDNA HVRI and HVRII sequences | 4 | 0.251 | 2.987 | 74.92 | 25.08 |
| Y chromosome HG data | 4 | 0.053 | **17.929** | **94.72** | 5.28 |
| Y chromosome 5 STR ht data | 4 | 0.047 | **20.106** | **95.26** | 4.74 |
| Y chromosome 8 STR ht data | 4 | 0.046 | **20.944** | **95.44** | 4.56 |
| **b. three central African populations[2]** |  |  |  |  |  |
| mtDNA subHG frequency | 3 | 0.014 | **71.307** | **98.62** | 1.38 |
| mtDNA HVRI sequences | 3 | 0.014 | **68.638** | **98.56** | 1.44 |
| mtDNA HVRI and HVRII sequences | 3 | 0.018 | **54.525** | **98.20** | 1.80 |
| Y chromosome HG data | 3 | 0.026 | 37.745 | 97.42 | 2.58 |
| Y chromosome 5 STR ht data | 3 | 0.037 | 26.307 | 96.34 | 3.66 |
| Y chromosome 8 STR ht data | 3 | 0.035 | 27.482 | 96.49 | 3.51 |
| **c. all comparative African populations[3]** |  |  |  |  |  |
| mtDNA subHG frequency | 41 | 0.214 | 3.677 | 78.62 | 21.38 |
| mtDNA HVRI sequences | 41 | 0.160 | **5.265** | **84.04** | 15.96 |
| Y chromosome HG data | 20 | 0.281 | 2.564 | 71.94 | 28.06 |
| Y chromosome 5 STR ht data | 17 | 0.120 | **7.334** | **88.00** | 12.00 |
| Y chromosome 8 STR ht data | 13 | 0.141 | **6.068** | **85.85** | 14.15 |

[1] Zambians, Ugandans, and Biaka and non-Pygmies from the Central African Republic.

[2] excluding the Biaka

[3] refer to Tables 2.2 and 2.3

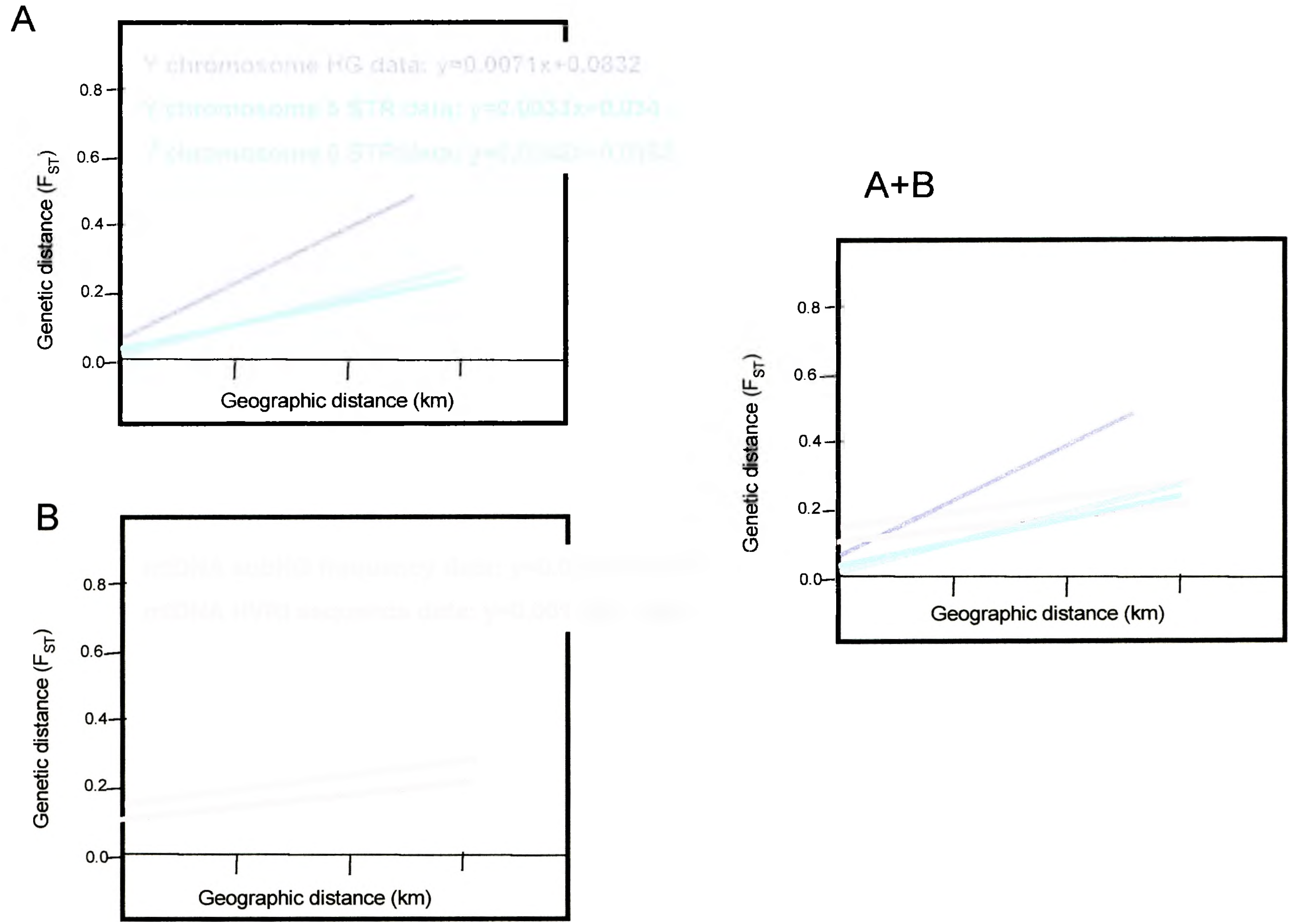


**Figure 5.1.** Graphs of regression lines showing the relationship between geographic distance and genetic distances in African populations calculated from (A) Y chromosome data and (B) mtDNA data.

The change in results from Table 5.1a to 5.1b indicated that the Pygmies of the C.A.R. may have experienced different migration patterns to other central African populations. MtDNA analyses in this population indicated that the Biaka mtDNA gene pool is highly population-specific, and that very little female gene flow either into or out of the Biaka population has occurred. Their Y chromosomes, on the other hand, indicated some similarity with other those of other central African populations, suggesting that male gene flow, and possibly migration, has been of greater extent than female migration in this population. This pattern is opposite to that observed above for other central African populations. The small sample size (three populations or one population) precluded the use of regression analysis to confirm these conclusions.

5.2 *Migration rates in pan-African populations*

Different types of mtDNA and Y chromosome data from populations spanning the African continent were used to compare male and female migration rates (Table 5.1c). The use of different data types from each locus yielded considerably different $F_{ST}$ and $N\upsilon$ estimates. Generally lower-resolution data (mtDNA subHG frequency data and Y chromosome HG data) produced a higher $F_{ST}$ values, lower migration rate, and lower within-population variance than values calculated from higher-resolution data (HVRI sequence data and Y chromosome STR ht data; Table 5.1c). These results highlight the concern that the use of different data types in such analyses may influence results (Stoneking 1998).

When the low-resolution mtDNA and Y chromosome data types were compared to each other (Table 5.1c), the estimate of $N\upsilon$ calculated from mtDNA subHG frequency data in African populations was 1.4 times greater than that calculated from Y chromosome HG data (Table 5.1c), indicating greater female than male migration across the African continent. This was similar to patterns observed elsewhere in populations from Africa, Europe (Seielstad et al. 1998) and Oceania (Kayser et al. 2001).

However the estimate of $N\upsilon$ calculated from high-resolution mtDNA HVRI sequence data was lower than estimates obtained using high-resolution Y chromosome STR ht data (Table 5.1c) –

opposite to the low-resolution analysis. In addition, all mtDNA types produced lower within-population variance than any Y chromosome data type (Table 5.1c), suggesting that high-resolution mtDNA variation is more geographically localized than Y chromosome STR variation in Africa. The conflicting scenarios may be a function of time, perhaps suggesting that in more ancient times (represented by low-resolution HG data), female migration exceeded male migration in Africa, whilst more recently (represented by high-resolution data), male migration has exceeded female migration. Again, these results emphasize that the use of different data types in such analyses can influence the outcome (Stoneking 1998).

Regression analyses of both mtDNA and Y chromosome data were also used to assess male and female migration rates in African populations. Both types of data showed positive relationships between genetic and geographic distances among African populations (fig. 5.1). Only the slope of the regression line drawn from Y chromosome HG data was significantly different to the slopes of lines drawn from Y chromosome 5-locus or 8-locus STR data (t-test, $P<0.05$); all other pairs of lines were not significantly different to each other. These results suggest that there has not been a significant difference in the rates of female vs. male migration in populations from across the African continent.

Overall, the comparison of mtDNA and Y chromosome data from African populations suggested that results are highly dependant on the type and resolution of data used and that it is unclear whether or not female and male migration rates in Africa differ; it is possible that they have differed over time.

## 6. SUMMARY AND CONCLUSIONS

In this study, Y chromosome and mtDNA variation were examined in four central African populations, including Ugandans, Zambians, and Biaka Pygmies and non-Pygmies from the C.A.R. Y chromosome variation was also examined in populations from the D.R.C. These data have contributed to overcoming the paucity of haploid data from central Africa. The data generated in this study were analysed in conjunction with published data in order to achieve the objectives of the study outlined in Section 1.4 (pg. 24).

### 6.1 *Haploid genetic variation in Africa*

Patterns of mtDNA and Y chromosome variation were explored in each central African population sampled, in order to contribute to the understanding of the overall patterns of haploid genetic variation in Africa.

### 6.1.1 *Y chromosome variation in Africa*

In Chapter 3, Y chromosome variation defined by 16 biallelic markers and 8 STR markers was examined in five central African populations (N=369). Eleven binary lineages (HGs A, B*, B-M150, B-M112, B-M211, E-M191, E-M2, E-M35, E-M40, FJ and R) and 174 compound HG-STR hts were identified (fig. 3.1, Table 3.1). The patterns of Y chromosome variation in central African populations were discussed in detail in Section 3.1.1.

As in other African populations (Underhill et al. 2000; Semino et al. 2002; Cruciani et al. 2002), HG E was the most common HG in central African populations, accounting for 87.2% of the total sample (fig. 3.1). However the high frequency of sublineage E-M191 (overall 49%) was specifically characteristic of central African populations (fig. 3.1, 3.9). HGs E-M2 and E-M191 were shown to have experienced founder effects (fig. 3.5), and have undergone large increases in frequency and geographic distribution that have had a massive impact on central and other African populations (figs. 3.9, 3.10). HGs E-M191 and E-M2 were suggested to be potential markers of the Eastern and Western streams of the recent "Bantu Expansion" respectively (fig. 3.9, 3.10), although it was also cautioned that this explanation may be overly simplistic. The term "Bantu Modal Haplotype" (Thomas et al. 2000; Pereira et al. 2002) was suggested to be inappropriate. HG E-M35 and HG E-M40 were relatively rare in the central African populations examined in this study (fig. 3.1). STR hts from E-M35 in central Africa may have been gained due to shared ancestry with or gene flow from populations from east Africa rather than from north African populations (fig. 3.4).

HG B was significantly more frequent in central African populations than in populations from other geographic regions (fig. 3.9). As in previous studies (Underhill et al. 2000; Cruciani et al. 2002; Semino et al. 2002), HG B-M112 was found mostly in hunter-gatherer populations from southern and central Africa, whilst HG B-M150 was found in other sub-Saharan African populations (fig 3.2). STR hts from these two HGs could be distinguished from each other (fig. 3.2), and dating indicated a more recent expansion of HG B-M150 than of B-M112.

This study showed for the first time that HG R is present in African populations outside of Cameroon and Egypt, but that it is present at low frequencies (fig. 3.1). HG R may have been introduced from Eurasia via north Africa to the present-day region of Cameroon and the C.A.R (Salas et al. 2002), and then perhaps was spread to other parts of sub-Saharan Africa via the western stream of the recent "Bantu Expansion". Recent European contribution of HG R Y chromosomes could not be excluded (fig. 3.7).

### 6.1.2 *mtDNA variation in Africa*

In Chapter 4, mtDNA variation defined by the mitochondrial COII/tRNA$^{Lys}$ intergenic 9-bp deletion, the 3592 *Hpa*I and 10397 *Alu*I RFLPs, and the two hypervariable regions of the mtDNA control region was examined in four central African populations (N=397). A total of 246 mtDNA types defined by HVRI and HVRII variation were identified and were classified into 18 mtDNA subHGs, including L1a, L1b, L1c, L1d, L1e, L1f, L1k, L2a, L2b, L2c, L2d, L3b, L3d, L3e, L3f, L3g, the newly described L3h, and the Eurasian subHG X2c (figs. 4.2-4.7, Tables 4.3-4.5). Ten mtDNA types were categorised as belonging to the as yet further undefined African portion of HG L3, L3A*. The patterns of mtDNA variation in central African populations were discussed in detail in Section 4.1.1.

HVRII sequence data from this study helped to extend the definitions of mtDNA subHGs found in African populations (Tables 4.3-4.5, fig. 4.8). As well as subHG L3h, several new clades within subHGs were newly described, including L1f1, L1c4, L1c5 and L2a2 (Tables 4.3, 4.4). It was shown that subHGs L1a and L1f could be paired on the basis of shared mutations at sites 73, 189, 195 and 16172, and that subHGs L1d and L1k could be paired on the basis of shared mutations at positions 146 and 263 (fig. 4.8). The mtDNA 9-bp deletion was confirmed to be consistently associated only with subHG L1a2; other independent origins of the deletion were observed in subHGs L1c, L1d, L1f and L3A* (Tables 4.3, 4.5). A triplication of the 9-bp repeat motif was observed to be [inconsistently] associated with subHG L3e1 (Table 4.5).

Analyses of the geographic distributions of subHGs in Africa showed that, with the exception of the widespread subHG L1a, subHGs within HG L1 were extremely geographically localised (fig. 4.10). SubHG L1b was found mostly in west Africa, L1c in central Africa, L1d and L1k in Khoisan populations form southern Africa, and L1e and L1f in east Africa (fig. 4.10). Within HG L2, subHGs L2b, L2c and L2d were mostly found in west Africa, whilst L2a was widespread (fig. 4.11). Within HG L3, L3b and L3d were common in west Africa, L3g and L3h were common in east Africa, and L3e and L3f were widely distributed (fig. 4.12). Neither the east African lineage M1 nor the Asian HG M was found in central African populations in this study. The distributions of subHGs L3b, L3d, L2b and L1a may have been influenced by the recent "Bantu Expansion".

### 6.2 *Genetic affinities of African populations*

The genetic affinities of central African (Pygmy and non-Pygmy) populations with each other (Sections 3.1.2 and 4.1.2) and with other African populations (Sections 3.1.4 and 4.1.3) were examined.

### 6.2.1 *Genetic affinities of central African populations*

When high-resolution haploid genetic data (Y chromosome HG-STR hts, or mtDNA HVRI and II sequence data) were used, each central African population examined was significantly different to each other population examined, with the exception of Zambia vs. D.R.C. populations using Y chromosome data (Tables 3.2, 4.7). These results emphasised the high genetic diversity of African populations (Tishkoff and Williams 2002). When lower-resolution mtDNA subHG frequency data were used, the populations remained significantly different to each other (Table 4.7), but when lower-resolution Y chromosome HG data were used, they did not (Table 3.2). Therefore some of populations of central Africa examined in this study appear to be more closely related to each other through their paternal lineages than through their maternal lineages. This finding may also be influenced by Y SNP ascertainment bias. The genetic affinities of each population examined are discussed below.

