ABSTRACT

Title of Document: COMPLEX GENETIC HISTORY OF EAST AFRICAN HUMAN POPULATIONS

Jibril B. Hirbo, Doctor of Philosophy, 2011

Dissertation directed by: Sarah A. Tishkoff, Ph.D.
Departments of Biology, University of Maryland, College Park

Results from disparate fields indicate that anatomically modern *Homo sapiens* originated in Africa ~200 thousand years ago (kya), and that East Africa is the likely source of the migration of modern humans out of Africa within the past 100 thousand years. However, the genetic diversity currently found in Africa, and especially East Africa, has not been well studied compared to non-African populations, due in large part to the fact that DNA samples from many remote regions of Africa are currently not available. The goal of this study was, therefore, to characterize genetic variation in previously unstudied East African human populations using mitochondrial and Y chromosome DNA from 1500 individuals. These data were then compared to independently collected data of the same populations from ~1327 nuclear markers (848 microsatellites and 479 insertion/deletion polymorphisms). The data were used to gain insight into patterns of genetic diversity, to construct past relationships of East African populations to each other and to other African populations, to clarify historical demographic events such as population expansion,

contraction, and migration that these populations might have experienced. Several independent analyses showed significant relationships between genetic and geographic/linguistic distances among East African populations. Genetic variation is more strongly correlated with geography than is linguistics. Overall, the correlations between genetic versus geographic/ linguistic variation is stronger for autosomal and Y chromosome than for *mt*DNA lineages. Y chromosome and *mt*DNA lineage distributions seem to cluster geographically and for some lineages, linguistically. Two major migration events, namely the migration of Bantu-speaking populations from Central/West Africa across sub-Saharan Africa and the migration of pastoralist populations from Sudan and Ethiopia within the past 5000 years have had a major influence on extant genetic patterns in East Africa.

.

COMPLEX GENETIC HISTORY OF EAST AFRICAN HUMAN POPULATIONS

by

JIBRIL BORU HIRBO

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the
degree of Doctor of
Philosophy 2011

Advisory Committee:
Professor Sarah A. Tishkoff, Chair
Professor Michael P. Cummings
Professor Gerald S. Wilkinson
Professor Eric S. Haag
Professor Stephen M. Mount

Acknowledgements

Sarah A. Tishkoff (Departments of Biology University of Maryland College Park, currently at Department of Genetics and Biology, School of Medicine and School of Arts and Sciences University of Pennsylvania) for offering me the opportunity to do this research at her laboratory and work under her tutelage.

Olivier Hanotte (University of Nottingham), Toomas Kivisild and Marta Lahr (University of Cambridge).

My Committee Members: Drs. Michael Cummings, Eric Haag, Steve Mount and Gerald Wilkinson.

Nikolay Eltsov and Natalya Volodko from the Siberian Branch of the Russian Academy of Science (SBRAS) who allowed me to use a beta version of their *mt*DNA analysis program MtPhyl. Mingyao Li from University of Pennsylvania – Philadelphia for suggestions on statistical tests used in the study, Stanley Ambrose from University of Illinois Urbana-Champaign (UIUC), Chris Ehret from the Department of History at University of California – Los Angeles (UCLA) & Chapurukha Kusimba of the Field Museum, University of Illinois Chicago (UIC) for useful discussion and advice on literature on archaeological, linguistic and Paleoclimatic data used for inferences of African population history. Tom Johnson from Large Lake Observatory at the University of Minnesota-Duluth (LLO - UM Duluth) for helpful advice regarding literature on Paleoclimatic conditions in Africa.

Tishkoff Lab:

Floyd Reed (now at Max Planck Institute for Evolutionary Biology, Germany) for help in formatting Marshfield data and offering advice on some statistical analysis; Alessia Ranciaro who helped with figures, typesetting of the thesis; Alessia Ranciaro and Karuna Panchapakesan who helped in sample DNA extraction and organization; Charla Lambert for advice on statistical tests. Tishkoff lab manager Bill Beggs for letting me do simulation runs on his desktop.

Collaborators to Dr. Tishkoff in Africa:

Sabah A. Omar (Kenya), Muntaser Ibrahim (Sudan), Thomas B. Nyambo (Tanzania), Alain Fremont (Cameroon and Chad).

Field assistants:

Lilian Alando, Daniel Kariuki, Eva Aluvalla and Fatiha Mohamed (Kenya), Abdallah Teia, Maha Osman and Mervat Osman (Sudan).

Family and Friends

I am overly indebted to my family for their unwavering support during challenging and good times of my study. My best friends Alex Tibwitta, George Ogutu and Hussein Musa for their moral support and were also very generous with their times and resources by helping with some logistics and errands during the field excursion in Kenya (2004 and 2006).

I am indebted to a number of individuals who have commented on ideas contained in the dissertation and on the thesis itself. These include: My supervisor, Dr. Sarah Tishkoff, Toomas Kivisild, Olivier Hanotte, Stanley Ambrose, Colin Rose (University of Southern

California). I would also like to acknowledge my colleagues in Tishkoff Laboratory who have been so generous with their time to offer comments and suggestions on the thesis, Alessia Ranciaro and Joe Jarvis. Their input has greatly improved the quality of this thesis and a more refined write-up could not have been possible without their feedback. I am solely responsible for the integration and interpretation of the ideas and evidence, and thus any error which might have slipped their review, in the thesis. Completion of this research was made possible by several grants to my supervisor Dr. Sarah Tishkoff: National Science Foundation (NSF), Wenner Gren, Leakey, Burroughs Wellcome Fund, and Lucille Packard. I was supported by Leakey Foundation's Baldwin Fellowship award, 2003-2005 and University of Maryland Graduate School Matching Fund, and Teaching Assistantship at the department of Biology.

TABLE OF CONTENTS

List of Tables

List of Figures

List of Initials and Abbreviations

MSA - Middle Stone Age
LSA - Later Stone Age
Ky – thousand years
Kya – thousand years ago
Kyo – thousand years old
IBD - Identity by descent
Hp – Helicobacter Pylori
DNA - Deoxyribonucleic Acid
*mt*DNA – Mitochondrial Deoxyribonucleic Acid
LGM – Last Glacial Maximum
TMRCA – The most recent common ancestor
SAK – South African Khoisan
NS – Nilo-Saharan
NK – Niger-Kordofanian
AA – Afroasiatic
AMHS - anatomically modern *Homo sapiens*
SNPs – Single Nucleotide Polymorphisms
STRs – Simple Tandem Repeats
HGDP–CEPH Human Genome Diversity Cell Line Panel representing global human populations
BC – Before Christ
AD – Anno Domini (after Christ)
HG – Hunter-gatherers
EAHG – East African Hunter-gatherers
UEPs – Unique Event Polymorphisms
Y-DNA – Y chromosome Deoxyribonucleic Acid
HVS-I – Hyper-variable sequence region I (Mitochondrial)
HVS-II – Mitochondrial Hyper-variable sequence region II (Mitochondrial)
D-loop – Displacement Loop (Mitochondrial region encompassing HVS-I & HVS-II)
SMM - stepwise mutation model
TRD - Trinucleotide Repeat Disorders
NRY – Non recombining (region) of Y chromosome
IRB - Internal Review Board
UK – United Kingdom
MDS – Multidimensional scaling
MCMC – Markov Chain Monte Carlo
CO - Cytochrome oxidase subunit
NADH - Nicotinamide adenine dinucleotide hydroxide
ND4 - Ubiquinone oxido-reductase subunit IV
AT6 – Adenine triophosphate synthase 6
RFLP – Restriction Fragment Length Polymorphisms
BSPs- Bayesian Skyline Plots
AMOVA – Analysis of Molecular Variance

Chapter 1: Introduction and Background

1.0: Aims of the study

Despite an overwhelming indication from disparate fields that Africa is the ancestral homeland of modern humans, Africa has not been intensively studied for human genetic diversity compared to non-African populations [1], probably due to the fact that DNA samples from many regions of Africa, especially East Africa have not been available. In general there is a dearth of information on genetic variation in East Africa, specifically southern Ethiopia, Sudan, Uganda, Tanzania and Kenya, a region of great linguistic and cultural diversity. The focus of my dissertation research, eastern Africa (Kenya, Tanzania, Sudan and Ethiopia), has populations speaking languages belonging to the four major linguistic African families: Khoisan, Niger-Kordofanian, Nilo-Saharan and Afroasiatic. This study therefore, aims to address the dearth in genetic population data from a region that is probably the cradle of humanity and as acted as a major corridor of human migration within and outside Africa.

Because African populations have lived in close contact with each other for thousands of years, it is unlikely that any population would remain so isolated that it has not undergone admixture with neighboring groups [2]. Therefore, detailed sampling and determination of the amount and distribution of genetic diversity among ethnically diverse East African populations will thus enable a clearer understanding of historic migration events within and out of East Africa, as well as more recent admixture events among East African populations. Additionally, genetic data may clarify existing interpretations of the history of East African populations based on cultural, linguistic and archaeological data.

In order to shed light on East African population history, mitochondrial DNA (*mt*DNA) and Y chromosome genetic variation were characterized in approximately 1500 individuals from previously unstudied and under-studied East African populations. These data were then compared to independently collected data from the same populations using ~1327 nuclear markers (848 microsatellites and 479 insertion/deletion polymorphisms) [3, 4]. The data were used to gain insight into patterns of genetic diversity, to construct past relationships between East African populations and their relationships with other African populations, and to clarify historical demographic events such as population expansion, contraction, and migration events. Four specific hypotheses were tested:

(1) Geographic and genetic distances between populations in East Africa are correlated

(2) Linguistic and genetic distances between populations in East Africa are correlated

(3) Distribution of *mt*DNA and Y chromosome lineages are correlated with geography/language

4) Effective population sizes and population range expansions correlate with changes in climate and/or technology based on the archeological record

1.1: Background

Results from disparate fields indicate that anatomically modern *Homo sapiens* (AMHS) originated in Africa ~200 thousand years ago (kya), and that East Africa is the likely source of migration of modern humans out of Africa within the past 100 kya [5-25]. Eastern Africa is the geographical region that provides the best evidence for both the earliest modern humans and for a relatively continuous morphological transition from late archaic to early modern humans [26]. According to paleobiological and archeological evidence, indicators of modern human behavior such as symbolism in art and artifacts, specialized tool technologies and long-range exchange of raw materials, in Africa predate their earliest occurrence outside of Africa. [22]. Previously, the earliest AMHS fossils were those found in the Klasies River Mouth Caves in South Africa which date to ~130 kya [19] and those from the Levant, at Qafzeh and Skhul which date to ~130-90 kya [19]. However, recent fossil evidence from Ethiopia indicates the presence of early AMHS there between ~195-154 kya [23, 24]. After the initial appearance of AMHS in the Levant ~130-90 kya, they do not reappear in that region or in Europe until ~50-40 kya [27, 28]. As Neanderthal remains appear continuously in Levantine sites over this time, the fossil evidence suggests a lack of continuity of AMHS in this region since the first exodus [29]. The earliest evidence of AMHS presence outside of Africa comes from Australia where the oldest modern human fossils have been dated to ~46 kya [30]. However, recent archaeological evidence from pre-Toba and post Toba (74-77 kya) artefacts from the Indian subcontinent show closer affinities to African Middle Stone Age traditions (such as Howieson's Poort), and may indicate modern humans might have reached the Indian sub-continent by 70 kya [31]. Based on these patchy findings, several

dispersal scenarios have been proposed from East Africa across the globe; an early "southern" route through the Horn of Africa/Arabia towards South Asia ~100–60 kya and a later "northern" route through North Africa/Middle East towards Eurasia ~70–40 kya [17, 25, 28, 32]. Recent analyses of *mt*DNA data from aborigines of South East Asia [5-7] support the contention that dispersal along the coastal southern route, approximately ~70–40 kya was the most likely first route of modern humans out-of Africa.

Several population genetic studies exploring the relationships between genetic diversity, geography and language on a global scale [33, 34], continental scale [35, 36] and regional scale [37] have shown correlations between genetic variation and language/geography in human populations. However, Wood *et al.,* [38] analyzed 40 African populations for *mt*DNA and Y chromosome lineages and showed that for Y chromosomes, there was correspondence between genetic and linguistic distances but no significant correlation between genetic and geographic distances, whereas *mt*DNA variation was weakly correlated with both language and geography. Analyses of autosomal genetic data (mainly SNPs and STRs) have shown geographic patterns, typically 'clines', in which the genetic distance between populations increases with geographic distance at both continental and global scales. Thus, human populations appear to follow a pattern of 'isolation by distance' (IBD) [8, 9]. Prugnolle *et al.,* [8] and Ramachandran *et al.,* [9] showed that the average heterozygosity at 783 autosomal microsatellite loci decreases linearly with geographical distance from East Africa in 1048 individuals from the HGDP–CEPH Human Genome Diversity Cell Line Panel representing global human populations. The authors explained this pattern by positing a 'serial founder effect', on the basis of the pattern's similarity to the heterozygosity

patterns produced by simulations of the spread of genotypes along a linear path [8, 9, 39, 40]. Overall the extant genetic pattern in the global human population fits a serial founder effect model with increasing genetic distance correlated with geographic distance from East Africa indicating the cumulative effect of genetic drift as humans expanded into the rest of the world [41, 42].

Studies have also shown that global patterns of phenotypic variation can be explained by isolation by distance model. For example, Betti *et al.,* [10] showed that distance from sub-Saharan Africa is correlated with the levels of human within-population cranial phenotypic diversity. They were able to explain over half of the worldwide variation in cranial phenotypic diversity by looking at cranial measurements from male and female individuals from 105 and 39 worldwide human populations, respectively [10]. Using the same samples measured for 37 morphometric characteristics, Manica *et al.,* [40] showed a strong pattern of isolation by distance in craniometric inter-population differences [40]. The observed global pattern also shows evidence for an African origin, with the strongest clinal relationship between within-population phenotypic variability (corrected for climate) and distance on land (with its origin in Central/southern Africa [40]).

Demographic studies of modern human populations indicate that there were several population expansion events within Africa and subsequently out of Africa. These expansion events occurred during the late Middle or early Late Pleistocene (150 – 50 kya), with the oldest expansion occurring in Africa [43-47]. This observation partly explains the greater effective population size typically observed in African populations

relative to the rest of the world, and the subsequent conclusion that a population bottleneck occurred when populations left Africa [11, 17, 34, 48-55].

1.2: East African Linguistic diversity

To discern ancient events that might have recently shaped the extant genetic pattern in East Africa, I focus primarily on archaeological, linguistic and paleoclimatic data from the last 20 ky, with particular emphasis on the Holocene era (in the last 10 ky). This is because: a) the limit of resolution of the time to most recent common ancestry (tMRCA) of all language families in Africa is ~20 kya [56, 57]; b) Late Stone Age (LSA) industries in East Africa that are associated with specific groups of hunter-gatherer populations, for example the "Wilton" (that is associated with East African Khoisan speakers) [48, 56] and the Eburran (that is associated with other East African Holocene hunter-gatherers) [57] are estimated to have begun approximately 17 kya and 13 kya, respectively [57]; and (c) major events that are due to change in subsistence pattern such as development and spread of pastoralism [58, 59] and agriculture [60-64] took place within the last 10 kya.

Eastern Africa, including Tanzania, Kenya, Ethiopia, and Sudan, is a region with a large amount of ethnic and linguistic diversity. Populations in this region speak languages belonging to all four of the major language families in Africa: Afroasiatic, Nilo-Saharan, Niger-Kordofanian and Khoisan. In fact, based on linguistic analysis, Ehret [65] hypothesizes that; (1) all indigenous African language families originated in northeastern Africa [65], (2) that the Khoisan languages originated >20 kya in East Africa [66], and (3) that the Afroasiatic, Niger-Kordofanian and Nilo-Saharan languages may be more recent, at around 15 kya. Even though the standard linguistic methods used to infer

the age of extant human languages are limited to a depth of 20 kya, Ehret [66] argues that common ancestry of languages can be dated back to 25 – 30 kya, or even earlier in the case of Khoisan [66].

1.2.1: Niger-Kordofanian language family

The Niger-Kordofanian language family [67-69] is Africa's largest language family in terms of geographical area, number of speakers, and number of distinct languages **(Figure 1.2.3)**. There are over 1532 members of this language family according to recent estimates [70] spread contiguously across sub-Saharan Africa. However, besides the Kordofanian branch, which is exclusively found in Sudan, the majority of the Niger-Kordofanian language families [71, 72] are concentrated in West Africa with only one sub-branch of the Benue Congo, the Bantu, predominating over the area that extends from Cameroon to eastern and southern Africa **(Figure 1.2.1, Figure 1.2.3)** [62, 64, 71, 73]. Based on linguistic studies, "Bantu-speakers" moved gradually through the Congo forest from their putative origin in the Cameroon-Gabon area [64, 74], expanded to East Africa around 3 kya and subsequently moved farther South within the past 2 ky [75]. Some historians suggest that the current Bantu speakers in East Africa and South Africa absorbed many of the hunter-gatherer and pastoralist populations they encountered during their expansion [74-77]. For example, according to Dundas [77], based on oral and cultural evidence, the populations that constitute the present Kikuyu are a collection of farming groups that expanded and absorbed several neighboring hunter-gatherer and pastoral communities including the Dorobo and Maasai, respectively [77].

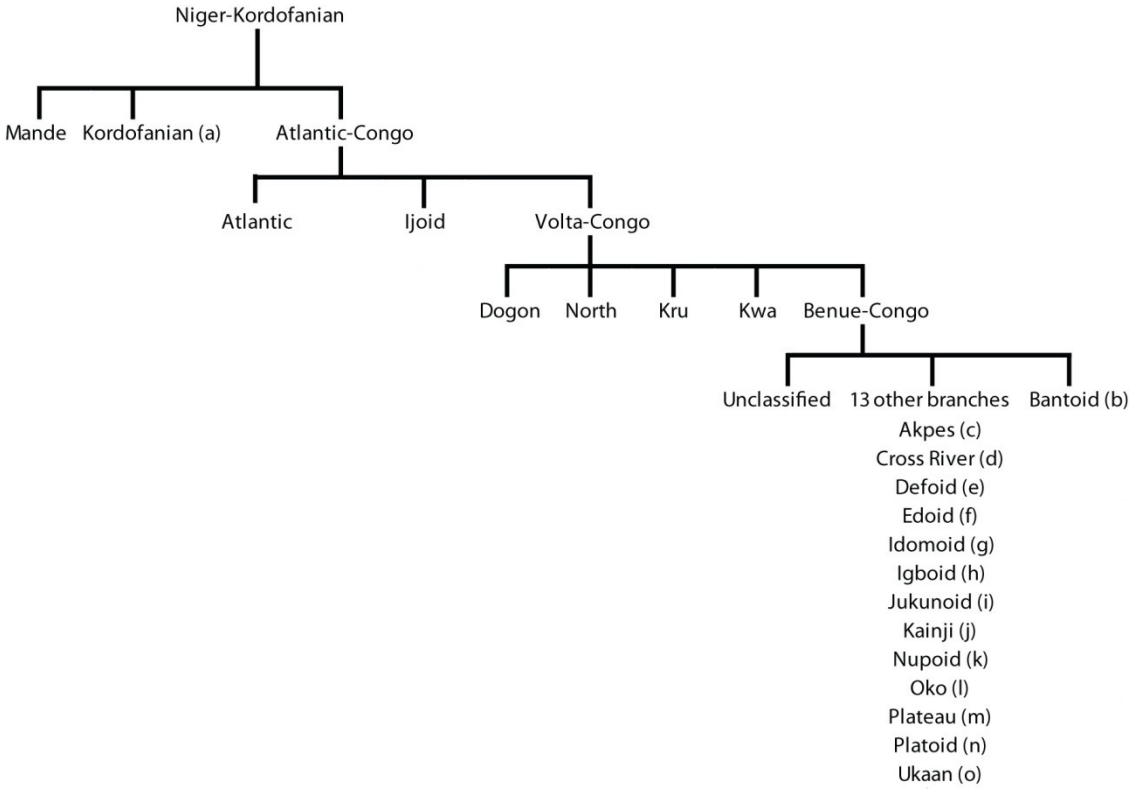


Figure 1.2.1: Primary splits in the Niger-Kordofanian language family adapted from Ethnologue.com [70] showing the position of the Bantoid in Niger-Kordofanian classification. Kordofanian branch (a) consists of 24 languages exclusively found in Sudan. Besides Bantoid (b) that consist of 691 Bantu languages identified so far, all the other sub-branches and classes are found in West Africa. Benue-Congo, the largest branch of the Niger-Kordofanian language family, is classified into 14 sub-branches (with one other unclassified). The other 13 sub-branches that are all found in Nigeria and Benin are indicated with number of distinct language genera identified so far in brackets are; c) Akpes (1), d) Cross River (68), e) Defoid (17), f) Edoid (33), g) Idomoid (9), h) Igboid (7), i) Jukunoid (20), j) Kainji (58), k) Nupoid (11), l) Oko (1), m) Plateau (55), n) Platoid (1) and o) Ukaan (1).

.

1.2.2: Nilo-Saharan language family

Nilo-Saharan languages are distributed across the middle portion of the African continent, but the distribution is rather fragmented **(Figure 1.2.3)**, a pattern used to argue that the language family was formerly spoken over a more extensive area [2]. Some historians have even argued that the language spoken from 8th Century BC until about

350 AD in the Meroitic kingdom in modern Sudan was Nilo-Saharan [71, 78]. The most widespread branch of the Nilo-Saharan language, Sudanic, consists of languages spoken across the African continent from the West African Songhay language spoken by populations on the Niger River to the Nilotic languages of Kenya and Tanzania in East Africa [79] **(Figure 1.2.2)**, but its actual geographic distribution is somewhat fragmentary. The second primary branch of the Nilo-Saharan language, Koman **(Figure 1.2.2)**, by contrast, is limited to areas just north and south of the Blue Nile at the western fringes of the Ethiopian highlands [79] **(Figure 1.2.2)**.

Although there is dispute on the precise source for the origin of Nilo-Saharan languages, linguists agree that Nilo-Saharan speakers originated in northeastern Africa, with Blench [80] proposing the southern Chad/western Sudan region, and Ehret [65] suggesting eastern Sudan, perhaps the Blue Nile as the site of origin. This language family is considered among the oldest in Africa, after the Khoisan language family, because it shows evidence of a pre-agricultural vocabulary and greater internal diversity than the other two non-Khoisan African language families – Afroasiatic and Niger-Kordofanian [81]. Initial dispersal of Nilo-Saharan speaking populations that practiced hunting and fishing is thought to have taken place about 18 kya [71]. All the Nilo-Saharan languages spoken in East Africa, Sudan and northern Tanzanian are of the East Sudanic sub-family [67, 82, 83] and the most notable ethnic groups from the family include the Luo, Maasai, Kalenjin and Dinka, who speak languages belonging to the Nilotic sub-branch **(Figure 1.2.2)**. Nilotic speakers are thought to have migrated into Kenya, Tanzania, and Southwestern Ethiopia within the past 1000 – 3000 years, although there may have been an earlier migration of fishing Nilo-Saharan speaking people into

northern Kenya between 8.7–11 kya [84, 85]. Eastern Nilotic groups are thought to have originated in the region west of Lake Turkana: the early Nilotic populations who might have been their ancestors were fishermen and pastoralists, a contention supported by 'Lopoy' excavations at the shore of Lake Turkana [86].

The Maasai and the Kalenjin are two Kenyan ethnic groups that are broadly known internationally. Familiarity with the Maasai can be attributed largely to their distinctive customs, dress, and residence near most of the game parks in East Africa, while the long distance running prowess of the Kalenjin has earned Kenya the major honors in international athletics at distances from 800 meters to the marathon since the mid 1960's. The Maasai are a collection of semi-nomadic populations that speak a Maa branch of eastern Nilotic languages **(Figure 1.2.2)** and are found in a nearly contiguous area from the northern part of Kenya up through northern Tanzania. The Kalenjin are an ethnic group consisting of eight culturally related populations that speak southern Nilotic languages **(Figure 1.2.2)** and are found in Kenya's western highlands and part of the Kenyan Rift Valley, and eastern Uganda. Considering that the Kalenjin and the Maasai currently occupy a large portion of Kenya and some part of both Uganda and Tanzania, they might have had substantial historical influence on the genetic landscape of Kenya and East Africa in general.

The Luo speak a western Nilotic language **(Figure 1.2.2)** and are found predominantly in the western part of Kenya, but also in parts of eastern Uganda and the upper tip of Tanzania. Ethno-linguistic studies [87] propose that the original Luo-speaking population split from a population that spoke a Nuer-Dinka precursor language approximately two thousand years ago. The Luo migrated from southern Sudan and

assimilated with populations they encountered in northern Uganda and Northwestern Kenya with consequent loss of some original grammar and vocabulary relative to the more "archaic" Shilluk dialect [88, 89]. Although the Luo speak a Nilo-Saharan language, previous study of 13 autosomal SNPs on genes associated with risk of or protection against severe malaria infection [90], and blood group studies [91] done in East African populations found no significant difference in allele frequencies between the Luo and populations from Kenya and Uganda that speak Bantu languages [92, 93]. However, blood group analysis showed similarity between the three other western Nilotic speaking populations of the Nuer, Dinka and Shilluk [94, 95].

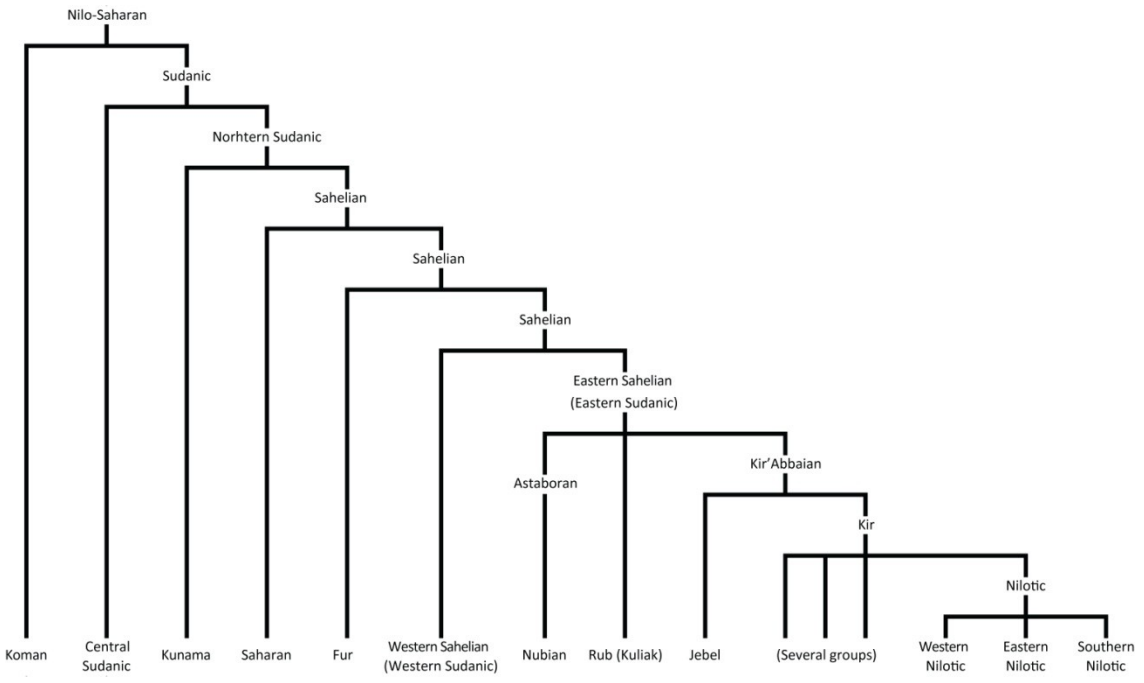


Figure 1.2.2: Classification of Nilo-Saharan language family adapted from Ehret [65, 96, 97] showing the position of the East African Nilotic speakers. Some literature [67, 82, 98] on Nilo-Saharan languages use the classification of Sudanic instead of Ehret's Sahelian thus Eastern Sahelian is also Eastern Sudanic.

1.2.3: Afroasiatic language family

Afroasiatic languages are classified into two primary branches [66, 99, 100]. The first branch, "Erythraic" is widespread across Africa and southwestern Asia [66, 99, 100] **(Figure 1.2.3)**. The Erythraic branch includes the Cushitic languages, spoken from the Red Sea hills of Sudan to central Tanzania, the Chadic languages of Niger, Nigeria, Chad and northern Cameroon, Berber speakers from the Sahara and North Africa, ancient Egyptian, and the Semitic languages of southwestern Asia and parts of northern Africa [66]. In contrast, the second primary branch, the Omotic, is entirely restricted to the southwestern Ethiopia, adjacent to the current distribution of the Koman branch of Nilo-Saharan speakers [79]. The Cushitic languages are further subdivided into northern, central, eastern and southern Cushitic roughly corresponding to the geographical distribution within northeastern Africa [96, 101]. Northern Cushitic is spoken in northern Sudan and southern Egypt, southern Cushitic is spoken in northern Tanzania and southern Kenya, and eastern and central Cushitic languages are spoken in Kenya and Ethiopia [96, 101]. Although the geographic distribution of the Bantu languages across Africa is quite broad in comparison to the Cushitic language distribution [102], the Cushitic language subgroup contains greater internal diversity than the Bantu. Differences among Cushitic subfamilies are much greater than among Bantu subfamilies, although the number of dialects and closely related languages are much fewer [102]. According to Fleming [102], the spread of the Cushitic languages occurred earlier than the spread of the Bantu languages. He infers that the Cushitic geographical continuity suggests that the whole geographical area it covers was occupied by Cushitic speakers, prior to more recent events, such as migration of different groups into the region, that

created gaps [102] in the distribution. Afroasiatic speakers in East Africa are thought to have originated in southwestern Ethiopia ~5 kya and to have migrated into Kenya and Tanzania between 3-5 kya [85]. During their initial dispersal southward (into Kenya, Tanzania and eastern part of Uganda) of Nilotic speakers from Sudan via southwest Ethiopia, Phillipson [2] suggests that there was contact between the Nilotic and Cushitic speakers which might have resulted in both cultural/linguistic [103] and/or genetic exchange.

Most of the populations included in this study sampled from northern Kenya, East of Lake Turkana, speak languages that belong to the lowland branch of the eastern Cushitic subfamily. But they exhibit some cultural differences among each, and their languages can be classified into three mutually unintelligible "languages/dialect clusters". Specifically, the Rendille speak a Somaloid language, while the Gabra, Sakuye and Gareeh have abandoned their original "Somaloid" language for Borana [104, 105]. There is also an overlap of clan names, rituals and beliefs among these historically "Somaloid" populations and a third set of populations speaking various Somali dialects [104, 105]. The putative center of origin of the eastern Cushitic speakers (including the eastern highland Cushitic speakers that are mostly found in Ethiopia) is in southern Ethiopia [106]. Eastern Cushitic populations have had cultural (and potentially genetic) influence on both southern Nilotic speaking populations and some eastern Nilotic populations like the Teso and Maasai around 3 kya, and more recently (about 0.5 kya) during the famous Orma expansion [102]. The excavations of Kalokol burial sites in Northwest Kenya dated to about 2.5 kya show evidence of funeral practices consistent with those practiced by modern eastern Cushitic speaking people, the Borana and Konso, suggesting evidence of

occupation of the area by an eastern Cushitic group around that period [107]. This Northwest part of Kenya is currently occupied by the eastern Nilotic Turkana.

Southern Cushitic speakers (represented by populations such as the Iraqw, Fyome and Burunge in Tanzania), which according to some historians represent descendants of the population that practiced the Savannah Pastoral Neolithic culture [108-111] **(section 1.3)**, are thought to have migrated into Kenya and Tanzania from southern Ethiopia from about 5 kya [111, 112]. But, the ancestral population/s that gave rise to the eastern Cushitic populations such as the Boni, Somali and Rendille started expanding much later, about 2 kya [113, 114] from northern Kenya to eastern parts of Kenya and Somalia (Somali) and southeastern Kenya (Boni), with the Rendille population remaining in northern Kenya. While the Dahalo population (from southeastern Kenya), which shares language features with southern Cushitic populations, are thought to have intermixed with some eastern Cushitic and Khoisan speaking populations [115, 116], their language might have descended directly from late proto-Cushitic language and been modified due to close interaction with the people speaking other proto-southern Cushitic languages [102].

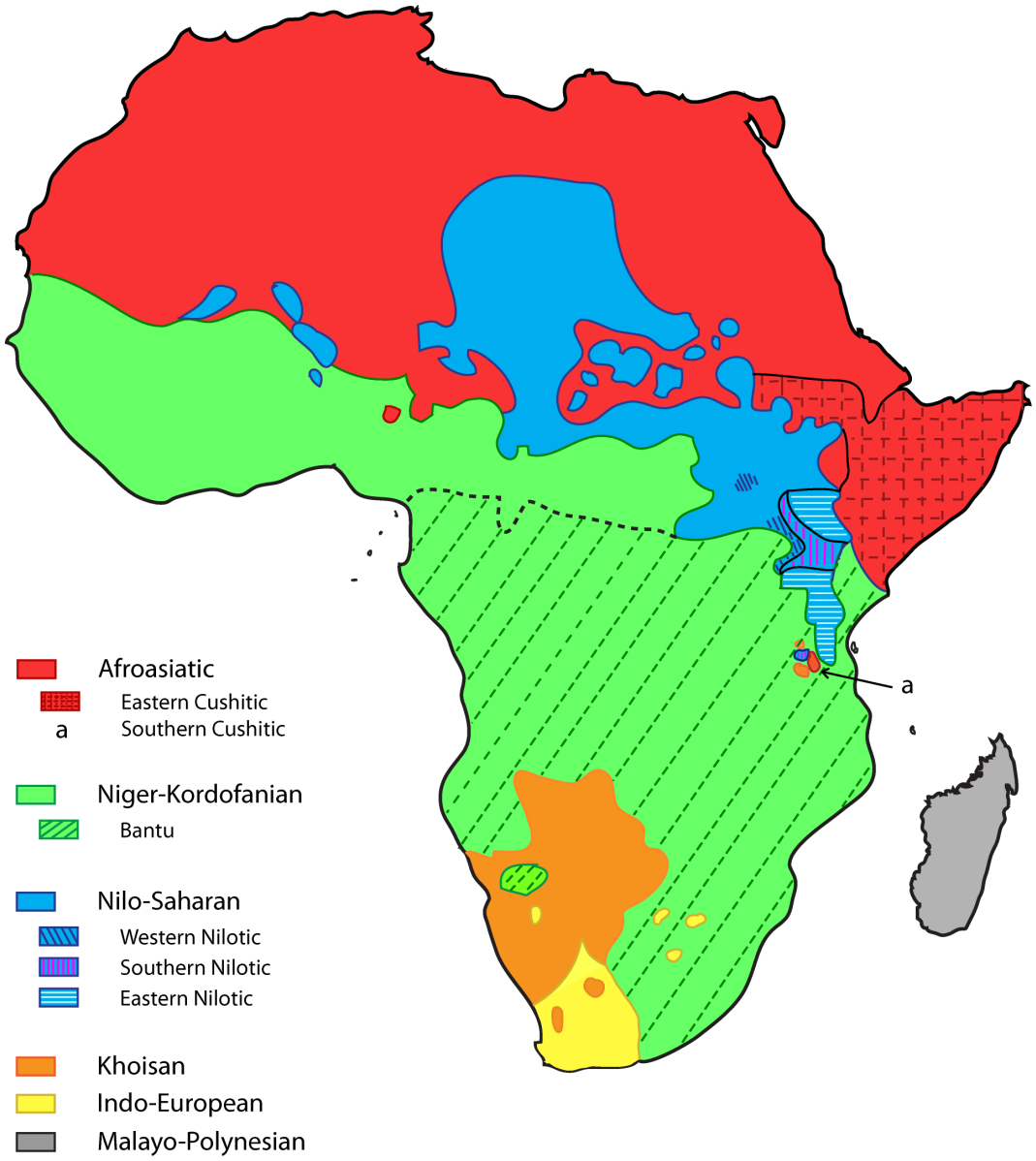


Figure 1.2.3: The geographical distribution of the major Linguistic groups in Africa. The Map was drawn using the information (geographical locations of ethnic speakers in Africa that are base on published sources) from Haspelmath *et al.,* 2008 [117] (http://wals.info/), http://www.ethnologue.com/[118], Greenberg 1963, 1972 [67, 69], Vansina 1995 [62], Ehret 1971, 1993, 1995, 2001 [64, 65, 87, 101]. The World Atlas of Language Structures (WALS) (http://wals.info/) is a database of structural (phonological, grammatical, lexical) properties of languages gathered from descriptive materials (such as reference grammars) by a team of leading authorities on the subject. The online version is a joint project of the Max Planck Institute of Evolutionary Anthropology and the Max Planck Digital Library. Ethnologue: Languages of the World (http://www.ethnologue.com/) is a web (updated periodically) and print publication of SIL International (the Summer Institute of Linguistics) that contains published statistics; information in which countries a language is spoken in, number of speakers, literacy,

dialect names etc of over 7300 global languages. SIL international is a Christian linguistic service organization, which studies lesser-known languages, primarily to provide the speakers with Bibles in their native language. Locations of linguistic subfamilies within the larger four major African linguistic phyla, Khoisan, Niger-Kordofanian, Nilo-Saharan and Afroasiatic, mentioned in the text are also shown.

1.2.4: East African hunter-gatherer and Khoisan speaking populations

In East Africa all the hunter-gatherer populations speak the languages of their neighboring pastoral Nilotic and Cushitic populations with the notable exception of two Khoisan speaking populations of northern Tanzania, the Hadza and the Sandawe **(Figure 1.2.3)**. The Khoisan languages are the "click languages" that are spoken only in limited area in southern Africa by the San Bushmen (Kalahari Desert, primarily in Namibia and Botswana, and some part of South Africa) and eastern Africa (Central Tanzania) [119]. Speakers of these languages are speculated to have been widespread throughout a contiguous area that encompasses southern and eastern Africa regions prior to the Bantu expansion [120, 121]. A previous linguistic link between southern African click speaking populations and East African click speakers (Sandawe and Hadza found in the northern part of Tanzania) has been posited but deemed controversial ([71] and references therein).

In the current study I will use the definition of hunter-gatherers as mode of production [108, 119]. However, hunter and gatherer populations are defined by anthropologists in economic terms as being (1) in a stage of economic development before achieving a "higher" level of production [108, 119] (pastoralism, agriculture and agro-pastoralism); or (2) as a retrograde step from a food producing economy [120-122]. Some anthropologists and linguists suggest that these populations may have been the

original inhabitants of East Africa [113, 119]. Alternatively, East African hunter-gatherers could be collections of individuals who originally practiced pastoralism but lost their herds due to natural calamities such as drought and disease [120-122]. Others might have joined the hunting-gathering societies after losing their herds through intertribal war and some after being socially excommunicated from their original pastoralist communities [122, 123]. Unfortunately, most of the East African hunter gatherers have been viewed and named through the eyes and languages of their pastoral or agricultural neighbors who generally consider them with disdain, e.g. (Athi (Gikuyu), Asi (Taita), Asa (Tanzanian Bantu), Iltorobo (Maasai – one without wealth or tsetse fly), Sanye (Swahili - gatherers), Dahalo (Aweer – stupid or worthless person), Liangulu (Giriama – game eaters) Bon - for Boni (Somali – low caste) [108, 113, 122] Weyto (Amhara) [123]. However, pastoralists consider neighboring hunter-gatherer populations as valuable economic assets and thus maintain close interactions with them [122, 124]. For example; hunter-gatherers often participate in the local economy of neighboring populations as blacksmiths and game hunters, they conduct important rituals, they are often consulted for information on pasturage, and occasionally provide assistance in time of famine or plague. In general, hunter-gatherer populations appear to change their cultures to conform to the organizational and behavioral mode of communities with whom they have tight symbiotic relationships in order to facilitate amicable interactions [124, 125] with some exceptions such as the Hadza.

The Khoisan languages share a distinctive sex/gender system with Afroasiatic languages [71]. Moreover, the southern Cushitic languages have some phonological overlap with East African Khoisan languages [71]. Other linguistic studies have shown

that some of the current East African hunter-gatherer populations like the Ogiek [108] and Sanye [115] have Khoisan click words in their dialects. These facts have led some historians to speculate that Khoisan speakers once inhabited the contiguous area from southern Ethiopia to South Africa and were slowly pushed to the fringes or absorbed into immigrant Bantu speaking and pastoral communities [74, 126].

1.3: Early late stone-age cultures in East Africa

According to the archaeologists, the current East African hunter-gatherer populations once practiced related Holocene stone tool traditions collectively called Kenya Capsian [108, 127, 128], the most famous among them being the Eburran tradition, which was found in the central Rift valley [108]. Eburran is considered among the earliest refined stone tool tradition in East Africa [108, 113]; the earliest phase of this tradition is thought to have originated between 6 – 12 kya [108, 128]. The Eburran tradition had a recent phase from ~1.3 - 3.3 kya and coexisted in the same geographical range with two advanced Late Stone Age cultural traditions that are associated with pastoralism in East Africa, collectively called the Late Stone Age Neolithics or 'Pastoral Neolithic' [109, 127, 129-133]. The earliest of these pastoralist traditions has been called Savanna Pastoral Neolithic (SPN) [127, 130-132] and the more recent one is called Elmenteitan Pastoral Neolithic [108, 109, 127, 129]. These traditions were separated both in space and time but in some cases overlapped in one or both. The spatial area inhabited by the makers of Savannah and Elmenteitan Pastoral Neolithic cultures overlap but do not co-occur [110] suggesting a separation in time. Some historians have speculated that these traditions represented peoples with different origins and cultures [108-111]. The

SPN existed 5.0 - 1.3 kya in Kenya and northern Tanzania, and is thought to have been practiced by southern Cushitic speakers who originated in Ethiopia [108-111]. Elmenteitan Neolithic existed 3 - 2 kya and was found in western Kenya and the central Rift Valley and is thought to have been practiced by southern Nilotic speakers [108-111]. The Turkwel Neolithic tradition [132] is more recent and was practiced in Northwestern Kenya by eastern Cushitic speakers who also originated in Ethiopia [108-111]. According to Ambrose [110] there was consistent interaction between Eburran hunter gatherers and SPN which led to some Eburran hunter-gatherers taking up a pastoral lifestyle. By contrast, the Elmenteitan Pastoralist groups may have competitively replaced the Eburran hunter-gatherers [110]. The pastoral Neolithic was distributed as far south as Zimbabwe and other parts of southern Africa [63]. The pottery and iron age traditions that are associated with the Bantu migration into East Africa, Urewe [134, 135] Lelesu [136], Kwale [137] and Maore wares have been dated to 2.25 - 0.6 kya [110, 134, 138, 139]. The Lanet and Sirikwa wares [140] are traditions dated to 1.2 kya and are associated with eastern Nilotic speakers [110].

Besides the Eburran tradition [108], two other Capsian industries have been associated with Holocene hunter-gatherers in East Africa. The "Wilton" industries (no relation to Wilton in southern Africa), have been associated with Khoisan speakers [56] and began as early as 17 kya [109] but with most sites younger than 7 kya [109]. This industry was found in western Kenya [141], southeastern Uganda [141], northern and central Tanzania [56], where it persisted until about 2.5 kya [56, 109]. The other industry is a later stone-age aquatic related tradition dated 5 – 8.4 kya observed in the Turkana basin in northern Kenya (around Lake Turkana) that is associated with hunter-gatherers

adapted to foraging and exploiting aquatic resources [109]. The makers of this industry might have been influenced by Nilo-Saharan speakers, because their industry bears some resemblance to the early Khartoum pottery in the Sudan [84] that has been associated with Nilo-Saharan speakers. Historical linguists speculate that the producers of late stone-age cultural traditions associated with East African hunter-gatherers may have spoken Khoisan languages [74].

1.4: The role of paleoclimatic conditions on human demography

The demographic history of human populations has been influenced by large-scale climatic fluctuations [23]. This is because fluctuating resources that might result due to extreme climatic changes will impact human population densities [142]. Thus, the distribution and density of human populations in Africa has likely been regulated by the availability of water [143]. Since climates exhibit a broadly zonal pattern with varying seasonal distributions of precipitation [143], human populations have commonly used migration as a means of adaptation to relatively long term climate change and to reduce ecological stress [144]. The climatic changes that have occurred during the Pleistocene are thought to have affected the global distribution of humans into different ecological regions, with a series of range shifts, range contractions and expansions, isolation, local extinctions, demographic bottlenecks, and demographic expansions [145]. The human habitat shrunk during cold periods and increased during warmer periods, with ancestral hunting and gathering populations expanding and contracting in response to shifting resource opportunities as the climate changed [145].

The climate of tropical Africa is dominated by variability in effective moisture (precipitation minus evaporation), rather than temperature as in higher latitudes, and is driven by the circulation of the African monsoon and the seasonal migration of the Inter-tropical Convergence Zone [146]. Precipitation recharges groundwater and creates lakes, thus facilitating cultural contacts between regions previously isolated either partially or totally as a result of severe aridity [147]. Labuda *et al.,* [50] speculated that population dispersal and subsequent isolation induced by climatic events and local adaptation may account for the complexity of the fossil record and the genetic composition of modern human populations. Beside favorable paleoclimate conditions [17, 48], initial migration of anatomically modern humans across the globe was facilitated by advancement in socio-cultural and technological ability [148]. Improved technological ability may have led to competition between adjacent groups for resources due to population increases [148]. In fact, population densities of upper palaeolithic humans may have played a role in determining the timing of the spread of modern human behavior, which correlated with the spread of anatomically modern humans outside of Africa [148, 149].

1.5: Molecular Markers used in the study

Despite the important role of East Africa in human evolutionary and linguistic history, there are still relatively few genetic studies of this region [15, 16, 150-156]. Most of these studies examined genetic diversity for only one or two marker systems; *e.g.* mitochondrial DNA [*mt*DNA] and/or Y chromosome haplotypes defined by a set of single nucleotide polymorphisms [SNPs] in the non-recombining portion of the Y chromosomes [NRY]) in one or a few populations (*e.g.* the study by Kivisild *et al.,* [16] included only four Ethiopian populations, with most samples originating from the

frequently studied Amharas and Oromos). Behar *et al.,* [157] did complete sequencing of the *mt*DNA genome in several individuals from East Africa from unknown ethnicity to make inferences about the possible structure of ancestral populations in Africa [157].

In order to explore possibilities of sex-biased gene flow, genetic diversity within and between populations and genetic correlations with linguistic and geographic regions were examined for both *mt*DNA and Y chromosome data. *Mt*DNA and the NRY are the only two genetic regions experiencing no recombination in humans. In contrast to autosomal loci which experience recombination, the effects of demography are confounded with those of natural selection and genetic drift only in *mt*DNA and Y chromosome loci. For this reason, patterns of variation based on the maternally inherited *mt*DNA genome, paternally inherited non-recombining portion of the Y chromosome (NRY) and previously collected autosomal microsatellites data were compared.

1.5.1: Y Chromosome

The Y chromosome is particularly informative for reconstructing human evolution due to uni-parental inheritance via males and the absence of recombination. The NRY, which encompasses 95% of the Y chromosome (the other 5% constitutes the pseudoautosomal region that undergoes recombination with the homologous X region with the same name), has a much lower level of polymorphism than most regions of the human genome [158]. The slower mutation rates of insertion or deletion (indels) and SNPs give rise to unique event polymorphisms (UEPs) which are unlikely to undergo homoplasy. Unique-event polymorphisms are markers that correspond to a mutation that occurs so infrequently that all the individuals who share the marker in global populations will probably have inherited it from the same common ancestor and is therefore used to

identify deep lineages of the NRY [158]. Faster mutating microsatellites, in combination with more stable site markers (UEPs) analyzed as haplotypes, are particularly informative for reconstructing recent evolutionary events [158].

African populations exhibit greater NRY diversity than other continents when diversity is measured by the mean number of pair-wise differences among haplogroups [158]. Thomson *et al.,* [159] observed continental structure in the NRY tree, with the oldest clade present only in Africans and the younger clades present in both African and non-African populations. Moreover, studies of Y chromosome microsatellite polymorphisms revealed substantial continental structure [160, 161]. The Y chromosome consortium established a system of defining Y-DNA haplogroups by letters A through to T, with further subdivisions using numbers (from 1) and lower case letters (from a) (Yconsortium 2002). Based on the latest Y consortium's classification of Y haplogroups (Yconsortium 2002), haplogroups A, B, E and J (haplotype in F) **(Figure 1.5.1)** are found in Africans, while C, D and the groups descending from haplogroup F (that encompasses haplogroups F to T) are found in most of the global population, but almost exclusively outside of sub-Saharan Africa.

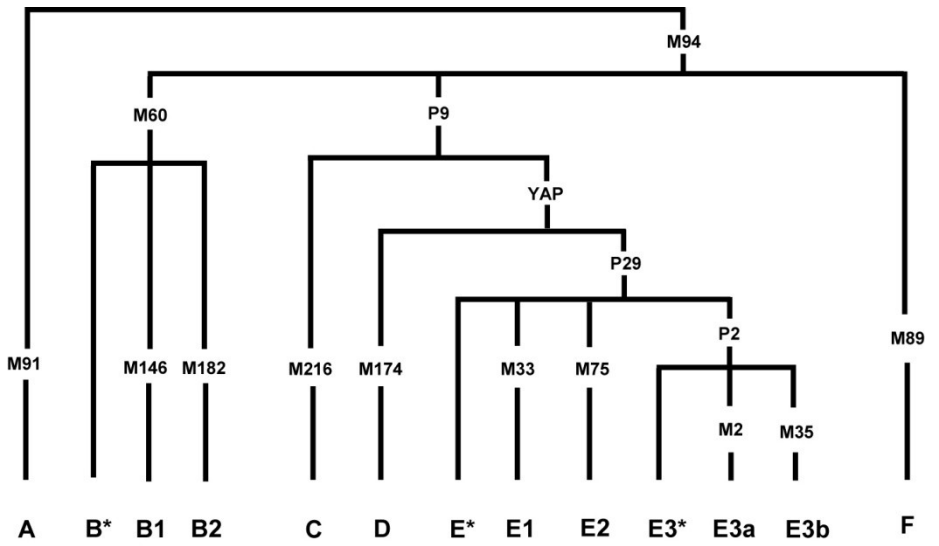


Figure 1.5.1: Nomenclature for major lineages of Y chromosome haplotypes. Note that this is not a phylogenetic tree, but a scheme to represent the classification of the haplogroups and their relationship to each other. The schematogram is based on a detailed one in the results section that follows the scheme of Y chromosome haplogroup trees inferred by Maximum parsimony method rooted using non-human primate sequences [14] and maximum likelihood from the 2870 variable sites identified in 1000 genome project [162]. The M#, P# and YAP labels leading to haplogroup/s are SNPs and indels that are used to define these haplogroups. Haplogroup F encompasses haplogroups F to T. Haplogroups A, B and E are mainly found in Africa while the rest are found mainly outside Africa

Nearly all of the sub-Saharan African Y chromosome lineages belong to haplogroups A, B, and E. Haplogroup A, the haplogroup with the largest average number of mutations per lineage, and B are Africa-specific [158]. Haplogroup A has the highest diversity among all haplogroups observed in Africa [158] suggesting it is the oldest. The nearly exclusive observations of the A haplogroup among the Khoisan and a few East African populations prompted Semino *et al.,* [155] to suggest that Khoisan speaking populations of South Africa were originally from East Africa. One of the most common B lineages, B2b, are most prevalent in Central and southern African populations whereas, the other common B lineages, B2a, are more common in populations from Sudan and Ethiopia [163]. Haplogroup E represents the great majority of Y chromosome haplotypes

found in sub-Saharan Africa [164]. The E3a-M2 haplogroup is typically the most wide-spread in sub-Saharan Africa, and is the most common haplotype in Central and West African populations [156, 160, 163]. Luis *et al.* [156] show a West-East as well as a North-South clinal distribution of the E3a-M2 haplogroup, with highest frequency in Central Africa. The spatial pattern and high frequencies of the E3a-M2 haplogroup indicate that this may be associated with the agricultural expansion of Bantu speakers from Nigeria/Cameroon within the past 5 ky [163]. However, the East-West clinal distribution of the E3b-M35 haplogroup is inverse to that displayed by E3a- M2 [156] and it has been suggested that the E3b-M35 haplogroup originated in East Africa [165]. F derived haplogroups, typically found outside Africa, are generally absent in sub-Saharan Africa, except in a few East African populations [156].

*1.5.2: mt*DNA

Mitochondrial DNA has proven to be a useful tool in studying the evolutionary history of human populations due to several key characteristics such as a high copy number (up to 5,000 – 10,000 times that of single-copy nuclear sequences), a lack of recombination [166-168], a high substitution rate [169], and, as mentioned above, a maternal mode of inheritance [170]. Most phylogenetic analyses of *mt*DNA have been based on sequence variation in the hyper-variable segments (HVS-I and HVS-II) of the control region D-loop [171, 172]. Both high mutation rate (over 10 times that of the protein coding region of *mt*DNA [45, 173]) and variation in the substitution rate among sites make the evolution of HVS-I and HVS-II complex. Parallel mutations at some sites (homoplasy) [166, 174] and mutational "hotspots" [175] make phylogenetic analyses of *mt*DNA control region data prone to ambiguity and haplotype networks are often difficult

to resolve. Moreover, phylogenetic trees constructed from *mt*DNA D-loop sequences can have incorrect topology compared to trees constructed from complete *mt*DNA sequences [166, 174, 175].

Mitochondrial DNA surveys in worldwide populations have also demonstrated a continental structure in the distribution of *mt*DNA lineages [167, 176-179]. Just like in the case of Y chromosome haplogroups, African populations are characterized by having haplogroups from the oldest lineages, L0-L6 [34, 151, 152, 180-187], as well as the M1 haplogroup, an L3 sub-lineage **(Figure 1.5.2)**. The estimated TMRCA of L0 and L1 lineages is over 130 kya, while the TMRCA of L2 and L3 *mt*DNA haplogroups, which represents more than two-thirds of the extant *mt*DNA lineages in Africa, is ~85 kya [152]. A subset of L3 lineages, haplogroups M and N radiated out of Africa ~60 kya through East Africa, giving rise to all *mt*DNA haplogroups outside of Africa [15, 152, 179] **(Figure 1.5.2)**.

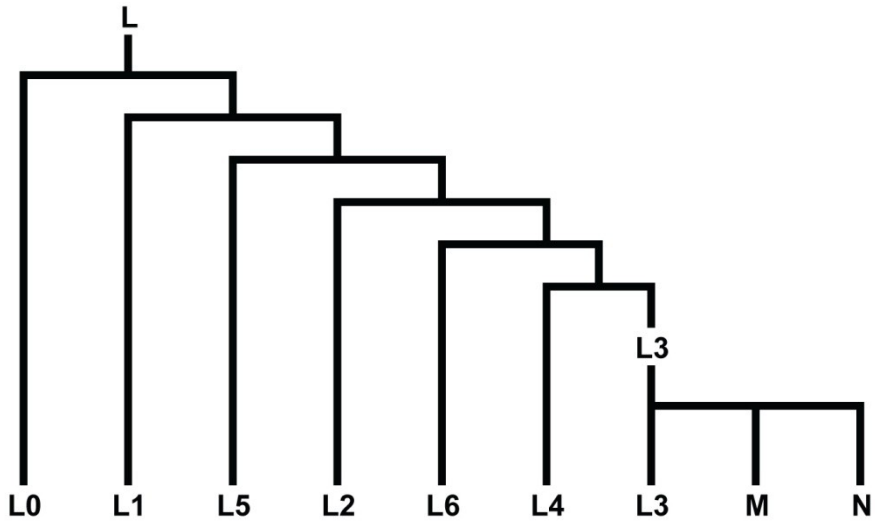


Figure 1.5.2: Overview of mtDNA haplogroup phylogeny. The tree, based on both coding and control region mutations, is abstracted from the detailed one in Results (Fig. 3.4.1) that was consistent with trees generated using Bayesian approach [188, 189] **(Appendix 5)** and with previously published trees generated using a maximum likelihood [190] and a PCA based method [191]. In the *mt*DNA haplogroup nomenclature, the letter names of the haplogroups run from A to Z, with further subdivisions using numbers (from 0) and lower case letters (from a). The naming was done in the order of their discovery and does not reflect the actual genetic relationships. Haplogroup M and N encompasses all the haplogroups lettered A to Z excluding haplogroup L. Haplogroups L (L0-L6) are mainly found in Africa while the rest are found mainly outside Africa [15, 16, 157, 171, 172, 185].

African *mt*DNA profiles inferred from studies of populations from mostly other parts of Africa and a few populations from East Africa indicate that the oldest lineages, L0, originated in East Africa and derived lineages (**Figure 1.5.2)**, L1 and L2 originated and expanded in Central/West Africa [185]. L0a may have originated in East Africa [185], and L0a2, L1c and L1b may have originated in Central Africa [185]. Haplogroup L0d represents about half of the total haplogroup composition for the southern African Khoisan (SAK) speaking populations [151, 183, 192]. However, based on phylogeographic and linguistic analyses, as well as the presence of L0d sister clades in East Africa, an East African origin for the SAK has been proposed [157]. Haplogroup L5

is mostly observed in East Africa, with only low frequency of the lineage observed in Central African hunter-gatherers (Mbuti Pygmies) and southeastern Bantu speakers [34, 183, 185, 193]. L0k lineages have exclusively been found in southern African Khoisan-speaking populations [34, 183, 185, 193]. L2 haplogroups, L2a-L2d are found in Central/West Africa and, thus, their origin has been suggested by Salas [185] to be in Central/West Africa.

1.5.3: Microsatellites and Diallelic Indels

Insertion or deletion polymorphisms (indels), consisting of one or more bases in the genome [194], are typically di-allelic [194]. Indels vary in length from a few nucleotides to several kilobases [194]. The indels data from the Marshfield Institute that were analyzed in the current study have allele-length differences of a few nucleotides [4]. Microsatellites, or short tandem repeats (STRs) [194] consist of 1 – 6 bp tandem repeat units which vary in repeat number between individuals. They are generally assumed to be selectively neutral, and are highly variable owing to their high rate of gain and loss of repeat units by DNA replication slippage (slipped-strand mis-pairing). According to the stepwise mutation model (SMM) [195], a model that defines microsatellite mutation patterns, a single repeat unit is either gained or lost leading to an expansion or a contraction, respectively. However, rarely insertions or deletions involving more than one repeat unit also occur [196]. Contractions are more likely for microsatellites with >20 repeats while expansions are more common for shorter repeats [189]. Repeat numbers are rarely over 30 units [197]. Rare exceptions include certain tri-nucleotide repeat loci implicated in a set of human genetic disorders called “Trinucleotide Repeat Disorders”

(TRD) [198-200]. Microsatellites are inherited in a Mendelian fashion and are widely distributed across the human genome, with an estimated average density of one microsatellite per 2 – 30 kb [201]. The variation and abundance of microsatellites make them excellent markers to answer population genetic and demographic questions (*i.e.* genetic diversity, population structure, migration patterns, population expansions, *etc* [47, 202, 203]).

The rationale for using different markers from the *mt*DNA genome, Y chromosome, and autosomes is based on the fact that previous studies using the (NRY) [153], *mt*DNA [164, 204], and autosomal SNPs and STRs [205, 206], have had contrasting, and sometimes conflicting results in regard to the genetic distances inferred between populations, migration patterns and estimates of the time to most recent common ancestor (TMRCA) of lineages. Contrasting patterns of genetic relationships described in these studies for the same populations may have been due to differences in the effective population sizes of the *mt*DNA, Y chromosome and autosomal DNA [207]. NRY and *mt*DNA behave as single locus systems and are subject to large stochastic effects [207]. They also have a four-fold smaller effective population size relative to autosomal regions due to uni-parental inheritance and their haploid nature, enhancing the effects of genetic drift [208]. Although *mt*DNA and Y chromosome lineages have similar effective population sizes, they show different patterns of variation. They also give different estimates for the TMRCA of modern human, with estimates of TMRCA with *mt*DNA lineages twice that of NRY lineages; ~200 kya versus ~100 kya respectively [158, 159, 167, 209, 210]. Under the expectation of neutrality, *mt*DNA and the NRY loci TMRCA should be primarily a function of effective population (*Ne)* and, consequently, similar

TMRCA estimates should be obtained for the two loci [207], assuming certain conditions such as equal male and female *Ne*, and equal male and female migration. Unequal migration rates and higher rate of female migration (due to patrilocality) in comparison to male migration rate have been suggested to explain the difference in TMRCA estimates [211, 212]. Wilder *et al.,* [207] hypothesizes that the disparity between TMRCA estimates is due to a higher effective sex ratio of females in relation to males, an idea that concurs with the fact that polygamy was common in earlier human populations and practiced by most extant native populations [213]. As part of my thesis, I examined if there are concordance in patterns of genetic variation between different marker systems (*mt*DNA, NRY and autosomal microsatellites) in order to distinguish male and female demographic histories.

Chapter 2: Materials and methods

2.1: Sample collection

In this study, Y chromosome SNP and *mt*DNA D-loop sequence data were generated for ~1500 individuals from 55 East and Central African populations and compared to previously collected microsatellite variation [3, 4] in the same populations **(Appendix 1 and Figure 2.1)**. In addition, whole *mt*DNA sequence variation was examined in a subset of 222 individuals with diverse *mt*DNA haplotypes **(Appendix 2)**. These data were compared with previously published data from other regions of Africa and southwestern Asia **(Appendix 1, Appendix 6)**. The samples included in this study are from East African populations speaking languages belonging to four African language families; Khoisan, Nilo-Saharan, Afroasiatic and Niger-Kordofanian from Kenya, Sudan, Ethiopia and Tanzania **(Appendix 1)**. Populations that represent hunter-gatherer groups in Africa; *e.g.* Pygmies and click speaking South African Khoisan populations, which might be representative of early occupants of sub-Saharan Africa, were also included in this study.

Members of the Tishkoff lab collected samples in Tanzania, Kenya, Cameroon and Sudan over the course of several field expeditions. Two collections were done in each country: Dr. Tishkoff and collaborators led sampling expeditions, in 2001 and 2002 in Tanzania and in 2002 and 2004 in Cameroon, while I led expeditions to Kenya in 2004 and 2006, and Sudan in 2003 and 2004. Internal Review Board (IRB) approval was received through the University of Maryland (IRB# 00992) as well as proper governmental authorities and appropriate informed consent obtained from all participants.

Local researchers, nurses or medical officers, and translators were hired to assist explaining to the local officials/leaders, and subsequently to the participants in the study, the goals of our research. After ensuring that all potential participants (limited to individuals 18 years or older) have understood the purpose, risks, and benefits of our study, and have given signatures/thumbprints on a translated informed consent form, we sampled blood and/or or cheek cell swabs based on individual participant's choice. Cheek cells were sampled by subjects brushing the inside of their cheek about 5 times each with a sterile brush and then storing the brush in its container for later analysis. Blood samples were collected (10 ml or less of blood) from each consenting subject by venupuncture. Additionally, information about the geographic origin and a three generation pedigree of each individual was recorded for all individuals sampled. Individuals filled out a questionnaire designed to gather information regarding the linguistic and ethnic groups to which subjects belong, and their samples were assigned an anonymous numerical identifier.

DNA from white cells was extracted from whole blood samples using a salting out process [214]. In the field, we first added a red blood cell lysis buffer, and spun down the white cell pellet in a portable centrifuge. After several washes intended to remove red blood cells (RBC), we added a white cell lysis buffer which stabilizes the DNA at room temperature. Samples were brought back to the laboratory (in the U.S.A) in 2 ml tubes. DNA was extracted by a modified salting out process using the Gentra Puregen DNA extraction kit (Gentra® systems). DNA was quantified using pico-green reagent in a fluorescent plate reader (Victor^2® by Perkin Elmer).

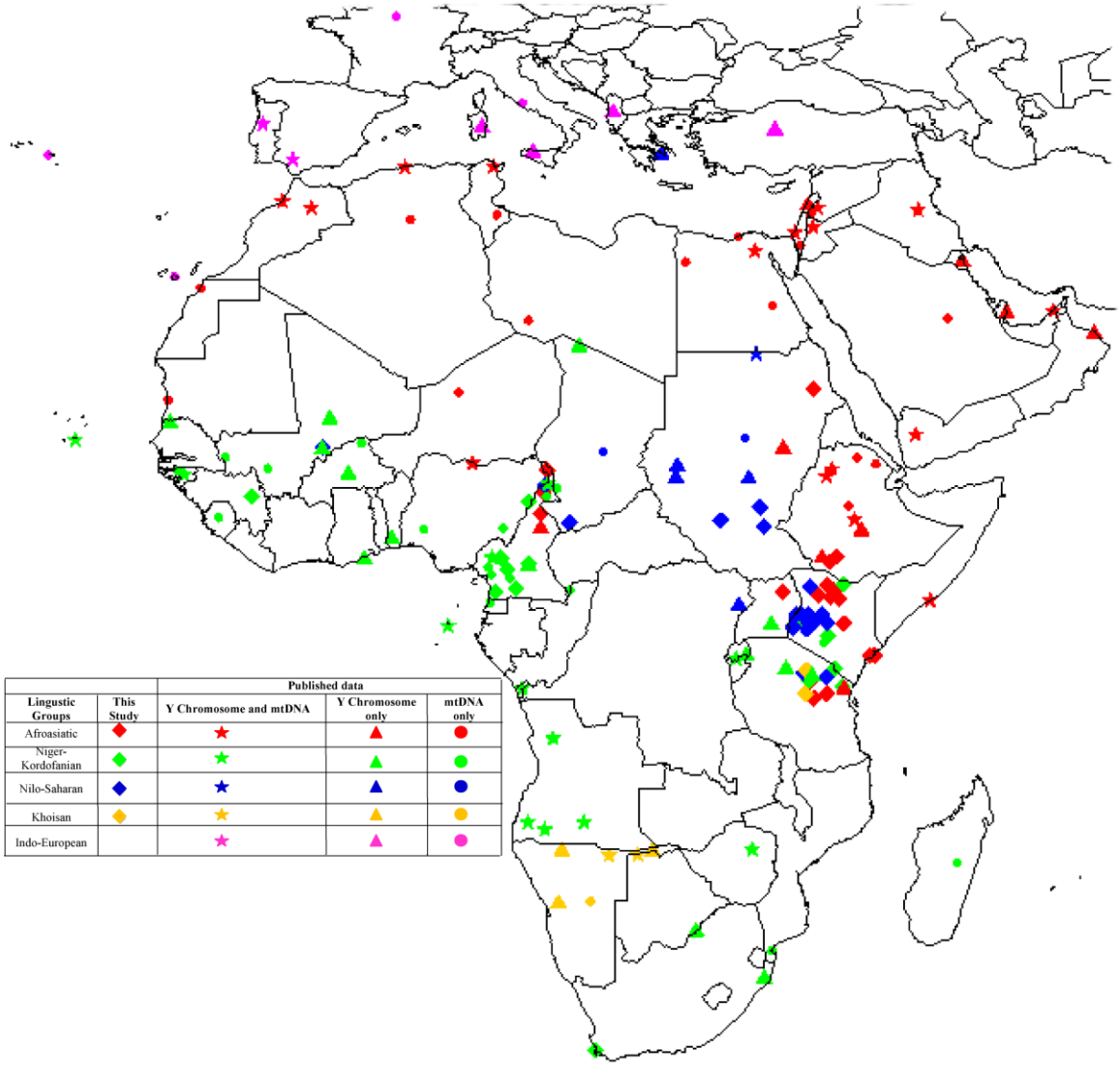


Figure 2.1: Geographical locations for populations genotyped in this study **(Appendix 1)** and published data included in the study **(Appendix 6, 12, 13)**. Label shape distinguishes marker systems (Y chromosome and *mt*DNA), and population genotyped in the current study and published data. Color codes represent language family of the population. Diamond labels represent populations genotyped in this study. For published data, star labels represent populations where both Y chromosome and *mt*DNA data are available, while triangles and full circles represent populations where only Y chromosome and *mt*DNA data are available, respectively. In each case the position of the label on the map is based on geographical coordinates of the populations as in **Appendix 1** and **Appendix 6**.

2.2: DNA Amplification and Sequencing

Y chromosome markers were genotyped in 1500 individuals to ascertain Y chromosome lineages using 55 previously studied binary markers and an insertion polymorphism, the YAP (a Y-specific polymorphic *Alu* insertion [215]), that define African Y haplotype lineages (according to the Y Chromosome Consortium (YCC) standardized nomenclature [216] (**Figure 2.2.1)**. For the Y chromosome study, an additional 16 STR loci were genotyped using ABI's Yfiler® kit; a hexanucleotide repeat DYS448, a pentanucleotide repeat DYS438, seven GATA tetranucleotide STRs; (DYS439, DYS389I, DYS389II, DYS390, DYS19, DYS391 and DYS393); one AGAT tetranucleotide DYS456, two GAAA, tetranucleotide STRs DYS458 and DYS385, and two other tetranucleotides, DYS437 and GATA_H4. Three of the STRs, GATA_H4, DYS456 and DYS439, failed to amplify in most samples and so were removed from subsequent analyses.

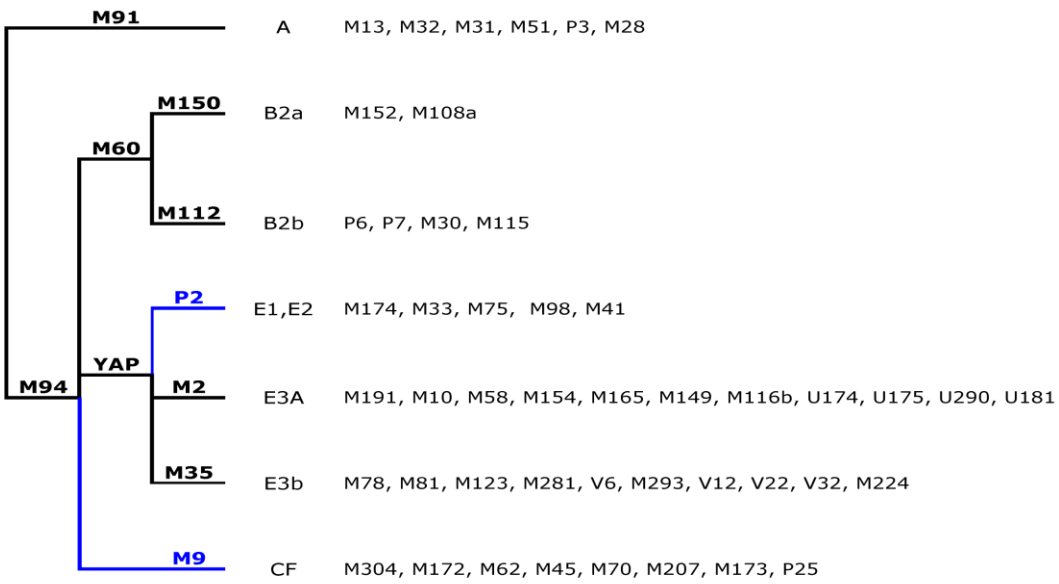


Figure 2.2.1: Strategy used to genotype the Y chromosome. Initially 8 primers; M91, M60, M150, M112, M94, YAP, M2, M35 were used to classify individuals into 5 common base clades found in Africa. Next, samples were hierarchically genotyped for the remaining primers (including the two that define the clades E1, E2 and CF – in blue - that are not common in Africa) until all samples were assigned to a specific terminal haplotype.

The *mt*DNA "D-loop" region was sequenced in 1500 individuals. The "D-loop" sequenced in this study constitutes an ~1300 bp region (base positions 15900-640 in the genome) that spans both sides of an arbitrary "0" basepair (bp) position of the mitochondrial genome and encompasses two hypervariable regions; HVR-I (Hyper variable control region 1) and HVR-II which are from positions 16024 to 16383 bp and from positions 73 to 340 bp, respectively (**Figure 2.2.2, Appendix 3**) [217]. Details of primers used for D-loop sequencing are shown in **Table 2.2.1**. Sequencing of 300-500 bp fragments of coding regions that have haplogroup/haplotype diagnostic SNPs (**Appendix 4**), was subsequently performed in cases of haplogroup/haplotype ambiguity or to confirm designations based solely upon the rapidly evolving HVI and HVII mutations. Primers used to characterize SNPs in the coding region are indicated in **Appendix 3**.

mtDNA CR PCR primers[a]	Sequence (5'-3')
L15793	TCATTGGACAAGTAGCATCC
H16571	GGGATGCTTGCATGTGTAATC
mtDNA CR Sequencing Primers[a]	**Sequence (5'-3')**
L15793	TCATTGGACAAGTAGCATCC
L16325	ACCGTACATAGCACATTACA
L00158	TTTATCGCACCTACGTTCAAT

[a]**Primer positions are relative to their location in the Cambridge Reference Sequence.**

Table 2.2.1: List of *mt*DNA control region (CR) primers used in this study. L and H primers represent forward and reverse primers respectively. L15793 and H16571 correspond to D-loop primers 23F and 709R in **Figure 2.2.2,** respectively.

Whole *mt*DNA genomes (~16,600 bp) were further sequenced in a subset 222 samples, representing 36 haplotypes (**Appendix 2**). Individuals were selected for whole *mt*DNA genome sequencing based on a proportional representation of all of the major *mt*DNA haplogroup clades, as determined by *mt*DNA D-loop sequencing. This was done by first amplifying the *mt*DNA genome using 3 primer pairs and sequencing overlapping 6 kb PCR fragments using 48 primers that produced overlapping sequence reads (**Figure 2.2.2, Appendix 3**) [187, 218]. For *mt*DNA complete genome sequences, quality control was assured by initially determining each base pair with both forward and reverse primers. This was followed by additional independent PCR and sequencing reactions for any ambiguous base calls, and finally all sequences being examined by another independent investigator (Dr. Toomas Kivisild, University of Cambridge, UK). In each case no multiple amplifications of fragments were obtained, as visualized on agarose gels. No inconsistent positions were detected in alignments, which were assembled from

overlapping sequences (from both forward and reverse primer sequencing). Cross-contamination between individuals was also pre-empted as well, since all negative controls revealed no amplifications and randomly repeated PCRs for the same individual produced identical sequences. About 700 previously published complete sequences [157, 167, 177, 178, 219-230] were also included in the final analysis for this study. The haplogroup nomenclature used in the current study follows that of previous analyses [15, 16, 157, 171, 172, 185].

Successful amplification of Y chromosome markers and *mt*DNA primer products was verified through electrophoresis on a 2% agarose gel stained in 1% ethidium bromide. Sequencing was performed using BigDye Terminator chemistry (Applied Biosystems) on a 3730xl DNA Analyzer (Applied Biosystems), and the resulting sequences were aligned by the program SEQUENCHER ver. 4.8 [231] and DNA alignment Software (Fluxus Engineering). Mutations were scored relative to the revised Cambridge Reference Sequence (rCRS) [232].

The autosomal microsatellite/indel Marshfield data analyzed in the current study was previously genotyped at the Marshfield Clinic Research Foundation, Marshfield, WI 54449, USA, and consisted of a total of 1327 autosomal markers (848 microsatellites and 479 insertion/deletion polymorphisms). The initial analyses of this dataset, and further description of the data, are published in Tishkoff *et al.,* [4]. The data is available at www.med.upenn.edu/tishkoff/Supplemental/files.html and at http://chgr.mc.vanderbilt.edu/page/supplementary-data. Data for some markers and individuals were removed from the original data because they either did not amplify properly or were from related individuals. Numbers of same samples genotyped for the

three marker systems (autosomal, *mt*DNA D-loop and Y chromosome) are listed in **Appendix 1**.

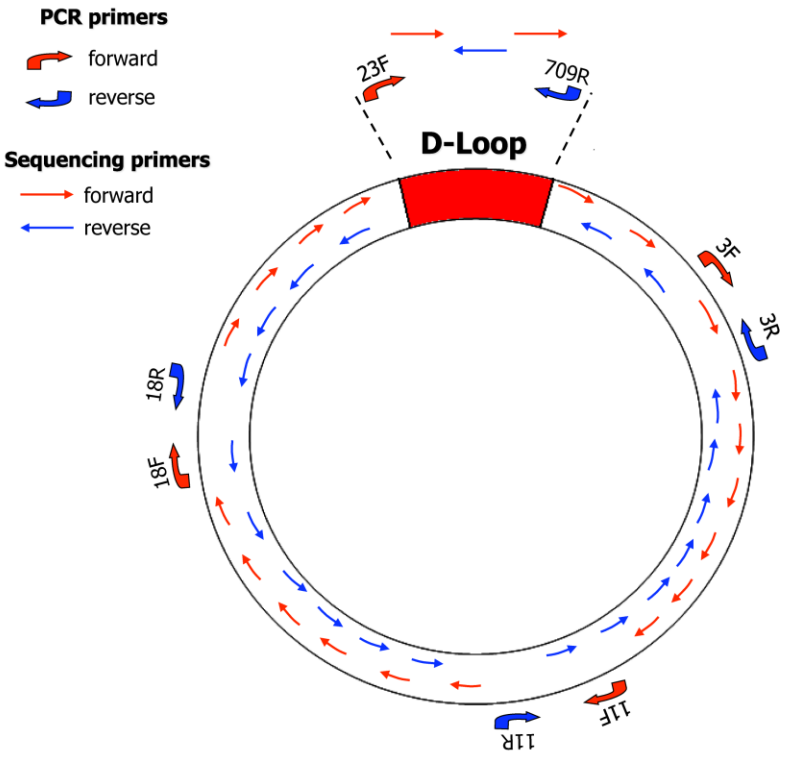


Figure 2.2.2: Sequencing strategy of *mt*DNA genomes used in this study. Coding region was sequenced by initially amplifying the genome into three ~6kb fragments using three pairs of primers; 3F/11R, 11F/18R, 18F/3R. These are followed by complete sequencing of the three fragments using a total of 48 forward (small red arrows) and reverse primers (small blue arrows) **(Appendix 3).**

2.3: Statistical Analysis

Models of population genetic analysis are limited by their assumptions which can include constant population size, an infinite sites model (unlimited number of sites at which mutational changes can take place), non-overlapping generations, random mating, selective neutrality and no recombination. However in most natural populations, one or more of these assumptions are frequently violated [233] and no single statistical test or model can simultaneously account for all of the possible evolutionary forces that have likely influenced the observed patterns of polymorphism. Therefore several complementary tests and models to investigate within- and between-population genetic diversity, the relationships among populations (using both phylogenetic reconstruction and multidimensional scaling analysis), population structure, demographic parameters (population size and growth rate, admixture), and gene flow were used in order to facilitate overall inference regarding the evolutionary histories of East African populations.

2.3.1: Tests of Neutrality

Several tests of neutrality were applied to complete mitochondrial sequences: Tajima's ***D*** statistic [234] and Fu and Li's ***D**** *and* ***F**** statistics [235] were calculated using the program DNAsp [236]. These tests are based on the difference between alternative estimates of the mutational parameter $\theta = 4N\mu$ ($\theta = 2N\mu$ for haploid *mt*DNA and Y chromosome) where N is the effective population size and μ is the per generation mutation rate. Tajima's ***D*** is based on the difference between two estimates of θ, one based on the number of segregating sites (***S***) [237] and the other based on the number of

average pairwise differences in a population sample ($\boldsymbol{\pi}$) [238]. Tajima's D test statistic is based on the fact that purifying selection is expected to affect the number of segregating sites more than the average number of nucleotide differences, while both balancing selection and positive directional selection are expected to affect the average number of nucleotide differences more than the number of segregating sites [234, 239]. Fu and Li's *F* statistics are based on a comparison between the number of "young" mutations (singletons) with other measures of genetic variation, $\boldsymbol{\pi}$ *and* $\boldsymbol{S}$ [235]. In a constant-size population at neutral equilibrium, the expectation of Tajima's $\boldsymbol{D}$ is nearly zero because the expectations of both estimates are the same. Under balancing selection, Tajima's $\boldsymbol{D}$ tends to be positive. On the other hand, purifying selection and positive directional selection can generate negative values of Tajima's $\boldsymbol{D}$.

Besides selection, any other violation of the assumptions of population equilibrium would lead to non-zero estimates of these statistics. For example, changes in population size can affect estimates of both Tajima's and Fu and Li's statistics. For example, recovery from a recent bottleneck event [240] and population expansion [235] can mimic the effect of selection because these demographic events result in high numbers of rare alleles (nucleotide sites with low frequency) [233, 235] and a reduction of the number of segregating sites [241, 242]. Both Tajima's $\boldsymbol{D}$ and Fu and Li's $\boldsymbol{D^*}$ statistics will have negative values in the case of demographic expansion and positive values under recent demographic contractions [212]. Tajima's $\boldsymbol{D}$ will also have negative values when there is undetected population structure [212], which could result in an excess of rare variants if structured populations are inadvertently pooled [212].

The numbers of synonymous (S) and nonsynonymous (A) substitutions in the 13 protein-encoding *mt*DNA genes for each of the haplogroups were compared using Elson's *et al.'s* [243] approach in MtPhyl [244]. Elson *et al*., [243] used a modified intraspecies index that is based on the interspecies Neutrality index (NI) [245], a parameter that provides a measure of the direction and magnitude of a gene's departure from neutral expectations [245]. It is calculated as the ratio of polymorphic replacement/silent sites to fixed replacements/silent sites (within the species between *mt*DNA haplogroup lineages). Therefore, the ratio of nonsynonymous divergence (dN) to synonymous divergence (dS) between lineages is expected to be equal to the ratio of nonsynonymous diversity to synonymous diversity between lineages within a species. N.I. values range from 0 to ∞, and under strict neutrality the ratios in the numerator and denominator are expected to be equal (equal ratios of polymorphism to divergence for both replacement and silent sites i.e. 1) [245, 246]. NI scales with selection where values less than 1 indicate an excess of amino acid fixations, or positive selection and NI values greater than 1 indicate an excess of amino acid polymorphisms, or negative (purifying) selection [247-249]. The test is analogous to a McDonald-Kreitman test (MK test) [250], but while the MK test provides a statistical statement of the departure from neutral expectations, the NI test yields values that may indicate positive or negative selection [249].

2.3.2: Genetic Diversity and AMOVA Analysis

Pairwise population genetic distances were calculated in ARLEQUIN ver3.1 [251] for autosomal data [252], *mt*DNA data and Y haplotype data [253, 254], and 10,000 permutations were used to determine significance in each case. F statistics [255] have traditionally been used to test the measure of differentiation in sampled populations and are usually calculated from allele frequencies at many independently inherited loci. For *mt*DNA data and Y haplotype data differentiation among populations was summarized by generating $\boldsymbol{\Phi_{ST}}$ statistics (explained below), while for microsatellite data R_{ST} statistics [252], a statistic that is analogous to Wrights F_{ST} [255], but which incorporates the rapid mutation rate and stepwise changes in repeat units (stepwise model of mutation) characteristic of microsatellites to weight distances [252], was generated. Estimate of genetic diversity [256] were computed for *mt*DNA sequence data and genetic distances between pairs of populations were estimated for *mt*DNA and NRY haplotypes using hierarchical analysis of molecular variance (AMOVA) [253] implemented in ARLEQUIN ver3.1 [251]. Excoffier [253] described a method akin to Wright's F statistics, but more appropriate for the analysis of restriction site and sequence data from non-independently inherited non-recombining loci such as *mt*DNA and the Y chromosome. The authors construct a hierarchical analysis of molecular variance directly from the matrix of squared-distances between all pairs of haplotypes (based on a simple count of the nucleotide differences between a pair of haplotypes [253]). This is achieved by initially treating all haplotypes as a matrix of presence or absence of SNP or Indel alleles. Distances between haplotypes can then be calculated by subtracting matrices for pairs of haplotypes. Squared count distances are further generated for all combinations of

pairwise arrangements of haplotypes, which are then arranged into a matrix, and partitioned into submatrices corresponding to subdivisions within the populations. These arrangements are structured such that the submatrices on the diagonal of the larger matrix are pairs of individuals in the same population while those outside the diagonal represent pairs of individuals from different populations. The sums of the diagonals in the matrix and submatrices yield sums of squares, related F-statistic analogues which are denoted as correlation statistics $\boldsymbol{\Phi}$, for three hierarchical levels of the population: $\boldsymbol{\Phi_{ST}}$ is the correlation of random haplotypes within populations, relative to that of random pairs of haplotypes drawn from all the populations; $\boldsymbol{\Phi_{CT}}$ is the correlation of random haplotypes within a group of populations, relative to that of random pairs of haplotypes drawn from all the populations, and $\boldsymbol{\Phi_{SC}}$ is the correlation of the molecular diversity of random haplotypes within populations, relative to that of random pairs of haplotypes drawn from a particular geographical region. The data were analyzed for differentiation within and among populations as well as within and among linguistic groups in East Africa. I compared $\boldsymbol{\Phi_{ST}}$ for *mt*DNA and NRY to test for differences in male and female migration rates.

2.3.3: mtDNA complete sequence phylogenetic Analysis

The high mutation rate of the *mt*DNA D-loop is expected to occasionally generate homoplasies. This leads to irresolvable phylogenies using traditional maximum parsimony and maximum likelihood tree building methods. Therefore, the complete *mt*DNA sequence phylogeny was reconstructed in MtPhyl [244]. MtPhyl uses a novel maximum parsimony-based algorithm for reconstruction of the human *mt*DNA

phylogeny, which consists of three successive stages: identification of potential homoplastic mutations, analysis and identification of true homoplastic mutations, resolution of back and parallel mutations and sequences. Sequences with the least mutational differences are grouped into haplogroups/haplotypes [244]. The topology of the worldwide *mt*DNA tree constructed using MtPhyl [244] was consistent with trees generated using Bayesian approach [188, 189] **(Appendix 5)** and with previously published trees generated using a maximum likelihood [190] and a PCA based method [191] in most respects, with certain minor exceptions in internal branches. In addition to 222 newly collected *mt*DNA sequences, 700 additional complete genomes from the literature were used for the tree reconstruction. For phylogeny construction, the length variation in the poly-C stretches at nucleotides 16180–16193 and 309–315 were not used.

For some haplotypes where complete sequencing was not done in this study, L1 and L2, control region networks were constructed using the Network program version 4.1.0.9 [257, 258]. Networks are generated by partitioning groups of haplotypes based on character differences. The characters are inferred by first generating a matrix table and all samples (sequences) are compared with a reference sequence. For each variable site (mostly a transition) a simple presence or absence code is assigned. However, if the variant site has three nucleotide states including a transition, a transversion or an indel a special coding can be used to distinguish them from other characters. Further reduction of characters is achieved by omitting ambiguous sequences/sites. Haplotypes are next partitioned into manageable components consisting of closely related haplotypes. A network will then be constructed by connecting haplotypes with each variant site representing a node. In the final network, each haplotype is represented by a circle whose

size reflects it's frequency in the samples analyzed. If there is more than one step (two or more site differences) between observed characters, the hypothetical intermediate variants are represented as points rather than circles [182, 257, 259]. The most parsimonious tree of haplotypes for the sequences is then subsequently inferred.

2.3.4: Haplotype Age Estimates and Lineage Expansion Times

a) *mt*DNA

*mt*DNA haplogroup/haplotype ages were inferred using the *rho* statistic (*ρ)* [260], an unbiased estimator, which is the average transitional distances of individual sequences from the inferred root haplotype. Coalescence ages of haplogroups/haplotypes and their standard errors were calculated based on the network, by means of the average transitional distance from their respective root haplotypes, ρ. The dating method used is based on the average number of mutations accumulated relative to the ancestral sequence (calibrated with reference to the human-chimpanzee divergence) as a linear function of time and mutation rate [261]. The mutational rates calibration used for the *mt*DNA control region is 7.68×10^{-7} mutations/site/year [260] and for the coding region is 1.7×10^{-8} mutations/site/year [167]. Forster *et al.*'s [260] method uses only substitutions within the control region and estimates haplotype age by taking 1 transitional step between nucleotide positions 16090 and 16365 as equal to 20,180 years [260], while Mishmar *et al.*'s [178] and Kivisild *et al.*'s [220] methods use substitutions in the coding region, based on a mutation rate of 1 base substitution between nucleotide positions (nps) 577 and 16023 equal to 5,140 years [178] and synonymous substitutions changes at protein-coding sites at a rate of 1 transition per 6,764 years [220], respectively. Age estimates

were also obtained based on a new mutation rate estimate by Soares *et al.,* [190] that is corrected for the confounding effects of saturations and purifying selection, with the average rate for a synonymous mutation taking place in the coding region estimated to be 1 per 7884 years. The synonymous substitution rate estimate [220] has been used as complementary method in cases where excess nonsynonymous substitutions were detected within *mt*DNA haplogroups. This occurs mostly for young haplotypes where Mishmar *et al.*'s method has high variance. The standard deviation of the *rho* estimate was calculated as in Saillard *et al.,* [261].

b) Y chromosome

To evaluate whether the observed geographic distribution for each Y chromosome binary haplotype is attributable either to recent or more ancient events, haplogroup-specific networks were constructed using the Median Joining Network method [182, 259] based on variation at the 13 STR loci on a background of NRY haplotypes. The expectation is that lineages with ancient common ancestry will have high microsatellite variance, whereas those which have recent common ancestry will have less microsatellite variation. TMRCA estimates and expansion times for Y chromosome haplotypes were estimated by calculating the allelic diversity at 13 STR loci and then averaging across all loci (based on the method by Zhivotovsky *et al.*, [262]), with the average multiplied by an STR mutation rate (most studies use ~2.80 x 10^{-3} per 25 year generation [159, 263-266]).

c) Lineage Expansion Times based on complete mitochondrial genomes

Bayesian skyline plots (BSP) [267-270] were constructed using a coalescent Markov chain Monte Carlo (MCMC) method to infer population size (Ne) through time from gene sequence data given a specified nucleotide-substitution model. Therefore, BSPs are graphical displays of the demographic information in genealogies reconstructed from gene sequences. Under coalescent theory, the probability of two lineages coalescing during a generation is inversely proportional to population size at that time. Bayesian Skyline plots estimate effective population sizes during the intervals between coalescence events based on interval lengths [268], which is correlated to the number of substitutions. This method incorporates uncertainty in the substitution model parameters underlying genealogy and coalescent times. Unlike previous approaches to estimating population size, the BSP does not require a pre-specified parametric growth model, and, because the method is applied directly to a set of sequence data rather than to pairwise sequence differences, it avoids the loss of information associated with distance-based methods [271]. BSPs were generated for a large dataset of complete human *mt*DNA sequences to estimate the relative size and timing of human *mt*DNA lineage expansions in Africa. Plots of effective population sizes (*Ne*) against time were drawn using BEAST ver1.4.6 [270] for (*i*) representative Sub-Saharan sequences encompassing all of the major African *mt*DNA haplogroups (n=113) *(ii)* haplogroup L0 (n=144 sequences), (*iii*) haplogroup L1 (n=93) (*iv*) haplogroup L2 (n=160) (*v*) haplogroup L3 (n=223), (*vi*) haplogroup L4 (n=21) (*vii*) haplogroup L5 (n=49), and (*viii*) M1a (n=86). The analyses were run for 50 to 250 million iterations [200, 200, 200, 200, 250, 50, 50 and 200 million iterations for analysis (*i*) - (*viii*), respectively], using the following parameters and priors; GTR

substitution model without site heterogeneity and partitioned into codon positions, strict molecular clock model, tree prior; coalescent, Bayesian skyline, with 10 groups and constant skyline model, operators; auto-optimize, log parameters; every 10,000 iterations. In order to assign a time scale to the population size estimates, the rate of molecular evolution must be calibrated, so mutation rates were fixed to 5.5278×10^{-8} substitutions per site per year based on an average of the mutation rate for the control regions and the coding region as determined by Henn *et al.,* [272]. A second rate estimate used by Atkinson *et al.,* [273] of 6.19×10^{-8} was also applied. Recent studies [272, 274] have shown that previous estimates of mutation rate for *mt*DNA [178, 220, 260] were time dependent, with estimates for older lineages showing underestimation due to saturation and effects of purified selection. So a simple average of the mutation rate for the control regions and coding region estimates derived by Henn *et al.,* [272] estimates that are corrected for purified selection and time dependence was used. Results of the analyses were visualized using Tracer v1.4 [275]. Convergence of the chains was systematically confirmed by visual inspection of plotted posterior estimates. For each haplogroup/haplotype three different runs were done to check for replication.

2.3.5: Test of independence of mtDNA and Y chromosome lineage frequencies, language, and geography

Considering that the data from autosomal markers indicated that the genetic diversity in East Africa (the current study) and by extension, across African populations [4] is correlated with language and geography, it is hypothesized here that some Y chromosome/*mt*DNA lineages are associated with specific linguistic groupings or geographical regions. Fisher's exact test [276-278] was initially used to check for

independence of lineage frequency distribution from language and/or geography. The null hypothesis of no difference in the frequencies among the language families/geographic regions against the alternative, that frequency of the lineages are not equal in all the language groups or geographic regions was done. The analysis was done on pooled data from the current study (~1500) (**Appendix 1)** and previously published data from the rest of Africa - (~3300 for Y chromosome and ~9800 for *mt*DNA data) (**Appendix 6)**. For this analysis, populations were grouped based either on language family (Niger-Kordofanian, Nilo-Saharan, Afroasiatic, Khoisan) or geographic location (central/west, eastern, north/northwestern and southern Africa). In this case, the null hypothesis is as follows:

H_0: frequency among the Khoisan = frequency among the Nilo-Saharans = frequency among the Niger-Kordofanian = frequency among the Afroasiatic (or East Africa = Central/West Africa = southern Africa = northern Africa)

Rejection of the null model would indicate significant differences in frequency of the lineages among all the language families/geographic regions. However, it is not possible to infer which language family or geographic region has significantly higher frequency of the particular lineage. A multiple comparison approach, the Marascuillo procedure [279], made it possible to do pairwise tests for lineage frequency differences among the populations classified based on language family or geography. The Marascuillo procedure is done in three steps: (1) compute the difference in frequency among all population pairs (2) pick a significance level and compute the corresponding critical values for the procedure, and (3) use the absolute values of pairwise frequency differences to compare with the critical value. Those pairs that have a test statistic that

exceeds the critical value are significant at the α level. Both the Fisher's exact test and Marascuillo procedure were implemented in R [280] with codes kindly provided by Dr. Mingyao Li (Department of Biostatistics and Epidemiology at the Center of Clinical Epidemiology and Biostatistics, University of Pennsylvania).

A significant result from the Marascuillo test between a group and each of the three other groups implies that the lineage is significantly correlated with a language family and/or geographic region. An increased frequency of a specific lineage in a linguistic group or a geographic region compared with that in others implies that the lineage might have arisen in that population or region. However, this inference has a caveat in that genetic drift can cause lineages to increase to high frequency or become lost in a population or group of populations. In order to help distinguish among these possibilities the presence of ancestral and sister lineages in the population group in question was checked. This is because of the expectation that if a lineage arose in a particular group or region, one would expect to observe both the ancestral and sister lineage in that group or region. For example, the Y chromosome family of E3a haplotypes (E3a*-ancestral, E3a1-8 being sister haplotypes) **(Figure 2.3.5.1)** are predominantly observed among populations that speak Niger-Kordofanian languages, as expected this lineage arose among the Niger-Kordofanians. Moreover, for each of the *mt*DNA and Y chromosome lineages, the lineage diversity and microsatellite variances within a language family or geographic region, respectively, were also checked. The prediction is that if a lineage rose recently to high frequency due to drift, it would have low microsatellite variance.

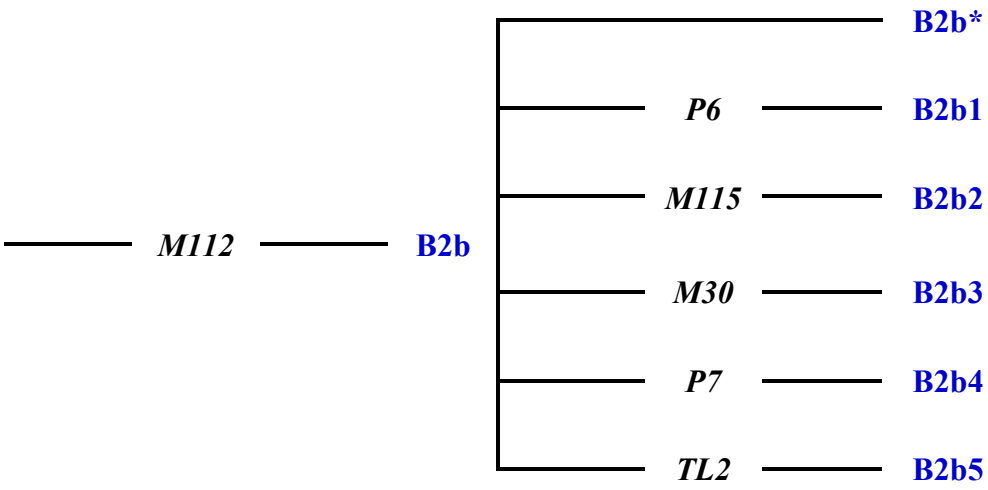


Figure 2.3.5.1: Splits in the Y chromosome B2b lineage are shown in order to clarify the terms "ancestral" and "sister" in the context of discussing lineages in this study. All the B2b lineages above have mutations M94, M60, M182 (not shown – see **Figure 3.3.2**) and M112 (shown in italics in the figure) beside a mutation that distinguishes each of them from B2b***** (which is the root B2b without additional mutations) which is termed here as the "ancestral" form. The mutations that distinguishes each of the lineages in the figure are (lineage in parenthesis) P6 (B2b1), M115 (B2b2), M30 (B2b3), P7 (B2b4) and TL2 (B2b5). B2b1-B2b5 lineages are referred here as "sister" lineages. Note that mitochondrial lineages may have more than one mutation that separate "sister" lineages. In the context of comparing *mt*DNA and Y chromosome haplogroups, "ancestral" will refer to those lineages that split off from others earlier in the lineage "tree" i.e. L0, L1 , L5 and L2 (*mt*DNA) and A and B (Y chromosome).

2.3.6: Multidimensional Scaling and Regression Analysis

Several methods were used to determine population clustering and structure. Multidimensional scaling (MDS) plots [281] were constructed for all the data based on pairwise population distance values obtained for this study using the software program STATISTICA ver8 (Statsoft) to identify linguistic and/or geographic clustering. The statistical significance of the correlation between genetic distances based on autosomal data, *mt*DNA and Y haplotype data and, geography and language were evaluated using

the Mantel test [282] implemented in XLSTAT (Addinsoft). Matrices of natural log values of geographical distances between pairs of population were generated and compared with corresponding pairwise population matrices. The geographic distances between population pairs were calculated using the following Great Circle Distance Formula using decimal degrees;

$$r*acos[sin(lat10*sin(lat2+cos(lat1)*cos(lat2)*cos(lon2-lon1)]$$

Where r is the radius of the earth in kilometers, r=6378.7

For the Mantel test for correlation between genetic and linguistic distance, language matrices were generated as follows: Pairs of populations belonging to the same language family were assigned a value 1.0, pairs belonging to the same subfamily in a language family were assigned a value of 2.0, and pairs of populations belonging to different language families were assigned a value of zero (0). These matrices was then checked to see if they are correlated with corresponding pairwise genetic distances for all the marker systems.

2.3.7: Population Structure

Population structure among East and Central African populations, and a few representative West African, southern African, and southwestern Asia populations, was assessed using autosomal microsatellite/indel Marshfield data, consisting of a total of 1327 autosomal markers (848 microsatellites and 479 insertion/deletion polymorphisms) (**Table 2.3.6**) [3, 4]. Analysis was implemented using the program STRUCTURE [283] which uses a clustering algorithm, which identifies subgroups that have distinctive allele frequencies by placing individuals into a specific number of *K* clusters using a Bayesian algorithm, where *K* is chosen in advance but can be varied across independent runs. The analysis uses linkage disequilibrium (non-random allelic associations) among unlinked loci to estimate admixture from parental population by assuming that the parental populations are in Hardy-Weinberg equilibrium. The estimate of *K* is dependent on the number of individuals that exist within populations, the number of loci sampled, and the amount of differentiation between populations. *K* provides a rough guide for determining how many discrete parental populations would be required to explain the variation in the data [284]. Exploratory runs were initially performed to determine the optimal number of clusters (K) using the approach described by Evanno *et al.,* [285], which utilizes the admixture model in STRUCTURE and the option of correlated allele frequencies between populations as implemented in Falush *et al.,* [286]. The number of populations (K) in the model was systematically varied from 1-14, with arbitrary 6 replicates performed for each K value. Markov chain Monte Carlo (MCMC) burn-in time and replication number of steps were set to 20,000 and 80,000, respectively, for each run. Then a parameter, $\Delta \boldsymbol{K}$ was inferred**.** $\Delta \boldsymbol{K}$ is the mean of the absolute values of the

'estimated Likelihood (*ln*) of data' for a *K* averaged over several runs (6 for this study), divided by the standard deviation of *ln* for that *K*, $\Delta K = m(|L''(K)|)/s[L(K)]$. The value of K which produces the highest "stable" ΔK [285] (value after which subsequent ΔK values do not change significantly), the lowest stable α value [287] (value after which subsequent α values do not change significantly) was chosen as the "correct" K value (**Figure 2.3.6**). I then did a final run for K=2-11 using an MCMC burn-in time of 100,000 and a replication step of 500,000. Considering that for the correlated model the authors' [287] advise that divergent populations should be removed from the analysis to achieve better inference, separate STRUCTURE runs were done for the following groups of populations; (1) Populations speaking the major language families present in East Africa (Niger-Kordofanian, Nilo-Saharan, and Afroasiatic) and (2) Hunter-gatherer populations. In each case geographically neighboring populations from different language families were included (population classification as in **Appendix 1**). I then visualized the individual estimated membership coefficients using the program DISTRUCT [288].

Populations	Linguistic group
Biaka	Niger-Kordofanian
Mbuti	Nilo-Saharan
Druze	Afroasiatic
Bedouin	Afroasiatic
Palestinian	Afroasiatic
Gogo	Niger-Kordofanian
Mandenka	Niger-Kordofanian
Tutsi/Hutu	Niger-Kordofanian
Venda	Niger-Kordofanian
!Xun/Kxoe	Khoisan
Italian	Indo-European
Parsi	Indo-European

Table 2.3.6: Published population data [3] used for STRUCTURE analysis and their language classification.

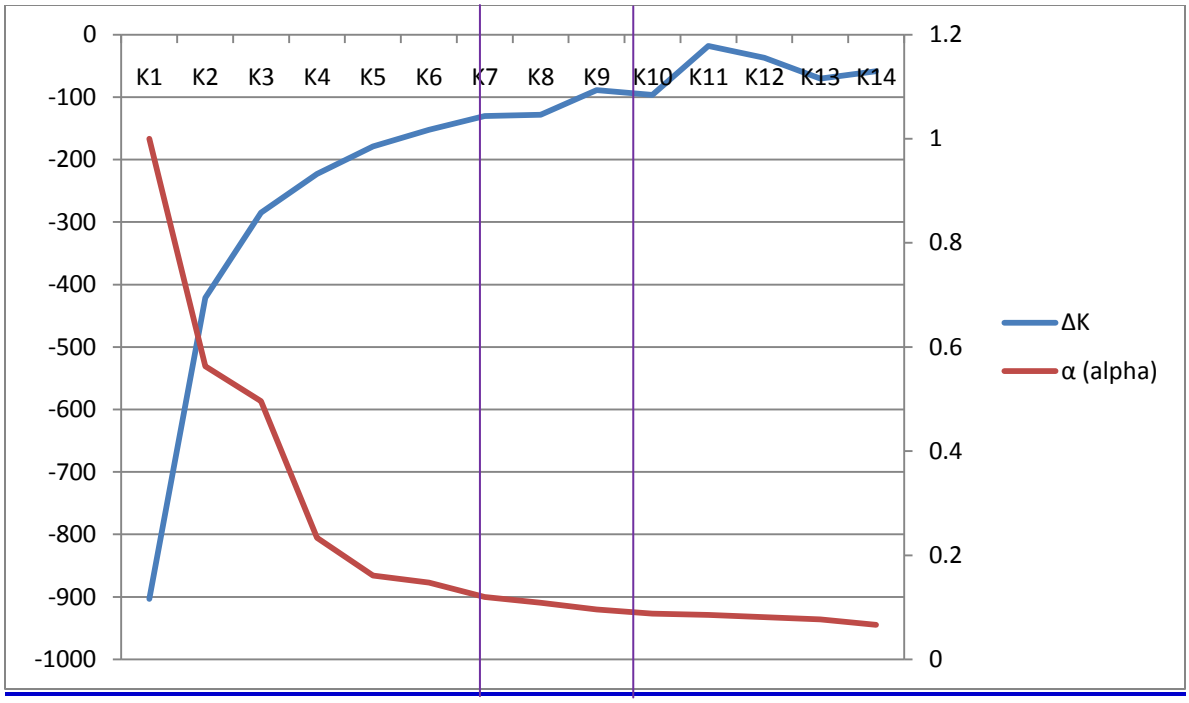


Fig. 2.3.6: Graphical presentation of determination of correct K value. In each case average Δ***K*** and **α** (Y axes) values are plotted against K values (X axes). Δ***K*** is the mean of the absolute values of 'estimated Likelihood (*ln*) of data' for a *K* averaged over several runs (6 for this study) divided by the standard deviation of *ln* for that *K* (generated as output from Structure run), and **α** is (also generated as output during each structure run). The two vertical purple lines delineate the range at which there is no significant change in values of Δ***K*** (blue graph) and **α** (red graph) with change in K values (before and after which the values are continuously rising or are unstable respectively).

Chapter 3: Results

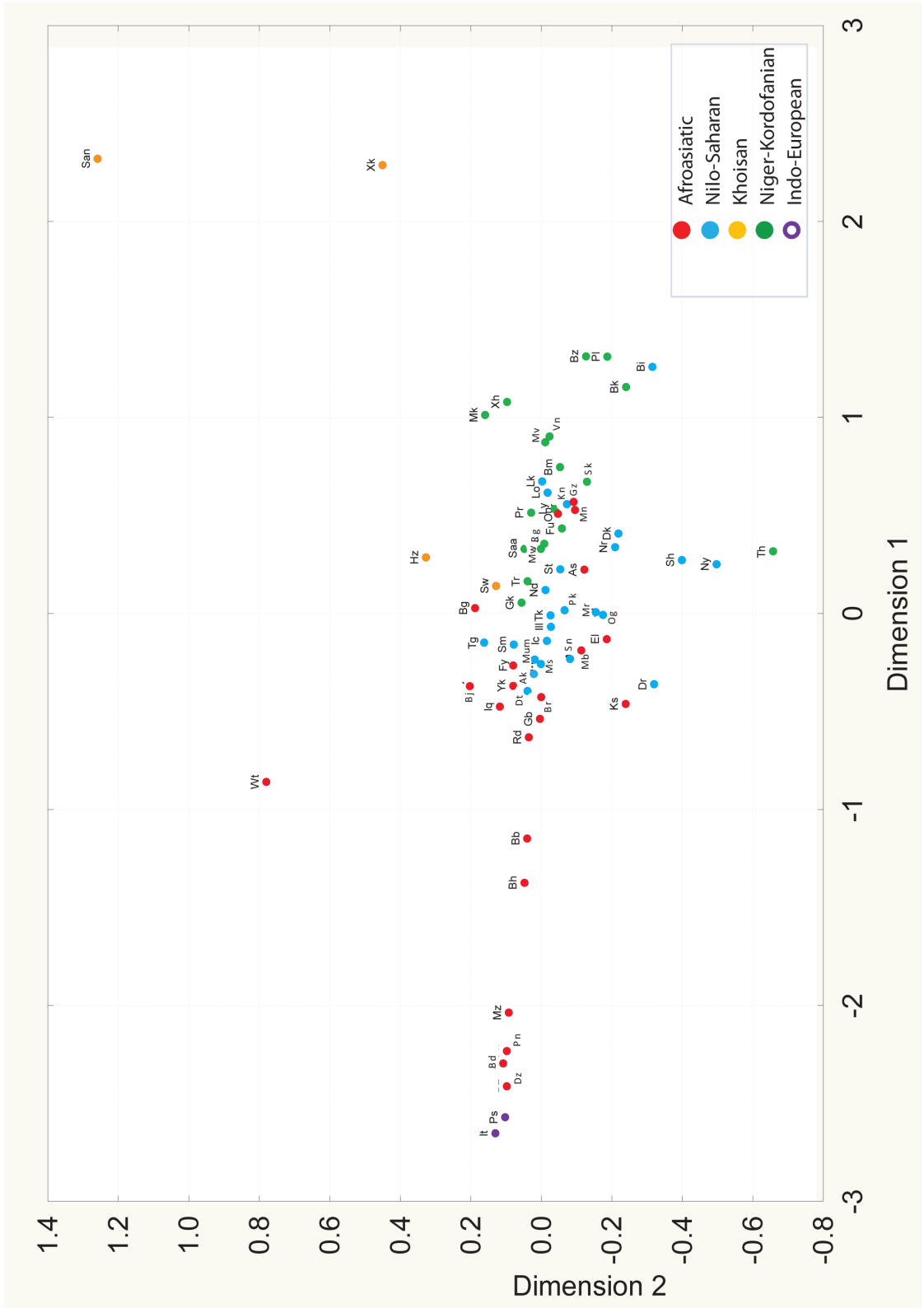
In this chapter, general comparative trends in population relationships based on genetic distance measures using different markers systems will be described: Multidimensional scaling plots, regression analysis and analysis of molecular variance. Subsequent sections in the chapter will address results for each of the marker systems used in the study: autosomal data, Y chromosome data and then mitochondrial DNA data.

3.1: General comparative trends in population relationships

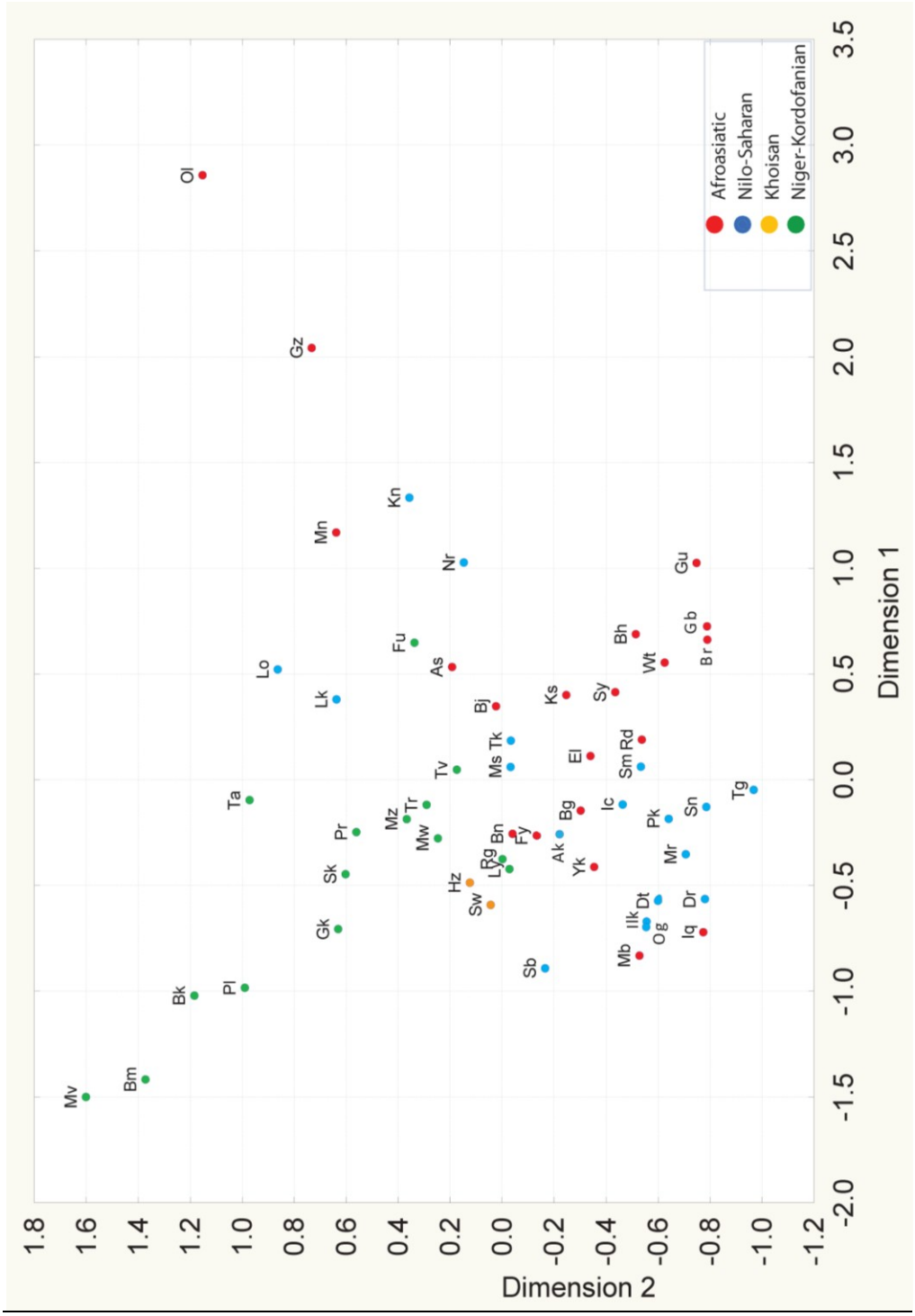
3.1.1: *Analyses of population clustering using multidimensional scaling plots and correlations between genetics, geography and language*

In order to characterize how populations cluster based on genetic variation, multidimensional scaling (MDS) based on pairwise population genetic distances for autosomal, Y chromosome, and *mt*DNA data was performed. Results are shown in Figures 3.1.1a-c. Generally, most populations cluster together based on linguistic classification for autosomal and Y chromosome data **(Figure 3.1.1.1a-b, Appendix 11)**. A few populations fall outside their linguistic clusters for these two systems. These outliers include southern Cushitic speakers, the Central African Chadic speakers and East African Cushitic speaking hunter-gatherer populations **(Figure 3.1.1.1a-b, Appendix 11)**. For *mt*DNA D-loop sequence data, all the populations cluster together except the hunter-gatherer populations of East Africa who are outliers **(Figure 3.1.1.1c)**. But when outliers are removed from the analysis **(Figure 3.1.1.1d, Appendix 11)**, only Kalenjin speaking Nilo-Saharans seem to cluster together, and other populations appear to cluster based primarily on geographic origin. **(Appendix 11)**.

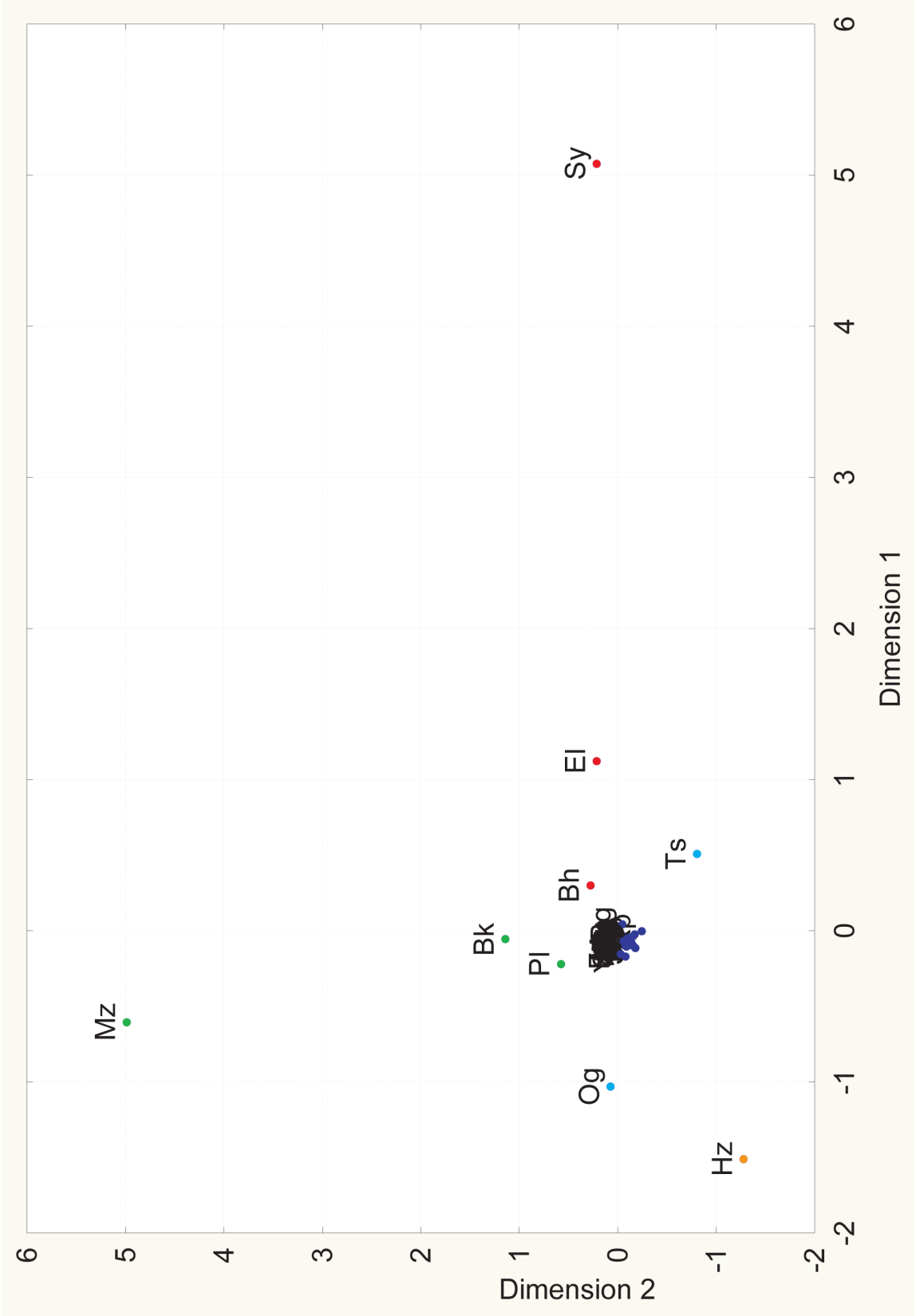
Figure 3.1.1.1a-c: MDS plot derived from population pairwise distances for (A) autosomal data (B) Y – chromosomal data and (C) *mt*DNA d-loop data. Population codes are as follows; **Khoisan:** Hz (Hadza), Sw (Sandawe), Xk (!Xun/Kxoe), San (San) **Afroasiatic:** Iq **(**Iraqw), Bg (Burunge), Mb (Mbugu), Fy (Fyome), Ol (Ouldeme), Gz (Giziga), Mn (Mandara), Bg (Baggara), Dz (Druze), Bd (Bedouin), Pn (Palestinian), Mz (Mozabite), Bb (Banuamir), Yk (Yaaku), Bh (Handadawa in autosomal plot and Beja in Y chromosome and *mt*DNA plots), Bj (Burji), Wt (Wata), El (El-Molo), Gb (Gabra), Rd (Rendille), Br (Borana), Ks (Konso), **Niger-Kordofanian:** Tr (Turu), Mw (Mbugwe), Rg (Rangi), Pr (Pare), Bk (Baka), Pl (Bakola), Bz (Bedzan), Mv (Mvae), Fu (Fulani), Bm (Bamoun), Mk (Mandenka), Sk (Sukuma), Saa (Samba'a), Ly (Luhya), Gk (Gikuyu), Th (Tutsi/Hutu), Vn (Venda), Xh (Xhosa) **Nilo-Saharan:** Ak (Akie), Ms (Maasai), Kn (Kanuri), Bi (Biaka), Dk (Dinka), Dt (Datog), Mum (Maasai Mumonyot), Ill (Maasai Il'ngwesi), Sm (Samburu), Dr (Dorobo), Tg (Tugen), Mr (Marakwet), Lo (Luo), Sn (Sengwer), Og (Ogiek), Nd (Nandi), St (Sabaot), Tk (Turkana), Pk (Pokot), Ic (Ilchamus), Lk (Laka), Nr (Nuer), Sh (Shilook), Ny (Nyimang) **Indo-European:** It **(**Italian), Ps (Parsi). **Figure 3.1.1.1d)** MDS plot derived from population pairwise F_{ST} distance for *mt*DNA d-loop data without hunter-gatherer outliers. Kalenjin populations include Tugen (Tg), Keiyo (Ky), Sengwer (Sn), Marakwet (Mr), Nandi (Nd), Kipsigis (Kp), Pokot (Pk) and Sabaot (Sb).



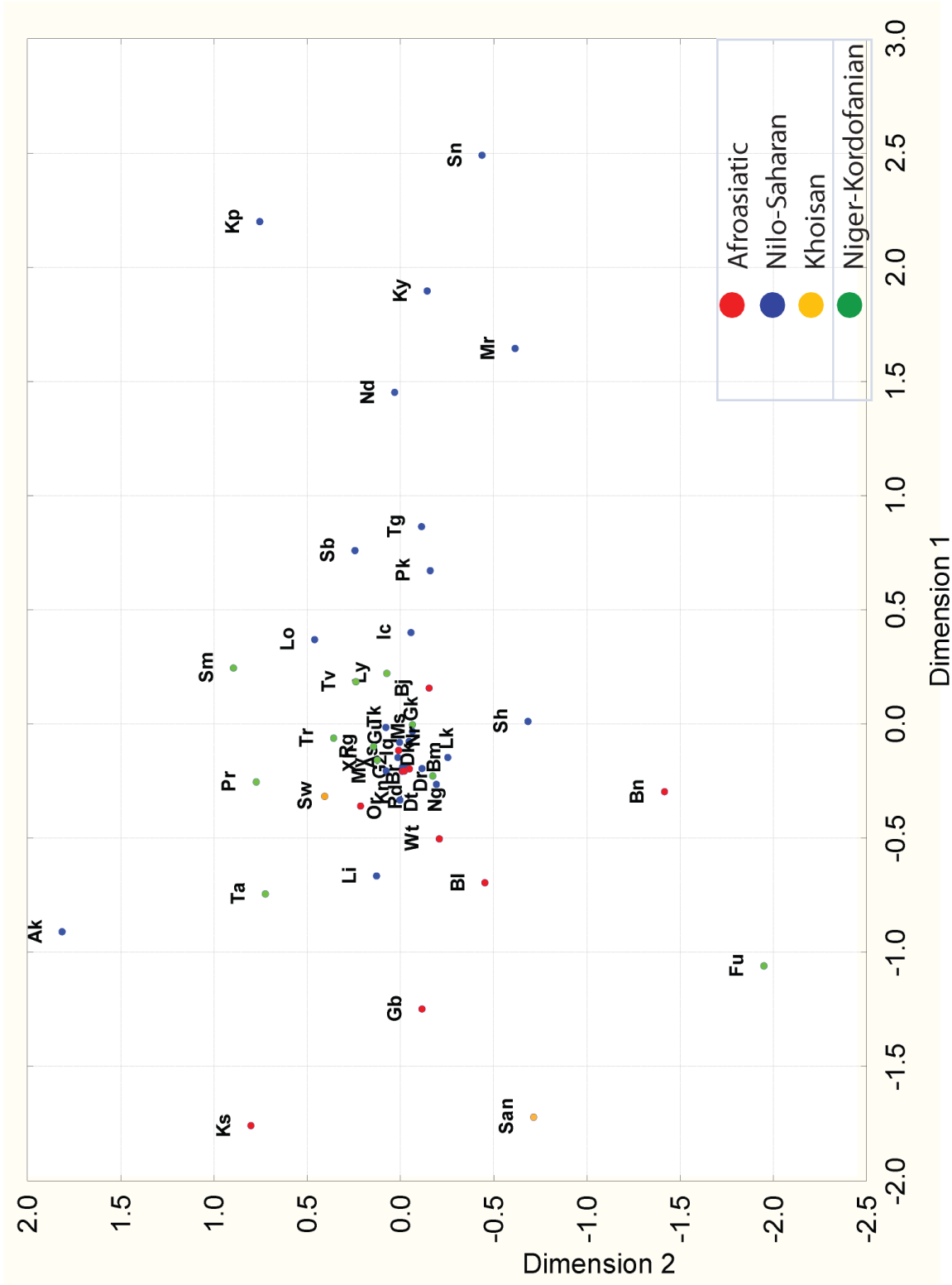
a) Autosomal MDS plot



b) Ychromosome MDS plot



c) *mt*DNA MDS plot



d) *mt*DNA d-loop data without hunter-gatherer outliers

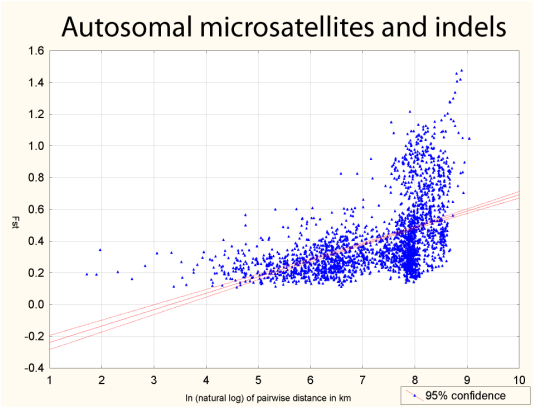
3.1.2: Regression Analysis, AMOVA and Mantel Tests

In order to further test for associations between genetic distance and geography, I did regression analysis of pairwise genetic distances plotted against geographic distances. Moreover, association between genetic distance and language was determined using Mantel test. The strongest correlation between genetic and geographic distance was for autosomal data where approximately ~26% of the variation was due to geography (**Figure 3.1.2.1a**). For Y chromosome data, the correlation with geography was about ~15% (**Figure 3.1.2.1b**) while for *mt*DNA, the correlation was less than 1% (**Figure 3.1.2.1c**). Generally for *mt*DNA data there is very low level of differentiation among the populations studied. Indeed a few estimates yielded negative genetic distance values (mostly for populations that speak related languages). The negative distance values mean that the true genetic distance values for those pairs of populations are extremely small and are probably not significantly different from zero (and also likely reflect the imprecision of the algorithm used by the software to estimate this value). Moreover, this means that there is more differences between two random individuals from the same population, rather than between two random individuals from different populations from the same language group [289]. Mantel tests for pairwise genetic distance versus geographic and linguistic distances indicated statistically significant correlations for both geography and language in all the marker systems (**Figure 3.1.2.1**). Consistent with the regression analysis, the results from Mantel tests and partial Mantel tests indicate that variation in autosomal and Y-linked data can be explained by linguistic patterns, independent of geography, while for mitochondrial data this is much less true.

