Genetic Patterns of Y-chromosome and Mitochondrial DNA Variation, with Implications to the Peopling of the Sudan

Hisham Yousif Hassan Mohamed

B.sc. (Honors) of Zoology, University of Khartoum, 1999
Upgrading to PhD. Institute of Endemic Diseases, 2004

Thesis Submitted for the Fulfillment of Requirements
for Philosophy Degree of Science in Molecular Biology

Supervisor: Prof. Muntaser Eltayeb Ibrahim
Department of Molecular Biology
Institute of Endemic Diseases
University of Khartoum

July 2009

DEDICATION

To my beloved family

LIST OF CONTENTS

ACKNOWLEDGEMENTS

Firstly, I would like to thank all scientists from all round the world who have contributed to the understanding of our past through the growing gene trees of modern humans since the first publication about the origin of human by Rebecca Cann and Alan Wilson and their colleagues in the mid eighties. At the top of this list of acknowledgements I would like to thank my supervisor Muntaser E. Ibrahim who encouraged me to follow the route of my interest in human diversity, and for giving me his time over several years to explain to me the genetic view of human diversity, archeology, and history of the Sudan, indeed without his efforts, guidance and supervision the study would not have been possible. I thank all participants form different areas in the Sudan who generously donated DNA samples so that we might learn more about their population history. The financial assistance of the Institute of Endemic Diseases (IEND) towards this research is hereby acknowledged.

Secondly, there are other scientists who helped me to complete this study, Himla Soodyall who gave me the opportunity to do the analysis of mitochondrial DNA in her lab at the National Health Laboratory Service, Johannesburg, SA, my thanks are extended to her crew Heeran Makkan and Carina Schlebusch for helping me in the lab work. Peter Underhill at Stanford University who offered me very useful advice and good counseling during the analysis of Y-chromosome. Salah M. Ahmed, Mahmoud S. Bashir, Yves Lecocinte, Jacques Reinold and Khider Abdelkarim who provided the ancient bone samples from the Sudan National Museum and the Dinder National Park. My thank also goes to Sheima Abdallah, a valuable help and advice in extraction of ancient DNA during the initial stages of this study.

I would like to express my gratitude to my colleagues at the Institute of Endemic Diseases for the time spent patiently explaining things to me in the lab, and offering their own insights. My friends Assad Haboub, Sami Saeed and the staff at the Archeology Department, Faculty of Arts were a great source of references and publications where available. I thank Hiba Abu Turkey from Directorate General of Pharmacy for technical assistance in graphics. There are a number of others not in the above groups who have helped me. Finally, thanks, and appreciations are extended to my family for providing emotional and financial support. Most especially, the wisdom, insight and patience of my mother must be recognized. Sincere thank goes to my uncle Abdel Azeim Hassan, my brother and sisters.

ABSTRACT

The area known today as Sudan may have been the scene of pivotal human evolutionary events, both as a corridor for ancient and modern migrations, as well as the venue of crucial past cultural evolution. Several questions pertaining to the pattern of succession of the different groups in early Sudan have been raised. To shed light on these aspects, ancient DNA (aDNA) and present DNA collection were made and studied using Y-chromosome markers for aDNA, and Y-chromosome and mtDNA markers for present DNA. Bone samples from different skeletal elements of burial sites from Neolithic, Meroitic, Post-Meroitic and Christian periods in Sudan were collected from Sudan National Museum. aDNA extraction was successful in 35 out of 76 samples, PCR was performed for sex determination using Amelogenin marker. Fourteen samples were females and 19 were males. To generate Y-chromosome specific haplogroups A-M13, B-M60, F-M89 and Y Alu Polymorphism (YAP) markers, which define the deep ancestral haplotypes in the phylogenetic tree of Y-chromosome were used. Haplogroups A-M13 was found at high frequencies among Neolithic samples. Haplogroup F-M89 and YAP appeared to be more frequent among Meroitic, Post-Meroitic and Christian periods. Haplogroup B-M60 was not observed in the sample analyzed.

For extant DNA, Y-chromosome and mtDNA haplogroup variations were studied in 15 Sudanese populations representing the three linguistic families in Sudan by typing the major Y haplogroups in 445 unrelated males, and 404 unrelated individuals were sequenced for the mitochondrial hypervariable region. Y-chromosome analysis shows Sudanese populations falling into haplogroups A, B, E, F, I, J, K, and R in frequencies of 16.9, 8.1, 34.2, 3.1, 1.3, 22.5, 0.9, and 13% respectively. Haplogroups A, B, and E occur

mainly in Nilo-Saharan speaking groups including Nilotics, Fur, Borgu, and Masalit; whereas haplogroups F, I, J, K, and R are more frequent among Afro-Asiatic speaking groups including Arabs, Beja, Copts, and Hausa, and Niger-Congo speakers from the Fulani ethnic group. Mantel test reveal a strong correlation between genetic and linguistic structures (r= 0.30, p= 0.007), and a similar correlation between genetic and geographic distances (r= 0.29, p= 0.025) that appears after removing nomadic pastoralists of no known geographic locality from the analysis. For mtDNA analysis, a total of 56 haplotypes were observed, all belonging to the major sub-Saharan African and Eurasian mitochondrial macrohapolgroups L0, L1, L2, L4, L5, L3A, M and N in frequencies of 12.1, 11.9, 22, 4.2, 6.2, 29.5, 2, and 12.2% respectively. Haplogroups L6 was not observed in the sample analyzed. The considerable frequencies of macrohaplogroup L0 in Sudan is interesting given the fact that this macrohaplogroup occurs near the root of the mitochondrial DNA tree. Afro-Asiatic speaking groups appear to have sustained high gene flow form Nilo-Saharan speaking groups. Mantel test reveal no correlations between genetic, linguistic (r = 0.12, p = 0.14), and geographic distances (r = -0.07, p = 0.67).

Accordingly, through limited on number of aDNA samples, there is enough data to suggest and to tally with the historical evidence of the dominance by Nilotic elements during the early state formation in the Nile Valley, and as the states thrived there was a dominance by other elements particularly Nuba/Nubians. In Y-chromosome terms this mean in simplest terms introgression of the YAP insertion (haplogroups E and D), and Eurasian Haplogroups which are defined by F-M89 against a background of haplogroup A-M13. The data analysis of the extant Y-chromosomes suggests that the bulk of genetic diversity appears to be a consequence of recent migrations and demographic events

mainly from Asia and Europe, evident in a higher migration rate for speakers of Afro-Asiatic as compared to the Nilo-Saharan family of languages, and a generally higher effective population size for the former. While the mtDNA data suggests that regional variation and diversity in mtDNA sequences in Sudan is likely to have been shaped by a longer history of in-situ evolution and then by human migrations form East, west-central and North Africa and to a lesser extent from Eurasia to the Nile Valley.

المستخلص

المنطقة التى تعرف الآن بالسودان لعبت دوراً هاماً فى الأحداث المرتبطة بتاريخ الإنسان التطورى، و كذلك كمعبر للعديد من الهجرات القديمة و الحديثة و دورها فى تطور الثقافات القديمة. هنالك العديد من الأسئلة المتعلقة بتعاقب المجموعات السكانية فى السودان على مدى التاريخ البشرى. لإلقاء الضوء على هذه المظاهر، جُمعت عينات من الحمض النووى القديم و الحمض النووى الحديث. تمت دراسه الحمض النووى القديم على مستوى الكروموسوم الذكرى، أما الحمض النووى الحديث فقد تمت دراسته على مستوى الكروموسوم الذكرى و الميتوكزندريا. عينات العظام التى أستخدمت فى تحليل الحمض النووى القديم تم جمعها سابقاً بواسطة البعثات الأثرية بمتحف السودان القومى فى أجزاء هيكلية مختلفة من مدافن اثرية تمثل فترات العصر الحجرى، الفترة المروية، فترة ما بعد مروى و الفترة المسيحية.

عملية إستخلاص الحمض النووى كانت ناجحة فى عدد 35 عينة من مجموع 76 عينة. تحديد الجنس من حيث الذكورة و الأنوثة تمت بواسطة معلم وراثى يسمى Amelogenin بإستخدام تقنية التفاعل الأنزيمى المتبلمر. أربعة عشر عينة كانت من الإناث بينما 19 عينة من الذكور. لتحديد المجموعات الوراثية للكروموسوم الذكرى فى عينات الحمض الوراثى القديم، أستخدمت المعلمات الوراثية M13، M60، M89 و YAP. هذه المعلمات الوراثية تحدد المجموعات الوراثية الأساسية فى الشجرة الوراثية للكروموسوم الذكرى. المعلم الوراثي M13 وجد بتكرار مقدر فى مجموعة العصر الحجرى. أما المعلمان الوراثيان M89 و YAP فقد ظهرا بتكرار أعلى فى فترة مروى و الفترة ما بعد مروى و الفترة المسيحية. لم يلاحظ وجود المعلم الوراثى M60 فى العينات التى تم تحليلها.

بالنسبة للحمض النووى الحديث، فقد تمت دراسة التباين على مستوى المجموعات الوراثية للكروموسوم الذكرى و الميتوكوندريا فى 15 مجموعة اثنية فى السودان يمثلون ثلاث مجموعات لغوية فى السودان. و ذلك بدراسة 445 عينة من الرجال غير الأقارب للكروموسوم الذكرى و 404 عينة من أشخاص غير أقارب للميتوكوندريا.

أظهرت نتائج الكروموسوم الذكرى أن السودانيين يقعون فى المجموعات الوراثية A، B، E، F، I، J، K و R بتكرار 16.9، 8.1، 34.2، 3.1، 1.3، 22.5، 0.9 و 13% على التوالى. المجموات الوراثية A، B، E وجدت بتكرار أعلى فى المجموعات التى تتحدث اللغات النيلية الصحراوية مثل النيليين، الفور، البرقو و المساليت. بينما المجموعات الوراثية F، I، J، K و R فقد وجدت أكثر تكراراً فى المجموعات التى تتحدث اللغات الأفروأسيوية مثل

العرب، البجة، الأقباط و الهوسا، و كذلك مجموعة الفولانى التى تتحدث لغة تنتمى الى مجموعة اللغات النيجركردفانية.

إختبار مانتل اوضح وجود علاقة قوية بين التركيبة الوراثية و اللغة ($r = 0.30, P = 0.007$) و التركيبة الوراثية و التوزيع الجغرافى ($r = 0.29, P = 0.025$) على مستوى الكروموسوم الذكرى. هذا الإرتباط بين التركيبة الوراثية و التوزيع الجغرافى ظهر عند إستبعاد المجموعات الرعوية التى لا يوجد لها مكان جغرافى ثابت تستقر فيه.

بالنسبة للميتوكوندريا، فقد لوحظ وجود 56 معلم وراثى فى العينات التى تم تحليلها. تنتمى هذه المعلمات الوراثية للمجموعات الأساسية L0، L1، L2، L4، L5، L3A، M و N بتكرار 12.1، 11.9، 22، 4.2، 6.2، 29.5، 2 و 12.2% على التوالى. المجموعة الوراثية L6 لم يلاحظ وجودها فى العينات المحللة. التكرار المقدر للمجموعة الوراثية L0 فى السودان و التى تقع بالقرب من أصل الشجرة الوراثية للميتوكوندريا يعطى هذه النتائج أهمية خاصة. أوضحت نتائج هذه الدراسة أن المجموعات التى تتحدث اللغات الأفروأسيوية استقبلت كمية مقدرة من الجينات من الجموعات التى تتحدث اللغات النيلية الصحراوية. أوضح إختبار مانتل عدم وجود إرتباط بين التركيبة الوراثية و اللغة ($r = 0.12, P = 0.14$)، او التركيبة الوراثية و التوزيع الجغرافى ($r = -0.07, P = 0.63$).

من خلال ما ذكر، و خلال العدد الموجود من العينات التى تم تحليلها على مستوى الحمض النووى القديم، أوضحت النتائج سيادة المجموعات النيلية فى منطقة حوض النيل خلال العصر الحجرى، أعقبتها سيادة مجموعات أخرى خاصة النوبة و النوبيين خلال الفترات اللاحقة. تم الإستدلال على ذلك بظهور المعلمات الوراثية YAP و M89 على حساب المعلم الوراثي M13 و الذى كان سائدا خلال العصر الحجرى. النتائج على مستوى الحمض النووى الحديث للكروموسوم الذكرى اوضحت أن التباين الوراثى فى السودان هو نتيجة هجرات حديثة متعاقبة من أسيا و أوروبا، و هو ما تم الإستدلال عليه بمعدل الهجرات العالى و الحجم المؤثر للمجموعة الذى يميز المجموعات التى تتحدث اللغات الأفروأسيوية مقارنة بتلك التى تتحدث اللغات النيلية الصحراوية. بينما أوضحت النتائج أن التباين الوراثى على مستوى الميتوكوندريا فى السودان من الأرجح أن يكون قد تشكل على مدى تاريخ طويل من التطور فى نفس المنشأ ثم بالهجرات البشرية من شرق، غرب و وسط أفريقيا و بشكل أقل من أوروبا و أسيا الى منطقة وادى النيل.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AD	Anno Domini (Medieval Latin: In the Year of the Lord Jesus)
aDNA	Ancient DNA
AMOVA	Analysis of molecular variance
ARMS	Amplification refractory mutation system
BC	Before Christ
BP	Before present
bp	Base pair
CR-I	Control region I
CR-II	Control region II
D-loop	Displacement loop
DNA	Deoxyribonucleic acid
dNTPs	Deoxy nucleoside triphosphate
EDTA	Ethylene diamine tetra acetic acid
Fig	Figure
F_{ST}	Whright's fixation index
G6PD	Glucose-6-phosphate dehydrogenase
GnSCN	Silica-Guanidine theocyanate
HP	Haplogroup
HVR-I	Hypervariable region I
HVR-II	Hypervariable region II
ME	Minimum evolution
min	Minute

mtDNA	Mitochondrial DNA
NC	Niger-Congo
NJ	Neighbor-joining tree
NRY	Non-recombining region of Y-chromosome
NS	Nilo-Saharan
OTUs	Operational taxonomic units
PBS	Phosphate buffer saline
PCA	Principal component analysis
PCR	Polymerase chain reaction
RBCs	Red blood cells
RFLP	Restriction fragment length polymorphism
SD	Standard deviation
SNP	Single nucleotide polymorphism
SDS	Sodium didocyl sulphate
SE	Sodium chloride EDTA
STE	Sodium chloride–Tris–EDTA
MRCA	Most recent common ancestor
Taq	*Thermus aquaticus*
TMRCA	Time of most recent common ancestor
TBE	Tris-boric Acid-EDTA
UEP	Unique event polymorphism
UPGMA	Unweighted pair group method using arithmetic averages
UV	Ultra violet

YAP	Y-chromosome Alu polymorphism
YCC	Y-chromosome Consortium
Y-STRs	Y-chromosome short tandem repeats

CHAPTER **1**

INTRODUCTION & LITERATURE REVIEW

1.1 General Introduction

Analysis of human remains, as well as ancient objects play a crucial role in studying the human origin, tracing human migration and reconstructing ancient events in human history (McBrearty and Brooks, 2000; White *et al.*, 2003; McDougall and Fleagle 2005; Trinkaus, 2005). Archaeological and fossil evidences suggest that modern humans originated in Africa during the Middle Stone Age around 115,000–130,000 years ago or somewhat earlier (Grün and Stringer, 1991). And there are signs of modern human behavior from ~ 70,000 years ago (Deacon, 1989), which become more fully apparent from ~ 40,000 years ago, with the onset of the Late Stone Age (Clark, 1994). In West Africa, the Middle Stone Age and Late Stone Age are poorly understood, but East, Central, and southern Africa provide evidence of regional traditions dating back to the Acheulian ~ 120,000 years ago (Phillipson, 1993). North Africa has had a distinct history; climatic fluctuations would have played an important role in establishing the substrate of human habitation in the region. Widespread aridity occurred during the Last Glacial Maximum about 12,000 years ago in North Africa followed by a Holocene Lacustral Phase (Street and Grove, 1976). The beginning of sedentary settlements in Africa can be dated to ~ 18,000 years ago, in the favorable environment afforded by the Nile Valley (Phillipson, 1993).

Anthropologists were quick to realize the potential of new techniques in molecular biology to provide additional lines of evidence on questions of interest that could not

have been addressed by traditional anthropological techniques (Kaestle and Horsburgh, 2002). Studies of protein polymorphisms, as well as studies of mitochondrial DNA (mtDNA), Y-chromosomal, autosomal and X-chromosomal DNA variation indicate that African populations have a rich and important history from which we can learn much about our origins as a species (Cann *et al.*, 1987; Cavalli-Sforza *et al.*, 1994; Zietkiewicz *et al.*, 1997; Seielstad *et al.*, 1999; Chen *et al.*, 2000; Jorde *et al.*, 2000; Underhill *et al.*, 2000) and the way in which variation affects human phenotypes, including diseases. Studies using mitochondrial and nuclear DNA markers consistently indicate that Africa is the most genetically diverse region of the world (Calafell *et al.*, 1998; Ingman *et al.*, 2000; Alonso and Amour, 2001). However, most studies report only a few markers in divergent African populations, which make it difficult to draw general conclusions about the levels and patterns of genetic diversity in these populations. Historically human population genetic studies have relied on very few numbers of African populations as being representative of African diversity, but recent studies showed extensive genetic variation among even geographically close African populations, which indicates that there is not a single representative African population (Tishkoff and Williams, 2002).

1.2 Out of Africa

The study of human evolution provides a powerful illustration of the coalescence principle applied to the derivation of most recent common ancestors. According to Cann *et al.* (1987) all humans today can be traced along maternal lines of descent to a woman who lived about 200,000 years ago, in Africa. Modern humans arose in one place and spread elsewhere out of Africa to Eurasia and the rest of the world, analysis of the Y-chromosome showed that all extant human males are traced back to a most recent common ancestor who lived 40,000-140,000 years ago (Thomson *et al.*, 2000).

Therefore, there are thousands of years that separate mtDNA and Y-chromosome MRCAs due to the genetic bottleneck which has occurred for Y-chromosome probably before ~50,000 years ago (Olson, 2002). Humans spread to Asia through East Africa around 100-86 years ago (Underhill *et al.*, 2000; Cavalli-Sforza and Feldman, 2003). Then, between 40-60 years ago there was a second expansion, most probably from a descendant population and again into Asia. Two possible routes out of Africa have been proposed: down the Nile and then by land across the Sinai Peninsula, or by the Red Sea across the Gate of Tears Strait to present-day Yemen (Tchernov, 1992; Cavalli-Sforza *et al.*, 1993; Lahr and Foley, 1994). Then humans expanded from Asia to Australia ~ 50,000 years ago, Europe ~ 35,000 years ago and America ~ 15,000 years ago (Fagan, 1987; Stringer, 2000).

1.3 African Languages

Because of the rather poor state of archaeological understanding, especially within the tropical forest zone, linguistics has played a large role in understanding African prehistory. Approximately 1400 languages are spoken today in Africa (Ruhlen, 1987). These languages belong to four major phyla: Niger-Congo (including the Atlantic, Mande, Voltaic, Kwa, Adamawa, and Bantu families), Nilo-Saharan (including east and central Sudanic, Saharan, and Songhai), Afro-Asiatic (Semitic, Berber, Cushitic, and Chadic) and Khoisan including San and Khoi-Khoi (Greenberg, 1963). It has been suggested that the initial development of the first three families took place somewhere between the Sahara and the equatorial forest (Blench, 1993); Niger-Congo and Nilo-Saharan may even share a common ancestor (Phillipson, 1993). The distribution of Khoisan languages may have extended, before the Bantu Diasporas, to present-day Ethiopia and Sudan. This is surmised in the presence of small groups of people speaking

click-language isolates in Tanzania (Hadza and Sandawe) (Greenberg, 1963; Blench, 1993) and, more controversially, from the presence of click consonants in some languages in Kenya and Ethiopia (Cavalli-Sforza *et al.*, 1994). It is resuming to note that the patterns of linguistics variation in space parallel those of genetic or geographic variation, there are in fact good reasons why cultural and genetic pools have close similarities; both genetic and cultural contacts take place by the same routes; they respond to the same geographic and ecological barriers; and they also can influence each other, in the sense of mutual reinforcement (Cavalli-Sforza and Wang, 1986).

1.4 Y-chromosome

The human Y-chromosome has important properties for evolutionary studies. It is inherited paternally with no recombination. The Y-chromosome DNA sequence preserves a unique record of mutational events that occurred in ancestral generations. Studies of polymorphisms in the non-recombining portion of the Y-chromosome (NRY) have been proposed for detecting male-mediated migration events and for reconstructing paternal history (Casanova *et al.*, 1985; Ngo *et al.*, 1986; Malaspina *et al.*, 1990).

These features make its DNA sequence variation an invaluable tool for the study of modern human evolution. Haploidy and patrilinearity translate into increased levels of population subdivision compared with the autosomes, and the lack of recombination permits the reconstruction of an unequivocal haplotype phylogeny, which can be related to the geographic distribution of the Y haplotypes (Avise *et al.*, 1987; Underhill *et al.*, 2001). The Y-chromosome Consortium YCC (2002) has reported a large number of Y-chromosome biallelic polymorphisms, which has provided a detailed phylogeographic portrait of contemporary global population structure and past population movements and

interactions. The availability of these highly geographically structured sets of markers has stimulated the analysis of more restricted areas and the Y-chromosome hierarchical nomenclature is generated to supersede and unify past nomenclatures (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; Karafet *et al.*, 2001) to allow the inclusion of additional mutations and haplogroups yet to be discovered.

1.4.1 Y-chromosome Haplogroups

The results of a nested cladistic analysis indicated that these geographical associations arose through a combination of processes, including restricted, recurrent gene flow (isolation by distance) and range expansions. One of the oldest events in the nested cladistic analysis was a range expansion out of Africa which resulted in the complete replacement of Y-chromosomes throughout the Old World, a finding consistent with many versions of the Out of Africa Replacement Model. A second and more recent range expansion brought Asian Y-chromosomes back to Africa without replacing the indigenous African male gene pool. Thus, the previously observed high levels of Y-chromosomal genetic diversity in Africa may be due in part to bidirectional population movements (Hammer *et al.*, 1998).

The Y-chromosome tree (Y-chromosome Consortium, 2002) consists of 18 different major haplogroups A-R (Fig. 1.1). Haplogroups A and B are entirely restricted to African populations; while the rest of the haplogroups are present outside of Africa. This pattern supports the hypothesis of an African origin of human NRY diversity (Hammer *et al.*, 1998; Underhill *et al.*, 2000). Haplogroup A is closest to the root of the tree and is found most frequently in the Khoisan and Nilotics. Haplogroup B

chromosomes are most frequently observed among Pygmies in Central Africa, with B2a* and B2b* being nearly exclusive to this group. Haplogroup E is overwhelmingly the most common haplogroup in Africa (Wood *et al.*, 2005). Additional region-specific major haplogroups include Asia (D, H, and P) and Oceania. Paragroup F and haplogroups L and N are shared primarily between Asia, Europe and the Middle East, whereas haplogroup O is shared by Asia and Oceania. Haplogroups G and R are widely shared outside of Africa. Haplogroup J is primarily distributed throughout Europe, Middle East, Asia and Africa; whereas haplogroup I is most often found in Europe, Middle East and paragroup K occurs predominantly within Oceania. The number of haplogroups by region ranges from 8 in Africa and the Americas to 15 in Asia. Three of the regions are characterized by predominant haplogroups: E in Africa (60%), R in Europe, Middle East region (45%) and Q in the Americas (80%). Both the large number of Asian haplogroups and their comparatively even distribution underscore the centrality of Asia for human dispersals (Hammer *et al.*, 2001).

1.4.2 Y Alu Polymorphism (YAP)

The Y-chromosome Alu polymorphism (YAP), first described by Hammer (1994), is a unique event polymorphism (UEP) marker defining a deep-rooting clade of the Y-chromosome genealogical tree that originated some 50,000-60,000 years ago (Hammer *et al.*, 1998; Karafet *et al.*, 1999) although the origin of YAP is still controversial. Initial surveys showed that the YAP clade could be split into two main haplogroups worldwide, haplogroup E and haplogroup D, the first one is confined to sub-Saharan Africa, middle

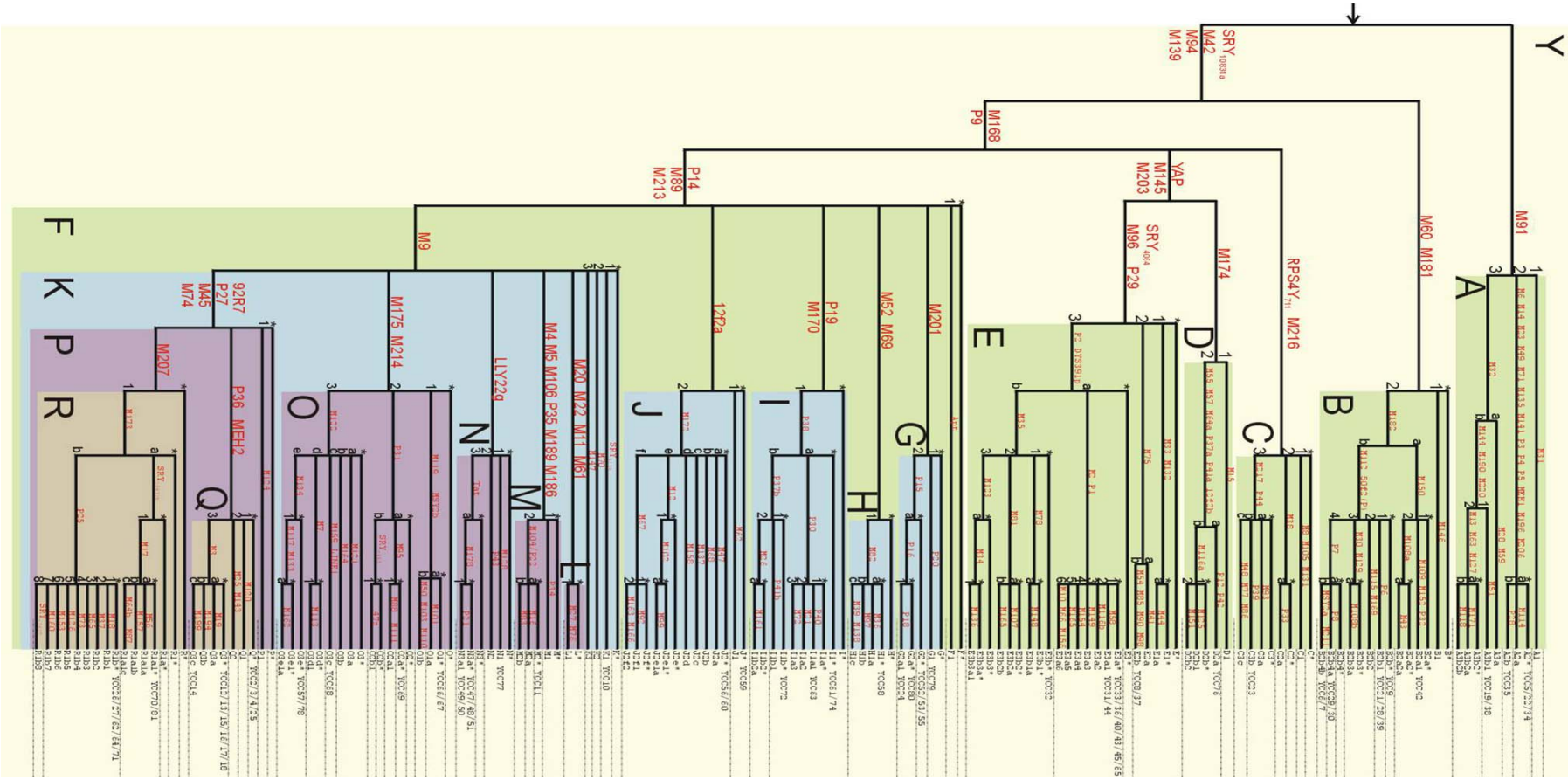


Fig. 1.1. Y-chromosome Phylogenetic Tree. The arrow shows the root of the tree which is the most recent common ancestor (MRCA). Source: Y Chromosome Consortium (2002).

eastern and southern European populations, while the second one is specific to Japan and other south/east Asian populations (Weale *et al.*, 2003).

The frequency of Y-chromosomes carrying YAP insertion is the highest among sub-Saharan African populations (82–95%), followed by North African populations (50–70%), and is rare among Europeans 1–10% (Hammer *et al.*, 1998; Underhill *et al.*, 2000; Jobling and Tyler-Smith, 2003). Most Asian populations examined so far lack this element. An exception to this pattern is the presence of YAP insertion in the Middle East and central Asian populations 12–27% (Deka *et al.*, 1996; Al-Zahery *et al.*, 2003). YAP element is present among Japanese populations in relatively higher frequency 27–85.7% (Tajima *et al.*, 2002). Interestingly, YAP individuals in these two diverse geographical areas of Africa and Japan are found to have different lineages (Agrawal *et al.*, 2005).

1.5 Mitochondrial DNA (mtDNA)

Because mitochondrial genome is inherited maternally, it offers a very different perspective of human evolution. Variation in mtDNA in fact provides a reliable record of the maternal lineage of human species. In contrast to autosomal DNA, the non-coding mitochondrial D-loop control region or hypervariable region (HVR) is especially rich in polymorphisms, and the mutation rate is estimated to be approximately 2 to 3×10^{-7} per nucleotide per generation (Horai *et al.*, 1995). The distribution of the major indigenous African mtDNA patterns across the different regions of the continent and their divergence times are well discussed in the literature (Salas *et al.*, 2002, 2004; Gonder *et al.*, 2007; Tishkoff *et al.*, 2007).

1.5.1 mtDNA Heteroplasmy

Heteroplasmy is the presence of two or more populations of mtDNA in a single individual, and is expected to be more widespread than has been reported in the control

region (Gill *et al.*, 1994; Bendall *et al.*, 1996). Knowledge on the frequencies of new mutations in maternal lineages is of central importance for the application of mtDNA in forensic identity investigations (Stoneking *et al.*, 1991; Piercy *et al.*, 1993). Furthermore, heteroplasmy has been described in association with several diseases (Wallace, 1992; Marchington *et al.*, 1996). The nucleotide position could be considered as heteroplasmic if a secondary peak of more than about 40% is present, which could be confirmed in the reverse sequencing reaction (Hühne *et al.*, 1998).

Heteroplasmy in the first and the second hypervariable segments of the mitochondrial D-loop seems to increases with age (Calloway *et al.*, 2000). In addition the homopolymeric cytosine tracts (poly C tracts) in both hypervariable regions (HVR-I and HVR-II) which represent hot spots for mutations and replication slippage, is suspected to be the cause of this length polymorphism (Hauswirth *et al.*, 1984; Hauswirth and Clayton, 1985). In the second hypervariable region (HVR-II) length heteroplasmy is due to cytosine insertions between the positions 303 and 309 of the reference sequence (Anderson *et al.*, 1981). These heteroplasmic mixtures in the HVR-II cytosine stretch usually have an identifiable prominent type (Parson *et al.*, 1998).

1.5.2 Mitochondrial DNA Haplogroups

The differences in mtDNA haplogroup distribution were attributed to founder effects, specifically the colonizing of new geographic regions by only a few immigrants that contributed a limited number of mtDNAs. However, this model is difficult to reconcile with the fact that northeastern Africa harbors all of the African specific mtDNA lineages as well as the progenitors of the Eurasian radiation (Mishmar *et al.*, 2003). Macrohaplogroup L (Fig. 1.2) is geographically limited to sub-Saharan Africa and has been divided into haplogroups L0-L6 (Mishmar *et al.*, 2003; Salas *et al.*, 2004; Kivisild *et*

al., 2006). The phylogeny of macrhaplogroup L is largely based on D-loop sequence and RFLP analysis. In particular, African mtDNAs that belong to L0 and L1 fall into several distinctive sub-haplogroups, but their history is complex and poorly understood (Pereira *et al.*, 2001; Kivisild *et al.*, 2006).

The haplogroup L0 include the most recent common ancestor (MRCA) of the human mtDNA, which is at least 150,000-170,000 years old (Horai *et al.*, 1995). Haplogroup L0 is divided into sub-haplogroups L0a, L0d, L0f and L0k (Salas *et al.*, 2002; Mishmar *et al.*, 2003; Salas *et al.*, 2004; Kivisild *et al.*, 2006). Haplogroup L0a probably originated in eastern Africa and is common in eastern, central and southeastern Africa, but is almost absent in northern, western and southern Africa (Salas *et al.*, 2002). Haplogroup L0d is found almost exclusively among Khoisan people (Vigilant *et al.*, 1991; Bandelt and Forster, 1997; Behar *et al.*, 2008), Turkana of Kenya (Watson *et al.,* 1997), Mozambicans (Pereira *et al.*, 2001; Salas *et al.,* 2002) and Sandawe and Burunge of Tanzania (Tishkoff *et al.,* 2007). Similarly, haplogroup L0k is found exclusively among the Khoisan (Salas *et al.*, 2002). However, the phylogenetic relationship of haplogroups L0d and L0k is uncertain (Watson *et al.*, 1997; Kivisild *et al.*, 2006). Haplogroup L0f is rare and appears to be geographically confined to East Africa (Salas *et al.*, 2002). Haplogroup L1 is composed of L1b and L1c, L1b is concentrated in western Africa but it also occurs in central and northern Africa (Watson *et al.*, 1997; Rosa *et al.*, 2004). Haplogroup L1c occurs frequently among central African Bantu populations (Vigilant *et al.*, 1991; Destro-Bisol *et al.*, 2004a), and probably originated in the Atlantic coast in western equatorial Africa (Salas *et al.*, 2004). Both L1b and L1c are nearly absent in eastern and southern Africa. Haplogroup L5 previously referred to as L1e

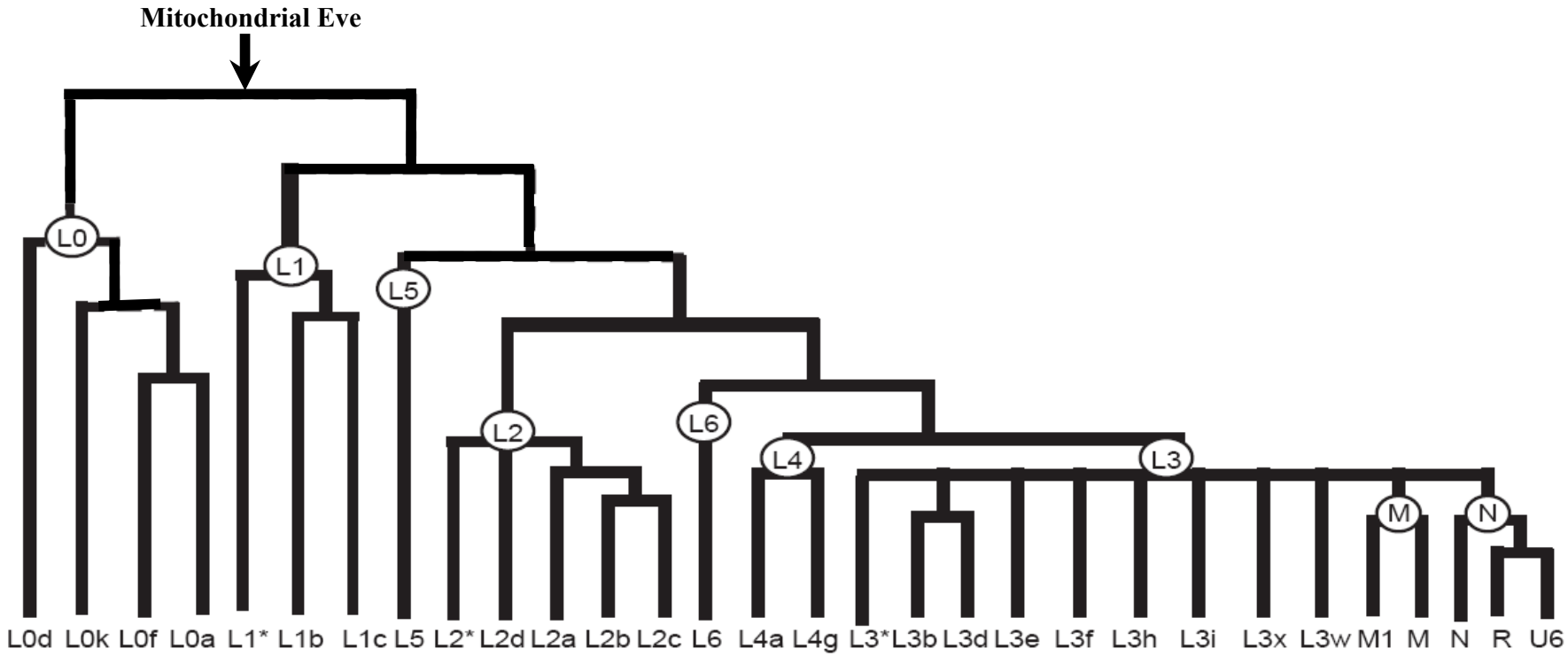


Fig. 1.2. Mitochondrial DNA phylogenetic tree, sub-clades of M and N haplogropus were not shown. Arrow indicates the root of the tree. Source: Gonder *et al.* (2006) with minor modifications.

(Kivisild *et al.*, 2004) has been observed at low frequency only in eastern Africa, Egypt and among the Mbuti Pygmies (Salas *et al.*, 2002; Stevanovitch *et al.*, 2004; Kivisild *et al.*, 2006).

Haplogroup L2 is generally subdivided into five subclades, the main subclade L2* consist of four subclades L2a, L2b, L2c and L2d (Chen *et al.*, 2000; Pereira *et al.*, 2001; Torroni *et al.*, 2001). According to Torroni *et al.* (2001), L2* mtDNAs are not present among southern Africans. The great majority belong to L2a which is the most frequent and widespread mtDNA cluster in sub-Saharan African (nearly a quarter of all indigenous types) as well as in African Americans at ~ 19%. The wide distribution of L2a in Africa makes identifying geographical origins of lineages difficult (Salas *et al.*, 2004). Both L2a1a and L2a1b subclades are well presented in southeastern Africans, and they appear to have an origin in West Africa which has been indicated by the distribution of matching or neighboring types, and to have undergone dramatic expansion either in southeastern Africa or in a population ancestral to present-day southeastern African which may suggest a signature for the Bantu expansions (Pereira *et al.*, 2001).

It appears that the founder ages for L2a are significantly older than for L1a, consistent with the phylogeographical picture, with an earlier West African origin for the lineages of southeastern Africa and a more recent East Africa origin for the L1a lineages (Phillipson, 1993).

The other subclades of L2 (L2b, L2c, and L2d) have a clear western distribution in Africa. Complete sequence data indicates that subclade L2d is the oldest of the four subclades of haplogroup L2 (Torroni *et al.*, 2001). A possible solution would be an origin for L2a somewhere between East and West Africa, followed by dispersals in both

directions along the Africa's Sahel corridor. Therefore, an origin for all three types in West and west/central Africa seems likely. The estimated divergence times range from ~120,000 years for L2d, through 55,000 years, for L2a, and ~30,000 years for L2b and L2c, with an estimated overall age for L2* of ~70,000 years. In the light of this, it is scarcely surprising that tracing its place of origin is problematic. At such an age, it seems perhaps unlikely that subclade L2d should have diverged in West Africa, but, given the period of potential drift and extinction, the data are certainly consistent with a central African origin (Salas *et al.*, 2002).

All non-Africans today are descended from macrohaplogroup L3, which gave rise to two (haplogroups M and N) founder types outside Africa. This macrohaplogroup consists of sub-haplogroups L3b, L3d, L3e, L3f, L3h, L3i, L3w, L3x, M and N. Haplogroups L3b and L3e lineages are widely spread in West African populations and Bantu-speaking southeastern Africans (Salas *et al.*, 2002; Rosa *et al.*, 2004), and haplogroup L3d, with a predominantly West African distribution (Salas *et al.*, 2002). The geographic distribution of L3f, L3h, L3i, and L3x lineages in East Africa, and Yemen is more consistent with a recent gene flow from Ethiopia or the Nile Valley than from southeastern Africa (Kivisild *et al.*, 2004). Haplogroup L3w occurs in East and north/eastern Africa at low frequencies (Watson *et al.*, 1997; Krings *et al.*, 1999b).

An ancient presence of the current haplogroup M types in Africa is debatable because they are similar to an Indian M type (Maca-Meyer *et al.*, 2001) and because haplogroup M in Africa is mostly restricted to Afro-Asiatic speaking groups, which may indicate a reverse migration in the past 20,000 years (Foster, 2004). Haplogroup N consists of haplogroups A, H, I, J, T, U, V, W, X and R. The R descendant haplogroup

consists of haplogroups B, U, F, pre-HV, H, V, J and T. Haplogroup M is divided into C, Z, D, Q, G, and E (Bermisheva *et al.*, 2002). In Asia, haplogroups N and M contributed equally to mtDNA radiation. Among Europeans, haplogroup N make up more than 98% of the mtDNAs. In Siberia, only six mtDNA haplogroups A, C, D, G, Z, and Y, which belong to haplogroup M make up more than 75% of the mtDNAs. In Native American populations, only five Old World mtDNA haplogroups A, B, C, D, and X encompass 100% of the mtDNA variation (Wallace *et al.*, 1999). Haplogroup L4 previously named L3g (Torroni *et al.*, 2006) reveals high haplotype and sequence diversity in Ethiopians. The coalescent calculations suggest that L4 lineages diversified from their founder 68,800 ±18,300 years ago (Kivisild *et al.*, 2004).

Haplogroup L6 a novel sister clade of L2, L3 and L4, defined by six coding transitions and one control region transition, was found in Yemeni and Ethiopians (Kivisild *et al.*, 2004). An East African origin of haplogroup L6 seems most likely, given its presence in Ethiopians and the fact that its sister haplogroups L2, L3, and L4 are all diverse and frequent there. Given the lack of an exact match from the African database for the southern Arabian L6 samples and the relatively deep time depth of its variation in Ethiopians and Yemenis taken together, 36,600 ± 23,400 years it is possible that this haplogroup has been preserved in isolation in the Ethiopian Highlands and southern Arabia for tens of thousands of years (Kivisild *et al.*, 2004).

1.6 Correlation between Genetics and Linguistics

Several studies indicate strong correlations between genetic and linguistic relationships among globally distributed human populations (Cavalli-Sforza *et al.*, 1988; Chen *et al.*, 1995). At the subcontinental scale, correlations between genetic variation and linguistic or geographic variation differ substantially. Y-chromosome studies have shown

that geographic distances correlate with genetic affinities among populations in Europe (Rosser et al., 2000), the Americas (Zegura *et al.*, 2004) and Austronesia (Hurles *et al.*, 2002) whereas language better explains Y-chromosome relationships in Siberia (Karafet *et al.*, 2002). Mitochondrial DNA (mtDNA) studies suggest that linguistic relationships are better correlated with genetic affinities among South American populations (Fagundes *et al.*, 2002) while both geography and language are correlated with maternal variation in Austronesia (Lum *et al.*, 1998).

In Africa the Y-chromosome, reveal a strong partial correlation between genetic and linguistic distances and no correlation between genetic and geographic distances. In contrast, mtDNA variation is weakly correlated with both language and geography. These data suggest that patterns of differentiation and gene flow in Africa have differed for males and females in the recent evolutionary past (Wood *et al.*, 2005).

1.7 Ancient DNA (aDNA)

Physical anthropologists had long used molecular characters of modern populations to elucidate human variability and human prehistory (Wilson and Sarich, 1969). Ancient DNA (aDNA) refers to DNA recovered from clinical, museum, archaeological, and paleontological specimens that ranges in age from less than 100 years to tens of millions of years. aDNA could be extracted and characterized and was first demonstrated in nonhuman material by Higuchi and colleagues (Higuchi *et al.*, 1984), a year later Pääbo *et al.* (1985) obtained DNA sequence data from a 2,400 years old Egyptian mummy. This results was surprising due to the large fragment sequenced >3kb (O'Rourke *et al.*, 2000). Ancient DNA techniques have also been used for different applications including; genetic sexing, population continuity and replacement, phylogenetic reconstruction, paleopathology and paleoparasitology (Kaestle and Horsburgh, 2002). For example

ancient DNA analysis was used to investigate mtDNA variation among Neanderthals to address the questions of whether they were similar to humans in having much variation and to obtain an indication of whether an expansion took place (Krings *et al.*, 1997, 1999a). Traditional paleopathological methods for inferring the presence of disease based on diagnostic skeletal lesions. While genomes of pathogens are present in multiple copies in affected individuals, and this has proven beneficial to paleopathologic studies, as many disorders have no skeletal traces. Accessing genomes of infectious agents avoids the difficulties of traditional methods (O'Rourke *et al.*, 2000). The molecular identification of bacterial DNA, such as *Mycobacterium tuberculosis* has been isolated and identified from ancient bones and mummified tissues using PCR techniques (Spigelman and Lemma, 1993).

1.7.1 Sources of aDNA

Bone is generally considered an optimal aDNA source due to the binding of DNA to hydroxyapatite (the binding between the phosphate group of the DNA and calcium), which make the aDNA preserved from longer time compared to soft tissue (Tuross, 1994). Skeletal tissue can be sampled in several ways, small fragmentary pieces can be used, or long bones can be sampled by drilling. Skeletal elements without lesions should be chosen because lesions provide an avenue for contamination. Using teeth as a source of aDNA has the advantage of multiple, independent samples per individual. Teeth without caries should be chosen because dental caries allow contaminating DNA to enter the pulp cavity. Using undegraded teeth further reduces this risk (O'Rourke *et al.*, 2000).

The second source that could be used as a source for aDNA is soft tissue, in this case subsurface tissue should be selected whenever possible to reduce contamination from handling. Desiccation may protect DNA from hydrolytic damage, although it is still

susceptible to oxidative damage (Pääbo, 1989). Also hairs have been used to obtain DNA in forensic cases (Wilson *et al.*, 1995).

1.7.2 Ancient DNA Degradation

DNA normally becomes degraded after death by endogenous nucleases. Under fortunate circumstances, such as rapid desiccation, low temperatures or high salt concentrations, nucleases can themselves become destroyed or inactivated before all nucleic acids are reduced to mononucleotides. If this is the case, slower but still relentless processes start affecting the DNA (Hofreiter *et al.*, 2001). Oxidation, as well as the direct and indirect effects of background radiation, will modify the nitrous bases and the sugar-phosphate backbone of the DNA (Fig. 1.3). Furthermore, deamination, depurination and other hydrolytic processes will lead to destabilization and breaks in DNA molecules. All these processes create problems for the retrieval of ancient DNA sequences. For example, a high proportion of cytosine and thymine residues in extracts of ancient tissues are modified to oxidation products of the pyrimidine called Hydantoins, which block DNA polymerases and thus the PCR (Höss *et al.*, 1996). Furthermore, deamination products of cytosine, for example, are common in ancient DNA and cause incorrect bases to be inserted during the PCR (Pääbo, 1989). After a long enough time, the cumulative effects of damage to the DNA will become so extensive that no useful molecules remain. Assuming physiological salt concentrations, neutral pH and a temperature of 15°C, it would take about 100,000 years for hydrolytic damage to destroy all DNA that could reasonably be retrieved (Pääbo and Wilson, 1991; Lindahl, 1993). Some environmental conditions, such as lower temperatures will extend this time limit (Smith *et al.*, 2001), whereas other conditions will reduce it, and amplification of aDNA, when possible, is usually limited to fragments less than 300–500 bp in length (O'Rourke *et al.*, 2000).

Bone is generally considered an optimal aDNA source because the binding between DNA phosphate group and calcium which is called hydroxyapatite, slows DNA degradation. Experimental results support that DNA yields from bone exceed those from soft tissue (Tuross, 1994). Environmental factors such as temperature (4°–37°C), humidity (20%–98%), pH (3.0–10.0), do not significantly affect DNA yields from teeth samples (Schwartz *et al.*, 1991). However, to consider amplification of DNA molecules older than one million years of age is overly optimistic (Hofreiter *et al.*, 2001).

1.7.3 Authenticity of Ancient DNA

Due to the sensitivity of the PCR and the degraded nature of DNA in ancient samples, the contamination of samples and laboratory preparations by exogenous DNA is a constant concern. Such contamination can derive from a variety of sources (Kaestle and Horsburgh, 2002). Further more, X-raying bones can increase the fragmentation of the endogenous DNA (Gotherstrom *et al.*, 1995). Several decontamination procedures have been described to decontaminate the samples before extraction (Handt *et al.*, 1994; Richard *et al.*, 1995; Kaestle and Horsburgh, 2002). Authenticity of aDNA is therefore of paramount concern, and efforts to assure that research results reflect endogenous target sequences rather than modern contaminations have received considerable attention (Richards *et al.*, 1995; Handt *et al.*, 1996; Hirano *et al.*, 2002).

1.7.4 aDNA and Sex Determination

The most common method of genetic sexing takes advantage of differences in the Amelogenin gene, which is present on both X and Y chromosomes, but with slightly varying sequences. The favored protocol involves amplifying a specific segment of the

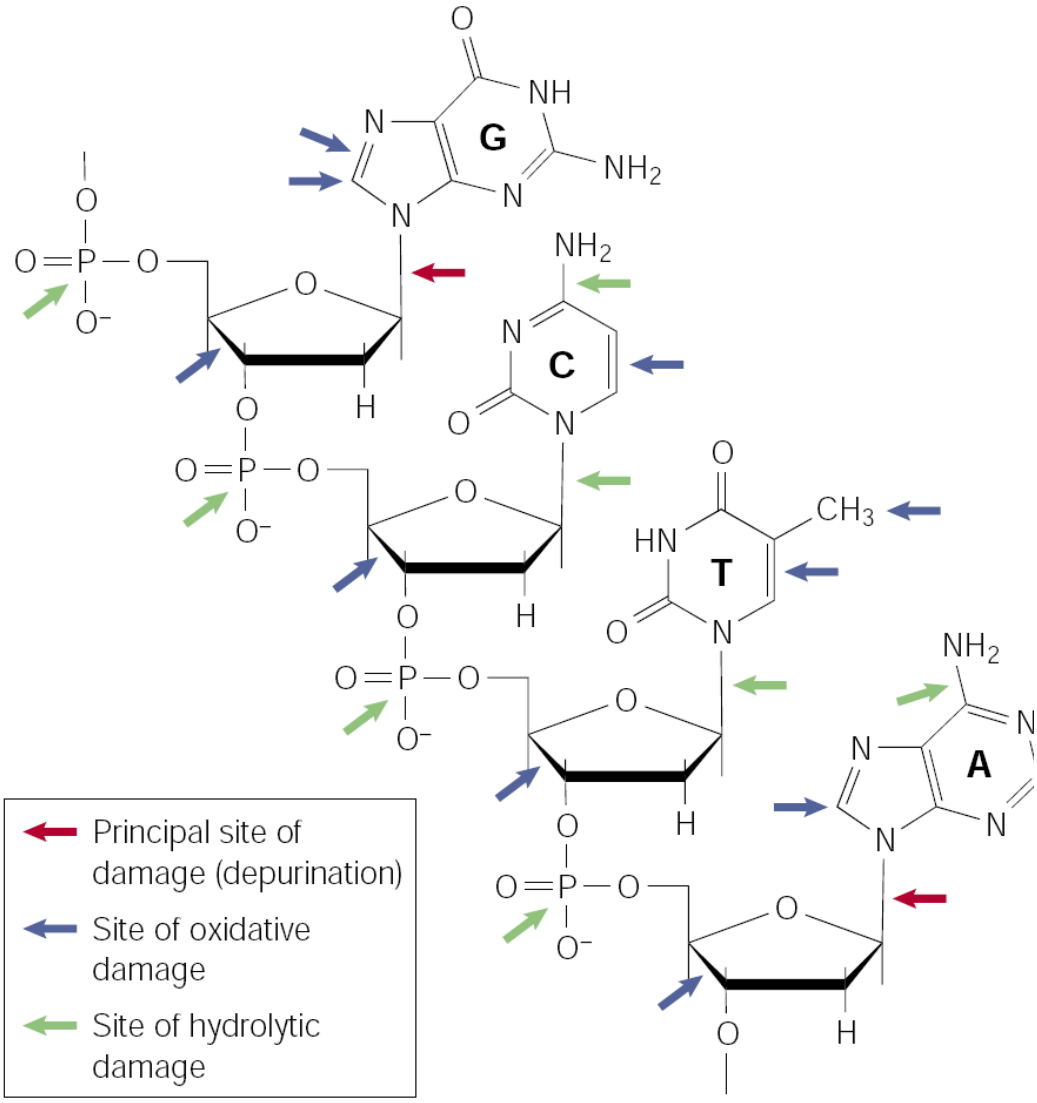


Fig. 1.3. Ancient DNA damage. Principal sites of damage are indicated by red arrows. Sites susceptible to hydrolytic attack are indicated by green arrows and those prone to oxidative damage by blue arrows (Hofreiter *et al.*, 2001).

Amelogenin gene that contains a 6 bp deletion in the copy on the X-chromosome, when compared with Y-chromosome. Thus, the DNA fragment amplified from a Y-chromosome is 218 bp long; while that from an X is only 212 bp. Amplifications from a male individual will therefore contain DNA fragments of two sizes, while those from a female individual will contain DNA fragments of only one size (Sullivan *et al.*, 1993).

1.8 Methods of Analysis

Many statistical analyses can be performed to compare populations with one another, or to derive estimates of important population genetic variables from observed data. Most of these methods rely on estimating total variation within a population sample, and comparisons of variation within and between samples.

1.8.1 Genetic Distances

The most common statistics used for these studies is the F_{ST} or Wright's fixation index (Wright, 1951) which is an estimate of differentiation between populations. F_{ST} calculated from genetic data such as allelic or haplotype data, or from sequence data (Nei, 1975; Jorde, 1980). As two populations become more differentiated from each other, F_{ST} increases. Population differentiation is influenced by the elapsed time since common ancestry and the level of migration between populations, with higher levels of migration leading to reduced differentiation through the action of gene flow (Cockerham and Weir, 1993). Genetic distance could be affected by effective population size *Ne,* gene flow, genetic drift, geographic distance and migration rate. Thus, more complicated analytical and simulation models have been developed in an attempt to estimate F_{ST} (Slatkin, 1985; Hunley, 2002). In addition, most population genetics methods assume that a sample has been drawn from a single or limited number of generations in a population. This is

generally not the case when dealing with ancient samples, which may span hundreds or thousands of years (Kaestle and Horsburgh, 2002).

1.8.2 Neutrality Test and Diversity Measures

The test statistics Tajima's *D* (Tajima, 1989) is one of the most popular tests employed to detect departures from the null model of selection or neutrality. The Tajima's *D* test measures the allele frequency distribution of nucleotide sequence data. This statistic can be influenced by both population history and natural selection. *D* is defined as:

$$D = \frac{\hat{\theta}_{\pi} - \hat{\theta}_{S}}{\sqrt{Var(\hat{\theta}_{\pi} - \hat{\theta}_{S})}}$$

Where $\theta = 2Ne\mu$, Ne is effective population size and μ is mutation rate. θ indices are providing an alternative picture of past evolutionary processes. $\theta\pi$ is heavily influenced by deep genealogical structures, and the product of ancient demographic events (Rogers and Harpending, 1992). θs is primarily influenced by lineage sorting in the external branches of genealogies and reflects recent demographic history (Helgason *et al.*, 2000). The θs estimator is based on the number of observed segregating sites (Watterson, 1975), whereas $\theta\pi$ is equivalent to the mean number of pairwise differences between sequences (Tajima, 1983). If, according to coalescence theory, population growth produces genealogies that are star-like, then an excess of low frequency mutations or singletons defined by external network branches will be observed, inflating θs relative to $\theta\pi$ because of the smaller influence of low frequency variants on the value of π. Hence, a population that has undergone a size expansion should produce negative *D* values.

Conversely, population bottlenecks tend to create gene genealogies that are highly fragmented with deep internal branches that are mostly comprised of high and intermediate frequency variants, which will increase $\theta\pi$ relative to θs and thus resulting in positive *D* values. However, significant *D* scores can also be the product of selection and mutation rate heterogeneity (Tajima, 1989; Aris-Brosou and Excoffier, 1996).

The difference between these two estimators of θ is scaled by the standard deviation of their difference. A positive value indicates an excess of intermediate frequency (polymorphic) alleles, while a negative value indicates an excess of rare alleles. The null hypothesis of the Tajima's *D* test is neutral evolution in an equilibrium population. This implies that no selection is acting at the locus and that the population has not experienced any recent growth or contraction (Tajima, 1989). The significance of the statistics could be tested by generating random samples under selective neutrality and population equilibrium using a coalescent simulation algorithm, with *p*-values representing the proportion of random statistics less or equal to the observation (Hudson, 1990).

1.8.3 Pairwise Differences

The pairwise difference (mismatch distribution) is the distribution of the observed number of differences between pairs of haplotypes. Pairwise mismatch distributions were routinely used to infer recent demographic population expansion, this distribution is usually multimodal in samples drawn from populations at demographic equilibrium, as it reflects the highly stochastic shape of gene trees, but it is usually unimodal (or smoothly decreasing) and has a single peak in populations having passed through a recent demographic expansion (Rogers and Harpending, 1992) or through a range expansion with high levels of migration between neighboring demes (Ray *et al.*, 2003).

1.8.4 Phylogenetic Trees

A phylogenetic tree is an evolutionary hypothesis about the proportional relatedness of individuals, populations or species (Hillis *et al.*, 1996). The central premise of phylogenetic reconstruction is that measures of similarity that in some way reflects common ancestry. There are currently a multitude of methods for estimating phylogenies, which essentially break down along two line when are come to the type of input data analyzed; character methods and distance methods. Raw sequence data are discrete, and they can be analyzed as such, using each nucleotide site during the analysis. Alternatively they can be converted to a distance matrix, in which sequence divergence between each pair of sequences is calculated. There are several different methods for calculating pairwise sequence divergences, each taking into account observed patterns of molecular evolution, distance methods nonetheless result in a significant loss of information (Hillis *et al.*, 1996).