#### 6.2.1.1 *Non-Pygmy populations from the C.A.R.*

The frequencies and distributions of certain mtDNA lineages (L1b, L2c, L2d, L3h) indicated possible stronger maternal genetic links between Ubangian-speakers from the C.A.R. and populations from west Africa, than between other central African populations from this study and west African populations (figs. 4.10-4.12). This relationship was also detectable in a PC plot drawn from mtDNA HVRI sequence data (fig. 4.13), and was similar to the observation made by Salas et al. (2002) that mtDNA types from Sao Tome and Bioko islanders and from Fang individuals from Equatorial Guinea were closely related to mtDNA types from west African populations. The Y chromosomes of Ubangian-speakers from the C.A.R. were similar to those of other south, central and west African populations (fig. 3.8, 3.11-3.13).

#### 6.2.1.2 *Ugandans*

The frequencies and distributions of certain mtDNA (L1a, L2a2, L3g, L3h) and Y chromosome lineages (HG E-M35) indicated a closer genetic link between Ugandans and populations from east Africa, than between other central and east African populations (figs.

4.10-4.12, 3.9) This is consistent with the geographic position of Uganda on the border between central and east Africa. However this relationship was detectable only in a PC plot drawn from HVRI sequence data (not from Y chromosome data) and only included certain east African populations (figs. 3.11-3.13, 4.13-4.14). This suggests (a) that the genetic link between Ugandans and east Africans may only exist with specific east African populations; (b) that mtDNAs and Y chromosomes of Ugandans are still more closely related to those of other central and southern African populations than they are to those of east African populations. Y chromosome data from the populations from Tanzania and Kenya may help to more fully understand the genetic affinities of Ugandan populations.

#### 6.2.1.3 *Zambians*

Y chromosome analyses suggested that, paternally, Zambian populations are quite closely related to populations from west-central Africa, and that they are especially closely related to populations from the D.R.C. (figs. 3.8, 3.11-3.13) However mtDNA analyses suggested that, maternally, Zambian populations are quite closely related to east and south-east African populations such as Ugandans and Mozambiquans (figs. 4.9, 4.13, 4.14). These results are interesting since it is know that the western and eastern streams of the "Bantu Expansion" intermingled in the Zambia region about 2000 years ago (Vansina 1984; Phillipson 1985; Huffman 1989). It is possible the Y chromosomes of populations living in present-day Zambia mostly came from speakers of Western Bantu languages, whilst their mtDNAs predominantly came from speakers of Eastern Bantu languages. Although some archaeological evidence (e.g. Jenkins 1988) suggests that Khoisan-speaking populations once extended as far as Zambia, Y chromosome and mtDNA data in present-day Zambians suggest either (a) that that Khoisan populations were not the original inhabitants of the present-day Zambia region or (b) that indigenous Zambian populations were almost completely maternally and paternally replaced by non-Khoisan populations.

#### 6.2.1.4 *Biaka Pygmies*

MtDNA and Y chromosome variation in Biaka Pygmies differed substantially to mtDNA and Y chromosome variation in other central African populations (figs. 3.8, 4.9, Tables 3.2, 4.7). Very limited maternal gene flow between Biaka and other central (or other) African populations was observed in either direction (figs. 4.1, 4.3, 4.13, 4.14), suggesting almost complete genetic isolation of Biaka females. On the other hand, an extensive amount of male gene flow, mostly from non-Pygmy populations to the Biaka, was inferred (figs. 3.1, 3.11-3.13). A greater male than female rate in Biaka Pygmies was also inferred (Section 5.1.1). Therefore the extensive levels of gene flow inferred from non-Pygmy to Biaka populations using autosomal data (Cavalli-Sforza 1986; Wijsman 1986; Cavalli-Sforza et al. 1994; Coia et al. 2004) may be attributed to male-specific gene flow. The observed patterns of male and female gene flow are consistent with local social constraints which allow marriages only between Pygmy women and non-Pygmy men, but not vice versa (Cavalli-Sforza 1986), but implies that male offspring of such unions (or of extra-marital liaisons) have returned to Biaka society instead of staying with their non-Pygmy father's society (as otherwise expected). As shown previously by Salas et al. (2002), mtDNA from central African Mbuti Pygmies resembled east African populations more than they resembled the Biaka or other central African populations, suggesting that Biaka and Mbuti Pygmy populations have different maternal ancestry. Very limited Y chromosome data from Mbuti Pygmies is available (N=12, HG frequencies only); however based on the high frequency of HG B-M112 in this Mbuti sample, it is still possible that Biaka and Mbuti Pygmies share common ancestry through their paternal heritage. More detailed HG B markers need to be examined to test this hypothesis.

### 6.2.2 *Genetic affinities of pan-African populations*

Using both mtDNA and Y chromosome data from this study and from published studies (Tables 2.2, 2.3), AMOVA analyses showed that African populations were best grouped using geographic rather than linguistic criteria (Tables 3.4, 4.8). In specific cases, such as for Khoisan-speaking populations and for Biaka Pygmies, linguistic / ethnic criteria were also useful for identifying genetically distinct groups of populations. The overall structuring of both Y chromosome and mtDNA genetic variation in African populations by geography was confirmed by the significant positive correlations between geographic and genetic distances among them found in this study (fig. 5.1).

Using PC plots (figs. 3.11-3.13, 4.13-4.14), more groups of populations defined by geographic-ethnic criteria could be identified using mtDNA data (seven groups: north-east, north-west, east, west, south-central, Khoisan, Biaka) than using Y chromosome data (four groups: north, west, south-central, Khoisan). This may suggest that mtDNA variation is more finely structured in Africa than Y chromosome variation. These observations contrast with the expectations that Y chromosomes should show greater amounts of genetic drift than mtDNA due to the smaller effective Y chromosome population size. One explanation is that results have been affected by the numbers and distributions of populations examined in the analyses (e.g. see Cruciani et al. 2002 for a different number of groups identified using Y chromosome data and different African populations). Another explanation could be higher male versus female migration rates in Africa, or more recent male than female migrations (Chapter Five).

### 6.3 *Use of haploid markers with different mutation rates in assessing population affinities*

In this study it was questioned whether haploid markers with different mutation rates were capable of revealing similar population affinities. Three different types of Y chromosome data (HG frequency data, HG-STR haplotypes, and STR haplotypes) were compared. Genetic distances among five central African populations calculated from the different Y chromosome data types were well-correlated with each other (Section 3.1.3), supporting the statement by Bosch et al. (1999) that "variation in STRs is deeply structured by genetic background on the human Y chromosome". The three data types also produced similar genetic affinities among African populations in PC plots (figs. 3.11-3.13) and AMOVA analyses (Table 3.4), despite the fact that different populations with different sample sizes were used in each analysis. These results suggested that datasets in which Y chromosome STRs only were typed may still be useful for resolving (African) population affinities. However it was cautioned that the association between a HG and the STR variation within it may break down when closely related and more detailed SNP lineages are examined; for instance in this study a great amount of STR ht homoplasy was observed among HGs E-M2 and E-M191. The use of different Y chromosome data types produced significantly different regression lines showing the relationship between genetic and geographic distances among African populations (Table 5.1a). Caglia et al. (2003) also recently demonstrated that whilst STR data can be used successfully to reveal interpopulation relationships, results can be highly affected by choice of genetic distance method.

MtDNA types defined by HVR sequences diverge much faster than mtDNA subHGs, which may be defined by coding region SNPs or by the accumulation of several HVR mutations. Although so far subHG frequencies in African populations have all been extrapolated from HVR sequence data, it is feasible that in the future mtDNA subHG frequencies may be estimated only by typing of relevant coding-region SNPs. Therefore it was tested whether different mtDNA data types (HVR sequence data vs. subHG frequencies) were capable of revealing similar population affinities. Genetic distances among four central African populations calculated from mtDNA HVR sequence data were very similar to and significantly correlated with distances predicted from mtDNA subHG frequency data (Table 4.7). The two data types also produced similar genetic affinities among African populations in PC plots (figs. 4.13, 4.14) and AMOVA analyses (Table 4.8), although the differentiation of populations from different geographic regions was less clear when subHG frequency data were used. These results suggested that both data types are useful for resolving (African) population affinities, and may reveal different facets of population relationships.

### 6.4 *Comparison of male and female migration rates in African populations*

Chapter 5 dealt with the comparison of male and female migration rates estimated from Y chromosome and mtDNA data in central African and pan-African populations. The female migration rate was greater than the male migration rate in non-Pygmy central African populations in this study, whilst the opposite was found in Biaka Pygmies (Section 5.1). In pan-African populations (Section 5.2), low-resolution Y chromosome HG frequency and mtDNA subHG frequency data suggested greater female than male migration across Africa, consistent with observations made by Seielstad et al. (1998). However high-resolution Y chromosome STR ht and mtDNA HVRI sequence data suggested greater male than female migration across Africa, consistent with observations made Hammer et al. (2001). Male and female migration rates in Africa may have differed over time, and may not have been significantly different.

### 6.5 *The prehistory of central African populations from a haploid genetic perspective*

The most ancestral African Y chromosome lineage (HG A) and the most ancient African mtDNA lineages (L1a, L1f, L1d and L1k) were generally absent or rare in central African populations in this study (figs. 3.1, 4.1). When such lineages were found in central Africans (e.g. mtDNA subHG L1a), it was shown that they had not evolved in central Africa, and were likely to have been gained by migration or gene flow to central African populations after they had evolved elsewhere (Section 4.2.1). These observations were consistent with archaeological evidence (Cahen 1982; Phillipson 1985) suggesting that most of central African region (perhaps excluding the Great Lakes area) was populated at a later date (Middle Stone Age, ~80 kya) than eastern or southern Africa (populated since the Early Stone Age).

Nevertheless, Y chromosome and mtDNA lineages that are still very deeply placed in phylogenetic trees were identified that seem to have evolved in central African populations, showing that central Africa populations still have very ancient origins. These lineages included Y chromosome HG B, and mtDNA subHGs L1c and L3e (figs. 3.1, 4.1, 4.8). Dating of these lineages showed that at least one of them, mtDNA subHG L1c (TMRCA of 95 139 years), may have evolved in central Africa during the Middle Stone Age (Table 4.8). Y chromosome HG B may also have evolved during the Middle Stone Age: Underhill et al. (2001) proposed that Y chromosome HG B may have been involved in population expansions occurring in Africa~ 70 000 years ago. Estimates of TMRCA of HG B in this study were substantially younger than this date (between 6 395 and 9948 years ago, Section 3.1.1.3), but the discrepancy may be due to the need for more robust methods of dating Y chromosome HGs (see Underhill et al. 2001 for a discussion of why estimates of the age of human Y chromosome variation are less than the estimated archaeological dates of events in human history).

Archaeological evidence shows that during the later stages of the MSA, populations in central Africa began to diversify and subdivide (Lahr and Foley 1998). Such processes were probably exacerbated by the arid environmental conditions associated with the Last Glacial Maximum ~18 kya (Oliver 1999; Vansina 1990), when environmental fragmentation may have caused population isolation. MtDNA data from this study showed that ancestors of Biaka Pygmies are an example of a central African population who may have begun to diverge (at least maternally) from other autochthonous central African populations ~20 kya (Section 4.2.1)

Populations that evolved in central Africa probably expanded geographically during environmental warm conditions, mimicking the expansion of their rainforest habitat. Particularly warm time phases include the Stage 5 interglacial 130-90 000 years ago (Lahr and Foley 1998), and the Holocene Wet Phase ~10 000 years ago (Maley 1995). Some evidence for central African participation in such migrations may be observed in the wide-spread geographic distribution of specific haploid sublineages that are thought to have evolved in central Africa, including Y chromosome HG B-M150 and mtDNA subHGs L1c2, L1c3 and L3e (Sections 3.2.1, 4.2.1).

There is substantial archaeological and linguistic evidence that central Africa has been greatly affected by recent migrations of populations speaking Bantu and Ubangian languages (Saxon 1982; Phillipson 1985; Vansina 1990). This was corroborated by evidence that both the mtDNA and Y chromosome gene pools of central African populations were extensively influenced by gene flow from migrating populations originating in other African geographic regions (Sections 3.2.2, 4.2.2). Y chromosome data especially suggested an overwhelming replacement of previously existing Y chromosome diversity in central African populations within the very recent past (~3 000 years). This replacement is associated with a massive increase in frequency and diversity of Y chromosomes from HGs E-M2 and E-M191 in central African populations (Section 3.2.2.1), and has been suggested to be associated with the recent "Bantu Expansion" (Thomas et al. 2000; Underhill et al. 2001; Pereira et al. 2002). However Y chromosome replacement may also be associated with other population migrations, such as the equally recent and parallel migrations of speakers of Ubangian languages (Section 3.2.2.1).

Both Bantu and Ubangian languages are suggested to have originated in west Africa (e.g. Saxon 1982; Vansina 1990) before being dispersed across central Africa; Bantu languages spread as far as southern Africa. The close genetic affinities of west, central and southern African populations suggested from Y chromosome data in this study confirmed recent shared male ancestry of population from these geographic regions (figs. 3.11-3.13). The high frequency of HG E-M2 in west African populations, coupled with the high frequency of its derivative lineage E-M191 in central African populations (fig. 3.9) is consistent with the evolution of the HG as it was carried from west to central Africa by either Bantu-speaking or Ubangian-speaking (or other) migrating populations. In addition, the observation of a very recent secondary founder effect and expansion in males from the region of Uganda (associated with specific Y chromosome STR haplotypes from HG E-M191, figs. 3.5, 3.10) might be linked with archaeological evidence for a secondary expansion of speakers of eastern Bantu languages from the Lake Victoria region ~2000 years ago (Ehret 1998). Thus there is strong Y chromosome evidence supporting the role of migratory males from west Africa in the process of the recent replacement of indigenous central African Y chromosomes.

MtDNA in central African populations did not show the same degree of reduction in diversity as seen with the Y chromosome (fig. 4.1), and therefore it may be assumed that women did not play as great a role as men in the recent migration of Bantu-or Ubangian-speaking populations. However this study supported a higher female than male migration rate in three non-Pygmy central African populations (Section 5.1). One explanation could be the practice of patrilocality (Stoneking 1998; Hammer et al. 2001; Oota et al. 2001) which is known to occur in many Bantu-speaking populations (Preston-Whyte 1974). Another explanation could be that females migrated more than males in the past (Chapter 5). Certainly a substantial amount of mtDNA gene flow (and assumed migration) into central Africa was detected from both east and west African sources (Section 4.2.2).

Neither mtDNA nor Y chromosome data showed much evidence of gene flow among populations living in north-west or north-east Africa, and populations living in central Africa, which may be attributed to the Sahara Desert acting as a genetic barrier between north and central Africa (Sections 3.2.4, 4.2.4). Despite the large influence of European colonial countries on the politics and economies of central African populations, Eurasians were estimated to have contributed to approximately 3% of the central African Y chromosome gene pool, and less than 1% of the central African mtDNA gene pool (Sections 3.2.2.2, 4.2.2.4).

## 6.6 *Future studies*

In conclusion, the study of haploid genetic variation in populations from central Africa in this study has facilitated a greater understanding of the haploid genetic variation in Africa, and of the genetic affinities of African populations. Haploid genetic data were shown to be remarkably concordant with suggestions about the prehistory of central African populations based on non-genetic studies.