Consistent with both regression analysis and the results from Mantel test, the results of Analysis of Molecular Variance (AMOVA) **(Table 3.1.2.1)** indicate that both language and geography are important predictors of population genetic variation in East Africa. The proportion of within and between group variance is larger when populations are grouped according to geography rather than languages for all the marker systems **(Table 3.1.2.1)**. However, the percentages of variance explained by autosomal and Y chromosome data were much higher (>20 times) compared to the *mt*DNA lineages **(Table 3.1.2.1)**.

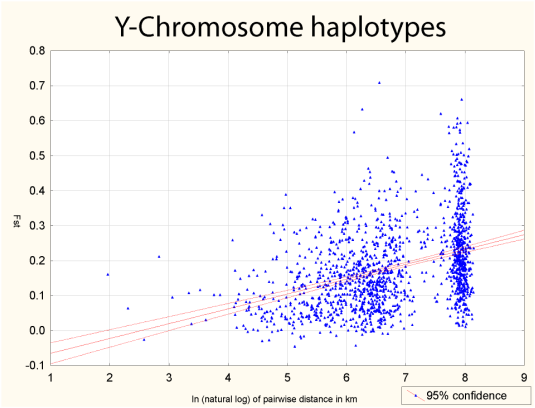
Figure 3.1.2.1 (next page): Regression of pairwise genetic distance vs. natural log (ln) of geographical distances between respective pair of populations for the three marker systems. For each system the significance level for the Mantel test is indicated; a) autosomal b) Y chromosome and c) *mt*DNA D-loop sequence data

A



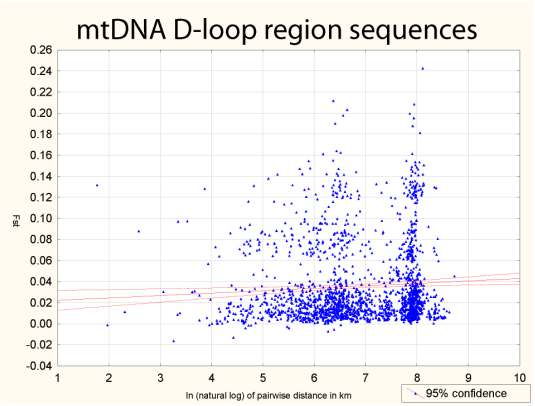
R^2=0.2553.
Mantel test geography p=0.001.
Partial Mantel Test Language p=0.0001.

B



R^2=0.1498.
Mantel test geography p=0.001.
Partial Mantel Test Language p=0.0001.

C



R^2=0.0044.
Mantel test geography p=0.004.
Partial Mantel Test Language p=0.032.

			Within populations		Among populations within groups		Among groups	
	Group	# of groups	Variance		Variance		Variance	
			%	ΦVc	%	ΦVb	%	ΦVa
Y-chrom	Linguistic	4	81.7	0.3884	15.16	0.07205	3.14	0.01495
	Geographic cluster	2	79.56	0.3884	15.03	0.07339	5.41	0.02641
D-Loop	Linguistic	4	96.61	0.48299	3.2	0.01599	0.19	0.00096
	Geographic cluster	3	96.57	0.48323	3.21	0.01604	0.23	0.00114
Autosomal	Linguistic	5	85	362.415	12	53.49	3	12.571
	Geographic cluster	4	83	272.32	9	31.08	8	26.52

Table 3.1.2.1: Analysis of Molecular Variance (AMOVA) for the three marker systems; autosomal data, *mt*DNA D-loop data and Y chromosome data (Y-chrom). Group refers to populations grouped based on language families or geographical regions. The number of groups in each of the two categories, language and geography, used for analysis in the three marker systems differed. For Y chromosome data the four African language families; Khoisan, Niger-Kordofanian, Nilo-Saharan and Afroasiatic and two geographical regions; East and Central Africa are represented. While for *mt*DNA D-loop data the four African language families and an additional geographical region, South Africa, besides the two mentioned for Y chromosome data were used. For autosomal data additional language family and geographic region, Indo-European and North Africa/Mediterranean/southwest Asia, respectively, besides those mentioned for *mt*DNA data were used.

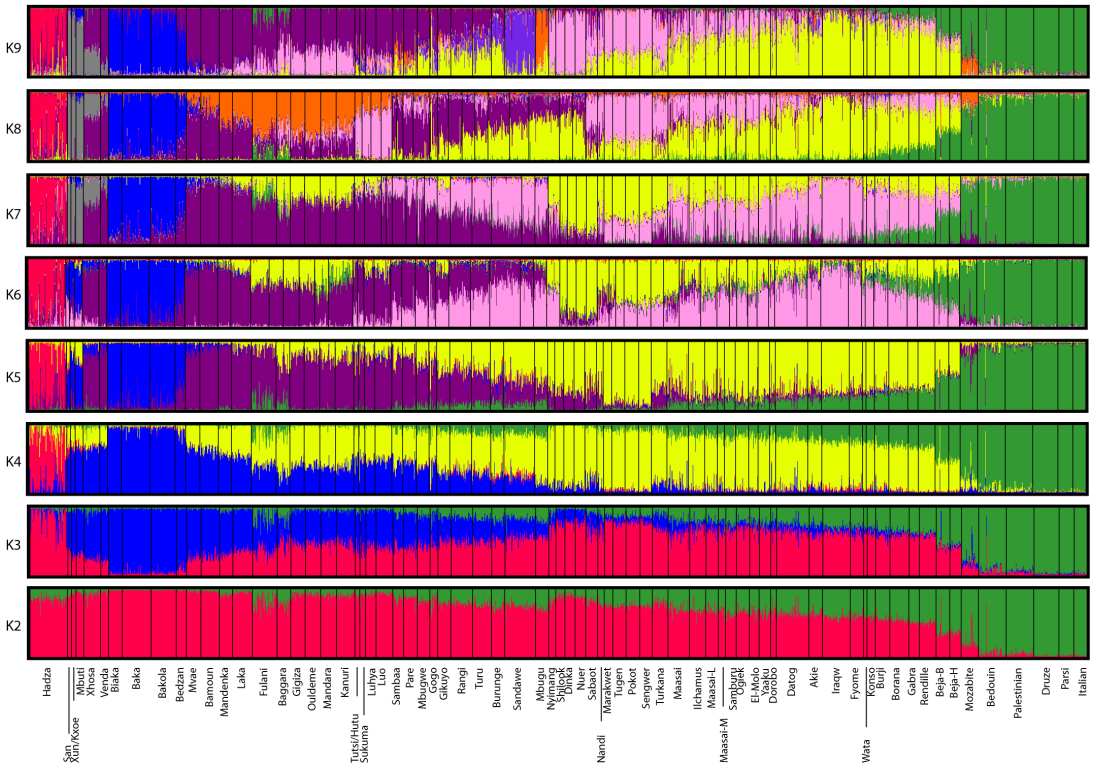
3.2: Population Structure as inferred from Autosomal Data

Population structure was inferred using the program STRUCTURE [283], and the results were visualized using the program DISTRUCT [288]. Based on a combination of Evanno *et al.'s* [285] and Falush *et al.'s* [286] approaches, the data analyzed in this study yielded nine population clusters **(Figure 3.2.1a-b)**. At K=2 individuals are divided into African and non-African clusters with populations from North Africa and East Africa showing ancestry from the African and non-African clusters. At K=3 the African component splits into East and Central/Western African clusters. At K=4, the Hadza, a Khoisan speaking population from northern Tanzania, separates from the East African cluster. At K=5 the pygmy populations from Central Africa and South African Khoisan populations separate from the Central/West African population cluster. Each value of K beyond 5 introduces a new cluster that tends to be associated with populations from a language family, and in some cases geographic proximity, or, especially at higher K values, by membership in a small population isolate (e.g. Ogiek at K=12 in the STRUCTURE analysis of hunter-gatherers, and at K=9 in the STRUCTURE analysis of Nilotic populations **Fig., 3.2.1b**). At K=6, Nilo-Saharan ancestry is distinguished, with the highest ancestry observed in the western Nilotic speaking populations of Sudan: Shilluk, Nuer, and Dinka **(Figure 3.2.1a)**. The proportion of Nilo-Saharan cluster (shown in red) decreases in northern Kenyan and northern Tanzanian Nilotic populations **(Figure 3.2.1a, 3.2.1b (iv))**. At K=7 individuals from other African language groups, Khoisan (blue) **(Figure 3.2.1a)**, Niger-Kordofanian (red) and Cushitic (yellow) **(Figure 3.2.1a)** cluster based on language classification. At K=8 the Central African Chadic and Nilo-

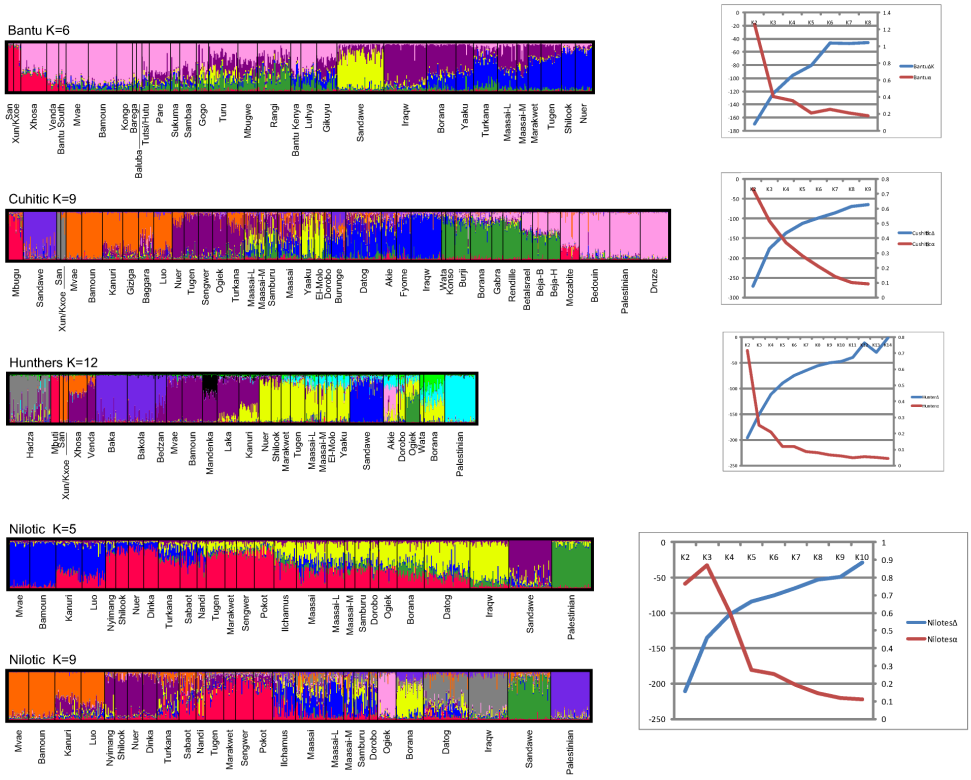
Saharan populations become distinct. At K=9 Sandawe individuals, the other Khoisan speaking populations from Tanzania are distinguished.

Figure 3.2.1a: Population structure as inferred using the program STRUCTURE [283]. Colors represent inferred ancestral populations. Each individual is represented by a vertical bar, partitioned into colored segments with the length of each segment representing the proportion of the individual's genome from K = 2 - 9 ancestral populations. **Figure 3.2.1b *i-v*:** Population structure inferred for each linguistic and subsistence grouping, and corresponding graph used to infer number of ancestral clusters for each group. **Khoisan:** Hadza, Sandawe, !Xun/Kxoe, San **Afroasiatic:** Iraqw, Burunge, Mbugu, Fyome, Ouldeme, Giziga, Mandara, Baggara, Druze, Bedouin, Palestinian, Mozabite, Beja-B (Banuamir), Yaaku, Beja-H (Handadawa), Burji, Wata, El-Molo, Gabra, Rendille, Borana, Konso, **Niger-Kordofanian:** Turu, Mbugwe, Rangi, Pare, Baka, Bakola, Bedzan, Mvae, Fulani-CA, Bamoun, Mandenka, Sukuma, Samba'a, Luhya, Gikuyu, Tutsi/Hutu, Venda, Xhosa **Nilo-Saharan:** Akie, Maasai, Kanuri, Biaka, Dinka, Datog, Maasai-m (Mumonyot), Maasai-L (Il'ngwesi), Samburu, Dorobo, Tugen, Marakwet Luo, Sengwer, Ogiek, Nandi, Sabaot, Turkana, Pokot, Ilchamus, Laka, Nuer, Shilluk, Nyimang **Indo-European:** Italian, Parsi. Populations are separated by black lines.

A


K9
K8
K7
K6
K5
K4
K3
K2
Hadza
San
Xun/Kxoe
Mbuti
Xhosa
Venda
Biaka
Baka
Bakola
Bedzan
Mvae
Bamoun
Mandenka
Laka
Fulani
Baggara
Gigiza
Ouldeme
Mandara
Kanuri
Tutsi/Hutu
Sukuma
Luhya
Luo
Samba
Pare
Mbugwe
Gogo
Gikuyo
Rangi
Turu
Burunge
Sandawe
Mbugu
Nyimang
Shilook
Dinka
Nuer
Sabaot
Nandi
Marakwet
Tugen
Pokot
Sengwer
Turkana
Maasai
Ilchamus
Maasai-L
Samburu
Maasai-M
Ogiek
El-Molo
Yaaku
Dorobo
Datog
Akie
Iraqw
Fyome
Konso
Wata
Burji
Borana
Gabra
Rendille
Beja-B
Beja-H
Mozabite
Bedouin
Palestinian
Druze
Parsi
Italian

B


Bantu K=6
Cuhitic K=9
Hunthers K=12
Nilotic K=5
Nilotic K=9
K2
K3
K4
K5
K6
K7
K8
K9
K10
NilotesΔ
Nilotesα

3.3: Y chromosome haplotypes

A total of 1500 male individuals from 55 Tanzanian, Kenyan, Sudanese, Cameroonian and Chadian populations **(Appendix 1)** were genotyped for the following Y chromosome markers; 55 SNPs and an *Alu* insertion polymorphism in a hierarchical manner, using 8 markers to define the base haplogroups followed by other appropriate markers to define each haplotype. A total of 40 haplotypes that belong to Haplogroups A, B, E, J and R were observed at varying frequencies (**Figure 3.3.1).** Collated frequencies are shown for populations genotyped in this study in **Figure 3.3.2**. Median-joining network of Y-STR on different haplotype backgrounds for individuals in East African populations are below in this section, **Appendix 9** and **Appendix 18**. Comparative data from published sources for other African and Middle Eastern populations are given in **Appendix 6a** and were used to make inferences about origin and migrations of East African populations. Age estimates for haplogroups are shown in **Table 3.3.1**.

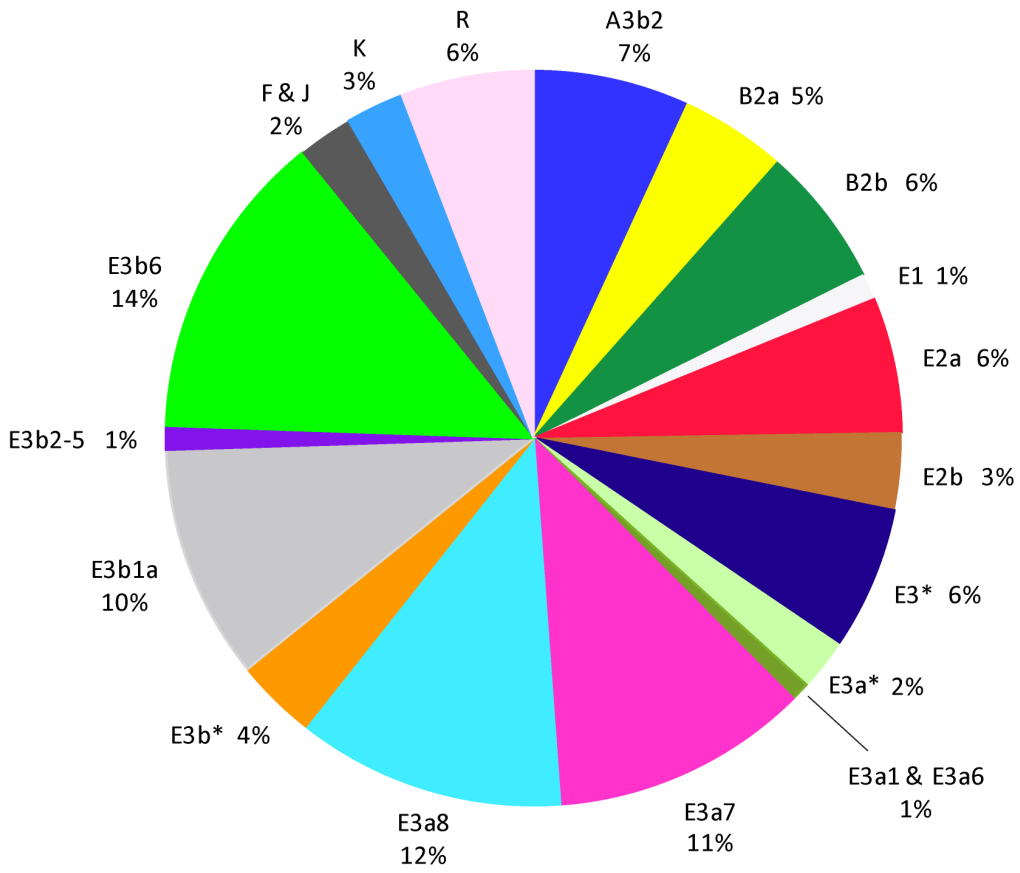


Figure 3.3.1: Relative frequencies for Y chromosome haplogroups/haplotypes observed in the current study populations. E3 (E3*, E3a & E3b) haplotypes **(Figure 3.3.2)** are the most common haplotypes in East Africa, constituting about two third of all the male lineages observed in the region **(details in Figure 3.3.2, Appendix 6, Appendix 9).**

Figure 3.3.2 (next page): Maximum parsimony tree of Y-chromosome haplogroups defined by binary markers and an Alu Insertion with relative frequencies of haplotypes from East and Central African populations based on genotyping and sequencing of samples for the current study. The scheme of Y haplogroup classification is based on a scheme proposed by the Y consortium. Specific haplotypes are defined by compound SNPs that consist of all SNPs that link the haplotype to the base of the "tree" (for example E3b6 by SNP M293 and all the SNPs, M35, P2, P29, Alu insertion YAP, P9 and M94 that link the haplotype to the base). Three new SNPs were identified in this study; TL1 that defines haplotype B2a3 (following the Y consortium scheme) that has near exclusive distribution among the Yaaku population from Kenya, TL2 that defines B2b5 that is mostly observed in low frequencies in East African populations, and TL3 that further subdivides the E3a7 haplotype.

* Ilaikipiak = Ilngwesi & Mumonyot

Haplotypes	n	AGE in KY	SE
E3b	581	31.80	6.50
E3b1	244	22.25	4.57
E3b1a1b	75	14.82	4.16
E3b1a3	33	14.25	2.91
E3b3	63	17.21	3.68
E3b5	4	12.94	4.29
E3b6	170	21.09	6.96
E3a	1231	21.96	4.57
E3a7	149	16.00	3.23
E3a7a	141	13.55	2.79
E3a8a	54	13.58	2.41
E3a8	154	16.84	4.08
E3a & E3b	1753	38.06	8.04
E3	1831	38.05	8.06
B2	252	60.75	14.50
B2b	116	71.15	16.29
B2b5	13	62.97	17.28
B2a	137	29.65	6.31
B2a1	40	27.89	5.62
B2a3	14	10.73	3.47
A3b2	123	23.54	4.88
A	159	51.81	12.91
E2a	90	23.10	5.95
E2b	55	21.23	5.82
E2	195	39.37	9.90
J*	17	25.34	7.78
K2	102	23.99	5.01
R1b	115	18.32	4.59

Table 3.3.1: TMRCA estimates for Y chromosome lineages identified in East African populations. The age estimates were computed from Y chromosome microsatellite variation on each UEP haplotype background using the method described by Zhivotovsky *et al.,* [262]. KY = thousand years. SE = standard error. Age estimates for haplogroups A and B yield oldest estimates while E and F haplogroups yield relatively younger estimates.

Table 3.3.2: Y chromosome lineage frequency heterogeneity among language families in Africa. The table shows results for the Marascuillo procedure to check if there were significant differences in the proportions of the Y chromosome lineages among the four African language families, Ks – Khoisan, AA – Afroasiatic, NS – Nilo-Saharan and NK – Niger-Kordofanian. The Marascuillo procedure does pairwise comparison of lineage frequencies between groups and compares the difference of the two frequencies with a critical value determined by a pre-specified significance threshold. If the difference is bigger than the critical value, then it is an indication that the difference between the two proportions is significantly different at the pre-specified significance level. Note that the test is adjusted for multiple comparison, in this case significance level = 0.05/47 (there are 47 tests because there are 47 rows representing lineages). The number in each cell is calculated as: difference between the two proportions - corresponding critical value. Note that only comparisons that yielded significant values are shown – meaning those tests with positive values (blank cells indicate negative or 0 values). Each of the four African language families have significantly higher frequencies of the following respective lineages **(in bold)** relative to the other three language families; Khoisan – A2, A3b1; Nilo-Saharan – A3b2, B2a, E2a; Niger-Kordofanian – E1, E3a*, E3a7; Afroasiatic – E3b*, E3b1*, E3b1a*, E3b1a1, E3b2, E3b3, Eb6, J and K2. n = number of individuals, N^a = number of populations represented from each group, [b]other lineages refers to F lineages that are observed outside Africa.

Language	Ks	AA	NS	NK	Fisher's exact test	Marascuillo test					
N[a]	**7**	**33**	**25**	**38**							
n	**281**	**1655**	**787**	**2087**	***p-value***	**Ks vs AA**	**Ks vs NS**	**Ks vs NK**	**AA vs NS**	**AA vs NK**	**NS vs NK**
A*	0	0	0	3	0.4125						
A1	0	2	0	18	0.00061						
A2	**28**	**0**	**0**	**4**	**4.55E-32**	**0.0231**	**0.0231**	**0.0211**			
A3a	0	0	0	0	1						
A3b1	**37**	**0**	**0**	**7**	**2.13E-41**	**0.0453**	**0.0453**	**0.0418**			
A3b2	**2**	**77**	**159**	**14**	**5.17E-80**	**0.0085**	**0.13**		**0.0903**	**0.0164**	**0.1336**
B*	0	3	0	6	0.4929						
B1	0	0	0	2	0.7014						
B2*	2	0	3	5	0.01525						
B2a*	**0**	**6**	**26**	**6**	**1.58E-11**		**0.0058**		**0.0014**		**0.0024**
B2a1	2	20	45	46	1.58E-09		0.0086		0.0078		
B2a3	0	13	8	0	7.88E-06						
B2b*	24	15	31	28	3.18E-14	0.0043					
B2b1	8	0	0	0	1.23E-10						
B2b2	0	0	2	0	0.04926						
B2b4	4	0	10	41	4.93E-10					0.0066	
B2b5	3	4	4	2	0.01607						
E*	2	2	0	10	0.03058						
E1	**0**	**11**	**0**	**65**	**9.09E-14**			**0.0149**		**0.0061**	**0.0149**
E2*	2	0	0	2	0.0145						
E2a	**0**	**8**	**87**	**7**	**5.07E-53**		**0.0627**		**0.0573**		**0.059**
E2b	1	14	19	86	6.07E-11			0.0136		0.0118	
E3*	6	73	27	27	4.66E-08					0.0071	
E3a*	**43**	**32**	**23**	**822**	**5.24E-245**	**0.0406**	**0.0283**	**0.1381**		**0.3265**	**0.3121**
E3a1	2	0	2	27	4.76E-07					0.0023	
E3a4	1	0	0	7	0.03047						
E3a6	0	0	2	1	0.1797						
E3a7	**29**	**14**	**49**	**423**	**4.19E-98**	**0.0164**		**0.0131**	**0.0157**	**0.1553**	**0.0877**
E3a8*	13	24	15	61	0.001314						
E3a8a*	6	10	15	26	0.009292						
E3b*	**3**	**95**	**24**	**81**	**0.000126**	**0.0108**					
E3b1*	**0**	**293**	**62**	**7**	**3.36E-106**	**0.1369**	**0.0377**		**0.0408**	**0.1332**	**0.034**
E3b1a*	**0**	**28**	**0**	**0**	**7.04E-12**	**0.0033**			**0.0033**	**0.0033**	
E3b1a1	**0**	**109**	**12**	**2**	**3.87E-39**	**0.0398**			**0.0185**	**0.0386**	
E3b1a2	0	0	0	0	1						
E3b1a3	0	12	14	0	2.52E-08						
E3b2	**0**	**182**	**0**	**11**	**6.46E-71**	**0.077**			**0.077**	**0.0711**	
E3b3	**0**	**34**	**1**	**0**	**9.55E-14**	**0.0056**			**0.0034**	**0.0056**	
E3b4	0	3	1	0	0.1722						
E3b5	0	3	2	0	0.103						
E3b6	**41**	**79**	**82**	**16**	**6.05E-42**	**0.0053**		**0.0477**	**0.0047**	**0.0162**	**0.0492**
F*	0	24	6	5	0.000111	0.0019					
J	**2**	**242**	**24**	**41**	**5.71E-59**	**0.0962**			**0.0702**	**0.0872**	
K2	**1**	**71**	**6**	**7**	**2.15E-19**	**0.0131**			**0.0102**	**0.0175**	
R1b*	0	88	24	63	5.11E-07	0.0296	0.0043	0.0142			
R1b3	2	15	0	16	0.02211						
others[b]	16	47	3	93	5.39E-10				0.0047		0.0192

Table 3.3.3: Y chromosome lineage frequency heterogeneity among African geographic regions. The table shows results for the Marascuillo procedure to test if there were significant differences in the proportions of the Y chromosome lineages among the four African regions, EA – East Africa, SA – South Africa, CWA – Central/West Africa and NWA – North West Africa. The Marascuillo procedure does pairwise comparison of lineage frequencies between groups and compares the difference of the two frequencies with a critical value determined by a pre-specified significance threshold. If the difference is bigger than the critical value, then it is an indication that the difference between the two proportions is significantly different at the pre-specified significance level. Note that the test is adjusted for multiple comparison, in this case significance level = 0.05/47 (there are 47 tests because there are 47 rows representing lineages). The number in each cell is calculated as: difference between the two proportions - corresponding critical value. Note that only comparisons that yielded significant values are shown – meaning those tests with positive values (blank cells indicate negative or 0 values). Each of the four African regions have significantly higher frequencies of the following respective lineages **(in bold)** relative to the other three geographical regions; South Africa – A2, A3b1; East Africa – A3b2, B2a3, E2a, E3*, [x]E3a8*, [x]E3a8a*, E3b1a1; Central-West Africa – E1, E3a7; North/Northwestern Africa – E3b1a*, E3b2 and J. [x]probably so because previous studies of central-West African populations did not type mutations that define these two lineages, n = number of individuals, N[a] = number of populations represented from each group, [b]other lineages are those F lineages that are observed outside Africa

Region	EA	SA	CWA	NWA	Fisher's exact test	Marascuillo test					
N[a]	**60**	**11**	**29**	**6**							
n	**2236**	**453**	**1701**	**635**	***p-value***	**EA vs SA**	**EA vs CWA**	**EA vs NWA**	**SA vs CWA**	**SA vs NWA**	**CWA vs NWA**
A*	0	0	3	0	0.190015						
A1	0	0	18	2	5.67E-07						
A2	**0**	**28**	**4**	**0**	**5.94E-27**	**0.0134**			**0.0108**	**0.0134**	
A3a	1	0	0	0	1						
A3b1	**2**	**44**	**0**	**0**	**8.12E-45**	**0.0366**			**0.0376**	**0.0376**	
A3b2	**251**	**0**	**33**	**7**	**1.13E-52**	**0.0837**	**0.0609**	**0.0676**	**0.0051**		
B*	3	0	6	0	0.280119						
B1	0	0	2	0	0.500732						
B2*	1	2	8	0	0.010503						
B2a*	27	1	10	0	0.002299			0.0022			
B2a1	62	24	31	7	7.49E-05						
B2a3	**21**	**0**	**0**	**0**	**3.18E-06**	**0.0007**	**0.0007**	**0.0007**			
B2b*	**72**	**6**	**20**	**0**	**1.27E-09**		**0.0009**	**0.0162**			**0.0006**
B2b1	0	8	0	0	4.12E-09						
B2b2	2	0	0	0	0.698685						
B2b4	0	4	51	0	3.93E-21		0.0123				0.0123
B2b5	13	0	0	0	0.001054						
E*	1	7	5	1	5.81E-05						
E1	**0**	**0**	**72**	**4**	**8.05E-30**		**0.0214**		**0.0214**		**0.0112**
E2*	0	3	1	0	0.003175						
E2a	**118**	**0**	**0**	**0**	**8.04E-40**	**0.0325**	**0.0325**	**0.0325**			
E2b	45	37	48	0	3.61E-15	0.005		0.0074		0.0266	0.011
E3*	**134**	**0**	**19**	**0**	**5.11E-30**	**0.0384**	**0.0247**	**0.0384**	**0.0003**		**0.0003**
E3a*	**84**	**156**	**668**	**16**	**3.14E-231**	**0.2097**	**0.3016**			**0.22**	**0.3103**
E3a1	14	0	17	0	0.006842						
E3a4	0	1	7	0	0.006603						
E3a6	0	0	3	0	0.190015						
E3a7	**200**	**87**	**225**	**2**	**2.32E-36**	**0.0193**		**0.0588**		**0.1091**	**0.0927**
E3a8*	**93**	**0**	**20**	**0**	**7.43E-18**	**0.0235**	**0.0086**	**0.0235**	**0.0006**		**0.0006**
E3a8a*	**48**	**0**	**9**	**0**	**3.90E-09**	**0.0083**	**0.001**	**0.0083**			
E3b*	125	7	70	20	0.000136	0.0081					
E3b1*	322	0	27	73	4.76E-69	0.1122	0.0938		0.0029	0.0608	0.0434
E3b1a*	**1**	**0**	**3**	**24**	**5.98E-18**			**0.0049**		**0.0054**	**0.0033**
E3b1a1	**117**	**0**	**5**	**1**	**5.08E-32**	**0.0322**	**0.0285**	**0.0295**			
E3b1a2	0	0	0	0	1.00E+00						
E3b1a3	24	0	0	2	2.39E-06	0.0014	0.0014				
E3b2	**3**	**0**	**12**	**178**	**8.78E-151**			**0.2026**		**0.204**	**0.1965**
E3b3	29	0	0	14	3.75E-10	0.0027	0.0027				
E3b4	4	0	0	0	0.282969						
E3b5	5	0	0	0	0.175446						
E3b6	200	20	0	0	3.99E-62		0.0636	0.0636	0.0028	0.0028	
F*	20	3	7	9	0.071927						
J	**104**	**3**	**49**	**179**	**4.06E-83**	**0.0148**		**0.1566**		**0.1971**	**0.1747**
K2	59	0	6	20	3.56E-12	0.0119	0.0071			0.0018	
R1b*	8	8	140	19	7.21E-42		0.0497		0.0257		0.0117
R1b3	0	4	14	15	2.55E-10						
others[b]	26	0	89	42	2.78E-22	0.0019	0.0156	0.0112	0.0292	0.0239	

3.3.1: Phylogeography of East African Y chromosome Variations

All the individuals sequenced/genotyped were assigned to a Y chromosome lineage **(Figure 3.3.2)** according to the Y chromosome Consortium's recommendations. In this section the results of the phylogeographic analysis of the Y chromosome lineages in East Africa are briefly described for each of the lineages in the context of the lineages' TMRCA estimates **(Table 3.3.1),** and the lineages' distribution within Africa and the Near East **(Tables 3.3.2, 3.3.3)**.

Haplogroup A

Five haplotypes in haplogroup A have previously been identified in Africa [158] (**Figure 3.3.2):** A1, A2, A3a, A3b1, A3b2**.** Haplotype A3b2, defined by the M13 mutation (**Figure 3.3.2),** was observed in most of the East African populations analyzed in the current study and previous studies [38, 155, 164, 290, 291] **(Appendix 6a)**. This haplotype (A3b2) was also observed at low to moderate frequencies in some Central African populations from this study and prior studies [38, 292] **(Figure 3.3.2 and Appendix 6a**). The highest frequency of this haplotype was observed in the Nilo-Saharan speaking populations in Kenya and Sudan [290] (**Figure 3.3.2 and Appendix 6a**). The A3b2 haplotype is significantly higher in frequency among Nilo-Saharan speaking populations **(Table 3.3.2)** relative to populations from other language families in Africa. All the chromosomes that carry this haplotype in Kenya and Tanzania cluster together, while those from western Nilotic speaking Luo and Nuer, cluster with those from Central Africa **(Figure 3.3.3)**.

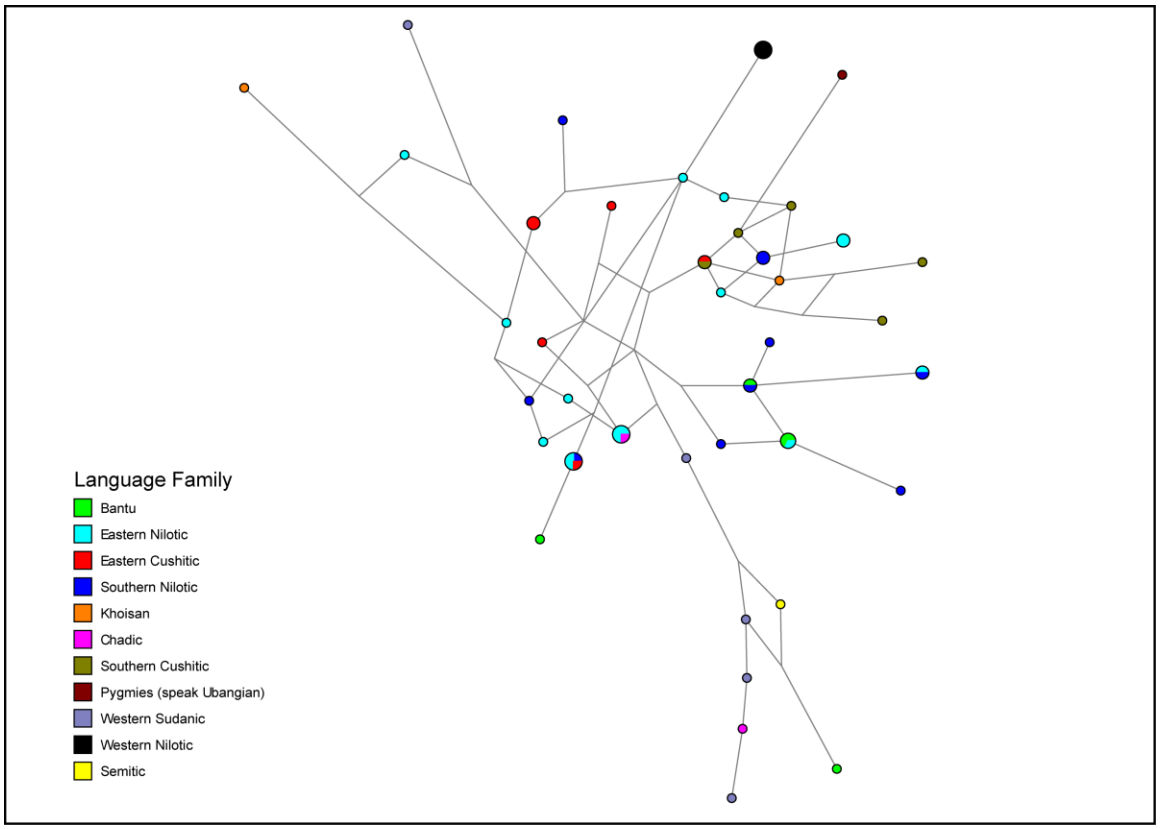


Figure 3.3.3: Median-joining network of 10-loci (DYS389I, DYS389II, DYS390, DYS19, DYS437, DYS448, DYS391, DYS392, DYS393 and DYS635) Y-STR haplotypes for A3b2-M13 haplotypes in East African populations. The network was generated using Network 4.1.1.1 [258]. Networks were processed first by the reduced-median method, and then by the median-joining method [182, 259] without weighting any of the STR loci. Areas of circles are proportional to absolute frequencies and colored according to which language family/branch an individual is from.

The TMRCA estimate for the A3b2 haplotype inferred from variance of microsatellite alleles [262] (described in methods section) from samples genotyped in this study is 26.24+/- 4.61 kya (**Table 3.3.1**). When recently published data from the rest of Africa are included (that include the A1 and A3b1 haplotypes) [293-295] the age estimate is 51.81+/- 12.91 kya.

Haplogroup B

The B haplotypes also show deep lineages in Africa and have TMRCA estimates that are comparable to A haplotypes (**Table 3.3.1**). Haplotype B2a* was observed in the current study at low frequency among the East African Cushitic speaking Konso and Sanye, Afroasiatic populations from southwestern Ethiopia and southeastern Kenya, respectively (**Figure 3.3.2**). The B2a1 haplotype, defined by mutation M152 (**Figure 3.3.2)**, was observed in East African populations in this study, and in low frequencies in earlier studies among the pygmies [38] and neighboring populations in Cameroon [38, 292], and several southern African Bantu speaking populations [38] **(Appendix 6aii)**. The B2a1 haplotype is mostly observed among the Nilo-Saharan speaking and neighboring populations across Africa with comparatively higher frequencies observed among the Nilo-Saharan speakers **(Table 3.3.2)**. In the current study I observed a previously undefined B2a haplotype which I have classified as B2a3 (following the Y chromosome Consortium scheme). The B2a3 haplotype is defined by an A to G transition at position 31 of the M108a fragment that is used to define the B2a2 haplotype (labeled here as TL1 – Tishkoff's lab 1 **Figure 3.3.2**). This new haplotype was found in only four populations: Pokot, El-Molo, Turkana and Yaaku from Kenya (**Figure 3.3.2)**.

Haplotype B2b*, the ancestral form B2b haplotype, is observed in most hunter-gatherer populations in East and Central Africa, and neighboring populations (**Figure 3.3.2, Appendix 6aii)**. The highest frequency of this haplotype was observed in the Khoisan-speaking populations from Tanzania (**Figure 3.3.2, Appendix 6aii)**. This haplotype is also found at moderate frequencies in the southern Cushitic-speaking populations of Tanzania (**Figure 3.3.2)**. Network analysis shows that there are two large

clusters of this haplotype **(Figure 3.3.4)**. One cluster includes mainly lineages from the East African Khoisan populations, and the other includes virtually all other chromosomes from within Africa and the Central African Pygmy individuals who speak a language belonging to the Ubangian branch of Niger-Kordofanian languages.

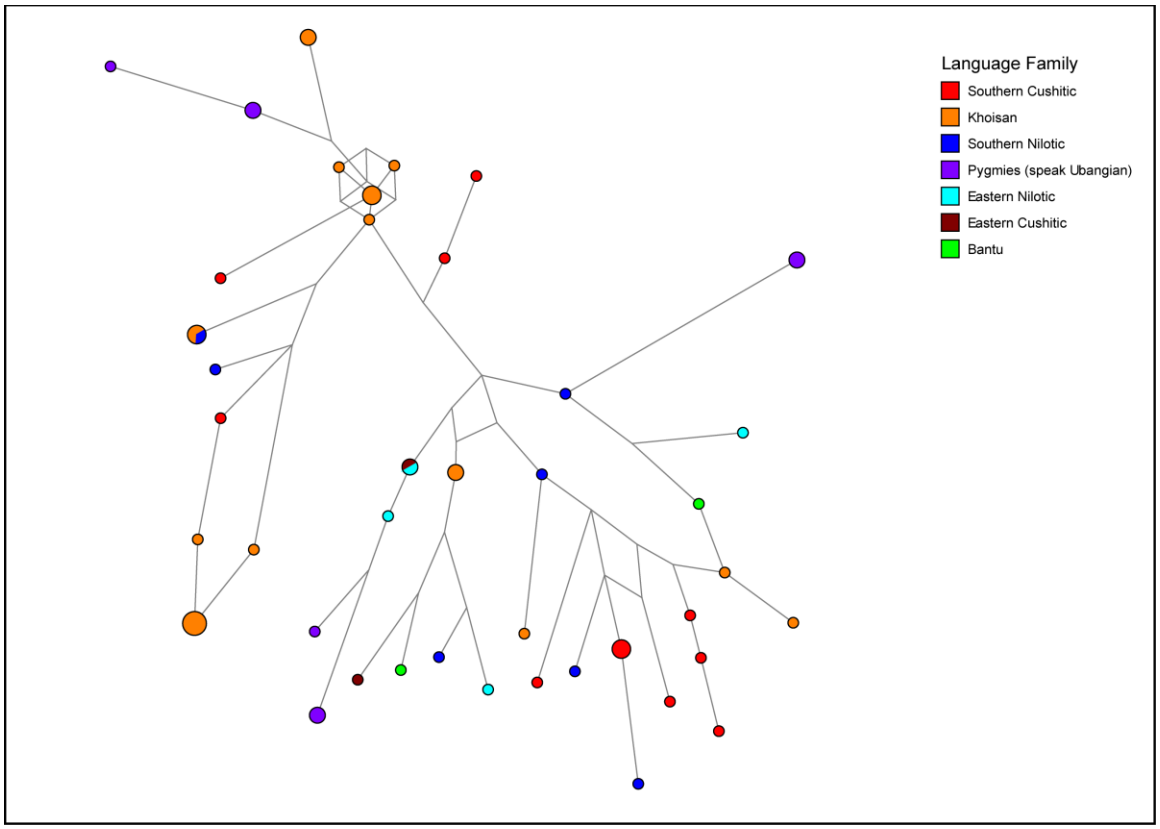


Figure 3.3.4: Median-joining network of 10-loci (DYS389I, DYS389II, DYS390, DYS19, DYS437, DYS448, DYS391, DYS392, DYS393 and DYS635) Y-STR haplotypes for B2b* lineages in East and Central African populations. The network was generated using Network 4.1.1.1 [258]. Networks were processed first by the reduced-median method, and then by the median-joining method [182, 259] without weighting any of the STR loci. Areas of circles are proportional to absolute frequencies and colored according to which language family/branch an individual is from.

Another previously unobserved haplotype was also identified and classified in this study as B2b5 (**Figure 3.3.2)**. It is defined by a G to A transition at position 385 of the M35 fragment that is used to define the E3b haplotypes, labeled as TL2 – Tishkoff's lab 2 (**Figure 3.3.2)**. This haplotype is found at low frequency in several populations from Kenya and Tanzania including hunter-gatherers (Ogiek, Hadza and Sandawe) and neighboring populations (Borana, Taita, Iraqw, Burunge, Maasai and Turu) (**Figure 3.3.2)**. Haplotype B2b4 is observed only among Central African pygmy populations and hunter-gatherer populations of southern Africa: it is uniquely shared among the Pygmy populations and the Namibian Khoisan speaking !Kung/Sekele and Tsumkwe populations (**Figure 3.3.2, Appendix 6a)**.

The TMRCA estimated age of the B2a (29.65 +/- 6.31 kya) and B2b haplotypes is 71.15 +/- 16.29 kya and the age of the M182 mutation, the most common recent ancestral mutation that is shared by B2a and B2b (**Figure 3.3.2)**, is 60.75 +/- 14.5 kya (**Table 3.3.1**).

Haplogroup E

Y chromosome lineages E and F are observed mostly in East and Central/West Africa, and populations that migrated out of these regions in the last 5 kya [57, 59, 62, 74, 87, 111, 296]. E1, defined by the M33 mutation, is found in West Africa, especially in Mali and Burkina Faso [38] (**Appendix 6a)**. The Fulbe (also called Peul, Fulani), a Niger-Kordofanian speaking population from northern Cameroon, has the highest frequency of this haplotype (**Figure 3.3.2, Appendix 6a)**. The E2a haplotype is found predominantly among the Nilo-Saharan speaking populations of East Africa **(Table 3.3.2,**

Table 3.3.3) that constitute the Sudanic branch, and in low frequencies among a few populations neighboring them (**Figure A9.1.4)**. Over a quarter to more than half of individuals from the Kalenjin-speaking populations (Tugen, Marakwet, Pokot and Sengwer) and the neighboring Maa speaking populations carry this haplotype (**Figure 3.3.2, Appendix 6a)**. In contrast, the E2b haplotype is found at high frequency among the Niger-Kordofanian speaking populations in current (**Figure 3.3.2)** and previous studies (**Appendix 6a)** and is found at low frequency in most other populations all across Africa [38].

The E3* haplotype, defined by the presence of the P2 mutation (without the E3a and E3b defining M2 and M35 mutations, respectively) is observed mostly among the East African populations (**Figure 3.3.2**) and at very low frequencies in some populations from Central/West Africa [155, 164] (**Figure 3.3.2, Appendix 6a)**. The E3a haplotypes are observed mostly among the Niger-Kordofanian speaking populations and at low to moderate frequencies in non-Niger-Kordofanian populations that neighbor them (**Figure 3.3.2, Appendix 6a, Appendix 1)**. The E3a7 haplotype, the most common E3a haplotype, has a pattern of decreasing frequency from Central Africa to eastern and southern Africa (**Figure 3.3.2, Appendix 9, Appendix 6a)**. The E3a7 haplotype was observed at high frequency in the Central African populations and at relatively lower frequency in Bantu speakers from eastern and southern Africa **(Figure 3.3.2, Figure A9.1.6, Appendix 6a, Table 3.3.2, Table 3.3.3)**.

The E3b haplotypes have a different distribution pattern to that of the E3a haplotypes, and are observed mostly among the Afroasiatic and Nilo-Saharan populations of East Africa and at low frequency in neighboring Niger-Kordofanian populations

(**Figure 3.3.2, Appendix 6a, Table 3.3.2, Table 3.3.3)**. The E3b1 haplotype, the most common E3b haplotype (**Figure 3.3.2, Appendix 6a)** is observed at highest frequencies among the Cushitic speaking populations of East Africa and at relatively lower frequencies among Semitic speaking populations [156] (**Figure 3.3.2, Appendix 6a)**. The E3b3, E3b4 and E3b5 haplotypes are observed in East African Cushitic speakers at low frequency (**Figure 3.3.2, Appendix 6a)** [155]. The E3b6 haplotype has the highest frequency among the southern Cushitic speaking populations of East Africa (**Figure 3.3.2, Appendix 6a)**. The lineage is also found at moderate to high frequency in East African hunter-gatherer and Nilotic speaking populations (**Figure 3.3.2, Appendix 6a)**. In the current study, the E3b6 haplotype was not observed among the western Nilotic speaking populations of Sudan but is observed at low frequency in eastern Cushitic populations that were originally from Ethiopia (Konso and Borana - **Figure 3.3.2, Appendix 6a)**. In the network analysis **(Figure 3.3.5)**, lineages from Cushitic speaking individuals are observed in both of the two major clusters observed.

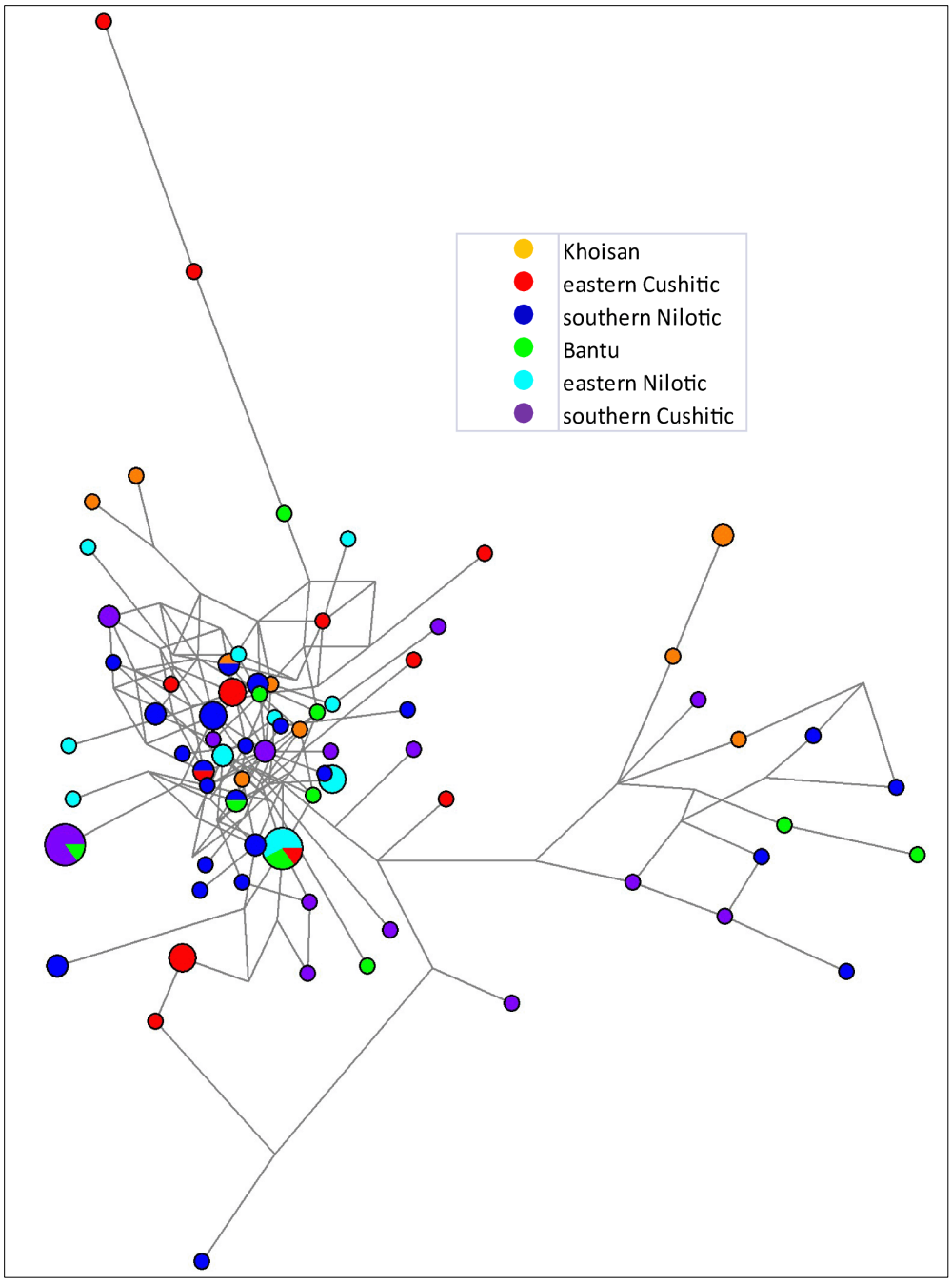


Figure 3.3.5: Median-joining network of 12-loci (DYS389I, DYS389II, DYS390, DYS19, DYS458, DYS437, DYS438, DYS448, DYS391, DYS392, DYS393 and DYS635) Y-STR haplotypes for E3b6 positive individuals in East African populations. The network was generated using Network 4.1.1.1 [258]. Networks were processed first by the reduced-median method, and then by the median-joining method [182, 259] without weighting any of the STR loci. Areas of circles are proportional to absolute frequencies and colored according to which language family/branch an individual is from.

Haplogroup F

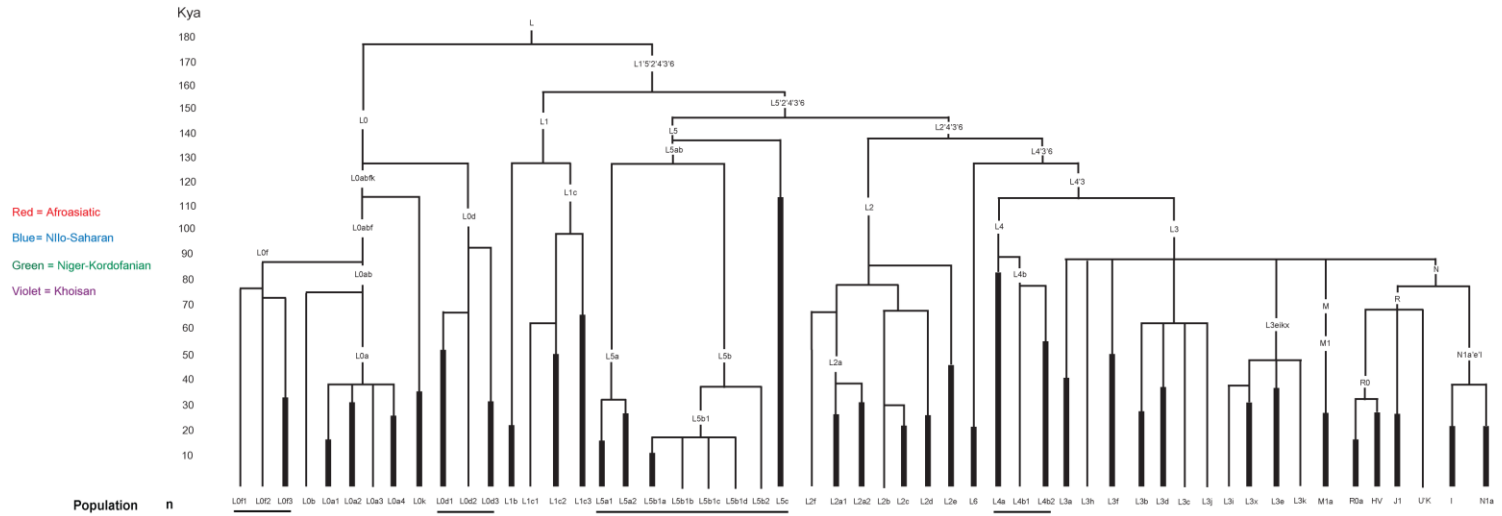
The F haplotypes (J, K and R) are mainly observed at low frequencies among the Afroasiatic speakers and neighboring populations, both in East and Central Africa (**Figure 3.3.2, Appendix 6a).** An unusually high frequency of the R1b* haplotype has been observed among populations from northern Cameroon [38]. A recent study of Y chromosome lineages in western and southern Cameroonian populations showed low to moderate frequencies of R1b* that extend into Gabon [294]. The current study confirms the results of previous work [38, 294] that demonstrates that the R1b* haplotype is at its highest frequency in Chadic speaking northern Cameroonian populations (**Figure 3.3.2, Appendix 6a)**. Moreover, this haplotype was found at moderate frequency in the Laka, a Nilo-Saharan speaking population from Chad. However, the haplotype was not found in any population in East Africa (**Figure 3.3.2, Appendix 6a)**.

Overall the distributions of Y chromosome lineages in Africa show distinct geographic and linguistic distribution patterns. Y chromosome haplotype results indicate that East African populations have the derived variants of the most ancestral lineages (A-A3b2, B-B2) and share lineages with both Sub-Saharan Africa and the Near East (as exemplified by haplotypes E3b, and K/R/J).

3.4: mtDNA

Analysis of *mt*DNA sequences was done in four steps; 1) first, an approximately 1300 bp region of the *mt*DNA genome encompassing the two hypervariable regions (HVI & HVII) of the *mt*DNA D-loop was sequenced in a total of 1,500 individuals from 61 populations from Tanzania, Kenya, Sudan, South Africa, Cameroon and Chad, 2) then approximately 16,600 bp of the complete mitochondrial genome was sequenced for a subsample of 222 individuals (from the 1,500) with proportional representation of the haplogroups identified by the D-loop sequencing (identification based solely on D-loop sequences), 3) then whole *mt*DNA genome sequence phylogenies for published (~900 sequences) and newly sequenced sequences (222) from this study were constructed, 4) based on lineage defining SNPs identified by analyzing the whole genome sequences, SNP genotyping was done for the coding region for the remaining 1278 individuals to unambiguously determine haplotype identities (which were initially based solely on D-loop sequences). Population frequencies of *mt*DNA haplotypes from the current study are shown in **Figure 3.4.1**. Haplogroup/haplotype assignations are shown in **Appendix 8**. Collated frequencies from ~20,000 individuals from over 100 previously genotyped populations from the rest of Africa and the Near East (**Appendix 6bi**) were compared with the current data. A total of 954 *mt*DNA haplotypes from the nine major haplogroups L0-L6, M and N were observed in the following proportions; L0 (22%), L1 (5%), L2 (16%), L3 (33%), L4 (9%), L5 (4%), L6 (1%), M (5%) and N (5%), with L1 and L2 mostly observed in Central African populations and the other haplotypes mostly observed in East African populations.

Figure 3.4.1: Maximum parsimony tree of *mt*DNA haplogroups with relative frequencies of haplogroups/haplotypes from East, Central and South African populations sequenced/genotyped in this study. For most of the branch leading to a haplotype the beginning of the position at which it thickens corresponds to the TMRCA age estimate for the haplotype as estimated from whole genome *mt*DNA sequences. A total of 1,500 individuals from 61 Tanzanian, Kenyan, Sudanese, Cameroonian, South African and Chadian populations **(Appendix 1)** were sequenced for the 1300 bp D-loop region **(Methods section)** and each individual genotyped for haplogroup/haplotype defining coding region's SNP for unambiguous classification. The scheme of *mt*DNA haplogroup classification is based on previously published (900) and newly sequenced (in the current study) 222 complete mitochondrial genomes **(materials and methods section)**.



*[1] Il'Laikipiak includes Il'ngwesi and Mumonyot

*[2] Kalenjin includes 6 Keiyo, 7 Kipsigis, 13 Nandi individuals

[a] includes an individual who is M*, [b] M8, [c] H, [d] U9, [e] U6, [f] U3, [g] T2

Table 3.4.1: *mt*DNA lineage frequency heterogeneity among language families in Africa. The table shows results for the Marascuillo procedure to check if there was significance difference in the proportions of the *mt*DNA lineages among the four African language families, Ks – Khoisan, AA – Afroasiatic, NS – Nilo-Saharan and NK – Niger-Kordofanian. The Marascuillo procedure does pairwise comparison of lineage frequencies between groups and compares the difference of the two frequencies with a critical value determined by a pre-specified significance threshold. If the difference is bigger than the critical value, then it is an indication that the difference between the two proportions is significantly different at the pre-specified significance level. Note that the test is adjusted for multiple comparison, in this case significance level = 0.05/36 (there are 36 tests because there are 36 rows representing lineages). The number in each cell is calculated as: difference between the two proportions - corresponding critical value. Note that only comparisons that yielded significant values are reported – meaning those tests with positive values. Each of the four African language families have significantly higher frequencies of the following respective lineages **(in bold)** relative to the other three language families; Khoisan – L0d1&L0d2, L0k; Nilo-Saharan – L3h, L3i, L5; Niger-Kordofanian – L1b, L1c, L2c; Afroasiatic – other lineages. n = number of individuals, N[a] = number of populations represented from each group, [b]other lineages include N1a, K1, R0a, I, J1, HV1 and U.

Language	Ks	NS	NK	AA	Fisher's exact test	Marascuillo test					
N[a]	*5*	*26*	*91*	*50*							
n	***148***	***901***	***6260***	***2949***	***p-value***	***Ks vs NS***	***Ks vs NK***	***Ks vs AA***	***NS vs NK***	***NS vs AA***	***NK vs AA***
L0a1	4	83	325	95	2.88E-11					0.0171	0.0016
L0a2	4	58	201	36	3.87E-15					0.0167	0.0072
L0a3	0	3	0	5	0.000872						
L0a4	1	4	1	10	1.59E-05						
L0b	0	7	2	2	4.64E-05						
L0d1&2	**54**	**0**	**45**	**0**	**2.18E-83**	**0.1982**	**0.191**	**0.1982**	**0.0027**		**0.0027**
L0d3	1	1	1	7	0.001344						
L0f	3	16	21	53	6.73E-13						0.0039
L0k	**20**	**0**	**2**	**0**	**1.25E-35**	**0.0168**	**0.0165**	**0.0168**			
L1b	**1**	**15**	**458**	**93**	**8.36E-25**		**0.0348**		**0.0338**		**0.0222**
L1c	**0**	**2**	**1324**	**37**	**6.52E-248**		**0.1898**	**0.0039**	**0.1866**		**0.1756**
L2a	0	4	48	14	0.326105		0.003				
L2a1	9	91	913	236	1.50E-20		0.0002				0.0376
L2a2	0	40	13	7	8.51E-27	0.0155			0.0133	0.0129	
L2b	4	5	190	29	9.70E-13				0.0109		0.0086
L2c	**0**	**1**	**280**	**18**	**1.39E-37**		**0.0337**	**0.0001**	**0.0317**		**0.0261**
L2d	0	3	74	10	5.19E-05		0.0061				0.0011
L2e	0	1	17	8	0.883277						
L2f	0	0	5	0	0.409489						
L3*	0	3	30	17	0.822083		0.0011				
L3a	0	13	11	26	4.88E-08			0.0016			
L3b	8	35	407	84	1.61E-13						0.0181
L3c	0	3	4	5	0.088468						
L3d	0	32	316	60	2.79E-13	0.0095	0.0388	0.0094			0.0141
L3e	15	36	877	138	1.65E-55				0.067		0.0686
L3f	0	45	306	135	0.009848	0.0194	0.0374	0.0296			
L3h	**1**	**49**	**49**	**35**	**3.26E-18**	**0.005**			**0.0144**	**0.0096**	
L3i	**0**	**45**	**4**	**36**	**9.84E-37**	**0.0194**		**0.0037**	**0.0187**	**0.006**	**0.0029**
L3j	0	0	2	1	1.00E+00						
L3k	0	0	3	3	0.682068						
L3x	1	15	7	46	1.56E-17						0.0047
L4	19	89	53	87	8.90E-54		0.004		0.0482	0.0254	0.007
L5	**1**	**81**	**19**	**33**	**1.28E-54**	**0.034**			**0.0466**	**0.0378**	
L6	0	4	0	8	2.61E-05						
M1a	**0**	**46**	**5**	**130**	**2.19E-66**	**0.0202**		**0.0282**	**0.0193**		**0.0273**
others[b]	**2**	**61**	**247**	**1445**	**< <<2.2e-16**	**0.0009**		**0.4208**		**0.3699**	**0.4104**

Table 3.4.2: *mt*DNA lineage frequency heterogeneity among African geographic regions. The table shows results for the Marascuillo procedure to check if there was significance difference in the proportions of the *mt*DNA lineages among the four African regions, EA – East Africa, SA – South Africa, CWA – Central-West Africa and NWA – North West Africa. The Marascuillo procedure does pairwise comparison of lineage frequencies between groups and compares the difference of the two frequencies with a critical value determined by a pre-specified significance threshold. If the difference is bigger than the critical value, then it is an indication that the difference between the two proportions is significantly different at the pre-specified significance level. Note that the test is adjusted for multiple comparison, in this case significance level = 0.05/36 (there are 36 tests because there are 36 rows representing lineages). The number in each cell is calculated as: difference between the two proportions - corresponding critical value. Note that only comparisons that yielded significant values are reported – meaning those tests with positive values. Each of the four African regions have significantly higher frequencies of the following respective lineages **(in bold)** relative to the other three geographical regions; South Africa – L0d1&2, L0k; East Africa – L0f, L3a, L3h, L3i, L3x, L4, L5, M1a; Central-West Africa – L0a2, L1b, L1c, L2b, L2c, L3b, ; North/Northwestern Africa – others. n = number of individuals, N^a = number of populations represented from each group, [b]other lineages include N1a, K1, R0a, I, J1, HV1 and U.

Region	EA	SA	CWA	NWA	Fisher's exact test	Marascuillo Test					
N[a]	56	27	77	15							
n	1800	1649	5913	1904	*p-value*	EA vs SA	EA vs CWA	EA vs NWA	SA vs CWA	SA vs NWA	CWA vs NWA
L0a1	151	84	256	34	3.22E-20		0.0109	0.0357		0.0069	0.0085
L0a2	**106**	**127**	**85**	**0**	**3.30E-65**		**0.0203**	**0.0355**	**0.0342**	**0.0494**	**0.0079**
L0a3	8	0	0	0	6.27E-07						
L0a4	16	0	0	0	2.12E-13						
L0b	11	0	0	0	2.33E-09						
L0d1&2	**0**	**437**	**0**	**0**	**<<2.2e-16**	**0.2192**			**0.2192**	**0.2192**	
L0d3	9	1	0	0	2.31E-07						
L0f	**86**	**1**	**0**	**6**	**3.72E-63**	**0.0258**	**0.0266**	**0.0228**			
L0k	**0**	**22**	**0**	**0**	**3.87E-19**	**0.0014**			**0.0014**	**0.0014**	
L1b	**14**	**23**	**505**	**73**	**1.48E-63**		**0.06**	**0.0101**	**0.0519**	**0.0022**	**0.023**
L1c	**10**	**95**	**1286**	**23**	**3.05E-263**	**0.0268**	**0.1882**		**0.1268**	**0.0192**	**0.1805**
L2a	8	2	60	5	8.24E-06				0.0024		0.0001
L2a1	156	213	848	132	1.93E-22		0.0229			0.0173	0.0429
L2a2	22	0	37	1	8.18E-09	0.0013		0.0006	0.0019		0.0009
L2b	**10**	**20**	**201**	**12**	**2.80E-22**		**0.0161**		**0.0068**		**0.0152**
L2c	**1**	**21**	**323**	**12**	**3.66E-58**	**0.0003**	**0.0414**		**0.0249**		**0.0337**
L2d	4	4	71	9	8.30E-07		0.0022		0.0017		
L2e	0	0	22	5	0.001382		0.0004		0.0004		
L2f	0	3	2	0	0.086281						
L3*	8	4	31	8	0.533734						
L3a	**49**	**0**	**0**	**1**	**1.68E-38**	**0.0111**	**0.0111**	**0.0104**			
L3b	**46**	**52**	**432**	**42**	**1.62E-29**		**0.0263**		**0.0185**		**0.0309**
L3c	3	0	4	5	0.046567						
L3d	49	60	321	21	2.64E-20		0.0067			0.0035	0.0273
L3e	52	182	847	86	6.31E-70	0.045	0.089			0.027	0.0703
L3f	79	34	329	54	4.30E-12				0.0157		0.0069
L3h	**85**	**1**	**41**	**9**	**8.75E-36**	**0.0254**	**0.0187**	**0.0204**	**0.0011**		
L3i	**76**	**0**	**1**	**8**	**4.36E-54**	**0.0223**	**0.0221**	**0.0171**			
L3j	3	0	0	0	0.007204						
L3k	1	0	6	2	0.759878						
L3x	**59**	**0**	**4**	**6**	**5.29E-38**	**0.0151**	**0.0144**	**0.0111**			
L4	**205**	**6**	**37**	**5**	**1.14E-118**	**0.0781**	**0.0758**	**0.0793**			
L5	**120**	**4**	**8**	**5**	**6.46E-77**	**0.039**	**0.0405**	**0.0388**			
L6	10	0	0	2	1.22E-07						
M1a	**120**	**0**	**4**	**57**	**4.65E-85**	**0.0419**	**0.0412**	**0.007**		**0.0135**	**0.0128**
others[b]	**223**	**254**	**152**	**1281**	**<<2.2e-16**		**0.0643**	**0.493**	**0.0899**	**0.46**	**0.601**

3.4.1: Phylogeography of East African mtDNA Variation

All the individuals sequenced/genotyped were assigned to a *mt*DNA lineage type **(Figure 3.4.1)** according to the haplogroup nomenclature used in previous studies [15, 16, 157, 171, 172, 185]. All sequences were first classified on the basis of control regions (encompassing HVS-I & HVS-II motifs **(Appendix 8)**), and coding region haplotype SNPs used to further confirm assignation unambiguously. In this section the results of the phylogeographic pattern of the *mt*DNA lineages in East Africa are briefly

described for each of the lineages in the context of the lineages' TMRCA estimates **(Appendix 7b),** and the lineages' distribution within Africa and the Near East **(Table 3.4.1, 3.4.2)**.

Haplogroup L0

There is a distinct distribution pattern of L0 haplotypes with L0a, L0b and L0f mostly observed in East and Central/West Africa, while L0d and L0k are observed mostly in South African San populations (**Figure 3.4.1, Appendix 6b, Table 3.4.3)**. Haplotype L0b was observed at low frequency among the Luhya, Sengwer and Wata populations from Kenya (**Figure 3.4.1, Appendix 6b)**.

The distribution of the L0f haplotype is restricted to East Africa mostly among the Afroasiatic speaking hunter-gatherer populations of East Africa (**Figure 3.4.1, Appendix 6b)**. Three sub-haplotypes have so far been identified within the L0f haplotype, L0f1-3 **(Appendix 9, Appendix 8)**. The L0f1 sub-haplotype was observed at moderate to high frequency among the hunter-gatherer populations from southern Kenya (Boni and Sanye), Khoisan speaking Sandawe and neighboring southern Cushitic speaking Burunge populations from northern Tanzania, and low frequency among several Kenyan and Tanzanian populations. The L0f2 and L0f3, sub-haplotypes were observed at moderate to high frequencies among the southern Cushitic speakers and at low frequencies among the eastern Nilotic, eastern Cushitic and hunter-gatherer populations from northern Kenya **(Appendix 9, Appendix 8)**.

Haplotypes L0d and L0k are observed mostly among South African Khoisan, South African colored populations and neighboring Bantu speaking populations [185,

295, 297] (**Figure 3.4.1, Appendix 6b, Table 3.4.2)**. A younger clade of L0d, L0d3 was previously found in a South Africa Khoisan individual [157] **(Figure 3.4.2)**. In the current study the L0d3 haplotype was observed among East African hunter gatherer populations from the Kenyan Coast (Boni) and southern Cushitic speakers (Burunge) and hunter-gatherer populations (Akie and Sandawe) from northern Tanzania (**Figure 3.4.1, Appendix 6b)**. This haplotype was also observed at low frequencies in the Nilo-Saharan speaking Turkana population [151, 152] and the Bantu speaking population of Ronga (from southern Mozambique and Swaziland) [185, 298] **(Appendix 6b)**.

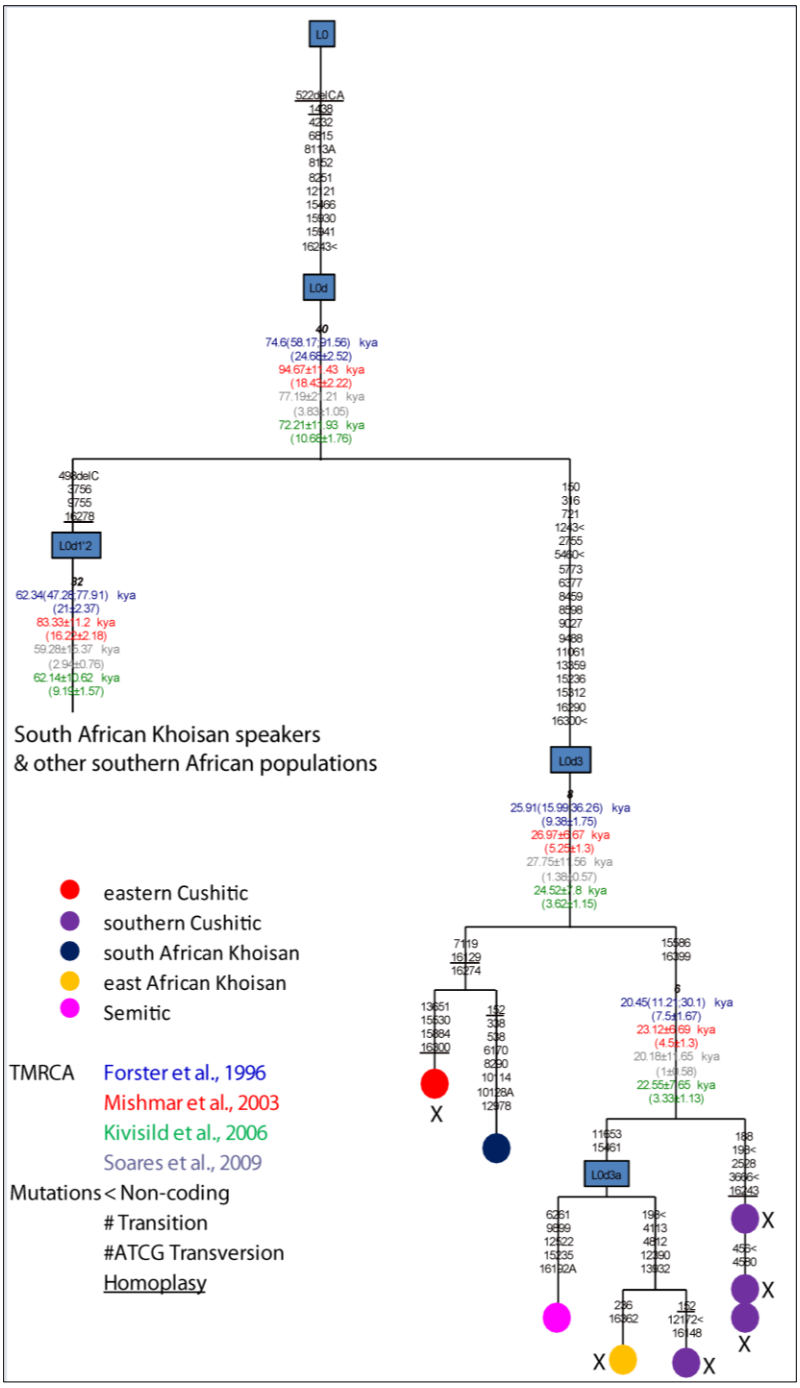
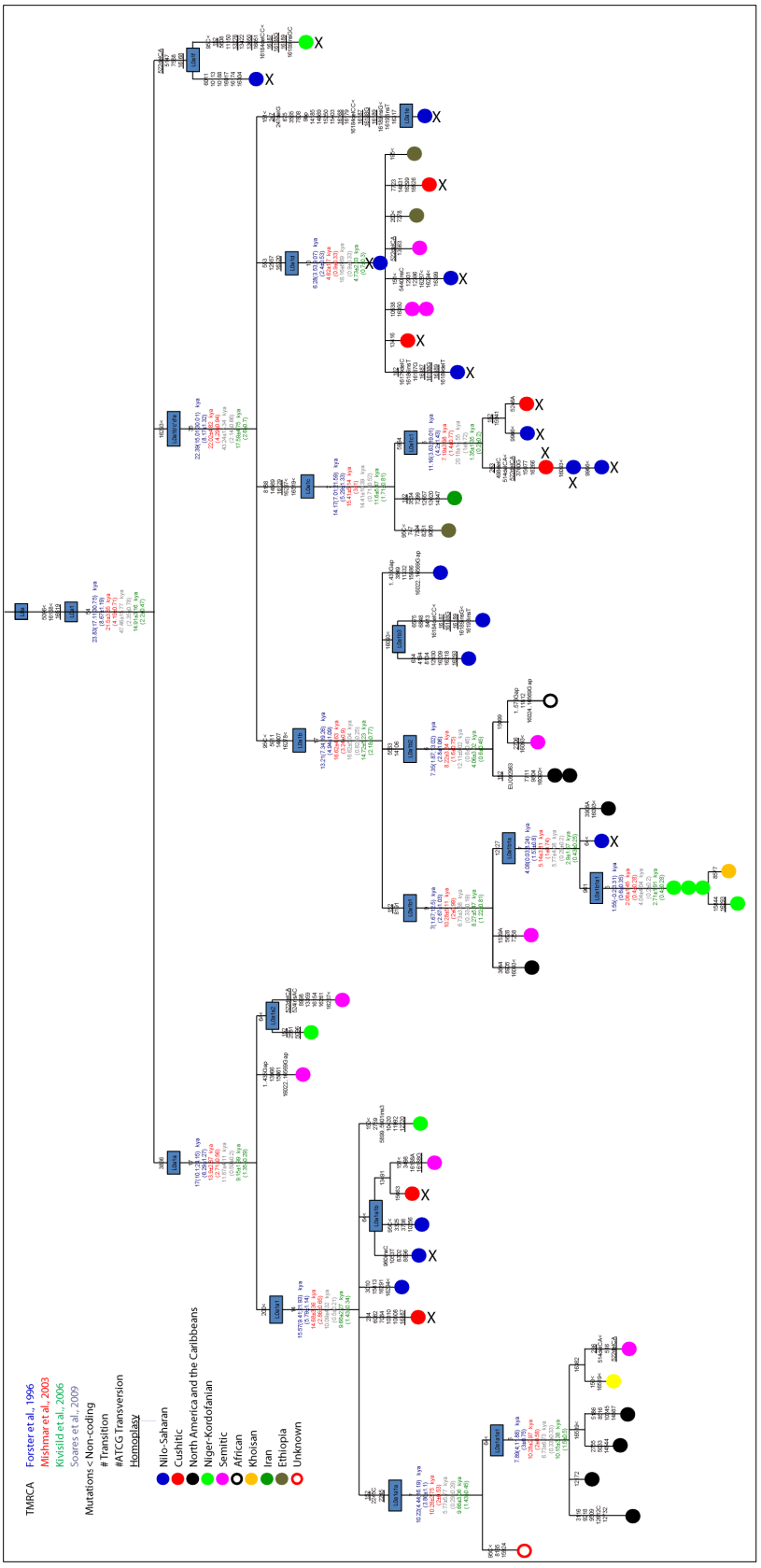


Figure 3.4.2: Phylogeny of *mt*DNA complete genome sequences that belong to the L0d3 lineage. Samples marked **X** were sequenced in this study. Linguistic group/family of the population/s the sequences were sampled from in the current study, and published sequences whose linguistic affiliations are known are shown. For published data where linguistic affiliations are not known, the country/region in which the sequences are sampled is indicated. L0d1 and L0d2 are observed exclusively in southern African populations. The frequency distributions of the lineage in African populations are shown in **Appendix 6b.**

Haplotype L0a1 is the most widely distributed L0 haplotype in Africa with varied regional frequencies of its sub-haplotypes L0a1a-L0a1d in Africa, the Caribbean and among African Americans [180, 299-301]. The L0a2 haplotype is observed mostly in East and Central African populations **(Figure 3.4.3, Appendix 6b).** In East Africa, the L0a2 haplotype has nearly the same distribution pattern as L0f **(Appendix 9)** with the highest frequency found in southern Cushitic speaking populations and moderate frequencies in southeastern Kenyan hunter-gatherer populations (Boni and Sanye) and northern Tanzanian Bantu populations **(Figure 3.4.3, Appendix 8)**. Most of the East Africans that have L0a2 lineages belong to the L0a2c sub-haplotype, which is defined by a T->C transition at base pair position 95, with an average TMRCA age estimate of 37 kya **(Figure 3.4.3, Appendix 7b, Appendix 8)**. The L0a2b sub-haplotype is observed at low frequencies among the hunter-gatherer populations from northern Tanzania and southeastern Kenya (Boni, Akie and Hadza) (**Appendix 8).**

Figure 3.4.3: Phylogeny of *mt*DNA complete genome sequences that belong to the L0a2 lineages. Samples marked **X** were sequenced in this study. Linguistic group/family of the population/s the sequences were sampled from in the current study, and published sequences whose linguistic affiliations are known are shown. For published data where linguistic affiliations are not known, the country in which the sequences are sampled is indicated. The frequency distribution of the lineage in African and non-African populations are shown in **Appendix 6.**


TMRCA
Forster et al., 1996
Mishmar et al., 2003
Kivisild et al., 2006
Soares et al., 2009
Mutations < Non-coding
Transition
#ATCG Transversion
Homoplasy
Nilo-Saharan
Cushitic
North America and the Caribbeans
Niger-Kordofanian
Semitic
African
Khoisan
Iran
Ethiopia
Unknown

Haplogroup L1

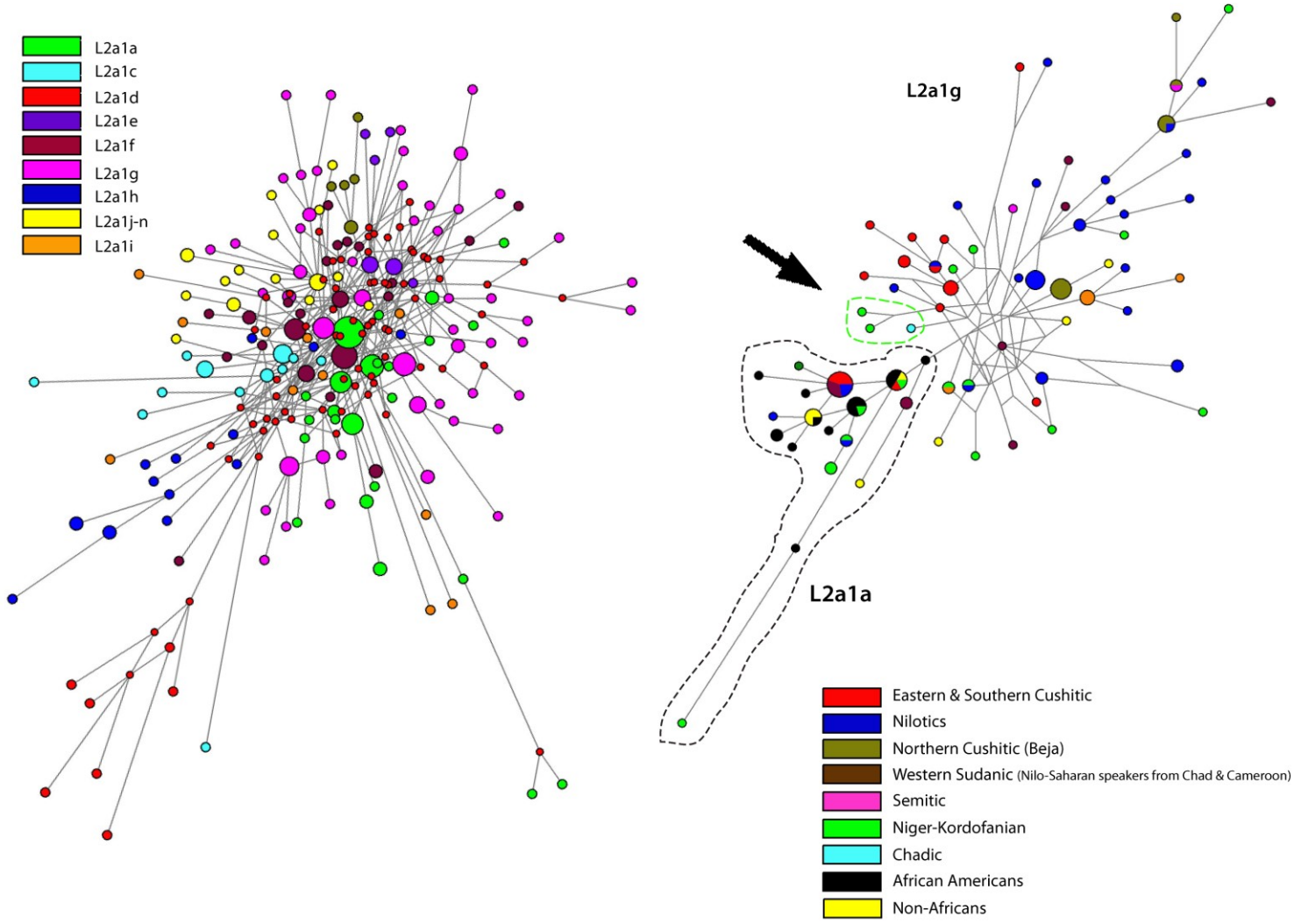
The L1 haplogroup is most prevalent in Central/West Africa (**Table 3.4.3)**, with the L1c haplotype being at highest frequency among pygmy populations of Central Africa **(Appendix 6b, Appendix 9, Figure A9.2.5)**. Virtually all of the individuals from the three pygmy populations sequenced in this study as well as a pygmy populations analyzed in a previous study [302] have the L1c1a sub-haplotype; Medzan (100%), Baka (96%) and Bakola (96%) (this study), and Mbenzele pygmy population (96%) from Central African Republic [302] **(Appendix 9 Figure A9.2.5, Appendix 6b, Appendix 8)**. Haplotype L1b is mostly observed in West African populations **(Appendix 6b, Appendix 9, Figure A9.2.6, Table 3.4.2, Table 3.4.3)**

Haplogroup L2

The L2a1 haplotype is distributed widely across Africa **(Figure A9.2.7),** with highest frequency in the Central African Baggara population (24%), an Afroasiatic speaking group that trace their ancestry to the Middle East [303] **(Figure 3.4.1, Appendix 6b)**. Fourteen clades have been defined so far from the L2a1 haplotype and are labeled L2a1a-n. Nearly three quarter of the L2a1 individuals genotyped in this study belong to three clades: L2a1g (50%), L2a1a (11%) and L2a1h (11%) **(Appendix 8)**. In addition to the L2a1 defining control region's mutation motif, these clades have unique control region mutations that are diagnostic for distinguishing them **(Appendix 8)**. Two of these clades, L2a1g and L2a1h, are found mostly among the Afroasiatic and Nilo-Saharan populations, whereas L2a1a lineages are observed mostly among Niger-Kordofanian speakers, Central African populations and African Americans **(Figure 3.4.4,**

Appendix 8) [226, 299]. Interestingly, an individual from the Chadic speaking Giziga population and two individuals from the Fulani (Cameroon) population that have the L2a1g haplotype cluster together with the east African Cushitic speakers in a network analysis **(Figure 3.4.4).**

Figure 3.4.4 (previous page): Network for *mt*DNA d-loop sequences that belong to the L2a1 haplotype constructed using the Network program version 4.1.0.9 [257, 258]. The network on the left is for all the L2a1 individuals including published sequences. The network on the right is for the most common clades identified in east African population data in this study (L2a1a and L2a1g). The arrow points to sequences of a Chadic and two Fulani individuals in L2a1g that cluster with east African Cushitic speakers.


L2a1a
L2a1c
L2a1d
L2a1e
L2a1f
L2a1g
L2a1h
L2a1j-n
L2a1i
L2a1g
L2a1a
Eastern & Southern Cushitic
Nilotics
Northern Cushitic (Beja)
Western Sudanic (Nilo-Saharan speakers from Chad & Cameroon)
Semitic
Niger-Kordofanian
Chadic
African Americans
Non-Africans

Haplogroup L3

The *mt*DNA L3 haplotypes are mostly observed among populations in East and Central/West Africa **(Figure 3.4.1, Appendix 6b)**. L3 haplotype include L3a-L3k plus M and N lineages. Note that older studies [152] defined L3A (different from L3a) as encompassing all the L3 haplotypes, L3a-L3k, excluding the M and N haplotypes. L3a haplotypes (**Figure A9.2.12)** were observed at highest frequencies in the following populations; Yaaku (26%), El-Molo (25%), Sanye (15%), and neighboring populations such as Samburu (23%) and Orma (15%) **(Figure 3.4.1, Appendix 6b)**.

The L3b haplotype is observed mainly among the Niger-Kordofanian and Nilo-Saharan speaking populations **(Figure 3.4.1, Appendix 6b)**. Frequency comparisons between major geographical regions in Africa (North/Northwest, East, Central and West Africa) for the L3b haplotype **(Figure 3.4.1, Appendix 6b)** show significantly higher frequency observed in central west Africa **(Table 3.4.3)** relative to other African regions **(Figure 3.4.1, Appendix 6b).** The L3d and L3e haplotypes are also observed mainly among the Niger-Kordofanian speakers and neighboring populations **(Appendix 9, Figure A9.2.14, Figure A9.2.15)**. Frequencies of these two haplotypes decrease with distance from Central Africa **(Figure 3.4.1, Appendix 6b, Appendix 9 Figure A9.2.14, Figure A9.2.15).** Comparison of L3d frequencies among populations speaking languages belonging to different language families in Africa **(Figure 3.4.1, Appendix 6b)** demonstrated a significantly higher frequency among the Niger-Kordofanian speakers **(Table 3.4.2)** relative to populations speaking languages belonging to the other three African language families**.** Overall, *mt*DNA lineages, L3b, L3d and L3e, are observed at significantly higher proportion among the Niger-Kordofanian relative to the Nilo-Saharan

and Afroasiatic speaking populations but not relative to the Khoisan (**Table 3.4.2)**. Moreover, they are also observed at significantly higher proportion in central-west Africa relative to East Africa and north-west Africa but not southern Africa **(Table 3.4.3).**

Haplotype L3f is widely distributed among East and Central African Afroasiatic and Nilo-Saharan speaking populations, and is at low frequency in neighboring Niger-Kordofanian populations. The highest frequency was observed in the Afroasiatic speaking Beja (27%) from Northeast Sudan **(Figure 3.4.1, Appendix 6b, Appendix 9)**. Most of the east African individuals analyzed in the current study belong to the L3f1 clade, however, most of the Chadic speaking individuals from central Africa in the study belong to the L3f3 clade **(Appendix 8)**. The L3f2b sub-haplotype defined by a combination of control region mutations at nucleotide positions 16311 (HVI) and 152 (HVII), are observed exclusively in Cushitic speaking populations from East Africa **(Appendix 8)**.

The L3h haplotype is observed at high frequency in Nilo-Saharan speaking populations **(Table 3.4.2)** and at low frequency in East African Afroasiatic speaking populations **(Figure 3.4.1, Appendix 6b)**. Outside of East Africa, the L3h haplotype is observed at low frequencies in a few populations that live in the vicinity of Nilo-Saharan speaking populations [16, 151, 152, 204, 219, 227, 297, 304-318] **(Figure 3.4.1, Appendix 6b)**.

The L3i haplotype is mostly observed among the Nilo-Saharan and Afroasiatic speaking populations from Kenya and Ethiopia **(Figure 3.4.1, Appendix 6b, Appendix 9 Figure A9.2.17, Table 3.4.2, Table 3.4.3)**. Interestingly, the L3i1 sub-haplotype

(Figures 3.4.5, Appendix 8) is prevalent among the Nilo-Saharan speaking populations, while L3i2 is mostly observed among Afroasiatic speaking populations **(Appendix 8)**.

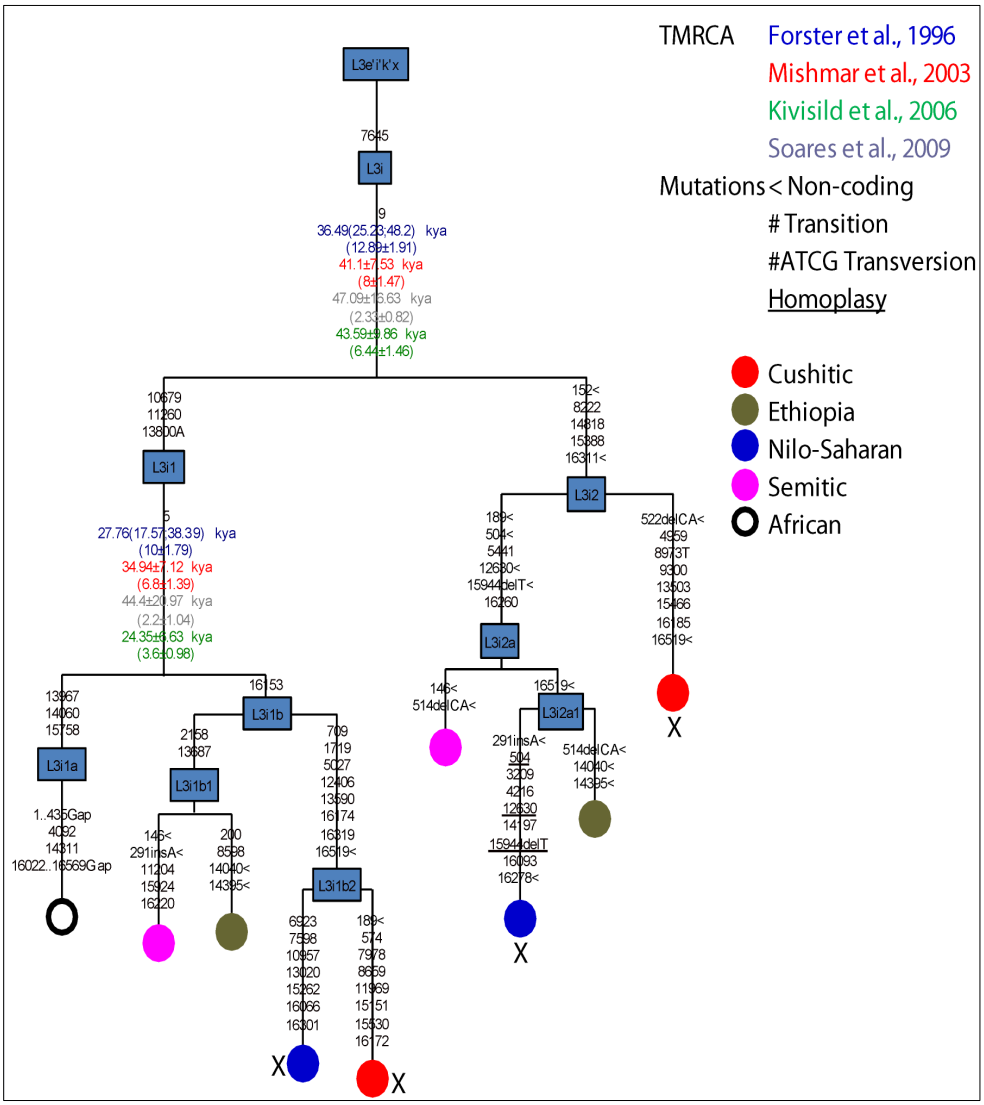


Figure 3.4.5: Phylogeny of *mt*DNA complete genome sequences that belong to the L3i lineages. Samples marked **X** were sequenced in this study. Linguistic group/family of the population/s the sequences were sampled from in the current study, and published sequences whose linguistic affiliations are known are shown. For published data where linguistic affiliations are not known, the country in which the sequences are sampled is indicated. L3i1 sequences are mainly observed among the Nilo-Saharan speakers, while L3i2 are mainly observed among the Afroasiatic speakers (**Appendix 8**). The frequency distribution of the lineage in African and non-African populations are shown in **Appendix 6.**

The L3x haplotype is commonly observed among the Afroasiatic and Maa (eastern Nilotic) speaking populations from Kenya and Tanzania **(Figure 3.4.1, Appendix 6b, Figure A9.2.18)**. Its frequency maximum is centered in southern Ethiopia **(Appendix 9)**, with the highest frequency among East highland Cushitic speaking populations of northern Kenya and southern Ethiopia (**Figure 3.4.1, Appendix 6b)**.

Haplogroup L4

The L4 haplotype, previously defined as L3g [185], is mainly restricted to eastern and northeastern Africa **(Figure 3.4.1, Appendix 6b, Figure A9.2.19, Table 3.4.3)**. Frequency comparisons between East African countries show that this haplotype is most commonly observed in Kenya and Tanzania with decreasing frequency moving into North Africa [16, 185, 312] **(Figure 3.4.1, Appendix 6b, Figure A9.2.19)**. The L4 haplotype is also observed at low frequencies in a few Central African populations [204, 302] and southeastern Arabian populations [16, 219, 312, 313, 319] **(Figure 3.4.1, Appendix 6b)**. Over two-thirds of Hadza individuals analyzed in this study carry the L4 haplotype **(Figure 3.4.1)**. Moreover, most of the other hunter-gatherer populations in East Africa carry it in moderate to high frequencies **(Figure 3.4.1)**. Most of the haplotypes found in East African hunter-gatherer populations are a derived variant, L4b2a2b defined by HVI & HVII's nucleotide base positions 16172 and 244, respectively (in addition to 16293-16355-16399-146 motif that defines L4b2) **(Figure 3.4.6)**.

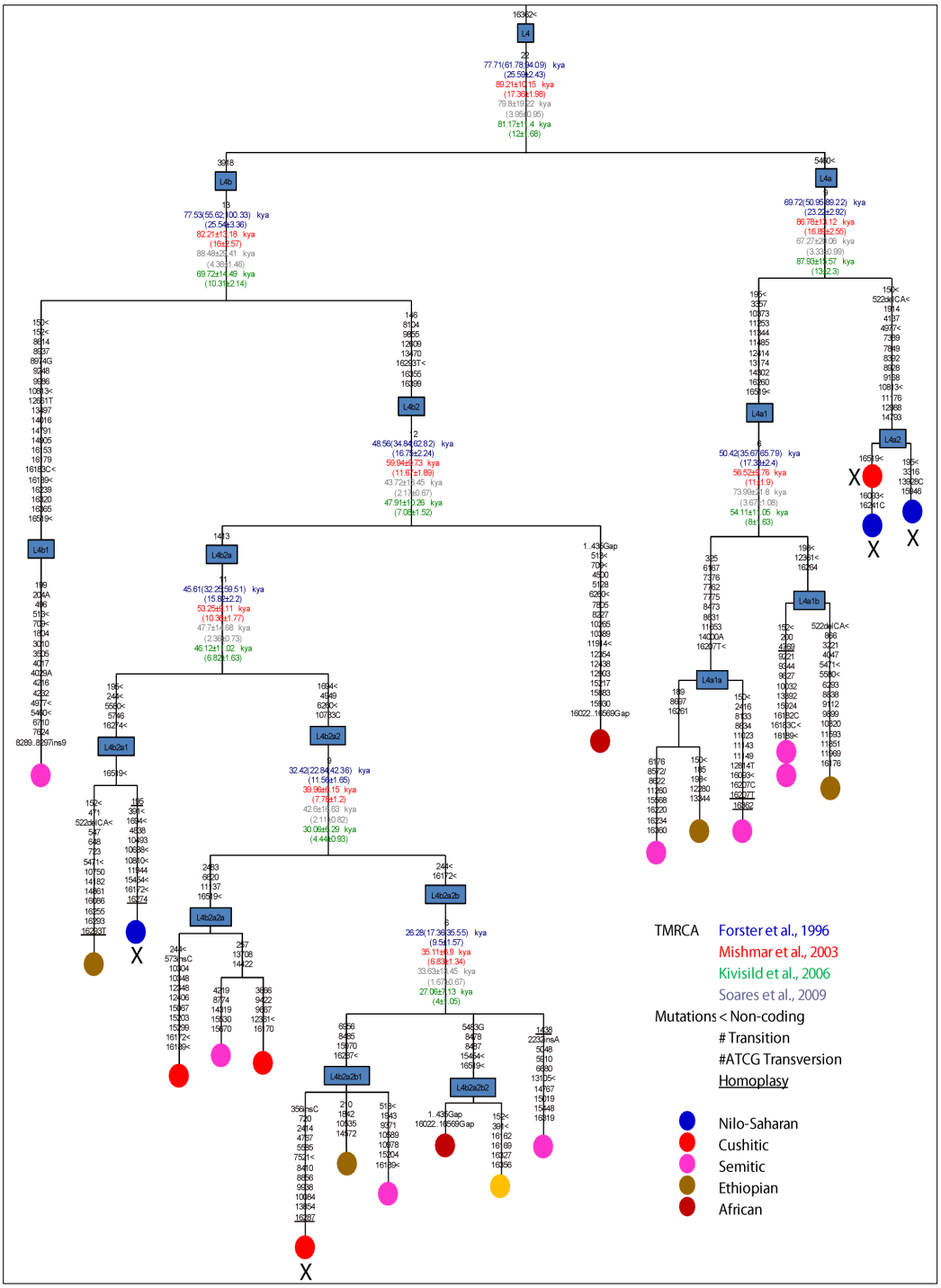
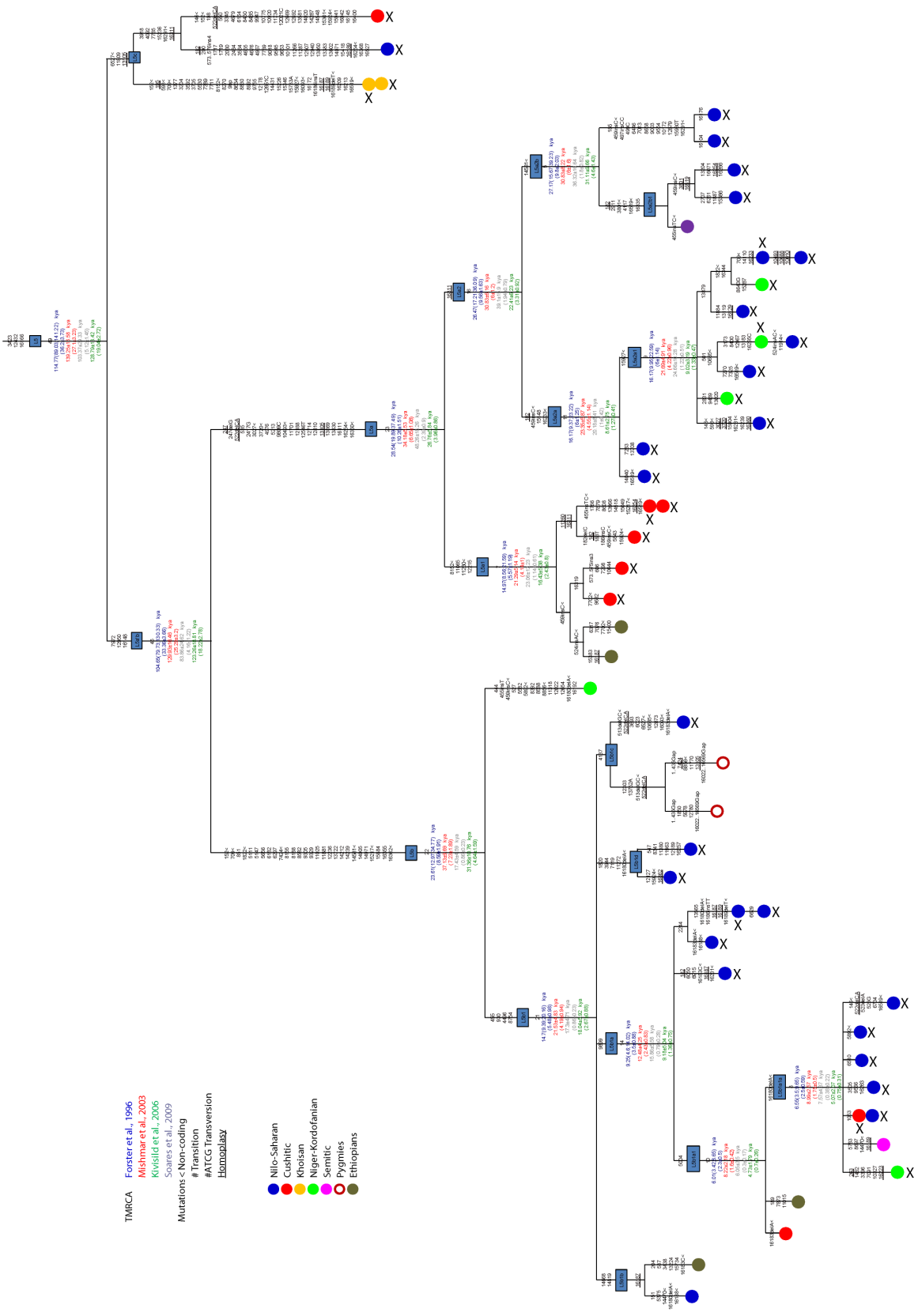


Figure 3.4.6: Phylogeny of *mt*DNA complete genome sequences that belong to the L4 lineages. Samples marked **X** were sequenced in this study. Linguistic group/family of the population/s the sequences were sampled from in the current study, and published sequences whose linguistic affiliations are known are shown. For published data where linguistic affiliations are not known, the country in which the sequences are sampled is indicated. The frequency distribution of the lineage in African and non-African populations are shown in **Appendix 6.**

Haplogroup L5

Based on extensive analysis among Nilotic speakers in this study and published data, the L5 haplotype is observed at highest frequency among the Nilo-Saharan speaking populations (**Table 3.4.2),** specifically among southern Nilotic speakers in Kenya **(Figure 3.4.1, Appendix 6b)**. Most of the individuals that have lineages classified as L5 are from L5a and L5b sublineages **(Figure 3.4.1, Figure 3.4.7, Appendix 8)**. A sub-haplotype, L5a1, defined by transitions at 8152, 11065, 11260 and 12215 (and either 16311del or 459insC mutations in the control region) **(Figure 3.4.7)** is observed exclusively among the Cushitic speaking populations of northern Kenya and southern Ethiopia (and other Afroasiatic speakers from Ethiopia (**Appendix 8)** [16]) at low frequencies. The L5c sublineage, a very old and distinct lineage, is observed at low frequencies in eastern Africa; among Khoisan, Afroasiatic and Nilo-Saharan individuals **(Figure 3.4.1, Figure 3.4.7, Appendix 6, Appendix 8)**.

Figure 3.4.7: Phylogeny of *mt*DNA complete genome sequences that belong to the L5 lineages. Samples marked **X** were sequenced in this study. Linguistic group/family of the population/s the sequences were sampled from in the current study, and published sequences whose linguistic affiliations are known are shown. For published data where linguistic affiliations are not known, the country in which the sequences are sampled is indicated. L5a1 sequences are mainly observed among the Afroasiatic speakers (**Appendix 6)**. The frequency distribution of the L5 lineage in African and non-African populations are shown in **Appendix 6.**


TMRCA
Forster et al., 1996
Mishmar et al., 2003
Kivisild et al., 2006
Soares et al., 2009
Mutations < Non-coding
Transition
#ATCG Transversion
Homoplasy
Nilo-Saharan
Cushitic
Khoisan
Niger-Kordofanian
Semitic
Pygmies
Ethiopians

Haplogroup M and N

The M haplogroup lineages are generally observed outside of Africa, with most members, M2-M64, found in the Indian sub-continent [7, 320, 321], aboriginal populations from South-East Asia [7, 322, 323] and the South West Pacific [324]. The M1a subhaplotype is most common among the Cushitic speaking populations in East Africa **(Figure 3.4.1, Appendix 6b, Table 3.4.3).** The N haplotypes are also generally observed outside of Africa, but some N-clade haplotypes, N1a, I, J1 and K1 and U were found at low frequency in Afroasiatic speaking populations of East Africa **(Figure 3.4.1, Appendix 9, Appendix 12)**, and implication of this will be addressed in the discussion section.

Similar to the Y chromosome lineages' distribution pattern, the *mt*DNA lineages seem to cluster based on geographic classification, and in some cases based on linguistic classification. Overall, the distribution of *mt*DNA lineages in Africa indicates that East African populations carry some of the most ancestral lineages in Africa (L0b, L0f, L0a, L5) and share some lineages with populations from the Near East as exemplified by haplotypes L3x, M1a, and N-clades (J1, K1a and U9).

3.4.2: Neutrality Tests for Mitochondrial Genomes

Tests of neutrality were done for complete *mt*DNA sequences to check if natural selection had an effect on the level of variation in the *mt*DNA genomes. Results from the current study indicate that there has been both gene-specific and lineage-specific selection. All the lineages analyzed here except L4 and L5 have significant negative values for both Tajima's D and Fu & Li's D* (**Table 3.4.2.1**). A neutrality index (NI) was computed for the 13 genes encoded in the mitochondria. Only L0 and M1a lineages showed significant NI values with the former indicating a signature of negative selection and, the latter showing a signature of weakly positive selection.

Hap	n	Tajima's D	Fu & Li's D*	Ni[1]	Gene[2]	Ni[3]
L0	178	-2.13123*	-4.05685*	1.68**	COIII	30***
L1	120	-1.82542*	-4.37743*	1.33		
L2	207	-2.49883***	-7.79301*	1.39		
L3	291	-2.57614***	-7.42726*	1.23	ND5	0.43*
L4	20	-1.89791*	-2.40851	1.58		
L5	53	-1.55188	-2.72479*	1.17		
M1a	92	-2.56307***	-5.47493**	2.15*		
N	226	-2.33426**	-8.78122*	1.29	COI	9.21*

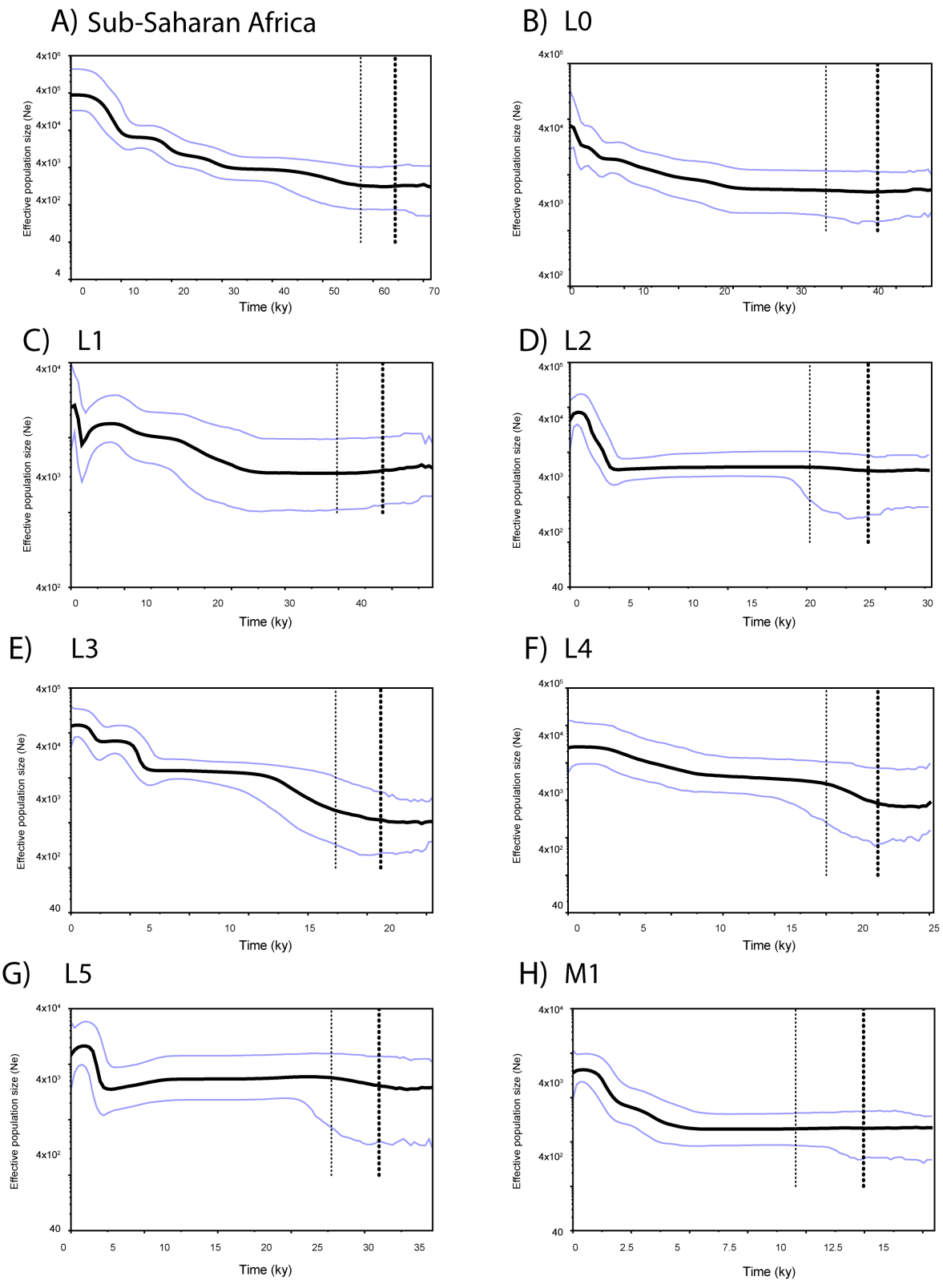
Table 3.4.2.1: Neutrality tests for mitochondrial complete coding region sequences (mitochondrial genome region between basepair positions 577-16023). Only three genes in three different haplogroups indicated significant values of NI; COIII (L0), ND5 (for L3) and COI (for N). NI scales with selection where values less than 1 indicate an excess of amino acid fixations, or positive selection and NI values greater than 1 indicate an excess of amino acid polymorphisms, or negative selection [245, 246]. Discussions on the implications of these results are found in **Appendix 10.**

[1]Neutrality index for 13 genes in each lineage, stars indicate significance levels.
[2]Lineage specific gene(s) with statistically significant Ni.
[3]Statistically significant Neutrality index for the three lineage specific genes.
NS (values without any stars P>0.05, * P<0.05, ** P<0.01, *** P<0.001)

3.4.3: Human demographic expansion related to mtDNA Lineages

Bayesian skyline plots (BSP) [267-270], graphical display of estimates of effective population sizes through time, were generated for a large dataset of complete human *mt*DNA sequences to estimate the relative size and timing of human *mt*DNA lineage expansions in Africa (which correspond to periods of human population expansions). Initial major expansions in African *mt*DNA lineages correspond to ~80-90 kya [273], a period before expansion of human population outside Africa and subsequent expansions of *mt*DNA lineages happened ~55, ~35, ~22 and ~12 kya **(Table 3.4.3.1, Figures 3.4.3.1a-h)**. The expansion that occurred 55 kya corresponds to a period of expansion of modern humans outside Africa [39].

Figures 3.4.3.1: Bayesian Skyline Plots (BSPs) of effective population size through time for (A) Haplogroups found in Sub-Saharan Africa (n=113) (B) haplogroup L0 (n=144), (C) haplogroup L1 (n=93), (D) haplogroup L2 (n=160) (E) haplogroup L3 (n=223) (F) haplogroup L4 (n=21) (G) haplogroup L5 (n=49) (H) haplogroup M1 (n=108). The bold black line represents the median posterior effective population size through time. The blue lines delimit the 95% highest posterior density for effective population size, accounting for uncertainty in the reconstructed phylogeny and substitution model parameters. Effective population size is plotted on the X axis and assumes a generation time of 25 years as used elsewhere in population genetic studies [159, 325-327]. The two black dotted vertical lines demarcate the 95% period of initial onset of population growth. Correspondence of these *mt*DNA lineages‟ expansion dates with Paleoclimatic conditions are discussed in **Appendix 14**.


A) Sub-Saharan Africa
B) L0
C) L1
D) L2
E) L3
F) L4
G) L5
H) M1
Effective population size (Ne)
Time (ky)

Haplogroups/haplotypes	n	Initial expansion events*	Later expansion events
Sub-Saharan African	122	62	55, 35, 22, 12
L0	144	53	32, 6, 2.5
L0a1	50	44	4.5
L0a2	17	45	12, 2.5
L0d1 & 2	27	45	21, 2
L0f	20	52	32
L1	93	53	35, 2
L1b	41	38	4
L1c	53	46	32, 14, 2
L2	160	34	5
L2a1	117	28	5.5
L3	223	24	24, 7, 2.5
L3a	13	42	24
L3b	36	18	6
L3d	22	17	12, 3
L3e	78	16	5.5
L3f	47	17	17, 8
L3h	20	22	22, 9
L3x	14	18	14
L4	21	27.5	13
L5	49	39	5
M1	108	24	12, 5
M1a	93	20	11, 4.5
U6	49	12	5

Table 3.4.3.1: Timing of major expansion events related to *mt*DNA haplogroups/haplotypes in kya. Mutation rate used is average of control and coding region rates corrected for time dependence of initial estimates due to saturation effect and purifying selection that is common in the *mt*DNA genome [272]. These dates correspond to time when the effective population size in the Bayesian Skyline Plots (BSPs) curve undergoes a sudden increase as depicted in **Figure 3.4.3.1** and **Appendix 14**. Atkinson *et al.,* [273] estimated periods of expansions from double the number of sequences used in this study for Sub-Saharan Africa, with their graph indicating initial demographic expansion at 80-90 kya [273]. *Initial timing of population growth as inferred from the haplotypes' BSP, corresponding to that is the mean of the 95% period of initial onset of population growth as demarcated by two dotted lines plotted in **Figure 3.4.3.1 (below)** and **Appendix 14.** Correspondence of these *mt*DNA lineages expansion dates with Paleoclimatic conditions are discussed in **Appendix 14**.

Chapter 4: Discussions

The aim of this study is to characterize autosomal, mitochondrial DNA and Y chromosome variation among diverse East African populations in order to reconstruct East African population history and to (1) test for correlations between genetic variation with both geography and language (2) to examine the phylogeographic pattern of variation of *mt*DNA and Y chromosome lineages and test for correlations with geography and language (3) to infer changes in effective population size and population range expansions and examine correlations with changes in climate and/or technology based on the archeological records and (4) to characterize genetic structure of populations in East Africa and to determine the influence of major historical events such as the spread of pastoralism and the Bantu expansion over the last 5 ky on patterns of genetic variation in the region. The data from the current study will also add to the growing knowledge on human population genetics that will ultimately help in deciphering modern humans' genetic history in Africa. To achieve these aims, Y chromosome SNP variation, *mt*DNA D-loop sequence and autosomal microsatellite variation in 1500 individuals from 55 East (Kenya, Tanzania, Sudan) and Central African (Cameroon & Chad) populations, who speak languages belonging to the four major African linguistic groupings (Khoisan, Nilo-Saharan, Afroasiatic and Niger-Kordofanian) were analyzed. In addition, whole genome *mt*DNA variations in a subset of 222 individuals with diverse *mt*DNA haplotypes were examined. Sequencing of 300-500 bp fragments of mitochondrial coding regions that have haplogroup/sub-haplogroup/haplotype diagnostic SNPs determined by complete sequencing were subsequently used to classify all the 1500 sequences, in cases of haplogroup/haplotype ambiguity or to confirm designations based solely upon rapidly

evolving HVI and HVII mutations. The data were compared with previously published *mt*DNA data from over 20,000 individuals from over 100 populations, and Y chromosome data from over 10,000 individuals from over 80 populations from other regions of Africa, the Near East, Central Asia and Europe.

Tests of neutrality were applied to complete mitochondrial sequences to check if selection might have influenced patterns of genetic diversity in the mitochondrial genomes. The *mt*DNA complete sequence data showed statistically significant negative values of Tajima's D and Fu & Li's D*, indicating the possibility that these negative values are due to population demographic expansions. Population phylogenetic analysis, MDS plots, regression analysis and Mantel's tests of genetic distance versus geographic and linguistic distances, STRUCTURE analysis and AMOVA analysis all showed genetic correlation with geography and language in East African populations. The *mt*DNA and Y chromosome lineage distributions seem to cluster geographically, and for some lineages, based on linguistic classification. The correlation between genetic distances with geography is stronger than it is with language classification. Additionally, correlation between genetic variation and geography/language is stronger for autosomal and Y chromosome lineage data than for *mt*DNA lineages.

The expected pattern of genetic variation under a model of female biased migration (meaning higher levels of female gene flow relative to male gene flow) is that we would observe higher levels of structure for Y chromosome and autosomal variation relative to *mt*DNA variation. Consistent with this expectation, previous human population genetic studies of Y chromosome, *mt*DNA, X chromosome and autosome genetic variation have also inferred female-biased migration patterns for modern humans [38,

211, 328-340]. Two common cultural practices may result in a female-biased migration pattern: polygyny and patrilocality. Several studies have independently shown sex-biased migration that are consistent with patrilocality in the populations studied; for example in Central Asia [328, 329, 339], Europe [330, 331], south east Asia [336] and sub-Saharan Africa [38, 340]. According to Wilkins and Marlowe [341], the observed female-biased migration patterns are due to the fact that the transition to a food-producing lifestyle 10 kya was probably accompanied by an increase in patrilocality [342]. Differential selective effects on the Y chromosome (and X chromosome) relative to *mt*DNA (and autosomes) have also been considered to explain the patterns discussed above [161, 333, 334, 338]. For example, a selective sweep will reduce the diversity on the Y chromosomes and, thus, could result in lower estimates of Ne and higher levels of genetic structure for Y chromosome variation. Additionally, background selection on the X chromosome [338] is expected to increase the effective population size of the X chromosome and result in lowered levels of population structure [338]. However, these studies indicate that a sex-biased migration model is a better explanation for the observed patterns of variation, even though the effect of selection could not be entirely ruled out [161, 333, 334].

Overall, studies of worldwide human variation using autosomal markers have shown three trends in summary statistics as a function of increasing geographic distance from East Africa: a decrease in genetic heterozygosity [8, 9, 343], an increase in linkage disequilibrium [344], and a decrease in frequency of the ancestral genetic alleles/lineages [41, 343]. Consistent with this observation, the results from the current study show that most of the ancestral *mt*DNA and Y chromosome lineages are found in East Africa **(Appendix 15)**. As discussed in **Appendix 9, 15** and below, East Africa appears to be the

cradle of modern humans and has served as a corridor of major migration events within and out of Africa. Here, I integrate the genetic data with data from archaeology, palaeobiology, linguistics and geology to shed light on East African population history and historic migration events, including the migration of pastoralists from Ethiopia and Sudan, and of Bantu speaking agriculturalists from Central/West Africa into East Africa.

4.1: Interactions among Holocene East African Hunter-Gatherers and Southern Cushitic Speakers

Ambrose [57] speculated that Late Stone Age (LSA) hunter-gatherer populations in East Africa, represented by makers of the Eburran industry in the Rift Valley and the "Wilton" industry and Kansyore ware in lowland Tanzania, spoke Khoisan languages [57]. Moreover, based on loanwords [115, 116, 345] some linguists [74, 115] argue that the Khoisan languages was spoken as far east and north as Somalia and the southern edge of Ethiopia, respectively. Therefore, it appears there were two "groups" of Holocene hunter-gatherers in East Africa from whom the current hunter-gatherer populations arose; that is, the makers of Eburran and "Wilton"/Kansyore wares, who might all have spoken Khoisan languages [57, 115].