The second dichotomy in phylogenetic analysis is that between clustering methods and search methods. Clustering methods use distance data and implement an algorithm which judges the best edge to which to join the next sequence. Clustering methods are fast and always produce a single tree (Page and Holmes, 1998). In contrast, search or optimality methods scan many trees, judging them against an objective criterion, such as parsimony or likelihood. There are several search methods, but they all tend to have the same faults. They are slow, requiring considerable computational power, and while still preferable to a clustering algorithm, they are only as good as the objective criterion chosen. It remains difficult to assess the accuracy of phylogenetic trees, that why it is sometimes essential to compare trees topology of various methods. Some methods are employed which are able to give a sense of the reliability of a tree. Bootstrapping

(Felsenstein, 1985) is the most popular of these methods. It is a method adopted from statistics to produce pseudo- replicates of the data, in which data columns are randomly selected with replacement to manufacture a varied data set form the information contained among the true data. The bootstrap pseudo-replicate data sets are analyzed to generate a phylogenetic tree for each of them, and then a consensus tree is calculated with each node, representing the proportion of pseudo-replicates that generated that split. It is important to appreciate that a bootstrap value for a particular node is no indication of the truth of that split, but simply reflects the likelihood that the split will be retained (Felsenstein, 1985).

1.8.4.1 Unweighted Pair Group Method Using Arithmetic Averages (UPGMA)

UPGMA method (Sneath and Sokal, 1973) assumes that the rate of nucleotide substitution is the same for all evolutionary lineages. An interesting aspect of this method is that it produces a tree that resembles a species tree, with the branch lengths for two Operational Taxonomic Units (OTUs) being the same after their separation. Because of the assumption of a constant rate of evolution, this method produces a rooted tree.

1.8.4.2 Minimum Evolution (ME)

The ME method uses distance measures that correct for multiple hits at the same sites; it chooses a topology showing the smallest value of the sum of all branches (*S*) as an estimate of the correct tree. However, construction of an ME tree is time-consuming because, in principle, the *S* values for all topologies must be evaluated. Because the number of possible topologies (unrooted trees) rapidly increases with the number of taxa, it becomes very difficult to examine all topologies (Rzhetsky and Nei, 1992).

1.8.4.3 Neighbor-Joining (NJ) Method

NJ method is a simplified version of the minimum evolution (ME) method. In the case of the NJ method, the *S* value is not computed for all or many topologies, but the examination of different topologies is embedded in the algorithm, so that only one final tree is produced. The algorithm of the NJ method is somewhat complicated (Nei and Kumar, 2000). The NJ method produces an unrooted tree because it does not require the assumption of a constant rate of evolution. Finding the root requires an outgroup taxon. In the absence of outgroup taxa, the root is sometimes given at the mid-point of the longest distance connecting two taxa in the tree, which is referred to as mid-point rooting (Saitou and Nei, 1987).

1.8.5 Networks

Another way to approach evolutionary relationships is a network, which allows reticulation or cycling. In many cases, this may be a more appropriate representation of our knowledge or of reality than a tree, which presumes dendritic evolution. Network allows illustrating these ambiguities in concentrations between nodes in a topology by connecting nodes through multiple pathways. The same type of data can be used to construct networks as are used to construct trees. Network-building algorithms generally begin with raw sequence or haplotype data, rather than distance data, and generate their own genetic distances from these data (Huson, 1998; Clement *et al.*, 2000). Because it allows for reticulation, the median-joining approach to the inference of haplotype relationships is appropriate for the analyses of human mtDNA control region sequences and Y chromosome short tandem repeat polymorphism haplotypes, which exhibit high levels of homoplasy (Posada and Crandall, 2001).

1.8.6 Principal Component Analysis (PCA)

Principal component analysis (Jollife, 1986) is almost as popular as trees for phylogenetic analysis. It permits to collect in a few simple graphs a large fraction of the information contained in all the genes tested. Principal component analysis (PCA) involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability (Cavalli-Sforza *et al.*, 1994).

1.8.7 Effective Population Size (*Ne*)

The effective size of populations *Ne*, is the size of a homogeneous population of equally reproductive individuals that would generate the same rate of genetic drift as is observed in the real population of total census size N. *Ne* is one of the most fundamental parameters determining how much genetic variation can be maintained in populations that are in mutation-drift equilibrium, and determine how quickly allele frequencies change with genetic drift (Kalinowski, 2002).

1.8.8 Human Migration Rate (Nm)

Migration may have a great role in shaping the genetic variation within populations and the differences among them; migration is synonymous with the gene flow which may change allele frequencies within a population if the immigrants differ genetically. Human migration rate is a measure of gene flow among populations (Maynard, 1989). The remarkable contrast between Y-chromosome and mtDNA, combined with the increasing availability of data on DNA variations, has triggered a number of studies addressing males and females migration history (Cavalli-Sforza and Feldman, 2003).

From a worldwide comparison, Seielstad *et al.* (1998) reported a large contrast in the geographic stratification between the two sex-specific markers. The analysis was based on both published (Underhill *et al.*, 1997) and original data of various genetic systems such as RFLP, SNPs and Y-STRs. The F_{ST}, was 0.65 for Y-chromosome and 0.19 for mtDNA loci. After discarding various alternative scenarios like polygyny, natural selection, and mutation rate, concluded that only a large sex differential in migration rate could account for the discrepancy. Seielstad *et al.* (1998) found the migration rate to be higher for women compared to that of men by a factor of eight in most human societies, throughout most of human history. The migrations referred to here are matrimonial migrations, by which individuals leave their community of origin to marry, settle, and reproduce in a new community. Others also showed that in the Hindu caste system, migration rates were substantially higher in maternally than in paternally inherited markers, with more upward movement between Hindu castes among women than among men (Bamshad *et al.*, 1998).

In order to explain the counterintuitive view that women have a higher migration rate than men, the widespread practice of patrilocality among human populations was called upon: in about 70% of documented societies, women typically move from their communities of origin to that of their husbands (Murdoch, 1967; Divale, 1984). The conclusions of Seielstad et al. (1998) have been challenged in the last few years (Wilder *et al.*, 2004; Shriver, 2005). First, polygyny was brought back into the picture as a possible explanation. Without directly addressing the notion of a much higher global migration rate for women, Dupanloup *et al.* (2003) suggested that a greater impact of genetic drift at Y-chromosome markers would be enough to explain apparent inconsistencies among mtDNA and Y-chromosome variations. The large-scale practice of

polygyny would have maintained the male effective population at a much smaller size than that of females during much of human history. Taking the issue further, Wade and Shuster (2004) showed how polygyny leads to a sex difference in the opportunity for selection and how this, through the mediating impact of varying genetic drift, results in different patterns of diversity in the two sex-specific genetic systems.

Introducing a new strategy that allowed for direct sequence comparisons between mtDNA and the Y-chromosome in 389 individuals from 10 populations, Wilder *et al.* (2004) found the overall between group variation F_{ST} to be 0.334 and 0.382 for the Y-chromosome and mtDNA, respectively. And using a high density of SNPs, at the global and continental scales, Y-chromosome variants are not more geographically stratified than mtDNA variants (Gagnon *et al*., 2006), which may indicate that globally the migration rate of males is similar to that of female although this is not the case at the continental level (Simoni *et al*., 2000; Fagundes *et al.*, 2002; Destro-Bisol *et al*., 2004b).

1.9 Prehistory of the Nile Valley

Archeology in the Sudan has so far been the archeology of the Nile Valley, with overwhelming emphasis on the historical periods. It has become increasingly apparent that the sequence of events in the Nile Valley cannot be fully understood, without considering neighboring regions (Abbas and Ali, 1981).

During past millennia the Nile Valley has been the corridor that connected the Mediterranean world with sub-Saharan Africa, allowing the movement of people from north to south and vice-versa. Following the progressive desertification of the Sahara from around 10,000 BC, the Nile has become the most important path in north-south migrations in Africa, as it represents a constant source of water. Therefore, the Nile Valley was one of the most heavily inhabited areas of antiquity. Consequently,

archaeologists have recovered a vast quantity of skeletal material from different sites and periods all along the Nile Valley (Fox, 1997).

Physical anthropologists have attempted to study the peopling of the valley, but conclusions have been sometimes contradictory. Traditionally, some morphological traits have been used to attribute the human samples from the Nile Valley to a Caucasian or sub-Saharan African stock, depending on the predominance of traits that presumably defines one or another ethnic group (Batrawi, 1945, 1946; Strouhal, 1969; Nielsen, 1970; Berry and Berry, 1972). Nasal width, facial prognathism and frontal prominence are the widely used traits. The southern populations of the valley have been regarded as being of mainly sub-Saharan type (thus displaying more sub-Saharan African traits, especially wide nasal apertures and large prognathism), and the northern as mainly Caucasian type. However, in Nubia (a historical region south of the 1st Cataract, and actually included in Egypt and Sudan), the area where the admixture between African and Caucasoid types is more apparent, even in the present-day populations, these ethnic attributions have sometimes become confusing and controversial (Fox, 1997).

Roughly, two different hypotheses try to explain the characteristics of the Nubian populations. Several authors (Carlson, 1976; Carlson and Van Gerven, 1979; Van Gerven, 1982; Brace *et al.*, 1993) described a remarkable degree of temporal biological continuity among Nubian groups. According to these authors, this suggests a long-term regional continuity, at least from the Neolithic. In contrast, other authors emphasize the cultural discontinuities and the morphological changes, supporting a multiple-migratory view of Nubian prehistory (Crichton, 1966).

The Paleolithic (200,000-30,000 BP) located west of the East Africa Rift Valley; where the first traces of modern human exists, probably do not extend back further than

190,000 years ago (Cann *et al.*, 1987). There were hand axes of flint from a slightly differentiated early Acheulean era. This was followed during a moister period about 70,000 years ago by the middle Paleolithic, giving rise to the culture known as the Levallois and Mousterian. Hunting and fishing were the primary sources of survival. The late Paleolithic comes to a close around 40,000 BP with characteristic types of stone blades and microliths. The cool and dry weather of that era concentrated human activities in the river valley, as is evident from the finds at sites along the Atbara and the Blue Nile, and junction of Blue Nile with White Nile (Wilding, 1997).

During the Mesolithic (30,000-10,000 BP), a fundamental shift in climate conditions to moist environment occurred in the eighth millennium BC, giving rise on both sides of the Nile Valley to a grassland. The Wadi Howar in northwest Sudan became a succession of rivers and lakes; that climate allowing human habitation in what are now desert zone, partially in stable settlements along river courses and partially in seasonally occupied areas. Fishing, hunting and gathering of plants and fruits were primary sources of survival (Wilding, 1997).

The Neolithic Revolution (8000-4500 BP) is the term given for the first agricultural revolution, describing the transition from hunting and gathering to agriculture, as first adopted by various independent prehistoric human societies, the Neolithic era in Sudan was located between the south Egypt and central Sudan around the Nile Valley. The second half of the sixth millennium BC is distinguished by an increasingly dry climate. The origins and groupings of the population are difficult to reconstruct, since well-preserved skeletal remains have been discovered (Abbas and Ali, 1982). Around the middle of the fifth millennium BC, Neolithic groups that practiced a predatory economy can be identified at the 2nd Cataract in Lower Nubia. A group of

Neolithic people probably migrated and displays similarities to early Khartoum (Wilding, 1997). In the A-Group era (3800-3100 BC) there was a distinctive combination of indigenous and imported Egyptian material that has indeed long defined the late Neolithic of Lower Nubia as a discrete regional culture (Gatto, 1997).

1.9.1 The Kerma Culture (2300-1550 BC)

The archaeology of the third and second millennia BC remains dominated by the development of the first larger-scale polities in the Middle Nile, centered on the Kerma-Dongola region and their relations with Egypt. By ~2500 BC a settlement and religious site at Kerma began to develop into a political center that over the next 1000 years came to dominate ~1000 km of the Nile Valley and its vicinity (Bonnet, 1990). Indigenous populations also reestablished themselves in Lower Nubia and known there as the C-Group, this group was located in southern Egypt, southward to the modern Sudanese boarder, in approximately the same region as the earlier A-Group. Cattle were extremely important to the C-Group people, as they are today to other African cultures. The kingdom of Kush, as the Egyptians knew it, centered on Kerma, and was able by the second millennium BC to create a significant threat to Pharaonic Egypt. The threat was removed only by the conquest of Kerma by the New Kingdom pharaohs of Egypt around 1450 BC (Edwards, 2007).

Kerma settlement in northern Sudan has left a rich burial and settlement record, focused on the remarkable settlement of ancient Kerma and its associated cemetery. The settlement, occupied for at least 1000 years and ultimately covering an area of ~20 hectares, became the first urban community of the Sudan. Centered on a great religious complex, it remains as yet the only such settlement so far discovered in the Middle Nile (Bonnet, 2004).

The history and archaeology of the northern Kushite populations (the C-Group) again provide a remarkable opportunity to study early colonial interactions in those areas of Nubia conquered by the Pharaonic Egyptian state (Adams, 1984; Smith, 1995). During later periods of weak Egyptian royal power, the kings of Kerma were in turn able to re-enter Lower Nubia, at times raiding into Egypt itself ~1575–1550 BC (Davies, 2003). Revitalized Egyptian royal and military power during the New Kingdom (1550–1070 BC) reoccupied Lower Nubia and destroyed Kerma itself, and for 400 years the Pharaohs of Egypt dominated Nubia (Smith, 1995).

1.9.2 Napata and Meroe Culture (1100 BC-300 AD)

The nature of social changes that followed the end of Egyptian rule in Nubia remains unclear, however, by the 8th Century BC, a new Kushite kingdom had appeared in Nubia, centered on the Napata region (Vincentelli, 2006).

The history of this revitalized Kushite kingdom (Welsby, 1996) has traditionally been divided into two main periods, the Napatan and Meroitic. A Kushite conquest of Egypt during the mid-eighth Century BC and their subsequent century-long rule there, known as the 25th Dynasty (Morkot, 2000), ensured some historical record of a kingdom whose origins remain largely invisible in the archaeological record. The Napatan kings adapted many Egyptian cultural practices, contemporary and ancient, which became important parts in creating a distinctive new Kushite culture. Most obvious among the practices was pyramid burial, which they transferred to their Sudanese homelands (Edwards, 2007).

Around 300 BC a shift in the royal cemeteries from the Napata region to Meroe defines a new and distinct Meroitic period, settlement sites are known from riverine areas from the Blue Nile south of Khartoum to northern Nubia as well as within the western

Butana (Edwards, 1998). These changes are seen in the influence of external contacts, in cultural and material links with Ptolemaic and then Roman Egypt, and in indigenous cultural traditions rooted in Sudanic Africa (Zabkar, 1975). The symbolic collection in Meroitic arts indicated the coexistence of Sudanic, Egyptian and Roman cultures during that period. In the only well-studied part of the Meroitic kingdom, its northern margins in Lower/Middle Nubia, settlement seems likely to have been very atypical with regard to the maintenance of communication along the Nile route across the Sahara to Egypt (Edwards, 1996).

The Kushite period also saw technological innovation, with evidence of significant Iron working at Meroe by the middle of the first millennium BC. Its control may have been a significant source of royal power, and indeed its spread into many areas of the Sudan may have been slow (Rehren, 2001). This period also saw the earliest written indigenous language in the Sudan, when the indigenous Meroitic language which is a north Eastern Sudanic language began to be written in its own unique script, in the 3rd Century BC (Rilly, 2007).

Explaining the disappearance of the Meroitic kingdom as a political unit some time around 300–350 AD remains problematic. Suggestions that Nubian invasions destroyed Meroe or that Axumite kings might have conquered Meroe have little evidence to support them. Instead, how political collapse may or may not be related to observed cultural changes in the late and post-Meroitic centuries is an area of increasing debate. Meroe and other urban settlements declined and disappeared as centers of population. None seems to have survived into the medieval period. The use of Meroitic writing also ended by the 5th Century. Although historical records are lacking the archaeological evidence suggests

that with the disappearance of a central authority and the unifying imperial culture it generated, new and more diverse regional cultures developed (Edwards, 2007).

During the third to sixth centuries AD, several rival groups occupied Nubia and there was dominance by the X-Group which is an interesting mixture of ancient Egyptian, Greek, and Nubian traditions.

1.9.3 The Medieval Period: Post-Meroitic and Christian period (550-1500 AD)

The new regional cultural forms in turn somewhat correspond with an emerging series of smaller political units, consolidated into three regional powers by the 6th Century AD, Nobadia in Lower Nubia, Makuria in the Dongola Reach, and Alodia in central Sudan, with Makuria later absorbing Nobadia. The Nobadians, on the Romano-Egyptian frontier, retained close contacts with late Roman Egypt during that period (Farid, 1963). This new political order was soon closely linked with conversion of all three kingdoms during the 6th Century to Christianity (Kirwan, 1987).

The arrival of Christianity also encouraged new links with the north and remains the defining cultural development of the medieval period within riverine Nubia, bringing it within the empire of eastern Christendom. Pastoral opportunities were also transformed with the spread of camel pastoralism through more dry areas, probably closely linked with new Arab populations entering the region over extended periods (Edwards, 2007).

In the Medieval period Greek was already in use as an official language, at least in the north, and Coptic also came to be quite widely used in the Church. Some time in the early medieval period, varieties of Nile Nubian languages also became the primary common languages of central riverine Sudan, by the late 8th Century; people also began to call themselves Nubians (Browne, 2002).

As well as the Nubian kingdoms of the riverine Sudan, other kingdoms were developing in the later medieval period in Darfur, as they were elsewhere in Sudanic Africa (O'Fahey and Spaulding, 1974). However, archaeological research has as yet been able to add little to historical descriptions from that region. Nonetheless, some claims of material evidence for links between the Nubian kingdoms and those of Darfur now seem unfounded (McGregor, 2001). The history of the eastern Sudan was rather different during that period. Along the Red Sea coast, several important ports developed, operating within Islamic trading networks linking Egypt and the Near East with Arabia and the Indian Ocean. The Nubian kings may have controlled some parts of the east, but most seem to have remained in the hands of the nomadic Beja (Kawatoko, 1993).

1.10 Arabic Influence in the Nile Valley

Contacts between the Nile Valley and Arabia were in existence long before Islam, and cultural influence of Arab in Nile Valley was the direct result of the penetration of large numbers of Arab males over long period of time. Traffic followed two main directions: the first was either from south-west Arabia across the Gate of Tears into Abyssinia and thence northward, or directly across the Red Sea; the second across the Sinai desert, through Nile Valley and into the Sudan; (Hassan, 1973). Nothing is mentioned of any population wandering southwards from Libya, Algeria, or Morocco into the western kingdoms and thence eastwards into the Sudan (MacMichael, 1967).

The Arab conquest of Egypt in the 7th Century A.D. opened a lengthy history of relations between the Arab and the inhabitants of the Sudan. Immediately prior to the Arab invasion of Egypt in 641, the northern region of the Sudan was occupied by a Hamitic-speaking people. The southern region was inhibited by dark-skinned peoples.

The Hamitic-speaking Beja populations, though perhaps not influenced to the same degree by the influx of dark-skinned populations, were nevertheless affected by slight Arab influences (Hassan, 1973). The Arab genealogies have always been purely patrilinear, and little account is taken of the maternal lines (MacMichael, 1967).

1.11 Peopling of the Present Sudan

The area known today as Sudan may have been the scene of pivotal human evolutionary events, both as a corridor for ancient and modern migrations, as well as the venue of crucial past cultural evolution. Such legacy is attested by a population that encompasses some 90 discrete ethnic groups with around 140 spoken languages, of those, 134 are living languages and 8 are extinct (Gordon, 2005).

The languages spoken today in the Sudan belong to the three major African linguistic families: Nilo-Saharan, Afro-Asiatic, and Niger-Congo (Greenberg, 1963). Tantamount ethnic and cultural diversity exist, rendering the study of genetic diversity of human populations an interesting and appealing endeavor. Sudan is dominated by tropical climate and it can be divided into three different types of terrain: the desert in the north, which comprises approx. 30% of the country's total area; the semi-desert Sahel more to the West and East; and the Savannas in the South. Sudan is populated by different ethnic groups, the populations along the southern part of the Sudan are Nilotics (or Nilotes) and Nuba, Nilotics is a population of unusually large stature, live in the savannas of the Upper Nile region. They comprise the Dinka, the Nuer, the Shilluk, Anwak, Baria and a number of small ethnic groups. Dinka and Nuer are pastroralists and Shilluk are agriculturists, Nilotics speak languages within the Nilo-Saharan family (Bayoumi *et al.*, 1982). Nuba are settled in Nuba Moutains and speak languages beloging to Nilo-Saharan and Niger-Congo and vary considerably in their culture and social organization. The

westernmost part is inhabited mainly by different Nilo-Saharan groups, Arabs, Hausa and both are Afro-Asiatic, and in part Niger-Congo including Fulani and Zandi, The Nothern part is dominated mainly by Arabs, Nubians and to some extent Copts, all of them are agriculturists settled in the Nile Valley. Copts are known to be the most ancient population of Egypt, the history of Copts in Nubia began after the conversion of Nubian from polytheism to Christianity in the 6th Century (Sudan studies Association, 2006). After the 7th Century, both Arabs and Copts gained importance in the Nubian kingdoms, especially as a medium for commerce. Beja people live in the eastern desert of Sudan, Egypt across the Red Sea and possibly Eritrea. They today consist of four subgroups. These subgroups are relatively loosely integrated confederations of endogamous lineages based on assumptions of shared descent and cohabitation in an ancestral territory (Dahl and Hjort-af-Ornas, 2006). They speak languagues that belongs to the Afro-Asiatic family of languages. They have good contacts with Ethiopia (Passarino *et al.*, 1998). Both Nubians and Beja were influenced by Arabs as a direct result of the penetration of large numbers of Arabs into the Nile Valley over long period of time following the arrival of Islam around 651 A.D (MacMichael, 1967). The Central part of the Sudan is almost combination of different cultural and ethnic backgrounds.

1.12 Genetic Diversity among Sudanese Populations

Few studies have addressed the genetic variation among Sudanese populations. In one study 15 ancient samples from Meroitic period were screened for *Hpal* mitochondrial marker, four of them (26.7%) displayed the sub-Saharan African marker, due to the presence of *Hpal* at a frequency of 68.7% in sub-Saharan Africa, the results suggest the presence of sub-Saharan influence in Meroitic Nubian population, and that the morphological changes observed historically in the Nubian populations are more likely to

be due to the existence of south-north gene flow through the Nile Valley than to in-situ evolution (Fox, 1997). Mitochondrial DNA (mtDNA) analysis gave evidence that the Nile Valley has been a corridor for human migrations between Egypt and sub-Sahara Africa, and migrations had occurred bidirectional along the Nile Valley within the past few hundred to few thousand years and that the migration from north to south was either earlier or lesser in the extent of gene flow than the migration from south to north (Krings *et al.*, 1999b).

Bereir *et al.* (2003) highlighted the importance of inter-population stratification for polymorphic markers by typing four single nucleotide polymorphisms (SNPs) and a variable number of tandem repeats (VNTR) polymorphism located within disease associated/causing gene in four Sudanese populations of different ethnic affiliation including Nilotics, Fur, Masalit and Riverian. Population sub-structuring according to ethnic/linguistic group indicated departure from Hardy–Weinberg equilibrium (HWE). Differences in allele frequencies and genotype distribution between groups were also noted as a result of potential influence of evolutionary history and ethnic variation of loci, in the general distribution of SNPs and other polymorphisms.

The distribution of the sickle cell gene (S gene) among various ethnic and linguistic groups in the Sudan is well documented, Afro-Asiatic speaking groups including nomadic groups of Arab and non-Arab descent that migrated to the Sudan in various historical era showed a higher frequency of Hb^S compared to Nilo-Saharan speaking groups (Bayoumi *et al.*, 1985, 1987; Mohammed *et al.*, 2006), although recent study showed that Masalit which are Nilo-Saharan speaking group displayed high frequency of S gene (Salih *et al.*, 2006 unpublished data). A recent study based on analysis of Y-chromosome haplogroups and sickle cell genes among different ethnic groups across the

Africa's Sahel, suggests that the sickle gene, one of the major protective polymorphisms known in malaria, has in fact found its way from western Sahel only recently to the gene pool of the populations in eastern Sahel (Bereir *et al.*, 2007).

Genetic distance analysis among seven populations and based on the basis of six polymorphic loci (ABO, Rhesus blood groups, haemoglobin, G6PD, serum haptoglobin and transfenin polymorphisms) showed that Nuba and Nilotics have sub-Saharan African genetic characteristics; while Fur are intermediate between the Arabs and sub-Saharan populations. Beja and Gaalien have more pronounced Arab genetic characteristics than Hawazma and Meseria, who have a great deal of African population, these findings suggest the importance of language as influential tool in determining population genetic profile (Tay and Saha, 1988).

Studies on adult lactase phenotypes showed a higher prevalence of lactase persistence in Beja and Fulani compared to Nilotics and Nuba which have a higher proportion of lactose malabsorption, this indicate that life style plays a crucial role in shaping population genetic structure (Bayoumi *et al*., 1982).

A study based on the variation at 10 Y-STR loci analyzed among 506 males representing 49 populations, showed that Dinka from southern Sudan have a higher genetic diversity when compared with Beja from eastern Sudan, as there are several possible reasons for an excess of genetic diversity in sub-Saharan African. They are older and have been accumulating genetic variation for a longer period of time, sub-Saharan African populations have maintained a higher long-term effective population size, and gene flow into Africa has been higher than into other continents. Also the results suggest that Beja are more related to Ethiopian and Asian populations rather than sub-Saharan African (Seielstad *et al.*, 1999).

1.13 Objective

The aim of this study is to draw a portrait of the genetic patterns of peopling of the Sudan through identifying and characterizing the skeletal remains found in archeological burial sites in Sudan through ancient DNA analysis, and investigating the genetic structure of the Sudanese populations, using Y-chromosome and mitochondrial DNA markers.

CHAPTER **2**

MATERIALS & METHODS

2.1 Ancient DNA

2.1.1 Samples Collection and Identification

A total of 76 bone samples from different skeletal elements representing different historical periods including Neolithic, Meroitic, post-Meroitic and Christian periods including soft and hard tissues from different anatomical parts of Nuri mummy [this mummy is a naturally mummified remain found in Nuri archeological site in northern Sudan in March 2002] were collected form Sudan National Museum in Khartoum, to isolate aDNA. Identification and dating was performed by physical anthropologists in Sudan National Museum. Also 18 bone samples of Meroitic period were collected from eastern Sudan from Dinder National Park (Fig. 2.1).

2.1.2 Samples Preparation

In order to prevent possible contamination all stages of the work were carried out under sterile conditions following Handt *et al.* (1994) criteria, using latex gloves, mouth masks and lab coats. All samples were exposed to UV (254nm) for 5 min according to Cone and Fairfax (1993), and the outer surface has been removed using sterile surgical blade. Small aliquots were pulled off from tissue samples with sterile forceps, sterile pestle and mortar were used for powdery material. Bones were powdered by drilling. DNA extraction has been done in laminar airflow cabinet (Biohazard®). Post-PCR activities were spatially separated in the lab. Negative controls were used routinely to monitor contamination. All samples have been replicated to confirm the initial results.

Fig. 2.1. A map of the Sudan showing some archeological sites where the human remains analyzed in this study have been collected. These remains are now stored in Sudan National Museum in Khartoum.

2.1.3 Preparation of Silica for DNA Extraction

Silica was prepared for ancient DNA extraction according to Boom *et al.* (1990). Six grams of silica oxide (Sigma) were added to 50ml dH_2O in a 50ml measuring cylinder. Cylinder was covered with nescofilm and inverted to suspend. Vortexed and allowed to settle for 24 hours at room temperature. Volume of 43ml of supernatant was removed from cylinder using 25ml pipette. Demineralized water was added to a volume of 50ml. Mix was resuspended by stirring with 25ml pipette and mixed by inversion, allowed to settle for 5 hours at room temperature. Volume of 44ml of supernatant was removed as before. Mix was shaked and stirred with pipette to resuspend the residual 6ml and transferred to 100ml glass bottle. pH was adjusted to 2.0 by adding 60µl HCl 32% (w/v) and autoclaved. Silica was aliquoted into clean sterile plastic tubes (1.5ml), and exposed to UV light for at least 30 min. Silica was stored at 4°C in refrigerator, and discarded after 6 months of storage.

2.1.4 aDNA Extraction

Silica-guanidine theocyanate (GnSCN) protocol was used as the analysis of aDNA is complicated due to the DNA degraded nature as well as the presence of PCR inhibitors in aDNA extracts, which limit the amount of amplifiable DNA available. This extraction method is based on the protocols as described by Boom *et al.* (1990) and modified by Donghue *et al.* (1998). The silica method extracts DNA in a high concentration of guanidinium thiocyanate (GuSCN). GuSCN has the ability to lyse proteins, and acts as a chaotropic agent facilitating the binding of DNA to silica particles. The advantage of this protocol is that it removes all co-extract PCR inhibitors; but silica itself is a strong PCR inhibitors, that why care has been taken to remove all silica during the washes (O'Rourke *et al.*, 2000).

For each sample, 100µl demineralization solution was dispensed into an eppendorf tube containing glass beads. An extraction negative tube was always included. Samples were added using a small blue loop (1µl), quickly vortexed, and then vortexed at medium speed for 50 seconds. Tubes were placed in a 56°C water bath and mixed in a bead beater twice daily for 48 hours or until specimens were dispersed completely. Tubes were mixed with a bead beater and 250µl lysis buffer (L6) was added to each tube. Tubes were vortexed and placed in a 37°C water bath for 2 hours. Tubes were vortexed, and then spun in microcentrifuge for 1 min at maximum speed. 25µl silica suspension was added to all tubes. Tubes were placed on a vortex Genie-2 Mixer, shaker setting No. 3, for 1 hour. Tubes were spun for 1 min and silica supernates were collected into clean sterile tubes to process separately. The silica was washed by adding 100µl washing buffer to each sample, vortex mixed, then spun for 1 min. Supernates were discarded to hycolin, and the wash was repeated when necessary until supernate was colorless, 200µl 70% ethanol (-20°C) was added, samples were vortex mixed then spun. Supernates were discarded and the ethanol wash was repeated once again, 200µl acetone (-20°C) was added, samples were vortex mixed then spun and supernates were discarded. Tubes were drained on a clean absorbent paper in the laminar hood having ensured all tubes were properly labeled before removing lids. Tubes were dried in a 56°C heating block for two hours with loose lids. DNA was eluted by adding 100µl molecular biology water to each tube, vortex mixed and placed in a 65°C water bath for 1 hour. Samples were spun for 3 min. Elutes were aliquoted into sterile 0.5ml tubes and stored at -20°C, and working DNA was stored in 4°C.

In some of the negative cases where DNA was absent from the silica pellet, silica supernates were processed for confirmation. Samples were processed up to the washing

buffer stage as described previously, 200µl Puregene Protein Precipitation Solution was added to each tube and vortex mixed vigorously for at least 20 seconds, then spun for 3 min. 600µl isopropanol (-20°C) per samples was dispensed into clean, sterile tubes. Supernates (to maximum volume 550µl) were added to isopropanol, mixed by inversion 50 times then spun for 3 min. Supernates were discarded and tubes were drained on clean, absorbent paper, having ensured tubes fully labeled before removing the lids, 600µl 70% ethanol (20°C) was added to each tube, vortex mixed and spun for 2 min. Supernates were discarded and tubes were drained on clean absorbent paper. Tubes were placed with loosened lids in 56°C heating block for 2 hours with loose lids. DNA was eluted by adding 100µl molecular biology water to each tube, vortex mixed and placed in a 65°C water bath for 1 hour. Samples were vortex mixed, quickly spun, aliquoted into sterile 0.5ml tubes and stored at -20°C.

2.1.5 aDNA Quantification

PicoGreen®dsDNA quantification regent (a molecular probe product) is an ultra-sensitive fluorescent nucleic acid stain for quantification of double-stranded DNA (dsDNA) in solution. This reagent allows quantification of small amount (e.g. 25pg/ml) of dsDNA using a fluorescent microplate reader and fluorescent excitation and emission wavelengths. This method also minimizes the fluorescence contribution of RNA and single-stranded DNA (ssDNA). The PicoGreen dsDNA quantification reagent is provided as a 1 ml concentrated dye solution in high-quality, anhydrous dimethyl sulfoxide C_2H_6OS (DMSO). A dilution series (800, 400, 200, 100, 50, 25, 12.5, and 6.25ng/ml) of standard 16,000 µg/ml DNA was prepared by dilution with TE buffer, 50µl of TE buffer as blank was added into well 1A and 1B of a costar 96 wells plate following by 50 µl from the prepared standard DNA in row A and B. 1:100 dilutions were prepared for all

DNA samples and 5µl were transferred to the plate from row C to H. PicoGreen working solutions A and B were prepared and covered with foil to protect from light as they are susceptible to photodegradation, 50µl of PicoGreen solution A was added to sample DNAs in row C to H. the plate was wrapped in foil immediately after the addition of PicoGreen and read using FluoStar® Optima. Cytocalculator program was used to get the final concentrations.

2.1.6 Sex Identification and Y-chromosome Biallelic Markers

PCR was performed for sex determination using Amelogenin sex markers, and Y-chromosome biallelic markers M13, M60, M89. Nested-PCR was used to genotype Y Alu polymorphism YAP (Table 2.1, and Appendix II). Bovine Serum Albumin (BSA) was added to overcome such inhibitory processes.

For sex identification Amelogenin marker was used according to Sullivan *et al.* (1993), Amelogenin marker is specific for a sequence of DNA found in both the X and Y-chromosomes. However the sequence captured by this marker is slightly shorter in the X-chromosome subsequently if the sample amplified using this marker was from a male subject the PCR is expected to amplify two nearly identical fragments of DNA with the exception of the variation of fragment size due to longer sequence on the Y-chromosome of male as result when running the PCR product on the gel. Amelogenin was amplified revealing a sex-specific length dimorphism. For the detection of the Y-chromosome, two polymorphic Y-specific and X-chromosome markers (Table 2.2) were used as following:

DNA	7µl
10x Buffer	2.5µl
Primer F (10uM)	1µl
Primer R (10uM)	1µl

dNTPs (2.5mM)	1µl
BSA (10mg/ml)	2.5µl
Taq polymerase	0.5µl
dH_2O	9.5µl
Total volume	25µl

Touchdown program has been used for all Y-chromosome markers. Touchdown composes of two phases, phase 1 consist of five min initial activation at 95ºC was followed by 15 cycles of 20 seconds denaturation at 94ºC, 45 seconds annealing at 63ºC (temperature of annealing decreased by 0.5ºC per cycle), and 1 min extension at 72ºC. Phase 2 consist of 35 cycles, 20 seconds denaturation at 94ºC, 45 seconds annealing at 56ºC, 1.30 min extension at 72ºC, and 5 min final extension at 72ºC.

2.1.7 Gel Electrophoresis

Electrophoresis was done for determination of the amplified DNA product size. Agarose gel was prepared by dissolving 1.5gms agarose in 75ml of 1x Tris Boric EDTA (TBE) to make 2% agarose gel according to the size of products of the same marker being tested. The mixture was boiled in a microwave oven, and then was cooled to 60ºC at room temperature. After that it was stained with 1.5µl of 10mg/ml ethedium bromide. The mixture was poured into a tray with a suitable sized combs and left for 45 min to set and solidify. PCR products were loaded after mixing with loading dye. The gel was run in 1x TBE buffer for one hour and then visualized by UV light, using SYNGENE® Chemi Genius gel documentation system.

Table 2.1. Oligonucleotides used to Generate the different Y-chromosome Haplogroups

PCR Conditions			Polymorphism Conditions		
Marker	Primer pairs (5'-3')	Primer References	Enzymes	Ancestral alleles (bp)	Derived alleles (bp)
YAP outer	F: tcacataatttcattttccc R: caagttagctgtccatactg	(Thangaraj *et al.*, 2003)	-	Outer primer for ancient DNA analysis	
YAP inner	F: caggggaagataaagaaata R: actgctaaaaggggatggat	(Hammer& Horai, 1995)	-	YAP- (150)	YAP+ (450)
M2	F1: tttcattgttaacaaaagtcca F2: tttcattgttaacaaaagtccg R: taatttgtgtaaatggacttga	Present Study	-	A (203)	G (203)
M9	F1: ggcctaagatggttgaatg F2: ggcctaagatggttgaatc R: tatgtaagacattgaacgtttg	Present Study	-	G (71)	C (71)
M11	F: ccctccctctctccttgtattctacc R: ttcatcacaaggagcataaacaa	(Qamar *et al.*, 2002)	MspI (NEB)	A (215)	G (193+22)
M13	F: tcctaacctggtggtctttc R: tgagccatgattttatccaac	(Underhill *et al.*, 1997)	MboI (NEB)	G (156+77)	C (233)
M23	F1: gtagaaacattctttctttctaga F2: gtagaaacattctttctttctagg R: ctcttttaaactctcaattcagaatatt	Present Study	-	A (327)	G (327)
M33	F1: agtttatctcataagttactagttaa F2: agtttatctcataagttactagttac R: atcataaacactttatttgcag	Present Study	-	A (328)	C (328)
M40	F1: cccttcgagaggtcaaggcg F2: cccttcgagaggtcaaggca R: ctgagttaagtgccctgcagtt	Present Study	-	G (440)	A (440)
M42	F: aaagcgagagattcaatccag R: ttttagcaagttaagtcaccagc	(Shen *et al.*, 2000)	AluI (NEB)	A (340)	T (295+45)
M51	F1: tctatctcctgaagcagagtagacacag F2: tctatctcctgaagcagagtagacacaa R: catatttctgtcttctagcccctgt	Present Study	-	G (339)	A (339)

Continued

M52	F: ttaatacctataagaatattgcctgcac R: tttcagactaaaatgatatagttttcc	Present Study	HpyCH41V (NEB)	A (164)	C (138+26)
M60	F: gcactggcgttcatcatctg R: atgttcattatggttcaggagg	(Shen *et al*., 2000)	MboI (NEB)	GATC (241+147)	+ 1bp insertion G**T**ATC (389)
M74	F: atgctataataactaggtgttgaag R: aattcagcttttaccacttctgaa	(Underhill *et al*., 2000)	RsaI (NEB)	G (385)	A (195+190)
M78	F1: cacttaacaaagatacttctttcc F2: cacttaacaaagatacttctttct R: attactttcctaggttctccca	Present Study	-	C (319)	T (319)
M89	F: acagaaggatgctgctcagctt R: gcaactcaggcaaagtgagacat	(Akey *et al*., 2001)	NlaIII (NEB)	C (20+67)	T (87)
M170	F: aaataatttcacgtttgttcaaataa R: acacaacccacactgaaaaaca	Present Study	NlaIII (NEB)	A (109+20)	C (129)
M172	F: atcccccaaacccattttgatgcat R: ggatccatctccatcttcactcaatgttg	(Nebel *et al*., 2001)	NlaIII (NEB)	T (148)	G (122+26)
M173	F1: tcaagggcatttagaaca F2: tcaagggcatttagaacc R: tgcaaagctttttaaaacaa	Present Study	-	A (205)	C (205)
M174	F1: tgaataccttctggagtgccct F2: tgaataccttctggagtgcccc R: tggtgaatacatcaaagagagga	Present Study	-	T (243)	C (243)
M175	F: ttgagcaagaaaaatagtaccca R: ctccattcttaactatctcaggga	(Underhill *et al*., 2000)	MboII (Amersham)	No del (370+74)	Del (444)

Continued

12f2	F: ctgactgatcaaaatgcttacagatc R: ggatcccttccttacaccttatac	(Rosser *et al*., 2000)	-	No del (+88)	Del (-88)
P25	F1: ctgcctgaaacctgcctgc F2: ctgcctgaaacctgcctga R: tccctgtaatcaagacaaaggc	Present Study	-	C (248)	A (248)
V12	F: caaagtttattttcaaaggggaga R: ccataaagttgggttgaaggag	(Cruciani *et al.*, 2006)	BsgI (NEB)	A (439)	G (237+202)
V13	F: ggatgctgctaaacatcctaca R: atcccatctcaatcccttaaca	(Cruciani *et al.*, 2006)	AciI (NEB)	G (45+190)	A (235)
V22	F: aatgcctcaacttacagaaatgg R: cactgaccagaaacagcatgag	(Cruciani *et al*., 2006)	MmeI (NEB)	T (289)	C (107+182)
V32	F: gcaaaatcccagaacatcatt R: tcattgacccaaagcagaca	(Cruciani *et al*., 2006)	MnlI (Fermentas)	G (150+205)	C (355)
V65	F: cctcaacctactaaatgtgaccatg R: atggccacacaattctccat	(Cruciani *et al*., 2007)	Sequencing	G (349)	T (349)

Table 2.2. Forward and Reverse Primers of Sex Identification.

Primer name	Primer sequence (5'-3')	Primer size	Primer reference
Amelogenin	ACCTCATCCTGGGCACCCTGG AGGCTTGAGGCCAACCATCAG	X = 212 bp Y = 218 bp	(Sullivan *et al.*, 1993)

2.2 Extant DNA

2.2.1 Sample Collection and Preparation

A total of 445 unrelated individual samples belonging to different 15 Sudanese populations were collected (Fig. 2.2). Appropriate informed consent was obtained from all participants. Sample sizes, geographic origin, and linguistic affiliation for each population are reported in Table (2.3). A fresh toothbrush was employed for each individual to improve the yield of cells from lining of the buccal cavity. The individual was first asked to wash his or her mouth out gently with water to remove large particles. The toothbrush was then moved over the inner surface of the cheeks to scrape the cells lining the mouth cavity. Two milliliter of phosphate buffer saline (PBS) were used to rinse the mouth, and the PBS with cells collected into a 50ml falcon tube. A cell pellet was collected by centrifugation at 3000rpm for 5 min, the supernatant discarded, and the cells resuspended in lysis buffer.

2.2.2 Blood Samples

Five milliliter venous blood samples were collected in vaccutainer EDTA tubes, using sterile needles and alcohol sawb and each blood sample was transferred to a 15ml falcon tube and 10ml of Red Blood Cell Lysis Buffer (RBCLB) was added to the blood, mixed gently by inversion, and spun in centrifuge at 3,000rpm. The supernatant was discarded carefully; and the washing has been repeated until a white pellet was reached. Then a volume of 2ml of White Blood Cell Lysis Buffer (WBCLB) was added to the pellet.

2.2.3 DNA Extraction

DNA extraction was employed following Chloroform method which has been described by Sambrook *et al.* (1990) with minor modifications. Ten µl of 10mg/ml

Table 2.3. Sample Size, Linguistic Affiliation and Geographic Location of Populations Analyzed.

Populations		Socio-economical activities	Sample size		Linguistic Affiliation[a]		Location	Latitude/ Longitude
Group	**Sub-groub**		**Y-chromosome**	**mtDNA**	**Family**	**Level**		
Nilotics	Dinka	Pastoralists	26	29	Nilo-Saharan	Eastern Sudanic	Malakal	9N 31E
	Shullik	Agri-pastoralists	15	16	Nilo-Saharan	Eastern Sudanic	Malakal	9N 31E
	Nuer	Pastoralists	12	16	Nilo-Saharan	Eastern Sudanic	Bentiu	9N 29E
Nubian	-	Agriculturists	39	29	Nilo-Saharan	Eastern Sudanic	Wadi-Halfa	21N 31E
Borgu	-	Agriculturists	26	25	Nilo-Saharan	Maban	EL-Genena	13N 22E
Masalit	-	Agriculturists	32	41	Nilo-Saharan	Maban	EL-Genena	13N 22E
Fur	-	Agriculturists	32	21	Nilo-Saharan	Fur	El-Fashir	13N 25E
Fulani	-	Nomadic Pastoralists	26	28	Niger-Congo	Atlantic	-	-
Hausa	-	Agriculturists	32	35	Afro-Asiatic	Chadic	Elgadarif	12N 8E
Nuba	-	Agri-pastoralists	28	33	Nilo-Saharan + Niger-Congo	Eastern Sudanic + Kordofanian	Kadugli	11N 29E
Copts	-	Agriculturists	33	29	Afro-Asiatic	Ancient Egyptian	Qena	26N 32E
Arab	Gaalien		50	22	Afro-Asiatic	Semetic	Shendi	16N 33E
	Meseria	Agriculturists	28	15	Afro-Asiatic	Semetic	-	-
	Arakien	Nomadic Pastoralists	24	17	Afro-Asiatic	Semetic	Medani	4N 33E
Beja	-	Agriculturists	42	48	Afro-Asiatic	Cushitic	Kassala	15N 36E
Total			**445**	**404**				

[a] According to ethnologue.

proteinase K, 1ml of 1M guanidine chloride, and 300µl of NH_4 acetate were added to each sample and incubated at 37°C over night. The samples were cooled to room temperature, transferred to another 15ml polypropylene falcon tubes containing 2ml of pre-chilled chloroform, vortexed and after 1 min spun at 1400g for 3 min. The upper layers from each sample were collected and transferred to a new falcon tube containing 10ml cold absolute ethanol. The tubes were inverted, shaken gently to precipitate the DNA and kept at -20°C for at least 2 hours. The tubes were centrifuged at 2000g for 15 min, the supernatants discarded. Four milliliter of 70% ethanol was added to the pellet in each tube for washing, spun at 2000g for 15 min, the supernates discarded and pellet was left for 2-3 hours to dry, 200µl of TE buffer were added to re-suspend the pellet. The DNA was left for 2 days at 4°C and then quantified.

2.2.4 Quantification of DNA and Genotyping

DNA was quantified using Picogreen kits as described previously. Several different methods for detecting polymorphisms at marker loci were employed for both Y-chromosome and mtDNA. All of these relied on PCR amplification of DNA using primers specific for the region.

2.2.5 Y-chromosome Genotyping

The Biallelic variability at Y-chromosome specific polymorphisms M1 (YAP), M2, M11, M13, M23, M33, M40, M42, M51, M52, M60, M74, M78, M89, M170, M172, M173, M174, M175, P25, and 12f2 (Table 2.1, and Appendix II) were used to generate male specific haplotypes. Also, to investigate the distribution of M78 binary subclades among the Sudanese populations, all 114 Y-chromosomes carrying the E-M78 derived T allele were further genotyped for 5 binary markers V12, V13, V22, V32, and V65 (Fig. 2.3) according to Cruciani *et al.* (2006, 2007).

Fig. 2.2. A map of the Sudan showing the approximate locations of 13 populations typed for Y-chromosome markers in this study. The nomadic populations, Meseria and Fulani were not shown in the map due to their wide distribution in west, east, and south-west of the country. Circles indicate the geographic regions.

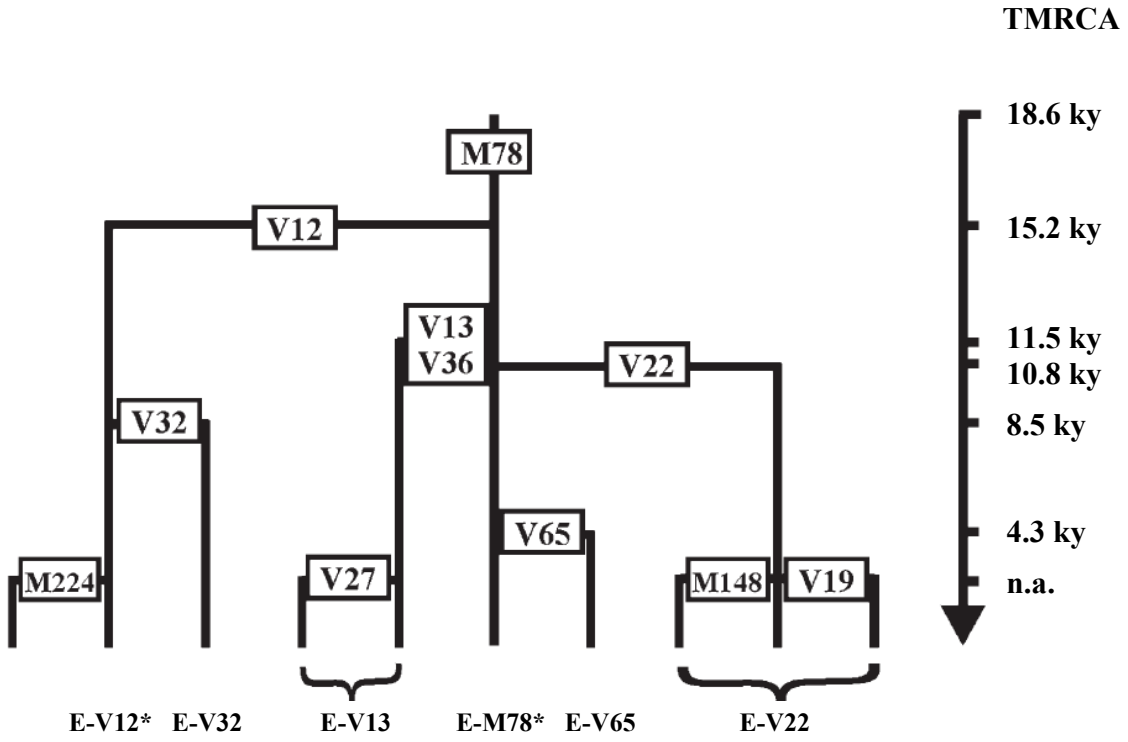


Fig. 2.3. Maximum parsimony phylogeny of haplogroup E-M78, based on 10 UEPs five of which are used here (V12, V13, V22, V32 and V65). Coalescent estimates for the haplogroup E-M78 and major subhaplogroups are shown on the right. Haplogroup nomenclature as cited in the text is reported at the bottom of the tree; n.a.- not available (rare haplogroups). **Source:** Cruciani *et al*. (2007).

2.2.6 Amplification Refractory Mutation System (ARMS) PCR Primer Design Program

For M2, M9, M23, M33, M40, M51, M78, M174, and P25, tetra-primer ARMS-PCR program (Ye *et al.*, 2001) was used for primer design to detect the ancestral and derived haplotypes. The program is accessible through the internet at (http://cedar.genetics.soton.ac.uk/public_html/primer1.html). PCR reactions were set up for two forward primers with common reverse primer. Touchdown PCR was carried out to detect the presence or absence of the derived haplotypes. To check for successful PCR amplification, human growth hormone gene has been co-amplified with all ARMS and 12f2 PCR as internal control, using primers HGHF (5´-GCCTTCCCAACCATT CCCTTA-3´) and HGHR (5´-TCACGGATTTCTGTTGTGTTTC-3´) which provide a product of 429bp (Perrey *et al.*, 1999). The PCR reactions were carried out on MJ Research PTC-225 Peltier Thermal Cycler. The PCR mix consisted of oligonucleotide primers, deoxynucleotide triphosphates (dNTPs) and the thermostable Taq DNA polymerase in 10x PCR buffer with a DNA samples as following:

dNTPs (equal mix of 100mM)	1.0µl
10x PCR buffer	2.5µl
$MgCl_2$ (50mM)	2.0µl
Forward primer (10µM)	1.50µl
Reverse primer (10µM)	1.5µl
Genomic DNA (100µg/ml)	1.0µl
H_2O	15.5µl

Touchdown program has been used for all Y-chromosome markers as described previously. Overall amplification success was determined by (2-4%) agarose gel electrophoresis.

2.2.7 Restriction Fragment Length Polymorphism (RFLP) for Single Nucleotide Polymorphism (SNP) Typing

For M11, M13, M42, M52, M60, M74, M89, M170, M172, and M175 markers, RFLP method was used for genotyping SNP genetic markers. The PCR was generated to flank the polymorphic site of interest. PCR products were subjected to digestion using the appropriate restriction enzyme, HpyCH4IV (NEB), MboII (Amersham), MspI (NEB), RsaI (NEB), AluI (NEB), BsgI (NEB), AciI (NEB), MmeI (NEB) and MnII (Fermentas) were used to digest M52, M175, M11, M74, M42, V12, V13, V22 and V32 respectively, also MboI (NEB) was used to digest both M60 and M13 as well as NlaIII (NEB) for M170 and M89 (Table 2.1 and Appendix II). The digestion has been carried out as follows:

	Per reaction	for 96 reactions
PCR product	8.5µl	
Enzyme	0.25µl	24µl
10x buffer	1.25µl	120µl

The reactions were performed in 96 microtitre plate. The plate was sealed and incubated at 37ºC for two hours in a PCR thermocycler. The presence or absence of the alleles was determined by agarose gel electrophoresis.

2.2.8 Mitochondrial DNA Genotyping

The mitochondrial control region or D-loop consists of three non-coding hypervariable regions (HVR-I, HVR-II, and HVR-III). To determine the sequencing of

the control region, by mean of Polymerase Chain Reaction (PCR); all of which is amplified in one PCR using L15996 and H408 primers (Table 2.4) to obtain approximately 1kb fragment as follows:

DNA (100ng)	1.00µl
10x Buffer	5.00µl
Primer1 (L15996)(10µM)	2.00µl
Primer2 (H408)(10µM)	2.00µl
dNTPs (2.5mM)	2.00µl
BSA (10mg/ml)	5.00µl
Taq polymerase	0.20µl
dH_2O	32.8µl
Total volume	50.0µl

Five min initial activation at 95ºC was followed by 30 cycles of 1 min denaturation at 94ºC, 1 min annealing at 56°C and 1 min extension at 72ºC, followed by a final extension for 10 min at 72°C in thermal cycler (Applied Biosystems). To detect the presence or absence of the 1kb specific bands electrophoresis was performed using 2% agarose gel, in presence of 1kb DNA marker.

2.2.9 Purification of PCR Product

The Nucleospin gel purification protocol has been followed for purification of DNA. PCR product was loaded on a 4.3% Nusieve gel in 1x TAE buffer, the band of interest has been cut out and placed in an Eppendorf tube. The weight of the gel was determined, then 200µl buffer NT were added for each 100mg of agarose gel, sample was incubated for 5-10 min at 50°C until the gel slices were dissolved, during incubation

Table 2.4. Oligonucleotide Pairs of Mitochondrial HVR-I and HVR-II.

Primer name	Primer sequence (5'-3')	Region	Primer Reference
L15996	CTC CAC CAT TAG CAC CCA AAG C	HVR-I	(Vigilant *et al*., 1989)
H16401	TGA TTT CAC GGA GGA TGG TG		
L29	GGT CTA TCA CCC TAT TAA CCA C	HVR-II	
H408	CTG TTA AAA GTG CAT ACC GCC A		

sample was vortexed briefly every 2-3 min. NucleoSpin® Extract II column was placed into 2ml collecting tube and sample was loaded. Sample was centrifuged for 1 min at 11,000g, flow was discarded and the NucleoSpin® Extract II column was placed back into the collecting tubes. 600µl buffer NT3 were added, sample was centrifuged for 1 min at 11,000g. Flow was discarded and the NucleoSpin® Extract II column was placed back into the collecting tube.

Sample was centrifuged for 2 min at 11,000g to remove buffer NT3 quantitatively. Sample was removed carefully from centrifuge to avoid the contact of the tube with the flow-through. Collecting tube was discarded. The residual ethanol from buffer NT3 has been removed by incubation of NucleoSpin® Extract II column for 2-5 min at 70°C prior to elution.

NucleoSpin® Extract II column was placed into a clean 1.5ml micro-centrifuge tube, 15-50µl elution buffer NE were added and sample was incubated at room temperature for 1 min to increase the yield of eluted DNA. Sample was centrifuged for 1 min at 11,000g. Eluted DNA was collected by centrifugation.

2.2.10 Cycle Sequencing

DNA sequencing was carried out for both HVR-I and HVR-11 using primers L15996 and L29 respectively (Fig. 2.4), in the presence of fluorescently labeled dideoxynucleotides (ddNTPs), which are dideoxy analogues of dNTPs that differ by the lack of a hydroxyl group attached to C-3 of the sugar. The absence of this OH group prohibits phosphodiester bond formation to the subsequent dNTP therefore incorporation of the analogue terminates chain synthesis. An oligonucleotide is used to prime sequence from the specific region of template DNA required. A volume of 5µl of the purified PCR product was run on a 2% agarose gel to determine how much of template required in -

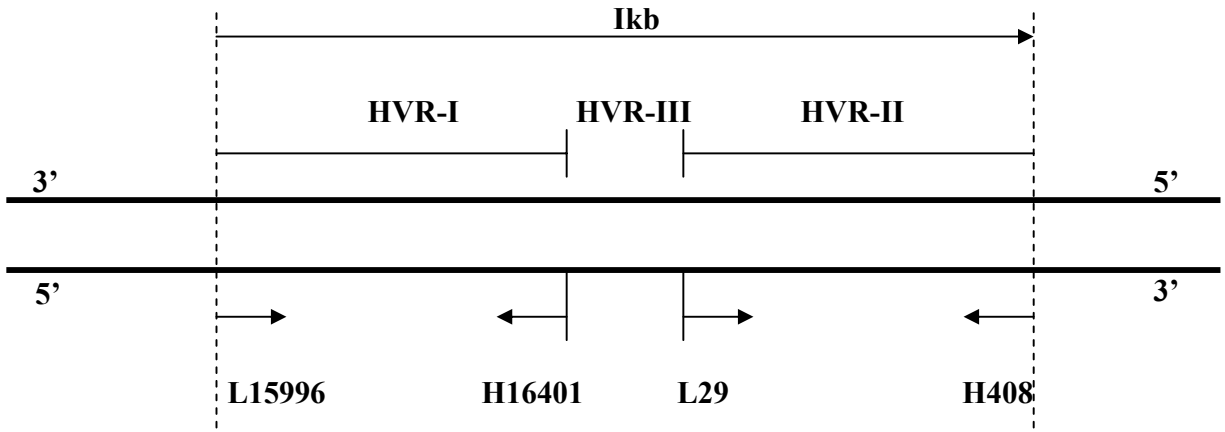


Fig. 2.4. A diagram shows the amplification of HVR, forward and reverse cycle sequencing of HVR-I, and HVR-II.

order to obtain ~100ng of DNA. The ABI PRISM Big Dye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems) was used for sequencing. The cycle sequencing was performed as following:

Big Dye premix	4µl
Primer L15996 (3.3uM)	1µl
dH_2O to make up to	20µl
Total	20µl

Cycling parameters consists of 25 cycles with 96°C for 30 seconds, 50°C for 15 seconds and 60°C for 4 min in a thermal cycler (Applied Biosystems). Product was stored in a covered box at 4°C until ready to purify.