However data from this study also emphasised that there is great genetic diversity in African populations and our understanding of the patterns of this diversity is far from complete. There is still a lack of both Y chromosome and mtDNA data from Nilo-Saharan-speaking African populations, from populations from the west coast of sub-Saharan Africa, from populations living in the Sahel region, and from non-Khoisan southern Africa. Overall many more populations and regions of Africa are represented in mtDNA databases than in Y chromosomes databases, and in particular, Y chromosome data from east Africa are lacking.

Compound HG-STR Y chromosome haplotypes and whole mtDNA genome sequences would be the most informative data types for understanding haploid genetic evolution in Africa. The discovery of more Y chromosome SNPs will also facilitate this process. As the number of populations with available haploid genetic data increases, it should become possible to understand the structuring of male and female genetic variation in smaller geographic regions and in more defined linguistic subgroups. Also, because mtDNA and Y chromosomes each represent single loci in the human genome, haploid data and the conclusions drawn from them should, wherever possible, be compared to data from other independent (nuclear) loci.

# 7. REFERENCES

Adcock GJ, Dennis ES, Easteal S, Huttley GA, Jermiin LS, Peacock WJ, Thorne A (2001) Mitochondrial DNA sequences in ancient Australians: Implications for modern human origins. Proc Natl Acad Sci USA 98:537-542

Alves-Silva J, da Silva Santos M, Guimaraes PEM, Ferreira ACS, Bandelt H-J, Pena SDJ, Ferreira Prado V (2000) The ancestry of Brazilian mtDNA lineages. Am J Hum Genet 67:444–461

Ambrose SH (1982). Chapter 7: Archaeology and linguistic reconstructions of history in east Africa. In: The Archaeological and Linguistic Reconstruction of African History. Eds: C. Ehret and M. Posnansky. University of California press: Los Angeles

Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290 (5806): 457-65

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23:147

Armour JA, Anttinen T, May CA, Vega EE, Sajantila A, Kidd JR, Kidd KK, Bertranpetit J, Paabo S, Jeffreys AJ (1996) Minisatellite diversity supports a recent African origin for modern humans. Nat Genet 13(2): 154-60

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

Bandelt H-J, Forster P (1997) The myth of bumpy hunter-gatherer mismatch distributions. Am J Hum Genet 61: 980-983

Bandelt H-J, Forster P, Sykes B, Richards M (1995) Mitochondrial portraits of human populations using median networks. Genetics 141: 743-753

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

Bandelt H-J, Alves-Silva J, Guimaraes PEM, Santos MS, Brehm A, Pereira L, Coppa A, Larruga JM, Rengo C, Scozzari R, Torroni A, Prata MJ, Amorim A, Prado VF, Pena SDJ

(2001) Phylogeography of the human mitochondrial haplogroup L3e: a snapshot of African prehistory and Atlantic slave trade. Ann Hum Genet 65: 549-563

Batzer MA, Stoneking M, Alegria-Hartman M, Bazan H, Kass DH, Shaikh TH, Novick GE, Ioannou PA, Scheer WD, Herrera RJ, et al. (1994) African origin of human-specific polymorphic *Alu* insertions. Proc Natl Acad Sci U S A 91(25): 12288-92

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

Bianchi NO, Catanesi CI, Bailliet G, Martinez-Marignac VL, Bravi CM, Vidal-Rioja LB, Herrera RJ, Lopez-Camelo JS (1998) Characterisation of ancestral and derived Y-chromosome haplotypes of New World Native populations. Am J Hum Genet 63: 1862-1871

Bleek WHI (1862) A comparative grammar of South African languages. Part I. Phonology. London

Bosch E, Calafell F, Santos FR, Perez-Lezaun A, Comas D, Benchemsi N, Tyler-Smith C, Bertranpetit J (1999) Variation in short tandem repeats is deeply structured by genetic background on the human Y chromosome. Am J Hum Genet 65: 1623 – 1638

Bosch E, Calafell F, Perez-Lezaun A, Clarimon J, Comas D, Mateu E, Martinez-Arias R, Morera B, Brakez Z, Akhayat O, Sefiani A, Harit G, Cambon-Thomsen A, Bertranpetit J (2000) Genetic structure of north-west Africa revealed by STR analysis. Eur J Hum Genet 8: 360-366

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

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368(6470): 455-7

Bradley DG, MacHugh DE, Cunningham P, Loftus RT (1996) Mitochondrial diversity and the origins of African and European cattle. Proc Natl Acad Sci USA 14: 5131-5135

Brehm A, Pereira L, Bandelt H-J, Prata MJ, Amorim A (2002) Mitochondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of the Atlantic slave trade. Ann Hum Genet 66:49–60

Briggs P (1996) Guide to Uganda, 2nd ed. Bradt Publications, UK

Brown WM, Prager EM, Wang A, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. Proc Natl Acad Sci USA 76: 1967-1971

Brown W (1980) Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. Proc NAtl Acad Sci USA 77: 3605-3609

Brown MD, Sun F,Wallace DC (1997) Clustering of Caucasian Leber hereditary optic neuropathy patients containing the 11778 or 14484 mutations on an mtDNA lineage. Am J Hum Genet 60:381–387

Brown MD, Starikovskaya E, Derbeneva O, Hosseini S, Allen JC, Mikhailovskaya IE, Sukernik RI, Wallace DC (2002) The role of mtDNA background in disease expression: a new primary LHON mutation associated with Western Eurasian haplogroup J. Hum Genet 110(2): 130-8

Burckhardt F, von Haesler A, Meyer S (1999) HvrBase: compilation of mtDNA control region sequences from primates. Nucleic Acids Research 27: 138-142

Burgoyne PS (1982) Genetic homology and crossing over in the X and Y chromosomes of mammals. Hum Genet 61(2): 85-90

Caglia A, Tofanelli S, Coia V, Boschi I, Pescarmona M, Spedini G, Pascali V, Paoli G, Destro-Bisol G (2003) A study of Y-chromosome microsatellite variation in sub-Saharan Africa: a comparison between F(ST) and R(ST) genetic distances. Hum Biol 75(3): 313-30

Cahen D (1982) The Stone Age in the south and west. In: The archaeology of central Africa, ed. F. van Noten. Akademische Druck -u. Verlagsanstalt: Austria, pp 41-55

Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. Nature 325: 31-36

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

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

Cavalli-Sforza LL (1986) Chapter 26. African Pygmies: an evaluation of the state of research. In: Cavalli-Sforza LL (ed) African Pygmies. Academic Press Inc, New York, pp 361-426

Cavalli-Sforza LL and Bodmer WF (1971) The Genetics of Human Populations. Freeman, San Francisco

Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, New Jersey

Chen Y-S, Torroni A, Excoffier L, Santachiara-Benerecetti AS, Wallace DC (1995) Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. Am J Hum Genet 57: 133 -149

Chen Y-S, Olckers A, Schurr TG, Kogelnik AM, Huoponen K, Wallace DC (2000) mtDNA variation in the South African Kung and Khwe – and their genetic relationships to other African populations. Am J Hum Genet 66: 1362-1383

Chima SC, Ryschkewitsch CF, Stoner GL (1998) Molecular epidemiology of human polyomavirus JC in the Biaka Pygmies and Bantu of central Africa. Mem Inst Oswaldo Cruz 93: 615-623

Clark JD (1968) The Prehistoric Origins of African Culture. J Af Hist 5(2):161-182

Clark JD (1982) The cultures of the Middle Palaeolithic/Middle Stone Age. In: Cambridge History of Africa, Vol I: From the earliest times to c. 500 BC. Cambridge: Cambridge University Press pp. 248-341

Coia V, Caglia A, Arredi B, Donati F, Santos FR, PAndya A, Tagliolo L, Paoli G, Pascali V, Destro-Bisol G, Tyler-Smith C (2004) Binary and microsatellite polymorphisms of the Y-Chromosome in the Mbenzele Pygmies from the Central African Republic. Am J Hum Biol 16:57-67

Comas D, Francesc Calafell F, Mateu E, Perez-Lezaun A, Bosch E, Martýnez-Arias R, Clarimon J, Facchini F, Fiori G, Luiselli D, Pettener D, Bertranpetit J (1998) Trading genes along the Silk Road: mtDNA Sequences and the origin of central Asian populations. Am J Hum Genet 63:1824–1838

Corte-Real HB, Macaulay VA, Richards MB, Hariti G, Issad MS, Cambon-Thomsen A, Papiha S, Bertranpetit J, Sykes BC (1996) Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. Ann Hum Genet 60 ( Pt 4): 331-50

Corte-Real F, Carvalho M, Andrade L, Anjos MJ, Pestoni C, Lareu MV, Carracedo A, Vieira DN, Vide MC (2000) Chromosome Y STRs analysis and evolutionary aspects for Portuguese spoken countries. In: Progress in Forensic Genetics 8 (eds GF Sensabaugh, P. J. Lincoln & B. Olaisen). Amsterdam: Elsevier Science, pp. 272-274

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

David N (1982) Prehistory and historical linguistics in central Africa: points of contact. In: The Archaeological and Linguistic Reconstruction of African History. Eds: C. Ehret and M. Posnansky. University of California Press: Los Angeles

Denaro M, Blanc H, Johnson MJ, Chen KH, Wilmsen E, Cavalli-Sforza LL, Wallace DC (1981) Ethnic variation in *HpaI* endonuclease cleavage patterns of human mitochondrial DNA. Proc Natl Acad Sci USA 78(9): 5768 - 5772

Diamond J (1997) Guns, germs and steel. Norton, New York

Diamond J and Bellwood P (2003) Farmers and their languages: the first expansion. Science 300: 597-603

Dorit RL, Akashi H, Gilbert W (1995) Absence of polymorphisms at the ZFY locus on the human Y chromosome. Science 268:1183-1185

Ehret C (1982) Linguistic inferences about early Bantu history. In: The Archaeological and Linguistic Reconstruction of African History. Eds: C. Ehret and M. Posnansky. University of California Press: Los Angeles

Ehret C (1998) An African classical age: eastern and southern Africa in world history, 1000 BC to AD 400. The University Press of Virginia, Charlottesville and James Currey, Oxford.

Ehret C, Posnansky M (1982) The archaeological and linguistic reconstruction of African history. University of California Press, California

Finnila S, Lehtonen MS, Majamaa K (2001) Phylogenetic Network for European mtDNA. Am J Hum Genet 68:1475–1484

Felsenstein J (1975) Confidence limits on phylogenies: an approach using the bootstrap. Evol 39: 783-791

Forster P, Harding R, Torroni A, Bandelt H-J (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. Am J Hum Genet 59:935–945

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

Fox CL (1997) mtDNA analysis in ancient Nubians supports the existence of gene flow between sub-Sahara and north Africa in the Nile Valley. Annals Hum Biol 24 (3): 217-227

Giles RE, Blanc H, Cann HM and Wallace DC (1980) Maternal inheritance of human mitochondrial DNA. Proc Natl Acad Sci USA 77: 6715-6719

Goldstein D, Ruiz-Linares A, Cavalli-Sforza LL, Feldman M (1995) An evaluation of genetic distances for use with microsatellite loci. Genetics 139: 463-471

Goldstein DB, Zhivotovsky LA, Nayar K, Ruiz-Linares A, Cavalli-Sforza LL, Feldman M (1996) Statistical properties of variation at linked microsatellite loci: Implications for the history of human Y chromosomes. Mol Biol Evol 13:1213-1218

Graven L, Passarino G, Semino O, Boursot P, Santachiara-Benerecetti S, Langaney A, Excoffier L (1995) Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. Mol Biol Evol 12:334–345

Greenberg JH (1963) The languages of Africa. The Hague: Mouton.

Greenberg JH (1972) Linguistic evidence concerning Bantu origins. J Afr Hist 13:189-216

Grimes BF (ed) (2001) Ethnologue: languages of the world, 13th ed. Summer Institute of Linguistics, Dallas

Gurven M (2000) How can we distinguish between mutational “hot spots” and “old sites” in human mtDNA samples? Human Biology 72: 455-471

Guthrie M (1967-71). Comparative Bantu: an introduction to the comparative linguistics and prehistory of the Bantu languages. (Vol. 1: 1967, Vol. 2: 1971, Vols. 3 & 4: 1970). Farnborough: Gregg International Publishers

Guthrie M (1948) The classification of the Bantu languages. London: IAI/OUP

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

Hammer MF (1995) A recent common ancestry for human Y chromosomes. Nature 378: 376-378

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

Hammer MF, Karafet T, Rasanayagam A, Wood RT, Altheide TK, Jenkins T, Griffiths RC, Templeton AR, Zegura AL (1998) Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. Mol Biol Evol 15: 427-441

Hammer M, Karafet T, Redd A, Jarjanazi H, Santachiara-Benerecetti S, Soodyall H, Zegura S (2001) Hierarchical patterns of global human Y-chromosome diversity. Mol Biol Evol 18: 1189-1203

Hammond-Tooke WD (1974) The Bantu-speaking peoples of southern Africa, 2nd ed. Routledge and Kegan Paul: London

Handt O, Meyer S, von Haeseler A (1998). Compilation of human mtDNA control region sequences. Nucleic Acids Res 26(1):126-9 and http://db.eva.mpg.de/hvrbase/

Helgason A, Sigurdardóttir S, Nicholson J, Sykes B, Hill EW, Bradely DG, Bosnes V, Gulcher JR, Ward R, Stefánsson K (2000) Estimating Scandinavian and Gaelic ancestry in the male settlers of Iceland. Am J Hum Genet 67:697-717

Herrnstadt C, Elson JL, Fahy E, Preston G, Turnbull DM, Anderson C, Ghosh SS, Olefsky JM, Beal MF, Davis RE, Howell N (2002) Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. Am J Hum Genet 70:1152–1171

Hertzberg M, Mickleson KN, Serjeantson SW, Prior JF, Trent RJ (1989) An Asian-specific 9-bp deletion of mitochondrial DNA is frequently found in Polynesians. Am J Hum Genet 44(4): 504-10

Holden CJ (2002) Bantu language trees reflect the spread of farming across sub-Saharan Africa: a maximum parsimony analysis. Proc. R. Soc. Lond. B 269(1493):793-9

Horai S, Hayasaka K, Kondo R, Tsugane K, Takahata N (1995) Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs.

Proc Natl Acad Sci USA 7:532–536

Horai S, Murayama K, Hayasa K, Matsubayashi S, Hattori Y, Fucharoen G, Harihara S, Park KS, Omoto K, Pan I-H (1996) MtDNA polymorphism in east Asian populations, with special reference to the peopling of Japan. Am J Hum Genet 59: 579-590

Horai S, Hayasaka K (1990) Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. Am J Hum Genet 46: 828-842

Howell N, Smejkal CB, Mackey DA, Chinnery PF, Turnbull DM, Herrnstadt C (2003) The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates. Am J Hum Genet 72:659–670

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

Huffman TN (1986) Iron age settlement patterns and the origins of class distinctions in southern Africa. Advances in World Archaeology 5: 291-338

Huffman TN (1989) Iron Age migrations: the ceramic sequence in southern Zambia. Witwatersrand University Press, South Africa

Hurles ME, Veitia R, Arroyo E, Armenteros M, Bertranpetit J, Peréz-Lezaun, Bosch E, Shlumukova M, Cabon-Thomsen A, McElreavey K, López de Munain L, Röhl A, Wilson IJ, Singh L, Pandya A, Santos FR, Tyler-Smith C, Jobling MA (1999) Recent male-mediated gene flow over a linguistic barrier in Iberia, suggested by analysis of a Y chromosomal DNA polymorphism. Am J Hum Genet 65:1437-1448

Ingman M, Kaessmann H, Pääbo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. Nature 408: 708-713

Jakubiczka S, Arneman J, Cooke HJ, Krawczak M, Schmidtke J (1989) A search for restriction fragment length polymorphism on the human Y chromosome. Hum Genet 84: 86-88

Jenkins T (1988) The peoples of southern Africa: studies in diversity and disease. 24th Raymond Dart Lecture, Institute for the Study of Man in Africa. Witwatersrand University Press: Johannesburg.