The genetic signature of these two groups of hunter-gatherer populations might be reflected in the distribution patterns of genetic lineages in East Africa. The current study shows that there are both *mt*DNA and Y chromosome lineages that seem to be affiliated with East African hunter-gatherer populations. For example, the Y chromosome lineages most commonly observed among East African hunter-gatherers are predominantly derived from B haplogroups, specifically B2a and B2b. Eastern Cushitic speaking hunter-

gatherer populations, such as the Yaaku and Sanye have a high frequency of B2a, while the Khoisan speaking Hadza and Sandawe have a high frequency of B2b. Fully one half of the sampled Yaaku individuals carry the B2a haplotype. However, all individuals carry a unique variant of the B2a haplotype, B2a3 that is found at low frequency among several Nilo-Saharan populations, the Pokot, the Karamajong [346] and the Turkana, as well as among the El-Molo, another eastern Cushitic speaking hunter-gatherer population **(Table 3.3.2, Appendix 6a)**. The Turkana and Karamajong populations occupy geographic areas that overlap with both the El-Molo and Pokot, respectively, so the presence of the B2a3 haplotype in these two Nilo-Saharan populations may represent gene flow from the El-Molo or, alternatively, from hunter-gatherer populations that preceded them in the area. One third of the Sanye chromosomes analyzed in this study carry the ancestral form of the B2a haplotypes, B2a*. Other populations showing the B2a haplotype (B2a1 sub-haplotype) at high frequency are the western Nilotic speaking populations from Sudan, and a southern Nilotic speaking population, the Sengwer from Kenya. Even though there is no formal published literature on Sengwer history, Sengwer is listed among hunter-gatherer populations in Kenya (according to the World Directory of Minorities and indigenous peoples' site – http://www.minorityrights.org/3956/kenya/huntergatherers.html).

The frequency pattern of the B2a Y chromosome haplotypes in East African eastern Cushitic speaking hunter-gatherers is mirrored by the *mt*DNA haplotype, L3a, where a moderate to high proportion of the *mt*DNA lineages of Sanye, Yaaku and El-Molo are classified as L3a haplotypes. Two other groups, the eastern Nilotic speaking Samburu and the eastern Cushitic Orma have this haplotype at moderate frequency. The

sharing of the L3a *mt*DNA and Y chromosome B2a3 haplotypes among several eastern Cushitic speaking hunter-gatherer populations is reflected in the STRUCTURE analysis of Cushitic populations using autosomal markers at K = 9 **(Figure 3.2.1b)**, where the El-Molo and Yaaku form a cluster (yellow). Eastern Nilotic and eastern Cushitic speaking individuals show low levels of ancestry from this cluster, indicating admixture with these East African hunter-gatherers (Yaaku and El-Molo).

Among Khoisan speaking and pygmy populations, the ancestral form B2b haplotype (B2b*) is the most common. Haplotype B2b* is also observed at moderate frequency among southern Cushitic speaking populations and at low frequencies among other non-Khoisan East African hunter-gatherer and eastern Nilotic speaking populations. Network analysis shows that there are two large clusters of this haplotype **(Figure 3.3.4)**; one includes East African Khoisan populations, and the other includes virtually all other B2b* chromosomes from within East and Central Africa. The only non-Khoisan speaking East African hunter-gatherer population that does not carry this haplotype is the Boni. Considering that in the current study the B2b haplotype was not found in populations from Sudan (putative origin of Nilo-Saharan speakers, **Appendix 6aii)** nor in eastern Cushitic populations (who originated in Ethiopia), it may represent a signature of introgression from previous hunter-gatherer populations that existed in East Africa (Kenya and Tanzania) before the expansion of pastoralist southern Cushitic [110] and eastern Nilotic speaking populations [347]. This finding suggests that there has been Y chromosome gene flow from East African hunter-gatherers into expanding pastoralist populations. This observation is in contrast to what was observed for pygmy populations

that have had limited or no Y chromosome gene flow into neighboring Bantu speaking agriculturalist populations [340].

Frequency patterns for the L4 *mt*DNA lineage, specifically of the derived form, L4b2a2 found in East African hunter-gatherer populations (**Appendix 9 - A9.2.3**) and two other *mt*DNA lineages, L0d3 and L0f, support the argument of gene-flow between East African hunter-gatherers and pastoralist populations, specifically the southern Cushitic speaking pastoralist populations. The frequency profile for L4b2a2 clearly shows that most of the non-hunter-gatherer populations that carry the haplotype are those living in the vicinity of East African hunter-gatherer populations, and are likely to have had historical interactions **(Appendix 9 - A9.2.3)** [108, 113, 121, 122, 348] with them. However, L0f is found at highest frequency among the southern Cushitic speaking populations, followed by East African hunter-gatherer populations who have the haplotype at moderate frequencies **(Appendix 9)**. This observation is consistent with the assertion that the East African hunter gatherers had most extensive interaction with southern Cushitic speaking populations, who represent the Savannah pastoral traditions [110, 133]. Interestingly, consistent with the possibility of two ancestral "groups" of hunter-gatherer populations in eastern Africa, the Khoisan speaking and neighboring populations mainly have L0f1 sub-haplotypes while other east Africans hunter-gatherer and neighboring populations have the L0f2 and L0f3 sub-haplotypes. The L0d3 haplotype is found at low frequency in East Africa in the Sandawe, a Khoisan speaking population, the neighboring southern Cushitic speaking Burunge and other hunter-gatherer populations including the Akie (northern Tanzania) and Boni (South-eastern Kenya). This haplotype, which shares a common ancestry with the L0d1/2 lineages which

are predominantly observed among southern African Khoisan speaking populations, was previously found at low frequency among the eastern Nilotic Turkana (northern Kenya) [151, 152] and South-eastern Bantu speaking Ronga population (southern Mozambique) [185, 298]. Moreover, a recent study found that the *mt*DNA lineages which I observed to be most common observed among East African hunter-gatherer populations, L4b2a2 (7.1%) and L3a (4.8%), at low frequency among the Hutu populations from Rwanda [349]. That study also found the L4b2a2 haplotype at low frequency (1.7%) among the Bantu speaking Shona population from Zimbabwe [349]. These results indicate that East African hunter gatherers from northern Kenya to northern Tanzania share some *mt*DNA lineages, and that the area they occupied may have once extended into Rwanda [59].

Previous studies of *mt*DNA and Y chromosome markers have shown that hunter–gatherer groups typically have reduced genetic diversity and high frequencies of unique *mt*DNA and Y chromosome haplotypes relative to neighboring agriculturalist or pastoralist populations [6, 46, 183, 350-352]. Except for the Hadza, East African hunter-gatherers have several haplotypes at moderate to high frequencies (*mt*DNA L0f, L4, L3a and Y chromosome B2a and B2b) **(Appendix 6a, Appendix 6b)** but also have a low frequency of haplotypes that are mostly observed among neighboring agricultural and pastoralist communities, likely due to geneflow **(Appendix 6a, b)**. Despite this evidence of geneflow from neighboring groups, the genetic diversity of hunter-gatherers in East Africa is generally lower than that of other populations in East Africa that practice other subsistence methods **(Appendix 7a**, **Figure S2b** [4]**),** an observation that is consistent with previous studies [6, 46, 183, 350-352].

Besides the Boni and Sengwer, all current East African hunter-gatherers share common haplotypes (*mt*DNA L4, L3a, L0d3 and Y chromosome B2a and B2b), suggesting that the current East African hunter-gatherers might represent remnants of the previous pre-Holocene and early Holocene population in the region. The Boni and the Sengwer appear to have had substantial levels of gene flow from other neighboring populations that appear to have the hunter-gatherer "genetic signature". In fact, the Orma and Kalenjin populations that neighbor the Boni and the Sengwer, respectively, carry hunter-gatherer haplotypes at moderate frequencies indicating possible "reverse" gene flow. Alternatively, the Boni and Sengwer might have initially been pastoral populations that took up a hunter-gatherer lifestyle after losing their herds due to raids or natural calamities such as drought or disease [121, 122]. This latter possibility is consistent with the conclusion of a recent study of the origin of a hunter-gatherer population in Thailand, where mitochondrial control region [350] and SNP genome-wide analysis [353] have shown an origin of the Mlabri, a Khmuic Austro-Asiatic language speaking hunter-gatherer population, originating from an extreme founder event from an agricultural group that speaks a closely related language [353].

The southern Cushitic populations included in this study show differing levels of interaction with populations from other linguistic groupings, in addition to the extensive interaction with the East African Holocene hunter-gatherers **(discussed above)**. In fact, the Gorowa (Fyome), Iraqw and Burunge have all had long historical interactions with Nilotic speaking Datog and exhibit similar rituals; for example the removal of two front central teeth in the Gorowa, which is a practice traditionally associated with Nilotic-speaking populations [354]). Consistent with this socio-cultural evidence, all the southern

Cushitic populations besides the Mbugu have a low frequency of Y chromosome lineages commonly observed among the Nilo-Saharan speakers (B2a1, E2a and A3b2), probably obtained through admixture/geneflow. The southern Cushitic speaking populations also carry Y chromosome lineages commonly observed among the Bantu speakers, E3a. For example, 40% of the Mbugu analyzed have the E3a lineages. The remaining 60% of the Mbugu chromosomes consist of haplotype E3b6, which is common among the southern Cushitic population and is associated with the expansion of pastoralism in East and South Africa [355] **(discussed below in section 4.4)**. This observation is consistent with linguistic evidence suggesting that the Mbugu speak a mixed language that exhibit Cushitic lexicon [102] and a Bantu grammatical structure [356]. Tucker and Bryan [356] speculate that the Mbugu have had a long period of contact with the Pare and historical interactions with other Bantu speakers.

Overall, even though not absolute, it appears that the distribution pattern of Y chromosome and *mt*DNA lineages reflect two ancestral hunter-gatherer populations in East Africa: the hunter-gatherers in northern Tanzania and southeastern Kenya who carry *mt*DNA L0f1 and L42b2a2 subhaplotypes and Y chromosome B2b haplotypes at moderate to high frequencies, and the rest of East African hunter-gatherer populations who carry *mt*DNA L3a haplotype and L0f2-3 subhaplotypes and Y chromosome B2a haplotype at moderate to high frequencies. Based on the current results, there has been extensive genetic interaction between the East African hunter-gatherers and southern Cushitic speaking populations. There also has been some admixture with later immigrants into the region including the Nilo-Saharan and Bantu speaking populations. Consistent with the findings in this study which suggest geneflow from southern Cushitic speakers

into east African hunter-gatherer populations, there is historical evidence of some southern Cushitic populations like the Alagwa **(not analyzed in this study)**, living among the Rangi [357] (Bantu speakers) and Sandawe [348, 354] (Khoisan hunter-gatherers), and adopting their traditions.

4.2: Genetic history of Nilo-Saharan speaking populations

Prior archaeological, linguistic and paleoclimatic studies suggest that southern Sudan was the site of origin and migration of Nilo-Saharan speaking populations into East and West/Central Africa. These data indicate that there was an initial northward shift of the inter-tropical convergence zone (ITCZ) in North Africa around 10 kya that increased precipitation in North Africa and led to a shift in desert margins northward relative to the tropics [358] (meaning shrinking of the Sahara desert in the region that borders the tropics). The reduction in area covered by the Sahara allowed the movement of hunter-gatherers northward from Central Sudan. This movement was associated with "wavy line" and "dotted wavy line" motif pottery traditions and the cultural "Khartoum Neolithic" (also called "Khartoum Mesolithic") traditions which spread as far north as Middle Egypt [84, 359]. According to Ehret [79, 96, 97], this group constitutes the "northern Sudanic" **(Figure 1.2.2)** speaking population which later split into several sub-groups, among them the eastern Sudanic (the family that Nilotic speakers of East Africa belong to). Subsets of this "northern Sudanic" group **(Figure 1.2.2)** expanded west along the Sahara-Sahel belt [79, 96, 97], and might be the makers of the "Khartoum Mesolithic" traditions observed in West Africa ([360] and references therein). Some of the Nilo-Saharan speaking populations that reached southern Egypt moved back south

gradually after an initial shift to cultivation in the Egyptian Nile Valley around 6 kya and reached lower Nubia by 4 kya [84, 359].

Previous cranial studies showed that the Egyptian dynasty was founded by an indigenous population in Upper Egypt (which presumably spoke Afroasiatic languages) which later expanded into Lower Egypt and grew in part due to immigration [361-363]. These cranial studies suggest that the immigrants might have been the Nilo-Saharan that initially expanded into middle Egypt. Nilo-Saharan speaking populations that moved into northern Sudan subsequently set up kingdoms centered in Kerma/Dongola (in Modern Sudan) 3.5 - 4.5 kya which was in rivalry with the Pharaonic kingdom in the north (Egypt). Edwards [359] further speculates that the kingdom at Kerma was formed emphasizing livestock, as evinced by burial sites suggesting a strong pastoral component in subsistence. Subsequent fall of Kerma (due to the rivalry with Pharaonic kingdom) might have been a catalyst for nomadic pastoralism among the Nilotic speakers [359]. Moreover, the later kingdom at Meroe further south in Sudan (1.65 – 2.8 kya) was also established by Nilo-Saharan speakers [364], an assertion that is supported by a study of cranial traits in six Nubian groups, that concluded that the Meroites were Nilo-Saharans [71, 78, 365]. The Nilotic populations that settled in Kenya and Tanzania might have originated from the southern portion of this larger pastoralist group, that initially expanded through Southwestern Ethiopia [102] and had extensive interaction with the eastern Cushitic groups before subsequently migrating south into Kenya and Tanzania [87]. If these scenarios are correct, the Nilo-Saharan populations described above may originally migrated northward 8 kya from Central Sudan and then later migrated and settled in southern Sudan and East Africa. The Nilo-Saharan populations that speak

languages belonging to the Nilotic sub-subfamily are restricted to East Africa, southern Sudan, Uganda, Kenya and northern Tanzania [117, 118].

All the marker systems used in the current study show that Nilo-Saharan speaking populations can be distinguished from populations from other linguistic groups (autosomal - **Figure 3.3.1,** Y chromosome **- Appendix 6a,** and *mt*DNA **- Appendix 6b**). The Y chromosome lineages that seem to reflect a signature of Nilo-Saharan expansion are A3b2, B2a1 and E2a **(Table 3.3.2)**. The *mt*DNA lineages that seem to reflect a signature of the Nilo-Saharan expansion are L3h (**Appendix 9 Figure A9.2.21)** and L5 **(Appendix 9 Figure A9.2.26, Appendix 6b)**. Interestingly, inferred expansion times of the L3h and L5 *mt*DNA lineages correspond to 18-20 kya and 5 kya, respectively (**Figure 3.6.1, Table 3.6.1, Appendix14).** These two time periods correspond to the proposed time of origin of the Nilo-Saharan languages based on linguistic data 18-20 kya [71] and geographic expansion of populations speaking languages belonging to the Nilotic branch of the Nilo-Saharan language family ~5 kya from Sudan [71, 83]. An early expansion of Nilo-Saharan speaking populations is also supported by archeological data indicating the presence of populations that practiced an aquatic economy and culture associated with Nilo-Saharan speakers that extended across the Sahel up to the middle Nile Valley by 9 kya [84].

In East Africa, the Y chromosome lineages that seem to reflect a signature of Nilo-Saharan expansion, have different frequency patterns within the western Nilotic, southern Nilotic and eastern Nilotic groups. Western Nilotic speakers carry the A3b2 and B2a1 haplotypes at high frequency whereas the southern Nilotic speakers carry the E2a haplotype at high frequency. All the chromosomes that carry the Y chromosome A3b2

haplotype in Kenya and Tanzania cluster together based on STR analysis, with the exception of the A3b2 lineages from western Nilotic speaking Luo, which cluster with Y chromosome lineages from southern Sudanese and Central African populations **(Figure 3.3.2)**. This observation indicates that southern and eastern Nilotic speaking populations have had a stronger influence on the East African (Kenyan and Tanzanian) Y chromosome genetic landscape compared to the western Nilotic speakers, who are more recent migrants into the area [88, 89]. By contrast, the Y chromosome E2a haplotype is mostly restricted to the southern Nilotic speaking populations, reflecting a shared common ancestry of these populations (**Appendix 6a, Appendix 9 Figure A9.1.4)**.

There are several *mt*DNA and Y chromosome lineages that are restricted to East African populations. The *mt*DNA L5 haplotype is absent in Nilo-Saharan speakers outside northeastern Africa, and is observed predominantly in East Africa **(Appendix 6b)**. The other *mt*DNA haplotype associated with the Nilo-Saharan expansion, is the L3h haplotype. L3h lineages in East African populations mainly belong to the L3h1a2 sub-haplotype [219, 349, 366], whereas all the non-East African individuals who have the L3h haplotype belong to the L3h1b sub-haplotype [151, 152, 204, 297, 305, 307, 309-311, 313-315, 317, 318, 351, 367, 368] **(Appendix 6b, Appendix 8, Appendix 9 Figure A9.2.20).** So these lineages (*mt*DNA L5, L3h1a2, and Y chromosome E2a) appears to reflect the migration of Nilotic speaking populations from a southern Sudan homeland into Kenya and Ethiopia, within the past 3kya [65, 87] **(Figure 4.4.1)**.

Consistent with linguistic and archaeological evidence of initial expansion of Nilo-Saharan speakers into Egypt, the *mt*DNA and Y chromosome lineages that might reflect the expansion of the Nilo-Saharan populations across Africa are observed at low

frequency among the Egyptians [305, 369]. Specifically, the *mt*DNA L5 haplotype was observed at low frequency in one of the putative modern descendant populations of the Pharaonic era, namely the Gurna of the Nile valley **(Appendix 6b)** [369]. Moreover, the *mt*DNA L5 haplotype was also found among populations from northern Egypt at low frequency **(Appendix 6b)** [369]. *Mt*DNA haplotype L3h was observed in Egyptian populations from northern Egypt at low frequency [305], but was not observed among the Gurna populations [369]. Hassan *et al.,* [370] also reported moderate frequency of the Y chromosomes B haplogroup in the Coptic population from northern Sudan **(Appendix 6b)**. The Coptic language is a later version of the ancient Egyptian language [371, 372] presumably spoken 200 AD until the Middle Ages, by descendants of the Pharaonic Egyptians (a Coptic dialect is still used as a liturgical language in the Coptic Church). The B haplotype to which these Sudanese populations belong was not determined in that study [370]. However, only the B2a1 haplotype has been shown to be present in the Nuer, a Sudanese western Nilotic speaking population included in the current study. As observed in the current study, all the Nilo-Saharan populations included in the study by Hassan *et al.,* [370] carry the A3b2 haplotype and the B haplogroup at moderate to high frequency. Considering that all other B haplotypes were found only among the pygmies from central Africa, South African Khoisan and East African hunter-gatherers, and that the B2a1 haplotype is the predominant lineage observed in Nilo-Saharan speakers, it seems likely that the B lineages in Hassan *et al.'s* study [290] can be classified as B2a1. This result indicates that either there was genetic contribution by Nilo-Saharan speakers into the Egyptian Coptic population or that perhaps the ancestral Coptic population shares some genetic ancestry with Nilo-Saharan speakers. Most of the other non-Nilo-Saharan

speaking populations that carry the Y chromosome B2a1 haplotype, usually at low frequency, are from Central and East Africa and thus may have acquired the lineage through gene flow. The Niger-Kordofanian populations outside East Africa that carry this haplotype currently live in the vicinity of Nilo-Saharan speaking and pygmy populations, or might have had historical contacts with Nilo-Saharan speaking populations. For example, the Fali live in the northern part of Cameroon, an area which is ethnically heterogeneous and might have been historically influenced by Nilo-Saharans from western Sudan, while the Ewondo and the Cabindans neighbor pygmy populations of Central Africa. South-African Bantu groups, including the Zulu, Shona and Xhosa, carry the B2a1 haplotype at moderate frequency and have forged a strong agro-pastoralist lifestyle. Thus, their ancestors might have interacted with Nilotic speakers in East Africa during the Bantu expansion through East Africa into southern Africa **(discussed in section 4.4)**.

In the STRUCTURE analysis of Nilotic populations based on autosomal data, at K=9 individuals from populations speaking the southern, eastern, and western Nilotic languages can be distinguished. The STRUCTURE results from East Africa show that the eastern Nilotic populations have experienced substantial genetic admixture with the eastern Cushitic speaking populations **(Figure 3.2.1)**, with the Cushitic component higher in eastern Nilotic speakers relative to other Nilotic speakers. However, the Turkana, an eastern Nilotic speaking population, seems to cluster with western Nilotic speakers **(Figure 3.2.1)**. This clustering is consistent with the fact that nearly a third of the Turkana Y chromosome analyzed in this study belong to the A3b2 haplotype **(Appendix 6a)**, which is prevalent in western Nilotic speakers. Moreover, based on their geographic

location in northwestern Kenya, the Turkana might have had extensive interaction with western Nilotic speakers that inhabit the southern part of Sudan. I also observed a strong Cushitic genetic influence in southern Nilotic speakers from Tanzania, the Datog. Moreover, genetic influence from Bantu speakers among the Nilotic populations in East Africa is apparent, mostly in Tanzanian populations and other eastern Nilotic speakers. The southern Nilotic speakers in Kenya that seem to have experienced Bantu admixture are the Sabaot and the Nandi **(Figure 3.2.1)**, who might have recently had interaction with neighboring Bantu speakers from the western part of Kenya.

In fact, in the current study I observe high levels of mixed ancestry in many of the East African populations. In several cases, populations may consist of composites of individuals with mixed ethnicity who have adopted one cultural identity (**Appendix 9** - as previously argued for the Gabra of Kenya [123] and Ndebele of Zambia and Zimbabwe [373]). This point is exemplified by many populations that exist on the periphery of the contiguous geographical areas occupied by a linguistic family, or isolated from other populations of the same linguistic family, which have absorbed or been absorbed by the other populations that they encountered during their expansion. For example, this appears to be the case of the Luo and Datog who, despite speaking Nilo-Saharan languages, appear genetically like other Bantu and southern Cushitic speakers, respectively. Ambrose [374] considers the Datog to be the southernmost survivors of expansion of Elmenteitan Neolithic populations; he posits that they should have absorbed pre-existing southern Cushitic populations, consistent with my observations based on genetic data. The proto-Datog were cut off from the rest of the southern Nilotic groups by a Maasai expansion that probably began about 1.2 kya [374].

According to the STRUCTURE analyses of the Marshfield autosomal dataset, the western Nilotic speakers have a larger proportion of Bantu ancestry, with the Luo population showing far higher Bantu than Nilotic ancestry **(Figure 3.2.1)**. There is also evidence for sex-biased gene flow in the Luo population. Specifically, the Luo population is genetically similar to other Nilotic-speaking populations based on their Y chromosome lineages (two third of the chromosomes belong to A3b2). However, the Luo *mt*DNA lineages cluster with Bantu-speaking populations (two thirds of the lineages are those mainly observed among Bantu speaking populations L3b, L3d and L3e) **(Appendix 6b)**. These results suggest that the paternal lineage of the Luo is of Nilotic ancestry whereas their maternal lineage is of Bantu ancestry. The observation that the Luo have likely had substantial *mt*DNA gene flow from Bantu-speaking populations can be attributed to a combination of three factors that are related to their migration route and cultural practices: 1) during the expansion from southern Sudan their migration route passed through an area predominantly occupied by Bantu speakers [88, 89], 2) marriage of Luo men to multiple women (in some cases up to a dozen), and 3) the Luo tendency to assimilate other ethnic groups they encounter [88, 375]. The latter point is exemplified by the Luo Basuba, an ethnic group closer to the Kuria (a Bantu group) in terms of language and culture but have adopted a Luo ethnic identity [376, 377].

Overall, the genetic evidence shows that there has been substantial gene flow from Cushitic and Bantu populations into the eastern Nilotic, but limited gene flow into the southern Nilotic and western Nilotic speaking population, save for the Datog and Luo, respectively. Despite historical evidence that shows southern Nilotic speakers migrating into Kenya and Tanzania prior to other Nilotic speakers [87], eastern Nilotic

populations seem to have had more genetic admixture with populations that belong to other linguistic groupings (Cushitic and Bantu speakers) compared to southern Nilotic speakers. Consistent with the historical account, the western Nilotic populations have had limited geneflow with Kenyan and Tanzanian populations except for the Luo [89] who appear to have admixed considerably with the Bantu speakers.

4.3: History of Afroasiatic speakers

Analyses of *mt*DNA and Y chromosome lineages are informative for inferring the population history of Afroasiatic speakers. The genetic data from the current study were used, together with archaeological and linguistic evidence, to understand the genetic history of Afroasiatic speakers in Africa.

4.3.1: Chadic speaking populations

The Chadic languages belong to the Afroasiatic language phylum and according to recent estimates, comprises 195 languages [70] spoken by populations found in Chad (East Chadic), northern Cameroon (Central Chadic - Biu-Mandara and Masa), southern Niger and northern Nigeria (West Chadic). According to Ehret [101], the proto-Afroasiatic language phylum split into two groups; Omotic speakers and a group that constitute all the others referred to as "Erythraean". The latter divides into Cushitic and "North Erythraean", which itself consists of Chadic, Egyptian, Berber, and Semitic. A subsequent split gave rise to proto-Chado-Berber and "Boreafrasian" languages; this last group split into Egyptian and Semitic languages [96, 378]. According to this model, Chadic languages share a more common recent ancestry with the Berber languages than

with any of the other four Afroasiatic branches; Semitic, Omotic, Cushitic, and Egyptian [96, 378]. Chadic languages show great internal differences [379] and are thought to have spread west/south from a putative center of Afroasiatic expansion in the region from the Nile Valley to the Ethiopian highlands [101, 380] about 11 kya [381].

STRUCTURE analysis of autosomal markers genotyped in a larger African population dataset [4], as well as results from the current study **(Figure 3.2.1)**, show that Chadic speaking populations are genetically more similar to Nilo-Saharan speaking populations than to other African Afroasiatic speaking populations, with some Niger-Kordofanian admixture. Considering that linguistically, the Chadic classification as a language branch of the Afroasiatic phylum [382] is indisputable, the Nilo-Saharan and Niger-Kordofanian ancestry among Chadic speakers observed in the STRUCTURE analysis is likely the result of extensive historical admixture between proto-Chadic speaking populations with Nilo-Saharan and later Niger-Kordofanian groups in the last 8 ky [380]. Besides the presence of both Y chromosome and *mt*DNA lineages commonly observed among the Niger-Kordofanian and Nilo-Saharan speakers **(Figure 3.3.2, Figure 3.4.1, Appendix 6),** I also observe genetic evidence for some shared ancestry with other Afroasiatic speakers, which is summarized as follows;

(i) The *mt*DNA L3f3 haplotype is observed predominantly among the Chadic and East African Cushitic speakers **(Appendix 8)**. This haplotype has also been observed at low frequency among the Sierra Leoneans [317], Moroccans [309], Tunisians [309] and Nigerians [151, 152] (areas possibly influenced genetically by proto-Chado-Berber populations).

(ii) The inferred dates of demographic expansion associated with the *mt*DNA L3f lineage are 13-15 kya and 6-8 kya (**Table 3.6.1, Appendix 14**). These two periods possibly correspond to the expansion of Afroasiatic speakers [101] from the Nile Valley/northern fringe of the Ethiopian highlands 14 kya, and initial dispersal of pastoralist populations in North Africa 8 kya [380], respectively.

(iii) The Y chromosome haplotype inferred here to be a reflection of the signature of Afroasiatic expansion, E3b (also speculated in Lancaster [383]), was either observed at low frequencies or was absent among the Chadic speaking populations. Moreover, the E3b1 clade that is inferred here to be associated with Cushitic expansions, E3b1a1 (E-M78γ [153]), is found at low frequencies only among a subset of Chadic speakers (the Giziga and Mandara **(Table 3.3.2, Appendix 6a)**, but is absent in 33 Hausa male individuals from Nigeria **(data not shown)**. However, the E3b haplotypes were observed at low frequencies in other Chadic speakers in previous studies: Ouldeme [164], Hausa [290, 384] and Podokwo [38] **(Appendix 6a)**.

(iv) The *mt*DNA L2a1*g* lineage, found mostly among the Cushitic speakers in this study, was found at low frequencies in the Chadic speakers **(Appendix 8)**. Network analysis based on *mt*DNA control region sequences indicates that the Chadic speaking individuals that have the L2a1g lineage clustered with Cushitic speakers **(Figure 3.4.4)**.

(v) The *mt*DNA L3e5 sub-haplotype is observed predominantly among the Chadic speaking populations and the Berbers (and Arabs from North African countries who might have gotten this haplotype due to gene flow from the Berbers since this haplotype is not observed in the Middle East) **(Appendix 9 Table A9.2.2, Figure A9.2.15f)**. This observation is consistent with the linguistic evidence that show that Chadic languages share a more common recent ancestry with the Berbers than with languages belonging to any of the other four Afroasiatic branches [96, 378].

(vi) Haplotype U6, which is mainly observed among the Berber speakers and Northwestern African Arabic populations, is observed at low frequencies in East African populations and Chadic populations **(Appendix 9)**.

Previously, Ehret [380] suggested that the proto-Chadic were originally Afroasiatic speakers that occupied the northern part of Central Saharan Africa around 8 kya and subsequently moved south and settled in Chad after changing their subsistence pattern to pastoralism. During this expansion, Ehret [380] proposes that they mixed with several Nilo-Saharan speaking populations, borrowed Nilo-Saharan words into their Chadic languages and picked up agriculture. Based on cognate vocabularies for livestock terms that are shared between Cushitic and Chadic speakers, Blench [382] also suggested that proto-Chadic speakers are essentially "Cushitic pastoralists" that wandered westward along what he calls the "inter-Saharan corridor". This claim is supported by the presence of livestock terms in Nilo-Saharan languages spoken in the geographical region between the Nile and Lake Chad that are similar to Cushitic-Chadic terms [382]. The suggested

corridor of migration is the area that covers modern Sudan and Chad, which might have been occupied by Nilo-Saharan populations. Furthermore, a study of Hausa languages (West Chadic) found similarities in some vowels between East Chadic, Hausa languages and Cushitic languages [385]. Therefore, despite the result from autosomal data indicating a large proportion of Chadic speakers' ancestry being Nilo-Saharan (and some Niger-Kordofanian) [4], there are genetic signatures of Afroasiatic based on analyses of Y chromosome and *mt*DNA lineages in this and previous studies [38, 96, 164, 290, 309, 378, 384].

4.3.2: Cushitic speaking populations

With the exception of the Beja languages which are classified as northern Cushitic and spoken in northern Sudan up to southern Egypt, all the Cushitic languages are mainly spoken in eastern Africa (Ethiopia, Somalia, Kenya and Tanzania). The Cushitic branch is considered the second oldest Afroasiatic language branch after the Omotic (formerly Western Cushitic) branch [101]. In the current study, populations speaking the southern and eastern Cushitic languages **(Appendix 1)** were analyzed and the result obtained compared with published data of central and western Cushitic speakers, and other Afroasiatic speaking populations (Semitic, Chadic and Egyptian) **(Appendix 6)** in order to reconstruct the genetic history of the Cushitic and Afroasiatic speaking populations in general.

Based on patterns of Y chromosome microsatellite diversity, a previous study suggested that the Y chromosome E3b1 lineage originated in eastern Africa (which exhibited the highest diversity levels) ~23.2 kya [153]. This and other studies have shown

that the E3b1 family of haplotypes shows a restricted geographic distribution with E3b1a* (E-M78β) found only in northern/northwestern Africa, E3b1a2 (E-M78α) found only in Europe and E3b1a1b (E-M78γ) found only in East Africa [153, 386] **(Table 3.3.2, Appendix 6a)**. The fourth sub-haplotype in the E3b1 family, E3b1a3 (E-M78δ) is widely distributed at low frequency in Europe, southwest Asia and Africa [153, 386] **(Table 3.3.2, Appendix 6a)**. Based on its widespread distribution, Cruciani *et al.,* [153], have speculated that E3b1a3, and the European specific haplotype (E3b1a2), reflect the first dispersal(s) of E3b1 carrying populations from eastern Africa into northern Africa and the Near East around 14.7±2.7 kya. The age estimate for the E3b1a3 and E3b1a1b lineages in this study, based on a larger dataset that includes East Africans **(Table. 3.3.1)**, are 14.3±2.9 kya and 14.8±4.2, respectively which concurs with Cruciani *et al.*'s [153] estimate. This time period corresponds to the time when modern humans changed subsistence patterns in North Africa and the Near East, during the late Paleolithic and Neolithic times, and began exploiting more of the food resources provided by the seeds of wild grasses, corms, roots and other plant material [387]. This change in subsistence pattern is related to "intensive grass collection" that is taken as evidence for an initial attempt at domestication and transition to a farming lifestyle [101] and has been archeologically shown to have started earlier in North Africa relative to the Near East [388]. For example, there is evidence of initial exploitation of plant and grass resources in southern Egypt at Wadi Kubbaniya sites [388], 20-16 kya [389]. The Wadi Kubbaniyan tradition, even though not associated with a fully sedentary lifestyle, arose nearly 4-8 ky earlier than the well known Neolithic Natufian culture that began in the southern Levant [390]) in the Near East ~12.5 kya [391], and subsequently expanded across Mesopotamia

~11 kya [101]. In fact, there is archaeological evidence that shows contacts between people from the Nile Valley and the Levant, namely North African derived lithic chipping techniques found in the Mushabians (an industry associated with people in southern Levant 14-12 kya), which might have been a precursor to emergence of the Natufian culture [392, 393].

Olszewski [394], argues that the available archaeological evidence reflect a collecting strategy used by both the Wadi Kubbaniya of the Nile Valley and Natufian populations in the Mediterranean, who exploited a wide variety of available plant foods, and who seem to have decreased their dependence on animal protein. Moreover, cereals/grasses were just part of this wider strategy, but not the exclusive source, nor even a major constituent of the diet [394]. Ehret [395] argued that the archaeological data described above and the distribution of words related to grains in eastern and northeastern Africa and the Middle East indicate an expansion of populations speaking Afroasiatic languages from northeastern Africa into the Near East ~15 kya. This proposed time of Afroasiatic expansion also coincides with the beginning of warmer moist conditions 14 kya after a long dry period that started during the Last Glacial Maximum (LGM) 23-24 kya [393]. From 25 kya to the early Holocene (10 kya), there is evidence of continued occupation in northern Africa, specifically lower Nubia and Upper Egypt, with more than five cultural traditions reported ([393] and references therein).

The Y chromosome E3b1 haplotype may therefore reflect a signature of Afroasiatic expansion from northeastern Africa into central Africa, northwestern Africa and the Near East. The expansion of speakers of Afroasiatic languages from an East African origin may also explain the current distribution pattern of *mt*DNA haplotype M1

in Africa and the Near East. Based on data from extensive sampling in Europe and the Near East both Olivieri *et al.,* [222] and Gonzalez *et al.,* [396] argued that M1 originated in southwest Asia because of a similar distribution pattern of the *mt*DNA U6 haplotype (haplogroup N haplotype), which has a similar TMRCA as M1a and also appears to have originated in southwest Asia [222, 396]. However, results from the current study indicate that the M1 haplotype is more likely African in origin.

Firstly, the *mt*DNA M1 haplotype, specifically the M1a lineage is present at high frequency in the Horn of Africa and appears to be observed predominantly among Northeast Africans **(Table 3.4.2, Table 3.4.3, Appendix 17 Table A17.3)**. In fact, even with data from the above mentioned studies which include a few East African populations [396], the highest frequency and diversity of M1a appears to be in eastern Africa [396]. With a better sampling from East Africa in the current study, the frequency maximum for M1a is clearly centered in East Africa **(Appendix 9 Figure A9.2.21)**, where it is mostly observed among Cushitic speaking populations **(Table 3.4.2, Table 3.4.3, Appendix 17 Table A17.3)**. M1a contains nine sub-branches (M1a1-9), with two new sub-branches identified in this study, in addition to the seven identified in the previous study [222] **(Appendix 9)**. All the sub-branches are found in East Africa with at least five being nearly exclusive to eastern Africa (M1a4-9). Only one branch of M1a (M1a3) seems to be observed exclusively outside Africa. The M1a1 lineage is the most diverse sub-branch in the M1a haplotype, with at least 13 M1a1 clades. Africa is home to 11 of the 13 clades, with nine distributed almost exclusively in East Africa. The M1a2 lineage has three clades, two of which are mostly African **(Appendix 9)**. In Olivieri *et al.'s* study [222] they report elevated frequency of the M1a haplotype in two Egyptian

populations, Gurna and Adaima, a fact they use to support their contention. However, all the M1a lineages observed in Gurna [369] and other Egyptian populations belong to the M1a1 sub-haplotype (because they have all undergone a transition mutation at nucleotide position 16359 in addition to the M1 defining mutations) [154, 228, 305, 314, 369]. Thus the M1a haplotype might have reached an elevated frequency in these populations (Egyptians) as a result of genetic drift.

Second, the fact that previous studies did not observe presence of M1b (the sister clade to M1a lineages) in African populations was cited by Olivieri *et al.,* [222] to argue for a Eurasian origin of the M lineages (Olivieri *et al.,* [222] reported M1b as being present only in the Mediterranean area). However the M1b lineage has been reported in later studies to be present in most of the whole geographical range where M1a was observed; East Africa [16, 397], Southern Arabia [312, 313], North Africa [171, 309, 398-400] and Northwest Africa [309, 401, 402], and Gonzalez *et al.,* [396] who labeled it as M1c).

Third, Gonzalez *et al.,* [381] and Olivieri *et al.,* [221] studies used coalescence ages for M1(a) lineages to infer back migration from Eurasian to Africa 30 kya [222, 396]. Inference in the current study concurs with Olivieri *et al.'s* [222] and Gonzalez *et al.'s* [396], on ancient divergence of the M1 lineage from other M lineages, observed today predominantly in Indian sub-continent [7, 320, 321], aboriginal populations from Southeast Asia [7, 322, 323] and the South West Pacific [324]. However, the conclusion here differs in detail of the pattern and timing of the expansion associated with the M1 lineage. The inferred time of demographic expansion of the M1 lineage in the current study based on analysis of ~100 complete *mt*DNA sequences [270, 271, 275, 403], is

estimated to be ~13 kya **(Figure 3.6.1, Table 3.6.1)**, a time period that corresponds to the putative age of the migration of Afroasiatic speaking populations across northern Africa, eastern Africa and the Near East [395]. This time estimate is also comparable to the median coalescent ages of M1a and M1b sub-haplotypes **(Appendix 9)**. The demographic expansion associated with M1a seems to have begun around 8-10 kya **(Appendix 14),** a time period that corresponds with the proposed time of the major expansion and migration of Cushitic speakers from northeastern to eastern Africa. Thus, the migration that conforms to the inferred time of expansion of M1 lineages seems to be more recent (10-20 kya) than that inferred by Olivieri *et al.,* [222] (45 kya) and Gonzalez *et al.,* [396] (20-30 kya).

Fourth, the *mt*DNA U6 haplotype which has been used to support previous studies [222, 396] contention of back-migration from Eurasia to Africa, has nearly the same distribution pattern as two related Y chromosome lineages that seem to have expanded in north/northwestern Africa (among the Berber speakers), E3b2 and E3b1a* [153] **(Appendix 9 Figure A9.2.22, Table A9.2.3)**. Cruciani *et al.,* [153] argued that after the initial expansion of E3b1 lineages from northeastern Africa about 14 kya, which is associated with the Afroasiatic expansion, the northwest African lineages subsequently, expanded about 5 kya. The major demographic expansion associated with *mt*DNA haplotype U6 in the current study **(Appendix 14 Figure A14.2p),** ~12.5 kya and later expansion around 5 kya is consistent with Cruciani *et al.'s* [153] inferences based on Y chromosome data.

Fifth, other *mt*DNA haplotypes with widespread distributions but at low frequency in the same geographic regions in which M1a is observed are HV1, I and N1a

(Appendix 12). The frequency pattern for these *mt*DNA lineages (HV1, I and N1a) might have been subject to the same migration events that are associated with Afroasiatic and later Neolithic expansion events. Previously only N1a (among HV1, I and N1a) has been associated with the Neolithic expansion in the Near East and Europe [404, 405]. Haak *et al.,* [404] and Bramanti *et al.,* [406] analyzed *mt*DNA from 4.3 – 15 kyo skeletal remains representing Central and northern European post-Last Glacial Maximum (LGM) hunter-gatherers and putative early European farmers and concluded that the first farmers into Europe were not the descendants of local hunter-gatherers, but rather, immigrated into Central Europe from southwestern Asia at the onset of the Neolithic about 7.5 kya [404, 406]. The predominant *mt*DNA haplotypes found in the ancient samples from Central Europe (mainly N1a (25%), and also H, HV, J, K, T, V, and U3 types at low frequencies) are found in the extant European populations at lower frequencies, and may have originated in southwestern Asia [404, 406]. Consistent with this conclusion, a recent study of DNA polymorphisms in an 8 kb intronic segment flanking exon 44 (*dys44*) of the human dystrophin gene on Xp21 in European populations [407] concluded that migration by Neolithic farmers from the Near East has shaped the contemporary European gene pool [407]. Besides detecting an underlying 'ancient' east–west cline across the Eurasian continent, which they inferred as the signature of an initial out-of-Africa migration event through the northern route, they also detected a signature of recent migration into Europe that they concluded happened during the Neolithic era from the Near East [407].

In addition, the Neolithic expansion from the Near East 7 – 10 kya also influenced the genetic landscape of Central Asia.. Recent studies of Uyghurs [408, 409], a Turkic

speaking Central Asian population, and other Central Asian populations [409] showed mixed European and East Asian ancestry [408, 409]. Moreover, recent studies of Indian populations using autosomal markers [410] indicate that Northern Indians are genetically close to Middle Easterners, Central Asians, and Europeans than to south Indian populations considered to have been derived from ancestral populations that left Africa via a southern migration route within the past ~50 – 70 kya [410]. This observation is consistent with a recent study that concluded that East and southeastern Asian populations are derived primarily from the initial out-of-Africa southern migration route whereas Europeans, Central Asians and north Indians are mainly derived from the northern migration event [411]. Studies of *mt*DNA [223, 412-416] and Y chromosome [412, 417-420] variation in Indian [412, 415, 417, 418], Pakistani [416, 419] and Central Asian [223, 413, 414, 420] populations also indicated low levels of gene flow from the Near East in the recent past. Moreover, Ricaut *et al.,* [405] found a 2.5 kyo fossil excavated in the Altai region of Russia that carried the *mt*DNA N1a haplotype, which is associated with the Neolithic expansion. Presence of moderate frequencies of the genetic variant associated with lactase persistence (LP) in European population ($T_{\_13910}$) [421] was also observed in India, mainly among the northern Indian populations [412, 422]. However, the lactase persistence associated variant that is predominant in southern Arabia ($T_{\_13915}$) [423] was not observed, ruling out significant gene flow from Arabia [412]. This observation is consistent with a previous study that also showed the European LP associated $T_{\_13910}$ variant at low frequencies in Pakistani and Central Asian populations [424, 425]. Considering that the age of the European $T_{\_13910}$ variant has been estimated to be 5 - 12 kya [425], the spread of this variant might be during the Neolithic

expansion from the Fertile Crescent ~7 – 10 kya. The Fertile Crescent is an ancient area of fertile soil that runs along the western Mediterranean and the important Middle Eastern rivers, Tigris and Euphrates. It encompasses part of modern Iraq, Jordan, Lebanon, Syria, Israel and the Palestinian Territories, the southeastern fringe of Turkey and the western portion of Iran. Therefore, the genetic influence of the Neolithic expansion from the Fertile Crescent has been far and wide, encompassing the Near East, Central Asia and Europe. Therefore, the HV1, I and N1a *mt*DNA haplotypes might have been found in small bands of expanding Neolithic farmers that were absorbed into larger populations in Europe and Central Asia.

In summary, I observe different distribution patterns of both *mt*DNA and Y chromosome lineages in Africa, Near East, and Europe. In North/Northwest Africa we observe U6 and M1b *mt*DNA lineages and E3b2, E3b1a*, R1b* Y chromosomes lineages **(Appendix 9)** whereas in Northeastern Africa we observe M1a *mt*DNA lineage and E3b3, E3b1a1b, E3b6 and the ancestral E3b* Y chromosome lineages. However, there are widespread distributions (including Northeast Africa, the Near East and Europe), albeit at low frequencies, of HV1, N1a, I (*mt*DNA), and E3b1a3 (Y chromosome). It is only Y chromosome lineages E3b1a2 (E-M78α) and R1b3-R-M269 **(Appendix 9)** that are observed predominantly in Europe. Moreover, TMRCA estimates for the four *mt*DNA haplotypes (M1a, HV1, I, and N1a) and Y chromosome E3b1 and E3b3 haplotypes range between 18 – 32 kya **(Table 3.3.1; Appendix 7b)**, which coincides with around the beginning of LGM 24 kya. Subsequent population expansions of Afroasiatic speakers ~15 kya appears to correlate with the onset of warm weather that happened in Northeastern Africa ~14 kya [96, 101]. The time of the proposed Afroasiatic expansion

also correlates with the possible spread of grass collecting strategy 16 - 20 kya [389]. Ehret [96, 101], argues that the initial split and expansion of the Afroasiatic languages was from a geographical homeland between east of the Nile Valley and the northern fringes of the Ethiopian highlands. The premise of this argument is the "principle of least movement" of populations that spoke ancestral forms [426-428] of Afroasiatic languages to explain the current language distribution pattern. Based on this model, one would expect to find the greatest linguistic diversity in the region where the language originated [426-428]. In the case of the Afroasiatic language family, five of the six branches of the Afroasiatic languages are found in Africa, with only Semitic found in Southwest Asia, suggesting that the root of this language family is in Africa. Additionally, the branch with highest linguistic diversity which he proposes as the root of the Afroasiatic split, the Omotic languages, are spoken only by populations from the southwestern part of Ethiopia [96, 101, 426-428]. So these Y chromosome and *mt*DNA lineages would have been found in populations across a broad geographic range that includes the putative center of Afroasiatic expansion from the fringes of the Ethiopian highland into parts of southern Levant. The extant distribution pattern of these lineages might reflect the ancient distribution in populations that occupied this region 15 - 20 kya: for example U6, M1b and E3b2 (that are mainly observed in Northwest Africa) might have been found in the populations that occupied the western portion of this area while L3x, M1a, E3b*, E3b1a1b and E3b3 (those lineages currently observed predominantly in northeastern Africa) might have been found in populations that occupied the southeastern portion of the area **(Figure 4.3.1)**. The lineages that are widespread in Africa, southwest Asia and Europe at low frequencies (*mt*DNA HV1, N1a, I and Y chromosome E3b1a3 lineages)

might have been widespread in the whole region or originated in the central portion of this putative Afroasiatic homeland **(Figure 4.3.1)**. The scenarios described above might also explain the puzzling prevalence of Y chromosome R1b* among the Chadic speakers, with the derived R1b3 (M269) [429-431] commonly observed in the Mediterranean, Europe and the Near East **(Appendix 6a)**.

Lastly, the contention by Olivieri *et al.,* [222] and Gonzalez *et al.,* [396] that back migration from southwestern Asia into Africa occurred, is based on the premise that there was only a single initial migration of modern humans out-of-Africa via the southern route [432, 433]. Under the latter scenario both N and M were part of the same colonization process [5], and the current European populations are due to westward expansion of modern humans from southwest Asia [5]. However, there is both archaeological and genetic evidence **(Appendix 16)** that supports later movement from Africa into Europe through a northern route [393].

Overall, the genetic evidence for late Pleistocene era expansion out of Africa around 15 kya [153] as reflected by the distribution patterns of the *mt*DNA lineages M1, HV1, I, and N1a, which are associated with the expansion of Afroasiatic speakers as described above and subsequent Neolithic expansion from the Near East into the Mediterranean and the Balkans about 7 - 10 kya [153, 434], is also supported by archaeological evidence. Moreover, the current distribution of the African specific haplotypes L0-L6, at low frequencies in the Near East and Europe **(Appendix 6b)** [16, 219, 312-316, 319, 435-445] might be due to these hypothesized migration events: the expansion of Afroasiatic speakers from Africa to the Near East 15 kya and later Neolithic expansions from the Near East to Europe and Central Asia 7 - 10 kya. In fact, the *mt*DNA

African lineages, L0-L6, are absent in current Central Asian [223, 413, 414], Indian [415] and South East Asian populations [322, 323] except in Pakistan [416]. Considering that nearly two thirds of the European *mt*DNA lineages belong to the R0 family (H, HV, and V), while the Near Easterners and Central Asians have predominantly the other N haplogroups at low to medium frequency, there was probably one major initial out-of-Africa expansion into Europe and Central Asia from East/Northeast Africa 40 - 60 kya through the northern route. Migrants from this expansion event that had not already reached central/west Europe and central Asia may have contracted back to the southern Levant and the Nile Valley during the onset of extreme climatic conditions of the Last glacial maximum after 24 kya [393, 446] that lasted until 11.5 kya [446]. When warm conditions began in North Africa 15 - 14 kya [393, 447], minor expansion of Afroasiatic speakers from East/Northeast Africa to the Near East ~15 kya [96, 101, 153] and subsequent expansion during the Neolithic from the Near East into central Asia and Europe [153] might have taken place.

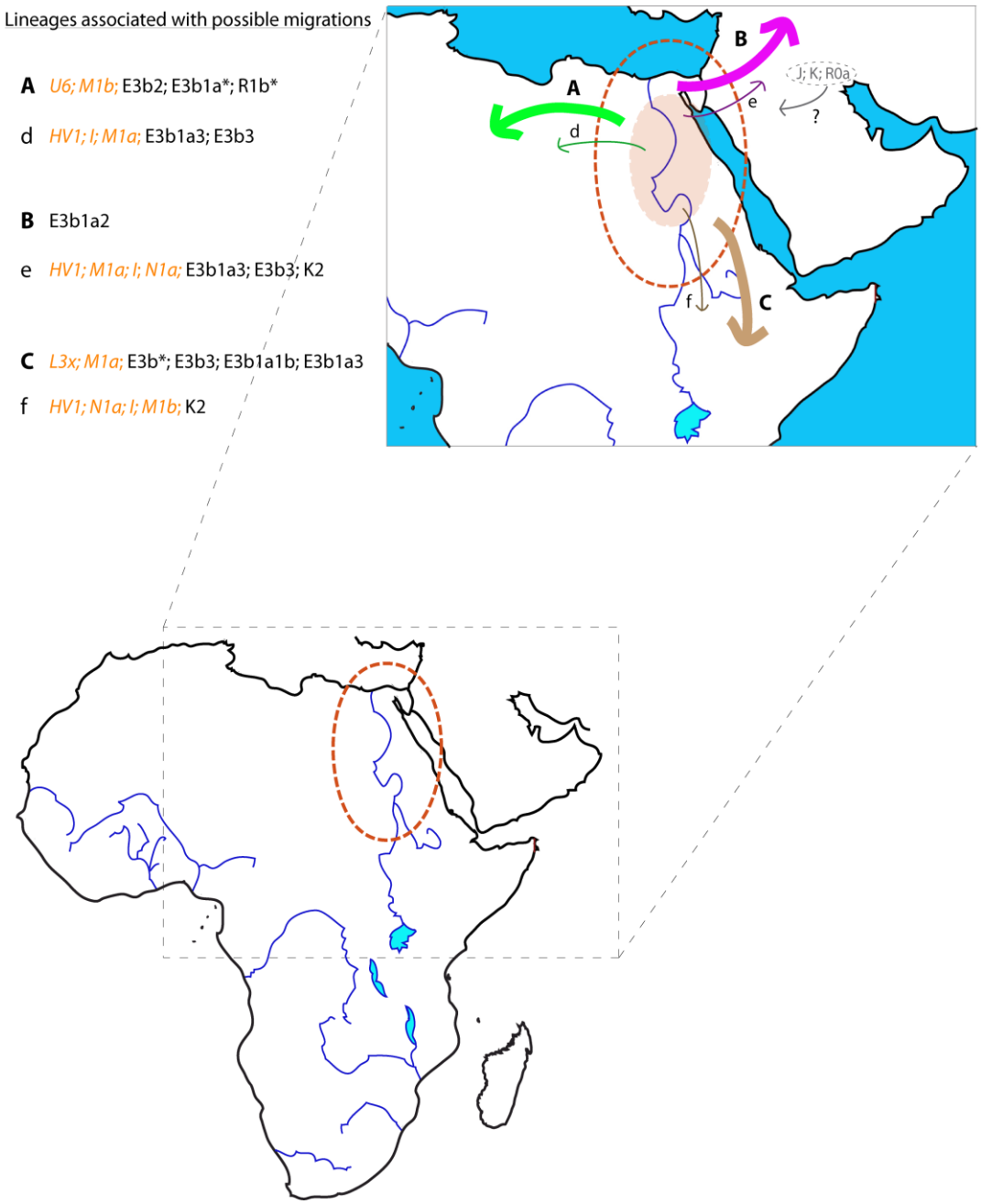


Figure 4.3.1: Model used to explain the initial expansion of Afroasiatic speakers from a Northeast African homeland (dotted circle): A and d represent proto-Chadic/Berber, B and e, proto-Semitic, while C and f represent Omotic and proto-Cushitic speakers. The bigger (capital letters) and smaller (lower cases) arrows are used to explain movement of populations with largest (major migrations) and smaller (minor migrations) proportions of *mt*DNA and Y chromosome lineages, respectively. *mt*DNA lineages are colored orange while Y chromosome lineages are colored black. Explanation of the migrations is in the text.

4.3.3: History of Back-migration from the Near East into Africa and Origins of Ethiopian Semitic Languages

Both *mt*DNA and Y chromosome data also indicate that there might have been back-migration into northeastern/East Africa from the Near East. The *mt*DNA haplotypes J, K and R0a, currently common in the Near East and Europe **(Appendix 13)**, are observed at low frequencies in East African populations (**Figure 3.4.1, Appendix 8)** and at moderate to high frequencies in southeastern Arabia, which may reflect back-migration from the Arabian peninsula across the Red sea during the Axum kingdom era (1.9 – 1.3 kya) [448] or earlier [449].

Early texts indicate that all of Arabia up to Syria was occupied by nomadic tribes and only two major kingdoms, the Ma'ian and the Sabaen, are mentioned before 2.5 kya [450]. The earliest documented Ethiopian contact with southern Arabia was around 3 kya [448, 451]. Extensive contact took place in the form of trade and migration between southern Arabia and Ethiopia by 2.7 kya [448, 452]. The trade contacts seem to have existed before the purported major migration of the Semitic people from southern Arabia into Ethiopia around 2.1 kya that founded the Axum kingdom in northern Ethiopia that lasted from 1.9 – 1.3 kya [448]. There is also evidence of navigation of the Indian Ocean by 'Ethiopians' from the African Horn and the Axumite Kingdom in their own ships about 2 kya [453] (most probably related to trade). In fact, according to Butzer [451], Axum owed its power to its partial control of an international exchange network started by Pharaonic Egypt three millennia earlier [451] meaning that Ethiopia and especially the Red sea coast have been corridors of human movement since around 5 kya. At its peak, Axum kingdom trading contacts extended from the Mediterranean and Byzantine world to India [449, 452, 454], and its rule at times encompassed parts of southern Arabia [448,

449, 451, 452]. Ethiopian Semitic languages [71, 455] might have been formed during these historical events 2 - 3 kya [455]. Ethiopian Semitic languages are hypothesized to have resulted through a process of "interference through language shift" [456], where indigenous Cushitic speakers picked up Semitic language/s of immigrants from present-day Yemen (most probably because of the political prestige of the conquering Semites) [457, 458], adapting it to their phonology and syntax [457, 459, 460]. Due to the north-to-south historical path of linguistic diffusion of the Semitic languages [459], the Semitic linguistic influence increases from north to south, with the more southern Ethiopian Semitic languages like Gurage and Harari exhibiting less Semitic (and more Cushitic) influence than the northern Ethiosemitic languages like Tigray [457, 458, 460].

The J haplogroup (J-M304) is the predominant Y chromosome lineage in the Near East region **(Appendix 6a)**. The overall distribution of the Y-chromosome J haplotype in the Near East and in Northeast Africa might also reflect recent back-migration from southern Arabia to Africa during the Axumite and Islamic era because its frequency maximum occurs in the Middle East **(Appendix 6a)**. The frequency of the J haplotype gradually decreases in regions northwest and southwest of the Fertile Crescent **(Appendix 6a)**. The haplogroup splits into two main subclades, J1 and J2 defined by the M267 and M172 SNPs, respectively. The frequency pattern of the J1–M267, a variant of the J haplotype, shows that its highest frequency and diversity is centered in the Near East. The TMRCA of the J1–M267 subhaplotype was estimated to be 4.3 - 8.7 ky [434] while a more refined estimate suggests 5.5 - 7.2 Ky [461]. Both studies concur with the interpretation that the haplotype originated in the Near East and was introduced into southern Arabia and Europe about 5 kya [434, 461]. These dates coincide perfectly with

the expansion of Semitic speaking populations in the Levant 4.4 - 7.4 kya [455]. After expansion of Semitic speakers from the Levant southward to the Arabian Desert in the Early Bronze Age (5.5 kya) [455], this haplotype might have been introduced into East Africa during the pre-Axumite and Axumite era. This is consistent with a recent study's [462] assertion that the current distribution pattern of a subset of J1-M267 subhaplotype defined by the P58 SNP, J1e lineages, might be due to expansion of Semitic pastoralists into arid habitats coinciding with the spread of Arabic and other Semitic-speaking populations after the initial Neolithic expansion about 9 - 10 kya [462].

Mitochondrial DNA haplotype L3x is observed at moderate to high frequencies among East African Cushitic speaking populations, and at low frequency in Yemen and Saudi Arabian populations. The presence of this haplotype and other East Africa-specific *mt*DNA lineages L0d3 [157], L3i, L4, L3h and L5 in Arabia might be related to pre-Axumite and Axumite era events discussed above as well as slave trade during the Islamic era. An initial excursion of Islamic crusaders entered North and Northeast Africa 1.3 kya, and by 0.8 - 0.7 kya the expansion had reached the West African forest zone and sub-Saharan savannah, called "the land of the blacks" [463]. The other contact with the Islamic world was through the trading posts along the coast of East Africa from Somalia to Madagascar that started as early as 1.3 kya [463]. During both of these contacts, animist populations in the African interior acted as a source of slaves [463]. The Christian and Jewish northeastern Africans (such as the Eritreans, Amharas and populations from North Africa) that constituted part of the Roman empire were exempt from enslavement according to Islamic law that recognized them as tax-paying citizens [463]. A recent observation by Richards *et al.,* [464] that substantial gene-flow to the

southern Arabian gene pool from East Africa through female slaves may have occurred during the recent past is also consistent with these results. Moreover, classical blood polymorphism studies show that Yemenite Jews and Arabic populations from Arabia carried some African-specific forms of blood factor [465]. Therefore, the genetic signature of some Africa lineages in southern Arabia is most probably attributed to slavery.

Pastoralism related gene flow might also have occurred from southern Arabia into East Africa around the pre-Axumite era during the introduction of camels into East Africa from Arabia through the Horn of Africa 2.5 kya [466]. Recent studies identified a variant, T/G_{13915} within a previously determined enhancer region in intron 13 of the MCM6 gene, 13,915 bp upstream from the LCT gene that is associated with lactase persistence in Saudi Arabian [423, 467] and East African populations [468]. The variant showed a high selection coefficient and functional analyses confirmed that the SNP might play a critical regulatory role [423]. Based on it's estimated age, 4 kya, and the fact that the camel was domesticated in Arabia around the same time period [466], Enattah *et al.,* [423] speculated that the T/G_{13915} variant arose independently in the Arabian Peninsula in response to camel milk consumption. The T/G_{13915} variant is found at highest frequency (~57%) in Saudi Arabia [423]. The variant is found in moderate frequencies in mostly Afroasiatic populations in Sudan, northern Kenya and Ethiopia [467, 469], and might have been introduced into the northeastern African region during the Axumite era.

In summary, there is a long history of interaction between the populations from Northeast/East Africa and the Near East, and most of the genetic signature observed in the extant human populations of these regions is a reflection of these historical events.

Initial expansion from northeastern Africa might have been associated with modern human expansion outside Africa into the Levant through the northern route around 40 - 60 kya. Later expansion might have been associated with Afroasiatic speaking populations into the Near East around 13 - 15 kya. Considering that this expansion (expansion of the Afroasiatic speakers) took place during the late Pleistocene and early Holocene [395] it might have been driven by climatic factors and a change in subsistence pattern. Subsequent expansions might have occurred during the expansions of famous kingdoms (Pharaonic, Roman and Axumite) and later religious expansions, leading to bi-directional migration between Africa and the Near East. These latter migrations, during which initial formation of Ethiopian Semitic languages took place, in most cases were coupled with inter-continental trade across the Red sea between Northeastern/East Africa and the Near East.

4.4: Bantu and Pastoralist migration to southern Africa

After the initial domestication, the spread of livestock reflects the migration of humans. Thus, the phylogeographic patterns of cattle genetic diversity should to a large extent reflect human movements, specifically the pastoralist populations [470]. In Africa, domestic animals appear to have spread from north to south in several brief episodes separated by long periods of no expansion. This pattern, may be due to climatic changes [471, 472], though the spread from East Africa to southern Africa may not have been associated with any obvious environmental factors [133].

Evidence that indicates pastoralism in Africa originated from a "single source", likely to be an area in Northeast Africa, before later expanding to the rest of Africa includes the following;

1. Linguistic evidence that shows the root words related to pastoralism originating in northeastern African among the Cushitic and Eastern Sudanic speakers. Examples of these includes; a) shared cattle related words between the Chadic, Berbers, southern Cushitic and northern Cushitic speakers - populations that are separated geographically, which indicates a knowledge of cattle that dates back to the ancestral Afroasiatic language [71, 96, 97, 101], b) different but related words for "cattle" are used among the Nilo-Saharan and Bantu, who may have borrowed them from the Sudanic Nilo-Saharans that neighbor them [59, 64, 473], c) moreover, sharing of words for "cattle" in southern Africa with the East African Sandawe hunter-gatherer population [59, 474] who in turn might have borrowed

them from Cushitic speakers [112], d) the root word for "sheep" in southern Africa related to the Sudanic word for "sheep" [59, 474].

2. Shared pastoral husbandry practices (like stimulating the flow of milk by blowing into the vulva of the cow [475, 476]) and shared rituals (for example southern Bantu populations adhere to Sudanic practices of keeping women away from cattle [473]).
3. The chronological order of archeological evidence of pastoralism that starts in North Africa and ends in South Africa **(Appendix 17)**.
4. Genetic evidence that shows a pattern of *B. indicus* introgression into native *B. taurus* cattle that begins in Northeast Africa and ends in western and southern Africa [477, 478].

Based on the distribution of livestock terms in Africa, Bender [78] concluded that pastoralism spread from the putative center of domestication in Egypt to the horn of Africa after the initial divergence of Afroasiatic populations, around the time of the Cushitic-Omotic split (8 - 10 kya), and later spread into East Africa (about 3 kya) by Nilotic speakers from the putative center of proto-Nilotic expansion in Sudan [78]. These assertions are consistent with archaeological evidence indicating that pastoralism might have been initially introduced into East Africa by people classified as Pastoral Neolithic (5.0 - 1.3 kya), associated with southern Cushitic speakers, who made Late Stone Age (LSA) tools and pottery and herded domestic cattle, sheep and goats [479]. The Elmenteitan (3 - 2 kya), another pastoral tradition associated with southern Nilotic speakers is thought to be a reflection of the second wave of pastoralist movement within East Africa [108, 109]. Based on archeological evidence of the existence of possible

proto-Nilotic speakers in northern Kenya [87] and possibly southern Sudan, Robbins [480] argues that pastoralism spread into East Africa by socio-cultural mechanisms such as trading, bride-wealth exchange and raiding.

Two models are used to explain the spread of pastoralism from East Africa into southern Africa in the last 3 ky. The first and oldest model, speculates that the South African Khoekhoe herders are descendents of eastern Sudanic speaking pastoralists who moved southwards with their herds from their East African homelands [481]. The preferred and more recent model suggests that pastoralism spread south from the great Lakes region by southern movement of a mixture of Bantu farmers and eastern Sudanic pastoralists aided by iron technology [121, 477]. These groups then came into contact with the Khoisan speaking hunter gatherers in southern Zambia/northern Botswana, who subsequently adopted a pastoral lifestyle. This model portends that only the domestic stock and the knowledge associated with herding diffused into southern Africa from East Africa [481, 482]. Herd transfer might have taken place [481] in southern Zambia/northern Botswana. However, Ehret [59] argued that besides the initial introduction of livestock by the Khoisan speaking populations (speculated to be Sandawe [59]), the Bantu speakers that migrated into southern Africa later, might have also played a crucial role in the spread of pastoralism into South Africa [59]. But, according to Fleming [112] the Sandawe's vocabulary of livestock suggests that they acquired herding-from southern Cushitic speakers. Based on the conflicting assertions above, the human population movement associated with these livestock movements is still not well understood.

Southern African Bantu speaking populations seem to have forged a strong agro-pastoral and/or pastoralist economic way of life. Apart from the well known example of populations like the Herero of Botswana, Namibia and Angola [483], the Ila and Lozi of Zambia, the Ambos and Nyanyeka-Humbe [483] of Angola and the South African Nguni, several other southern African populations have marked similarities in economy, law, ritual and symbolism with the East African Nilo-Saharan pastoralist populations [484]. Holden and Mace [485] argued that acquiring cattle led formerly matrilineal Bantu-speaking cultures, especially in East and southern Africa, to change to patrilineal or mixed descent. Moreover, archeological evidence of widespread adoption of an originally Nilotic pottery tradition [486], in lieu of Urewe (Bantu type of pottery tradition) happened in East Africa [484]. Despite the linguistic and cultural [484-486] evidence that supports Ehret's contention that livestock was introduced by Sudanic speaking populations into southern Africa [473], there are no Nilo-Saharan speaking populations in southern Africa. In fact, aside from Khoisan speaking groups, all the populations in African countries south of Tanzania (latitude 6S) are Bantu speakers [70, 487].

*mt*DNA and Y chromosome genetic evidence of expansion of the Bantu speaking populations into eastern and southern Africa has not been conclusive. For example, *mt*DNA L0a [34, 182] and L2a [181] lineages have been proposed as markers of the Bantu expansion. However, timing of the population expansion associated with these haplotypes [273] inferred using simulation to determine expansion events of these two *mt*DNA lineages pre-dates the Bantu expansion [273]. Analysis in the current study, which includes both extensive sequencing and genotyping of populations from both Central and East Africa as well as published data from African populations, shows

several mitochondrial and Y chromosome lineages with frequency maxima in Central Africa and a gentle cline away from the putative center of Bantu expansion **(Appendix 9)**. Y chromosome haplogroups E2b and E3a, especially E3a7 haplotype **(Appendix 9)**, conform to the postulated expansion of Bantu speakers from Central Africa to East Africa followed by subsequent expansion south from secondary centers in East Africa. Mitochondrial lineages that conform to this pattern include L3b, L3d and L3e, specifically L3b and L3e **(Appendix 9)**. These *mt*DNA lineages, L3b, L3d and L3e, are observed at significantly higher proportion among the Niger-Kordofanian relative to the Nilo-Saharan and Afroasiatic populations but not Khoisan (**Table 3.4.2)**. It is interesting that despite being mainly observed among the Niger-Kordofanian and showing clinal frequency patterns from Central/west Africa to the rest of Africa **(Appendix 6b, Appendix 9),** *mt*DNA L3b and L3e lineages did not show statistically significant higher proportion among Niger-Kordofanian relative to the Khoisan speaking populations (**Table 3.4.2),** and L3d and L3e lineages did not show statistically significant higher proportion in Central/west Africa relative to southern Africa **(Table 3.4.3)**. These observations might be due to small sample sizes for Khoisan speakers used in the analysis, and the possibility of genetic drift in the southern African Khoisan populations. The timing of the population expansion associated with *mt*DNA L3b and L3e lineages (4.5 - 5.5 kya) **(Appendix 14)**, inferred using simulations, corresponds with the time period estimated for the Bantu expansion [62, 71, 74]. But the estimated time of expansion for the L3d lineage (8 - 12 kya) pre-dates the Bantu expansion **(Appendix 14)**. However, this estimate corresponds with an initial expansion event of the Niger-Kordofanian populations in Central/West Africa [71]. In fact the interpolated frequency

map for L3d is consistent with a model of expansion of Niger-Kordofanian speakers from a recent Central/West Africa homeland **(Appendix 9)**. The phylogeographic pattern suggests that L3d initially expanded in a proto-Niger-Kordofanian population in Central/West Africa and was later carried into East Africa by Bantu speaking populations.

Considering the level of cultural influence of pastoralist populations on the Bantu speaking populations in East and southern Africa, it seems likely that the Bantu speaking populations had longstanding interactions with pastoral communities in East Africa before expansion to southern Africa. Alternatively, it is possible that the expanding Bantu population, armed with iron tools, absorbed all the pastoral communities in southern African that spoke other languages besides Khoisan (i.e. Nilo-Saharan or southern Cushitic) [488]. Consistent with this contention, Fleming [112] showed some linguistic borrowing of southern Cushitic words into Bantu languages of Zimbabwe. Moreover, there is archaeological evidence that shows that the Pastoral Neolithic cultural tradition (that is associated with southern Cushitic populations [488]) was distributed in parts of southern Africa as far south as Zimbabwe 2.5 kya [63, 112].

The fact that there are no southern Cushitic or Nilotic speaking populations south of Tanzania today indicates that the genetic signature of Cushitic and Nilotic speakers in southern Africa were overwhelmed by a large Bantu movement into the region. However, from the frequency pattern of Y chromosome lineage E3b6, Henn *et al.,* [355] showed evidence of migrations from East Africa into southern Africa just over 2 kya, independent of initial movement of Bantu speakers into the region after 2 kya [62, 64, 73]. Moreover, there is archaeological evidence indicating that there were livestock, cattle and sheep, and

use of pottery (which were presumably introduced by non-Khoisan speaking populations) in southern African prior to the arrival of the Bantu speakers in the area [489].

Henn *et al.,* [355] concluded that the Y chromosome E3b6 haplotype reflects a signature of the first pastoralist migration from East Africa into southern Africa. Henn *et al.,* [355] analyzed 13 populations from southern and eastern Africa. Results from the current study, which includes 1500 individuals belonging to 55 populations from Kenya, Tanzania, Sudan and Central Africa, indicate that the E3b6 haplotype is absent in Central Africa and the Sudan. The haplotype was previously found at low frequency in mixed samples of Afroasiatic and Nilotic speakers [355] from Ethiopia **(Appendix 6a)**. Two populations of Ethiopian descent, namely the Konso and the Borana, sampled in the northern part of Kenya also carry the E3b6 haplotype at low frequency **(Table 3.3.2, Appendix 6a)**. This lineage is observed at highest frequencies among the southern Cushitic speaking populations **(Figure 3.3.1, Appendix 9 Figure A9.1.8)**. Moreover, southern Cushitic populations exhibit the highest Y-chromosome microsatellite average variance on this haplotype background relative to other populations **(Table 3.3.2)**. The TMRCA estimate for the E3b6 haplotype in East Africa is 10 kya **(Table 3.3.1)** and, therefore, it might have arisen in Cushitic speaking populations possibly in Southwest Ethiopia. Based on the TMRCA estimate and distribution pattern of E3b6 haplotype, I speculate that the E3b6 haplotype probably arose in a proto eastern-southern Cushitic populations ~8 - 10 kya [115]. The initial expansion of pastoralism to southern Africa from East Africa appears to have involved southern Cushitic speaking populations. Interestingly, the average variance of Y-chromosome microsatellites on the E3b6 haplotype background in East African hunter-gatherer populations is comparable to that

of southern Cushitic speakers **(Table 3.3.2)**. This might be due to longstanding interaction between southern Cushitic speakers and east Africa hunter-gatherer populations since the migration of southern Cushitic populations into Kenya and Tanzania 5 kya [109, 110].

STRUCTURE analysis of autosomal data in the current study indicates that the southern Cushitic genetic influence extends from Kenya and Tanzania to the Hutu/Tutsi population of Rwanda **(Figure 3.2.1b Bantu)** which is at the eastern edge of the Central African region. Interestingly, a recent study found two *mt*DNA lineages mostly observed among Cushitic speakers and East African hunter-gatherer populations, at moderate (L0f) and low (L3x) frequency among the Hutu populations of Rwanda **(Appendix 6b)** [349]. This might represent the western limit of Savannah Pastoral Neolithics (Neolithic southern Cushitic) and possibly East African hunter-gatherer populations [109, 110, 490]. Moreover, a recent study found *mt*DNA lineages L4 (specifically L4b2a2) and L3a at low frequency among the Hutu populations of Rwanda [349]. The *mt*DNA L3a haplotype is another lineage (besides L4b2a2) that is observed at moderate to high frequency in mainly Cushitic speaking hunter-gatherer populations in East Africa (**Appendix 6b**). *Mt*DNA lineages L4b2a2 and L0d3 (which is also mostly observed in East African hunter-gatherer populations) have been observed at low frequency in two southern African Bantu populations, the Ronga (southern Mozambique) [185] and the Shona (Zimbabwe) [349]. These findings are consistent with the assertion based on archaeological evidence that southern Cushitic speakers and East African hunter-gatherer populations co-existed and likely admixed [109, 110]. Here I speculate that these admixed Cushitic/hunter-gatherer population(s) migrated into southern Africa in the last

2 kya, and may have introduced pastoralism, as well as some genetic lineages, into the southern African populations.

Other genetic evidence that supports the contention that the expansion of pastoralism from East Africa into southern Africa involved southern Cushitic speakers is the distribution pattern of an East African-specific single nucleotide polymorphism (SNP) variant associated with lactase persistence at position -14010 upstream from the lactase gene (the -14010C allele) [468, 469] The -14010C variant is found at highest frequency (~41%) among the southern Cushitic and eastern Nilotic speakers from East Africa [469]. Interestingly, just like the Y chromosome E3b6 lineage, the -14010C variant was not observed in Sudan and Central Africa [467, 469]. The -14010C variant was found at low to moderate frequency in populations sampled in the northern part of Kenya who are of Ethiopian ancestry and in the Sandawe from Tanzania, and at low frequency in South African Khoisan and Bantu speakers [469]. Coelho *et al.,* [297] showed low frequencies, 1 - 6%, of the -14010C variant among the Angolan Bantu speakers but did not find the variant in a sample of 111 individuals from several populations in Mozambique [297]. Considering that 7% of these Angolan populations' *mt*DNA lineages consist of Khoisan specific haplotypes, L0d and L0k [297], these Bantu populations might have acquired the -14010C variant mostly through gene-flow from Khoisan speaking pastoralist populations (who might have acquired the variant themselves from southern Cushitic speakers and/or East African hunter-gatherer populations). Moreover, microsatellite analysis of the LCT variant background showed that the -14010C variant found in South African populations appeared on the same microsatellite haplotype background as that of a variant found in the northern part of Kenya among the Cushitic speaking populations

[469]. Interestingly based on STRUCTURE analysis, the South African Cape Town admixed (self-identified as "Cape Colored") population show low levels of Cushitic (0.03) ancestry [4].

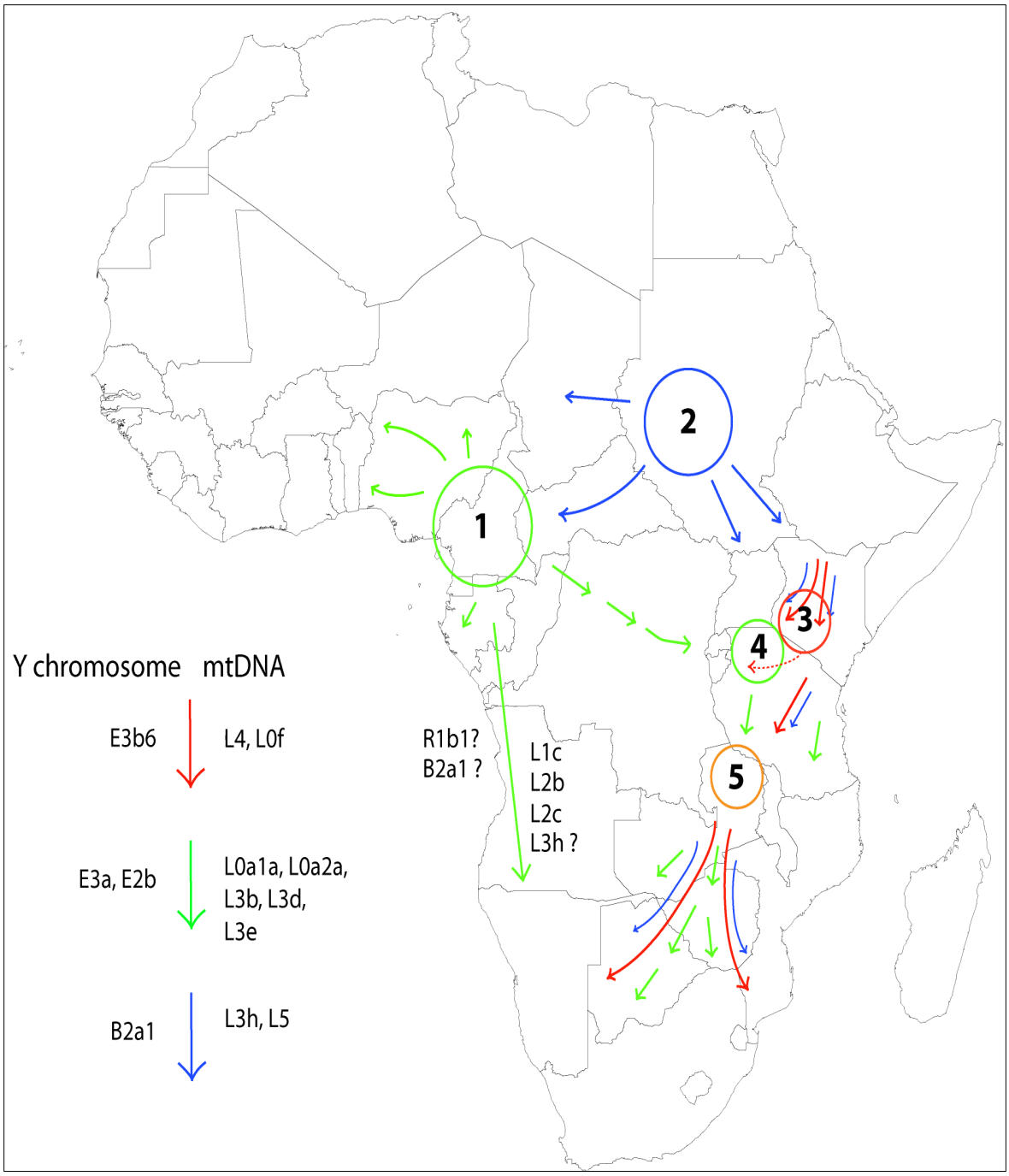
Even though, there is some genetic signature of the Nilo-Saharan speakers south of Tanzania **(Appendix 6),** later movement of pastoralism might have mainly been through cultural transmission to Bantu speakers who practiced agro-pastoralism. This assertion is exemplified by the fact that despite Nilo-Saharan culture having had a large influence on animal husbandry in southern Africa [473, 484], the level of genetic influence looks minimal **(Appendix 6)**. In STRUCTURE analysis the Nilotic speaking populations in Tanzania, the Datog, Maasai and Dorobo, analyzed in this and a larger African study [4] show substantial genetic similarity with Afroasiatic groups which is most probably attributed to the southern Cushitic speaking populations **(Figures 3.2.1a, 3.2.1b)**. The lineages that might reflect expansion of Nilotic speakers from their Sudan homeland, Y chromosome haplotypes A3b2 & E2a and *mt*DNA haplotypes L3h and L5, are found at low frequency among a few populations in southern Africa [38, 185, 295, 298]. The mitochondrial L5 lineage and Y chromosome A3b2 lineages have been reported at low frequency in Bantu speaking Mozambican populations [185, 298], in the South African colored population [295] and in some South African Bantu populations [38]. So despite lack of Cushitic and Nilo-Saharan speaking populations south of Tanzania today, there is some genetic evidence showing a signature of Cushitic and Nilo-Saharan expansion into southern Africa in the past. Moreover, the limited Nilo-Saharan genetic contribution into southern Africa shown in the current study (and from results of previous studies [38, 185, 295, 297, 298, 349]) indicate that the Bantu-speaking

populations that expanded to southern Africa via East Africa might have been influenced more culturally than genetically by the Nilotic speaking populations.

The stages in Bantu migration from central/West Africa into East Africa and subsequently to southern Africa seems to have coincided fairly closely with the spread of early Iron Age culture [73, 491]. Based on evidence of three chronological period of Iron Age culture movement into southern Africa (Early Iron age, Later iron age expansion and West Zambezi river), Vansina [62] speculates that there were at least three different dispersal events of Bantu speaking populations from Central/West Africa to southern Africa [62] (two via East Africa). Interestingly, southern African Khoisan languages have a single root word meaning "sheep" but three different words for "cattle" [71] (raising the possibility that three different groups introduced the cattle into southern Africa). There may have been an initial expansion to South Africa of a mixture of southern Cushitic speaking populations and hunter-gatherer populations from East Africa. This initial expansion might have consisted of pastoralists with mainly small ruminants specifically sheep, and possibly some longhorn Taurine cattle [489]. Considering the lag in cattle movement south relative to the small ruminants because of the "tsetse barrier" [492], the initial movement of cattle into southern Africa might have been low scale. Based on the published chronology of evidence of livestock south of Tanzania, the subsequent expansion of livestock was mostly by early iron-age Bantu speakers who had been culturally influenced by Nilo-Saharan populations that kept goats and cattle [493] **(Figure 4.4.1)**. Considering that there is evidence of humped cattle in Ngamuriak, a site associated with the Elmenteitan Neolithic (associated to southern Nilotic [109, 110]), by

2kya [494] the cattle typical of expanding Bantu pastoralists might have been the more trypanotolerant Sanga, a cross between humped zebu and the humpless Taurines [489].

Figure 4.4.1: Model of possible lineages associated with human population migration from Central and East Africa into southern Africa. The initial movement of the Bantu speakers from Central Africa **(1)** into East Africa **(4)** and subsequently southern Africa would have involved *mt*DNA lineages L0a1a, L0a2a, L3b, L3d and L3e, Y chromosome lineages E3a and movement of E2b **(Green arrows, Text and Appendix 9)**. In addition to these lineages the Bantu migration that followed the "western route" might have also involved *mt*DNA lineages L1c, L2b, L2c **(Appendix 9)** and possibly L3h, and probably Y chromosome lineages B2a1 and R1b1 **(Appendix 9)**. Genetic signature of pastoralist migration from East Africa into southern Africa that is associated with Cushitic/and or East African hunter-gatherers **(3)** might have involved *mt*DNA L4 and L0f haplotypes, and the Y chromosome E3b6 haplotype **(Red arrows)**. Transfer of livestock and possible admixture of East Africans with Khoisan speakers probably happened in Zambia **(5).** There is signature of migration of Nilo-Saharan speakers from Sudan **(3)** into Kenya and Tanzania (*mt*DNA L3h, L5 and Y chromosomes A3b2, B2a1 and E2a **(Text and Appendix 9)**). However, only Y chromosome haplotype B2a1 is observed south of Tanzania (thus, there is a limited signature of Nilo-Saharan expansion into Southern Africa **(Blue arrows)**.


Y chromosome mtDNA
E3b6
L4, L0f
E3a, E2b
L0a1a, L0a2a,
L3b, L3d,
L3e
B2a1
L3h, L5
R1b1?
B2a1 ?
L1c
L2b
L2c
L3h ?
1
2
3
4
5

Conclusions

In the current study, Y chromosome SNP and STR variation, *mt*DNA D-loop sequence, and autosomal microsatellite variation was collected and analyzed from 1500 individuals originating from 55 East (Kenya, Tanzania, Ethiopia, and Sudan) and Central African (Cameroon & Chad) populations. These populations speak diverse languages, representing the four major African linguistic families (Khoisan, Nilo-Saharan, Afroasiatic and Niger-Kordofanian) and practice a broad range of subsistence patterns (hunter-gatherers, pastoralists, agro-pastoralists, and agriculturalists). In addition, whole genome *mt*DNA variations in a subset of 222 individuals with diverse *mt*DNA haplotypes were examined. This is the first large scale Y chromosome and *mt*DNA analysis of East African and comparative African human populations **(Appendix 6)**. East African populations have been under-represented in previous studies of *mt*DNA and Y chromosome variation in Africa. Earlier studies of mitochondrial and Y chromosome lineages have observed elevated frequencies of some lineages in certain geographical regions of Africa or among populations speaking languages belonging to particular language families, but most of these studies were focused on either few populations or distributions of a specific lineage **(Appendix 6)**. In the current study I show that autosomal, mitochondrial and Y chromosome data are correlated with geographical regions within Africa or with populations speaking languages belonging to certain language families. In addition, I was able to show that the distributions of some lineages are also statistically correlated with geography/language. Consistent with previous studies [50, 55, 157, 495-499], mitochondrial and Y chromosome results from the current study may also indicate structured ancestral human populations in Africa **(Appendix 15)**.

Moreover, the current study was able to show that human population expansion occurred at periods in the past that seem to correspond to periods of warm and wet climatic conditions **(Appendix 14)**. The population expansions detected in Bayesian Skyline Plots of complete mitochondrial sequences at time periods 80 - 90 [273], 55, 35, 22, 12, 8 - 10, 5 - 6 and 2 - 3 kya fall strikingly in the periods that have been shown to be warm and wet. However, the inferred relation in the current study between demographic expansions with paleoclimatic conditions is tempered by the fact that there are not enough *mt*DNA lineage expansion time data points to statistically test correlations **(Appendix 14)**. Overall, from the extensive analysis of East African populations in the current study, several general conclusions were inferred;

- There is genetic correlation with both geography and language in eastern Africa and elsewhere in Africa. Moreover, there seems to be some Y chromosome and *mt*DNA lineages that are specific to populations speaking particular languages or belonging to clusters of language families. The correlation of genetic distance with linguistic classification in East Africa is especially interesting considering the high level of admixture between populations from different linguistic groups. Comparison of genetic and geographic/linguistic distance in a larger African population dataset [4] found that the smallest genetic versus geographic correlations were observed in eastern Africa. However, the study [4] and present reanalysis shows that population structure within East Africa is primarily due to language classification. This observation might be due to a combination of several related factors: (1) the fact that in East Africa all four African language families (Khoisan, Nilo-Saharan, Niger-Kordofanian and Afroasiatic) are represented, (2)

populations speak languages belonging to these language families mostly practice specific subsistence patterns – Khoisan (hunter-gatherers), Nilo-Saharan (pastoralist and agro-pastoralist), Niger-Kordofanian (agriculturalist and agro-pastoralists) and Afroasiatic (pastoralists), (3) East Africa has a diverse range of environments (ecological/climatic conditions) including high rainfall in the highlands (exploited mainly by agriculturalist), montane forest (exploited by hunter-gatherers), Savannah grasslands (exploited by pastoralists and hunter-gatherers) and semi-arid shrub lands (exploited by pastoralists). This environmental diversity is especially accentuated in the central Rift Valley which contains all the different ecological condition in a small geographic region, thus facilitating the possibility of these different groups practicing different economies to live in a relatively small geographical area [109].

- East Africa exhibits a complex pattern of genetic variation that is likely the result of a complex history of migrations into the region. East African populations also have elevated frequencies of several ancient lineages; L0f, L0a2 and L0b (*mt*DNA), A3b2 and B2b, B2a (Y Chromosome) some of which are shared with hunter-gatherer populations from central Africa (pygmies) and southern Africa (Khoisan speakers) which may represent ancestry from indigenous populations in the region.
- It appears that the distribution pattern of Y chromosome and *mt*DNA lineages reflect two sets of hunter-gatherer populations with different genetic ancestries in East Africa: 1) the Khoisan speaking Hadza and Sandawe populations and 2) other foragers from central and northern Kenya that generally speak the languages

of the neighboring herders. The genetic signature of hunter-gathering populations might represent initial East African inhabitants as previously indicating by archaeological and historical linguistic evidence. However, there have been extensive genetic interactions between the East African hunter-gatherers and southern Cushitic speaking populations (the first migrants into East Africa that practiced pastoralism), and also there has been some admixture with later immigrants including the pastoralist Nilo-Saharan and agriculturalist Bantu speaking populations.

- The genetic evidence shows that there has been substantial gene flow from Cushitic and Bantu speaking populations into the eastern Nilotic speaking populations (Maasai and related groups), but limited geneflow into the southern Nilotic (Kalenjins) and western Nilotic speaking populations (except Luo). Despite linguistic and archeological evidence indicating that southern Nilotic speakers migrated into Kenya and Tanzania prior to other Nilotic speakers, eastern Nilotic speaking populations seem to have had more genetic admixture with populations that belong to other linguistic groupings (Cushitic and Bantu speakers) compared to southern Nilotic speakers. Consistent with historical accounts, the western Nilotic populations have had limited geneflow with Kenyan and Tanzanian populations except for the Luo who appear to have had high levels of genetic exchange with Bantu speakers.
- There is a long history of interaction between the populations from Northeast/East Africa and the Near East, consistent with genetic signatures observed in the extant human populations of these regions. Initial expansion from northeastern Africa

might have been associated with modern human expansion outside of Africa through the northern route around 40-60 kya. Later expansion might have been associated with Afroasiatic speaking populations into the Near East around 13 – 15 kya. Considering that the expansion of the Afroasiatic speakers from northeast Africa took place during the late Pleistocene and early Holocene, it might have been driven by climatic factors and a change in subsistence pattern. Subsequent expansions might have occurred during the famous kingdoms: Pharaonic, Roman, Axumite and later religious expansions, with these events also leading to back-migration into Africa from the Near East. These migrations, which also resulted in the formation of Ethiopian Semitic languages, might have in most cases occurred through inter-continental trade across the Red sea between northeastern/East Africa and the Near East.

- The two major events that occurred due to change in subsistence patterns, namely the Bantu migration (associated with expansion of agriculturalist from central-west Africa) and the migration of pastoralist populations in Africa have had major influences on extant genetic patterns in East Africa.

Based on archaeological evidence, East Africa shows continuous human occupation after the initial emergence of modern humans ~160 – 200 kya [26]. In the more recent past, archaeological and linguistic evidence suggest that there have been at least six different migration events into East Africa during the Holocene [108-111, 127, 129-140]. There were two separate migrations associated with Afroasiatic speakers from Ethiopia into Kenya and Tanzania: the first 5 kya consisted of pastoralist southern Cushitic

speakers associated with the Savannah pastoral Neolithic culture [127, 130-132] and the second consisting of eastern Cushitic speakers beginning 2.5 kya that were associated with Turkwel traditions [108-111, 132]. There were two separate migrations associated with Nilo-Saharan speakers from Sudan, possibly via southwestern Ethiopia into Kenya and Tanzania: the first 3 kya consisted of pastoralist southern Nilotic speakers associated with Elmenteitan pastoral Neolithic culture [108, 109, 127, 129] and the second consisted of eastern Nilotic speakers beginning 1.2 kya that has been associated with Lanet and Sirikwa traditions [110, 140]. The most recent migration of Nilo-Saharan speakers into Kenya involved the western Nilotic speaking Luo population from southern Sudan through Uganda [88, 89]. Lastly, the migration of Bantu speaking populations into East Africa has been associated with several pottery traditions: Urewe, Lelesu, Kwale and Maore that are dated from 2.5 – 0.6 kya [110, 134-139]. Therefore, the frequency pattern of *mt*DNA and Y chromosome lineages in East Africa is a reflection of these historical human population movements. The unique genetic profiles among the populations speaking languages belonging to the four major linguistic families in East Africa may have been reinforced by a diverse range of environments (ecological/climatic conditions) that makes it possible for populations with different subsistence patterns to live in a relatively small geographic area.

In the future, as there has been a recent decrease in the costs of generating whole-genome SNP data and next-generation sequencing data, it will be informative to study genome-wide variation across ethnically and geographically diverse African populations. Future studies that could help clarify the demographic history of East African populations, include simulation and modeling approaches to analyze autosomal data from

unlinked loci to test a number of hypotheses raised by the current study. For example, an Isolation and Migration (IM) method can be used to distinguish time of common ancestry and migration events between populations based on non-recombining regions of the genome [500, 501]. Another approach for inferring demographic parameters and testing evolutionary hypotheses is Approximate Bayesian Computation (ABC) [502, 503], in which several scenarios with several parameters such as population divergence, admixture (migration rate) and effective population size can be modeled by comparing summary statistics estimated from simulated and observed data.

Given the large amount of ethnic diversity and high level of admixture exhibited by East African populations in the current study, additional sampling particularly from under-represented eastern African countries like Uganda, Rwanda, Burundi, Congo (particularly the eastern part), Ethiopia and Sudan will be crucial in deciphering fine scale genetic history of this genetically, culturally, and linguistically diverse regions. Moreover, among published data there is an underrepresentation of populations from Southern African countries like Malawi, Zambia, Zimbabwe and Botswana, and sampling from populations in these countries will help to fill in gaps about the initial expansion of Khoisan speakers, migration of Bantu speaking populations and spread of pastoralism from East Africa into Southern Africa. Considering the crucial role East Africa has played in human evolutionary history, future large-scale resequencing and genotyping of these Eastern and southern African populations will be informative for reconstructing both ancient and recent human evolutionary history.

Appendices

Introductions

In this section I give further information referred to in the main text of the thesis. Information on samples genotyped in the current study is given in **Appendix 1** and **Appendix 2**, and those for published data in **Appendix 6**. Primer details for PCR and sequencing for *mt*DNA analysis, including conditions, are given in **Appendices 3** and **4**. **Appendix 5** contains *mt*DNA complete genome sequence tree generated using Bayesian method in MrBayes that was used to confirm the topology obtained using MtPhyl program. In **Appendix 7,** genetic diversity generated for both *mt*DNA and Y chromosome data are given. Age estimates inferred for *mt*DNA lineages using the three methods mentioned in materials and method section, are also given. In **Appendix 8**, control region sequence and coding region SNPs genotypes for the individuals analyzed in the current study are given. Information on few individuals with mixed parentage is also given. In **Appendix 9** the trend in frequency profiles of both Y chromosome and *mt*DNA lineages (based on both results from the current study and published data) are discussed in details, in the context of the lineages' TMRCA estimates **(Table 3.3.1, Appendix 7),** and the lineages' distribution within Africa and the Near East **(Appendix 6)**. The results on neutrality test of *mt*DNA genomes are further discussed in **Appendix 10**. In **Appendix 11** extended description and discussion of results from STRUCTURE and Neighbor joining tree reconstruction analysis is done. **Appendix 12** and **Appendix 13** shows global frequency patterns of N1a, I and HV1, and J, K and R0a *mt*DNA haplotypes, respectively, used to understand the distribution pattern of these N lineages in northeastern Africa. In **Appendix 14**, detailed discussion on inference of human

demographic expansions from mitochondrial genome sequences and correspondence with paleoclimatic conditions is done. In **Appendix 15** implications of unique *mt*DNA and Y chromosome lineage distributions across Africa vis-à-vis ancestral human population structure is discussed. Archeological and genetic evidence that support "northern route" of human out-of-Africa Expansion is discussed in detail in **Appendix 16**, while the chronology of historical movement of livestock based on archaeological and paleoclimatic evidence is listed in **Appendix 17**. Median-joining network of Y-STR haplotypes for some paternal lineages in East African populations (which were not included in the results section and **Appendix 9)**, are shown in **Appendix 18**.

Appendix 1: Information on samples genotyped in the current study. Columns show African region, the country, the language family, language genus, and ethnicity which population samples are sourced from, number of individuals tested for each of the marker systems (Y chromosome Unique event polymorphisms that define the lineages, Y chromosome microsatellite –YSTR, *mt*DNA D-loop sequences, and autosomal microsatellite and indel data from Marshfield – MarsDT), and the geographic location where the population is sampled from (longitude and latitude). Numbers in parentheses in the MarsDT column are for same samples genotyped for all the markers, [y]Because of admixture between different neighboring populations some samples genotyped in Y and *mt*DNA are not the same **(Appendix 8)**.

Region	Country	Linguistic Group	Genus	Ethnicity	UEP	STR	D-loop	MarsDT	Longitude	Latitude
East Africa	Kenya	Afroasiatic	E. Cushitic	Borana	58	40	59	32(31)[y]	38.01	2.21
		Afroasiatic	E. Cushitic	Burji	23	27	21	24(14)[y]	37.78	5.60
		Afroasiatic	E. Cushitic	Gabra	29	22	31	17(6)	37.63	2.93
		Afroasiatic	E. Cushitic	Wata	9	10	24	6(6)	37.07	3.32
		Afroasiatic	E. Cushitic	Konso	17	16	17	14(14)	37.30	5.15
		Afroasiatic	E. Cushitic	El-Molo	15	20	24	16(5)	36.42	2.50
		Afroasiatic	E. Cushitic	Rendille	31	30	31	28(16)	37.35	2.30
		Afroasiatic	E. Cushitic	Garreh	27	27	26	N/A	33.51	2.79
		Afroasiatic	E. Cushitic	Yaaku	24	22	19	19(16)	38.41	0.24
		Afroasiatic	E. Cushitic	Boni	21	20	30	N/A	40.92	-2.27
		Afroasiatic	S. Cushitic	Sanye	12	13	20	N/A	40.50	-2.37
		Afroasiatic	E. Cushitic	Orma	N/A	N/A	20	N/A	40.45	-2.39
		Nilo-Saharan	W. Nilotic	Luo	27	27	20	28(16)	34.32	-0.07
		Nilo-Saharan	E. Nilotic	Teso	N/A	N/A	6	N/I	34.35	0.72
		Nilo-Saharan	E. Nilotic	Turkana	32	30	39	26(9)	35.67	3.17
		Nilo-Saharan	S. Nilotic	Sabaot	22	23	17	20(14)	34.62	0.95
		Nilo-Saharan	S. Nilotic	Keiyo	N/A	N/A	6	N/I	35.58	0.75
		Nilo-Saharan	S. Nilotic	Kipsigis	N/A	N/A	7	N/I	35.28	-0.37
		Nilo-Saharan	S. Nilotic	Marakwet	20	20	22	14(10)	35.57	0.98
		Nilo-Saharan	S. Nilotic	Pokot	38	40	39	23(18)	36.05	0.41

Region	Country	Linguistic Group	Genus	Ethnicity	UEP	YSTR	D-loop	MarsDT	Longitude	Latitude
		Nilo-Saharan	S. Nilotic	Sengwer	29	28	27	21(20)[y]	35.00	1.00
		Nilo-Saharan	S. Nilotic	Nandi	N/A	N/A	13	11(9)	35.18	0.12
		Nilo-Saharan	S. Nilotic	Tugen	34	30	43	22(22)	35.52	0.31
		Nilo-Saharan	S. Nilotic	Ogiek	24	23	22	22(10)	35.49	-0.22
		Nilo-Saharan	E. Nilotic	Ilchamus	28	28	20	27(15)	36.07	0.50
		Nilo-Saharan	E. Nilotic	Il'Laikipiak	12	20	25	33(17)[y]	37.08	0.26
		Nilo-Saharan	E. Nilotic	Samburu	12	10	30	18(15)[y]	36.63	1.00
		Niger-Kordofanian	Bantoid	Luhya	19	21	28	17(9)	34.77	0.33
		Niger-Kordofanian	Bantoid	Gikuyu	24	25	26	22(15)	37.15	-0.72
		Niger-Kordofanian	Bantoid	Taita	29	22	29	N/A	38.37	-3.40
		Niger-Kordofanian	Bantoid	Taveta	15	17	19	N/A	37.68	-3.40
	Sudan	Afroasiatic	N. Cushitic	Beja	12	14	51	42(30)[y]	36.00	19.00
		Nilo-Saharan	W. Nilotic	Nuer	14	14	22	18(15)	32.00	8.00
		Nilo-Saharan	W. Nilotic	Dinka	8	N/A	33	17(14)	28.5	8.5
		Nilo-Saharan	W. Nilotic	Shilluk	PDA	N/A	20	20(20)	31.66	9.54
		Nilo-Saharan	Nyimang	Nyimang	N/A	N/A	12	12(7)	30.45	15.07
	Tanzania	Afroasiatic	S. Cushitic	Iraqw	47	30	23	46(38)[y]	35.37	-3.53
		Afroasiatic	S. Cushitic	Fyome	19	20	13	22(18)[y]	35.67	-4.42
		Afroasiatic	S. Cushitic	Burunge	48	30	25	22(18)[y]	35.97	-5.71
		Afroasiatic	S. Nilotic	Akie	14	13	20	23(21)[y]	37.04	-5.41
		Afroasiatic	S. Cushitic	Mbugu	20	20	N/A	22(11)	38.43	-4.85
		Nilo-Saharan	S. Nilotic	Dorobo	6	8	13	10(9)[y]	35.50	-4.50
		Nilo-Saharan	S. Nilotic	Datog	50	30	30	54(29)[y]	35.17	-3.69
		Nilo-Saharan	E. Nilotic	Maasai	36	30	41	36(5)	37.00	-4.00
		Niger-Kordofanian	Bantoid	Pare	21	21	37	23(12)	38.30	-4.76
		Niger-Kordofanian	Bantoid	Mbugwe	21	23	N/A	21(14)	35.86	-3.78
		Niger-Kordofanian	Bantoid	Rangi	32	30	27	36(19)[y]	35.66	-4.30

Region	Country	Linguistic Group	Genus	Ethnicity	UEP	YSTR	D-loop	MarsDT	Longitude	Latitude
		Niger-Kordofanian	Bantoid	Turu	29	25	24	32(22)	35.15	-5.10
		Niger-Kordofanian	Bantoid	Sukuma	6	6	N/A	10(3)	33.75	-3.25
		Niger-Kordofanian	Bantoid	Sambaa	N/A	N/A	N/A	18	38.28	-4.78
		Niger-Kordofanian	Bantoid	Gogo	N/A	N/A	N/A	13	34.90	-5.78
		Khoisan	Hatsa	Hadza	42	30	19	63(15)	35.37	-3.47
		Khoisan	Sandawe	Sandawe	70	30	25	51(31)[y]	35.32	-5.39
Central African	Cameroon	Afroasiatic	Semitic	Baggara	28	25	21	23(23)[y]	14.5	12.5
		Afroasiatic	Chadic	Ouldoume	29	25	PDA	26(20)	14.00	9.00
		Afroasiatic	Chadic	Mandara	24	25	PDA	26(14)	14.00	8.00
		Afroasiatic	Chadic	Giziga	26	20	26	24(15)	14.25	10.75
		Nilo-Saharan	Western Saharan	Kanuri	30	25	26	31(26)[y]	14.30	11.30
		Niger-Kordofanian	Northern Atlantic	Fulani	40	30	23	41(24)	13.00	10.00
		Niger-Kordofanian	Bantoid	Bamun	19	25	18	31(16)	10.80	5.50
		Niger-Kordofanian	Bantoid	Mvae	16	25	23	24(15)[y]	12.00	3.00
		Niger-Kordofanian	Ubangian	Medzan	17	18	28	17(12)[y]	11.30	4.50
		Niger-Kordofanian	Ubangian	Bakola	26	25	22	42(13)[y]	10.30	2.75
		Niger-Kordofanian	Ubangian	Baka	30	25	26	48(16)	13.00	5.00
	Chad	Nilo-Saharan	Bongo-bagirmi	Laka	36	30	38	33(26)	16.30	8.30
		Nilo-Saharan	Bongo-bagirmi	Boulala	N/A	N/A	22	23(22)	19	14
	South Africa	Khoisan	Southern Khoisan	San	N/A	N/A	11	6(0)	15.68	-17.78
		Niger-Kordofanian	Bantoid	Xhosa	PDA	N/A	17	28(0)	26.53	-31.90

N/A: Not done or data not available; N/I – data not included in the current study, PDA – Published data available. S. Cushitic - Southern Cushitic; E. Cushitic – Eastern Cushitic; N. Cushitic – Northern Cushitic; W. Nilotic – Western Nilotic; E. Nilotic – Eastern Nilotic; S. Nilotic – Southern Nilotic.

Appendix 2: Information on samples for complete *mt*DNA genome sequencing in the current study. Samples that represent language families in East Africa that belong to M1a, L5, L0, N, L3 & L4 haplotypes were sequenced. Columns starting from the second column show individual sequence ID **(Appendix 8)**, *mt*DNA haplotype the sequence belongs to, ethnicity which individual sequences are sourced from, the language family and language genus.

a. M1a

	Sample	Haplotype	Ethnicity	Linguistic Group	Family
1	KEBJ009	M1a2b	Burji	Afroasiatic	Eastern Cushitic
2	KEBJ020	M1a1d	Burji	Afroasiatic	Eastern Cushitic
3	KEBR026	M1a4	Borana	Afroasiatic	Eastern Cushitic
4	KEBR034	M1a5	Borana	Afroasiatic	Eastern Cushitic
5	KEBR044	M1a1d	Borana	Afroasiatic	Eastern Cushitic
6	KEGB027	M1a4	Gabra	Afroasiatic	Eastern Cushitic
7	KEGB056	M1a4	Gabra	Afroasiatic	Eastern Cushitic
8	KEKS001	M1a2b	Konso	Afroasiatic	Eastern Cushitic
9	KEKS002	M1a4	Konso	Afroasiatic	Eastern Cushitic
10	KEKS012	M1a2c	Konso	Afroasiatic	Eastern Cushitic
11	KEOR014	M1a1j	Orma	Afroasiatic	Eastern Cushitic
12	KERD067	M1a1*	Rendille	Afroasiatic	Eastern Cushitic
13	KEWT012	M1a2c	Wata	Afroasiatic	Eastern Cushitic
14	KEWT032	M1a1j	Wata	Afroasiatic	Eastern Cushitic
15	KEYK003	M1a1*	Yaaku	Afroasiatic	Eastern Cushitic
16	KEYK021	M1a5	Yaaku	Afroasiatic	Eastern Cushitic
17	TZIQ025	M1a1h	Iraqw	Afroasiatic	Southern Cushitic
18	SDBA011	M1a5	Beja	Afroasiatic	Northern Cushitic
19	SDBA027	M1a1d	Beja	Afroasiatic	Northern Cushitic
20	SDBA032	M1a4	Beja	Afroasiatic	Northern Cushitic
21	SDHD009	M1a1*	Beja	Afroasiatic	Northern Cushitic
22	Keil021	M1a1i	Il'Laikipiak	Nilo-Saharan	Eastern Nilotic
23	Keil025	M1a1i	Il'Laikipiak	Nilo-Saharan	Eastern Nilotic
24	Kemn009	M1a2c	Il'Laikipiak	Nilo-Saharan	Eastern Nilotic
25	Kesm002	M1a1b	Samburu	Nilo-Saharan	Eastern Nilotic
26	Kesm024	M1a1h	Samburu	Nilo-Saharan	Eastern Nilotic
27	TZMS058	M1a1d	Maasai	Nilo-Saharan	Eastern Nilotic
28	KEND009	M1a9	Nandi	Nilo-Saharan	Southern Nilotic
29	KEND010	M1a1b	Nandi	Nilo-Saharan	Southern Nilotic
30	KENJ001	M1a1h	Ilchamus	Nilo-Saharan	Eastern Nilotic
31	KENJ009	M1a5	Ilchamus	Nilo-Saharan	Eastern Nilotic
32	KENJ019	M1a5	Ilchamus	Nilo-Saharan	Eastern Nilotic
33	KEOG001	M1a1f	Ogiek	Nilo-Saharan	Southern Nilotic
34	KEOG032	M1a1f	Ogiek	Nilo-Saharan	Southern Nilotic
35	KESB015	M1a9	Sabaot	Nilo-Saharan	Southern Nilotic
36	KESN069	M1a1*	Sengwer	Nilo-Saharan	Southern Nilotic
37	KETG030	M1a5	Tugen	Nilo-Saharan	Southern Nilotic
38	TZDT004	M1a1j	Datog	Nilo-Saharan	Southern Nilotic
39	TZDT035	M1a1i	Datog	Nilo-Saharan	Southern Nilotic
40	SDNY003	M1a7	Nyimang	Nilo-Saharan	Nyimang

b. L5					
	Sample	**Haplotype**	**Ethnicity**	**Linguistic Group**	**Family**
1	KEBJ006	L5a1	Burji	Afroasiatic	Eastern Cushitic
2	KEBJ030	L5a1	Burji	Afroasiatic	Eastern Cushitic
3	KEBN041	L5c	Boni	Afroasiatic	Eastern Cushitic
4	KEBR008	L5a1a	Borana	Afroasiatic	Eastern Cushitic
5	KEGU011	L5a1	Garreh	Afroasiatic	Eastern Cushitic
6	KEKS013	L5a1a	Konso	Afroasiatic	Eastern Cushitic
7	KEYK035	L5b1a1a	Yaaku	Afroasiatic	Eastern Cushitic
8	TZSW017	L5c	Sandawe	Khoisan	Khoisan
9	TZSW053	L5c	Sandawe	Khoisan	Khoisan
10	KEGK011	L5a3a1	Gikuyu	Niger-Kordofanian	Bantoid
11	KELY013	L5a3a1	Luhya	Niger-Kordofanian	Bantoid
12	KELY028	L5a3a1	Luhya	Niger-Kordofanian	Bantoid
13	KESB020	L5c	Luhya	Niger-Kordofanian	Bantoid
14	TZPR008	L5b1a1a	Pare	Niger-Kordofanian	Bantoid
15	KEIL005	L5a3a1	Il'Laikipiak	Nilo-Saharan	Eastern Nilotic
16	KESM037	L5a2	Samburu	Nilo-Saharan	Eastern Nilotic
17	KETK008	L5b1c	Turkana	Nilo-Saharan	Eastern Nilotic
18	KETS002	L5b1a2	Teso	Nilo-Saharan	Eastern Nilotic
19	TZMS053	L5a3a1	Maasai	Nilo-Saharan	Eastern Nilotic
20	KEND012	L5a3a	Nandi	Nilo-Saharan	Southern Nilotic
21	KENJ002	L5a	Ilchamus	Nilo-Saharan	Eastern Nilotic
22	KENJ013	L5a3a	Ilchamus	Nilo-Saharan	Eastern Nilotic
23	KEOG005	L5a3	Ogiek	Nilo-Saharan	Southern Nilotic
24	KEPK004	L5b1a2	Pokot	Nilo-Saharan	Southern Nilotic
25	KEPK010	L5b1d	Pokot	Nilo-Saharan	Southern Nilotic
26	KEPK053	L5a3	Pokot	Nilo-Saharan	Southern Nilotic
27	KESB014	L5b1a1a	Sabaot	Nilo-Saharan	Southern Nilotic
28	KESN003	L5a3a1	Sengwer	Nilo-Saharan	Southern Nilotic
29	KESN026	L5a2	Sengwer	Nilo-Saharan	Southern Nilotic
30	KESN049	L5a2	Sengwer	Nilo-Saharan	Southern Nilotic
31	KETG024	L5b1a2	Tugen	Nilo-Saharan	Southern Nilotic
32	KETG043	L5a	Tugen	Nilo-Saharan	Southern Nilotic
33	KETG053	L5b1a1a	Tugen	Nilo-Saharan	Southern Nilotic
34	SDDN016	L5b1d	Dinka	Nilo-Saharan	Western Nilotic
35	SDDN020	L5b1a1a	Dinka	Nilo-Saharan	Western Nilotic
36	SDDN027	L5b1a1a	Dinka	Nilo-Saharan	Western Nilotic
37	SDNR028	L5b1a2	Nuer	Nilo-Saharan	Western Nilotic
38	SDSH004	L5b1a1a	Shilluk	Nilo-Saharan	Western Nilotic
39	SDNY008	L5a3a1	Nyimang	Nilo-Saharan	Nyimang

c. L0					
	Sample	**Haplotype**	**Ethnicity**	**Linguistic Group**	**Family**
1	KEBJ005	L0f3a	Burji	Afroasiatic	Eastern Cushitic
2	KEBJ011	L0a3a	Burji	Afroasiatic	Eastern Cushitic
3	KEBJ037	L0f2a	Burji	Afroasiatic	Eastern Cushitic
4	KEBJ038	L0a1a	Burji	Afroasiatic	Eastern Cushitic
5	KEBN003	L0f1b1	Boni	Afroasiatic	Eastern Cushitic
6	KEBN018	L0d3b	Boni	Afroasiatic	Eastern Cushitic
7	KEBN019	L0a1c	Boni	Afroasiatic	Eastern Cushitic
8	KEBN022	L0f2a1	Boni	Afroasiatic	Eastern Cushitic
9	KEBN023a	L0f1b2	Boni	Afroasiatic	Eastern Cushitic
10	KEBR001	L0a2d2	Borana	Afroasiatic	Eastern Cushitic
11	KEBR013	L0a1d	Borana	Afroasiatic	Eastern Cushitic
12	KEBR062	L0a1d	Borana	Afroasiatic	Eastern Cushitic
13	KEBR100	L0a1a	Borana	Afroasiatic	Eastern Cushitic
14	KEBR104	L0a1c	Borana	Afroasiatic	Eastern Cushitic
15	KEEL005	L0a1d	El-Molo	Afroasiatic	Eastern Cushitic
16	KEEL022	L0f2a	El-Molo	Afroasiatic	Eastern Cushitic
17	KESY003	L0a4a1	Sanye	Afroasiatic	Southern Cushitic
18	KESY004	L0f1b1	Sanye	Afroasiatic	Southern Cushitic
19	KEWT005	L0b1a1	Wata	Afroasiatic	Eastern Cushitic
20	TZIQ059	L0f3b	Iraqw	Afroasiatic	Southern Cushitic
21	TZIQ075	L0a2d1	Iraqw	Afroasiatic	Southern Cushitic
22	TZIQ083	L0a3a1	Iraqw	Afroasiatic	Southern Cushitic
23	TZWF021	L0f3b	Fyome	Afroasiatic	Southern Cushitic
24	TZWF041	L0a2d1	Fyome	Afroasiatic	Southern Cushitic
25	TZSW109	L0d3a2	Sandawe	Khoisan	Khoisan
26	CABM037	L0a5	Bamoun	Niger-Kordofanian	Bantoid
27	KEGK022	L0f3b	Gikuyu	Niger-Kordofanian	Bantoid
28	KELY037a	L0a4a2	Luhya	Niger-Kordofanian	Bantoid
29	KELY040	L0b1a	Luhya	Niger-Kordofanian	Bantoid
30	TZTR029	L0f2a	Turu	Niger-Kordofanian	Bantoid
31	CHBU017	L0a5	Boulala	Nilo-Saharan	Bongo-bagirmi
32	KELO002	L0b1	Luo	Nilo-Saharan	Western Nilotic
33	KELO084	L0a3a1	Luo	Nilo-Saharan	Western Nilotic
34	SDDN006	L0a1e	Dinka	Nilo-Saharan	Western Nilotic
35	KEKY002	L0f3a	Keiyo	Nilo-Saharan	Southern Nilotic
36	KEMR019	L0a4a2	Marakwet	Nilo-Saharan	Southern Nilotic
37	KENJ028	L0a1d	Ilchamus	Nilo-Saharan	Eastern Nilotic
38	KESB027	L0a1c	Sabaot	Nilo-Saharan	Southern Nilotic
39	KESN015	L0b1a1	Sengwer	Nilo-Saharan	Southern Nilotic
40	KETG044	L0a1a	Tugen	Nilo-Saharan	Southern Nilotic
41	KETG063	L0f1b2	Tugen	Nilo-Saharan	Southern Nilotic
42	TZBG031	L0d3a2	Burunge	Nilo-Saharan	Southern Nilotic
43	TZBG057	L0a3a1	Burunge	Nilo-Saharan	Southern Nilotic
44	TZDT012	L0a4a1	Datog	Nilo-Saharan	Southern Nilotic
45	KEMN007	L0a1c	Il'Laikipiak	Nilo-Saharan	Eastern Nilotic
46	KEMN011	L0a1d	Il'Laikipiak	Nilo-Saharan	Eastern Nilotic
47	KETK009	L0a2d2	Turkana	Nilo-Saharan	Eastern Nilotic
48	KETK042	L0a1c	Turkana	Nilo-Saharan	Eastern Nilotic
49	TZAK003	L0d3a1	Akie	Nilo-Saharan	Southern Nilotic
50	TZAK004	L0d3a1	Akie	Nilo-Saharan	Southern Nilotic
51	TZAK006	L0d3a1	Akie	Nilo-Saharan	Southern Nilotic
52	TZDR007	L0a1b	Dorobo	Nilo-Saharan	Southern Nilotic

d. N					
	Sample	**Haplotype**	**Ethnicity**	**Linguistic Group**	**Family**
1	CAAS016	H1	Baggara	Afroasiatic	Semitic
2	CAGZ021	U6b	Giziga	Afroasiatic	Chadic
3	SDBA019	J1a	Beja	Afroasiatic	Northern Cushitic
4	KEEL001	I	El-Molo	Afroasiatic	Eastern Cushitic
5	KERD050	K1a	Rendille	Afroasiatic	Eastern Cushitic
6	KEBJ021	N1a	Burji	Afroasiatic	Eastern Cushitic
7	KEGB037	N1a	Gabra	Afroasiatic	Eastern Cushitic
8	KEBR020	N1a	Borana	Afroasiatic	Eastern Cushitic
9	KEBR018	N1a	Borana	Afroasiatic	Eastern Cushitic
10	KEGU042	N1a	Garreh	Afroasiatic	Eastern Cushitic
11	KEBR018	N1a	Borana	Afroasiatic	Eastern Cushitic
12	KEBR101	R0a2	Borana	Afroasiatic	Eastern Cushitic
13	KEBR021	U9	Borana	Afroasiatic	Eastern Cushitic
14	KEOR023	U9	Orma	Afroasiatic	Eastern Cushitic
15	TZIQ079	N1a	Iraqw	Afroasiatic	Southern Cushitic
16	TZBG033	I	Burunge	Nilo-Saharan	Southern Nilotic
17	KEIL018	HV	Il'Laikipiak	Nilo-Saharan	Eastern Nilotic
18	Kesm014	J1b	Samburu	Nilo-Saharan	Eastern Nilotic
19	TZDR002	R0a2	Dorobo	Nilo-Saharan	Southern Nilotic
20	CHLA013	U6a	Laka	Nilo-Saharan	Bongo-bagirmi

e. L3 & L4					
	Sample	**Haplotype**	**Ethnicity**	**Linguistic Group**	**Family**
1	KEBJ001	L3x2a	Burji	Afroasiatic	Eastern Cushitic
2	KEBJ003	L3e2a	Burji	Afroasiatic	Eastern Cushitic
3	KEBJ014	L3x2a	Burji	Afroasiatic	Eastern Cushitic
4	KEBJ040	L3x1a	Burji	Afroasiatic	Eastern Cushitic
5	KEBN017	L3j	Boni	Afroasiatic	Eastern Cushitic
6	KEBR035	L3i2	Borana	Afroasiatic	Eastern Cushitic
7	KEBR040	L3x2a	Borana	Afroasiatic	Eastern Cushitic
8	KEBR048	L3a1a2	Borana	Afroasiatic	Eastern Cushitic
9	KEBR056	L4b2a2	Borana	Afroasiatic	Eastern Cushitic
10	KEBR070	L3x1b	Borana	Afroasiatic	Eastern Cushitic
11	KEBR093	L3i1c	Borana	Afroasiatic	Eastern Cushitic
12	KEEL009	L3a1a1	El-Molo	Afroasiatic	Eastern Cushitic
13	KEGU060	L3h1a2a	Garreh	Afroasiatic	Eastern Cushitic
14	KESY016	L3a1	Sanye	Afroasiatic	Southern Cushitic
15	KESY018	L3a1	Sanye	Afroasiatic	Southern Cushitic
16	KEWT033	L3a1a1	Wata	Afroasiatic	Eastern Cushitic
17	KEYK010	L3f1b	Yaaku	Afroasiatic	Eastern Cushitic
18	KEYK014	L4c	Yaaku	Afroasiatic	Eastern Cushitic
19	TZIQ042	L3f2	Iraqw	Afroasiatic	Southern Cushitic
20	TZIQ057	L3a2	Iraqw	Afroasiatic	Southern Cushitic
21	TZWF028	L3j	Fyome	Afroasiatic	Southern Cushitic
22	CABM003	L3e1a	Bamoun	Niger-Kordofanian	Bantoid
23	CABM033	L3e1	Bamoun	Niger-Kordofanian	Bantoid
24	KEGK006	L3f2a	Gikuyu	Niger-Kordofanian	Bantoid
25	KEGK007	L3j	Gikuyu	Niger-Kordofanian	Bantoid
26	KEGK035	L3a1a2	Gikuyu	Niger-Kordofanian	Bantoid
27	KELY011	L3b1a	Luhya	Niger-Kordofanian	Bantoid
28	KETV066	L3a1	Taveta	Niger-Kordofanian	Bantoid
29	TZTR015	L3b1a	Turu	Niger-Kordofanian	Bantoid
30	CAKN036	L3f2b	Kanuri	Nilo-Saharan	Western Saharan
31	KELO003	L3h1a1	Luo	Nilo-Saharan	Western Nilotic
32	KELO009	L3e1g	Luo	Nilo-Saharan	Western Nilotic
33	KENJ051	L3b1a	Ilchamus	Nilo-Saharan	Eastern Nilotic
34	KEPK040	L3i2a	Pokot	Nilo-Saharan	Southern Nilotic
35	KEPK097	L4c	Pokot	Nilo-Saharan	Southern Nilotic
36	KETG007	L3h1a1	Tugen	Nilo-Saharan	Southern Nilotic
37	KETG055	L3f4	Tugen	Nilo-Saharan	Southern Nilotic
38	KETG075	L3h1a1	Tugen	Nilo-Saharan	Southern Nilotic
39	KETG099	L3a1a	Tugen	Nilo-Saharan	Southern Nilotic
40	TZDT015	L3h1b	Datog	Nilo-Saharan	Southern Nilotic
41	TZDT076a	L3h1a1	Datog	Nilo-Saharan	Southern Nilotic
42	TZDT081	L4b2a1	Datog	Nilo-Saharan	Southern Nilotic
43	KETK017	L3h1a1	Turkana	Nilo-Saharan	Eastern Nilotic
44	KETK061	L4c	Turkana	Nilo-Saharan	Eastern Nilotic
45	TZDR005	L3i1c	Dorobo	Nilo-Saharan	Southern Nilotic
46	SDDN024	L3h1a2a	Dinka	Nilo-Saharan	Western Nilotic
47	SDSH009	L3d1b	Shilluk	Nilo-Saharan	Western Nilotic

Appendix 3: Oligonucleotide primers used for complete *mt*DNA genome PCR amplification[a] and DNA sequencing[b].