2.2.11 DNA Purification for Sequencing

The DNA purification has been performed following QIAGEN DyeEx 2.0 Spin Kit Protocol for sequence analysis using ABI 3100 Genetic Analyzer. All centrifugation steps were performed using Eppendorf Centrifuge 5415C. DyeEx 2.0 spin column was gently vortexed to resuspend the resin. The cap of the column was loosen quarter turn to avoid a vacuum inside the spin column. The bottom closure of the spin column was snapped off and the spin column was placed in a 2ml collection tube. Spin column was centrifuged for 3 min at 3000rpm. Spin column was carefully transferred to a clean centrifuge tube. And the sequencing reaction (10-20µl) was applied to the gel bed by pipetting the sequencing reaction directly onto the center of the slanted gel-bed surface. Spin column containing the sequencing reaction was centrifuged for 3 min at 3000rpm, and spin column was removed from the micro-centrifuge tube, the elute contains the purified DNA. Samples were dried in a vacuum centrifuge for 40 min to 1 hour and stored in a

covered box until ready to use. Samples were resuspended in 10-25µl HI-DI Formamide (ABI), and transferred to ABI 96 microplate.

2.2.12 DNA Sequencing

All samples were sequenced using ABI 3100 Genetic Analyzer, which is a high throughput machine using a capillary system instead of gel electrophoresis, with a computer workstation for instrument operation and data analysis. It can automatically analyze multiple runs of 96 samples, with a high performance, producing sequence read lengths greater than 550bp at 98.5% accuracy. All relevant options on the computer were selected according to the manufacture before preparing the samples sheet, and samples were run.

For those sequences containing a homopolymeric cytosine stretch which usually associated with length heteroplasmy, additional amplification and reverse sequencing were performed for HVRI and HVRII using primers H16401 and H408 respectively (Fig. 2.4).

2.3 Data Analysis

2.3.1 mtDNA Sequences Alignment

The mtDNA alignment for HVR-I (Fig. 2.5) and HVR-II (Fig. 2.6) sequences was performed using BioEdit software which is a biological sequence alignment editor (Hall, 1999). BioEdit DNA alignment software can be downloaded free from the website: http://www.mbio.ncsu.edu/BioEdit/bioedit.html

2.3.2 mtDNA Haplotypes Classification

The HVS-I and HVS-II sequences were classified into mtDNA different haplogroups. Classification of subclades of haplogroups L0, L1, L2, and L3 (Table 2.5) following the nomenclature of Torroni *et al.* (2001), Salas *et al.* (2002, 2004), and Shen

et al. (2004). Haplogroups L4, L5, and L6 following the nomenclature of Kivisild *et al.* (2004).

2.3.3 Genetic Distances

mtDNA Genetic distance between populations were calculated using Molecular Evolutionary Genetics Analysis (MEGA) version 3 based on Tajima's Nei model (Tajima and Nei, 1984). MEGA is an integrated tool for automatic and manual sequence alignment, inferring phylogenetic trees, mining web-based databases, estimating rates of molecular evolution (Kumar *et al.*, 2004) the program is available at: http://www.megasoftware.net/

2.3.4 Molecular Diversity

To locate such an expansion in time among Sudanese populations, mismatch distribution between individuals was performed based on HVR-I sequences using ARLIQUIN 3.0, which is a program for population genetics analysis (Excoffier *et al.*, 2005), available online at: http://lgb.unige.ch/arlequin/. Since small sample sizes could potentially affect the reliability of the analysis, all populations whose sample size is < 20 were pooled. Pooling was performed based on linguistic affiliation and/or geographical proximity. Accordingly, Arakien and Meseria were merged with Gaalien as Arab, Nuer and Shilluk were merged with Dinka as Nilotics. Mismatch distributions were not performed for Y-chromosome data due to the small number of polymorphisms. Also nucleotide diversity, polymorphic sites, transitions and transversions in HVR-I within population were calculated based on pairwise difference method using ARLIQUIN 3.0. Pairwise distances for Y-chromosome and mtDNA were performed to describe the short term genetic distance between populations, using Markov chain with 10,000 steps.

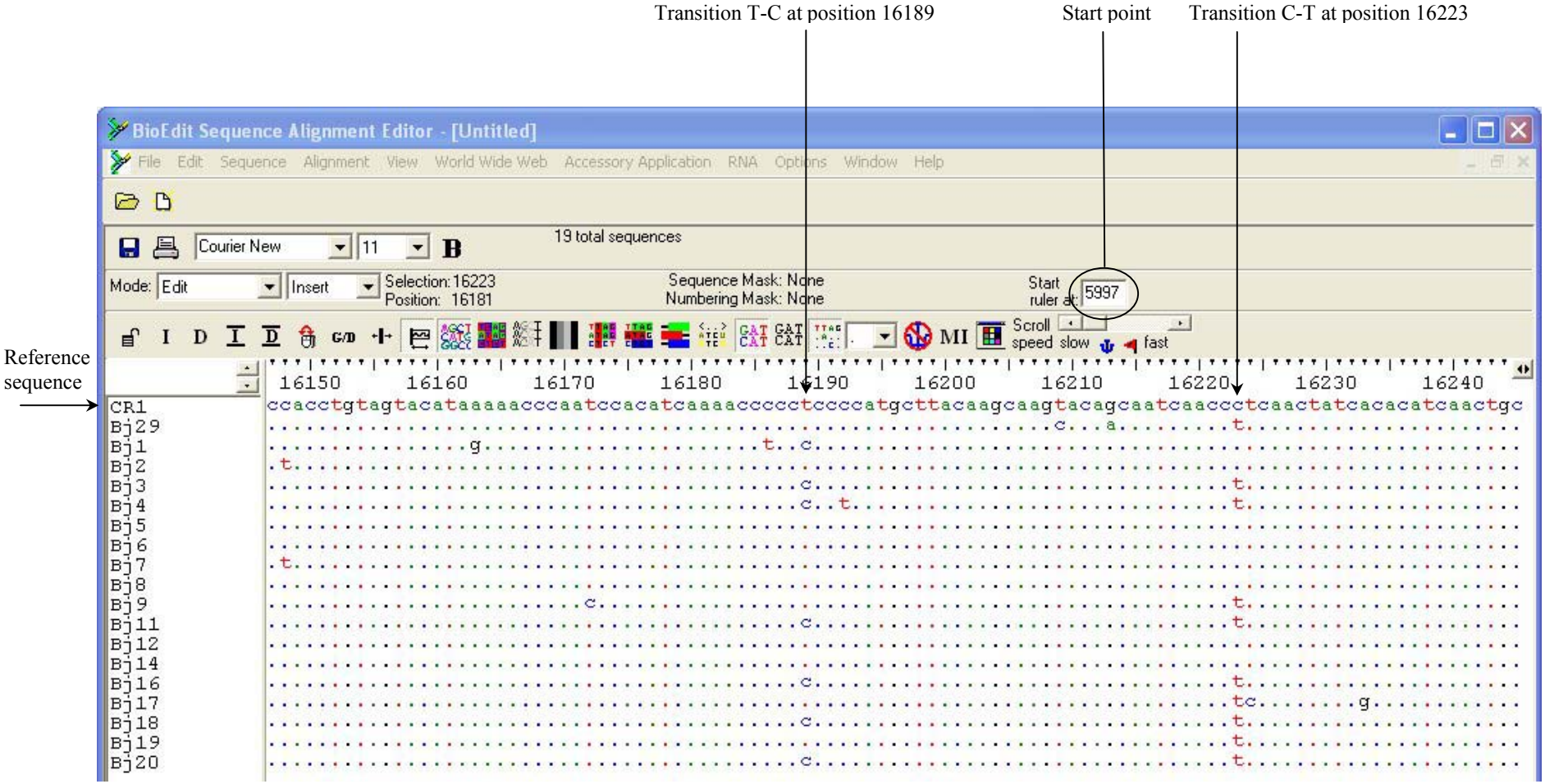


Fig. 2.5. Showing DNA multiple alignment of HVR-I using BioEdit software.

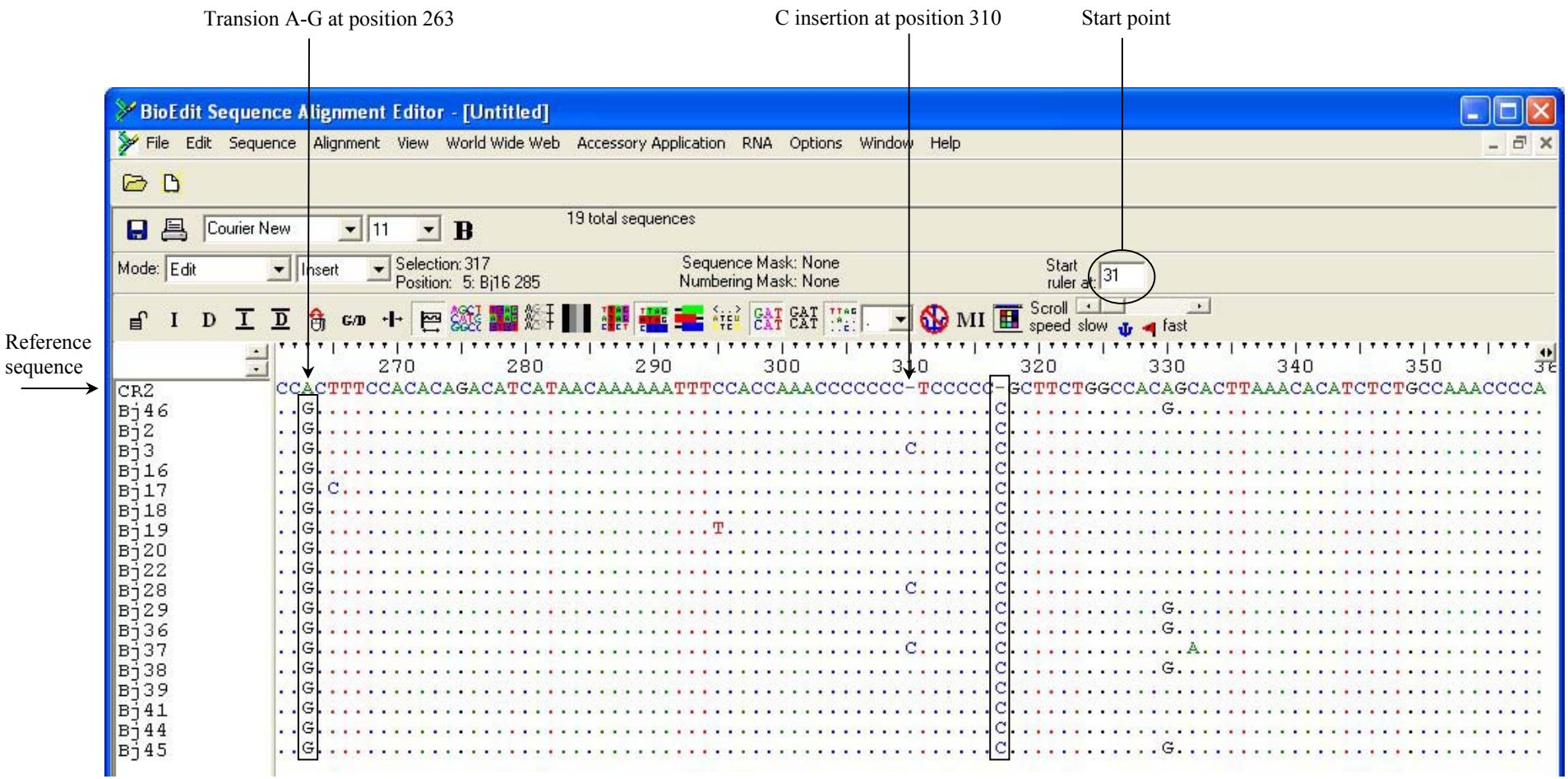


Fig. 2.6. Showing DNA multiple alignment of HVR-II using BioEdit software.

Table 2.5. mtDNA Motifs used to Classify Sequences Types in this Study

Haplotype	**HVR-I and HVR-II motifs**	**73 status**
L0a	16129 16148 16172 16187 16188(G) 16189 16223 16230 16311 16320	A
L0a1	16168	G
L0a1a	16278	G
L0a2	16148 16172 16187 16188 16189 16223 16230 16311 16320	G
L0d	16129 16243	G
L0f	16169 16187 16189 16209 16223 16230 16278 16311 16327	G
L1*	16187 16189 16223 16230 16278 16311	G
L1b	16126 16187 16189 16223 16264 16270 16278 16311	G
L1b1	16293	G
L1c	16129 16187 16189 16223 16278 16294 16311 16360	G
L1c1	16129 16187 16189 16223 16278 16293 16294 16311 16360	G
L1c1a1	16214 16223 16234	G
L1c3	16189 16215	G
L2*	16223 16278 16390	G
L2a	16294	G
L2a1	16309	G
L2a1a	16286	G
L2b	16114A 16129 16213 16223 16278	G
L2b1	16362	G
L2c	325 16223 16278 16390	G
L2c1	325 16318	G
L2c2	325 16264	G
L2d1	16129 16189 16223 16300 16354	G
L2d2	16111 16145 16239 16292 16355	G
L3*	16223 16278	G
L3b	16124 16223 16278 16362	G
L3b1	16124	G
L3b2	16311	G
L3d	16124 16223 16111 16311	G

Continued

Continued

L3d1	16319	G
L3d2	16256	G
L3d3	16189 16278 16304 16311	G
L3e1a	16185 16223 16327	G
L3e2	16223 16320	G
L3e2b	16172 16189	G
L3e5	16041G 16223T	G
L3f	16209 16223 16311	G
L3f1	16292	G
L3h	16129 16223 16256A 16311 16362	G
L3i	16124 16223 16319 16153	G
L3w	16223 16260	G
L3x2	16169 16193 16195 16223	G
L4a	16223 16260 16362	G
L4g	16223 16293(T) 16311 16355 16362	G
L5a	16129 16148 16166 16187 16189 16223 16311	G
M/D	16223 16362	G
M1	16129 16189 16223 16249 16311	G
M1a	16129 16189 16223 16249 16311 16359	G
N/J1a	16126 16145 16231 16261	G
N/J1b	16126 16145 16222 16261	G
N/J2	16126 16193	G
N1a	16147G 16172 16223 16248 16355	G
preHV1	16126 16362	G
R/U3	16343	G
R/U6a1	16172 16189 16219 16278	G
R/T1	16126 16163 16186 16189 16294	G

2.3.5 Y-chromosome Phylogenetic Trees

Y-chromosome phylogenetic tree of the 15 Sudanese populations and their haplogroups was designed by hand according to YCC nomenclature (YCC, 2002). Neighbor-joining tree of Y-chromosome was generated using Genetic Data Analysis (GDA) and TreeView software based on Nei (1978) genetic distance. GDA is a population genetics program (Weir, 1990), available online at http://hydrodictyon.eeb.uconn.edu/people/plewis/software.php.

2.3.6 mtDNA Phylogeny

2.3.6.1 mtDNA Phylogenetic Trees

Three types of phylogenetic trees neighbor-joining (NJ), minimum evolution (ME) and Unweighted Pair Group Method Using Arithmetic Averages (UPGMA) were constructed for the 15 Sudanese populations in the present study using MEGA 3 program (Kumar *et al.*, 2004) based on pairwise genetic distances of mtDNA HVR-I. To convert the raw sequence data into a genetic distance matrix; individual sequences were grouped into the different 15 populations, and pairwise genetic distances were calculated based on the model of Tamura and Nei (1993) using 500 bootstarp replicates, this model corrects for multiple hits, taking into account the differences in substitution rate between nucleotides and the inequality of nucleotide frequencies. It distinguishes between transitional substitutions rates between purines and transversional substitution rates between pyrimidines. It also assumes equality of substitution rates among sites. The genetic matrix was used to generate the phylogenetic trees.

2.3.6.2 mtDNA Phylogenetic Networks

To portray haplotypes variation and frequencies of mitochondrial haplogroups among Sudanese populations, sequences of mtDNA HVR-I were used to draw a median

network (Bandelt *et al*., 1999) for each haplogroup independently using Network 4.5 software (available online at http://www.fluxus-engineering.com) with minor modifications.

2.3.7 Analysis of Molecular Variance (AMOVA)

AMOVA test was performed to verify statistical differences between groups of populations. Both haplotypes frequencies and molecular differences of Y-chromosome and mtDNA among haplogroups were taken into account. The populations were grouped into the three linguistic families. All calculations were performed using the ARLIQUIN 3.0.

2.3.8 F-statistics and Migration Rate

The theory based on the island model (Wright, 1951) of migration indicates that F_{ST} is equal to 1/(1+4Nm) for diploid markers, and to 1/(1+Nm) for haploid markers (Y-chromosome and mtDNA). The effective number of migrants (Nm) for both Y-chromosome and mtDNA, which is the product of the population effective size (Ne) and the migration rate (m) when the mutation rate is negligible, was estimated based on the F_{ST} values assuming an island model (Wright, 1969). This model assumes that all populations have the same effective size and the sexual proportion is 1:1. To estimate F_{ST} and then migration rate (Nm), Nilo-Saharan speaking groups were divided into Nilotic and Non-Nilotic, and Afro-Asiatic were divided into Arabs and Non-Arabs; Niger-Congo was excluded from this test due to the small sample size.

2.3.9 Principal Component Analysis (PCA)

For the populations analyzed in this study; two runs of PCA were performed for F_{ST} genetic distance matrices of Y-chromosome and mtDNA separately using PAST software (available online at: http://folk.uio.no/ohammer/past). The first run was performed for

Sudanese samples from the present study, whereas in the second run published data of 4 populations including Amhara, Oromo, Senegalese (Semino *et al.,* 2002) and Turks (Sanchez *et al.,* 2005) were used alongside population data from the present study to generate PC plots of Y-chromosome. Also two runs of PCA were performed for mtDNA data, the first for Sudanese populations from the present study and the second to compare between Sudanese populations with 6 neighboring populations including Egyptian (Stevanovitch *et al.,* 2004), Amhara and Oromo (Kivisild *et al.,* 2004), Berber (Fadhlaoui-Zid *et al.,* 2004), Guineans (Rosa *et al.,* 2004), Chadi (Hajek *et al.,* 2008), Khoisan (Behar *et al.,* 2008) and Saudi Arabian (Abu-Amero *et al.,* 2006). Some of the populations used in the comparison are single tribal groups, whereas others are pools of populations.

2.3.10 Mantel Test

The correlation among genetic, linguistic, and geographic distances for both Y-chromosome and mtDNA as well as the correlation between pairwise distances for Y-chromosome and mtDNA was assessed by the Mantel test (Mantel, 1967; Smouse *et al.*, 1986), employing ARLEQUIN 3.0 (Excoffier *et al.*, 2005). This test is a statistical procedure that allows measuring the correlation coefficient between two matrices while holding a third one constant. To test whether statistically significant associations between linguistic and genetic affiliations reflect the same events in population history or parallel, but separate isolation by distance processe; genetic distances matrices were based on Slatkin's distance linearized F_{ST} values which is consider as incorporating molecular distances among haplogroups (Slatkin, 1995). To find if there is any correlation between genetic and linguistic, Sudanese populations were classified according to the linguistic affiliation into the main families, Afro-Asiatic, Nilo-Saharan, and Niger-Congo. Nuba

populations were excluded from the analysis because they are composes of two linguistic families Niger-Congo and Nilo-Saharan. Also to find if there is any correlation between genetic and geographic relations; geographic distances between populations were calculated and converted to a matrix using approximate latitude and longitude data using Latitude/Longitude Distance Calculation program. This program is available online at: http://jan.ucc.nau.edu/~cvm/latlongdist.html assuming that the earth is a perfect sphere and consider the distance between two points as a straight line. Fulani and Meseria were excluded form the analysis because they are widely spread nomadic populations. And linguistic distances matrices were calculated by hand, since the linguistic classification of Greenberg (1963) is hierarchical. In this way, the different languages can be arranged in a tree-like diagram, the languages at the tips of the tree being connected by a series of branches and nodes. Thus, if persons from two populations speak the same language, their distance is zero. This distance increases by one unit as we advance one node in the direction of the root of the diagram. This approach was already used by Excoffier *et al.* (1991), Llop (1996), and Wood *et al.* (2005).

2.4 Reference Management

The management of the references has been done manually following the style of the American Journal of Physical Anthropology (AmJPA).

CHAPTER 3

RESULTS

3.1 Results of aDNA

3.1.1 Results of aDNA Extraction

After the application of the appropriate extraction procedure, all types of skeletal elements including spongy and compact bone provided a considerable amount of aDNA ranged from 1.8–25.2 ng/µl, using Silica-Guanidine Theocyanate (GnSCN) method. No differences concerning DNA quality could be detected within the different types and origins of ancient bone samples. The highest aDNA concentration was found in a female teeth sample from a post-Meroitic grave from the 4th Cataract, this was qualitatively obtained from the Micro-plate reader (FLOstar-Optima) results.

3.1.2 Results of Genotyping

Thirty five out of 76 samples which have been collected from northern Sudan were positive for DNA extraction. Thirty three samples out of the 35 samples were positive for PCR when using sex Amelogonin marker, 14 samples of them were females and 19 were males. Three samples from Neolithic were positive for haplogroup A-M13, and 2 samples were positive for YAP insertion. For Meroitic and post-Meroitic samples, 5 of them were carrying YAP insertion, while a single post-Meroitic sample was positive for haplogroup A-M13. One rib sample from Haraz Cemetry was negative for both M13 and YAP and failed to give any result for M60 and M89. Three samples from Christian period were positive YAP insertion, and 2 samples were carrying M89 mutation, M13 and M60 were not shown among Christian period (Table 3.1). Gel electrophoresis results of sex determination are illustrated in Fig (3.1), while the YAP-PCR results of Nuri mummy

and some of aDNA samples are illustrated in Fig (3.2) and Fig (3.3) respectively. aDNA extarction was negative for the all 18 samples which have been collected from the Dinder National Park in eastern Sudan and supposed to be from Meroitic period according to the archaeological investigations.

3.2 Results of Extant DNA

3.2.1 Y-chromosome Genotyping

Gel electrophoresis results of PCR and RFLP genotyping of some of Y-chromosome specific haplotypes are illustrated in Fig (3.4–3.5). and the results of PCR of the mitochondrial HVR are shown in Fig (3.6).

3.2.2 Y-chromosome Haplotypes Diversity

Y-chromosome haplotype frequencies in 15 Sudanese populations are shown in Fig (3.7) following YCC nomenclature (2002). Eight haplogruops A, B, E, F, I, J, K and R have been observed among Sudanese populations in frequencies of 16.9, 8.1, 34.2, 3.1, 1.3, 22.3, 0.9, and 13% respectively. Haplogroups A-M13 and B-M60 are present at high frequencies in Nilo-Saharan speaking groups except Nubians, and with low frequencies in Afro-Asiatic speaking groups although notable frequencies of B-M60 were found in Hausa (15.6%) and Copts (15.2%). Both haplogroups were absent in Fulani. Haplogroup E which is defined by M40 (4 different haplotypes) accounts for the majority (34.2%) of the chromosomes and is widespread in Sudan. E-M78 represent 80% of haplogroup E, the highest frequencies found to be in Masalit and Fur populations, E-M33 (5.3%) is present in Fulani and Hausa, whereas E-M2 (5.3%) is restricted to Hausa, and E-M215 (9.4%) seems to occur more in Nilo-Saharan as compared to Afro-Asiatic speaking groups. In contrast haplogroups F-M89, I-M170, J-12f2 and J-M172 were found to be more frequent in Afro-Asiatic speaking groups. Haplogroups J-12f2 and J-M172

Table 3.1. Genotyping of aDNA from Different Archeological Periods, samples with negative PCR results are not shown in the table.

No	Period	Collection site	Skeletal element	aDNA (ng/ul)	Genotyping				
					Sex	M13	M60	YAP	M89
1	Neolithic	Kadruka	Femur	5.0	F				
2	Neolithic	Kadruka	Ulna	7.2	M	+ve	-	-	-
3	Neolithic	Kadruka	Femur	8.4	M	+ve	-	-	-
4	Neolithic	Kadruka	Tibia	3.6	M	-	-	+ve	-
5	Neolithic	Kadruka	Tibia	3.5	F				
6	Neolithic	Kadruka	Femur	9.6	M	+ve	-	-	-
7	Neolithic	Kadruka	Femur	8.6	F				
8	Neolithic	Kadruka	Femur	3.5	M	-	-	+ve	-
9	Meroitic	Haraz Cemetery	Rib	8.7	F				
10	Meroitic	Haraz Cemetery	Tarsal bone	8.7	M	-	-	+ve	-
11	Meroitic	Haraz Cemetery	Skull	4.1	M	-	-	+ve	-
12	Meroitic	Haraz Cemetery	Rib	8.7	M	-ve	-	-ve	-
13	Meroitic	Haraz Cemetery	Rib	8.7	M	-	-	+ve	-
14	Post-Meroitic	4th Cataract	Teeth	25.2	F				
15	Post-Meroitic	Haraz Cemetery	Lumber vertebra	3.4	M	-	-	+ve	-
16	Post-Meroitic	4th Cataract	Tarsal bone	9.4	F				
17	Post-Meroitic	Haraz Cemetery	Rib + Ulna	8.7	F				
18	Post-Meroitic	Haraz Cemetery	Skull	10.1	M	-	-	+ve	-
19	Post-Meroitic	Haraz Cemetery	Skull	5.7	M	+ve	-	-	-
20	Christian	4th Cataract	Thoracic vertebra	8.7	F				
21	Christian	Meroe Island	Patella	6.3	M	-ve	-	+ve	-
22	Christian	Meroe Island	Thoracic vertebra	8.7	M	-ve	-	+ve	-
23	Christian	Meroe Island	Rib	5.0	F				
24	Christian	Meroe Island	Lumber vertebra	13.8	F				
25	Christian	Meroe Island	Rib	13.1	M	-	-	-	+ve
26	Christian	Meroe Island	Rib	5.9	M	-ve	-	-ve	-
27	Christian	4th Cataract	Rib	7.7	F				
28	Christian	Meroe Island	Teeth	5.3	M	-ve	-	-ve	-
29	Christian	Meroe Island	Pelvic girdle	9.9	F				
30	Christian	4th Cataract	Rib	10.7	M	-	-	-	+ve
31	Christian	Meroe Island	Shull	1.8	F				
32	Christian	Meroe Island	Rib	11.2	F				
33	Christian	Nuri	Skin (Mummy)	2.4	M	-	-	+ve	-

(-) indicates negative PCR results

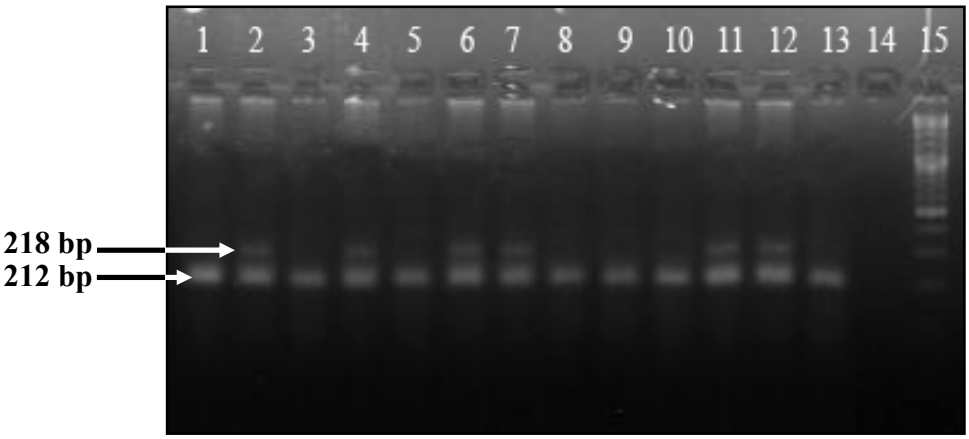


Fig. 3.1. Amplification of Amelegonin gene for aDNA sex determination; Lane 1, 3, 5, 8, 9, 10 and 13 show the absence of Y-chromosome (female samples); while lane 2, 4, 6, 7, 11 and 12 show the presence of Y-chromosome (male samples). Lane 15, 123bp DNA marker; lane 14, negative control. The products were run on 2% agarose gel electrophoresis.

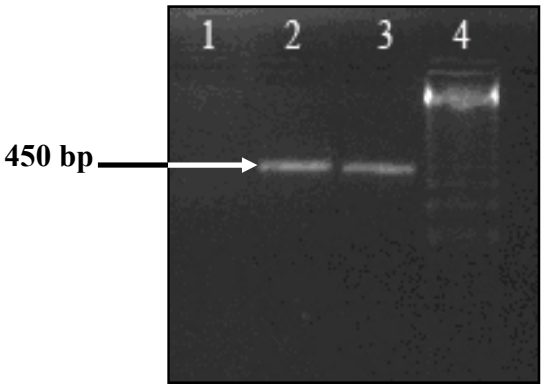


Fig. 3.2. Y-chromosome analysis of Nuri mummy; Lane 1, negative control; lanes 2 and 3 show YAP+; lane 4, 100bp DNA marker. The products were run on 2% agarose gel electrophoresis.

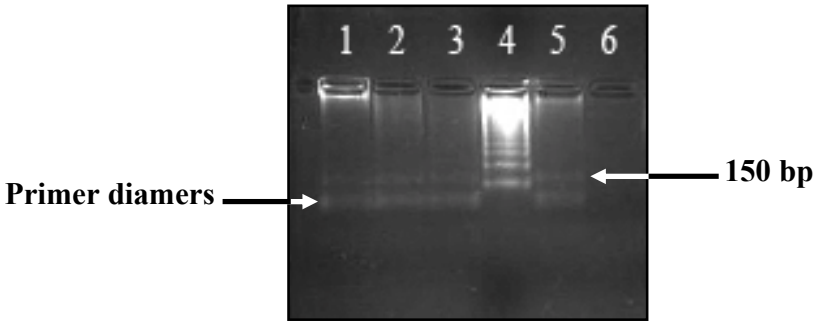


Fig. 3.3. Amplification of aDNA samples. Lanes 1, 2, 3 and 5 show YAP –ve; lane 4, 100bp DNA marker; lane 6, negative control. The products were run on 2% agarose gel electrophoresis.

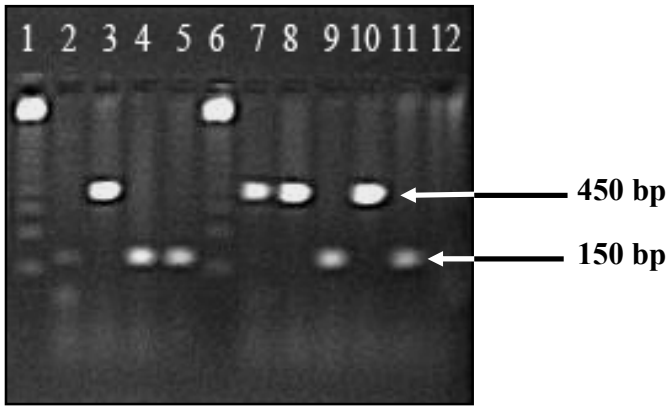


Fig. 3.4. PCR amplification of YAP; Lanes 1 and 6 100bp DNA marker; lanes 2, 4, 5, 9 and 11 show YAP-ve; Lanes 3, 7, 8 and 10 show YAP+ve; lane 12, negative control. The products were run on 2% agarose gel electrophoresis.

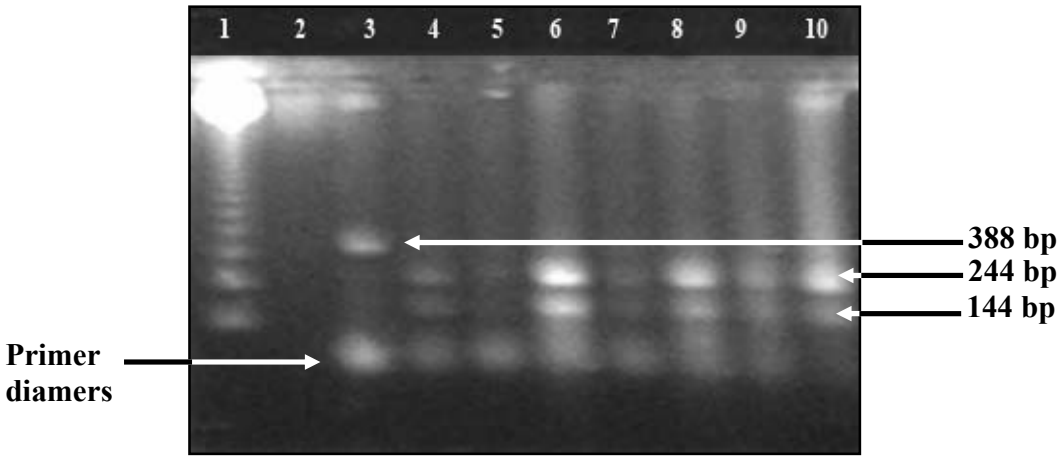


Fig. 3.5. RFLP analysis of M60; Lane 1, 100bp DNA marker; lane 2, negative control; lane 3 positive for M60; lanes 4, 5, 6, 7, 8, 9, and 10 negative for M60. The products were run on 2% agarose gel electrophoresis.

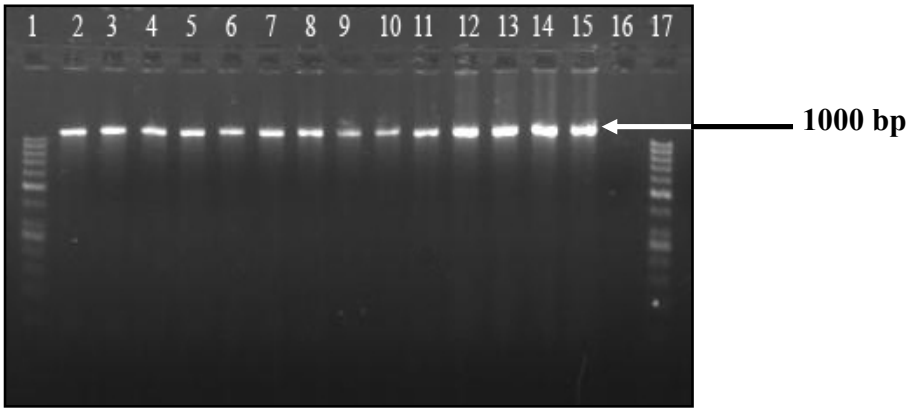
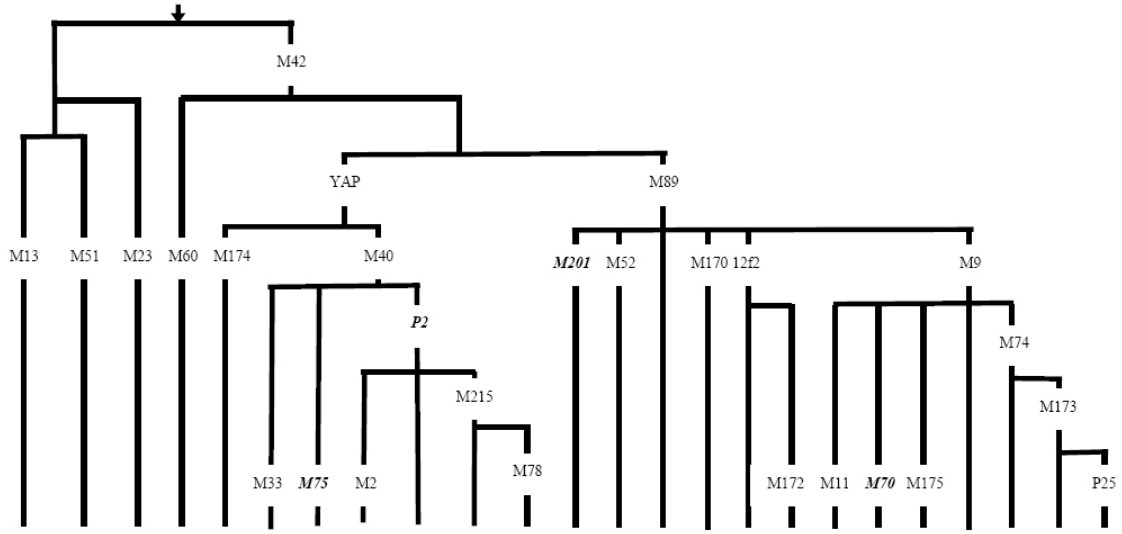


Fig. 3.6. PCR amplification of mitochondrial hypervariable region; Lanes 1 and 17, 50bp DNA marker; lane 16, negative control, the products were run on 2% agarose gel electrophoresis.



Haplogroups	A3b2	A3b1	A2	B	D	E1	E2	E3a	E3	E3b	E3b1	G	H	F	I	JI	J2	L	K2	O	K*	P	R1	R1b
Dinka (26)	16			6							4													
Shilluk (15)	8			4						1	2													
Nuer (12)	4			6							2													
Borgu (26)	9									10	4													3
Nuba (28)	13			4						4	7													
Masalit (32)	6			1							23					2								
Fur (32)	10			1							19					2								
Nubians (39)				3						3	6			4	2	16	1							4
Fulani (26)						3					9												14	
Hausa (32)	4			5		5		4			1													13
Copts (33)				5						2	5					13	2				1			5
Beja (42)	2									7	15					15	1							2
Gaalien (50)	3										9			5	2	18	2				3		1	7
Meseria (28)											4			3	2	12								7
Arakien (24)											4			2		16								2
Turks (59)										7	1	5			4	1	16		1		2	2	7	12
Oromo (78)	8			1			1		12	21	28					2	1		4					
Amhara (48)	7			1					5	6	11					15	1		2					
Senegalese (139)						7	4	113	4	8	1			2										

Fig. 3.7. Phylogenetic distribution of the Y-chromosome haplotypes and their frequencies in 15 Sudanese populations in the present study, compared with the frequencies in the Turks (Sanchez *et al.*, 2005), two Ethiopian groups (Oromo and Amhara), and Senegalese (Semino *et al.*, 2002). Numbering of mutations and haplogroups nomenclature are according to YCC (2002): those previously reported in the literature are shown in italics. The arrow shows the root of the tree.

represent 94% and 6% respectively of haplogroup J with high frequencies among Nubians, Copts and Arabs. Haplogroup K-M9 is restricted to Hausa and Gaalien at low frequencies, and absent in Nilo-Saharan and Niger-Congo speaking groups. Haplogroup R-M173 appears to be the most frequent haplogroup in Fulani (53.8%), while haplogroup R-P25 has the highest frequency in Hausa but found to be common also in Arabs and Copts, and with low frequencies in Nilo-Saharan speaking groups. Haplogroups A-M51, A-M23, D-M174, H-M52, L-M11, O-M175, and P-M74 were found to be completely absent from the populations sample analyzed.

3.2.3 E-M78 Subclades

The distribution of E-M78 subclades among Sudanese is shown in Table (3.2). Only 2 chromosomes fell under the paragroup E-M78*. E-V65 and E-V13 were completely absent in the samples analyzed, whereas the other subclades were relatively common. E-V12* accounts for 19.3% and is widely distributed among Sudanese. E-V32 (51.8%) is by far the most common subclades among Sudanese. It has the highest frequency among populations of western Sudan and Beja. E-V22 accounts for 27.2% and its highest frequency appears to be among Fulani, but it is also common in Nilo-Saharan speaking groups.

3.2.4 mtDNA Haplotypes Frequencies in Sudanese Populations

A total of 404 samples from different Sudanese populations were sorted into their respective clades and sub-clades based on mtDNA sequencing of HVR (Fig. 3.8–Fig. 3.18), samples that were tested for SNP variation in hypervariable region but did not fit into known haplogroup were classified into L1*, L2* and L3* (Appendix III). Those sequences containing homopolymeric cytosine stretch (Fig. 3.19) were usually associated with length heteroplasmy (Fig. 3.20). A total of 56 haplotypes were observed (Table 3.3),

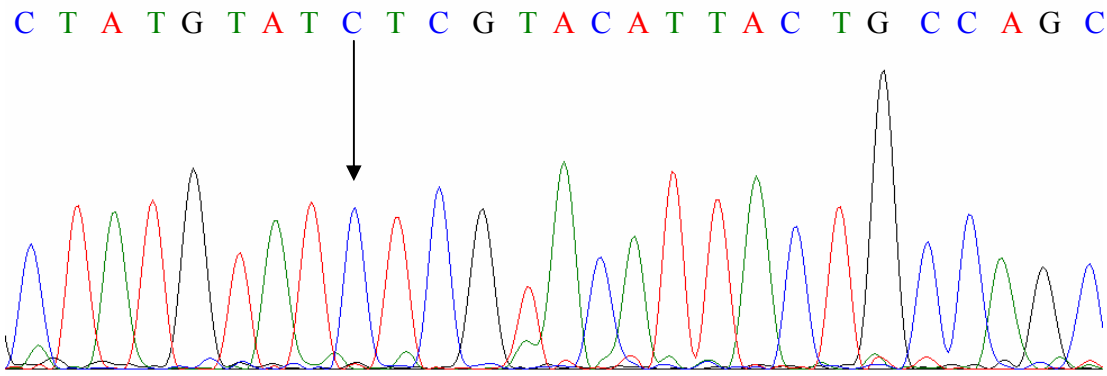


Fig. 3.8. Transition (T → C) at position 16093 HVR-I.

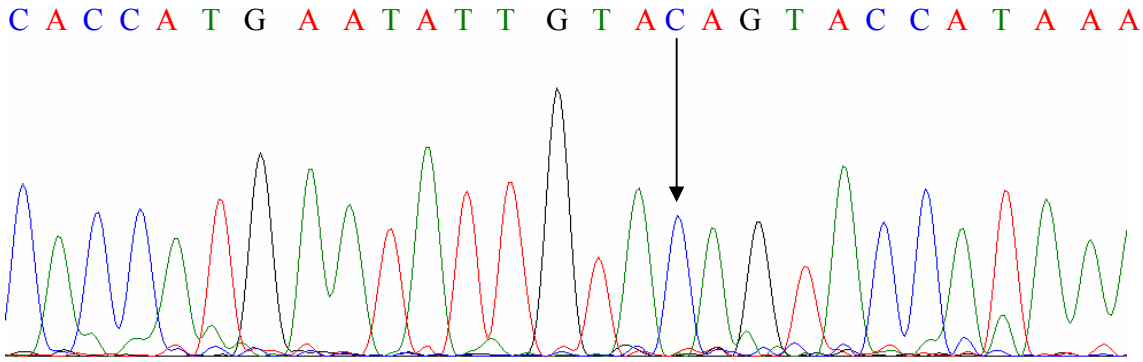


Fig. 3.9. Transversion (G → C) at position 16126 HVR-I.

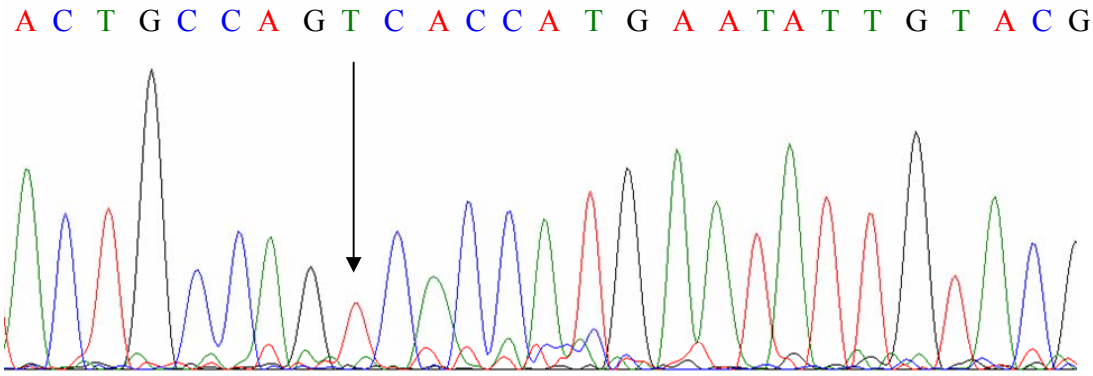


Fig. 3.10. Transition (C → T) at position 16111 HVR-I.

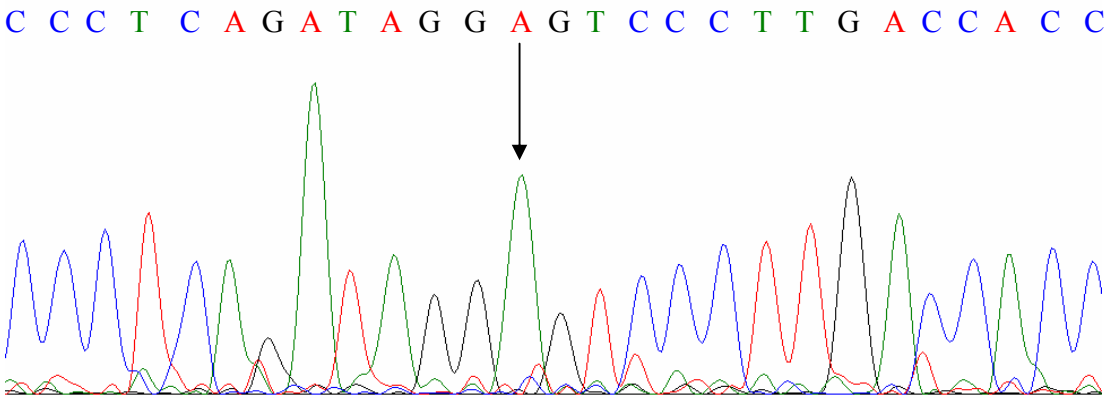


Fig. 3.11. Transition (G → A) at position 16390 HVR-I.

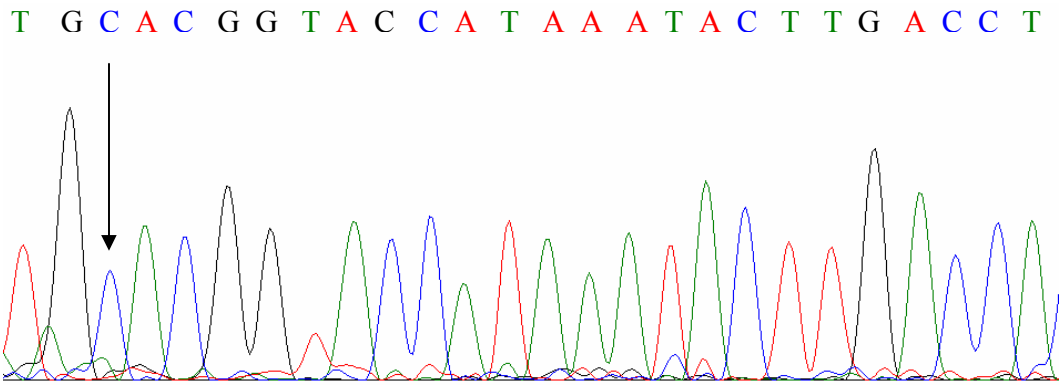


Fig. 3.12. Transition (T → C) at position 16126 HVR-I.

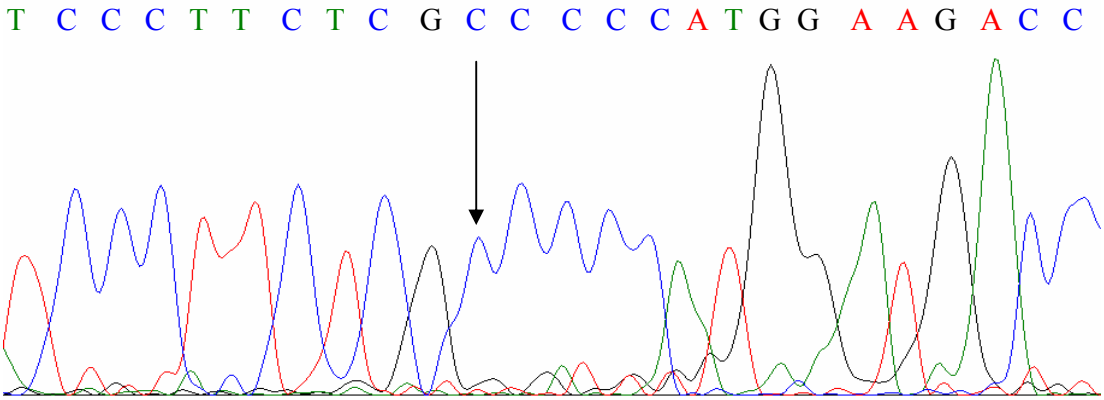


Fig. 3.13. Transition (T → C) at position 16362 HVR-I.

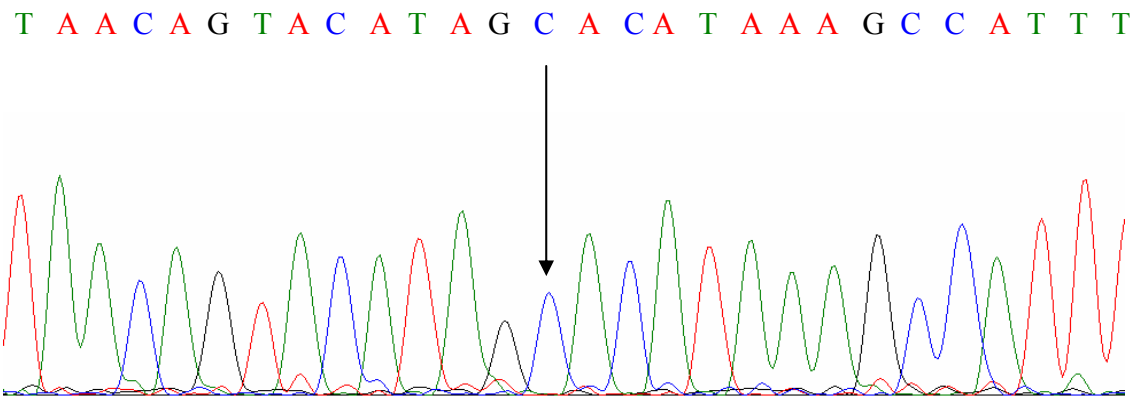


Fig. 3.14. Transition (T → C) at position 16311 HVR-I.

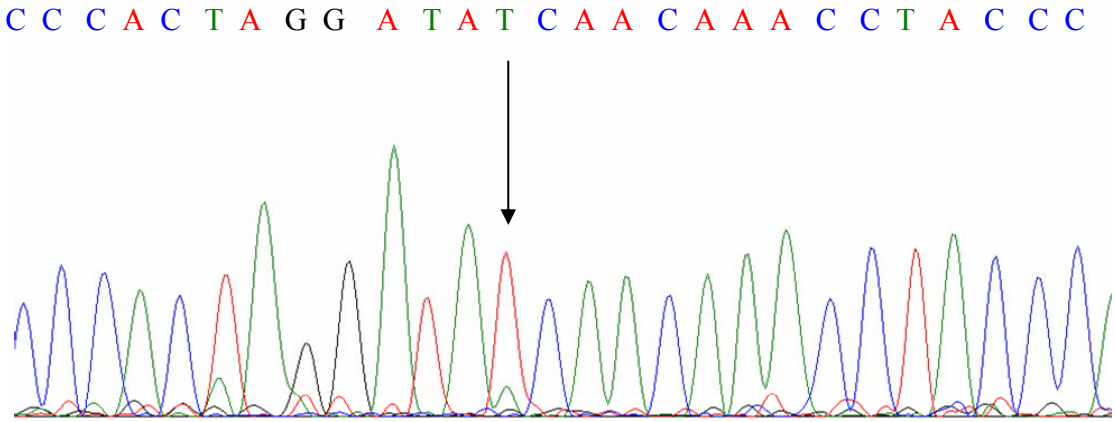


Fig. 3.15. Transition (C → T) at position 16278 HVR-I.

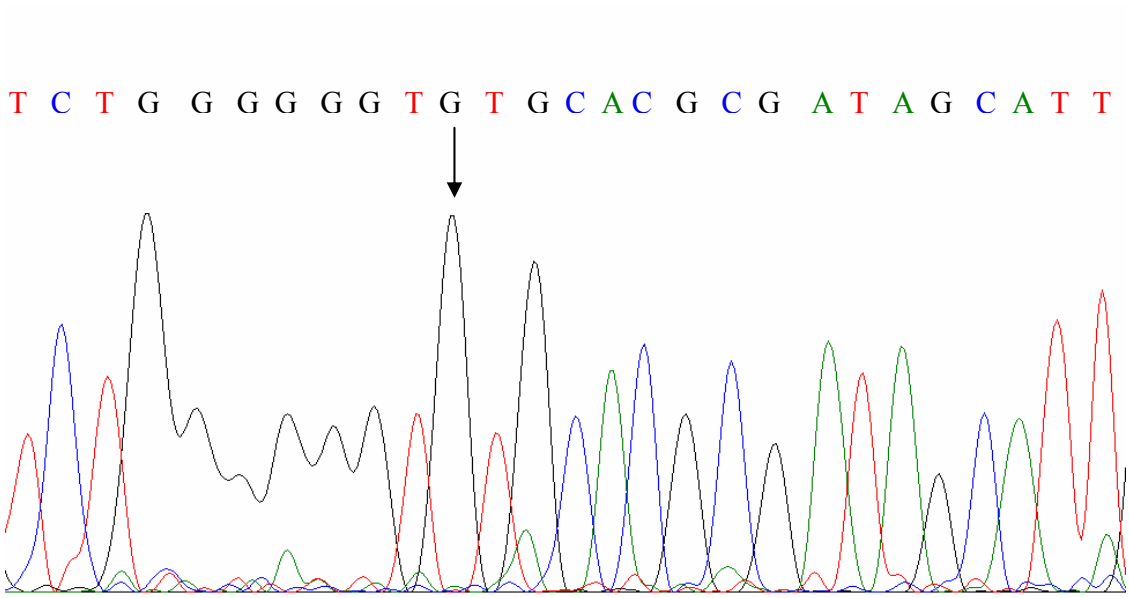


Fig. 3.16. Transition (A → G) at position 73 HVR-II.

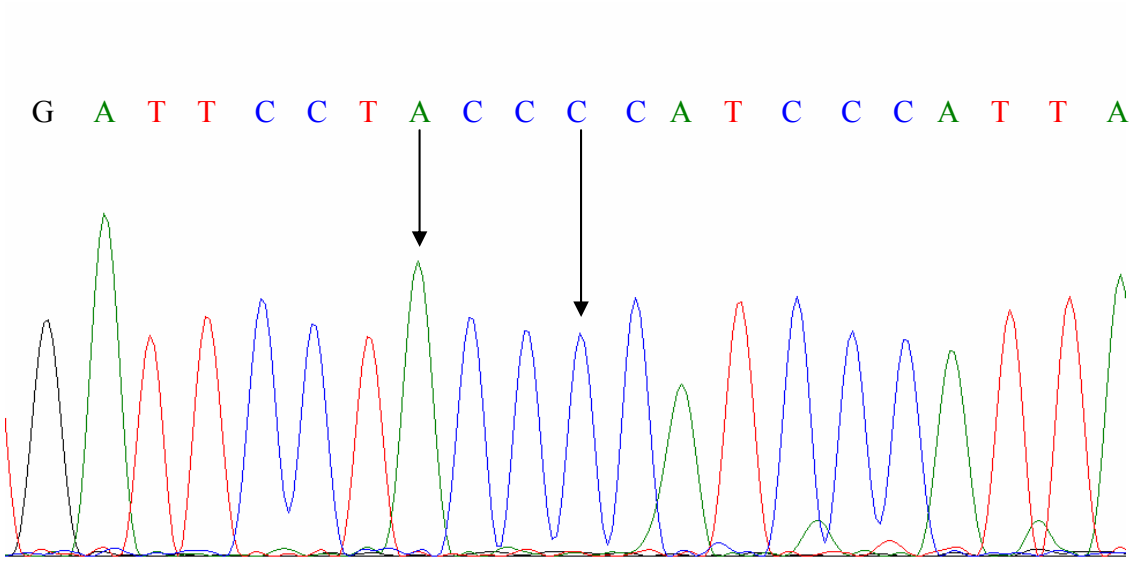


Fig. 3.17. Transition (G → A) at position 143, and transition (T to C) at position 146 HVR-II.