Jobling MA (2001) Y-chromosomal SNP haplotype diversity in forensic analysis. Forensic Sci Int 118(2-3): 158-62.

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

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

Johnson MJ, Wallace DC, Ferris SD, Rattazzi MC, Cavalli-SForza LL (1983) Radiation of human mitochondrial DNA types analysed by restriction endonuclease cleavage patterns. J Mol Evol 19: 255-271

Johnston HH (1913) A survey of the ethnography of Africa: and the former racial and tribal migrations of that continent. Journal of the Royal Anthropological Institute XLIII: 391-2

Jorde LB, BAmshad MJ, Watkins WS, Zenger R, Fraley AE, Karkowiak PA, Carpenter KD, Soodyall J, Jenkins T, Rogers AR (1995) Origins and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data. Am J Hum Genet 57: 523-538

Jorde LB, Rogers AR, Bamshad M, Watkins WS, Krakowiak P, Sung S, Kere J, Harpending HC (1997) Microsatellite diversity and the demographic history of modern humans. Proc Natl Acad Sci U S A 94(7): 3100-3

Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Seielstad MT, Batzer MA (2000) The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. Am J Hum Genet 66(3): 979-88

Kaladjieva L, Calafell F, Jobling MA, Angelicheva D, de Knijff P, Rosser ZH, Hurles ME, Underhill P, Tournov I, Marushiakova E et al (2001). Patterns of inter- and intra-group genetic diversity in the Vlax Roma as revealed by Y chromosome and mtDNA lineages. Eur J Hum Genet 9: 97-104

Kayser M, Caglia A, Corach D, Fretwell N, Gehrig C, Graziosi G, Heidorn F, et al. (1997) Evaluation of Y-chromosomal STRs: a multicenter study. Int J Legal Med 110: 125-133

Kayser M, Roewer L, Hedman M, Henke L, Henke J, Brauer S, Kruger C, Krawczak M, Nagy M, Dobosz T, Szibor R, de Knijff P, Stoneking M, Sajantila A (2000) Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. Am J Hum Genet 66: 1580-1588

Kayser M, Krawczak M, Excoffier L, Dieltjes P, Corach D, Pascali V, Gehrig C, Bernini LF, Jespersen J, Bakker E, Roewer L, de Knijff P (2001) An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. Am J Hum Genet. 68(4): 990-1018

Knight A, Underhill PA, Mortensen HM, Zhivotovsky LA, Lin AA, Henn BM, Louis D, Ruhlen M, Mountain JL (2003) African Y chromosome and mtDNA divergence provides insight into the history of click languages. Curr Biol. 13(6): 464-73

Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, Paabo S (1997) Neandertal DNA sequences and the origin of modern humans. Cell 11:19–30

Krings M, Salem AE, Brauer K, Geisert H, Malek AK, Chaix L, Simon C, Welsby D. Di Rienzo A, Utermann G, Sajantila A, Pääbo S, Stoneking M (1999) mtDNA analysis of Nile River Valley populations: a genetic corridor or a barrier to migration? Am J Hum Genet 64: 1166-1176

Krings M, Capelli C, Tschentscher F, Geisert H, Meyer S, von Haeseler A, Grosschmidt K, Possnert G, Paunovic M, Paabo S (2000) A view of Neandertal genetic diversity. Nat Genet 26:144–146

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: Molecular Evolutionary Genetics Analysis software, Arizona State University, Tempe, Arizona, USA

Lahr MM, Foley RA (1998) Towards a theory of modern human origins: geography, demography and diversity in recent human evolution. Yb Phys Anthrop 41: 137-176

Lane A, Soodyall H, Arndt S, Ratshikhopha E, Jonker E, Freeman C, Young L, Morar B, Toffie L (2002) Genetic substructure in South African Bantu-speakers: evidence from autosomal DNA and Y chromosome studies. Am J Phys Anthropol 119:175-85

Lafranchi R, Ndanga J, Zana H (1998) New carbon 14C dating of iron metallurgy in the central African dense forest. Yale F&ES Bulletin 102: 41 – 50

Letouzey R (1976) Contribution de la Botanique au Problème d'Une Eventuelle Langue Pygmie. Bibliothéque de la Selaf, Paris

Lucotte G, Ngo NY (1985) p49f, a highly polymorphic probe that detects TaqI RFLPs on the human Y chromosome. Nucleic Acids Res 13:8285

Maca-Meyer N, Gonzalez AM, Larruga JM, Flores C, Cabrera VM (2001) Major genomic mitochondrial lineages delineate early human expansions. BMC Genet 2:13

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

Malaspina P, Persichetti F, Novelletto A, Iodice C, Terrenato L, Wolfe J, Ferraro M, Prantera G (1990) The human Y chromosome shows a low level of DNA polymorphism. Ann Hum Genet 54:297-305

Maley J (1995) Chapter 2: The climatic and vegetational history of the equatorial regions of Africa during the Upper Quaternary. In: The archaeology of Africa: food, metal and towns, eds. T Shaw, P Sinclair, B Andah, A Okpoko. Routledge: London pp 43-52

Malyarchuk, BA, Rogozin IB, Berikov VB, Derenko MV (2002) Analysis of phylogenetically reconstructed mutational spectra in human mitochondrial DNA control region. Hum Genet 111 :46–53

Mateu E, Comas D, Calafell F, Perez-Lezaun A, Abade A, Bertranpetit J (1997) A tale of two islands: population history and mitochondrial DNA sequence variation of Bioko and Sao Tome, Gulf of Guinea. Ann Hum Genet 61:507–518

Maynard Smith J (1990) Models of a dual inheritance system. J Theor Biol 143(1): 41-53

Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favourable gene. Genet Res 23: 23-35

Merriwether DA, Clark AG, Ballinger SW, Schurr TG, Soodyalll H, Jenkins T, Sherry ST, Wallace DC (1991) The structure of human mitochondrial DNA variation. J Mol Evol 33: 543-555

Miller S, Dyk D, Pelesky H (1988) A simple salting-out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16: 1215

Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, Brandon M, Easley K, Chen E, Brown MD, Sukernik RI, Olckers A, Wallace DC (2003) Natural selection shaped regional mtDNA variation in humans. Proc Natl Acad Sci U S A 100(1): 171-6.

Mitchell RJ and Hammmer MF (1996) Human evolution and the Y chromosome. Curr Opin Genet Dev. 6(6): 737-42

Nachman MW (1998) Y chromosome variation of mice and men. Mol Biol Evol. 15(12): 1744-50.

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

Nei M, Jin L (1989) Variances of the average numbers of nucleotide substitutions within and between populations. Mol Biol Evol 6(3): 290-300

Nurse D and Tucker I (2002) A Survey Report for the Bantu Languages. In: SIL International, http://www.sil.org/silesr/2002/016/SILESR2002-016.htm

Oliver R (1999) The African experience: from Olduvai Gorge to the 21st century. Phoenix Press: London

Oliver R, Fagan BM (1975) Africa in the Iron Age, c. 500 BC to AD 1400. London: Cambridge University Press

Oota H, Settheetham-Ishida W, Tiwawech D, Ishida T, Stoneking M (2001) Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. Nat Genet 29(1): 20-1

Ovchinnikov IV, Gotherstrom A, Romanova GP, Kharitonov VM, Liden K, Goodwin W (2000) Molecular analysis of Neanderthal DNA from the northern Caucasus. Nature 30: 490–493

Paabo S (1995) The Y chromosome and the origin of all of us (men). Science 268: 1141-1142

Parsons TJ, Coble MD (2001) Increasing the forensic discrimination of mitochondrial DNA testing through analysis of the entire mitochondrial DNA genome. Croat Med J 42(3): 304-309

Passarino G, Semino O, Quintana-Murci L, Excoffier L, Hammer MF, Sanatachiara-Benerecetti AS (1998) Different genetic components in the Ethiopian population, identified by mtDNA and Y-chromosome polymorphisms. Am J Hum Genet 62: 420-434

Pereira L, Macaulay V, Torroni A, Scozzari R Prata MJ, Amorim A (2001) Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu Expansions and the slave trade. Ann Hum Genet 65: 439-458

Pereira L, Gusmao L, Alves C, Amorim A, Prata MJ (2002) Bantu and European Y-lineages in Sub-Saharan Africa. Ann Hum Genet 66(Pt 5-6): 369-78

Pérez-Lezaun A, Calafell C, Comas D, Mateu E, Bosch E, Martínez-Arias R, Clarimón J, Fiori G, Luiselli D, Facchini F, Pettener D, Bertranpetit J (1999) Sex-specific migration patterns in central Asian populations, revealed by analysis of Y chromosome short tandem repeats and mtDNA. Am J Hum Genet 65:208-219

Phillipson DW (1977) The spread of the Bantu language. Sci Am 236: 106-114

Phillipson DW (1985) African archaeology. Cambridge University press: Cambridge.

Pinto F, Gonzalez AM, Hernandez M, Larruga JM, Cabrera VM (1996) Genetic relationship between the Canary Islanders and their African and Spanish ancestors inferred from mitochondrial DNA sequences. Ann Hum Genet 60 (Pt 4): 321-30

Preston-Whyte E (1974) Chapter 6: Kinship and marriage. In: The Bantu-speaking peoples of southern Africa, 2nd ed. Ed: WD Hammond-Tooke. Routledge and Kegan Paul: London

Pritchard JK, Seielstad MT, Perez-Lezuan A, Feldman MW (1999) Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Mol Biol Evol. 16(12): 1791-8

Quintana-Murci L, Semino O, Bandelt H-J, Passarino G, McElreavey K, Santachiara-Benerecetti AS (1999) Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. Nat Genet 23:437–441

Quintana-Murci L, Krausz C, Zerjal T, Sayar H, Hammer MF, Mehdi SQ, Ayub Q, Qamar R, Mohyuddin A, Radhakrishna U, Jobling MA, Tyler-Smith C, McElreavey K (2001) Y-chromosome lineages trace diffusion of people and language in southwestern Asia. Am J Hum Genet 68: 537- 542

Rando JC, Pinto F, Gonzalez AM, Hernandez M, Larruga JM, Cabrera VM, Bandelt J-H (1998) Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, Near-Eastern and sub-Saharan populations. Ann Hum Genet 62: 531-550

Raymond M, Rousset F (1995) An exact test for population differentiation. Evolution 49:1280-1283

Redd AJ, Takezaki N, Sherry ST, McGarvey ST, Sofro AS, Stoneking M (1995) Evolutionary history of the COII/tRNA$^{Lys}$ intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. Mol Biol Evol 12(4): 604-15

Redd AJ and Stoneking M (1999) Peopling of Sahul: mtDNA variation in Aboriginal Australians and Papua New Guinean populations. Am J Hum Genet 65: 808-828

Reidla M, Kivisild T, Metspalu E, Kaldma K, Tambets K, Tolk H-V, Parik J, Loogvali E-L, Derenko M, Malyarchuk B et al. (2003) Origins and diffusion of mtDNA haplogroup X. Am J Hum Genet 73:1178–1190

Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C et al. (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. Am J Hum Genet 67:1251-1276

Rieder J, Taylor SL, Tobe VO, Nickerson DA (1998) Automating the identification of DNA variations using quality-based fluorescence re-sequencing: analysis of human mitochondrial genome. Nucleic Acids Research 26: 967-973

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

Rohlf FJ (1997) NTSYS-pc: Numerical taxonomy and multivariate analysis system, V1.50. Exeter Publishers, New York, USA

Ruhlen M (1987) A Guide to the World's Languages, volume 1. Stanford University Press, Stanford, California

Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC (2004) Effects of purifying and adaptive selection on regional variation in human mtDNA. Science 303(5655): 223-6

Saitou N, Nei M (1987) The neighbour-joining method: a new method for constructing phylogenetic trees. Mol Biol Evol 4: 406 –425

Salas A, Richards M, De la Fe Tomas, Lareu M, Sobrino B, Sanchez-Diz P, Macaulay V, Carracedo (2002) The making of the African mtDNA landscape. Am J Hum Genet 71: 1082-1111

Santos FR, Bianchi NO, Pena SD (1996) Worldwide distribution of human Y-chromosome haplotypes. Genome Res 6(7): 601-11

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

Saxon DE (1982) Linguistic evidence for the eastward spread of Ubangian peoples. In: The Archaeological and Linguistic Reconstruction of African History. Eds: C. Ehret and M. Posnansky. University of California Press: Los Angeles

Schneider S, Roessli D, Excoffier L (2000) ARLEQUIN v2.000: a software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Geneva

Scozzari R, Cruciani F, Santolamazza P, Malaspina P, Torroni A, Sellito D, Arredi B, Destro-Bisol G, De Stefano G, Rickards O, Martinez-Labarga C, Modiano D, Biondi G, Moral P, Olckers A, Wallace DC, Novoletto A (1999) Combined use of biallelic and microsatellite Y-chromosome polymorphisms to infer affinities among African populations. Am J Hum Genet 65: 829-846 (erratum: 66: 346)

Seielstad MT, Hebert JM, Lin AA, Underhill PA, Ibrahim M, Vollrath D, Cavalli-Sforza LL (1994) Construction of human Y-chromosomal haplotypes using a new polymorphic A to G transition. Hum Mol Genet 3(12): 2159-61

Seielstad MT, Minch E, Cavalli-Sforza LL (1998) Genetic evidence for a higher female migration rate in humans. Nature Genet 20:278-280

Seielstad M, Bekele E, Ibrahim M, Toure A, Traore M (1999) A view of modern human origins from Y chromosome microsatellite variation. Genome Res 9: 558 – 567

Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill, PA (2000) The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: A Y chromosome perspective. Science 290:1155-1159

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

Shaw T, Sinclair P, Andah B, Okpoko A (1995) The archaeology of Africa: food, metals, and towns 1st ed. Routledge: London

Shen P, Wang F, Underhill PA, Franco C, Yang W-H, Roxas A, Sung R, Lin AA, Hyman RW, Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ (2000) Population genetic implications from sequence variation in four Y chromosome genes. Proc Natl Acad Sci USA 97:7374-7359

Sherry S (1991) GELIN: an ASCII-based DNA sequence management programme, V 1.0. Pennsylvania State University, USA

Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S et al. (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423(6942): 825-37

Simoni L, Calafell F, Pettener D, Bertranpetit J, Barbujani G (2000) Geographic patterns of mtDNA diversity in Europe. Am J Hum Genet 66: 262-278

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457-462

Soodyall H, Jenkins T (1992) Mitochondrial DNA polymorphisms in Khoisan populations from Southern Africa. Ann Hum Genet 56:315–324

Soodyall H, Vigilant L, Hill AV, Stoneking M, Jenkins T (1996) mtDNA control-region sequence variation suggests multiple independent origins of an "Asian-specific" 9-bp deletion in sub-Saharan Africans. Am J Hum Genet 58: 595-608

Soper R (1971) A general review of the early Iron Age in the southern half of Africa. Azania 6: 5-37

Spurdle AB, Jenkins T (1992) The Y chromosome as a tool for studying human evolution. Curr Opin Genet Dev 2(3): 487-91

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

Stoneking M, Soodyall H (1996) Human evolution and the mitochondrial genome. Curr Opin Genet Dev 6:731-736