Primer name[a]	Primer Sequence 5'-3'	3' Position[c]
1F	CTCCTCAAAGCAATACACTG	611
1R	TGCTAAATCCACCTTCGACC	1411
2F	CGATCAACCTCACCACCTCT	1245
2R	TGGACAACCAGCTATCACCA	2007
3F	GGACTAACCCCTATACCTTCTGC	1854
3R	**GGCAGGTCAATTTCACTGGT**	**2669**
4F	**AAATCTTACCCCGCCTGTTT**	**2499**
4R	AGGAATGCCATTGCGATTAG	3346
5F	TACTTCACAAAGCGCCTTCC	3169
5R	ATGAAGAATAGGGCGAAGGG	3961
6F	TGGCTCCTTTAACCTCTCCA	3796
6R	AAGGATTATGGATGCGGTTG	4654
7F	ACTAATTAATCCCCTGGCCC	4485
7R	CCTGGGGTGGGTTTTGTATG	5420
8F	CTAACCGGCTTTTTGCCC	5255
8R	ACCTAGAAGGTTGCCTGGCT	6031
9F	GAGGCCTAACCCCTGTCTTT	5855
9R	ATTCCGAAGCCTGGTAGGAT	6642
10F	CTCTTCGTCTGATCCGTCCT	6469
10R	AGCGAAGGCTTCTCAAATCA	7315
11F	**ACGCCAAAATCCATTTCACT**	**7148**
11R	**CGGGAATTGCATCTGTTTTT**	**8095**
12F	ACGAGTACACCGACTACGGC	7937
12R	TGGGTGGTTGGTGTAAATGA	8797
13F	TTTCCCCCTCTATTGATCCC	8621
13R	GTGGCCTTGGTATGTGCTTT	9397
14F	CCCACCAATCACATGCCTAT	9230
14R	TGTAGCCGTTGAGTTGTGGT	10130
15F	TCTCCATCTATTGATGAGGGTCT	9989
15R	AATTAGGCTGTGGGTGGTTG	10837
16F	GCCATACTAGTCTTTGCCGC	10672
16R	TTGAGAATGAGTGTGAGGCG	11472
17F	TCACTCTCACTGCCCAAGAA	11314
17R	GGAGAATGGGGGATAGGTAT	12076
18F	**TATCACTCTCCTACTTACAG**	**11948**
18R	**AGAAGGTTATAATTCCTACG**	**12772**
19F	AAACAACCCAGCTCTCCCTAA	12571
19R	TCGATGATGTGGTCTTTGGA	13507
20F	ACATCTGTACCCACGCCTTC	13338
20R	AGAGGGGTCAGGGTTCATTC	14268
21F	GCATAATTAAACTTTACTTC	14000

Primer name[a]	Primer Sequence 5'-3'	3' Position[c]
21R	AGAATATTGAGGCGCCATTG	14998
22F	TGAAACTTCGGCTCACTCCT	14856
22R	AGCTTTGGGTGCTAATGGTG	15978
23F	TCATTGGACAAGTAGCATCC	15811
23R	GAGTGGTTAATAGGGTGATAG	5
24F	CACCATTCTCCGTGAAATCA	16420
24R	AGGCTAAGAGTTTTGAGCTG	775

[a]PCR primers are shown in bold. Primers pairs used in PCR amplifications are 4F/11R, 11F/18R and 18/3R. PCR amplification of mtDNA genomes were completed in 3 ≈6.0kb reactions. Cycles conditions used for the amplification of the Control Region and some coding regions SNPs in the mtDNA of the East African samples are as indicated in **Appendix 4**. All the amplification products were done in PTC 9600/2700 thermocyclers. Each PCR cycles consisted of the following;

Denaturation	Cycles		Final extension
94°C-45seconds	94°C-30seconds 68°C-30seconds 72°C-60seconds for each 1kb	×30	72°C-60seconds

[b]Modified from Gonder *et al*., [218]
[c]Primer positions are relative to their location in the Cambridge Reference Sequence [232].

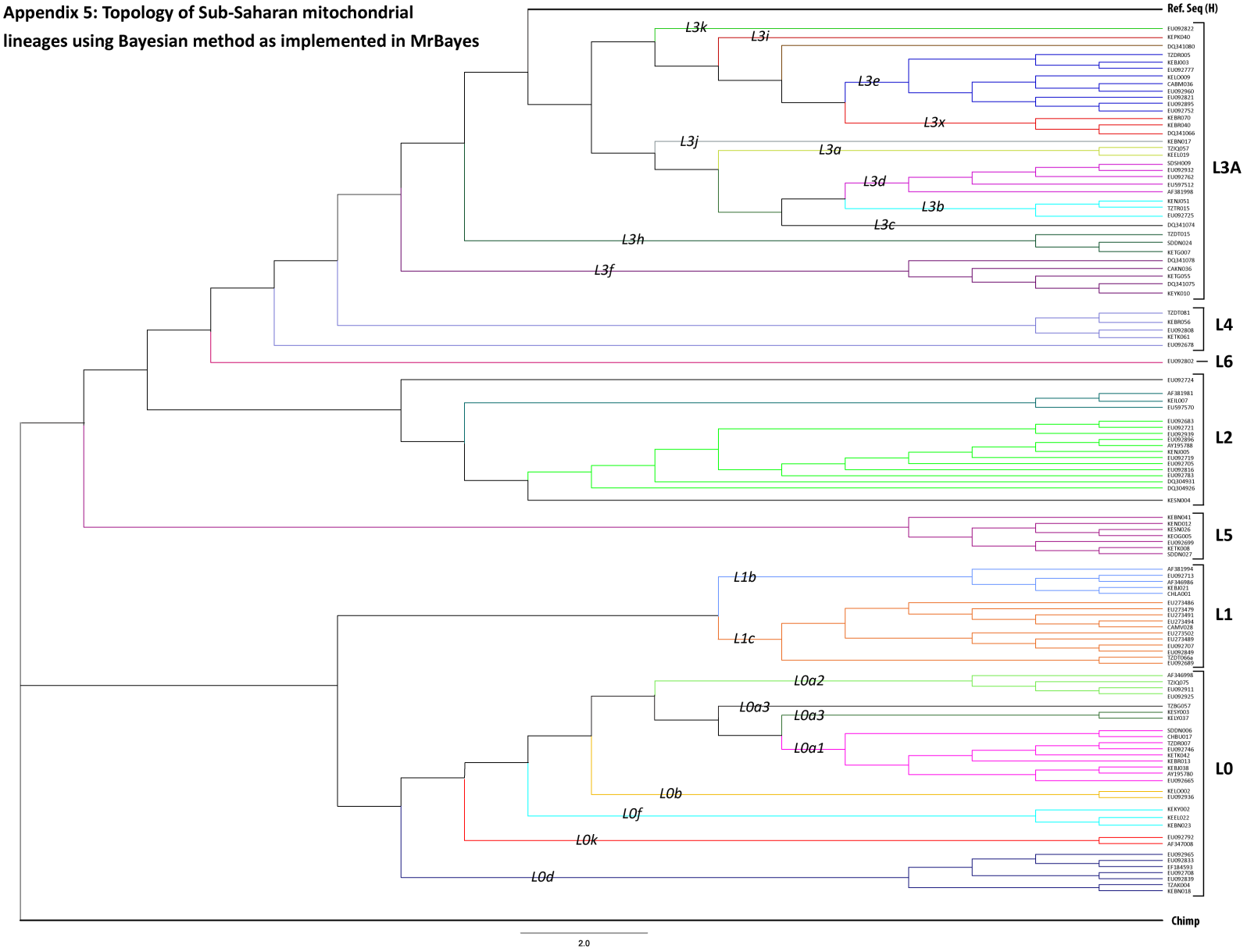
Appendix 4: Coding Region haplogroup diagnostic SNPs genotyped. The columns show *mt*DNA haplogroups the sequences belong to, haplotype the sequence belongs to, the haplotype diagnostic position sequences are tested for, primer used and genome region in the *mt*DNA genome covered by the fragment amplified for testing.

Haplogroup	Haplotypes	Diagnostic positions	Primers used[1]	Genome Region
L0	L0afbk	T4586C	6F/7R	3796-5420
	L0abf	C5603T	8F/9R	5255-6642
	L0a	G5231A	6F/7R	3796-5420
	L0a1	T5096C	6F/7R	3796-5420
	L0a1b	C5911T	8F/9R	5255-6642
	L0a2	A5711G, G6257A	8F/9R	5255-6642
	L0a2d	A5581G	8F/9R	5255-6642
	L0a3	T6050C, C6689A	8F/9R, 10F/11R	5255-6642, 6469-8095
	L0b	T6719C	10F/11R	6469-8095
	L0d3	G4580A, G5773A, C6277T	6F/7R, 8F/9R	3796-5420, 5255-6642
	L0f	C4964T, T7148C	6F/7R, 10F/11R	3796-5420, 6469-8095
	L0f2a	A4562G	6F/7R	3796-5420
	L0f2a	T4688C, 7061G	6F/7R, 10F/11R	3796-5420, 6469-8095
	L0f2b	T6152C	8F/9R	5255-6642
	L0f2b1	T6176C	8F/9R	5255-6642
	L0f2b2	C6164T, A6923G	8F/9R, 10F/11R	5255-6642, 6469-8095
L2a				
	L2a	7175C, 7274T, 7028T, 7771G	10F/11R	6469-8095
	L2a	7175C, 7274C, , 7028T, 7771G	10F/11R	6469-8095
	L2a2	A6752G	10F/11R	6469-8095
	L2a2a	C6656T	10F/11R	6469-8095
	L2a1a	A6663G	10F/11R	6469-8095

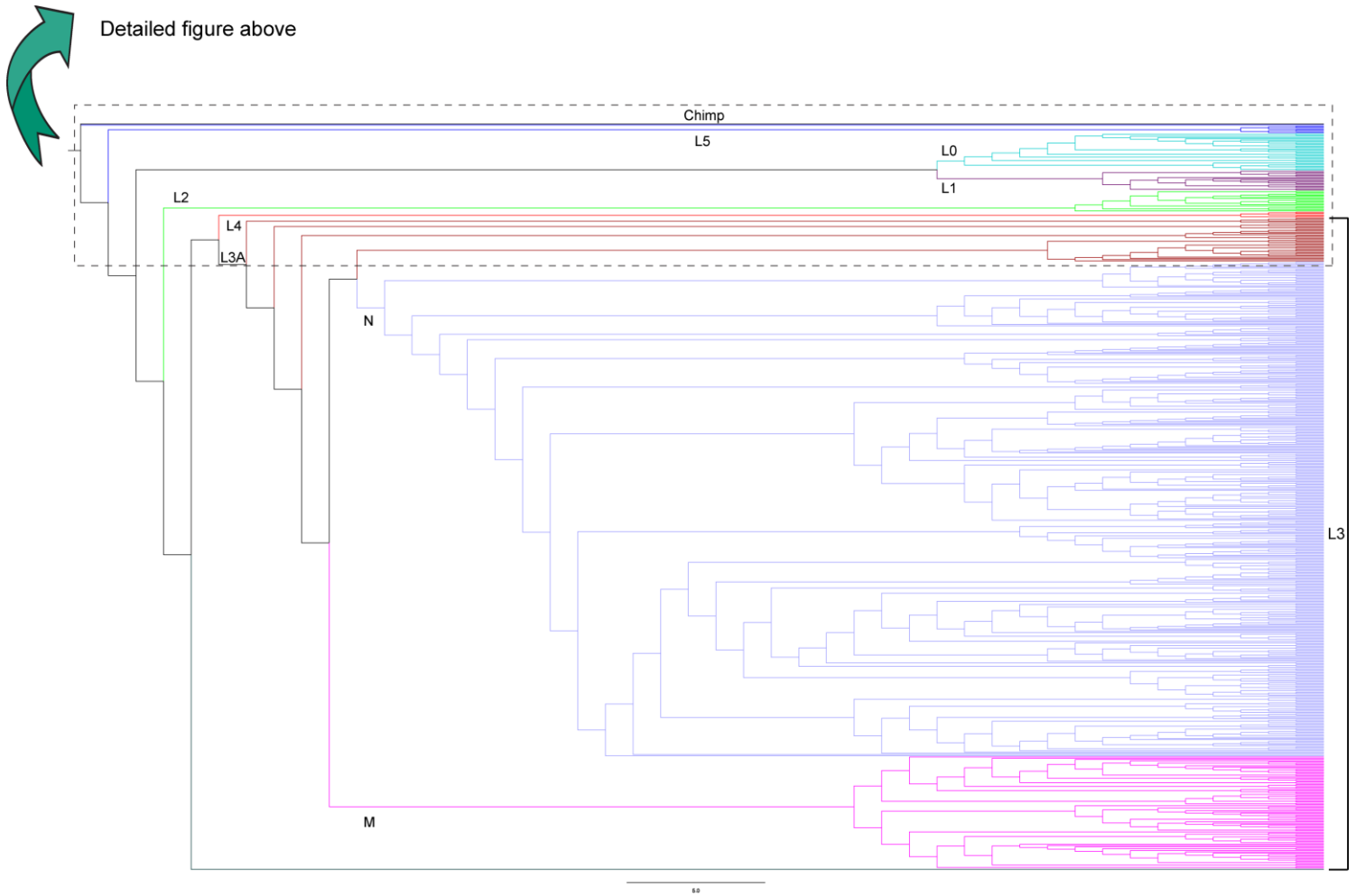
Haplogroup	Haplotypes	Diagnostic positions	Primers used[1]	Genome Region
L3	L3b	A10086G	L09965: TGTCTCCATCTATTGATGAG	9966-85
			H10302: ACTATTAGTGGCAGGTTAGT	10303-22
	L3d	T8618C	L08412: TACTCCTTACACTATTCCTC	8413-32
			H08702: TCCTTTAGTGTTGTGTATGG	8722-41
	L3e	T2352C	L02264: AATTGGACCAATCTATCACC	2266-85
			H02437: AACCTTTCCTTATGAGCATG	2439-58
	L3h	G9575A	14F/16R	9230-11472
	L3h1	G1719A	2F/2R	1245-2007
	L3h1	A4388G	6F/7R	3796-5420
	L3h1	T9509C	14F/16R	9230-11472
	L3h1	A11590G	17F/17R	11314-12076
	L3eikx	A10819G	14F/16R	9230-11472
	L3i1	G13800A	20F/20R	13338-14268
	L3h/L3x	G3483A	5F/5R	3169-3961
	L3x	A13708A	20F/20R	13338-14268
M1	M1	A10398G	14F/16R	9230-11472
	M1	C10400T	14F/16R	9230-11472

[1]Primers without sequence fragment details are the same ones used in complete mitochondrial sequencing in Appendix 3.

Appendix 5: Topology of Sub-Saharan mitochondrial lineages using Bayesian method as implemented in MrBayes



The figure below depicts the topology of the global mitochondrial lineages while the one above show detailed topology of Sub-Saharan mitochondrial lineages. The *mt*DNA L3 haplogroup includes L3a-L3k plus M and N (as shown in topology of global mitochondrial lineages), while previously defined L3A haplogroup encompasses all the L3 haplotypes, L3a-L3k, excluding the M and N haplotypes.



Topology of global mitochondrial lineages inferred from complete sequences using Bayesian method as implemented in MrBayes

Appendix 6a (i): Published Y chromosome data included in the study. Columns show global region, the country, the language family, language genus, and the population name (or nationality), number of individuals tested for each of the marker systems, the geographic location where the population is sampled from (longitude and latitude) and the primary source references.

Region	Country	Language family	Genus	Population	n	longitude	latitude	Sources
East Africa	Somalia	Afroasiatic	E. Cushitic	Somali	205	45.37	2.07	[504]
	Ethiopia	Afroasiatic	E. Cushitic	Arsi	85	39.83	7.83	[291]
		Mixed		Ethiopians	95	36.65	5.65	[291]
		Mixed		Ethiopian	88	36.65	5.65	[14, 155, 355]
		Afroasiatic	W. Cushitic	Ethiopian Jews	39	37.47	12.60	[153, 164, 505]
		Afroasiatic	E. Cushitic	Oromo	78	39.27	8.55	[155]
		Afroasiatic	Semitic	Amhara	48	37	12	[155]
	Rwanda	Niger-Kordofanian	Bantoid	Hutu	69	29.74	-2.60	[156]
		Niger-Kordofanian	Bantoid	Tutsi	94	30.54	-2.16	[156]
	Uganda	Niger-Kordofanian	Bantoid	Ganda	26	32.57	0.32	[38]
		Nilo-Saharan	E. Nilotic	Karamajong	118	34	3.33	[346]
	Comoros	Niger-Kordofanian	Bantoid	Comorian	381	4.42	-12.25	[506]
	Sudan	Nilo-Saharan	western Nilotic	Dinka	26	28.5	8.5	[370]
		Nilo-Saharan	Fur	Fur	32	25	13	[370]
		Nilo-Saharan	western Nilotic	Masalit	32	24.88	12.05	[370]
		Nilo-Saharan	W. Nilotic	Shilluk	15	31.66	9.54	[370]
		Nilo-Saharan/Niger-Kordofanian		Nuba	28	30.75	12.00	[370]
		Afroasiatic	Semitic	Copts	33	33.53	14.40	[370]
		Nilo-Saharan	Nubian	Nubians	39	31.35	21.80	[370]
Central Africa	C.A.R.[b]	Nilo-Saharan	Central Sudanic	Mbuti	47	29.97	1.85	[38]
	Sao Tome	Creole	Creole	Sao Tome	138	6.53	0.03	[507]
	Cameroon	Niger-Kordofanian	Adamawa	Fali	39	14.56	11.40	[160, 164, 508]

Region	Country	Linguistic	Family	Population	n	longitude	latitude	Sources
		Niger-Kordofanian	Ubangian	Baka	48	13.00	5.00	**A**; [38]
		Niger-Kordofanian	Bantoid	Bakola	59	10.30	2.75	**A**; [38]
		Niger-Kordofanian	Ubangian	Biaka	31	13.00	5.00	[38]
		Niger-Kordofanian	Bantoid	Bamileke	48	10.07	5.45	[160, 508]; [164]
	Angola	Niger-Kordofanian	Bantoid	Cabinda[c]	74	12.50	-5.00	[292]
Southern Africa	South Africa	Khoisan	Northern Khoisan	!Kung	64	19.50	-18.33	[160, 164, 355, 508]
		Niger-Kordofanian	Bantoid	Xhosa	80	26.53	-31.90	[38]
		Niger-Kordofanian	Bantoid	Zulu	29	32.00	-28.00	[38]
	Namibia	Khoisan	Central Khoisan	Khwe	26	21.8	-18.3	[160, 164, 355, 508]
		Khoisan	Northern Khoisan	!Kung/Sekele	32	23.00	-17.83	[38]
		Khoisan	Central Khoisan	Dama	18	15.42	-22.00	[38]
		Khoisan	Central Khoisan	Tsumkwe	29	15.68	-17.78	[38]
	Botswana	Niger-Kordofanian	Bantoid	Sotho-Tswana	28	26.50	-24.25	[38]
		Niger-Kordofanian	Bantoid	Herero	24	17.08	22.57	[38]
	Zimbabwe	Niger-Kordofanian	Bantoid	Shona	49	31.03	-17.86	[38]
West Africa	Nigeria	Afroasiatic	Chadic	Hausa	32	8.48	13.05	[370]
	Benin	Niger-Kordofanian	Kwa	Fon	100	1.98	7.18	[156]
	Guinea Bissau	Niger-Kordofanian	N. Atlantic	Guinea Bissau	276	-9.30	10.38	[165]
	Burkina Faso	Niger-Kordofanian	Gur	Mossi	49	-1.52	12.37	[160, 164, 508]
	Burkina Faso	Niger-Kordofanian	Mixed	Rimaibe	37	-3.01	16.77	[160, 164, 508]

Region	Country	Linguistic	Family	Population	n	longitude	latitude	Sources
	Senegal	Niger-Kordofanian	N. Atlantic	Senegalese	139	-15.86	16.39	[155]
		Niger-Kordofanian	W. Mande	Mandinka	39	-14.75	12.17	[38]
		Niger-Kordofanian	N. Atlantic	Wolof	34	-15.86	16.39	[38]
	Ghana	Niger-Kordofanian	Kwa	Fante	32	-0.22	5.55	[38]
	Cape Verde	Creole		Cape Verde	201	-23.52	14.92	[165]
	Mali	Niger-Kordofanian	Dogon	Dogon	55	-3.61	14.35	[38]
North & North West Africa	Morocco	Afroasiatic	Semitic	Arabs	49	-6.83	34.02	[153, 160, 164, 509]
		Afroasiatic	Berber	Berbers	64	-4.50	33.50	[153, 164, 509]
	Tunisia	Afroasiatic	Semitic	Tunisians	148	10.18	36.80	[510]
	Algeria	Afroasiatic	Semitic	Algerians	102	3.05	36.75	[511]
	Egypt	Afroasiatic	Semitic	Egyptians	239	31.25	30.05	[38, 156]
Middle East	Iraq	Afroasiatic	Semitic	Iraqis	139	44.39	33.34	[512]
	Qatar	Afroasiatic	Semitic	Qataris	72	51.53	25.29	[513]
	Yemen	Afroasiatic	Semitic	Yemenis	62	44.21	15.35	[513]
	Saudi Arabia	Afroasiatic	Semitic	Saudis	157	46.77	24.64	[514]
	U.A.E.[d]	Afroasiatic	Semitic	Emirates	164	55.28	25.25	[513]
	Lebanon	Afroasiatic	Semitic	Lebanese	914	35.51	33.87	[515]
	Oman	Afroasiatic	Semitic	Omanis	121	58.59	23.61	[156]
	Jordan	Afroasiatic	Semitic	Jordanians	146	35.93	31.95	[516]
	Syria	Afroasiatic	Semitic	Syrians	554	36.3	33.5	[517, 518]
	Palestine	Afroasiatic	Semitic	Palestinians	101	34.47	31.50	[518]
	Kuwait	Afroasiatic	Semitic	Kuwaitis	42	47.97	29.37	[518]
	Jordan	Afroasiatic	Semitic	Jordanians	273	35.93	31.95	[518]
	Israel	Afroasiatic	Semitic	Jews[e]	952	34.76	32.07	[519]

Region	Country	Linguistic	Family	Population	n	longitude	latitude	Sources
	Iran	Indo-European	Iranian	Iranians	150	51.42	35.67	[520]
Mediterranean and Anatolia	Turkey	Altaic	Turkic	Turks	523	32.86	39.93	[521]
	Greece	Indo-European	Greek	Greeks	93	23.73	37.98	[522]
	Albania	Indo-European	Albanian	Albanians	119	19.82	41.33	[522]
	Italy	Indo-European	Romance	Sicilians	236	13.37	38.12	[523]
		Indo-European	Romance	Sardinians	930	9.25	40.12	[327]
	Spain	Indo-European	Romance	Spanish	978	-5.99	37.38	[524]
	Portugal	Indo-European	Romance	Portuguese	795	-8.42	40.20	[524, 525]

E. – eastern, W.- western, N. – northern, **A** – This study, C.A.R.[b]; Central African Republic, [c]Cabinda; either Kikongo or Yombe which are both Bantu, U.A.E.[d] - United Arab Emirates, [e]Jewish males representing the major Jewish communities across the Jewish Diaspora 215 Cohanim, 738 Israelites.

Appendix 6a (ii) part I: Frequency of Y chromosome lineages in Africa, Mediterranean and the Near East. Data for some lineages that are commonly observed outside Africa are not shown, thus frequency does not add up to 100% for all populations.

Population	n	A*	A1	A2	E3a	A3b1	A3b2	B*	B1	B2*	B2a*	B2a1	B2a3	B2b*	B2b1	B2b2	B2b4	B2b5
Garreh	27																	
Boni	21						4.8							4.8				
Borana	58																	1.7
Burji	23																	
Konso	19						5.3											
El-Molo	15						13.3					6.7	6.7					
Gabra	29						3.4											
Rendille	31						6.5											
Sanye	12											8.3						
Wata	9																	
Yaaku	24												50					
Iraqw	47						2.1							6.4				2.1
Fyome	19						10.5							15.8				
Burunge	48						8.3							14.6				4.2
Mbugu	20													5.0				
Akie	14																	
Somali	205						0.5	1[j]										
Ethiopian Jews	39						41.0											
Arsi	85						18.0											
Oromo	78						10.3	1.3										
Amhara	48						14.6					2.1						
Ethiopians[l]	95					2.0	18.0					2[e]						
Mixed Ethiopians	88				1.1		12.5			1.1		9.1						
Beja	12																	
Giziga	26						7.7					3.8						
Mandara	24						4.2											
Oldoume	29											10.3						
Hausa	32						12.5					15.63[e]						

Population	n	A*	A1	A2	E3a	A3b1	A3b2	B*	B1	B2*	B2a*	B2a1	B2a3	B2b*	B2b1	B2b2	B2b4	B2b5
Baggara	28						7.1					3.6						
Copts	33											15.15[e]						
Egyptians	239						2.9					0.8						
Tunisians	148																	
Algerians	102																	
Moroccan	49																	
Arabs																		
Moroccan	64		3.0															
Berbers																		
Iraqis	139																	
Qataris	72											1.4[e]		1.4[f]				
Saudis	157										0.64			1.27				
Yemen	62																	
Emirates	164																	
Lebanese	914																	
Omanis	121						1.0											
Jordanians	146						0.7											
Syrians	554						0.5											
Palestinians	101						1.0											
Kuwaitis	42																	
Jordanians	273																	
Jewish	952																	
Turks	523						0.4											
Greeks	93																	
Albanians	119																	
Sicilians	236																	
Sardinians	376						0.3											
Spanish	978																	
Portuguese	795																	
Hadza	42													42.9				4.8
Sandawe	70						2.9					2.9		14.3				1.4
!Kung	64			8.0		28.0								8.0				
Khwe	26					12.0												
!Kung/Sekele	32			25.0		22.0									3.0		6.0	

Population	n	A*	A1	A2	E3a	A3b1	A3b2	B*	B1	B2*	B2a*	B2a1	B2a3	B2b*	B2b1	B2b2	B2b4	B2b5
Dama	18			6.0		6.0				6.0								
Tsumkwe	29			45.0		21.0									24.0		7.0	
Gikuyu	24											4.2						
Luhya	19						15.8					5.3						
Taita	29																	3.4
Taveta	15											6.7						
Mbugwe	21											4.8		4.8				
Pare	21						4.8											
Rangi	32						9.4							6.3				
Sukuma	6																	
Turu	29																	3.4
Ganda	26						8.0											
Comorian	381											1.57[e]						
Hutu	69											3[e]		1[d]				
Tutsi	94											1[e]		14[d]				
Bamoun	19																	
Mvae	16																	
Fali	39											18.0						
Bamileke	48								4.0									
Fulani	40						7.5					2.5						
Medzan	17													5.9				
Baka	48						4.2				2.1						52.1	
Bakola	59	5.1								3.4		8.5		6.8			1.7	
Biaka	31									3.0		3.0		3.0			45.0	
Sao Tome	138							2.9				0.7		2.90[d]				
Cabinda (Angola)	74										1.4	8.1						
Herero	24																	
Shona	49											10.0						
Xhosa	80					5.0						5.0						
Zulu	29					3.0						14.0		3.0				
Sotho-Tswana	28					7.0						18.0						
Cape Verde	201		0.5	2.0														
Fon	100																	

Population	n	A*	A1	A2	E3a	A3b1	A3b2	B*	B1	B2*	B2a*	B2a1	B2a3	B2b*	B2b1	B2b2	B2b4	B2b5
Guinea Bissau	276		5.1															
Mossi	49							2.0										
Rimaibe	37																	
Senegalese	139																	
Mandinka	39		5.0					3.0										
Wolof	34																	
Fante	32																	
Dogon	55		2.0								7.0							
Datog	50						2.0					2.0		4.0				
Dorobo	6																	
Il'Laikipiak	12						16.7											
Ilchamus	28													3.6				
Maasai	36						27.8							2.8				5.6
Marakwet	20						10.0					5.0		5.0	10.0	10.0		
Nuer	14						42.9					57.1						
Ogiek	24						8.3											8.3
Pokot	38						13.2					10.5	2.6	2.6				
Sabaot	22																	
Samburu	12						16.7							8.3				
Sengwer	29						3.4					17.2						
Tugen	34						2.9											
Turkana	32						28.1					6.3	6.3	3.1				
Luo	27						63.0											
Karamajong	118						33.1				16.9		4.2	11.0				
Dinka	26						61.5					23.1[e]						
Shilluk	15						53.3					26.7[e]						
Fur	32						31.3					3[e]						
Masalit	32						18.8					3[e]						
Nuba	28						46.4					14.29[e]						
Nubians	39											7.69[e]						
Mbuti	47						2.0			6.0	11.0			21.0			21.0	
Kanuri	30						26.7					10.0						
Laka	36											5.6						

Appendix 6a (ii) part II: Frequency of Y chromosome lineages in Africa, Mediterranean and the Near East

Population	n	E*	E1	E2*	E2a	E2b	E3*	E3a*	E3a1	E3a4	E3a6	E3a7	E3a8*	E3a8a*	E3b*	E3b1*
Garreh	27						3.7								14.8	
Boni	21											9.6	19.0	14.3		
Borana	58				1.7	1.7	8.6								5.2	
Burji	23						34.8					8.7	4.3		13.0	
Konso	19						36.8								15.8	
El-Molo	15											13.4			6.7	
Gabra	29				3.4		3.4								6.9	
Rendille	31						12.9								19.4	
Sanye	12					8.3										
Wata	9				11.1		22.2									
Yaaku	24							4.2				4.2	8.3			
Iraqw	47				4.3							2.1	4.3	4.3	2.1	
Fyome	19							10.5				5.3	15.8		10.5	
Burunge	48					2.1	29.2	2.1					8.3			
Mbugu	20											5.0	35.0			
Akie	14				7.1			7.1				7.1	7.1	7.1	14.3	
Somali	205				0.5[j]			1.50[a]							1.5	77.60[c]
Ethiopian Jews	39						18.0								7.7	
Arsi	85						8.0								18.0	25[c]
Oromo	78				1.3[k]		15.4								19.2	35.9[c]
Amhara	48						10.4								10.4	22.9[c]
Ethiopians[l]	95				1[k]		4.0								15.0	22[c]
Mixed Ethiopians	88				17[k]		18.2	3.4							4.8	22.7[c]
Beja	12														50.0	
Giziga	26											3.8		7.7		
Mandara	24					41.7								4.2		
Oldoume	29															
Hausa	32		15.6					12.5[a]								3.13[c]
Baggara	28					3.6		14.3						3.6	3.6	
Copts	33														6.1	15.15[c]

Population	n	E*	E1	E2*	E2a	E2b	E3*	E3a*	E3a1	E3a4	E3a6	E3a7	E3a8*	E3a8a*	E3b*	E3b1*
Egyptians	239	0.4	0.4					0.8				0.8			5.0	22.59[c]
Tunisians	148		0.7					1.35[a]							3.4	5.41[c]
Algerians	102							7.84[a]								5.88[c]
Moroccan Arabs	49														2.0	
Moroccan Berbers	64		2.0					5.0								
Iraqis	139							1.4[a]							10.8[b]	
Qataris	72					2.8						2.8				
Saudis	157	1.27						3.82				3.82			2.55	
Yemen	62							3.2							3.2	
Emirates	164											5.5				
Lebanese	914	0.6	0.2					0.66[a]							0.2	10.50[c]
Omanis	121					2.0						2.0				
Jordanians	146															10.27[c]
Syrians	554		0.4					1.08[a]							12.09[b]	
Palestinians	101							1[a]							26.73[b]	
Kuwaitis	42							4.76[a]							14.29[b]	
Jordanians	273							0.73[a]							23.08[b]	
Jewish	952	0.1					0.1								2.1	4.60[c]
Turks	523						0.4	0.2							5.0	
Greeks	93															
Albanians	119															1.7
Sicilians	236															
Sardinians	376														3.19[b]	4.79[c]
Spanish	978							0.20[a]							0.9	3.68[c]
Portuguese	795		0.5					0.25[a]							1.1	4.15[c]
Hadza	42					4.8	7.1		2.4			16.7	4.8	2.4		
Sandawe	70					1.4	4.3	2.1				18.6	15.7	7.1	2.9	
!Kung	64					6.0						16.0				
Khwe	26					4.0										
!Kung/Sekele	32							19.0				13.0				
Dama	18	6.0		6.0		6.0		33.0				22.0				
Tsumkwe	29														3.0	

Population	n	E*	E1	E2*	E2a	E2b	E3*	E3a*	E3a1	E3a4	E3a6	E3a7	E3a8*	E3a8a*	E3b*	E3b1*
Gikuyu	24					12.5		12.5				37.5	25.0			
Luhya	19				5.3	10.5						15.8	15.8	10.5		
Taita	29					10.3							65.5	24.1		
Taveta	15					20.0	20.0	13.3	6.7				20.0	6.7		
Mbugwe	21					4.8	33.3					23.8		9.5	9.5	
Pare	21					4.8		4.8	4.8			14.3	23.8	33.3		
Rangi	32				3.1			3.1				12.5	21.9	3.1	6.3	
Sukuma	6					33.3						50.0				
Turu	29					6.9	20.7	3.4	3.4			24.1	17.2	3.4	3.4	
Ganda	26				8.0	4.0						46.0				
Comorian	381	1.3				14.4		9.97				22.1		9.5		
Hutu	69				4.0	4.0		22[f]	10.0			51.0			3.0	
Tutsi	94					4.0		32[f]				48.0			1.0	
Bamoun	19							10.6				73.8		15.8		
Mvae	16											87.6	6.3	6.3		
Fali	39							26.0				33.0				
Bamileke	48								25.0	15.0		56.0				
Fulani	40					7.5		15.0			2.5		5.0		2.5	
Medzan	17					5.9		11.8	17.6			17.7	23.5			
Baka	48											39.6	2.1			
Bakola	59					3.4		13.56*				49.2	6.8	1.7		
Biaka	31							6.0				39.0				
Sao Tome	138							69.57[a]							2.2	2.90[c]
Cabinda (Angola)	74							45.95[f]				32.4				
Herero	24											33.0				
Shona	49			2.0				51.0				37.0				
Xhosa	80	4.0				28.0		34.0				20.0			5.0	
Zulu	29	3.0				21.0		34.0				21.0				
Sotho-Tswana	28	4.0				4.0		36.0				21.0			7.0	
Cape Verde	201	2.5	5.0				0.5	15.92[a]							17.41[h]	
Fon	100					5.0		38[f]				57.0				
Guinea Bissau	276					0.72[g]	0.7	71.38[a]							6.16[h]	

Population	n	E*	E1	E2*	E2a	E2b	E3*	E3a*	E3a1	E3a4	E3a6	E3a7	E3a8*	E3a8a*	E3b*	E3b1*
Mossi	49					4.0	2.0	67.0				22.0			2.0	
Rimaibe	37					27.0	3.0	51.0	5.0			8.0				
Senegalese	139		5.0			2.9[g]	2.9	80.6				0.7			5.0	0.7[c]
Mandinka	39		3.0	3.0				79.0								5[c]
Wolof	34					3.0	3.0									
Fante	32		3.0				3.0	44.0				41.0			3.0	
Dogon	55		45.0			2.0		38.0				5.0				
Datog	50				14.0	2.0	12.0					12.0		2.0	4.0	
Dorobo	6				16.7		16.7									
Il'Laikipiak	12											16.7				
Ilchamus	28				32.1	14.3			7.1			3.6				
Maasai	36				8.3	2.8	5.6					5.6	11.1	8.3	5.6	
Marakwet	20				25.0											
Nuer	14															
Ogiek	24											25.0		8.3	4.2	
Pokot	38				31.6		2.6							13.2		
Sabaot	22				9.1							36.3	4.5	9.1		
Samburu	12															
Sengwer	29				37.9							3.4			3.4	
Tugen	34				55.9	5.9	17.6					8.8			2.9	
Turkana	32				15.6		9.4	3.1				3.1	6.3	3.1		
Luo	27					3.7						25.9		3.7		
Karamajong	118				11.0	2.5		6.8				5.1			5.9	0.8[c]
Dinka	26															15.38[c]
Shilluk	15														6.7	13.33[c]
Fur	32															59.38[c]
Masalit	32															71.88[c]
Nuba	28														14.3	25[c]
Nubians	39														7.7	15.38[c]
Mbuti	47					4.0		28.0				6.0				
Kanuri	30						6.7								6.7	
Laka	36					11.1	16.7	2.8			5.6	8.3	22.2			

Appendix 6a (ii) part III: Frequency of Y chromosome lineages in Africa, Mediterranean and the Near East

Population	n	E3b1a*	E3b1a1	E3b1a2	E3b1a3	E3b2	E3b3	E3b4	E3b5	E3b6	F*	J	K2	R1b*	R1b3
Garreh	27		74.1									3.7	3.7		
Boni	21				4.8		4.8			19.0		14.3		4.8	
Borana	58		50.0		1.7					6.9			3.4		
Burji	23		8.7		4.3		8.7							4.3	
Konso	19		15.8		10.5					5.3					
El-Molo	15		20.0		6.7					13.3					
Gabra	29		58.6		3.4				10.3	3.4		3.4	3.4		
Rendille	31		29.0							19.4			9.7		
Sanye	12		41.7							8.3					
Wata	9		55.6							11.1					
Yaaku	24						8.3			25.0					
Iraqw	47				2.1					51.1			12.8		
Fyome	19									21.1			10.5		
Burunge	48									22.9			4.2		
Mbugu	20									55.0					
Akie	14				7.1					21.4			21.4		
Somali	205					1.5	0.5					3.0	10.4		
Ethiopian Jews	39		10.86[m]				12.8					5.1	5.1		
Arsi	85						5.0					26.0			
Oromo	78						5.1	2.6				3.9	5.1		
Amhara	48						2.1					33.3	4.2		
Ethiopians[l]	95						6.0					25.0			
Mixed Ethiopians	88						2.3			2.0		3.4		**X**	
Beja	12	8.3	33.3		8.3										
Giziga	26	7.7												57.7	
Mandara	24		4.2											37.5	
Oldoume	29												3.4	86.2	
Hausa	32													40.63[i]	
Baggara	28	3.6										17.9	3.6	17.9	
Copts	33											45.5		15.15[i]	

Population	n	E3b1a*	E3b1a1	E3b1a2	E3b1a3	E3b2	E3b3	E3b4	E3b5	E3b6	F*	J	K2	R1b*	R1b3
Egyptians	239					6.7	4.2				0.8	28.5	8.0	1.7	1.3
Tunisians	148					37.8	2.7				4.7	35.8	0.7	6.8	
Algerians	102					45.1						27.5			10.8
Moroccan Arabs	49	34.7	2.0		4.1	32.7						20.0			2.0
Moroccan Berbers	64	11.0				69.0						6.0			
Iraqis	139										0.7	58.3	7.2		10.8
Qataris	72		1.4		1.4		1.4					66.7			1.4
Saudis	157		0.64				4.46					57.98	5.1		1.19
Yemen	62						8.1					82.2			
Emirates	164		0.6		6.7		3.1					45.1	4.9	0.6	3.7
Lebanese	914					1.2	4.3					46.1	4.7		7.3
Omanis	121											48.0	8.0		1.0
Jordanians	146					2.7	13.0				0.7	43.8		**X**	
Syrians	554											55.2		2.71[i]	
Palestinians	101											58.4		1[i]	
Kuwaitis	42											42.9		9.52[i]	
Jordanians	273											50.2		9.16[i]	
Jewish	952					0.3	8.8				0.4	44.8	2.4	2.8	7.5
Turks	523					0.2	5.5					33.5	2.5	0.4	14.5
Greeks	93			16.3	1.1		2.2					22.9	4.3		17.4
Albanians	119			29.4			1.7					22.7			18.5
Sicilians	236		1.3	5.9	3.8	2.1	4.7					29.7	5.5		24.6
Sardinians	376						2.9				0.8				19.2
Spanish	978					4.4	0.9				0.1	8.5		0.2	65.6
Portuguese	795					5.7	1.3				0.1	10.2	1.9		58.4
Hadza	42									14.3					
Sandawe	70									21.4			1.4		
!Kung	64									11.0					
Khwe	26									31.0					
!Kung/Sekele	32									13*					
Dama	18											6.0			6.0
Tsumkwe	29														

Population	n	E3b1a*	E3b1a1	E3b1a2	E3b1a3	E3b2	E3b3	E3b4	E3b5	E3b6	F*	J	K2	R1b*	R1b3
Gikuyu	24									8.3					
Luhya	19														
Taita	29														
Taveta	15									6.7					
Mbugwe	21									9.5					
Pare	21														
Rangi	32									18.8			12.5	3.1	
Sukuma	6														
Turu	29		3.4							3.4			3.4		
Ganda	26														
Comorian	381				0.5		1.1			0.8	0.5	15.2			
Hutu	69													1.0	
Tutsi	94														
Bamoun	19														
Mvae	16														
Fali	39													23.0	
Bamileke	48														
Fulani	40		2.5											17.5	
Medzan	17														
Baka	48														
Bakola	59														
Biaka	31														
Sao Tome	138					1.5									10.1
Cabinda (Angola)	74										1.4			9.5	
Herero	24											4.0		16.0	
Shona	49														
Xhosa	80														
Zulu	29														
Sotho-Tswana	28										4.0				
Cape Verde	201					3.0						19.9	1.0	17.41[i]	
Fon	100														
Guinea Bissau	276														

Population	n	E3b1a*	E3b1a1	E3b1a2	E3b1a3	E3b2	E3b3	E3b4	E3b5	E3b6	F*	J	K2	R1b*	R1b3
Mossi	49														
Rimaibe	37														
Senegalese	139										1.4				
Mandinka	39					3.0									
Wolof	34														
Fante	32													6.0	
Dogon	55														
Datog	50				2.0					38.0			4.0		
Dorobo	6									66.7					
Il'Laikipiak	12				16.7					50.0					
Ilchamus	28				28.6					10.7					
Maasai	36				2.8					8.3		2.8	2.8		
Marakwet	20		5.0		5.0					35.0					
Nuer	14														
Ogiek	24									45.8					
Pokot	38		2.6						2.6	18.4					
Sabaot	22		4.5							36.4					
Samburu	12		25.0		8.3				8.3	25.0					
Sengwer	29		13.8							20.7					
Tugen	34									5.9					
Turkana	32		3.1				3.1			3.1					
Luo	27		3.7												
Karamajong	118									1.9			0.8		
Dinka	26														
Shilluk	15														
Fur	32														
Masalit	32														
Nuba	28														
Nubians	39										10.3	43.6		10.26[i]	
Mbuti	47														
Kanuri	30											6.7	6.7	36.7	
Laka	36													25.0	

[a]consist of all the individuals who carry M2 (E3a)

[b]consist of all the individuals who carry M35 (E3b)

[c]consist of all the individuals that carry M78 (E3b1)

[d]subhaplotype in B2b not determined

[e]subhaplotype in B2a not determined

[f]E3a8 defining U81 not typed

[g] subhaplotype in E2 not determined but since only E2b have been observed in population outside East Africa – this frequency might represent that of E2b in the population

[h]All possible M35 carrying haplotypes beside E3b2 (E3b,xE3b2)

[i]M269 and other downstream SNPs not typed

[j]haplotype in B not determined

[k]subhaplotype in E2 not determined

[l]students at Kotebe Teaching College, Addis Ababa intended to be representative of the general Ethiopian population

[m]consist of all the individuals that carry M78 (E3b1), but since all the chromosomes from population of Ethiopian Jews from E3b1 that were analyzed in Cruciani *et al.,* 2004 are E3b1a1, it is assumed here that these individuals belong to this haplotype (E3b1a1)

X – M173 present but haplotypes within not determined

Appendix 6b(i): Published *mt*DNA data included in the study. Columns show global region, the country, the language family, language genus, and the population name (or nationality), number of individuals tested for each of the marker systems, the geographic location where the population is sampled from (longitude and latitude) and the primary source reference.

Region	country	Linguistic	Genus	Group	n	Longitude	Latitude	Source
East Africa	Kenya	Afroasiatic	E. Cushitic	Daasanech	49	35.80	4.62	[526]
		Nilo-Saharan	E. Nilotic	Turkana	47	35.67	3.17	[151, 152, 526]
		Mixed		Nairobians	100	36.82	-1.28	[366]
	Ethiopia	Afroasiatic	Semitic	Amhara	120	37.0	12.0	[16]
		Afroasiatic	Semitic	Tigray	53	39.5	13.5	[16]
		Afroasiatic	Semitic	Gurage	21	38.8	9.7	[16]
		Afroasiatic	C. Cushitic	Jews	70	37.5	12.6	[219, 397]
		Afroasiatic	E. Cushitic	Afar	16	41.0	13.0	[16]
		Afroasiatic	E. Cushitic	Oromo	33	39.3	8.6	[16]
		Mixed		Ethiopians	169	38.7	9.0	[222]
		Nilo-Saharan	Nilotic	Nyangatom	112	35.38	5.53	[526]
	Somalia	Afroasiatic	E.Cushitic	Somalia	27	45.4	2.1	[151, 152]
	Sudan	Nilo-Saharan	Nubian	Nubian	80	31.4	21.8	[154, 314]
	Rwanda	Niger-Kordofanian	Bantoid	Hutu	42	29.74	-2.60	[349]
Central Africa	Cameroon	Afroasiatic	Chadic	Chadic[d]	227	14.8	12.2	[204, 527]
		Niger-Kordofanian	Bantoid	Bamilike	47	10.07	5.45	[302]
		Niger-Kordofanian	Bantoid	Ewondo	78	11.52	3.87	[302, 351]
		Niger-Kordofanian	Adamawa	Mbum[e]	39	14.44	10.40	[204]
		Niger-Kordofanian	Adamawa	Fali	41	14.56	11.40	[204]
		Niger-Kordofanian	Ubangian	Baka	127	13.00	5.00	[351]
		Niger-Kordofanian	Bantoid	Bakola	88	10.30	2.75	[351]
		Niger-Kordofanian	Bantoid	Fang	105	11.37	1.88	[351]
		Niger-Kordofanian	Bantoid	Ngoumba	88	10.05	2.55	[351]

Region	country	Linguistic	Family	Group	n	Longitude	Latitude	Source
		Niger-Kordofanian	Bantoid	Bassa	46	9.98	4.13	[204]
		Niger-Kordofanian	Bantoid	Bakaka	50	9.75	4.72	[204]
		Niger-Kordofanian	N. Atlantic	Fulani	76	13.00	10.00	[204, 528]
		Niger-Kordofanian	Ubangian	Mbenzele	57	16.32	2.88	[302]
		Niger-Kordofanian	Bantoid	Tigar	35	11.13	2.33	[351]
		Niger-Kordofanian	Bantoid	Tikar	34	10.92	5.72	[368]
		Niger-Kordofanian	Bantoid	Aghem	115	10.07	6.38	[368]
		Niger-Kordofanian	Bantoid	Bamoun	107	11.50	6.08	[368]
	[c]D.R.C.	Nilo-Saharan	Mangbetu	Mbuti	52	29.97	1.85	[151, 152, 351]
	[c]C.A.R.	Niger-Kordofanian	Bantoid	Biaka	73	15.32	3.90	[151, 152, 351]
	Gabon	Niger-Kordofanian	Bantoid	Akele	48	12.50	-1.20	[351]
		Niger-Kordofanian	Bantoid	Ateke	54	13.58	-1.63	[351]
		Niger-Kordofanian	Bantoid	Duma	47	12.48	-1.13	[351]
		Niger-Kordofanian	Bantoid	Eshira	40	11.88	-1.78	[351]
		Niger-Kordofanian	Bantoid	Eviya	38	10.60	-1.22	[351]
		Niger-Kordofanian	Bantoid	Galoa	71	9.83	-1.72	[351]
		Niger-Kordofanian	Bantoid	Kota	56	11.93	-0.10	[351]
		Niger-Kordofanian	Bantoid	Makena	45	13.93	1.02	[351]
		Niger-Kordofanian	Bantoid	Mitsogo	64	11.93	-1.88	[351]
		Niger-Kordofanian	Bantoid	Ndumu	39	13.70	-1.70	[351]
		Niger-Kordofanian	Bantoid	Nzebi	63	12.20	-0.18	[351]
		Niger-Kordofanian	Bantoid	Obama	47	13.78	-0.68	[351]
		Niger-Kordofanian	Bantoid	Punu	52	9.55	-1.13	[351]
		Niger-Kordofanian	Bantoid	Shake	51	12.20	0.87	[351]

Region	country	Linguistic	Family	Group	n	Longitude	Latitude	Source
		Niger-Kordofanian	Ubangian	Bakoya	31	13.98	1.03	[351]
		Niger-Kordofanian	Ubangian	Benga	50	9.45	0.38	[351]
		Niger-Kordofanian	Ubangian	Bongo	45	11.50	-2.20	[351]
	Sao Tome	Creole		Sao Tome	153	6.53	0.03	[529, 530]
	Equatorial Guinea	Niger-Kordofanian	Bantoid	Bubi	45	9.90	1.80	[529]
	Chad	Niger-Kordofanian	N. Atlantic	Fulani	49	15.30	11.03	[528]
West Africa	Guinea-Bissau	Niger-Kordofanian	N. Atlantic	Bayot	50	-16.2	12.0	[307]
		Niger-Kordofanian	N. Atlantic	Balanta	62	-15.3	11.3	[307]
		Niger-Kordofanian	W. Mande	Malinke	31	-11.4	13.5	[307]
		Niger-Kordofanian	N. Atlantic	Manjaku	77	-15.3	12.1	[307]
		Niger-Kordofanian	N. Atlantic	Fulani	77	-9.30	10.38	39]
	Mali	Niger-Kordofanian	W. Mande	Bambara	52	-8.0	12.7	[306]
		Niger-Kordofanian	N. Atlantic	Malinke	31	-11.40	13.54	[307]
	Senegal	Niger-Kordofanian	Mande	Mandenka	110	-14.8	12.2	[151, 152, 184]
		Niger-Kordofanian	N. Atlantic	Senegalese	170	-15.86	16.39	[299, 401]
	Sierra Leone	Niger-Kordofanian	W. Mande/ N. Atlantic	Sierra Leone[f]	166	-11.95	8.72	[317]
	Burkina Faso	Niger-Kordofanian	N. Atlantic	Fulani	97	-0.46	14.68	[528]
	Nigeria	Niger-Kordofanian	Defoid	Yoruba	33	4.57	7.77	[151, 152]
		Niger-Kordofanian	Cross River	Anang	107	7.72	5.05	[368]
		Niger-Kordofanian	Cross River	Efik	145	7.98	5.17	[368]
		Niger-Kordofanian	Bantoid	Ejagham	133	8.32	4.95	[368]
		Niger-Kordofanian	Cross River	Ibibio	509	7.87	4.72	[368]

Region	country	Linguistic	Family	Group	n	Longitude	Latitude	Source
		Niger-Kordofanian	Igboid	Igbo	201	7.48	6.43	[368]
		Niger-Kordofanian	Cross River	Oron	98	8.23	4.83	[368]
		Niger-Kordofanian	N. Atlantic	Fulani	60	10.97	7.86	[151, 152]
		Afroasiatic	Chadic	Hausa	20	8.48	13.05	[151, 152]
	Ghana	Niger-Kordofanian	Kwa	Akan	151	-2.27	6.33	[368]
		Niger-Kordofanian	Kwa	Ewe	87	0.47	6.60	[368]
	Niger	Afroasiatic	Berber	Tuaregs	23	7.39	18.74	[151, 152]
	Cape Verde	Creole		Cabo Verde	292	-23.52	14.92	[318]
	Madeira-Azores	Indo-European	Romance	Madeiran	490	-25.67	37.74	[440, 441]
	Canary Island	Indo-European	Romance	Canarian	300	-15.50	28	[531]
Southern Africa	Mozambique	Niger-Kordofanian	Bantoid	Bantu[g]	416	32.6	-26.0	[298, 307]
	Angola	Niger-Kordofanian	Bantoid	Cabindans[a]	109	12.5	-5.0	[292]
		Niger-Kordofanian	Bantoid	Bantu[b]	365	12.2	-15.2	[297]
	Zimbabwe	Niger-Kordofanian	Bantoid	Shona	59	31.03	-17.86	[349]
	Madagascar	Niger-Kordofanian	Bantoid	Malagasy	170	47.52	-18.92	[532, 533]
	South Africa	Khoisan		!Kung	62	19.50	-18.33	[183, 193]
		Khoisan		Kwe	31	15.68	17.78	[183]
		Mixed		SAC	563	18.42	-33.92	[295]
North/NW Africa	Egypt	Afroasiatic	Semitic	Egyptians[h]	126	31.25	30.05	[154, 228, 314, 369]
		Afroasiatic	Semitic	Gurna	34	32.6	25.7	[369]

Region	country	Linguistic	Family	Group	n	Longitude	Latitude	Source
		Afroasiatic	Berber	Siwa	78	25.7	29.2	[402]
		Afroasiatic	Semitic	Egyptian[1]	277	29.92	31.20	[305]
	Libya	Afroasiatic	Berber	Tuaregs	129	13.00	24.50	[534]
	Algeria	Afroasiatic	Semitic	Arabs	47	3.1	36.8	[398]
		Afroasiatic	Berber	Mozabite	85	3.50	32.58	[171, 399]
	Tunisia	Afroasiatic	Berber	Berbers	182	10.5	32.9	[308, 310, 535]
		Afroasiatic	Semitic	Arabs	200	10.2	36.8	[227, 308-310]
	Morocco	Afroasiatic	Semitic	Arabs	169	-6.8	34.0	[309, 401, 536]
		Afroasiatic	Berber	Berbers	327	-4.50	33.50	[401, 402] [537]
	Western Sahara	Afroasiatic	Berber	Saharawi	81	-13.42	27.10	[398, 401]
	Mauritania	Afroasiatic	Semitic	Mauritanians	94	-16.0	18.1	[306, 401]
Middle East	Iraq	Afroasiatic	Semitic	Jews	135	44.4	33.3	[219]
		Afroasiatic	Semitic	Arabs	168	44.4	33.3	[314, 512]
	Jordan	Afroasiatic	Semitic	Arabs	145	-9.1	38.7	[442]
	Israel	Afroasiatic	Semitic	Bedouins	87	34.9	30.5	[219, 367, 538]

Region	country	Linguistic	Family	Group	n	Longitude	Latitude	Source
		Afroasiatic	Semitic	Druze	388	35.05	32.69	[438]
	Palestine	Afroasiatic	Semitic	Palestine	227	34.5	31.5	[219, 367]
	Yemen	Afroasiatic	Semitic	Arabs	350	45.2	12.8	[16, 312, 397]
		Afroasiatic	Semitic	Jews	164	44.2	15.4	[219, 397]
	Saudi Arabia	Afroasiatic	Semitic	Arabs	553	46.77	24.64	[313]
	UAE	Afroasiatic	Semitic	Dubai	249	55.28	25.25	[319]
	Syria	Afroasiatic	Semitic	Arabs	69	36.3	33.5	[314]
Mediterranean/Iberia	Spain	Indo-European	Romance	Spanish	800	-5.99	37.38	[315, 316]
	Italy	Indo-European	Romance	Italians	797	12.48	41.90	[314, 367, 539-541]
	France	Indo-European	Romance	French	788	2.35	48.85	[435]
	Portugal	Indo-European	Romance	Portuguese	449	-8.42	40.20	[436, 437]

E. – eastern, C. – Central, W. – Western, N. – Northern, SAC – South African colored populations. [a]Cabinda: either Kikongo or Yombe which are both Bantu, [b]Bantu: Ovimbundu, Kuvale, Nyaneka and Ganguela, [c]C.A.R.: Central African Republic, [c]D.R.C. – Democratic Republic of Congo, [d]Daba, Mandara, Ouldeme, Podokwo, Masa, Hide, Mafa, Kotoko, [e]Mbum: Tali, Tupuri, [f]Sierra Leone: Temne & Limba (Southern Atlantic), Loko & Mende (Mande), [g]Bantu: Yao, Tonga, Shangaan, Chopi, Chwabo, Lomwe, Makonde, Makhuwa, Ndau, Nguni, Nyungwe, Nyanja, Ronga, Shona, Sena, Tswa, [h]Egyptians: Egyptians sampled along the Nile Valley and west of the Nile, [i]Egyptians: Sampled in in port city of Alexandria

Appendix 6b (ii) part I: Frequency of *mt*DNA lineages in Africa, Mediterranean and the Near East

Group	n	L0a1	L0a2	L0a3	L0a4	L0b	L0d1&2	L0d3	L0f	L0k	L1b	L1c	L2a	L2a1	L2a2	L2b	L2c	L2d
Borana	59	8.47	3.39						5.08					6.78				
Gabra	31	3.23							6.45					6.45				
Garreh	26	7.69												15.38				
Burji	21	4.76	4.76								4.80							
Daasanech	49	8.16	4.08		2.04	2.04			16.33					6.12				
Konso	17		5.88											5.88				
Rendille	31	16.13	3.23						6.45					9.68				
Orma	20	10.00	5.00											5				
Oromo	33	3.03		3.33									3.33	6.06				
El-Molo	24	25.00	4.17						8.33									
Wata	24					4.2			4.17					12.50				
Yaaku	19								5.26									
Boni	30	6.67	26.67					3.30	23.33					16.67	3.33			
Afar	16	6.25									18.75			18.75				
Somali	27	3.74												14.81	11.11			
Sanye	20				45				10.00									
Akie	20		15.00					25.00	10.00									
Fyome	13	7.69	15.38	7.70					30.77					7.69				
Iraqw	23	4.35	8.70	4.30					17.39									
Burunge	25	4.00	24.00	4.00				4.00	28.00					4.00				
Beja	51	5.88							1.96					23.53				
Ethiopian Jews	70	2.86	7.14						2.86					8.57	1.43	1.43		
Amhara	120	5.83	2.5	0.83					0.83		0.83		2.50	12.5		3.33		
Gurage	21										4.76			9.52				
Tigray	53	3.77												5.66		1.89		
Baggara	21										4.8	4.76		23.81	4.76			
Giziga	26											3.85		23.08				
Chadic	227	7.49									5.29	4.85	2.20	12.78		5.29	2.64	0.44
Hausa	20										5.00	5.00		30.00				
Egypt Gurna	34	11.76												2.94				
Egypt (Arabs)	126	5.56									0.79	1.59		3.17	0.79	0.79		
Egyptian (Alexandria)	277	1.81							0.36		0.36	2.17	0.36	2.89		0.36		
Siwa	78	1.28									1.28			5.13				
Libyan Tuaregs	129	6.2									3.88			10.08		1.55		
Tuareg (Niger)	24	4.17									12.50			37.50			4.17	

Group	n	L0a1	L0a2	L0a3	L0a4	L0b	L0d1&2	L0d3	L0f	L0k	L1b	L1c	L2a	L2a1	L2a2	L2b	L2c	L2d
Tunisian Arabs	200	0.5									3	1	0.5	10.5		2		
Tunisian Berbers	182	1.10									2.20	3.85	1.10	8.24				2.20
Algeria (Arabs)	47	2.13									4.26			6.38			2.13	
Mozabites	85													5.88				
Moroccan Arabs	169	0.59									6.51	1.18		5.33			0.59	1.76
Moroccan Berbers	327								1.22		5.81			2.75		0.92	0.61	0.31
Western Saharans	81										3.70			9.88			6.17	1.23
Mauritanians	94										18.09	4.26	1.06	11.7		1.06	2.16	
Yemen	350	3.71	4.57							0.57	0.57	1.14	0.86	3.71		0.57		0.57
Yemenite Jews	164																	
Saudi Arabia	553	0.90			0.18						0.18	0.36	0.36	1.81	0.54	0.18	0.18	
Bedouin	87	4.60									2.30			3.45		2.30		
Palestine	227	0.44												4.41				
Druze	388													2.32				
Jordan	145	1.38									0.69		0.69	2.07			0.69	
Dubai	249	2.01	3.21								0.40	1.20		4.82				0.40
Iraq	168		0.60									1.19		1.79				1.19
Iraq Jews	135																	
Syria	69	1.45									1.45			2.90				
Portuguese	449										2.9	0.45	0.22	1.11		0.45		
Spanish	800										1.2		0.1	0.1		0.1		
Canarians	300										2		0.33	1.33				
French	788										0.25			0.38				
Italians	797	0.13									0.13			0.13	0.13			
South Italians	341										0.59							
Madeira-Azores	490										2.24	0.20	0.20	0.61		0.20	0.41	
Hadza	19		5.26											21.05				
Sandawe	25	8.00	12.00		4.00			4.00	8.00					8.00				
San	11						81.82			9.09								
Kung	62						64.52		1.61	19.35						3.23		
Kwe	31	6.45					16.13			22.58	3.23			9.68		6.45		
Samburu	30	3.33	3.33						10.00					3.33				
Il'Laikipiak	25	12.00	4.00						4.00					12.00				
Maasai	41	14.63	4.88		2.40				2.44					12.20				
Ilchamus	20	25.00	5.00											10.00				
Turkana[a]	86	8.14	11.63		1.16	3.49		1.16	4.65			1.16		8.14	4.65		1.16	
Teso	8	25.00																
Nyangatom	112	5.36	7.14			2.68			0.89		0.89		0.89	7.14		3.57		2.68
Kalenjin	26		7.69															
Marakwet	22	9.09							4.55					9.09				

Group	n	L0a1	L0a2	L0a3	L0a4	L0b	L0d1&2	L0d3	L0f	L0k	L1b	L1c	L2a	L2a1	L2a2	L2b	L2c	L2d
Tugen	43	23.26	4.65						2.33					9.30				
Pokot	39	17.95	15.38											15.38	2.56			
Sabaot	17	41.18												17.65				
Sengwer	27	3.70	3.70			3.70			3.70					11.11	3.70			
Ogiek	22													13.64				
Dorobo	13	7.69	15.38											15.38				
Datog	30	16.67	6.67		6.70				6.67			3.33		3.33				
Nubians	80	12.5							1.25		4		3.75	7.5				
Luo	21		4.76	4.80										4.76				
Dinka	33	3.03	6.06											15.15	12.12			
Nuer	22	9.09		4.50										13.64	4.55			
Shilluk	20			5							20				5.00			
Nyimang	12	16.67												16.67				
Kanuri[a]	40	5.00									10.00			15.00		2.50		
Laka	38	2.63	2.63								7.90			18.42	10.53			
Boulala	22	9.10												13.64	4.55			
Nairobi (Kenya)	100	8	7						8		2	2		9			1	
Gikuyu[a]	50	6	4.00						4.00					8.00				
Luhya	28	25.00			3.60	7.10								3.57	7.14			
Taita	29	3.45	27.59						17.24			10.34		10.34	10.34			
Taveta	19	21.05	5.26									5.26		5.26				
Pare	37	13.51	16.22									8.11		13.51				
Rangi	27	7.41	7.41						13.79					22.22				
Turu	24		16.67						12.50					8.33				
Hutu	42	11.9	4.76						16.67			2.38		7.14		2.38		2.38
Mozambique	416	9.61	14.66				4.81	0.24			1.20	5.00		32.94		1.44	0.72	0.72
Shona	59	6.78	20.34				3.39				3.39	8.47	3.39	16.95		1.69	1.69	
Kuvale	54	11.11	20.37				22.22				1.85	24.07					1.85	
Ganguela	21		28.57									23.81		14.29		4.76		
Nyaneka-Nkhumbi	153	10.46	3.27				5.88			0.65	2.61	16.99		9.15		2.61	8.50	0.65
Ovimbundu	92	4.35	5.43				2.17			1.09	7.61	17.39		14.13		4.35	2.17	
Xhosa	17	11.76	5.88									5.88		5.88				
Malagasy	170	1.76	4.71									1.18		3.53				
South African Colored	563	1.24	3.2				60.04				0.53	1.07		4.62			0.18	
Mbum	39	2.56									5.13	2.56		17.95		2.56		
Fali	41	11.80										4.92	7.32	24.39			4.88	
Medzan pygmies	28.0											96.43		3.57				
Mbenzele pygmies	57		3.51									96.49						
Baka pygmies	153	1.96	1.31									94.12			0.65			
Bakola pygmies	110											100.00						

Group	n	L0a1	L0a2	L0a3	L0a4	L0b	L0d1&2	L0d3	L0f	L0k	L1b	L1c	L2a	L2a1	L2a2	L2b	L2c	L2d
Bakoya pygmies	31											96.77						
Benga pygmies	50	10.00									10.00	44.00		4.00			2.00	
Biaka pygmies	73	1.37	9.59									86.30						
Bongo pygmies	45		2.22									82.22						2.22
Tigar pygmies	35											100.00						
Mbuti pygmies	52		30.77											15.38	44.23			
Mvae	23	4.35										52.17		17.39	4.35			
Bamileke	47	6.38	4.26								2.13	6.38		29.79		4.26	2.13	4.26
Tikar	34	2.94	11.76								2.94	11.76		26.47		2.94	2.94	
Aghem	115	5.22	2.61								6.96	12.17	0.87	11.30		2.61		
Bassa	46	4.35									6.52	23.91		15.22				4.35
Bakaka	50	10	10								6	14		10		2	4	
Bubi	45		8.89									22.22		17.78				17.78
Bamoun	125	13.6	3.2								4.8	9.6	3.2	14.4		0.8		
Akele	48	8.33									4.17	39.58		14.58			2.08	
Ateke	54	7.41	5.56								11.11	27.78		12.96			3.70	
Duma	47	6.38	6.38								8.51	34.04		12.77		2.13		2.13
Eshira	40	7.50	5.00								12.50	40.00		7.50			2.50	5.00
Eviya	38	15.79									2.63	31.58		21.05		13.16	7.89	
Ewondo	78	10.26	1.28								7.69	29.49	1.28	11.54			2.56	5.13
Fang	105	3.81									10.48	33.33		14.29				0.95
Galoa	71	5.63	4.23									32.39		21.13			7.04	1.41
Kota	56	10.71	1.79								3.57	35.71		5.36		5.36		3.57
Makena	45	4.44	2.22								13.33	48.89		8.89				
Mitsogo	64	3.13	1.56									39.06		26.56		3.13		1.56
Ndumu	39	5.13	7.69								7.69	28.21		10.26	10.26		5.13	2.56
Ngoumba	88	6.82	5.68								6.82	30.68		11.36				
Nzebi	63	3.17	1.59								7.94	39.68		4.76		3.17	1.59	
Obama	47	6.38									4.26	31.91		17.02			2.13	2.13
Punu	52	3.85										28.85		15.38		1.92	1.92	3.85
Shake	51	3.92	5.88								3.92	43.14		17.65		1.96	1.96	
Angola-Cabinda	109	10.09	2.75								2.75	24.77		4.59		6.42	1.83	
Sao Tome Principe	153	5.88	0.65								14.38	16.34		15.03		1.96	5.88	0.65
Fulani (Cameroon)	99	2.02									19.19	4.04		5.05	2.02	5.05	5.05	1.01
Fulani (Chad)	49										28.57			6.12		6.12	4.08	
Yoruba	33	3.03									18.18	6.06		21.21		3.03		
Anang	107	2.80									4.67	16.82	0.93	14.02		4.67	2.80	1.87
Efik	145	5.52									11.03	18.62	0.69	20.00		6.21	2.07	0.69
Ejagham	133	5.26									7.52	9.77	1.50	14.29		4.51	3.01	1.50
Ibibio	509	4.91	0.39								10.22	16.70	1.38	19.25		3.93	2.75	1.38

Group	n	L0a1	L0a2	L0a3	L0a4	L0b	L0d1&2	L0d3	L0f	L0k	L1b	L1c	L2a	L2a1	L2a2	L2b	L2c	L2d
Igbo	201	6.97									11.44	15.42	1.00	16.92		4.48	2.99	0.50
Oron	98	4.08	1.02								7.14	12.24	2.04	11.22		6.12	7.14	3.06
Akan	151	1.32									8.61	10.60	5.30	27.15		2.65	5.30	
Ewe	87										8.05	4.60	1.15	27.59		5.75	2.30	
Fulani (Burkina Faso)	97										28.87			4.12		9.28	11.34	3.09
Fulani (Nigeria)	60										18.33			15.00			3.33	
Mandenka	110	1.82									20	1.82		11.82		2.73	39.09	
Malinke	31	3.23									16.13			25.81		6.45	22.58	
Sierra Leone	276	3.26									14.49	5.8	0.72	19.2		2.71	12.32	4.35
Fulani (Guinea-Bissau)	77	1.30									16.88	5.19		22.08			15.58	2.60
Guinea-Bissau	295	6.10									8.47	5.08	2.71	12.54		9.83	15.93	1.69
Bambara	52	3.85									5.77	9.62	1.92	19.23		1.92	11.54	1.92
Senegalese	170										15.88	3.53	1.18	16.47		10.59	8.82	
Cape Verde	292	0.68									7.88	6.85	3.08	17.47		4.11	16.44	

Appendix 6b (ii) part II: Frequency of *mt*DNA lineages in Africa, Mediterranean and the Near East

Group	n	L2e	L2f	L3*	L3a	L3b	L3c	L3d	L3e	L3f	L3h	L3i	L3j	L3k	L3x	L4	L5	L6	M1a
Borana	59				3.39			5.08		5.08	3.39	15.25			8.47	10.17	3.4	1.7	10.17
Gabra	31									3.23	6.45	3.23			6.45	9.68			22.58
Garreh	26				3.85			7.69		15.38	3.85					7.69	3.8		19.23
Burji	21				4.76						4.76				33.33	19.05	9.5	4.76	9.5
Daasanech	49								2.04		8.16	6.12			2.04	18.4	20.4		4.08
Konso	17									11.76	5.88				29.41	11.76	5.9		23.53
Rendille	31				3.23					3.23	6.45				3.23	12.90		3.23	9.68
Orma	20				15			10.00	5.00	10.00					10.00	10.00			10.00
Oromo	33							3.03		9.09		3.03			12.12	3.03	3.03		6.06
El-Molo	24				25.00						8.33								
Wata	24				4.17			4.17		20.83		4.17				4.17			29.17
Yaaku	19				26.32			5.26		5.26						31.58	5.26		15.79
Boni	30					3.33				3.33					6.67		3.3		
Afar	16							12.5		6.25						6.25	6.25		
Somali	27			7.41				3.74		7.47		11.11	3.74	3.74	7.47				11.11
Sanye	20				15.00											30.00			
Akie	20									5.00	5.00				5.00	35.00			
Fyome	13															23.08			7.69
Iraqw	23				8.70					13.04	8.70					13.04			8.70
Burunge	25							4.00	8.00	4.00	4.00				4.00	4.00			
Beja	51					1.96		1.92		27.45		3.92			1.96	1.96		1.96	9.80
Ethiopian Jews	70					2.86					4.29	1.43			2.86	4.29	7.14	4.29	15.71
Amhara	120							2.5		5	1.67	5.83			2.50	6.67	3.33	0.83	10.00
Gurage	21									4.76		4.76			4.76	4.76	9.52	4.76	9.52
Tigray	53									5.66		3.77			1.89	11.32			11.32
Baggara	21					4.76		28.57	14.28	4.76									
Giziga	26					19.23		11.54	3.85	19.23	7.69				3.85				
Chadic	227	1.32		1.32		13.22		5.73	17.62	14.54	0.44					2.20			

Group	n	L2e	L2f	L3*	L3a	L3b	L3c	L3d	L3e	L3f	L3h	L3i	L3j	L3k	L3x	L4	L5	L6	M1a
Hausa	20			10.00		20.00			20.00	5.00									
Egypt Gurna	34						2.94								2.94		5.88		17.65
Egypt (Arabs)	126			0.79	0.79	1.59			0.79	0.79		1.59			1.59		1.59		2.38
Egyptian (Alexandria)	277			1.08		1.44	0.36	1.44	2.89	4.69	0.36				0.36				4.69
Siwa	78						3.85		11.54			1.28				1.28			16.67
Libyan Tuaregs	129								7.75	1.55		2.32							1.55
Tuareg (Niger)	24					4.17		4.17	8.33	4.17	4.17								
Tunisian Arabs	200	2				3		1.5	4	5.5	2.5					1		0.5	
Tunisian Berbers	182	0.55				3.30		0.55	3.85	1.10	0.55			1.10					1.10
Algeria (Arabs)	47					2.13		2.13	10.64										10.64
Mozabites	85					3.53		1.18	2.35										
Moroccan Arabs	169					1.18		2.96	2.37	4.14	0.59					0.59			2.37
Moroccan Berbers	327			0.61		2.45		0.31	8.56								0.31		1.22
Western Saharans	81			1.23		4.94		3.70	1.23	3.70					1.23				
Mauritanians	94			1.06		4.26			1.06										
Yemen	350			0.29		1.72		3.71	3.14	2.29	1.14	1.14			0.28	0.86		4.00	1.14
Yemenite Jews	164						3.05			1.22					9.75	2.44		1.22	1.22
Saudi Arabia	553					0.54	0.18	0.54		0.72	0.36	0.36			0.72	0.36	0.72	0.18	2.89
Bedouin	87			1.15															2.30
Palestine	227			1.32					2.20	0.44	0.44				0.44				1.32
Druze	388													0.52					1.29
Jordan	145					3.45		0.69		1.38		0.69			0.69				1.38
Dubai	249					2.81		2.41	2.41	0.80		0.40				1.20			
Iraq	168									1.19									
Iraq Jews	135																		2.22
Syria	69									1.45									
Portuguese	449			0.67		0.45			1.34	0.67					0.45				0.22
Spanish	800							0.5			0.1				0.1				0.1
Canarians	300					0.67		1	0.67	0.33					0.33				
French	788			0.13					0.13	0.13									

Group	n	L2e	L2f	L3*	L3a	L3b	L3c	L3d	L3e	L3f	L3h	L3i	L3j	L3k	L3x	L4	L5	L6	M1a
Italians	797									0.13									0.13
South Italians	341					0.29			0.59							0.29			0.59
Madeira-Azores	490							0.20	1.02	0.41					0.61				1.84
Hadza	19										5.26					68.42			
Sandawe	25					4.00			16.00						4.00	24.00	4.0		
San	11					9.091													
Kung	62					8.06			3.23										
Kwe	31					3.23			29.03										
Samburu	30				23.33			6.67				13.33				10.00	3.3		10.00
Il'Laikipiak	25							4		4.00		4.00			4.00	16.00	4.0		24.00
Maasai	41				2.44			4.88	9.76	2.44	7.32	2.44			4.88	7.32	7.3		4.88
Ilchamus	20									15.00		5.00			5.00	10.00	10.0		15.00
Turkana[a]	86				1.16	1.16	2.33	3.49	1.16	3.49	4.65	2.33			3.49	13.95	12.8		2.32
Teso	8					62.50											12.5		
Nyangatom	112					5.36	0.89	4.46	2.68	3.57	1.79	13.4			6.25	9.82	11.6	2.68	0.89
Kalenjin	26				3.80	3.85					15.38	3.85				34.62	19.2		11.54
Marakwet	22				4.55				9.09		9.09	9.09				31.82	9.09		
Tugen	43				2.33	2.33			2.33	2.33	11.63	6.98				9.30	14.0		4.65
Pokot	39				2.56					2.56	5.13	10.26				12.82	12.8		
Sabaot	17					11.76			5.88		11.76	5.88							5.88
Sengwer	27							3.70			7.41	3.70				11.11	29.6		
Ogiek	22									4.55		4.55				22.73	4.5		36.36
Dorobo	13										23.08	7.69				23.08			
Datog	30								3.33	3.33	20.00					13.33		3.33	10.00
Nubians	80									10	1.25	7.5				1.25	3.75		10
Luo	21					33.33		14.29	14.29	4.76	4.76					9.52			
Dinka	33					3.03		3.03		3.03	24.24					6.06	18.2		
Nuer	22					4.55			9.09	9.09	13.64					18.18	13.64		
Shilluk	20					5.00		10.00		5.00	5.00					20.00	15.0		16.25
Nyimang	12								8.33	8.33						8.33	16.67		16.67

Group	n	L2e	L2f	L3*	L3a	L3b	L3c	L3d	L3e	L3f	L3h	L3i	L3j	L3k	L3x	L4	L5	L6	M1a
Kanuri[a]	40			5.00		10.00		10.00	22.50	15.00									
Laka	38	2.50				7.89		15.79	15.79	7.89					2.63				
Boulala	22					9.09		9.09	9.09	27.27		4.55							
Nairobi (Kenya)	100				2	7			10	9	6	1			2	16	3		4
Gikuyu[a]	50			6.00	4.00	4.00			16.00	6.00	10.00	6.00	4.00			8.00	8.00		2.00
Luhya	28				3.57	7.14		7.14	7.14	3.57	3.57	3.57			3.57		14.3		
Taita	29					3.45			3.45	3.45					6.90	3.45			
Taveta	19				10.53	5.26		21.05	5.26						5.26	5.26	10.5		
Pare	37				8.11	2.70		2.70	13.51	2.70						8.11	8.1		
Rangi	27							7.41	14.81	3.70	7.41					11.11			
Turu	24				4.17	12.50		8.33	4.17	4.17	4.17					20.83			
Hutu	42				4.76	16.67		2.38	7.14		7.14				2.38	7.14	4.76		
Mozambique	416			0.96		2.88		5.52	15.12	2.4							0.48		
Shona	59					6.78		6.78	18.64							1.69			
Kuvale	54		1.85						9.26	5.56	1.85								
Ganguela	21							9.52	19.05										
Nyaneka-Nkhumbi	153					3.27		4.58	22.88	7.84									
Ovimbundu	92		2.17			3.26		6.52	22.83	5.43									
Xhosa	17							5.88	23.53										
Malagasy	170					12.35		0.59	4.71	0.59									
South African Colored	563							2.84	3.55	0.53						0.89	0.36		
Mbum	39			7.69		5.13		15.38	17.95	17.95	5.13								
Fali	41	2.44				12.20		12.20	2.44	4.88						7.32	4.88		
Medzan pygmies	28.0																		
Mbenzele pygmies	57																		
Baka pygmies	153							1.31	0.65										
Bakola pygmies	110																		
Bakoya pygmies	31								3.23										
Benga pygmies	50					2.00		4.00	20.00	4.00									
Biaka pygmies	73		2.74																

Group	n	L2e	L2f	L3*	L3a	L3b	L3c	L3d	L3e	L3f	L3h	L3i	L3j	L3k	L3x	L4	L5	L6	M1a
Bongo pygmies	45									13.33									
Tigar pygmies	35																		
Mbuti pygmies	52																9.62		
Mvae	23							4.35	8.70	4.35									
Bamileke	47					10.64		6.38	10.64	8.51									
Tikar	34					2.94		11.76	17.65	2.94									
Aghem	115	5.22				4.35		7.83	25.22	3.48	6.09					2.61			
Bassa	46			2.17		6.52		10.87	13.04	10.87						2.17			
Bakaka	50						2	2	30	8						2			
Bubi	45								33.33										
Bamoun	125	0.8				4	0.8	11.2	25.6	4						1.60			
Akele	48							2.08	20.83	8.33									
Ateke	54							3.70	24.07	3.70									
Duma	47							8.51	14.89	4.26									
Eshira	40							2.50	5.00	5.00	7.50								
Eviya	38					2.63		2.63	2.63										
Ewondo	78					7.69		6.41	5.13	5.13						2.56			
Fang	105					9.52		1.90	17.14	6.67						1.90			
Galoa	71					1.41		11.27	8.45	5.63						1.41			
Kota	56					1.79		5.36	7.14	17.86	1.79								
Makena	45					4.44			11.11	2.22						2.22			
Mitsogo	64							4.69	18.75	1.56									
Ndumu	39					5.13			17.95										
Ngoumba	88					1.14		5.68	12.50	12.50	4.55					2.27			
Nzebi	63					1.59		6.35	22.22	6.35						1.59			
Obama	47					2.13		2.13	21.28	8.51				2.13					
Punu	52					9.62		1.92	23.08	9.62									
Shake	51							3.92	13.73	1.96						1.96			
Angola-Cabinda	109					3.67		5.5	22.02	12.84						0.92			
Sao Tome Principe	153			0.65		4.58		3.92	24.84	3.27				0.65			0.7		

Group	n	L2e	L2f	L3*	L3a	L3b	L3c	L3d	L3e	L3f	L3h	L3i	L3j	L3k	L3x	L4	L5	L6	M1a
Fulani (Cameroon)	99			4.04		23.23		7.07	13.13	2.02				1.01					
Fulani (Chad)	49					34.69			12.24										
Yoruba	33					12.12		6.06	21.21	6.06									
Anang	107	0.93		0.93		8.41		5.61	25.23	7.48									
Efik	145	1.38				4.14		2.07	21.38	4.14						1.38			
Ejagham	133			0.75		3.76		8.27	27.07	10.53						0.75			
Ibibio	509	0.20		1.38		6.48		3.14	17.68	7.86	0.20					1.57			
Igbo	201	1.00		0.50		3.48	0.50	2.49	21.39	9.45					0.50				
Oron	98	1.02				5.10		1.02	26.53	11.22									
Akan	151	1.32		0.66		5.96		3.31	19.87	5.96	0.66								
Ewe	87			2.30		11.49		8.05	17.24	4.60									
Fulani (Burkina Faso)	97					31.96				1.03									
Fulani (Nigeria)	60					21.67		8.33	15.00	8.33									
Mandenka	110					8.18		9.09	4.55										
Malinke	31					9.68		9.68	6.45										
Sierra Leone	276			0.36		11.59	0.36	9.78	7.61	5.07	0.72				0.36				
Fulani (Guinea-Bissau)	77					12.99		5.19	2.60	1.30	1.30								
Guinea-Bissau	295					7.46		9.83	8.47	2.37	4.07								1.36
Bambara	52					21.15		9.62	9.62	1.92									
Senegalese	170			0.59		14.12		7.65	6.47	3.53	1.18			0.59					
Cape Verde	292	0.34				10.62		7.19	14.73	0.68	0.68			0.68					

Note that the lineage frequencies do not add up to 100 for all the populations, especially those from East Africa, the Near East and the Mediterranean. The remaining frequency constitutes the N and M lineages that are mainly observed outside Africa **(including U6, HV1, N1a, I, K, R0a, J - Appendix 8, Appendix 12).**

Appendix 7: *mt*DNA D-loop and Y chromosome diversity in East and representative central African populations, and Haplotype age estimates for *mt*DNA haplotype from Africa

Table A7ai: *mt*DNA D-loop diversity in East and representative central African populations generated using Arlequin Program [251]. Hunter-gatherer populations are colored blue.

Population	Genic Diversity
Baggara	0.9952 +/- 0.0165
Bamoun	0.9869 +/- 0.0229
Fulani	0.9605 +/- 0.0272
Giziga	0.9969 +/- 0.0117
Kanuri	0.9938 +/- 0.0126
Mvae	0.9881 +/- 0.0163
Baka	0.8800 +/- 0.0340
Bakola	0.9437 +/- 0.0245
Medzan	0.7619 +/- 0.0738
Boulala	0.9784 +/- 0.0213
Laka	0.9929 +/- 0.0084
Burji	0.9947 +/- 0.0178
Boni	0.9701 +/- 0.0171
Borana	0.9942 +/- 0.0044
El-Molo	0.8498 +/- 0.0426
Gabra	0.9718 +/- 0.0186
Gikuyu	0.9943 +/- 0.0119
Garreh	0.9969 +/- 0.0117
Il'Laikipiak	0.9800 +/- 0.0175
Kipsigis	0.9524 +/- 0.0955
Konso	0.9632 +/- 0.0328
Luo	0.9952 +/- 0.0165
Luhya	0.9974 +/- 0.0104
Marakwet	0.9740 +/- 0.0217
Ilchamus	0.9842 +/- 0.0205
Ogiek	0.8658 +/- 0.0402
Orma	0.9842 +/- 0.0205
Pokot	0.9879 +/- 0.0092
Pare	0.9775 +/- 0.0115
Rendille	0.9957 +/- 0.0095
Sabaot	0.9853 +/- 0.0252
Samburu	0.9793 +/- 0.0154
Sengwer	0.9573 +/- 0.0230
Sanye	0.7526 +/- 0.0740
Taita	0.9754 +/- 0.0163
Tugen	0.9834 +/- 0.0097
Turkana	0.9933 +/- 0.0074
Taveta	0.9942 +/- 0.0193
Wata	0.9819 +/- 0.0184
Yaaku	0.9532 +/- 0.0305
San	0.9636 +/- 0.0510
Beja	0.9216 +/- 0.0283
Dinka	0.9962 +/- 0.0086
Nuer	0.9957 +/- 0.0153
Nyimang	0.9848 +/- 0.0403
Shilluk	0.9789 +/- 0.0245
Akie	0.9632 +/- 0.0255
Burunge	0.9433 +/- 0.0366
Dorobo	0.9872 +/- 0.0354
Datog	0.9885 +/- 0.0114
Hadza	0.8246 +/- 0.0621
Iraqw	0.9921 +/- 0.0154
Maasai	0.9951 +/- 0.0066
Rangi	0.9943 +/- 0.0119
Sandawe	0.9833 +/- 0.0153
Turu	0.9891 +/- 0.0152
Fiome	0.9487 +/- 0.0506
Xhosa	0.9926 +/- 0.0230

Table A7aii: Nei [542] Genic diversity of Y Chromosomes in East African populations based on haplogroup frequencies generated using Arlequin Program [251].

Population	Genic Diversity
Akie	0.9231 +/- 0.0500
Arabi Shuwa	0.9127 +/- 0.0278
Baka	0.6092 +/- 0.0490
Bakola	0.7600 +/- 0.0738
Bamoun	0.5205 +/- 0.1229
Beja	0.6818 +/- 0.1019
Boni	0.9143 +/- 0.0347
Borana	0.8088 +/- 0.0425
Burji	0.8538 +/- 0.0557
Burunge	0.8387 +/- 0.0299
Datog	0.8204 +/- 0.0432
El-Molo	0.9524 +/- 0.0403
Fulani	0.8103 +/- 0.0448
Fyome	0.9064 +/- 0.0334
Gabra	0.7833 +/- 0.0715
Garreh	0.5869 +/- 0.1000
Gikuyu	0.8188 +/- 0.0486
Giziga	0.6646 +/- 0.1006
Hadza	0.7735 +/- 0.0522
Il'Laikipiak	0.7273 +/- 0.1086
Ilchamus	0.8042 +/- 0.0440
Iraqw	0.7243 +/- 0.0667
Kanuri	0.7931 +/- 0.0497
Konso	0.8304 +/- 0.0630
Laka	0.8571 +/- 0.0291
Luhya	0.9006 +/- 0.0342
Luo	0.5926 +/- 0.1027
Maasai	0.9000 +/- 0.0348
Mandara	0.7029 +/- 0.0623
Marakwet	0.8263 +/- 0.0607
Mbugu	0.6000 +/- 0.0771
Mbugwe	0.8381 +/- 0.0554
Medzan	0.9118 +/- 0.0424
Mvae	0.3500 +/- 0.1478
Nuer	0.5275 +/- 0.0638
Ogiek	0.7681 +/- 0.0766
Ouldoume	0.2537 +/- 0.1000
Pare	0.8333 +/- 0.0546
Pokot	0.8393 +/- 0.0354
Rangi	0.8931 +/- 0.0273
Rendille	0.8796 +/- 0.0230
Sabaot	0.7792 +/- 0.0606
Samburu	0.8939 +/- 0.0627
Sandawe	0.8758 +/- 0.0180
Sanye	0.8485 +/- 0.0744
Sengwer	0.7882 +/- 0.0483
Sukuma	0.8667 +/- 0.1291
Taita	0.5563 +/- 0.0863
Taveta	0.9048 +/- 0.0456
Tugen	0.6595 +/- 0.0794
Turkana	0.8952 +/- 0.0384
Turu	0.9187 +/- 0.0300
Wata	0.6944 +/- 0.1470
Yaaku	0.6993 +/- 0.0780

Table A7aiii: Nei [542] Genic diversity of Y Chromosomes in East African populations based on 14 Y chromosome microsatellites generated using Arlequin Program [251].

	Population	Average genic diversity over loci	
1	Burunge	0.627466 +/-	0.323596
2	Datog	0.617930 +/-	0.320031
3	Hadza	0.631579 +/-	0.326331
4	Iraqw	0.575729 +/-	0.298853
5	Akie	0.579710 +/-	0.310779
6	Maasai	0.627484 +/-	0.325369
7	Mbugu	0.310204 +/-	0.173118
8	Mbugwe	0.567771 +/-	0.299851
9	Pare	0.546886 +/-	0.289289
10	Rangi	0.646753 +/-	0.335235
11	Sandawe	0.609584 +/-	0.313949
12	Turu	0.562152 +/-	0.294481
13	Fyome	0.602537 +/-	0.315550
14	Boni	0.561473 +/-	0.296000
15	Borana	0.618042 +/-	0.320216
16	Burji	0.605935 +/-	0.316865
17	El-Molo	0.658385 +/-	0.349665
18	Gabra	0.543186 +/-	0.290068
19	Gareeh	0.526108 +/-	0.281697
20	Gikuyu	0.508213 +/-	0.269570
21	Ilchamus	0.582610 +/-	0.304549
22	Illaikipiak	0.605991 +/-	0.320119
23	Konso	0.655971 +/-	0.343857
24	Luhya	0.625275 +/-	0.327346
25	Luo	0.446414 +/-	0.239187
26	Marakwet	0.619793 +/-	0.325140
27	Ogiek	0.577031 +/-	0.305362
28	Pokot	0.674593 +/-	0.346705
29	Rendille	0.572257 +/-	0.300572
30	Sabaot	0.530051 +/-	0.280436
31	Samburu	0.601113 +/-	0.322709
32	Sanye	0.598681 +/-	0.319034
33	Sengwer	0.644904 +/-	0.334826
34	Taveta	0.570549 +/-	0.305175
35	Taita	0.370257 +/-	0.204309
36	Tugen	0.544655 +/-	0.285243
37	Turkana	0.646812 +/-	0.334847
38	Yaaku	0.477847 +/-	0.254860
39	Beja	0.604686 +/-	0.321029
40	Nuer	0.542627 +/-	0.289144
41	Dorobo	0.433281 +/-	0.246430

Appendix 7b: Haplotype age estimates for *mt*DNA haplotype from Africa

*Mt*DNA haplogroup/haplotype ages were inferred using the rho statistic (ρ) [260], an unbiased estimator, which is the average transitional distances of individual sequences from the inferred root haplotype. Coalescence ages of haplogroups/haplotypes and their standard errors were similarly calculated based on the network, by means of the average transitional distance from their respective root haplotypes, ρ using four different methods suggested by the following studies; Forster *et al.* [260], Mishmar *et al.* [178], Kivisild *et al.* [220] and Soares *et al.* [190]. Based on phylogenetic analysis of a total of 5515 complete sequences collated from published data including 222 from this study, the human mitochondrial TMRCA is 130 - 210 kya, which is slightly older than the 111 - 148 kya estimated by Watson [152] and the estimates of 101 - 133 kya based on inter-specific calibration [543] using complete *mt*DNA sequences and African RFLP data [34]. However, the range in TMRCA age estimates in this study is consistent with the oldest archeological evidence of anatomically modern human in Africa (from Omo- Kibish) that dates to around 150 - 195 kya [544].

Haplogroups[a]	n	[b]Mishmar *et al.*, 2003	[c]Kivisilid *et al.*, 2006	[d]Soares *et al.*, 2009
L	5515	193.34±23.84	155.41±24.3	173.32±42.77
L1'2'3'4'5'6	5344	168.41±21.75	134.98±22.16	173.68±44.12
L2'3'4'5'6	5205	148.09±19.81	128.28±21.71	175.03±45.29
L2'3'4'6	5156	117.25±15.55	94.33±15.86	114.98±29.48
L3'4'6	4939	107.34±14.49	87.06±15.09	115.43±30.74
L3'4	4932	91.9±11.46	66.72±9.54	95.35±23.25
L0	171	128.63±14.05	112.58±15.35	141.97±30.54
L1	139	128.86±14.35	112.17±15.48	123.26±27.97
L2	217	83.08±13.13	78.8±16.08	84.44±23.74
L3 (L3s + N & M)	4910	81.64±8.93	66.65±9.58	75.15±11.75
M	1416	55.76±4.36	45.11±4.53	73.3±9.64
N	3192	62.1±6.09	47.91±6.98	76.9±17.54
L4	22	89.21±10.15	81.17±11.4	79.8±19.22
L5	49	139.25±16.58	128.79±18.42	103.37±29.33
L6	7	23.49±6.05	18.36±5.88	28.83±13.52
L0a'b'f'k	131	114.6±15.78	98.98±16.47	135.41±33.2
L0a'b'f	122	86.5±12.31	82.66±14.72	137.62±35.51
L0a'b	102	61.66±11.18	53.38±12.5	119.3±36.62
L0a	97	39.46±5.63	31.59±5.87	59.71±16.05
L0b	5	78.1±15.72	62.23±16.9	60.54±20.97
L0d	40	94.67±11.43	72.21±11.93	77.19±21.21
L0f	20	88.37±9.52	81.17±10.52	67.6±15.53
L0k	9	39.39±10.82	42.09±13.98	65.02±31.47
L0a1	54	21.5±3.65	14.91±3.16	47.46±15.77
L0a2	32	33.88±6.23	32.97±7.62	30.27±8.08
L0a1a	17	13.9±2.87	9.15±1.99	11.87±4.11
L0a1a1	14	14.68±3.36	9.66±2.27	10.09±4.32
L0a1b	17	16.62±4.63	14.72±5.23	16.62±5.04
L0a1b1	9	10.28±5.11	8.27±5.47	6.73±3.88
L0a1b-d	35	22.02±4.82	17.59±4.75	43.24±13.34
L0a1c	7	15.41±5.14	11.6±5.47	14.41±10.39
L0a1d	10	4.62±1.7	4.73±2.03	18.16±6.69
L0a2a	17	21.76±5.84	23.48±7.48	15.43±4.6
L0a2a1	7	18.35±5.83	14.49±6.48	31.71±10.39
L0a2a2	10	10.28±4.05	11.5±5.2	4.04±2.85
L0a2b	9	7.99±2.68	3.01±1.5	17.94±8.39
L0a2c	6	29.12±6.41	25.93±7.22	60.54±24.72

Haplogroups[a]	n	[b]Mishmar *et al.*, 2003	[c]Kivisilid *et al.*, 2006	[d]Soares *et al.*, 2009
L0a3	6	37.68±9.22	34.95±10.87	30.27±17.48
L0d1'2	32	83.33±11.2	62.14±10.62	59.28±15.37
L0d1	22	51.61±8.67	37.82±8.61	75.22±21.32
L0d2	10	65.77±11.14	52.08±11.18	24.22±14.83
L0d3	8	26.97±6.67	24.52±7.8	27.75±11.56
L0d1c	11	24.29±7.53	17.83±7.99	38.53±21.32
L0d3a	6	23.12±6.69	22.55±7.65	20.18±11.65
L0f1	6	77.07±13.32	71.02±14.91	67.27±23.78
L0f2	9	75.93±12.96	62.38±12.78	53.81±22.2
L0f3	5	33.91±8.16	35.17±9.57	32.29±15.1
L1b	60	22.44±7.36	18.83±7	9.08±3.28
L1c	79	94.44±10.29	84.34±11.35	92.98±20.67
L1b1	59	17.16±5.44	18.8±7.12	9.23±3.33
L1b1a	56	12.11±2.51	12.32±3.21	9.37±3.49
L1b1a3	23	13.85±5.49	15.29±7.16	4.39±1.96
L1b1a3a	21	8.07±3.05	7.73±3.87	2.88±1.66
L1c1'2'4'6	67	87.58±10.59	84.9±12.99	98.19±24.06
L1c1	46	63.33±10.44	60.88±12.56	85.55±26.19
L1c1a	38	51.38±10.08	48.24±11.61	91.34±31.44
L1c1a1	26	34.78±10.39	36.94±13.3	103.23±40.2
L1c1a1a	25	29.39±9.49	30.84±12.06	44.4±22.9
L1c1a1a1	23	14.74±5.15	12.06±5.82	26.32±14.47
L1c1a2	12	23.98±7.24	24.24±8.75	1.68±1.68
L1c1b'c'd	8	71.93±10.97	56.65±10.73	58.02±19.7
L1c2'4	20	97.62±14.46	81.51±15.53	82.74±26.62
L1c2	16	52.99±8.24	43.12±8.95	42.88±13.82
L1c2a	10	38.02±7.48	35.85±8.95	46.41±21.07
L1c2a1	6	30.83±10.13	31.57±12.75	16.82±16.82
L1c3	12	73.22±10.57	47.35±9.8	63.9±21.67
L1c3a	7	35.23±6.14	25.12±5.63	31.71±11.89
L2a'd'f	214	77.55±12.28	71.72±14.83	83.83±24.06
L2a'f	164	67.48±14.51	69.91±17.75	86.63±30.73
L2a	163	36.82±7.4	36.19±9.5	66.73±23.43
L2b'c	44	68.9±11.63	40.12±9.63	64.21±22.99
L2b	25	30.83±6.73	25.43±7.27	34.71±17.7
L2c	19	22.44±4.35	14.95±3.1	23.37±13.27
L2d	6	26.55±8.35	21.42±8.66	30.27±15.41
L2a1	150	24.94±3.26	22.01±3.72	43.19±15.06
L2a2	13	34.78±8.02	30.18±8.42	24.84±15.52

Haplogroups[a]	n	[b]Mishmar *et al.*, 2003	[c]Kivisilid *et al.*, 2006	[d]Soares *et al.*, 2009
L2a1a	38	13.25±2.61	11.04±2.78	18.59±11
L2a1a2	20	15.93±4.54	13.19±4.72	9.08±4.16
L2a1a2a	10	16.96±6.95	16.23±8.82	18.16±8.32
L2a1b'f'i'j'k'l	76	17.58±2.68	11.48±1.42	26.29±5.75
L2a1b'f'i-n	67	16.95±2.98	10.5±1.49	17
L2a1c'd'e'h	34	32.04±5.51	28.05±4.56	20.77±5.28
L2a1i'j'k'l	35	17.76±2.06	14.69±2.29	32.86±9.46
L2a1b	8	10.28±6.02	6.76±5.21	73.15±33.94
L2a1c	20	19.01±3.02	16.91±3.1	15.13±5.8
L2a1e	9	21.69±6.66	22.55±8.3	-
L2a1f	32	18.3±5.61	9.3±1.84	7.57±2.68
L2a1i'g	12	20.12±3.85	17.47±4.26	38.68±16.22
L2a1g	6	18.84±5.42	14.66±5.64	30.27±18.11
L2a1i	5	21.58±6.25	18.94±7.16	12.11±6.99
L2a1l	9	13.13±4.16	12.02±5.21	11.21±5.01
L2b1	17	15.41±3.32	12.73±3.77	23.74±15.84
L2b1a	13	15.02±4.11	11.97±4.68	-
L2c2	12	17.55±3.85	14.66±4.29	3.36±2.38
L3e'i'k'x	121	49.13±7.37	42.32±7.48	55.2±12.73
L3b'c'd'j	93	65.25±10.16	52.66±10.51	66.18±21.84
L3c'd'j	40	68.21±12.44	54.11±12.02	51.46±20.07
L3a	11	45.77±8.69	39.97±9.26	47.7±16.81
L3b	53	27.05±6.68	24.5±8.66	16.75±4.14
L3d	37	41.94±6.89	33.27±5.5	29.45±7.67
L3e	97	37.45±5.29	33.54±6.05	51.18±15.34
L3f	54	55.19±8.76	44.47±9.03	59.79±17.58
L3h	23	92.48±13.43	77.93±13.97	78.97±24.03
L3i	9	41.1±7.53	43.59±9.86	47.09±16.63
L3x	13	33.2±5.75	29.14±6.66	74.51±19.76
L3a2	9	33.68±6.97	31.57±8.3	44.84±19.55
L3b1	50	21.89±4.85	17.59±6.17	15.34±4.04
L3b1a	43	16.49±2.1	10.85±2.2	17.83±4.69
L3d1-5	36	37.11±4.86	33.26±5.63	30.27±7.89
L3d1	19	36.51±5.89	32.04±6.19	18.06±8.02
L3e3'4'5	23	48.25±9.29	40.88±9.89	21.06±11.84
L3e1	30	15.59±2.6	14.2±2.91	33.63±9.56
L3e2	44	21.37±3.8	17.68±4.05	44.95±21.59
L3e1a	11	17.75±4.62	19.68±5.7	18.35±11.01
L3e2a	16	15.74±3.35	15.22±3.87	7.57±3.09

Haplogroups[a]	n	[b]Mishmar *et al.*, 2003	[c]Kivisilid *et al.*, 2006	[d]Soares *et al.*, 2009
L3e2b	27	15.98±2.55	11.52±2.45	8.22±3.58
L3e2b4	10	9.76±2.57	9.47±3.02	-
L3e3'4	18	50.81±10.55	46.6±12.44	21.3±14.83
L3f1	30	43.16±10.73	33.59±11.02	60.54±26.86
L3f1b	27	15.22±2.72	13.03±3.23	26.16±8.68
L3f2	7	49.91±7.56	34.79±7.1	34.59±9.99
L3f3	16	8.03±2.47	8.03±3.08	23.96±8.46
L3h1	22	67.03±8.01	57.8±8.66	59.62±14.88
L3h1a	13	68.77±10.43	56.71±10.85	41.91±12.12
L4b	13	82.21±13.18	69.72±14.49	88.48±29.41
L4b2	12	59.94±9.73	47.91±10.26	43.72±13.45
L4b2a	11	53.25±9.11	46.12±11.02	47.7±14.68
L5a'b	45	129.93±16.46	123.26±18.81	83.86±24.62
L5a	23	34.18±5.53	26.76±5.84	48.26±18.26
L5b	22	37.13±9.69	31.36±10.76	17.43±4.59
L5a2	16	30.83±6.16	22.41±6.23	39.1±15.9
L5a2a	11	23.35±5.87	8.61±2.75	20.18±8.41
L5b1	21	21.53±4.83	18.04±5.92	17.3±4.71
L5b1a	14	12.48±4.25	9.18±5.04	15.86±5.58
M1	108	36.06±7.44	25.18±6.97	52.69±11.45
M1a	93	27.18±4.54	18.69±4.25	53.6±12.98
M1a1	50	17.26±2.5	10.15±1.92	33.5±9.51
M1a1a	5	6.17±3.25	2.71±2.71	24.22±13.98
M1a1b	9	15.41±3.83	7.52±2.6	22.42±11.43
M1a1c'd	11	13.55±4.88	7.38±3.01	20.18±9.17
M1a1d	8	5.78±1.93	5.07±2.07	27.75±12.61
M1a1e-h	17	12.69±2.22	7.56±2.07	17.81±7.41
M1a2	14	19.08±4.76	12.08±4.18	63.42±18.12
M1a3	8	17.34±4.59	12.68±5.28	63.06±21.25
M1a4	10	15.93±5.36	16.23±6.83	50.45±26.54
M1a5	8	8.35±3.34	5.92±3.05	12.61±8.37
M1b	15	22.26±4.98	16.68±4.4	47.09±18
M1b1	5	15.41±6.08	9.47±4.88	16.14±9.89
M1b2	10	17.98±3.67	13.53±3.83	52.47±24.55
N1a	11	24.29±5.57	12.3±4.6	38.53±17.69
HV1	19	24.07±4.61	16.73±4.57	19.12±7.8
I	50	24.15±4.16	20.43±4.85	24.22±10.54

Table A7b: TMRCA estimates for maternal lineages found in Africa. [a]Includes haplogroups/haplotype split. The age estimates and standard errors were calculated using the method described by [b]Mishmar *et al.,* [178] that uses all the mutation changes in the coding region (16023-577) and the method described by [c]Kivisild *et al.,* [220] that uses only synonymous substitutions in the coding region. [d]Adjusted based on new mutation rate estimate by Soares *et al.* [190] that is corrected for the confounding effects of saturations and purifying selection, with the average estimate for a synonymous mutation taking place in coding region per 7884 years [190] compared with Kivisild *et al.,* [220] synonymous transition rate estimate of 1 per 6764 years [220], thus yield 16.56% older dates than that for Kivisild *et al.,* [220]. All estimates are in thousand years.

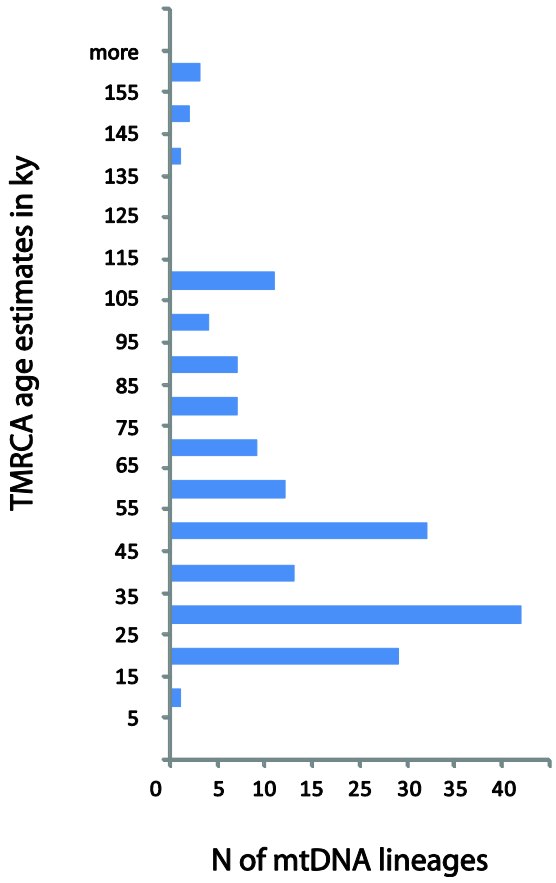


Figure A7c: Distribution of TMRCA age estimates of mtDNA lineages. The TMRCA age has been adjusted based on new mutation rate estimate by Soares [190] *et al.,* that is corrected for the confounding effects of saturations and purifying selection.

Appendix 8: Control region sequences and coding region SNPs genotypes for the individuals analyzed in this study. The data provided shows the haplotype assignations, control region sequence motifs found for each of the haplotype identified, and absence (0) or presence (1) of coding region SNPs checked for each haplogoups (blank indicates that the individual was not tested).