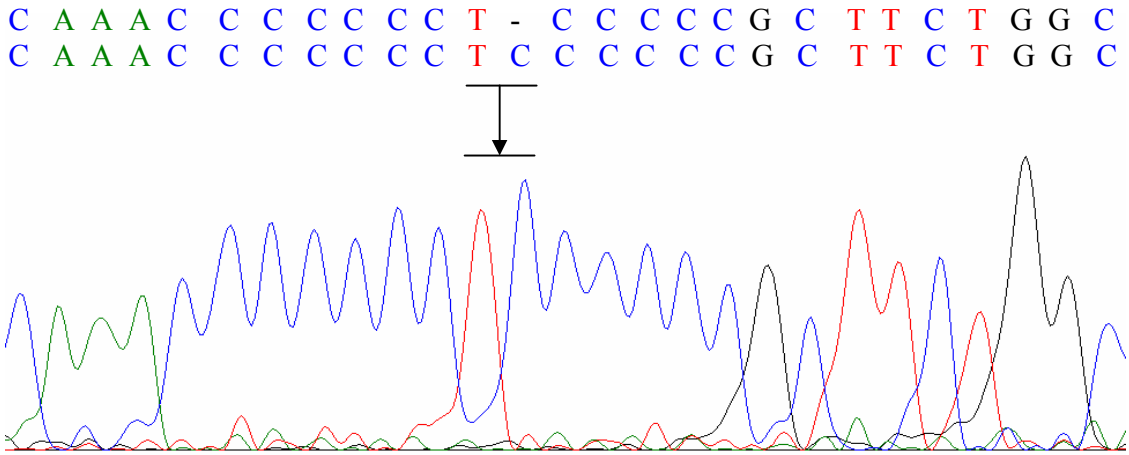


Fig. 3.18. Insertion T → TC at position 310 HVR-II, the upper sequence shows the ancestral and the lower is the derived.

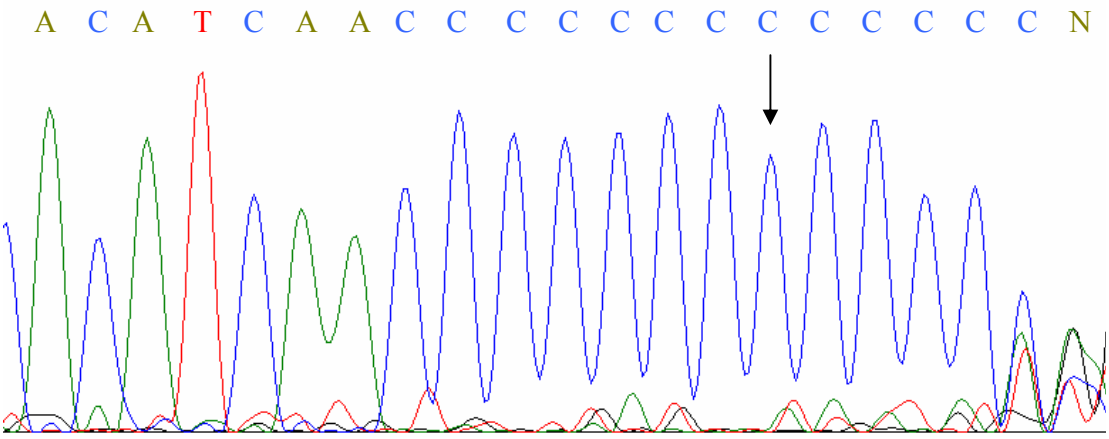


Fig. 3.19. mtDNA Cytosine stretch at position 16,188 of HVR-II due to transition T to C.

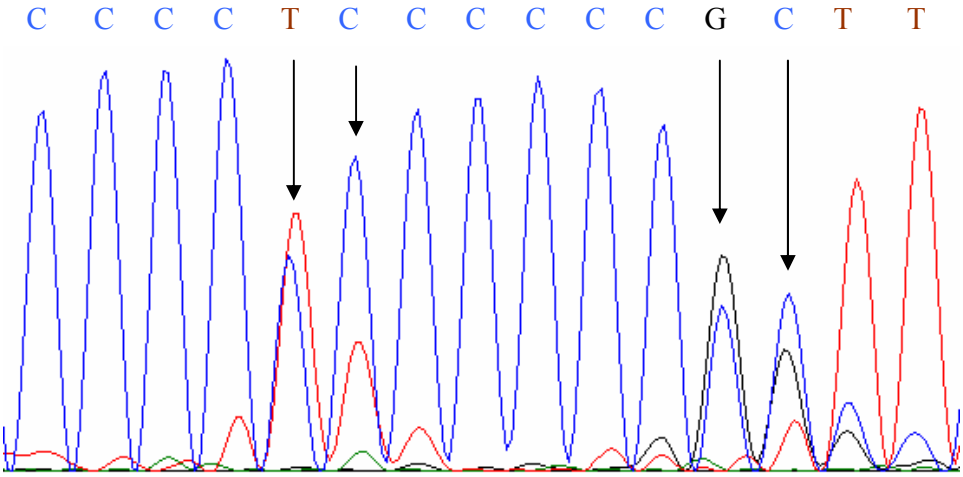


Fig. 3.20. Heteroplasmy in HVR-I, the arrows show the heteroplasmic sites.

Table 3.2. Frequencies (N) of the Y-chromosome M78 Subclades in 15 Sudanese Populations

Populations	**M78**	**M78***	**V12***	**V32**	**V13**	**V22**	**V65**
1. Dinka	4	–	1	–	–	3	–
2. Shilluk	2	–	–	–	–	2	–
3. Nuer	2	–	2	–	–	–	–
4. Borgu	4	–	–	3	–	1	–
5. Nuba	7	1	1	3	–	2	–
6. Masalit	23	1	–	17	–	5	–
7. Fur	19	–	–	13	–	6	–
8. Nubians	6	–	5	1	–	–	–
9. Fulani	9	–	–	1	–	8	–
10. Hausa	1	–	–	1	–	–	–
11. Copts	5	–	5	–	–	–	–
12. Beja	15	–	2	13	–	–	–
13. Gaalien	9	–	3	3	–	3	–
14. Meseria	4	–	1	3	–	–	–
15. Arakien	4	–	2	1	–	1	–
Total N (%)	**114**	**2 (1.8)**	**22 (19.3)**	**59 (51.8)**	**–**	**31(27.2)**	**–**

Table 3.3. mtDNA Haplotypes Frequencies in 15 Sudanese Populations.

HP	Dinka	Shilluk	Nuer	Borgu	Masalit	Fur	Nubians	Nuba	Fulani	Arakien	Beja	Copts	Gaalien	Hausa	Meseria
L0a	3 (10.3)	-	-	-	-	-	-	6 (18.2)	-	-	-	-	-	1 (2.9)	-
L0a1	-	-	-	1 (4.2)	6 (14.6)	4 (15.3)	5 (17.2)	3 (9.0)	-	-	1 (2.1)	-	-	1 (2.9)	2 (13.3)
L0a1a	-	-	-	-	2 (4.9)	2 (7.7)	-	-	-	-	-	-	-	-	-
L0a2	1 (3.4)	-	-	1 (4.2)	-	-	1 (3.4)	1 (3.0)	-	-	-	-	-	-	-
L0d	-	-	2 (12.5)	2 (8.3)	-	-	-	1 (3.0)	-	-	-	-	-	-	-
L0f	-	-	-	-	-	-	-	1 (3.0)	-	-	2 (4.2)	-	-	-	-
L1*	1 (3.4)	-	-	-	1 (2.4)	-	-	-	-	-	-	-	-	-	-
L1b	-	1 (6.3)	-	-	-	-	-	-	1 (3.6)	-	-	-	-	2 (5.7)	-
L1b1	-	2 (12.5)	-	-	-	-	2 (6.9)	1 (3.0)	8 (28.6)	-	-	-	-	1 (2.9)	-
L1c	-	-	-	1 (4.2)	5 (12.2)	3 (11.5)	-	3 (9.1)	-	-	4 (8.3)	-	-	1 (2.9)	-
L1c1	-	-	-	-	-	-	-	-	-	-	-	-	-	1 (2.9)	1 (6.7)
L1c1a1	1 (3.4)	-	-	-	-	1 (3.8)	-	-	-	-	-	2 (6.9)	1 (4.5)	-	-
L1c3	-	-	-	1 (4.2)	1 (2.4)	-	-	2 (6.1)	-	-	-	-	-	-	-
L2*	-	-	-	-	-	-	-	-	-	-	2 (4.2)	-	-	-	-
L2a	-	-	-	3 (12.5)	3 (7.3)	-	2 (6.9)	2 (6.1)	-	-	-	-	2 (9.0)	-	2 (13.3)
L2a1	1 (3.4)	-	2 (12.5)	1 (4.2)	3 (7.3)	2 (7.7)	-	3 (9.1)	-	-	1 (2.1)	-	-	3 (8.6)	-
L2a1a	-	-	-	-	-	-	1 (3.4)	-	-	-	-	-	-	-	-
L2b	-	1 (6.3)	-	1 (4.2)	1 (2.4)	-	-	3 (9.1)	-	-	-	-	-	1 (2.9)	-

Continued

HP	Dinka	Shilluk	Nuer	Borgu	Masalit	Fur	Nubians	Nuba	Fulani	Arakien	Beja	Copts	Gaalien	Hausa	Meseria
L2b1	-	-	-	1 (4.2)	2 (4.9)	1 (3.8)	-	-	1 (3.6)	-	-	-	-	2 (5.7)	1 (6.7)
L2c	-	-	-	-	-	-	-	-	3 (10.7)	-	-	-	-	-	-
L2c1	1 (3.4)	-	-	1 (4.2)	-	-	-	-	-	-	-	-	-	-	-
L2c2	-	-	-	1 (4.2)	1 (2.4)	-	-	-	-	-	-	-	-	-	-
L2d1	2 (6.9)	1 (6.3)	6 (37.5)	-	-	1 (3.8)	-	1 (3.0)	-	9 (52.9)	5 (10.4)	-	-	3 (8.6)	2 (13.3)
L2d2	1 (3.4)	-	-	-	-	-	-	1 (3.0)	1 (3.6)	-	-	-	-	1 (2.9)	-
L3*	-	-	-	-	-	-	4 (13.8)	-	-	-	-	-	-	-	-
L3b	-	1 (6.3)	-	1 (4.2)	2 (4.9)	1 (3.8)	-	-	3 (10.7)	-	-	-	-	1 (2.9)	2 (13.3)
L3b1	-	-	-	2 (8.3)	3 (7.3)	1 (3.8)	1 (3.4)	-	5 (17..9)	-	2 (4.2)	-	-	3 (8.6)	-
L3b2	-	2 (12.5)	-	-	-	-	-	-	1 (3.6)	-	1 (2.1)	-	-	-	3 (20)
L3d	-	-	-	1 (4.2)	3 (7.3)	-	-	-	-	-	-	-	-	1 (2.9)	-
L3d1	-	-	-	-	-	-	-	-	-	1 (5.9)	-	-	-	-	-
L3d2	-	-	-	-	-	-	-	-	-	1 (5.9)	-	-	2 (9.0)	-	-
L3d3	-	-	-	1 (4.2)	-	-	-	-	-	-	-	-	-	-	-
L3e1a	-	-	-	-	-	-	-	-	-	-	-	-	-	1 (2.9)	-
L3e2	-	-	-	-	-	-	-	-	-	-	-	-	-	1 (2.9)	1 (6.7)
L3e2b	-	1 (6.3)	-	-	1 (2.4)	1 (3.8)	1 (3.4)	-	-	-	2 (4.2)	-	-	4 (11.4)	-
L3e5	-	-	-	1 (4.2)	2 (4.9)	-	-	-	-	-	-	-	1 (4.5)	1 (2.9)	-
L3f	5 (17.2)	-	-	1 (4.2)	1 (2.4)	-	1 (3.4)	-	-	2 (11.8)	11 (23)	-	10 (45.4)	1 (2.9)	-
L3f1	-	1 (6.3)	2 (12.5)	-	1 (2.4)	1 (3.8)	1 (3.4)	1 (3.0)	2 (7.2)	-	-	-	2 (9.0)	2 (5.7)	-

Continued

HP	Dinka	Shilluk	Nuer	Borgu	Masalit	Fur	Nubians	Nuba	Fulani	Arakien	Beja	Copts	Gaalien	Hausa	Meseria
L3h	-	1 (6.3)	-	-	-	-	-	-	-	-	-	-	-	1 (2.9)	-
L3i	1 (3.4)	-	-	-	-	-	-	-	-	-	-	-	-	-	-
L3w	1 (3.4)	-	-	-	-	-	1 (3.4)	-	-	-	1 (2.1)	-	-	-	-
L3x2	-	-	-	-	1 (2.4)	-	-	-	-	-	-	-	-	-	-
L4a	-	-	-	1 (4.2)	-	-	-	1 (3.0)	-	-	-	-	-	-	-
L4g	2 (6.9)	3 (18.8)	-	2 (8.3)	2 (4.9)	2 (7.7)	2 (6.9)	1 (3.0)	-	-	1 (2.1)	-	-	-	-
L5a	9 (31)	2 (12.5)	4 (25)	1 (4.2)	-	1 (3.8)	2 (6.9)	2 (6.1)	-	1 (5.9)	1 (2.1)	-	2 (9.0)	-	-
M/D	-	-	-	-	-	-	1 (3.4)	-	-	-	-	-	-	-	-
M1	-	-	-	-	-	-	1 (3.4)	-	-	-	1 (2.1)	3 (10.3)	-	-	-
M1a	-	-	-	-	-	-	-	-	-	-	1 (2.1)	1 (3.4)	-	-	-
N/J1a	-	-	-	-	-	-	-	-	-	-	-	3 (10.3)	-	-	-
N/J1b	-	-	-	-	-	-	-	-	3 (10.7)	1 (5.9)	1 (2.1)	-	-	1 (2.9)	1 (6.7)
N/J2	-	-	-	-	-	-	-	-	-	-	-	3 (10.3)	-	-	-
N1a	-	-	-	-	-	-	1 (3.4)	-	-	1 (5.9)	-	-	-	-	-
preHV1	-	-	-	-	-	-	1 (3.4)	-	-	1 (5.9)	4 (8.3)	4 (13.8)	2 (9.0)	-	-
R/T1	-	-	-	-	-	-	-	-	-	-	1 (2.1)	5 (17.2)	-	-	-
R/U3	-	-	-	-	-	-	-	-	-	-	1 (2.1)	-	-	-	-
R/U6a1	-	-	-	-	-	-	1 (3.4)	-	-	-	5 (10.4)	8 (27.6)	-	1 (2.9)	-

all belonging to the major sub-Saharan and Eurasian mitochondrial macrohapolgroups L0, L1, L2, L4, L5, L3A, M and N in frequencies of 12.1, 11.9, 22, 4.2, 6.2, 29.5, 2, and 12.2% respectively (Fig. 3.21). The frequencies of the mtDNA haplotypes and their subclades are shown in Fig (3.22). The frequency of the eastern Africa haplogroups among Sudanese is 35.6%, these haplotypes include L0a, L0f, L5, L3*, L4, L3f, L3h, L3i, L3x and L3w. While the frequencies of west/central Africa (L1b, L2b, L2c, L2d, L3b, L3d and L3e), northern Africa (U6 and M1), southern Africa (L0d), and Eurasian (D, J1, J2, N1a, preHV1, T1, and U3) haplotypes among Sudanese are 49, 5.5, 1.2 and 8.7% respectively. The frequency of the sub-Saharan Africa haplotypes among Arabs including Gaalien, Meseria and Arakien is 88.9%, while the frequency of the Eurasian haplotyes among them is 11.1%, the northern Africa haplogroups U6 and M1 are completely absent in this group. Haplogroup L6 was not observed in the sample analyzed.

3.2.5 Sub-Saharan African Macrohaplogroup L Lineages

3.2.5.1 Macrohaplogroup L0

Haplogroup L0 (Table 3.3) is represented in Sudanese samples primarily by its sub-clades L0a, L0d, and L0f. Haplogroup L0a and its daughter clades L0a1, L0a1a, and L0a2, which accounts for the majority of L0a lineages in Sudanese. Haplogroup L0a1a is restricted to Fur and Masalit, while haplogroup L0f has the lowest frequencies of haplogroup L0. Haplogroup L0d which is defined by 16129 and 16243 substitutions, is the most prominent haplogroup among southern African Khoisan speakers (Vigilant *et al.,* 1991; Bandelt and Forster, 1997; Behar *et al.,* 2008), and uncommon in East Africa (Watson *et al.*, 1997; Tishkoff *et al.*, 2007); five Sudanese samples found to be related to

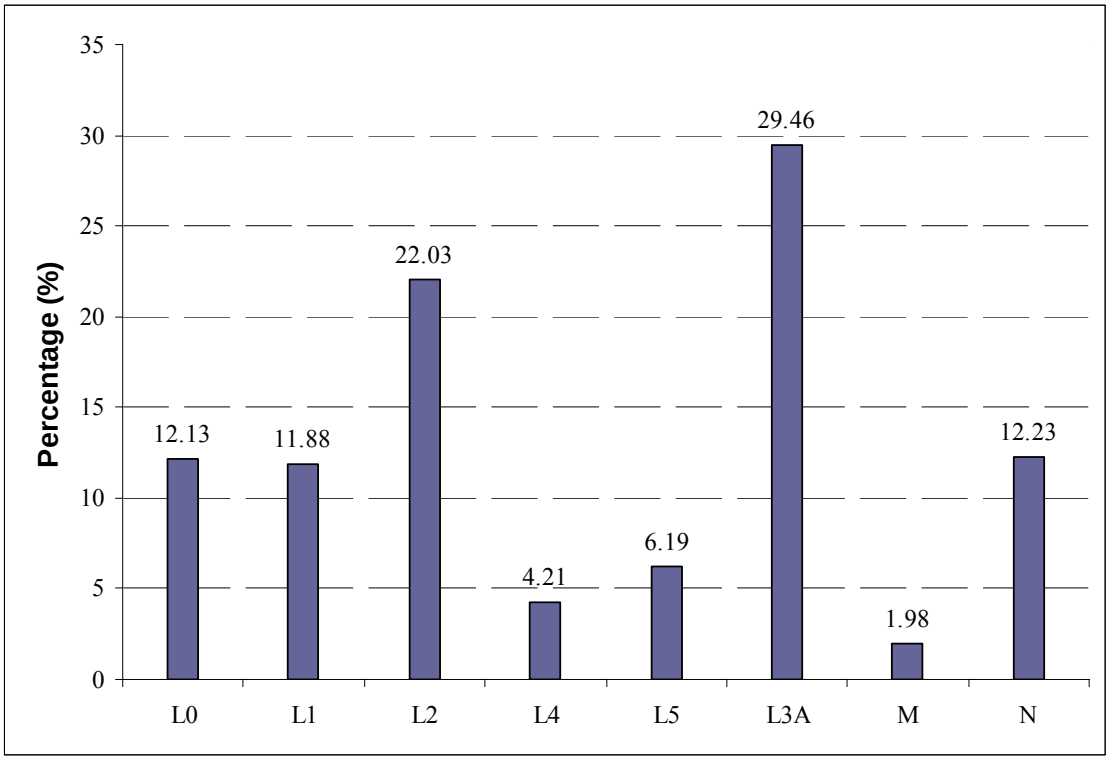


Fig. 3.21. The frequencies of the main mtDNA haplogroups in Sudanese populations.

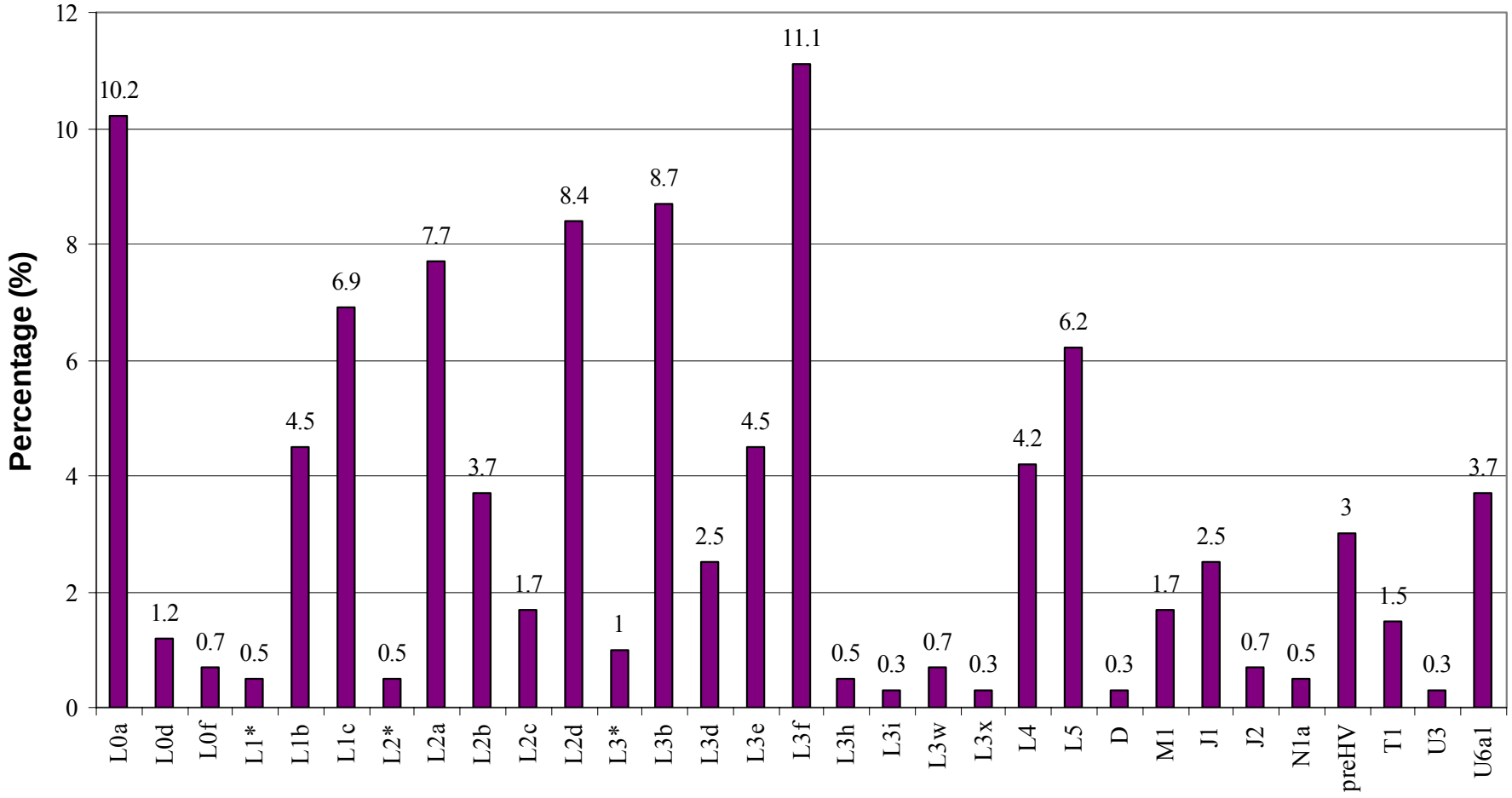


Fig. 3.22. The frequencies of different mtDNA haplotypes in Sudanese populations (haplotypes sub-clades were not shown in the chart)

L0d (Table 3.3). L0f which is rare and appears to be geographically confined to east Africa has the lowest frequencies of L0 in Sudanese (Fig. 3.22). Haplogroup L0k which has been found exclusively among Khoisan speaking populations is completely absent in Sudanese samples.

3.2.5.2 Macrohaplogroup L1

Haplogroup L1, which is more frequent and diverse in west and central Africa than in East Africa (Salas *et al.*, 2002) is represented in Sudanese particularly in people residing in western Sudan including Fulani, Fur, Hausa, and Masalit by its sub-clades L1*, L1b (and its daughter clade L1b1) and L1c (and its daughter clades L1c1, L1c1a and L1c3). L1c has the highest frequencies of L1 sub-clades. All Sudanese L1b1 lineages share the 16287 and 16293 substitutions and associated with relatively low downstream variation.

3.2.5.3 Macrohaplogroup L2

All Sudanese L2 lineages can be seen as derived from L2*, L2a including two further minor sub-clades L2a1 and L2a1a; L2b including L2b1, L2c including L2c1 and L2c2, and L2d including L2d1 and L2d2. Haplogroup L2* was found in two Beja samples. All 16 Sudanese L2a1 sequences share a mutation at position 16309, 10 of them share a mutation at position 16189, L2a1a defined by a substitution at position 16286 (Salas *et al.*, 2002) and occurs at its highest frequency in southern Africa (Pereira *et al.*, 2001) was found in a single Nubian sample. Out of 15 Sudanese L2b sequences which is predominantly West African clade, a single Masalit sample share a transition at position 16145 with Ethiopian L2b clade. L2c which is defined by HVR-II mutation at position 325 was found only in Fulani, and its sub-clades L2c1 and L2c2 are observed in

Sudanese including Borgu, Dinka and Masalit. L2d which appear to be largely confined to west and western central Africa occurs in most Sudanese populations by its sub-clades L2d1 and L2d2. L2d1 has the highest frequencies of L2d compared to L2d2 (Table 3.3).

3.2.5.4 Macrohaplogroup L3

Nearly thirty percent (29.5%) of Sudanese L lineages are captured by diversed sub-clades of L3A. Haplogroup L3* which is defined by substitution at positions 16223 and 16278 was observed only in Nubians. Haplogroup L3f, which is defined by a combination of transitions at position 16209 (Salas *et al.*, 2002), and as expected is the highest frequent mitochondrial haplotype in Sudanese, particularly in Beja, Dinka and Gaalien. L3f1 which is defined by substitution at position 16292 was found in 13 Sudanese. L3b and L3e lineages are widely spread in West African populations and Bantu-speaking southeastern Africans (Rosa *et al.*, 2004), L3f2 which has been described for the first time by Cerny *et al.* (2007) and most probably autochthonous in Chad Basin was not observed in the population sample analyzed. L3b was found mainly in population of western Sudan, Borgu, Fulani, Fur, Masalit and Meseria, in addition to a single Shilluk sample. The occurrence of the main sub-clades L3b1 and L3b2 was in west Sudan, L3b1 is also present in Beja and Nubian, and L3b2 in Beja and Shilluk in low frequencies. Eighteen Sudanese samples belonged to L3e sub-clades, L3e1a, L3e2, L3e2b, and L3e5. The frequency of L3e2b was the highest is Hausa (11.4%). Haplogroup L3d, which is also a predominantly West African (Salas *et al.*, 2002), was found in 10 Sudanese, sub-clade L3d1 was found in a single Arakien sample, L3d2 was found only in Arakien and Gaalien, and L3d3 was found in a single Borgu sample.

Haplogroups L3i, L3x, and L3w which have been described by Kivisild *et al.* (2004), were observed in Sudanese at low frequencies. Haplogroup L3i is characterized by transition at positions 16129, 16153, 16223 and 16319, was observed in a single Dinka sample sharing the mutation at position 16319 with all L3i Ethiopian samples. Haplogroup L3x2 subclade is defined by transitions at positions 16169, 16223, 16193, and 16195 was found in a single Masalit sample. Haplogroup L3w occurred in three Sudanese samples. This clade is defined by substitutions at positions 16260, 16223, and 16260. Haplogroup L3h which is defined by HVR-I motif at positions 16129, 16223, 16256A, 16311, and 16362, was first reported at a moderate frequency in Guinea-Bissau populations of West Africa (Rosa *et al.*, 2004). In the present study, two different L3h haplotypes were found in Sudanese, one sample shared transversions at positions 16129, 16223, 16256A, and 16311 with West Africans, the other related sequence shares the 16223, and 16311 with Yemeni L3h haplotype substitutions, but lacks 16165, and 16192 substitutions (Krings *et al.*, 1999b).

Haplogroup L4 (previously named "L3g") whose descendants share the ancestral states at position 16362, is a combination of three control region markers splits haplogroup L4 into two major subclades haplogroup L4a and L4g. Haplogroup L4a is defined by substitutions at position 16260 and occurs in two Sudanese, haplogroup L4g which is defined by substitutions at positions 16293T and 16355 was observed in 15 samples and it is widely spread among Nilo-Saharan speaking groups. Both L4a and L4g reveal high haplotype and sequence diversity in Sudanese. Haplogroup L5a (previously named "L1e) which is defined by substitutions at positions 16129, 16148 and 16166, and found to be restricted to East African (Salas *et al.*, 2002) was found in 25 Sudanese, the

geographic spread of the L5a is more among southern Sudanese population including Dinka, Nuer, and Shilluk.

Haplogroup M1 lineages which are defined by substitutions at positions 16189, 16223, 16249, and 16311, constitute (1.7%) of the Sudanese mtDNA sequences, one subclade, M1a which can be distinguished by a transition at position 16359 is also present in only two Sudanese. The higher frequency of M1 haplogroup found to be in Copts (13.7 %). Also haplogroup D which is characterized by substitutions at positions 16223 and 16362 was found in a single Nubian sample.

Lineages that belong to haplogroup N and cover virtually all mtDNA sequences in western Eurasia (Richards *et al.*, 2000) were showed substantial frequencies (12.2%) in the Sudanese mtDNA pools. Ten of the 13 haplogroup J lineages in Sudanese belong to haplogroup J1, subclade J1a which is defined by substitutions at positions 16145, 16231, and 16261 is restricted to Copts (10.3%), subclade J1b which is characterized by substitutions at positions 16145, 16222 and 16261 was found with remarkable amount in Fulani (10.7%). The other three Sudanese J sequences, present in Copts, belonged to a subclade of J2 that is defined by a substitution at position 16193. Most of the Sudanese J sequences, share the combination of 16145 and 16261 mutations in haplogroup J1b, which is a common motif of J lineages in populations from the Near East and all over western Eurasia (Richards *et al.*, 2000).

Haplogroup N1a was observed at marginal frequencies in Sudanese. It occurs in two samples, both samples share the substitutions at positions 16147G, 16172, 16248, and 16355, with published N1a sequence from Egypts (Krings *et al.*, 1999b). All Sudanese haplogroup T sequences clustered with T1 subclade, which is defined by

substitutions at positions 16163, 16189 and 16294. T1 haplogroup constitute (17.2%) in Copts and (2.1%) in Beja.

Haplogroup U is by far the most frequent (6.7%) subclade in the Sudanese N cluster. And all U sequences clustered with either U3 or U6a1 subclades. Subclade U3 which is defined by substitution at position 16343 was observed in a single Beja sample. U6a1 occurs in remarkable frequencies in Copts (27.6%) and Beja (10.4%) and in low frequencies in Hausa (2.9%) and Nubians (3.4%). All Sudanese U6a1 sequences shares the substitutions at positions 16172, 16189 and 16287.

Haplogroup preHV1 which is defined by substitution at positions 16126 and 16362, was observed in 12 Sudanese the highest frequency found in Copts (13.8%), and Beja (8.3%). Sudanese preHV1 lineages match founder haplotypes common to North African populations (Krings *et al.*, 1999b), the only single Nubian samples that belong to this haplogroup share the substitution at position 16304 with an Ethiopian sample. Haplogroup HV1 which has been reported previously in populations of the Arabian Peninsula (Di Rienzo and Wilson, 1991; Richards *et al.*, 2000), southern Egypt, northern Sudan (Krings *et al.*, 1999b), and Ethiopia (Kivisild *et al*., 2004), was not observed in the sample analyzed.

3.2.6 Pairwise Genetic Distances

Table (3.4) and Table (3.5) contain genetic distances of Y-chromosome and mtDNA variations respectively. Y-chromosome genetic distances were significantly greater than zero for 86 of the 105 pairwise comparisons (81.9%). Only 19 of the 105 pairwise comparisons (18.1%) were not significant. Most of the non significant genetic distances occur between populations of the same linguistic group, only 4 of the 19 were found to be between populations of different linguistic affiliation (21.1%). For mtDNA,

Table 3.4. Y-chromosome Pairwise Genetic Distances of Different 15 Sudanese Populations

List of labels for population samples used in this table: 1: Arakien, 2: Beja, 3: Borgu, 4: Copts, 5: Dinka, 6: Fulani, 7: Fur, 8: Gaalien, 9: Hausa, 10: Masalit, 11: Meseria, 12: Nuer, 13: Nuba, 14: Nubians, and 15: Shilluk.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
[1]	0.0000														
[2]	0.1368	0.0000													
[3]	0.3727	0.1369	0.0000												
[4]	0.0265*	0.0662	0.2325	0.0000											
[5]	0.4868	0.3252	0.1602	0.3370	0.0000										
[6]	0.2834	0.1295	0.2198	0.1500	0.4004	0.0000									
[7]	0.4155	0.1393	0.1045	0.2816	0.2043	0.2460	0.0000								
[8]	0.0109*	0.0906	0.2591	0.0071*	0.3641	0.1492	0.2991	0.0000							
[9]	0.2326	0.1457	0.1360	0.1018	0.2360	0.0757	0.2058	0.1328	0.0000						
[10]	0.4668	0.1522	0.1808	0.3221	0.3353	0.2693	-0.0006*	0.3417	0.2508	0.0000					
[11]	0.0174*	0.1446	0.3209	0.0202*	0.4413	0.1831	0.3827	-0.0098*	0.1388	0.4296	0.0000				
[12]	0.4561	0.2815	0.1898	0.2518	0.0624*	0.3529	0.2232	0.3294	0.1660	0.3239	0.3990	0.0000			
[13]	0.4158	0.2008	0.0332*	0.2706	0.0221*	0.2903	0.0670*	0.3033	0.1769	0.1707	0.3777	0.0724*	0.0000		
[14]	0.0081*	0.0615	0.2422	-0.0158*	0.3613	0.1560	0.2907	-0.0070*	0.1308	0.3295	0.0069*	0.2969	0.2891	0.0000	
[15]	0.4539	0.2718	0.1041	0.2817	-0.0449*	0.3533	0.1661	0.3251	0.1812	0.2956	0.4011	-0.0011*	-0.0150*	0.3114	0.0000

*Not significant at multiple tests adjusted *P*-value.

Markov chain length: 10,000 steps.

Table 3.5. mtDNA Pairwise Genetic Distances of Different 15 Sudanese Populations

List of labels for population samples used in this table:

1: Arakien, 2: Beja, 3: Borgu, 4: Copts, 5: Dinka, 6: Fulani, 7: Fur, 8: Gaalien, 9: Hausa, 10: Masalit, 11: Meseria, 12: Nuer, 13: Nuba, 14: Nubians, and 15: Shilluk.

	[1	2	3	4	5	6	7	8	9	10	11	12	13	14	15]
[1]	0.0000														
[2]	0.0742	0.0000													
[3]	0.0762	0.0333	0.0000												
[4]	0.2898	0.1407	0.2296	0.0000											
[5]	0.0957	0.0475	0.0186*	0.2570	0.0000										
[6]	0.1593	0.0643	0.0538	0.2713	0.1099	0.0000									
[7]	0.1360	0.0547	-0.0013*	0.2546	0.0494	0.0539	0.0000								
[8]	0.1876	0.0594	0.0941	0.3267	0.1001	0.1477	0.1306	0.0000							
[9]	0.0749	0.0236	0.0154*	0.2341	0.0593	0.0413	0.0282*	0.0819	0.0000						
[10]	0.1088	0.0444	-0.0168*	0.2416	0.0404	0.0657	-0.0133*	0.1073	0.0179*	0.0000					
[11]	0.0827	0.0401	-0.0086*	0.2598	0.0662	0.0337*	-0.0112*	0.1295	0.0022*	-0.0010*	0.0000				
[12]	0.0109*	0.0854	0.0497	0.3114	0.0295*	0.1598	0.1069	0.1903	0.0732	0.0865	0.0825	0.0000			
[13]	0.1302	0.0948	0.0187*	0.2586	0.0461	0.1365	0.0369	0.2016	0.0830	0.0190*	0.0649	0.0738	0.0000		
[14]	0.1153	0.0256	0.0006*	0.1997	0.0174*	0.0741	0.0062*	0.0979	0.0316	0.0001*	0.0237*	0.0889	0.0311	0.0000	
[15]	0.1017	0.0248*	0.0044*	0.2253	0.0229*	0.0171*	0.0258*	0.0974	0.0099*	0.0321*	0.0165*	0.0767	0.0788	0.0103*	0.0000

*Not significant at multiple tests adjusted *P*-value.

Markov chain length: 10,000 steps.

76 of 105 pairwise comparisons (73.4%) were significantly greater than zero, 29 of the total pairwise comparisons were not significant, and 13 of them were found between populations of different linguistic affiliation (44.8%).

Also the genetic distances of Y-chromosome and mtDNA of Sudanese populations compared to some other African and Asian populations are shown in Table (3.6) and Table (3.7) respectively.

3.2.7 Analysis of Molecular Variance (AMOVA)

The overall Y-chromosome F_{ST} for the 15 populations is 0.24 (Table 3.8), a value that is similar to that found when populations are grouped into five geographic regions, and less than when populations are grouped into three linguistic families (F_{ST} = 0.25). When populations are grouped according to geographic regions, the proportion of among-group variance (F_{CT} =0.13) is higher than when populations are grouped according to linguistic affiliation (F_{CT} = 0.11).

AMOVA results for the mtDNA data are also presented in Table (3.9). The overall F_{ST} is 0.09, which is very similar when populations are placed in either linguistic or geographic groups. When populations are grouped according to linguistic families, the proportion of among group variance (F_{CT} =0.02) is higher than when populations are grouped according to geographic regions (F_{CT} = 0.01).

3.2.8 Pairwise Differences and Nucleotide Diversity

Mismatch distributions were computed for the Sudanese populations after pooling of the populations whose sample size is < 20 into larger groups. Unimodal distributions were observed in all Sudanese populations except Fulani which they exhibited multimodal distribution, and share a higher frequency of the low differences classes 0-3

Table 3.6 Y-chromosome Pairwise Genetic Distances of Different 15 Sudanese Populations Compared with some other African Populations and Turks.

List of labels for population samples used in this table: 1: Dinka, 2: Shilluk, 3: Nuer, 4: Borgu, 5: Nuba, 6: Masalit, 7: Fur, 8: Nubians, 9: Fulani, 10: Hausa, 11: Copts, 12: Beja, 13: Gaalien, 14: Meseria, 15: Arakien, 16: Turks, 17: Oromo, 18: Amhara, 19: Senegalese.

[	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19]
[1]	0.00000																		
[2]	0.04355	0.00000																	
[3]	0.05407	0.01009	0.00000																
[4]	0.15112	0.08567	0.12673	0.00000															
[5]	0.02678	0.01680	0.04157	0.03358	0.00000														
[6]	0.35133	0.30225	0.29353	0.17697	0.17849	0.00000													
[7]	0.20988	0.16219	0.17767	0.10512	0.06828	0.00035	0.00000												
[8]	0.37101	0.31411	0.27238	0.22555	0.29691	0.33151	0.29274	0.00000											
[9]	0.43851	0.38126	0.35230	0.24225	0.35454	0.34867	0.32673	0.21597	0.00000										
[10]	0.28710	0.22823	0.19693	0.14927	0.23811	0.31712	0.27316	0.16004	0.07262	0.00000									
[11]	0.34715	0.28565	0.23545	0.20853	0.27949	0.32077	0.28177	0.01924	0.17779	0.11773	0.00000								
[12]	0.31852	0.25569	0.22824	0.12611	0.19625	0.14837	0.13602	0.05484	0.22178	0.19257	0.06153	0.00000							
[13]	0.35667	0.30941	0.28193	0.23398	0.30059	0.33547	0.29545	0.00550	0.17542	0.13444	0.00644	0.08972	0.00000						
[14]	0.43597	0.38345	0.35320	0.28462	0.37606	0.41916	0.37638	0.01637	0.17143	0.12202	0.00687	0.14525	0.01054	0.00000					
[15]	0.49591	0.45222	0.42824	0.34265	0.42104	0.46320	0.41258	0.00439	0.30951	0.24539	0.01588	0.12549	0.01426	0.03004	0.00000				
[16]	0.40671	0.36043	0.33529	0.27316	0.36378	0.41095	0.37923	0.08618	0.13629	0.11072	0.06300	0.19697	0.05245	0.02547	0.12127	0.00000			
[17]	0.28013	0.21225	0.17814	0.07831	0.12727	0.08954	0.08665	0.24670	0.30348	0.25873	0.24262	0.09866	0.28196	0.34895	0.36268	0.35464	0.00000		
[18]	0.23293	0.17288	0.15740	0.09036	0.12916	0.15225	0.11310	0.06497	0.24029	0.18719	0.07203	0.00055	0.09677	0.16167	0.13768	0.20776	0.08085	0.00000	
[19]	0.71171	0.69543	0.69204	0.53781	0.60728	0.59403	0.58196	0.62886	0.66649	0.57106	0.62957	0.52669	0.62615	0.70279	0.73895	0.63579	0.39826	0.49965	0.0000

Table 3.7. mtDNA Pairwise Genetic Distances of Different 15 Sudanese Populations Compared with some other African Populations and Saudi.

List of labels for population samples used in this table: 1: Arakien, 2: Beja, 3: Borgu, 4: Copts, 5: Dinka, 6: Fulani, 7: Fur, 8: Gaalien, 9: Hausa, 10: Masalit, 11: Meseria, 12: Nuer, 13: Nubians, 14: Nuba, 15: Shilluk, 16: Saudi, 17: Guinea, 18: Egypt, 19: Chad, 20: Khoisan, 21: Oromo, 22: Amhara, 23: Berber.

[	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23]
[1]	0.00000																						
[2]	0.10994	0.00000																					
[3]	0.12799	0.04535	0.00000																				
[4]	0.30875	0.11794	0.23654	0.00000																			
[5]	0.13753	0.02796	0.02445	0.23627	0.00000																		
[6]	0.18282	0.05720	0.04535	0.22805	0.09330	0.00000																	
[7]	0.15134	0.06447	0.00080	0.25422	0.05046	0.05511	0.00000																
[8]	0.23577	0.03897	0.11546	0.23019	0.06882	0.14750	0.15175	0.00000															
[9]	0.09305	0.02051	0.00499	0.20255	0.03795	0.03181	0.02050	0.09127	0.00000														
[10]	0.14758	0.05262	0.01612	0.23735	0.04508	0.04289	0.02205	0.12239	0.00782	0.00000													
[11]	0.08780	0.03787	0.00547	0.23451	0.05413	0.02728	0.00742	0.14954	0.00601	0.00006	0.00000												
[12]	0.00213	0.08921	0.04798	0.31963	0.06780	0.14300	0.07168	0.19560	0.04816	0.07583	0.04319	0.00000											
[13]	0.17104	0.01541	0.00900	0.17854	0.01321	0.06706	0.01245	0.08327	0.02310	0.00999	0.02920	0.10446	0.00000										
[14]	0.17078	0.11065	0.01647	0.31311	0.08410	0.10488	0.00581	0.21377	0.05559	0.00958	0.04581	0.07343	0.06333	0.00000									
[15]	0.15486	0.02108	0.01767	0.21426	0.01434	0.02666	0.03819	0.08912	0.00951	0.03332	0.01418	0.09247	0.01399	0.10549	0.00000								
[16]	0.35958	0.18309	0.31614	0.04464	0.29932	0.29877	0.35031	0.27590	0.28201	0.31975	0.30987	0.38704	0.24644	0.38959	0.28945	0.00000							
[17]	0.16697	0.09481	0.01050	0.29936	0.09379	0.08075	0.06737	0.18495	0.03519	0.03003	0.04113	0.09970	0.08342	0.05717	0.08956	0.35063	0.00000						
[18]	0.25443	0.06836	0.15958	0.06262	0.14383	0.17761	0.19165	0.16072	0.14060	0.16787	0.16657	0.24854	0.09177	0.24302	0.13969	0.06728	0.22444	0.00000					
[19]	0.14694	0.03202	0.00978	0.19308	0.04604	0.04226	0.04533	0.07343	0.00132	0.01679	0.01820	0.09214	0.01941	0.07828	0.02122	0.24174	0.03906	0.12312	0.00000				
[20]	0.34697	0.22896	0.15023	0.42064	0.23003	0.24799	0.13182	0.35838	0.19135	0.14524	0.20750	0.23480	0.19105	0.11276	0.26058	0.46496	0.22943	0.35325	0.19919	0.00000			
[21]	0.19705	0.02914	0.09657	0.08582	0.08334	0.13121	0.13201	0.08230	0.08754	0.10986	0.10792	0.17199	0.04288	0.18284	0.07890	0.13186	0.16461	0.01395	0.07403	0.29703	0.00000		
[22]	0.16903	0.02717	0.06236	0.10379	0.05853	0.11271	0.09852	0.08407	0.06864	0.07818	0.07783	0.13383	0.01810	0.13258	0.06158	0.15134	0.11635	0.02995	0.05567	0.23344	0.00658	0.00000	
[23]	0.24281	0.08014	0.17177	0.04540	0.17575	0.17285	0.20317	0.17129	0.14807	0.18058	0.16760	0.24486	0.12084	0.24864	0.15380	0.04294	0.22354	0.01007	0.13356	0.35207	0.03841	0.05864	0.00000

Table 3.8 Analysis of Molecular Variance (AMOVA)

Y-chromosome

Groups	No. of groups	Within populations		Among populations within groups		Among groups	
		Variance (%)	Φ_{ST}	Variance (%)	Φ_{SC}	Variance (%)	Φ_{CT}
Overall	1	76.3	0.24	15.7	0.17	8.0	0.08
Linguistic groups	3	75.30	0.25	14.62	0.16	10.08	0.11
Geographic groups	5	75.94	0.24	11.28	0.13	12.79	0.13

Table 3.9 Analysis of Molecular Variance (AMOVA)

mtDNA

Groups	No. of groups	Within populations		Among populations within groups		Among groups	
		Variance (%)	Φ_{ST}	Variance (%)	Φ_{SC}	Variance (%)	Φ_{CT}
Overall	1	91.3	0.091	7.64	0.08	1.1	0.01
Linguistic groups	3	91.03	0.090	7.00	0.07	1.97	0.02
Geographic groups	5	91.48	0.090	8.07	0.08	0.45	0.01

All Φ-statistics. *P*-values are less than 0.03.

with Hausa, whereas most of the Sudanese populations show a higher frequency of the high differences classes 4-12. Nuba and Nubian show a highest frequency of the low differences class 3. Both Arabs and Nilotics group show similar unimodal distributions (Fig. 3.23–Fig. 3.33).

Although Beja has the greatest sample size, but they are found to harbor the lowest range of nucleotide diversity (0.0255 ± 0.0158). Shilluk showed the highest range of nucleotide diversity (0.0732 ± 0.0409). Gaalien, Fur, Fulani, Borgu, Masalit, Dinka, Arakien, Meseria and Hausa showed similar mean values of nucleotide diversity ranged from 0.0306 – 0.0364. Mean values of nucleotide diversity of the Sudanese populations and their standard deviations are shown in Table (3.10) and Fig (3.34). The highest value of mean of parirwise differences was observed in Nuer (13.989 ± 6.5042) and the lowest value was found to be in Beja (2.299 ± 1.2814). Mean values and their standard deviations of pairwise differences are shown in Table (3.11).

3.2.9 Mantel Test

For Y-chromosome, Mantel test (Table 3.12) reveal strong correlation between genetic and linguistic structures when geography is held constant (r= 0.31, p= 0.007), and a slightly weaker correlation between genetic and geographic distances when linguistics is held constant (r= 0.29, p= 0.025) that appears only after the removal of nomadic pastoralists, Fulani and Meseria. Those groups are known to have migrated recently to the country and are widely spread in different areas in Sudan. In contrast mtDNA showed no correlations between genetic, linguistic (r= 0.15, p= 0.08), and geographic distances (r= -0.05, p= 0.63). Also Mantel test indicates no correlation between Y-chromosome and mtDNA pairwise genetic distances (r= 0.08, p= 0.20), the correlation between Y-chromosome and mtDNA pairwise genetic distances is illustrated in Fig. (3.35).