Stoneking M, Fontius JJ, Clifford SL, Soodyall H, Arcot SS, Saha N, Jenkins T, Tahir MA, Deininger PL, Batzer MA (1997) *Alu* insertion polymorphisms and human evolution: evidence for a larger population size in Africa. Genome Res 7(11): 1061-71

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

Stoneking M (2000) Hypervariable sites in the mtDNA control region are mutational hotspots. Am J Hum Genet 67(4): 1029-32

Su B, Xiao J, Underhill P, Deka R, Zhang W, Akey J, Huang W, Shen D, Lu D, Luo J, Chu J, Tan J, Shen P, Davis R, Cavalli-Sforza L, Chakraborty R, Xiong M, Du R, Oefner P, Chen Z, Jin L (1999) Y-Chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age. Am J Hum Genet 65(6): 1718-24

Sudoyo H, Suryadi H, Lertrit P, Pramoonjago P, Lyrawati D, Marzuki S (2002) Asian-specific mtDNA backgrounds associated with the primary G11778A mutation of Leber's hereditary optic neuropathy. J Hum Genet 47(11): 594-604

Sykes B, Leiboff A, Low-Beer J, Tetzner S, Richards M (1995) The origins of the Polynesians: an interpretation from mitochondrial lineage analysis. Am J Hum Genet 57: 1463-1475

Thomas MG, Skorecki K, Ben-Ami H, Parfitt T, Bradman N, Goldstein DB (1998) Origins of Old Testament priests. Nature 394:138-140

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

Thomas MG, Parfitt T, Weiss DA, Skorecki K, Wilson JF, le Roux M, Bradman N, Goldstein DB (2000) Y chromosomes travelling south: the Cohen Modal Haplotype and the origins of the Lemba-the "Black Jews of Southern Africa". Am J Hum Genet 66:674-86

Thomas MG, Weale ME, Jones AL, Richards M, Smith A, Redhead N, Torroni A, Scozzari R, Gratrix F, Tarekegn A, Wilson JF, Capelli C, Bradman N, Goldstein DB (2002) Founding mothers of Jewish communities: geographically separated Jewish groups were independently founded by very few female ancestors. Am J Hum Genet 70(6): 1411-20

Thomson R, Pritchard JK, Shen P, Oefner PJ, Feldman MW (2000) Recent common ancestry of human Y chromosomes: Evidence from DNA sequence data. Proc Natl Acad Sci USA 97:7360-7365

Tishkoff SA, Pakstis AJ, Stoneking M, Kidd JR, Destro-Bisol G, Sanjantila A, Lu RB, Deinard AS, Sirugo G, Jenkins T, Kidd KK, Clark AG (2000) Short tandem-repeat polymorphism/ *Alu* haplotype variation at the PLAT locus: implications for modern human origins. Am J Hum Genet 67(4): 901-25

Tishkoff SA, Williams SM (2002) Genetic analysis of African populations: human evolution and complex disease. Nat Rev Gen 3: 611-620

Tofanelli S, Boschi I, Bertoneri S, Coia V, Taglioli L, Franceschi MG, Destro-Bisol G, Pascali V, Paoli G (2003) Variation at 16 STR loci in Rwandans (Hutu) and implications on profile frequency estimation in Bantu-speakers. Int J Legal Med 117 : 121–126

Torroni A, Schurr TG, CabellMF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM, Wallace DC (1993) Asian Affinities and continental radiation of the four founding Native American mtDNAs. Am J Hum Genet 53:563–590

Torroni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, Obinu D, Savontaus M-L,Wallace DC (1996) Classification of European mtDNAs from an analysis of three European populations. Genetics 144:1835–1850

Torroni A, Rengo C, Guida V, Cruciani F, Sellitto D, Coppa A, Luna Calderon F, Simmionati B, Valle G, Richards M, Macaulay V, Scozzari R (2001) Do the four clades of the mtDNA haplogroup L2 evolve at different rates? Am J Hum Genet 69: 1348-56

Trovoada MJ, Alves C, Gusmao L, Abade A, Amorim A, Prata MJ (2001) Evidence for population sub-structuring in Sao Tome e Principe as inferred from Y-chromosome STR analysis. Ann Hum Genet 65:271-83

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

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

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

Van Noten F (1982) The archaeology of central Africa. Akademische Druck –u. Verlagsanstalt: Austria.

Vansina J (1984) Western Bantu Expansion. Journal of African History 25: 129-145

Vansina JC (1990) Paths in the rainforest. Towards a history of political tradition in equatorial Africa. Currey: London

Vigilant L, Pennington R, Harpending H, Kocher TD, Wilson AC (1989). Mitochondrial DNA sequences in single hairs from a southern African population. Proc Natl Acad Sci USA 86: 9350-4

Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. Science 253: 1503-1507

Vogel JO (1994) Eastern and south-central African Iron Age. In: Vogel JO (ed) Encyclopedia of precolonial Africa. Alta-Mira Press, Walnut Creek, pp 439–444

Wallace DC (1995) 1994 William Allan Award Address: Mitochondrial DNA variation in human evolution, degenerative disease and ageing. Am J Hum Genet 57: 210-223

Watson E, Brauer K, Aman R, Weiss G, von Haeseler A, Paabo S (1996) mtDNA sequence diversity in Africa. Am J Hum Genet 59: 437-444

Watson E, Forster P, Richards M, Bandelt H-J (1997) Mitochondrial footprints of human expansions in Africa. Am J Hum Genet 61: 691-704

Weale ME, Yepiskoposyan L, Jager RF, Hovhannisyan N, Khudoyan A, Burbage-Hall O, Bradman N, Thomas MG (2001) Armenian Y chromosome haplotypes reveal strong regional structure within a single ethno-national group. Hum Genet 109(6): 659-74

Whitfield LS, Sulston JE, Goodfellow PN (1995) Sequence variation of the human Y chromosome. Nature 378(6555): 379-80

Wijsman EM (1986) Chapter 25. Estimation of genetic admixture in Pygmies. In: Cavalli-Sforza LL (ed) African Pygmies. Academic Press Inc, New York, pp 349 -356

Wrischnik LA, Higuchi RG, Stoneking M, Erlich HA, ARnheim N, Wilson AC (1987) Length mutations in human mitochondrial DNA: direct sequencing of enzymatically amplified DNA. Nucleic Acids Research 15: 529-542

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

# 8. APPENDICES

## 8.1 RECIPES AND SOLUTIONS

### 10% APS

Dissolve 1 g of APS in 10 ml of $dH_2O$ and store at 4°C for a maximum of 2 weeks.

### Bromophenol blue Ficoll dye

Dissolve 50g sucrose, 1.86g EDTA, 0.1g bromophenol blue and 10g Ficoll in ~50ml $dH_2O$. Adjust volume to 100 ml with $dH_2O$, stir overnight , pH to 8.0 and filter through Whatmann filter paper. Store at room temperature.

### 10 mg/ml BSA

Dissolve 1 g of BSA in 10 ml of $dH_2O$. Aliquot into 1 ml amounts and store at –20° C.

### Dextran/formamide dye

Dissolve 20 mg Dextran (Sigma) in 1 ml formamide.

### 2.5mM dNTPs

Use 100mM premade stocks (GibcoBRL) of dATP, dGTP, dCTP and dTTP. Add 10ul of each stock dNTP to 360µl sterile d $dH_2O$ to make 400µl of 2.5.mM dNTPs.

### 0.5 M EDTA

Mix 186.1 g of EDTA with 800 ml of $dH_2O$. While stirring adjust the pH to 8.0 with a 10M NaOH solution. When EDTA is completely dissolved, adjust volume to 1 litre.

### 10mg/ml Ethidium bromide (EtBr)

Add 1 g of ethidium bromide to 100 ml of $dH_2O$. Stir for several hours until completely dissolved and store wrapped in aluminium foil at 4°C.

**80% isopropanol**

Add 80 ml of 100% isopropanol to 20 ml of $dH_2O$ and mix thoroughly (makes 100 ml).

**2% Metaphor Gel**

Add 2g of Metaphor gel powder to 100 ml of 1X TBE buffer and stir over heat to dissolve.

**2% NuSieve Gel**

Add 2g of Nusieve to 100 ml of 1X TAE buffer and stir to dissolve.

**3.4% polyacrylamide gel**

Dissolve 36g urea in 10.6 ml of deionized 40% bis-acrylamide, 10 ml of 10XTBE and ~50ml of $ddH_2O$. Stir over heat; when translucent bring volume to 100ml. Filter through Nalgene filtration system under vacuum pressure. Store covered in foil at -4°C.

Sequencing gel preparation in 36cm X 25cm plates: mix 50ml of 4.3% polyacrylamide gel with 30µl TEMED (as a catalyst) and 250µl 10% APS (to facilitate cross-binding) immediately before pouring.

Y-STR gel preparation in 24cm X 25cm plates: mix 20ml of 4.3% polyacrylamide gel with 12µl TEMED and 100µl 10% APS immediately before pouring.

**100 uM primer stocks**

Determine spectrophotometrically the concentration of the stock primers from an appropriate dilution and calculate its concentration in µm. Dilute this with $ddH_2O$ to 100µm and store at –20°C.

**1kb size standard**

Add 285µl 1kb ladder (GibcoBRL) and 143ul Ficoll dye to 2 400µl 1XTE.

**2.5Mm Spermidine (Sigma)**

Add 6 887ml $ddH_2O$ to 1g of Spermidine to make a 1M stock. Dilute 1 in 400 to 2.5mM for use.

**50X TAE**

Dissolve 121g Tris in 28.55 ml of glacial acetic acid and 50 ml of 0.5 M EDTA pH 8.0. Make up to 500ml volume with $dH_2O$ .

**1X TAE**

Mix 20 ml of 50XTAE stock with 980 ml of $dH_2O$ to make 1 litre of 1X TAE.

**10XTBE**

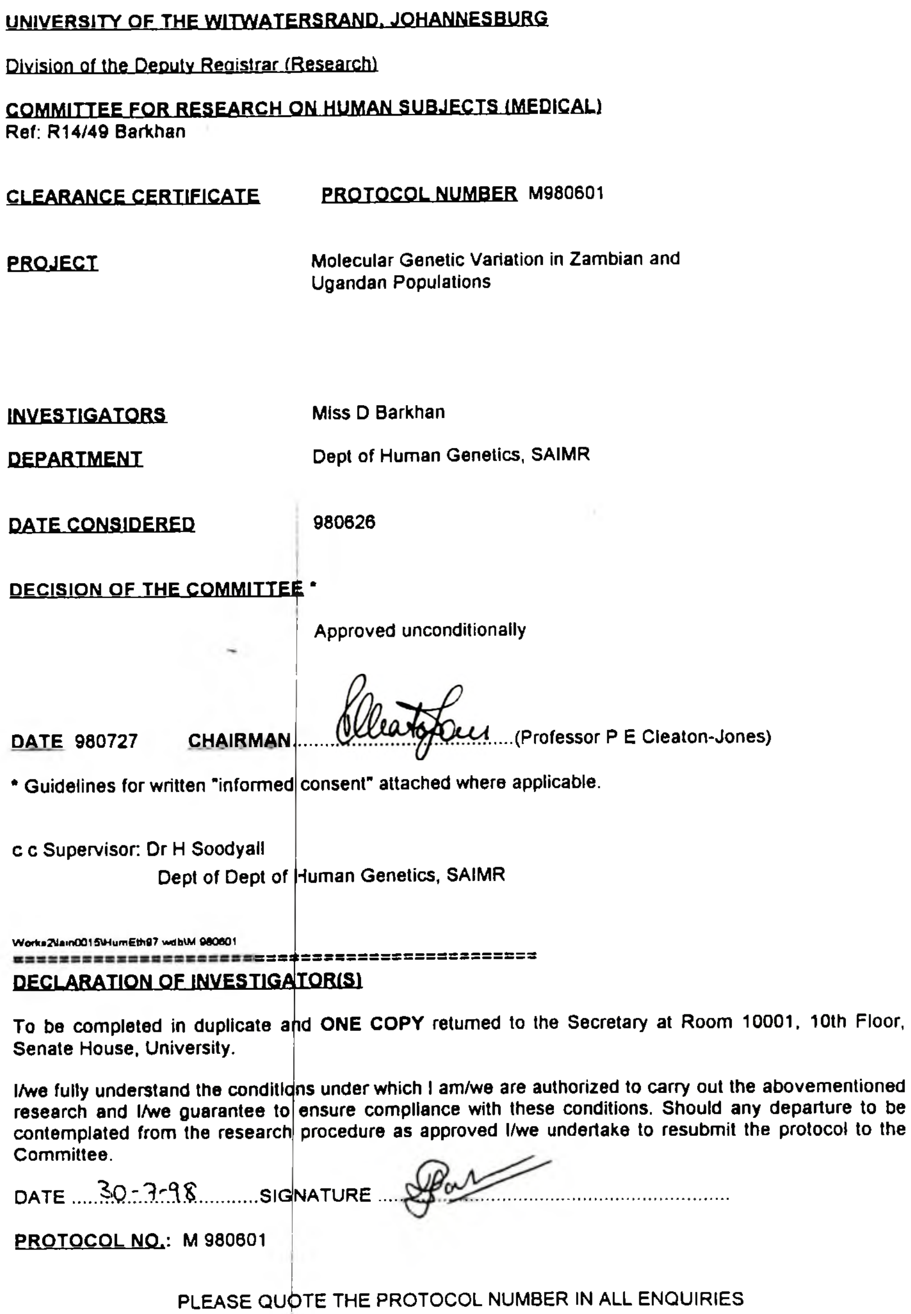
Dissolve 108 g of Tris base and 55 g boric acid in 800 ml of $dH_2O$. Add 40 ml of 0.5 M EDTA (pH 8.0) and make up to 1 litre. The pH of this stock should be approximately 8.3.

**1X TBE**

Mix 100 ml of 10X TBE stock and 900 ml of $dH_2O$ to make 1 litre of 1X TBE solution with final concentrations of 89mM Tris, 89mM boric acid and 2mM EDTA.

**1 M TRIS (pH 8.)**

Add 121.1.g of Tris base to 800ml $dH_2O$. While stirring adjust pH with 1 M HCl. Adjust final volume to 1 litre.



UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

Division of the Deputy Registrar (Research)

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

CLEARANCE CERTIFICATE PROTOCOL NUMBER M980601

PROJECT Molecular Genetic Variation in Zambian and Ugandan Populations

INVESTIGATORS Miss D Barkhan

DEPARTMENT Dept of Human Genetics, SAIMR

DATE CONSIDERED 980626

DECISION OF THE COMMITTEE *

Approved unconditionally

DATE 980727 CHAIRMAN ..............................(Professor P E Cleaton-Jones)

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

c c Supervisor: Dr H Soodyall
Dept of Dept of Human Genetics, SAIMR

Works2\Jain0015\HumEth97 wdb\M 980601

==============================================

DECLARATION OF INVESTIGATOR(S)

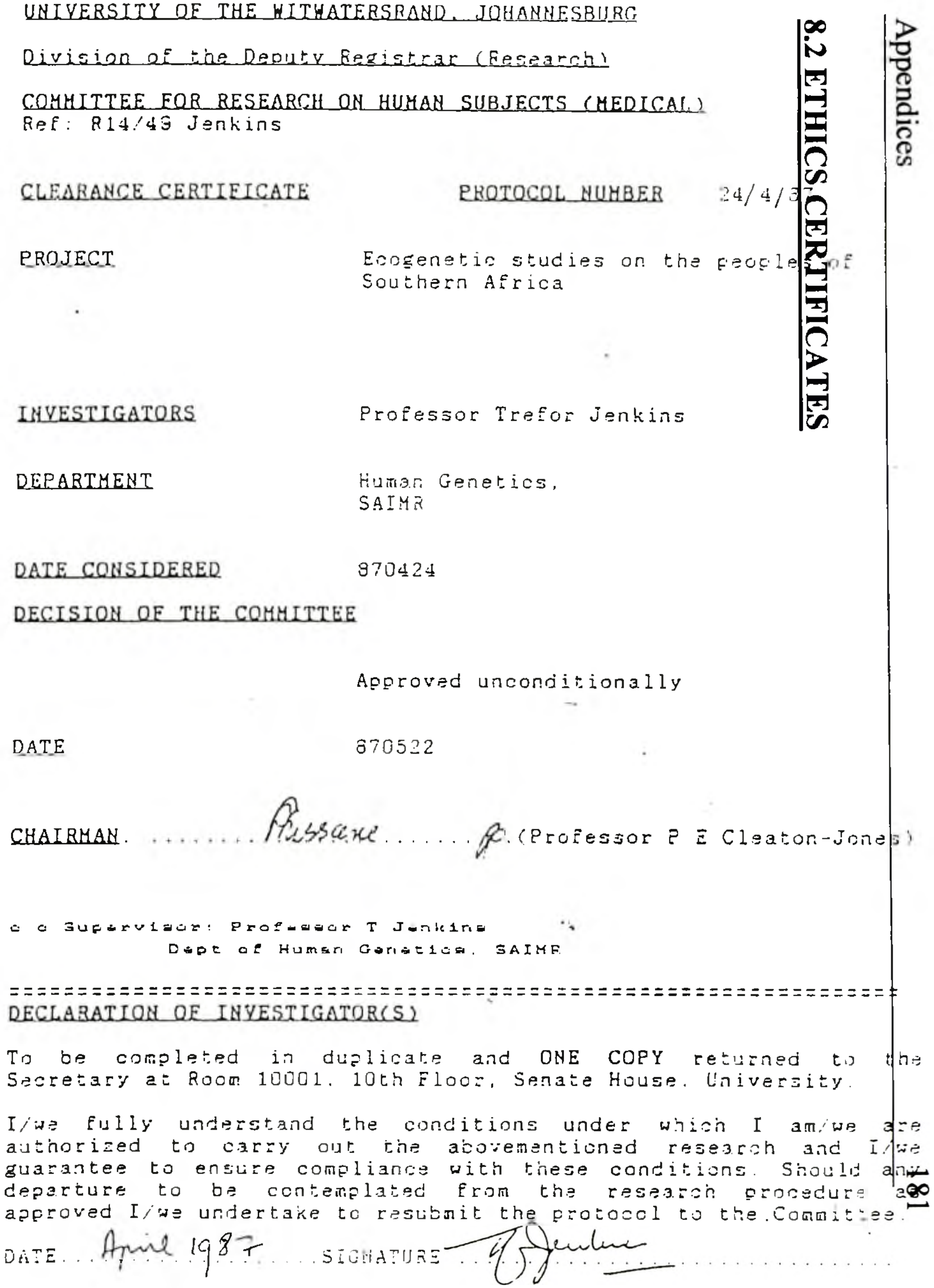
To be completed in duplicate and **ONE COPY** returned to the Secretary at Room 10001, 10th Floor, Senate House, University.

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

DATE ....30-7-98..........SIGNATURE ...............................................

PROTOCOL NO.: M 980601

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES



UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

Division of the Deputy Registrar (Research)

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

CLEARANCE CERTIFICATE PROTOCOL NUMBER 24/4/3

PROJECT Ecogenetic studies on the peoples of Southern Africa

INVESTIGATORS Professor Trefor Jenkins

DEPARTMENT Human Genetics, SAIMR

DATE CONSIDERED 870424

DECISION OF THE COMMITTEE

Approved unconditionally

DATE 870522

CHAIRMAN. .................. .(Professor P E Cleaton-Jones)

c c Supervisor: Professor T Jenkins
Dept of Human Genetics, SAIMR

=================================================================

DECLARATION OF INVESTIGATOR(S)

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

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

DATE...April 1987....SIGNATURE.......................................

**Appendix 8.3** Y chromosome STR allele sizes and number of repeats

| locus | size (bp) | size (repeats) |
|---|---|---|
| DYS19 (394) | 180 | 11 |
| | 184 | 12 |
| | 188 | 13 |
| | 192 | 14 |
| | 196 | 15 |
| | 200 | 16 |
| | 204 | 17 |
| | 208 | 18 |
| DYS388 | 120 | 9 |
| | 123 | 10 |
| | 126 | 11 |
| | 129 | 12 |
| | 132 | 13 |
| | 135 | 14 |
| | 138 | 15 |
| | 141 | 16 |
| | 144 | 17 |
| DYS389I | 242 | 10 |
| | 246 | 11 |
| | 250 | 12 |
| | 254 | 13 |
| | 258 | 14 |
| | 262 | 15 |
| | 265 | 16 |
| DYS389II | 353 | 25 |
| | 357 | 26 |
| | 361 | 27 |
| | 365 | 28 |
| | 369 | 29 |
| | 373 | 30 |
| | 377 | 31 |
| | 381 | 32 |
| | 385 | 33 |
| | 389 | 34 |
| DYS390 | 189 | 17 |
| | 193 | 18 |
| | 197 | 19 |
| | 201 | 20 |
| | 205 | 21 |
| | 209 | 22 |
| | 213 | 23 |
| | 217 | 24 |
| | 221 | 25 |
| DYS391 | 277 | 8 |
| | 281 | 9 |
| | 285 | 10 |
| | 289 | 11 |
| | 293 | 12 |
| DYS392 | 242 | 8 |
| | 245 | 9 |
| | 248 | 10 |
| | 251 | 11 |
| | 254 | 12 |
| | 257 | 13 |
| | 260 | 14 |
| | 263 | 15 |
| DYS393 | 124 | 13 |
| | 128 | 14 |
| | 132 | 15 |
| | 136 | 16 |

**Appendix 8.4a.** List of 20 populations used for Y HG analysis, showing geographic sampling points.

| | population | country | point on map used for distance estimates (reason) |
|---|---|---|---|
| 1 | Amhara | Ethiopia | Addis Ababa (capital) |
| 2 | Arabs | Morocco | Rabat (capital) |
| 3 | Bamileke | south Cameroon | Yaounde (centre of south Cameroon) |
| 4 | Berbers | Morocco | Rabat (capital) |
| 5 | Biaka | CAR | Bambio (sample collection point) |
| 6 | CAR Pygmies | CAR | Bambio (sample collection point) |
| 7 | CAR | CAR | Dula (half-way between two sample collection points) |
| 8 | DRC | DRC | Luozi (sample collection point) |
| 9 | Ewondo | south Cameroon | Yaounde (centre of south Cameroon) |
| 10 | Fali | north Cameroon | Ngandoure (centre of north Cameroon) |
| 11 | Fulbe | Burkina Faso and Cameroon | Kagara in Nigeria (halfway between capitals of the two countries) |
| 12 | Khwe | Caprivi Strip | Kongola (estimated sample collection point) |
| 13 | Kung | Caprivi Strip | Kongola (estimated sample collection point) |
| 14 | Mbuti | DRC | Nia-Nia (estimated sample collection point) |
| 15 | Mossi | Burkina Faso | Ougadougou (capital) |
| 16 | Oromo | Ethiopia | Addis Ababa (capital) |
| 17 | Rimaibe | Burkina Faso | Ougadougou (capital) |
| 18 | Senegal | Senegal | Dakar (capital) |
| 19 | Uganda | Uganda | Kabale (sample collection point) |
| 20 | Zambia | Zambia | Lusaka (capital) |

**Appendix 8.4b.** List of 13 populations used for Y 8-STR analysis, showing geographic sampling points.

| | population | country | point on map used for distance estimates (reason) |
|---|---|---|---|
| 1 | Arabs | Morocco | Rabat (capital) |
| 2 | Berbers | Morocco | Rabat (capital) |
| 3 | Biaka | CAR | Bambio (sample collection point) |
| 4 | CAR Pygmies | CAR | Bambio (sample collection point) |
| 5 | CAR | CAR | Dula (half-way between two sample collection points) |
| 6 | DRC | DRC | Luozi (sample collection point) |
| 7 | Ethiopia | Ethiopia | Addis Ababa (capital) |
| 8 | Mali | Mali | Bamako (capital) |
| 9 | Saharawis | Western Sahara | Villa Cisneros (capital) |
| 10 | San | Caprivi Strip | Kongola (estimated sample collection point) |
| 11 | South Africa | South Africa | Johannesburg (probable sample collection point) |
| 12 | Uganda | Uganda | Kabale (sample collection point) |
| 13 | Zambia | Zambia | Lusaka (capital) |

**Appendix 8.4c.** List of 17 populations used for Y 5-STR analysis, showing geographic sampling points.

| | population | country | point on map used for distance estimates (reason) |
|---|---|---|---|
| 1 | Angola | Angola | Luanda (capital) |
| 2 | Arabs | Morocco | Rabat (capital) |
| 3 | Berbers | Morocco | Rabat (capital) |
| 4 | Biaka | CAR | Bambio (sample collection point) |
| 5 | CAR | CAR | Dula (half-way between two sample collection points) |
| 6 | Cape Verde | Cape Verde | Praia (capital) |
| 7 | DRC | DRC | Luozi (sample collection point) |
| 8 | Ethiopia | Ethiopia | Addis Ababa (capital) |
| 9 | Guinea Bissau | Guinea Bissau | Bissan (capital) |
| 10 | Mali | Mali | Bamako (capital) |
| 11 | Mozambique | Mozambique | Maputo (capital) |
| 12 | Saharawis | Western Sahara | Villa Cisneros (capital) |
| 13 | San | Caprivi Strip | Kongola (estimated sample collection point) |
| 14 | Sao Tome | Sao Tome | Sao Tome (capital) |
| 15 | South Africa | South Africa | Johannesburg (probable sample collection point) |
| 16 | Uganda | Uganda | Kabale (sample collection point) |
| 17 | Zambia | Zambia | Lusaka (capital) |

**Appendix 8.5a**. Matrix of geographic distances among 20 populations used for Y chromosome HG analysis
Distance is given in cm, scale of map was 1cm: 125km

| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **population** | Amhara | Arabs | Bamileke | Berbers | Biaka | CAP | CAR | DRC | Ewondo | Fali | Fulbe | Khwe | Kung | Mbuti | Mossi | Oromo | Rimaibe | Senegal | Uganda | Zambia |
| 1 | Amhara | 0.0 | | | | | | | | | | | | | | | | | | | |
| 2 | Arabs | 43.2 | 0.0 | | | | | | | | | | | | | | | | | | |
| 3 | Bamileke | 24.5 | 30.0 | 0.0 | | | | | | | | | | | | | | | | | |
| 4 | Berbers | 43.2 | 0.0 | 30.0 | 0.0 | | | | | | | | | | | | | | | | |
| 5 | Biaka | 19.8 | 32.3 | 4.8 | 32.3 | 0.0 | | | | | | | | | | | | | | | |
| 6 | CAP | 19.8 | 32.3 | 4.8 | 32.3 | 0.0 | 0.0 | | | | | | | | | | | | | | |
| 7 | CAR | 16.6 | 33.5 | 7.9 | 33.5 | 3.0 | 3.0 | 0.0 | | | | | | | | | | | | | |
| 8 | DRC | 25.0 | 38.0 | 8.0 | 38.0 | 8.0 | 8.0 | 10.0 | 0.0 | | | | | | | | | | | | |
| 9 | Ewondo | 24.5 | 30.0 | 0.0 | 30.0 | 4.8 | 4.8 | 7.9 | 8.0 | 0.0 | | | | | | | | | | | |
| 10 | Fali | 23.3 | 28.2 | 3.5 | 28.2 | 4.5 | 4.5 | 6.6 | 10.8 | 3.5 | 0.0 | | | | | | | | | | |
| 11 | Fulbe | 28.8 | 23.1 | 7.2 | 23.1 | 11.1 | 11.1 | 13.5 | 15.0 | 7.2 | 7.0 | 0.0 | | | | | | | | | |
| 12 | Khwe | 27.3 | 51.8 | 21.7 | 51.8 | 19.7 | 19.7 | 19.8 | 14.0 | 21.7 | 23.7 | 28.8 | 0.0 | | | | | | | | |
| 13 | Kung | 27.3 | 51.8 | 21.7 | 51.8 | 19.7 | 19.7 | 19.8 | 14.0 | 21.7 | 23.7 | 28.8 | 0.0 | 0.0 | | | | | | | |
| 14 | Mbuti | 11.7 | 40.0 | 14.5 | 40.0 | 9.7 | 9.7 | 7.0 | 13.4 | 14.5 | 13.7 | 20.6 | 17.9 | 17.9 | 0.0 | | | | | | |
| 15 | Mossi | 35.5 | 19.3 | 13.7 | 19.3 | 18.0 | 18.0 | 20.2 | 20.5 | 13.7 | 13.8 | 7.0 | 34.6 | 34.6 | 27.6 | 0.0 | | | | | |
| 16 | Oromo | 0.0 | 43.2 | 24.5 | 43.2 | 19.8 | 19.8 | 16.6 | 25.0 | 24.5 | 23.3 | 28.8 | 27.3 | 27.3 | 11.7 | 35.5 | 0.0 | | | | |
| 17 | Rimaibe | 35.5 | 19.3 | 13.7 | 19.3 | 18.0 | 18.0 | 20.2 | 20.5 | 13.7 | 13.8 | 7.0 | 34.6 | 34.6 | 27.6 | 0.0 | 35.5 | 0.0 | | | |
| 18 | Senegal | 48.2 | 20.0 | 26.3 | 20.0 | 30.8 | 30.8 | 33.4 | 31.9 | 26.3 | 26.8 | 20.0 | 45.0 | 45.0 | 40.3 | 13.2 | 48.2 | 13.2 | 0.0 | | |
| 19 | Uganda | 11.8 | 43.3 | 17.2 | 43.3 | 12.4 | 12.4 | 10.2 | 14.5 | 17.2 | 16.6 | 23.5 | 15.7 | 15.7 | 3.2 | 30.5 | 11.8 | 30.5 | 43.0 | 0.0 | |
| 20 | Zambia | 23.5 | 41.9 | 22.2 | 41.9 | 19.9 | 19.9 | 19.8 | 15.4 | 22.2 | 23.5 | 29.6 | 4.6 | 4.6 | 14.7 | 35.9 | 23.5 | 35.9 | 47.1 | 12.4 | 0.0 |