					L0afbk	L0abf	L0f	L0f2a/L0f3a	L0f2a	L0f2b	L0f2b1	L0f2b2	L0f2a3	L0f1b	L02b	L0f3a	L0a	L0a1		L0a1c1	L0a2	L0a2d	L0a1b	L0a3	L0b	L0d3
L0	Hg	Control region motifs	Abs. Freq	Samples	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	L0a1	16114A, 16129A, 16148T, 16168T, 16172C, 16184T, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 93G, 152C, 185A, 189G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	1	TZWF033	0	0											1									
2	L0a1	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16234T, 16278T, 16311C, 16319A, 16320T, 16519C, 93G, 185A, 189G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	1	KEBN031	1	1											1									
3	L0a1	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16234T, 16283T, 16311C, 16319A, 16320T, 93G, 185A, 189G, 236C, 247A, 263G, 315.1C, 522-3D	1	KEPR030	1	1											1									
4	L0a1	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16234T, 16311C, 16319A, 16320T, 93G, 185A, 189G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	1	TZBG023	1	1											1									
5	L0a1	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16234T, 16311C, 16319A, 16320T, 93G, 185A, 189G, 236C, 247A, 263G, 315.1C, 522-3D	3	KEPR003, KEPR004, KEPR013	1	1																				
6	L0a1	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16256T, 16311C, 64T, 93G, 152C, 189G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	1	CAKN036	0												0									
7	L0a1	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16278T, 16311C, 16320T, 93G, 95C, 185A, 189G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	1	CAMV028		0											0									
8	L0a1	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 93G, 185A, 189G, 236C, 247A, 263G, 315.1C, 522-3D	1	KELY010	1	1											1	1								
9	L0a1	16051G, 16129A, 16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 93G, 95C, 185A, 189G, 236C, 247A, 263G, 309.1C, 315.1C	1	CABM036,	0	0											0									
10	L0a1	16017C, 16129A, 16148T, 16172C, 16174T, 16187T, 16188G, 16189C, 16223T, 16230G, 16304C, 16311C, 16320T, 93G, 152C, 185A, 189G, 236C, 247A, 263G, 309.1C, 315.1C	1	CHBU017													1	1	1							
11	L0a1	16051G, 16129A, 16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 93G, 95C, 185A, 189G, 236C, 247A, 263G, 309.1C, 315.1C	1	CABM037	1	1											1	1	1							
12	L0a1a1	16129A, 16148T, 16168T, 16172C, 16187T, 16188A, 16189C, 16223T, 16230G, 16311C, 16320T, 64T, 93G, 146C, 152C, 185A, 189G, 200G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	1	SDNR004													1									
13	L0a1a1	16129A, 16148T, 16168T, 16172C, 16187T, 16188A, 16189C, 16223T, 16230G, 16311C, 16320T, 64T, 93G, 152C, 185A, 189G, 200G, 236C, 247A, 263G, 315.1C, 522-3D	1	SDBA022	1	1																				
14	L0a1a1	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16234T, 16311C, 16320T, 64T, 93G, 152C, 185A, 189G, 200G, 236C, 247A, 263G, 315.1C, 522-3D	1	KETV004	1	1											1	1	1							
15	L0a1a1	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16318G, 16320T, 64T, 93G, 152C, 185A, 189G, 200G, 236C, 247A, 263G, 315.1C, 522-3D	3	TZDT036, TZDT064, TZIQ068	1	1											1	1								
16	L0a1a1	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 64T, 93G, 146C, 185A, 189G, 200G, 236C, 247A, 263G, 315.1C, 522-3D	1	KEPR029	1	1											1	1								

17	L0a1a1	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 64T, 93G, 151T, 152C, 185A, 189G, 200G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	1	KESB011																								
18	L0a1a1	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 64T, 93G, 152C, 185A, 189G, 200G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	8	KEMR015, KEMR023, KETG044, KETG052, KETG065, KETG074, KETV018, KEPK041a																								
19	L0a1a1	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 64T, 93G, 152C, 185A, 189G, 200G, 236C, 247A, 263G, 315.1C, 522-3D	2	KEBR100, KESB029																								
20	L0a1a1	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 64T, 93G, 185A, 189G, 200G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	1	CABM003,	0	0											1											
21	L0a1a1	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 64T, 93G, 185A, 189G, 200G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	1	CAFU002	1	1																						
22	L0a1a1	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 64T, 93G, 185A, 189G, 200G, 236C, 247A, 263G, 315.1C, 522-3D	1	CAKN028	1	1																						
23	L0a1a1	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 64T, 93G, 95C, 152C, 185A, 189G, 200G, 236C, 247A, 263G, 315.1C, 522-3D	1	SDNY018	1	1											1	1	1									
24	L0a1a1	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 93G, 152C, 185A, 189G, 200G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	2	SDHD003, SDHD041	1	1											1	1										
25	L0a1a1	16129A, 16148T, 16168T, 16172C, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 93G, 152C, 185A, 189G, 200G, 234G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	1	KEGU053	1	1											1	1										
26	L0a1a1	16129A, 16148T, 16168T, 16172C, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 93G, 152C, 185A, 189G, 200G, 234G, 236C, 247A, 263G, 315.1C, 522-3D	1	KEBJ038																								
27	L0a1b	16093C, 16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16209C, 16223T, 16230G, 16278T, 16293G, 16311C, 16320T, 93G, 95C, 152C, 185A, 189G, 200G, 207A, 236C, 247A, 263G, 315.1C, 522-3D	1	CHLA072	1	1																					1	
28	L0a1b	16093C, 16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16209C, 16223T, 16230G, 16278T, 16293G, 16311C, 16320T, 93G, 95C, 152C, 185A, 189G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	1	CAFU024	1	1																					1	
29	L0a1b	16093C, 16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16209C, 16223T, 16230G, 16278T, 16293G, 16311C, 16320T, 93G, 95C, 152C, 185A, 189G, 236C, 247A, 263G, 315.1C, 522-3D, 634C	1	CHBU019													1	1									1	
30	L0a1b	16129A, 16148T, 16168T, 16172C, 16184T, 16187T, 16188G, 16189C, 16223T, 16230G, 16278T, 16293G, 16311C, 16320T, 93G, 95C, 146C, 185A, 189G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	1	SDNY001																								
31	L0a1b	16129A, 16148T, 16168T, 16172C, 16187T, 16188A, 16189C, 16223T, 16230G, 16278T, 16293G, 16311C, 16320T, 93G, 95C, 185A, 189G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	1	KETS003	1	1											1	1					0	1				
32	L0a1b	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16278T, 16293G, 16311C, 16320T, 64T, 93G, 95C, 185A, 189G, 236C, 247A, 263G, 315.1C, 522-3D	1	TZDR007																								
33	L0a1b	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16278T, 16293G, 16311C, 16320T, 89C, 93G, 95C, 185A, 189G, 236C, 247A, 263G, 309.1C, 309.2C, 315.1C, 522-3D	1	TZDT001	1	1											1	1			0		1					

34	L0a1b	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16278T, 16293G, 16311C, 16320T, 93G, 95C, 152C, 185A, 189G, 236C, 247A, 263G, 315.1C, 522-3D	1	KESB028	1	1											1	1					1			
35	L0a1b	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16278T, 16293G, 16311C, 16320T, 93G, 95C, 185A, 189G, 236C, 247A, 263G, 309.1C, 309.2C, 315.1C	1	KELY001	1	1											1	1				0	1			
36	L0a1b	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16278T, 16293G, 16311C, 16320T, 93G, 95C, 185A, 189G, 236C, 247A, 263G, 309.1C, 309.2C, 315.1C, 522-3D	1	XH023																		0				
37	L0a1b	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16278T, 16293G, 16311C, 16320T, 93G, 95C, 185A, 189G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	2	KETA019, XH020	1	1											1	1			0		1			
38	L0a1b	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16278T, 16293G, 16311C, 16320T, 93G, 95C, 185A, 189G, 236C, 247A, 263G, 315.1C, 522-3D	1	TZRG033	1	1											1	1			0		1			
39	L0a1b	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16278T, 16293G, 16311C, 16320T, 93G, 95C, 185A, 189G, 236C, 247A, 263G, 315.1C, 522-3D, 593C	1	CAKN005	1	1											1	1				0	1			
40	L0a1b	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16278T, 16293G, 16311C, 89C, 93G, 95C, 185A, 189G, 236C, 247A, 263G, 315.1C, 522-3D	1	KETV061																			1			
41	L0a1b	16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16278T, 16293G, 16311C, 16320T, 89C, 93G, 95C, 185A, 189G, 236C, 247A, 263G, 315.1C, 522-3D	2	TZMS001, TZMS059	1	1											1	1			0		1			
42	L0a1c1a	15941C, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16287T, 16293G, 16311C, 16320T, 16519C, 93G, 185A, 189G, 236C, 247A, 263G, 309.1C, 309.2C, 315.1C, 522-3D	1	KESB027 (KESB027A)																						
43	L0a1c1a	15941C, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16287T, 16293G, 16311C, 16320T, 16519C, 93G, 185A, 189G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	7	KENJ014, KEPK007, KESB013, KESN059, KETG049, KELY008, KETG082													1	1		1						
44	L0a1c1a	15941C, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16287T, 16293G, 16311C, 16320T, 16519C, 93G, 185A, 189G, 236C, 247A, 263G, 315.1C, 522-3D	3	KEBR104, TZDT067, TZDT070																						
45	L0a1c1b	16093C, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16266T, 16287T, 16293G, 16311C, 16320T, 16519C, 93G, 152C, 185A, 189G, 236C, 247A, 315.1C, 498D, 522-3D	5	KEOR021a, KEOR021b, KETK042, Kemn007, TZMS060	1	1											1	1		1						
46	L0a1c1b	16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16266T, 16287T, 16293G, 16311C, 16320T, 16519C, 93G, 152C, 185A, 189G, 236C, 247A, 315.1C, 498D, 522-3D	2	KEBN019, TZMS036	1	1											1	1								
47	L0a1d	16114A, 16129A, 16148T, 16162G, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16293G, 16311C, 16320T, 93G, 152C, 185A, 189G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	1	KESB019													1	1								
48	L0a1d	15924G, 16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 93G, 152C, 185A, 189G, 236C, 247A, 263G, 315.1C, 522-3D, 553T	1	SDDN008	1	1											1	1								
49	L0a1d	16071T, 16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 93G, 152C, 185A, 189G, 236C, 247A, 263G, 309.1C, 315.1C, 327T, 522-3D, 553T	1	KEPK005	1	1											1	1								
50	L0a1d	16093C, 16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16293G, 16311C, 16519C, 89C, 93G, 95C, 152C, 185A, 189G, 195C, 236C, 247A, 263G, 315.1C, 522-3D, 553T	1	KEPK082	1	1																				
51	L0a1d	16129A, 16148T, 16168T, 16172C, 16179:, 16187T, 16188G, 16189C, 16223T, 16230G, 16293G, 16311C, 93G, 185A, 189G, 236C, 247A, 263G, 315.1C, 522-3D, 553T	6	KEEL011, KEEL005, KEEL029, KEEL031, KEEL034, KEEL043																						

52	L0a1d	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16261T, 16293G, 16311C, 93G, 152C, 185A, 189G, 195C, 236C, 247A, 263G, 315.1C, 522-3D, 553T	2	KETS001, TZWF025	1	1											1	1								
53	L0a1d	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16261T, 16293G, 16311C, 93G, 152C, 185A, 189G, 225A, 236C, 247A, 263G, 315.1C, 522-3D, 553T	1	KETV058	1	1											1	1								
54	L0a1d	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16261T, 16293G, 16311C, 93G, 152C, 185A, 189G, 236C, 247A, 263G, 315.1C, 522-3D, 553T	3	TZMS054, TZSW010, tzsw121	1	1											1	1								
55	L0a1d	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16287T, 16293G, 16294T, 16311C, 16399G, 93G, 151T, 152C, 153G, 185A, 189G, 236C, 247A, 263G, 315.1C, 522-3D, 553T	1	KEPK080	1	1											1	1								
56	L0a1d	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16287T, 16293G, 16294T, 16311C, 16399G, 93G, 151T, 152C, 185A, 189G, 236C, 247A, 263G, 315.1C, 522-3D, 553T	5	KENJ028, KENJ046, KEPK009, KESB018, KETG008																						
57	L0a1d	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16293G, 16299G, 16311C, 16526:A, 93G, 152C, 185A, 189G, 236C, 247A, 263G, 315.1C, 522-3D, 553T	1	KEBR013																						
58	L0a1d	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16293G, 16311C, 16519C, 89C, 93G, 95C, 152C, 185A, 189G, 195C, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D, 553T	2	KELY027, KELY033	1	1											1	1								
59	L0a1d	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16293G, 16311C, 16519C, 89C, 93G, 95C, 152C, 185A, 189G, 195C, 236C, 247A, 263G, 315.1C, 522-3D, 553T	4	KELY007, KENJ006, KENJ041, KESN033	1	1											1	1								
60	L0a1d	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16293G, 16311C, 89C, 93G, 185A, 189G, 236C, 247A, 263G, 315.1C, 522-3D, 553T	1	KEGU039	1	1											1	1								
61	L0a1d	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16293G, 16311C, 93G, 151T, 152C, 185A, 189G, 236C, 247A, 263G, 315.1C, 522-3D, 553T	2	KERD048, Kesm021																						
62	L0a1d	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16293G, 16311C, 93G, 151T, 152C, 185A, 189G, 236C, 247A, 263G, 315.1C, 522-3D, 553T, 573.1C	1	TZMS038	1	1											1	1								
63	L0a1d	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16293G, 16311C, 93G, 152C, 185A, 189G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D, 553T	2	KEOR003, Kemn011	1	1																				
64	L0a1d	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16293G, 16311C, 93G, 152C, 185A, 189G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D, 553T, 654G	1	KETK002	1	1											1	1								
65	L0a1d	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16293G, 16311C, 93G, 152C, 185A, 189G, 236C, 247A, 263G, 315.1C, 522-3D, 553T	4	KEBR025, KEBR062, KENJ079, KERD082	1	1											1	1								
66	L0a1d	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16293G, 16311C, 93G, 185A, 189G, 236C, 247A, 263G, 315.1C, 522-3D, 553T	2	KENJ055, kesm012	1	1											1	1								
67	L0a1d	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16519C, 93G, 152C, 185A, 189G, 236C, 247A, 263G, 315.1C, 522-3D, 553T	1	Kemn004													1	1								
68	L0a1d	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 93G, 152C, 185A, 189G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D, 553T	1	KERD022	1	1											1	1								
69	L0a1d	16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 93G, 152C, 185A, 189G, 236C, 247A, 263G, 315.1C, 522-3D, 553T	3	KEGB033, KERD034, KERD088	1	1											1	1								

70	L0a1d	16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16287T, 16293G, 16294T, 16311C, 16399G, 93G, 151T, 152C, 153G, 185A, 189G, 236C, 247A, 263G, 315.1C, 522-3D, 553T	1	KEPK104	1	1																				
71	L0a1e	16129A, 16148T, 16172C, 16179T, 16187T, 16188G, 16189C, 16192T, 16223T, 16230G, 16293G, 16311C, 16317G, 16320T, 93G, 151T, 152C, 185A, 189G, 236C, 247D, 263G, 315.1C, 522-3D, 675	1	SDDN006	1	1												1	1							
72	L0a2	15916C, 15924G, 15942C, 16148T, 16172C, 16173T, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16399G, 16519C, 93G, 152C, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	1	TZWF012														1		1		1				
73	L0a2	16071T, 16111T, 16148T, 16172C, 16187T, 16188G, 16189C, 16192T, 16223T, 16230G, 16234T, 16290T, 16311C, 16320T, 16519C, 73G, 152C, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	1	TZAK018	1	1												1		1		1				
74	L0a2	16071T, 16111T, 16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16234T, 16290T, 16311C, 16320T, 16519C, 73G, 152C, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	1	TZAK023	1	1												1		1		1				
75	L0a2	16093C, 16148T, 16172C, 16187T, 16188A, 16189C, 16223T, 16230G, 16255A, 16261T, 16283G, 16311C, 16320T, 16390A, 16519C, 64T, 93G, 236C, 247A, 263G, 315.1C	1	KEKS015	1	1												1		1		1				
76	L0a2	16129A, 16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16519C, 93G, 152C, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	1	KETA032	1	1												1				1				
77	L0a2	16148T, 16172C, 16187T, 16188A, 16189C, 16223T, 16230G, 16233G, 16311C, 16320T, 16390A, 16519C, 64T, 93G, 236C, 247A, 263G, 309.1C, 315.1C	1	KETK078	1	1												1		1		1				
78	L0a2	16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16304C, 16311C, 16320T, 16465T, 16519C, 64T, 93G, 152C, 236C, 247A, 263G, 315.1C, 522-3D	1	tzsw036	1	1														1		1				
79	L0a2	16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16465T, 16519C, 64T, 93G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	2	KEBN014, KEBN038	1	1																1				
80	L0a2a2	15900C, 16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16519C, 64T, 93G, 152C, 189G, 204C, 207A, 236C, 247A, 263G, 315.1C, 522-3D	1	KETA003	1	1												0		1		1				
81	L0a2a2	15938T, 16086C, 16129A, 16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 93G, 152C, 185A, 189G, 204C, 207A, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	1	KEBJ011																						
82	L0a2a2	16129A, 16148T, 16154C, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16260T, 16311C, 16320T, 93G, 152C, 189G, 204C, 207A, 236C, 247A, 263G, 315.1C, 522-3D	2	TZPR011, TZPR038	1	1												1								
83	L0a2a2	16129A, 16148T, 16172C, 16187T, 16188G, 16189C, 16192T, 16223T, 16230G, 16311C, 16320T, 16362C, 16519C, 73G, 93G, 95C, 152C, 185A, 189G, 204C, 207A, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	3	KEBN008, KEBN039, KEBN002	1	1																				
84	L0a2a2	16129A, 16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16260T, 16311C, 16320T, 93G, 152C, 185A, 189G, 204C, 207A, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	1	KETA069														1								
85	L0a2a2	16129A, 16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16260T, 16311C, 16320T, 93G, 152C, 185A, 189G, 204C, 207A, 236C, 247A, 263G, 315.1C, 522-3D	1	KEPR043														1								
86	L0a2a2	16129A, 16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16260T, 16311C, 16320T, 93G, 152C, 189G, 204C, 207A, 236C, 247A, 263G, 315.1C, 522-3D	3	KETA008, KETA010, KETA012	1	1												1								
87	L0a2a2	16129A, 16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 93G, 152C, 185A, 189G, 204C, 207A, 236C, 247A, 263G, 315.1C, 522-3D	2	KEOR011, TZBG041	1	1												1								

88	L0a2a2	16129A, 16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 93G, 152C, 189G, 204C, 207A, 236C, 247A, 263G, 315.1C, 522-3D	1	tzsw133													1									
89	L0a2a2	16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16399G, 16519C, 64T, 93G, 152C, 189G, 204C, 207A, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	1	CABM027	0	0											0									
90	L0a2a2	16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16519C, 64T, 93G, 150T, 152C, 189G, 204C, 207A, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	1	KEND004	1	1											1	1		1						
91	L0a2a2	16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16519C, 64T, 93G, 151T, 152C, 189G, 204C, 207A, 236C, 247A, 263G, 315.1C, 522-3D	1	TZTR002	1	1											1	1		1						
92	L0a2a2	16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16519C, 64T, 93G, 152C, 189G, 204C, 207A, 236C, 247A, 263G, 315.1C, 522-3D	7	KETK091, TZBG035, TZBG037, TZPR029, TZRG008, TZTR038, XH022	1	1											1	1		1						
93	L0a2a2	16148T, 16187T, 16188G, 16189C, 16223T, 16230G, 16265G, 16311C, 16320T, 16519C, 64T, 93G, 152C, 189G, 204C, 207A, 236C, 247A, 263G, 315.1C, 522-3D	1	KETV003	1	1											1	1		1						
94	L0a2b	15927A, 16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16519C, 93G, 152C, 189G, 236C, 247A, 263G, 315.1C, 522-3D	1	CHLA073	1	1											1			1						
95	L0a2b	16111T, 16148T, 16172C, 16187T, 16188A, 16189C, 16223T, 16230G, 16311C, 16519C, 64T, 93G, 152C, 189G, 236C, 247A, 263G, 309.1C, 309.2C, 315.1C, 522-3D	1	KEBN011	1	1											1	1		1						
96	L0a2b	16129A, 16148T, 16172C, 16187T, 16188G, 16189C, 16217C, 16223T, 16230G, 16311C, 16320T, 93G, 152C, 189G, 236C, 247A, 263G, 315.1C, 522-3D	1	KEBN034	1	1											1			0						
97	L0a2b	16129A, 16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16519C, 93G, 152C, 189G, 207A, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	1	KENJ054	1	1											1			1						
98	L0a2b	16129A, 16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16519C, 93G, 152C, 189G, 236C, 247A, 263G, 309.1C, 309.2C, 315.1C, 522-3D	2	TZPR028, TZPR007	1	1											1			0						
99	L0a2b	16148T, 16166:, 16172C, 16180G, 16187T, 16188G, 16189C, 16223T, 16230G, 16289G, 16311C, 16320T, 73G, 93G, 143A, 152C, 189G, 236C, 247A, 263G, 315.1C, 522-3D	1	KEBN026	1	1											1			1						
100	L0a2b	16148T, 16169T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16234T, 16311C, 16320T, 16519C, 64T, 93G, 152C, 189G, 207A, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	1	SDDN029													1	1		1						
101	L0a2b	16148T, 16172C, 16186T, 16187T, 16188G, 16189C, 16223T, 16230G, 16256T, 16311C, 16320T, 16519C, 73G, 150T, 152C, 185A, 189G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	1	tzhz250	1	1											1	1		1						
102	L0a2b	16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16519C, 64T, 93G, 152C, 189G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	1	SDDN012	1	1											1	1		1						
103	L0a2b	16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16519C, 73G, 93G, 95C, 151T, 152C, 189G, 236C, 247A, 263G, 315.1C, 522-3D, 538C	1	TZAK033	1	1											1	1		1						
104	L0a2c	15900C, 16172C, 16187T, 16188G, 16189C, 16213A, 16223T, 16230G, 16311C, 16320T, 16437C, 16519C, 64T, 93G, 95C, 152C, 236C, 247A, 263G, 315.1C, 522-3D	2	TZBG050, TZBG056	1	1											1	1		1						
105	L0a2c	15916C, 15924G, 15942C, 16148T, 16172C, 16173T, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16399G, 16519C, 16527T, 93G, 95C, 152C, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	1	KEPK008	1	1											1	1		1						

106	L0a2c	15916C, 15924G, 15942C, 16148T, 16172C, 16173T, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16399G, 16519C, 93G, 95C, 152C, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	2	KEPK035, KETG048	1	1											1		1		1					
107	L0a2c	15916C, 15924G, 15942C, 16148T, 16172C, 16173T, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16399G, 16519C, 93G, 95C, 152C, 236C, 247A, 263G, 315.1C, 522-3D	1	KEBR003													1		1		1					
108	L0a2c	16148T, 16172C, 16187T, 16188A, 16189C, 16223T, 16230G, 16311C, 16320T, 16497G, 16519C, 16527T, 93G, 95C, 152C, 195C, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	1	Keil017	1	1															1					
109	L0a2c	16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16234T, 16311C, 16320T, 16519C, 93G, 152C, 195C, 236C, 247A, 263G, 309.1C, 315.1C	1	KETA075	1	1											1		1		1					
110	L0a2c	16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16519C, 64T, 93G, 152C, 195C, 236C, 247A, 263G, 315.1C, 522-3D	1	TZBG054	1	1											1		1							
111	L0a2c	16148T, 16172C, 16187T, 16188A, 16189C, 16214T, 16223T, 16230G, 16234T, 16311C, 16519C, 64T, 93G, 95C, 152C, 189G, 236C, 247A, 263G, 309.1C, 309.2C, 315.1C, 522-3D	2	KEBJ037x, KEEL003																						
112	L0a2c	16129A, 16148T, 16172C, 16184T, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16519C, 64T, 89C, 93G, 95C, 152C, 189G, 236C, 247A, 263G, 315.1C, 522-3D	1	KEKP008													1		1		1	1				
113	L0a2c	16148T, 16172C, 16183G, 16184T, 16185T, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16519C, 64T, 93G, 95C, 152C, 189G, 236C, 247A, 263G, 315.1C, 522-3D	2	KETK033, KETK071	1	1											1		1		1	1				
114	L0a2c	16148T, 16172C, 16184T, 16185T, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16519C, 64T, 93G, 95C, 152C, 189G, 236C, 247A, 263G, 315.1C, 522-3D	1	KETK003	1	1											1		1		1	1				
115	L0a2c	16148T, 16172C, 16184T, 16187T, 16188G, 16189C, 16223T, 16230G, 16266T, 16311C, 16320T, 16519C, 64T, 93G, 95C, 152C, 189G, 236C, 247A, 263G, 315.1C, 522-3D	1	KELO086	1	1											1		1		1	1				
116	L0a2c	16148T, 16172C, 16184T, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16519C, 64T, 93G, 95C, 152C, 189G, 236C, 247A, 263G, 315.1C, 522-3D	1	KEPK096	1	1											1				1	1				
117	L0a2c	16148T, 16172C, 16184T, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16519C, 64T, 93G, 95C, 152C, 189G, 236C, 247A, 263G, 315.1C, 522-3D, 571T	1	KERD041	1	1											1		1			1				
118	L0a2c	16148T, 16172C, 16184T, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16519C, 64T, 93G, 95C, 152C, 189G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	1	Keil011	1	1											1				1	1				
119	L0a2c	16172C, 16184T, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16519C, 64T, 93G, 95C, 152C, 189G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	2	kemn001, kesm023	1	1											1		1		1	1				
120	L0a2c	16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16292T, 16311C, 16320T, 16519C, 89C, 93G, 95C, 152C, 189G, 200G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D, 533G	1	TZDT079	1	1											1		1		1	1				
121	L0a2c1	16086C, 16148T, 16172C, 16187T, 16188G, 16189C, 16212G, 16223T, 16230G, 16311C, 16320T, 16519C, 73G, 93G, 146C, 152C, 189G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D, 538C	1	TZRG016	1	1											1		1		1	0				
122	L0a2c1	16148T, 16166G, 16172C, 16188A, 16189C, 16212G, 16223T, 16230G, 16311C, 16320T, 16519C, 73G, 93G, 95C, 146C, 151T, 152C, 189G, 236C, 247A, 263G, 315.1C, 522-3D, 538C	1	tzsw014	1	1											1		1		1	0				
123	L0a2c1	16148T, 16172C, 16187T, 16188G, 16189C, 16212G, 16223T, 16230G, 16283T, 16311C, 16320T, 16519C, 73G, 93G, 146C, 152C, 189G, 236C, 247A, 263G, 315.1C, 522-3D, 538C	1	TZTR021													1				1	0				

124	L0a2c1	16148T, 16172C, 16187T, 16188G, 16189C, 16212G, 16223T, 16230G, 16311C, 16320T, 16519C, 73G, 93G, 146C, 152C, 189G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D, 538C	1	TZTR001	1	1											1		1		1	0				
125	L0a2c1	15924G, 16093C, 16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16519C, 64T, 93G, 146C, 152C, 189G, 195C, 236C, 247A, 263G, 315.1C, 522-3D	1	KEPK073	1	1											1		1		1	1				
126	L0a2c1	15924G, 16093C, 16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16519C, 64T, 93G, 146C, 152C, 189G, 236C, 247A, 263G, 315.1C, 522-3D	1	KEPK014													1		1		1	1				
127	L0a2c1	16129A, 16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16519C, 64T, 93G, 95C, 146C, 152C, 189G, 236C, 247A, 263G, 315.1C, 523:, 525G, 526:, 573.1C, 573.2C, 573.3C	1	TZDR001	1	1											1				1	1				
128	L0a2c1	16148T, 16153A, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16519C, 64T, 93G, 146C, 152C, 189G, 236C, 247A, 263G, 315.1C, 522-3D	1	KEGK050													1		1		1	1				
129	L0a2c1	16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16519C, 64T, 93G, 146C, 152C, 189G, 200G, 236C, 247A, 263G, 315.1C, 522-3D	1	KEGK021	1	1																1				
130	L0a2c1	15924G, 16148T, 16172C, 16187T, 16188A, 16189C, 16223T, 16230G, 16311C, 16320T, 16519C, 64T, 93G, 146C, 152C, 189G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	1	KESN007	1	1											1		1		1	1				
131	L0a2c1	15924G, 16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16519C, 64T, 93G, 146C, 152C, 189G, 236C, 247A, 263G, 315.1C, 522-3D	5	KEPK028, KEPK032, TZDR003, TZDR009A (ZDR009B), TZDR012A (TZDR012)	1	1											1		1		1	1				
132	L0a2c1	16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16519C, 64T, 93G, 146C, 150T, 152C, 189G, 193G, 200G, 236C, 247A, 263G, 315.1C, 522-3D	1	TZIQ075																						
133	L0a2c1	16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16519C, 64T, 93G, 146C, 150T, 152C, 189G, 200G, 236C, 247A, 263G, 315.1C, 522-3D	2	TZWF026, TZWF038													0		1			0				
134	L0a2c1	16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16519C, 64T, 93G, 146C, 150T, 152C, 189G, 200G, 236C, 247D, 263G, 309.1C, 315.1C, 522-3D	1	KETA020	1	1											1		1		1	1				
135	L0a2c1	16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16519C, 93G, 146C, 150T, 152C, 189G, 200G, 236C, 247A, 263G, 315.1C, 522-3D	1	TZIQ005	1	1											1		1		1	1				
136	L0a2c2	16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16519C, 64T, 93G, 95C, 152C, 185A, 189G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	1	KETK009	1	1											1		1		1	1				
137	L0a2c2	16148T, 16187T, 16188G, 16189C, 16223T, 16230G, 16293G, 16311C, 16320T, 16519C, 64T, 93G, 95C, 152C, 189G, 236C, 247A, 263G, 315.1C, 522-3D	1	KEBR001	1	1											1		1		1	1				
138	L0a3	15941C, 16129A, 16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16287T, 16311C, 16320T, 93G, 95C, 152C, 185A, 189G, 207A, 236C, 247A, 263G, 309.1C, 315.1C, 318C, 522-3D	2	SDNR029, SDSH003	1	1											1						1			
139	L0a3a1	16129A, 16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16519C, 93G, 152C, 189G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	3	TZBG057, TZIQ083, TZWF019													1									
140	L0a3a1	15937G, 16129A, 16148T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T, 16519C, 93G, 152C, 189G, 236C, 247A, 263G, 309.1C, 309.2C, 315.1C, 522-3D	1	KELO084	1	1																				
141	L0a4a	16129A, 16148T, 16166:, 16172C, 16180G, 16187T, 16188G, 16189C, 16223T, 16230G, 16289G, 16311C, 16320T, 16362C, 73G, 93G, 143A, 146C, 152C, 185A, 189G, 195C, 236C, 247A, 263G, 315.1C, 522-3D	3	TZDT012, TZDT037, TZMS041	1	1											1									

159	L0d1c	15930A, 15941C, 15951G, 16167T, 16187T, 16189C, 16223T, 16230G, 16234T, 16242T, 16243C, 16311C, 16519C, 73G, 146C, 152C, 195C, 198T, 247A, 315.1C, 456T, 498D, 522-3D, 593C	1	Hgdp01036																							
160	L0d1c	15930A, 15941C, 15951G, 16189C, 16223T, 16230G, 16234T, 16243C, 16249C, 16311C, 16519C, 73G, 146C, 152C, 195C, 247A, 315.1C, 456T, 498D	1	XH017																							
161	L0d2a	15930A, 15941C, 16093C, 16129A, 16187T, 16189C, 16212G, 16223T, 16230G, 16243C, 16311C, 16390A, 16519C, 73G, 146C, 152C, 195C, 198T, 247A, 315.1C, 498D, 522-3D, 597T	1	XH021																							
162	L0d2a	15930A, 15941C, 16129A, 16187T, 16189C, 16212G, 16223T, 16230G, 16243C, 16311C, 16390A, 16519C, 73G, 146C, 152C, 195C, 198T, 247A, 315.1C, 498D, 597T	1	XH018																							
163	L0d3	15884A, 15930A, 15941C, 16187T, 16189C, 16223T, 16230G, 16243C, 16274A, 16278T, 16290T, 16311C, 16519C, 73G, 146C, 150T, 152C, 195C, 247A, 315.1C, 316A	1	KEBN018	0	0	0																				**1**
164	L0d3	15930A, 15941C, 16129A, 16148T, 16187T, 16189C, 16223T, 16230G, 16243C, 16278T, 16290T, 16300G, 16311C, 16399G, 16519C, 73G, 146C, 150T, 195C, 198T, 247A, 315.1C, 316A	1	TZBG031																				0			
165	L0d3	15930A, 15941C, 16129A, 16187T, 16189C, 16223T, 16230G, 16278T, 16290T, 16300G, 16311C, 16399G, 16519C, 73G, 146C, 150T, 152C, 188G, 195C, 198T, 247A, 309.1C, 309.2C, 315.1C, 316A, 456T	2	TZAK004, TZAK007																		0				**1**
166	L0d3	15930A, 15941C, 16129A, 16187T, 16189C, 16223T, 16230G, 16278T, 16290T, 16300G, 16311C, 16399G, 16519C, 73G, 146C, 150T, 152C, 188G, 195C, 198T, 247A, 309.1C, 315.1C, 316A	1	TZAK003	0	0	0	0	0							0	0	0								**1**
167	L0d3	15930A, 15941C, 16129A, 16187T, 16189C, 16223T, 16230G, 16278T, 16290T, 16300G, 16311C, 16399G, 16519C, 73G, 146C, 150T, 152C, 188G, 195C, 198T, 247A, 309.1C, 315.1C, 316A, 456T	2	TZAK017, TZAK006	0	0	0	0	0							0	0	0								**1**
168	L0f1b	16169T, 16172C, 16187T, 16189C, 16223T, 16230G, 16278T, 16311C, 16319A, 16327T, 16354T, 16368C, 93G, 95C, 151T, 152C, 185A, 189G, 247A, 263G, 315.1C, 523:, 525G, 526:, 573.1C, 573.2C, 573.3C	1	KEYK023	1	1	1							1												
169	L0f1b	16169T, 16172C, 16187T, 16189C, 16214A, 16223T, 16230G, 16278T, 16311C, 16327T, 16368C, 16465T, 16519C, 73G, 152C, 185A, 189G, 207A, 247A, 263G, 309.1C, 315.1C, 522-3D	1	TZTR043	1	1	1							1		0										
170	L0f1b	16169T, 16172C, 16187T, 16189C, 16223T, 16230G, 16278T, 16301T, 16311C, 16327T, 16368C, 16519C, 73G, 152C, 185A, 189G, 207A, 247A, 263G, 315.1C, 522-3D	2	tzsw031, tzsw034	1	1	1							1		0										
171	L0f1b	16169T, 16172C, 16187T, 16189C, 16223T, 16230G, 16265C, 16270T, 16278T, 16311C, 16327T, 16368C, 16465T, 16519C, 73G, 152C, 185A, 189G, 207A, 247A, 263G, 315.1C, 391C, 522-3D	7	TZBG026, TZBG030, TZBG034, TZBG036, TZBG040, TZBG048, TZRG025	1	1	1	0	0	1	0	1		1												
172	L0f1b	16169T, 16172C, 16187T, 16189C, 16223T, 16230G, 16265G, 16278T, 16311C, 16327T, 16354T, 16368C, 93G, 151T, 152C, 185A, 189G, 247A, 263G, 309.1C, 315.1C, 522-3D	2	KEBN010, KEBR047	1	1	1							1												
173	L0f1b	16169T, 16172C, 16187T, 16189C, 16223T, 16230G, 16265G, 16278T, 16311C, 16327T, 16354T, 16368C, 93G, 151T, 152C, 185A, 189G, 247A, 263G, 309.1C, 315.1C, 522-3D	1	KEBN025	1	1	1							1												
174	L0f1b1	16169T, 16172C, 16187T, 16189C, 16223T, 16230G, 16265G, 16278T, 16311C, 16327T, 16354T, 16368C, 93G, 151T, 152C, 185A, 189G, 247A, 263G, 315.1C, 522-3D	4	KEBN003, KEBN007, KESY004, KESY008										1												
175	L0f1b2	16169T, 16172C, 16187T, 16189C, 16223T, 16230G, 16265G, 16278T, 16311C, 16327T, 16354T, 16368C, 89C, 93G, 95C, 151T, 152C, 185A, 189G, 247A, 263G, 315.1C, 523:, 525G, 526:, 573.1C, 573.2C, 573.3C	1	KEBN023	1	1	1							1												
176	L0f1b2	16169T, 16172C, 16187T, 16189C, 16223T, 16230G, 16278T, 16311C, 16319A, 16327T, 16354T, 16368C, 93G, 95C, 151T, 152C, 185A, 189G, 247A, 263G, 315.1C, 522-3D, 573.1C, 573.2C	1	KETG063										1												

177	L0f2a	16129A, 16169T, 16172C, 16187T, 16189C, 16223T, 16230G, 16239T, 16278T, 16311C, 16327T, 16368C, 16519C, 143A, 146C, 152C, 185A, 189G, 247A, 263G, 315.1C	1	KERD001			1										0										
178	L0f2a	16129A, 16147A, 16169T, 16172C, 16187T, 16189C, 16223T, 16230G, 16278T, 16311C, 16327T, 16368C, 16390A, 16519C, 44.1C, 146C, 152C, 185A, 189G, 207A, 247A, 263G, 309.1C, 315.1C, 511T	1	KEBN012	1	1	1	1																			
179	L0f2a1	16129A, 16169T, 16172C, 16173T, 16187T, 16189C, 16223T, 16230G, 16239T, 16278T, 16311C, 16327T, 16359C, 16368C, 16519C, 143A, 146C, 152C, 185A, 189G, 247A, 263G, 315.1C	1	TZBG028	1	1	1	1																			
180	L0f2a1	16129A, 16169T, 16172C, 16173T, 16187T, 16189C, 16223T, 16230G, 16239T, 16278T, 16311C, 16327T, 16368C, 16519C, 143A, 146C, 152C, 185A, 189G, 195C, 247A, 263G, 315.1C	2	TZTR025, TZTR029	1	1	1	1																			
181	L0f2a2	16129A, 16147A, 16169T, 16172C, 16187T, 16189C, 16223T, 16230G, 16278T, 16311C, 16327T, 16368C, 16390A, 16519C, 44.1C, 146C, 152C, 185A, 189G, 204C, 207A, 247A, 263G, 309.1C, 315.1C, 511T	1	KEBN022																							
182	L0f2a2	16129A, 16169T, 16172C, 16187T, 16189C, 16213A, 16223T, 16230G, 16278T, 16311C, 16327T, 16368C, 16519C, 44.1C, 146C, 152C, 185A, 189G, 204C, 207A, 247A, 263G, 315.1C, 511T	1	Kemn014			1										0										
183	L0f2a2	16129A, 16169T, 16172C, 16187T, 16189C, 16223T, 16230G, 16278T, 16311C, 16327T, 16368C, 16519C, 44.1C, 146C, 152C, 185A, 189G, 204C, 207A, 247A, 263G, 315.1C, 511T, 524.1A, 525.1C	1	KETA076													1										
184	L0f2a2	16129A, 16169T, 16172C, 16187T, 16189C, 16274A, 16278T, 16311C, 16327T, 16352C, 16368C, 16519C, 44.1C, 73G, 146C, 152C, 189G, 204C, 207A, 247A, 263G, 315.1C, 511T	5	TZIQ042, TZIQ049, TZWF001, TZWF022, TZWF047																							
185	L0f2a3	16093C, 16129A, 16169T, 16172C, 16187T, 16189C, 16223T, 16230G, 16278T, 16311C, 16325C, 16327T, 16354T, 16368C, 73G, 146C, 152C, 185A, 189G, 199C, 204C, 207A, 247A, 263G, 309.1C, 315.1C	1	KEWT015	0												0										
186	L0f2a3	16129A, 16169T, 16172C, 16187T, 16189C, 16216G, 16223T, 16230G, 16278T, 16311C, 16325C, 16327T, 16354T, 16368C, 73G, 146C, 152C, 185A, 189G, 199C, 207A, 247A, 263G, 315.1C	4	KEBR031, KEEL004, KEEL028, KEBJ037a	1	1	1	1					1				0										
187	L0f2a3	16129A, 16169T, 16172C, 16187T, 16189C, 16223T, 16230G, 16278T, 16311C, 16325C, 16327T, 16354T, 16368C, 73G, 146C, 152C, 185A, 189G, 199C, 207A, 247A, 263G, 315.1C	2	TZAK002, TZAK028	1	1	1	1																			
188	L0f2b	15884A, 15927A, 16169T, 16172C, 16187T, 16189C, 16218T, 16223T, 16230G, 16278T, 16291T, 16311C, 16327T, 16354T, 16368C, 16519C, 146C, 183G, 184A, 185A, 189G, 207A, 247A, 263G, 315.1C	1	SDHD002	1	1	1	0	0	1	1	0		1													
189	L0f2b	15927A, 16129A, 16169T, 16172C, 16187T, 16189C, 16218T, 16223T, 16230G, 16278T, 16291T, 16311C, 16327T, 16354T, 16368C, 16519C, 146C, 152C, 183G, 184A, 185A, 189G, 195C, 247A, 263G, 309.1C, 315.1C	2	KETA034, KETA013	1	1	1	0	0	1	1	0		1													
190	L0f3a	16052T, 16129A, 16169T, 16176T, 16187T, 16189C, 16223T, 16230G, 16278T, 16290T, 16311C, 16325C, 16327T, 16368C, 73G, 146C, 152C, 185A, 189G, 204C, 247A, 263G, 309.1C, 315.1C, 522-3D	2	KEKY002, KESN011											1												
191	L0f3a	16052T, 16129A, 16169T, 16187T, 16189C, 16223T, 16230G, 16278T, 16290T, 16311C, 16325C, 16327T, 16368C, 73G, 146C, 152C, 185A, 189G, 204C, 247A, 263G, 315.1C, 522-3D	1	Keyk045	1	1	1	1	1						1												
192	L0f3a	16052T, 16129A, 16169T, 16187T, 16189C, 16223T, 16230G, 16278T, 16290T, 16293T, 16304C, 16311C, 16325C, 16327T, 16354T, 16368C, 16519C, 146C, 152C, 185A, 189G, 247A, 263G, 309.1C, 315.1C, 522-3D	1	KEBJ005											1												
193	L0f3a	16052T, 16129A, 16169T, 16187T, 16189C, 16223T, 16229A, 16230G, 16278T, 16290T, 16311C, 16325C, 16327T, 16344T, 16354T, 16368C, 16519C, 143A, 146C, 152C, 185A, 189G, 247A, 263G, 309.1C, 315.1C, 522-3D	1	KEGB011	1	1	1		1						1												
194	L0f3a	16052T, 16129A, 16169T, 16187T, 16189C, 16223T, 16230G, 16278T, 16290T, 16311C, 16325C, 16327T, 16354T, 16368C, 16519C, 143A, 146C, 152C, 185A, 189G, 247A, 263G, 309.1C, 315.1C, 522-3D	1	KEYK005	1	1	1		1						1												

195	L0f3a	16052T, 16114A, 16169T, 16187T, 16189C, 16223T, 16230G, 16278T, 16290T, 16311C, 16325C, 16327T, 16368C, 16519C, 143A, 146C, 152C, 185A, 189G, 247A, 263G, 309.1C, 315.1C, 522-3D	1	TZIQ036	1	1	1		1							1	0		1							
196	L0f3a	16051G, 16052T, 16129A, 16169T, 16187T, 16189C, 16223T, 16230G, 16278T, 16290T, 16293G, 16311C, 16325C, 16327T, 16354T, 16368C, 16519C, 143A, 146C, 152C, 185A, 189G, 247A, 263G, 315.1C, 523:, 525G, 526:, 573.1C, 573.2C, 573.3C	1	TZRG055	1	1	1		1							1	0		1							
197	L0f3a	16052T, 16129A, 16169T, 16182.1C, 16183C, 16189C, 16223T, 16230G, 16278T, 16290T, 16311C, 16325C, 16327T, 16354T, 16368C, 16519C, 143A, 146C, 152C, 185A, 189G, 247A, 263G, 309.1C, 309.2C, 312.1C, 522-3D	1	TZDT039	1	1	1		1							1	0		1							
198	L0f3a	16052T, 16129A, 16169T, 16183C, 16189C, 16223T, 16230G, 16278T, 16290T, 16311C, 16325C, 16327T, 16354T, 16368C, 16519C, 143A, 146C, 152C, 185A, 189G, 247A, 263G, 309.1C, 309.2C, 315.1C, 522-3D	1	TZRG038	1	1	1		1							1	1		1							
199	L0f3a	16052T, 16129A, 16169T, 16187T, 16189C, 16223T, 16230G, 16243C, 16278T, 16290T, 16311C, 16325C, 16327T, 16354T, 16368C, 16519C, 51C, 73G, 146C, 152C, 185A, 189G, 247A, 263G, 315.1C, 522-3D	1	TZRG011	1	1	1	1	1							1	1		1							
200	L0f3a	16052T, 16129A, 16169T, 16187T, 16189C, 16223T, 16230G, 16278T, 16290T, 16311C, 16325C, 16327T, 16354T, 16360T, 16368C, 16519C, 143A, 146C, 152C, 189G, 247A, 263G, 315.1C, 522-3D	1	KETA067	1	1	1		1							1	0		1							
201	L0f3a	16052T, 16129A, 16169T, 16187T, 16189C, 16223T, 16230G, 16278T, 16290T, 16311C, 16325C, 16327T, 16354T, 16368C, 16399G, 16519C, 143A, 146C, 152C, 185A, 189G, 247A, 263G, 309.1C, 315.1C, 522-3D	1	TZAK034	1	1	1		1							1	0		1							
202	L0f3b	16052T, 16093C, 16129A, 16169T, 16187T, 16189C, 16223T, 16230G, 16278T, 16290T, 16311C, 16316G, 16325C, 16327T, 16354T, 16368C, 16519C, 143A, 146C, 185A, 189G, 199C, 204C, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	1	KEGK022	1	1	1																			
203	L0f3b	16052T, 16093C, 16129A, 16169T, 16187T, 16189C, 16230G, 16278T, 16290T, 16311C, 16316G, 16325C, 16327T, 16354T, 16368C, 16519C, 143A, 146C, 185A, 189G, 199C, 204C, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	2	KETK076, KESM026	1	1	1										0		1							
204	L0f3b	16052T, 16129A, 16169T, 16187T, 16189C, 16223T, 16230G, 16278T, 16290T, 16311C, 16316G, 16325C, 16327T, 16354T, 16368C, 16519C, 143A, 146C, 185A, 189G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	2	TZIQ059, TZWF021																						
205	L0f3b	16052T, 16129A, 16169T, 16187T, 16189C, 16223T, 16230G, 16278T, 16290T, 16311C, 16316G, 16325C, 16327T, 16354T, 16368C, 16519C, 16527T, 143A, 146C, 185A, 189G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	1	KERD083	1	1	1										0		1							
206	L0f3b	16052T, 16129A, 16169T, 16187T, 16189C, 16207G, 16223T, 16230G, 16278T, 16290T, 16311C, 16316G, 16325C, 16327T, 16354T, 16368C, 16519C, 143A, 146C, 185A, 189G, 236C, 247A, 263G, 309.1C, 315.1C, 522-3D	1	TZIQ009	1	1	1	1																		
207	L0k1	16166C, 16172C, 16187T, 16189C, 16209C, 16214T, 16223T, 16230G, 16278T, 16291G, 16311C, 16519C, 73G, 146C, 152C, 189G, 195C, 198T, 204C, 207A, 247A, 309.1C, 315.1C, 522-3D	1	Hgdp00992																						

SNPs genotyped of L0 haplogroup		
	Snp	**Haplotypes defined**
1	T4586C	L0afbk
2	C5603T	L0abf
3	C4964T, T7148C	L0f
4	A4562G	L0f2a/L0f3a
5	T4688C, 7061G	L0f2a
6	T6152C	L0f2b
7	T6176C	L0f2b1
8	C6164T, A6923G	L0f2b2
9	G4655A	L0f2a3
10	4695C, 5276G, 6923G	L0f1b
11	6152C, 6176C	L02b
12	T4688C, T5063C, A7061G	L0f3a
13	G5231A	L0a
14	T5096C	L0a1
15	G5147A	
16	5964C	L0a1c1
17	A5711G, G6257A	L0a2
18	A5581G	L0a2d
19	C5911T	L0a1b
20	T6050C, C6689A	L0a3
21	T6719C	L0b
22	G4580A, G5773A, C6277T	L0d3

L1	Hg	Control region motif	Abs. Freq	Samples
1	L1b	16093C, 16126C, 16139T, 16187T, 16189C, 16223T, 16264T, 16270T, 16278T, 16293G, 16311C, 16519C, 73G, 135.1T, 152C, 182T, 185T, 189G, 195C, 199C, 247A, 263G, 315.1C, 357G, 522-3D	1	CHLA058
2	L1b	16093C, 16126C, 16187T, 16189C, 16223T, 16264T, 16270T, 16278T, 16293G, 16311C, 16519C, 73G, 152C, 182T, 185T, 195C, 247A, 263G, 315.1C, 357G, 522-3D	4	CAFU010, CAFU011, CAFU014, CAFU016
3	L1b	16126C, 16187T, 16189C, 16223T, 16264T, 16270T, 16274A, 16278T, 16293G, 16311C, 16519C, 73G, 152C, 182T, 185T, 189G, 195C, 199C, 247A, 263G, 315.1C, 357G, 522-3D	3	SDSH007, SDSH008, SDSH020
4	L1b	16126C, 16187T, 16189C, 16223T, 16264T, 16270T, 16278T, 16293G, 16311C, 16519C, 73G, 152C, 182T, 183G, 185T, 195C, 247A, 263G, 315.1C, 357G, 522-3D	1	CABM010
5	L1b	16126C, 16187T, 16189C, 16223T, 16264T, 16270T, 16278T, 16293G, 16311C, 16519C, 73G, 152C, 182T, 185T, 189G, 195C, 199C, 247A, 263G, 315.1C, 357G, 522-3D	1	CAKN007
6	L1b	16126C, 16187T, 16189C, 16223T, 16264T, 16270T, 16278T, 16293G, 16311C, 16519C, 73G, 152C, 182T, 185T, 189G, 195C, 247A, 263G, 315.1C, 357G, 522-3D	1	CHLA037
7	L1b	16126C, 16187T, 16189C, 16223T, 16264T, 16270T, 16278T, 16293G, 16311C, 16519C, 73G, 152C, 182T, 185T, 195C, 247A, 263G, 315.1C, 357G, 522-3D	4	CAFU023, CAFU025, CAFU028, CAKN020
8	L1b	16126C, 16189C, 16223T, 16264T, 16270T, 16293G, 16311C, 16519C, 73G, 146C, 152C, 185T, 234G, 247A, 263G, 315.1C, 357G, 522-3D	1	CAAS034
9	L1b1a	16126C, 16187T, 16189C, 16223T, 16264T, 16270T, 16278T, 16311C, 16519C, 16525G, 73G, 152C, 182T, 185T, 189G, 195C, 199C, 247A, 263G, 309.1C, 315.1C, 357G, 522-3D	1	SDSH015
10	L1b1a	16126C, 16187T, 16189C, 16223T, 16264T, 16270T, 16278T, 16311C, 16519C, 73G, 151T, 152C, 182T, 185T, 189G, 247A, 263G, 315.1C, 357G, 522-3D	1	CAKN013
11	L1b1a	16126C, 16187T, 16189C, 16223T, 16264T, 16270T, 16278T, 16311C, 16519C, 73G, 151T, 152C, 182T, 185T, 195C, 247A, 263G, 315.1C, 357G, 522-3D	2	CHLA001, CABM047
12	L1b1a	16126C, 16187T, 16189C, 16223T, 16264T, 16270T, 16278T, 16311C, 16519C, 73G, 152C, 182T, 185T, 189G, 195C, 204C, 247A, 263G, 315.1C, 357G, 522-3D	1	CABM031
13	L1b1a	16126C, 16187T, 16189C, 16223T, 16264T, 16270T, 16278T, 16311C, 16519C, 73G, 152C, 182T, 185T, 195C, 247A, 263G, 315.1C, 357G, 522-3D	1	CABM041
14	L1b1a2	16075C, 16108T, 16126C, 16187T, 16189C, 16223T, 16264T, 16270T, 16278T, 16289G, 16293G, 16311C, 16519C, 73G, 151T, 152C, 182T, 185T, 195C, 247A, 315.1C, 357G, 522-3D	1	KEBJ021a
15	L1c1	15970C, 16129A, 16172C, 16173T, 16188A, 16189C, 16223T, 16256T, 16278T, 16293G, 16294T, 16311C, 16360T, 16368C, 16519C, 73G, 151T, 152C, 182T, 186A, 189G, 195C, 198T, 247A, 263G, 297G, 315.1C, 316A, 522-3D	1	KETV029
16	L1c1	16051G, 16129A, 16187T, 16189C, 16214T, 16234T, 16249C, 16258G, 16278T, 16293G, 16294T, 16311C, 16360T, 16519C, 44.1C, 73G, 151T, 152C, 186A, 189C, 195C, 247A, 263G, 297G, 315.1C, 316A, 467T, 522-3D	3	CAMV039, CAPL050, CAPL095
17	L1c1	16187T, 16189C, 16214T, 16234T, 16249C, 16278T, 16293G, 16294T, 16311C, 16360T, 16519C, 16527T, 44.1C, 73G, 151T, 152C, 182T, 186A, 189C, 195C, 204C, 247A, 263G, 297G, 309.1C, 315.1C, 316A, 522-3D	2	CAPL016, CAPL036
18	L1c1	16187T, 16189C, 16214T, 16234T, 16249C, 16278T, 16293G, 16294T, 16311C, 16360T, 16519C, 44.1C, 73G, 151T, 152C, 182T, 186A, 189C, 195C, 204C, 247A, 263G, 297G, 315.1C, 316A, 522-3D	1	CAPM001
19	L1c1a	15881.1A, 15887.1A, 16051G, 16129A, 16187T, 16189C, 16214T, 16234T, 16258G, 16274A, 16278T, 16293G, 16294T, 16311C, 16360T, 16519C, 44.1C, 73G, 151T, 152C, 186A, 189C, 195C, 247A, 263G, 297G, 315.1C, 316A, 467T, 522-3D	1	CAPM038
20	L1c1a	16051G, 16129A, 16187T, 16189C, 16214T, 16234T, 16249C, 16258G, 16274A, 16278T, 16293G, 16294T, 16311C, 16360T, 16512C, 16519C, 16529C, 44.1C, 73G, 151T, 152C, 186A, 189C, 195C, 247A, 263G, 297G, 315.1C, 316A, 467T, 522-3D	1	CAPL059

21	L1c1a	16051G, 16129A, 16187T, 16189C, 16214T, 16234T, 16249C, 16258G, 16274A, 16278T, 16293G, 16294T, 16311C, 16360T, 16519C, 44.1C, 73G, 146C, 151T, 152C, 186A, 189C, 195C, 198T, 247A, 263G, 297G, 309.1C, 315.1C, 316A, 467T, 522-3D	1	CAPB004
22	L1c1a	16051G, 16129A, 16187T, 16189C, 16214T, 16234T, 16249C, 16258G, 16274A, 16278T, 16293G, 16294T, 16311C, 16360T, 16519C, 44.1C, 73G, 151T, 152C, 186A, 189C, 195C, 198T, 247A, 263G, 297G, 309.1C, 315.1C, 316A, 467T, 522-3D	1	CAPB049
23	L1c1a	16051G, 16129A, 16187T, 16189C, 16214T, 16234T, 16249C, 16258G, 16274A, 16278T, 16293G, 16294T, 16311C, 16360T, 16519C, 44.1C, 73G, 151T, 152C, 186A, 189C, 195C, 198T, 247A, 263G, 297G, 315.1C, 316A, 467T, 522-3D	2	CAPL038, CAPL043
24	L1c1a	16051G, 16129A, 16187T, 16189C, 16214T, 16234T, 16249C, 16258G, 16274A, 16278T, 16293G, 16294T, 16311C, 16360T, 16519C, 44.1C, 73G, 151T, 152C, 186A, 189C, 195C, 247A, 263G, 297G, 309.1C, 315.1C, 316A, 467T, 522-3D	3	CAPL017, CAPL053, CAPL056
25	L1c1a	16051G, 16129A, 16187T, 16189C, 16214T, 16234T, 16249C, 16258G, 16274A, 16278T, 16293G, 16294T, 16311C, 16360T, 16519C, 44.1C, 73G, 151T, 152C, 186A, 189C, 195C, 247A, 263G, 297G, 315.1C, 316A, 467T	1	CAGZ005
26	L1c1a	16051G, 16129A, 16187T, 16189C, 16214T, 16234T, 16249C, 16258G, 16274A, 16278T, 16293G, 16294T, 16311C, 16360T, 16519C, 44.1C, 73G, 151T, 152C, 186A, 189C, 195C, 247A, 263G, 297G, 315.1C, 316A, 467T, 522-3D	10	CAMV001, CAMV033, CAPB079, CAPB088, CAPB089, CAPL013, CAPM027, CAPM032, CAPM043, CAPM045
27	L1c1a	16051G, 16129A, 16187T, 16189C, 16214T, 16234T, 16249C, 16258G, 16274A, 16278T, 16293G, 16294T, 16311C, 16519C, 44.1C, 73G, 151T, 152C, 186A, 189C, 195C, 198T, 247A, 263G, 297G, 315.1C, 316A, 467T, 522-3D	1	CAPB067
28	L1c1a	16051G, 16129A, 16187T, 16189C, 16214T, 16234T, 16249C, 16258G, 16274A, 16278T, 16294T, 16311C, 16360T, 16519C, 44.1C, 73G, 151T, 152C, 186A, 189C, 195C, 247A, 263G, 297G, 315.1C, 316A, 467T, 522-3D	2	CAPB018, CAPB043
29	L1c1a	16051G, 16129A, 16187T, 16189C, 16214T, 16234T, 16258G, 16274A, 16278T, 16293G, 16294T, 16311C, 16360T, 16519C, 44.1C, 73G, 151T, 152C, 186A, 189C, 195C, 247A, 263G, 297G, 315.1C, 316A, 467T, 522-3D	13	CAPL060, CAPM005, CAPM014, CAPM015, CAPM017, CAPM018, CAPM020, CAPM024, CAPM031, CAPM036, CAPM037, CAPM042, CAPM051
30	L1c1a	16051G, 16129A, 16187T, 16189C, 16214T, 16249C, 16258G, 16274A, 16278T, 16293G, 16294T, 16311C, 16360T, 16519C, 44.1C, 73G, 151T, 152C, 186A, 189C, 195C, 247A, 263G, 297G, 309.1C, 315.1C, 316A, 467T, 522-3D	1	CAPM023
31	L1c1a	16086C, 16129A, 16187T, 16189C, 16223T, 16241G, 16274A, 16278T, 16291T, 16294T, 16311C, 16360T, 16519C, 73G, 151T, 152C, 182T, 186A, 189C, 195C, 198T, 247A, 263G, 297G, 315.1C, 316A	2	CAMV015, CAMV021
32	L1c1a	16129A, 16187T, 16189C, 16214T, 16234T, 16249C, 16274A, 16278T, 16294T, 16311C, 16360T, 16519C, 44.1C, 73G, 151T, 152C, 183G, 186A, 189C, 195C, 204C, 247A, 263G, 297G, 315.1C, 316A, 467T, 522-3D	1	CAPB078
33	L1c1a	16129A, 16187T, 16189C, 16214T, 16234T, 16249C, 16274A, 16278T, 16294T, 16311C, 16360T, 16519C, 44.1C, 73G, 151T, 152C, 186A, 189C, 195C, 204C, 247A, 263G, 297G, 315.1C, 316A, 467T, 522-3D	9	CAMV025, CAMV028, CAPB005, CAPB019, CAPB022, CAPB046, CAPB056, CAPL001, CAPL033, CAPL077
34	L1c1a	16129A, 16187T, 16189C, 16223T, 16274A, 16278T, 16293G, 16294T, 16311C, 16360T, 16519C, 64T, 73G, 93G, 95C, 152C, 182T, 186A, 189C, 195C, 236C, 247A, 263G, 297G, 315.1C, 316A, 522-3D	1	CAMV041
35	L1c1a	16129A, 16187T, 16189C, 16223T, 16274A, 16278T, 16293G, 16294T, 16311C, 16360T, 16519C, 73G, 93G, 95C, 152C, 182T, 186A, 189C, 195C, 214W, 236C, 247A, 263G, 297G, 315.1C, 316A, 522-3D	1	CAPM019
36	L1c1a	16129A, 16187T, 16189C, 16223T, 16274A, 16278T, 16293G, 16294T, 16311C, 16360T, 16519C, 73G, 93G, 95C, 152C, 182T, 186A, 189C, 195C, 236C, 247A, 263G, 297G, 315.1C, 316A, 522-3D	4	CAPL035, CAPL041, CAPL049, CAPM028

37	L1c1a	16129A, 16187T, 16189C, 16223T, 16274A, 16278T, 16293G, 16294T, 16311C, 16360T, 16519C, 73G, 93G, 95C, 182T, 186A, 189C, 195C, 236C, 247A, 263G, 297G, 309.1C, 315.1C, 316A, 522-3D	9	CAMV018, CAPB015, CAPB033, CAPB074, CAPB080, CAPB082, CAPB090, CAPL020, CAPL096
38	L1c1a	16129A, 16187T, 16189C, 16223T, 16274A, 16278T, 16293G, 16294T, 16311C, 16360T, 16519C, 73G, 93G, 95C, 182T, 186A, 189C, 195C, 236C, 247A, 263G, 297G, 315.1C, 316A, 522-3D	11	CAAS004, CAPB066, CAPB068, CAPB069, CAPB071, CAPB098, CAPL062, CAPM009, CAPM016, CAPM050, CAPM059
39	L1c1a	16129A, 16187T, 16189C, 16223T, 16274A, 16278T, 16293G, 16294T, 16311C, 16360T, 73G, 93G, 95C, 152C, 182T, 186A, 189C, 195C, 236C, 247A, 263G, 297G, 315.1C, 316A, 522-3D	2	CAPL089, CAPM007
40	L1c1a	16187T, 16189C, 16214T, 16234T, 16249C, 16274A, 16278T, 16294T, 16311C, 16360T, 16519C, 44.1C, 73G, 151T, 152C, 186A, 189C, 195C, 204C, 247A, 263G, 297G, 315.1C, 316A, 467T, 522-3D	1	CAPM011
41	L1c1a1b	16129A, 16187T, 16189C, 16241T, 16245T, 16270T, 16278T, 16293G, 16294T, 16311C, 16519C, 73G, 151T, 152C, 182T, 186A, 189C, 195C, 198T, 247A, 263G, 297G, 315.1C, 316A, 522-3D	1	CABM017
42	L1c2	15924G, 16129A, 16187T, 16189C, 16214T, 16223T, 16265C, 16278T, 16286A, 16291T, 16294T, 16311C, 16360T, 16519C, 16527T, 73G, 151T, 152C, 182T, 183G, 186A, 189C, 195C, 198T, 247A, 263G, 297G, 309.1C, 315.1C, 316A, 513A	1	CAMV003
43	L1c2	16129A, 16148T, 16187T, 16189C, 16213A, 16223T, 16265C, 16278T, 16286G, 16294T, 16311C, 16360T, 16519C, 16527T, 73G, 151T, 152C, 182T, 186A, 189C, 195C, 198T, 247A, 263G, 297G, 309.1C, 315.1C, 316A, 522-3D	1	CAMV030
44	L1c2	16129A, 16163G, 16187T, 16189C, 16259T, 16265C, 16278T, 16286G, 16294T, 16311C, 16320T, 16360T, 16519C, 16527T, 73G, 151T, 152C, 182T, 186A, 189C, 195C, 198T, 247A, 263G, 297G, 315.1C, 316A, 522-3D	1	KETA009
45	L1c2	16129A, 16169T, 16187T, 16189C, 16223T, 16265C, 16278T, 16286G, 16294T, 16311C, 16360T, 16519C, 16527T, 73G, 151T, 152C, 182T, 186A, 189C, 195C, 198T, 247A, 263G, 297G, 309.1C, 315.1C, 316A, 471C, 522-3D	1	KEPR001
46	L1c2	16129A, 16187T, 16189C, 16223T, 16265C, 16278T, 16286G, 16294T, 16311C, 16360T, 16519C, 16527T, 73G, 151T, 152C, 182T, 186A, 189C, 195C, 198T, 247A, 263G, 297G, 309.1C, 315.1C, 316A, 471C, 522-3D	1	KEPR010
47	L1c2	16172C, 16187T, 16189C, 16223T, 16265C, 16278T, 16286G, 16294T, 16311C, 16360T, 16390A, 16519C, 16527T, 73G, 151T, 152C, 182T, 186A, 189C, 195C, 198T, 247A, 263G, 297G, 309.1C, 315.1C, 316A, 385G, 471C, 522-3D	1	XH009
48	L1c3	15905C, 15978T, 16017C, 16129A, 16163G, 16187T, 16189C, 16209C, 16223T, 16278T, 16293G, 16294T, 16298C, 16311C, 16360T, 16519C, 73G, 152C, 182T, 186A, 189C, 194T, 198T, 247:, 263G, 309.1C, 315.1C, 316A, 522-3D, 629C	1	CAMV032
49	L1c3	15905C, 15978T, 16017C, 16129A, 16163G, 16187T, 16189C, 16209C, 16223T, 16278T, 16293G, 16294T, 16311C, 16360T, 16519C, 73G, 151T, 152C, 182T, 186A, 189C, 247A, 263G, 315.1C, 316A, 522-3D, 629C	2	CAMV038, TZPR031
50	L1c3	15905C, 15978T, 16093C, 16129A, 16183C, 16189C, 16215G, 16223T, 16278T, 16294T, 16311C, 16355T, 16360T, 16390A, 16519C, 73G, 151T, 152C, 182T, 186A, 189C, 247A, 263G, 309.1C, 315.1C, 316A, 522-3D	1	KETA068
51	L1c3	15905C, 15978T, 16129A, 16183C, 16189C, 16215G, 16223T, 16278T, 16294T, 16311C, 16320T, 16355T, 16360T, 16519C, 73G, 151T, 152C, 182T, 183G, 186A, 189C, 204C, 247A, 257G, 263G, 309.1C, 315.1C, 316A, 522-3D	1	TZDT066
52	L1c6	16117C, 16129A, 16172C, 16173T, 16188A, 16189C, 16223T, 16256T, 16278T, 16291T, 16294T, 16311C, 16360T, 16368C, 16519C, 73G, 151T, 152C, 182T, 186A, 189G, 195C, 198T, 247A, 263G, 297G, 309.1C, 315.1C, 316A, 522-3D	1	KETA065

					L2a	L2a	L2a2	L2a2a	L2a1a				
L2	Hg	Controlregion	Abs. Freq	Samples	1	2	3	4	5	6	7	8	9
1	L2f	16223T, 16224C, 16278T, 16295T, 16309G, 16390A, 16519C, 73G, 146C, 152C, 182T, 195C, 215G, 249:, 263G, 315.1C, 511T, 537T	1	KEPR017									
2	L2f	16223T, 16224C, 16278T, 16309G, 16390A, 16399G, 16519C, 16523G, 73G, 146C, 152C, 182T, 183G, 186T, 263G, 309.1C, 315.1C, 513:, 514:	1	KEMR017									
3	L2f	16223T, 16224C, 16278T, 16309G, 16390A, 16399G, 16519C, 73G, 146C, 152C, 182T, 183G, 186T, 263G, 309.1C, 315.1C, 513:, 514:	1	SDSH010									
4	L2f	16223T, 16224C, 16278T, 16309G, 16390A, 16399G, 16519C, 73G, 146C, 152C, 182T, 183G, 186T, 263G, 315.1C, 513:, 514:	5	KELO104, KESN004, KESN008, KESN054, KESN061									
5	L2a1	15924G, 16189C, 16213A, 16223T, 16278T, 16294T, 16309G, 16390A, 73G, 146C, 152C, 195C, 263G, 315.1C, 385G	1	CHBU010									
6	L2a1	16189C, 16223T, 16278T, 16294T, 16309G, 16390A, 73G, 146C, 152C, 195C, 263G, 309.1C, 315.1C	1	CAGZ012	1					1			
7	L2a1	16189C, 16192T, 16223T, 16278T, 16294T, 16309G, 16390A, 73G, 146C, 152C, 195C, 263G, 309.1C, 315.1C	1	KESB009	1					1			
8	L2a1a1	16223T, 16278T, 16294T, 16309G, 16368C, 16390A, 16519C, 73G, 146C, 189G, 195C, 263G, 309.1C, 315.1C, 524.1A, 524.2C, 524.3A, 524.4C, 524.5A, 525.1C	1	KETA005	1								
9	L2a1a	16223T, 16278T, 16294T, 16309G, 16390A, 16519C, 73G, 146C, 152C, 195C, 263G, 309.1C, 315.1C	9	CHLA040, CHLA068, CHLA069, KEBN001, KEBN006, KEBN013, KEBN036, KEPK062, KEPK109	1								
10	L2a1a2	16092C, 16223T, 16278T, 16286T, 16294T, 16309G, 16390A, 16519C, 73G, 146C, 152C, 195C, 263G, 315.1C	2	KEPK071, TZRG003	1					1			
11	L2a1a2	16223T, 16278T, 16286T, 16294T, 16309G, 16390A, 16519C, 73G, 146C, 152C, 195C, 198T, 263G, 309.1C, 315.1C	1	KETG073	1					1			
12	L2a1a	16183C, 16189C, 16223T, 16278T, 16294T, 16309G, 16390A, 16519C, 73G, 146C, 152C, 195C, 198T, 263G, 309.1C, 309.2C, 315.1C	1	CHLA070	1								
13	L2a1a	16183C, 16189C, 16223T, 16278T, 16294T, 16309G, 16390A, 16519C, 73G, 146C, 152C, 195C, 263G, 315.1C	2	SDDN022, TZRG049	1					1			
14	L2a1a	16223T, 16278T, 16294T, 16309G, 16390A, 16519C, 73G, 146C, 152C, 195C, 263G, 315.1C	2	CAFU012, KEBN004			1			1			
15	L2a1b	16182C, 16183C, 16189C, 16223T, 16278T, 16290T, 16294T, 16309G, 16390A, 16465A, 73G, 146C, 152C, 195C, 263G, 315.1C	1	KEPR002	1								
16	L2a1b	16182C, 16183C, 16189C, 16223T, 16278T, 16290T, 16294T, 16309G, 16390A, 73G, 146C, 152C, 195C, 263G, 315.1C	1	XH010									
17	L2a1c	16086C, 16223T, 16278T, 16294T, 16309G, 16390A, 73G, 143A, 146C, 152C, 195C, 198T, 263G, 315.1C	2	CAMV010, CAMV014	1	1				1			
18	L2a1c	16086C, 16223T, 16278T, 16294T, 16309G, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 315.1C	1	TZAK030									
19	L2a1c	16223T, 16278T, 16294T, 16309G, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 309.1C, 315.1C	2	TZRG027, TZRG036	1					1			
20	L2a1c	16223T, 16278T, 16294T, 16309G, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 309.1C, 315.1C, 385G	1	TZMS013									
21	L2a1c	15884A, 16223T, 16278T, 16294T, 16309G, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 309.1C, 315.1C	1	CAAS009	1				1				
22	L2a1c	16129A, 16223T, 16278T, 16294T, 16309G, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 309.1C, 315.1C	1	TZBG047	1				1				
23	L2a1c	16157C, 16192.1T, 16223T, 16278T, 16294T, 16390A, 16400T, 73G, 143A, 146C, 195C, 263G, 315.1C, 502:, 522-3D	1	SDDN005	1					1			
24	L2a1d	16175C, 16209C, 16223T, 16278T, 16294T, 16301T, 16354T, 16390A, 73G, 143A, 146C, 152C, 182T, 189G, 195C, 263G, 309.1C, 315.1C, 522-3D	1	CAAS005									

25	L2a1d	16189C, 16209C, 16223T, 16278T, 16294T, 16301T, 16354T, 16390A, 73G, 143A, 146C, 152C, 182T, 189G, 195C, 263G, 315.1C, 522-3D	1	SDNR009	1					1		1		
26	L2a1d	16209C, 16223T, 16278T, 16294T, 16301T, 16354T, 16390A, 16519C, 73G, 143A, 146C, 152C, 182T, 189G, 195C, 249:, 263G, 315.1C, 522-3D	1	SDNR002	1					1		1		
27	L2a1d	16209C, 16223T, 16278T, 16294T, 16301T, 16354T, 16390A, 73G, 146C, 152C, 189G, 195C, 200G, 263G, 309.1C, 309.2C, 315.1C, 522-3D	1	KEBR074	1					1		1		
28	L2a1e	15883A, 16218G, 16223T, 16278T, 16294T, 16309G, 16390A, 16519C, 73G, 143A, 146C, 152C, 195C, 263G, 315.1C	1	KETK018	1					1				
29	L2a1e	16093C, 16223T, 16278T, 16294T, 16309G, 16390A, 16519C, 73G, 143A, 146C, 152C, 195C, 263G, 315.1C	1	CHLA046	1									
30	L2a1e	16223T, 16231C, 16278T, 16294T, 16309G, 16390A, 16399G, 16519C, 73G, 143A, 146C, 152C, 195C, 263G, 309.1C, 315.1C, 398C	1	SDNY017	1					1				
31	L2a1e	15924G, 16093C, 16213A, 16223T, 16278T, 16294T, 16309G, 16390A, 16519C, 73G, 143A, 146C, 152C, 195C, 263G, 309.1C, 315.1C	1	CAAS017	1				1					
32	L2a1e	16223T, 16278T, 16294T, 16309G, 16390A, 16519C, 73G, 143A, 146C, 151T, 152C, 195C, 198T, 263G, 309.1C, 315.1C	2	KEGU033, KEMG023	1				1					
33	L2a1f	16183C, 16189C, 16209C, 16223T, 16278T, 16294T, 16390A, 16519C, 73G, 146C, 152C, 195C, 263G, 309.1C, 315.1C, 522-3D	2	KENJ047, KENJ005	1					1				1
34	L2a1f	16189C, 16192T, 16223T, 16278T, 16294T, 16309G, 16390A, 16519C, 73G, 146C, 151T, 152C, 195C, 263G, 309.1C, 315.1C	1	CAGZ013	1					1		1		
35	L2a1f	16090C, 16189C, 16192T, 16223T, 16278T, 16294T, 16309G, 16390A, 16519C, 73G, 146C, 152C, 195C, 263G, 309.1C, 315.1C	1	TZPR009	1									1
36	L2a1f	16093C, 16189C, 16192T, 16223T, 16278T, 16294T, 16309G, 16390A, 16519C, 73G, 146C, 152C, 195C, 263G, 309.1C, 315.1C	1	KETK040	1					1				
37	L2a1f	16189C, 16192T, 16223T, 16249C, 16278T, 16294T, 16309G, 16390A, 16519C, 73G, 146C, 152C, 195C, 263G, 309.1C, 315.1C	1	KELO006	1					1				
38	L2a1f	16189C, 16192T, 16223T, 16271C, 16278T, 16294T, 16309G, 16390A, 16519C, 73G, 146C, 152C, 195C, 263G, 309.1C, 315.1C	1	CAPM054										
39	L2a1f	16189C, 16192T, 16223T, 16278T, 16294T, 16309G, 16311C, 16390A, 16519C, 73G, 146C, 195C, 263G, 309.1C, 309.2C, 315.1C	1	TZTR008	1					1				
40	L2a1f	16189C, 16192T, 16223T, 16278T, 16294T, 16309G, 16390A, 16519C, 73G, 146C, 152C, 195C, 198T, 263G, 309.1C, 315.1C	1	laka										
41	L2a1f	16189C, 16192T, 16223T, 16278T, 16294T, 16309G, 16390A, 16519C, 73G, 146C, 152C, 195C, 263G, 309.1C, 315.1C	1	CAMV006										
42	L2a1f	16189C, 16192T, 16223T, 16278T, 16294T, 16309G, 16390A, 16519C, 73G, 146C, 152C, 195C, 263G, 315.1C	1	CAMV020			1							
43	L2a1g	15924G, 16092C, 16183C, 16189C, 16223T, 16278T, 16291T, 16294T, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 315.1C	1	KEBR080	1					1				
44	L2a1g	15924G, 16183C, 16189C, 16192T, 16223T, 16239T, 16278T, 16291T, 16294T, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 315.1C	1	KEGB012	1									
45	L2a1g	15924G, 16183C, 16189C, 16223T, 16278T, 16291T, 16294T, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 309.1C, 315.1C	1	KEGU056										
46	L2a1g	15924G, 16183C, 16189C, 16223T, 16278T, 16291T, 16294T, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 315.1C	2	KEBN032, KEMN018	1					1				
47	L2a1g	15924G, 16183C, 16189C, 16223T, 16278T, 16294T, 16295T, 16390A, 16519C, 73G, 143A, 146C, 152C, 195C, 263G, 315.1C	1	KEGU043										
48	L2a1g	15924G, 16183C, 16189C, 16223T, 16278T, 16294T, 16390A, 73G, 143A, 146C, 152C, 195C, 239C, 263G, 315.1C	2	KEWT001, KEWT002a										
49	L2a1g	15924G, 16183C, 16189C, 16223T, 16278T, 16294T, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 315.1C	3	KEBR004, KEGU031, KEKS010	1									
50	L2a1g	16086C, 16183C, 16189C, 16223T, 16278T, 16294T, 73G, 143A, 146C, 152C, 195C, 263G, 315.1C	1	CAGZ035										

51	L2a1g	15924G, 16182C, 16183C, 16189C, 16223T, 16278T, 16294T, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 309.1C, 315.1C	1	TZPR001	1										1
52	L2a1g	16092C, 16182C, 16183C, 16189C, 16223T, 16278T, 16290T, 16294T, 16309G, 16390A, 73G, 146C, 152C, 195C, 263G, 315.1C	1	KEPR009	1										
53	L2a1g	16041G, 16189C, 16223T, 16278T, 16294T, 16309G, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 309.1C, 315.1C, 518T, 573.1	1	TZDR013	1						1				
54	L2a1g	16041G, 16189C, 16223T, 16278T, 16294T, 16309G, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 315.1C, 518T	3	tzhz015, tzhz031, tzhz051	1										
55	L2a1g	16041G, 16189C, 16223T, 16278T, 16294T, 16309G, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 315.1C, 518T, 573.1	1	TZDT030	1						1				
56	L2a1g	16092C, 16189C, 16192T, 16223T, 16278T, 16292T, 16294T, 16309G, 16354T, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 315.1C, 522-3D	1	CAKN010	1						1				
57	L2a1g	16093C, 16171T, 16189C, 16223T, 16256T, 16278T, 16294T, 16309G, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 309.1C, 315.1C	1	KEPK020	1						1				
58	L2a1g	16124C, 16183C, 16189C, 16223T, 16278T, 16294T, 16309G, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 315.1C	1	KERD085	1										
59	L2a1g	16129A, 16169T, 16189C, 16192T, 16223T, 16256T, 16278T, 16294T, 16309G, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 309.1C, 315.1C	1	KESN012	1										
60	L2a1g	16129A, 16189C, 16192T, 16223T, 16256T, 16278T, 16294T, 16309G, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 309.1C, 315.1C	1	KEPK074	1						1				
61	L2a1g	16172C, 16189C, 16223T, 16278T, 16294T, 16309G, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 315.1C	6	SDBA026, SDHD005, SDHD011, SDHD013, SDHD015, SDHD023	1						1				
62	L2a1g	16183C, 16189C, 16192T, 16223T, 16278T, 16292T, 16294T, 16309G, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 315.1C, 522-3D	1	SDNR026	1						1				
63	L2a1g	16189C, 16192T, 16223T, 16256T, 16278T, 16294T, 16309G, 16318CT, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 309.1C, 315.1C	1	KETA011											
64	L2a1g	16189C, 16192T, 16223T, 16256T, 16278T, 16294T, 16309G, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 309.1C, 315.1C	2	KEPK017, KESN047	1						1				
65	L2a1g	16189C, 16192T, 16223T, 16278T, 16292T, 16294T, 16309G, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 315.1C, 522-3D	4	SDBA004, SDBA024, SDDN036, SDHD018	1						1				
66	L2a1g	16189C, 16192T, 16223T, 16278T, 16294T, 16309G, 16320T, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 279C, 309.1C, 315.1C	1	CHBU007											
67	L2a1g	16189C, 16192T, 16223T, 16278T, 16294T, 16309G, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 309.1C, 309.2C, 315.1C	1	CHBU015											
68	L2a1g	16189C, 16192T, 16223T, 16278T, 16294T, 16309G, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 309.1C, 315.1C, 522-3D	1	KERD056	1						1				
69	L2a1g	16189C, 16192T, 16223T, 16278T, 16294T, 16309G, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 315.1C	5	KEOG029, KEOG046, TZMS020, TZMS063, KEOG019	1						1				
70	L2a1g	16189C, 16223T, 16256T, 16278T, 16294T, 16309G, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 309.1C, 315.1C	1	KESN050	1						1				
71	L2a1g	16189C, 16223T, 16278T, 16292T, 16294T, 16309G, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 315.1C, 522-3D	2	CAAS003, SDBA018	1										
72	L2a1g	16189C, 16223T, 16278T, 16292T, 16294T, 16309G, 16390A, 73G, 143A, 146C, 152C, 263G, 315.1C, 522-3D	1	SDBA020	1						1				
73	L2a1g	16189C, 16223T, 16278T, 16294T, 16309G, 16390A, 73G, 143A, 146C, 150T, 152C, 195C, 263G, 315.1C, 522-3D	1	KETV047											
74	L2a1g	15784C, 15884C, 16093C, 16183C, 16189C, 16278T, 16294T, 16309G, 16390A, 16519C, 73G, 143A, 146C, 152C, 195C, 263G, 309.1C, 315.1C, 522-3D, 769A		KEGK011-ale											
75	L2a1g	15911G, 16111T, 16189C, 16192T, 16223T, 16278T, 16294T, 16309G, 16390A, 16519C, 73G, 143A, 146C, 152C, 195C, 263G, 309.1C, 315.1C	1	SDDN011	1						1				

76	L2a1g	16183C, 16189C, 16223T, 16278T, 16294T, 16309G, 16390A, 16519C, 16527T, 73G, 143A, 146C, 150T, 152C, 195C, 263G, 309.1C, 315.1C, 522-3D	2	KEMR007, KEMR028		1				1			
77	L2a1g	16183C, 16189C, 16223T, 16278T, 16294T, 16309G, 16390A, 16519C, 73G, 143A, 146C, 152C, 195C, 200G, 263G, 315.1C, 573.1	1	TZTR037	1								
78	L2a1g	16183C, 16189C, 16223T, 16278T, 16294T, 16309G, 16390A, 16519C, 73G, 143A, 146C, 152C, 195C, 263G, 309.1C, 315.1C	1	CAKN001	1								
79	L2a1g	16183C, 16189C, 16278T, 16294T, 16309G, 16390A, 16519C, 73G, 143A, 146C, 152C, 195C, 263G, 309.1C, 309.2C, 315.1C	1	TZRG017	1					1			
80	L2a1g	16183C, 16189C, 16278T, 16294T, 16309G, 16390A, 16519C, 73G, 143A, 146C, 152C, 195C, 263G, 309.1C, 315.1C	2	KENJ068, KESB016	1					1			1
81	L2a1g	16189C, 16223T, 16278T, 16292T, 16294T, 16309G, 16390A, 16519C, 73G, 143A, 146C, 152C, 195C, 263G, 309.1C, 315.1C	1	KESM036	1								
82	L2a1g	16189C, 16223T, 16278T, 16294T, 16309G, 16390A, 16519C, 73G, 143A, 146C, 152C, 195C, 200G, 263G, 315.1C	2	TZTR020, tzsw128	1					1			
83	L2a1g	16182C, 16183C, 16189C, 16223T, 16278T, 16294T, 16390A, 16519C, 73G, 143A, 146C, 152C, 195C, 263G, 315.1C, 534T	1	CAFU009									
84	L2a1g	16092C, 16145A, 16183C, 16189C, 16223T, 16224C, 16278T, 16294T, 16390A, 73G, 143A, 146C, 152C, 189G, 195C, 199C, 263G, 309.1C, 315.1C	1	TZDR017a	1								
85	L2a1g	16145A, 16183C, 16189C, 16223T, 16224C, 16278T, 16294T, 16390A, 73G, 143A, 146C, 152C, 189G, 195C, 199C, 263G, 309.1C, 309.2C, 315.1C	1	TZWF013	1					1			
86	L2a1g	16145A, 16183C, 16189C, 16223T, 16278T, 16294T, 16390A, 16519C, 64T, 73G, 143A, 146C, 152C, 189G, 195C, 263G, 309.1C, 315.1C	1	KETG014	1					1			
87	L2a1g	16183C, 16189C, 16219G, 16223T, 16278T, 16294T, 16390A, 16519C, 73G, 143A, 146C, 152C, 195C, 263G, 315.1C, 534T	1	CAFU008	1					1			
88	L2a1g	16183C, 16189C, 16223T, 16278T, 16294T, 16390A, 16519C, 73G, 143A, 146C, 152C, 195C, 198T, 263G, 315.1C	1	KELO004	1					1			
89	L2a1g	16183C, 16189C, 16223T, 16278T, 16294T, 16390A, 16519C, 73G, 143A, 146C, 152C, 195C, 263G, 315.1C	1	KEWT008	1								
90	L2a1h	16129A, 16223T, 16278T, 16294T, 16390A, 73G, 143A, 146C, 152C, 182T, 195C, 263G, 309.1C, 315.1C, 522-3D	2	KEOR030a, KEOR030B	1								
91	L2a1h	16223T, 16234T, 16249C, 16278T, 16294T, 16295T, 16390A, 73G, 146C, 152C, 195C, 263G, 309.1C, 315.1C	1	KEPR011	1								
92	L2a1h	16223T, 16234T, 16249C, 16278T, 16294T, 16295T, 16390A, 73G, 146C, 195C, 263G, 309.1C, 315.1C	1	TZRG019	1					1			
93	L2a1h	16223T, 16278T, 16294T, 16368C, 16390A, 16519C, 73G, 146C, 189G, 195C, 234G, 263G, 309.1C, 315.1C, 524.1A, 524.2C, 524.3A, 525.1C	1	KETA006	1								
94	L2a1h	16041G, 16189C, 16223T, 16278T, 16294T, 16390A, 73G, 143A, 152C, 195C, 263G, 315.1C, 518T	1	tzhz025	1					1			
95	L2a1h	16129A, 16223T, 16256T, 16264T, 16278T, 16294T, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 309.1C, 315.1C	1	TZMS017	1					1			
96	L2a1h	16189C, 16223T, 16245T, 16278T, 16294T, 16390A, 73G, 143A, 146C, 263G, 309.1C, 315.1C	1	CABM009									
97	L2a1h	16223T, 16234T, 16249C, 16264T, 16278T, 16294T, 16295T, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 315.1C	2	Keil013, Keil022	1					1			
98	L2a1h	16223T, 16294T, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 309.1C, 315.1C	1	KESB038									1
99	L2a1h	16223T, 16234T, 16249C, 16278T, 16294T, 16295T, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 315.1C	1	CAMV011	1				1				
100	L2a1h	16223T, 16256T, 16264T, 16278T, 16294T, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 315.1C	1	KERD062	1				1				
101	L2a1h	16189C, 16223T, 16278T, 16294T, 16390A, 16519C, 73G, 143A, 146C, 152C, 195C, 200G, 263G, 315.1C, 573.1	1	KETG026	1					1			
102	L2a1h	16189C, 16223T, 16278T, 16294T, 16390A, 16519C, 73G, 143A, 146C, 152C, 195C, 263G, 315.1C	1	KEWT020	1								

103	L2a1h	16183C, 16189C, 16192T, 16223T, 16278T, 16294T, 16311C, 16318CT, 16390A, 73G, 152C, 195C, 263G, 315.1C	1	KETG098	1					1				
104	L2a1i	16103G, 16175G, 16189C, 16192T, 16223T, 16263C, 16278T, 16294T, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 315.1C	1	SDNR005										
105	L2a1i	16186.1T, 16189C, 16192.1T, 16223T, 16278T, 16294T, 16390A, 73G, 143A, 146C, 152C, 263G, 315.1C	1	SDNY016	1					1				
106	L2a1i	16189C, 16192T, 16223T, 16266T, 16278T, 16291T, 16294T, 16355T, 16390A, 16427.1C, 73G, 143A, 146C, 152C, 195C, 263G, 309.1C, 315.1C	1	CAAS020	1									
107	L2a1i	16189C, 16192T, 16223T, 16278T, 16291T, 16294T, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 315.1C	1	CAGZ024	1					1				
108	L2a1i	16189C, 16192T, 16223T, 16278T, 16292T, 16294T, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 309.1C, 309.2C, 315.1C	1	SDHD035										
109	L2a1i	16189C, 16192T, 16223T, 16278T, 16294T, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 315.1C	1	CAGZ018	1					1				
110	L2a2	16170G, 16189C, 16223T, 16229C, 16264T, 16278T, 16294T, 16311C, 16390A, 16519C, 73G, 152C, 182T, 263G, 315.1C, 513:, 514:, 593A	1	KETK013	1					1				
111	L2a2	16183C, 16189C, 16223T, 16229C, 16264T, 16278T, 16294T, 16311C, 16390A, 16519C, 73G, 152C, 182T, 263G, 309.1C, 309.2C, 315.1C, 346C, 498.1C, 513:, 514:, 593C	1	KELO028	1					1				
112	L2a2	15939T, 16189C, 16223T, 16229C, 16278T, 16291T, 16311C, 16390A, 16519C, 73G, 152C, 182T, 195C, 263G, 309.1C, 309.2C, 315.1C	1	CAGZ011										
113	L2a2	15894A, 15939T, 16189C, 16223T, 16229C, 16278T, 16291T, 16294T, 16311C, 16390A, 16519C, 73G, 152C, 182T, 194T, 195C, 263G, 315.1C	1	CHLA064	1		1							
114	L2a2	15939T, 16066G, 16183C, 16189C, 16223T, 16229C, 16278T, 16291T, 16294T, 16311C, 16354T, 16390A, 16519C, 73G, 152C, 182T, 195C, 263G, 309.1C, 315.1C	1	CHBU014										
115	L2a2	15939T, 16067T, 16189C, 16223T, 16229C, 16278T, 16291T, 16294T, 16298C, 16311C, 16390A, 16519C, 73G, 152C, 182T, 183G, 195C, 263G, 315.1C	1	CAFU033	1		1							
116	L2a2	15939T, 16093C, 16189C, 16223T, 16229C, 16278T, 16291T, 16294T, 16311C, 16390A, 16519C, 73G, 152C, 182T, 195C, 263G, 315.1C	1	CHLA041	1		1							
117	L2a2	15939T, 16111T, 16183C, 16189C, 16223T, 16229C, 16278T, 16291T, 16294T, 16311C, 16390A, 16519C, 73G, 152C, 182T, 263G, 315.1C	1	KENJ076	1					1				
118	L2a2	15939T, 16111T, 16189C, 16223T, 16229C, 16278T, 16291T, 16294T, 16311C, 16390A, 16519C, 73G, 152C, 182T, 263G, 309.1C, 315.1C	1	KELY015	1					1				
119	L2a2	15939T, 16182C, 16183C, 16189C, 16215G, 16223T, 16229C, 16260T, 16278T, 16291T, 16294T, 16311C, 16390A, 16519C, 73G, 152C, 182T, 263G, 309.1C, 309.2C, 315.1C	1	KEPK091	1					1				
120	L2a2	15939T, 16183C, 16189C, 16223T, 16229C, 16260T, 16278T, 16291T, 16294T, 16311C, 16390A, 16519C, 73G, 152C, 182T, 263G, 309.1C, 315.1C	1	KESN002	1		1							
121	L2a2	15939T, 16183C, 16189C, 16223T, 16229C, 16278T, 16291T, 16294T, 16311C, 16390A, 16519C, 73G, 152C, 182T, 195C, 263G, 309.1C, 315.1C	1	CHLA005	1		1							
122	L2a2	15939T, 16183C, 16189C, 16223T, 16229C, 16278T, 16291T, 16294T, 16311C, 16390A, 16519C, 73G, 152C, 182T, 195C, 263G, 309.1C, 315.1C, 522-3D	1	CAPB075	1		1							
123	L2a2	15939T, 16183C, 16189C, 16223T, 16229C, 16278T, 16291T, 16294T, 16311C, 16390A, 16519C, 73G, 152C, 182T, 195C, 263G, 309.1C, 315.1C, 573.1C, 573.2C, 573.3	1	CAAS030	1		1							
124	L2a2	15939T, 16189C, 16223T, 16229C, 16270T, 16278T, 16291T, 16294T, 16311C, 16390A, 16519C, 73G, 152C, 182T, 263G, 315.1C, 522-3D, 534T	3	KETA004, KETA022, KETA073	1		1							
125	L2a2	15939T, 16189C, 16223T, 16229C, 16278T, 16291T, 16294T, 16311C, 16390A, 16519C, 73G, 152C, 182T, 195C, 198T, 263G, 315.1C	1	SDDN035	1					1				
126	L2a2	15939T, 16189C, 16223T, 16229C, 16278T, 16291T, 16294T, 16311C, 16390A, 16519C, 73G, 152C, 182T, 195C, 263G, 309.1C, 315.1C	2	SDDN023, SDDN026	1					1				

127	L2a2	15939T, 16189C, 16223T, 16229C, 16278T, 16291T, 16294T, 16311C, 16390A, 16519C, 73G, 152C, 182T, 257G, 263G, 309.1C, 315.1C, 663G	1	SDDN032	1					1				
128	L2a2	15939T, 16189C, 16229C, 16278T, 16291T, 16294T, 16311C, 16390A, 16519C, 73G, 152C, 182T, 263G, 309.1C, 315.1C	1	SDSH002	1					1				
129	L2a2a	16189C, 16192T, 16223T, 16278T, 16294T, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 315.1C, 518T	1	CHLA033	1			1						
130	L2a2b	16189C, 16223T, 16278T, 16294T, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 315.1C	1	SDNR008	1					1	1			
131	L2b	15924G, 15944.1T, 16114A, 16129A, 16145A, 16213A, 16223T, 16278T, 16390A, 16519C, 73G, 146C, 150T, 152C, 182T, 195C, 198T, 199C, 204C, 263G, 315.1C, 522-3D	1	CHLA055										
132	L2b	15944.1T, 16114A, 16129A, 16145A, 16213A, 16223T, 16278T, 16298C, 16390A, 73G, 146C, 150T, 152C, 182T, 195C, 198T, 199C, 204C, 263G, 315.1C	1	KEGU054										
133	L2b	15944.1T, 16114A, 16129A, 16145A, 16213A, 16223T, 16278T, 16309G, 16390A, 16519C, 73G, 146C, 150T, 152C, 182T, 195C, 198T, 199C, 204C, 263G, 315.1C	1	Keil007										
134	L2b	15944.1T, 16114A, 16129A, 16145A, 16213A, 16223T, 16278T, 16311C, 16390A, 73G, 146C, 150T, 152C, 182T, 195C, 198T, 199C, 204C, 263G, 315.1C	1	TZMS087										
135	L2b	16114A, 16129A, 16213A, 16218T, 16223T, 16278T, 16390A, 16519C, 73G, 146C, 150T, 152C, 182T, 183G, 186T, 195C, 198T, 204C, 263G, 315.1C	2	CHBU005, CHBU006										
136	L2b	16114A, 16129A, 16213A, 16223T, 16278T, 16291T, 16354T, 16390A, 16519C, 73G, 146C, 150T, 152C, 182T, 195C, 198T, 204C, 263G, 315.1C	1	SDSH016										
137	L2b	16114A, 16129A, 16213A, 16223T, 16278T, 16354T, 16390A, 73G, 146C, 150T, 152C, 182T, 195C, 198T, 204C, 263G, 315.1C	2	CHBU011, San49										
138	L2b	16114A, 16129A, 16213A, 16223T, 16278T, 16355T, 16362C, 16390A, 73G, 150T, 152C, 182T, 195C, 198T, 204C, 207A, 263G, 315.1C, 418T, 522-3D	1	CAAS007										
139	L2b	16114A, 16129A, 16213A, 16223T, 16278T, 16355T, 16362C, 16390A, 73G, 150T, 152C, 182T, 195C, 198T, 204C, 263G, 315.1C, 418T, 522-3D	1	CAKN016										
140	L2b	16114A, 16129A, 16213A, 16223T, 16278T, 16390A, 73G, 146C, 150T, 152C, 182T, 183G, 195C, 198T, 204C, 263G, 315.1C, 522-3D	1	SDNY011										
141	L2c	15883A, 16223T, 16278T, 16318G, 16390A, 16527T, 73G, 93G, 146C, 150T, 152C, 182T, 195C, 198T, 263G, 315.1C, 325T, 522-3D	1	SDDN010										
142	L2c	16093C, 16223T, 16278T, 16294T, 16390A, 73G, 93G, 146C, 150T, 152C, 182T, 195C, 198T, 263G, 315.1C, 325T, 522-3D	1	CHLA065										
143	L2c	16223T, 16278T, 16318G, 16390A, 73G, 93G, 146C, 150T, 152C, 182T, 195C, 198T, 207A, 263G, 315.1C, 325T, 522-3D	1	CHLA025										
144	L2c	16223T, 16278T, 16318G, 16390A, 73G, 93G, 146C, 150T, 152C, 182T, 195C, 198T, 263G, 315.1C, 325T, 522-3D	1	CAMV016										
145	L2c	16223T, 16278T, 16390A, 16519C, 73G, 93G, 150T, 152C, 182T, 195C, 198T, 263G, 315.1C, 325T, 499A, 522-3D	1	SDHD042										
146	L2d	15937G, 16129A, 16189C, 16278T, 16300G, 16354T, 16390A, 16399G, 16519C, 73G, 146C, 150T, 195C, 263G, 309.1C, 315.1C, 456T	1	CAGZ039										
147	L2d	16093C, 16129A, 16183C, 16189C, 16278T, 16300G, 16311C, 16354T, 16390A, 16399G, 16519C, 73G, 146C, 150T, 195C, 263G, 315.1C, 456T	2	TZRG013, tzsw127										
148	L2d	16093C, 16129A, 16189C, 16278T, 16300G, 16311C, 16354T, 16390A, 16399G, 16519C, 73G, 146C, 150T, 195C, 263G, 309.1C, 315.1C, 456T	1	TZTR026										
149	L2e	15889C, 16111A, 16145A, 16184T, 16223T, 16239T, 16278T, 16355T, 16390A, 16399G, 16400T, 16519C, 73G, 146C, 150T, 152C, 263G, 315.1C, 479G	1	SDDN003										

SNPs genotyped of L0 haplogroup		
	Snp	**Haplotypes defined**
1	7175C, 7274	L2a
2	7175C, 7274	L2a
3	A6752G	L2a2
4	C6656T	L2a2a
5	A6663G	L2a1a
6	Gl3590A	
7	14118G	
8	13395G, 13934T	
9	C3594T	

L3a	Hg	Controlregion	Abs. Freq	Samples	13708A	T3618C	2357T
1	L3a	16221T, 16223T, 16316G, 73G, 152C, 263G, 315.1C, 522-3D, 573T	1	KEYK028			
2	L3a	16223T, 16256T, 16316G, 73G, 150T, 152C, 195C, 263G, 315.1C, 522-3D	1	Kemn008	1		
3	L3a	16223T, 16316G, 73G, 150T, 152C, 195C, 263G, 315.1C, 522-3D	1	KETV001		1	
4	L3a1	16169T, 16223T, 16316G, 16519C, 73G, 152C, 263G, 315.1C, 522-3D, 573T	3	Kemn015, kesm005, kesm013			
5	L3a1	16223T, 16278T, 16316G, 16519C, 73G, 152C, 263G, 309.1C, 315.1C, 522-3D, 573T	3	KEOR016, KEOR024a, KETV066			
6	L3a1	16223T, 16278T, 16316G, 16519C, 73G, 263G, 309.1C, 315.1C, 522-3D, 573T	2	KEPR027, KEPR035			
7	L3a1	16223T, 16316G, 16519C, 73G, 152C, 263G, 309.1C, 315.1C, 499A, 522-3D, 573T, 576G	1	KESB005			
8	L3a1a	16104T, 16223T, 16254G, 16292T, 16316G, 16519C, 73G, 152C, 195C, 198T, 247A, 263G, 315.1C, 522-3D, 573T	3	KESY016, KESY018, KESY022			
9	L3a1a	16223T, 16254G, 16316G, 16519C, 73G, 151T, 152C, 247A, 309.1C, 315.1C, 522-3D, 573T	1	KEYK029			
10	L3a1a	16254G, 16260T, 16316G, 16519C, 73G, 152C, 263G, 315.1C, 368G, 522-3D, 573T	1	KEGK035			
11	L3a1a	16254G, 16316G, 16519C, 73G, 152C, 199C, 263G, 309.1C, 315.1C, 522-3D, 573T	2	KEPK077, KETG099			
12	L3a1a	16254G, 16316G, 16519C, 73G, 152C, 263G, 315.1C, 522-3D, 573T	13	KEBR029, KEBR048, KEEL009, KEEL016, KEEL027, KEEL030, KEEL042, KEEL045, KEEL050, KEEL051, KEGU007, KERD021, KEWT033			
13	L3a2	16093C, 16223T, 16254G, 16294T, 16311C, 16316G, 73G, 152C, 195C, 263G, 315.1C, 522-3D	1	TZTR024			
14	L3a2	16129A, 16223T, 16254G, 16316G, 73G, 150T, 152C, 195C, 263G, 309.1C, 315.1C, 522-3D	1	TZIQ057			
15	L3a2	16148T, 16178C, 16223T, 16254G, 16311C, 16316G, 73G, 151T, 152C, 195C, 198T, 263G, 309.1C, 315.1C, 522-3D	1	KELY023			
16	L3a2	16178C, 16223T, 16254G, 16311C, 16316G, 73G, 151T, 152C, 195C, 198T, 200G, 263G, 315.1C, 522-3D	1	KEBJ018	1		
17	L3a2	16178C, 16223T, 16254G, 16311C, 16316G, 73G, 151T, 152C, 195C, 198T, 263G, 315.1C, 522-3D	5	KEPR005, KEYK004, Keyk012, TZIQ037, keyk008			
18	L3a2	16221T, 16223T, 16254G, 16316G, 73G, 152C, 263G, 315.1C, 522-3D, 573T	1	KEOR009			
19	L3a2	16223T, 16254G, 16316G, 73G, 152C, 200G, 263G, 309.1C, 315.1C, 522-3D	1	KEYK002			
20	L3a2	16223T, 16254G, 16316G, 73G, 152C, 263G, 309.1C, 315.1C, 522-3D, 573T	4	KEKP009, KESM031, KESM032, KESM033			
21	L3a2	16148T, 16178C, 16223T, 16254G, 16311C, 16316G, 16519C, 73G, 151T, 152C, 195C, 198T, 263G, 309.1C, 315.1C, 522-3D	1	KEMR018	1		1

					L3b
L3b	Hg	Controlregion	Abs. Freq	Samples	A10086G
1	L3b1	15883A, 15944D, 15950A, 16093C, 16124C, 16223T, 16278T, 16362C, 16519C, 73G, 263G, 315.1C, 522-3D	1	KELO012	1
2	L3b1	15883A, 15944D, 16093C, 16124C, 16166G, 16223T, 16278T, 16362C, 16519C, 73G, 263G, 315.1C, 522-3D	1	KELY034	1
3	L3b1	15883A, 15944D, 16093C, 16124C, 16223T, 16259T, 16278T, 16362C, 16519C, 73G, 263G, 315.1C, 522-3D	1	KEBN040	1
4	L3b1	15883A, 15944D, 16093C, 16124C, 16223T, 16278T, 16362C, 16519C, 73G, 263G, 309.1C, 315.1C, 522-3D	1	KETA074	1
5	L3b1	15883A, 15944D, 16093C, 16124C, 16223T, 16278T, 16362C, 16519C, 73G, 263G, 315.1C, 522-3D	11	KELO074, KELY025, KEND002, KEPR028, KESB017,	1
6	L3b1	15883A, 15944D, 16124C, 16223T, 16262T, 16278T, 16362C, 16519C, 73G, 263G, 315.1C, 522-3D	1	KELO035	1
7	L3b1	15883A, 15944D, 16124C, 16223T, 16278T, 16362C, 16519C, 73G, 263G, 315.1C, 522-3D	1	KELO014	1
8	L3b1	15944D, 16093C, 16124C, 16223T, 16278T, 16362C, 16519C, 73G, 263G, 315.1C, 522-3D	1	KELO011	1
9	L3b1	15944D, 16093C, 16223T, 16278T, 16295T, 16362C, 16519C, 73G, 263G, 315.1C, 522-3D	1	CAFU015	1
10	L3b1	15944D, 16093C, 16223T, 16278T, 16362C, 16519C, 73G, 263G, 315.1C, 522-3D	2	CAFU003, CAFU004	1
11	L3b1	15944D, 16124C, 16223T, 16278T, 16290T, 16311C, 16362C, 16519C, 73G, 263G, 315.1C, 522-3D	1	CHLA066	1
12	L3b1	15944D, 16124C, 16223T, 16278T, 16294T, 16311C, 16362C, 16519C, 73G, 263G, 315.1C, 522-3D	2	KESB010, KESB046	1
13	L3b1	15944D, 16124C, 16223T, 16278T, 16311C, 16362C, 16519C, 73G, 189G, 195C, 263G, 315.1C, 522-3D	1	CAGZ028	1
14	L3b1	15944D, 16124C, 16223T, 16278T, 16311C, 16362C, 16519C, 73G, 263G, 315.1C, 522-3D	5	CAKN004, CHLA023, KELO019, KELY011, SDNR017	
15	L3b1	15944D, 16124C, 16223T, 16278T, 16362C, 16519C, 73G, 146C, 263G, 309.1C, 315.1C, 522-3D	1	CHBU009	
16	L3b1	15944D, 16124C, 16223T, 16278T, 16362C, 16519C, 73G, 263G, 309.1C, 315.1C, 522-3D	1	CHLA004	1
17	L3b1	15944D, 16124C, 16223T, 16278T, 16362C, 16519C, 73G, 263G, 309.1C, 315.1C, 522-3D, 573.1C	1	CAKN019	1
18	L3b1	15944D, 16145A, 16223T, 16278T, 16362C, 16519C, 73G, 189G, 263G, 315.1C, 522-3D	1	CAGZ026	1
19	L3b1b	15883A, 15944D, 16124C, 16223T, 16278T, 16362C, 16519C, 73G, 152C, 263G, 315.1C, 522-3D	1	KELO016	1
20	L3b1b	15944D, 16086C, 16124C, 16223T, 16278T, 16362C, 16519C, 73G, 151T, 152C, 263G, 315.1C, 494A, 522-3D	1	CAFU005	1
21	L3b1b	15944D, 16093C, 16124C, 16223T, 16278T, 16311C, 16362C, 16519C, 73G, 152C, 263G, 309.1C, 315.1C, 522-3D	1	SDDN030	1
22	L3b1b	15944D, 16124C, 16170G, 16223T, 16278T, 16362C, 16519C, 73G, 152C, 263G, 309.1C, 315.1C, 522-3D	1	KEGK036	1
23	L3b1b	15944D, 16124C, 16170G, 16223T, 16278T, 16362C, 16519C, 73G, 152C, 263G, 315.1C, 522-3D	1	KENJ051	1
24	L3b1b	15944D, 16124C, 16189C, 16223T, 16266T, 16278T, 16362C, 16519C, 73G, 152C, 263G, 309.1C, 315.1C, 522-3D	1	CAGZ010	1
25	L3b1b	15944D, 16124C, 16223T, 16278T, 16290T, 16311C, 16362C, 16519C, 73G, 152C, 263G, 309.1C, 315.1C, 378T, 482C, 522-3D	1	CAAS028	1
26	L3b1b	15944D, 16124C, 16223T, 16278T, 16311C, 16362C, 16519C, 73G, 146C, 152C, 263G, 309.1C, 315.1C, 482C, 522-3D	1	SDSH005	1
27	L3b1b	15944D, 16124C, 16223T, 16278T, 16362C, 16519C, 73G, 152C, 263G, 309.1C, 315.1C, 522-3D	2	CAGZ032, CAKN030	1
28	L3b1b	15944D, 16124C, 16278T, 16311C, 16362C, 16519C, 73G, 152C, 263G, 315.1C, 482C, 522-3D	1	CAGZ006	1
29	L3b1a	15944D, 16223T, 16278T, 16311C, 16362C, 16519C, 73G, 152C, 263G, 315.1C, 522-3D	3	TZTR015, TZTR028, tzswl15	
30	L3b1a	15944D, 16223T, 16278T, 16311C, 16362C, 16519C, 73G, 263G, 309.1C, 315.1C, 522-3D	1	KEGK033a	
31	L3b1a	15944D, 16145A, 16223T, 16278T, 16362C, 16519C, 73G, 263G, 315.1C, 522-3D	1	CAFU006	
32	L3b2	15944D, 16124C, 16183C, 16189C, 16223T, 16278T, 16293G, 16304C, 16362C, 16519C, 16527T, 73G, 263G, 315.1C, 522-3D	1	SDBA016	1
33	L3b2	15944D, 16124C, 16183C, 16189C, 16223T, 16278T, 16311C, 16362C, 16527T, 73G, 263G, 315.1C, 522-3D	1	CHBU003	

					L3d
L3d	Hg	Controlregion	Abs. Freq	Samples	T8618C
1	L3d	16051G, 16124C, 16223T, 16256T, 16304C, 73G, 152C, 263G, 309.1C, 315.1C, 522-3D	1	CAFU026	1
2	L3d	16086C, 16111T, 16124C, 16223T, 73G, 152C, 199C, 263G, 315.1C, 522-3D	1	CAAS031	1
3	L3d	16086C, 16124C, 16223T, 16311C, 73G, 152C, 189G, 195C, 263G, 315.1C, 522-3D	1	KETV051	1
4	L3d	16111T, 16124C, 16223T, 16399G, 16519C, 73G, 152C, 199C, 263G, 309.1C, 315.1C	1	CAGZ023	1
5	L3d	16124C, 16183C, 16189C, 16223T, 16278T, 16304C, 16311C, 73G, 152C, 263G, 309.1C, 315.1C, 522-3D	1	KEPR018	1
6	L3d	16124C, 16183C, 16189C, 16223T, 16278T, 16304C, 16311C, 73G, 152C, 263G, 315.1C, 522-3D	3	CAMV024, KELO013, TZTR022	0
7	L3d	16124C, 16223T, 16256T, 16278T, 16368C, 16399G, 73G, 152C, 263G, 309.1C, 315.1C, 522-3D	1	SDSH018	1
8	L3d	16124C, 16223T, 16261T, 16288C, 16311C, 73G, 152C, 153G, 263G, 315.1C, 522-3D	1	CAKN025	1
9	L3d	16124C, 16223T, 16264T, 73G, 152C, 263G, 309.1C, 315.1C, 522-3D	1	CAGZ001	1
10	L3d	16124C, 16223T, 16288C, 73G, 150T, 152C, 153G, 228A, 263G, 315.1C, 522-3D	1	CHBU001	
11	L3d	16124C, 16223T, 16519C, 73G, 146C, 152C, 263G, 309.1C, 315.1C, 522-3D	1	KELY009	1
12	L3d	16124C, 16223T, 73G, 152C, 153G, 263G, 315.1C, 522-3D	1	CAAS008	1
13	L3d	16124C, 16223T, 73G, 152C, 189G, 195C, 263G, 315.1C, 522-3D	1	CAAS036	1
14	L3d	16124C, 16223T, 73G, 152C, 263G, 315.1C, 522-3D	2	KETV064, KEGU001	1
15	L3d	16126C, 16223T, 73G, 151T, 152C, 263G, 315.1C, 522-3D	1	CAAS006	1
16	L3d	16214A, 16223T, 73G, 151T, 152C, 263G, 315.1C, 522-3D	2	CHLA036, CHLA038	1
17	L3d	16223T, 16354T, 73G, 151T, 152C, 263G, 315.1C, 522-3D	1	CHLA003	1
18	L3d	16223T, 16519C, 73G, 151T, 152C, 263G, 315.1C, 522-3D	1	CHLA056	1
19	L3d	16223T, 73G, 151T, 152C, 263G, 309.1C, 315.1C, 522-3D, 573.1C	1	CHLA032	1
20	L3d	16223T, 73G, 151T, 152C, 263G, 315.1C, 522-3D	1	CHLA028	1
21	L3d	16124C, 16148T, 16192T, 16223T, 16311C, 16399G, 73G, 146C, 150T, 152C, 207A, 263G, 315.1C, 489C	1	KESB006	
22	L3dl	15944:, 16124C, 16223T, 16278T, 16362C, 16519C, 73G, 263G, 315.1C, 522-3D	2	CAAS011, KEMN016	1
23	L3dla	16124C, 16183C, 16189C, 16223T, 16257T, 16319A, 16519C, 73G, 152C, 263G, 315.1C, 522-3D	1	CABM045	
24	L3dla	16124C, 16189C, 16223T, 16311C, 16319A, 73G, 150T, 152C, 263G, 309.1C, 315.1C, 522-3D	1	TZBG052	1
25	L3dla	16124C, 16189C, 16223T, 16319A, 73G, 150T, 152C, 263G, 309.1C, 315.1C, 522-3D	1	KETK006	
26	L3dla	15892C, 16124C, 16223T, 16254G, 16319A, 16385G, 73G, 152C, 153G, 263G, 315.1C, 522-3D	3	KEPK011, Kesm008, Kesm009	1
27	L3dla	15941C, 16117C, 16124C, 16223T, 16319A, 73G, 152C, 263G, 315.1C, 522-3D	1	SDDN019	1
28	L3dla	15941C, 16124C, 16223T, 16319A, 73G, 152C, 263G, 315.1C, 522-3D	1	CHBU022	
29	L3dla	16124C, 16223T, 16304C, 16319A, 73G, 150T, 152C, 195C, 263G, 309.1C, 315.1C, 522-3D	1	KELO008	1
30	L3dla	16124C, 16223T, 16319A, 16468C, 73G, 150T, 152C, 263G, 309.1C, 315.1C, 522-3D	1	TZTR007	1
31	L3dla	16124C, 16223T, 16319A, 16519C, 73G, 152C, 263G, 315.1C, 522-3D	1	CAKN029	1
32	L3dla	16124C, 16223T, 16319A, 73G, 150T, 152C, 182T, 263G, 309.1C, 315.1C, 522-3D	1	XH014	
33	L3dla	16124C, 16223T, 16319A, 73G, 150T, 152C, 263G, 309.1C, 315.1C, 508G, 522-3D	2	KEOR022b, KEOR031	1
34	L3dla	16124C, 16223T, 16319A, 73G, 150T, 152C, 263G, 309.1C, 315.1C, 522-3D	8	KELY035, KETV033, KETV048, Keil020, Kesm011, TZRG007, TZRG053, KELO023	1
35	L3dla	16124C, 16223T, 16319A, 73G, 150T, 152C, 263G, 315.1C, 522-3D	2	CABM032, TZMS057	0
36	L3dla	16124C, 16223T, 16319A, 73G, 151T, 152C, 263G, 315.1C, 522-3D	1	KEBR073	1
37	L3dla	16124C, 16223T, 16319A, 73G, 152C, 263G, 309.1C, 315.1C, 522-3D	4	CAGZ038, KEGU006, KEWT007, SDBA028	1
38	L3dla	16124C, 16223T, 16319A, 73G, 152C, 263G, 315.1C, 522-3D	2	KEBR002, KEBR092	1
39	L3dlb	16223T, 73G, 151T, 152C, 263G, 309.1C, 315.1C, 522-3D	1	SDSH009	1

					T2352C L3e
L3e	Hg	Controlregion	Abs. Freq	Samples	
1	L3e1	15942C, 15961A, 16189C, 16223T, 16291T, 16327T, 73G, 150T, 189G, 200G, 263G, 315.1C	1	CAKN021	1
2	L3e1	15942C, 16183C, 16189C, 16223T, 16260T, 16327T, 73G, 150T, 189G, 200G, 207A, 220.1C, 263G, 309.1C, 315.1C	1	CABM038	
3	L3e1	15942C, 16183C, 16189C, 16223T, 16260T, 16327T, 73G, 150T, 189G, 200G, 207A, 263G, 309.1C, 315.1C	1	CABM033	
4	L3e1	15942C, 16185T, 16223T, 16327T, 16519C, 73G, 150T, 189G, 200G, 263G, 309.1C, 315.1C	1	TZMS082	1
5	L3e1	15942C, 16223T, 16327T, 73G, 150T, 152C, 189G, 193G, 200G, 263G, 309.1C, 315.1C	1	CAFU001	1
6	L3e1	15942C, 16223T, 16327T, 73G, 150T, 189G, 200G, 263G, 309.1C, 315.1C	1	CHLA021	1
7	L3e1	15942C, 16223T, 16327T, 73G, 150T, 189G, 263G, 309.1C, 315.1C	1	CABM039	1
8	L3e1	15942C, 16223T, 16327T, 73G, 150T, 189G, 200G, 215G, 235G, 263G, 315.1C	1	CHLA031	1
9	L3e1a	15942C, 16185T, 16223T, 16311C, 16327T, 73G, 150T, 185A, 189G, 263G, 315.1C	1	CABM003	
10	L3e1a	15942C, 16185T, 16223T, 16311C, 16327T, 73G, 150T, 185A, 189G, 200G, 263G, 315.1C	2	KETK036, TZBG051	1
11	L3e1a	15942C, 16185T, 16223T, 16311C, 16327T, 73G, 150T, 185A, 189G, 263G, 315.1C	2	CABM001, Keil003	
12	L3e1b	15942C, 16185T, 16209C, 16223T, 16327T, 73G, 150T, 152C, 189G, 195C, 200G, 207A, 263G, 315.1C	1	KELY030	1
13	L3e1b	15942C, 16185T, 16209C, 16223T, 16327T, 73G, 150T, 152C, 189G, 195C, 200G, 207A, 263G, 309.1C, 315.1C	1	TZMS037	
14	L3e1d	15942C, 15977T, 16176T, 16223T, 16234T, 16287T, 16291T, 16327T, 73G, 150T, 152C, 189G, 200G, 263G, 315.1C	1	TZRG012	0
15	L3e1d	15942C, 16176T, 16223T, 16234T, 16287T, 16291T, 16292T, 16327T, 73G, 150T, 152C, 189G, 200G, 263G, 315.1C	1	tzsw021	1
16	L3e1d	15942C, 16176T, 16223T, 16234T, 16287T, 16291T, 16311C, 16327T, 73G, 150T, 152C, 189G, 200G, 205A, 263G, 315.1C	1	TZSW007	1
17	L3e1d	15942C, 16176T, 16223T, 16234T, 16287T, 16291T, 16327T, 73G, 150T, 152C, 189G, 200G, 263G, 315.1C	1	TZTR009	1
18	L3e1g	15942C, 16223T, 16325D, 16327T, 73G, 150T, 185A, 189G, 263G, 309.1C, 315.1C	5	KELO009, KEPR006, KETV037, TZRG026, TZSW033	1
19	L3e1g	15942C, 16223T, 16325D, 16327T, 73G, 150T, 185A, 189G, 263G, 315.1C	1	tzsw131	
20	L3e1g	15942C, 16223T, 16239T, 16325D, 73G, 150T, 185A, 189G, 263G, 309.1C, 315.1C	1	XH008	
21	L3e2	15930A, 16223T, 16311C, 16320T, 16519C, 73G, 150T, 195C, 198T, 263G, 315.1C	1	SDNY002	1
22	L3e2	16051G, 16093C, 16223T, 16311C, 16320T, 16519C, 73G, 150T, 195C, 198T, 263G, 315.1C	1	CAMV029	1
23	L3e2	16051G, 16223T, 16311C, 16320T, 16519C, 73G, 150T, 195C, 198T, 263G, 315.1C	1	CAMV042	1
24	L3e2	16126C, 16223T, 16299G, 16320T, 16519C, 73G, 150T, 195C, 263G, 315.1C	1	SDNR019	1
25	L3e2	16223T, 16261T, 16295T, 16299G, 16320T, 16519C, 73G, 93.1A, 150T, 195C, 263G, 315.1C	3	KEBJ003, KEMR006, KESB024	1
26	L3e2	16223T, 16261T, 16295T, 16299G, 16320T, 16519C, 73G, 93.1A, 150T, 195C, 263G, 315.1C, 573.1C, 573.2C, 573.3C	1	KEMR003	
27	L3e2	16223T, 16311C, 16320T, 16519C, 73G, 150T, 195C, 198T, 200G, 263G, 315.1C, 573.1	1	CHLA043	1
28	L3e2	16223T, 16311C, 16320T, 16519C, 73G, 150T, 195C, 198T, 263G, 315.1C	1	CAFU032	1
29	L3e2	16223T, 16320T, 16399G, 16519C, 73G, 150T, 195C, 198T, 263G, 315.1C, 499A	3	CAKN018, CAKN031, KEPR024	1
30	L3e2b	16150T, 16172C, 16183C, 16189C, 16223T, 16311C, 16320T, 16519C, 73G, 150T, 195C, 263G, 309.1C, 315.1C, 522-3D	1	CAKN034	1
31	L3e2b	16172C, 16183C, 16189C, 16223T, 16320T, 16519C, 73G, 150T, 152C, 195C, 263G, 315.1C	1	KELO101	
32	L3e2b	16172C, 16183C, 16189C, 16223T, 16320T, 16519C, 73G, 150T, 195C, 263G, 309.1C, 309.2C, 315.1C	2	CAKN011, TZDT065	1
33	L3e2b	16172C, 16183C, 16189C, 16223T, 16320T, 16519C, 73G, 150T, 195C, 263G, 309.1C, 315.1C	6	CHBU018, KENJ075, KEPR016, KEPR019, TZPR013, XH026	
34	L3e2b	16172C, 16183C, 16189C, 16223T, 16320T, 16519C, 73G, 150T, 195C, 263G, 315.1C	3	KELO090, XH025, XH028	1
35	L3e3	16223T, 16265T, 16519C, 73G, 150T, 195C, 228A, 263G, 315.1C, 522-3D	1	KELY016	1
36	L3e3	16223T, 16265T, 16519C, 73G, 150T, 195C, 263G, 309.1C, 315.1C, 460C, 573.1C, 573.2C, 573.3C, 573.4	1	KEGK032a	1
37	L3e3	16223T, 16265T, 16519C, 73G, 150T, 195C, 263G, 309.1C, 315.1C, 522-3D, 573.1C, 573.2	2	KEGK018, KEGK024	1
38	L3e3	16223T, 16265T, 16519C, 73G, 150T, 195C, 263G, 309.1C, 315.1C, 523:, 525G, 526:, 573.1C, 573.2C, 573.3C, 573.4C	2	KETA027, TZMS084	1
39	L3e3	16223T, 16265T, 16519C, 73G, 150T, 195C, 263G, 309.1C, 315.1C, 523:, 525G, 526:, 573.1C, 573.2C, 573.3C, 573.4C, 573.5C	3	KEGK025, TZBG039, TZRG030	1
40	L3e3	16223T, 16265T, 16519C, 73G, 150T, 195C, 263G, 315.1C, 523:, 525G, 526:, 573.1C, 573.2C, 573.3C, 573.4C, 573.5C	1	TZRG006	1
41	L3e3	16223T, 16265T, 73G, 150T, 195C, 263G, 315.1C, 522-3D	1	KELO103	1
42	L3e3-5	16124C, 16223T, 16519C, 73G, 152C, 263G, 309.1C, 315.1C, 522-3D	1	CAAS015	1
43	L3e4	15924G, 16051G, 16145A, 16223T, 16264T, 16519C, 73G, 146C, 150T, 189G, 263G, 309.1C, 315.1C, 522-3D	1	KEOR018	1
44	L3e4	16051G, 16223T, 16256T, 16519C, 73G, 150T, 263G, 315.1C, 522-3D	2	CHLA012, CHLA063	1
45	L3e5	15884A, 15929G, 16041G, 16223T, 16519C, 73G, 150T, 263G, 315.1C, 398C, 522-3D	1	CHLA039	1
46	L3e5	16037G, 16041G, 16223T, 16519C, 73G, 150T, 263G, 315.1C, 398C, 522-3D	1	CAKN006	1
47	L3e5	16041G, 16192T, 16223T, 16519C, 73G, 150T, 152C, 263G, 309.1C, 315.1C, 398C, 522-3D	1	CHBU002	
48	L3e5	16041G, 16223T, 16519C, 73G, 150T, 195C, 263G, 315.1C, 398C, 522-3D	2	CAAS014, CAAS021	1

49	L3e5	16041G, 16223T, 16519C, 73G, 150T, 195C, 263G, 315.1C, 398C, 522-3D, 538C	1	CAAS010	0
50	L3e5	16041G, 16223T, 16519C, 73G, 150T, 263G, 309.1C, 315.1C, 398C, 522-3D	1	SDNR003	1
51	L3e5	16041G, 16223T, 16519C, 73G, 150T, 263G, 315.1C, 398C, 522-3D	2	CAKN008, CAKN015	
52	L3e5	16041G, 16223T, 16519C, 73G, 150T, 263G, 315.1C, 398C, 522-3D, 652	1	CAGZ025	1

							L3f	L3f
L3f	Hg	Control region	Abs. freq	Samples	3505G	3693A	T3396C	T4218C
1	L3f1a	15874G, 15944D, 16111A, 16209C, 16223T, 16311C, 73G, 263G, 309.1C, 315.1C	1	KETA042				
2	L3f	15944D, 16111A, 16209C, 16223T, 16311C, 73G, 263G, 309.1C, 315.1C	1	KEBR094				
3	L3f	15944D, 16209C, 16223T, 16234T, 16235G, 16256T, 16311C, 73G, 263G, 315.1C, 522-3D	1	SDNR024				
4	L3f	15944D, 16209C, 16223T, 16309G, 73G, 263G, 309.1C, 315.1C	1	KEOG025			1	1
5	L3f	15944D, 15946T, 16209C, 16213A, 16223T, 16311C, 16519C, 73G, 195C, 263G, 315.1C, 328G	1	SDHD016				
6	L3f	15944D, 16126C, 16172C, 16209C, 16223T, 16519C, 73G, 228A, 263G, 315.1C	1	KEOR010				
7	L3f	15944D, 16209C, 16223T, 16311C, 16519C, 73G, 189G, 263G, 315.1C	1	TZDT074				
8	L3f	15944D, 16209C, 16218T, 16223T, 16292T, 16311C, 73G, 189G, 263G, 309.1C, 315.1C	2	KERD077, Keil026				
9	L3f1a	15944D, 16209C, 16213A, 16223T, 16311C, 16519C, 73G, 195C, 263G, 315.1C, 328G	13	SDHD001, SDHD007, SDHD008, SDHD010, SDHD020, SDHD021, SDHD022, SDHD024, SDHD025, SDHD026, SDHD028, SDHD037, SDHD039		1		
10	L3f1a	15944D, 16209C, 16223T, 16234T, 16311C, 16519C, 73G, 263G, 315.1C, 522-3D	1	KELY039		1		
11	L3f1a	15944D, 16140C, 16209C, 16223T, 16519C, 73G, 152C, 263G, 309.1C, 315.1C	1	TZAK010				
12	L3f1a	15944D, 16140C, 16209C, 16223T, 16519C, 73G, 152C, 263G, 315.1C	1	SDNR013				
13	L3f1a	15944D, 16209C, 16223T, 16224C, 16519C, 73G, 152C, 200G, 263G, 309.1C, 315.1C, 491T, 522-3D	1	TZBG053			1	1
14	L3f1a	15944D, 16209C, 16223T, 16224C, 16519C, 73G, 152C, 200G, 263G, 315.1C, 491T, 522-3D	1	KENJ008			1	1
15	L3f1b	15944D, 16189C, 16209C, 16223T, 16311C, 16519C, 73G, 189G, 200G, 263G, 309.1C, 315.1C	1	KEYK010				
16	L3f1b	15944D, 16209C, 16223T, 16309G, 16519C, 73G, 200G, 263G, 309.1C, 315.1C	1	TZMS056				
17	L3f1b	15944D, 16209C, 16223T, 16311C, 16519C, 73G, 189G, 200G, 263G, 309.1C, 315.1C	1	KEWT003				
18	L3f1b	15884A, 15944D, 16209C, 16223T, 16292T, 16311C, 16519C, 73G, 200G, 263G, 309.1C, 315.1C, 522-3D	1	CAGZ015				
19	L3f1b	15944D, 16129A, 16209C, 16223T, 16292T, 16295T, 16311C, 16519C, 73G, 189G, 200G, 263G, 309.1C,	1	KELO065				
20	L3f1b	15944D, 16192T, 16209C, 16223T, 16292T, 16311C, 16519C, 73G, 189G, 195C, 200G, 263G, 315.1C	1	KEBR027				
21	L3f1b	15944D, 16192T, 16209C, 16223T, 16292T, 16311C, 16519C, 73G, 189G, 200G, 263G, 315.1C	1	KEPK047				
22	L3f1b	15944D, 16209C, 16223T, 16292T, 16311C, 16465T, 16519C, 73G, 189G, 200G, 263G, 315.1C	1	CHLA053				
23	L3f1b	15944D, 16209C, 16223T, 16292T, 16311C, 16519C, 73G, 185A, 189G, 200G, 263G, 315.1C, 522-3D	1	SDSH011				
24	L3f1b	15944D, 16209C, 16223T, 16292T, 16311C, 16519C, 73G, 189G, 193G, 200G, 263G, 309.1C, 315.1C, 335	1	KEKS004				
25	L3f1b	15944D, 16209C, 16223T, 16292T, 16311C, 16519C, 73G, 189G, 195C, 200G, 263G, 315.1C, 573.1C, 574C, 577C, 577.1A, 577.2G	1	SDNY009				
26	L3f1b	15944D, 16209C, 16223T, 16292T, 16311C, 16519C, 73G, 189G, 200G, 263G, 309.1C, 315.1C	1	CHBU012				
27	L3f1b	15944D, 16209C, 16223T, 16292T, 16311C, 16519C, 73G, 189G, 200G, 263G, 315.1C, 522-3D	3	CHBU004, CHBU020, CHBU021				
28	L3f1b	15944D, 16209C, 16223T, 16292T, 16311C, 16519C, 73G, 152C, 189G, 200G, 263G, 315.1C	1	KEBR030				
29	L3f1b	15944D, 16209C, 16223T, 16292T, 16311C, 16519C, 73G, 152C, 189G, 195C, 263G, 315.1C, 522-3D	1	KEBN021				
30	L3f1b	15944D, 16209C, 16223T, 16292T, 16311C, 16519C, 73G, 152C, 189G, 263G, 315.1C, 524.1A, 525.1C	1	KEGU029				
31	L3f1b3/4	15944D, 16093C, 16209C, 16223T, 16235G, 16311C, 73G, 150T, 263G, 309.1C, 315.1C, 522-3D	1	KEGK006				
32	L3f1b3/4	15944D, 16093C, 16209C, 16223T, 16235G, 16292T, 16311C, 16519C, 73G, 150T, 189G, 200G, 207A, 263G,	1	CAKN009				
33	L3f1b4	15944D, 16209C, 16223T, 16311C, 16519C, 73G, 150T, 189G, 200G, 263G, 309.1C, 315.1C, 522-3D	1	CAMV012	1			
34	L3f1b4	15930A, 15944D, 16075C, 16209C, 16223T, 16292T, 16311C, 16519C, 73G, 150T, 189G, 200G, 263G,	1	KEGK012	1			

35	L3f2a	15944D, 16145A, 16209C, 16223T, 16266T, 16294T, 16311C, 16519C, 73G, 146C, 200G, 263G, 315.1C,	2	CAAS029, CAKN036				
36	L3f2b	15944D, 16209C, 16223T, 16224C, 16311C, 16519C, 73G, 152C, 200G, 263G, 309.1C, 315.1C, 491T, 522-3D	2	TZIQ039, TZIQ042j (TZIQ042K)			1	
37	L3f2b	15944D, 16129A, 16209C, 16223T, 16295T, 16311C, 16519C, 73G, 152C, 189G, 200G, 263G, 309.1C,	1	TZPR039				
38	L3f2b	15944D, 16209C, 16223T, 16311C, 73G, 152C, 189G, 262T, 263G, 315.1C	3	TZIQ028, TZRG028, KETG055	1			
39	L3f2b	15874G, 15944D, 16111A, 16209C, 16223T, 16286T, 16311C, 73G, 152C, 263G, 309.1C, 315.1C	1	KEGU034				
40	L3f2b	15944D, 16093G, 16209C, 16223T, 16311C, 73G, 152C, 262T, 263G, 315.1C	2	KEWT009, KEWT030				
41	L3f2b	15944D, 16111A, 16209C, 16223T, 16286T, 16311C, 64T, 73G, 152C, 263G, 309.1C, 309.2C, 315.1C	1	KEGU059				
42	L3f2b	15944D, 16111A, 16209C, 16223T, 16286T, 16311C, 73G, 152C, 263G, 309.1C, 309.2C, 315.1C	1	KEGU045				
43	L3f2b	15944D, 16104T, 16209C, 16223T, 16266T, 16311C, 16519C, 73G, 152C, 210G, 263G, 309.1C, 309.2C,	1	KEKS018				
44	L3f2b	15944D, 16129A, 16176T, 16209C, 16223T, 16263C, 16266T, 16311C, 16519C, 73G, 152C, 200G, 263G, 309.1C, 315.1C, 522-3D	2	KEWT018, KEWT025				
45	L3f2b	15944D, 16209C, 16223T, 16266T, 16311C, 16519C, 73G, 152C, 195C, 263G, 315.1C, 466C, 522-3D	1	KEOR029				
46	L3f2b	15944D, 16209C, 16223T, 16266T, 16311C, 16519C, 73G, 152C, 200G, 263G, 309.1C, 315.1C	1	TZTR035				
47	L3f3	15940C, 15944D, 16176T, 16183C, 16189C, 16209C, 16223T, 73G, 195C, 263G, 315.1C, 318C	1	SDDN014			1	1
48	L3f3	15940C, 15944D, 16176T, 16188.1C, 16189C, 16209C, 16223T, 73G, 189G, 195C, 217C, 263G, 309.1C,	1	KEGB044			1	1
49	L3f3	15940C, 15944D, 16176T, 16189C, 16209C, 16223T, 16287T, 73G, 189G, 195C, 217C, 263G, 315.1C, 318C	2	KENJ011, KENJ052			1	1
50	L3f3	15941C, 15944D, 16176T, 16188T, 16209C, 16223T, 16234T, 16311C, 16355T, 73G, 143A, 189G, 214G,	1	CHLA007				
51	L3f3	15941C, 15944D, 16176T, 16188T, 16209C, 16223T, 16234T, 16362C, 73G, 189G, 215G, 263G, 315.1C,	1	CAKN017			1	1
52	L3f3	15941C, 15944D, 16176T, 16188T, 16209C, 16223T, 16234T, 73G, 189G, 263G, 279C, 309.1C, 315.1C,	1	CAGZ034			1	1
53	L3f3	15941C, 15944D, 16176T, 16209C, 16223T, 16234T, 73G, 146C, 189G, 195C, 263G, 309.1C, 315.1C, 318C,	1	CAKN023				
54	L3f3	15941C, 15944D, 16176T, 16209C, 16223T, 16234T, 73G, 185A, 189G, 263G, 315.1C, 318C, 522-3D	2	CHBU008, CHBU013				
55	L3f3	15941C, 15944D, 16176T, 16209C, 16223T, 16234T, 73G, 189G, 263G, 309.1C, 315.1C, 318C, 522-3D	1	CAGZ027			1	1
56	L3f3	15941C, 15944D, 16176T, 16209C, 16223T, 16234T, 73G, 189G, 263G, 315.1C, 318C, 522-3D	1	CAGZ009			1	1
57	L3f3	15941C, 15944D, 16176T, 16188T, 16209C, 16223T, 16234T, 16519C, 73G, 189G, 263G, 309.1C, 315.1C,	1	CAKN027				
58	L3f3	15941C, 15944D, 16176T, 16209C, 16223T, 16234T, 73G, 152C, 189G, 263G, 315.1C, 318C, 522-3D	1	CAGZ022			1	1
59	L3f3	15941C, 15944D, 16176T, 16188T, 16209C, 16223T, 16234T, 16311C, 16355T, 73G, 143A, 152C, 189G,	1	CHLA052				

					L3h	L3h1	L3h1	L3h1	L3h1				L3h1b	L3h1a2a
L3h	Hg	Controlregion	Abs. Freq	Samples	1	2	3	4	5	6	7	8	9	10	11
1	L3h1	16189C, 16223T, 16311C, 16359C, 73G, 93G, 95C, 146C, 246C, 263G, 315.1C	1	tzhz041	1	1	1	1	1						
2	L3h1a1a	16192T, 16195C, 16245T, 16298C, 16311C, 16390A, 16519C, 73G, 309.1C, 315.1C, 518T	1	KETG007	1	1	1	1	1						
3	L3h1a1a	16172C, 16195C, 16298C, 16311C, 16390A, 73G, 146C, 152C, 315.1C	1	KELO003	1	1	1	1	1						
4	L3h1a1a	16298C, 16311C, 73G, 263G, 309.1C, 315.1C	2	TZDT076, TZDT080	1	1	1	1	1						
5	L3h1a1a	16189C, 16192T, 16298C, 16311C, 73G, 263G, 309.1C, 315.1C	1	KETG075	1	1	1	1	1						
6	L3h1a1a	16192T, 16298C, 16311C, 73G, 263G, 309.1C, 315.1C	2	KEBJ030, KEEL008	1	1	1	1	1						
7	L3h1a1a	16192T, 16298C, 16311C, 73G, 263G, 309.1C, 315.1C, 368G, 573.1C	2	KETK021, KETK017	1	1	1	1	1						
8	L3h1a2a	16111T, 16165G, 16192T, 16223T, 16311C, 16399G, 73G, 146C, 263G, 315.1C	1	SDDN021	1	1		1	1						
9	L3h1a2a	16148T, 16154C, 16192T, 16223T, 16311C, 16399G, 73G, 146C, 263G, 315.1C	1	SDDN009	1	1	1	1	1						
10	L3h1a2a	16148T, 16172C, 16192T, 16214T, 16223T, 16292T, 16311C, 16399G, 73G, 93G, 146C, 152C, 263G, 315.1C	3	TZDR008, TZDR010A, TZDT073	1	1	1	1	1						
11	L3h1a2a	16148T, 16189C, 16192T, 16223T, 16311C, 16399G, 16519C, 73G, 146C, 152C, 263G, 309.1C, 315.1C	1	SDSH017	1	1	1	1	1						
12	L3h1a2a	16148T, 16192T, 16223T, 16311C, 16399G, 16519C, 73G, 146C, 263G, 309.1C, 315.1C	1	KEGB026	1	1	1	1	1						
13	L3h1a2a	16148T, 16192T, 16223T, 16311C, 16399G, 73G, 146C, 150T, 152C, 207A, 263G, 315.1C, 489C	4	KEND008, KEPK006, KETG071, KETK090	1	1	1	1	1						
14	L3h1a2a	16148T, 16192T, 16223T, 16311C, 16399G, 73G, 146C, 152C, 263G, 309.1C, 315.1C	3	KEPK025, KEPK030, KESN039	1	1	1	1	1						
15	L3h1a2a	16148T, 16192T, 16223T, 16311C, 16399G, 73G, 95C, 146C, 152C, 263G, 315.1C	1	KEKS022		1	1								
16	L3h1a2a	16148T, 16223T, 16242T, 16311C, 16399G, 73G, 146C, 150T, 152C, 263G, 315.1C	1	SDDN013	1	1	1	1	1						
17	L3h1a2a	16148T, 16223T, 16311C, 16399G, 73G, 146C, 150T, 152C, 263G, 315.1C	1	SDNR012	1	1	1	1	1						
18	L3h1a2a	16148T, 16223T, 16311C, 16399G, 73G, 146C, 152C, 263G, 309.1C, 315.1C	1	KELY036	1	1		1	1						
19	L3h1a2a	16165G, 16172C, 16189C, 16192.1T, 16223T, 16311C, 16399G, 73G, 146C, 263G, 315.1C	2	SDNR022, SDNR023	1	1	1	1	1						
20	L3h1a2a	16165G, 16192T, 16223T, 16311C, 16399G, 73G, 146C, 263G, 315.1C	1	KESB025	1	1	1	1	1						
21	L3h1a2a	16169T, 16223T, 16311C, 16354T, 16399G, 73G, 146C, 263G, 309.1C, 315.1C	2	KEBJ032, KEBR032	1	1	1	1	1						
22	L3h1a2a	16178C, 16223T, 16311C, 16399G, 73G, 146C, 153G, 263G, 309.1C, 315.1C, 499A	1	SDDN034	1	1	1	1	1						
23	L3h1a2a	16189C, 16223T, 16311C, 16362C, 16399G, 73G, 146C, 152C, 263G, 309.1C, 315.1C, 522-3D	1	SDDN007	1	1		1	1						
24	L3h1a2a	16192T, 16223T, 16311C, 16399G, 73G, 146C, 263G, 315.1C	2	KESN019, KESN042	1	1	1	1	1						
25	L3h1a2a	16192T, 16223T, 16311C, 16399G, 73G, 146C, 263G, 315.1C, 522-3D	1	SDNR010	1	1	1	1	1						
26	L3h1a2a	16218T, 16223T, 16311C, 16354T, 16399G, 73G, 146C, 153G, 263G, 315.1C	1	KEGK023	1	1	1	1	1						
27	L3h1a2a	16223T, 16311C, 16354T, 16362C, 16399G, 73G, 146C, 153G, 263G, 309.1C, 315.1C	1	TZWF020	1	1	1	1	1						
28	L3h1a2a	16223T, 16311C, 16354T, 16399G, 73G, 146C, 153G, 263G, 309.1C, 315.1C	1	TZIQ065	1	1	1	1	1						
29	L3h1a2a	16223T, 16311C, 16354T, 16399G, 73G, 146C, 153G, 263G, 315.1C	5	KEGK019, KEGK027, KEND011, TZBG032, TZIQ030	1	1	1	1	1						
30	L3h1a2a	16223T, 16311C, 16354T, 16399G, 73G, 146C, 263G, 309.1C, 315.1C	2	Keil004, TZBG049	1	1	1	1	1						
31	L3h1a2a	16223T, 16311C, 16354T, 16399G, 73G, 146C, 263G, 309.1C, 315.1C, 554T, 561T, 562T	1	TZTR027	1	1	1	1	1						
32	L3h1a2a	16223T, 16311C, 16356C, 16399G, 73G, 146C, 153G, 263G, 309.1C, 315.1C	1	KERD028	1	1	1	1	1						
33	L3h1a2a	16223T, 16311C, 16399G, 73G, 146C, 153G, 263G, 271T, 315.1C	1	KETG003	1	1	1	1	1						
34	L3h1a2a	16223T, 16311C, 16399G, 73G, 146C, 153G, 263G, 315.1C, 573.1	1	KEGK015	1	1	1	1	1						
35	L3h1a2a	16223T, 16311C, 16399G, 73G, 146C, 263G, 291G, 315.1C	2	KEPK067, KETG027	1	1	1	1	1						
36	L3h1a2a	16124C, 16148T, 16192T, 16223T, 16311C, 16399G, 73G, 146C, 150T, 152C, 207A, 263G, 315.1C, 489C	1	KESB006	1		1	1	1						