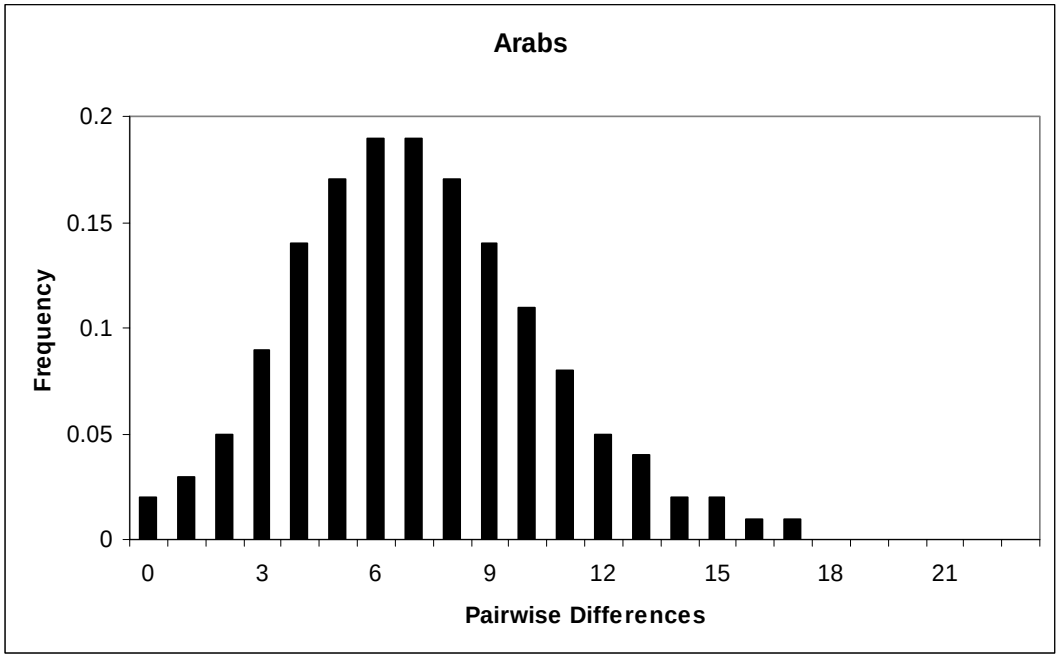


Fig. 3.23. Mismatch distribution of mtDNA HVR-I sequences in Arabs populations

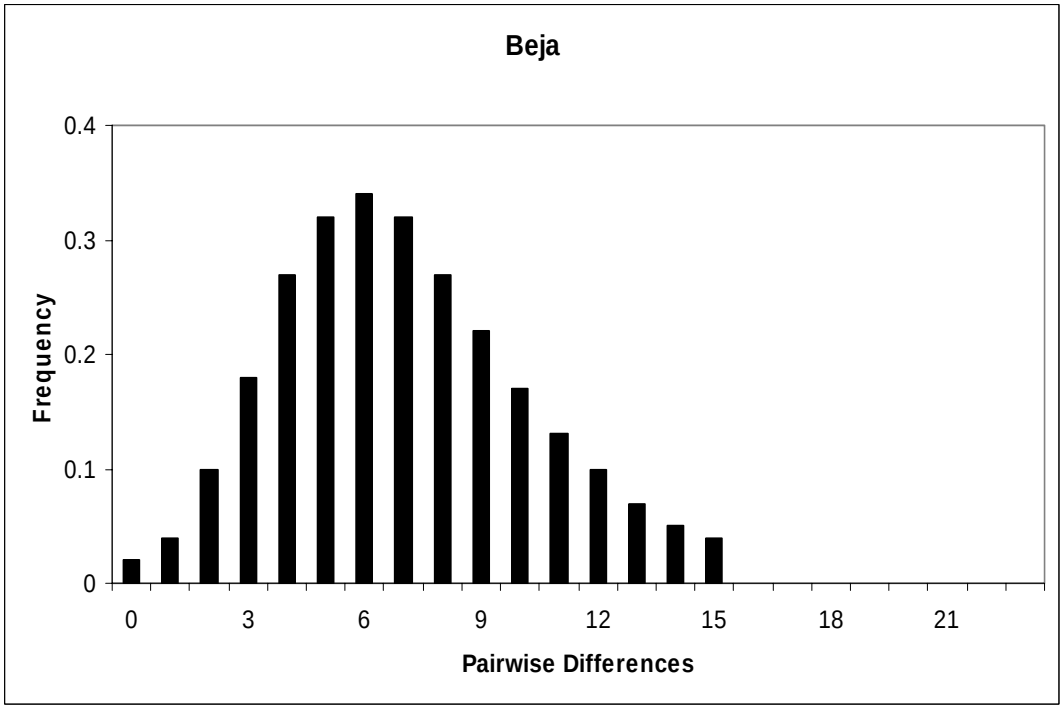


Fig. 3.24. Mismatch distribution of mtDNA HVR-I sequences in Beja

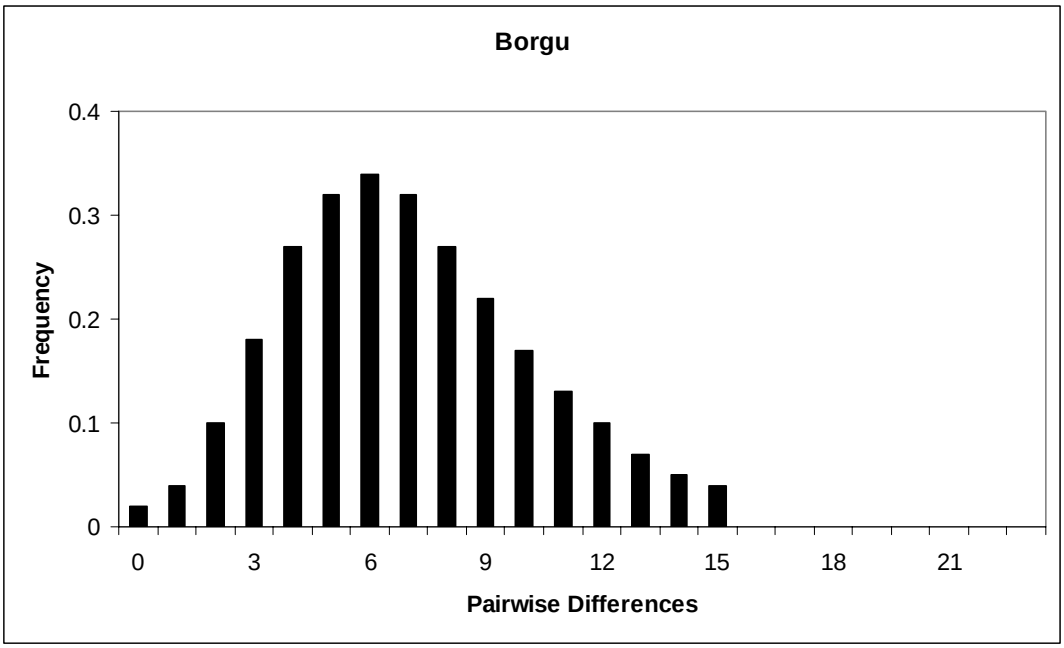


Fig. 3.25. Mismatch distribution of mtDNA HVR-I sequences in Borgu

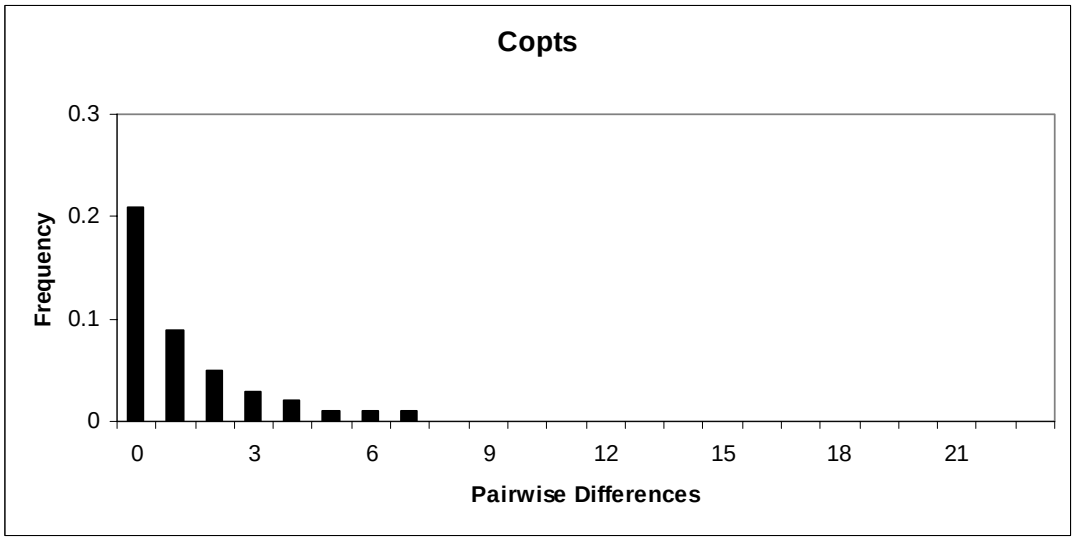


Fig. 3.26. Mismatch distribution of mtDNA HVR-I sequences in Copts

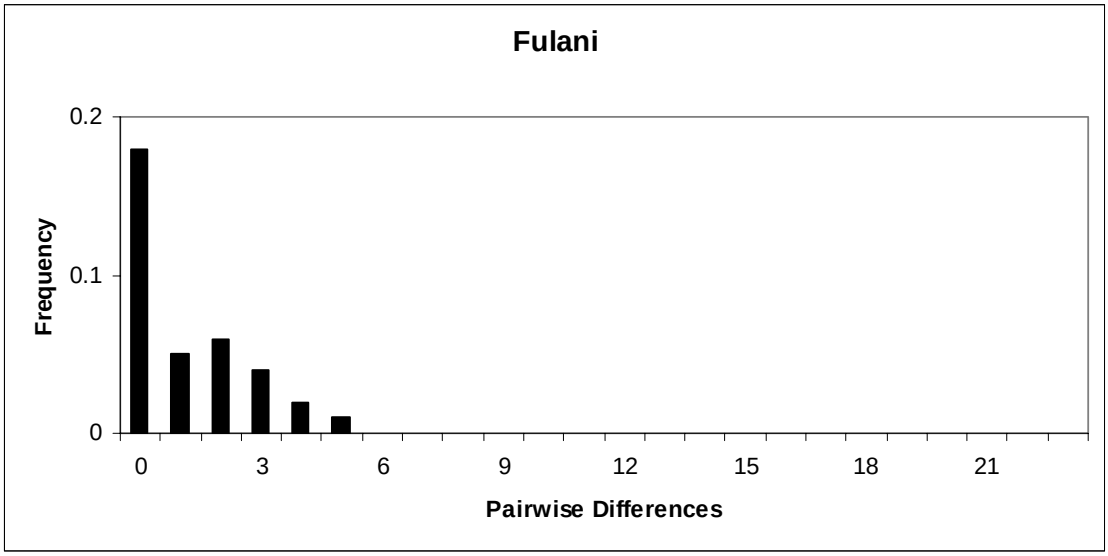


Fig. 3.27. Mismatch distribution of mtDNA HVR-I sequences in Fulani

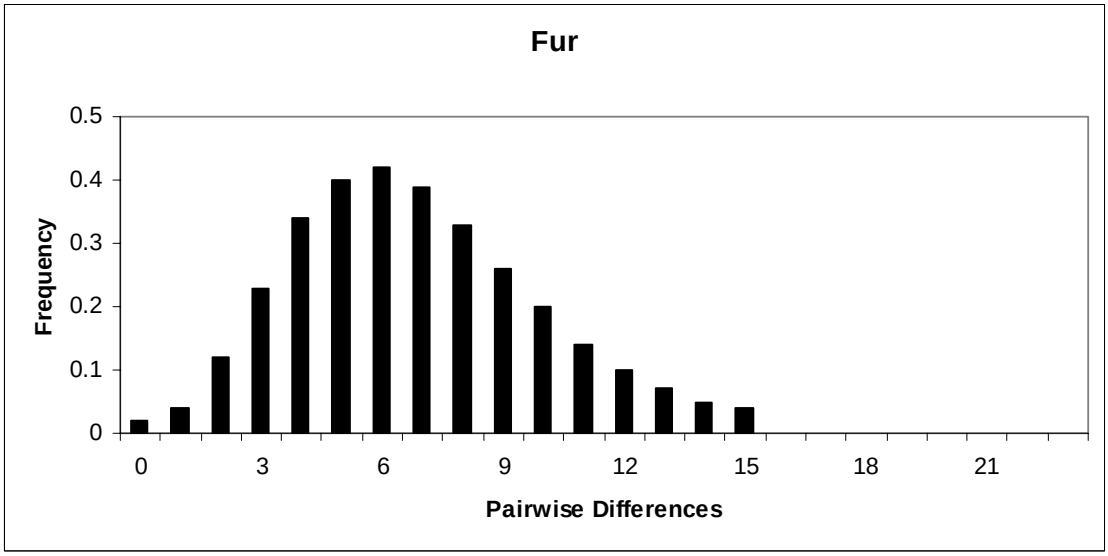


Fig. 3.28. Mismatch distribution of mtDNA HVR-I sequences in Fur

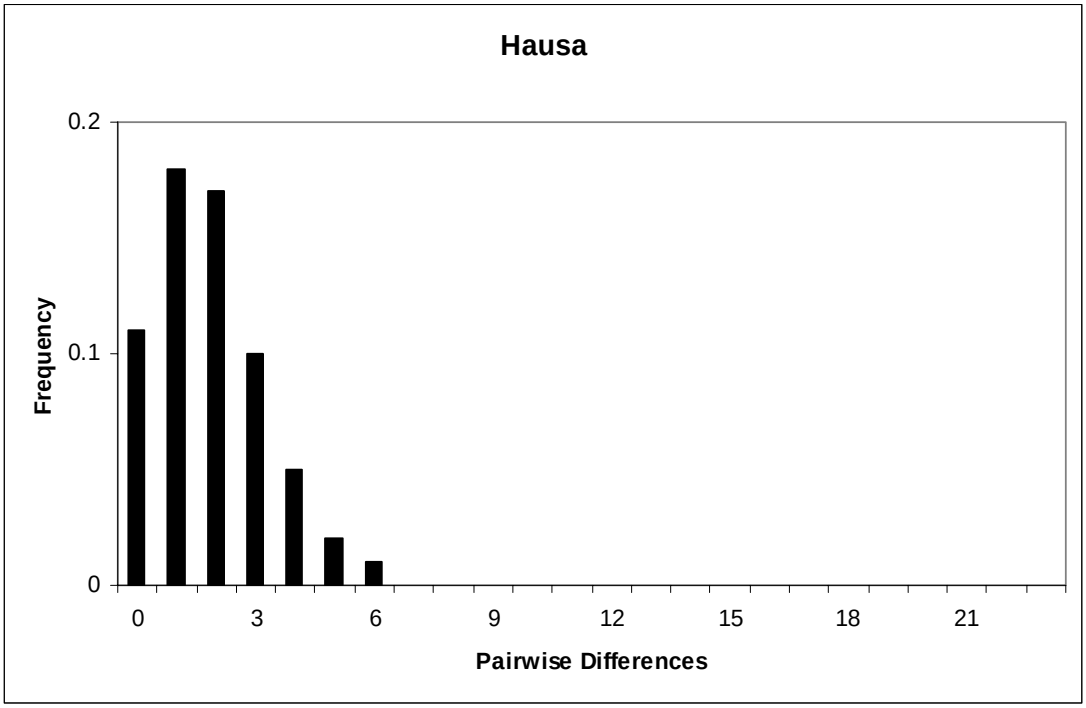


Fig. 3.29. Mismatch distribution of mtDNA HVR-I sequences in Hausa

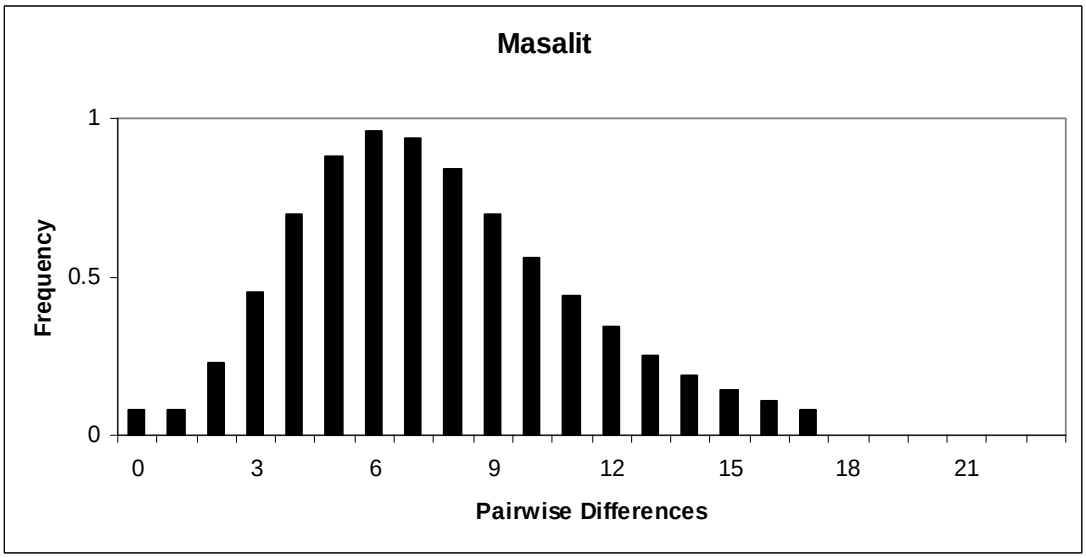


Fig. 3.30. Mismatch distribution of mtDNA HVR-I sequences in Masalit

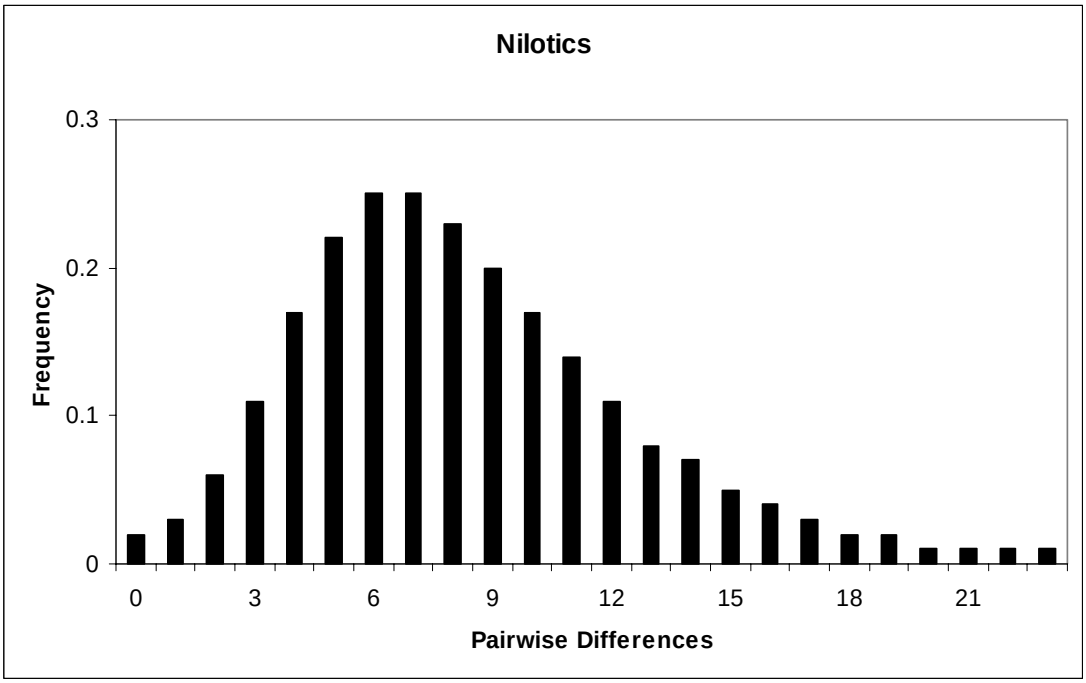


Fig. 3.31. Mismatch distribution of mtDNA HVR-I sequences in Nilotics

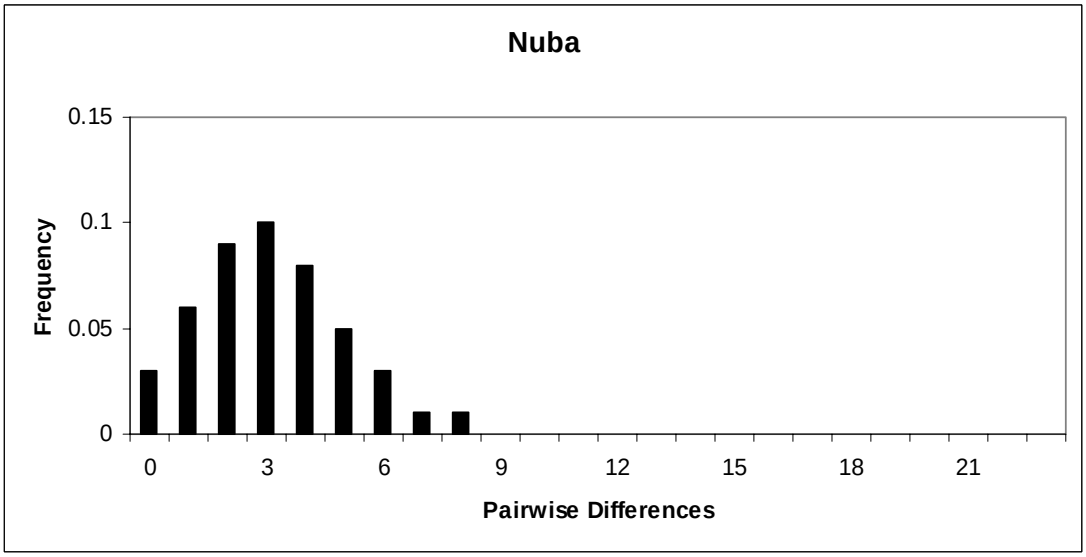


Fig. 3.32. Mismatch distribution of mtDNA HVR-I sequences in Nuba

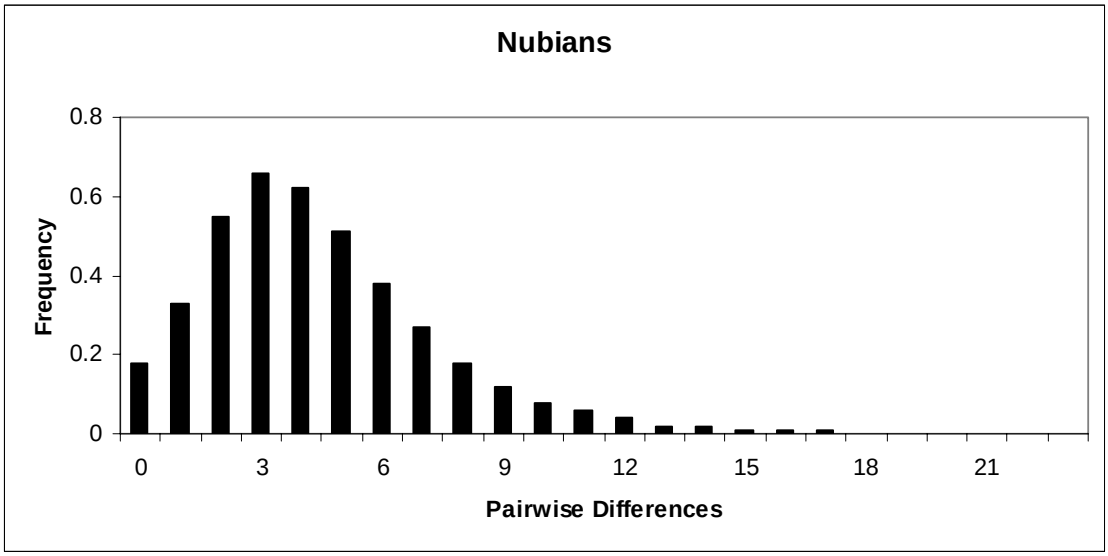


Fig. 3.33. Mismatch distribution of mtDNA HVR-I sequences in Nubians

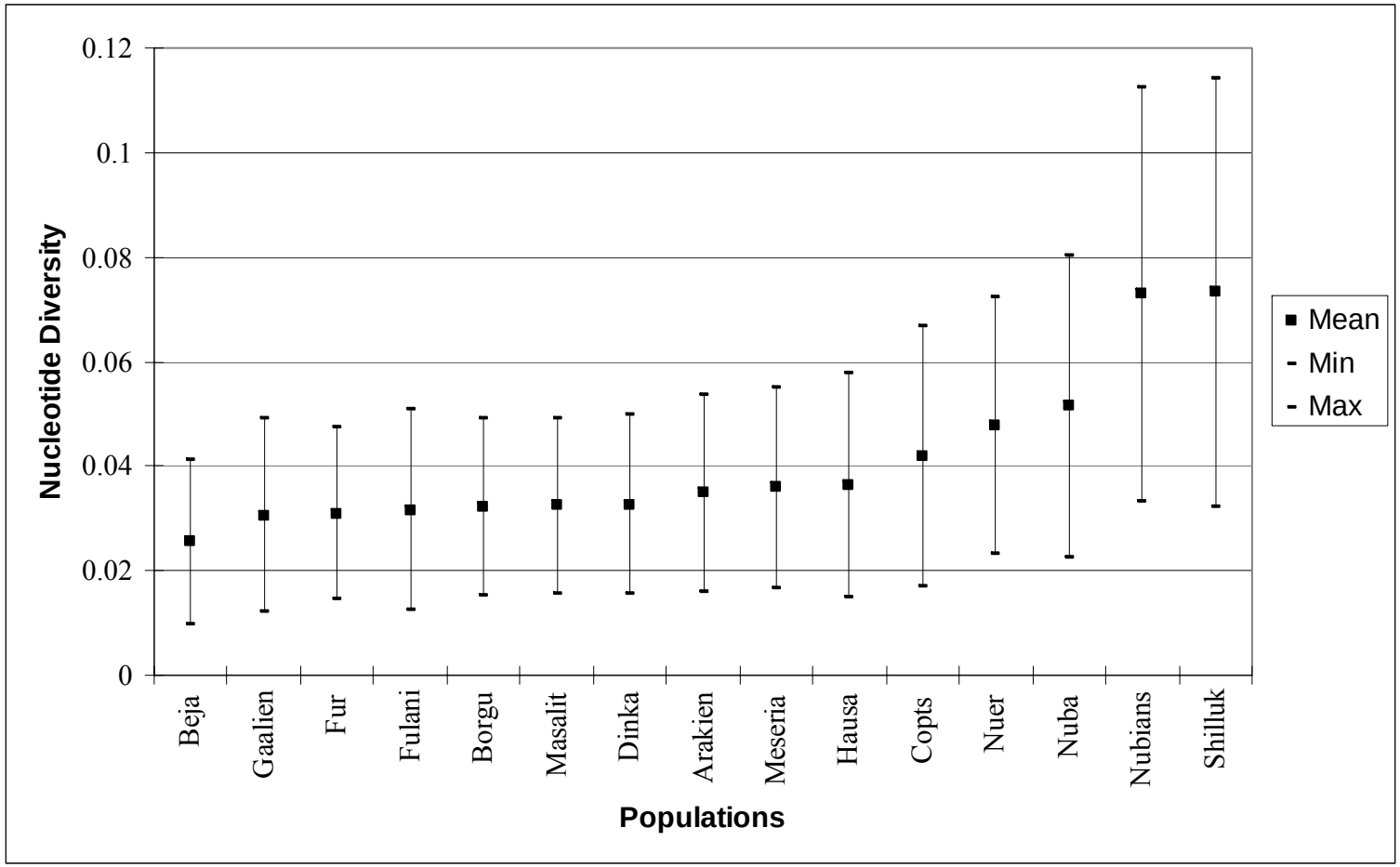


Fig. 3.34. Nucleotide diversity of HVR-I among 15 different Sudanese populations

Table 3.10. Nucleotide Diversity of HVR-I among 15 Sudanese Populations

Populations	**Mean ± SD**
Beja	0.0255 ± 0.0158
Gaalien	0.0306 ± 0.0186
Fur	0.0309 ± 0.0164
Fulani	0.0316 ± 0.0192
Borgu	0.0321 ± 0.0169
Masalit	0.0324 ± 0.0168
Dinka	0.0326 ± 0.0171
Arakien	0.0348 ± 0.0188
Meseria	0.0358 ± 0.0193
Hausa	0.0364 ± 0.0214
Copts	0.0419 ± 0.0248
Nuer	0.0476 ± 0.0246
Nuba	0.0515 ± 0.0289
Nubians	0.0729 ± 0.0396
Shilluk	0.0732 ± 0.0409

Table 3.11. Mean Values of Pairwise Differences based on HVR-I Sequences among 15 Sudanese Populations

Populations	Mean ± SD
Beja	2.2990 ± 1.2814
Gaalien	2.8116 ± 1.5375
Fulani	2.8148 ± 1.5311
Hausa	3.1714 ± 1.6797
Copts	3.2299 ± 1.7138
Nuba	4.5849 ± 2.3149
Nubians	6.4926 ± 3.1623
Shilluk	6.5083 ± 3.2491
Fur	9.0862 ± 4.3209
Borgu	9.4239 ± 4.4828
Masalit	9.5317 ± 4.4628
Dinka	9.5834 ±4.5249
Arakien	10.225 ± 4.9293
Meseria	10.533 ± 5.0685
Nuer	13.989 ± 6.5042

Table 3.12. Correlation Coefficients, *r* (*P*-value), between Genetic, Linguistic, and Geographic Distances in Sudanese Populations based on Y-chromosome and mtDNA Data.. And Correlation between Y-chromosome and mtDNA Genetic Distances.

	Y chromosome	**mtDNA**
Genetics and Linguistics, Geography held constant	0.30 (0.007)	0.15 (0.08)
Genetics and Geography, Linguistics held constant	0.29 (0.025)	-0.05 (0.63)
Y-chromosome *vs.* mtDNA	0.08 (0.20)	

3.2.10 F-statistics and Migration Rate

Table (3.13) shows the F_{ST} estimates and migration rate for the groups analyzed according to the island model. For Y-chromosome, Afro-Asiatic speaking groups had a higher Nm rather than the Nilo-Saharan speaking groups. Within Afro-Asiatic Arabs showed the highest migration rate (Nm = 5.02), followed by Non-Arabs (Nm = 4.86), for Nilo-Saharan, Non-Nilotics showed higher migration rate (Nm = 3.0) when compared with Nilotics (Nm = 2.33). In contrast for mtDNA Nilo-Saharan speaking groups showed very high migration rate compared to Afro-Asiatic (Table 3.13), the highest migration rate was found in Non-Nilotics (Nm = 87.89), followed by Nilotics (Nm = 38.26), for Afro-Asiatic again Non-Arabs showed a lower migration rate (Nm = 4.67) compared to Arabs (Nm = 7.40).

3.2.11 Gene Diversity and Neutrality Test

Table (3.14) shows the measures of genetic diversity estimated from Y-chromosome data; the highest number of haplotypes (different 9 haplotypes) was found to be in Gaalien, while the lowest number (3) of haplotypes was found to be in Dinka, Nuer and Fulani. The θ_π values ranged from 1.812–2.955 indicate that Afro-Asiatic speaking groups harbor the most Y-chromosome diversity. The θ_π values of Nilo-Saharan speaking groups ranged from 1.831–2.858 which is slightly smaller when compared to Afro-Asiatic speaking groups. Y-chromosome Tajima's *D* was slightly positive in most cases, whereas it was slightly negative in the case of Arakien, Gaalien, Hausa and Nubians; all *P* values were not significant.

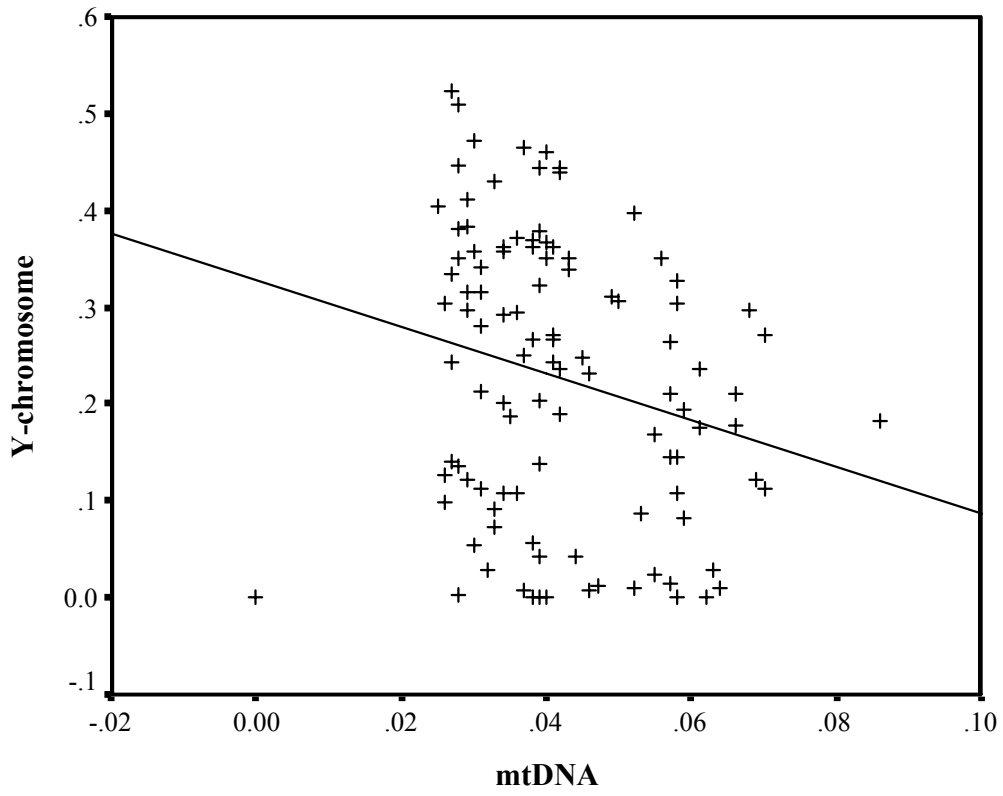


Fig. 3.35. Pairwise mtDNA and Y-chromosome F_{ST} values of the 15 Sudanese populations analyzed in this study, Mantel test indicates no significant correlation between pairwise mtDNA and Y-chromosome F_{ST} values - Mantel correlation coefficient $r = 0.08$; $p = 0.20$.

Table 3.13. F_{ST} and Nm Estimates of Y-chromosome and mtDNA in Two Linguistic Groups in Sudanese Populations

Linguistic groups		Subgroups	F_{ST}[1]	Nm
Afro-Asiatic	**Y-chromosome**	**Arabs**	0.166	5.02
		Non-Arabs	0.171	4.86
Nilo-Saharan		**Nilotics**	0.300	2.33
		Non-Nilotics	0.250	3.00
Afro-Asiatic	**mtDNA**	**Arabs**	0.119	7.403
		Non-Arabs	0.176	4.674
Nilo-Saharan		**Nilotics**	0.255	38.26
		Non-Nilotics	0.0113	87.89

[1]F_{ST} is a measure of interpopulation variability, while Nm is the effective number of migrants.

Table 3.14. Measures of Genetic Diversity and Neutrality Test Estimated from Y-chromosome Data.

Linguistic groups	Populations	N	Hp	θ_K[1]	θ_S (SD)[2]	θ_π (SD)[3]	D (*P*)[4]
Afro-Asiatic	**Arakien**	24	4	1.105	1.875 (0.901)	1.812 (1.205)	− 0.105 (>0.10)
	Beja	42	6	1.679	2.324 (0.965)	2.670 (1.612)	0.440 (>0.10)
	Copts	33	7	2.424	2.710 (1.124)	2.955 (1.764)	0.284 (>0.10)
	Gaalien	50	9	2.941	2.679 (1.045)	2.670 (1.605)	− 0.010 (>0.10)
	Hausa	32	6	0.976	1.986 (0.898)	1.831 (1.202)	− 0.235 (>0.10)
	Meseria	28	5	1.502	2.056 (0.940)	2.384 (1.488)	0.492 (>0.10)
Nilo-Saharan (NS)	**Borgu**	26	4	1.066	2.096 (0.965)	2.858 (1.732)	1.143 (>0.10)
	Dinka	26	3	0.645	1.572 (0.785)	2.265 (1.432)	1.303 (>0.10)
	Fur	32	4	0.976	1.986 (0.898)	2.250 (1.414)	0.397 (>0.10)
	Masalit	32	4	0.976	1.986(0.898)	1.831 (1.202)	− 0.235 (>0.10)
	Nuer	12	3	0.934	1.987(1.067)	2.667(1.719)	1.305 (>0.10)
	Nubians	39	8	2.763	2.838 (1.133)	2.729 (1.644)	− 0.119 (>0.10)
	Shilluk	15	4	1.428	2.153 (1.085)	2.629 (1.667)	0.796 (>0.10)
Niger-Congo (NC)	**Fulani**	26	3	0.645	1.310 (0.693)	2.234 (1.417)	1.997 (>0.10)
NS+NC	**Nuba**	28	4	1.032	1.799 (0.854)	2.677 (1.636)	1.467 (>0.10)

[1] θ_K based on the number of distinct haplotypes per sample size

[2] θ_S is based on number of variant sites.

[3] $\theta\pi$ is equivalent to mean number of pairwise differences between sequences (π).

[4] Tajima's D compares two θ estimates, θ_S values relative to $\theta\pi$.

Table (3.15) shows the measures of genetic diversity estimated from mtDNA data of 404 mitochondrial HVR-I sequences from Sudanese populations. The great majority of nucleotide polymorphisms were found to be biallelic. Transition ratio (A ↔ G or T ↔ C) in most of the population analyzed was higher than transversion ratio (A ↔ T,C or G ↔ T,C); the only exception is Copts which harbors a higher transversion ratio when compared to transition ratio. While only three populations, namely Gaalien, Shilluk and Fulani harbor equal numbers of transversion versus transition. The highest number of polymorphic sites (71) was found to be in Masalit and the lowest one (16) in Gaalien. On the whole, most of these populations exhibit high diversity, characterized by a great number of distinct mtDNA haplotypes. Hausa showed the highest genetic diversity in the Sudan (different 23 haplogroups). The two populations mutation rate parameters, θ_K (based on the number of distinct haplotypes per sample size) and θ_S (based on the number of polymorphic sequence positions per sample size), indicate that the largest diversity values belong to Nilo-Saharan speaking groups, ranged from 7.84–17.14 for θ_S and 3.015–40.121 for θ_K. Afro-Asiatic speaking groups have smaller diversity values ranged from 4.04–15.37 for θ_S and 4.07–28.06 for θ_K. Also Nilo-Saharan speaking groups have higher θ_π values ranged form 6.49–13.99, while θ_π values among Afro-Asiatic speaking groups ranged from 2.29–10.53.

All mtDNA Tajima's *D* scored negative values; three of them namely Hausa, Fur and Masalit were significantly negative at the 0.05 level. Masalit showed the highest negative value, and the lowest negative value was observed in Meseria. Fulani had an intermediate value between Nilo-Saharan and Afro-Asiatic speaking groups.

Table 3.15. Measures of Genetic Diversity and Neutrality Test Estimated from mtDNA Data

Linguistic groups	Population	N	No. of Transition	No. of Transversion	No. of Substitutions	No. of polymorphic	Hp	θ_K	θ_S (SD)	θ_π(SD)	D (*P*)
Afro-Asiatic	**Arakien**	17	32	23	55	51	8	5.283	15.37 (5.81)	10.23 (5.52)	− 1.419 (0.07)
	Beja	48	17	2	19	18	21	13.683	4.04 (1.44)	2.29 (1.42)	− 1.360 (0.07)
	Copts	29	9	11	20	18	9	4.083	4.54 (1.72)	3.23 (1.91)	− 0.996 (0.17)
	Gaalien	22	8	8	16	16	8	4.070	4.28 (1.71)	2.81 (1.71)	− 1.224 (0.09)
	Hausa	35	15	11	26	23	23	28.060	5.55 (1.97)	3.17 (1.87)	− 1.464 (0.04)
	Meseria	15	28	4	32	32	9	8.579	9.64 (3.77)	10.53 (5.68)	− 0.383 (0.68)
Nilo-Saharan	**Borgu**	25	37	21	58	55	19	40.121	14.73 (5.089)	9.42 (4.99)	− 1.403 (0.06)
	Dinka	29	51	13	64	56	14	10.018	14.26 (4.74)	9.58 (5.04)	− 1.236 (0.10)
	Fur	26	42	14	56	56	15	13.929	14.68 (4.99)	9.09 (4.81)	− 1.463 (0.04)
	Masalit	42	51	22	73	71	19	13.156	16.60 (5.09)	9.53 (4.96)	− 1.535 (0.04)
	Nuer	16	42	31	73	64	6	3.015	17.14 (5.86)	13.99 (7.25)	− 0.720 (0.25)
	Nubians	29	25	21	46	38	18	19.273	9.68 (3.33)	6.49 (3.52)	− 1.215 (0.11)
	Shilluk	16	14	14	28	26	11	14.271	7.84 (3.13)	6.51 (3.64)	− 0.695 (0.27)
Niger-Congo	**Fulani**	28	10	10	20	18	11	6.189	4.63 (1.77)	2.81 (1.70)	− 1.367 (0.08)
NS+NC	**Nuba**	33	17	12	29	29	17	13.391	6.51 (2.32)	4.58 (2.57)	− 1.049 (0.14)

Note: N = sample size; Hp = number of different haplogruops observed; SD = standard deviation; D = Tajima's D; *P* = *P* value for D.

3.2.12 Principal Component Analysis (PCA)

Y-chromosome PCA of Sudanese populations is shown in Fig (3.36) based on F_{ST} genetic distances, the first two principal components account for 81.4% of the total variation and revealed broad genetic affinities between populations. Three main clusters of genetically closely related populations are observed: the first of Nilotics and Nuba group, the second included Arab, Copts and Nubians, and the third cluster was strictly for Nilo-Saharan speaking groups, including Masalit, Fur, and Borgu. Data from this study alongside available data of African populations and Turks were analyzed in the second run of PCA (Fig. 3.37). The plot portrays broad genetic affinities reflected in two main clusters of genetically closely related populations. The first cluster groups Nilo-Saharan speaking groups from the Sudan together with Oromo from Ethiopia. The second cluster encompasses Afro-Asiatic speaking populations from the Sudan as well as Nubians and Amhara from Ethiopia, Fulani, and Turks. Senegalese fell relatively distant from both clusters in the plot.

Figure (3.38) displays the first run of PCA of mtDNA, for the first two principal components, which together account for 87.2% of the total variation. Both Nilo-Saharan and Afro-Asiatic speaking groups showed a high degree of interpopulation diversity, which is illustrated by more scattered distribution of the points in the plot. Most populations of west and south Sudan occupied the upper half of the plot; the only exceptions are Nuer and Fulani, who have closer genetic affinities with the populations from the Nile Valley. Both Nubians and Shilluk showed high genetic affinities with the population of western Sudan including Fur, Masalit and Borgu. Meseria were found to show close genetic affinities with Dinka. Figure (3.39) display the PCA of the Sudanese populations compared to other African populations and Saudi Arabian, the plot could be

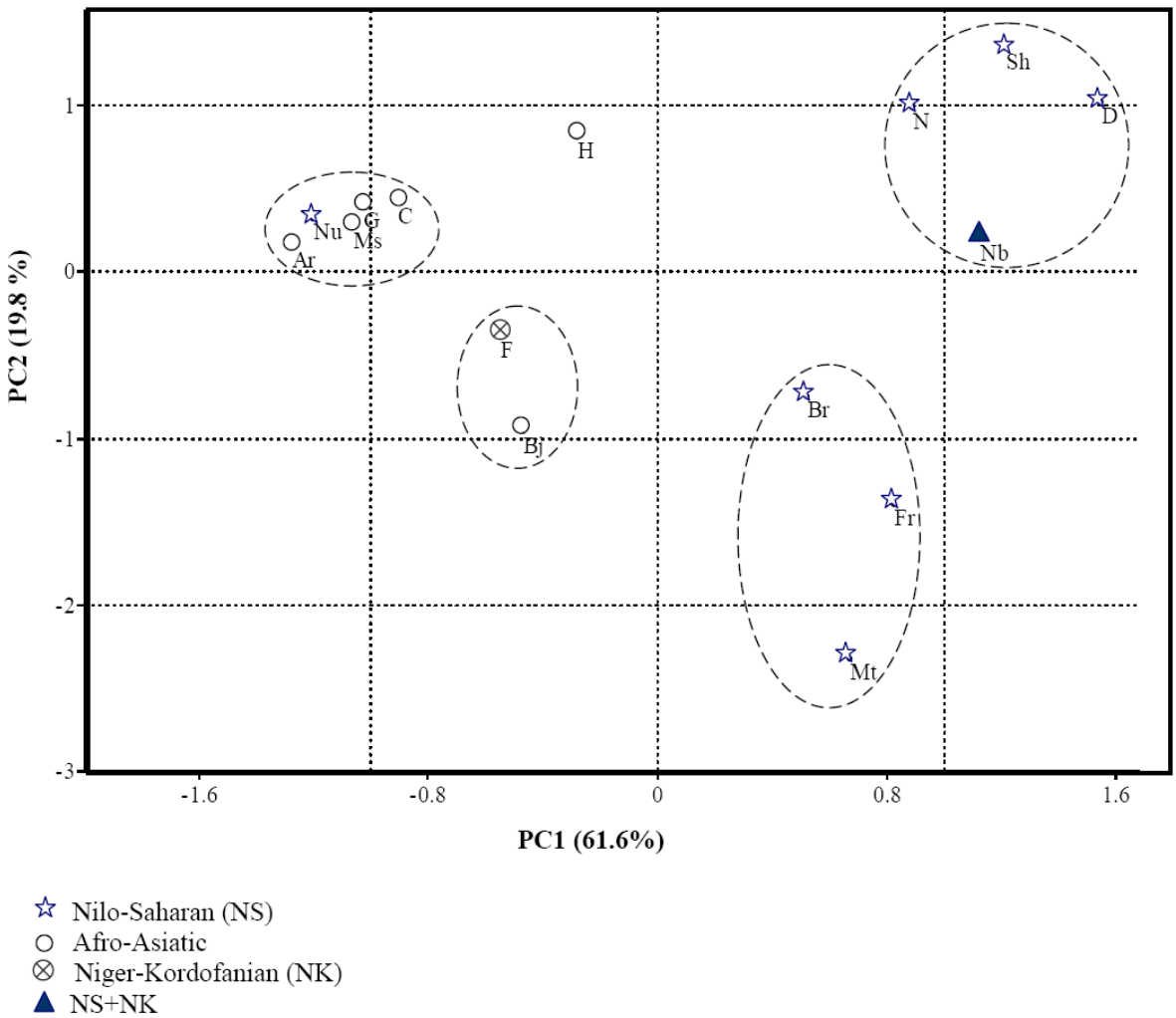


Fig. 3.36. PC plot of Y-chromosome of 15 Sudanese populations based on F_{ST} genetic distances; Codes as follows: **Ar**, Arakien; **Bj**, Beja; **Br**, Borgu; **C**, Copts; **D**, Dinka; **F**, Fulani; **Fr**, Fur; **G**, Gaalien; **H**, Hausa; **Mt**, Masalit; **Ms**, Meseria; **N**, Nuer; **Nb**, Nuba; **Nu**, Nubians; **Sh,** Shilluk. Circles indicate the genetic relationship between populations.

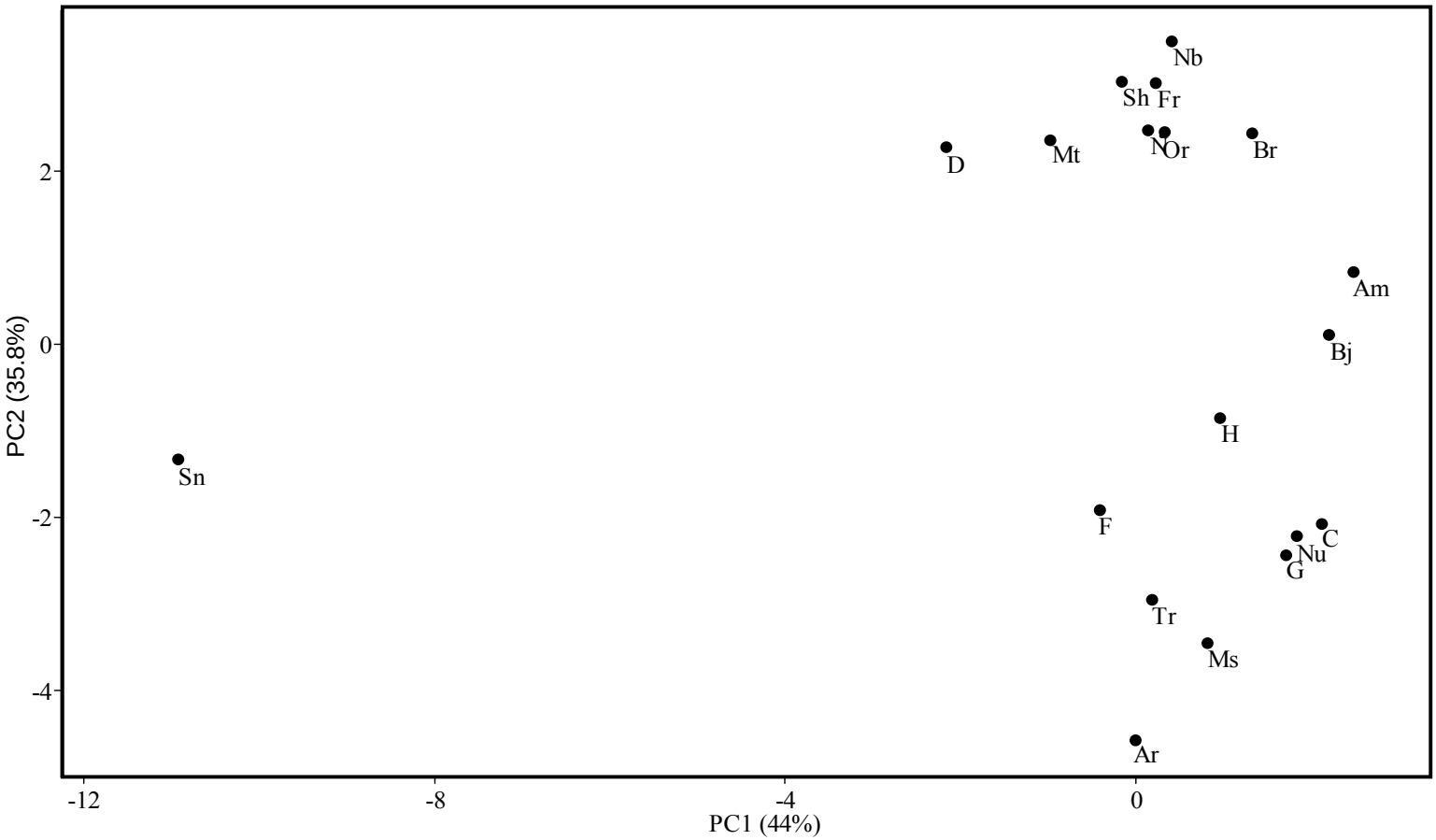


Fig. 3.37. PC plot of the Y-chromosome of Sudanese populations in the present study, compared with Turks (Sanchez *et al.* 2005), two Ethiopian groups (Oromo and Amhara), and Senegalese (Semino *et al.* 2002). Codes as follows: **Am**, Amhara; **Ar**, Arakien; **Bj**, Beja; **Br**, Borgu; **C**, Copts; **D**, Dinka; **F,** Fulani; **F**r, Fur; **G**, Gaalien; **H**, Hausa; **M**t, Masalit; **Ms**, Meseria; **N**, Nuer; **Nb**, Nuba; **Nu**, Nubians; **Or**, Oromo; **Sh**, Shilluk. **Sn**, Senegalese; **Tr**, Turks.

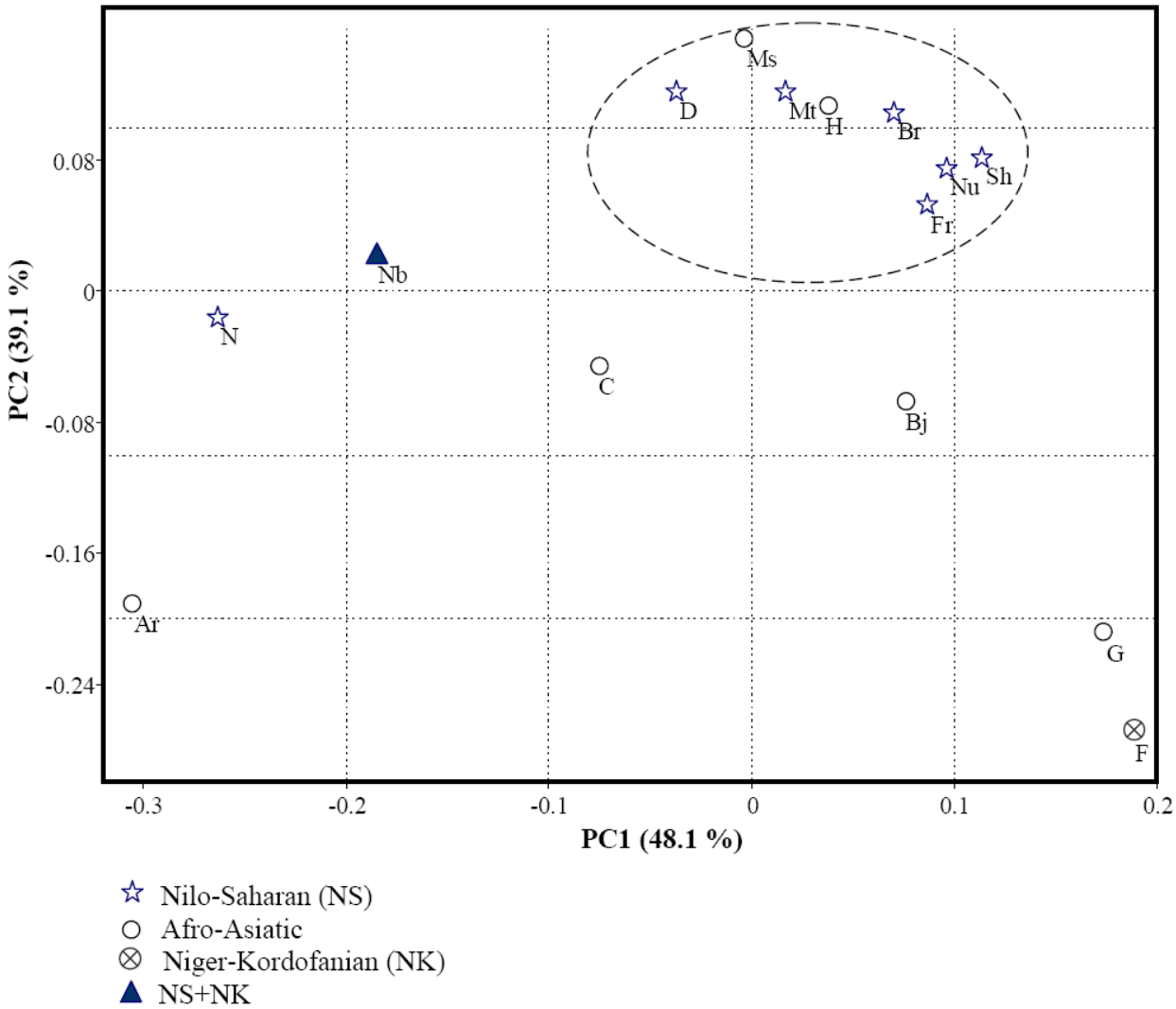


Fig. 3.38. PC plot of mtDNA of 15 Sudanese populations based on F_{ST} genetic distances; Codes as follows: **Ar**, Arakien; **Bj**, Beja; **Br**, Borgu; **C**, Copts; **D**, Dinka; **F**, Fulani; **Fr**, Fur; **G**, Gaalien; **H**, Hausa; **Mt**, Masalit; **Ms**, Meseria; **N**, Nuer; **Nb**, Nuba; **Nu**, Nubians; **Sh,** Shilluk. Circles indicate the genetic relationship between populations.

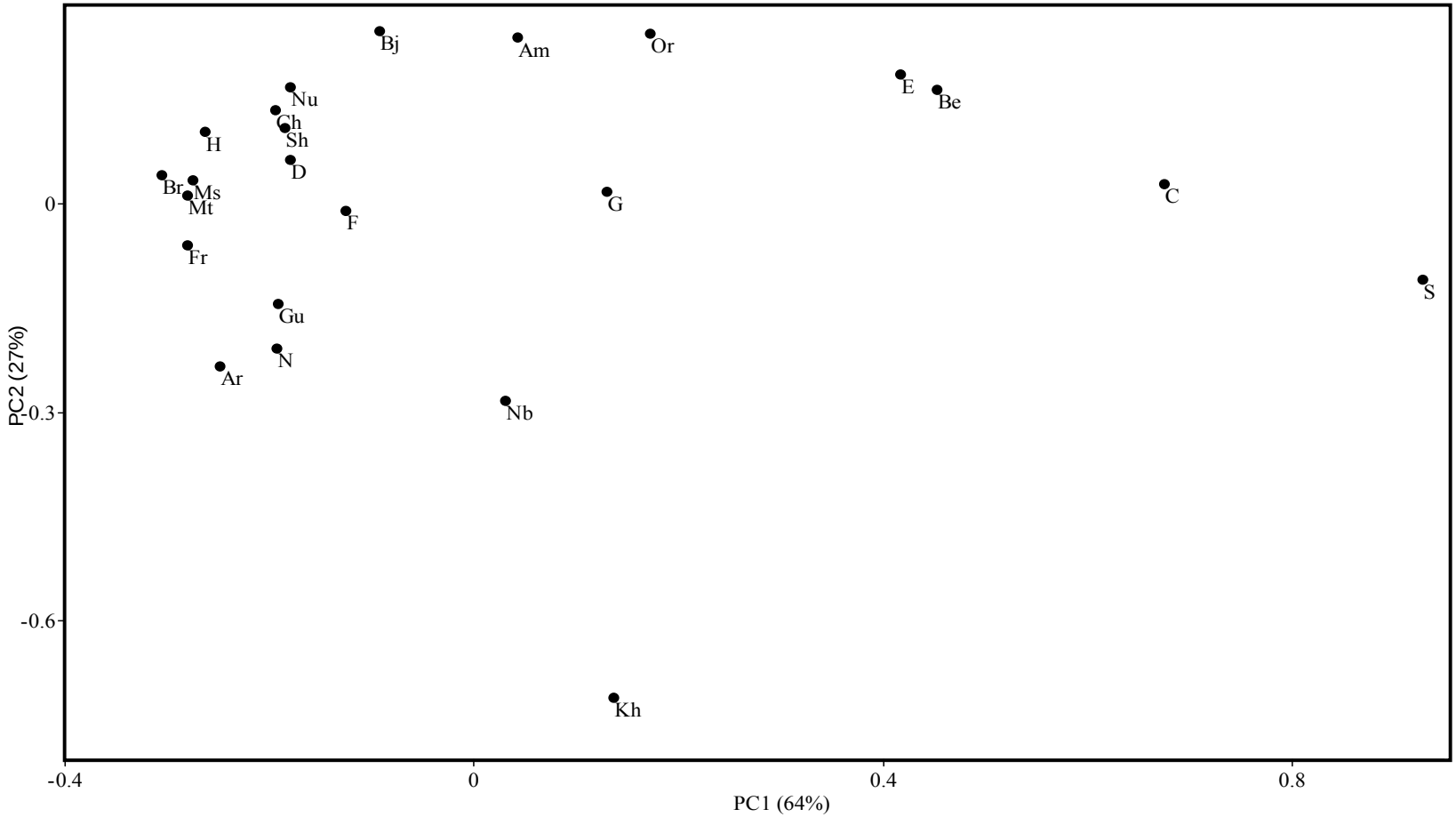


Fig. 3.39. PC plot of the mtDNA of Sudanese populations in the present study, compared with: Egyptian (Stevanovitch *et al.*, 2003), Berber (Fadhlaoui-Zid *et al.*, 2004), Amhara and Oromo (Kivisild *et al.*, 2004), Guineans (Rosa *et al.*, 2004), Saudi Arabian (Abu-Amero *et al.*, 2006), Khoisan (Behar *et al.*, 2008) and Chad (Hajek *et al.*, 2008). Codes as follows: **Ar**, Arakien; **Bj**, Beja; **Br**, Borgu; **C**, Copts; **D**, Dinka; **F**, Fulani; **Fr**, Fur**; G**, Gaalien**; H**, Hausa; **Mt**, Masalit; **Ms**, Meseria**; N**, Nuer; **Nb**, Nuba**; Nu**, Nubians; **Sh,** Shilluk; **Or**, Oromo; **Am**, Amhara**; Kh**, Khoisan; **S**, Saudi; **Ch**, Chad; **Be**, Berber**; E**, Egypt; **Gu**, Guinea.

divided into three main clusters the first one contains most of the Sudanese populations with Amhara and Oromo, the second cluster includes Egyptian, Berber, Copts and Saudi, and the third cluster includes Nuba and Khoisan of southern Africa.

3.2.13 Y-chromosome Phylogeny

Unrooted phenogram (Fig. 3.40) based on pairwise F_{ST}, shows five clusters, the first one included Dinka, Shilluk, Nuba, Nuer which is mainly related to southern Sudan, the second cluster contains the population of western Sudan including Fur, Masalit, and Borgu, the third one contains Hausa and Fulani which are West African, the fourth cluster contains Beja of eastern Sudan, the fifth cluster contains Arabs in addition to Nubians, and the sixth cluster contains Copts who are believed to display Egypt genetic legacy.

3.2.14 mtDNA Phylogeny

3.2.14.1 mtDNA Phylogenetic Trees

In all topologies of the trees each subtree is made of two branches or taxa. Both unrooted phylogenetic trees: neighbor-joining (NJ) and minimum evolution (ME) based on the F_{ST} of HVR-I sequences (Fig. 3.41) and (Fig. 3.42) respectively, showed similar topologies and branch lengths with few differences, also a good fit with PC plot was observed. Both trees obtained two main branches. In the first one most the populations of west Sudan were clustered together including Borgu, Fur and Masalit with Shilluk, Hausa and Nubians. While Beja was clustered with Gaalien and Fulani with long branch lengths. In the second branch, Meseria is clustered with Copts, Dinka with Nuba and Nuer with Arakien, in these groups long length of branches were observed between all populations. UPGMA obtained a rooted tree (Fig. 3.43); once again most of the populations of west Sudan were clustered together including Fur, Masalit, Borgu and Meseria with Nubians and Hausa. Shilluk with Dinka, Nuba, Copts and Fulani were stand

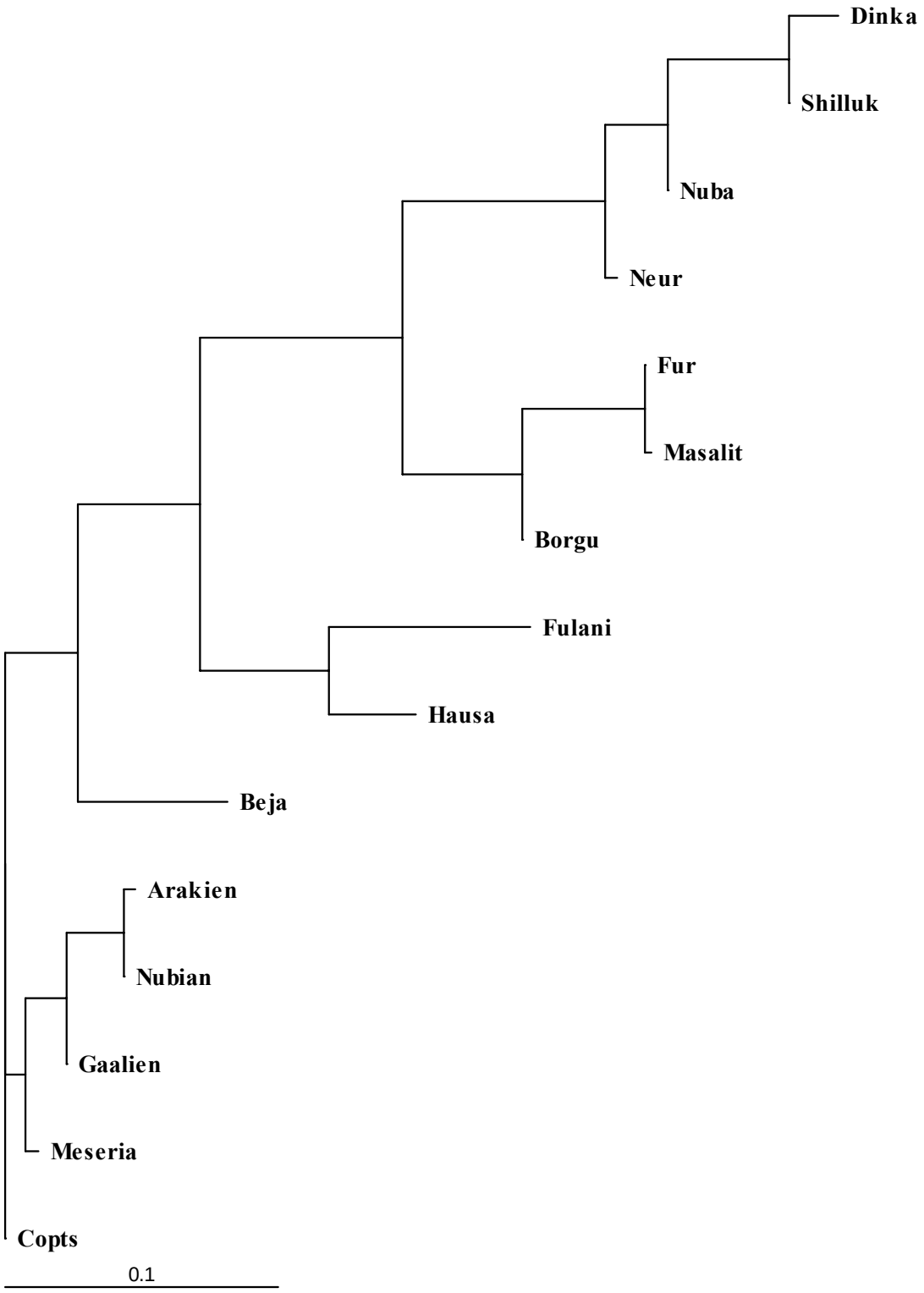


Fig. 3.40. Y-chromosome neighbor-joining tree based on pairwise distances of different 15 Sudanese populations

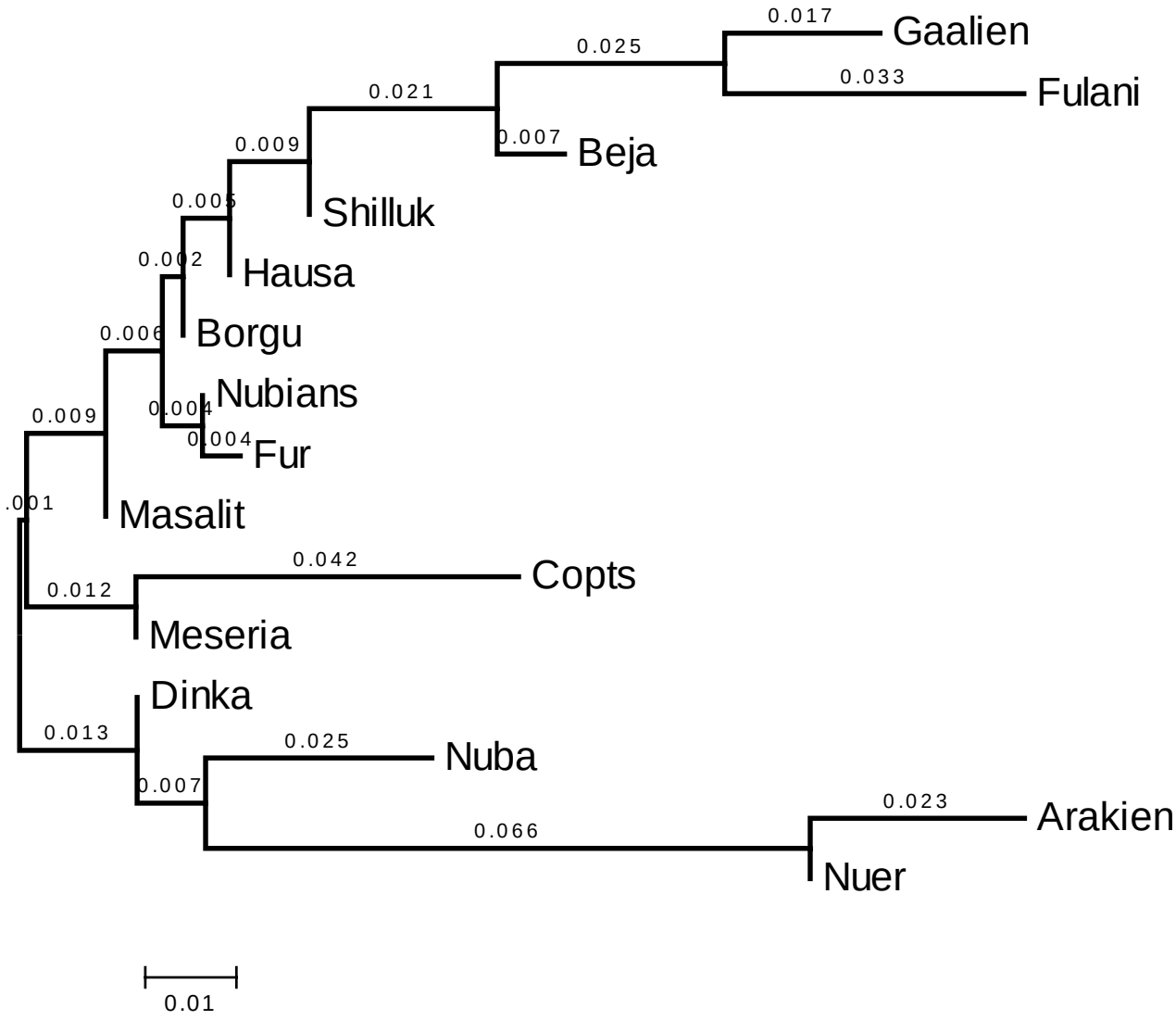


Fig. 3.41. Neighbor-joining (NJ) tree of 15 different Sudanese populations based on genetic distance of mtDNA HVR-I sequence. Numbers above branches correspond to branch lengths.