**Appendix 8.5b.** Matrix of geographic distances among 13 populations used for Y chromosome 8-STR analysis
Distance is given in cm; scale of map was 1cm: 125km

| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **population** | Arabs | Berbers | Biaka | CAP | CAR | DRC | Ethiopia | Mali | Saharawis | San | South Africa | Uganda | Zambia |
| 1 | Arabs | 0.0 | | | | | | | | | | | | |
| 2 | Berbers | 0.0 | 0.0 | | | | | | | | | | | |
| 3 | Biaka | 32.3 | 32.3 | 0.0 | | | | | | | | | | |
| 4 | CAP | 32.3 | 32.3 | 0.0 | 0.0 | | | | | | | | | |
| 5 | CAR | 33.5 | 33.5 | 3.0 | 3.0 | 0.0 | | | | | | | | |
| 6 | DRC | 38.0 | 38.0 | 3.0 | 3.0 | 10.0 | 0.0 | | | | | | | |
| 7 | Ethiopia | 43.2 | 43.2 | 19.8 | 19.8 | 16.6 | 25.0 | 0.0 | | | | | | |
| 8 | Mali | 19.0 | 19.0 | 23.1 | 23.1 | 25.8 | 24.9 | 40.8 | 0.0 | | | | | |
| 9 | Saharawis | 12.0 | 12.0 | 32.4 | 32.4 | 34.6 | 35.7 | 47.8 | 11.4 | 0.0 | | | | |
| 10 | San | 51.8 | 51.8 | 19.7 | 19.7 | 19.8 | 14.0 | 27.3 | 38.7 | 49.5 | 0.0 | | | |
| 11 | South Africa | 59.7 | 59.7 | 27.6 | 27.6 | 27.4 | 21.8 | 32.3 | 46.4 | 57.4 | 7.9 | 0.0 | | |
| 12 | Uganda | 43.3 | 43.3 | 12.4 | 12.4 | 10.2 | 14.5 | 11.8 | 35.7 | 44.8 | 15.7 | 21.7 | 0.0 | |
| 13 | Zambia | 41.9 | 41.9 | 19.9 | 19.9 | 19.8 | 15.4 | 23.5 | 40.3 | 50.8 | 4.6 | 9.2 | 12.4 | 0.0 |

**Appendix 8.5c.** Matrix of geographic distances among 17 populations used for Y chromosome 5-STR analysis
Distance is given in cm; scale of map was 1cm. 125km

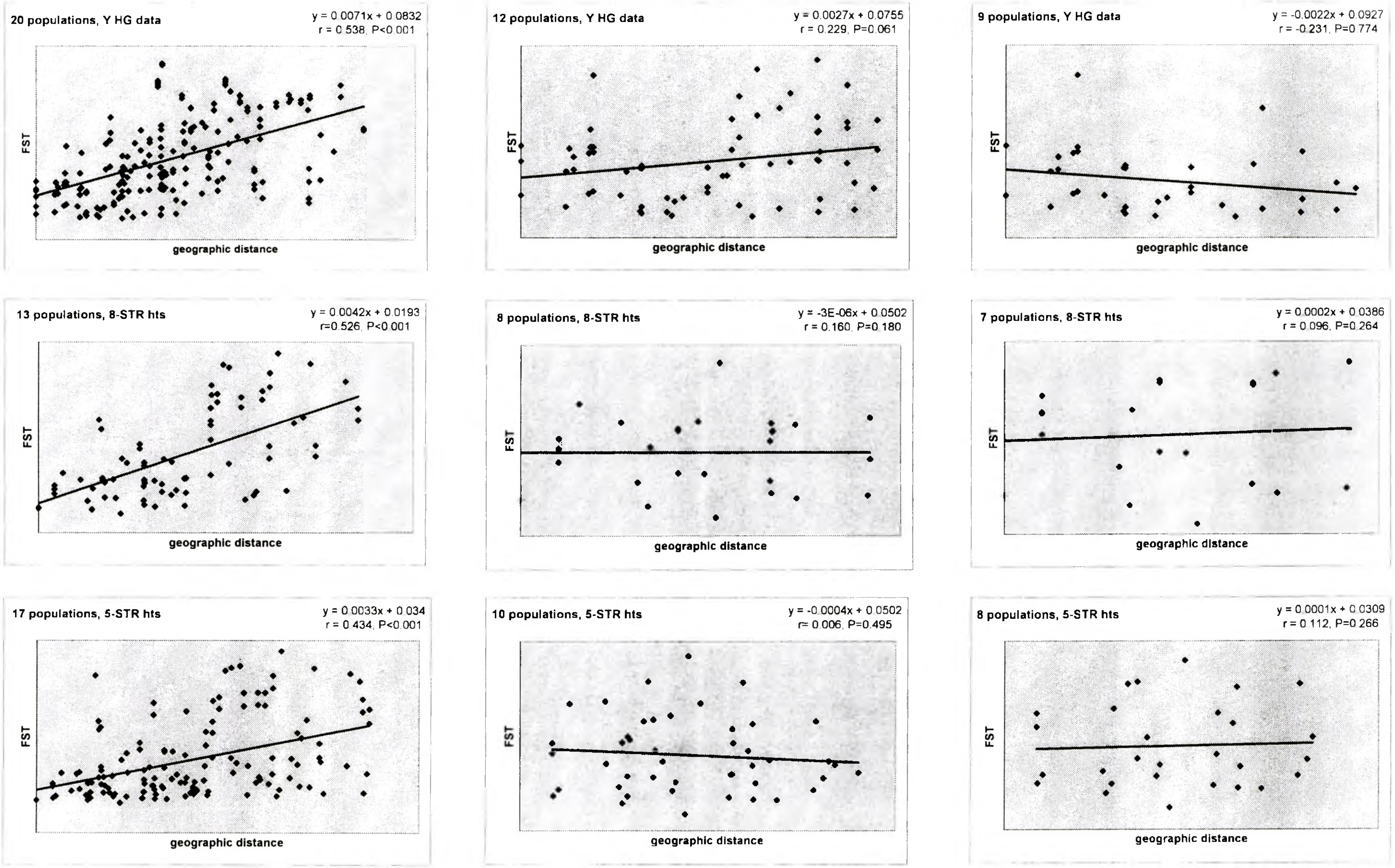
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **population** | Angola | Arabs | Berbers | Biaka | CAR | Cape Verde | DRC | Ethiopia | Guinea Bissau | Mali | Mozambique | Saharawis | San | Sao Tome | South Africa | Uganda | Zambia |
| 1 | Angola | 0.0 | | | | | | | | | | | | | | | | |
| 2 | Arabs | 40.6 | 0.0 | | | | | | | | | | | | | | | |
| 3 | Berbers | 40.6 | 0.0 | 0.0 | | | | | | | | | | | | | | |
| 4 | Biaka | 11.5 | 32.3 | 32.3 | 0.0 | | | | | | | | | | | | | |
| 5 | CAR | 13.3 | 33.5 | 33.5 | 3.0 | 0.0 | | | | | | | | | | | | |
| 6 | Cape Verde | 47.8 | 23.5 | 23.5 | 36.1 | 38.8 | 0.0 | | | | | | | | | | | |
| 7 | DRC | 3.1 | 38.0 | 38.0 | 3.0 | 10.0 | 36.7 | 0.0 | | | | | | | | | | |
| 8 | Ethiopia | 27.3 | 43.2 | 43.2 | 19.8 | 16.6 | 44.0 | 25.0 | 0.0 | | | | | | | | | |
| 9 | Guinea Bissau | 30.8 | 21.5 | 21.5 | 28.8 | 31.4 | 7.4 | 29.5 | 46.7 | 0.0 | | | | | | | | |
| 10 | Mali | 26.5 | 19.0 | 19.0 | 23.1 | 25.8 | 13.5 | 24.9 | 40.8 | 6.5 | 0.0 | | | | | | | |
| 11 | Mozambique | 22.0 | 60.9 | 60.9 | 28.8 | 28.3 | 60.5 | 24.0 | 30.9 | 52.6 | 48.2 | 0.0 | | | | | | |
| 12 | Saharawis | 37.3 | 12.0 | 12.0 | 32.4 | 34.6 | 11.5 | 35.7 | 47.8 | 10.8 | 11.4 | 59.1 | 0.0 | | | | | |
| 13 | San | 11.9 | 51.8 | 51.8 | 19.7 | 19.8 | 59.9 | 14.0 | 27.3 | 42.6 | 38.7 | 10.0 | 49.5 | 0.0 | | | | |
| 14 | Sao Tome | 10.0 | 31.1 | 31.1 | 9.5 | 12.6 | 28.9 | 8.0 | 29.1 | 21.5 | 16.6 | 31.6 | 27.5 | 21.7 | 0.0 | | | |
| 15 | South Africa | 19.5 | 59.7 | 59.7 | 27.6 | 27.4 | 56.8 | 21.8 | 32.3 | 50.0 | 46.4 | 3.6 | 57.4 | 7.9 | 29.4 | 0.0 | | |
| 16 | Uganda | 16.8 | 43.3 | 43.3 | 12.4 | 10.2 | 48.5 | 14.5 | 11.8 | 41.0 | 35.7 | 21.3 | 44.8 | 15.7 | 20.8 | 21.7 | 0.0 | |
| 17 | Zambia | 14.2 | 41.9 | 41.9 | 19.9 | 19.8 | 52.0 | 15.4 | 23.5 | 44.5 | 40.3 | 9.5 | 50.8 | 4.6 | 23.3 | 9.2 | 12.4 | 0.0 |

**Appendix 8.6** Graphs showing correlations between geographic and genetic distances using Y chromosome data

(A) full data sets

(B) populations from central and southern Africa only

(C) Niger-Congo-speaking populations from central and southern Africa only



**Appendix 8.7.** List of 41 populations used for mtDNA analysis, showing geographic sampling points.

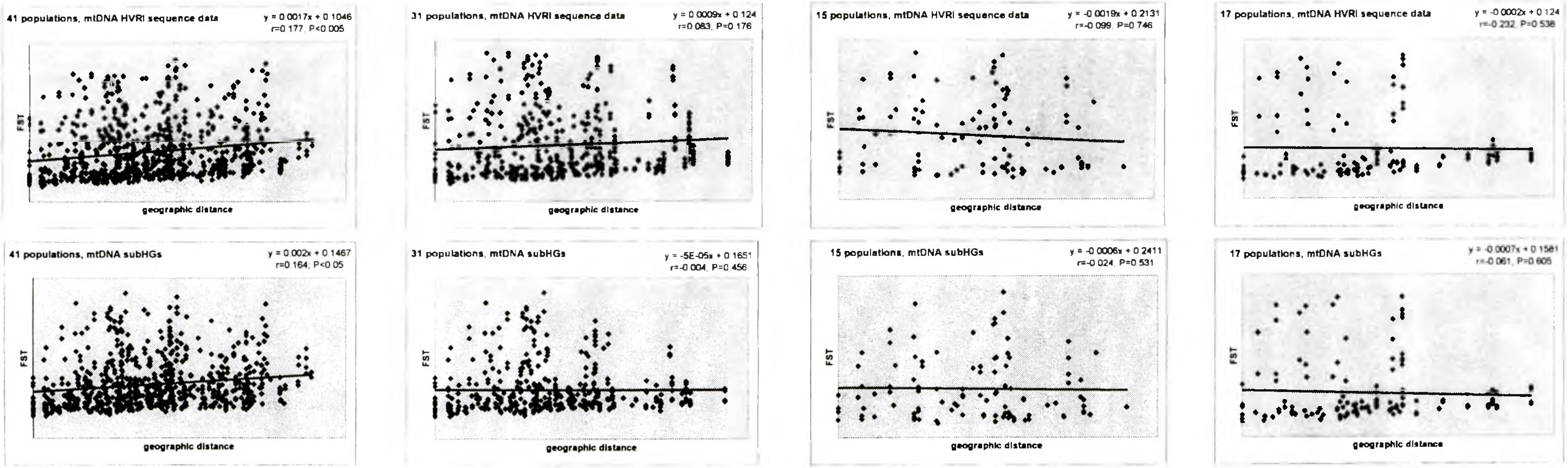
| | population | country | point on map used for distance estimates (reason) |
|---|---|---|---|
| 1 | Algeria | Algeria | Algiers (capital) |
| 2 | Berbers | Morocco | Rabat (capital) |
| 3 | Biaka Pygmies | Central African Republic | Bambio (sample collection point) |
| 4 | Bubi | Equatorial Guinea | Bata (probable origin of sample) |
| 5 | CAR Pygmies | Central African Republic | Bambio (sample collection point) |
| 6 | CAR | Central African Republic | Dula (half-way between two sample collection points) |
| 7 | Canary Islanders | Canary Islands | Tenerife (capital) |
| 8 | Datoga | Tanzania | Lake Eyasi (sample collection point) |
| 9 | Egyptians | Egypt | Cairo (capital) |
| 10 | Ethiopians | Ethiopia | Addis Ababa (capital) |
| 11 | Fang | Equatorial Guinea | Bata (probable origin of sample) |
| 12 | Fulbe | Burkina Faso and Cameroon | Kagara in Nigeria (halfway between capitals of the two countries) |
| 13 | Hadza | Tanzania | Lake Eyasi (sample collection point) |
| 14 | Hausa | Nigeria/Niger | Bambereke (halfway between two capitals) |
| 15 | Iraqw | Tanzania | Lake Eyasi (sample collection point) |
| 16 | Kanuri | Nigeria/Niger | Bambereke (halfway between two capitals) |
| 17 | Khwe | SA/ Nambia - Caprivi Strip | Kongola (estimated sample collection point) |
| 18 | Kikuyu | Kenya | Nairobi (capital) |
| 19 | !Kung 1 | SA/ Nambia - Caprivi Strip | Kongola (estimated sample collection point) |
| 20 | !Kung 2 | Botswana | Kai Kai (estimated sample collection point) |
| 21 | Mandenka | Senegal | Dakar (capital) |
| 22 | Mauritianians | Mauritiania | Nouakchott (capital) |
| 23 | Mbuti Pygmies | Democratic Republic of Congo | Nia-Nia (estimated sample collection point) |
| 24 | Morocco non-Berbers | Morocco | Rabat (capital) |
| 25 | Mozambique 1 | Mozambique | Maputo (capital) |
| 26 | Mozambique 2 | Mozambique | Maputo (capital) |
| 27 | Nubians | Sudan/Egypt | Wadi-Halfa (estimated sample collection point) |
| 28 | Sao Tome | Sao Tome / Principe | Sao Tome |
| 29 | Senegal | Senegal | Dakar (capital) |
| 30 | Serer | Senegal | Dakar (capital) |
| 31 | Somali | Somalia | Mogadishu (capital) |
| 32 | Songhai | Nigeria/Niger/Mali | Tenkodogo (middle of three capitals) |
| 33 | Sudanese | Sudan | Khartoum (capital) |
| 34 | Sukuma | Tanzania | Lake Eyasi (sample collection point) |
| 35 | Tuareg | Nigeria/Niger/Mali | Tenkodogo (middle of three capitals) |
| 36 | Turkana | Kenya | Nairobi (capital) |
| 37 | Ugandans | Uganda | Kabale (sample collection point) |
| 38 | Western Saharawis | Western Sahara | Villa Cisneros (capital) |
| 39 | Wolof | Senegal | Dakar (capital) |
| 40 | Yoruba | Nigeria | Lagos (capital) |
| 41 | Zambians | Zambia | Lusaka (capital) |