37	L3h1a2a	16165G, 16172C, 16189C, 16192.1T, 16223T, 16311C, 16399G, 73G, 146C, 263G, 315.1C	2	SDNR022, SDNR023	1	1	1	1	1						
38	L3h1a2a	16148T, 16223T, 16311C, 16399G, 73G, 146C, 152C, 263G, 315.1C	1	KEGU060, KERD079, TZDR006						1					1
39	L3h1a2a	16165G, 16172C, 16192T, 16223T, 16311C, 16399G, 73G, 146C, 263G, 309.1C, 309.2C, 315.1C	1	SDDN024	1			1	1			1			1
40	L3h1a2a	16148T, 16223T, 16265G, 16292T, 16311C, 16399G, 73G, 146C, 152C, 263G, 315.1C	1	SDNR016	1	1	1	1	1			1			1
41	L3h1a2a	16223T, 16311C, 16354T, 16399G, 73G, 146C, 263G, 315.1C	1	KEEL022											
42	L3h1b	16129A, 16174T, 16192T, 16218T, 16223T, 16256A, 16311C, 16317T, 73G, 151T, 152C, 189C, 195C, 263G, 294C, 315.1C, 420A, 522-3D, 606G	2	CAGZ019, CAGZ020	1	1	1	1	1						
43	L3h1b1	16179T, 16192T, 16215G, 16223T, 16256A, 16284G, 16311C, 73G, 189C, 195C, 263G, 309.1C, 315.1C, 522-3D	2	TZMS005, KEGK030	1	0	1	1	1						
44	L3h1b1	16179T, 16192T, 16223T, 16256A, 16284G, 16311C, 73G, 189C, 195C, 263G, 315.1C, 522-3D	2	TZDT015, TZDT016	1	0	1	1	1						
45	L3h1b1	16179T, 16223T, 16256A, 16284G, 16311C, 73G, 189C, 195C, 263G, 309.1C, 315.1C, 522-3D	1	TZDR014	1	1	1	1	1	1					
46	L3h1b1	16179T, 16223T, 16256A, 16289G, 16311C, 16354T, 73G, 150T, 152C, 189C, 195C, 263G, 309.1C, 309.2C, 315.1C, 522-3D	1	KEMR012	1	1	1	1	1		1				
47	L3h1b1	16179T, 16223T, 16256A, 16289G, 16311C, 16354T, 73G, 150T, 152C, 189C, 195C, 263G, 309.1C, 315.1C, 522-3D	3	KEND013, KESN052, TZMS050	1	1	1	1	1		1				
48	L3h1b1	16179T, 16223T, 16256A, 16289G, 16311C, 16354T, 73G, 150T, 152C, 189C, 195C, 263G, 315.1C, 522-3D	1	TZAK022	1	1	1	1	1		1				
49	L3h1b1	15924G, 16048A, 16153A, 16179T, 16192T, 16215G, 16223T, 16256A, 16284G, 16311C, 73G, 189C, 195C, 263G, 309.1C, 315.1C, 522-3D	1	TZDT054	1			1	1				1	1	
50	L3h2	15910T, 16111T, 16184T, 16223T, 16304C, 16311C, 16519C, 73G, 150T, 263G, 315.1C, 318C	1	SDDN001	1		0	0	0						
51	L3h2	16111T, 16129A, 16184T, 16223T, 16304C, 16519C, 73G, 150T, 195C, 263G, 315.1C, 318C, 522-3D	3	KEBR033, KEGB054, KERD080	1		0	0	0						

SNPs genotyped of L0 haplogroup			
	Snp	**Haplotypes defined**	
	1	G9575A	L3h
	2	G1719A	L3h1
	3	A4388G	L3h1
	4	T9509C	L3h1
	5	A11590G	L3h1
	6	C3594T	
	7	T3777C	
	8	C8781A, C8943T	
	9	A8767G	
	10	A10044G, G11963A	L3h1b
	11	T12175C, G12236A, T12519C	L3h1a2a

					L3eikx	L3i1
L3i	Hg	Controlregion	Abs. Freq	Samples	A10819G	G13800A
1	L3i1b	16129A, 16153A, 16223T, 73G, 150T, 263G, 309.1C, 309.2C, 315.1C	1	KENJ063	1	1
2	L3i1b	16153A, 16223T, 16292T, 73G, 150T, 263G, 315.1C	1	CHBU016	1	
3	L3i1c	15924G, 16066G, 16153A, 16172C, 16174T, 16223T, 16319A, 16519C, 73G, 150T, 189G, 263G, 315.1C	1	KETK005	1	1
4	L3i1c	16066G, 16153A, 16174T, 16223T, 16301T, 16319A, 16519C, 73G, 150T, 263G, 315.1C	1	TZDR005	1	
5	L3i1c	16153A, 16172C, 16174T, 16179T, 16189C, 16223T, 16319A, 16519C, 73G, 150T, 263G, 309.1C, 315.1C, 522-3D	1	KESM035	1	1
6	L3i1c	16153A, 16172C, 16174T, 16223T, 16319A, 16519C, 73G, 150T, 152C, 257G, 263G, 309.1C, 309.2C, 315.1C	1	KEOG053	1	1
7	L3i1c	16153A, 16172C, 16174T, 16223T, 16319A, 16519C, 73G, 150T, 263G, 315.1C, 316A	3	KEPK003, KEPK045, KEPK086	1	1
8	L3i1c	16153A, 16172C, 16174T, 16223T, 16319A, 16519C, 73G, 150T, 263G, 315.1C, 574G	7	KEMR024, KEMR025, KEND014, KEPK059, KESB023, KESN018, KETG006	1	1
9	L3i1c	16153A, 16172C, 16174T, 16223T, 16319A, 16519C, 73G, 150T, 189G, 263G, 315.1C, 574G	2	KEBR093, KEMN017	1	1
10	L3i1c	16153A, 16174T, 16223T, 16274A, 16319A, 16519C, 73G, 150T, 263G, 315.1C	1	KEGK029		
11	L3i1c	15924G, 16066G, 16153A, 16174T, 16223T, 16319A, 16519C, 73G, 150T, 263G, 315.1C	1	Keil002	1	1
12	L3i2a	15924G, 16185T, 16223T, 16260T, 16311C, 16519C, 73G, 150T, 152C, 200G, 263G, 315.1C, 462T, 522-3D	1	KEBR042	1	
13	L3i2a	16093C, 16168T, 16223T, 16260T, 16311C, 16519C, 73G, 150T, 152C, 189G, 200G, 207A, 263G, 315.1C, 634C	1	KEBR028	1	
14	L3i2a	16093C, 16223T, 16260T, 16278T, 16311C, 16519C, 73G, 150T, 152C, 189G, 263G, 291.1A, 315.1C	1	KEPK040	1	
15	L3i2a	16093C, 16223T, 16260T, 16278T, 16311C, 16519C, 73G, 150T, 152C, 189G, 263G, 315.1C	2	KELY045, KELY046	1	
16	L3i2a	16111T, 16184T, 16223T, 16260T, 16311C, 16519C, 73G, 150T, 152C, 189G, 199C, 263G, 309.1C, 315.1C, 522-3D	1	KEBR007	1	
17	L3i2a	16184T, 16223T, 16260T, 16311C, 16519C, 73G, 150T, 152C, 189G, 199C, 263G, 309.1C, 315.1C, 522-3D	1	KEBR005	1	
18	L3i2a	16184T, 16223T, 16260T, 16311C, 16519C, 73G, 150T, 152C, 189G, 263G, 315.1C, 324T, 522-3D	1	SDBA025	1	
19	L3i2a	16185T, 16223T, 16260T, 16311C, 16519C, 73G, 150T, 152C, 200G, 263G, 315.1C, 522-3D	4	KEBR049, Kesm020, Kesm022, kesm007	1	
20	L3i2a	16223T, 16260T, 16278T, 16311C, 16519C, 73G, 150T, 152C, 189G, 263G, 315.1C, 522-3D	1	KEGB010	1	
21	L3i2a	16223T, 16260T, 16311C, 16519C, 73G, 150T, 152C, 189G, 200G, 204C, 207A, 263G, 309.1C, 315.1C	1	KEWT019	1	
22	L3i2a	16223T, 16260T, 16311C, 16519C, 73G, 150T, 152C, 189G, 200G, 204C, 207A, 263G, 315.1C	1	SDBA031	1	
23	L3i2	16185T, 16223T, 16311C, 16519C, 73G, 150T, 152C, 263G, 309.1C, 315.1C, 522-3D	4	KEBR035, KEBR063, KEBR078, KENJ044		

					L3j	
L3j	Hg	Controlregion	Abs. Freq	Sample	G3516A,C3594T	
1	L3j	16192T,16218T,16223T,16303A,16360T,73G,146C,150T,152C,199C,263G,309.1C,315.1C,522-3D,596C	1	TZWF028	1	
2	L3j	16192T,16218T,16223T,16303A,16360T,73G,146C,149.1T,152:,263G,309.1C,315.1C,522-3D,596C	1	TZIQ032	1	
3	L3j	16192T,16218T,16223T,16303A,16360T,73G,146C,150T,152C,263G,309.1C,315.1C,522-3D,596C	1	KEGK007		
4	L3j	16192T,16218T,16223T,16303A,16360T,73G,146C,150T,152C,263G,315.1C,522-3D,596C	1	Keil006		
5	L3j	16192T,16209C,16218T,16223T,16303A,16360T,73G,146C,150T,152C,263G,309.1C,315.1C,522-3D,596C	1	KEBN017		

					L3eikx	L3k
L3k	Hg	Controlregion	Abs. Freq	Samples	A10819G	C13992T
1	L3k	15924G,16129A,16223T,16290T,16344T,16519C,73G,150T,152C,235G,263G,309.1C,315.1C,494T	1	CAFU027	1	1

					L3eikx	L3h/L3x	L3x
L3x	Hg	Controlregion	Abs. Freq	Samples	A10819G	G3483A	A13708A
1	L3x1	16169T, 16182.1C, 16183C, 16189C, 16223T, 16278T, 16298C, 16311C, 16519C, 73G, 150T, 199C, 263G, 309.1C, 309.2C, 315.1C	1	KEBJ008	1	1	
2	L3x1	16169T, 16223T, 16256T, 16278T, 16305G, 16311C, 16320T, 16519C, 73G, 150T, 152C, 200G, 263G, 315.1C	1	KEBR079	1	1	1
3	L3x1a	15924G, 16169T, 16171G, 16189C, 16223T, 16278T, 16292T, 16311C, 16519C, 73G, 150T, 204C, 263G, 309.1C, 315.1C	1	SDHD029	1		
4	L3x1a	16066G, 16169T, 16223T, 16278T, 16311C, 73G, 150T, 204C, 207A, 263G, 315.1C	2	CAGZ014, KEKS006	1	1	1
5	L3x1a	16066G, 16169T, 16223T, 16278T, 16311C, 73G, 150T, 204C, 207A, 263G, 315.1C	4	KEBJ036B, KEBJ034, KEBJ040, KEBR098	1	1	1
6	L3x1a	16169T, 16189C, 16223T, 16278T, 16293G, 16298C, 16311C, 16519C, 73G, 150T, 199C, 204C, 263G, 309.1C, 315.1C	1	KENJ004	1	1	
7	L3x1a	16169T, 16223T, 16256T, 16278T, 16311C, 16320T, 16519C, 73G, 150T, 204C, 207A, 263G, 294C, 309.1C, 315.1C	2	KEGB036, KEGB003	1	1	1
8	L3x1b	16066G, 16136C, 16169T, 16223T, 16278T, 16368C, 16422C, 16519C, 73G, 150T, 263G, 309.1C, 315.1C, 522-3D	1	KEBR070	1		
9	L3x1b	16066G, 16136C, 16169T, 16223T, 16278T, 16368C, 16519C, 73G, 150T, 263G, 309.1C, 315.1C, 522-3D	1	KERD086	1	1	1
10	L3x1b	16066G, 16136C, 16169T, 16223T, 16278T, 16519C, 73G, 150T, 263G, 315.1C, 522-3D	2	KETK001, KETK011	1	1	1
11	L3x1b	16169T, 16223T, 16519C, 73G, 150T, 204C, 207A, 263G, 315.1C, 522-3D	1	TZBG029	1	1	1
12	L3x1b	16169T, 16172C, 16223T, 16243C, 16278T, 16311C, 16519C, 73G, 150T, 152C, 189G, 199C, 204C, 228A, 263G, 315.1C, 522-3D	2	TZMS034, TZMS035	1	1	1
13	L3x1b	16169T, 16192T, 16223T, 16278T, 16519C, 73G, 150T, 152C, 189G, 199C, 204C, 207A, 263G, 309.1C, 315.1C, 522-3D	2	KETK043, KEBR014	1	1	1
14	L3x1b	16169T, 16203G, 16223T, 16278T, 16519C, 73G, 150T, 204C, 207A, 263G, 315.1C, 522-3D	1	KETV030	1	1	1
15	L3x1b	16169T, 16207G, 16223T, 16278T, 16519C, 73G, 150T, 204C, 207A, 263G, 315.1C, 522-3D	1	KEOR001	1	1	1
16	L3x1b	16169T, 16223T, 16278T, 16362C, 16519C, 73G, 150T, 204C, 263G, 315.1C, 522-3D	1	KEOR020B	1	1	1
17	L3x1b	16169T, 16223T, 16278T, 16519C, 73G, 150T, 204C, 207A, 243G, 263G, 315.1C, 522-3D	1	Kemn002	1	1	1
18	L3x1b	16169T, 16223T, 16278T, 16519C, 73G, 150T, 204C, 207A, 263G, 315.1C, 522-3D	4	KELY029, KETA044, KETA071, TZAK029	1	1	1
19	L3x1b	16169T, 16171G, 16223T, 16278T, 16519C, 73G, 150T, 204C, 207A, 263G, 315.1C, 522-3D	1	tzsw135			
20	L3x1b	16169T, 16223T, 16278T, 16362C, 16519C, 73G, 150T, 263G, 309.1C, 315.1C, 522-3D	2	KEBN033, KEBN020	1	1	1
21	L3x2a	15928A, 16169T, 16189C, 16193T, 16195C, 16223T, 16235G, 16243C, 16399G, 73G, 150T, 189G, 249D, 263G, 309.1C, 315.1C, 494T, 650C	1	CHLA059	1	1	1
22	L3x2a	16166G, 16169T, 16173T, 16193T, 16195C, 16264T, 73G, 150T, 249D, 263G, 309.1C, 315.1C, 494T	1	KEBR040	1		
23	L3x2a	16169T, 16173T, 16193T, 16195C, 16264T, 73G, 150T, 249D, 263G, 309.1C, 315.1C, 494T	1	KEBJ001			
24	L3x2a	16169T, 16193T, 16195C, 73G, 150T, 200G, 249D, 263G, 309.1C, 309.2C, 315.1C, 494T	1	KEKS011			
25	L3x2a	16169T, 16193T, 16195C, 73G, 150T, 200G, 249D, 263G, 309.1C, 315.1C, 494T	3	KEKS003, KEKS007, KEKS020			
26	L3x2a	16189C, 16193T, 16195C, 16362C, 73G, 150T, 249D, 263G, 315.1C, 494T	1	KEBJ014			

					L4a	L4	L4a	L4b2
L4	Hg	Controlregion	Abs. Freq	Samples	G3357A	A13470G	T11485C	A9855G
1	L4a1	15968C, 16093C, 16169T, 16207C, 16223T, 16260T, 16311C, 16362C, 16519C, 73G, 195C, 198T, 263G, 309.1C, 315.1C, 325T	1	TZRG002	1		1	
2	L4a1	15968C, 16169T, 16207C, 16223T, 16260T, 16311C, 16362C, 16519C, 73G, 195C, 198T, 263G, 309.1C, 315.1C, 325T	1	TZMS085				
3	L4a1	16093C, 16104T, 16207C, 16223T, 16260T, 16311C, 16519C, 73G, 152C, 195C, 198T, 263G, 309.1C, 315.1C, 325T	1	TZAK013	1		1	
4	L4a1	16104T, 16207C, 16223T, 16260T, 16311C, 16519C, 73G, 152C, 195C, 198T, 263G, 309.1C, 315.1C, 325T	3	TZAK001, TZAK008, TZAK011	1		1	
5	L4a1	16207C, 16223T, 16260T, 16274A, 16295T, 16311C, 16362C, 16519C, 73G, 152C, 195C, 198T, 263G, 309.1C, 315.1C, 325T	1	KEYK027	1			
6	L4a1a	16093C, 16169T, 16207T, 16223T, 16232T, 16260T, 16311C, 16519C, 73G, 189G, 195C, 198T, 263G, 309.1C, 315.1C, 325T	2	KEGU003, KEGU037	1		1	
7	L4a1a	16207T, 16220G, 16223T, 16260T, 16261T, 16311C, 16320T, 16362C, 16519C, 73G, 195C, 198T, 263G, 309.1C, 315.1C, 325T	2	KEKS008, KEKS009	1		1	
8	L4a1a	16207T, 16223T, 16260T, 16261T, 16311C, 16362C, 16519C, 73G, 150T, 189G, 195C, 198T, 263G, 315.1C, 325T	1	KEBJ019	1			
9	L4a1b	16093C, 16182C, 16183C, 16189C, 16223T, 16260T, 16264T, 16311C, 16362C, 16519C, 73G, 195C, 198T, 263G, 309.1C, 315.1C	1	KETV035	1			
10	L4a1b	16179T, 16182C, 16183C, 16189C, 16223T, 16260T, 16264T, 16311C, 16362C, 16519C, 73G, 195C, 198T, 263G, 303.1T, 315.1C, 524.1A, 525.1C	1	KEOR012			1	
11	L4a1b	16093C, 16207C, 16223T, 16260T, 16264T, 16311C, 16320T, 16362C, 16474A, 16519C, 73G, 151T, 195C, 198T, 263G, 309.1C, 309.2C, 315.1C, 325T	1	KEBJ002	1			
12	L4a1b	16093C, 16207C, 16223T, 16260T, 16264T, 16311C, 16320T, 16362C, 16519C, 73G, 195C, 198T, 263G, 309.1C, 315.1C, 325T	2	KEKS016, KEKS017	1		1	
13	L4a2	15946T, 16223T, 16311C, 16362C, 73G, 150T, 195C, 263G, 315.1C, 522-3D	1	KETK061				
14	L4a2	16183C, 16189C, 16223T, 16311C, 16362C, 73G, 195C, 263G, 315.1C, 522-3D, 575T	1	SDNR025				
15	L4a2	16093C, 16223T, 16241C, 16311C, 16362C, 16519C, 73G, 150T, 263G, 315.1C, 522-3D	1	KEPK097				
16	L4a2	16223T, 16311C, 16362C, 16519C, 73G, 150T, 263G, 315.1C, 522-3D	2	KEYK014, KEYK033				
17	L4a2	16223T, 16311C, 16362C, 73G, 150T, 204C, 263G, 309.1C, 315.1C, 356.1	1	SDSH006				
18	L4a2	16223T, 16311C, 16362C, 73G, 150T, 263G, 309.1C, 309.2C, 315.1C	1	SDNY004				
19	L4a2	16223T, 16311C, 16362C, 73G, 150T, 263G, 315.1C, 522-3D	1	SDNR011				
20	L4a2	16223T, 16241C, 16311C, 16362C, 16519C, 73G, 150T, 263G, 315.1C, 522-3D	6	KEKP003, KEMR011, KEND006, KENJ032, KENJ033, KEPK012				
21	L4b1	16179T, 16183C, 16189C, 16223T, 16239T, 16311C, 16320T, 16362C, 16519C, 73G, 150T, 199C, 204C, 263G, 309.1C, 315.1C, 522-3D	1	SDDN028				
22	L4b1	15944D, 16124C, 16223T, 16278T, 16311C, 16362C, 16519C, 73G, 152C, 263G, 315.1C, 522-3D	1	CABM048		1		1
23	L4b2	16222T, 16223T, 16248T, 16293T, 16311C, 16362C, 16399G, 73G, 146C, 199C, 244G, 263G, 315.1C, 522-3D	1	CABM049				
24	L4b2	16223T, 16287G, 16293T, 16311C, 16355T, 16362C, 16399G, 73G, 146C, 152C, 244G, 249.1A, 263G, 297G, 315.1C, 522-3D	1	SDSH019				
25	L4b2a1	15924G, 16111T, 16223T, 16293T, 16311C, 16355T, 16362C, 16399G, 73G, 146C, 195C, 244G, 263G, 315.1C, 398C, 522-3D	1	KEMR021				
26	L4b2a1	16093A, 16095T, 16223T, 16287A, 16293T, 16301T, 16311C, 16355T, 16362C, 16399G, 73G, 146C, 152C, 195C, 244G, 263G, 315.1C, 522-3D	1	KENJ003				
27	L4b2a1	16093C, 16223T, 16293T, 16311C, 16355T, 16362C, 16399G, 73G, 146C, 195C, 244G, 309.1C, 315.1C, 368G	4	KEPR015, KEPR020, KEPR022, KETA028				
28	L4b2a1	16093G, 16223T, 16287A, 16292.1C, 16293C, 16311C, 16355T, 16362C, 16399G, 73G, 146C, 152C, 195C, 200G, 244G, 263G, 309.1C, 315.1C, 522-3D	1	KEBR038				1

29	L4b2a1	16093G, 16223T, 16287A, 16293T, 16311C, 16355T, 16362C, 16399G, 73G, 146C, 152C, 195C, 244G, 263G, 309.1C, 315.1C, 522-3D	1	TZDR018					
30	L4b2a1	16093G, 16223T, 16287A, 16293T, 16311C, 16355T, 16362C, 16399G, 73G, 146C, 152C, 195C, 244G, 263G, 315.1C, 522-3D	1	TZWF016					
31	L4b2a1	16223T, 16274A, 16293T, 16311C, 16355T, 16362C, 16399G, 16519C, 73G, 146C, 195C, 244G, 263G, 309.1C, 315.1C	1	TZAK016A(TZAK016B)		1			
32	L4b2a1	16075C, 16223T, 16274A, 16293T, 16311C, 16355T, 16362C, 16399G, 73G, 146C, 150T, 195C, 244G, 263G, 315.1C	5	KEMR022, KEYK017, KEYK036, Kesm018, TZDT071					
33	L4b2a1	16223T, 16274A, 16293T, 16311C, 16355T, 16362C, 16368C, 16399G, 73G, 146C, 195C, 244G, 263G, 315.1C	3	TZTR023, tzsw013, tzsw023					
34	L4b2a1	16223T, 16274A, 16293T, 16311C, 16355T, 16362C, 16399G, 16519C, 73G, 146C, 152C, 195C, 244G, 263G, 309.1C, 315.1C, 471C, 547G	1	TZWF011					
35	L4b2a1	16223T, 16274A, 16293T, 16311C, 16355T, 16362C, 16399G, 73G, 146C, 152C, 195C, 244G, 263G, 309.1C, 315.1C, 471C, 524.1A, 525.1C, 547G	1	KEBJ004					
36	L4b2a1	16223T, 16274A, 16293T, 16311C, 16355T, 16362C, 16399G, 73G, 146C, 152C, 195C, 244G, 263G, 315.1C, 471C, 547G	1	KEBR039		1			1
37	L4b2a2a	15924G, 16093C, 16189C, 16192.1T, 16223T, 16293T, 16311C, 16344T, 16355T, 16362C, 16399G, 16519C, 73G, 146C, 182T, 183G, 244G, 249:, 263G, 309.1C, 315.1C, 522-3D	1	TZIQ081					
38	L4b2a2a	15924G, 16093C, 16189C, 16192T, 16223T, 16293T, 16311C, 16344T, 16355T, 16362C, 16399G, 16519C, 73G, 146C, 182T, 183G, 244G, 249:, 263G, 309.1C, 315.1C, 522-3D	2	TZIQ077, TZWF044					
39	L4b2a2a	15924G, 16111T, 16184T, 16223T, 16293T, 16311C, 16355T, 16362C, 16399G, 16519C, 73G, 146C, 244G, 263G, 315.1C, 398C, 522-3D	1	KEGK016					
40	L4b2a2a	15924G, 16111T, 16223T, 16269G, 16293C, 16311C, 16355T, 16362C, 16399G, 16519C, 73G, 146C, 185A, 244G, 263G, 315.1C, 398C, 522-3D	1	KELO001					
41	L4b2a2a	15924G, 16111T, 16223T, 16293T, 16294T, 16311C, 16355T, 16362C, 16399G, 16519C, 73G, 146C, 244G, 263G, 315.1C, 398C, 522-3D	1	KETK053					
42	L4b2a2a	15924G, 16111T, 16223T, 16293T, 16299G, 16311C, 16355T, 16362C, 16399G, 16519C, 73G, 146C, 244G, 263G, 309.1C, 315.1C, 398C, 522-3D	1	KEGB042					
43	L4b2a2a	15924G, 16111T, 16223T, 16293T, 16311C, 16355T, 16362C, 16399G, 16519C, 73G, 146C, 185A, 244G, 263G, 309.1C, 315.1C, 398C, 522-3D	1	KEMR002					
44	L4b2a2a	15924G, 16111T, 16223T, 16293T, 16311C, 16355T, 16362C, 16399G, 16519C, 73G, 146C, 244G, 263G, 309.1C, 315.1C, 398C, 522-3D	8	KEBJ039, KEKY001, KEMR008, KEND005, KEPK078, KESN001, KESN006, KEMR010		1			
45	L4b2a2a	15924G, 16111T, 16223T, 16293T, 16311C, 16355T, 16362C, 16399G, 16519C, 73G, 146C, 244G, 263G, 309.1C, 315.1C, 398C, 513A, 593C	1	TZAK031					
46	L4b2a2a	15924G, 16111T, 16223T, 16293T, 16311C, 16355T, 16362C, 16399G, 16519C, 73G, 146C, 244G, 263G, 315.1C, 398C, 522-3D	2	KEPK016, KEPK039					1
47	L4b2a2a	15924G, 16189C, 16223T, 16293T, 16311C, 16344T, 16355T, 16362C, 16399G, 16519C, 73G, 146C, 182T, 183G, 244G, 249:, 263G, 315.1C, 522-3D	1	TZDT063					
48	L4b2a2a	15924G, 16223T, 16293T, 16311C, 16355T, 16362C, 16399G, 16519C, 73G, 93G, 146C, 200G, 244G, 263G, 309.1C, 315.1C, 522-3D	2	KETK024, KETK073					
49	L4b2a2a	16093C, 16223T, 16293T, 16311C, 16355T, 16362C, 16399G, 16519C, 73G, 146C, 244G, 263G, 315.1C	2	TZDR015, TZDR016					
50	L4b2a2a	16223T, 16293T, 16301T, 16311C, 16355T, 16362C, 16399G, 16519C, 73G, 146C, 244G, 263G, 315.1C, 455.1T, 522-3D	5	KEYK006, KEYK032, KEYK046, Kemn003, TZTR046		1			
51	L4b2a2a	16223T, 16293T, 16301T, 16311C, 16355T, 16362C, 16399G, 16519C, 73G, 244G, 263G, 309.1C, 315.1C, 455.1T, 522-3D	1	Kesm003					
52	L4b2a2a	15924G, 16111T, 16223T, 16293T, 16311C, 16355T, 16360T, 16362C, 16399G, 16519C, 73G, 146C, 195C, 244G, 263G, 315.1C, 398C, 522-3D	1	KESN028					
53	L4b2a2a	16051G, 16114T, 16189C, 16192T, 16223T, 16293T, 16311C, 16316G, 16355T, 16362C, 16399G, 16519C, 73G, 146C, 152C, 195C, 244G, 263G, 315.1C, 340T, 522-3D	3	TZHZ088, tzhz001, tzhz091					
54	L4b2a2a	16183G, 16223T, 16261T, 16293T, 16311C, 16355T, 16362C, 16399G, 16519C, 73G, 146C, 244G, 257G, 263G, 294C, 309.1C, 315.1C	1	SDSH001					
55	L4b2a2a	16223T, 16293T, 16311C, 16355T, 16362C, 16399G, 16519C, 73G, 146C, 257G, 263G, 315.1C	3	KEGB053, SDDN017, KEGB032					

56	L4b2a2b	15884A, 16071T, 16172C, 16223T, 16293T, 16311C, 16355T, 16362C, 16399G, 16519C, 16T, 73G, 244G, 263G, 315.1C	1	Kesm001					
57	L4b2a2b	16172C, 16223T, 16293T, 16311C, 16355T, 16362C, 16399G, 16519C, 73G, 146C, 195C, 244G, 263G, 309.1C, 315.1C, 391C	1	KEMR020					
58	L4b2a2b	16172C, 16223T, 16293T, 16311C, 16355T, 16362C, 16399G, 16519C, 73G, 146C, 195C, 244G, 263G, 315.1C, 391C	2	SDNR006, SDNR015					
59	L4b2a2b	15884A, 15930A, 16172C, 16189C, 16223T, 16293T, 16311C, 16355T, 16362C, 16399G, 16519C, 73G, 146C, 152C, 244G, 263G, 315.1C, 391C	1	KEPK023					
60	L4b2a2b	15924G, 16172C, 16223T, 16293T, 16311C, 16319A, 16355T, 16362C, 16399G, 16519C, 73G, 146C, 152C, 244G, 263G, 315.1C, 391C	1	KEKY004					
61	L4b2a2b	15930A, 16071T, 16093C, 16172C, 16223T, 16293T, 16311C, 16355T, 16362C, 16399G, 16519C, 73G, 146C, 151T, 189G, 244G, 263G, 315.1C	1	KEGK013			1		
62	L4b2a2b	15930A, 16071T, 16172C, 16223T, 16293T, 16311C, 16355T, 16362C, 16399G, 16519C, 73G, 146C, 152C, 189G, 244G, 263G, 315.1C, 573+4C	1	KEGK034B					
63	L4b2a2b	16071T, 16172C, 16223T, 16293T, 16311C, 16355T, 16362C, 16399G, 16519C, 73G, 146C, 244G, 263G, 315.1C	5	KESY001, KESY002, KESY005, KESY007, TZBG055					
64	L4b2a2b	16111A, 16172C, 16223T, 16293T, 16311C, 16355T, 16362C, 16381C, 16399G, 73G, 146C, 152C, 244G, 263G, 309.1C, 315.1C	1	KEOG009					
65	L4b2a2b	16169T, 16172C, 16223T, 16293T, 16311C, 16355T, 16362C, 16399G, 16519C, 73G, 125G, 146C, 244G, 263G, 309.1C, 315.1C, 391C	1	KETG036					
66	L4b2a2b	16169T, 16172C, 16223T, 16293T, 16311C, 16355T, 16362C, 16399G, 16519C, 73G, 146C, 244G, 263G, 309.1C, 315.1C, 391C	2	KERD057, KETG064					
67	L4b2a2b	16169T, 16172C, 16223T, 16293T, 16311C, 16355T, 16362C, 16399G, 16519C, 73G, 146C, 244G, 263G, 315.1C, 391C	1	KETG088					
68	L4b2a2b	16172C, 16173T, 16192T, 16223T, 16293T, 16311C, 16319A, 16355T, 16362C, 16399G, 73G, 146C, 244G, 263G, 309.1C, 315.1C	2	KESY014, KESY015					
69	L4b2a2b	16172C, 16183C, 16189C, 16223T, 16287G, 16293T, 16311C, 16355T, 16362C, 16399G, 73G, 146C, 152C, 244G, 263G, 315.1C, 522-3D	1	KETK010					
70	L4b2a2b	16172C, 16183C, 16189C, 16223T, 16287G, 16293T, 16311C, 16355T, 16362C, 16399G, 73G, 146C, 152C, 244G, 263G, 315.1C, 522-3D	1	KEPK083					
71	L4b2a2b	16172C, 16209C, 16223T, 16292T, 16293T, 16311C, 16355T, 16362C, 16399G, 16519C, 73G, 146C, 244G, 263G, 315.1C, 391C	1	TZMS055					
72	L4b2a2b	16172C, 16213A, 16223T, 16266T, 16293T, 16311C, 16355T, 16362C, 16399G, 16519C, 73G, 146C, 244G, 263G, 309.1C, 315.1C, 391C	1	KEND001					
73	L4b2a2b	16172C, 16223T, 16278T, 16293T, 16311C, 16355T, 16362C, 16399G, 73G, 146C, 244G, 263G, 315.1C	1	KEKP001					
74	L4b2a2b	16172C, 16223T, 16293G, 16311C, 16355T, 16362C, 16399G, 16519C, 73G, 146C, 152C, 244G, 263G, 315.1C, 391C	1	KEKY003					
75	L4b2a2b	16172C, 16223T, 16293T, 16311C, 16327T, 16355T, 16362C, 16399G, 16519C, 73G, 146C, 244G, 263G, 315.1C, 391C	3	KEOR006, tzsw030, tzsw032			1		1
76	L4b2a2b	16172C, 16223T, 16293T, 16311C, 16355T, 16362C, 16399G, 16519C, 73G, 146C, 244G, 263G, 309.1C, 315.1C, 391C	6	KEOG026, KEOG042, KEOG050, KEOG051, KEND003, KEYK007					
77	L4b2a2b	16172C, 16223T, 16293T, 16311C, 16355T, 16362C, 16399G, 16519C, 73G, 146C, 244G, 263G, 315.1C, 391C	3	SDSH013, TZDT081, KESM017					
78	L4b2a2b	16172C, 16223T, 16293T, 16311C, 16355T, 16362C, 16399G, 16519C, 73G, 146C, 244G, 263G, 315.1C, 391C, 519G	1	KEWT006					
79	L4b2a2b	16172C, 16223T, 16293T, 16311C, 16355T, 16362C, 16399G, 73G, 146C, 244G, 263G, 315.1C	1	KELO077					
80	L4b2a2b1	15970C, 16172C, 16223T, 16234T, 16287T, 16293T, 16311C, 16355T, 16362C, 16399G, 73G, 146C, 244G, 263G, 309.1C, 315.1C, 522-3D	1	TZRG051					
81	L4b2a2b1	15970C, 16172C, 16223T, 16234T, 16287T, 16293T, 16311C, 16355T, 16362C, 16399G, 73G, 146C, 244G, 263G, 315.1C, 522-3D	1	TZIQ034					
82	L4b2a2b1	15970C, 16172C, 16223T, 16278T, 16293T, 16294T, 16311C, 16355T, 16399G, 73G, 146C, 244G, 263G, 309.1C, 315.1C, 606G	3	TZHZ046, TZHZ084, tzhz082			1		
83	L4b2a2b1	15970C, 16172C, 16223T, 16278T, 16293T, 16294T, 16311C, 16355T, 16399G, 73G, 146C, 244G, 263G, 315.1C, 606G	8	TZDT029, TZHZ074, tzhz004, tzhz026, tzhz033, tzhz043, tzhz092, tzhz248			1		1

84	L4b2a2b1	15970C, 16172C, 16223T, 16287T, 16293T, 16311C, 16355T, 16362C, 16399G, 73G, 146C, 244G, 263G, 315.1C	1	SDBA009					
85	L4b2a2b1	15970C, 16172C, 16223T, 16287T, 16293T, 16311C, 16355T, 16362C, 16399G, 73G, 146C, 244G, 263G, 315.1C, 522-3D	1	KERD049					
86	L4b2a2b1	15970C, 16172C, 16223T, 16293T, 16311C, 16355T, 16362C, 16399G, 73G, 146C, 199C, 244G, 263G, 315.1C	2	tzsw037, TZTR042					
87	L4b2a2b1	15970C, 16172C, 16223T, 16293T, 16311C, 16355T, 16362C, 16399G, 73G, 146C, 244G, 263G, 309.1C, 315.1C, 356.1C	2	KEBR056, KEBR095					
88	L4b2a2b1	15970C, 16172C, 16223T, 16293T, 16311C, 16355T, 16362C, 16399G, 73G, 146C, 244G, 263G, 315.1C	4	TZTR003a, TZRG054, TZSW018, TZTR034					
89	L4b2a2b1	15970C, 16172C, 16223T, 16293T, 16311C, 16355T, 16399G, 73G, 146C, 244G, 246C, 263G, 315.1C	2	TZAK019, TZAK021		1			

L5	Hg	Controlregion	Abs. Freq	Samples
1	L5c	16129A, 16166G, 16187T, 16223T, 16254G, 16278T, 16291T, 16368C, 16527T, 73G, 195C, 200G, 247A, 263G, 315.1C, 522-3D, 573.1C, 573.2C, 573.3C, 573.4C	1	KESB020
2	L5c	15924G, 15941C, 16042A, 16129A, 16145A, 16166G, 16187T, 16189C, 16223T, 16278T, 16291T, 16400T, 73G, 146C, 152C, 182T, 195C, 198T, 247A, 263G, 315.1C, 550G	1	KEBN041
3	L5c	15927A, 16093C, 16129A, 16166G, 16172C, 16187T, 16189C, 16209C, 16213A, 16223T, 16278T, 16311C, 16519C, 73G, 152C, 182T, 247A, 263G, 309.1C, 315.1C, 522-3D, 593C	1	tzsw017
4	L5a	15930T, 16104T, 16111T, 16129A, 16148T, 16166G, 16187T, 16189C, 16223T, 16254G, 16278T, 16291T, 16360T, 73G, 182T, 185A, 195C, 247D, 263G, 315.1C, 459.1C, 498+2C, 499C, 535T	1	KETG043
5	L5a	15930T, 16111T, 16129A, 16148T, 16166G, 16176T, 16187T, 16189C, 16223T, 16254G, 16278T, 16291T, 16360T, 73G, 182T, 185A, 195C, 247D, 263G, 315.1C, 459.1C, 498+2C, 499C, 535T	2	KENJ002, KEKY006
6	L5a	15930T, 16111T, 16129A, 16148T, 16166G, 16176T, 16187T, 16189C, 16223T, 16254G, 16278T, 16291T, 16360T, 73G, 182T, 195C, 247D, 263G, 315.1C, 459.1C, 535T	1	KETV046
7	L5a1	15924G, 16111T, 16129A, 16148T, 16166G, 16187T, 16189C, 16223T, 16254G, 16278T, 16360T, 73G, 182T, 183G, 185C, 195C, 247D, 263G, 309.1C, 315.1C, 459.1C, 535T	1	KEGU011
8	L5a1	16111T, 16129A, 16148T, 16166G, 16187T, 16189C, 16223T, 16278T, 16360T, 16519C, 73G, 182T, 195C, 247D, 263G, 315.1C, 455.1T, 459.1C, 535T	2	KEBJ006, KEBJ030a
9	L5a1a	16111T, 16129A, 16148T, 16166G, 16187T, 16189C, 16223T, 16254G, 16278T, 16311C, 16319A, 16360T, 16519C, 143A, 146C, 152C, 185A, 189G, 247A, 263G, 309.1C, 315.1C, 459.1C, 535T, 573.1C, 573.2C, 573.3C, 606G	1	KEBR008
10	L5a1a	16111T, 16129A, 16148T, 16166G, 16187T, 16189C, 16223T, 16254G, 16278T, 16311C, 16319A, 16360T, 73G, 182T, 195C, 247D, 263G, 309.1C, 315.1C, 459.1C, 535T	1	KEKS013
11	L5a2	16071T, 16111T, 16129A, 16148T, 16166G, 16187T, 16189C, 16223T, 16266T, 16278T, 16335G, 16360T, 73G, 195C, 247D, 263G, 315.1C, 459.1C, 535T	1	KESM037
12	L5a2	16111T, 16129A, 16148T, 16166G, 16187T, 16189C, 16223T, 16254G, 16278T, 16335G, 16360T, 73G, 152C, 195C, 247D, 263G, 315.1C, 459.1C, 535T	1	KESN026
13	L5a2	16111T, 16129A, 16148T, 16166G, 16187T, 16189C, 16223T, 16254G, 16278T, 16335G, 16360T, 73G, 195C, 247D, 263G, 315.1C, 459.1C, 535T	1	KESN049
14	L5a3	16111T, 16129A, 16148T, 16166G, 16187T, 16189C, 16223T, 16233G, 16254G, 16278T, 16360T, 16519C, 73G, 195C, 247D, 263G, 315.1C, 459.1C, 535T	1	KEPK053
15	L5a3	16111T, 16129A, 16148T, 16166G, 16187T, 16189C, 16223T, 16233G, 16254G, 16278T, 16360T, 73G, 195C, 247D, 263G, 315.1C, 459.1C, 535T	1	KEOG005
16	L5a3a	15927A, 16111T, 16129A, 16148T, 16166G, 16187T, 16189C, 16223T, 16254G, 16278T, 16344T, 16360T, 73G, 195C, 247D, 263G, 315.1C, 459.1C, 535T	2	KEND012, KENJ013
17	L5a3a1	15904T, 15927A, 16111T, 16129A, 16148T, 16166G, 16187T, 16189C, 16223T, 16231C, 16233G, 16239T, 16254G, 16278T, 73G, 146C, 195C, 247D, 263G, 315.1C, 459.1C, 535T, 593C	1	Keil005
18	L5a3a1	15927A, 16111T, 16129A, 16148T, 16166G, 16187T, 16189C, 16223T, 16233G, 16254G, 16265C, 16278T, 16360T, 73G, 195C, 247D, 263G, 315.1C, 459.1C, 501T, 524.1A, 525.1C, 535T	1	TZMS053
19	L5a3a1	15927A, 16111T, 16129A, 16148T, 16166G, 16187T, 16189C, 16223T, 16233G, 16254G, 16265C, 16278T, 16360T, 73G, 195C, 247D, 263G, 315.1C, 459.1C, 501T, 535T	2	KEGK011, KESN065
20	L5a3a1	15927A, 16111T, 16129A, 16148T, 16166G, 16187T, 16189C, 16223T, 16233G, 16254G, 16278T, 16344T, 16360T, 16488T, 73G, 195C, 247D, 263G, 315.1C, 459.1C, 535T	1	SDNR020
21	L5a3a1	15927A, 16111T, 16129A, 16148T, 16166G, 16187T, 16189C, 16223T, 16233G, 16254G, 16278T, 16344T, 16360T, 73G, 195C, 247D, 263G, 315.1C, 459.1C, 535T	1	KELY013
22	L5a3a1	15927A, 16111T, 16129A, 16148T, 16166G, 16187T, 16189C, 16223T, 16233G, 16254G, 16278T, 16360T, 16519C, 73G, 195C, 247D, 263G, 309.1C, 315.1C, 459.1C, 501T, 535T	8	KEKP002, KEKP004, KELO048, KEPK013, KESN003, KESN029, KETG009, KETS005
23	L5a3a1	15927A, 16111T, 16129A, 16148T, 16166G, 16187T, 16189C, 16223T, 16233G, 16254G, 16278T, 16360T, 73G, 195C, 247D, 263G, 315.1C, 459.1C, 535T	3	KELY028, SDDN004, SDNY014
24	L5a3a1	15927A, 16111T, 16148T, 16166G, 16187T, 16189C, 16223T, 16233G, 16254G, 16278T, 16360T, 73G, 195C, 247D, 263G, 315.1C, 459.1C, 535T	1	SDNY008
25	L5b1	15884A, 15924G, 16129A, 16148T, 16166G, 16183:, 16187T, 16189C, 16223T, 16278T, 16311C, 16355T, 73G, 152C, 182T, 195C, 247A, 263G, 315.1C, 455+2T, 459.1C, 522-3D	1	SDDN016

26	L5b1	15884A, 16129A, 16148T, 16166G, 16183:, 16187T, 16188T, 16189C, 16223T, 16278T, 16311C, 16355T, 16362C, 73G, 152C, 182T, 195C, 247A, 263G, 309.1C, 315.1C, 455+2T, 459.1C, 522-3D	1	KETS002	
27	L5b1	15884A, 16129A, 16148T, 16166G, 16183:, 16187T, 16189C, 16223T, 16257T, 16278T, 16311C, 16355T, 16362C, 73G, 152C, 182T, 195C, 247A, 263G, 315.1C, 455+2T, 459.1C, 522-3D, 547G	7	KEBR071, KEMR026, KEPK010, KEPK034, KEPK038, KEPK055, KETK035	
28	L5b1	15884A, 16129A, 16148T, 16166G, 16183:, 16187T, 16189C, 16223T, 16278T, 16311C, 16355T, 16362C, 16519C, 73G, 146C, 152C, 182T, 195C, 247A, 263G, 315.1C, 455+2T, 459.1C, 522-3D	1	SDDN027	
29	L5b1	15884A, 16129A, 16148T, 16166G, 16183:, 16187T, 16189C, 16223T, 16278T, 16311C, 16355T, 16362C, 73G, 152C, 182T, 195C, 247D, 263G, 309.1C, 315.1C, 455+2T, 459.1C, 522-3D	1	KETV045	
30	L5b1	15884A, 16129A, 16148T, 16166G, 16183:, 16187T, 16189C, 16223T, 16278T, 16311C, 16355T, 16362C, 73G, 152C, 182T, 195C, 247A, 263G, 315.1C, 455+2T, 459.1C, 522-3D	7	KEGK028, KESB014, KESN031, KESN073, SDNR018, SDSH012, TZMS052	
31	L5b1	15884A, 16129A, 16148T, 16166G, 16183:, 16187T, 16189C, 16278T, 16311C, 16355T, 16362C, 73G, 152C, 182T, 195C, 247A, 315.1C, 455+2T, 459.1C, 522-3D	3	KEPR023, KEPR026, TZPR008	
32	L5b1	16129A, 16148T, 16166G, 16183:, 16187T, 16189C, 16223T, 16263C, 16278T, 16311C, 16355T, 16362C, 73G, 152C, 182T, 195C, 247A, 263G, 309.1C, 315.1C, 455+2T, 459.1C, 522-3D	2	SDSH004, SDSH014	
33	L5b1	16129A, 16148T, 16166G, 16183:, 16187T, 16189C, 16223T, 16278T, 16311C, 16355T, 16362C, 73G, 152C, 182T, 195C, 247A, 263G, 309.1C, 315.1C, 455+2T, 459.1C, 522-3D	1	SDDN020	
34	L5b1	16129A, 16148T, 16166G, 16183:, 16187T, 16189C, 16223T, 16278T, 16311C, 16355T, 16362C, 73G, 152C, 182T, 195C, 247A, 263G, 315.1C, 455+2T, 459.1C, 522-3D	1	KETG053	
35	L5b1a2	15884A, 16129A, 16148T, 16166G, 16183C, 16186T, 16187T, 16189C, 16223T, 16278T, 16311C, 16355T, 16362C, 73G, 152C, 182T, 195C, 247A, 263G, 309.1C, 315.1C, 455+2T, 459.1C, 522-3D	1	KEPK004	
36	L5b1a2	15884A, 16129A, 16148T, 16166G, 16183C, 16186T, 16189:, 16223T, 16278T, 16311C, 16355T, 16362C, 73G, 152C, 182T, 195C, 247A, 263G, 315.1C, 455+2T, 459.1C, 522-3D	4	KESN035, KESN060, KEYK035, SDDN018	
37	L5b1a2	15884A, 16129A, 16148T, 16166G, 16183C, 16189C, 16223T, 16231C, 16278T, 16311C, 16355T, 16362C, 73G, 152C, 195C, 247A, 263G, 315.1C, 455+2T, 459.1C, 522-3D	1	SDNR028	
38	L5b1a2	16129A, 16148T, 16166G, 16183C, 16186T, 16187T, 16189C, 16223T, 16278T, 16311C, 16355T, 16362C, 73G, 152C, 182T, 195C, 247A, 263G, 309.1C, 315.1C, 455+2T, 459.1C, 522-3D	1	KETG024	
39	L5b1c	15884A, 16093C, 16129A, 16148T, 16166G, 16183:, 16187T, 16189C, 16223T, 16278T, 16311C, 16355T, 16362C, 73G, 152C, 182T, 195C, 247A, 263G, 315.1C, 455+2T, 459.1C, 513A, 522-3D	1	KETK008	
					L6
L6	Hg	Controlregion	Abs. Freq	Samples	C9332T
1	L6	15924G,16048A,16153A,16179T,16192T,16215G,16223T,16256A,16284G,16311C,73G,189C,195C,263G,309.1C,315.1C,523:,525:	1	TZDT054	
2	L6	16048A,16093C,16223T,16224C,16278T,16311C,16519C,73G,146C,152C,182T,185C,263G,315.1C,316A	1	KEBJ023	1
3	L6	16048A,16223T,16311C,16343G,16368C,16519C,73G,146C,152C,182T,185C,263G,309.1C,315.1C	1	KERD053	1
4	L6a	15927A,16048A,16129A,16223T,16224C,16278T,16311C,16519C,73G,146C,152C,182T,185C,228A,263G,265C,315.1C	1	KEBR015	1
5	L6a	16048A,16173T,16184T,16223T,16224C,16278T,16311C,16362C,16399G,16519C,73G,146C,152C,182T,185C,207A,263G,265C,309.1C,315.1C	1	KERD040	
6	L6a	16048A,16223T,16224C,16233G,16278T,16311C,16357C,16519C,73G,146C,152C,182T,185C,263G,265C,315.1C,523:,525:	1	SDBA017	

					M1	M1
M	Hg	Controlregion	Abs. Freq	Samples	A10398G	C10400T
1	M*	16223T, 16294T, 16497G, 16519C, 73G, 189G, 195C, 263G, 315.1C, 489C	1	KEOR025		
2	M1	16129A, 16183C, 16189C, 16223T, 16249C, 16311C, 16519C, 73G, 195C, 263G, 309.1C, 309.2C, 315.1C, 489C	3	KEWT016, KEWT027, KEWT028	1	1
3	M1a1	15959A, 16129A, 16182C, 16183C, 16189C, 16223T, 16249C, 16311C, 16359C, 16519C, 73G, 195C, 263G, 309.1C, 315.1C, 489C	1	Keyk003	1	1
4	M1a1	16129A, 16148T, 16182C, 16183C, 16189C, 16223T, 16249C, 16294T, 16311C, 16359C, 16519C, 73G, 195C, 263G, 309.1C, 309.2C, 311.1C, 489C	1	Kesm024	1	1
5	M1a1	16129A, 16182C, 16183C, 16189C, 16223T, 16249C, 16294T, 16311C, 16359C, 16519C, 73G, 195C, 263G, 315.1C, 489C	1	KENJ001	1	1
6	M1a1	16129A, 16182C, 16183C, 16189C, 16223T, 16249C, 16311C, 16359C, 16519C, 73G, 195C, 207A, 263G, 309.1C, 309.2C, 311.1C, 489C	2	KEOG035, KEOG001	1	1
7	M1a1	16129A, 16182C, 16183C, 16189C, 16223T, 16249C, 16311C, 16359C, 16519C, 73G, 195C, 207A, 263G, 309.1C, 315.1C, 489C	2	KEOG022, KEOG032	1	1
8	M1a1	16129A, 16182C, 16183C, 16189C, 16223T, 16249C, 16311C, 16359C, 16519C, 73G, 195C, 263G, 309.1C, 309.2C, 315.1C, 489C	1	TZDT004	1	1
9	M1a1	16129A, 16182C, 16183C, 16189C, 16223T, 16249C, 16311C, 16359C, 16519C, 73G, 195C, 263G, 309.1C, 315.1C, 489C	1	KEWT002	1	1
10	M1a1	16129A, 16183C, 16189C, 16213A, 16223T, 16249C, 16311C, 16359C, 16519C, 73G, 195C, 263G, 309.1C, 315.1C, 489C	1	SDHD009	1	1
11	M1a1	16129A, 16183C, 16189C, 16223T, 16249C, 16311C, 16359C, 16519C, 73G, 195C, 207A, 263G, 309.1C, 315.1C, 489C	1	KEWT032	1	1
12	M1a1	16129A, 16183C, 16189C, 16223T, 16249C, 16311C, 16359C, 16519C, 73G, 195C, 263G, 309.1C, 309.2C, 315.1C, 316A, 489C	1	KEGU025	1	1
13	M1a1	16129A, 16183C, 16189C, 16223T, 16249C, 16311C, 16359C, 16519C, 73G, 195C, 263G, 309.1C, 309.2C, 315.1C, 489C	2	KEYK019, TZIQ025	1	1
14	M1a1	16129A, 16183C, 16189C, 16223T, 16249C, 16311C, 16359C, 16519C, 73G, 195C, 263G, 309.1C, 315.1C, 316A, 489C	2	KERD067, KERD074	1	1
15	M1a1	16129A, 16183C, 16189C, 16223T, 16249C, 16311C, 16359C, 16519C, 73G, 195C, 263G, 309.1C, 315.1C, 489C	3	KEOR014, KEOR015,TZWF009	1	1
16	M1a1	16129A, 16183C, 16189C, 16223T, 16249C, 16311C, 16359C, 16519C, 73G, 195C, 263G, 315.1C, 489C	2	KESN069, SDHD012	1	1
17	M1a1	16129A, 16189C, 16213A, 16223T, 16249C, 16311C, 16359C, 16519C, 73G, 195C, 263G, 309.1C, 315.1C, 489C	1	KEND010	1	1
18	M1a1	16129A, 16189C, 16223T, 16249C, 16311C, 16359C, 16519C, 73G, 195C, 263G, 315.1C, 489C	1	Kesm002	1	1
19	M1a1d	16093C, 16129A, 16182.1C, 16183C, 16189C, 16223T, 16249C, 16311C, 16359C, 16519C, 73G, 150T, 189G, 195C, 198T, 263G, 310C, 312T, 315+3C, 489C, 522-3D	1	KEBJ020	1	1
20	M1a1d	16093C, 16129A, 16183C, 16189C, 16223T, 16249C, 16311C, 16359C, 16519C, 73G, 150T, 195C, 198T, 263G, 315.1C, 489C, 522-3D	2	KEBR061, KEBR044	1	1
21	M1a1d	16093C, 16129A, 16188.1C, 16189C, 16223T, 16249C, 16311C, 16359C, 16519C, 73G, 150T, 189G, 195C, 198T, 263G, 309.1C, 315.1C, 489C, 522-3D	1	KEGU032	1	1
22	M1a1d	16129A, 16183C, 16189C, 16223T, 16249C, 16311C, 16359C, 16519C, 73G, 150T, 189G, 195C, 198T, 263G, 309.1C, 315.1C, 489C, 522-3D	1	SDBA027	1	1
23	M1a1d	16129A, 16189C, 16223T, 16249C, 16311C, 16359C, 16519C, 73G, 150T, 152C, 189G, 195C, 198T, 263G, 309.1C, 315.1C, 489C, 522-3D	2	KEGU009, KEMG028	1	1
24	M1a1d	16129A, 16189C, 16223T, 16249C, 16311C, 16359C, 16519C, 73G, 150T, 189G, 195C, 198T, 263G, 309.1C, 315.1C, 489C, 522-3D	2	TZMS058 , TZMS074a	1	1
25	M1a1i	16129A, 16182C, 16183C, 16189C, 16223T, 16249C, 16311C, 16359C, 16519C, 73G, 183G, 195C, 263G, 309.1C, 315.1C, 489C	8	KEYK031, KEYK044, Keil021, Keil023, KEIL024, KEIL025a, TZDT069, keyk009	1	1
26	M1a1i	16129A, 16183C, 16189C, 16223T, 16249C, 16311C, 16359C, 16519C, 73G, 183G, 195C, 263G, 309.1C, 315.1C, 489C	1	TZDT035	1	1
27	M1a2b	16126C, 16182C, 16183C, 16189C, 16223T, 16249C, 16255A, 16260T, 16311C, 16320T, 16519C, 73G, 195C, 200G, 263G, 309.1C, 309.2C, 315.1C, 489C	1	KEKS001	1	1
28	M1a2b	16182C, 16183C, 16189C, 16223T, 16249C, 16255A, 16260T, 16311C, 16320T, 73G, 195C, 200G, 263G, 309.1C, 315.1C, 489C, 593C	1	KEBJ009	1	1
29	M1a2c	16129A, 16182G, 16189C, 16223T, 16249C, 16265C, 16311C, 16519C, 73G, 195C, 263G, 309.1C, 315.1C, 489C	1	KEWT012	1	1

30	M1a2c	15884A, 16129A, 16182G, 16186T, 16189C, 16223T, 16249C, 16265C, 16311C, 16519C, 73G, 152C, 195C, 263G, 309.1C, 315.1C, 489C	1	KEKS012	1	1
31	M1a2c	15884A, 16129A, 16182G, 16187T, 16189C, 16223T, 16249C, 16265C, 16311C, 16519C, 73G, 195C, 263G, 315.1C, 489C	5	KEGB028, KEGB043, KEMG004, KEMG005, KEMG027	1	1
32	M1a2c	15884A, 16129A, 16182G, 16189C, 16223T, 16249C, 16265C, 16311C, 16519C, 73G, 195C, 263G, 315.1C, 489C	1	Kemn009	1	1
33	M1a4	16129A, 16182C, 16183C, 16189C, 16223T, 16249C, 16311C, 16519C, 73G, 195C, 198T, 263G, 309.1C, 309.2C, 315.1C, 489C	1	KEBR026		
34	M1a4	16051G, 16129A, 16182C, 16183C, 16189C, 16223T, 16249C, 16266T, 16311C, 16519C, 73G, 195C, 263G, 309.1C, 309.2C, 315.1C, 489C	2	KEKS002, KEKS005	1	1
35	M1a4	16129A, 16182C, 16183C, 16189C, 16223T, 16249C, 16311C, 16519C, 73G, 189G, 195C, 200G, 207A, 263G, 309.1C, 315.1C, 489C, 513:	1	SDBA032	1	1
36	M1a4	16182C, 16183C, 16189C, 16223T, 16249C, 16311C, 16519C, 73G, 195C, 263G, 309.1C, 309.2C, 315.1C, 489C	1	KEGB056	1	1
37	M1a4	16182C, 16183C, 16189C, 16223T, 16249C, 16311C, 16519C, 73G, 195C, 263G, 309.1C, 315.1C, 489C	1	KEGU010	1	1
38	M1a4	16182C, 16183C, 16189C, 16223T, 16249C, 16311C, 16465T, 16519C, 73G, 195C, 263G, 309.1C, 315.1C, 489C	2	KEGB027, KEWT026	1	1
39	M1a5	16183C, 16189C, 16223T, 16249C, 16311C, 16519C, 73G, 195C, 263G, 309.1C, 315.1C, 489C, 523.1CA	2	KEGK038, KENJ009	1	1
40	M1a5	16183C, 16189C, 16223T, 16249C, 16311C, 16519C, 73G, 195C, 263G, 315.1C, 489C, 523.2CA	1	TZWF041	1	1
41	M1a5	15959A, 16183C, 16189C, 16223T, 16249C, 16311C, 16519C, 73G, 195C, 263G, 309.1C, 315.1C, 489C, 524.1A, 525.1C	1	KENJ019	1	1
42	M1a5	16182C, 16183C, 16189C, 16223T, 16249C, 16311C, 16519C, 73G, 189G, 195C, 263G, 309.1C, 309.2C, 315.1C, 489C	1	KEBR034	1	1
43	M1a5	16182C, 16183C, 16223T, 16311C, 16519C, 73G, 189G, 195C, 263G, 309.1C, 309.2C, 315.1C, 320G, 489C	1	KEBR024		
44	M1a5	16183C, 16189C, 16223T, 16249C, 16311C, 16365T, 16519C, 73G, 195C, 263G, 309.1C, 315.1C, 489C, 524.1A, 525.1C	2	KETG030, KETG059	1	1
45	M1a5	16183C, 16189C, 16223T, 16249C, 16311C, 16519C, 73G, 195C, 263G, 309.1C, 309.2C, 315.1C, 489C	1	SDBA011	1	1
46	M1a5	16183C, 16189C, 16223T, 16249C, 16311C, 16519C, 73G, 195C, 263G, 315.1C, 489C, 524.1A, 525.1C	1	KEYK021	1	1
47	M1a7	15860G, 16129A, 16189C, 16223T, 16249C, 16311C, 16519C, 55C, 57C, 73G, 152C, 195C, 263G, 309.1C, 315.1C, 489C	2	SDNY003, SDNY005	1	1
48	M1a9	16129A, 16189C, 16223T, 16249C, 16301T, 16311C, 16519C, 73G, 263G, 315.1C, 489C	1	KEND009	1	1
49	M1a9	16129A, 16189C, 16223T, 16249C, 16311C, 16399G, 16519C, 73G, 195C, 263G, 309.1C, 315.1C, 489C	1	KESB015 (KESB015A)	1	1
50	M8	16111T, 16136C, 16223T, 16260T, 16298C, 73G, 146C, 152C, 249:, 263G, 309.1C, 309.2C, 315.1C, 489C	1	XH016		

N	Hg	Controlregion	Abs. Freq	Samples
1	HV1b	15947T, 16067T, 16158G, 16177T, 16274A, 263G, 309.1C, 315.1C	1	KETG081
2	HV1b	16067T, 16158G, 16274A, 263G, 315.1C	3	KEEL013, KEEL037, KERD084
3	HV1b	16067T, 16274A, 16311C, 263G, 309.1C, 315.1C, 522-3D	2	KEMG024, KEMG015
4	HV1b	16067T, 16274A, 263G, 315.1C	1	KETG011
5	HV1b	16067T, 16274A, 263G, 315.1C	1	KEWT021
6	HV1b	16067T, 16274A, 263G, 315.1C, 522-3D	1	Keil018
7	HV1	16067T, 200G, 263G, 309.1C, 315.1C	1	SDBA008
8	I	15924G, 15951G, 16129A, 16223T, 16391A, 16519C, 73G, 152C, 199C, 204C, 250C, 263G, 309.1C, 315.1C, 573+5C	1	TZBG033
9	I	15924G, 16093C, 16129A, 16223T, 16325C, 16391A, 16519C, 73G, 199C, 250C, 263G, 315.1C, 573+3C	4	KEEL001B, KEEL025, KEEL033, KEEL036
10	I	15924G, 16093C, 16129A, 16223T, 16519C, 73G, 152C, 199C, 204C, 207A, 250C, 263G, 291+2A, 309.1C, 315.1C, 573+4C	1	KEEL049
11	I	15924G, 16129A, 16223T, 16391A, 16519C, 73G, 199C, 204C, 207A, 250C, 263G, 309.1C, 315.1C, 573+4C	1	KERD078
12	I	15924G, 16129A, 16223T, 16391A, 16519C, 73G, 199C, 204C, 207A, 250C, 263G, 309.1C, 315.1C, 573+5C	1	kesm004
13	I	15924G, 16129A, 16223T, 16519C, 73G, 146C, 199C, 204C, 207A, 250C, 263G, 291+2A, 309.1C, 315.1C, 324G, 573+4C	1	KERD003
14	I	15924G, 16129A, 16223T, 16519C, 73G, 152C, 199C, 204C, 207A, 250C, 263G, 291+2A, 309.1C, 315.1C, 573+3C	1	KETK041
15	I	15924G, 16129A, 16223T, 16519C, 73G, 152C, 199C, 204C, 207A, 250C, 263G, 291+2A, 309.1C, 315.1C, 573+4C	2	KEGB031, Keil028
16	I	15924G, 16129A, 16223T, 16519C, 73G, 199C, 204C, 207A, 250C, 263G, 291+2A, 309.1C, 315.1C, 573+3C	1	KEGB030
17	I	15924G, 16129A, 16223T, 16519C, 73G, 199C, 204C, 207A, 250C, 263G, 291+2A, 309.1C, 315.1C, 573+4C	3	KEGB035, KEMG026, KEWT011
18	I	15924G, 16129A, 16223T, 16519C, 73G, 199C, 204C, 207A, 250C, 263G, 291+2A, 309.1C, 315.1C, 573+5C	1	Kemn005
19	J1b	16069T, 16126C, 16136C, 16145A, 16222T, 16261T, 16519C, 73G, 263G, 295T, 315.1C, 462T, 489C	1	SDBA019
20	J1c	16069T, 16126C, 16193T, 16300G, 16301T, 16309G, 73G, 152C, 263G, 295T, 309.1C, 315.1C, 462T, 489C	1	Kesm014
21	J1c	16069T, 16126C, 16193T, 16300G, 16309G, 73G, 152C, 263G, 295T, 315.1C, 316A, 462T, 489C	1	TZDT002
22	J1c	16069T, 16126C, 16193T, 16300G, 16309G, 73G, 152C, 263G, 295T, 315.1C, 462T, 489C	1	KERD002
23	K1a	16093C, 16224C, 16311C, 16482G, 16519C, 73G, 263G, 315.1C, 497T, 524.1A, 525.1C	1	KEYK034
24	K1a	16224C, 16311C, 16482G, 16519C, 73G, 263G, 315.1C, 497T	1	TZIQ051
25	K1a	16224C, 16311C, 16519C, 73G, 204C, 263G, 309.1C, 315.1C, 485C, 497T, 523+2AC	3	KERD081, KEYK018, KEKS014
26	K1a	16224C, 16311C, 16519C, 73G, 204C, 263G, 315.1C, 485C, 497T, 523+2AC	1	KERD050
27	N1a2	16147G, 16172C, 16213A, 16223T, 16248T, 16325C, 16355T, 16519C, 73G, 199C, 204C, 263G, 309.1C, 315.1C, 573+5C	1	KEBR018
28	N1a2	16147G, 16172C, 16213A, 16223T, 16248T, 16355T, 16519C, 73G, 199C, 204C, 207A, 263G, 309.1C, 315.1C, 573+3C	1	KEBR052
29	N1a2	16147G, 16172C, 16213A, 16223T, 16248T, 16355T, 16519C, 73G, 199C, 204C, 207A, 263G, 309.1C, 315.1C, 573+5C	1	KEGB004
30	N1a2	16147G, 16172C, 16213A, 16223T, 16248T, 16355T, 16519C, 73G, 199C, 204C, 207A, 263G, 315.1C, 573+4C	2	KEBJ021, KEBJ022
31	N1a2	16147G, 16172C, 16220G, 16223T, 16248T, 16355T, 16519C, 73G, 199C, 204C, 263G, 315.1C, 573+5C	1	KEGB037
32	N1a2	16147G, 16172C, 16223T, 16248T, 16355T, 16519C, 73G, 150T, 151T, 199C, 204C, 263G, 309.1C, 315.1C, 573+5C	1	KEGU042
33	N1a2	16147G, 16172C, 16223T, 16248T, 16355T, 16519C, 73G, 151T, 199C, 204C, 263G, 315.1C, 573+3C	1	KEBR020
34	N1a2	16147G, 16172C, 16223T, 16248T, 16355T, 16519C, 73G, 199C, 204C, 263G, 309.1C, 315.1C, 573+5C	1	TZIQ079
35	R0a2a	15924G, 16126C, 16305T, 16362C, 57.1C, 64T, 263G, 309.1C, 309.2C, 315.1C	1	KEWT004
36	R0a2a	16114T, 16126C, 16362C, 16519C, 57.1C, 64T, 263G, 309.1C, 315.1C	1	KEYK022
37	R0a2c	16114T, 16126C, 16362C, 16519C, 58C, 64T, 263G, 309.1C, 315.1C, 522-3D	1	TZDR002
38	R0a2a	16126C, 16188T, 16362C, 57.1C, 64T, 185A, 261T, 263G, 315.1C	1	SDBA033
39	R0a2a	16126C, 16309G, 16362C, 57.1C, 64T, 263G, 309.1C, 315.1C	2	KEGB029, KEGU017
40	R0a2b	16126C, 16309G, 16362C, 60.1T, 64T, 263G, 309.1C, 315.1C, 324	1	KEBR101
41	R0a2a	16126C, 16362C, 16519C, 57.1C, 64T, 152C, 263G, 309.1C, 315.1C	3	SDBA005, SDBA006, SDHD033
42	R0a2a	16126C, 16362C, 57.1C, 64T, 263G, 267C, 309.1C, 309.2C, 315.1C	1	KEBR076
43	R0a2b	16126C, 16362C, 60.1T, 64T, 131C, 263G, 309.1C, 309.2C, 315.1C	1	SDBA014
44	T2f	15928A, 16126C, 16153A, 16294T, 16296T, 16519C, 73G, 140T, 150T, 152C, 263G, 309.1C, 315.1C, 522-3D	4	KEOG005, KEOG006, KEOG007, KEOG021

45	U3	15865G, 16343G, 73G, 150T, 195C, 263G, 315.1C	1	SDBA012
46	U6a5	15927A, 16172C, 16184T, 16219G, 16234T, 16278T, 16519C, 73G, 263G, 315.1C	1	CHLA013
47	U6	16172C, 16219G, 16311C, 16320T, 73G, 263G, 315.1C	2	CAFU036, CAGZ021
48	U9	15927A, 16239T, 16519C, 73G, 195C, 263G, 315.1C, 455.1T, 499A	1	KEBR021
49	H1	16145A, 16222T, 16519C, 263G, 315.1C, 524.1A, 525.1C	1	CAAS016
50	U9	16162G, 16184T, 16239T, 16519C, 73G, 263G, 315.1C, 499A	1	KEOR023A

Country and Sample ID codes used for the samples

	Sample ID codes	Population
1	CAAS	Baggara
2	CABM	Bamoun
3	CAFU	Fulani
4	CAGZ	Giziga
5	CAKN	Kanuri
6	CAMV	Mvae
7	CAPB	Baka
8	CAPL	Bakola
9	CAPM	Medzan
10	CHBU	Boulala
11	CHLA	Laka
12	San, Hgdp	San
13	KEBJ	Burji
14	KEBN	Boni
15	KEBR	Borana
16	KEOR	Orma
17	KEEL	El-Molo
18	KEGB, KEMG	Gabra
19	KEGK	Gikuyu
20	KEGU	Garreh
21	KEIL, KEMN	Il'Laikipiak (Il'ngwesi and Mumonyot)
22	KEKP, KEND, KEKY	Kalenjin (Kipsigis, Nandi, Keiyo)
23	KEKS	Konso
24	KELO	Luo
25	KELY	Luhya
26	KEMR	Marakwet
27	KENJ	Ilchamus
28	KEOG	Ogiek
29	KEPK	Pokot
30	TZPR, KEPR	Pare
31	KERD	Rendille
32	KESB	Sabaot
33	KESM	Samburu
34	KESN	Sengwer
35	KESY	Sanye
36	KETA	Taita
37	KETG	Tugen
38	KETK	Turkana
39	KETS	Teso
40	KETV	Taveta
41	KEWT	Wata
42	KEYK	Yaaku
43	SDBJ, SDHD	Beja
44	SDDN	Dinka
45	SDNR	Nuer
46	SDNY	Nyimang
47	SDSH	Shilluk
48	TZAK	Akie
49	TZBG	Burunge
50	TZDR	Dorobo
51	TZDT	Datog
52	TZHZ	Hadza
53	TZIQ	Iraqw
54	TZMS	Maasai
55	TZRG	Rangi
56	TZSW	Sandawe
57	TZTR	Turu
58	TZWF	Fyome
59	XH	Xhosa

The first two letters are country codes while the second two letters are population codes, the number represent sample ID – the country codes are as follows

	Country Codes	
1	KE	Kenya
2	TZ	Tanzania
3	SD	Sudan
4	CA	Cameroon
5	CH	Chad

Some individuals that belong to different population (paternally belong to the population classified as indicated by ID codes but the maternal lineage is from different populations) despite having above listed IDs are as follows (reassigned new IDs a in bracket)

	Sample ID	Population		Sample ID	Population
1	KEBJ005 (KEBR131)	Borana	49	KENJ051 (KETG114)	Tugen
2	KEBJ032 (KEBR132)	Borana/Oromo	50	KENJ063 (KETG115)	Tugen
3	KEKS008 (KEBR134)	Borana	51	KENJ075 (KETG116)	Tugen
4	KEKS009 (KEBR133)	Borana	52	KEPK004 (KETG117)	Tugen
5	KEKS012 (KEBR135)	Borana	53	KEPK035 (KETG118)	Tugen
6	KEBJ037 (KEGB071)	Gabra	54	KEPK038 (KETG119)	Tugen
7	KEKS014 (KEGB072)	Gabra	55	KEPK040 (KETG120)	Tugen
8	KEGB056	Garreh	56	KEEL022 (KETK123)	Turkana
9	KEBJ022	Garreh	57	KEMR022 (KETK122)	Turkana
10	Keyk003 (Keil042)	Il'Laikipiak	58	KENJ076 (KETK124)	Turkana
11	KEYK006 (Keil043)	Il'Laikipiak	59	KEPK011 (KETK125)	Turkana
12	KEYK017 (Keil044)	Il'Laikipiak	60	KEPK023 (KETK126)	Turkana
13	Keyk036 (Keil045)	Il'Laikipiak	61	KEPK025 (KETK127)	Turkana
14	KEND002 (KEKP011)	Kipsigis	62	KEPK030 (KETK128)	Turkana
15	KEND006 (KEKY011)	Keiyo	63	KEPK071 (KETK129)	Turkana
16	KELO002 (KELY051)	Luhya	64	Kesm021 (KETK130)	Turkana
17	KELO004 (KELY052)	Luhya	65	KELY025 (KETS010)	Teso
18	KELO028 (KELY053)	Luhya	66	KESB010 (KETS011)	Teso
19	KELO048 (KELY054)	Luhya	67	KESB037 (KETS012)	Teso
20	KESB018 (KELY055)	Luhya	68	KESB046 (KETS013)	Teso
21	KESB020 (KELY056)	Luhya	69	Kesm011 (KEYK051)	Yaaku
22	KEPK078 (KEMR031)	Marakwet	70	SDNR002 (SDDN041)	Dinka
23	KESN035 (KEMR032)	Marakwet	71	SDNR022 (SDDN042)	Dinka
24	KESN039 (KEMR033)	Marakwet	72	SDNR023 (SDDN043)	Dinka
25	KESN069 (KEND015)	Nandi	73	SDDN006 (SDNR031)	Nuer
26	KETS005 (KEND016)	Nandi	74	TZIQ036 (TZDT091)	Datog
27	KENJ068 (KEPK121)	Pokot	75	TZWF012 (TZDT092)	Datog
28	KESN065 (KEPK122)	Pokot	76	TZWF041 (TZDT093)	Datog
29	Kesm001 (KERD091)	Rendille	77	TZWF028 (TZIQ101)	Iraqw
30	KEYK021 (KERD092)	Rendille	78	Keil002 (TZMS101)	Maasai
31	KESN050 (KESB041)	Sabaot	79	Keil003 (TZMS102)	Maasai
32	KESN052 (KESB042)	Sabaot	80	Keil004 (TZMS103)	Maasai
33	KESN059 (KESB043)	Sabaot	81	Keil006 (TZMS104)	Maasai
34	kemn001 (KESM041)	Samburu	82	Keil007 (TZMS105)	Maasai
35	Kemn008 (KESM042)	Samburu	83	Keil017 (TZMS106)	Maasai
36	Kemn015 (KESM043)	Samburu	84	Keil018 (TZMS107)	Maasai
37	KEYK005 (KESM044)	Samburu	85	Keil020 (TZMS108)	Maasai
38	KEYK018 (KESM045)	Samburu	86	KESB005 (TZMS109)	Maasai
39	KEYK019 (KESM046)	Samburu	87	Kesm009 (TZMS110)	Maasai
40	KEYK023 (KESM047)	Samburu	88	kesm023 (TZMS111)	Maasai
41	KEYK028 (KESM048)	Samburu	89	TZAK030 (TZMS112)	Maasai
42	KEYK034 (KESM049)	Samburu	90	TZAK031 (TZMS113)	Maasai
43	KESB006 (KESN075)	Sengwer	91	TZAK034 (TZMS114)	Maasai
44	KESB009 (KESN076)	Sengwer	92	TZBG032 (TZRG061)	Rangi
45	KESB014 (KESN077)	Sengwer	93	TZWF020 (TZRG062)	Rangi
46	KENJ028 (KETG111)	Tugen	94	TZWF033 (TZRG063)	Rangi
47	KENJ032 (KETG112)	Tugen	95	TZTR020 (TZSW151)	Sandawe
48	KENJ046 (KETG113)	Tugen			

Appendix 9: Detailed descriptions and discussion of mtDNA and Y chromosome lineages observed in Africa

9.0: Introductions

Most genetic lineages are far much older than language families in general and it will be difficult to associate a lineage with a specific language family. In fact, because of the limit in the method of aging languages, all of the African linguistic groupings, besides the Khoisan spoken by hunter-gatherer populations have been estimated to be younger than 20 kya [71, 101, 545], while genetic lineage ages date back to over 100 kya. Overall, emergence of distinct language families seem to correspond to the time period when the transition from the Middle Stone Age (MSA) to the Late Stone Age (LSA) took place [48], 20 – 40 kya. The period is characterized by cultural shift which might have been climate driven **(Appendix 14)**, where initial disparate hunter-gatherer populations (bands made up of related families) started interacting, coming together and forming larger groups [22, 48, 133, 546]. This period also marked the appearance of long distance trading [22, 48, 133]. Ethnic group (or tribe – a higher level of population organization) that is usually made up of integrated bands of kin groups (which are themselves composed of related families) that generally share a common language and cultural norms [546], then expand into ecological areas that yield necessary resources due to population pressure [546, 547]. Over time, tribal populations become differentiated and intergroup competition ensues [546, 547].

The genetic lineages may have already been spatially demarcated due to isolation by distance among disparate hunter-gatherer populations. Linguistic innovations might have arisen by banding of hunter-gatherer populations that lived in geographical vicinity. Subsequent linguistic differentiation might have followed pre-existing geographically

enforced differentiation. And since, Pleistocene human migration might have been culturally mediated [548] meaning that movement of populations were initially "restricted" within a cultural sphere that might have been defined by attributes like language, rituals and socio-economic activities. The tendency to interact within a limited sphere might have maintained cultural defined population structure (thus a reflection of restricted distribution of lineages among the language families and geography **(Appendix 15)** [548]. Subsequent population expansion during the Holocene, due to a change in subsistence pattern shifted, increased or in some cases decreased the range of area occupied by populations from different language families. These migrations might have also led to geneflow between African populations. Thus the two major events that occurred due to change in subsistence patterns, namely the Bantu migration and the shift to pastoralism, appear to have necessitated expansion and movement in search of fresh land for tilling and fresh pasture/water respectively, and thus have had major influences on extant genetic patterns in Africa.

In this section, details as appertains to the frequency observed in African, Mediterranean and Near eastern populations for the Y chromosome and *mt*DNA lineages (that are mainly observed in African populations) are discussed. For each of the lineage, speculation on a possible origin based on haplotype/haplogroup interpolation maps and related archaeological and linguistic evidence are done. The haplotype/halogroup interpolation maps are representation of the geographic spread of haplogroup/haplotype, which might be a reflection of its spatial expansion from its center of origin especially when the expanding/migrating population (that carried the specific lineage) moved rapidly with some admixture with populations they encountered. This is because it is

expected that the expansion of a population into an occupied territory leads to progressive introgression of lineages of the local populations that they encounter and that a gradient of lineage frequencies would develop along the expansion axis [145, 549], a phenomenon that is the basis of the theory of demic diffusion [550-553], which was developed in the context of the colonization of Europe by Neolithic farmers to explain gradients of genetic diversity between the Near East and the north Europe [550-553]. Alternatively the frequency gradient might be due to isolation by distance, whereby there is gene flow at equilibrium only among neighboring populations [554, 555]. In either case migration, that is "geographical displacement of populations" [556] due to maybe either overpopulation or environmental deterioration because of climatic changes (movement in a need to access resources elsewhere) [557], will lead to interbreeding of individuals in populations (with different lineage frequencies in different places) that are from different language family/or geographical area. However, for most lineages it might be difficult to infer their origin from global/continental frequency profiles due to lack of any clear pattern. This might be because of the phenomenon of "surfing" [558, 559], which is basically genetic drift [560], where new mutations (substitute lineage in this case because a mutation arises that define a lineage) occurring at the front of a range expansion travel with the wave of advance and be carried over long distances.

The haplotype/halogroup interpolation maps for Y chromosome and *mt*DNA haplotypes were done by displaying frequency distributions of the lineages as a contour map using the Kriging algorithm of Surfer 8 (Golden Software®). This was done using the geographic Locations of populations shown in **Appendix 1 and Appendix 6** – Longitude as Easting, Latitude as Northing and the frequency of the haplotype as the

elevation. Note that the lineages' ages estimated using all the methods listed in **Appendix 7** correspond to genealogical events and not to demographic events. However, the age estimates are used together with inferred dates of lineage expansions (**Section 3.4.3** and **Appendix 14**) as a pointer to speculate population history (demographic events) by incorporating archaeological and linguistic evidence.

9.1: Y Chromosomes

Haplogroup A

Of the five haplotypes in haplogroup A which have previously been identified in Africa [158] (**Figure 3.3.2):** A1, A2, A3a, A3b1, A3b2, haplotype A3b2 was mainly observed in East African populations. Haplotype A3b2 was observed at low frequencies in a few populations from North Africa [156, 290], the Middle East [156] and the Mediterranean [327, 561, 562] **(Figure 3.3.2 and Appendix 6a**). Overall, haplotype A3b2 was observed at moderate to high frequency in most Nilo-Saharan and Afroasiatic populations of East Africa, with the frequency maximum centered in southern Sudan (**Figure A9.1.1**). The remaining A haplotypes were observed by previous studies in populations from either southern [293, 294] and Central/West Africa [38]. Haplotypes A2 and A3b1, defined by the P3 and M51 SNP **(Figure 3.3.2)**, respectively, are observed among the Khoisan-speaking populations of South Africa (SAK), with the highest frequency observed in the Tsumkwe, San and Nama populations of Namibia [38] **(Appendix 6a, Table 3.3.2)**. Haplotype A1 is observed at low frequency in a few Central/West African [293, 294] populations and has also been reported in some individuals in Yorkshire, England [293]. This is likely a genetic legacy of the Trans-Atlantic slave trade [293]. The A1 haplotype does not show any obvious frequency maximum and the current distribution may be as a result of genetic drift or recent large scale population dispersion (migration). Based on deep lineages inferred from TMRCA estimate for the A haplogroup, the haplogroup might represent a set of haplotypes present in an ancestral population prior to the initial split that gave rise to the three African regional populations: East, Central/West and South African.

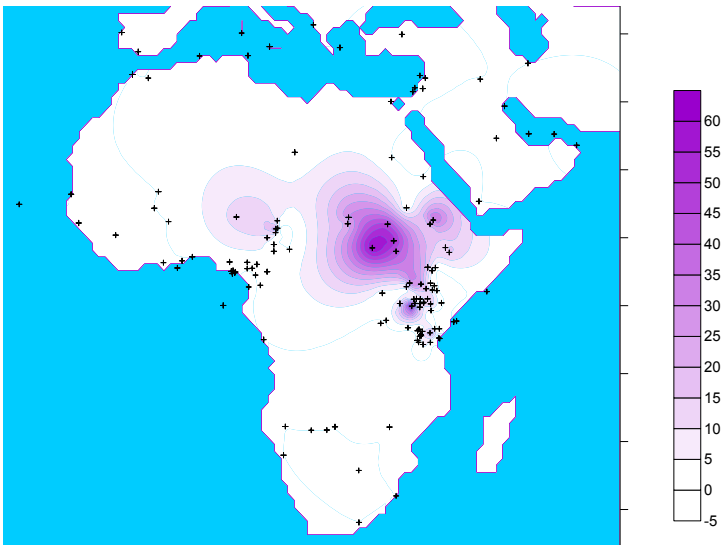


Figure A9.1.1: Contour maps representing geographic distributions of A3b2 frequencies across Africa created using the Kriging method in Surfer 8 (Golden Software). Areas of high frequency are represented by darker tone of the color used in the Map. Crosses represent approximate locations of sample collection for populations genotyped in this study or reported site of collection for published data. In cases where only country of sampling but no specific site in the country is listed for published data, capital cities of the country are used as the sampling locations.

Haplogroup B

Previous studies have shown that B*/B1 and B2* haplotypes are observed at low frequency among a few populations; specifically Khoisan-speaking (Nama and Hadza), and pygmy (Bakola and Biaka) populations [38, 294] (**Appendix 6a)**. Other studies [38] also indicate that the relatively younger B2a* haplotype defined by the M150 SNP **(Figure 3.3.2, Table 3.3.1)** is observed in Mbuti pygmies from eastern Democratic Republic of Congo (DRC) at low frequency (**Appendix 6a)**. The B2a1 haplotype was mainly observed in East African populations **(Table 3.3.3)**, and in low frequencies in earlier studies among the pygmies [38] and neighboring populations in Cameroon [38, 292], and several southern African Bantu speaking populations [38]. Haplotype B2a2 defined by the M108a SNP (**Figure 3.3.2)** has only been reported at low frequency in

Malian populations [163]. Mali is one of the West African countries where speakers of three out of the eight Songhay languages, a branch of Nilo-Saharan language family, are found [70]. Overall, the distribution patterns of B2a haplotypes seem to correspond with the distribution of Nilo-Saharan speakers and might reflect signature of Nilo-Saharan expansions in Africa (**Figure A9.1.2)**.

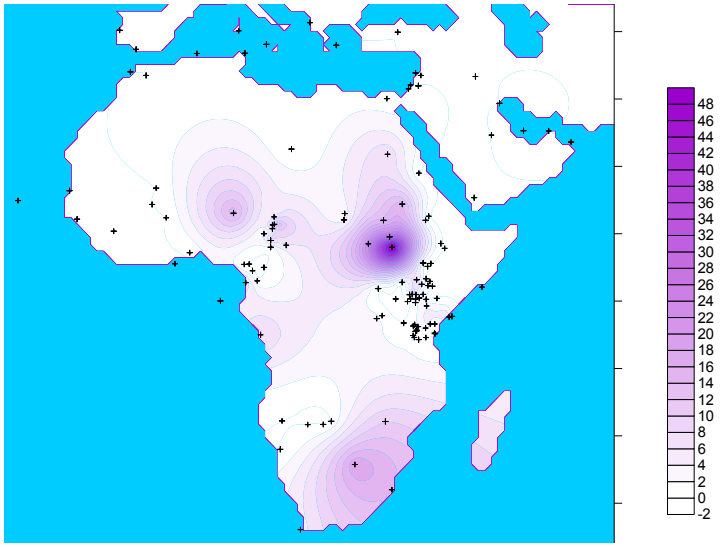


Figure A9.1.2: Contour maps representing geographic distributions of B2a frequencies across Africa created using the Kriging method in Surfer 8 (Golden Software). Areas of high frequency are represented by darker tone of the color used in the Map. Crosses represent approximate locations of sample collection for populations genotyped in this study or reported site of collection for published data. In cases where only country of sampling but no specific site in the country is listed for published data, capital cities of the country are used as the sampling locations.

Haplotype B2b*, which is observed in most hunter-gatherer populations in East, South and Central Africa, are also observed in low frequencies in neighboring populations (**Figure 3.3.2, Appendix 6a)** suggesting gene flow between hunter-gatherer populations and neighboring farming and pastoralist populations. The B2b distribution mostly among the hunter-gatherer populations in Africa (pygmies, Khoisan speakers in

East Africa and southern Africa) might indicate shared ancestry among these hunter-gatherer populations. And, since the age of shared SNP, M112 (B2b) (**Table 3.3.1**) is over 50 kya, it might be an indication that the time depth for divergence between East African hunter-gatherer, pygmy and southern African Khoisan speaking populations is about 50-80 kya (**Table 3.3.1**).

Populations from three geographical regions of sub-Saharan Africa (East, Central/West, and South) share the ancestral B2b* haplotype but also have B haplotypes that are unique to each region: B2b3 (South Africa), B2b2 (East Africa), B*, B1 and B2b1 (Central-West Africa) (**Figure 3.3.2, Appendix 6a)**. However, Central African populations carry two haplotypes, haplotype B2* and B2b4, that are shared with hunter-gatherer populations of southern Africa: for example the B2b4 haplotype is uniquely shared among the Pygmy populations and the Namibian Khoisan speaking !Kung/Sekele and Tsumkwe populations (**Figure 3.3.2, Appendix 6a)**. This observation indicates a possible ancestral relationship between Pygmy and Khoisan speaking populations, a conclusion also inferred from population structure analysis of autosomal markers [4].

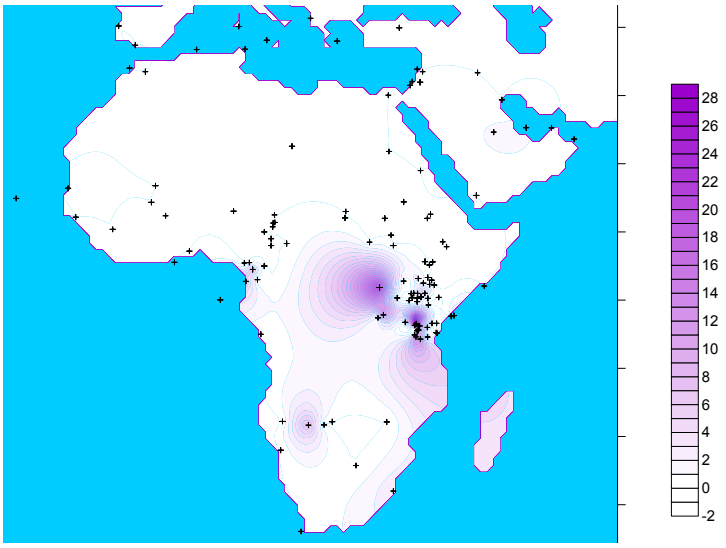


Figure A9.1.3: Contour maps representing geographic distributions of B2b frequencies across Africa created using the Kriging method in Surfer 8 (Golden Software). Areas of high frequency are represented by darker tone of the color used in the Map. Crosses represent approximate locations of sample collection for populations genotyped in this study or reported site of collection for published data. In cases where only country of sampling but no specific site in the country is listed for published data, capital cities of the country are used as the sampling locations.

Haplogroup E

Y chromosome lineages E and F, observed mostly in East and Central/West Africa, might represent human migrations within Africa and out-of-Africa into the Near East (gene flow) in the last 5 – 20 kya [57, 59, 62, 74, 87, 111, 296]. The E2a haplotype has frequency maximum centered in East Africa (**Figure A9.1.4)** (possibly reflected migration of the Nilotic speakers) while the E2b haplotype has frequency profile with its maximum centered in Central Africa, suggesting the distribution pattern of this haplotype might be a signature of the Bantu expansion (**Figure A9.1.5)**.

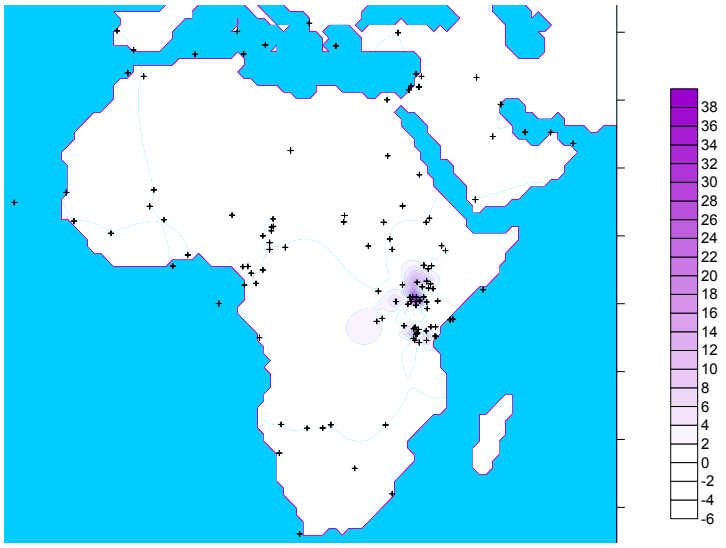


Figure A9.1.4: Contour maps representing geographic distributions of E2a frequencies across Africa created using the Kriging method in Surfer 8 (Golden Software). Areas of high frequency are represented by darker tone of the color used in the Map. Crosses represent approximate locations of sample collection for populations genotyped in this study or reported site of collection for published data. In cases where only country of sampling but no specific site in the country is listed for published data, capital cities of the country are used as the sampling locations.

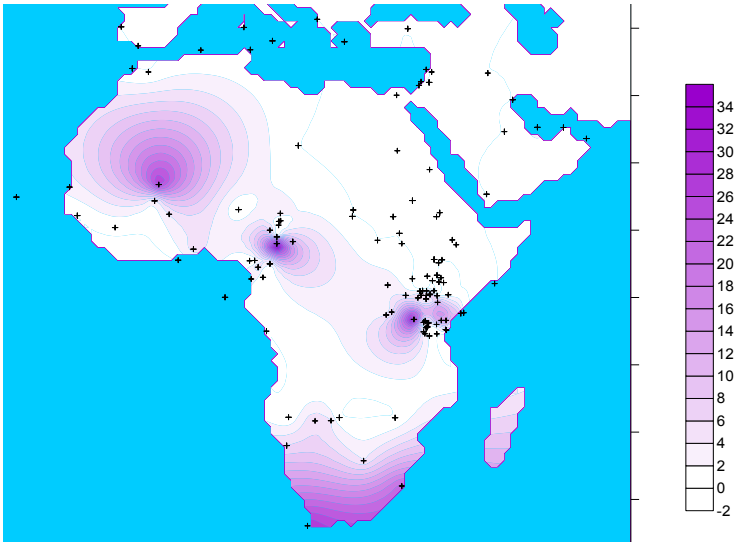


Figure A9.1.5: Contour maps representing geographic distributions of E2b frequencies across Africa created using the Kriging method in Surfer 8 (Golden Software). Areas of high frequency are represented by darker tone of the color used in the Map. Crosses represent approximate locations of sample collection for populations genotyped in this study or reported site of collection for published data. In cases where only country of sampling but no specific site in the country is listed for published data, capital cities of the country are used as the sampling locations.

The E3a haplogroup (E3a1-8 haplotypes) (**Figure 3.3.2**) is thought to represent a signature of Bantu expansion [38, 156]. The ancestral form of this family of haplotypes, E3a* (without mutations that define terminal haplotypes in the phylogeny - **Figure 3.3.2**) is observed among the Niger-Kordofanian speaking populations with highest frequencies observed in Senegal and other West African countries [155, 164] (**Appendix 6a)**. However, the frequency indicated in previous studies [155, 164] might include the recently defined mutation that defines the E3a8 haplotypes [563] (**Figure 3.3.2)**. The E3a haplotypes are observed at low to moderate frequencies in non-Niger-Kordofanian populations which might have had gene exchange with Niger-Kordofanian speaking populations (**Figure 3.3.2, Appendix 6a)**. The E3a7 haplotype has a pattern of decreasing frequency from Central Africa to eastern and southern Africa, making it concordant with a suggested pattern of Bantu expansion originating in Nigeria/Cameroon in Central/West Africa [61, 62] (**Figure 3.3.2, Figure 3.1.4, Appendix 6a)**.

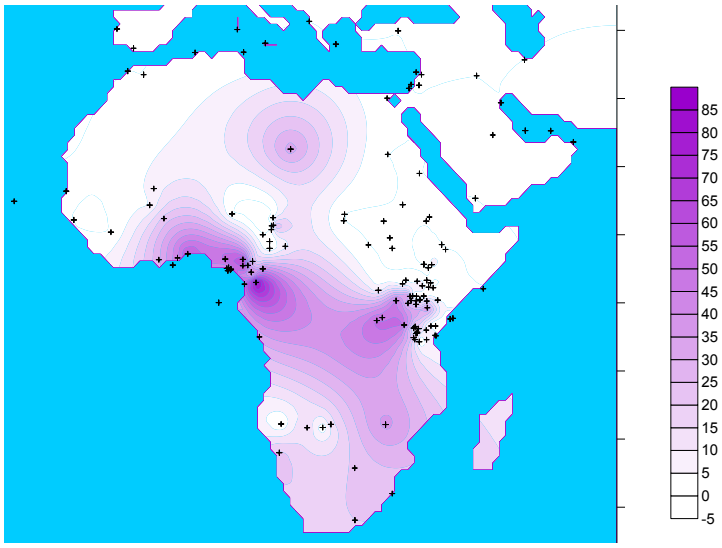


Figure A9.1.6: Contour maps representing geographic distributions of E3a7 frequencies across Africa created using the Kriging method in Surfer 8 (Golden Software). Areas of high frequency are shaded in a darker tone. Crosses represent approximate locations of sample collection for populations genotyped in this study or reported site of collection for published data. In cases where only country of sampling but no specific site in the country is listed for published data, capital cities of the country are used as the sampling locations.

The E3b haplotypes have a different distribution pattern to that of the E3a haplotypes, and are observed mostly among the Afroasiatic and Nilo-Saharan populations from East Africa and at low frequency in neighboring Niger-Kordofanian populations with whom they might have had gene exchange (**Figure 3.3.2, Appendix 6a)**. Based on collated population frequencies from the current study and previous studies (**Figure 3.3.2, Appendix 6a),** the E3b1 haplotype seems to reflect a signature of Afroasiatic expansion, with highest frequencies among the Cushitic speaking populations of East Africa and relatively lower frequencies among Semitic speaking populations [156] (**Figure 3.3.2, Appendix 6a)**. The E3b2 haplotype is at highest frequency among the Berber populations of Morocco and gradually decreases in frequency among Arab populations to the East in Northeast Africa [156, 510, 564] (**Appendix 6a)**.

The E3b3 haplotype, defined by the M123 mutation, was previously suggested to have originated in the Middle East and is found at low frequencies both in the Middle Eastern and East African populations [153, 434] (**Appendix 6a)**. This conclusion was based on the fact that the E3b3 haplotype was observed only in Ethiopian samples from among the nine East African populations analyzed in a previous study [153], and lower STR variances in the Ethiopian E3b3 samples than in the Middle East [153]. However, extensive analysis of East African populations in the current study shows that this haplotype is found in Kenyan and Tanzanian Cushitic speakers as well, albeit, at low frequency (**Figure 3.3.2, Appendix 6a)**. Its frequency maximum is centered in northeastern Africa **(Table A9.1.1, Figure A9.1.7)**. Considering that the highest frequency is observed among the Ethiopian Jews [164] **(Appendix 6a)** a population that has been shown to be paternally [505, 565] and maternally [219] distinct from other Jewish populations, and genetically most similar to Sub-Saharan Africans [219, 505, 565], and the highest variance is observed in African populations **(Table A9.1.2)**, the origin of E3b3 will most probably be among Cushitic/Omotic speaking populations of Southwest and Central Ethiopia. This is because the Ethiopian Jews have been shown to be Cushitic speakers according to historical [566, 567], ethno-musicological [568] and linguistic [569] studies. They are western Cushitic speaking Agaw population who started adopting the Judaic tradition about 2 kya [566-569]. They were later joined by some parts of the people that rebelled from the Christian practicing Ethiopian Semitic Axumites [566-568]. By about 0.8 kya, they had gradually assumed an ethnic identity consisting of a unique form of Judaism that combines Judaic teaching and some Axumite orthodox rituals [566-568].

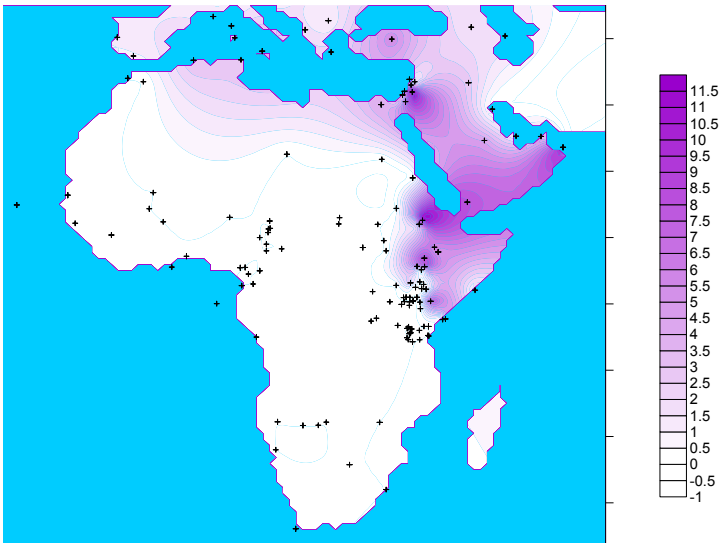


Figure A9.1.7: Contour map representing geographic distributions of Y chromosome E3b3 haplotype frequencies across Africa, the Near East and Europe created using the Kriging method in Surfer 8 (Golden Software). Areas of high frequency are represented by darker tone of the color used in the Map. Crosses represent approximate locations of sample collection for populations genotyped in this study or reported site of collection for published data. In cases where only country of sampling but no specific site in the country is listed for published data, capital cities of the country are used as the sampling locations.

Table A9.1.1: Frequency pattern of Y chromosome E3b3 haplotype. Columns show the language family then name (country) of each of the populations, number of individuals tested, frequency of E3b3 haplotype in the populations, geographic location where the population is sampled from (longitude and latitude), and the primary source reference. **A –** this study, [a]From Moldova.

Linguistic	Population	N	E3b3	Longitude	Latitude	Sources
Afro-asiatic	Oman	147	10.2	58.59	23.61	[153, 156]
Afro-asiatic	Emirates	164	3.1	55.28	25.25	[513]
Afro-asiatic	Qataris	72	1.4	51.53	25.29	[513]
Altaic	Azeris	97	2.1	49.88	40.4	[153]
Afro-asiatic	Somali	205	0.5	45.37	2.07	[504]
Indo-European	Georgians	66	1.5	44.79	41.73	[522]
Afro-asiatic	Iraqi	218	2.8	44.39	33.34	[434]
Afro-asiatic	Yemen	62	8.1	44.21	15.35	[513]
Afro-asiatic	Boni	21	4.76	40.92	-2.27	**A**
Afro-asiatic	Arsi	85	5	39.83	7.83	[291]
Afro-asiatic	Oromo	103	5.8	39.27	8.55	[153, 155]
Afro-asiatic	Yaaku	24	8.33	38.69	0.4	**A**
Afro-asiatic	Burji	23	8.7	37.78	5.6	**A**
Afro-asiatic	Wolayta	12	8.3	37.75	6.9	[153]
Afro-asiatic	Ethiopian Jews	39	12.82	37.47	12.6	[153, 164, 505]
Afro-asiatic	Amharas	82	11	37	12	[153, 155]
	Ethiopians	95	6	36.65	5.65	[291]
	Mixed Ethiopians	88	2.3	36.65	5.65	[14, 155, 355]
Afro-asiatic	Jordanians	146	13.01	35.93	31.95	[516]
Afro-asiatic	Druze	28	3.6	35.75	33	[153]
Nilo-Saharan	Turkana	32	3.125	35.67	3.17	**A**
Afro-asiatic	Lebanese	914	4.27	35.51	33.87	[515]
Afro-asiatic	Bedouins	28	7.1	34.92	30.5	[153]
Afro-asiatic	Jews	952	8.8	34.76	32.07	[519]
Altaic	Turks	523	5.54	32.86	39.93	[521]
Afro-asiatic	Egyptians	239	4.18	31.25	30.05	[38, 156]
Indo-European	Ukrainians	93	1.1	30.52	50.43	[434]
Altaic	Gaugazes[a]	89	2.2	28.86	47.01	[570]
Indo-European	Greeks	93	2.2	23.73	37.98	[522]
Indo-European	Bulgarians	116	0.9	23.32	42.7	[153]
Indo-European	Albanians	119	1.68	19.82	41.33	[522]
Indo-European	Hungarians	53	1.9	19.08	47.5	[434]
Indo-European	Croats	89	1.1	16	45.8	[522]; [434]
Indo-European	Sicilians	236	4.66	13.37	38.12	[523]
Afro-asiatic	Tunisians	148	2.7	10.18	36.8	[510]
Indo-European	Sardinians	376	2.93	9.25	40.12	[327]
Indo-European	Corsicans	140	1.4	8.73	41.92	[153]
Indo-European	Asturians	90	1.1	6	43.33	[153]
Afro-asiatic	Algerian	32	3.1	3.05	36.75	[434]
Indo-European	Spanish	978	0.92	-5.99	37.38	[524]
Indo-European	Portuguese	795	1.26	-8.42	40.2	[524, 525]

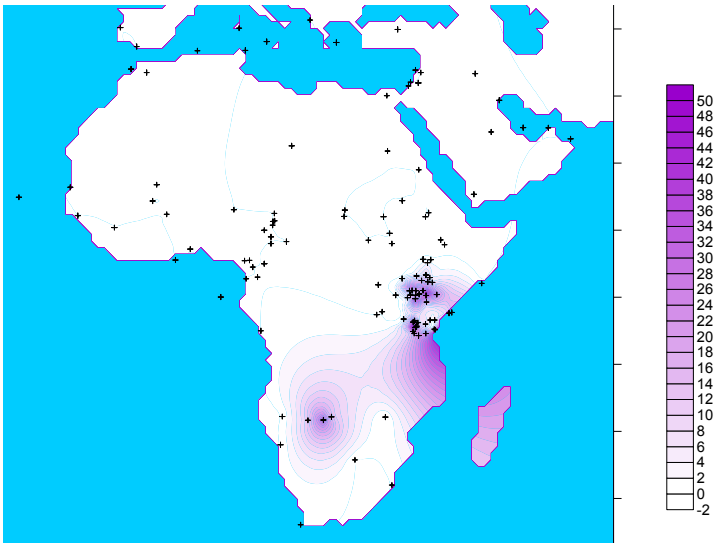


Figure A9.1.8: Contour map representing the geographic distribution of Y chromosomal E3b6 haplotype frequencies across Africa created using the Kriging method in Surfer 8 (Golden Software). Areas of high frequency are represented by darker tones of color. Crosses represent approximate locations of sample collection for populations genotyped in this study or reported site of collection for published data. In cases where only country of sampling but no specific site in the country is listed for published data, capital cities of the country are used as the sampling locations.

The E3b6 haplotype, which potentially shows a signature of pastoralist movement from East Africa [355] to southern Africa, has the highest frequency among the southern Cushitic speaking populations of East Africa (**Figure 3.3.2, Appendix 6a)**. The lineage has two frequency maxima, one centered in East Africa and the other centered in southern Africa (**Figure A9.1.8)**.

The ancestral E3* haplotype is found at maximum frequency in North East Africa, while E3a and E3b have frequency maxima in Central/West and East Africa respectively. Since the E3 haplogroup includes most of the Y chromosome lineages observed in Africa (60% in this study), this pattern might be an indication of human population expansion in Africa originating in North East Africa (after the original splits in population that yielded African regional populations – East, Central/West and South

Africa as indicated by haplotypes A and B above – **Appendix 15**), and a subsequent expansion from Central/West and East/North-east Africa, respectively. The expansions from East/North-east and Central/West Africa may correspond to Afroasiatic (and pastoralism) and Bantu expansion respectively. The TMRCA estimate for all E3 chromosomes is 38.1+/- 8.1 **(Table 3.3.1)** and, thus suggests that the ancestral Y lineages in proto-population that gave rise to current Niger-Kordofanian (Central and West Africa) and Nilo-Saharan and Afroasiatic speaking populations might have split 30 kya.

Haplogroup F

The F haplotypes (J, K and R) in Africa are mostly observed in North-East Africa and Central Africa (**Figure 3.3.2, Appendix 6a)** and have been suggested to be due to back migrations from Eurasia [155, 164]**.** The expansion of Haplogroup J into Europe, especially eastern Europe, has been suggested as a signature of early expansions of the early farming populations from the Fertile Crescent [434, 571] since the overall pattern of the J haplogroup shows a frequency maximum in and around that area [156, 504, 512, 571, 572] **(Appendix 6a)**. Furthermore, it gradually decreases in frequency in regions Northwest and Southwest of the Fertile Crescent (North and West Africa, East Africa and Italy) with especially low frequencies occurring among the Berbers in Morocco and Sandawe in Tanzania (**Figure 3.3.2, Appendix 6a)**.

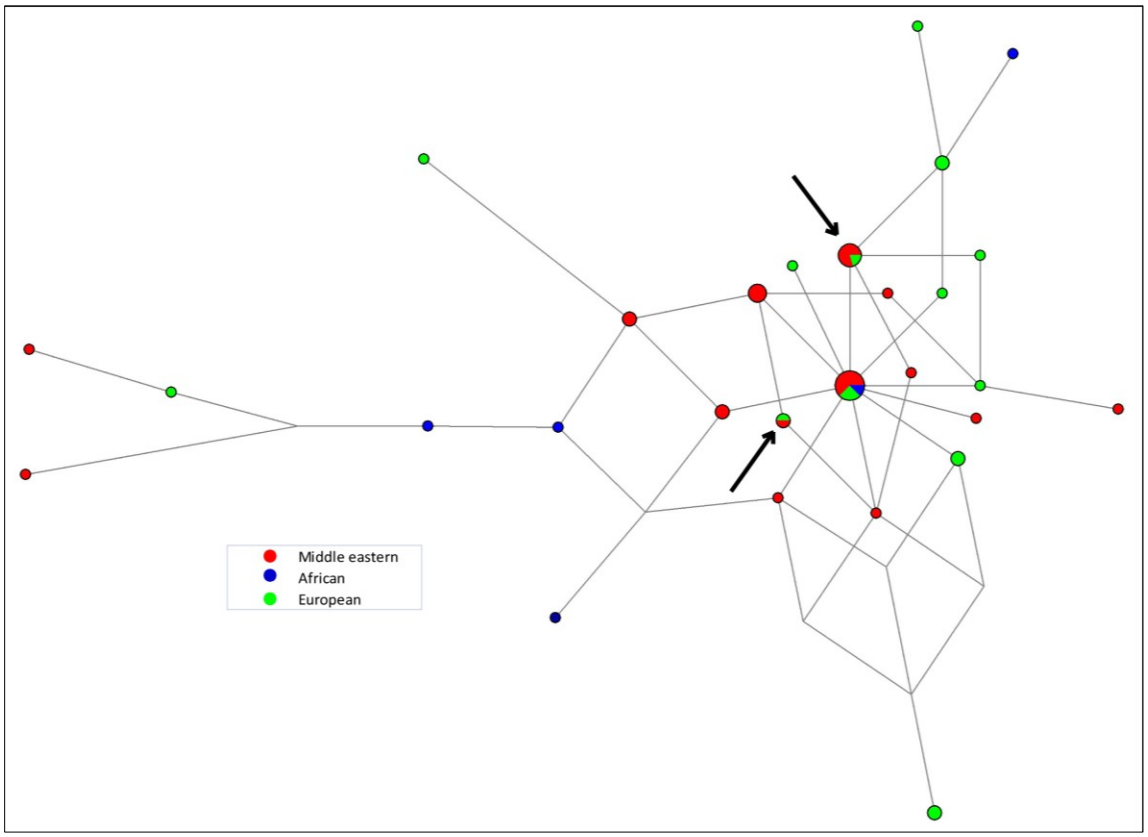
The other Y chromosome haplotypes in Africa that have been suggested to have been due to back migration from West Eurasia are K2 and R1b* [38, 573]. The ancestral haplogroups from which K2 and R1b* are derived, K and R respectively, have their highest frequencies in South [574, 575] and Southeast Asia (R) and in Melanesian

populations (K) [572]. However, the distribution of R1 (R1a and R1b) is restricted to Africa, Europe, the Middle East and part of Central Asia, and this is precisely where the K2 haplotype has been detected at low frequency (**Figure 3.3.2, Appendix 6a)** [576]. On the other hand, the majority of European R1b chromosomes carry the marker M269, which is absent in most Africa populations that belong to R1b haplotype. In fact a recent study [431] showed that the geographical distribution of R1b1b2 (R-M269) microsatellite diversity is best explained by spread from a single source in the Near East via Anatolia due to the Neolithic expansion [431].

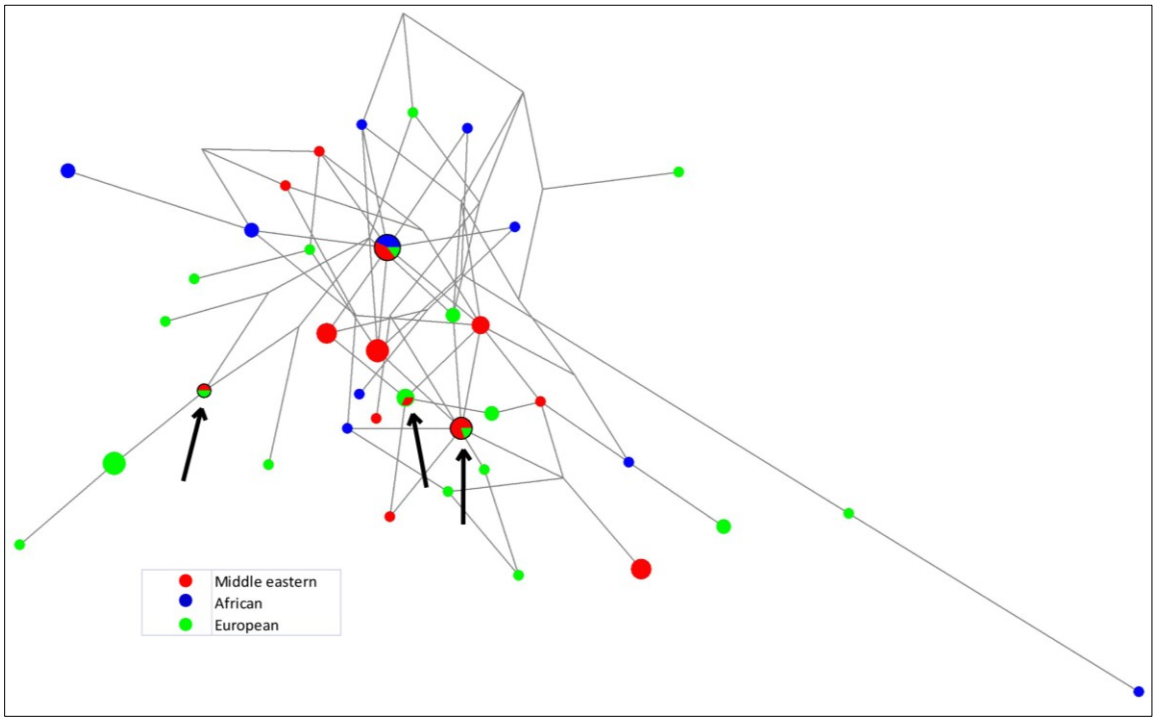
A recent study of Y chromosome lineages in western and southern Cameroonian populations showed low to moderate frequencies of R1b* that extend into Gabon [294]. Moreover, a recent fine scale sequencing/genotyping of terminal markers in R1b haplotype observed R-V88 sub-haplotype in virtually all the Africans that belong to R1b haplotype [573]. The current study confirms the results of previous work [38, 294] that demonstrates that the R1b* haplotype (meaning R-V88 sub-haplotype) is at its highest frequency in Chadic speaking northern Cameroonian populations (**Figure 3.3.2, Appendix 6a)**. However, the haplotype was not found in any population in East Africa (**Figure 3.3.2, Appendix 6a)**. Berniell-Lee *et al.,* [294] concluded that the frequency pattern of the R1b* haplotype within Africa, Central Africa in particular, might reflect the Bantu expansion. Considering that the sister haplotype that carries the M269 SNP has been reported at low frequency in populations in Angola and Namibia [38], the profile may reflect a signature of a western route of Bantu expansion from their putative homeland in Central Africa into southern Africa [61, 62].

Figure A9.1.9: Median-joining network of 8-loci (DYS389I, DYS458, DYS437, DYS448, DYS391, DYS392, DYS393 and DYS635) Y-STR haplotypes for a) E3b3 b) K2 (T-M70) positive individuals in East African populations. Network was generated using Network 4.1.1.1 [80]. Networks were processed first by the reduced-median method, and then by the median-joining method [81, 82] without weighting any of the STR loci. Areas of circles are proportional to absolute frequencies and colored according to which language family/branch an individual is from. In both E3b3 and K2 lineages, individuals sharing microsatellite background are observed between the three regions, Africa, Europe and the Middle East, but also there is background sharing between Europeans and Middle Easterners **(pointed out by arrows in the figures)**.

a)



b)



Region	E3b3		K2 (T-M70)	
	n	Variance	n	Variance
Africa	6	0.54	29	0.76
Near East	30	0.14	30	0.18
Europe	22	0.47	35	0.47

Table A9.1.2: Variation in E3b3 and T-M70 Y chromosome haplotypes in the three main regions where the lineages are observed

Y chromosome haplotype results indicate that East African populations have the derived variants of the most ancestral lineages (A3b2, B2) and represent a transition between Sub-Saharan Africa and the near east (as exemplified by haplotypes E3b, and K/R as argued in the discussion section). There is also some evidence of gene flow from the Near East (J) into northeast and eastern Africa.

Appendix 9.2: mtDNA

Haplogroup L0

There is a unique distribution pattern of L0 haplotypes with L0a, L0b and L0f mostly observed in East and Central/West Africa, while L0d and L0k are observed mostly in South African San populations (**Figure 3.4.1, Appendix 6b)**. Haplotype L0b was previously defined in an individual from Ethiopia [157] by complete *mt*DNA genome sequencing. In the current study the L0b haplotype was found at low frequency among the Luhya, Sengwer and Wata populations from Kenya (**Figure 3.4.1, Appendix 6b)**. Moreover, previous studies [151, 152, 526] reported the L0b haplotype in low to moderate frequencies in populations from northern Kenya (Turkana [151, 152] and Daasanech [526]) and Southwestern Ethiopia (Nyangatom [526]) **(Appendix 6b)**. The Luhya, a Bantu speaking population, has had extensive interaction with southern Nilotic speaking populations and has adopted some of their cultural norms and myths [89]. The Sengwer are a hunter-gatherer population that speaks a southern Nilotic language, a branch whose speakers have had extensive interaction with eastern Cushitic populations before expanding into Kenya 3 kya [87]. The Turkana (Nilo-Saharan) and Daasanech (Afroasiatic) populations live in the northern part of Kenya and continue to have interaction with populations from Southeastern Sudan as well as southwestern Ethiopia through intertribal cattle raids and trade [577]. Therefore, it is possible that the L0b lineage originated in the general area of southwestern Ethiopia.

The distribution of the L0f haplotype is mainly restricted to East African populations mostly among the Afroasiatic speaking hunter-gatherer populations of East Africa (**Figure 3.4.1, Appendix 6b)**. Three sub-haplotypes have so far been identified

within the L0f haplotype, L0f1-3 **(Figure A9.2.2, Appendix 8)**. In addition to L0f defining control region motif 16169-16327-16354-16368-207, L0f1 have two additional HVRII mutations at mitochondrial genome nucleotide positions 146 and 151 (**Figure A9.2.2, Appendix 8).** The L0f1 sub-haplotype was mostly observed among the hunter-gatherer populations from southern Kenya (Boni and Sanye), Khoisan speaking Sandawe and neighboring populations from northern Tanzania **(Appendix 8, Table A9.2.1)**. However, the other two sub-haplotypes, L0f2 and L0f3, besides being observed in moderate to high frequencies among the southern Cushitic speakers, is observed in low frequencies among the eastern Nilotic, eastern Cushitic and hunter-gatherer populations from northern Kenya **(Appendix 8, Table A9.2.1, Figure A9.2.3)**. All the individuals who carry the L0f haplotype outside Kenya and Tanzania (beside the ones observed in South Africa [157]) have the L0f2 sub-haplotype **(Table A9.2.1, Figure A9.2.3)** [16, 154, 219, 305, 314, 349, 402, 526]. The frequency pattern of the L0f haplotype, mostly observed among the southern Cushitic and East African hunter-gatherer populations, attests to the longstanding interaction between southern Cushitic speaking populations and East African hunter-gatherer populations [57, 115]. Behar [157] found two individuals that carry the L0f1 lineage in South African populations **(Figure A9.2.2)**. Moreover, the L0f haplotype is observed at low frequencies in the South African Khoisan speaking Kung population [183, 193] (**Appendix 6b, Table A9.2.1)**. This probably indicates that there has been recent geneflow into South Africa from East African populations (southern Cushitic speaking and/or East African hunter-gatherer populations).

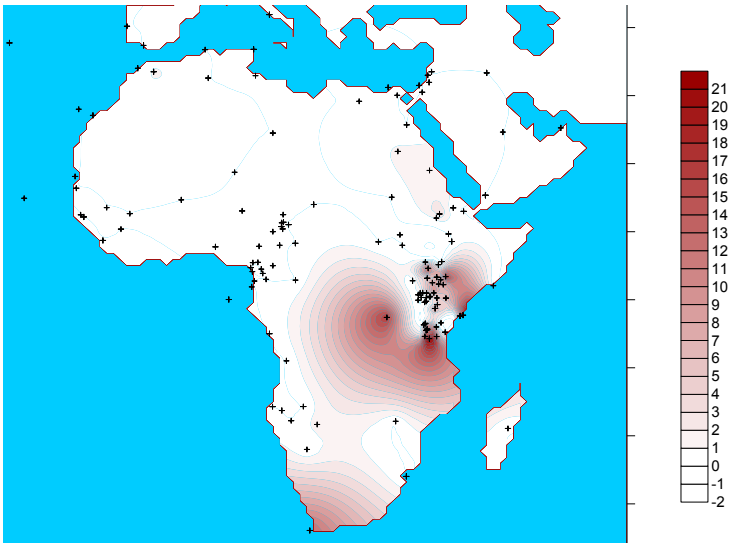


Figure A9.2.1: Contour map representing geographic distributions of mitochondrial L0f haplotype frequencies across Africa, the Near East and Europe created using the Kriging method in Surfer 8 (Golden Software). Areas of high frequency are represented by darker tone of the color used in the Map. Crosses represent approximate locations of sample collection for populations genotyped in this study or reported site of collection for published data. In cases where only country of sampling but no specific site in the country is listed for published data, capital cities of the country are used as the sampling locations.