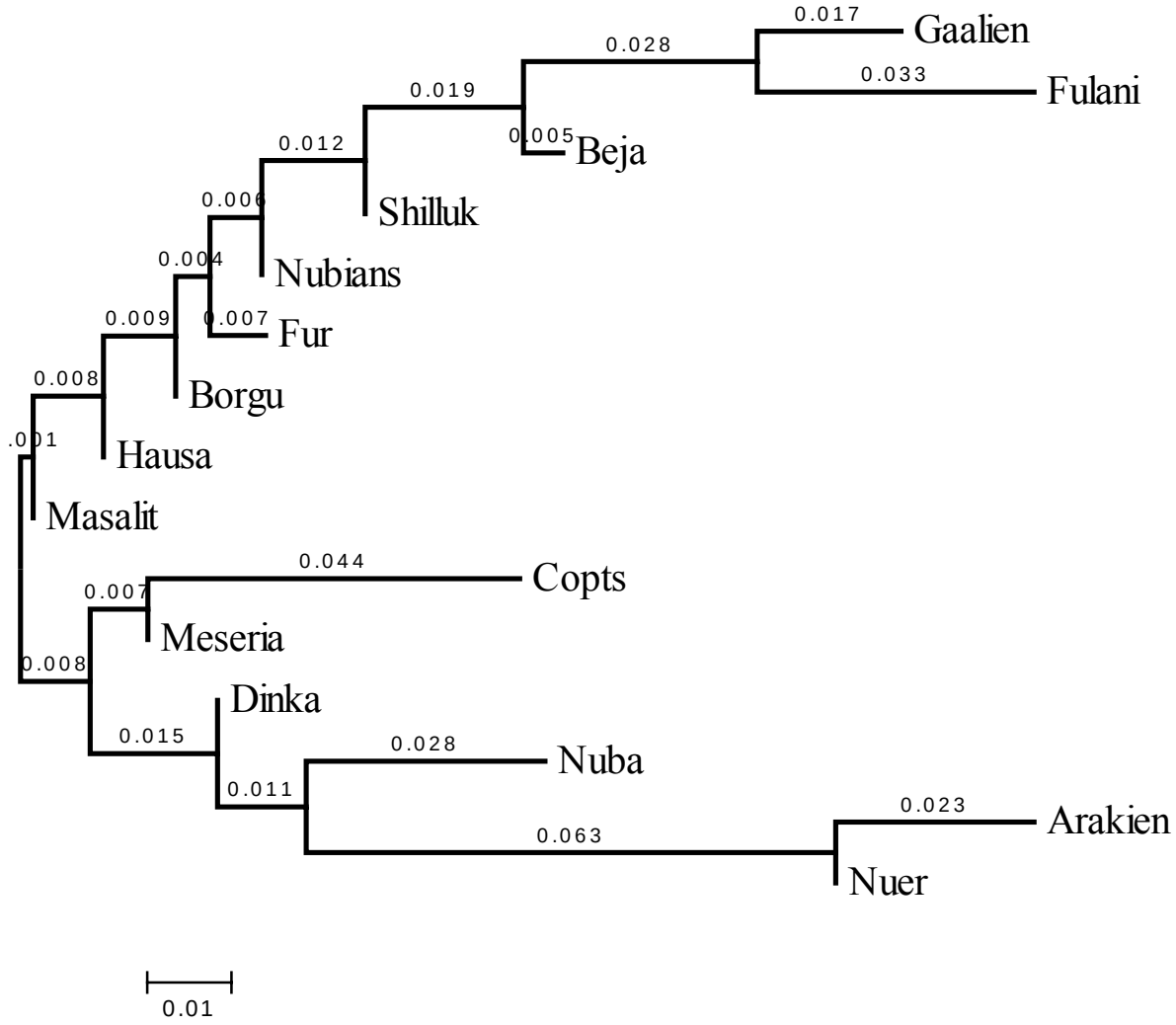


Fig. 3.42. Minimum evolution (ME) tree of 15 different Sudanese populations based on genetic distance of mtDNA HVR-I sequence. Numbers above branches correspond to branch lengths.

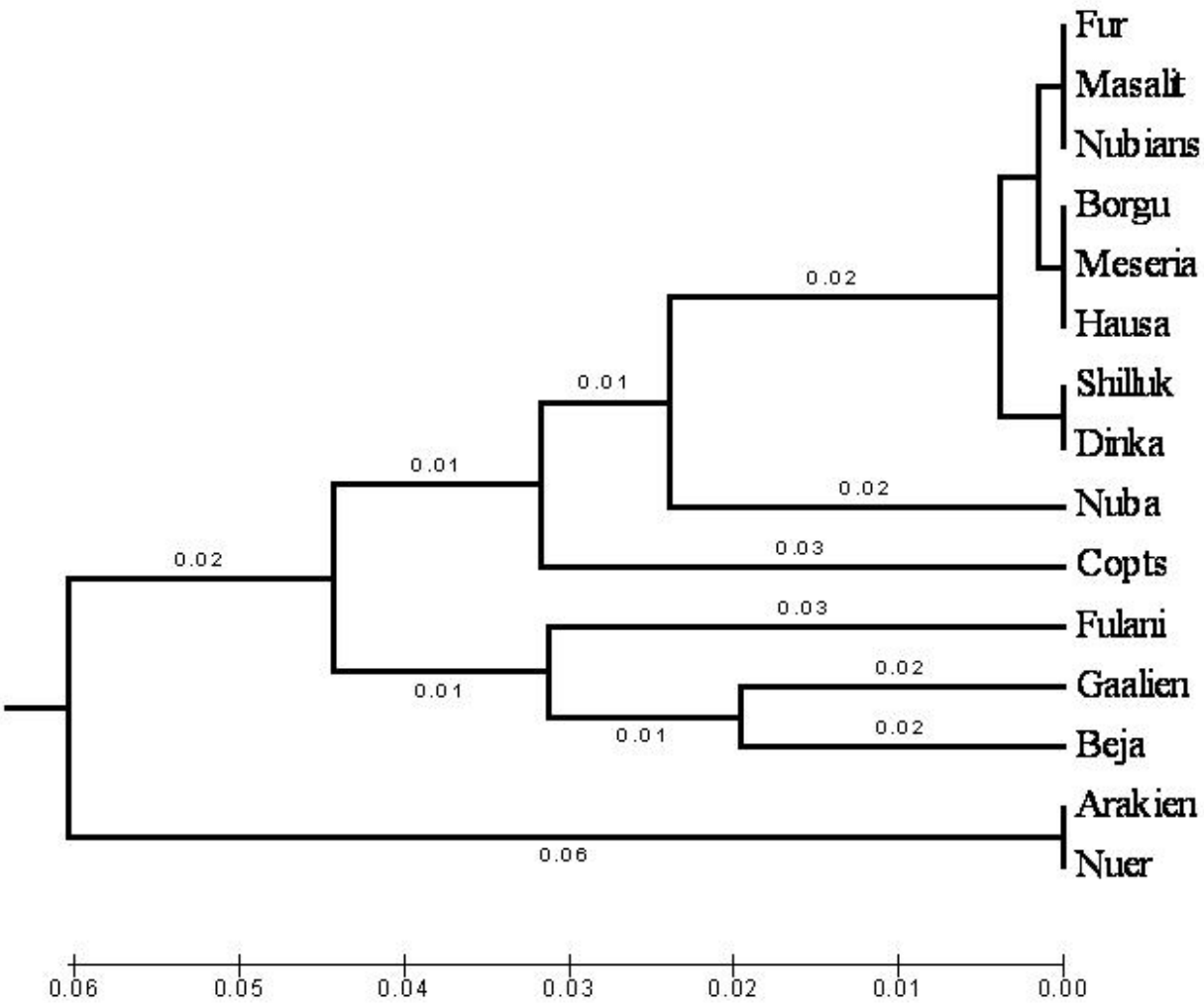


Fig. 3.43. Unweighted Pair Group Method Using Arithmetic Averages (UPGMA) tree of 15 different Sudanese populations, based on the genetic distance of mtDNA HVR-I sequence. Numbers above branches correspond to branch lengths.

alone in the tree. Gaalien was clustered with Beja and again Arakien with Nuer as an out-group connecting to the root of the tree.

3.2.14.2 mtDNA Phylogenetic Networks

Using median network analysis, both the substitutions in the mitochondrial HVR-I that occur along main branches of the network and the substitutions that occur along the terminal branches of the network have been investigated (Fig. 44 – Fig. 49). The main branches of the L0/L1 networks are well supported by several substitutions across the mitochondrial HVR-I. L0a and L1c haplogroups contain highly divergent lineages that are separated by several substitutions. L2 and L3 are the most divergent haplogroups among Sudanese; L2a, L2d, L3b, L3d and L3e haplogroups with their subclades are supported by several terminal branches. In contrast, L2c, L3i, L3h, L3w, L3x haplogroups showed low divergent lineages. Although haplogroup L3f has the highest frequency of all mtDNA lineages among Sudanese; it has a low haplotype variation and 67.4% of the sequences belonging to this haplogroup contain only the ancestral substitutions, especially among Beja (Appendix III). L4 and L5 haplogroups are represented by their subclades L4a and L5a respectively, although both subclades have low degree of divergence as well as the Eurasian haplogroup M. While haplogroup N is well represented by its subclades. Subclades J with its daughter clades J1a and J1b showed a remarkable haplotype variation.

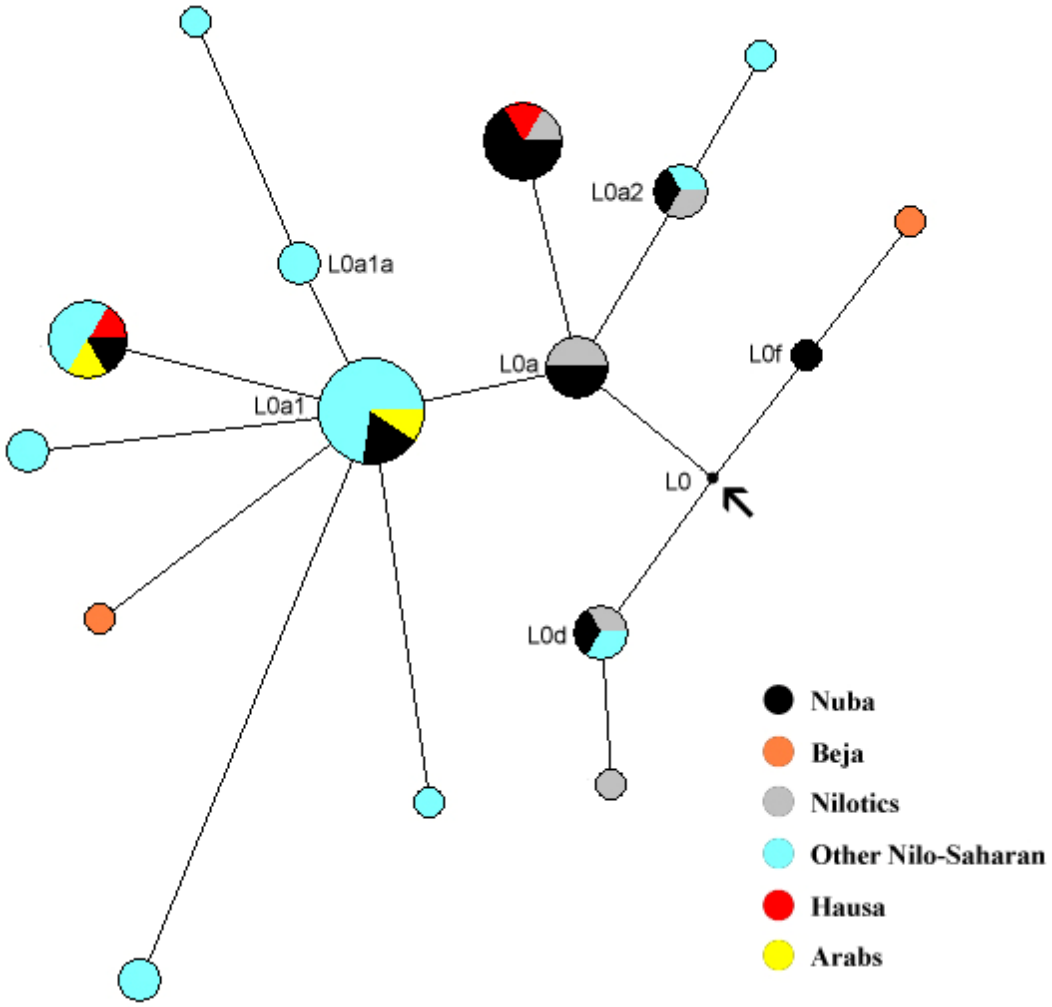


Fig. 3.44. Median network of Sudanese populations derived from mtDNA sequences of haplogroup L0. Node sizes are proportional to haplotypes frequencies, mutation positions are not shown. Arrow indicates the root of the network.

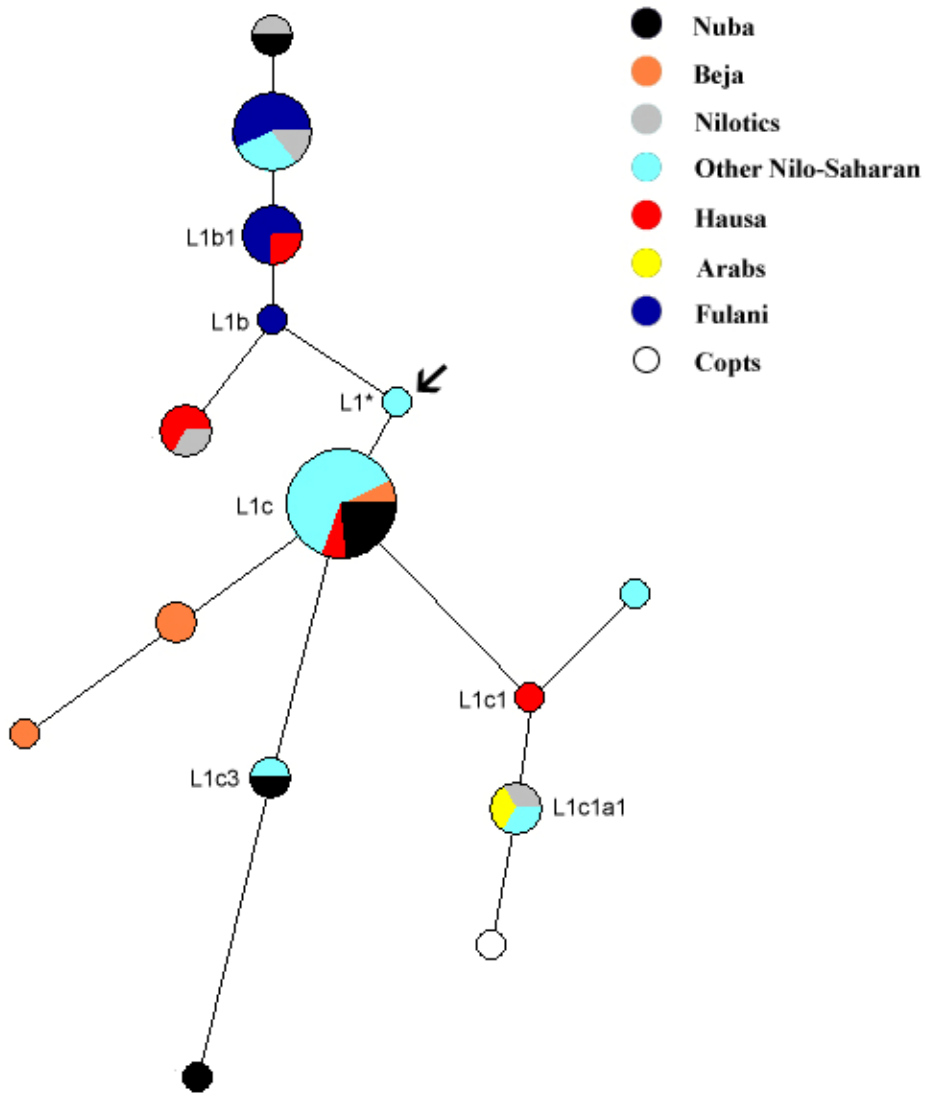


Fig. 3.45. Median network of Sudanese populations derived from mtDNA sequences of haplogroup L0. Node sizes are proportional to haplotypes frequencies, mutation positions are not shown. Arrow indicates the root of the network.

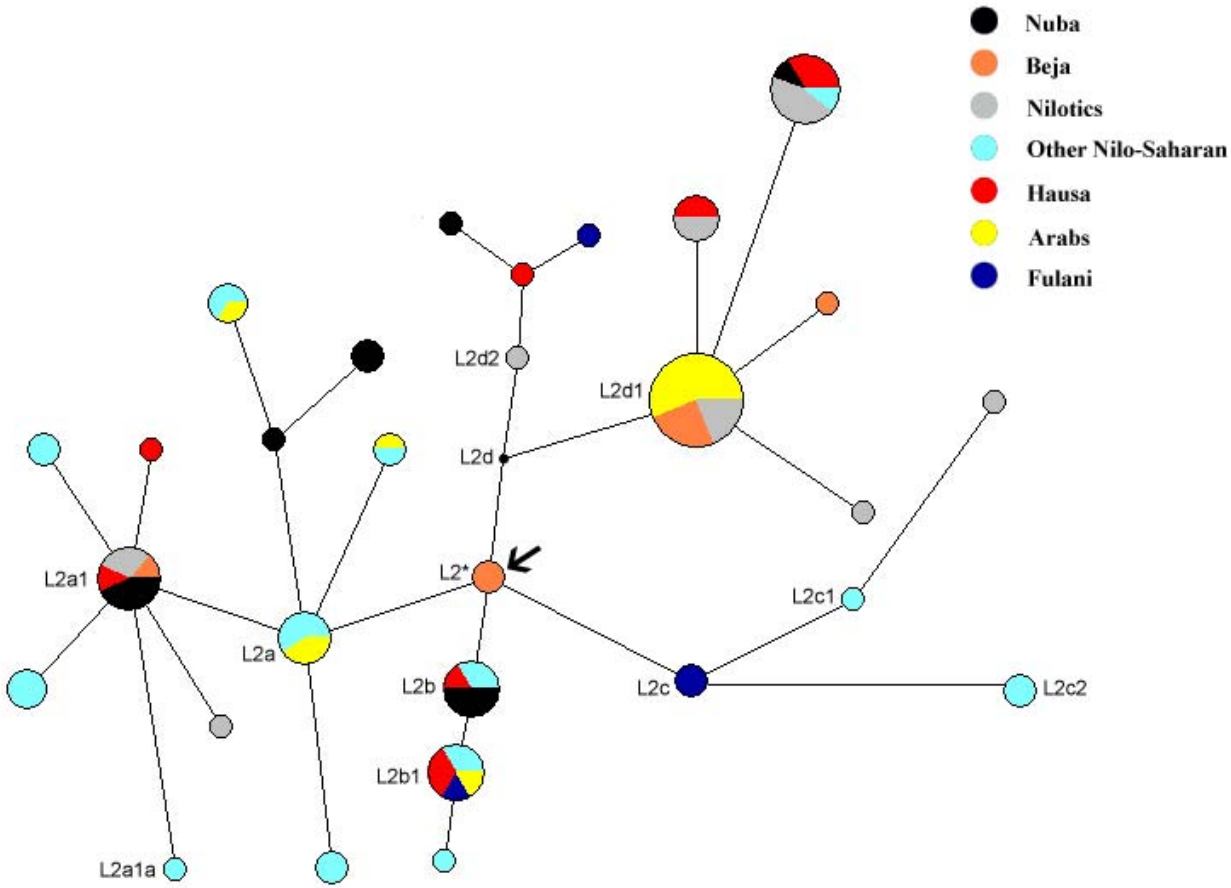


Fig. 3.46. Median network of Sudanese populations derived from mtDNA sequences of haplogroup L2. Node sizes are proportional to haplotypes frequencies, mutation positions are not shown. Arrow indicates the root of the network.

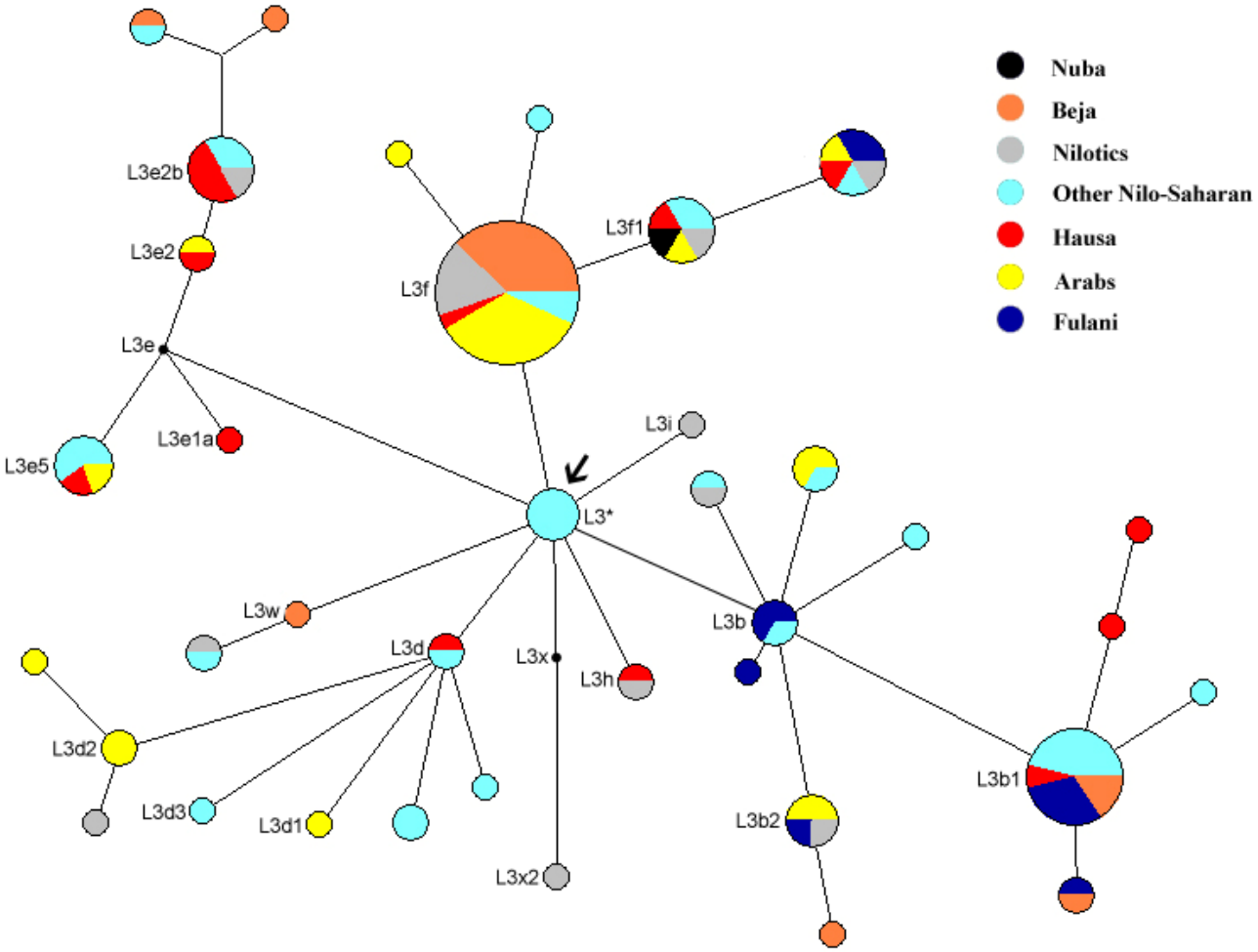


Fig. 3.47. Median network of Sudanese populations derived from mtDNA sequences of haplogroup L3. Node sizes are proportional to haplotypes frequencies, mutation positions are not shown. Arrow indicates the root of the network.

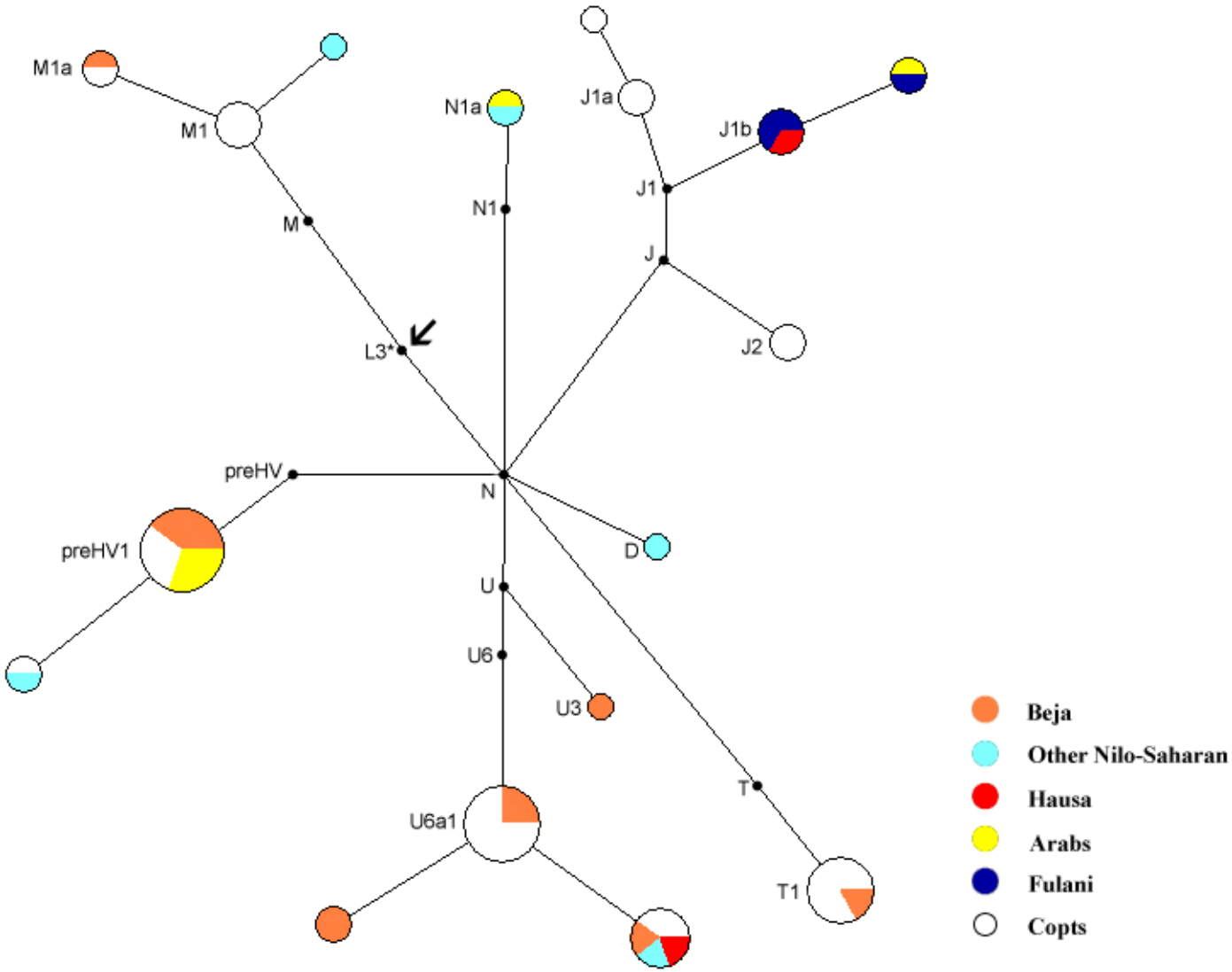


Fig. 3.48. Median network of Sudanese populations derived from mtDNA sequences of haplogroup M plus N. Node sizes are proportional to haplotypes frequencies, mutation positions are not shown. Arrow indicates the root of the network.

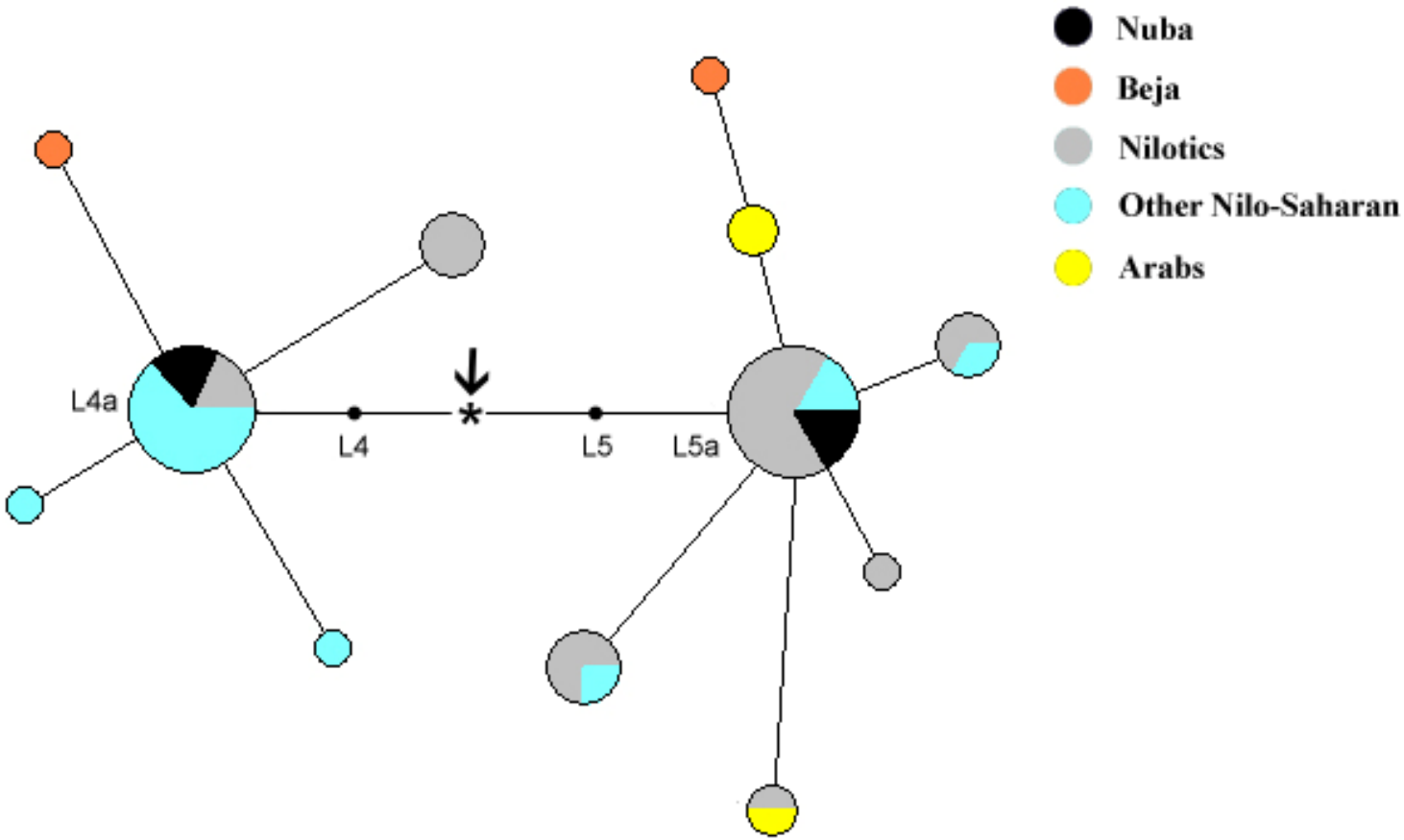


Fig. 3.49. Median network of Sudanese populations derived from mtDNA sequences of haplogroup L4 plus L5. Node sizes are proportional to haplotypes frequencies, mutation positions are not shown. Arrow indicates the root of the network.

CHAPTER 4

DISCUSSION

Several questions pertaining to the pattern of succession of the different groups and cultures in early Sudan have been raised with the hope of acquiring clues into the history of the Nile Valley, state formation and main demographic and migration events. The aim of this study is to draw a portray of the pattern of peopling of the Sudan through identifying and characterizing the skeletal remains found in archeological burial sites in Sudan by ancient DNA analysis, and investigating the genetic structure of the present Sudanese populations, using Y-chromosome and mitochondrial DNA markers. The results of the extraction of aDNA from the naturally mummified Nuri mummy from Christian period showed that both bone and soft tissue samples are suitable for aDNA analysis, but skeletal remains obtained a better quantity of aDNA rather than soft tissue; this could be interpreted as a result of binding of DNA to hydroxyapatite which slows DNA degradation. Silica-Guanidine Theocyanate (GnSCN) method showed that it is a suitable procedure to isolate aDNA although caution must be taken during aDNA extraction to remove all silica during the washes because silica itself is a strong PCR inhibitor. One possible explanation for the negative results of aDNA extraction of all samples collected from Dinder National Park in eastern Sudan is the geographical nature of that part of the country, Dinder Park is a humid and muddy area, and environmental factors may lead to DNA degradation through processes such as oxidation and hydrolysis. In contrast northern Sudan is a part of the great Sahara, this area is the most arid region in the Sudan, the rapid desiccation of the samples may protect DNA from hydrolytic damage, Nuri mummy is a good example for this process.

Haplogroup A-M13 of Y-chromosome which defines Nilotic and other populations in East Africa (Underhill *et al.*, 2000) was found at high frequencies among Neolithic samples. YAP insertion appeared to be more frequent among Meroitic, post-Meroitic and Christian periods. Haplogroup F-M89 which defines the Eurasian haplotypes was observed among samples from Christian period. Though limited in Number, there is enough data to suggest and to tally with the historical evidence of the dominance by Nilotic elements during the early state formation in the Nile Valley, followed by influx of north/eastern African populations during Meroitic and post-Meroitic, and then by arrival of Eurasian populations throughout the Christian period. In Y-chromosome terms this means in the simplest terms introgression of the YAP insertion (haplogroups E and D), and Eurasian haplogroups which are defined by the ancestral haplogroup F-M89 against a background of haplogroup A-M13.

For extant DNA, the PCA plot of Y-chromosome defines two major genetic episodes for Sudanese populations who speak Nilo-Saharan langauges , one is the predominace of haplogroup A-M13 and B-M60 among those groups and the other is defined by the other important haplogroup which is the E-M78. The frequency and wide spread of haplogroups A-M13 and B-M60 makes an excellent marker for early events in population affinities. And to test for hypothesis such as whether the Nilotics have a continuous history in the Sudan and East Africa or whether there was a re-entery to the Nile basin from the Great Sahara following the deterioration of the climate in that region. Haplogroup B-M60 appears to be more characteristic to the Nile Valley as being more associated with populations along the Nile and may equally give clues into past demographic events. Haplogroup E-M78 the other defining haplogroup in the PCA, has an African origin in northeastern African, with a corridor for bidirectional migrations between northeastern and eastern Africa (Cruciani *et al.*, 2007). Although this haplogroup is common to most

Sudanese populations it has exceptionally high frequency in few populations like those coming from western Suadn and particularly Darfur area and the Beja. The analysis of M78 subclades among Sudanese suggests that two subclades, E-V12 and E-V22, which are very common in northern African (Cruciani *et al.*, 2007), might have been brought to Sudan from North Africa after the progressive desertification of the Sahara around 6,000-8,000 years ago. Sudden climate change might have forced several Neolithic cultures/people to shift northwards to the Mediterranean and southwards to the Sahel and Nile Valley (Dutour *et al.*, 1988; Rando *et al.*, 1998). E-V32 which is supposed to be originated in East Africa is the most frequent subclade among Sudanese. The Masalit possesses by far the highest frequency of haplogroup E-M78 and of the E-V32 subclade, suggesting either a recent bottleneck in the population or a proximity to the origin of this subclade. Both E-V13, which is believed to originate in western Asia with its low frequency in North Africa, and E-V65 of North African origin (Cruciani *et al.*, 2007), were not found among Sudanese, these data suggest that haplogroup E-M78 has been brought to Sudan by two episodes; one from the north after the progressive desertification of the Sahara around 8,000-6,000 years ago, the second one may have come from Ethiopia after the collapse of the kingdom of Meroe on the Nile by the Axumite people during the 4th Century AD (Kobishchanov, 1979). The other group to show high frequency of haplogroup E-V32 are the Beja who shows genetic affinity with Ethiopians including Oromo and Amhara populations (Semino *et al.*, 2002) which consent the historical evidence of contact between Ethiopia and Sudan (Hassan, 1968, 1973; Passarino *et al.*, 1998) and the fact that they speaks two languages that branches from the Afro-Asiatic family, thus reinforcing the strong correlation between linguistic and genetic variation in this study.

A third cluster in the PCA is one that includes disparate groups including most Sudanese populations who speaks languges other than Nilo-Saharan family of languages, with exception of the Nubians. The population size of this group and the migration rate seem to be the highest compared to other Sudanese.

In the PCA plot Turks and African populations defines two main genetic episodes that feature striking concordance with linguistic and geographic variations. One cluster relates to populations who speak languages of the Nilo-Saharan family, the predominant linguistic family in the Sudan across the millennia. This cluster is defined by the predominance of the ancestral haplogroups A-M13 and B-M60, as well as the common and most widely distributed haplogroups E-M78. The second grouping encompasses populations who are essentially speakers of languages belonging to the Afro-Asiatic family, with the exception of Nubians. The placement of the Oromo, who speak a language of the Afro-Asiatic family, in the first cluster is probably due to their possession of high frequencies of A-M13. Although the PC plot places the Beja and Amhara from Ethiopia in one sub-cluster based on shared frequencies of the haplogroup J1, the distribution of M78 subclades indicates that the Beja are perhaps related as well to the Oromo on the basis of the considerable frequencies of E-V32 among Oromo in comparison to Amhara (Cruciani et al., 2007). These findings affirm the historical contact between Ethiopia and eastern Sudan (Hassan, 1968, 1973; Passarino et al., 1998), and the fact that these populations speak languages of the Afro-Asiatic family tree reinforces the strong correlation between linguistic and genetic diversity (Cavalli-Sforza, 1997). Senegalese who speak languages belong to Niger-Congo family fell relatively distant from both clusters in the plot due to the high frequency of haplogroup E-M2 among them which is almost rare among East African populations (Underhill *et al.*, 2000; Semino *et al.*, 2002; Sanchez *et al.*, 2005).

Both PCA and Mantel test indicate a pronounced divide and correlation across linguistic and geographic lines. It seems that gene flow is not only recent but largely of focal nature. According to historical and linguistic data, it has been suggested that the Sudanese populations espically Nubians and Beja were strongly affected by Arab migrations since the coming of Islam from the Arabian Peninsula through Egypt around 651 A.D (Hassan, 1968) and the subsequent fall or disintigration of the old Nubians states. This can explain the high frequency of haplogroup J1 (J-12f2) in the North and East and in a part west Sudan which was obviously brought by migrating Arab tribes from Asia. As haplogroup J-12f2 is the prominent haplogroup among Arabs of Iraq (Al-Zahery *et al*., 2003), North Africa (Arredi *et al.*, 2004; Luis *et al.,* 2004), Levant (Flores *et al.,* 2005) and Arabia (Cadenas *et al.,* 2008).

The group that have the highest population size in fact was the Gaalien from central north Sudan who occupies both a trading cross road and the land of the ancient kingdom of Meroe. The Gaalien exhibits a Y profile that give insights into past and recent migrations to the Sudan, intersetingly they still maintain a low frequency of haplogroup A-M13 and E-M78 and a frequency of haplogroup J-12f2 that is consistent with their Arabic oral tradition and descent. The remarkable presence of J-M172 chromosome in the Nile Valley seems to be a legacy of the influence of Turkish rule in Sudan during 1820-1881 A.D. (Holt, 1961) and Egypt for centuries before that. This haplogroup is quite frquent in Turkey (Al-Zahery *et al.*, 2003; Di Giacomo *et al.*, 2004; Semino *et al.*, 2004).

The strong correlation with geography was evident when strictly nomadic pastoralist groups who are know to have migrated recently to the country were removed from the analysis. These include Meseria and Fulani both nomadic tribes that used to traverse the Sahel and who have managed to settle only recently in the Sudan.

Most of the speakers of Nilo-Saharan languages the major linguistic family spoken in the country show very little evidence of gene flow, with exception of the Nubians, who appear to have a considerable gene flow from Asia and Europe together with the Beja. The Beja who were featured in Egyptian temples, pocessess high frequeny of E-M78 consistent with their East African origin contacts with Ethiopia (Passarino *et al.*, 1998). Beja are subgroups of relatively loosely integrated confederations of endogamous lineages based on assumptions of shared descent and cohabitation in an ancestral territory (Dahl and Hjort-af-Ornas, 2006). Both Beja and Nubians lie at entering ports of the Sudan, the Beja in the Red Sea area where past and recent settelments are evident including Turks and Arabs and the Nubians along the Nile in boarder with Egypt where successive waves of migration and conquest of the Sudan have passed. Beja and Nubians were influenced by Arabs as a direct result of the penetration of large numbers of Arabs into the Nile Valley over long period of time (MacMichael, 1967).

The other groups with a large population size are the Hausa and Copts. The Hausa possesses interestingly large frequencies of the haplogroup R-P25. Other groups with varying frequencies of this haplogroups like the Borgu and Meseria appear to have it from the Hausa by gene flow. A recent study showed that this haplogroup is strongly associated the Sickle cell gene and both markers might have co-introgressed recently to eastern Sahel (Bereir *et al.*, 2007). Copts however have no evidence of recent contact with the Hausa, which raises the possibility of a past presence for the Hausa in Egypt, Hausa, one of the best farming societies in Africa, might have an old history relating to Egypt, also a likely port for a back to Africa migration from Asia. Such presence could be more consolidated with the fact that Hausa speaks a language that is a part of the Afro-Asiatic family of languages and that is closely related to Copts and ancient Egyptian.

Another population with a relatively high effective population size that is unlikely to have been influenced by their recent past history in the Sudan are the Copts. Although the history of Copts in Nubia began earlier after the conversion of Nubians from polytheism to Christianity in the 6th Century (Sudan Studies Association, 2006). After the 7th Century barriers of languages and genes kept both populations relatively isolated.

The Copts population had a most intersting Y profile enough as much as that of the Gaalien for the Sudan, to suggest that they actually represent a living record of the peopling of Egypt. The significant frequency of the B-M60 in this groups might be a relic of a history of coloniziation of southern Egypt probably by Nilotics in the early state formation something that conforms to the recorded history and mythology of Egypt. The interesting genetic structure of Fulani, who possess the lowest population size in this study, effectively consisting of two haplogroups or founding lineages. One of the lineages is R-M173 (54%) as an evidence for back migration form Asia to sub-Saharan Africa. Cruciani *et al.* (2002), suggets either a recent migration of this group or a restricted gene flow due to a linguistic or cultural barrier. The Fulbi language is classified as one of the Niger-Congo of languages which is more prevalent in West Africa and among Bantu speakers, although their Y-chromosome showed very little evidence of western Africa genetic affiliation. The equally high frequencies of the two major haplogroups E-M78 and R-M173 suggest an amalgamation of two populations/cultures that take place some time in past into eastren or central Africa. This is also evident from the frequency of the 'T' allele of lactase persistance gene that is uniquly present in considerable frequencies among the Fulani (Mulcare *et al.*, 2004).

It seems however that the effective size of the pastoralists and nomadic pastoralists is much samller than the groups of sedintary agriculturalists nature. This is intriguing in the sense that one would expects nomadic tribes to be more able to admix, spread and

receive genes than their sedentary counterparts, however this data might point to the fact that population size in human history is largely affected by culture including the formation of states. Both Hausa and Copts descended from long established cultures of city states and empires that have historically expanded drawing into its spin other groups and populations.

Analysis of molecular variance indicate that Y-chromosome variation is significantly partitioned among both geographic and linguistic groups. Therefore, both language and geography are probably important for Sudanese genetic structure. For all studyed populations, *P* values of Tajima's *D* were not significant so the hypothesis of expansion could be rejected at the *P*, 0.05 level, this may be due to the recent expantion of the Sudanese populations. Nm support a larger effective size migration rate and expansion of Afro-Asiatic speaking groups as compared to Nilo-Saharan.

In general, the networking results are consistent with haplogroup designations based on HVR-I sequences and SNP analysis reported in East Africa in the previous studies (Kivisild *et al*. 2004; Gonder *et al*., 2006; Tishkoff *et al*., 2007). Networking suggests that the wide geographic distribution and haplotype diversity of L0a when compared with L0d and L0f could be as a result of recent population growth and/or a recent expansion of haplogroup L0a, and could perhaps reflect the expansion from western Africa into eastern Africa within the past few thousand years (Salas *et al*., 2002). Haplogroup L0d is characteristic of the Khoisan of southern Africa (Bandelt and Forster, 1997; Behar *et al.*, 2008). Interestingly, the presence of this haplogroup in 5 Sudanese samples (1.2%) among Borgu, Nuer and Nuba may represent a remain of a former wider distribution of the Khoisan in East African in the past which has been described previously (Underhill *et al.*, 2000, 2001; Cruciani *et al.*, 2002; Semino *et al.*, 2002; Knight *et al.*, 2003). Nuba harboring the highest frequency of L0 (36.4%) in Sudan as well as the Khoisan (67.6%) in

southern Africa (Behar *et al.*, 2008); this may suggest that both Nuba and Khoisan divergence dates to an early stage in the human history.

The first PC plot shows a one cluster including Hausa, Meseria and most of the Nilo-Saharan speaking groups indicating the high similarity of these ethnic groups, as suggested by their distribution in the phylogenetic tree, their closer neighbors in PC1 are Beja, Gaalien and Fulani, this cluster includes also Hausa and Meseria indicating a probable amalgamation of these populations. The other Afro-Asiatic groups are scattered in the plot with Fulani, Nuba and Nuer. Maternal sub-Sahara African lineages were observed in very high frequencies among Arabs (88.9%) and Beja (70.8%) respectively, whereas Copts, by contrast, carried lower frequency (6.9%) of sub-Saharan lineages suggesting that gene flow has been preferentially into Arab and Beja populations form sub-Saharan Africa populations as a result of female assimilation into the populations. In contrast there is little evidence for male-mediated gene flow from sub-Saharan Africa in the Arabs, taken together; these results are consistent with substantial male migration from Arabia into Sudan. The good example for the gene flow form sub-Saharan populations into Arabs is the high frequency of haplogroup L3f of East African origin (Watson *et al.*, 1997) in Gaalien (45.5%), haplogroup L3f was also observed in high frequencies among Beja (23%) and Dinka (17.2%), which could suggest the gene flow between these groups. The extensive female-mediated gene flow from sub-Saharan Africa into Arab populations 10-20% was observed previously in Arabian Peninsula (Richards *et al.*, 2003; Abu-Amero *et al.*, 2008). The genetic affinity between Nuba and Khoisan has been displayed in the second PC plot, and again most of the Sudanese populations were clustered together as a result of a high maternal gene flow from Nilo-Saharan speaking groups into Afro-Asiatic speaking groups. Nothern African populations with Copts show a genetic similarity with Saudi population; this could be explained as a result of a recent influx from Asia to North

Africa during the last 20,000 years (Foster and Romano, 2007), which is reflected in their lighter skin and European/Middle Eastern physical features (Cavalli-Sforza *et al*., 1994; Olivieri *et al.,* 2006). The closer neighbors in the plot to the Ethiopian (Amhara and Oromo) are Gaalien and Beja; indicating the genetic similarity of the ethnic groups; both Gaalien and Beja harbor a high frequencies of haplogroup L3f which is also common among Ethiopian (Kivisild *et al.,* 2004). Hausa showed the highest genetic diversity in the Sudan, this result is consistent with the Y-chromosome profile of this group, and this could be explained as a result of gene flow from other neighboring African populations in this West African group over time. The high frequency of west/central Africa haplotypes among Sudanese suggests the importance of the Africa's Sahel as a corridor for maternal bidirectional migrations connecting East and West Africa; this hypothesis has been established recently using mtDNA markers (Cerny *et al.*, 2006).

Haplogroup preHV1 has a high frequency in Arabian Pensula as a back flow from India (Abu-Amero *et al.*, 2008), and the remarkable amount of haplogroup preHV1 in Beja, Copts, Arabs and Nubians could be explained as a result of reverse migration from Asia to East Africa and could be brought to Sudan by two events one from North Africa, and the second from Ethiopia (Krings *et al.*, 1999b; Kivisild *et al.*, 2004).

Haplogroup U6 which is the most ancient haplogroup observed in northern Africa is rather frequent among Berbers of Algeria, Morocco and Muritania (Rando *et al.*, 1998; Plaza *et al.*, 2003) but it is rare or absent in sub-Saharan Africa, the remarkable amount of this haplogroup among Beja (10.4%) may be the result of fairly ancient interactions between this group and Berber of north/western Africa, the relation between Beja and Berber has been discribed previously based on genetic analysis by Cavalli-Sforza *et al.* (1994). Both Berber and Beja speak languages belong to Afro-Asiatic family. The possible explanation of the complete absence of haplogroups L6 and HV1 in the sample

analyzed in this study could be due to the low frequencies of these haplogroups among Sudanese as reported previously (Krings *et al.*, 1999b).

Sudanese populations showed unimodal distributions of pairwise differences with one exception of Fulani group, unimodal distributions are interpreted as signs of demographic expansions, although it may also be due to processes other than population expansions; selection or high level of homoplasy may produce the same pattern, because both reduce the correlation between sequences (Excoffier, 1990; Lundstrom *et al.*, 1992), while multimodal distributions are interpreted as signs of constant population size over time (Harpending *et al.*, 1993). Accordingly the unimodal distributions in Sudanese populations could be explained as signs of demographic expansions among these populations, and the multimodal distributions in Fulani suggested constant population sizes over time in this population. Moreover, the peaks observed at 0-3 classes in the mismatch distributions in Fulani and Hausa indicated genetic bottlenecks in these populations in the past (Excoffier and Schneider, 1999), mtDNA analysis of Fulani showed that four haplogroups L3d, U5, H and V which have been detected in nomadic Fulani population in Chad, Cameroon and Burkina Faso (Cerny *et al.*, 2006) were not observed in this group, this may indicate a genetic drift among Fulani in Sudan as a result of migration from West Africa, also Y-chromosome data showed that Fulani has reduced genetic diversity and were passed through a genetic bootleneck. A remarkable finding of genetic diversity of mtDNA among populations analyzed in this study that the great majority of nucleotide polymorphisms are transitions, whereas in autosomes there may be a greater relative frequency of transversions. This agrees with the simple hypothesis that transitions are more likely to occur than transversions as they involve less chemical difference, although the exact reason must involve a more precise biochemical explanation (Cavalli-Sforza, 2007). The difference in the ratio of transitions to transversions between

mtDNA and the autosomes, is likely to reflect the differences between uniparental and bi-parental transmission. DNA under uni-parental transmission has a higher evolutionary rate under genetic drift, as the basic unit of drift: effective population size is four times smaller than for autosomes.

For all analyzed populations , *P* values of Tajima's *D* were weakly significant only in Hausa, Fur and Masalit so the hypothesis of expansion could be accepted at the *P*, 0.05 level, in these groups and rejected for the other Sudanese populations, this could be explained as a result of the recent expansion of Sudanese populations which has been obtained also by mismatch distribution. Nm support a higher migration rate of Nilo-Saharan speaking groups as compared to Afro-Asiatic, this could be interpreted as a result of the larger effective population size of the former.

Analysis of molecular variance showed that when populations are grouped according to linguistic affiliation the level of maternal genetic among-group variation (F_{CT} = 0.02) is low but higher compared to that when populations are grouped according to geographic location (F_{CT} = 0.01), the low values of variation among geographic and linguistic groups indicate that both geography and language have no significant role in shaping the genetic structure of the Sudanese populations. For mtDNA Mantel test reveal no correlation between genetic and linguistic distances, or between genetic and geographic distances, these discrepancies between linguistic and maternal genetic patterns could be explaned through a number of processes: languages can be transmitted form a population to another horizentally without singnificant genetic change, and/or genetic and linguistic evolution may proceed at heterogenious rates (Cavalli-Sforza and Feldman, 1981; Barbujani, 1997; Diamond and Bellwood, 2003). Also Mantel test reveal no correlation between F_{ST} of mtDNA and Y-chromosome of Sudanese populations, this could suggest that matrilineal and patrilineal patterns in Sudan may have had different genetic history, although this

result should be taken with caution as mtDNA sequencing detects high numbers of polymorphisms compared to Y-RFLP method which allows to genotype only few numbers of polymorphisms.

Conclusion

Y-chromosome and mtDNA among Sudanese populations showed that regional variation in Y-chromosome sequences in Sudan is likely to have been shaped by human migrations some of which occurred in the recent past. For example high effective population size of Afro-Asiatic males could be explained by recent higher migration rate and diversity among populations in Afro-Asiatic as compared to Nilo-Saharan speaking groups, and the low male migration rate in Nilo-Saharan may be due to the long history of insitu evolution for Nilo-Saharan speaking groups in East Africa, in contrast the Afro-Asiatic speaking groups are widely distributed in Africa and Asia. The strong concordance between the language and genetics (P=0.007) suggest that language and cultural traits may have played a role in the genetic structure of Sudanese populations. Mitocondrial DNA data suggest that genetic variation among Sudanese population is probable to have been formed by regional insitu evolution of mtDNA in East Africa over time followed by gene flow from Eurasia with Afro-Asiatic groups and from west/central Africa with Fulani. These results suggest that genetic patterns and gene flow in Sudan have been different for males and females both in the old and recent evolutionary past.

REFERENCES

Abbas S, Ali M. 1981. Archaeological survey in the Wadi Hawar Basin. Current Anthropol 22:176-178.

Abbas S, Ali M. 1982. The Neolithic period in the Sudan: c.6000-2500 B.C. Cambridge: BAR International Series 139. Oxford.

Abu-Amero KK, González AM, Larruga JM, Bosley TM, Cabrera VM. 2006. Eurasian and African mitochondrial DNA influences in the Saudi Arabian population. BMC Evol Biol 7:32.

Abu-Amero KK, Larruga JM, Cabrera VM, Gonzalez AM. 2008. Mitochondrial DNA structure in the Arabian Peninsula. BMC Evol Biol 8:45.

Adams WY. 1984. The first colonial empire: Egypt in Nubia 3200-1200 BC. Compar Stud Soc Hist 26:36-71.

Agrawal S, Khan S, Pandey A, Tripathi M, Herrera RJ. 2005. YAP, signature of an African-Middle Eastern migration into northern India. Curr Sci 88:1977-1980.

Akey JM, Sosnoski D, Parra E, Dios S, Hiester K, Su B, Bonilla C, Jin L, Shriver MD. 2001. Melting curve analysis of SNPs (McSNP): A gel-free and inexpensive approach for SNP genotyping. Biotechniques 30:358-367.

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

Alonso S, Amour JA. 2001. A high variable segment of human subterminal 16p reveals a history of population growth for modern humans outside Africa. Pro Natl Acad Sci USA 98:864-869.

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Rose BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG. 1981. Sequence and organization of the human mitochondrial genome. Nature 290:457-465.

Aris-Brosou S, Excoffier L. 1996. The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. Mol Biol Evol 13:494-504.