**Appendix 8.8.** Matrix of geographic distances among 41 populations used for mtDNA analysis
Distance is given in cm, scale of map was 1cm 125km

| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Algeria | Berber | Biaka | Bubi | CAR Pygmies | CAR | Canary Islands | Datoga | Egypt | Ethiopia | Fang | Fulbe | Hadza | Hausa | Iraqw | Kanuri | Khwe | Kikuyu | Kung1 | Kung2 |
| 1 | Algeria | 0.0 | | | | | | | | | | | | | | | | | | | |
| 2 | Berber | 7.9 | 0.0 | | | | | | | | | | | | | | | | | | |
| 3 | Biaka | 30.7 | 32.3 | 0.0 | | | | | | | | | | | | | | | | | |
| 4 | Bubi | 31.0 | 31.0 | 6.5 | 0.0 | | | | | | | | | | | | | | | | |
| 5 | CAR Pygmies | 30.7 | 32.3 | 0.0 | 6.5 | 0.0 | | | | | | | | | | | | | | | |
| 6 | Central African Republic | 31.4 | 33.5 | 3.0 | 9.6 | 3.0 | 0.0 | | | | | | | | | | | | | | |
| 7 | Canary Islands | 16.5 | 9.1 | 34.4 | 31.3 | 34.4 | 36.1 | 0.0 | | | | | | | | | | | | | |
| 8 | Datoga | 34.3 | 47.8 | 17.4 | 23.0 | 17.4 | 14.9 | 51.2 | 0.0 | | | | | | | | | | | | |
| 9 | Egypt | 22.4 | 29.8 | 25.8 | 30.5 | 25.8 | 23.9 | 37.5 | 29.8 | 0.0 | | | | | | | | | | | |
| 10 | Ethiopia | 38.4 | 43.2 | 19.8 | 26.3 | 19.8 | 16.6 | 48.4 | 11.6 | 19.5 | 0.0 | | | | | | | | | | |
| 11 | Fang | 31.0 | 31.0 | 6.5 | 0.0 | 6.5 | 9.6 | 31.3 | 23.0 | 30.5 | 26.3 | 0.0 | | | | | | | | | |
| 12 | Fulbe | 23.9 | 23.1 | 11.0 | 8.2 | 11.0 | 13.4 | 23.6 | 28.4 | 27.3 | 28.8 | 8.2 | 0.0 | | | | | | | | |
| 13 | Hadza | 34.3 | 47.8 | 17.4 | 23.0 | 17.4 | 14.9 | 51.2 | 0.0 | 29.8 | 11.6 | 23.0 | 28.4 | 0.0 | | | | | | | |
| 14 | Hausa | 23.2 | 21.8 | 13.8 | 9.8 | 13.8 | 16.3 | 21.6 | 31.2 | 29.8 | 31.5 | 9.8 | 3.2 | 31.2 | 0.0 | | | | | | |
| 15 | Iraqw | 34.3 | 47.8 | 17.4 | 23.0 | 17.4 | 14.9 | 51.2 | 0.0 | 29.8 | 11.6 | 23.0 | 28.4 | 0.0 | 31.2 | 0.0 | | | | | |
| 16 | Kanuri | 23.2 | 21.8 | 13.8 | 9.8 | 13.8 | 16.3 | 21.6 | 31.2 | 29.8 | 31.5 | 9.8 | 3.2 | 31.2 | 0.0 | 31.2 | 0.0 | | | | |
| 17 | Khwe | 50.6 | 51.8 | 19.7 | 21.0 | 19.7 | 19.8 | 52.4 | 16.3 | 42.3 | 27.3 | 21.0 | 28.8 | 16.3 | 30.4 | 16.3 | 30.4 | 0.0 | | | |
| 18 | Kikuyu | 43.6 | 47.3 | 18.3 | 24.1 | 18.3 | 15.6 | 51.2 | 2.6 | 27.9 | 9.3 | 24.1 | 28.8 | 2.6 | 31.6 | 2.6 | 31.6 | 18.9 | 0.0 | | |
| 19 | Kung1 | 50.6 | 51.8 | 19.7 | 21.0 | 19.7 | 19.8 | 52.4 | 16.3 | 42.3 | 27.3 | 21.0 | 28.8 | 16.3 | 30.4 | 16.3 | 30.4 | 0.0 | 18.9 | 0.0 | |
| 20 | Kung2 | 51.4 | 52.1 | 20.8 | 21.3 | 20.8 | 21.1 | 52.4 | 18.9 | 44.0 | 29.7 | 21.3 | 29.3 | 18.9 | 30.7 | 18.9 | 30.7 | 2.5 | 21.4 | 2.5 | 0.0 |
| 21 | Mandenka | 26.3 | 20.0 | 30.7 | 25.8 | 30.7 | 33.4 | 12.8 | 48.1 | 42.6 | 48.4 | 25.8 | 20.2 | 48.1 | 17.1 | 48.1 | 17.1 | 45.4 | 48.9 | 45.4 | 45.0 |
| 22 | Mauritiania | 22.9 | 16.7 | 30.5 | 26.1 | 30.5 | 33.0 | 9.5 | 38.0 | 40.4 | 47.2 | 26.1 | 19.6 | 38.0 | 16.8 | 38.0 | 16.8 | 46.3 | 48.3 | 46.3 | 46.0 |
| 23 | Mbuti | 37.0 | 40.0 | 9.7 | 15.7 | 9.7 | 7.0 | 43.1 | 7.9 | 25.2 | 11.7 | 15.7 | 20.4 | 7.9 | 23.4 | 7.9 | 23.4 | 17.9 | 8.4 | 17.9 | 19.5 |
| 24 | Morocco | 7.9 | 0.0 | 32.3 | 31.0 | 32.3 | 33.5 | 9.1 | 47.8 | 29.8 | 43.2 | 31.0 | 23.1 | 47.8 | 21.8 | 47.8 | 21.8 | 51.8 | 47.3 | 51.8 | 52.1 |
| 25 | Mozambique1 | 59.5 | 61.2 | 28.9 | 30.8 | 28.9 | 28.3 | 62.4 | 19.4 | 48.6 | 31.0 | 30.8 | 38.5 | 19.4 | 40.4 | 19.4 | 40.4 | 10.1 | 21.9 | 10.1 | 10.6 |
| 26 | Mozambique2 | 59.5 | 61.2 | 28.9 | 30.8 | 28.9 | 28.3 | 62.4 | 19.4 | 48.6 | 31.0 | 30.8 | 38.5 | 19.4 | 40.4 | 19.4 | 40.4 | 10.1 | 21.9 | 10.1 | 10.6 |
| 27 | Nubia | 26.0 | 32.0 | 20.0 | 25.6 | 20.0 | 17.8 | 38.6 | 22.8 | 7.1 | 13.0 | 25.6 | 23.8 | 22.8 | 26.7 | 22.8 | 26.7 | 35.4 | 20.9 | 35.4 | 37.4 |
| 28 | Sao Tome / Principe | 32.1 | 31.3 | 9.5 | 3.0 | 9.5 | 12.6 | 30.7 | 25.5 | 33.2 | 29.3 | 3.0 | 8.8 | 25.5 | 9.4 | 25.5 | 9.4 | 21.7 | 26.8 | 21.7 | 21.5 |
| 29 | Senegal | 26.3 | 20.0 | 30.7 | 25.8 | 30.7 | 33.4 | 12.8 | 48.1 | 42.6 | 48.4 | 25.8 | 20.2 | 48.1 | 17.1 | 48.1 | 17.1 | 45.4 | 48.9 | 45.4 | 45.0 |
| 30 | Serer | 26.3 | 20.0 | 30.7 | 25.8 | 30.7 | 33.4 | 12.8 | 48.1 | 42.6 | 48.4 | 25.8 | 20.2 | 48.1 | 17.1 | 48.1 | 17.1 | 45.4 | 48.9 | 45.4 | 45.0 |
| 31 | Somalia | 46.7 | 51.4 | 25.0 | 31.2 | 25.0 | 22.1 | 56.2 | 10.2 | 27.6 | 8.5 | 31.2 | 35.1 | 10.2 | 38.0 | 10.2 | 38.0 | 26.0 | 7.9 | 26.0 | 28.6 |
| 32 | Songhai | 22.2 | 19.9 | 16.7 | 12.6 | 16.7 | 19.2 | 18.9 | 34.0 | 31.0 | 33.8 | 12.6 | 5.8 | 34.0 | 2.9 | 34.0 | 2.9 | 33.1 | 34.3 | 33.1 | 33.2 |
| 33 | Sudan | 30.0 | 35.4 | 17.0 | 23.3 | 17.0 | 14.3 | 41.1 | 11.3 | 12.7 | 7.9 | 23.3 | 23.4 | 11.3 | 26.5 | 11.3 | 26.5 | 30.4 | 15.3 | 30.4 | 32.3 |
| 34 | Sukuma | 34.3 | 47.8 | 17.4 | 23.0 | 17.4 | 14.9 | 51.2 | 0.0 | 29.8 | 11.6 | 23.0 | 28.4 | 0.0 | 31.2 | 0.0 | 31.2 | 16.3 | 2.6 | 16.3 | 18.9 |
| 35 | Tuareg | 22.2 | 19.9 | 16.7 | 12.6 | 16.7 | 19.2 | 18.9 | 34.0 | 31.0 | 33.8 | 12.6 | 5.8 | 34.0 | 2.9 | 34.0 | 2.9 | 33.1 | 34.3 | 33.1 | 33.2 |
| 36 | Turkana | 43.6 | 47.3 | 18.3 | 24.1 | 18.3 | 15.6 | 51.2 | 2.6 | 27.9 | 9.3 | 24.1 | 28.8 | 2.6 | 31.6 | 2.6 | 31.6 | 18.9 | 0.0 | 18.9 | 21.4 |
| 37 | Uganda | 40.2 | 43.3 | 12.4 | 18.3 | 12.4 | 10.2 | 46.4 | 4.4 | 27.4 | 11.8 | 18.3 | 23.5 | 4.4 | 26.2 | 4.4 | 26.2 | 15.7 | 5.9 | 15.7 | 17.9 |
| 38 | Western Sahara | 19.3 | 12.0 | 32.4 | 28.6 | 32.4 | 34.6 | 4.3 | 39.5 | 38.4 | 47.8 | 28.6 | 21.3 | 39.5 | 19.0 | 39.5 | 19.0 | 49.5 | 49.5 | 49.5 | 49.2 |
| 39 | Wolof | 26.3 | 20.0 | 30.7 | 25.8 | 30.7 | 33.4 | 12.8 | 48.1 | 42.6 | 48.4 | 25.8 | 20.2 | 48.1 | 17.1 | 48.1 | 17.1 | 45.4 | 48.9 | 45.4 | 45.0 |
| 40 | Yoruba | 26.4 | 25.2 | 12.2 | 7.1 | 12.2 | 15.1 | 24.6 | 29.6 | 31.3 | 31.0 | 7.1 | 4.2 | 29.6 | 3.4 | 29.6 | 3.4 | 27.8 | 30.2 | 27.8 | 28.0 |
| 41 | Zambia | 50.1 | 41.9 | 19.9 | 22.1 | 19.9 | 19.8 | 53.2 | 12.0 | 39.8 | 23.5 | 22.1 | 29.6 | 12.0 | 41.8 | 12.0 | 41.8 | 4.6 | 14.5 | 4.6 | 7.2 |

| | | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Mandenka | Mauritiania | Mbuti | Morocco | Mozambique1 | Mozambique2 | Nubia | Sao Tome | Senegal | Serer | Somalia | Songhai | Sudan | Sukuma | Tuareg | Turkana | Uganda | Western S | Wolof | Yoruba | Zambia |
| 21 | Mandenka | 0.0 | | | | | | | | | | | | | | | | | | | | |
| 22 | Mauritiania | 3.5 | 0.0 | | | | | | | | | | | | | | | | | | | |
| 23 | Mbuti | 40.5 | 40.0 | 0.0 | | | | | | | | | | | | | | | | | | |
| 24 | Morocco | 20.0 | 16.7 | 40.0 | 0.0 | | | | | | | | | | | | | | | | | |
| 25 | Mozambique1 | 55.5 | 56.5 | 24.0 | 61.2 | 0.0 | | | | | | | | | | | | | | | | |
| 26 | Mozambique2 | 55.5 | 56.5 | 24.0 | 61.2 | 0.0 | 0.0 | | | | | | | | | | | | | | | |
| 27 | Nubia | 41.6 | 39.8 | 18.2 | 32.0 | 41.6 | 41.6 | 0.0 | | | | | | | | | | | | | | |
| 28 | Sao Tome / Principe | 24.0 | 24.7 | 18.7 | 31.3 | 31.6 | 31.6 | 28.5 | 0.0 | | | | | | | | | | | | | |
| 29 | Senegal | 0.0 | 3.5 | 40.5 | 20.0 | 55.5 | 55.5 | 41.6 | 24.0 | 0.0 | | | | | | | | | | | | |
| 30 | Serer | 0.0 | 3.5 | 40.5 | 20.0 | 55.5 | 55.5 | 41.6 | 24.0 | 0.0 | 0.0 | | | | | | | | | | | |
| 31 | Somalia | 55.3 | 44.3 | 15.4 | 51.4 | 27.0 | 27.0 | 21.5 | 34.0 | 55.3 | 55.3 | 0.0 | | | | | | | | | | |
| 32 | Songhai | 14.3 | 13.9 | 26.2 | 19.9 | 43.2 | 43.2 | 28.8 | 11.6 | 14.3 | 14.3 | 40.4 | 0.0 | | | | | | | | | |
| 33 | Sudan | 42.4 | 41.0 | 13.1 | 35.4 | 36.1 | 36.1 | 5.6 | 26.3 | 42.4 | 42.4 | 16.4 | 28.8 | 0.0 | | | | | | | | |
| 34 | Sukuma | 48.1 | 38.0 | 7.9 | 47.8 | 19.4 | 19.4 | 22.8 | 25.5 | 48.1 | 48.1 | 10.2 | 34.0 | 11.3 | 0.0 | | | | | | | |
| 35 | Tuareg | 14.3 | 13.9 | 26.2 | 19.9 | 43.2 | 43.2 | 28.8 | 11.6 | 14.3 | 14.3 | 40.4 | 0.0 | 28.8 | 34.0 | 0.0 | | | | | | |
| 36 | Turkana | 48.9 | 48.3 | 8.4 | 47.3 | 21.9 | 21.9 | 20.9 | 26.8 | 48.9 | 48.9 | 7.9 | 34.3 | 15.3 | 2.6 | 34.3 | 0.0 | | | | | |
| 37 | Uganda | 43.3 | 43.0 | 3.2 | 43.3 | 21.4 | 21.4 | 20.4 | 20.8 | 43.3 | 43.3 | 13.6 | 28.8 | 15.0 | 4.4 | 28.8 | 5.9 | 0.0 | | | | |
| 38 | Western Sahara | 8.5 | 5.1 | 41.5 | 12.0 | 59.4 | 59.4 | 38.7 | 27.7 | 8.5 | 8.5 | 45.0 | 16.0 | 40.7 | 39.5 | 16.0 | 49.5 | 44.8 | 0.0 | | | |
| 39 | Wolof | 0.0 | 3.5 | 40.5 | 20.0 | 55.5 | 55.5 | 41.6 | 24.0 | 0.0 | 0.0 | 55.3 | 14.3 | 42.4 | 48.1 | 14.3 | 48.9 | 43.3 | 8.5 | 0.0 | | |
| 40 | Yoruba | 18.8 | 19.0 | 21.9 | 25.2 | 37.8 | 37.8 | 27.7 | 6.0 | 18.8 | 18.8 | 37.0 | 5.7 | 26.7 | 29.6 | 5.7 | 30.2 | 24.5 | 21.6 | 18.8 | 0.0 | |
| 41 | Zambia | 47.3 | 47.9 | 14.7 | 41.9 | 9.6 | 9.6 | 32.8 | 23.5 | 47.3 | 47.3 | 21.4 | 34.2 | 27.3 | 12.0 | 34.2 | 14.5 | 12.4 | 50.8 | 47.3 | 29.0 | 0.0 |

**Appendix 8.9** Graphs showing correlations between geographic and genetic distances using mtDNA data



(A) full data sets

(B) excluding populations from north-east and north-west Africa

(C) populations from central and southern Africa only

(D) Niger-Congo-speaking populations only