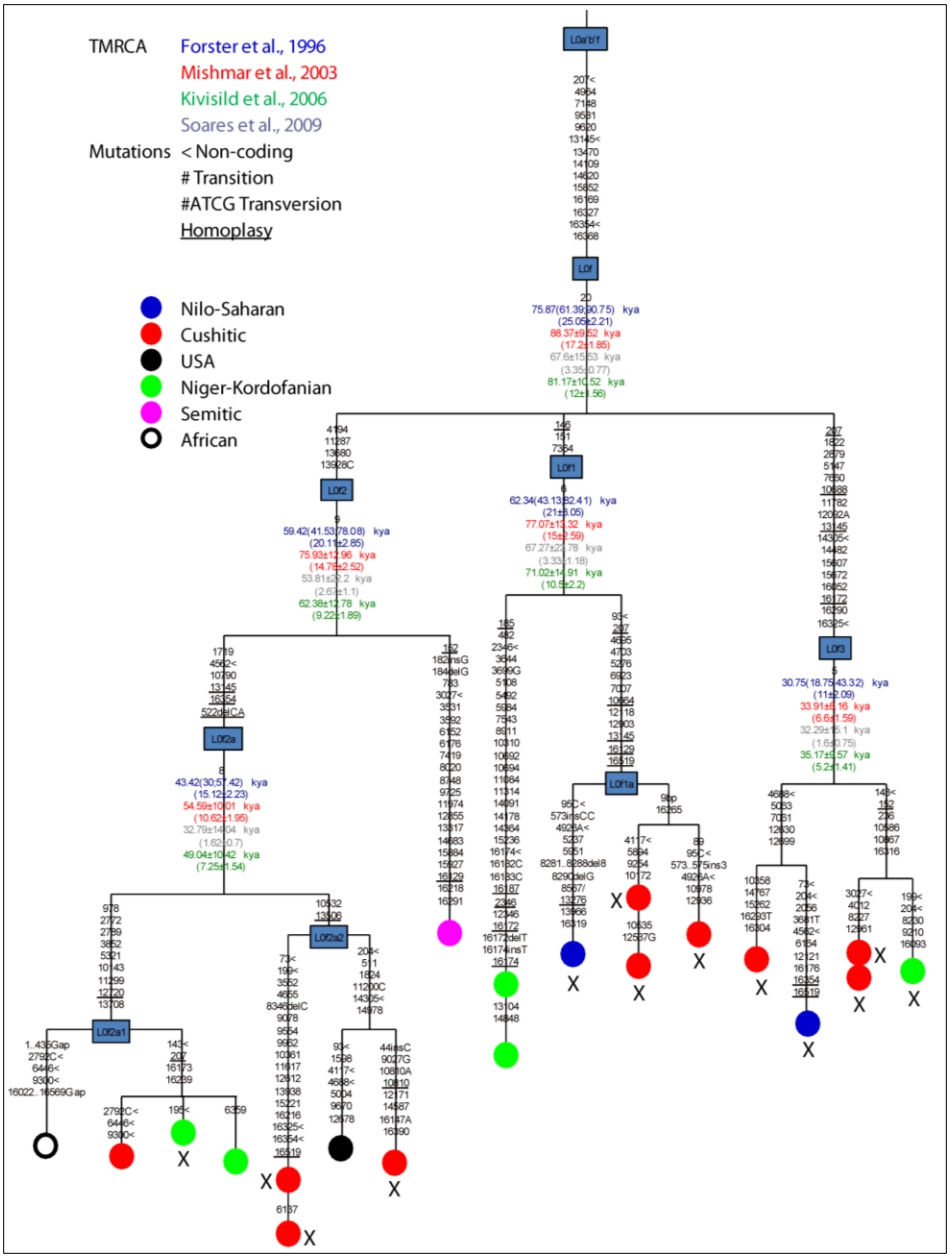


Figure A9.2.2: Phylogeny of *mt*DNA complete genome sequences that belong to the L0f lineage. Samples marked **X** were sequenced in this study. Linguistic group/family of the population/s the sequences were sampled from in the current study, and published sequences whose linguistic affiliations are known are shown. For published data where linguistic affiliations are not known, the country/region in which the sequences are sampled is indicated. The frequency distribution of the lineage in African and non-African populations are shown in **Table A9.2.1, Appendix 6b.**

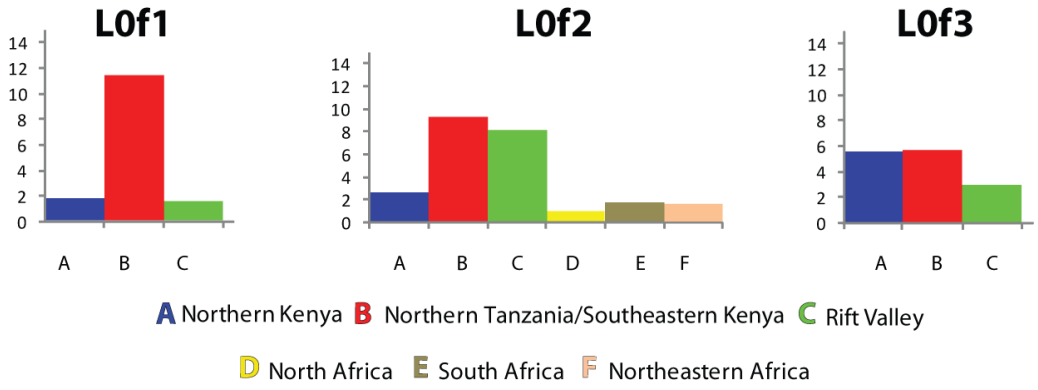


Figure A9.2.3: Histogram representing frequency distribution of L0f subhaplotypes in the *ad hoc* sub-region classifications within East Africa (Kenya and Tanzania) and other part of African continent. The sub-regions within East Africa are based on the fact that descendants of populations that represent migrations into East Africa within the last 5 ky settled mainly in specific regions; Southern Cushitic speakers who are represented by populations from southeastern Kenya and Northern Tanzania, southern Nilotic speakers are represented by Kalenjin populations who mainly found in the Rift valley region of Kenya, and Northern Kenya is occupied by later Cushitic migrants [38, 101-103]. The figure was drawn from information in the table above (**Table A9.2.1).**

Table A9.2.1: Frequency of L0f sub-haplotypes in African populations. Columns show the name (country) of each of the populations, number of individuals tested, frequency of three L0f subhaplotypes in the population, where the population is sampled from (region of country or continent), and the primary source reference. **A;** This Study, Afroasiatic, Nilo-Saharan, Niger-Kordofanian, Khoisan. Egyptians samples are from city of Alexandria.

Population	n	L0f1	L0f2	L0f3	Location	Source
Hutu	42		16.67		eastern Africa	[349]
Ethiopian Jews	29		3.45		eastern Africa	[219]
Amhara	120		0.83		eastern Africa	[16]
Nubians	80		1.25		eastern Africa	[154, 314]
Beja	51		1.96		eastern Africa	A
Gabra	32		3.13	3.13	northern Kenya	A
El-Molo	24		8.33		northern Kenya	A
Il'Laikipikiak	25		4		northern Kenya	A
Rendille	31		3.23	3.23	northern Kenya	A
Wata	24		4.17		northern Kenya	A
Yaaku	19			5.26	northern Kenya	A
Daasanech	49		2.04	14.29	northern Kenya	[526]
Nyangatom	112		0.89		northern Kenya	[526]
Turkana	86	1.16		2.33	northern Kenya	A; [151, 152, 526]
Samburu	30	3.33		6.67	northern Kenya	A
Borana	59	1.69	1.69		northern Kenya	A
Boni	30	16.67	6.67	3.33	Southeastern Kenya	A
Sanye	20	10			Southeastern Kenya	A
Taita	29		10.34	3.45	Southeastern Kenya	A
Burunge	25	24	4		northern Tanzania	A
Rangi	27	3.7		11.11	northern Tanzania	A
Sandawe	25	8			northern Tanzania	A
Turu	24	4.17	8.33		northern Tanzania	A
Maasai	41			2.44	northern Tanzania	A
Akie	20		10		northern Tanzania	A
Iraqw	23		8.7	8.7	northern Tanzania	A
Fyome	13		23.08	7.69	northern Tanzania	A
Datog	30			6.67	northern Tanzania	A
Berbers	327		1.22		North Africa	[401, 402, 537]
Egyptians	277		0.36		North Africa	[305]
Tugen	43	2.33			Rift Valley	A
Marakwet	22			4.55	Rift Valley	A
Sengwer	27			3.7	Rift Valley	A
Gikuyu	50		2	2	Rift Valley	A; [151, 152]
Nairobi	100	1	4	3	Rift Valley	[366]
Kung	62		1.61		South Africa	[183, 193]

Haplotypes L0d and L0k are mostly observed among South African Khoisan **(Table 3.4.2)**, South African colored populations and neighboring Bantu speaking populations [185, 295, 297] and have been speculated to have originated among the proto-Khoisan populations [185]. L0d (L0d1 & L0d2) are observed in low frequencies in most of the Bantu speaking populations in southern and southeastern Africa **(Table 3.4.3)** [185, 297, 298, 349]. The results from *mt*DNA studies are consistent with widespread borrowing into southern African Bantu language phonological systems of the click consonants which is inherent in Khoisan languages [578]. In fact, around 5% of population samples representative of Bantu speaking populations from Mozambique [185, 298] and Angola [297] belong to L0d haplotypes (L0d1 & L0d2) **(Appendix 6b)**. Interestingly, the L0d haplotype was observed at low frequency among the Makonde, a Bantu speaking population that is found in northern Mozambique and southern Tanzania [185]. This observation might indicate that the range occupied by current South African Khoisan speaking populations (currently found in South Africa, Botswana and Namibia) likely extended up to southern Tanzania/northern Mozambique and remnants of these populations might have been absorbed by expanding Bantu speakers in the last 1 kya [62, 64, 73]. This contention is consistent with Ehret's suggestion that there might have been Khoisan speakers living in the area currently occupied by Makua speaking populations (Bantu speakers in southern Tanzania and northern Mozambique), because Makua consonant phonology is different from that of most Bantu languages and has some resemblance to the Khoisan language consonant system **(Ehret Unpublished)**. Ehret's assertion is consistent with the studies of serum protein that also indicate that the population that exhibit south African Khoisan admixtures extend from South

Africa/Botswana/Namibia (countries Khoisan speakers are currently found) into Zambia and Malawi [579].

Besides being observed among the southeastern Kenya and northern Tanzanian hunter gatherers, the younger L0d3 haplotype was observed in low frequencies in the Nilo-Saharan speaking Turkana population [151, 152] and the Bantu speaking population of Ronga (from southern Mozambique and Swaziland) [185, 298] **(Appendix 6b)**. Considering that most individuals found to carry this haplotype belong to East African hunter-gatherer and neighboring populations (**Figure 3.4.1)**, the distribution pattern of L0d3 might represent a signature of the initial pastoralist (who might have been mainly southern Cushitic speakers after long interaction with east African hunter-gatherers) movement from East Africa into southern Africa [59]. Alternatively, might represent a signature of the expansion of East African hunter-gatherer populations during the initial movement of livestock into southern Africa [59], after picking up cattle rearing traditions in East Africa from the pastoralist Cushitic speaking neighbors [110].

Haplotype L0a1 is the most widely distributed L0 haplotype in Africa with varied regional frequencies of its sub-haplotypes L0a1a-L0a1d in Africa, the Caribbean and among African Americans [180, 299-301]. L0a1 haplotype has three major distribution maxima, one in East Africa the second in northeast Africa, and a third in Central/West Africa **(Figure A9.2.4)** though the latter maximum is somewhat diffuse. This distribution pattern corresponds with that observed for L0a1 clades **(Figure A9.2.5)**; L0a1a and L0a1b are observed mostly in Central African [157, 292, 351] and East African Niger-Kordofanian speaking populations **(Appendix 8).** Moreover, all of the L0a1 lineages observed in the Caribbean and Americas [180, 226, 299-301, 580] belong to the L0a1a

(and to a lesser extent to the L0a1b) sub-haplotype. Sub-haplotype L0a1c is observed among the Nilo-Saharan speaking populations while L0a1d sub-haplotype is observed among both East African Nilo-Saharan and Afroasiatic speaking populations **(Appendix 8)**. Therefore, L0a1 might represent a haplotype that was found in proto-populations that gave rise to current populations from the three major linguistic groups in Africa: Niger-Kordofanian, Nilo-Saharan and Afroasiatic.

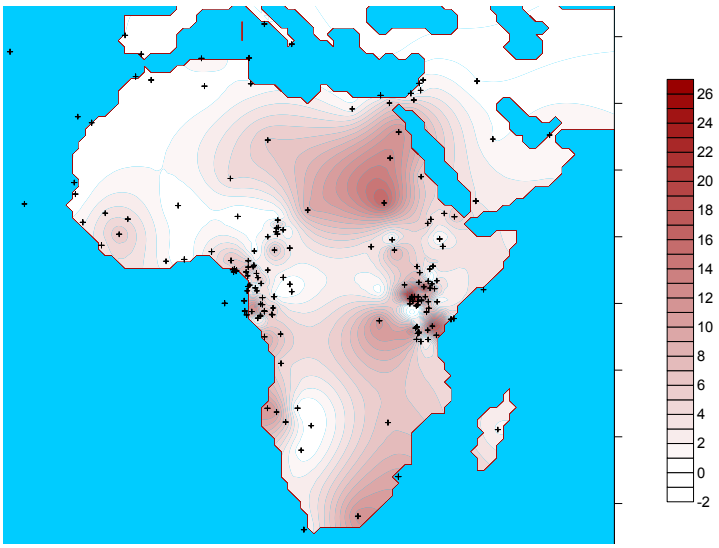
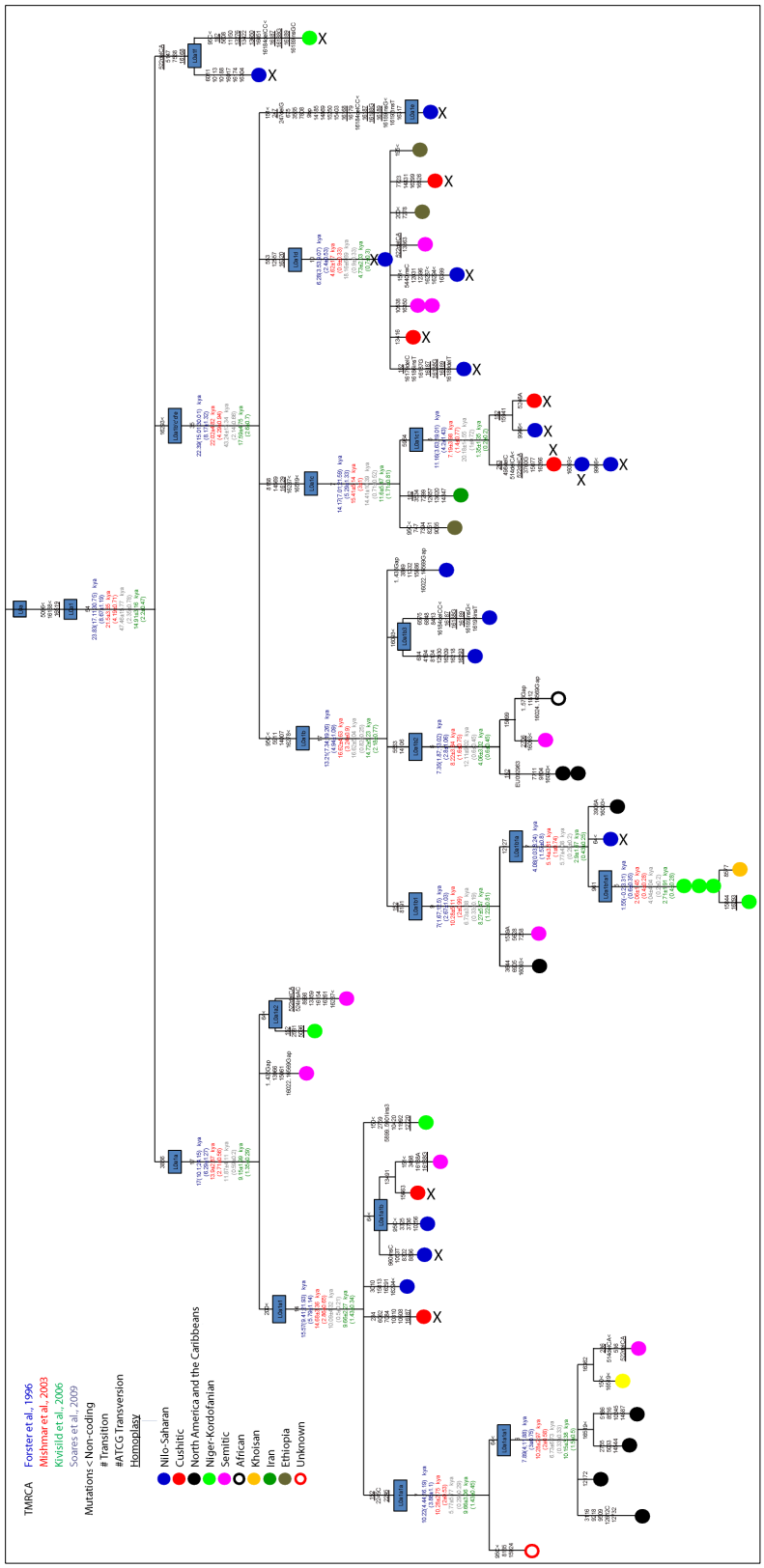


Figure A9.2.4: Contour map representing geographic distributions of L0a1 frequencies across Africa created using the Kriging method in Surfer 8 (Golden Software). Areas of high frequency are represented by darker tone of the color used in the Map. Crosses represent approximate locations of sample collection for populations genotyped in this study or reported site of collection for published data. In cases where only country of sampling but no specific site in the country is listed for published data, capital cities of the country are used as the sampling locations.

Figure A9.2.5: Phylogeny of *mt*DNA complete genome sequences that belong to the L0a1 lineage. Samples marked **X** were sequenced in this study. Linguistic group/family of the population/s the sequences were sampled from in the current study, and published sequences whose linguistic affiliations are known are shown. For published data where linguistic affiliations are not known, the country/region in which the sequences are sampled is indicated. The frequency distribution of the lineage in African and non-African populations are shown in **Appendix 6b.** Sequences that belong to L0a1a and L0a1b are predominantly from Niger-Kordofanian speaking populations or individuals from the Central/West and Northwestern African countries; Morocco, Algeria, Burkina Faso and Chad, and North America and the Caribbeans.


TMRCA
Forster et al., 1996
Mishmar et al., 2003
Kivisild et al., 2006
Soares et al., 2009
Mutations <-Non-coding
# Transition
#ATCG Transversion
Homoplasy
Nilo-Saharan
Cushitic
North America and the Caribbeans
Niger-Kordofanian
Semitic
African
Khoisan
Iran
Ethiopia
Unknown

Prior studies suggested that L0a2 haplotype originated in Central Africa [185] because of it being mainly observed among the Niger-Kordofanian Bantu speaking populations studied. The distribution of the L0a2 sub-haplotype, specifically L0a2a [185, 292, 297, 349], may support the contention that L0a2 might have originated in Central Africa because it is mostly observed among the Niger-Kordofanian speaking populations of East and southern Africa **(Appendix 8)** [185, 297, 298, 349]. Thus the observation of this haplotype in a population in Central American Belize [581] at 11% frequency might indicate that populations from Central Africa (L0a2 is not observed in West African populations) may also have acted as a source population for the trans-Atlantic slave trade. Behar [157] had previously defined L0a2b only among pygmy individuals of Zaire. Besides being observed among the pygmy populations [157], two of the L0a2 sub-haplotypes (L0a2b and L0a2c) are mostly observed among the southern Cushitic speakers and east African hunter-gatherers. However, the distribution pattern of L0a2a might be reflection of movement of Niger-Kordofanian speakers from Central/West Africa in the last 5 kya.

In East Africa, the L0a2 haplotype has nearly the same distribution pattern as L0f with the highest frequency found in southern Cushitic speaking populations and moderate frequencies in southeastern Kenyan hunter-gatherer populations (Boni and Sanye) and northern Tanzanian Bantu populations **(Figure 3.4.1, Appendix 8)**. Most of the East Africans from L0a2 lineages belong to the L0a2c sub-haplotype, which is defined by a T->C transition at base pair position 95, with an average TMRCA age estimate of 37 kya **(Figure 3.2.1.3, Appendix 7c, Appendix 8)**. L0a2b sub-haplotype is observed in low frequencies among the hunter-gatherer populations from northern Tanzania and

southeastern Kenya (Akie, Boni and Hadza, and few other populations; Pare, Ilchamus and Dinka) (**Appendix 8).** The frequency pattern of L0a2 (especially L0a2b and L0a2c sub-haplotypes) might reflect longstanding interaction between East African hunter-gatherers and Cushitic speaking pastoralist populations in the last 5 kya [109, 110, 115]. This is consistent with archaeological and linguistic evidence, which indicates that there have been extensive interactions between southern Cushitic speakers and East African hunter-gatherer populations [109, 110, 115].

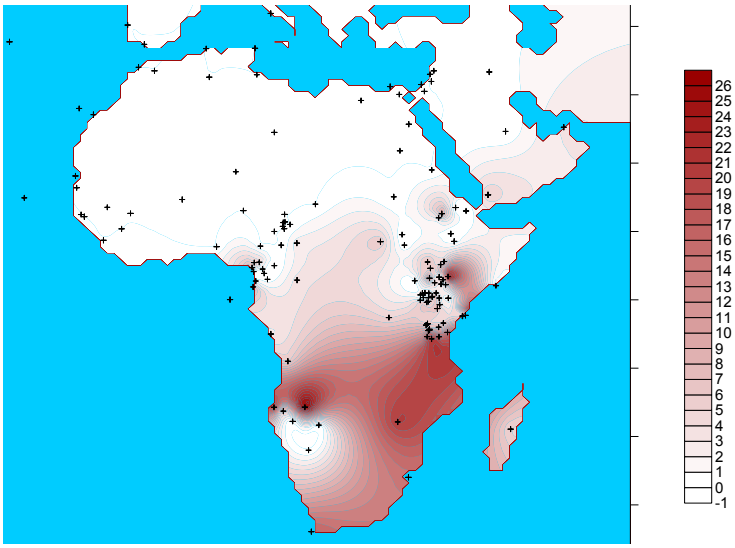


Figure A9.2.6: Contour map representing geographic distributions of L0a2 frequencies across Africa created using the Kriging method in Surfer 8 (Golden Software). Areas of high frequency are represented by darker tone of the color used in the Map. Crosses represent approximate locations of sample collection for populations genotyped in this study or reported site of collection for published data. In cases where only country of sampling but no specific site in the country is listed for published data, capital cities of the country are used as the sampling locations.

Haplogroup L1

The L1 haplogroup is most prevalent in Central/West Africa **(Table 3.4.3)**, with the L1c haplotype being at highest frequency among pygmy populations of Central Africa **(Appendix 6b, Table A9.2.2)**. Most of the individuals from the pygmy populations so far studied including in the current study belong to the L1c1a sub-haplotype [302, 351] **(Figure 3.4.1, Table A9.2.2, Appendix 6b, Appendix 8)**, probably leading to the L1c frequency maximum centered in Central Africa in interpolation map of the lineage **(Figure A9.2.7)**. Most of the L1c haplotypes reported in *mt*DNA lineages in Africa outside Central Africa belong to other L1c sub-haplotypes (L1c2-6) **(Table A9.2.2, Figure A9.2.7a-e)**. In the table below **(Table A9.2.2),** sequences previously defined as L1c5 subhaplotype based on HVI and HVII motif 16214–**16223**–16234–16249–16274 (**182**–204) [582], are reclassified as L1c1a1a based on complete *mt*DNA genome sequences since they share coding region mutations at base-pair positions 14088 and 14034 with L1c1a1 sequences. The frequency maxima observed for L1c and L1c1 **(Figure A9.2.7a,b)** is due to high frequency of this lineages among the pygmy populations **(Table A9.2.2)**. In fact majority of L1c observed in Central Africa belong to L1c1a indicating the magnitude of the pygmy populations' influence on the region's genetic landscape (with over a fifth of the maternal lineages of non-pygmy Central Africans constituting L1c1) **(Table A9.2.2)**. Moreover, most of L1c1 observed outside Central Africa carry the HVI mutation 16086 that defines lineages L1c1b-d [185, 292, 297, 298, 349, 368].

Haplotype L1b is mostly observed in West African populations **(Appendix 6b, Appendix 9, Figure A9.2.8, Table 3.4.3))**, an observation that led Salas [185] to

conclude that it originated in West Africa before subsequent expansion to other parts of Africa. Based on a phylogenetic analysis of the *mt*DNA sequences **(not shown)** [157], L1b and L1c are separated by up to 40 mutational steps indicating that the ancestral populations carrying these sister haplotypes separated a very long time ago (around 100 kya) **(not shown)** [157], and the split might have occurred before the initial expansion out of Central Africa.

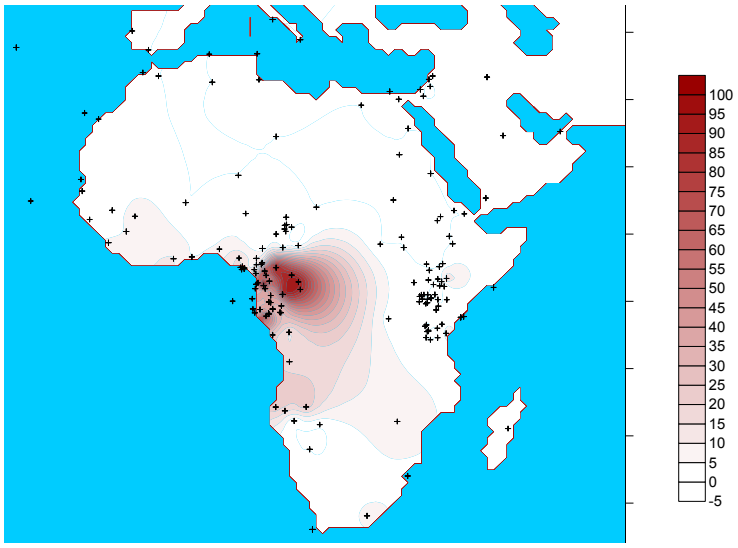
Table A9.2.2: Frequency distribution of L1c Sub-haplotypes in global human populations. Columns show the name (country) of each of the populations, number of individuals tested, frequency of LIc and its sub-haplotypes in the populations, and the primary source reference. Pygmy populations, whose maternal lineage are predominantly from L1c – specifically L1c1a1a are in blue. **A** – this study.

	population	n	L1c	L1c1	L1c2	L1c3	L1c4	L1c6	Sources
Central Africa	Bakola	159	100	100					**A**; [351, 582]
	Tigar	35	100	100					[351]
	Bakoya	31	96.77	96.77					[351]
	Medzan	28	96.43	96.43					**A**
	Mbenzele pygmies	57	94.74	89.47			5.26		[302]
	Baka	203	92.61	88.18	3.45		0.99		**A**; [351, 582]
	Biaka	73	86.3	65.75			20.55		[151, 152, 351]
	Babinga	44	86.36	15.91			70.45		[582]
	Babongo	45	82.22	80			2.22		[351]
	Mvae	23	56.52	39.13	8.7	8.7			**A**
	Ewondo	78	28.21	21.79	3.85	1.28	1.28		[302, 351]
	Ngoumba	132	25	16.67	2.27	5.3	0.76		[351, 582]
	Bassa	46	23.91	10.87	13.04				[204]
	Bakaka	50	14	6	8				[204]
	Bubi	45	22.22			22.22			[529]
	Tikar	34	11.76	5.88	2.94	2.94			[368]
	Bamoun	125	9.6	2.4	4	1.6		1.6	**A**; [368]
	Bamileke	47	6.38	2.13	4.26				[302]
	Fali	41	4.88			4.88			[204]
	Mbum	39	2.56			2.56			[204]
	Aghem	115	12.17	3.48	5.22	3.48			[368]
	Fulani (Cameroon)	99	4.04		1.01	1.01	1.01	1.01	**A**; [204, 528]
	Chadic	227	4.85	0.44	0.44	3.96			[204, 304]
	Giziga	26	3.85	3.85					**A**
	Baggara	21	4.76	4.76					**A**
	Akele	48	39.58	29.17	2.08	6.25	2.08		[351]
	Ateke	54	27.78	14.81	5.56	7.41			[351]
	Benga	50	44	34	4		6		[351]
	Duma	47	34.04	10.64	2.13	6.38	14.89		[351]
	Eshira	40	40	25	2.5	7.5	5		[351]
	Eviya	37	32.43	16.22		13.51	2.7		[351]
	Fang	105	33.33	27.62	1.9		3.81		[351]
	Galoa	71	28.17	21.13	4.23		2.82		[351]
	Kota	56	35.71	30.36			5.36		[351]
	Makina	45	48.89	37.78	4.44	6.67			[351]
	Mitsogo	64	39.06	29.69	1.56	4.69	3.13		[351]
	Ndumu	39	28.21	20.51	5.13	2.56			[351]
	Nzebi	63	38.1	22.22	4.76	6.35	4.76		[351]
	Obama	47	31.91	19.15		4.26	8.51		[351]
	Punu	52	28.85	21.15	1.92	3.85	1.92		[351]
	Shake	51	43.14	35.29	5.88		1.96		[351]
	Bateke	50	14	8			6		[582]
	Sao Tome Principe	153	16.34	7.19	5.88	3.27			[529, 530]

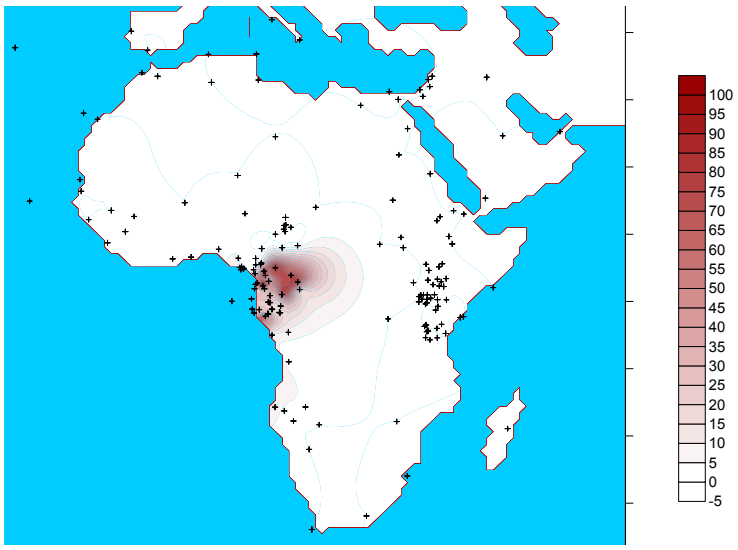
	population	n	L1c	L1c1	L1c2	L1c3	L1c4	L1c6	Sources
West Africa	Igbo	201	15.42	6.97	7.96	0.5			[368]
	Anang	107	16.82	6.54	7.48	1.87		0.93	[368]
	Ibibio	509	16.7	3.73	11.39	1.57			[368]
	Efik	145	18.62	3.45	13.79	1.38			[368]
	Oron	98	12.24	5.1	6.12	1.02			[368]
	Ejagham	133	9.77	0.75	9.02				[368]
	Yoruba	33	6.06	3.03		3.03			[151, 152, 193]
	Hausa	20	5	5					[151, 152]
	Akan	151	10.6	1.99		8.61			[368]
	Ewe	87	4.6	4.6					[368]
	Mandenka	110	1.82					1.82	[151, 152, 184]
	Sierra Leone	276	5.8	4.71		1.09			[317]
	Guinea-Bissau	372	5.11	3.23		1.88			[307]
	Bambara	52	9.62	1.92		7.69			[306]
	Senegalese (Wolof/Serer)	71	2.82			2.82			[401]
East Africa	Datog	30	3.33			3.33			**A**
	Turkana	86	1.16			1.16			**A**; [151, 152, 526]
	Nairobi (Kenya)	100	2			2			[366]
	Taveta	29	3.45	3.45					**A**
	Pare	37	8.11		5.41	2.7			**A**
	Taita	19	15.79		5.26	5.26		5.26	**A**
	Hutu	42	2.38		2.38				[349]
	Mozambique	416	4.09	1.44	1.44	1.2			[185, 298]
	Angola-Cabinda	109	23.85	7.34	11.93	4.59			[292]
South Africa	Kuvale	54	24.07	11.11	7.41	5.56			[297]
	Nyaneka-Nkhumbi	153	16.99	2.61	7.84	6.54			[297]
	Ovimbundu	92	17.39	4.35	6.52	5.43	1.09		[297]
	Ganguela	21	23.81	4.76	14.29	4.76			[297]
	Shona	59	8.47		5.08	3.39			[349]
	South African Colored	563	1.07		0.89	0.18			[295]
	Xhosa	17	5.88		5.88				**A**
	Malagasy	170	1.18	0.59	0.59				[532, 533]
N/NW Africa	Mauritanians	94	4.26	4.26					[306, 401]
	Moroccan Arabs	169	1.18			0.59	0.59		[309, 401, 536]
	Tunisian Arabs	200	1		1				[227, 308-310]
	Tunisian Berbers	182	3.85	0.55		3.3			[308, 310, 311]
	Egyptian (Alexandria)	277	2.17	0.36	0.72	1.08			[305]
	Madeira-Azores	490	0.41		0.41				[440, 441]
	Cape Verde	292	6.85	1.71		5.14			[318]
	Egypt (Arabs)	126	1.59			1.59			[154, 228, 369]
	Portuguese	449	0.45		0.45				[436, 437]
Near East	Yemen	300	1	0.67	0.33				[16, 312]
	Iraq	168	1.19		0.6	0.6			[314, 512]
	Saudi Arabia	553	0.36		0.18	0.18			[313]
	Dubai	249	1.2		0.4	0.8			[319]
	African American	465	9.25						[299, 580]

Figure A9.2.7a-e: Contour map representing geographic distributions of L1c and its subhaplotypes across Africa created using the Kriging method in Surfer 8 (Golden Software). Areas of high frequency are represented by darker tone of the color used in the Map. Crosses represent approximate locations of sample collection for populations genotyped in this study or reported site of collection for published data. In cases where only country of sampling but no specific site in the country is listed for published data, capital cities of the country are used as the sampling locations.

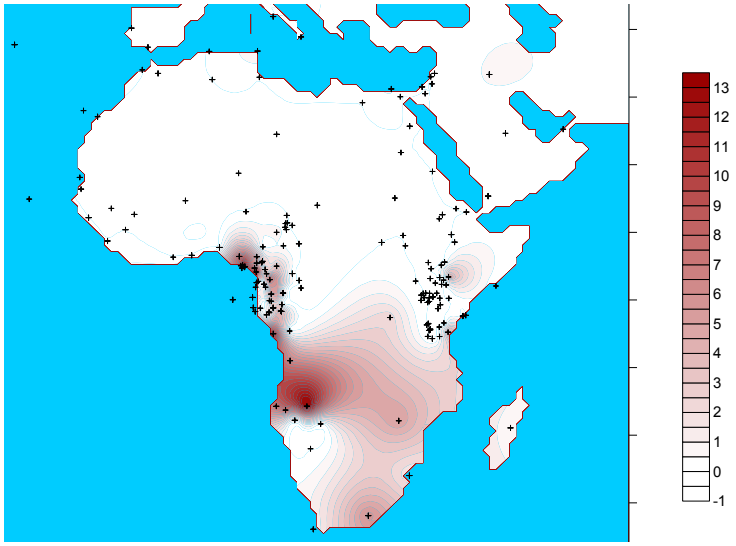
a) L1c



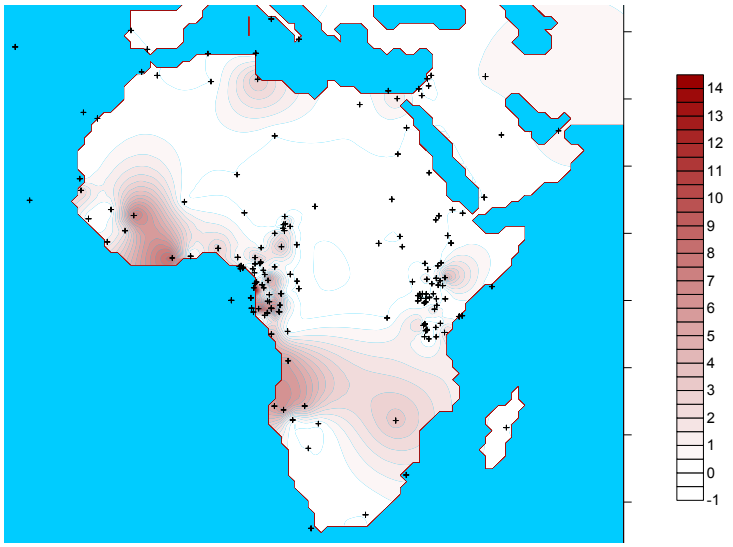
b) L1c1



c) L1c2



d) L1c3



e) L1c4

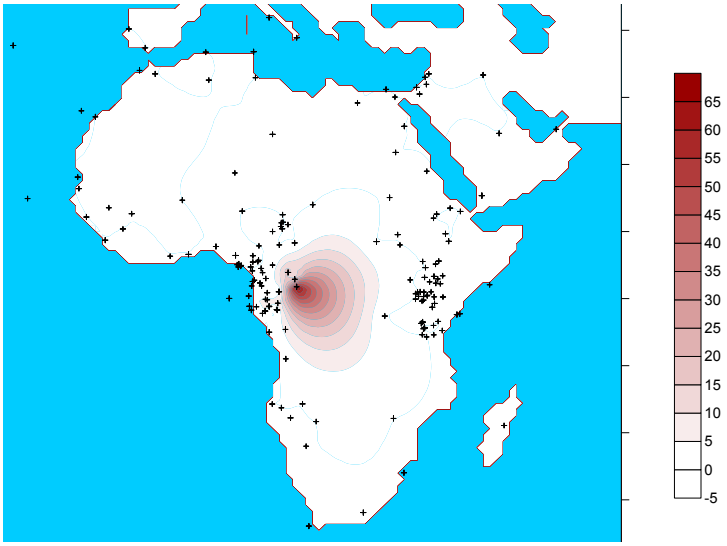
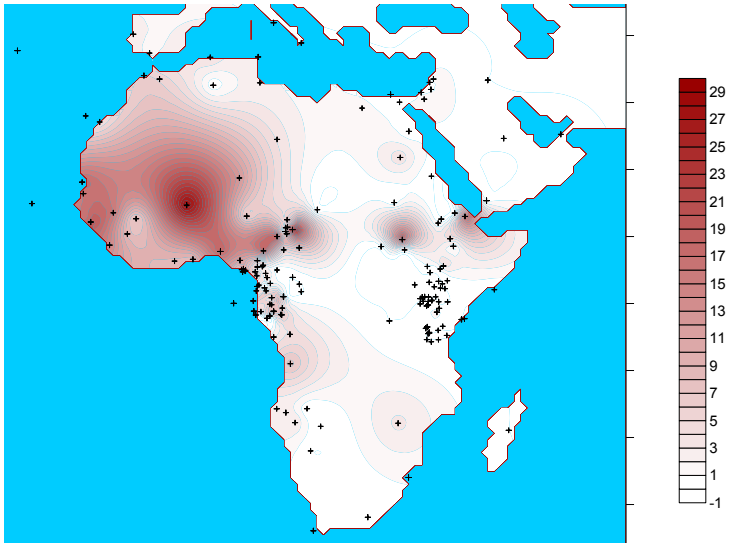


Figure A9.2.8: Contour map representing geographic distributions of L1b frequencies across Africa created using the Kriging method in Surfer 8 (Golden Software). Areas of high frequency are represented by darker tone of the color used in the Map. Crosses represent approximate locations of sample collection for populations genotyped in this study or reported site of collection for published data. In cases where only country of sampling but no specific site in the country is listed for published data, capital cities of the country are used as the sampling locations.



The L2a1 haplotype is the most common L2 haplotype and is widely distributed across Africa **(Figure A9.2.9, Figure 3.4.1, Appendix 6b)**. A previous study [185] argued that the L2a haplotype might have originated in Central Africa before expanding to the rest of Africa since its sister clades L2b-L2e are mainly observed in West Africa **(Appendix 6b, Appendix 9 Figure A9.2.11-13, Table 3.4.3)**. Based on phylogenetic analysis these clades initially split 60-80 kya followed by subsequent expansion just over 30 kya **(not shown)** [157]. As there is no archeological evidence from Central Africa [26] from the period around the initial split of these clades, it is possible that the split occurred in East Africa.

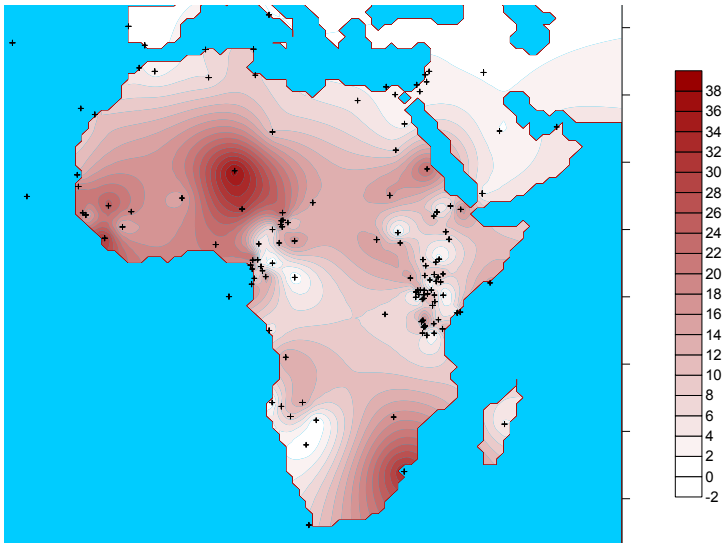


Figure A9.2.9: Contour map representing geographic distributions of L2a1 frequencies across Africa created using the Kriging method in Surfer 8 (Golden Software). Areas of high frequency are represented by darker tone of the color used in the Map. Crosses represent approximate locations of sample collection for populations genotyped in this study or reported site of collection for published data. In cases where only country of sampling but no specific site in the country is listed for published data, capital cities of the country are used as the sampling locations.

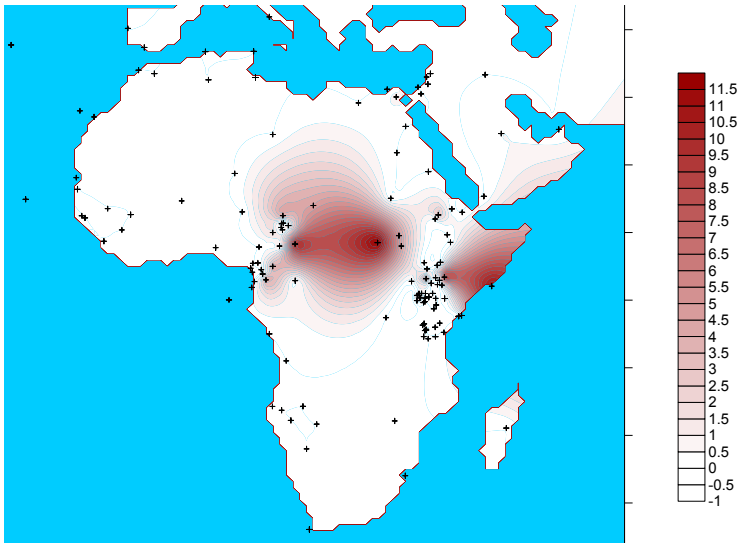


Figure A9.2.10: Contour map representing geographic distributions of L2a2 frequencies across Africa created using the Kriging method in Surfer 8 (Golden Software). Areas of high frequency are represented by darker tone of the color used in the Map. Crosses represent approximate locations of sample collection for populations genotyped in this study or reported site of collection for published data. In cases where only country of sampling but no specific site in the country is listed for published data, capital cities of the country are used as the sampling locations.

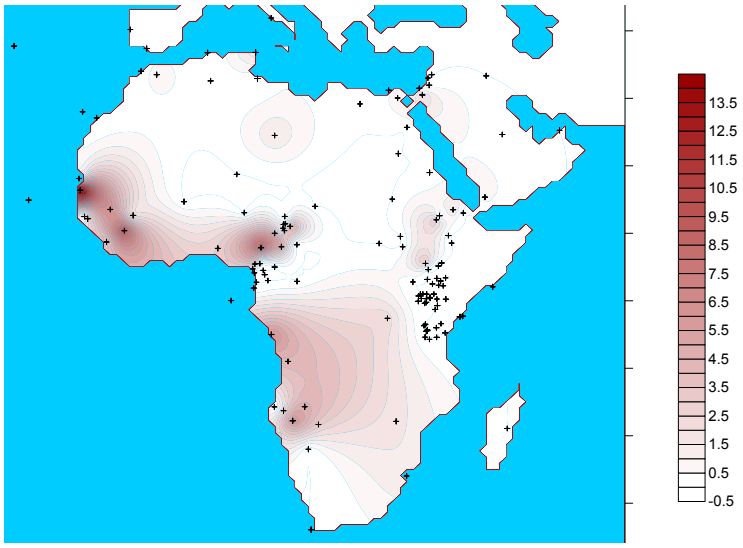


Figure A9.2.11: Contour map representing geographic distributions of L2b frequencies across Africa created using the Kriging method in Surfer 8 (Golden Software). Areas of high frequency are represented by darker tone of the color used in the Map. Crosses represent approximate locations of sample collection for populations genotyped in this study or reported site of collection for published data. In cases where only country of sampling but no specific site in the country is listed for published data, capital cities of the country are used as the sampling locations.

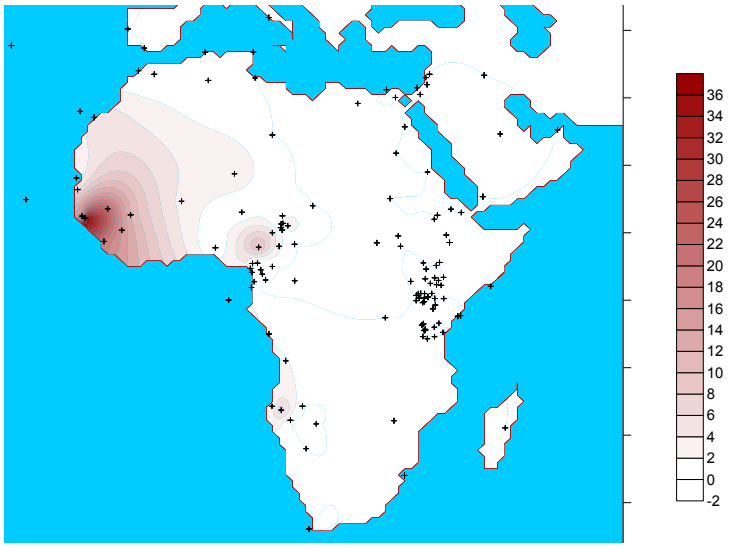


Figure A9.2.12: Contour map representing geographic distributions of L2c frequencies across Africa created using the Kriging method in Surfer 8 (Golden Software). Areas of high frequency are represented by darker tone of the color used in the Map. Crosses represent approximate locations of sample collection for populations genotyped in this study or reported site of collection for published data. In cases where only country of sampling but no specific site in the country is listed for published data, capital cities of the country are used as the sampling locations.

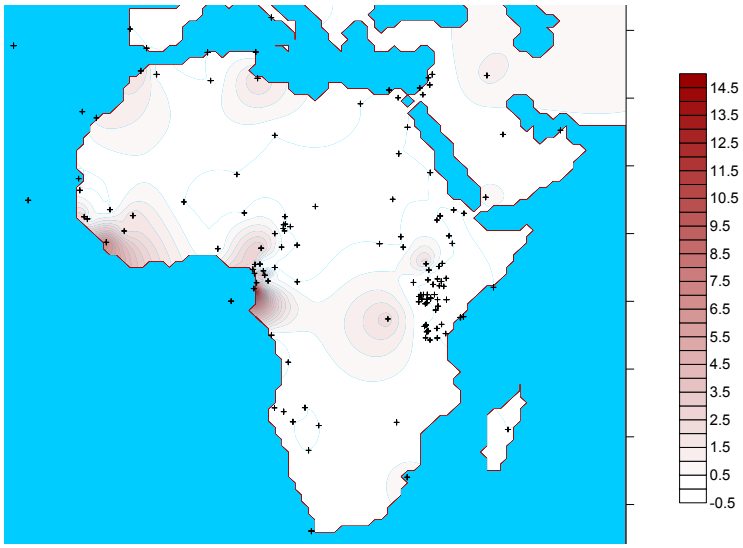


Figure A9.2.13: Contour map representing geographic distributions of L2d frequencies across Africa created using the Kriging method in Surfer 8 (Golden Software). Areas of high frequency are represented by darker tone of the color used in the Map. Crosses represent approximate locations of sample collection for populations genotyped in this study or reported site of collection for published data. In cases where only country of sampling but no specific site in the country is listed for published data, capital cities of the country are used as the sampling locations.

Haplogroup L3

The *mt*DNA L3 haplotypes are mostly observed among populations in East and Central/West Africa **(Figure 3.4.1, Appendix 6b)**. The L3a haplotype is mainly observed in East Africa among the Afroasiatic speaking hunter-gatherer populations (**Figure A9.2.14, Figure 3.4.1, Appendix 6b, Table 3.4.3)**. The TMRCA age estimate for L3a was 33-50 kya (**Figure A9.2.15)** and the limited distribution pattern of this haplotype may indicate that this lineage represents a residual signature of the current Cushitic speaking Holocene hunter-gatherer populations that expanded from northern Kenya/southern Ethiopia into central/southern Kenya.

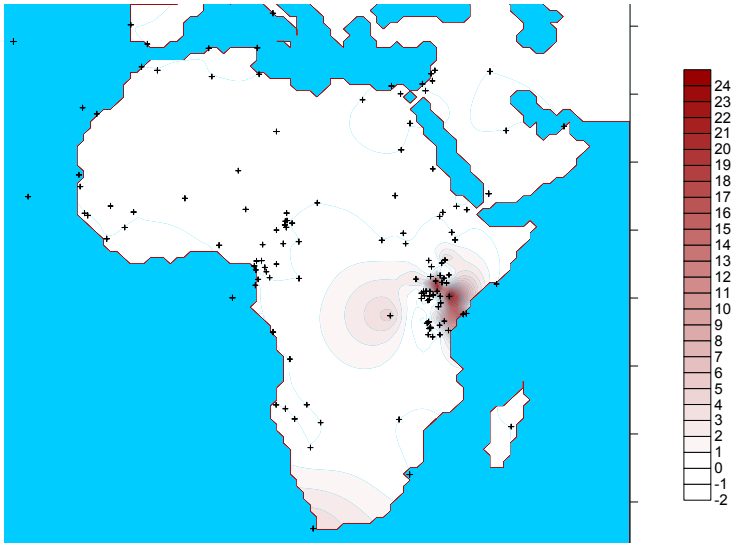


Figure A9.2.14: Contour map representing geographic distributions of L3a frequencies across Africa created using the Kriging method in Surfer 8 (Golden Software). Areas of high frequency are represented by darker tone of the color used in the Map. Crosses represent approximate locations of sample collection for populations genotyped in this study or reported site of collection for published data. In cases where only country of sampling but no specific site in the country is listed for published data, capital cities of the country are used as the sampling locations.

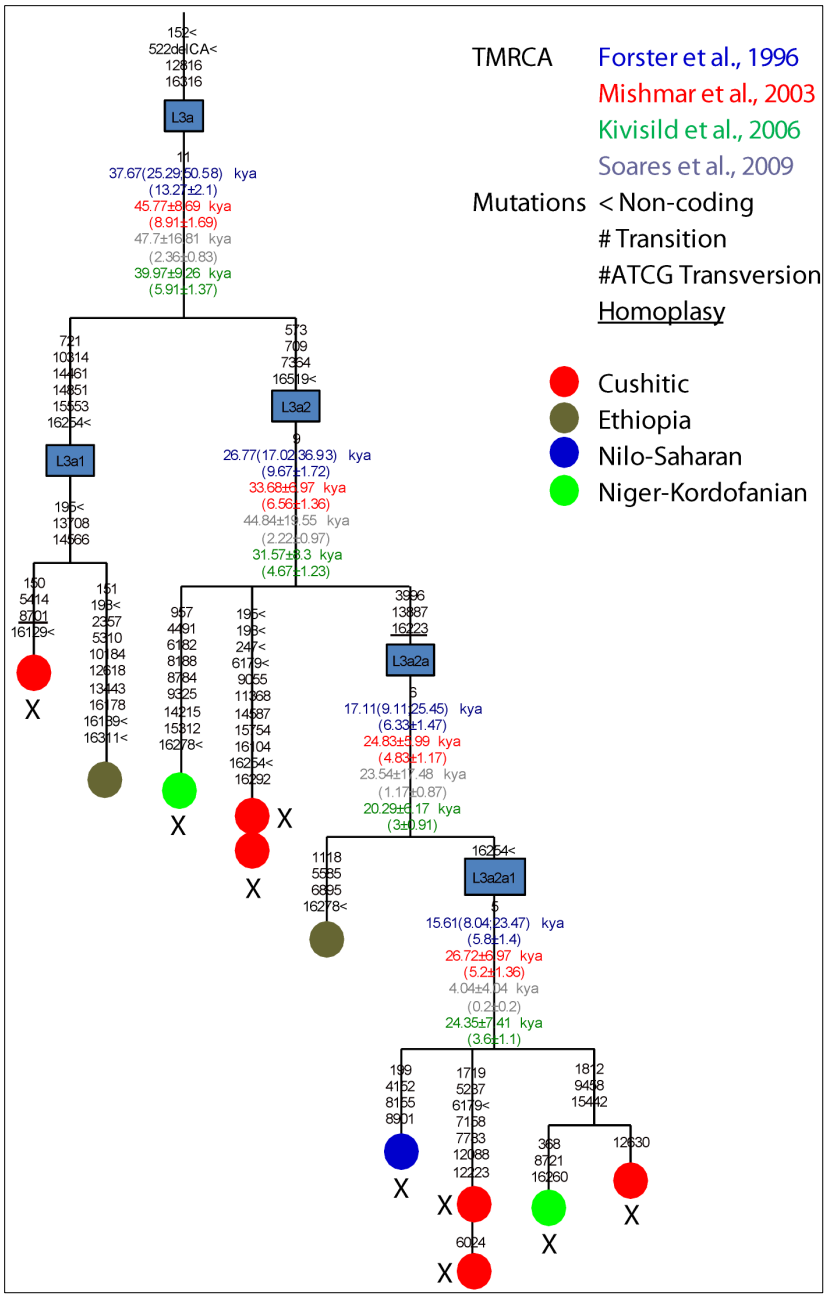


Figure A9.2.15: Phylogeny of *mt*DNA complete genome sequences that belong to the L3a lineage. Samples marked **X** were sequenced in this study. Linguistic group/family of the population/s the sequences were sampled from in the current study, and published sequences whose linguistic affiliations are known are shown. For published data where linguistic affiliations are not known, the country/region in which the sequences are sampled is indicated. The frequency distribution of the lineage in African and non-African populations are shown in **Appendix 6b.**

The L3b haplotype, which is mainly observed among the Niger-Kordofanian speaking populations **(Figure 3.4.1, Appendix 6b),** is also present among African Americans, and earlier studies have shown its predominantly West African distribution [151, 152, 185, 299, 306, 317, 401, 528]. The L3b haplotype has three frequency maxima, centered in West, Central and East Africa. Based on the frequency pattern (**Figure A9.2.16**) this haplotype appears to have expanded into East Africa carried by the Bantu speaking populations. The L3d and L3e haplotypes are also observed mostly among the Niger-Kordofanian speaking populations, with frequency maxima in Central Africa **(Figure A9.2.17, A9.2.18)**. Frequencies of these two haplotypes decrease with distance from Central Africa and might represent a signature of Bantu expansion into other parts of the African continent **(Figure A9.2.17, A9.2.18)**.

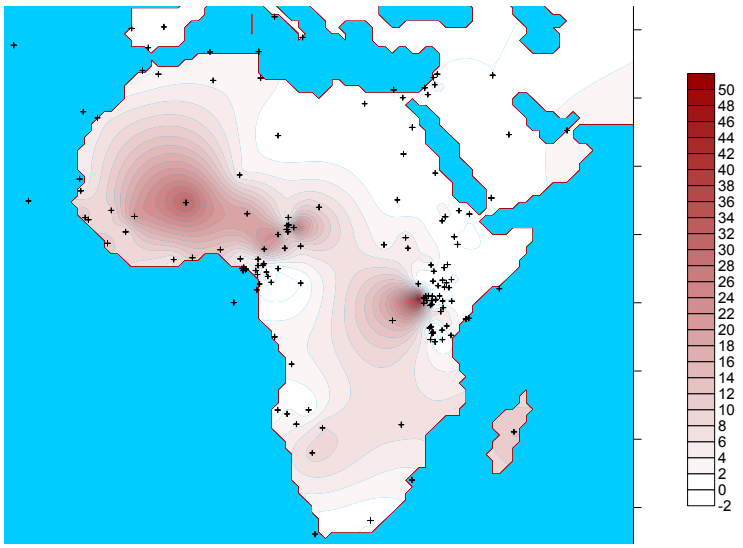


Figure A9.2.16: Contour maps representing geographic distributions of L3b frequencies across Africa created using the Kriging method in Surfer 8 (Golden Software). Areas of high frequency are represented by darker tone of the color used in the Map. Crosses represent approximate locations of sample collection for populations genotyped in this study or reported site of collection for published data. In cases where only country of sampling but no specific site in the country is listed for published data, capital cities of the country are used as the sampling locations.

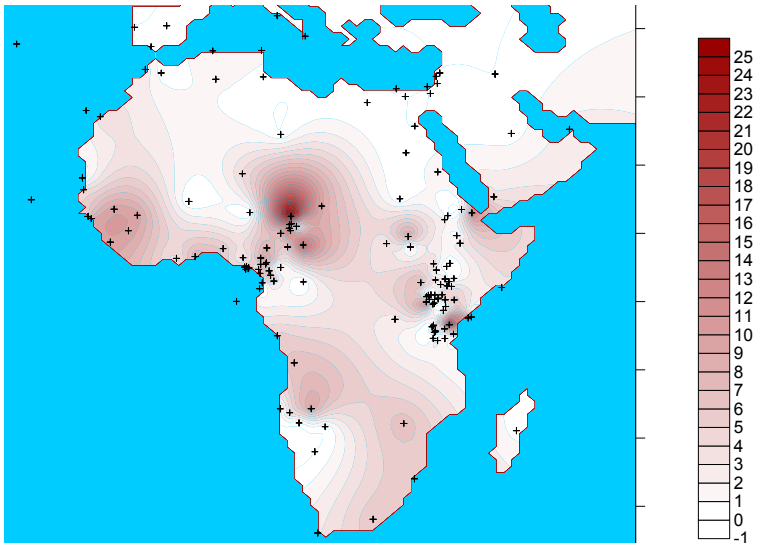


Figure A9.2.17: Contour maps representing geographic distributions of L3d frequencies across Africa created using the Kriging method in Surfer 8 (Golden Software). Areas of high frequency are represented by darker tone of the color used in the Map. Crosses represent approximate locations of sample collection for populations genotyped in this study or reported site of collection for published data. In cases where only country of sampling but no specific site in the country is listed for published data, capital cities of the country are used as the sampling locations.

Inspection of L3e profiles show overall that the lineage seem to correspond to expansion of Niger-Kordofanian from Central Africa to the rest of Africa **(figure A9.2.18a)**. Even though generally associated with Niger-Kordofanian speakers, L3e haplotype was also observed in elevated frequency among Afroasiatic speakers in North Africa. So, detailed analysis of L3e sub-haplotypes' frequency pattern was done in this study. Moreover, unlike other *mt*DNA haplotypes all the L3e sub-haplotypes can be distinguished using mutations within the control regions. When overall frequency profile for L3e haplotype is visualized, the frequency maximum is centered in Central/West Africa. However, when sub-haplotype frequency profiles are examined separately (based on the frequency in **Table A9.2.3**); sub-haplotype L3e1 **(Figure A9.2.18b)** seems to conform to Bantu expansion from Central Africa to East and southern Africa, sub-

haplotype L3e2 **(Figure A9.2.18c)** conforms to expansion of the large Niger-Kordofanian speakers to the rest of Africa (including possible migration of the Kordofanian branch to Sudan [71]), sub-haplotype L3e3 **(Figure A9.2.18d)** to migration of the Niger-Kordofanian speakers across the Sahel to North Africa, sub-haplotype L3e4 **(Figure A9.2.18e)** to a wave of Bantu expansion from Central into East Africa (consistent with inference of several expansion of Bantu speakers into eastern and southern Africa [62], while the sub-haplotype L3e5 has a frequency maxima centered in North Africa (among the Chadic speakers and population from North/northwest Africa) **(Table A9.2.3, Figure A9.2.18f).**

Considering that the L3e5 sub-haplotype was not found in the Middle East and the Mediterranean, this sub-haplotype might originally have been found among the original inhabitants of North Africa/Sahara (Berbers – Arab populations migrated into North Africa in the last 1 kya [303, 583-585]). According to Ehret [101], Afroasiatic initial split is between Omotic and the rest (a group he calls "Erythraean"). The latter divides into Cushitic and "North Erythraean, "which itself consists of Chadic, Egyptian, Berber, and Semitic. Subsequent split gave proto-Chado-Berber and "Boreafrasian"; this last group splits into Egyptian and Semitic [96, 378]. The distribution pattern of L3e5 is therefore consistent with the linguistic evidences that show that Chadic share a more common recent ancestry with the Berbers than with any of the other four Afroasiatic branches [96, 378].

Just like all other *mt*DNA haplotypes, which seems to be associated with movement of Niger-Kordofanian speakers (L1b, L1c, L2b-e, L3b, L3d – **above and main text**), L3e seems to have originated in Central Africa. Considering the TMRCA

estimate for the node that support L3e3-5 is 31 – 57 kya, proto-population that carried L3e5 might have migrated from Central Africa and got absorbed by inhabitants of North Africa (possible ancestors of current Chadic and Berber speakers) in the last 50 kya. Even though there is no consensus on the exact dates for its occurrence, Aterian tool industries might explain possible movement of human populations from Central Africa into Northern/North-West Africa in the last 50 kya. Most of the age estimates of the Aterian industry sites in North/northwestern Africa are between 30 – 50 kya [586-591]. There are few sites where slightly younger (20 – 30 kya during one of the moist period - Ghazalian) and older (60-70 kya) estimates have been recorded [588, 590]. The Aterian industry might have originated in Central Africa and then spread towards the Chad Basin and north to the Central Sahara (during the wet phase in the Sahara that coincides with Marine Isotope Stage 3, 40-50 kya [393, 592]), from which it later reached the Northwestern and Eastern Sahara and possibly the Nile Valley [586]. Thus, the industry has been found across most of North and Northwest Africa, from Atlantic coast to the Nile Valley even though its evidence/influences are few/decreases at the eastern most edge [586]. So, some of ancestral population that gave rise to the Niger-Kordofanian speakers might have migrated north (during initial expansion out of Africa) and joined some of the proto-Afroasiatic speakers. And the frequency profile of L3e5 observed might be reflection of later expansion of Berber-Chadic Afroasiatic speakers in North Africa about 8 – 10 kya [96, 101, 378, 380, 381].

Table A9.2.3: Frequency distribution of L3e Sub-haplotypes in global human populations. Columns show the language family then name (country) of each of the populations, number of individuals tested, frequency of L3e sub-haplotype in the populations, the geographic location where the population is sampled from (longitude and latitude), and the primary source reference. A – This study; [a]Niger, [b]Nigeria, [c]Libya, [d]Cameroon, [e]Chad, [f]Alexandria, [g]Morroco, [h]Tunisian, [i]Tupuri & Tale, [j]Nile Valley and west of the Nile, [k]Italians from Sicily (154), Basilicata (92) and Calabria (95) regions, [l]Fulani from Mali, [m]English speaking Caribbeans from Dominica, Grenada, St. Lucia, St. Kitts, St. Vincent and Trinidad

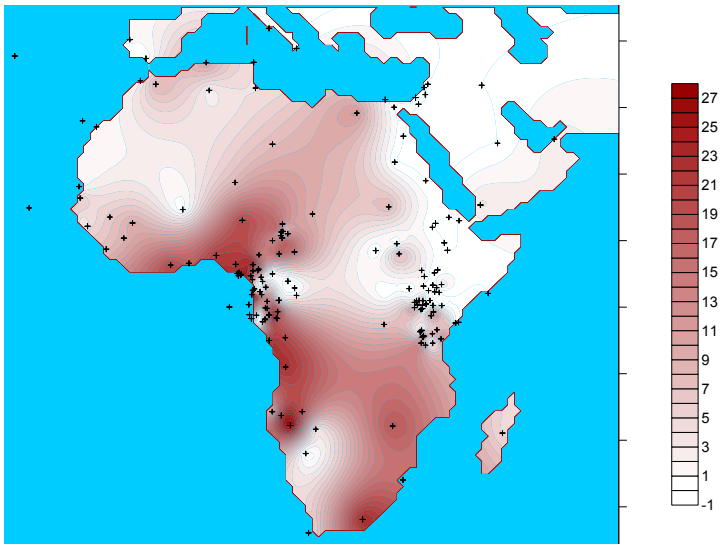
Population	n	L3e1	L3e2	L3e3	L3e4	L3e5	Longitude	Latitude	Source
Yemen	300	1.33	0.67	3.67			44.21	15.35	[16, 312]
Palestinian	227	1.76		0.44			34.47	31.5	[219, 314, 367]
Tuareg[a]	24	4.17	4.17				7.39	18.74	[151, 152]
Yoruba	33	3.03	12.12	6.06			4.57	7.77	[151, 152]
Gikuyu	50	4	2	10			37.15	-0.72	A; [151, 152]
Mandenka	110		0.91				-14.75	12.17	[151, 152]
Fulbe[b]	60		15				10.97	7.86	[151, 152]
Hausa	20		10	10			8.48	13.05	[151, 152]
Kanuri	40	2.5	12.5			7.5	14.3	11.3	A; [151, 152]
Sierra Leone	276	0.72	5.07		1.81		-11.95	8.72	[317]
Tuareg[c]	129	3.1	1.55	3.1			13	24.5	[534]
Kuvale	54	3.7	1.85				12.95	-15.72	[297]
Ganguela	21	9.52	9.52				17.47	-15.7	[297]
Nyaneka	153	9.15	9.15	3.27	1.31		14.32	-16.28	[297]
Ovimbundu	92	11.96	7.61	3.26			14.98	-8.98	[297]
Cabindans	109	11.93	6.42	0.92	0.92		12.5	-5	[292]
Mozambique	416	10.34	1.44	3.37	0.24		32.59	-25.97	[185, 298]
Fulani[d]	99	3.03	8.08			2.02	13	10	A; [204, 528]
Fulani[e]	49		10.2			2.04	15.3	11.03	[528]
Kung	62		3.23				19.5	-18.33	[183, 193]
Kwe	31	19.35	9.68				15.68	-17.78	[183]
South African Coloreds	563	2.49	0.53	0.53			18.42	-33.92	[295]
Sao Tome	103	17.48	11.65	1.94	5.83		6.53	0.03	[529, 530]
Bubi	45	11.11	20	2.22			9.9	1.88	[529]
Mauritania	94				1.06		-16.04	18.12	[306, 401]
Bambara	52	1.92	7.69				-8	12.65	[306]
Senegalese	170	0.59	3.53	0.59	1.76		-15.86	16.39	[299, 401]
Cape Verde	292		3.77		10.96		-23.52	14.92	[318]
Hutu	42	2.38	2.38		2.38		29.74	-2.6	[349]
Shona	59	6.78	11.86				31.03	-17.86	[349]
Egyptian[f]	277	1.44	1.08	0.36			29.92	31.2	[305]
Tunisians	200	3	0.5			2.5	10.18	36.8	[227, 308-310]
Dubai	249	1.61		0.4	0.4		55.28	25.25	[319]
Berbers[g]	327		1.53			7.03	-4.5	33.5	[401, 402, 537]
Berbers[h]	182	1.1			0.55	2.2	10.45	32.93	[308, 310, 311]
Algerians	47					10.64	3.05	36.75	[398]
Western Saharans	81	1.23					-13.42	27.1	[398, 401]
Malagasy	170	3.53	0.59	0.59			47.52	-18.92	[532, 533]
Siwa	78	11.54					25.67	29.17	[402]
Moroccans	169		2.37				-6.83	34.02	[309, 401, 536]
Nairobi	100	5	2	3			36.82	-1.28	[366]

Population	n	L3e1	L3e2	L3e3	L3e4	L3e5	Longitude	Latitude	Source
Mozabite	85		2.35				3.5	32.58	[171, 399]
Chadic	227	2.64	5.73	0.44		8.81	14	8	[204, 304]
Bassa	46	8.7	2.17	2.17			9.98	4.13	[204]
Bakaka	50	10	6	6	8		9.75	4.72	[204]
Bamileke	47	2.13	6.38	2.13			10.07	5.45	[204]
Ewondo	78	1.28	2.56		1.28		11.52	3.87	[204, 351]
Fali	41		2.44				14.56	11.4	[204]
Shake	51	1.96	11.06				12.2	0.87	[351]
Punu	52	13.46	5.77		3.85		9.55	-1.13	[351]
Obama	47	12.77	4.26	4.26			13.78	-0.68	[351]
Nzebi	63	1.59	9.52	11.11			12.2	-0.18	[351]
Ngoumba	88	3.41	6.82	2.27			10.05	2.55	[351]
Ndumu	39	5.13	2.56	10.26			13.7	-1.7	[351]
Mitsogo	64	6.25	4.69	7.81			11.93	-1.88	[351]
Makena	45		8.89	2.22			13.93	1.02	[351]
Kota	56		5.36	1.79			11.93	-0.1	[351]
Galoa	71	2.82	1.41	2.82	1.41		9.83	-1.72	[351]
Fang	105	4.76	9.52	1.9	0.95		11.37	1.88	[351]
Eviya	38	2.63					10.6	-1.22	[351]
Eshira	40	5					11.88	-1.78	[351]
Duma	47	14.89					12.48	-1.13	[351]
Benga	50	12	6	2			9.45	0.38	[351]
Bakoya	31	3.23					13.98	1.03	[351]
Ateke	54	7.41	11.11	1.85	3.7		13.58	-1.63	[351]
Baka	127	0.79					13	5	A; [351]
Akele	48	10.42	6.25	2.08	2.08		12.5	-1.2	[351]
Mbum[i]	39	2.56	7.69		5.13	2.56	14.44	10.4	[204]
Egyptians[j]	126		0.79				31.25	30.05	[154, 228, 314, 369]
Guinea Bissau	372		4.3		2.96		-9.3	10.38	[307]
Akan	151		15.89	3.97			-2.27	6.33	[368]
Ewe	87		13.79	3.45			0.47	6.6	[368]
Anang	107	10.28	13.08	1.87			7.72	5.05	[368]
Efik	145	8.28	10.34	2.07	0.69		7.98	5.17	[368]
Ejagham	133	9.77	11.28	6.02			8.32	4.95	[368]
Ibibio	509	6.09	8.84	1.38	1.38		7.87	4.72	[368]
Igbo	201	11.94	7.46	1.49	0.5		7.48	6.43	[368]
Oron	98	8.16	13.27	5.1			8.23	4.83	[368]
Tikar	34	5.88	11.76				10.92	5.72	[368]
Aghem	115	5.61	20	0.87			10.07	6.38	[368]
Bamoun	125	10.34	13.6	3.2			10.8	5.5	A; [368]
Laka	38	5.26	2.63		5.26	2.63	16.3	8.3	A
Maasai	41	7.32		2.44			37	-4	A

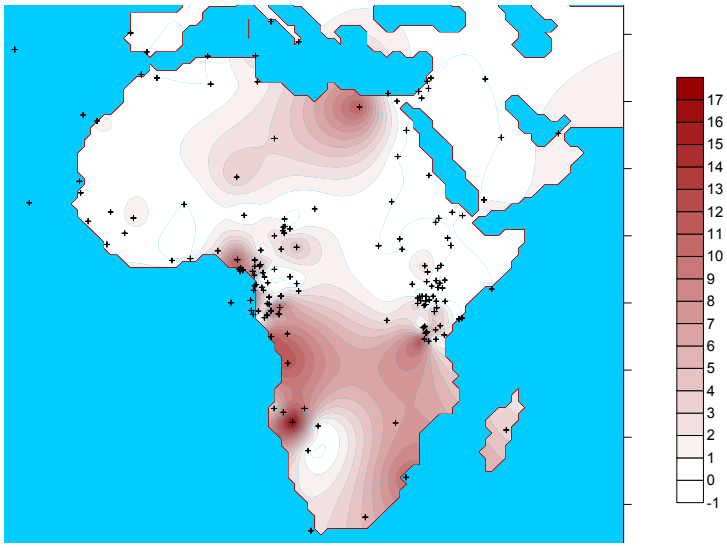
Population	n	L3e1	L3e2	L3e3	L3e4	L3e5	Longitude	Latitude	Source
Turkana	86	1.16					35.67	3.17	A; [151, 152, 526]
Burunge	25	4		4			35.97	-5.71	A
Luhya	28	3.57		3.57			34.77	0.33	A
Rangi	27	7.41		7.41			35.66	-4.3	A
Sandawe	25	16					35.32	-5.39	A
Turu	24	4.17					18.42	-33.92	A
Luo	21	4.76	9.52				34.32	-0.07	A
Pare	37	2.7	10.81				38.3	-4.76	A
Taveta	29	3.45					37.68	-3.4	A
Xhosa	17	5.88	17.65				18.42	-33.92	A
Mvae	23		8.7				12	3	A
Marakwet	22		9.09				35.57	0.98	A
Sabaot	17		5.88				34.62	0.95	A
Nyangatom	112	2.68					35.38	5.53	[526]
Daasanech	49	2.04					35.8	4.62	[526]
Nuer	22		4.55			4.55	32	8	A
Nyimang	12		8.33				30.45	15.07	A
Boulala	22		4.55			4.55	19	14	A
Tugen	43		2.33				35.52	0.31	A
Datog	30		3.33				35.17	-3.69	A
Taita	19			5.26			38.37	3.4	A
Orma	20				5		40.45	-2.39	A
Baggara	21					14.29	14.5	12.5	A
Giziga	26					3.85	14.25	10.75	A
Southern Italians[k]	341	0.29	0.29				16.6	38.89	[593]
Portuguese	449	0.22	0.45	0.45		0.22	-8.42	40.2	[436, 437]
French	788		0.13				2.35	48.85	[435]
Madeira-Azores	490		0.61	0.2	0.2		-25.67	37.74	[440, 441]
Malinke	31		3.23		3.23		-11.4	13.54	[306]
Peul[l]	15					6.67	0.05	16.06	[306]
Canarians	300					0.67	-15.5	28	[531]
African Americans	465	2.8	9.68	2.8	0.43				[299, 580]
The Caribbeans[m]	314	3.18	9.55	0.96	1.59				[594]
Brazilians	297	4.71	4.38	1.01		0.34			[180, 300]
Colombians	185	9.73	0.54	25.41	0.54				[595]
Cubans	245	0.82	4.49	0.82	0.41				[301]

Figure A9.2.18a-g: Contour maps representing geographic distributions of L3e haplotypes and its sub-haplotypes frequencies across Africa created using the Kriging method in Surfer 8 (Golden Software). Areas of high frequency are represented by darker tone of the color used in the Map. Crosses represent approximate locations of sample collection for populations genotyped in this study or reported site of collection for published data. In cases where only country of sampling but no specific site in the country is listed for published data, capital cities of the country are used as the sampling locations. a) overall L3e frequencies, b) L3e1, c) L3e2, d) L3e3, e) L3e4 and f) L3e5.

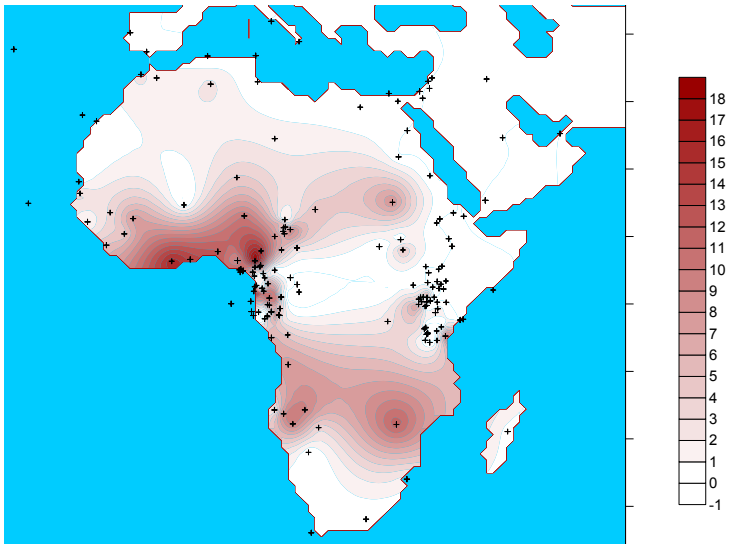
a) L3e



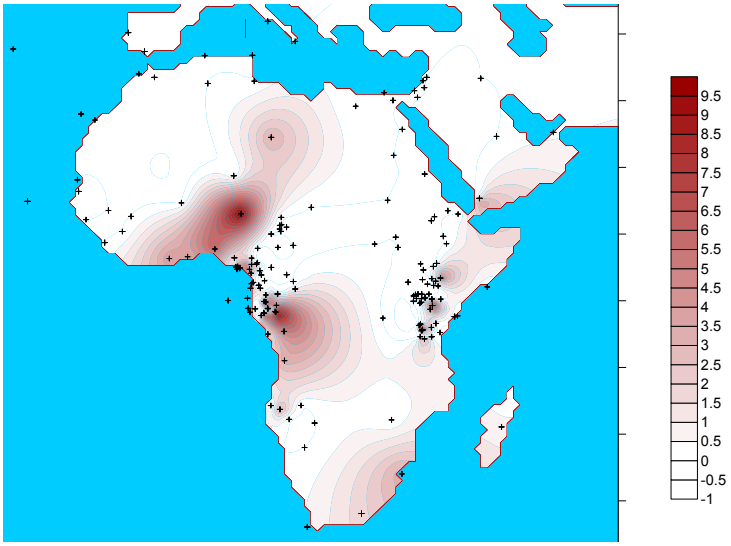
b) L3e1



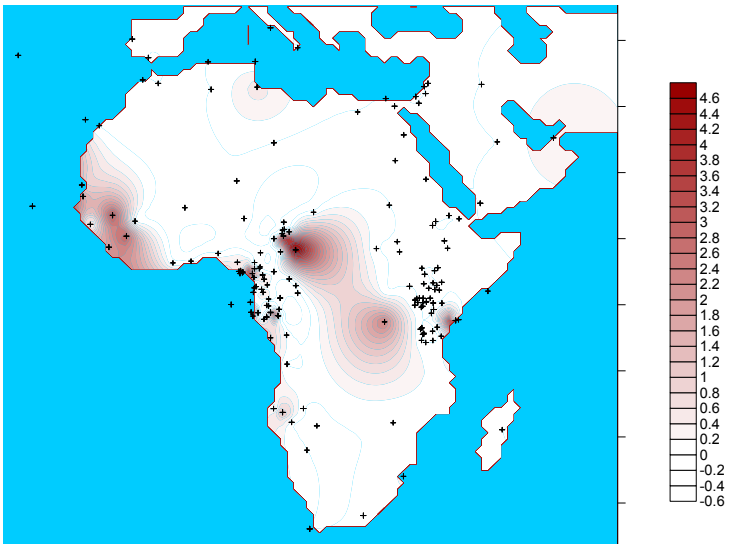
c) L3e2


18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
-1

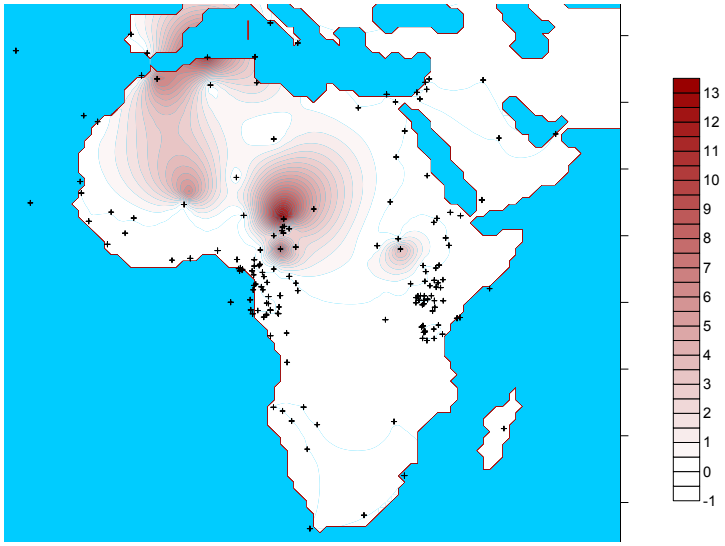
d) L3e3


9.5
9
8.5
8
7.5
7
6.5
6
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
-0.5
-1

e) L3e4


4.6
4.4
4.2
4
3.8
3.6
3.4
3.2
3
2.8
2.6
2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6

f) L3e5


13
12
11
10
9
8
7
6
5
4
3
2
1
0
-1

Haplotype L3f is widely distributed among East and Central African Afroasiatic and Nilo-Saharan speaking populations, and is at low frequency in neighboring Niger-Kordofanian populations **(Figure 3.4.1, Appendix 6b)**. The interpolation map shows that there two frequency maxima, one in Central Africa and the second one in the northeastern Africa **(Figure A9.2.19)**. This haplotype is also observed at low frequency among African Americans [226, 299, 580], and in the Caribbean [594], Central and South America (Cuban [301], Colombian [595] and Brazilian [180, 300]) populations. Based on observation of the L3f haplotype in the few representative East African populations from previous studies [45, 151, 152, 154, 596], Salas [185] speculated that this haplotype was of East African origin. Most of the individuals analyzed in the current study belong to the L3f1 clade, however, most of the Chadic speaking individuals in the study belong to the L3f3 clade **(Appendix 8)**. The L3f2b sub-haplotype defined by a combination of control region mutations at nucleotide positions 16311 (HVI) and 152 (HVII), are observed exclusively in Cushitic speaking populations from East Africa **(Appendix 8)**. A recent study of complete *mt*DNA genome sequencing of L3f3 individuals among the Chadic speaking populations [597] concluded that the variant might represent the expansion of Chadic speakers into central Africa from East Africa about 8 kya [597]. Consistent with linguistic evidence [96, 101, 378], the sharing of this haplotype (L3f3) between the Chadic and the East African Cushitic speakers **(Figure 3.2.1.4)** may indicate that they share recent common ancestry.

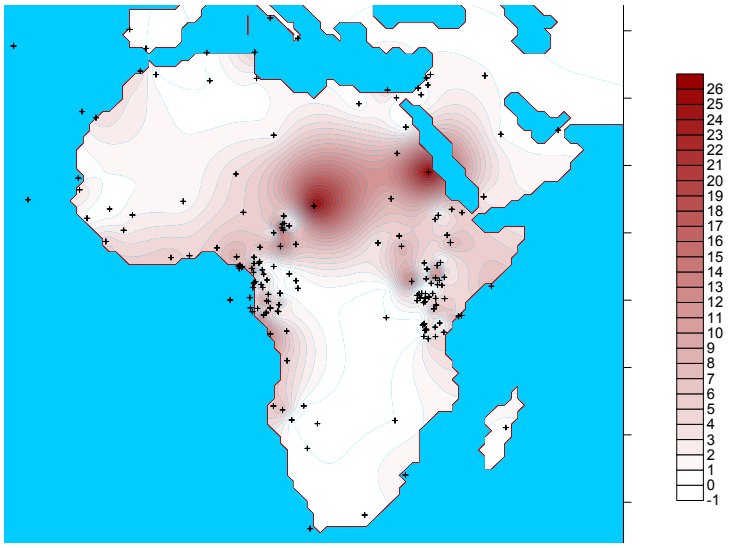


Figure A9.2.19: Contour maps representing geographic distributions of L3f haplotypes frequencies across Africa created using the Kriging method in Surfer 8 (Golden Software). Areas of high frequency are represented by darker tone of the color used in the Map. Crosses represent approximate locations of sample collection for populations genotyped in this study or reported site of collection for published data. In cases where only country of sampling but no specific site in the country is listed for published data, capital cities of the country are used as the sampling locations.

The L3h haplotype has a frequency maximum in southern Sudan. Outside of East Africa, the L3h haplotype is observed at low frequencies in a few populations that live in the vicinity of Nilo-Saharan speaking populations who might have had historical interaction with them [16, 151, 152, 204, 219, 227, 297, 304-318] **(Figure 3.4.1, Appendix 6b)**. However, most of all the non-East African individuals who have the L3h haplotype belong to the L3h1b sub-haplotype [151, 152, 204, 297, 305, 307, 309-311, 313-315, 317, 318, 351, 367, 368] **(Figure A9.2.20, Appendix 8, Appendix 6b)** while most of the individuals that carry the L3h haplotype in East African populations belong to L3h1a2 sub-haplotype [219, 349, 366] **(Figure A9.2.20, Appendix 8, Appendix 6b)**. Therefore, the distribution pattern of L3h might reflect the signature of Nilo-Saharan

ancestry in Africa. This observation is consistent with the assertion based on archaeological evidence, indicating that an aquatic culture that extended from Northwest Kenya up to the Atlantic coast in west Africa was associated with Nilo-Saharan speakers [57, 84]. The L3h1a2 sub-haplotype distribution appears to reflect the migration of Nilotic speaking populations from a southern Sudan homeland into Kenya and Ethiopia, within the past 3kya [65, 87]. Therefore, considering the L3h haplotype is found at low frequency in populations that live in the vicinity of Nilo-Saharan speaking populations **(Figure 3.4.1, Appendix 6b)**, it might be a signature of Nilo-Saharan expansion to the rest of Africa.

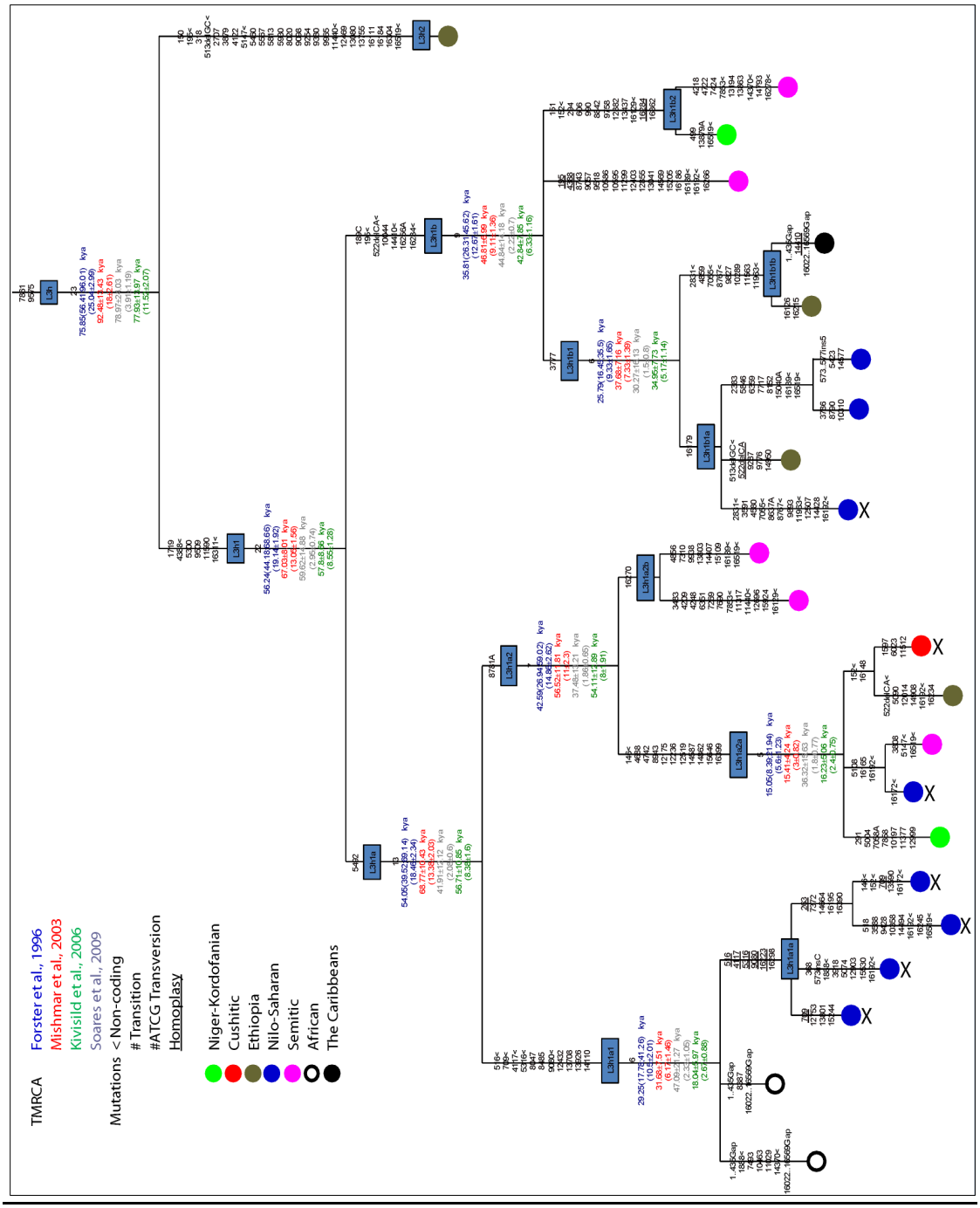


Figure A9.2.20: Phylogeny of *mt*DNA complete genome sequences that belong to the L3h lineage. Samples marked **X** were sequenced in this study. Linguistic group/family of the population/s the sequences were sampled from in the current study, and published sequences whose linguistic affiliations are known are shown. For published data where linguistic affiliations are not known, the country/region in which the sequences are sampled is indicated. The frequency distribution of the lineage in African and non-African populations are shown in **Appendix 6b.**

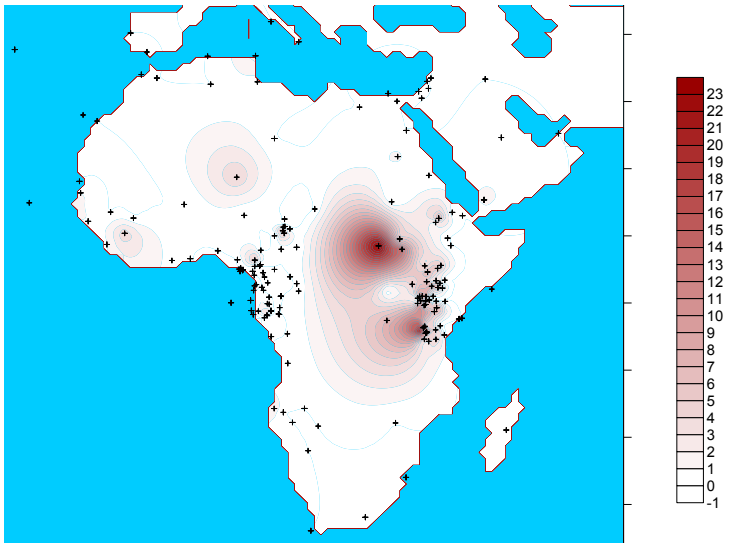


Figure A9.2.21: Contour maps representing geographic distributions of L3h haplotypes frequencies across Africa created using the Kriging method in Surfer 8 (Golden Software). Areas of high frequency are represented by darker tone of the color used in the Map. Crosses represent approximate locations of sample collection for populations genotyped in this study or reported site of collection for published data. In cases where only country of sampling but no specific site in the country is listed for published data, capital cities of the country are used as the sampling locations.

The L3i haplotype is mostly observed among the Nilo-Saharan and Afroasiatic speaking populations from Kenya and Ethiopia **(Figure 3.4.1, Appendix 6b)**. Interestingly, the L3i1 sub-haplotype **(Figures 3.2.1.5, Appendix 8)** is prevalent among the Nilo-Saharan speaking populations while L3i2 is mostly observed among Afroasiatic speaking populations **(Appendix 8)**. This pattern suggests that the L3i haplotype was found in the proto-population which existed prior to the Nilo-Saharan and Afroasiatic population split. Alternatively, proto-populations that spoke Nilo-Saharan or Afroasiatic that carried either one of the sub-haplotype got absorbed into proto-populations that spoke Afroasiatic or Nilo-Saharan respectively, before subsequent expansion of speakers of both linguistic families. The TMRCA estimate for L3i is 34-48 kya, and the time

might mark an upper bound to the time of population split for proto-Nilo-Saharan and proto-Afroasiatic speaking populations.

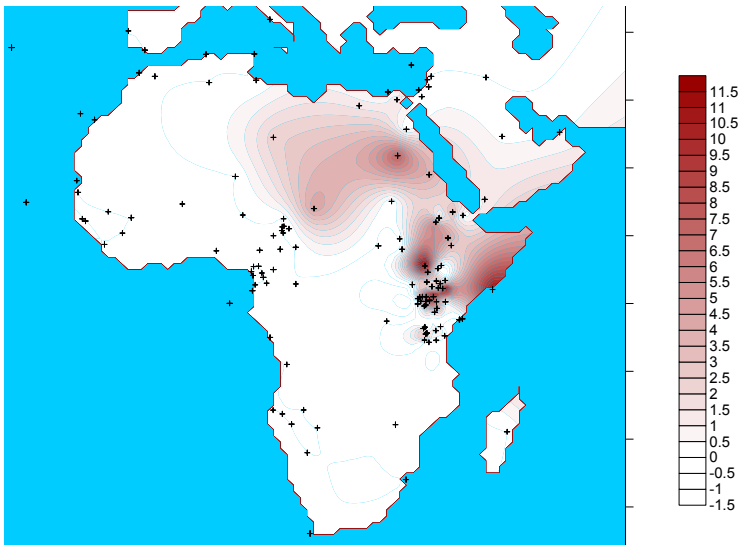


Figure A9.2.22: Contour maps representing geographic distributions of L3i haplotypes frequencies across Africa created using the Kriging method in Surfer 8 (Golden Software). Areas of high frequency are represented by darker tone of the color used in the Map. Crosses represent approximate locations of sample collection for populations genotyped in this study or reported site of collection for published data. In cases where only country of sampling but no specific site in the country is listed for published data, capital cities of the country are used as the sampling locations.

The L3x haplotype is commonly observed among the East African Cushitic speakers (with frequency maximum centered in East Africa – **Figure A9.2.24, Table 3.4.3**), and virtually all of the eastern Nilotic speaking populations at low frequency, indicating that there has been substantial gene flow into these populations from Cushitic speaking populations [79, 347, 374]. This inference is also consistent with analysis of autosomal markers genotyped in a larger African population dataset [4], as well as results from the current study. Linguistic evidence also attests to this interaction: the Maa languages contain a mixture of Cushitic features, Cushitic-like pronouns, noun and verb

vocabulary [112]. Moreover, archeological excavations of a burial sites called “Kalokol”, in Northwest Kenya, an area currently occupied by the eastern Nilotic Turkana, dated to about 2.5 kya, are consistent with eastern Cushitic speaking people’s funeral practices (Borana, Konso) [86, 107], indicating the possible region where interaction between Cushitic speakers and eastern Nilotic speakers may have taken place.

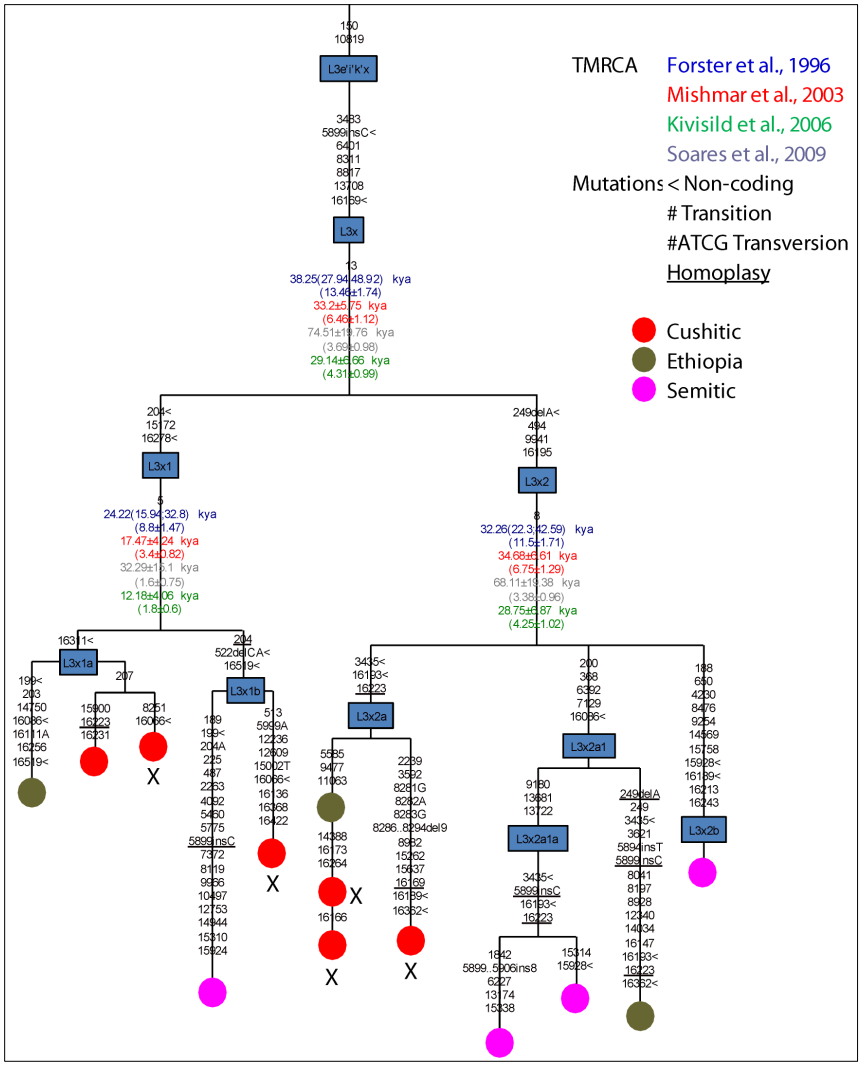


Figure A9.2.23: Phylogeny of *mt*DNA complete genome sequences that belong to the L3x lineage. Samples marked **X** were sequenced in this study. Linguistic group/family of the population/s the sequences were sampled from in the current study, and published sequences whose linguistic affiliations are known are shown. For published data where linguistic affiliations are not known, the country/region in which the sequences are sampled is indicated. The frequency distribution of the lineage in African and non-African populations are shown in **Appendix 6b.**

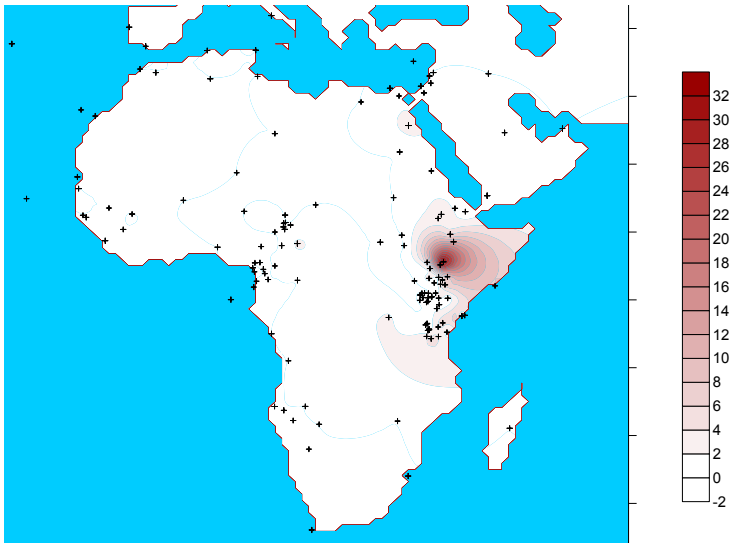


Figure A9.2.24: Contour maps representing geographic distributions of L3x haplotypes frequencies across Africa created using the Kriging method in Surfer 8 (Golden Software). Areas of high frequency are represented by darker tone of the color used in the Map. Crosses represent approximate locations of sample collection for populations genotyped in this study or reported site of collection for published data. In cases where only country of sampling but no specific site in the country is listed for published data, capital cities of the country are used as the sampling locations.

Haplogroup L4

The L4 haplotype, previously defined as L3g [185], is mainly restricted to eastern and northeastern Africa (**Table 3.4.3)**. The L4 haplotype is also observed at low frequencies in few Central African populations [204, 302] and southeastern Arabian populations [16, 219, 312, 313, 319] **(Figure 3.4.1, Appendix 6b)**. Previous studies had shown the L4 haplotype at low frequency in African American [299, 580], Caribbean [594], Central and South American (Cuban [301] and Brazilian [180, 300]) populations. Interestingly, an individual carrying the L3a haplotype, another haplotype (besides L0f and L4 discussed below) commonly observed among East African hunter-gatherer populations **(above)**, was observed in Brazilian samples [300], indicating that eastern Africa might have also acted as a source population for the trans-Atlantic slave trade.

Most of the haplotypes found in East African hunter-gatherers are a derived variant, L4b2a2b defined by HVI & HVII's nucleotide base positions 16172 and 244 respectively (in addition to 16293-16355-16399-146 motif that defines L4b2) **(Figure 3.4.6)**. The frequency profile for L4b2a2 clearly shows that most of the non-hunter-gatherer populations that carry the haplotype are those living in the vicinity of East African hunter-gatherer populations who have had historical interactions (**Table A9.2.4**) [108, 113, 121, 122, 348] with them. L4b2a2b was also observed in low frequencies among the Bantu speaking Shona populations from Zimbabwe [349], South African Khoisan speaking populations [157] and South African colored population [295] **(Figure 3.4.6, Appendix 6b)**, and just as in the case of L0d3, L4 haplotype might represent a signature of the expansion of East African hunter-gatherer populations during the initial movement of livestock into southern Africa [59].

Linguistic group	Ethnicity	n	Freq.	Subsistence/comments
Khoisan	Hadza	19	52.6	Hunter-Gatherers
Afroasiatic	Sanye	20	30	Hunter-Gatherers
Nilo-Saharan	Ogiek	22	22.7	Hunter-Gatherers
Khoisan	Sandawe	24	16.7	Hunter-Gatherers
Nilo-Saharan	Nandi	13	15.4	Admixed
Niger-Kordofanian	Turu	24	12.5	Admixed
Afroasiatic	Akie	20	10	Hunter-Gatherers
Nilo-Saharan	Shilluk	21	9.5	
Niger-Kordofanian	Gikuyu	26	7.7	Admixed
Niger-Kordofanian	Rangi	27	7.4	Admixed
Nilo-Saharan	Tugen	44	6.8	
Nilo-Saharan	Datog	30	6.7	
Afroasiatic	Gabra	31	6.5	Admixed
Afroasiatic	Rendille	31	6.5	
Afroasiatic	Yaaku	19	5.3	Hunter-Gatherers
Nilo-Saharan	Pokot	39	5.1	
Afroasiatic	Orma	20	5	Admixed
Nilo-Saharan	Luo	21	4.8	
Afroasiatic	Iraqw	23	4.3	Admixed
Afroasiatic	Wata	24	4.2	Hunter-Gatherers
Afroasiatic	Burunge	26	3.8	Admixed
Afroasiatic	Borana	59	3.4	Admixed
Nilo-Saharan	Samburu	30	3.3	Admixed
Nilo-Saharan	Dinka	33	3	
Nilo-Saharan	Turkana	39	2.6	Admixed
Nilo-Saharan	Maasai	41	2.4	Admixed

Table A9.2.4: Relative frequency of *mt*DNA lineage L4b2a2 among East African populations. Populations labeled admixed (is a misnomer) in the subsistence column is used here to refer to populations that are currently living in the vicinity of East African hunter-gatherer populations or had historical contacts with the hunter-gatherer populations.

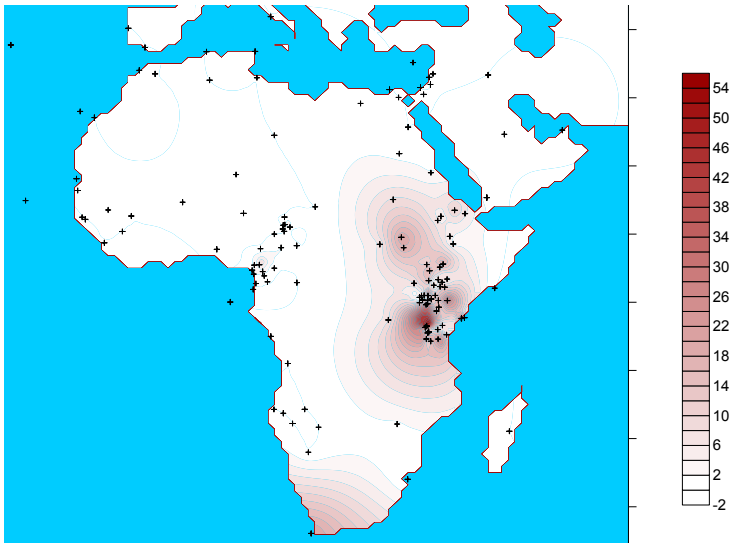


Figure A9.2.25: Contour maps representing geographic distributions of L4 haplotypes frequencies across Africa created using the Kriging method in Surfer 8 (Golden Software). Areas of high frequency are represented by darker tone of the color used in the Map. Crosses represent approximate locations of sample collection for populations genotyped in this study or reported site of collection for published data. In cases where only country of sampling but no specific site in the country is listed for published data, capital cities of the country are used as the sampling locations.

Haplogroup L5

The L5 haplotype was initially observed at low frequency in the Central African Mbuti pygmy population and in two Bantu speaking populations (Ronga and Shangaan) from southern Mozambique [185, 298]. L5 haplotype has a frequency maximum in southern Sudan and may reflect a signature of Nilo-Saharan expansion **(Figure A9.2.26)**. However, a sub-haplotype under this haplotype, L5a1, defined by transitions at 8152, 11065, 11260 and 12215 (and either 16311del or 459insC mutations in the control region) **(Figure 3.4.7)** is observed exclusively among the Cushitic speaking populations of northern Kenya and southern Ethiopia (and other Afroasiatic speakers from Ethiopia [16]). The L5a1 clade has a TMRCA age estimate of 16 – 26 kya **(Figure 3.4.7).** This might mark the lower bound of the time of population split between proto-Nilo-Saharan

and proto-Afroasiatic. Alternatively, an offshoot of Nilo-Saharan speakers might have been absorbed into Afroasiatic speakers in the recent past (in the last 20 kya). The fact that there is low frequency of this Nilo-Saharan-specific haplotype (L5) in Mozambique [16, 185] may be an indication that Nilo-Saharan speakers had a role in later expansion of pastoralism into southern Africa [59].

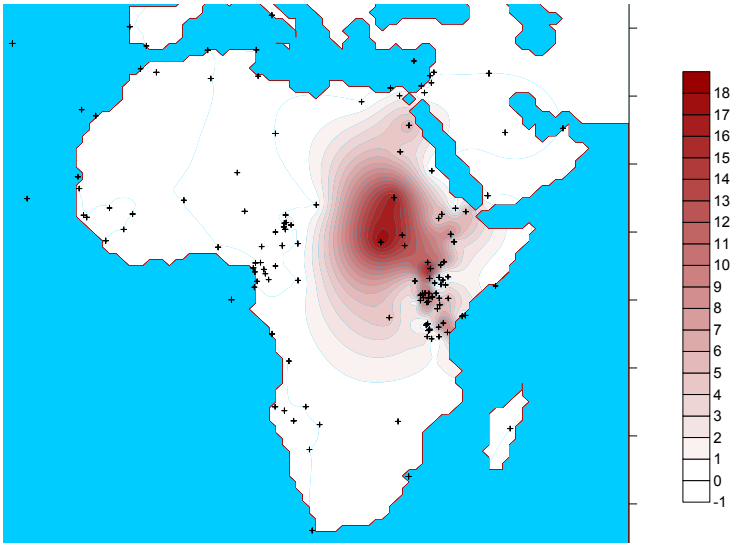


Figure A9.2.26: Contour maps representing geographic distributions of L5 haplotypes frequencies across Africa created using the Kriging method in Surfer 8 (Golden Software). Areas of high frequency are represented by darker tone of the color used in the Map. Crosses represent approximate locations of sample collection for populations genotyped in this study or reported site of collection for published data. In cases where only country of sampling but no specific site in the country is listed for published data, capital cities of the country are used as the sampling locations.

Haplogroup M

M haplotypes are generally observed outside Africa, with most members of the M haplogroup, M2-M64, found in the Indian sub-continent [7, 320, 321], aboriginal populations from Southeast Asia [7, 322, 323] and the South West Pacific [324]. By contrast, M1 seems to be restricted to the Near East, southern Europe, East Africa and North Africa [222, 396]. The debate to explain the current distribution pattern of M1 lineage [15, 16, 222, 396, 464, 598-600] is not yet resolved, with some of the previous studies arguing that the lineage originated in eastern Africa [15] while others asserting back-migration from southwestern Asia [16, 222, 396, 464, 598-600]. Among the two most recent studies [222, 396] based on data from extensive sampling in Europe and the Near East both Olivieri *et al.,* [222] and Gonzalez *et al.,* [396] argued that M1 originated in southwest Asia because of a similar distribution pattern of the *mt*DNA U6 haplotype (from haplogroup N), which has a similar age to M1a (the main M1 clade) and also appears to have originated in southwest Asia [222, 396]. However, based on the more extensive sampling in East Africa in the current study, the frequency maximum for M1a appears to be centered in East Africa **(Figure A9.2.27, Appendix 6b, Table 3.4.3)**. The M1a haplotype is most common among the Cushitic speaking populations in East Africa **(Figure 3.4.1, Appendix 6b).** Some N-clade haplotypes, N1a, I, J1 and K1 and U were found at low frequency in Afroasiatic speaking populations of East Africa, and implication of this is addressed in the discussion section.

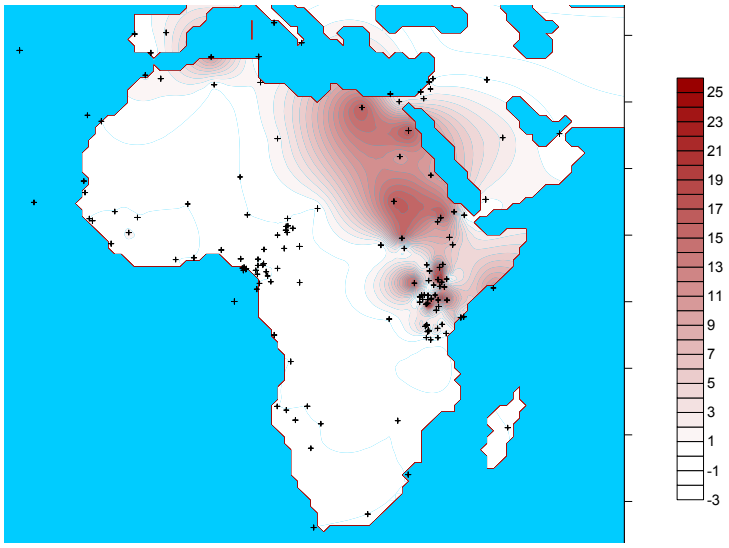
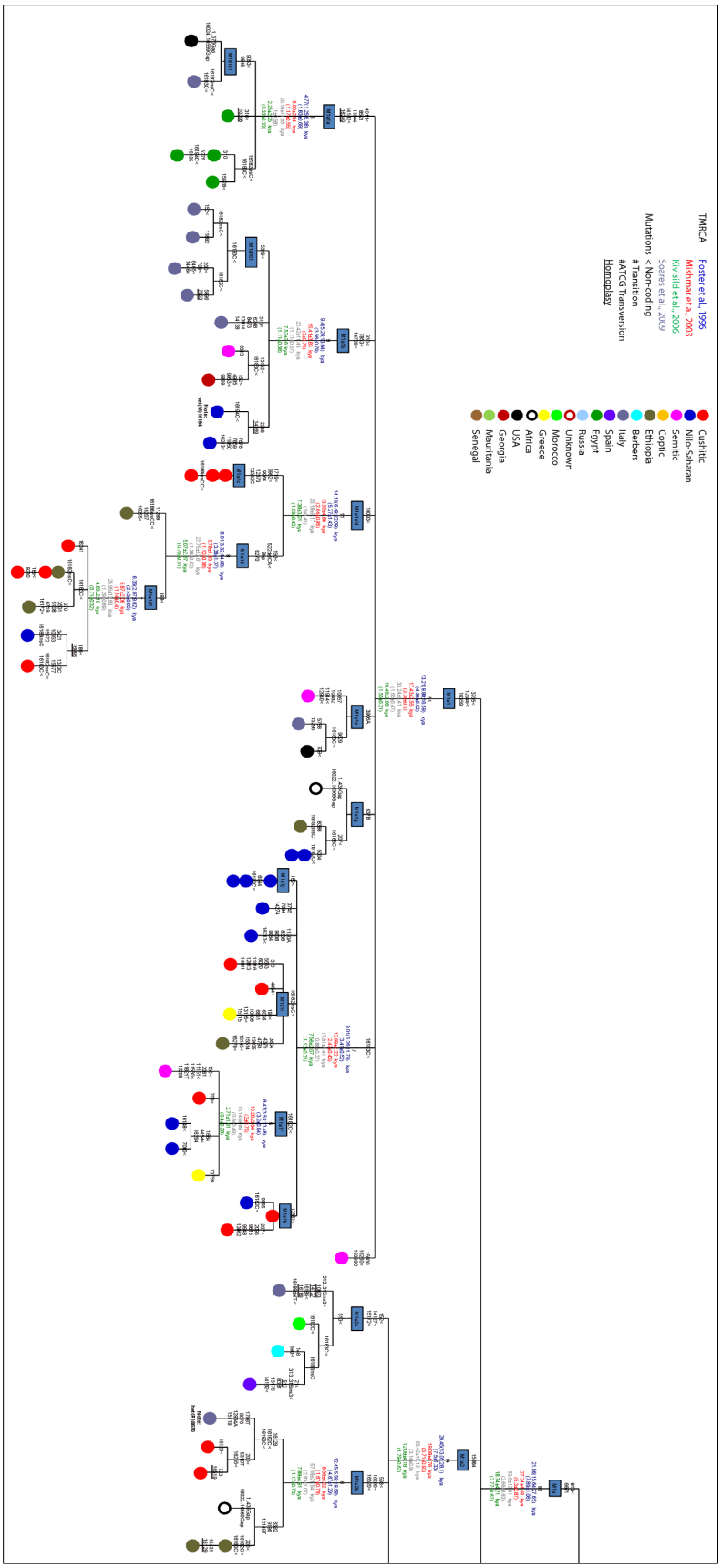
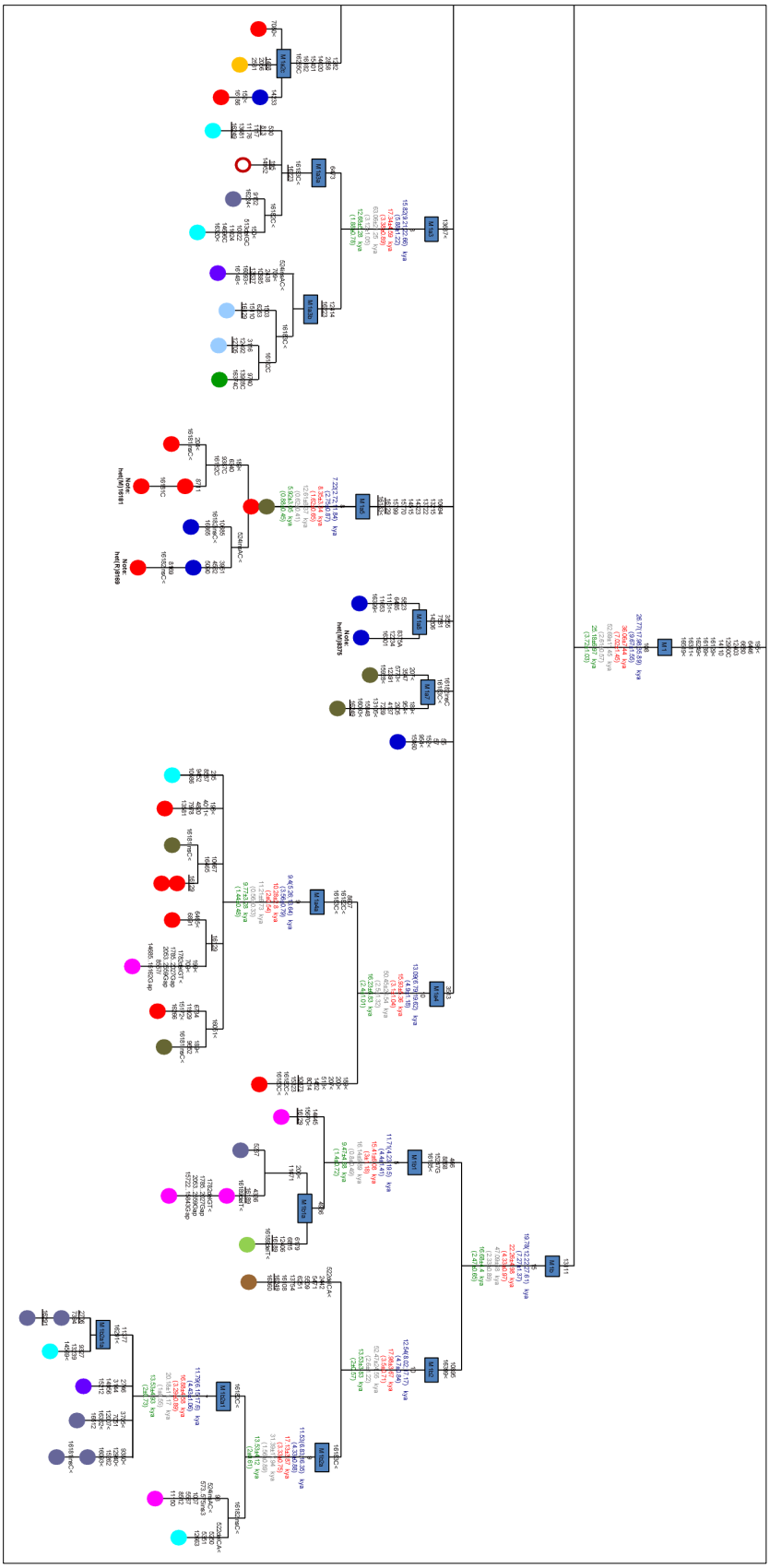


Figure A9.2.27: Contour maps representing geographic distributions of M1a haplotypes frequencies across Africa created using the Kriging method in Surfer 8 (Golden Software). Areas of high frequency are represented by darker tone of the color used in the Map. Crosses represent approximate locations of sample collection for populations genotyped in this study or reported site of collection for published data. In cases where only country of sampling but no specific site in the country is listed for published data, capital cities of the country are used as the sampling locations.

Figure A9.2.28: Phylogeny of *mt*DNA complete genome sequences that belong to the M1 lineage. Samples marked **X** were sequenced in this study. Linguistic group/family of the population/s the sequences were sampled from in the current study, and published sequences whose linguistic affiliations are known are shown. For published data where linguistic affiliations are not known, the country/region in which the sequences are sampled is indicated. The frequency distribution of the lineage in African and non-African populations are shown in **Appendix 6b.**


TMRCA
Foster et al., 1996
Mishmar et al., 2003
Kivisild et al., 2006
Soares et al., 2009
Mutations < Non-coding
Transition
#ATCG Transversion
Homoplasy
Cushitic
Nilo-Saharan
Semitic
Coptic
Ethiopia
Berbers
Italy
Spain
Egypt
Russia
Unknown
Morocco
Greece
Africa
USA
Georgia
Mauritania
Senegal



Haplotype U6

The *mt*DNA U6 haplotype is mostly observed among the Northwestern African populations, especially Berber speakers but also at moderate and low frequency among Arab speaking populations that border them and other Afroasiatic speakers in Africa/the Near East respectively **(Table A9.2.5)**. Haplotype U6, which has been used to support previous studies [222, 396] contention of back-migration, has nearly the same distribution pattern as two related Y chromosome lineages that seem to have expanded in north/northwestern Africa, E3b2 and E3b1a* [153] **(Figure A9.2.29, Table A9.2.5)**. The U6 haplotype has frequency maximum centered on northeastern Africa **(Figure A9.2.29)**.

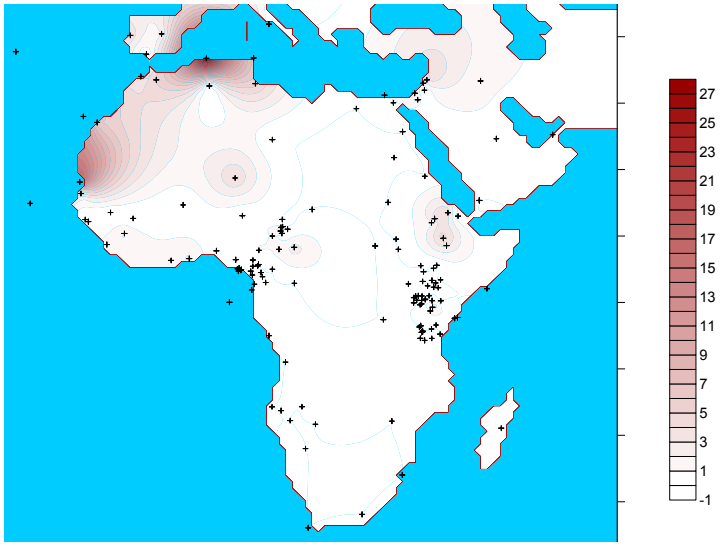


Figure A9.2.29: Contour maps representing geographic distributions of U6 frequencies across Africa and the Near East created using the Kriging method in Surfer 8 (Golden Software). Areas of high frequency are represented by darker tone of the color used in the Map. Crosses represent approximate locations of sample collection for populations genotyped in this study or reported site of collection for published data. In cases where only country of sampling but no specific site in the country is listed for published data, capital cities of the country are used as the sampling locations.

Overall, besides showing that autosomal, mitochondrial and Y chromosome data are correlated with geographical regions/language families, this study was able to show that the distributions of some lineages are also correlated with geography/language. There are a few cases where sister clades; i.e. L0a1a-d (L0a1), L3i1 & L3i2 (L3i), that coalesce in the last 20 – 40 kya are found among populations speaking languages that belong to different linguistic groupings. These lineages were likely to have been found in an ancestral population or populations that split and gave rise to extant populations that belong to different linguistic groups.