Arredi B, Poloni ES, Paracchini S, Zerjal T, Fathallah DM, Makrelouf M, Pascali VL, Novelletto A, Tyler-Smith C. 2004. A predominantly Neolithic origin for Y-chromosomal DNA variation in North Africa. Am J Hum Genet 75:338-345.

Avise JC, Arnold J, Ball RM, Bermingham E, lamb T, Neigel JE, Reeb CA, Saunders NC. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annu Rev Ecol Syst 18:489-522.

Bamshad MJ, Watkins WS, Dixon ME, Jorde LB, Rao BB, Naidu JM, Ravi Prasad BV, Rasanayagam A, Hammer MF. 1998. Female gene flow stratifies Hindu castes. Nature 395:651.

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

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

Barbujani G. 1997. DNA variation and language affinities. Am J Hum Genet 61:1011-1014.

Batrawi AM. 1945. The racial history of Egypt and Nubia. Part I. The craniology of lower Nubia from predynastic times to sixth century A.D. J Roy Anthropol Inst 75:81-101.

Batrawi AM. 1946. The racial history of Egypt and Nubia. Part II. The racial relationship of the ancient and modern populations of Egypt and Nubia. J Roy Anthropol Inst 76:131-156.

Bayoumi RA. 1987. The sickle-cell trait modifies the intensity and specificity of the immune response against P. falciparum malaria and leads to acquired protective immunity. Med Hypotheses 22:287-98.

Bayoumi RA, Flatz SD, Kuhnau W, Flatz G. 1982. Beja and Nilotes: Nomadic pastoralist groups in the Sudan with opposite distributions of the adult lactase phenotypes. Am J Phys Anthropol 58:173-178.

Bayoumi RA, Taha TS, Saha N. 1985. A study of some genetic characteristics of the Fur and Baggara tribes of the Sudan. Am J Phys Anthropol 67:363-70.

Behar DM, Villems R, Soodyall H, Blue-Smith J, Pereira L, Metspalu E, Rosaria S, Makkan H, Tzur S, Comas D, Bertranpetit J, Quintana-Murci L, Tyler-Smith C, Wells S, Rosset S, The Genographic Consortium. 2008. The Dawn of Human Matrilineal Diversity. Am J Hum Genet 82:1-11.

Bendall KE, Macaulay VA, Baker JR, Sykes BC. 1996. Heteroplasmic point mutations in the human mtDNA control region. Am J Hum Genet 59:1276-1287.

Bereir RE, Hassan HY, Salih NA, Underhill PA, Cavalli-Sforza LL, Hussain AA, Kwiatkowski D, Ibrahim ME. 2007. Co-introgression of Y-chromosome haplogroups and the sickle cell gene across Africa's Sahel. Eur J Hum Genet 15:1183-1185.

Bereir RE, Mohamed HS, Seielstad M, El Hassani AM, Khalil EA, Peacock CS, Blackwell JM, Ibrahim ME. 2003. Allele frequency and genotype distribution of polymorphisms within disease-related genes is influenced by ethnic population sub-structuring in Sudan. Genetica 119:57-63.

Bermisheva M, Tambets K, Villems R, Khusnutdinova EK. 2002. Diversity of mitochondrial DNA haplogroups in ethnic populations of the Volga-Ural region. Mol Biol 6:990-1001.

Berry AC, Berry RJ. 1972. Origins and relationships of the ancient Egyptians based on a study of non-metrical variations in the skull. J Hum Evol 1:199-208.

Blench R. 1993. Recent developments in African language classification. In: Shaw T, Sinclair P, Andah B, Okpoko A (eds). The archaeology of Africa: foods, metals and towns. Routledge, London.

Bonnet C. 1990. Kerma, Royaume de Nubie. Geneva: Muse´e d'Art et d'Histoire.

Bonnet C. 2004. Le Temple Principal de la Ville de Kerma et son Quartier Religieux. Paris: Errance.

Boom R, Sol CJ, Salimans MM, Jansen C, Wertheim-van PM, van der Noordac J. 1990. Rapid and simple method for purification of nucleic acid. J Clin Microbiol 28:495-503.

Brace CL, Tracer DP, Yaroch LA, Robb J, Brandt K, Nelson R. 1993. Clines and clusters versus 'race': a test in ancient Egypt and the case of a death on the Nile. Am J Phys Anthropol 36:1-31.

Browne G. 2002. Old Nubian Grammar. Munich: Lincom Europa.

Calafell F, Shuster A, Speed WC, Kidd JR, Kidd KK, 1998. Short tandem repeat polymorphism evolution in humans. Eur J Hum Genet 6:38-49.

Calloway CD, Reynolds RL, Herrin GL, Anderson W. 2000. The frequency of heteroplasmy in the HVII region differs across tissue types and increases with age. Am J Hum Genet 66:1384-1397.

Cadenas AM, Zhivotovsky LA, Cavalli-Sforza LL, Underhill PA, Herrera RJ. 2008. Y-chromosome diversity characterizes the Gulf of Oman. Euro J Hum Genet 16:374-386.

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

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

Carlson DS. 1976. Temporal variation in prehistoric Nubian crania. Am J Phys Anthropol 45:467-84.

Carlson DS, Van Gerven DP. 1979. Diffusion, biological determinism, and biocultural adaptation in the Nubian corridor. Am Anthropologist 81:561-80.

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. 1997. Genes, peoples and languages. PNAS 94:7719--7724.

Cavalli-Sforza LL. 2007. Human evolution and its relevance for genetic epidemiology. Annu Rev Genomics Hum Genet 8:1-15.

Cavalli-Sforza LL, Feldman MW. 1981. Cultural Transmission and Evolution. Princeton: Princeton University Press.

Cavalli-Sforza LL, Feldman MW. 2003. The application of molecular genetic approaches to the study of human evolution. Nat Genet 33:266-75.

Cavalli-Sforza LL, Menozzi P, Piazza A. 1993. Demic expansions and human evolution. Science 259:639-646.

Cavalli-Sforza LL, Menozzi P, Piazza A. 1994. History and geography of human genes. Princeton Uni. Press, Princeton, NJ.

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

Cavalli-Sforza LL, Wang WS. 1986. Spatial distance and lexical replacement. Language 62:38-55.

Cerny V, Hajek M, Bromova M, Cmejla R, Diallo I, Brdicka R. 2006. The mtDNA of Fulani nomads and their genetic relationships to neighbouring sedentary populations. Hum Biol 78:9-27.

Cerny V, Salas A, Hajek M, Zaloudkova MZ, Brdicka R. 2007. A Bidirectional Corridor in the Sahel-Sudan Belt and the Distinctive Features of the Chad Basin Populations: A History Revealed by the Mitochonderial DNA Genome. Ann Hum Genet 71:433-452.

Chen JT, Sokal RR, Ruhlen M. 1995. Worldwide analysis of genetic and linguistic relationships of human-populations. Hum Biol 67:595-612.

Chen YS, Olckers A, Schurr T G, 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.

Clark JD. 1994. Africa from the appearance of Homo sapiens sapiens to the beginnings of food production. In: De Laet SJ (ed) The history of humanity. Vol. I. Routledge, London, pp 191-206.

Clement M, Posada D, Crandall KA. 2000. TCS: a computer program to estimate gene genealogies. Mol Evol 9:1667-1660.

Cockerham CC, Weir BS. 1993. Estimation of gene flow from F-statistics. Evolution 47:855-863.

Cone RW, Fairfax MR. 1993. Protocol for ultraviolet irradiation of surface to reduce PCR contamination. PCR Methods Appl 3:15-17.

Crichton JM. 1966. A multiple discriminant analysis of Egyptian and African Negro crania. Papers of the Peabody Museum of Archaeology and Ethnology. 57:43-67.

Cruciani F, Fratta RL, Torroni A, Underhill PA, Scozzari R. 2006. Molecular dissection of the Y chromosome haplogroup E-M78 (E3b1a): a posteriori evaluation of a microsatellite-network-based approach through six new biallelic markers. Hum Mutat 27:831-832.

Cruciani F, Fratta RL, Trombetta B, Santolamazza P, Sellitto D, Colomb EB, Dugoujon JM, Crivellaro F, Benincasa T, Pascone R, Moral P, Watson E, Melegh B, Barbujani G, Fuselli S, Vona G, Zagradisnik B, Assum G, Brdicka R, Kozlov A, Efremov GD, Coppa A, Novelletto A, Scozzari R. 2007. Tracing past human male movements in northern/eastern Africa and western Eurasia: new clues from Y-chromosomal haplogroups E-M78 and J-M12. Mol Biol Evol 24:1300-1311.

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

Dahl G, Hjort-af-Ornas A. 2006. Precolonial Beja: A Periphery at the crossroads. Nordic J Afr Stud 15:473-498.

Davies WV. 2003. Sobeknakht of ElKab and the coming of Kush. Egypt Archaeol. 23:3-6.

Deacon HJ. 1989. Late Pleistocene palaeolecology and archaeology in the Southern Cape, South Africa. In: Mellars P, Stringer C (eds) The human revolution: behavioural and biological perspectives in the origins of modern humans. Edinburgh University Press, Edinburgh. p 547-564.

Deka R, Jin L, Shriver MD, Mei Yu L, Saha N, Barrantes R, Chakraborty R, Ferrell RE. 1996. Dispersion of human Y chromosome haplotypes based on five microsatellites in global populations. Genome Res 6:1177-1184.

Destro-Bisol G, Coia V, Boschi I, Verginelli F, Caglia A, Pascali V, Spedini G, Calafell F. 2004a. The analysis of variation of mtDNA Hypervariable region 1 suggests that Eastern and Western Pygmies diverged before the Bantu expansion. Am Nat 163:212-226.

Destro-Bisol G, Donati F, Coia V, Boschi I, Verginelli F, Caglià A, Tofanelli S, Spedini G, Capelli C. 2004b. Variation of female and male lineages in sub-Saharan populations: the importance of sociocultural factors. Mol Biol Evol 21:1673-1682.

Di Giacomo F, Luca F, Popa LO, Akar N, Anagnou N, Banyko J, Brdicka R, Barbujani G, Papola F, Ciavarella G, Cucci F, Di Stasi L, Gavrila L, Kerimova MG, Kovatchev D, Kozlov AI, Loutradis A, Mandarino V, Mammi´ C, Michalodimitrakis EN, Paoli G, Pappa KI, Pedicini G, Terrenato L, Tofanelli A, Mapaspina P, Novelletto A. 2004. Y chromosome haplogroup J as a signature of the post-neolithic colonization of Europe. Hum Genet 115:357-374.

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

Diamond J, Bellwood P. 2003. Farmers and their languages: the first expansions. Science 300:597-603.

Divale W. 1984. Matrilocal residence in pre-literate society. Ann Arbor: UMI Research Press.

Donoghue HD, Sigelman M, Zias J, Gernaey-Child Am, Minnikin DE. 1998. Mycobacterium tuberculosis complex DNA in calcified pleura from remains 1400 years old. Lett Appl Microbiol 27:265-269.

Dupanloup I, Pereira L, Bertorelle G, Calafell F, Prata MJ, Amorim A, Barbujani G. 2003. A recent shift from polygyny to monogamy in humans is suggested by the analysis of worldwide Y-chromosome diversity. J Mol Evol 57:85-97.

Dutour O, Vernet R, Aumassip G. 1988. Le peuplement pre! historique du Sahara. In Milieux, hommes et techniques du Sahara preU historique. Proble[mes actuels (eds. Aumassip, G. et al.), Paris.

Edwards DN. 1996. The Archaeology of the Meroitic State: New Perspectives on its Social and Political Organisation. Oxford: Archaeopress.

Edwards DN. 1998. Meroe and the Sudanic kingdoms. J Afr Hist 39:175-93.

Edwards DN. 2007. The archaeology of Sudan and Nubia Ann Rev Anthropol 36:211-28.

Excoffier L. 1990. Evolution of human mitochondrial DNA: Evidence for departure from a pure neutral model of populations at equilibrium. J Mol Evol 30:125-139.

Excoffier L, Harding RM, Sokal RR, Pellegrini B, Sanchez-Mazas A. 1991. Spatial differentiation of Rh and Gm haplotype frequencies in sub-Saharan Africa and its relation to linguistic affinities. Hum Biol 63:276-307.

Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1:47-50.

Excoffier L, Schneider S. 1999. Why hunter-gatherer populations do not show signs of Pleistocene demographic expansions. Proc Natl Acad Sci USA 96:10597-10602.

Fadhlaoui-Zid K, Plaza S, Calafell F, Ben Amor M, Comas D, Bennamar El Gaaied A. 2004. Mitochondrial DNA heterogeneity in Tunisian Berbers. Ann Hum Genet 68: 222-233.

Fagan BM. 1987. The great journey: The peopling of ancient America. Thames and Hudson, London.

Fagundes NJ, Bonatto SL, Callegari-Jacques SM, Salzano FM. 2002. Genetic, geographic, and linguistic variation among South American Indians: possible sex influence. Am J Phys Anthropol 117: 68-78.

Farid S. 1963. Excavations at Ballana 1958-1959. Cairo: Gov Print Off.

Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783-791.

Flores C, Maca-Meyer N, Larruga JM, Cabrera VM, Karadsheh N, Gonzalez AM. 2005. Isolates in a corridor of migrations: a high-resolution analysis of Y-chromosome variation in Jordan. J Hum Genet 50:435-441.

Forster P. 2004. Ice Ages and the mitochondrial DNA chronology of human dispersals: Philos Trans R Soc Lond B Biol Sci 359:255-64.

Forster P, Romano V. 2007. Timing of a back-migration into Africa. Science. 316:50-51.

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

Gagnon A, Toupance B, Tremblay M, Beise J, Heyer E. 2006. Transmission of migration propensity increases genetic divergence between populations. Am J Phys Anthropol. 129:630-636.

Gatto M. 1997. Regional variations in the so-called “A-group” culture of lower Nubia. In dynamics of populations, movements and responses to climatic change in Africa. p 105-11. Rome: Bonsignori Editore.

Gill P, Ivanov PL, Kimpton C, Piercy R, Benson N, Tully G, Evett I, Hagelberg E, Sullivan K. 1994. Identification of the remains of the Romanov family by DNA analysis. Nat Genet 6:130-135.

Gonder MK, Mortensen HM, Reed FA, de Sousa A, Tishkoff SA. 2007. Whole-mtDNA genome sequence analysis of ancient African lineages. Mol Biol Evol 24:757-768.

Gordon RG. 2005. Ethnologue: Languages of the World, Fifteenth edition. Dallas, Tex. SIL International. Online version: http://www.ethnologue.com/

Gotherstrom A, Fischer C, Linde´n K, Lide´n K. 1995. X-raying ancient bone: a destructive method in connection with DNA analysis. Lab Arkeol 8:26-28.

Greenberg JH. 1963. The languages of Africa. Indiana University Press, Bloomington.

Grün R, Stringer CB. 1991. Electron spin resonance dating and the evolution of modern humans. Archaeometry 33:153-199.

Hájek M, Cerný V, Brůžek J. 2008. Mitochondrial DNA and craniofacial covariability of Chad Basin females indicate past population events. Am J Hum Biol 20:465-474.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp 41:95-98.

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, Horai S. 1995. Y chromosomal DNA variation and the peopling of Japan. Am J Hum Genet 56:951-962.

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

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

Handt O, Hoss M, Krings M, Pääbo S. 1994. Ancient DNA: methodological challenges. Experientia 50:524-529.

Handt O, Krings M, Ward RH, Pääbo S. 1996. The retrieval of ancient human DNA sequences. Am J Hum Genet 59:368-376.

Harpending HC, Sherry ST, Rogers AR, Stonking M. 1993. The genetic structure of ancient human populations. Curr Anthropol 34:483-496.

Hassan YF. 1968. Sudan in Africa, Khartoum University press.

Hassan Y F. 1973. The Arabs and the Sudan, Khartoum University press.

Hauswirth WW, Clayton DA. 1985. Length heterogeneity of a conserved displacement-loop sequence in human mitochondrial DNA. Nucleic Acids Res 13:8093-8104.

Hauswirth WW, Van de Walle MJ, Laipis PJ, Olivo PD. 1984. Heterogeneous mitochondrial DNA D-loop sequences in bovine tissue. Cell 37:1001-1007.

Helgason A, Siguroardottir S, Gulcher JR, Ward R, Stefansson K. 2000. mtDNA and the origin of the Icelanders: Deciphering signals of recent population history. Am J Hum Genet 66:999-1016.

Higuchi R, Browman B, Freiberger M, Ryder OA, Wilson AC. 1984. DNA sequences from the quagga, an extinct member of the horse family. Nature 312:282-284.

Hillis DM, Mable BK, Moritz C. 1996. Applications of Molecular Systematics: The state of the Field and a look to the future. In Molecular Systematics, 2nd ed. Sinauer Associates Inc. Sunderland, MA.

Hirano T, Haque M, Utiyama H. 2002. Theoretical and experimental dissection of competitive PCR from accurate quantification of DNA. Anal Biochem 303:57-65.

Hofreiter M, Serre D, Poinar HN, Kuch M, Pääbo S. 2001. Ancient DNA. Nat Rev Genet 2:353-359.

Holt PM. 1961. A modern history of the Sudan. From the Funj sultanate to the present day. New York, Grove Press. p 35-75.

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 92:532-536.

Höss M, Jaruga P, Zastawny TH., Dizdaroglu M, Pääbo S. 1996. DNA damage and DNA sequence retrieval from ancient tissues. Nucleic Acids Res 24:1304-1307.

Hudson RR. 1990. Gene genealogies and the coalescent process. Oxf Surv Evol Biol 9:1-44.

Hühne J, Pfeiffer H, Brinkmann B. 1998. Heteroplasmic substitutions in the mitochondrial DNA control region in mother and child samples. Int J Legal Med 112:27-30.

Hunley K, Merriwether DA. 2002. A method for understanding the causes of patterns of genetic diversity in human populations. Am J Phys Anthropol 34:87.

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

Huson DH. 1998. Splits Tree: analyzing and visualizing evolutionary data. Bioinformatics 14:68-73.

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

Jobling MA, Tyler-Smith C. 2003. The human Y chromosome: An evolutionary marker comes of age. Nat Rev Genet 4:598-612.

Jollife I. 1986. Principal Component Analysis. Springer-Verlag, New York.

Jorde LB. 1980. The genetic structure of subdivided human populations: A review. In: Mielke JH, Crawford MH, editors. Current development in anthropological genetics, volume 1: theory and methods. New York: Plenum Press. p 135-208.

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 66:979-988.

Kaestle FA, Horsburgh KA. 2002. Ancient DNA in anthropology: Methods, Applications, and Ethics. Am J Phys Anthropol 45:92-130.

Kalaydjieva L, Calafell F, Jobling MA, Angelicheva D, de Knijff P, Rosser ZH, Hurles ME, Underhill P, Tournev I., Marushiakova E, Popov V. 2001. Patterns of inter- and intra-group genetic diversity in the Vlax Roma as revealed by Y chromosome and mitochondrial DNA lineages. Eur J Hum Genet 9:97-104.

Kalinowski ST. 2002. Evolutionary and statistical properties of three genetic distances. Mol Ecol 11:1263-1273.

Karafet TM, Osipova LP, Gubina MA, Posukh OL, Zegura SL, Hammer MF. 2002. High levels of Y-chromosome differentiation among native Siberian populations and the genetic signature of a boreal hunter-gatherer way of life. Hum Biol 74:761-789.

Karafet TM, Xu L, Du R, Wang W, Feng S, Wells RS, Redd AJ, Zegura SL, Hammer MF. 2001. Paternal population history of East Asia: Sources, patterns, and microevolutionary processes. Am J Hum Genet 69:615-628.

Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, Harihara S, de Knijff P, Wiebe V, Griffiths RC, Templeton AR, Hammer MF. 1999. Ancestral Asian Source(s) of New World Y-Chromosome Founder Haplotypes. Am J Hum Genet 64:817-831.

Kawatoko M. 1993. On the tombstones found at the Badi' site, the Al-Rih island. Kush 16:186-202.

Kirwan L. 1987. The birth of Christian Nubia: some archaeological problems. Revista degli Stud Oriental 63:119-134.

Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, Pennarun E, Parik J, Geberhiwot T, Usanga E, Villems R. 2004. Ethiopian mitochondrial DNA heritage: tracking geneflow across and around the gate of tears. Am J Hum Genet 75:752-770.

Kivisild T, Shen P, Wall DP, Do B, Sung R, Davis KK, Passarino G, Underhill PA, Scharfe C, Torroni A, Scozzari R, Modiano D, Coppa A, de Knjiff P, Feldman MW, Cavalli-Sforza LL, Oefner PJ. 2006. The role of selection in the evolution of human mitochondrial genomes. Genetics 172:373-387.

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

Kobishchanov Y. 1979. Axum. University Park: Pennsylvania State University Press.

Krings M, Geisert H, Schmitz RW, Krainitzki H, Pääbo S. 1999a. DNA sequence of the mitochondrial hypervariable region II from the Neandertal type specimen. Proc Natl Acad Sci USA 96:5581-5585.

Krings M, Salem AH, Bauer K, Geisert H, Malek AK, Chaix L, Welsby D, Di Rienzo A, Utermann G, Sajantila A, Pääbo S, Stoneking M. 1999b. mtDNA analysis of Nile River Valley populations: A genetic corridor or barrier for migration. Am J Hum Genet 64:1166-1176.

Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, Pääbo S. 1997. Neandertal DNA sequences and the origin of modern humans. Cell 90:19-30.

Kumar S, Tamura K, Nei M. 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinform 5:150-163.

Lahr MM, Foley RA. 1994. Multiple dispersals and modern human origins. Evol Anthropol 3:48-60.

Lindahl T. 1993. Instability and decay of the primary structure of DNA. Nature 362:709-715.

Llop E. 1996. Genetic composition of Chilean aboriginal populations: HLA and other genetic marker variation. Am J Phys Anthropol 101:325-332.

Luis JR, Rowold DJ, Regueiro M, Caeiro B, Cinnioğlu C, Roseman C, Underhill PA, Cavalli-Sforza LL, Herrera RJ. 2004. The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. Am J Hum Genet 74:32-544.

Lum JK, Cann RL, Martinson JJ, Jorde LB. 1998. Mitochondrial and nuclear genetic relationships among Pacific Island and Asian populations. Am J Hum Genet 63:613-624.

Lundstrom R, Tavaré S, Ward RH. 1992. Estimating substitution rates from molecular data. Proc Natl Acad Sci USA 89:5961-5965.

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

MacMichael HA. 1967. A History of the Arabs in the Sudan. Cambridge University Press, Cambridge, vol. II.

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.

Mantel N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Res 27:209-220.

Marchington DR, Poulton J, Sellar A, Holt IJ. 1996. Do sequence variants in the major non-coding region of the mitochondrial genome influence mitochondrial mutations associated with disease. Hum Mol Genet 5:473-479.

Maynard SJ. 1989. Evolutionary Genetics. Oxford University Press, Oxford.

McBrearty S, Brooks A. 2000. The revolution that wasn't: a new interpretation of the origin of modern human behavior. J Hum Evol 39:453-563.

McDougall I, Fleagle JG. 2005. Stratigraphic placement and age of modern humans from Kibish, Ethiopia. Nature 433:733-736.

McGregor AJ. 2001. Darfur (Sudan) in the Age of Stone Architecture c.AD 1000-1750. Oxford: Archaeopress.

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 USA 100:171–176.

Mohammed AO, Attalla B, Bashir FM, Ahmed FE, El Hassan AM, Ibnauf G, Jiang W, Cavalli-Sforza LL, Karrar ZA, Ibrahim ME. 2006. Relationship of the sickle cell gene to the ethnic and geographic groups populating the Sudan. Community Genet 9:113-20.

Morkot R. 2000. The Black Pharaohs: Egypt's Nubian Rulers. London: Rubicon.

Mulcare CA, Weale ME, Jones AL, Connell B, Zeitlyn D, Tarekegn A, Swallow DM, Bradman N, Thomas MG. 2004. The T allele of a single-nucleotide polymorphism 13.9 kb upstream of the lactase gene (LCT) (C-13.9kbT) does not predict or cause the lactase-persistence phenotype in Africans. Am J Hum Genet 74:1102-1110.

Murdoch GP. 1967. Ethnographic atlas. Pittsburgh: University of Pittsburgh Press.

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

Nei M, Kumar S. 2000. Molecular Evolution and Phylogenetics. Oxford University Press, New York.

Nei M. 1975. Molecular population genetics and evolution. Amsterdam: North Holland/American Elsevier.

Nei M. 1978. Estimation of average heterozygosity 'and genetic distance from a small number of individuals. Genetics 89:583-590.

Ngo KY, Vergnaud G, Johnsson C, Lucotte G, Weissenbach J. 1986. A DNA probe detecting multiple haplotypes of the human Y chromosome. Am J Hum Genet 38:407-418.

Nielsen OV. 1970. Human remains: The Scandinavian Joint Expedition to Sudanese Nubia. Vol. 9. Odensee, Denmark: Scandinavian University Books.

O'Fahey R, Spaulding J. 1974. Kingdoms of the Sudan. London: Methuen.

Olivieri A, Achilli A, Pala M, Battaglia V, Fornarino S, Al-Zahery N, Scozzari R, Cruciani F, Behar DM, Dugoujon JM, Coudray C, Santachiara-Benerecetti AS, Semino O, Bandelt H–J, Torroni A. 2006. The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa. Science 314:1767–1770.

Olson S. 2002. Mapping Human History: Unraveling the mystery of Adam and Eve. Bloomsbury, London. p 28-29.

O'Rourke DH, Hayes MG, Carlyle SW. 2000. Ancient DNA studies in physical anthropology. Annu Rev Anthropol 29:217-242.

Pääbo S. 1985. Molecular cloning of ancient Egyptian mummy DNA. Nature 314:644-45.

Pääbo S. 1989. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. Proc Natl Acad Sci USA 86:1939-1943.

Pääbo S, Wilson AC. 1991. Miocene DNA sequences: a dream come true. Curr Biol 1:45-46.

Page RDM, Holmes EC. 1998. Molecular evolution: a phylogenetic approach. Oxford: Blackwell Science.

Parson W, Parsons TJ, Scheithauer R, Holland MM. 1998. Population data for 101 Austrian Caucasian mitochondrial DNA D-loop sequences: application of mtDNA sequence analysis to a forensic case. Int J Legal Med 111:124-132.

Passarino G, Semino O, Quintana-Murci L, Excoffier L, Hammer M, Santachiara-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.

Perrey C, Turner SJ, Pravica V, Howell WM, Hutchinson IV. 1999. ARMS-PCR methodologies to determine IL-10, TNF-alpha, TNF-beta and TGF-beta 1 gene polymorphisms. Transpl Immunol 7:127-8.

Phillipson DW. 1993. African archaeology. Cambridge University Press, Cambridge.

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

Plaza S, Calafell F, Helal A, Bouzerna N, Lefranc G, Bertranpetit J, Comas D. 2003. Joining the Pillars of Hercules: mtDNA sequences show multidirectional gene flow in the western Mediterranean. Ann Hum Genet 67:312-328.

Posada D, Crandall KA. 2001. Intraspecific gene genealogies: trees grafting into networks. Trends Ecol Evol 16:37–45.

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

Rando JC, Pinto F, González AM, Hernández M, Larruga JM, Cabrera VM, Bandelt HJ. 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.

Ray N, Currat M, Excoffier L. 2003. Intra-Deme Molecular Diversity in Spatially Expanding Populations. Mol Biol Evol 20:76-86.

Rehren T. 2001. Meroe, iron and Africa. Antike Sudan 12:102-109.

Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, Sellitto D, Cruciani F, Kivisild T, Villems R, Thomas M, Rychkov S, Rychkov O, Rychkov Y,

Gölge M, Dimitrov D, Hill E, Bradley D, Romano V, Calì F, Vona G, Demaine A, Papiha S, Triantaphyllidis C, Stefanescu G, Hatina J, Belledi M, Di Rienzo A, Novelletto A, Oppenheim A, Nørby S, Al-Zaheri N, Santachiara-Benerecetti S, Scozari R, Torroni A, Bandelt HJ. 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. Am J Hum Genet 67:1251-1276.

Richards M, Rengo C, Cruciani F, Gratrix F, Wilson JF, Scozzari R, Macaulay V, Torroni A. 2003. Extensive female-mediated gene flow from sub-Saharan Africa into Near Eastern Arab populations. Am J Hum Genet 72:1058-1064.

Richards M, Sykes BC, hedges REM. 1995. Authenticating DNA extracted from ancient skeletal remains. J Arch Sci 22:291-299.

Rilly C. 2007. La Langue du Royaume de Meroe: Un Panorama de la Plus Ancienne Culture Ecrite D'Afrique Subsaharienne. Paris: Honore Champion.

Rogers AR, Harpending H. 1992. Population growth makes waves in the distribution of pairwise genetic differences. Mol Biol Evol 9:552-569.

Rosa A, Brehm A, Kivisild T, Metspalu E, Villems R. 2004. MtDNA profile of West Africa Guineans: towards a better understanding of the Senegambia region. Ann Hum Genet 68:340-352.

Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, Amorim A, Amos W, Armenteros M, Arroyo E, Barbujani G, Beckman G, Beckman L, Bertranpetit J, Bosch E, Bradley DG, Brede G, Cooper G, Côrte-Real HB, de Knijff P, Decorte R, Dubrova YE, Evgrafov O, Gilissen A, Glisic S, Gölge M, Hill EW, Jeziorowska A, Kalaydjieva L, Kayser M, Kivisild T, Kravchenko SA, Krumina A, Kucinskas V, Lavinha J, Livshits LA, Malaspina P, Maria S, McElreavey K, Meitinger TA, Mikelsaar AV, Mitchell RJ, Nafa K, Nicholson J, Nørby S, Pandya A, Parik J, Patsalis PC, Pereira L, Peterlin B, Pielberg G, Prata MJ, Previderé C, Roewer L, Rootsi S, Rubinsztein DC, Saillard J, Santos FR, Stefanescu G, Sykes BC, Tolun A, Villems R, Tyler-Smith C, Jobling MA. 2000. Y-chromosomal diversity in Europe is

clinal and influenced primarily by geography, rather than by language. Am J Hum Genet 67:1526-1543.

Ruhlen M. 1987. A Guide to the Worlds Languages. Stanford, California: Stanford University Press.

Rzhetsky A, Nei M. 1992. A simple method for estimating and testing minimum-evolution trees. Mol Biol Evol 9:945-967.

Saitou N, Nei M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406-425.

Salas A, Richards M, De la Fé T, Lareu MV, Sobrino B, Sánchez-Diz P, Macaulay V, Carracedo A. 2002. The making of the African mtDNA landscape. Am J Hum Genet 71:1082-1111.

Salas A, Richards M, Lareu MV, Scozzari R, Coppa A, Torroni A, Macaulay V, Carracedo A. 2004. The African diaspora: mitochondrial DNA and the Atlantic slave trade. Am J Hum Genet 74:454-465.

Sambrook J, Fritsch EF, Maniatis T. 1990. Molecular Cloning: A laboratory Manual (2nd Edition). Cold Spring Harbor Laboratory, New York.

Sanchez J, Hallenberg C, BØrsting C, Hernandez A, Morling N. 2005. High frequencies of Y chromosome lineages characterized by E3b1, DYS19-11, DYS392-12 in Somali males. Euro J Hum Genet 13:856-866.

Schwartz TR, Schwartz EA, Mieszerski L, McNally L, Kobilinsky L. 1991. Characterization of deoxyribonucleic acid (DNA) obtained from teeth subjected to various environmental conditions. J Forensic Sci 36:979-990.

Seielstad MT, Bekele E, Ibrahim M, Touré A, Traoré M. 1999. A view of modern human origins from Y chromosome microsatellite variation. Genome Res 9:558-567.

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

Semino O, Magri C, Benuzzi G, Lin A, Al-Zahery N, Battaglia V, Maccioni L, Triantaphyllidis C, Shen P, Oefner PJ, Zhivotovsky LA, King R, Torroni A, Cavalli-Sforza LL, Underhill PA, Santachiara-Benerecetti AS. 2004. Origin, diffusion, and differentiation of Y-Chromosome haplogroups E and J: Inferences on the neolithization of Europe and later migratory events in the Mediterranean area. Am J Hum Genet 74:1023-1034.

Semino O, Passarino G, Oefner P, Lin A, Arbuzova S, Beckman L, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti A, Cavalli-Sforza LL, Underhill PA. 2000. The genetic legacy of paleolithic homo 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.

Shen P, Lavi T, Kivisild T, Chou V, Sengun D, Gefel D, Shpirer I, Woolf E, Hillel J, Feldman MW, Oefner PJ. 2004. Reconstruction of patrilineages and matrilineages of Samaritans and other Israeli populations from Y-chromosome and mitochondrial DNA sequence variation. Hum Mutat 24:248-260.

Shen P, Wang F, Underhill PA, Franco C, Yang WH, Roxas A, Sun 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 97:7354-7359.

Shriver MD. 2005. Female migration rate might not be greater than male rate. Eur J Hum Genet 13:131-135.

Simoni L, Calafell F, Pettener D, Bertranpetit J, Barbujani G. 2000. Geographic patterns of mtDNA diversity in Europe. Am J Hum Genet 66:1173-1177.

Slatkin M. 1985. Rare alleles as indicators of gene flow. Evolution 39:53-65.

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

Smith CI, Chamberlain AT, Riley MS, Cooper A, Stringer CB, Collins MJ. 2001. Neanderthal DNA: Not just old but old and cold. Nature 410:771-772.

Smith ST. 1995. Askut In Nubia: The Economics and Ideology of Egyptian Imperialism in the Second Millennium BC. London: Kegan Paul.

Smouse PE, Long JC, Sokal RR. 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. Syst Zool 35:627-632.

Sneath PA, Sokal RR. 1973. Numerical Taxonomy. W.H. Freeman and Company, San Francisco. p 230-234.

Spigelman M, Lemma E. 1993. The use of polymerase chain reaction to detect *Mycobacterium tuberculosis* in ancient skeletons. Int J Osteoarchaeol 3:137-143.

Stevanovitch A, Gilles A, Bouzaid E, Kefi R, Paris F, Gayraud RP, Spadoni JL, El-Chenawi F, Beraud-Colomb E. 2004. Mitochondrial DNA sequence diversity in a sedentary population from Egypt. Ann Hum Genet 68:23-39.

Stoneking M, Hedgecock D, Higuchi RG, Vigilant L, Erlich HA. 1991. Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence specific oligonucleotide probes. Am J Hum Genet 48:370-382.

Street FA, Grove AT. 1976. Environmental and climatic implications of late Quaternary lake-level fluctuations in Africa. Nature 261:385-390.

Stringer C. 2000. Coasting out of Africa. Nature 405:24-27.

Strouhal E. 1969. Temporal and spatial analysis of some craniometric features in ancient Egyptians and Nubians. In population biology of the ancient Egyptians. London and New York: Academic Press.

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

Sudan Studies Association. 2006. 24:2. p. 13-15.

Sullivan KM, Mannucci A, Kimpton CP, Gill P. 1993. A rapid and quantitative DNA sex test: fluorescence-based PCR analysis of X-Y homologous gene amelogenin. BioTechniques 15:637-641.

Tajima A, Pan IH, Fucharoen G, Fucharoen S, Matsuo M, Tokunaga K, Juji T, Hayami M, Horai S. 2002. Three major lineages of Asian Y chromosomes: implications for the peopling of east and southeast Asia. Hum Genet 110:80-88.

Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics 105:437-460.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585-595.

Tajima F, Nei M. 1984. Estimation of evolutionary distance between nucleotide sequences. Mol Biol Evol 1:269-285.

Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512-526.

Tay JS, Saha N. 1988. Genetic heterogeneity among the Negroid and Arab tribes of the Sudan. Am J Phys Anthropol 76:211-5.

Tchernov E. 1992. The Afro-Arabian component in the Levantine mammalian fauna - A short biogeographical review. Israel J Zool 38:155-192.

Thangaraj K, Singh L, Reddy AG, Rao VR, Sehgal SC, Underhill PA, Pierson M, Frame IG, Hagelberg E. 2003. Genetic affinities of the Andaman Islanders, a vanishing human population. Curr Biol 13:86-93.

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, Gonder MK, Henn BM, Mortensen H, Knight A, Gignoux C, Fernandopulle N, Lema G, Nyambo TB, Ramakrishnan U, Reed FA, Mountain JL. 2007. History of Click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation. Mol Biol Evol 24:2180-2195.

Tishkoff SA, Williams SM. 2002. Genetic analysis of African populations: human evolution and complex disease. Nat Rev Genet 3:611-621.

Torroni A, Achilli A, Macaulay V, Richards M, Bandelt HJ. 2006. Harvesting the fruit of the human mtDNA tree. Trends Genet 22:339-345.

Torroni A, Rengo C, Guida V, Cruciani F, Sellitto D, Coppa A, Calderon FL, Simionati 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 69:1348–1356.

Trinkaus E. 2005. Early modern humans. Annu Rev Anthropol 34:207-230.

Tuross N. 1994. The biochemistry of ancient DNA in bone. Experientia 50:530-535.

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

Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonne-Tamir B, Bertranpetit J, Francalacci P, Ibrahim ME, 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 population. Nat Genet 26:358-361.

Van Gerven DP. 1982. The contribution of time and local geography to craniofacial variation in Nubia's Batn el Hajar. Am J Phys Anthropol 49:307-316.

Vigilant L, Pennington R, Harpending HC, Kocher TD, Wilson AC. 1989. Mitochondrial DNA sequences in single hairs from a southern African population. Proc Natl Acad Sci USA 86:9350-9354.

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

Vincentelli I. 2006. Hillat El-Arab. The Joint Sudanese-Italian Expedition in the Napatan Region, Sudan. Oxford Archaeopress.

Wade MJ, Shuster SM. 2004. Estimating the strength of sexual selection from Y-chromosome and mitochondrial DNA diversity. Evol Int J Org Evol 58:1613-1616.

Wallace DC. 1992. Mitochondrial genetics: a paradigm for aging and degenerative diseases. Science 256:628-632.

Wallace DC, Brown MD, Lott MT. 1999. Gene, vol. 238. p 211-230.

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

Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7:256-276.

Weale ME, Shah T, Jones AL, Greenhalgh J,Wilson JF, Nymadawa P, Zeitlin D, Connell BA, Bradman N, Thomas M. 2003. Rare deep-rooting Y chromosome lineages in humans: lessons for phylogeography. Genetics 165:229-234.

Weir BS. 1990. Genetic data analysis. Sunderland, MA: Sinauer.

Welsby D. 1996. The Kingdom of Kush. London: The British Museum Press.

White TD, Asfaw B, DeGusta D, Gilbert H, Richards GD, Suwa G, Howell FC. 2003. Pleistocene Homo sapiens from Middle Awash, Ethiopia. Nature 423:742-747.

Wilder JA, Kingan SB, Mobasher Z, Pilkington MM, Hammer MF. 2004. Global patterns of human mitochondrial DNA and Y-chromosome structure are not influenced by higher migration rates of females versus males. Nat Genet 36:1122-1125.

Wilding D. 1997. Sudan: Ancient Kingdom of the Nile. Flammarion New York.

Wilson AC, Sarich VM. 1969. A molecular timescale for human evolution. Proc Natl Acad Sci USA 63:1088-1093.

Wilson MR, Polanskey D, Butler J, Dizinno JA, Replogle J, Budowle B. 1995. Extraction, PCR amplification and sequencing of mitochondrial DNA from human hair shaft. Biotechniques 18:662-669.

Wood ET, Stover DA, Ehret C, Destro-Bisol G, Spedini G, McLeod H, Louie L, Bamshad M, Strassmann BI, Soodyall H, Hammer MF. 2005. Contrasting patterns of Y

chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes. Eur J Hum Genet 13:867-876.

Wright S. 1951. The genetical structure of populations. Ann Eugen 15:323-354.

Wright S. 1969. Evolution and the genetics of populations: The theory of gene frequencies, vol. 2. University of Chicago Press, Chicago.

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

Ye S, Dhillon S, Ke X, Collins AR, Day IN. 2001. An efficient procedure for genotyping single nucleotide polymorphisms. Nucleic Acids Res 29:88-98.

Zabkar LV. 1975. Apedemak: Lion God of Meroe. Warminster: Aris and Phillips.

Zegura SL, Karafet TM, Zhivotovsky LA, Hammer MF. 2004. High resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas. Mol Biol Evol 21:164-175.

Zietkiewicz EV, Yotova M, Jarnik M, Korab-Laskowska, Kidd KK, Modiano D, Scozzari R, Stoneking M, Tishkoff S, Batzer M, Labuda D. 1997. Nuclear DNA diversity in worldwide distributed human populations. Gene 205:161-171.

APPENDIX I

Solutions & Materials

- Puregene Protein Precipitation Solution (Amonium Acetate).
- Isopropanol (stored at -20°C).
- Sodium Tris EDTA

  100mM sodium chloride

  10mM Tris-HCl

  1mM EDTA.
- 20% Sodium Didocyl Sulphate (SDS)

  SDS 20g

  dH_2O 100ml
- Lysis buffer

  STE+ 20% SDS = 9:1.
- SE buffer (1liter), pH 8.0

  4.39g Sodium Chloride

  0.34g EDTA.
- 1M Tris-HCl, pH 7.5.
- 10% SDS

  25g SDS in 250ml water.
- Proteinase K, 10mg/ml in H_2O.
- Chloroform/Extraction solvent.
- Absolute ethanol.
- DNA Markers (different ranges).
- Ethedium Broromide.
- 0.5 M EDTA

14.61g EDTA pH to 8.0 with NaOH pellets and make up to 100 ml with dH_2O will only loose milky, appearance once pH is correct.

- I0X TE, pH 7.5

  10mM Tris-HCl pH 7.6

  1 mM EDTA.

- 10X TBE

  216g Trisma base

  110g Boric acid

  14.88 EDTA

  Make up to 2 liters with dH_2O.

- 1M Tris-HCl

  12.11g Trisma hydrochloride

  pH to8.0 and make up to 100ml with dH_2O.

- Phosphate Buffered Saline (PBS)

  Dissolve one tablet in 100ml dH_2O

  Autoclave at 115ºC for 10 minutes.

- 10X (Tris EDTA) TE Buffer

  10 mM Tris.Cl (pH 7.4)

  1 mM EDTA (pH 8.0).

- 6X Loading dye

  30 %(v/v) Glycerol

  25% (w/v) Bromophinol Blue

  25% (w/v) Xyline cyanole)

  125% mM EDTA.

- 6M Guanidine hydrochloride (MW = 95.53)

57.318 g in 100ml ddH_2O

Autoclave and store at room temp.

- 7.5M NH_4 acetate (MW = 77.08)

  57.81g in 100ml ddH_2O

  Autoclave and store at room temperature.

- 1M Tris pH 7.5

2 mol/l Tris base	11.2ml
2 mol/l Tris/HCl	38.8ml
ddH_20	50ml

  Store at room temperature.

- 0.5M EDTA pH 8.0 (MW = 372.2)

  18.61g EDTA dissolved into 100ml ddH_2O

  Autoclave and store at room temperature.

- Red Blood Cells (RBCs) lysis buffer

  Dissolve the following in 800ml distilled H_2O

  8.3g NH_4Cl

  1.0g $KHCO_3$

  1.8ml of 5% EDTA

  Filter through 0.2um filtered, complete to 1000ml with distilled H_2O.

- Dimineralization solution:

Proteinase K (Sigma) @20mg/ml	120 ul
0.5 M EDTA pH 8.0 (Promega)	2.28 ml.

- M Tris-HCl Buffer pH 6.4:

Trizma base	15.67g

  Ultrapure water to 1000 ml

pH adjusted to 6.4 with concentrated HCl.

The solution was sterilized by autoclaving at 121°C for 15 minutes.

- Guanidine thiocynate Tris- HCl (GuSCN/ Tris-HCl) Buffer pH 6.4 (L2):

  Guanidine thiocyanate (GuSCN) 11.82g (to give 10M solution)

  0.1 M Tris-HCl Buffer pH 6.4 10ml

  Dissolved by warming

  Stored at room temperature in the dark.

- Lysis buffer (L6):

  10 M GuSCN/ Tris-HCl 10ml

  0.2 M EDTA pH 8.0 (Promga, 0.5M diluted to 0.2M) 2.2ml

  Triton x10 (Promega) 0.13ml (3 drops from pastette).

  Stored at room temperature in the dark.

- Washing buffer (L2):

  GuSCN/ Tris-HCl buffer pH 6.4 as above, without additives.

  Stored at room temperature in the dark.

- 70% ethanol:

  70ml absolute ethanol added to 30ml ultra pure water. Stored at -20° C.

- Acetone stored at –20° C.

APPENDIX II

Y-chromosome and mtDNA Sequences

IUB code defines polymorphic site

R = A or G

Y = C or T

K = G or T

M = A or C

S = G or C

W = A or T

H = A, C or T

M1 (YAP insertion) **site located at position 253 • within the *Alu element (italics)*. Large font bold alleles are specific to the DYS287 Alu element.**

TCACATAATTTCATTTTCCCTATTGCAGATATGTTTTCCTGCATTTGTTAAGGA
ACAAGGGTCTTGAGAGGGAGAGCTTTTTGTCTTAAAGGGGAAGAGATACTTCT
GTGAGGCTAAGAGTTGCCTTTGACTTTGGAGATCTTCACAGGGTATAATAAGA
CAAGCATCAAAGGTAATAGTTTGGGGTCAACTTGACCTGGTTACGTTAATAAG
GAGAGGACTAGCAATAGcaggggaagataaagaaataTA•*GGCCGGGCGCGGTGGCTCAC*
GCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGTGGATCACGAGGTCAGGAG
ATCGAGACCATCCTGGCTAACAAGGTGAAACCCCGTCTCTACTAAAAATACAAAAAA
TTAGCCGGGCGCGGTGGCGGGCGCCTGTATTCCCAGCTACTCGGGAGGCTGAGG
*CAGGAGAATGGCGTGAACCC**A**GGAAGCGGAGCTTGCAGTGAGCCGAGATTGCGC*
*CA**T**tgcagtccgcagtccggccGTCCGGCCTGGGC**A**ACAGAGCGAGACTCCGTCTCAAA…p*
olyA tail

M2 = (203 bp) A to G at position 22

TTTCATTGTTAACAAAAGTCC**R**TGAGATCTGTGGAGGATAAAGGGGGAGCTG
TATTTTCCATTAAAAAAAAAAAAAAAAAGAAAACAGTTTCCAGATTAGGTGG

ACCTAGCCTTCAGTGCAAGTAAAAAGTGCTCTCTGAAGAACAAAGGAAGTAG
AGACAGTTCCTGCACTGGCTC CCAATCAAGTCCATTTACACAAATTA

M13 = (233 bp) G to C at position 157

TCCTAACCTGGTGGTCTTTCATTGTTTTACAAAGGTGATTTAGTTTTGGGAAGG
ACTATTCTCCTTTAAACTATAGACTAAATTTTTCTCAAAGTTAGGTTAGTTTAT
GCCCAGGAATGAACAAGGGCAGTAGGTAGGTTAAGGGCAAGACGGTTA**S**ATC
AGTTCTCTGTTACTGTTATAATTTTCTCATTGTTATATTTTTTGCAAATGTGGTT
GGATAAAATCATGGCTCA

M42 = (340 bp) A to T substitution at position 297

AAAGCGAGAGATTCAATCCAGGATGACAGAATGCGTTCACCTTTAAAGGGAT
TAAAAGAAGTATAATACAGTCTGTATTATTAGATCACCCAGAGACACACAAA
ACAAGAACCGTGAATTGAATTAGTGGTATACTAATAGAGTGGTTTTACCTGAA
ATATTTACACATCAATCCTACTGAATTCTTACAACAAATGATTTAGATTAGCTA
TTGTATTCACCAGTTGAAAGAACAGAAAATATTGAGGGAGATAACTTGTGTCA
GTGCAACTTAATCAGATTTAGGACACAAAAGC**W**ACTACATAATGAAAAAGA
GAGCTGGTGACTTAACTTGCTAAAA

M60 = (388 bp ancestral); **+1 bp insertion** (389 bp = derived). Extra T inserted at position 242

GCACTGGCGTTCATCATCTGGGAGCAGCTCAAAAGCCTCTCGCTCAGCCTCCG
TGACGCCCTGGGGGTGTTCAACCCACATATACTGTAAAGACTAGGAGTAGGGT
TGTGGACACCCCACCTCAGCCAACACTGAGCCCTGATGTGGACTCAACCTTGT
AAGGAAAGCTGTAGAGAAATTGGAAGAAAAAATATAAACACATACAGACTCT
GTCTTTACATTTCAAAATGCATGACTTAAAG**T**ATCAGGCACACAGTGGTTACT
CAATGTTGGTCTGTGTCTCTGTAACGTAATATATGTGACTAAATCCCTAAGCTC
TGCTCTTGACCACCCACCTTCTCCAAAAGGGCCTTTCGTAGACGTCGCTCCTCC
TGAACCATAATGAACAT

M89 = (87 bp) C to T at position 64

ACAGAAGGATGCTGCTCAGCTTCCTGGATTCAGCTCTCTTCCTAAGGTTATGT
ACAAAAATCT**Y**ATGTCTCACTTTGCCTGAGTTGC

M170 = (129 bp) A to C at position 107

AAATAATTTCATGTTTGTTCAAATAATTGCAGCTCTTATTAAGTTATGTTTTCA
TATTCTGTGCATTATACAAATTACTATTTTATTTACTTAAAAATCATTGTTC**M**T
TTTTTTCAGT GTGGGTTGTGT

M175 = (444 bp) 5 bp deletion at interval 84-88 non coding

TTGAGCAAGAAAAATAGTACCCAAATCAACTCAACTCCAGTGATTTAAACTCT
CTGAATCAGGCACATGCCTTCTCACTTCTC**TTCTC**AAGAATGAACAGAAACA
AAGGTATCAGTAGAAAAAAAGGTATCATTAATATTCTTTACTCAAAAGTATTT
CATTTAAAAATACTTACTTTCAGCATTGGACAAAGTACATGGATTACAGTCAA
TCAAGGCTAACTGAAAATGCTGCAAGAGAAAAGTAAAAATATTAATGCACTA
AATTAAGAGTGCATAAAAGTACATTTTCTATTTTAGCCTTTCAATGTCTATCAT
AAAATAACAAAGCTATGCTATACACCAATGCACTACACTCGACCAAATAAAA
TTACTGTAATTCCAAATTTATTTTGAAAATGTAAGTGCTAATCAAGTTATTTCC
CTGAGATAGTTAAGAATGGAG

M11 = (216 p) A to G at position 27

CCCTCCCTCTCTCCTTGTATTCTAAC**R**GAAAGGTTTAGAACTTGCATAATTGGG
AAAGAAGCTGTTGCCTGAACTTACTGGGGGATTCAGCATTGTCATTTTGGACA
TGTCACTTATCCTCAGTATTTGCTTCCCCCAGGAGAGAGCTGTAATAAAAAAG
CATTGCAATTTAATACATAAGCTCAGTAAGTTCTTGTTTATGCTCCTTGTGATG
AA

M74 = (385 bp) G to A at position 195. Resembles M45

ATGCTATAATAACTAGGTGTTGAAGATAAAATCAGTTTAATTTAAATAAGAGG
ATAAAAGAAGTATGAGCAGAAAAAGGTTTTCAATATTAACTAGGAAAGTCTG
AAAAATAATCAGAAATTCTAAAGATAAAAACATAACATTAAAAATTATAAAC
TAAGTTGTTTAATAGATTAGGTATTTTAAAAACTGGT**R**CATTTTTAAGTTGCTT

TAAGTAAGTTACTTAAAAGACAACAGCAGCAAAAGAATTAAAAAAAAATGAA
AGGTGAAGAAACACATACAAGAGAACCTTAGAACAGTAAGGTTCTAGCTAAC
AGGAGAAATAAATTACAGACTGTAAAAGTTGATGACCAAGAATTTTTTCAGA
AGTGGTAAAAGCTGAATT

M23 = (327 bp) A to G at position 24

GTAGAAACATTCTTTCTTTCTAG**R**TGACTAGTGACCAGAATTAAATTGAATCC
TAGGCCACCCATTTATTGTCTTCTGCAGAATTGGCGAGAATGGAGAGGAATCC
TCACCTATCGGTGACCAGAGATGAAATATTCTGAATTGAGAGTTTAAAAGAG

M52 = (289bp) A to C at position 28

TTAATACCTATAAGAATATTGCCTGCA**M**GGATGTTTGATAGGTTTCTTGATAT
TTCATTCTCTCTTGAAATGTTTGCTTCGTCAATTTTTTATTTTTCATAAAAGCTT
TACTCAAGAATTAGGATGACCACTATGTTATCTAGTAGTATAATTTGGAATCA
TTTATTATAAATCTGAAACTGACTTGACTGATGAATCTTGAGCACTGCTTATTG
TGTTCTAATTTTGGTAGCTGTTTCAATAAAGTAGCTCTTGTAGAATAGGAAAA
CTATATCATTTTAGTCTGAAA

M51 = (339 bp) G to A at position 28

TCTATCTCCTGAAGCAGAGTAGACACA**R**GCTTCCAACAGGGATCAGAGTTTA
GGGATCTGGATAGGTATAGAATGGAGCAAAGGGACTAGGCCAAAGGAGATTG
AAAACTGGGGAACAGGGACAAGACTGGAGCTACAAGAAGGACAGGGGCTAG
AAGACAGAAATATG

M33 = (328 bp) A to C at position 26

AGTTTATCTCATAAGTTACTAGTTA**M**ATCTCAGACATATTATACTTTTGTAACT
GAGTGACTCCCATTGTAAGGATAACTACTTCAATGTGCGTATAAATGAGTCAG
TTGTCTCTCTTGGGGGCTTCAACAAATAAGCAAAGATAACCTCATTGTGGAGA
GCACTTCACATTTGTTTTTAGGGTTACATAGTCTACTCTGTATCCTTAAACACT
TGAAGATCTGTTATAACTACATCTGAGATAGTAGTCACAGTGTTTTCTCATGTT
AATGCCTGGCTTCCACCCAGGAGGCACATGTGGTGTGTCTGCAAATAAAGTGT
TTATGAT

M40 = (440 bp) G to A at position 24

CCAGCCCTTCGAGAGGTCAAGGC**R**AGCGGATCACAGGGTGAAGAGATTGAGA
CCATCCTGGCCAACATGGTGAAACTGGGTCTCTACTAAAAATACAAAAAATTA
GCTGGGCGTGGTGACATGTGCCTGTAATCCCAGTTACTCGGGAGGCTGAGGCA
GAAGAATCATTTGAACTCATGAGGCAGAGGTTGCAGTAAGCTGAGATTGCGC
CGCTGCACCCCAGCCTGGCAACAGAGCGAGACTTTGTCTCAAAAAAAATAAA
TAAATAAATAAATAAATAAACAATAATAAAAAAAGCGTAATAGCTAGCCTAT
CCTACCCTATATTCTAAAATTCAAAAGTAATGGTTTTTGTTATGAAATCTCGTA
AGTCTTGCCATAAAGAGAGTTTTCTTCAAATTCGGGTGACACCAACTGCAGGG
CACTTAACTCAGTTGGTGC

M172 = (148 bp) T to G at position 26

ATCCCCCAAACCCATTTTGATGCTT**K**ACTTAAAAGGTCTTCAATTATTATTTTC
TTAAATATTTTGAAAGTCCAAACTTTCTCTGTACCTGGCTGATATTTAAAACTG
GATAAACTGTTCCAAACCAACATGGAGTGAAGATGGATCC

M78 = (319 bp) C to T at position 24

CACTTAACAAAGATACTTCTTTC**Y**GCCCTTCCAAATATTTCAAAATAAGCTGG
TCATAGTACTTGCTTTTCATAAAAAGATGGTAAGCTTCCAATATTTAGATTTAA
GGAAAGGTGAAGGAACACTATAGCTCTTCATTGATAATATCAAGATTTATACT
GTTCCTTTTATCTATCTCCATCAGAGTTTCAAGAGAAAAAAAATACGATGACT
GTCCATATTCCGGTTCACTGACAGATCATGAACCAGTATCATCAAACCTCATT
ATTATCCGAAATATGTGAAGACAACACTGTGTGGGAGAACCTAGGAAAGTAA
T

M173 = (205 bp) A to C at position 18

TCAAGGGCATTTAGAAC**M**CTTTGTCATCTGTTAATATTCAGAAATGATAAGCCAGTGTTTTGTT
TTCAGGATCTGGGAAAACTGCAGCATTTCTTTTACCCATACTGAGTCAGATATATACAGATGGT
CCAGGAGAAGCTTTGAAGGCTGTGAAGGTAAAGGTTTTGTTATAAAATCAGACATTTTTGTTTT
AAAAAGCTTTGCA

M174 = (243 bp) T to C at position 22

TGAATACCTTCTGGAGTGCCC**Y**AGTGCAGAAGTGAGGGGTGCATTTGCAAAA
CTTATAGTGTTTATTGCACACTTTTCCTTGCAAGATGGGTCTTGTCCTTCTCCTT
TTGCATCTCCAGGACCTTCTAGTCAGGTAATTGCATGGCTTTTTTTTTTTTAA
GCAATTAAATACTTAGAAGTAAATTCTGTTTCTGCAAATTCATATTAGCAAAT
TTATGTTCCTCTCTTTGATGTATTCACCA

M9 = (71 bp) G to C at position 19

GGCCTAAGATGGTTGAAT**S**CTCTTTATTTTTCTTTAATTTAGACATGTTCAAAC
GTTCAATGTCTTACATA

P25 = (248 bp) C to A at position 19

CTGCCTGAAACCTGCCTG**M**AAATAGAATTGTGTCTCGTATCTCGGTCCATCAT
GTATGTGATGTGACAACCTTATACTGCCTTGGACAAGCACTTACAGTACATTG
TGACACATAAGTGTGCAATGCACCCAGGTGATGTGATTCTCCTTCCTGGGATC
TGCCAACAGGAAGCATTGTAATGTATCACTTGGCTCAACACCTAAGTTGTGTT
TCTTTTTTGCTTTAGCCTTTGTCTTGATTACAGGGA

12f2 = (+88bp when ancestral, -88bp deletion when derived). *Italics* refers to the fragment which should be absent in the derived sequences.

ATTCCTTTGCCTCCCATTGCCGCTGGTCCCCCATATGAGTTCCTTCACTTACAT
AGCATGAGGAAATTCCTAAGCTGCCAGCTATGTTTTCAGCTAGTTGAGCAAAT
AAGTTTTTGGTTGATAGGGGAAGCTCAGGCA*CTGACTGATCAAAATGCTTACAGA*
TCTAACATGTAAGTCTTTAATCCATCTCAAATTAATTTTTGTATAAGGTGTAAGGAAGG
*GATCC*AGTTTCAGCTTTCTACATATGGCTAGCCAGTTTTCCCAGCACCATTTAT
TAAATAGGGAATGCTTGTTTTTGTCAGGTTTGTCAAAGATCAGATAGTTGTAG
ATATGCGGCATTATTTTTGAGGGCTCTGCTCTGTTCCATTGGTCTATATCTCTG
TGTTGGTACCAGTACCATGCTGTTTTTGTTACTGTAGC

V12 = (439bp) A to G at position 222

CAAAGTTTATTTTCAAAGGGGAGACATAAAATTTGCTCTAATGTCAGTTCTCT
ACAGATGAGAACCAATACTACAGTAACTAAGTCTTGCCTTTTTATCAGTTGGT

TAGTTAGGATGTATATTAGAAAGAAATACTACTGAGTTGGATTGTTTTAAGTT
GAAATCTCAGAAATATACCATATTGAATAAATTCTTACTTATTCCAAGAAAGA
ATAGTGCA**R**ACTTTGAATTATCTACACAAAACATTAAAATGGAAAAAAATGT
ACAGCACATGAAGAGAGACCAAAGAGCTTTTTATGTCTTTTTTTGCTAGTTTCT
TTTTGGAGGGACAGAGGACTCTTTATTCTCCCTCTTGGATCATAAGCAAGCCC
AAGTCATTTAATTATTTTATTTTCATCAAAGGAAATGTTTTCAATCTCCTTCAA
CCCAACTTTATGG

V13 = (235bp) G to A at position 192

GGATGCTGCTAAACATCCTACAGTGTGCATCACAGCCCCCTGCCAACCAAGAC
ATTGCTGTTTATTCTACACTTCCAAGGCCCTGACTTGAAGACTAAATTTGATTT
CAGCCCAGGCTGAAGTTTGAAGCATCATTAGTGGGAAGTGAGATAGCACAAC
AGTGGAGGACAAAGCATGTTACTGTCAGTGCC**R**CAAGGGAAGTTTGAGCCAA
ATTGTTAAGGGATTGAGATGGGAT

V22 = (289bp) T to C at position 84

AATGCCTCAACTTACAGAAATGGAGAAGGGAAAGAATATTGTGCTTCAGCTA
GTTTTCTCGTTACCCAGAAAAGAAAAAAATC**Y**GACCTCTGAAACCTTGGGGAA
TGTCTTCTTAAGATGCCAGAGCCCATATGTCTGCCTAGTCCATGCAGAGCCTG
TACTTCTGCCTGGTCCAGGTGTATGCCACACTGGGCATCACCTCTCTCTGCCCA
AGTGGCACTGGACGATGTCCAAGTCTGGTCAGATCTGGATGGGTTTGATAAAT
AGCTCATGCTGTTTCTGGTCAGTG

V32 = (355bp) G to C at position 197

GCAAAATCCCAGAACATCATTCCCTGTTAGGTAATTGCCAGCAAGTACATGGT
AAGTCAGTGAGTCTGAATTTCTACCCAGAGAACCTGCCCCACTTGGACATACT
ACTGCTCTATAAACCCATAGAGAAGCCACTTCAGCATGTATGTAAGTTCAGAG
CATGCAAATGTTCCATGAATGGTGCACTAGTGGGTGG**S**AGGGTGTGTATATAC
ATGTGTGCACCAGTGAACAAATTCTTAAAGTGCCTCTCTGGCTGGAGAATCAG
ATCTAGCATTTAATGAAAGCTCTCCTAGACTATTAATGTCTCTTCTATTAAAAT
AGTGTGACATAAAGGTTGTCTGCTTTGGGTCAATGA

V65 = (349bp)G to T at position 77

ATGGCCACACAATTCTCCATCACCATATAGGTGTCCACTGAACAACAGAGGGA

TCTGGGCACCAGTCCCCTACACA**K**TCGAAAATCCATGTATAATATTTGACTTC
CCCAAAACATAACTACTAATAGGTTATTGTTGACTAGGATCCTTATTGATATC
ACAAACAGCCAATTAACACATATTTTGTATGTTATATGTATTATATATTCTATT
CTTACAGTAAATAAGCTAGGGAAAAGAAAGTGTTATTGAAAAAATATAAGGA
AAAATACCTTTACTAGTCACTAACTGGAACTGGATCATCACAAAGGTCTTCAT
CCTCTTCATGGTCACATTTAGTAGGTTGAGG

Amelogenin Y = (218pb)

ACCTCATCCTGGGCACCCTGGTTATATCAACTTCAGCTATGAGGTAATTTTCT
CTTTACTAATTTTGATCACTGTTTGCATTAGCAGTCCCCTGGGCTCTGTAAAGA
ATAGTGGGTGGATTCTTCATCCCAAATAAAGTGGTTTCTCAAGTGGTCCCAAT
TTTACAGTTCCTACCATCAGCTTCCCAGTTTAAGCTCTGATGGTTGGCCTCAAG
CCT

Amelogenin X = (212pb)

ACCTCATCCTGGGCACCCTGGTTATATCAACTTCAGCTATGAGGTAATTTTCT
CTTTACTAATTTTGACCATTGTTTGCGTTAACAATGCCCTGGGCTCTGTAAAGA
ATAGTGTGTTGATTCTTTATCCCAGATGTTTCTCAAGTGGTCCTGATTTTACAG
TTCCTACCACCAGCTTCCCAGTTTAAGCTCTGATGGTTGGCCTCAAGCCT

HVR-1 = (446bp)

CTCCACCATTAGCACCCAAAGCTAAGATTCTAATTTAAACTATTCTCTGTTCTT
TCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAA
CCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACC
ATAAATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCC
CTCCCCATGCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCA
ACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCCA
CCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAG
TCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGGGTCCCTTGA
CCACCATCCTCCGTGAAATCA

HVR-II = (423bp)

GGTCTATCACCCTATTAACCACTCACGGGAGCTCTCCATGCATTTGGTATTTTC
GTCTGGGGGGTATGCACGCGATAGCATTGCGAGACGCTGGAGCCGGAGCACC

CTATGTCGCAGTATCTGTCTTTGATTCCTGCCTCATCCTATTATTTATCGCACCT
ACGTTCAATATTACAGGCGAACATACTTACTAAAGTGTGTTAATTAATTAATG
CTTGTAGGACATAATAATAACAATTGAATGTCTGCACAGCCGCTTTCCACACA
GACATCATAACAAAAAATTTCCACCAAACCCCCCCTCCCCCCGCTTCTGGCCA
CAGCACTTAAACACATCTCTGCCAAACCCCAAAAACAAAGAACCCTAACACC
AGCCTAACCAGATTTCAAATTTTATCTTTTGGCGGTATGCA CTTTTAACAG

APPENDIX III

HVR-I and HVR-II Sequences of mtDNA in 404 Individuals belonging to 15 Different Sudanese Populations.

ID	HVR-I motifs	HVR-II motifs	Haplogroup
Arakien			
Ar 1	69, 126, 145, 222, 261	73, 152, 263, 295, 310	**J1b**
Ar 2	189, 209, 223, 245, 273, 278, 282, 294, 300, 306, 310, 311, 336,354, 390, 392	NA	**L2d1**
Ar 3	189, 223, 278, 294, 309, 382, 390, 392	73, 143, 146, 152, 195, 263	**L2d1**
Ar 4	189, 223, 278, 294, 309, 384, 385, 390	73, 143, 146, 152, 195, 263, 310, 324, 350, 366	**L2d1**
Ar 5	189, 223, 278, 294, 309, 375, 390, 392	73, 143, 146, 152, 195, 263	**L2d1**
Ar 6	179, 223, 278, 354, 390, 399	73, 152, 263	**L2d1**
Ar 7	176, 223, 278, 354, 390, 399	73, 152, 263	**L2d1**
Ar 8	186, 189, 192, 223, 278, 294, 309, 390	73, 143, 146, 152, 195, 263	**L2d1**
Ar 9	186, 189, 192, 223, 278, 294, 309, 390	73, 109, 143, 146, 152, 195, 263, 373, 374, 375, 380, 383, 389, 394, 396, 397	**L2d1**
Ar 10	194, 223, 278, 294, 309, 319, 372, 380, 382, 298	73, 107, 143, 146, 152, 195, 263	**L3d1**
Ar 11	124, 256, 352	146, 152, 263, 310	**L3d2**
Ar 12	209, 213, 223, 311	73, 104, 109, 121, 171, 195, 263, 275	**L3f**
Ar 13	209, 213, 223, 311	73, 195, 263, 328	**L3f**
Ar 14	209, 213, 223, 311	73, 104, 109, 121, 171, 195, 263, 275	**L3f**
Ar 15	93, 129, 148, 166, 223, 270, 311	73, 263, 310	**L5a**
Ar 16	147, 172, 223, 248, 355	73, 199, 204, 263	**N1a**
Ar 17	126, 362, 384	152, 263, 310	**preHV1**
Beja			
Bj 1	126, 163, 186, 189, 294	73, 152, 263, 310	**T1**
Bj 2	69, 126, 136, 145, 222, 261	73, 263, 295, 310	**J1b**
Bj 3	129, 148, 168, 172, 187, 188, 189, 223, 230, 311, 320, 372	93, 152, 185, 189, 200, 236, 247, 263, 310	**L0a1**
Bj 4	169, 172, 187, 189, 218, 223, 230, 278, 291, 311, 327, 354, 368	146, 183, 184, 185, 189, 207, 247, 263, 310	**L0f**
Bj 5	189, 223, 278, 292, 294, 311, 385, 398, 399, 400	73, 143, 146, 152, 195, 263, 310	**L1c**
Bj 6	189, 223, 278, 292, 294, 311, 385, 386, 389	73, 143, 146, 152, 263, 310	**L1c**
Bj 7	189, 223, 278, 294, 311, 386	73, 143, 146, 152, 195, 263, 316, 396	**L1c**
Bj 8	189, 223, 278, 294, 311	112, 143, 146, 152, 182, 263, 310	**L1c**
Bj 9	223, 224, 233, 278, 311, 357, 372	73, 98, 146, 152, 182, 189, 247, 265, 310	**L0f**
Bj 10	148, 265, 343, 390, 372	73, 150, 263, 310	**L2***
Bj 11	148, 265, 343, 390	73, 108, 109, 150, 293, 310	**L2***
Bj 12	223, 278, 294, 309, 390	73, 112, 146, 152, 195, 263, 303, 310	**L2a1**
Bj 13	129, 189, 223, 278, 300, 390	73, 150, 195, 263, 303, 310	**L2d1**
Bj 14	189, 192, 223, 278, 292, 294, 309, 390	73, 143, 146, 152, 195, 263, 310	**L2d1**

Continued

Bj 15	129, 189, 223, 249, 261, 311, 359, 380, 390	73, 146, 195, 263, 303, 310	**L2d1**
Bj 16	189, 192, 223, 278, 292, 294, 309, 390	143, 146, 152, 195, 263, 311	**L2d1**
Bj 17	172, 189, 223, 278, 292, 294, 309, 390	73, 143, 146, 152, 195, 263, 310	**L2d1**
Bj 18	124, 189, 223, 278, 293, 304, 319, 362, 399, 400	73, 263	**L3b1**
Bj 19	124, 223, 311, 362	73, 152, 263, 303, 310	**L3b1**
Bj 20	169, 171, 189, 194, 223, 278, 292, 311, 361, 362, 363, 368, 372	73, 150, 204, 263, 303, 310	**L3b2**
Bj 21	189, 223, 249, 311, 319, 320, 323, 324, 357, 357, 375, 384, 386, 396, 399	73, 103, 108, 109, 195, 263, 310, 312, 316	**L3e2b**
Bj 22	172, 189, 223, 278, 294, 309, 372, 375, 386, 400	73, 143, 146, 152, 195, 263, 310	**L3e2b**
Bj 23	209, 213, 223, 311	73, 195, 263, 310, 329	**L3f**
Bj 24	209, 213, 223, 311	73, 195, 263, 310, 329	**L3f**
Bj 25	209, 213, 223, 311	73, 195, 263, 310, 328	**L3f**
Bj 26	209, 213, 223, 311	73, 195, 263, 310, 328	**L3f**
Bj 27	209, 213, 223, 311	73, 195, 263, 310, 328	**L3f**
Bj 28	209, 213, 223, 311	73, 195, 263, 310, 328	**L3f**
Bj 29	209, 213, 223, 311	73, 195, 263, 310, 328	**L3f**
Bj 30	209, 213, 223, 311	73, 195, 263, 310, 328	**L3f**
Bj 31	209, 213, 223, 311	73, 195, 263, 310, 328	**L3f**
Bj 32	209, 213, 223, 311	73, 195, 263, 310, 328	**L3f**
Bj 33	209, 213, 223, 311	73, 195, 263, 310, 328	**L3f**
Bj 34	184, 223, 260, 311	73, 150, 152, 189, 263, 310, 324	**L3w**
Bj 35	172, 223, 287, 293, 362, 372, 399, 311, 355	73, 107, 146, 244, 263, 310	**L4g**
Bj 36	129, 148, 168, 172, 187, 188, 189, 223, 230, 311, 320	93, 152, 185, 189, 200, 236, 247, 263, 303, 310	**L5a**
Bj 37	129, 189, 213, 223, 249, 311, 359, 361	73, 195, 263, 303, 310, 330, 366, 389	**M1**
Bj 38	192, 189, 223, 249, 311, 359, 361, 397	73, 150, 189, 195, 198, 263, 303, 310, 324	**M1a**
Bj 39	126, 362, 375	103, 104, 106, 107, 108, 109, 110, 131, 206, 263, 310, 316	**preHV1**
Bj 40	126, 362	152, 263, 303, 310	**preHV1**
Bj 41	126, 362	103, 107, 109, 152, 263, 303, 310	**preHV1**
Bj 42	126, 362, 372	103, 107, 152, 263, 303, 310, 316, 329, 332, 338, 345, 379, 383, 388, 394, 397, 402	**preHV1**
Bj 43	343	73, 109, 110, 150, 195, 244, 263, 291, 310, 366, 379, 380	**U3**
Bj 44	172, 189, 223, 278, 294, 309	73, 108, 143, 146, 152, 195, 263, 310	**U6a1**
Bj 45	172, 189, 223, 278, 294, 309	73, 132, 143, 146, 152, 195, 263, 310	**U6a1**
Bj 46	172, 189, 223, 278, 294, 309	73, 143, 146, 152, 195, 263, 310	**U6a1**
Bj 47	172, 189, 194, 278	73, 143, 146, 152, 195, 263, 310	**U6a1**
Bj 48	172, 189, 194, 278	73, 143, 146, 152, 195, 263, 310	**U6a1**

Continued

Borgu			
Br 1	129, 148, 168, 172, 187, 188, 189, 223, 230, 311, 320	93, 152, 185, 189, 236, 247, 267, 263, 303, 310	**L0a1**
Br 2	129, 148, 168, 172, 184, 187, 187, 188, 189, 223, 230, 278, 293, 311, 320	93, 95, 152, 185, 189, 200, 236, 247, 263, 310	**L0a2**
Br 3	129, 189, 223, 243, 311	73, 145, 152, 247, 303, 310	**L0d**
Br 4	129, 187, 189, 223, 243, 311	73, 152, 195, 247	**L0d**
Br 5	169, 189, 223, 229, 278, 294, 311	73, 96, 103, 106, 152, 182, 195, 263, 275, 310, 329	**L1c**
Br 6	129, 189, 215, 223, 278, 290, 294, 311	73, 96, 151, 152, 182, 183, 186, 189, 204	**L1c3**
Br 7	189, 223, 229, 274, 278, 291, 294, 311, 321, 322, 383, 390, 400	73, 152, 182, 195, 263, 303, 310	**L2a**
Br 8	189, 192, 223, 266, 278, 294, 355, 390	NA	**L2a**
Br 9	172, 209, 223, 278, 294, 301, 390	143, 152, 182, 195, 263, 310, 324, 348, 366	**L2a**
Br 10	189, 223, 250, 278, 294, 309, 319, 320, 323, 355, 390	73, 109, 110, 143, 146, 152, 195, 263, 303, 310	**L2a1**
Br 11	114, 145, 213, 223, 362, 278, 390	73, 146, 150, 152, 182, 263, 303, 310, 366, 385	**L2b**
Br 12	114, 129, 213, 223, 278, 294, 355, 362, 390	101, 107, 109, 152, 189, 194, 200, 263, 310	**L2b1**
Br 13	223, 264, 278, 390	73, 93, 146, 150, 152, 182, 195, 198, 263, 310, 325	**L2c2**
Br 14	124, 223, 243, 278	73, 150, 263, 303, 310, 398	**L3b**
Br 15	124, 126, 156, 223, 278, 362, 384	73, 152, 263, 310	**L3b1**
Br 16	124, 148, 223, 257, 362	73, 151, 152, 229, 263, 303, 310	**L3b1**
Br 17	124, 223, 235, 311	73, 152, 199, 263, 310, 312	**L3d**
Br 18	124, 189, 223, 278, 304, 311	NA	**L3d3**
Br 19	223	73, 150, 263, 310, 398	**L3e5**
Br 20	136, 209, 223, 311	73, 185, 189, 200, 263, 310	**L3f**
Br 21	117, 154, 159, 260, 223, 311, 317, 347, 362	73, 150, 263, 303, 310, 398	**L4a**
Br 22	93, 114, 129, 213, 223, 278, 355, 362	73, 150, 152, 182, 195, 198, 204, 263, 310	**L4g**
Br 23	223, 293, 311, 355, 362, 399	73, 146, 257, 263, 310	**L4g**
Br 24	129, 148, 166, 187, 189, 223, 278, 311, 355, 362	73, 152, 182, 195, 247, 263, 303, 310	**L5a**
Br 25	223, 278, 318	73, 150, 152, 182, 195, 198, 263, 310, 325	**L2c1**
Copts			
C 1	69, 126, 145, 193, 231, 261, 300, 309	73, 152, 263, 295, 303, 310	**J1a**
C 2	69, 126, 145, 193, 231, 261, 293, 300	73, 152, 263, 295, 303, 310	**J1a**
C 3	69, 126, 145, 193, 231, 261, 300, 309	73, 152, 263, 295, 303, 310	**J1a**
C 4	69, 126, 193, 234, 300	150,152, 295, 310	**J2**
C 5	69, 126, 145, 193, 234, 300, 309	73, 150, 263, 310	**J2**
C 6	69, 126, 145, 193, 234, 309	150,152, 295	**J2**
C 7	129, 189, 223, 230, 243, 265, 269, 293, 310. 311	73, 146, 153, 264, 268, 303, 310	**L1c1a1**

Continued

C 8	129, 189, 223, 230, 243, 265, 269, 293, 310. 311	73, 152, 195, 263, 306, 310	**L1c1a1**
C 9	189, 223, 249, 311	73, 109, 146, 263, 310	**M1**
C 10	189, 175, 223, 249, 311	N/A	**M1**
C 11	189, 223, 249, 311	N/A	**M1**
C 12	192, 189, 223, 249, 263, 311	73, 195, 204, 310	**M1a**
C 13	126, 271, 291, 362	204, 263, 303, 310	**PreHV1**
C 14	126, 362	152, 263, 303, 310	**PreHV1**
C 15	126, 362	152, 263, 310	**PreHV1**
C 16	126, 362	152, 263, 303, 310	**PreHV1**
C 17	117, 126, 135, 163, 175, 176, 180, 294, 311	73, 227, 263, 303, 310	**T1**
C 18	126, 163, 186, 189, 271, 294, 311	146, 263, 310	**T1**
C 19	126, 163, 186, 189, 271, 294, 311	146, 263, 310	**T1**
C 20	117, 126, 135, 163, 175, 176, 180, 294, 311	73, 227, 263, 303, 310	**T1**
C 21	126, 163, 186, 189, 271, 294, 311	73, 146, 263, 310	**T1**
C 22	172, 129, 219, 368	263, 303, 310	**U6a1**
C 23	172, 219, 278	73, 150, 195, 263, 303, 310	**U6a1**
C 24	172, 219, 278	73, 143, 146, 152, 195, 263, 303, 310	**U6a1**
C 25	172, 219, 278	73, 150, 195, 263, 303, 310	**U6a1**
C 26	172, 189, 219, 256, 278, 399	73, 150, 263, 295, 310	**U6a1**
C 27	172, 219, 278	73, 143, 146, 152, 195, 263, 303, 310	**U6a1**
C 28	172, 129, 219, 278, 368	263, 303, 310	**U6a1**
C 29	172, 219, 278	73, 150, 195, 263, 303, 310	**U6a1**
Dinka			
D 1	148, 172, 187, 188, 189, 223, 230, 311, 320	64, 93, 152, 189, 263, 247, 263, 303, 310, 324	**L0a**
D 2	165, 172, 192, 223, 311, 320	73, 146, 263, 303, 310	**L0a**
D 3	189, 223, 239, 311, 320, 362, 380, 389, 395, 396, 399, 400	N/A	**L0a**
D 4	129, 148, 172, 179, 187, 188, 189, 192, 223, 230, 293, 311, 317, 320	64, 93, 151, 152, 185, 189, 236, 263, 310	**L0a2**
D 5	189, 223, 229, 278, 291, 294, 311	73, 152, 182, 195, 263, 303, 310	**L1***
D 6	178, 223, 293, 311, 360, 400	73, 146, 152, 264, 268, 303, 310	**L1c1a1**
D 7	153, 223, 278, 294, 309, 390	73, 143, 146, 152, 195, 263, 310, 398	**L2a1**
D 8	223, 278, 318, 390	73, 93, 146, 150, 152, 182, 195, 198, 263, 310, 325	**L2c1**
D 9	111, 189, 192, 223, 278, 294, 300, 390	73, 143, 146, 152, 195, 263, 263, 303, 310	**L2d1**
D 10	129, 223, 278, 294, 390, 399	73, 143, 146, 150, 195, 263, 310	**L2d1**
D 11	145, 184, 223, 239, 278, 355, 390, 399	73, 146, 150, 152, 263, 310	**L2d2**

Continued

D 12	176, 189, 209, 223, 311	73, 109, 195, 263, 310, 318	**L3f**
D 13	140, 209, 223	73, 152, 263, 310	**L3f**
D 14	189, 197, 209, 223, 311	73, 152, 182, 195, 198, 263, 310	**L3f**
D 15	209, 223, 311	73, 150, 263, 310	**L3f**
D 16	209, 223, 311	93, 103, 109, 152, 185, 189, 236, 247, 263, 310	**L3f**
D 17	117, 124, 153, 223, 319	73, 152, 263, 310	**L3i**
D 18	93, 124, 210, 223, 260, 310, 311, 362, 384	N/A	**L3w**
D 19	223, 293, 311, 355, 362, 399	73, 146, 257, 263, 310	**L4g**
D 20	223, 293, 311, 355, 362, 399, 400	N/A	**L4g**
D 21	129, 148, 166, 187, 189, 223, 278, 311, 355	73, 152, 182, 195, 247 263	**L5a**
D 22	148, 223, 242, 311, 399	73, 146, 150, 152, 263, 310	**L5a**
D 23	129, 148, 166, 187, 189, 223, 278, 311, 355, 362	73, 146, 262, 310	**L5a**
D 24	129, 148, 166, 187, 189, 223, 278, 311, 355, 362	73, 195, 263, 310	**L5a**
D 25	111, 129, 148, 166, 187, 189, 223, 233, 254, 271, 278, 260	73, 195, 263, 310	**L5a**
D 26	148, 154, 166, 187, 189, 223, 233, 254, 271, 278, 260	73, 103, 106, 146, 263, 310	**L5a**
D 27	129, 148, 166, 187, 189, 223, 278, 311, 355	73, 152, 182, 195, 247 263, 313	**L5a**
D 28	129, 148, 166, 187, 189, 223, 278, 311, 355, 362	73, 152, 182, 195, 238, 247, 263	**L5a**
D 29	129, 148, 166, 187, 189, 223, 278, 311, 355, 362	73, 152, 182, 195, 247, 263, 303, 310	**L5a**
Fulani			
F 1	69, 126, 145, 222, 293, 261	263, 310	**J1b**
F 2	114, 126, 187, 189, 223, 264, 270, 278, 293, 311	73, 152, 182, 185, 195, 247, 263, 303, 310, 357	**L1b1**
F 3	126, 187, 189, 223, 264, 270, 278, 293, 311	152, 171, 182, 185, 195, 247, 261, 263, 303, 310, 357, 380	**L1b1**
F 4	126, 187, 189, 223, 264, 270, 278, 293, 311	152, 171, 182, 185, 195, 247, 261, 263, 303, 310, 357, 380	**L1b1**
F 5	114, 126, 187, 189, 223, 264, 270, 278, 293, 311	73, 152, 182, 185, 195, 247, 263, 303, 310, 357	**L1b1**
F 6	126, 187, 189, 223, 264, 270, 278, 293, 311	73, 152, 182, 185, 195, 247, 263, 310, 357	**L1b1**
F 7	93, 126, 187, 189, 223, 264, 270, 278, 293, 311	73, 152, 182, 185, 195, 247, 263, 310, 357	**L1b1**
F 8	126, 223, 264	73, 150, 263, 303, 310	**L1b**
F 9	126, 187, 189, 223, 264, 270, 293, 311	73, 152, 182, 185, 195, 247, 263, 310, 357	**L1b1**
F 10	126, 187, 189, 223, 264, 270, 278, 293, 311	73, 152, 182, 185, 195, 247, 263, 310, 357	**L1b1**
F 11	114, 129, 213, 223, 278, 294, 355, 362, 390	73, 150, 152, 182, 195, 198, 204, 263, 310	**L2b1**
F 12	223, 278, 384, 390	73, 150, 152, 182, 195, 198, 263, 310, 325	**L2c**
F 13	223, 278, 390	73, 93, 150, 152, 182, 195, 198, 263, 310, 325	**L2c**
F 14	223, 278, 390	73, 93, 150, 152, 182, 195, 198, 263, 310, 325	**L2c**
F 15	139, 145, 164, 198, 223, 273, 278, 355, 390	NA	**L2d2**

Continued

F 16	93, 223, 278, 362	73, 263, 310	**L3b**
F 17	93, 223, 278, 262	73, 263, 310	**L3b**
F 18	93, 223, 278, 311, 362	73, 263, 310	**L3b**
F 19	124, 223, 278, 304, 362	73, 146, 152, 195, 263, 303, 310	**L3b1**
F 20	124, 223, 319, 362	73, 146, 152, 195, 263, 303, 310	**L3b1**
F 21	124, 223, 278, 320, 362	73, 151, 152, 263, 267, 303, 310	**L3b1**
F 22	124, 223, 278, 362	73, 151, 152, 263, 267, 310	**L3b1**
F 23	124, 223, 278, 320, 362	73, 151, 152, 263, 267, 310	**L3b1**
F 24	93, 223, 278, 311, 362	73, 263, 310	**L3b2**
F 25	129, 209, 223, 292, 295, 311	73, 189, 200, 263, 310	**L3f1**
F 26	129, 209, 223, 292, 295, 311	73, 189, 200, 263, 310	**L3f1**
F 27	69, 126, 145, 222	263, 310	**J1b**
F 28	69, 126, 145, 222	263, 275, 303, 310	**J1b**
Fur			
Fur 1	223, 293, 311, 355, 362, 399	73, 146, 257, 263, 310	**L4g**
Fur 2	129, 148, 168, 172, 187, 188, 189, 223, 230, 311, 320	93, 152, 185, 189, 236, 247, 267, 263, 303, 310	**L0a1**
Fur 3	114, 129, 148, 162, 168, 172, 187, 188, 189, 223, 230, 311, 320	93, 96, 152, 185, 189, 200, 236, 247, 263, 303, 310	**L0a1**
Fur 4	129, 148, 168, 172, 187, 188, 189, 223, 230, 311, 320	93, 96, 152, 185, 189, 200, 236, 247, 263, 303, 310	**L0a1**
Fur 5	129, 148, 168, 172, 187, 188, 189, 233, 230, 311, 320	93, 152, 184, 189, 200, 236, 247, 263, 303, 310	**L0a1**
Fur 6	93, 129, 148, 168, 172, 187, 188, 189, 209, 223, 230, 278, 293	112, 152, 189, 236, 263, 310	**L0a1a**
Fur 7	93, 129, 148, 166, 187, 189, 223, 247, 265, 278, 311, 355, 362	73, 152, 182, 195, 247, 263, 310	**L0a1a**
Fur 8	189, 223, 229, 278, 291, 294, 311, 320	152, 182, 195, 229, 263, 270, 303, 310	**L1c**
Fur 9	189, 223, 266, 278, 294, 311	73, 143, 146, 152, 195, 263, 303, 310	**L1c**
Fur 10	189, 223, 229, 278, 291, 294, 311	73, 152, 182, 195, 263, 310, 316	**L1c**
Fur 11	189, 223, 293, 311, 360, 400	73, 146, 152, 264, 268, 303, 310	**L1c1a1**
Fur 12	189, 223, 250, 278, 294, 309, 319, 320, 323, 355, 390	73, 109, 110, 143, 146, 152, 195, 263, 303, 310	**L2a1**
Fur 13	93, 189, 223, 278, 294, 309, 390	73, 103, 146, 152, 195, 198, 263, 303, 310, 316, 320	**L2a1**
Fur 14	114, 129, 213, 223, 278, 362, 390	73, 146, 150, 152, 182, 183, 195, 198, 204, 263, 303, 310	**L2b1**
Fur 15	129, 189, 223, 266, 278, 294, 354, 390	73, 143, 146, 152, 195, 263, 303, 310	**L2d1**
Fur 16	124, 204, 205, 223, 346, 362, 370	73, 152, 263, 310	**L3b**
Fur 17	116, 124, 223, 362	73, 152, 263, 310	**L3b1**
Fur 18	189, 194, 195, 223, 320	73, 143, 146, 152, 195, 263, 310	**L3e2b**
Fur 19	192, 209, 223, 292, 295, 311	73, 189, 200, 263, 303, 310	**L3f1**
Fur 20	223, 293, 311, 355, 362, 399	73, 146, 257, 263, 310	**L4g**

Continued

Fur 21	129, 148, 166, 187, 189, 223, 278, 311, 355	73, 152, 182, 195, 247 263	**L5a**
Gaalien			
G 1	136, 223	73, 150, 263, 303, 310, 398	**L3e5**
G 2	178, 223, 293, 311, 399, 400	73, 146, 153, 264, 268, 303, 310	**L1c1a1**
G 3	189, 223, 229, 274, 278, 291, 294, 311, 321, 322, 383, 390	73, 152, 182, 195, 263, 303, 310	**L2a**
G 4	209, 223, 278, 294, 301, 354, 390	73, 143, 146, 152, 189, 195, 263, 310	**L2a**
G 5	179, 223, 223, 256, 284, 311	73, 152, 195, 215, 263, 310	**L3d2**
G 6	124, 256, 352	146, 152, 263, 310	**L3d2**
G 7	157, 209, 213, 223, 311	73, 195, 263, 310	**L3f**
G 8	209, 213, 223, 310, 311	73, 195, 263, 310	**L3f**
G 9	209, 213, 223, 311	73, 195, 263, 310	**L3f**
G 10	209, 223, 292, 310, 311	73, 189, 200, 263, 310	**L3f**
G 11	209, 213, 223, 311	73, 195, 263, 310, 328	**L3f**
G 12	209, 223, 292, 310, 311	73, 189, 200, 263, 310	**L3f**
G 13	209, 213, 223, 311	73, 195, 263, 310, 328	**L3f**
G 14	194, 209, 213, 223, 311, 384	73, 195, 263, 310, 328	**L3f**
G 15	209, 213, 223, 311	73, 195, 263, 310, 328	**L3f**
G 16	209, 213, 223, 311	73, 195, 263, 310, 328	**L3f**
G 17	192, 209, 223, 292, 295, 311	73, 189, 200, 263, 303, 310	**L3f1**
G 18	209, 223, 292, 311	73, 107, 189, 200, 263, 310	**L3f1**
G 19	129, 148, 166, 187, 223, 311	152, 263, 310	**L5a**
G 20	129, 148, 166, 187, 189, 223, 233, 254, 265, 278, 360	73, 152, 195, 263, 310	**L5a**
G 21	126, 362	152, 263, 303, 310	**preHV1**
G 22	126, 362, 384	152, 263, 310	**preHV1**
Hausa			
H 1	129, 162, 172, 187, 223, 230, 311	73, 150, 189, 200, 263, 310	**L0a**
H 2	93, 129, 148, 168, 172, 187, 188, 189, 223, 230, 278, 293, 311, 320	73, 150, 188, 189, 200, 263, 310	**L0a1**
H 3	126, 223, 310, 327, 278, 311	73, 150, 189, 195, 200, 263, 310	**L1b**
H 4	126, 187, 189, 205, 213, 223, 264, 378, 311	73, 150, 189, 195, 200, 263, 310	**L1b**
H 5	126, 187, 223, 264, 270, 278, 293, 311	73, 151, 152, 182, 185, 195, 247, 263, 310, 357	**L1b1**
H 6	187, 189, 223, 264, 270, 278, 311	73, 152, 182, 185, 195, 247, 263, 303, 310, 357	**L1c**
H 7	129, 187, 189, 223, 265, 286, 293, 311, 360	73, 151, 152, 182, 186, 189, 195, 198, 200, 247, 263, 297, 310, 316	**L1c1**
H 8	223, 278, 294, 309, 390	73, 143, 146, 152, 195, 263, 310	**L2a1**
H 9	189, 192, 223, 245, 278, 292, 309, 384, 390, 391	73, 143, 146, 152, 195, 263, 310	**L2a1**

H 10	223, 278, 386, 294, 309, 390	73, 146, 152, 195, 263, 303, 310	**L2a1**
H 11	114, 129, 213, 223, 278, 311, 390	73, 146, 150, 152, 182, 195, 198, 204, 263, 310	**L2b**
H 12	129, 213, 223, 278, 355, 362, 390	73, 150, 152, 182, 195, 198, 204, 263, 310	**L2b1**
H 13	114, 129, 213, 223, 278, 362, 390	73, 146, 150, 152, 182, 183, 195, 198, 204, 263, 303, 310	**L2b1**
H 14	129, 223, 278, 294, 300, 354, 390	73, 150, 195, 263, 310	**L2d1**
H 15	93, 129, 188, 189, 278, 300, 354, 390, 399	73, 146, 150, 195, 263, 303, 310	**L2d1**
H 16	129, 189, 223, 300, 354, 390, 399	73, 150, 195, 263, 310	**L2d1**
H 17	223, 278, 294, 311, 355, 390	73, 143, 146, 152, 195, 263, 310	**L2d2**
H 18	189, 223, 278, 294, 309, 362, 385, 386, 389, 395, 397, 398, 399, 400	73, 146, 152, 195, 263, 303, 310	**L3b**
H 19	124, 223, 278, 311, 262	73, 263, 310	**L3b1**
H 20	124, 189, 223, 278, 362	73, 263, 303, 310	**L3b1**
H 21	124, 157, 223, 278, 362	152, 263, 303, 310	**L3b1**
H 22	124, 223, 311	73, 152, 263, 310	**L3d**
H 23	185, 223, 327	73, 150, 189, 200, 263, 310	**L3e1a**
H 24	223, 320	73, 150, 195, 263, 303, 310, 326, 366	**L3e2***
H 25	189, 194, 195, 223, 320	73, 146, 150, 152, 195, 198, 263, 303, 310	**L3e2b**
H 26	189, 194, 195, 223, 320	73, 143, 146, 150,152, 195, 263, 310	**L3e2b**
H 27	172, 189, 223, 264, 320	73, 150, 263, 310	**L3e2b**
H 28	172, 189, 194, 195, 223, 320	73, 150, 195, 263, 310	**L3e2b**
H 29	223	73, 150, 263, 310	**L3e5**
H 30	209, 223, 292, 311	73, 189, 200, 263, 310	**L3f**
H 31	192, 209, 223, 292, 295, 311	73, 189, 200, 263, 303, 310	**L3f1**
H 32	209, 223, 276, 292, 311	73, 189, 200, 263, 310	**L3f1**
H 33	129, 186, 189, 192, 223, 256, 284, 311	73, 189, 263, 310	**L3h**
H 34	69, 126, 145, 172, 222	242, 263, 295, 310	**J1b**
H 35	172, 183, 189, 194, 219, 278	73, 150, 195, 263, 310	**U6a1**
Masalit			
Mt 1	189, 223, 229, 278, 291, 294, 311	73, 152, 182, 195, 263, 310	**L1***
Mt 2	129, 148, 168, 172, 187, 188, 189, 223, 230, 311, 320	93, 152, 185, 189, 236, 247, 267, 263, 303, 310	**L0a1**
Mt 3	114, 129, 148, 162, 168, 172, 187, 188, 189, 223, 230, 311, 320	93, 96, 152, 185, 189, 200, 236, 247, 263, 303, 310	**L0a1**
Mt 4	129, 148, 168, 172, 187, 188, 189, 223, 230, 311, 320	93, 96, 152, 185, 189, 200, 236, 247, 263, 303, 310	**L0a1**
Mt 5	129, 148, 168, 172, 187, 188, 189, 233, 230, 311, 320	93, 152, 184, 189, 200, 236, 247, 263, 303, 310	**L0a1**
Mt 6	114, 129, 148, 168, 172, 184, 187, 188, 189, 223, 230, 293, 311	NA	**L0a1**
Mt 7	93, 129, 148, 168, 172, 176, 186, 187, 188, 189, 223, 230, 311	NA	**L0a1**

Continued

Mt 8	93, 129, 148, 168, 172, 187, 188, 189, 209, 223, 230, 278, 293	112, 152, 189, 236, 263, 310	**L0a1a**
Mt 9	93, 129, 148, 166, 187, 189, 223, 247, 265, 278, 311, 362	73, 152, 182, 195, 247, 263, 310	**L0a1a**
Mt 10	189, 223, 278, 294, 309, 311	73, 143, 146, 152, 195, 263, 303, 310	**L1c**
Mt 11	183, 189, 223, 278, 294, 309, 311	NA	**L1c**
Mt 12	189, 223, 229, 278, 291, 294, 311	73, 152, 182, 195, 263, 310	**L1c**
Mt 13	189, 223, 229, 278, 291, 294, 311, 320	152, 182, 195, 229, 263, 303, 310	**L1c**
Mt 14	169, 189, 223, 229, 278, 294, 311	73, 96, 103, 106, 152, 182, 195, 263, 275, 310, 329	**L1c**
Mt 15	129, 189, 215, 223, 278, 290, 294, 311	73, 96, 151, 152, 182, 183, 186, 189, 204	**L1c3**
Mt 16	189, 223, 229, 274, 278, 291, 294, 311, 321, 322, 383, 390	73, 152, 182, 195, 263, 303, 310	**L2a**
Mt 17	189, 192, 223, 266, 278, 294, 355, 390	NA	**L2a**
Mt 18	172, 209, 223, 278, 294, 301, 390	143, 152, 182, 195, 263, 310, 324, 348, 366	**L2a**
Mt 19	189, 223, 250, 278, 294, 309, 319, 320, 323, 355, 390	73, 109, 110, 143, 146, 152, 195, 263, 303, 310	**L2a1**
Mt 20	223, 278, 294, 309, 390	NA	**L2a1**
Mt 21	93, 189, 223, 278, 294, 309, 390	73, 103, 146, 152, 195, 198, 263, 303, 310, 316, 320	**L2a1**
Mt 22	114, 145, 213, 223, 278, 390	73, 146, 150, 152, 182, 263, 303, 310	**L2b**
Mt 23	114, 129, 213, 223, 278, 294, 355, 362, 386, 390	NA	**L2b1**
Mt 24	114, 129, 213, 223, 278, 294, 355, 362, 390	152, 189, 194, 200, 263, 310	**L2b1**
Mt 25	223, 264, 278, 390	73, 93, 146, 150, 152, 182, 195, 198, 263, 310, 325	**L2c2**
Mt 26	223, 243, 278, 362	73, 150, 263, 303, 310	**L3b**
Mt 27	93, 156, 223, 287, 300, 362	73, 150, 263, 303, 310	**L3b**
Mt 28	124, 223, 278, 311, 362	73, 152, 263, 310	**L3b1**
Mt 29	124, 189, 223, 278, 362	73, 152, 263, 310	**L3b1**
Mt 30	124, 148, 223, 278, 362	73, 151, 152, 229, 263, 303, 310	**L3b1**
Mt 31	111, 124, 223, 235, 311	73, 152, 199, 263, 310, 312	**L3d**
Mt 32	124, 223	73, 152, 263, 303, 310	**L3d**
Mt 33	124, 169, 223, 319	73, 152, 262, 303, 310	**L3d**
Mt 34	172, 189, 194, 197, 223, 320	NA	**L3e2b**
Mt 35	223	73, 150, 263, 310	**L3e5**
Mt 36	223	73, 152, 189, 195, 263, 303, 310	**L3e5**
Mt 37	136, 209, 223, 311	73, 185, 189, 200, 263, 310	**L3f**
Mt 38	209, 223, 261, 292, 311	73, 152, 189, 194, 200, 263, 310	**L3f1**
Mt 39	169, 193, 195, 223, 235, 243	73, 150, 263, 310	**L3x2**
Mt 40	223, 293, 311, 355, 362, 399	73, 146, 257, 263, 310	**L4g**
Mt 41	223, 293, 311, 355, 362, 299	73, 146, 257, 263, 310	**L4g**

Meseria			
Ms 1	114, 129, 148, 162, 168, 172, 187, 188, 189, 223, 230, 311,320	93, 152, 185, 189, 236, 247, 263, 303, 310	**L0a1**
Ms 2	129, 148, 168, 172, 187, 188, 189, 223, 230, 311, 320	93, 152, 185, 189, 200, 236, 247, 263, 303, 310	**L0a1**
Ms 3	189, 194, 223, 278, 294, 309, 360	N/A	**L1c1**
Ms 4	209, 223, 278, 294, 301, 354, 390	73, 143, 146, 152, 182, 189, 195, 263, 310	**L2a**
Ms 5	189, 192, 223, 263, 278, 294, 390	73, 143, 146, 152, 195, 263, 310	**L2a**
Ms 6	114, 129, 213, 223, 278, 294, 355, 362, 390	73, 150, 152, 182, 195, 198, 263, 303, 310	**L2b1**
Ms 7	129, 189, 223, 229, 278, 291, 294, 300, 390	73, 146, 150, 182, 263, 310	**L2d1**
Ms 8	129, 189, 223, 278, 300, 309, 390	73, 143, 146, 150, 195, 263, 310	**L2d1**
Ms 9	93, 223, 278, 287, 300, 362	239, 263, 303, 310	**L3b**
Ms 10	93, 223, 278, 287, 300, 362	239, 263, 303, 310	**L3b**
Ms 11	124, 223, 278, 311, 362	73, 152, 263, 310	**L3b2**
Ms 12	124, 223, 278, 311, 362	73, 152, 209, 231, 257, 264, 270, 275, 277, 281, 310, 404	**L3b2**
Ms 13	93, 223, 278, 311, 362	73, 263, 310	**L3b2**
Ms 14	124, 223, 278, 311, 320	73, 150, 263, 310	**L3e2**
Ms 15	126, 129, 145, 261	73, 152, 263, 295, 303, 310	**J1b**
Nuer			
N 1	129, 187, 189, 223, 243, 311	73, 146, 152, 263, 303, 311	**L0d**
N 2	129, 187, 189, 223, 243, 294, 311	73, 152, 195, 255	**L0d**
N 3	153, 223, 278, 294, 309, 390	73, 143, 146, 152, 195, 263, 310	**L2a1**
N 4	189, 192, 223, 278, 283, 292, 294, 309, 390, 391	73, 143, 146, 152, 195, 263, 310	**L2a1**
N 5	129, 187, 223, 278, 294, 300, 390	73, 143, 146, 150, 152, 195, 263, 310	**L2d1**
N 6	93, 143, 148, 157, 172, 187, 223, 278, 294, 300, 254, 390	N/A	**L2d1**
N 7	129, 187, 223, 278, 294, 300, 254, 390	N/A	**L2d1**
N 8	111, 129, 189, 192, 223, 278, 294, 300, 354, 390	73, 143, 146, 152, 195, 263, 263, 303, 310	**L2d1**
N 9	129, 157, 189, 278, 294, 300, 354, 390	73, 143, 146, 148, 150, 195, 263, 310	**L2d1**
N 10	116, 122, 129, 223, 266, 278, 300, 354, 390, 399	73, 146, 152, 195, 263, 303, 310	**L2d1**
N 11	192, 209, 223, 292, 295, 311	73, 189, 200, 263, 303, 310	**L3f1**
N 12	209, 223, 292, 311	73, 189, 200, 263, 310	**L3f1**
N 13	129, 148, 166, 187, 189, 223, 278, 304, 311, 317, 355, 362	73, 152, 182, 195, 247, 263, 310	**L5a**
N 14	129, 148, 166, 187, 189, 223, 278, 304, 311, 355, 362	73, 152, 182, 195, 247, 263, 303, 310	**L5a**
N 15	129, 148, 154, 166, 187, 189, 278, 311, 399	73, 103, 106, 146, 263, 310	**L5a**
N 16	129, 148, 166, 187, 189, 223, 278, 311, 355	73, 152, 182, 195, 247 263, 313	**L5a**
Nuba			

Continued

Nb 1	129, 148, 172, 187, 188, 189, 223, 230, 311, 368	73, 152, 200, 236, 247, 263, 303, 310	**L0a**
Nb 2	129, 148, 172, 187, 188, 189, 223, 311, 320	73, 150, 195, 198, 263, 310	**L0a**
Nb 3	148, 172, 187, 188, 189, 223, 230, 311, 320	93, 146, 150, 152, 185, 189, 200, 236, 247, 263, 310	**L0a**
Nb 4	148, 172, 179, 187, 188, 189, 192, 223, 230, 293, 311, 317, 320	73, 109, 150, 195, 198, 263, 310	**L0a**
Nb 5	129, 148, 168, 172, 184, 187, 187, 188, 189, 223, 230, 278, 293, 311, 320	93, 151, 152, 185, 189, 236, 263, 310	**L0a**
Nb 6	148, 172, 179, 187, 188, 189, 192, 223, 230, 311, 320	73, 109, 110, 143, 152, 189, 195, 198, 263, 303, 310, 366	**L0a**
Nb 7	129, 148, 168, 172, 184, 187, 187, 188, 189, 223, 230, 278, 293, 311, 320	93, 95, 146, 185, 189, 236, 247, 263, 303, 310	**L0a1**
Nb 8	129, 148, 168, 172, 187, 188, 189, 223, 230, 311, 320	73, 143, 146, 152, 263, 303, 310	**L0a1**
Nb 9	129, 148, 168, 172, 187, 188, 189, 223, 230, 311, 320	73, 143, 146, 152, 263, 303, 310	**L0a1**
Nb 10	129, 148, 168, 172, 184, 187, 188, 189, 223, 230, 278, 293, 311, 320	93, 95, 152, 185, 189, 200, 236, 247, 263, 310	**L0a2**
Nb 11	129, 187, 189, 223, 243, 311	73, 152, 195, 255	**L0d**
Nb 12	169, 172, 187, 189, 218, 223, 230, 278, 291, 311, 327, 354, 368,	146, 183, 184, 185, 189, 207, 247, 263, 303, 310	**L0f**
Nb 13	126, 187, 187, 189, 223, 239, 254, 264, 270, 278, 293, 311	93, 146, 150, 152, 189, 200, 236, 247, 263, 310	**L1b1**
Nb 14	129, 192, 223, 278, 294, 309, 311	73, 143, 146, 152 195, 263, 303, 310	**L1c**
Nb 15	129, 189, 194, 223, 278, 294, 311	73, 143, 146, 152, 195, 263, 303, 310	**L1c**
Nb 16	129, 189, 194, 223, 278, 311	73, 143, 146, 152, 195, 263, 310	**L1c**
Nb 17	129, 189, 215, 223, 294, 311	73, 195, 263, 303, 310, 316	**L1c3**
Nb 18	129, 189, 223, 245, 256, 258, 273, 278, 294, 309, 311	73, 152, 182, 189, 195, 247, 263, 310, 357	**L1c3**
Nb 19	189, 192, 223, 278, 294, 310, 311, 362, 390	NA	**L2a**
Nb 20	189, 192, 223, 294, 310, 311, 386, 390	NA	**L2a**
Nb 21	189, 192, 223, 278, 294, 309, 390	73, 143, 146, 152, 195, 207, 263, 310	**L2a1**
Nb 22	183, 189, 194, 223, 278, 294, 309, 390	73, 143, 146, 152, 195, 263, 303, 310	**L2a1**
Nb 23	129, 189, 194, 278, 294, 309, 390	73, 195, 263, 303, 310	**L2a1**
Nb 24	114, 129, 213, 223, 278, 390	73, 146, 150, 152, 182, 183, 195, 198, 204, 263, 310	**L2b**
Nb 25	114, 129, 213, 223, 278, 309, 390	73, 146, 150, 152, 182, 183, 195, 198, 204, 235, 263, 310	**L2b**
Nb 26	114, 129, 213, 223, 278, 390	73, 146, 150, 152, 182, 183, 195, 198, 204, 263, 310	**L2b**
Nb 27	189, 223, 278, 293, 300, 311, 354, 362, 390, 399	73, 146, 152, 182, 195, 257, 263, 310	**L2d1**
Nb 28	111, 145, 184, 192, 213, 223, 239, 278, 292, 304, 355, 390, 399	73, 146, 150, 152, 182, 263, 310	**L2d2**
Nb 29	209, 223, 292, 311	73, 189, 195, 200, 263, 310	**L3f1**
Nb 30	223, 311, 362	73, 150, 263, 303, 310	**L4a**
Nb 31	223, 261, 293, 311, 355, 362, 399	73, 146, 244, 257, 263, 310	**L4g**
Nb 32	129, 148, 166, 188, 189, 223, 278, 311, 355, 362	73, 152, 182, 195, 247, 263, 310	**L5a**
Nb 33	129, 148, 166, 187, 189, 223, 233, 254, 278, 362	73, 195, 247, 263, 310	**L5a**
Nubians			

Continued

Nu 1	146, 189, 192, 201, 205, 223, 249, 311	73, 109, 195, 263, 310	**M1**
Nu 2	129, 148, 168, 172, 187, 188, 189, 223, 230, 294, 311, 320	93, 146, 152, 185, 189, 200, 236, 247, 263, 310, 316	**L0a1**
Nu 3	129, 148, 168, 172, 187, 188, 189, 223, 230, 294, 311, 320, 392	93, 152, 185, 189, 200, 236, 247, 263, 303, 310	**L0a1**
Nu 4	129, 148, 168, 172, 187, 188, 189, 223, 230, 311, 320	93, 146, 152, 185, 189, 200, 236, 247, 263, 310, 316	**L0a1**
Nu 5	129, 148, 168, 172, 186, 187, 188, 189, 223, 230, 311, 320, 322, 329, 370, 398	NA	**L0a1**
Nu 6	129, 148, 164, 172, 183, 189, 223, 230, 311, 320, 322, 329	93, 152, 185, 189, 200, 236, 247, 263, 303, 310	**L0a1**
Nu 7	129, 148, 156, 159, 168, 172, 187, 188, 189, 223, 230, 294, 311, 320	93, 146, 152, 185, 189, 200, 236, 247, 263, 310	**L0a2**
Nu 8	126, 187, 189, 223, 264, 270, 278, 289, 293, 311	73, 152, 182, 185, 189, 195, 247, 263, 310, 357	**L1b1**
Nu 9	126, 187, 189, 223, 264, 270, 278, 289, 293, 311	73, 152, 182, 185, 189, 195, 247, 263, 310, 357	**L1b1**
Nu 10	209, 223, 278, 294, 301, 354, 390	73, 143, 146, 152, 182, 189, 195, 263, 310	**L2a**
Nu 11	189, 192, 223, 263, 278, 294, 390	73, 143, 146, 152, 195, 263, 310	**L2a**
Nu 12	183, 189, 192, 223, 278, 292, 294, 309, 390	152, 182, 195, 247, 263, 310	**L2a1a**
Nu 13	210, 224, 268, 271, 273, 304, 310, 311, 340, 347, 366, 369, 388, 400	NA	**L3***
Nu 14	189, 195	73, 146, 152, 263, 303, 310	**L3***
Nu 15	189, 194	73, 143, 146, 152, 195, 263, 310	**L3***
Nu 16	170, 189, 194, 195	73, 143, 146, 152, 182, 189, 195, 263, 310	**L3***
Nu 17	124, 157, 223, 278, 311, 362	73, 263, 310	**L3b1**
Nu 18	165, 172, 189, 223, 258, 271, 276, 282, 310, 311, 320, 384, 386, 399	NA	**L3e2b**
Nu 19	93, 126, 131, 133, 135, 147, 172, 197, 205, 223, 245, 248, 355	73, 183, 199, 204, 263, 303, 310	**L3f**
Nu 20	209, 223, 292, 311	73, 189, 200, 303, 310	**L3f1**
Nu 21	116, 119, 157, 176, 223, 260, 311, 335, 362	73, 150, 152, 263, 310	**L3w**
Nu 22	93, 287, 293, 301, 311, 355, 362, 399	NA	**L4g**
Nu 23	172, 223, 293, 311, 355, 362, 399	73, 146, 195, 244, 263, 310	**L4g**
Nu 24	129, 148, 189, 223, 278, 311, 399	73, 146, 150, 152, 247, 263, 310	**L5a**
Nu 25	129, 148, 156, 157, 166, 187, 189, 223, 278, 311, 347, 355, 362	73, 143, 146, 152, 195, 247, 263, 310	**L5a**
Nu 26	116, 154, 156, 157, 170, 173, 181, 182, 183, 191, 223, 362	73, 150, 263, 303, 310	**D5**
Nu 27	146, 168, 172, 223, 245, 248, 355	73, 150, 152, 204	**N1a**
Nu 28	126, 362	263, 310	**preHV1**
Nu 29	165, 172, 189, 219, 278	73, 146, 152, 263, 310	**U6a1**
Shilluk			
Sh 1	93, 126, 179, 187, 189, 223, 264, 270, 278, 311	73, 152, 182, 185, 189, 195, 199, 247, 263, 303, 310, 357	**L1b**
Sh 2	126, 131, 184, 187, 189, 223, 264, 270, 274, 278, 293, 311	73, 152, 182, 185, 189, 195, 199, 247, 263, 310, 357	**L1b1**
Sh 3	126, 187, 189, 223, 264, 270, 274, 278, 293, 311	73, 152, 182, 185, 189, 195, 199, 247, 263, 310	**L1b1**
Sh 4	114, 129, 213, 223, 278, 291, 354, 390	NA	**L2b**

Continued

Sh 5	129, 189, 223, 224, 228, 278, 300, 354, 390, 399	73, 146, 152, 182, 195, 263, 303, 310	**L2d1**
Sh 6	124, 223, 278, 311, 362	73, 152, 263, 303, 310	**L3b2**
Sh 7	124, 126, 131, 133, 153, 205, 220, 223, 256, 278, 268, 399	73, 152, 263, 303, 310	**L3d2**
Sh 8	172, 189, 194, 223, 320	73, 150, 195, 263, 303, 310	**L3e2b**
Sh 9	209, 223, 292, 311	73, 96, 185, 189, 200, 263, 310	**L3f1**
Sh 10	223, 256, 311, 386	73, 151, 195, 263, 303, 310	**L3h**
Sh 11	126, 131, 183, 184, 197, 205, 223, 261, 293, 311, 355, 362, 399	73, 146, 244, 257, 263, 294, 303, 310	**L4g**
Sh 12	172, 184, 197, 205, 223, 293, 310, 311, 355, 362, 399	73, 109, 146, 244, 263, 310	**L4g**
Sh 13	223, 252, 287, 293, 311, 355, 362, 399	73, 146, 152, 244, 263, 310	**L4g**
Sh 14	102, 105, 126, 129, 148, 163, 166, 187, 189, 223, 263, 278, 292, 311, 355, 362	73, 152, 182, 195, 247, 263, 303, 310	**L5a**
Sh 15	126, 129, 148, 166, 187, 189, 223, 263, 278, 310, 311, 339, 355, 362	73, 152, 182, 195, 247, 263, 303, 310	**L5a**
Sh 16	124, 223, 278, 301, 362	73, 150, 204, 263	**L3b**

Variant positions from the CRS are shown, from 16093 to 16400 (minus 16000) for HVR-I, and from 73 to 404 for HVR-II. Haplogroups were assigned according to HVR-I and HVR-II motifs as discussed in the text. 1*, L2* and L3* indicate samples that were tested for SNP variation of hypervariable region but did not fit into known haplogroups classifications. N/A = not available.