Table A9.2.5: Frequency of U6 haplotype in Global human populations. Columns show the name (or nationality) of each of the populations, number of individuals tested, frequency of U6 mitochondrial haplotype in the populations, the geographic location where the population is sampled from (longitude and latitude), and the primary source reference. a) **This study,** [a]Fulani from Burkina Faso, [b]Egyptians from city of Alexandria, [c]Samples from individuals from Nairobi Kenya with undetermined ethnicities, [d]Egyptians from cities in the Nile Valley and West of the Nile

	Population	n	U6	Longitude	Latitude	References
1	Mozabites	85	28.24	3.5	32.58	[171, 399]
2	Mauritanians	64	17.19	-16.04	18.12	[306]
3	Tunisians	200	6.5	10.18	36.8	[227, 308-310]
4	Western Saharans	81	6.17	-13.42	27.1	[398, 401]
5	Moroccan Berbers	327	5.2	-4.5	33.5	[401, 402, 537]
6	Gurage	21	4.76	38.78	9.67	[16]
7	Fulani (Cameroon)	99	1.01	13	10	[204, 528]; **a**
8	Syrians	69	4.35	36.3	33.5	[314]
9	Tuareg (Niger)	24	4.17	7.39	18.74	[151, 152]
10	Gikuyu	50	4	37.15	-0.72	**a;** [151, 152]
11	Portuguese	449	4	-8.42	40.2	[436, 437]
12	Giziga	26	3.85	14.25	10.75	**a**
13	Tigray	53	3.77	39.48	13.5	[16]
14	Canarians	300	14	-15.5	28	[531]
15	Ethiopian Jews	29	3.45	37.47	12.6	[219]
16	Moroccans	169	2.96	-6.83	34.02	[309, 401, 536]
17	Cape Verde	292	3.08	-23.52	14.92	[318]
18	Madeira-Azores	490	3.06	-25.67	37.74	[440, 441]
19	Oromo	33	3.03	39.27	8.55	[16]
20	Senegalese	170	3.53	-15.86	16.39	[299, 401]
21	Laka	38	2.63	16.3	8.3	**a**
22	Ibibio	509	0.2	7.87	4.72	[368]
23	Aghem	115	0.87	10.07	6.39	[368]
24	Akan	151	1.32	-2.27	6.33	[368]
25	Ewe	87	2.3	0.47	6.6	[368]
26	Fulani[a]	97	1.08	10.97	7.86	[528]
27	Tunisian Berbers	182	5.49	10.45	32.93	[308, 310, 311]
28	Sierra Leonians	276	1.45	-11.95	8.72	[317]
29	Palestinians	227	2.3	34.47	31.5	[172, 219, 367]
30	Guineans	372	2.15	-9.3	10.38	[307]
31	Spanish	886	1.75	-3.7	40.42	[314-316, 399, 601-605]
32	Amharas	121	1.67	37	12	[16]
33	Bedouins	87	1.32	34.92	30.5	[219, 314, 367]
34	Iraqis	168	1.19	44.39	33.34	[314, 512]
35	Egyptians[b]	277	1.08	29.92	31.2	[305]
36	Nairobians[c]	100	1	36.82	-1.28	[366]
37	Egyptians[d]	126	0.79	31.25	30.05	[154, 228, 314, 369]
38	Saudis	553	0.9	46.77	24.64	[313]
39	Druze	388	0.77	35.75	33	[438]
40	Dubai	249	0.4	55.28	25.25	[319]
41	Italians	797	0.25	12.48	41.9	[314, 367, 539-541, 606]
42	Chadic[d]	227	1.32	14	8	[204, 304]
43	French	998	0.1	2.35	48.85	[435, 443]

a)

East/ Northeastern Africa	Central/West Africa	South Africa	Northwest Africa
L0f	L1b	L0d1	U6
L3a	L1c	L0d2	
L3h	L2b	L0k	
L3i	L2c	L0a2	
L3x	L2d		
L4	L3b		
L5	L3d		
M1a	L3e		

b)

East /Northeastern Africa	Central/West Africa	South Africa	Northwest Africa
A3b2	A1	A3b1	E3b2
B2b*	E2b	A2	
B2a	E3a (E3a*, E3a7, E3a8)		
E2a	R1b*		
E3*			
E3b (E3b*, E3b1, E3b3, E3b6)			

Table A9.2.6: Summary of centers of haplotype maxima for *mt*DNA **(A)** and Y chromosome **(B)** lineages inferred by lineage frequency interpolation maps. L0a1 and L2a1 *mt*DNA lineages are widely distributed all over Africa.

Appendix 10.1: Neutrality Tests for Mitochondrial Genomes

A previous study [178] analyzed global samples of 104 complete *mt*DNA sequences by comparing the ratio of non synonymous to synonymous substitution for mitochondrial genes in different climatic regions and concluded that selection may have played a role in shaping regional *mt*DNA variation in human populations and that one of the selective influences was climate [178]. However Elson [243] analyzed a larger dataset of mitochondrial sequences and concluded that climate has not been a major selective force during human *mt*DNA evolution [243]. Elson [243] argued that there has been negative selection on *mt*DNA because their single-gene analyses showed significant departures from neutrality at the cytochrome oxidase subunit I (*CO1),* NADH: Ubiquinone oxido-reductase subunit IV (*ND4),* and, NADH: Ubiquinone oxido-reductase subunit VI (*ND6)* genes, but positive selection on the *AT6* gene. Results from the current study indicate that there has been both gene-specific and lineage-specific selection. All the lineages analyzed here have significant negative values for both Tajima's D and Fu & Li's D* (**Table 3.2.2.1**). The ratio between non-synonymous-to-synonymous substitutions was also analyzed in the 13 protein-encoding *mt*DNA genes. Only L0 and M1a lineages showed significant NI values with the former indicating a signature of negative selection and, the latter one showing a signature of weakly positive selection (**Table 3.2.2.1**). Only three genes in three different haplogroups indicated significant values of NI; COIII (L0), ND5 (for L3) and COI (for N) (**Table 3.2.2.1**). NI scales with selection where values less than 1 indicate an excess of amino acid fixations, or positive selection and NI values greater than 1 indicate an excess of amino acid polymorphisms, or negative selection [245, 246]. So it appears that there has been negative selection

(purifying selection) acting on COIII and COI in L0 and N haplogroups, respectively (**Table 3.2.2.1**).

Related to this, Shevchuk and Allard [607] constructed data matrices for 13 mitochondrial protein-coding genes for 41 mammals (34 species) to check phylogenetic congruence which might be due to differential selection pressures on mitochondrial genes. In the pair-wise gene comparisons, significant incongruence was detected for ND6, COII, or COIII genes partitioned individually against the rest of the genes. Omission of the ND6, COII, and COIII from the comparisons significantly improved congruence observed in their data matrix. They concluded that the incongruence is because the three genes don't have crucial functional roles but rather likely play a secondary role. Thus they speculated that mutational/selectional constraints on these genes might have been relaxed [607]. However, previous studies [190, 220, 243, 608-611] showed that there has been widespread purifying selection in human *mt*DNA. In the current study two different genes, COIII and COI, appear to have undergone purifying selection along the L0 and N *mt*DNA lineages, respectively (**Table 3.2.2.1**).

Human cytochrome c oxidase (also called complex IV) is composed of 13 subunits, with the largest three encoded by genes in the *mt*DNA genome and the rest by nuclear genes. Mutant *mt*DNA genes among those coding for (part or subunit/s of) the complex have been implicated in several mitochondrial disorders; severe infantile myopathies, cardiomyopathies, Lehigh syndrome [612], prostate cancer [613] and in the syndrome of mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes commonly abbreviated as MELAS [614]. Cytochrome oxidase (ComplexIV or COX complex), specifically COI is involved in the key reduction-oxidation (redox) reactions

that convert oxygen to water. In contrast, COIII does not contain any redox centers, but may play a role during the biosynthesis and maintenance of smaller subunits of the COX complex [607, 612]. Interestingly the non-synonymous mutations studied so far in the case of COI are pathologic and for COIII are mostly benign [613, 615-622] **(Tables A10.1, A10.2)**.

The effect of purifying selection ([190, 220, 243, 608-611, 623] and the current study) acting throughout the human *mt*DNA phylogenetic tree leads to relatively more loss of mutations in older branches than in younger branches [624], as shown in a recent study [609]. A highly significant inverse relationship between the ratio of nonsynonymous-to-synonymous divergence ω (dN/dS) and the age of human haplogroups was shown for comparisons of ω of human *mt*DNA haplogroups and species (human, Neanderthal and chimpanzees) as a function of the coalescence/divergence times [609]. The average ω values observed for the young haplogroups (<50 ky) were significantly higher than those estimated for the older haplogroups (>50 kya) [609]. Overall, coding regions of the *mt*DNA sequence data showed statistically significant negative values for both Tajima's and Fu & Li's D*, which may indicate that there has been demographic expansion associated with *mt*DNA lineage in African human populations [233, 235]. An inference that is consistent with a recent study [625] that showed significantly negative values for the overall values of Tajima's D or Fu and Li's F* across 20 non-coding genomic regions studied, an observation the authors argued to support the occurrence of at least a phase of population expansion among sub-Saharan Africans [625].

Table A10.1: Non-Synonymous mutations identified in COIII gene from total of 179 L0 sequences using MtPhyl [244]. Columns starting from the second column show; the position of mutation in overall *mt*DNA evolutionary tree with variants that are more ancient observed in multiple *mt*DNA haplotypes and placed in the internal branches (I), whereas those that are more recent observed a single sequence are on the terminal branches (T) of the tree [626, 627], the position in the *mt*DNA genome where the mutation took place, the number of individuals who carry the mutation among all the L0 sequences, amino acid changes and position of the change that results from the mutation, conservation index of the mutation position, specific L0 haplotypes these mutations are observed in, samples (populations) in which the mutation is observed, pathological condition so far associated with the mutation and primary source that reported the association. Deleterious mutations are usually eliminated quickly from the population by purifying selection over time and thus would be rare in the internal branches, but more common at terminal branches of the tree, while advantageous mutations retained by adaptive selection would be observed more in internal branches relative to terminal branches [626]. Conservation Index (CI) [626] is value derived by comparing these two kinds of mutations assuming that neutral mutations are uniformly distributed throughout the *mt*DNA tree [626]. CI indicates if a mutation is neutral or non-neutral (deleterious adaptive meaning the mutation may have a functional consequence). The index is from 1-100% with lower values (meaning the position is polymorphic) indicating neutral mutation while higher value (don't appear to be polymorphic) indicates non-neutral mutations. LHON - Leber hereditary Optic & peripheral neuropathy, PEG - Pseudo exfoliation glaucoma.

	NS Mutations type	Position	Mut Freq	Amino-acid Change	Conservation Index	Sub-Haplotypes	Samples/pops	Conditions	Source
1	I	9438	9	G78S	92.31	L0d1c1	Khoisan	LHON, Benign LHON, PEG	[619, 622, 628]
2	T	9210	1	T2A	79.49	L0f3	KEGK022		
3	T	9214	1	H3R	100	L0a2a2a	DQ112953		
4	T	9261	1	T19A	94.87	L0a2c	EU092913		
5	T	9300	1	A32T	30.77	L0f2a1	DQ112934		
6	T	9300	1	A32T	30.77	L0f2a1	EU092668	Benign Optic Neuritis	[617]
7	T	9670	2	N155S	84.62	L0f2a2	EU092964		
8	T	9804	1	A200T	92.31	L0b	KELO002		
9	T	9804	2	A200T	92.31	L0a1b2	DQ304897, DQ304898	Benign LHON	[618]
10	T	9966	1	V254I	79.49	L0a1c	KESB027		
11	T	9966	1	V254I	79.49	L0a1c	KETK042		

Table A10.2: Non-Synonymous mutations identified in COI gene from total of 229 N haplogroup sequences using MtPhyl [244]. Columns starting from the second column show; the position of mutation in overall *mt*DNA evolutionary tree with variants that are more ancient observed in multiple *mt*DNA haplotypes and placed in the internal branches (I), whereas those that are more recent observed a single sequence are on the terminal branches (T) of the tree [626, 627], the position in the *mt*DNA genome where the mutation took place, the number of individuals who carry the mutation among all the N sequences, amino acid changes and position of the change that results from the mutation, conservation index of the mutation position, specific N haplotypes these mutations are observed in, samples (populations) in which the mutation is observed, pathological condition so far associated with the mutation and primary source that reported the association.

	NS Mutations type	Position	Mut Freq	Amino-acid Change	Conservation Index	Sub-Haplotypes	Samples	Conditions	Source
1	I	6480	2	V193I	87.18	I2, K1a4a	EF657612, EU603401	LHON, prostate cancer	[613, 620]
2	T	5973	1	A24T	92.31	I2	EF657593		
3	T	6060	1	I53L	92.31	K1a1	AY495257		
4	T	6060	1	I53V	92.31	HV1a	EU935461		
5	T	6150	2	V83I	94.87	HV1a1	FJ210914		
6	T	6261	1	A120T	97.44	N1a	KEGB037	Leber hereditary optic neuropathy (LHON)-like optic neuropathies	[615]
7	T	6261	1	A120T	97.44	N1a	KEBR020		
8	T	6267	1	A122T	92.31	I1b	EU564849		
9	T	6345	1	F148L	100	J1b1	AY495238		
10	T	6366	1	V155I	74.36	K1a1b1a1	EF657731		
11	T	6465	1	V188L	84.62	K1a4a1	EF464682		

Appendix 11: Population Structure and population NJ tree

11.1: Population Structure

The genetic structure inferred from autosomal data seems to reflect historical interactions between populations from different language families within east Africa within the last 5 kya. The proportion of Nilo-Saharan cluster in Kenyan and Tanzanian Nilotic populations is relatively lower **(Figure 3.3.1a, 3.3.1b (iv))**. Most of the individuals of the Kalenjin speaking population (southern Nilotic – Marakwet, Sengwer, Tugen, Pokot) have admixture with Cushitic speaking populations (yellow) **(Figure 3.3.1.1a)** while the Sabaot and the Nandi show admixture with the Niger-Kordofanian speaking populations. This observation is consistent with Ehret's [87] assertion that prior to differentiation into Kalenjin and Datog – the proto–southern Nilotic speakers had a long period of extensive contact with the eastern Cushites in southwestern Ethiopia and northern Kenya, with a concurrent exchange of cultural practices. The current study demonstrates that this cultural exchange was also accompanied by genetic exchange. However, the Cushitic admixture was higher for eastern Nilotic speaking populations which includes the Turkana, and Maa speaking populations (Maasai, Ilchamus, Samburu, Il'ngwesi and Mumonyot) **(Figure 3.3.1a, 3.3.1b (iv))** which according to historical linguists, migrated into Kenya and Tanzania much later than the southern Nilotic speaking populations (about 0.5 kya compared to 3 kya) [347]. Nearly all populations in East Africa show a large amount of Cushitic admixture, with this Cushitic influence detected in populations from Kenya, Tanzania and as far as Rwanda **(Figure 3.3.1a, 3.3.1b (i) Bantu structure)**. This observation is consistent with Ehret's conclusions that

most of Eastern Africa was occupied by southern Cushitic speakers prior to later expansions of Nilotic and Bantu speakers into East Africa in the last 3 ky [115]. Afroasiatic speaking populations in Northeast and East Africa cluster together with the Cushitic speaking hunter-gatherer populations; Akie, Yaaku and El-molo **(Figure 3.3.1a)**. These hunter-gatherer populations also show Nilotic admixture **(Figure 3.3.1a, 3.3.1b (iii))**, indicating gene-flow from Nilo-Saharan speaking populations that currently neighbor them. In fact, like most hunter-gatherer populations, these populations generally appear to have changed their cultures (kinship systems, rituals) and language to conform to the organizational and behavioral mode of neighboring pastoral communities. In this case, according to cultural practices of Maa speaking populations; Maasai (Akie), Mumonyot, Il'ngwesi (Yaaku) and Samburu (El-molo) (neighboring Maa speakers and the hunter-gatherer population in bracket in each case) appear to have developed a tight symbiotic relationship in order to facilitate amicable interactions [124]. Another anomaly is the case of individuals from the Luo population, a population that speaks a western Nilotic branch of Nilo-Saharan languages, but essentially clusters among the Niger-Kordofanian speaking population of East Africa **(Figure 3.3.1a, 3.3.1b (i, iv, v))**.

11.2: Unrooted NJ trees.

To reconstruct the phylogenetic relationship between populations, population Neighbor-Joining trees [629] were constructed for *mt*DNA d-loop sequences and Y Chromosome haplotypes based on $\boldsymbol{\Phi}_{ST}$ genetic distances, and for microsatellites data based on R_{ST} genetic distances using PHYLIP (ver 3.69) [630]. The NJ trees are shown in **Figure 11.2a-c.**

The population relationship inferred from Neighbor joining tree is consistent with results obtained from population structure analysis and multidimensional scaling plots. For Autosomal data (**Figure 11.2a)**, two East African hunter-gatherer populations, Hadza and Ogiek, branch out separately from all the other groups. Most of the other populations fall into two super-branches; the first one supporting the Niger-Kordofanian (light green) while the second one supports the other two main linguistic groupings – Nilo-Saharan (blue) and Afro-asiatic (red). However, there were some populations that did not cluster according to their linguistic groupings. These include the Central African Chadic and Nilo-Saharan speaking populations that cluster with other Central/West African Niger-Kordofanian speakers. Luo, a population that speak western Nilotic branch of the Nilo-Saharan linguistic also clustered with other Niger-Kordofanian populations. The South African Khoisan speaking populations, San and !Xun/Khoe, branch off from the speakers of Niger-Kordofanian family, indicating some admixture with South African Bantu speakers. The Cushitic speaking hunter gatherer populations, El-Molo and Yaaku cluster in the Nilo-Saharan branch, indicating geneflow from neighboring Nilo-Saharan speaking populations. The Afro-asiatic speaking Baggara population cluster with the western

Nilotic speaking populations from Sudan, Nuer, Dinka, Shilluk and Nyimang. Two small sub-branches that carry collection of South Cushitic speaking populations, Iraqw, Fyome, Mbugu, Burunge and other populations, Datog, Nandi, Maasai (Nilo-Saharan), Gikuyu, Turu (Niger-Kordofanian), Sandawe (Khoisan) which have had substantial admixture with the southern Cushitic populations, stand out from the super-branch that supports Nilo-Saharan and Afro-asiatic branches.

Within Niger-Kordofanian super branch there is some loose geographical clustering; Niger-Kordofanian branch where East Africa Bantu cluster together (Tutsi, Pare & Mbugwe), while pygmy population cluster closely together, and the Central/West Africa populations including the Nilo-Saharan and Chadic speakers cluster together. In the Nilo-Saharan branch the Sudanese western Nilotic speakers and most of Southern Nilotic speakers (Sabaot, Tugen, Pokot, Marakwet and Sengwer) form two separate tight clusters. In the Afro-Asiatic branch the East African Cushitic speakers form a tight basal cluster in the branch while the Indo-European Italian and Parsi come out on terminal sub-branch of the entire branch.

For the Y chromosome lineages (**Figure 11.2b)** three branches stand out; the Afro-asiatic branch which also carry East Nilotic Samburu that indicate that this population have had substantial paternal geneflow from Cushitic speakers especially. The Niger-Kordofanian branch also has Boni and Sandawe. The other linguistic cluster consist of Southern Nilotic and an East Nilotic Ilchamus which split from their neighboring population; Tugen at the terminal sub-branch of the branch. There are two other branches with mixture of population from different linguistic grouping. The smaller

branch contains collection of Kenyan and Tanzanian Nilotic speakers and southern Cushitic speaking populations, Mbugu, Iraqw and Akie. The other branch contains a mixture of Central African Chadic and Nilo-Saharan speakers on one sub-branch and East African hunter-gatherer populations and pygmy Baka.

For maternal lineages (**Figure 11.2c)** the only linguistic clustering are three minor branches that carry the Southern Nilotic populations, Central African pygmy populations and the one that carry the East Cushitic populations from Northern part of Kenya. So overall, the effect of language is stronger in case of autosomal data and Y chromosome data than mtDNA D-loop data.

The results from the current study **(chapter 3 and this section)** indicate that the genetic correlation with geography is stronger than it is with linguistics. However, the correlation of genetic distances with linguistic/geography is stronger for Y chromosome lineages than for *mt*DNA lineages.

Figure 11.2a-c (below): Phylogenetic relationship between populations based on pairwise genetic distances: Neighbor joining tree reconstructing from pairwise populations genetic distances for a) for autosomal microsatellites data based on R_{ST} genetic distances b) Y Chromosome haplotypes and c) *mt*DNA d-loop sequences and based on $\boldsymbol{\Phi}_{ST}$ genetic distances. Populations are color coded as follows; orange (Khoisan), light green (Niger-Kordofanian), blue (Nilo-Saharan), red (Afroasiatic) and purple (Indo-Europeans).

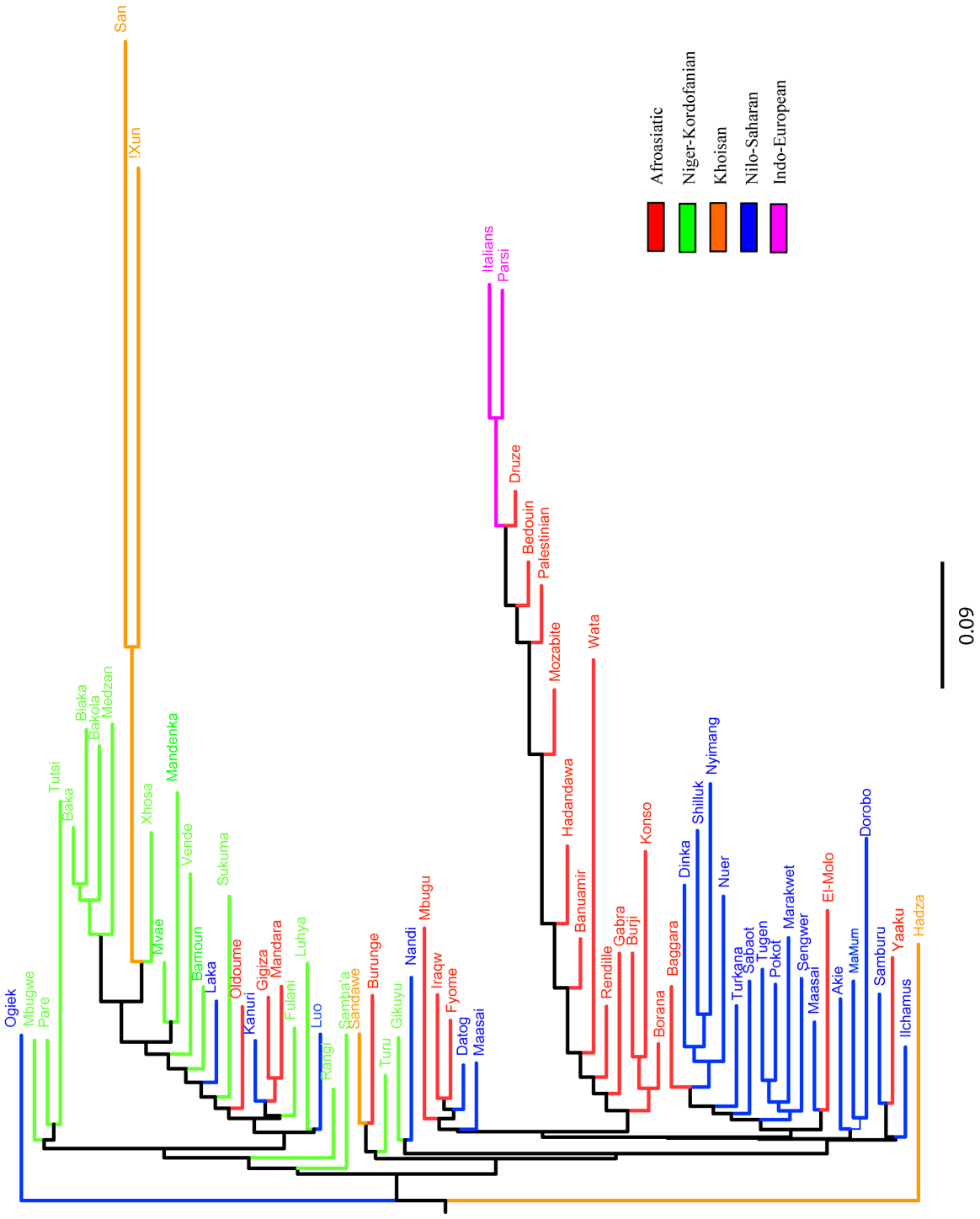


Figure A11.2a: Urooted NJ tree for autosomal data.

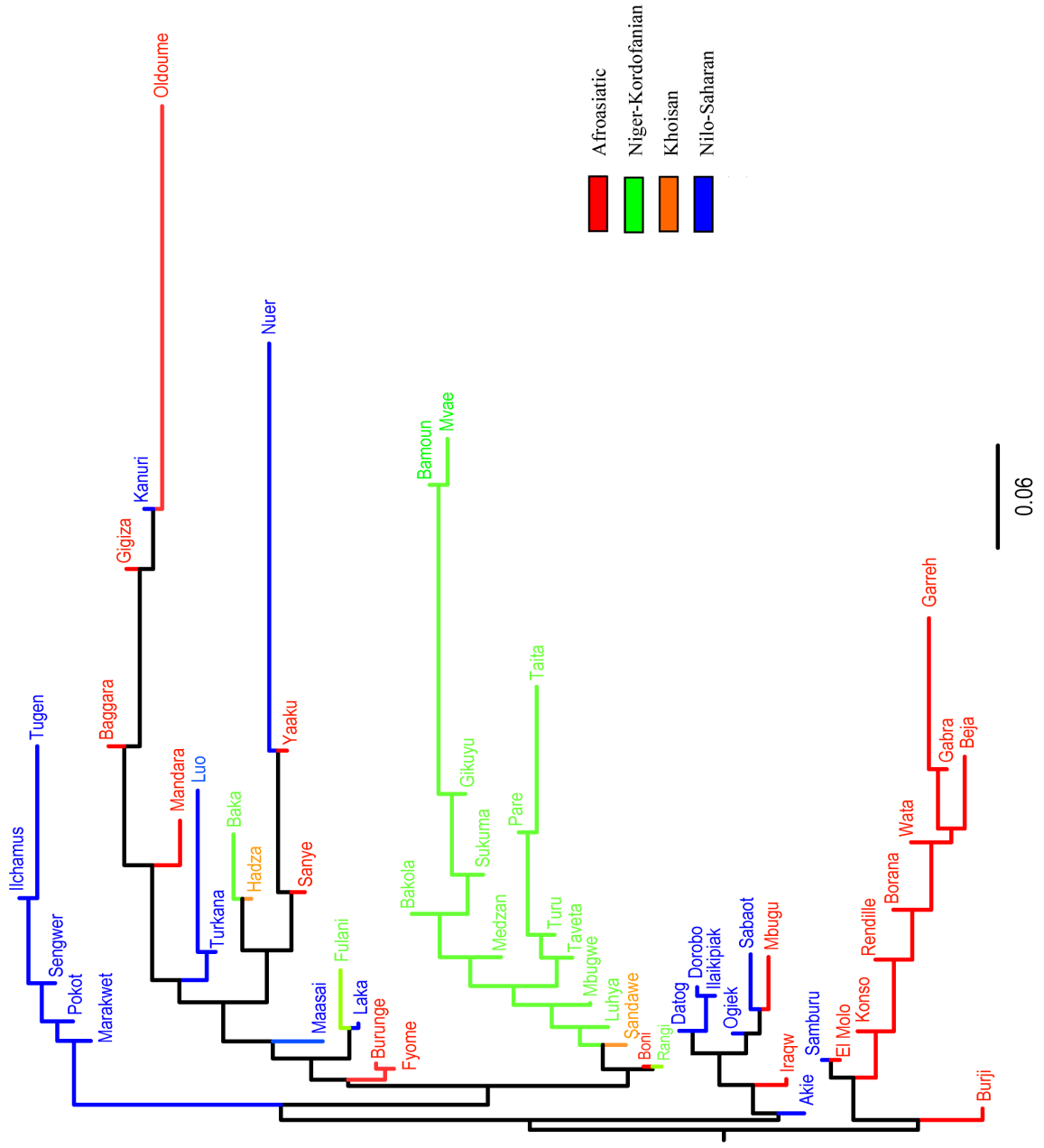


Figure A11.2b: Unrooted NJ tree for Y chromosome data.

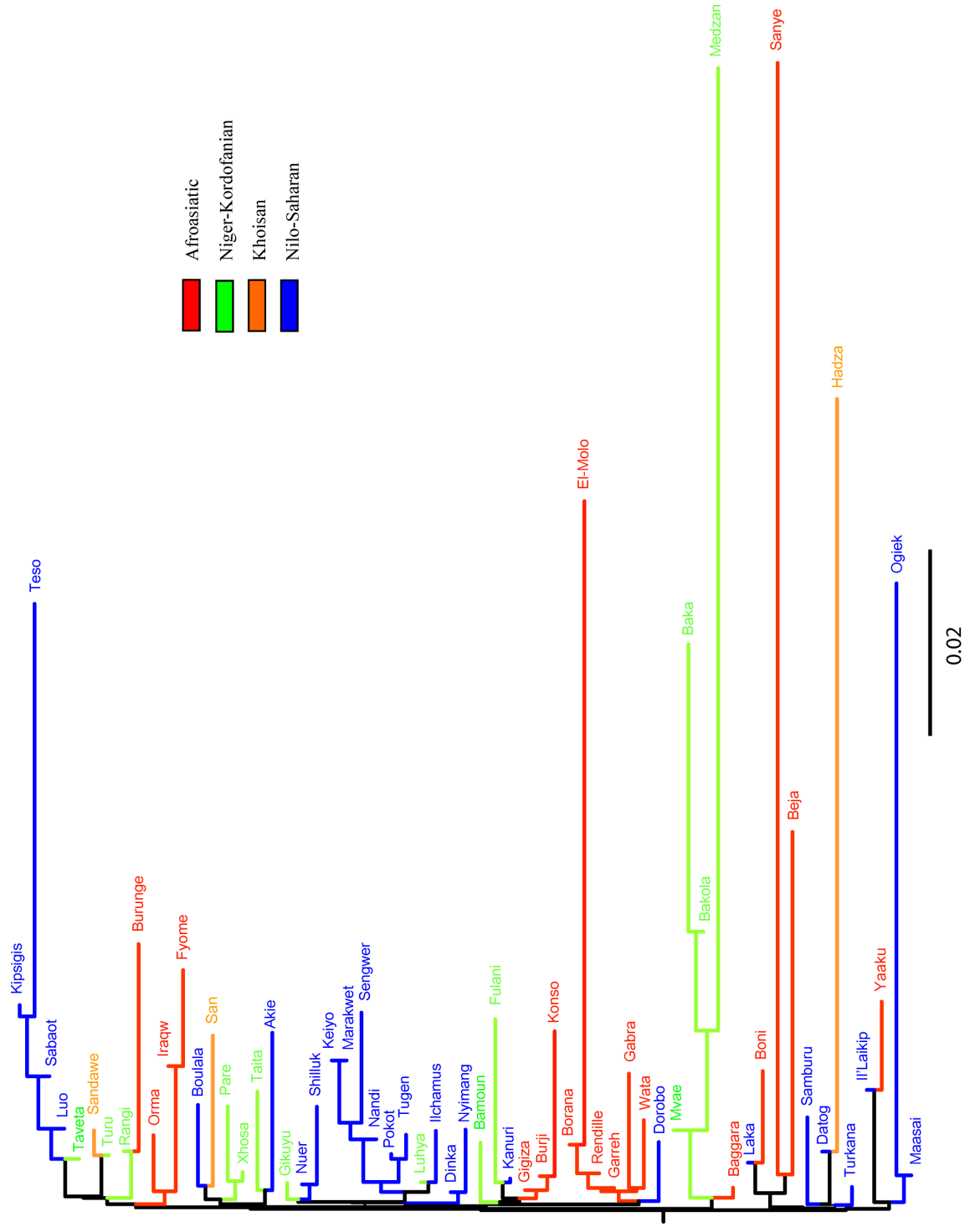


Figure 11.2c: Unrooted NJ tree for mtDNA D-loop data

Appendix 12: Global Frequency patterns of N1a, I and HV1 *mt*DNA haplotypes. Columns show the language family, name (or nationality) of each of the populations/language family, number of individuals tested, frequency of N1a, I and HV1 mitochondrial haplotypes in the populations, and the primary source reference.

Languages	Population/Country	n	N1a	I	HV1	Source
Afroasiatic	East African Cushitic[a]	369	2.71	3.52	2.17	**This study**; [16, 219, 397]
Afroasiatic	Ethio-Semitic[b]	245	2.45	0	1.63	**This Study**; [16]
Afroasiatic	NWA Arabs[c]	335	0	0.6	0.6	[227, 308-310, 398, 401]
Afroasiatic	NWA Jews[d]	289	0	4.15	3.81	[219]
Afroasiatic	Egyptians	403	0.99	3.23	3.23	[154, 228, 305, 314, 369]
Afroasiatic	Berbers[e]	680	0	0.59	0.29	[308, 310, 311, 398, 401, 402, 537, 631]
Afroasiatic	Jews[f]	338	0	3.55	0	[219]
Afroasiatic	Middle eastern[g]	1084	0.09	2.21	2.4	[219, 314, 367, 438, 442, 512, 538]
Afroasiatic	Arabians[h]	1336	2.17	2.4	2.02	[16, 219, 312, 313, 319, 439, 632]
Afroasiatic	Yemeni Jews	207	1.45	11.11	3.38	[219, 314, 367, 397]
Altaic	Russian altaics[i]	981	0.31	1.94	1.33	[223, 413, 414]
Altaic	Central Asian altaics[j]	354	0.85	4.8	1.13	[223, 413, 414, 416, 633]
Altaic	Turkic altaics[k]	306	0.33	2.61	1.31	[314, 416]
Austronesian	Sumatra	180	0	0.56	0	[322]
Indo-European	Alps[l]	218	0	1.83	0	[314, 634, 635]
Indo-European	Caucasus[m]	580	0.52	1.72	0.17	[314, 416, 512]
Indo-European	Kurdish[n]	105	0	2.86	5.71	[223, 314, 416]

Languages	Population/Country	n	N1a	I	HV1	Source
Indo-European	Mediterranean[o]	1764	0.06	1.81	0.28	[314-316, 367, 539-541, 606, 636]
Indo-European	North Central Europe[p]	545	0.18	1.83	0.18	[314, 635, 637, 638]
Indo-European	North West Europe[q]	2253	0.13	3.37	0.09	[314, 435, 443, 635, 639, 640]
Indo-European	North East Europe[r]	1356	0.74	2.95	0	[314, 635, 641-646]
Indo-European	North Caucasian[s]	208	0	1.92	0.96	[314]
Indo-European	South East Europe[t]	693	0	1.3	0.87	[314, 647-649]
Indo-European	Pakistan[u]	111		4.5		[416]
Indo-European	Iran[v]	177	0.57	3.95	2.26	[223, 314, 413, 416]
Indo-European	Portugal	449	0.22	2.84		[436, 437]
Indo-European	Scadanavian[w]	1907	0.31	4.07	0.21	[314, 444, 635, 642, 645, 650]
Indo-European	Balkans[x]	782	0.26	2.05	0.26	[445, 651, 652]
Nilo-Saharan	east African Nilotics[y]	322	0	1.24	1.55	**This study**; [151, 152, 314]
Indo-European	Canary island	300		1		[531]
Indo-European	Madeira-Azores	490		1.84		[440, 441]
Indo-European	Cape Verde	292	0.2	2.05		[318]
Niger-Kordofanian	Fulani	99		1.01		**This study**; [204, 528]

A: This Study

[a]East African Cushitics; Wata, Rendille, Borana, Burji, El-Molo, Gabra, Garreh, Iraqw, Burunge, Oromo, Ethiopian Jews

[b]Ethio-Semitic; Amhara, Tigray, Gurage, Beja

[c]NWA; North and North-West Africans (Tunisian, Algerian, Moroccan Arabs)

[d]NWA Jews; North and North-West African Jews (Libyan, Algerian, Tunisian, Moroccan)

[e]Berbers speaking populations from Tunisia, Libya, Egypt, Morocco, Western Sahara, Niger, Burkina Faso and Mali

[f]Jews; Jews from Azerbaijan, Bulgaria, Georgia, Iraqi

[g]Middle eastern; Bedouin, Druze, Iraqis, Syrian, Palestinians, Jordans

[h]South Arabians; Saudi Arabians, Yemenis, Emirates from Dubai (UAE), Soqotri, Yemenite jews

[i]Russian altaics; Khizki, Buryats, Tuvinian, kalmyks, Karakalpaks, Khamnigans, Shors, Tartar, Telenghits

[j]Central Asian altaics; Tajiks, Kyrgyz, Kazakhs, Turkmens, Uzbeks

[k]Turkic altaics; Azerbaijanis, Turks

[l]Alps; Swiss, Bavarians, Austrians

[m]Caucasus; Georgians, Armenians

[n]Kurdish; From Iran and Turkmenistan

[o]Mediterranean; Spainiads, Italians, Greeks, Albanians

[p]North central Europe; Poles, Czechs, Germans, Danes

[q]North West Europe; French, English, Welsh, Irish, Scottish

[r]North East Europe; Russians, Finns, Estonians, Lithuanians, Latvians

[s]North Caucasian; Ossetian, Chechens, Kabardians, Adygei

[t]South East; Bulgarians, Hungarians, Romanians

[u]Pakistan; Sindhi, Pathan, Burusho

[v]Iranians; Gilaki, Persians, Mazandarins

[w]Scandanavians; Norwegians, Swedish, Icelandic

[x]Balkans; Bosnians, Slovenians, Macedonians, Slovakians

[y]east African Nilotics; Maasai, Il'Laikipiak, Samburu, Turkana, Tugen, Nubia

Appendix 13: Frequency of *mt*DNA haplotypes J, K and R0a in Global human populations. Columns show name (or nationality) of each of the populations/language family, number of individuals tested, frequency of J, K and R0a mitochondrial haplotypes in the populations, and the primary source reference.

Populations	n	J	K	R0a	Sources
East African Cushitics[a]	370	0.73	1.7	6.81	**a**; [16, 151, 152, 219, 397]
East African Nilo-Saharan[b]	153	3.27	1.96	5.23	**a**; [154, 314]
Niger-Kordofanians[c]	206	3.88	0.97	0	[292, 528]
Ethio-Semitic[d]	194	2.06	1.03	0	[16]
Egyptians	387	7.49	4.13	1.81	[154, 228, 305, 314, 369]
Berbers[e]	768	4.69	4.56	0.78	[171, 308, 310, 311, 399, 401, 402, 537, 631]
North/North Western Africans[f]	535	4.86	4.3	1.5	[227, 306, 308-310, 398, 401, 536]
North/Northwestern African Jews[g]	289	6.57	9.69	2.77	[219]
Canary Island	300	7	4	0	[531]
Cape Verde	292	0	0	0.34	[318]
Madeira-Azores	490	6.12	4.69	0.2	[440, 441]
Middle Eastern[h]	1226	10.28	5.55	3.51	[171, 219, 314, 367, 438, 442, 512, 538]
South Arabians[i]	1424	15.24	4.07	8.36	[16, 219, 312-314, 319, 367, 397, 439]
Anatolians[j]	652	8.9	5.37	1.53	[223, 314, 416, 653-655]
Anatolian Jews[k]	340	16.18	3.24	0.29	[219]
Caucus[l]	337	8.01	6.53	0.89	[314, 416]
Caucus Jews[m]	203	21.67	6.4	0.99	[219]
Northern Caucus[n]	208	7.69	3.85	0	[171, 314]
Central Asian[o]	452	4.2	3.98	1.11	[223, 414, 416]
Eastern Mediterranean[p]	577	10.23	7.11	1.56	[314, 636, 656]
Italians	797	7.9	9.28	0.5	[314, 367, 540, 541]
Western Mediterranean[q]	1404	7.76	6.84	0.5	[314-316, 399, 436, 437, 605, 657]
Basque	156	2.56	4.49	0	[314]
Southeastern Europe[r]	1583	7.39	5.94	0.63	[314, 648, 649, 651, 652, 655]
Alps[s]	218	11.47	5.5	0.46	[314, 634, 635]
Northern Central Europeans[t]	838	8.47	4.89	0.6	[314, 635, 644, 658, 659]
Northwestern Europeans[u]	1465	13.45	8.05	0.2	[314, 435, 443, 635, 639, 660]
Northeastern Europeans[v]	1214	7.08	3.38	0	[314, 635, 641, 643-645]
Scadanavians[w]	1865	7.02	3.86	0	[314, 444, 635, 641, 642, 645, 650]
Southern Siberians[x]	1131	2.12	1.68	0	[223]
Pakistan	432	4.63	0.93	3.47	[416]
Mongolians	47	0	2.13	0	[223]
Gujarati	34	2.94	0	0	[416]

Adjusted TMRCA estimates [190] for these haplotypes, J, K and R0a, are 31.5, 15.9 and 29.4 ky respectively. Early European hunter-gatherer populations might have been predominantly from the *mt*DNA H, V, U4 and U5 [406, 661] haplotypes, thus these haplotype (J, K and R0a) might have been carried into Europe by Neolithic farmers 7 kya [404, 406], conclusion consistent with demic diffusion model [552]. The fossils of putative Neolithic farmers have elevated frequencies of J and K haplotypes [661]. Considering that populations from the Near East exhibit the highest diversity for this haplotypes, the moderate frequency in some European populations (that are comparable for those in the Near East) might be due to genetic drift.

a; This study

East African Cushitics[a]: Beja, Borana, Gabra, Garreh, Iraqw, Rendille, Wata, Yaaku, Somalis, Afar, Ethiopia Jews, Oromo

East African Nilo-Saharan[b]: Datog, Dorobo, Samburu, Nubians

Niger-Kordofanians[c]: Cabindans, Fulani (Burkina Faso)

Ethio-Semitic[d]: Amhara, Gurage, Tigray

Berbers[e]: Mozabites, Siwi, Saharawis, Moroccan and Tunisian Berbers, and Tuareg populations from Niger, Libya, Burkina Faso and Mali

North/Northwestern Africans[f]: Algerians, Mauritanians, Tunisians, Moroccan Arabs

North/Northwestern African Jews[g]: Algerian, Moroccan, Tunisian and Libyan Jews

Middle Eastern[h]: Bedouins, Druze, Iraqis, Jordanians, Palestinian, Syrians

South Arabians[i]: Saudi Arabians, Dubai, Yemenis, Yemenite Jews

Anatolians[j]: Turks, Iranians, Kurds

Anatolian Jews[k]: Turkish, Iranian and Iraqi Jews

Caucasus[l]: Armenians, Azerbaijans and Georgians

Caucasus Jews[m]: Georgian, Bulgarians, Azerbaijan Jews

Northern Caucasus[n]: Ossetian, Chechens, Kabardians, Adygei

Central Asian[o]: Tajiks, Kazakhs, Uzbeks, Karapalks, Turkmen

East Mediterranean[p]: Greeks, Albanians, Cypriots

Western Mediterranean[q]: Spanish and Portuguese

Southeastern Europe[r]: Bulgarians, Romanians, Hungarians, Roma from Hungary, Slovakians, Macedonia

Alps[s]: Swiss, Bavarians and Austrians

Northern Central Europeans[t]: Germans, Poles, Czech, Danes

Northwestern Europeans[u]: English, French, Scottish, Irish, Welsh

Northeastern Europeans[v]: Latvians, Estonians, Luthanians; Russians, Finnish

Scadanavians[w]: Swedish, Norwegians; Icelanders

Southern Siberians[x]: Kalmyks, Buryats, Khamnigans, Tuvinians, East Evenks, Shors, Khakassians, Altaians-Kizhi, Telenghits, Teleuts

Appendix 14: Human demographic expansions inferred from mitochondrial genome sequences

Initial major expansions in *mt*DNA lineages seem to correspond to 80-90 kya [273], a period before expansion of human population outside Africa and subsequent expansions of *mt*DNA lineages happened 55, 35, 22 and 12 kya **(Table 3.6.1 – Figures 3.6.1a-h, A14.2)**. The expansion that occurred 55 kya corresponds to a period of expansion of modern humans outside Africa [39]. Interestingly previous studies to determine the time of human population using *mt*DNA sequence in Africa have all yielded dates/times of, 110 - 70 kya [46, 271, 662, 663] and 60 – 20 kya [271, 663], that correspond with period of warm paleoclimatic conditions. Moreover, in a recent study, Cox *et al.,* [664] determined the timing of human population expansion by performing a multilocus analysis of over 20 unlinked autosomal non-coding regions among individuals from seven global human populations, four of which are Africans [664]. The data from all four sub-Saharan African populations fit a simple two-phase growth model (populations experienced exponential growth after a period of constant size), while populations from outside Africa did not fit with the simple growth model, speculated by the authors to be due to their more complex demographic histories that include bottlenecks [664]. In fact, among non-Africans genetic evidence of one (in Eurasians) or two (Americans) [665] major bottlenecks [666], the 'out-of-Africa' and plus the one through the Bering Strait respectively, have so far been demonstrated [666]. Consistent with the previous study [271] the African populations analyzed in the Cox *et al.,* [664] study best fit models of population growth that begin in the late Pleistocene 30 – 40 kya [664].

Previous studies of the *mt*DNA sequences (control region [46], *mt*DNA cytochrome c oxidase subunit III (COIII) gene [667]) indicated different signals of population expansion among African food-producing and hunter-gatherer populations

[46, 667]. Food-producing populations exhibited pattern consistent with past demographic expansions, while most hunter-gatherer showed pattern recent contraction in effective population size, speculated to have occurred because of food producing populations expanded into their territory [46, 667]. It is interesting that the BSP plots for *mt*DNA lineages that have been associated with hunter-gatherers in the current study; L0d (**Figure A14.3c**), L0f (**Figure A14.3d**), L1c (**Figure A14.3f**), L3a (**Figure A14.3h**) and L4 (**Figure 3.4.3.1f**) all show a signal of decrease or constancy in effective population sizes from around 2 kya. In contrast, BSP plots for *mt*DNA lineages that have been associated with food producing populations; Bantu agriculturalist, L3b (**Figure A14.3i**), L3d (**Figure A14.3j**) and L3e (**Figure A14.3k**), pastoralist – Nilo-Saharans – L5 (**Figure 3.4.3.1g**) and L3h (**Figure A14.3m**), and Afroasiatics – L3x (**Figure A14.3n**) and M1a (**Figure A14.3o**), exhibit continuous increase in effective population sizes from either 10 or 5 kya.

Several hypotheses propose that environmental changes that have been due to paleoclimatic conditions in the past were responsible for hominin speciation, shift to bipedality, increase in cranial capacity, behavioral adaptability, cultural innovations, and intercontinental migration events ([668] and references therein). These hypotheses are based on correlations between global-scale climate shifts documented in ice core/oceanic/lake sediments events in hominin evolution recorded in continental fossil bearing strata [668]. Paleoclimatic conditions are reconstructed using computer models from a combination of climatic imprints called proxies that include stable isotope **(Figure A14.1)**, pollen counts and mineral compositions in the ice core/oceanic/lake sediments, and ancient vegetation which are correlated with climatic fluctuations. In this study I will use the proxy Insolation (combination of the words *in*cident *sol*ar radi*ation,* a measure of solar radiation energy received on a given

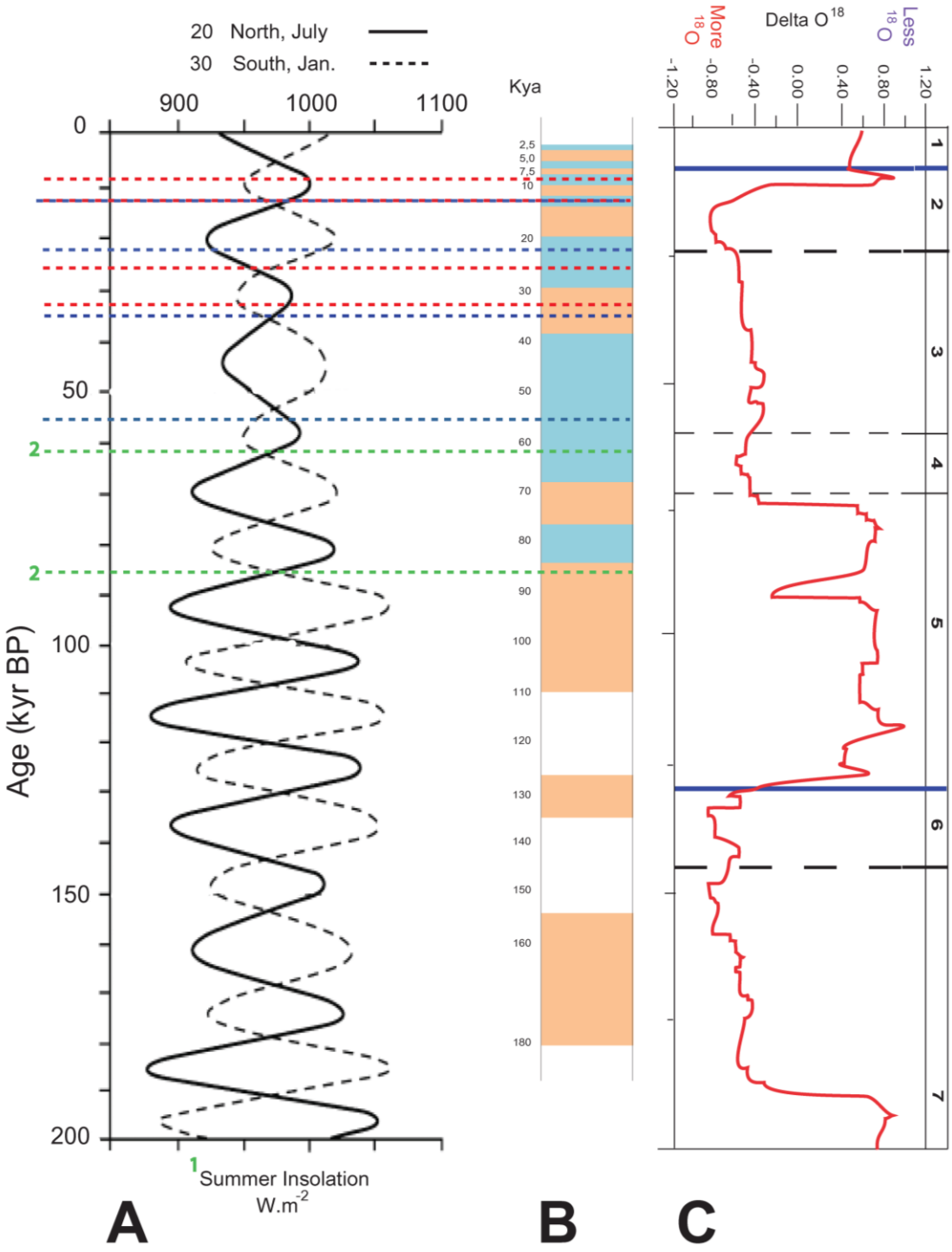
surface area in a given time). Earth surface temperature is correlated with amount of insolation received, and warm periods usually correspond to enhanced global precipitations **(Figure A14.1)**. The human habitat shrunk during cold periods and increased during warmer periods, with ancestral hunting and gathering populations expanding and contracting in response to shifting resource opportunities as the climate changed [145].

As indicated by several global paleoclimatic parameters, Greenland ice core records, Marine isotope stages and pollen count, climatic conditions in Africa were cold and dry during the Last Glacial Maximum (LGM) (after 24 kya until 15 kya) with a short wet hiatus recorded at some sites around 18 kya [143, 669-671]. The warming began early in Africa, 15-14 kya, with wet conditions reported until 12 ky before a short dry cold period around 13-12 kya that correspond to the Younger Dryas [669]. All the lineages in Africa seem to have expanded in the last 20 kya, with simulation results indicating major demographic signatures of expansion during the Holocene, 8-10 kya, 5-6 kya and most recently around 2-3 kya **(Table 3.6.1 – Figures 3.6.1a-h, A14.2).** Interestingly, these time periods of *mt*DNA related demographic expansion **(Figure 3.6.1, Table 3.6.1)** correspond to periods of increased global insolation (solar radiation received) **(Figures A14.1)**. Increased insolation is associated with increased precipitation that has been argued to correspond with periods of increase in population sizes and diversity [48]. During the late Pleistocene and Holocene, wet conditions are reported for periods 11.5 – 13 kya, 8.5 – 10.5 kya, 6.5 – 5.0 kya and 3.5 – 2.5 kya with a notable dry interregnum around 8.2 – 7.8 kya [143, 358, 670, 672-674].

Moreover, late Pleistocene and Holocene human population expansion might also correspond to periods of cultural innovation and change in subsistence patterns in

Africa. This hypothesis is consistent with evidence that showed increased postcranial biomechanical variability in early modern humans from the European Upper Paleolithic (UP), 40–20 ky BP ([675] and references therein) and Late Upper Paleolithic (LUP) of North Africa and Southeast Asia, 20–10 ky BP [675]. This observation has been interpreted as evidence that ecological changes associated with the last glacial maximum (LGM) played a critical role in cultural and biological adaptation. Shackelford *et al.,* [675] concludes that such patterns suggest changes in subsistence behavior and mobility after the LGM across the Old World most consistent with reduced mobility and broad-spectrum resource exploitation.

Figure A14.1: Paleoclimatic summer insolation/precipitations correspondence with *mt*DNA lineages' expansion dates. Curve for Summer insolation **(A)** was adapted from [143, 676] (full lines represent Northern hemispheric summer and the dotted lines for Southern hemispheric summer). Estimate of expansion events for 80-90 and 60 kya (green dotted lines 2) is as estimated in [273]. Red and blue dotted lines represent expansion time as estimated by analysis of individual sub-Saharan haplotypes and all sub-Saharan sequences respectively **(Table 3.2.3.1; Figures 3.2.3.1).** The light blue and orange "histogram" (blank – condition not determined) **(B)** represent wet and dry conditions, respectively, collated from published literatures: [143, 146, 147, 358, 492, 669, 670, 673, 674, 677, 678]. The curve on the right **(C)**, adapted from Martinson [679] represents the Marine isotope stages (MIS) up to 200 kya. MIS are alternating warm and cool periods in the Earth's paleoclimate, deduced from oxygen isotope data reflecting temperature curves derived from data from deep sea core samples. The cycles were found to correspond to terrestrial evidence of glacial and interglacial, meaning cooling and warming respectively. Each stage represents a glacial (even numbered) or interglacial (odd-numbered), starting from the present and working backward in time.


20 North, July
30 South, Jan.
900
1000
1100
0
50
100
150
200
Age (kyr BP)
2
2
1
Summer Insolation
W.m-2
A
Kya
2,5
5,0
7,5
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
180
B
Delta O18
More 18O
Less 18O
-1.20
-0.80
-0.40
0.00
0.40
0.80
1.20
1
2
3
4
5
6
7
C

A: Mean effective pop. size		B: mid-month insolation 15N for July in W/m^2			
Time in ya	**Pop. size**	**Years in kya**	**Insolation in W/m^2**	**Years in kya**	**Insolation in W/m^2**
Current	1.22E+07	0 (1950)	440.6	101	484.16
581	1.22E+07	1	442.48	102	490.57
1163	1.22E+07	2	445.62	103	494.36
1744	1.21E+07	3	449.77	104	495.08
2325	1.20E+07	4	454.6	105	492.54
2906	1.18E+07	5	459.67	106	486.88
3488	1.13E+07	6	464.53	107	478.53
4069	1.07E+07	7	468.75	108	468.21
4650	9852294	8	471.92	109	456.83
5231	8761044	9	473.76	110	445.39
5813	7550571	10	474.08	111	434.9
6394	6305471	11	472.84	112	426.28
6975	5094813	12	470.17	113	420.25
7556	4038352	13	466.28	114	417.31
8138	3191892	14	461.52	115	417.7
8719	2458812	15	456.31	116	421.39
9300	1835430	16	451.09	117	428.05
9881	1361442	17	446.3	118	437.14
10463	1030158	18	442.31	119	447.89
11044	847298	19	439.45	120	459.38
11625	765677	20	437.91	121	470.63
12207	727203	21	437.81	122	480.68
12788	708843	22	439.11	123	488.69
13369	700546	23	441.68	124	494
13950	695087	24	445.26	125	496.23
14532	689078	25	449.54	126	495.28
15113	680536	26	454.14	127	491.35
15694	667827	27	458.66	128	484.87
16275	646160	28	462.73	129	476.47
16857	612847	29	466.05	130	466.91
17438	571315	30	468.37	131	456.96
18019	523982	31	469.59	132	447.4
18600	471686	32	469.67	133	438.89
19182	415209	33	468.69	134	431.99
19763	367683	34	466.81	135	427.11
20344	327590	35	464.24	136	424.47

A: Mean effective pop. size		B: mid-month insolation 15N for July in W/m^2			
Time in ya	**Pop. size**	**Years in kya**	**Insolation in W/m^2**	**Years in kya**	**Insolation in W/m^2**
20925	295566	36	461.2	137	424.17
21507	273171	37	457.94	138	426.13
22088	255864	38	454.66	139	430.12
22669	242460	39	451.57	140	435.8
23250	232841	40	448.8	141	442.72
23832	224651	41	446.48	142	450.36
24413	216974	42	444.69	143	458.16
24994	209818	43	443.49	144	465.57
25576	201889	44	442.91	145	472.07
26157	193042	45	442.95	146	477.24
26738	183703	46	443.63	147	480.77
27319	172525	47	444.91	148	482.48
27901	160286	48	446.77	149	482.33
28482	148118	49	449.14	150	480.42
29063	137817	50	451.95	151	476.95
29644	128214	51	455.08	152	472.21
30226	119864	52	458.39	153	466.55
30807	113663	53	461.72	154	460.33
31388	108441	54	464.88	155	453.94
31969	104323	55	467.65	156	447.75
32551	101801	56	469.84	157	442.08
33132	99818	57	471.23	158	437.24
33713	98138	58	471.67	159	433.49
34294	96979	59	471.03	160	431.03
34876	96072	60	469.28	161	430
35457	95406	61	466.45	162	430.51
36038	94826	62	462.67	163	432.57
36620	94238	63	458.15	164	436.14
37201	93532	64	453.19	165	441.07
37782	92674	65	448.14	166	447.15
38363	91756	66	443.37	167	454.08
38945	90589	67	439.28	168	461.47
39526	89296	68	436.2	169	468.86
40107	87841	69	434.43	170	475.75
40688	86042	70	434.17	171	481.63

A: Mean effective pop. size		B: mid-month insolation 15N for July in W/m^2			
Time in ya	**Pop. size**	**Years in kya**	**Insolation in W/m^2**	**Years in kya**	**Insolation in W/m^2**
41270	83742	71	435.52	172	486.02
41851	81003	72	438.46	173	488.5
42432	78229	73	442.85	174	488.78
43013	75250	74	448.41	175	486.73
43595	72092	75	454.79	176	482.38
44176	68855	76	461.51	177	475.96
44757	65345	77	468.08	178	467.91
45338	61883	78	473.96	179	458.77
45920	58286	79	478.66	180	449.23
46501	55034	80	481.76	181	440
47082	52110	81	482.95	182	431.81
47664	49003	82	482.08	183	425.31
48245	46274	83	479.15	184	421.05
48826	43915	84	474.35	185	419.41
49407	41789	85	468.01	186	420.62
49989	39862	86	460.58	187	424.69
50570	38121	87	452.62	188	431.41
51151	36542	88	444.72	189	440.37
51732	35203	89	437.46	190	450.93
52314	33910	90	431.41	191	462.3
52895	32946	91	427.04	192	473.55
53476	32157	92	424.72	193	483.71
54057	31523	93	424.68	194	491.86
54639	30979	94	427.02	195	497.23
55220	30572	95	431.66	196	499.29
55801	30241	96	438.39	197	497.81
56382	29928	97	446.78	198	492.91
56964	29716	98	456.29	199	485.02
		99	466.22	200	474.83
		100	475.78		

Table A14.1: Table of effective population size estimates based on Bayesian Skyline plot from Sub-Saharan mtDNA sequences **(columns A)** as implemented in program BEAST ver4.10, and corresponding inferred mid-month insolation at 15N for July in W/m^2 **(columns B)** (data is from Berger 1991 [676].

Human population expansion is inferred here to have expanded during warm wet periods in the past. To make such an inference proxy data, including past inferred insolations were compared with period of increase in human effective populations. Below are the steps used to generate comparisons a portion in **Figure A14.1.**

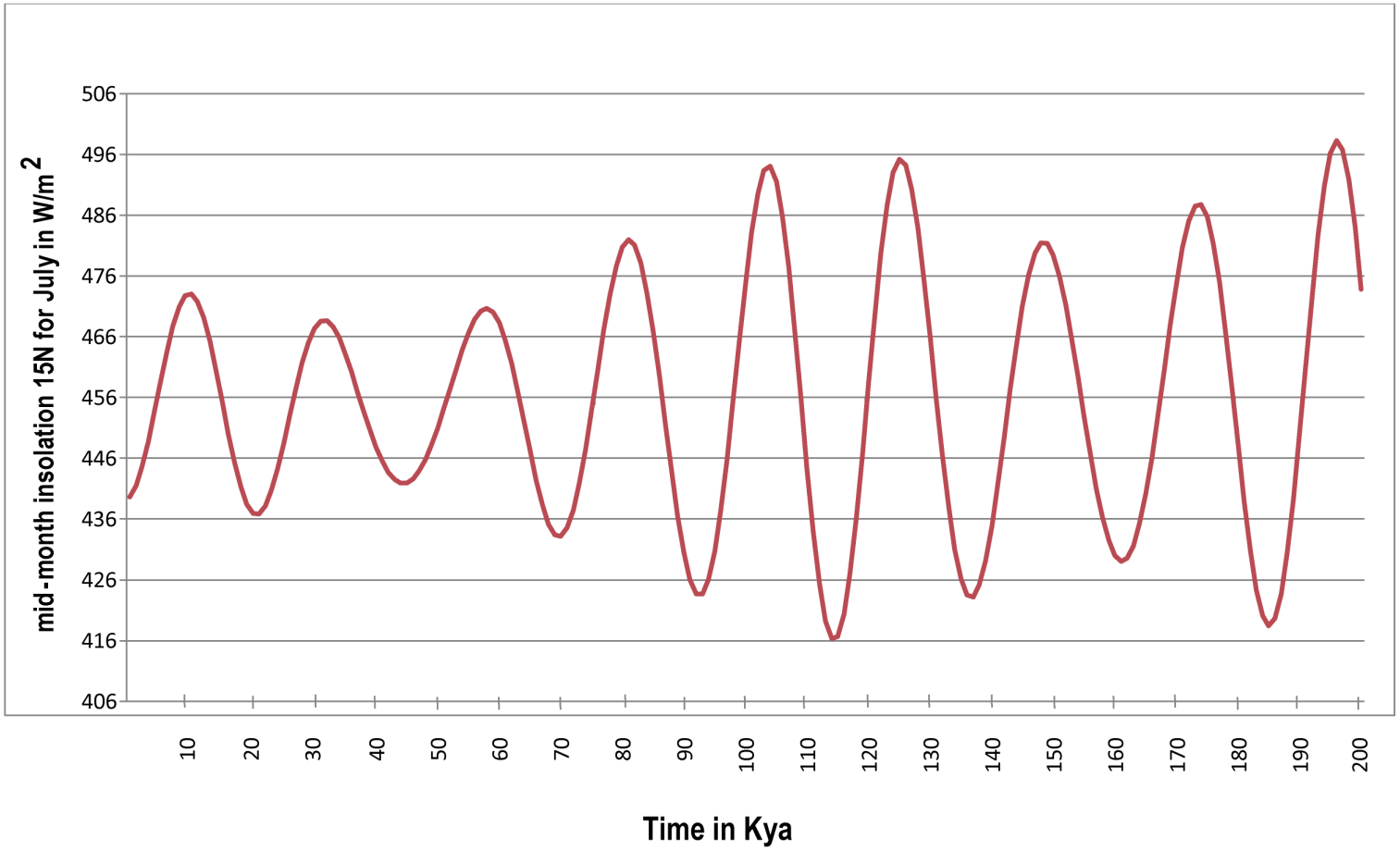


Figure A14.2a: Graph of inferred mid-month insolation at 15N for July in W/m^2

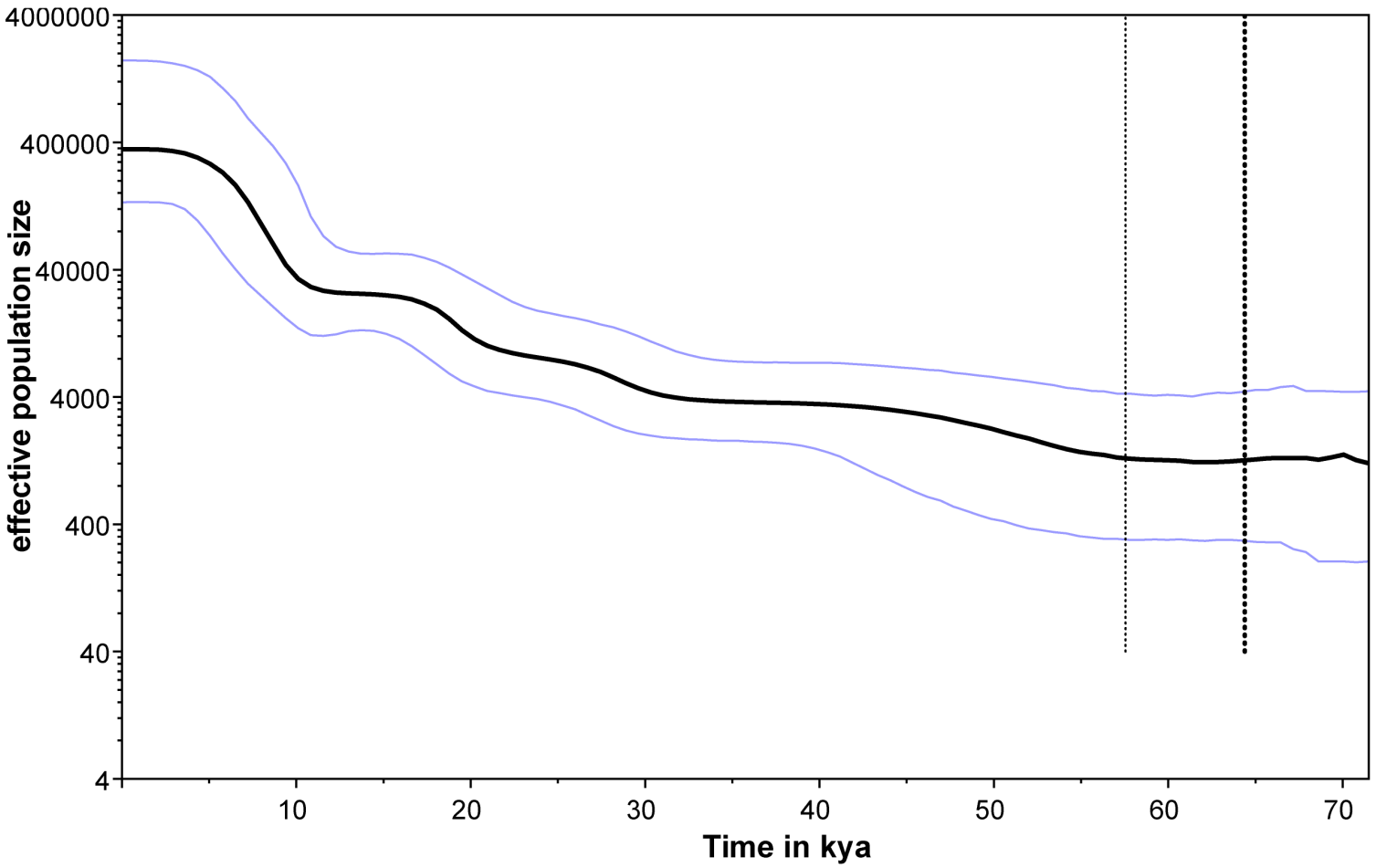


Figure A14.2b: Bayesian Skyline Plots (BSPs) [267-270] of effective population size through time for *mt*DNA sequences from Sub-Saharan Africa **(Table A14.1)**. The bold black line represents the median posterior effective population size through time. The blue lines delimit the 95% highest posterior density for effective population size, accounting for uncertainty in the reconstructed phylogeny and substitution model parameters. Effective population size is plotted on the X axis assumes a generation time of 25 years as used elsewhere in population genetic studies [159, 263-266]. The two black dotted vertical lines demarcate the 95% period of initial onset of population growth.

Then portion of insolation graph that correspond with BSP plot that is up to 50 kya was plotted so that it can be compared with the effective population graph, and subsequently slopes of the two graphs compared. High insolation values are estimated 30 and 10 kya, and these are reflected in increase of the slope in this parameter during these two periods **(Figure 14.2ci).**

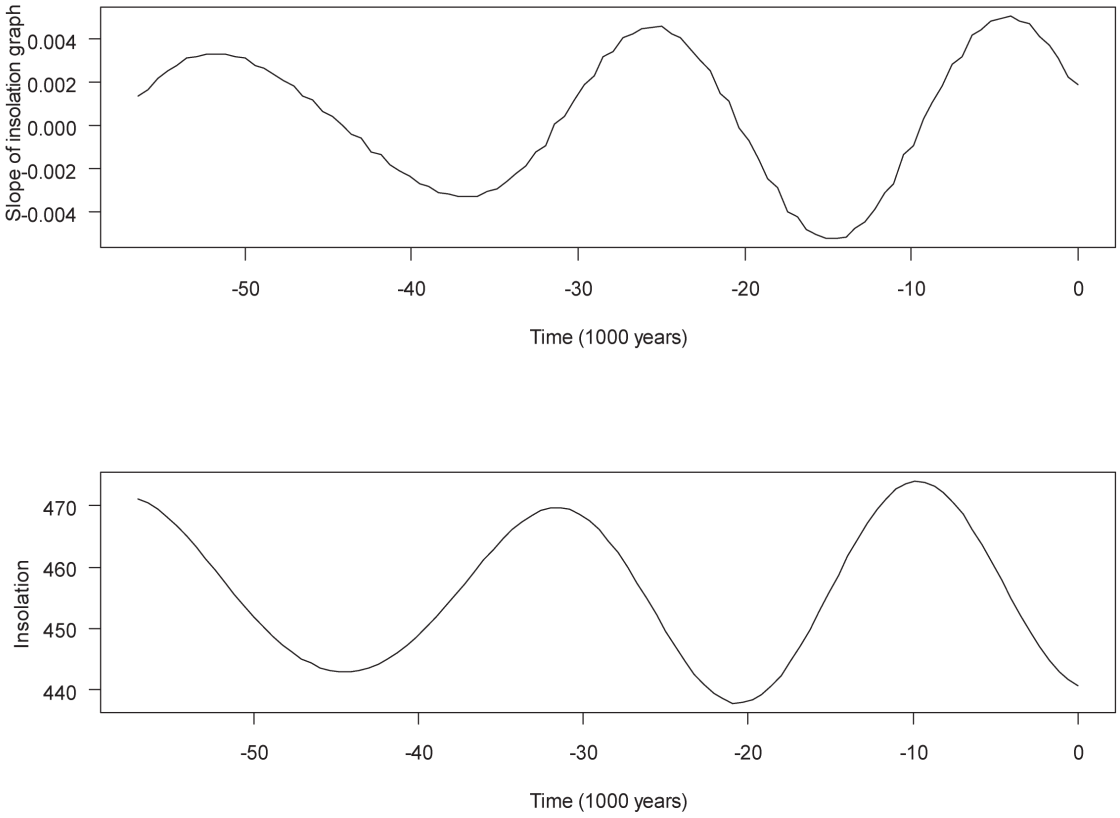


Figure A14.2ci: Graph of insolation received in mid-month July 15N for the last 50 kya and slopes at different time periods in the graph (top)

BSP indicates that the highest increase in effective population size happened around 10 kya **(Figure A14.2cii bottom),** with slope of the graph peaking just after 10 kya **(Figure A14.2cii top).** This was done for each of the lineage BSP in **Figure 3.6.1** And **Figures A14.3** to draw **Figure 14.1** comparisons.

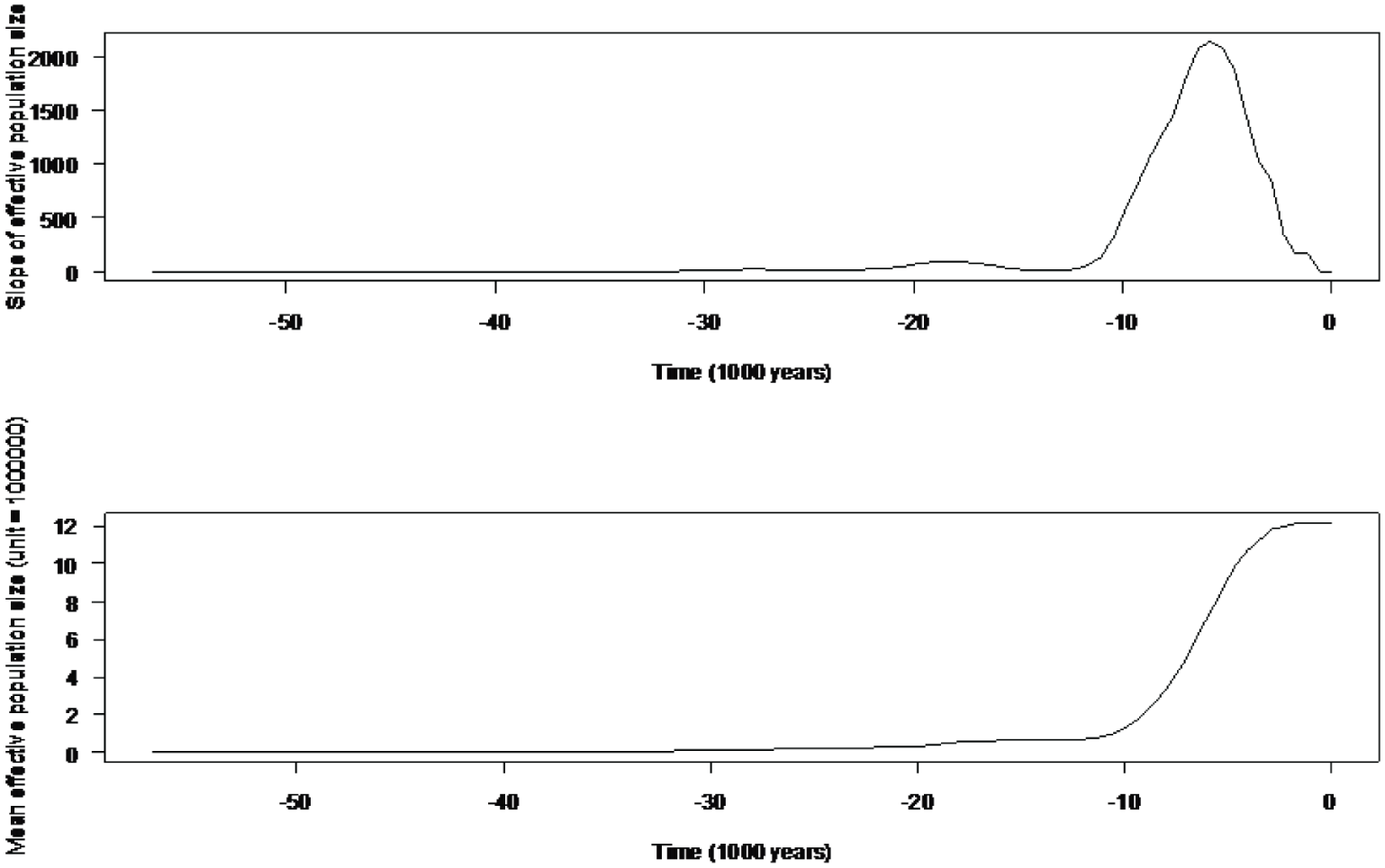
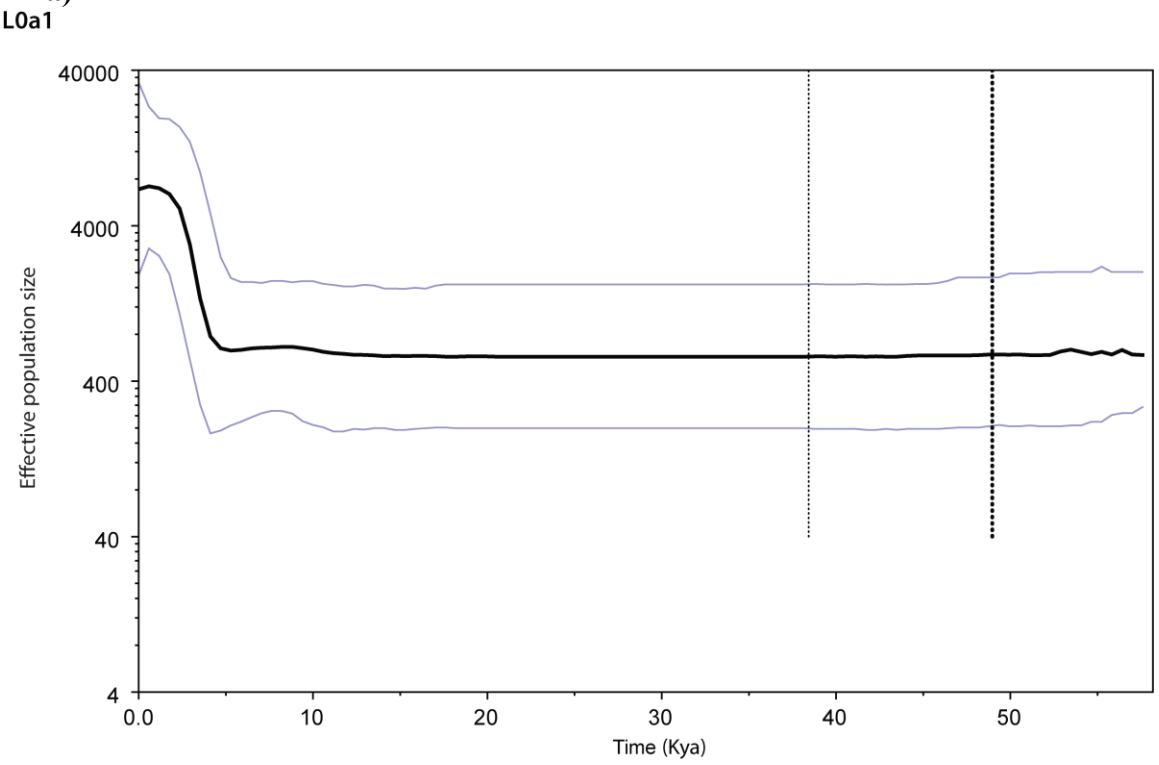


Figure A14.2cii: BSP [30-33] of effective population size through time for *mt*DNA sequences from Sub-Saharan Africa and slopes at different time periods in the graph (top)

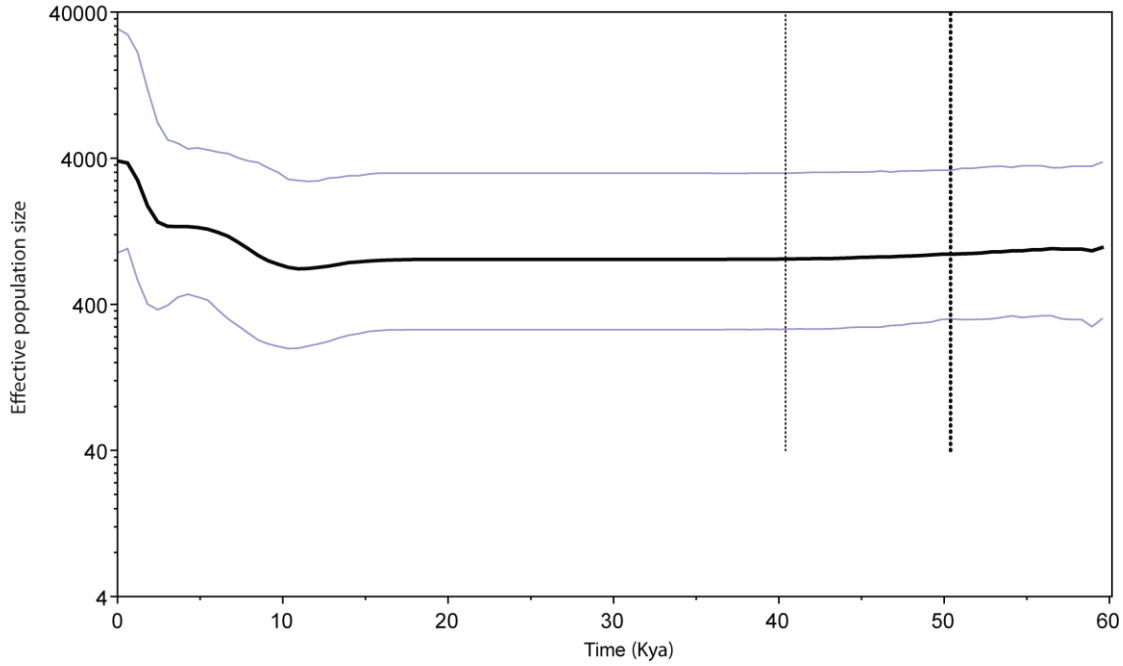
Figures A14.3: Bayesian Skyline Plots (BSPs) [267-270] of effective population size through time for haplotypes (a) L0a1 (n=49) (b) L0a2 (n=16), (c) L0d (n=26), (d) L0f (n=19) (e) L1b (n=40) (f) L1c (n=52) (g) L2a1 (n=116) (h) L3a (n=11) (i) L3b (n=35) (j) L3d (n=21) (k) L3e (n=77) (l) L3f (n=30) (m) L3h (n=19) (n) L3x (n=13) (o) M1a (n=93) (p) U6 (n=45). The bold black line represents the median posterior effective population size through time. The blue lines delimit the 95% highest posterior density for effective population size, accounting for uncertainty in the reconstructed phylogeny and substitution model parameters. Effective population size is plotted on the X axis assumes a generation time of 25 years as used elsewhere in population genetic studies [159, 263-266]. The two black dotted vertical lines demarcate the 95% period of initial onset of population growth.

a)



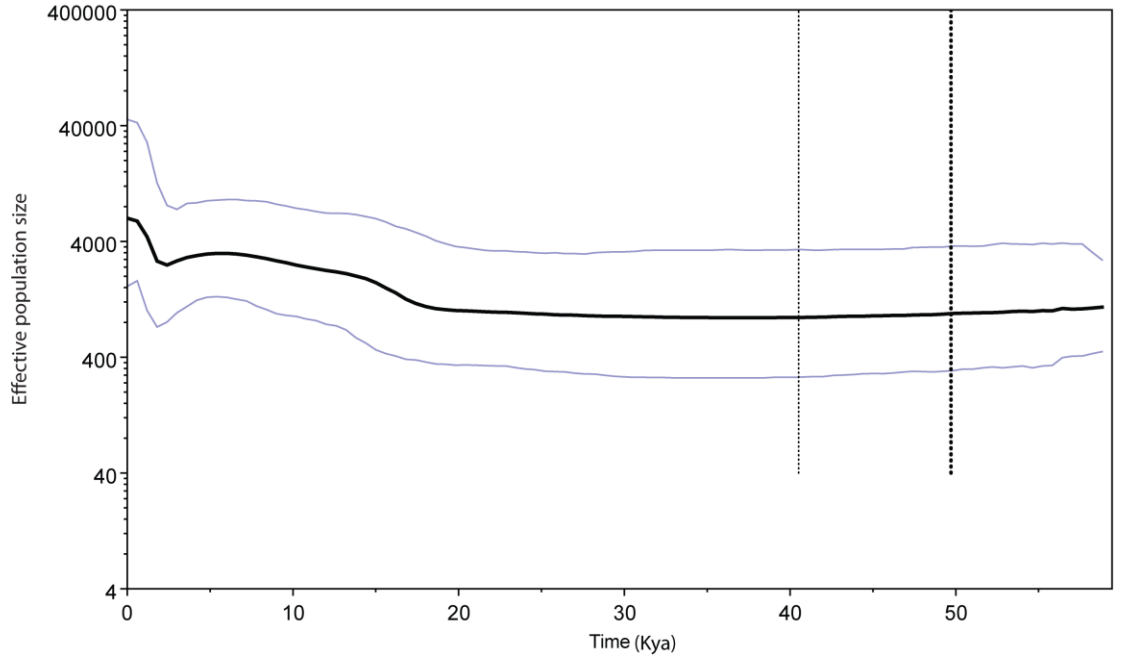
b)

L0a2


40000
4000
400
40
4
Effective population size
0
10
20
30
40
50
60
Time (Kya)

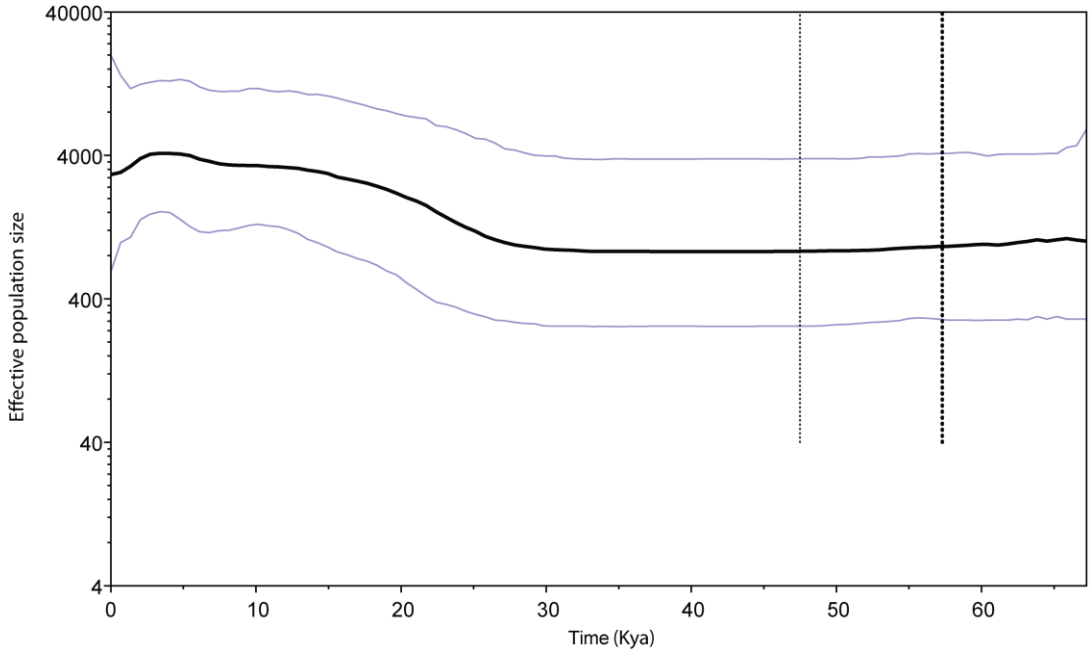
c)

L0d


400000
40000
4000
400
40
4
Effective population size
0
10
20
30
40
50
Time (Kya)

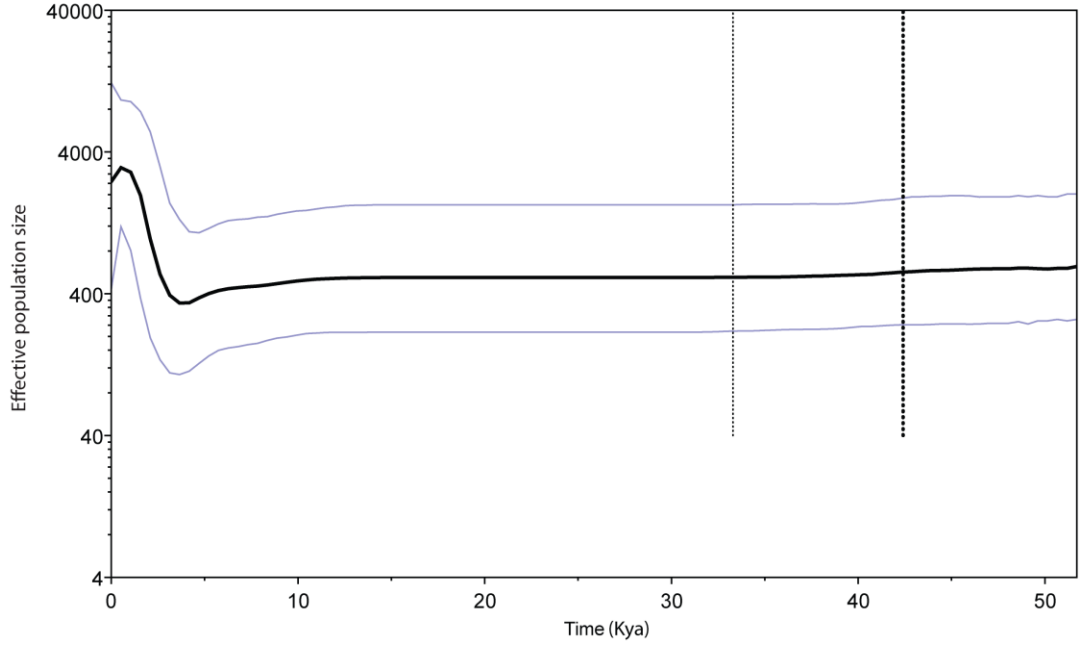
d)

L0f


40000
4000
400
40
4
Effective population size
0
10
20
30
40
50
60
Time (Kya)

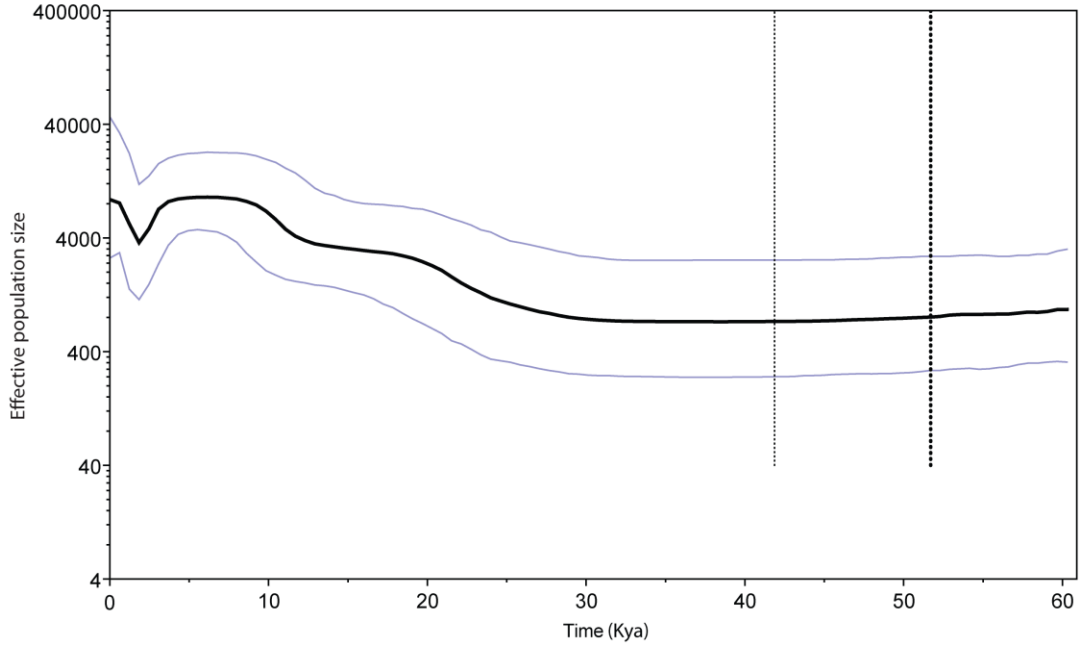
e)

L1b


40000
4000
400
40
4
Effective population size
0
10
20
30
40
50
Time (Kya)

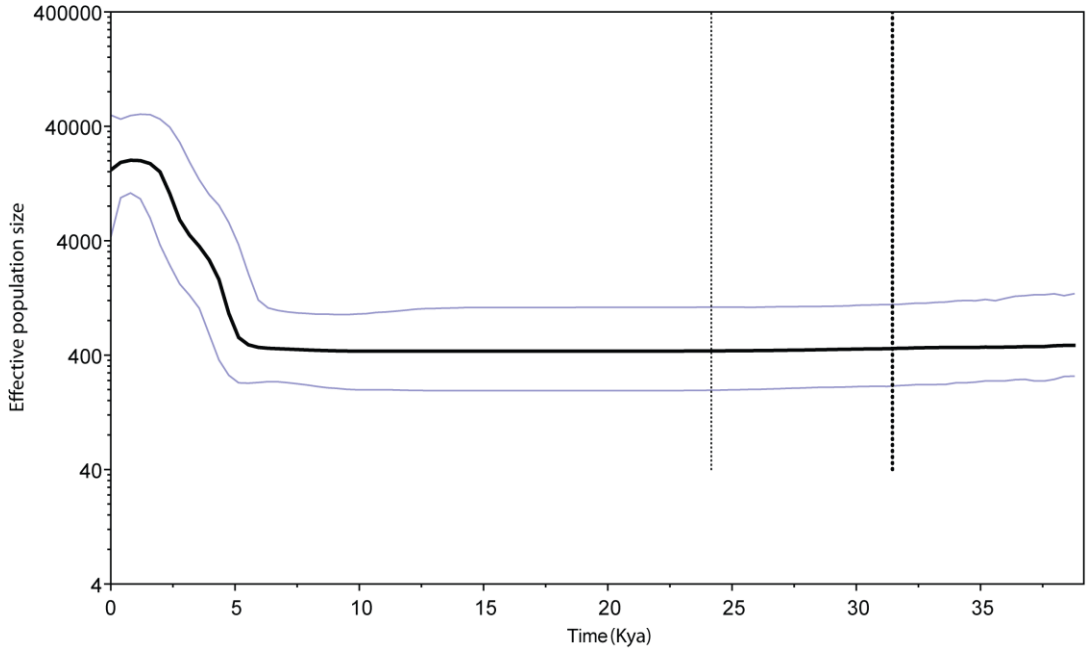
f)

L1c


400000
40000
4000
400
40
4
Effective population size
0
10
20
30
40
50
60
Time (Kya)

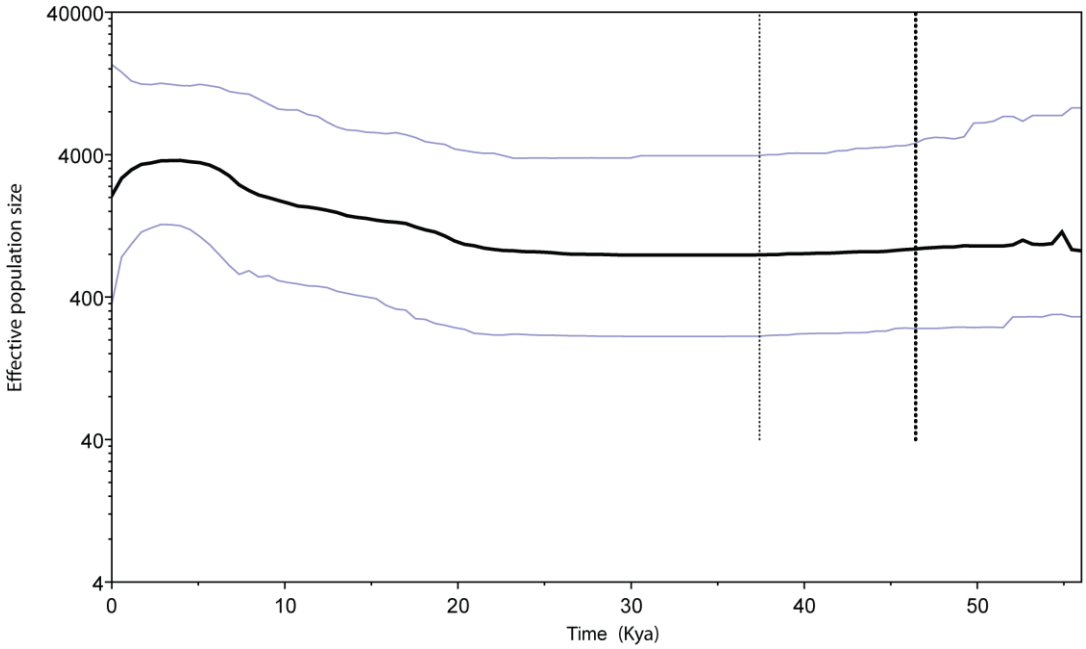
g)

L2a1


400000
40000
4000
400
40
4
Effective population size
0
5
10
15
20
25
30
35
Time (Kya)

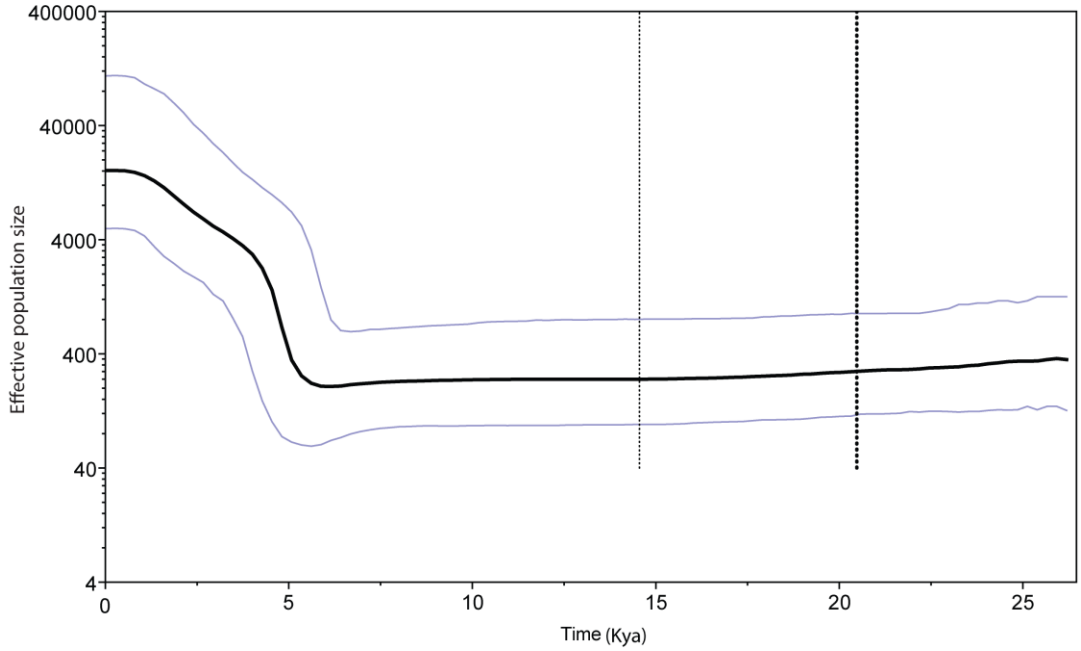
h)

L3a


40000
4000
400
40
4
Effective population size
0
10
20
30
40
50
Time (Kya)

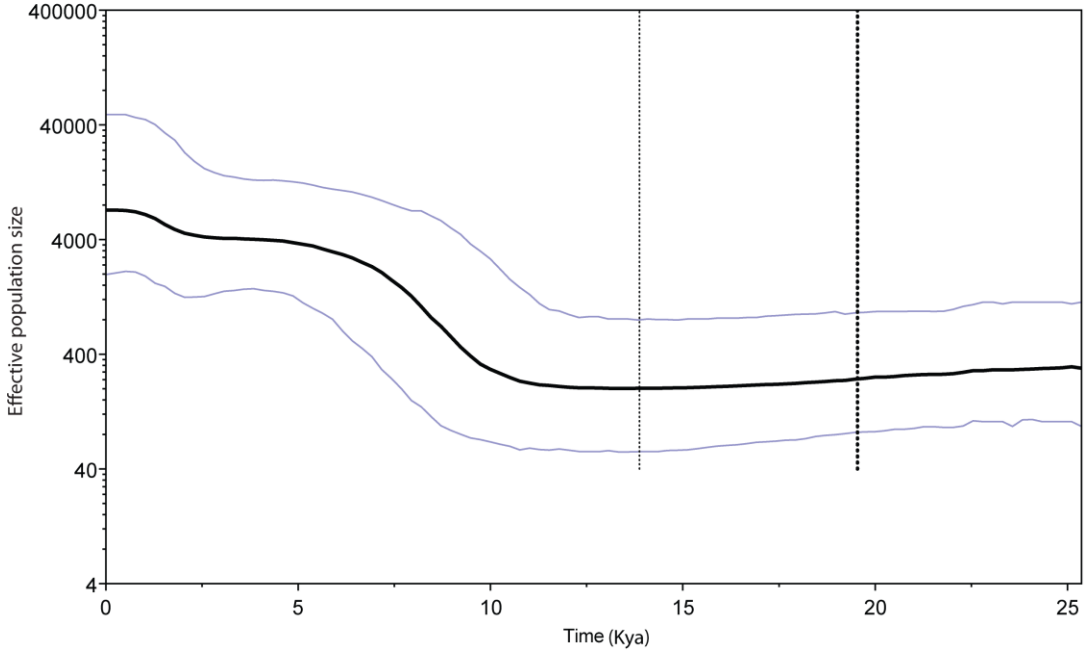
i)

L3b


400000
40000
4000
400
40
4
Effective population size
0
5
10
15
20
25
Time (Kya)

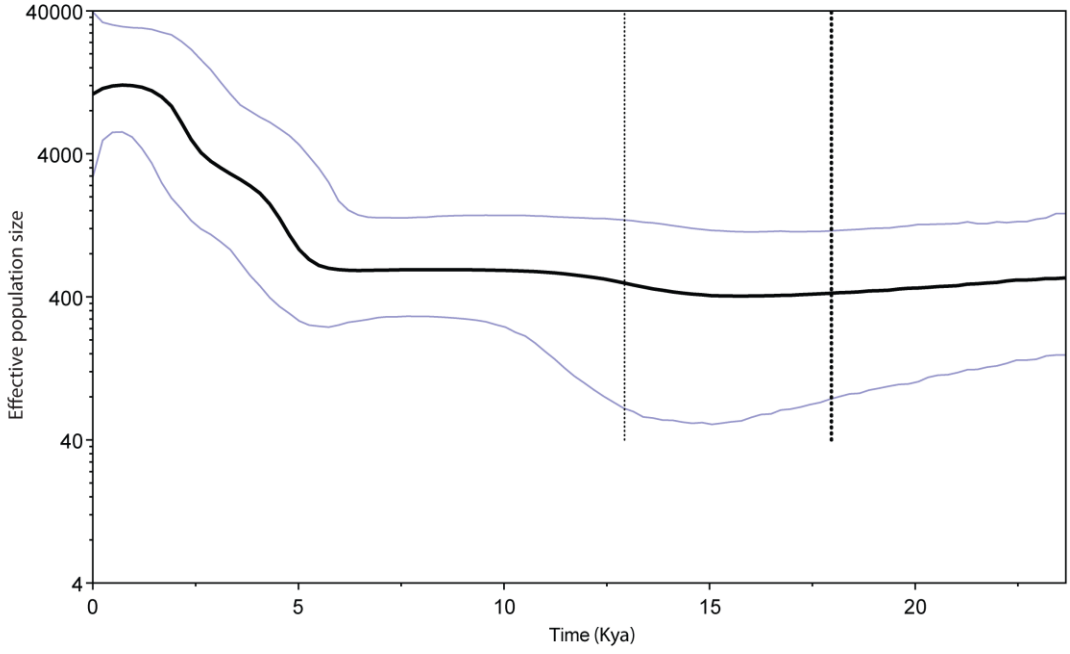
j)

L3d


400000
40000
4000
400
40
4
Effective population size
0
5
10
15
20
25
Time (Kya)

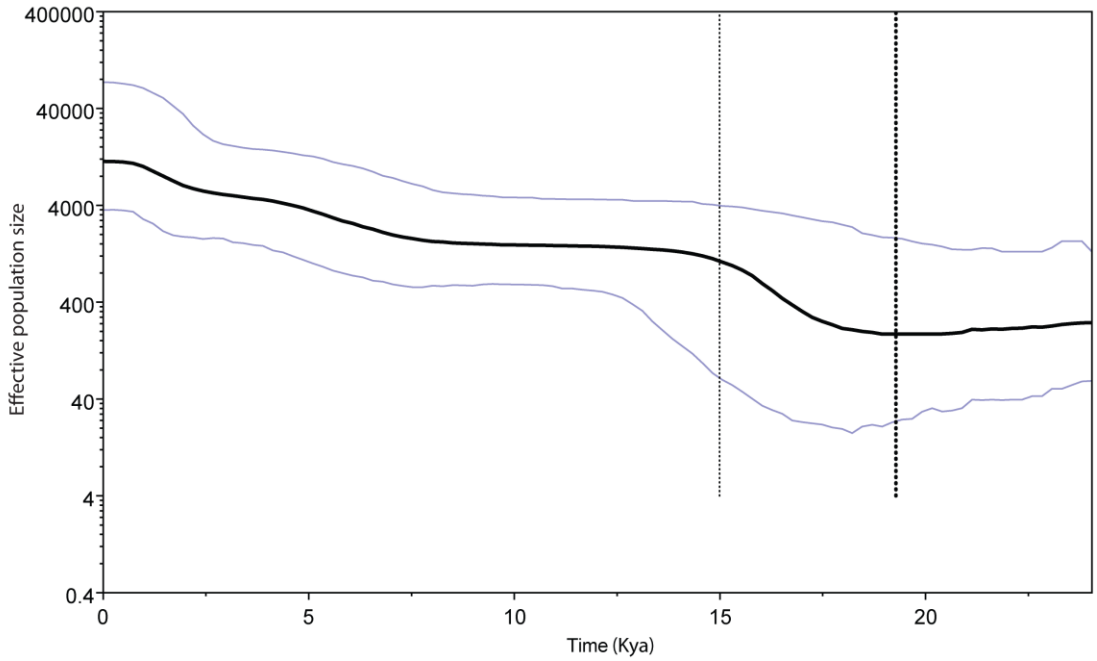
k)

L3e


40000
4000
400
40
4
Effective population size
0
5
10
15
20
Time (Kya)

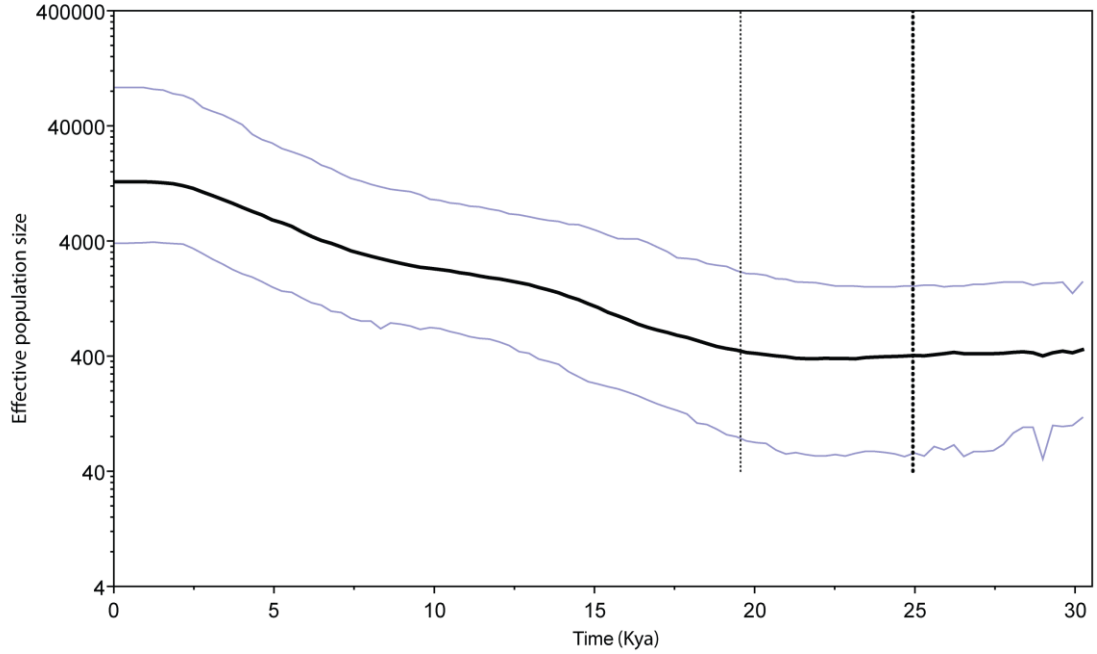
l)

L3f


400000
40000
4000
400
40
4
0.4
Effective population size
0
5
10
15
20
Time (Kya)

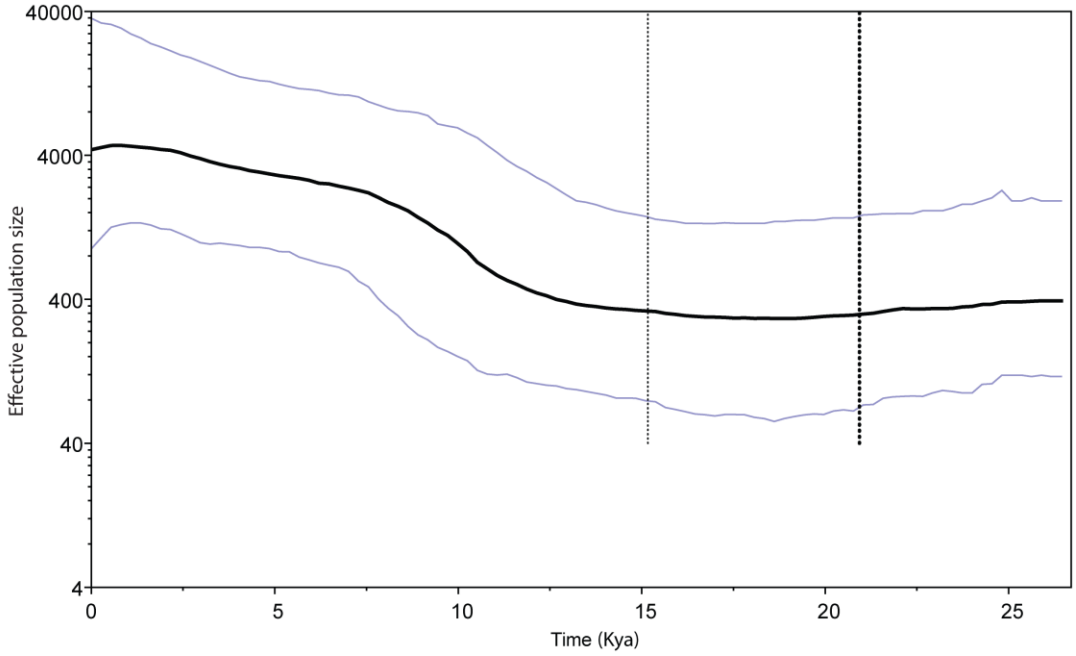
m)

L3h


400000
40000
4000
400
40
4
Effective population size
0
5
10
15
20
25
30
Time (Kya)

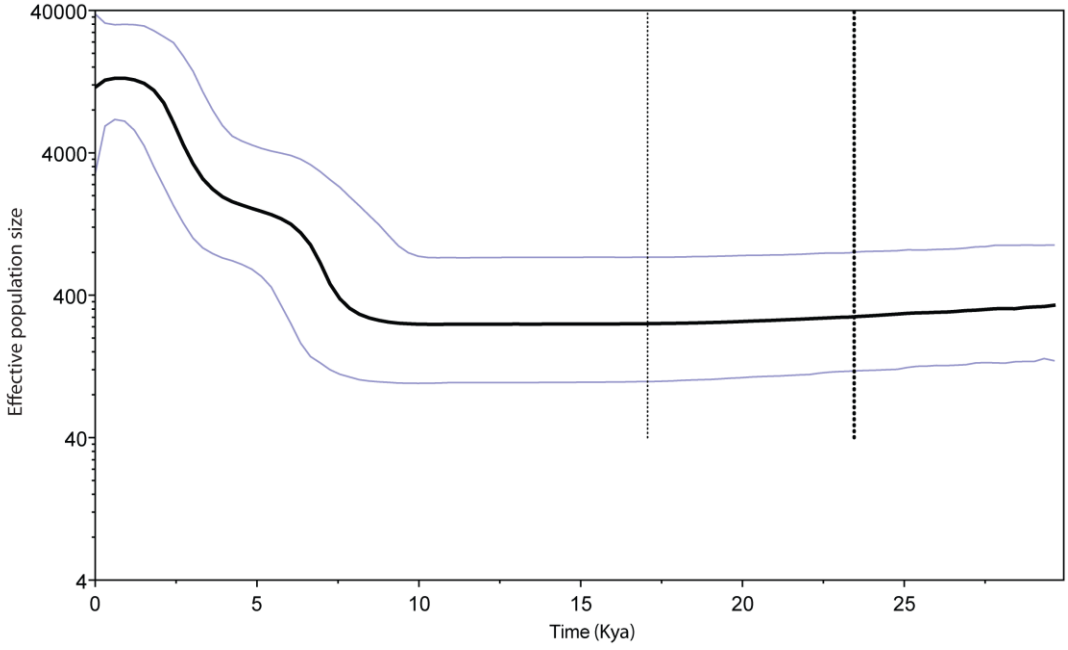
n)

L3x


40000
4000
400
40
4
Effective population size
0
5
10
15
20
25
Time (Kya)

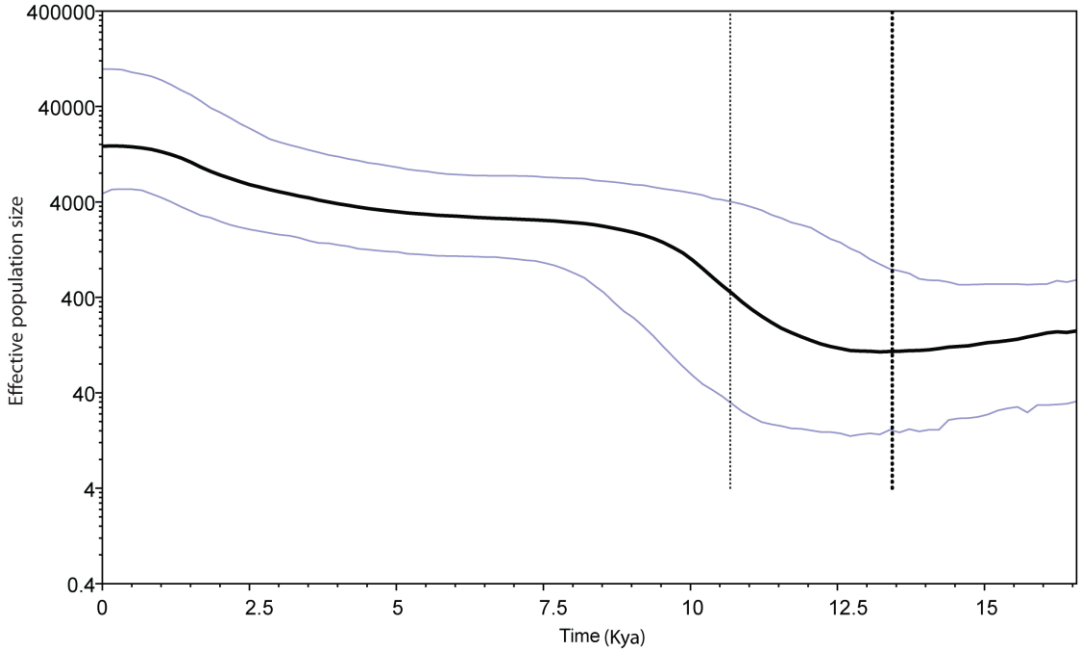
o)

M1a


40000
4000
400
40
4
Effective population size
0
5
10
15
20
25
Time (Kya)

p)

U6


400000
40000
4000
400
40
4
0.4
Effective population size
0
2.5
5
7.5
10
12.5
15
Time (Kya)

Appendix 15: Unique *mt*DNA and Y chromosome lineage distributions across Africa

Recent morphological studies of early modern human fossils concluded that early modern humans were already divided into different populations in Africa before expanding outside Africa [680]. This observation is consistent with molecular studies of a pseudogene [495, 496] and the DXS1238 microsatellite on the X chromosome [497] that suggest that ancestral modern humans might have arisen from a structured population, and isolation and admixture events likely occurred among ancient African subpopulations. Moreover, based on data from DNA polymorphisms in an 8 kb intronic segment flanking exon 44 (*dys44*) of the human dystrophin gene on Xp21, Labuda *et al.,* [50] suggested that the genetic pool of sub-Saharan Africans represents two lineages that had evolved separately for some period of time and eventually hybridized. They argued that their data provides evidence (i) for an expansion of one lineage throughout several different continents and (ii) that greater genetic diversity in sub-Saharan Africans as compared to other continental populations could be due to enrichment of their gene pool through admixture with independently evolving local population(s) [50, 55]. However the African populations included in their study were all of Central/West African origin. Recent studies [498, 499] used a model that introduced a new measure of linkage disequilibrium in sequence data from Yoruba, a West African population, showed some evidence of ancestral admixture in this population. However, a definitive source population could not be clearly determined [498]. The study by Behar *et al.,* [157] of *mt*DNA lineages suggests that the early settlement of humans in Africa was split into two matrilineally structured population clusters with each of the clusters being isolated and separately evolving populations. They [157] argued that one ancestral population cluster

was made up of ancestral populations that led to current Khoisan groups in South Africa while the other clusters represent ancestral populations that gave rise to all other modern human populations globally [157]. Moreover, the pattern where ancestral *mt*DNA lineages (L0 and L1) have a restricted geographic distribution has previously been interpreted [152] as footprints of an early spread across Africa before 100 kya, and the later major African re-expansion 60–80 kya [598].

However, based on results from the present study there might have been as many as three ancestral clusters. Both Y chromosome and *mt*DNA lineages clearly show restricted geographical distribution **(Appendix 6, Table 3.3.3, Table 3.4.2, Table A15.1)**. Derived variants of the putatively most ancestral Y chromosome haplogroup, A, form three separate geographical distributions. The A1 haplotype lineage is observed only in Central/West Africa, the A3b2 is observed commonly in East Africa while A2 and A3b1 lineages are observed only in southern African populations **(Table A15.1, Appendix 6a, Table 3.3.3)**. Moreover, the slightly younger Y chromosome B haplogroup shows a distribution consistent with three different ancestral population clusters corresponding to the same geographical clusters as observed for the A haplogroup **(Table A15.1, Appendix 6a, Table 3.3.3)**. Y chromosome haplotype lineages B* and B1 are observed mostly in Central Africa, while B2a is observed mostly in East African populations. However, the ancestral variant of B2b, B2b* is found in all the three regions (Central/western, eastern and southern Africa) but is at highest frequency in East African hunter-gatherers **(Table A15.1, Appendix 6a, Table 3.3.3)**. The younger Y chromosome lineages show clear bifurcation in their geographic distributions, with E2a and E3b observed mostly in East African populations while E1 and E3a are observed mostly in

Central/West Africa. Mitochondrial haplotypes with deep lineages, specifically L0d and L0k (southern Africa), L1 and L2 (Central/west Africa) and L0b, L0f, L5 and L4 (East Africa) mirror what is observed for Y chromosome haplogroup A **(Table A15.1, Appendix 6b, Table 3.4.2)**. Similar to what is observed for the Y chromosome lineages, the younger *mt*DNA L3 lineages are mostly observed in Central/West Africa (L3b, L3d, L3e) and in East Africa (L3a, L3h, L3i & L3x, L6) **(Table A15.1, Appendix 6b, Table 3.4.2)**.

According to the results from STRUCTURE analysis of autosomal data in the current study and larger analysis of the 121 African populations [4], at K=5, three of the five clusters might represent signatures of African regional populations: one cluster is composed of pygmy and Khoisan populations (Khoisan with some East African ancestry) from the Central Africa and South Africa, respectively, and other two clusters represent East and Central/Western African populations **(Figure 3.3.1a, Figure S15** [4]**)**. The fourth cluster represents non-African populations, which are a subset of an East African cluster [11, 34, 48-55] and mostly observed among the Afroasiatic speakers in the larger African dataset [4]. The fifth cluster represents the Hadza, a Khoisan speaking population from Tanzania. The distinctive genetic profile of the Hadza might be due to long time isolation and genetic drift. This is consistent with linguistic evidence that shows the Hadza language being very different from both South African Khoisan and from the nearby Sandawe language in northern Tanzania [102]. These results are consistent with the possibility that there was an initial split that gave rise to three ancestral population clusters that formed the proto-populations that expanded to extant populations in Central/West Africa, East Africa and South Africa. Later migration events and gene-flow

took place between populations originating from Central/West Africa and East Africa during the late Pleistocene and early Holocene [65, 84, 380]).

East and North-East Africa	Central & West Africa	South Africa
L0f	L0a1a	L0d1
L0f1	L0a1b	L0d2
L0a1c	L0a2a	L0k
L0a1d	L0a2b	L0d3
L0a2c	L1b	L0f1
L0a3	L1c	
L0a4	L2	
L0b	L3b	
L0d3	L3d	
L3a	L3e	
L3h		
L3i		
L3x		
L4		
L5		
M1a		

East and North-East Africa	Central & West Africa	South Africa
A3b2	A1	A3b1
B2b*	B2a2	A2
B2a1	B2b2	B2b1
B2a3	B2b3	B2b4
B2b2	B2b4	E3b6
E2a	E1	
E3*	E2b	
E3b (E3b*, E3b1, E3b3, E3b6)	E3a (E3a*, E3a7, E3a8)	
K2		

Table A15.1: Centers of haplotype maxima for both *mt*DNA (A) and Y chromosome (B) lineages; Center of haplotype frequency maximum; In red: haplotypes shared between regions most probably due to gene-flow. In blue: haplotypes found in very low frequency or might have been lost due to genetic drift

Expansion of modern humans outside of Africa appears to have occurred from a small subset of the ancestral eastern African population cluster as previously shown [11, 34, 48-55]. These inferences correspond to recent studies that showed concordance between human genetic diversity and gut microbe *Helicobacter pylori* (*Hp*) [681, 682]. Linz *et al.,* [682] analyzed 716 isolates of *Hp* from 51 representative global human populations and identified five clusters that they inferred represent *Hp's* ancestral sources. The study showed strong phylogeographic structure in *Hp* concordant to human phylogeography [681, 682]. Three of the five clusters identified are from Africa: hpAfrica1 (West Africa and South Africa), hpAfrica2 (South Africa), hpNEAfrica (East Africa) and the remaining two are from Europe and Asia, respectively. The hpNEAfrica cluster was predominant among isolates from Ethiopia, Somalia, Sudan and Nilo-Saharan speakers in northern Nigeria [682]. The three clusters may correspond to three ancestral structured populations in Africa from which current human populations have descended. Interestingly, assignments of individual isolates to populations using a linkage model shows that the proportions of ancestry, and relatedness of concatenated sequences from the five ancestral sources almost form a continuum between the clusters save for hpAfrica2, the south African variant, which was highly distinct. HpAfrica1 and HpNEAfrica formed a continuum [682]. This observation is consistent with the conclusion inferred in this study that after the initial three ancestral geographical populations, East, Central/West and South Africa, there were later migrations in populations from East and Central/West Africa.

In addition, hpAfrica1, which is shared by West Africans and South Africans might be a reflection of Bantu expansions to southern Africa originating from

Central/western Africa. No populations were sampled from the Great Lake Regions, the areas that have been shown by archeologists and linguistic historians to have served as a corridor of Bantu expansion to the South. The Great Lake Regions is an area encompassing East African countries, Kenya, Tanzania, Uganda, Rwanda, Burundi and northeastern Congo, bordered by a series of lakes in and around the East African Rift Valley: Lakes Victoria, Tanganyika, Malawi, Turkana, Albert, Kivu and Edward. Sampling of Hp isolates from these regions will confirm or reject this assertion.

Appendix 16: Archeological and genetic evidence that support "northern route" of human out-of-Africa Expansion

All archeological evidence of hominids in Eurasia, including the Levant, prior to about 75 kya are thought to be of archaic hominids, Neanderthals or initial unsuccessful out-of-Africa migrations of modern humans [29, 683-686]. In fact, there is no evidence of *Homo sapiens* in the Levant between 75–45 kya, with only Neanderthal remains associated with Tabun B-Type industry found in that region [686]. Archeological evidence supporting the existence of modern human populations in transitional areas from Africa to Europe, namely Egypt and Israel 50- 60 kya is scanty [393, 686]. The fossil records associated to *Homo sapiens* in the Levant are those that occur only after 45 kya [686]. There are sites from 35 - 60 kya in northern Egypt (35-44 kya male burial at Nazlet khater 4; 45- 65 kya late Mousterian in Wadi Kubbaniya - attributed to modern human; 55 kya Taramsa I burial; 60 ky chert quarrying at Nazlet Safaha [393, 687-690]), that offer possible evidence of precursors to modern human expansion from Africa into Europe.

Middle Paleolithic tool industries of northeast Africa, which might have served as transition technology to the Near East, have been broadly classified into two complexes: Nubian that is riverine that probably expanded northward from the Sudan, and the Lower Nile valley complex [586]. The Lower Nile valley complex is described as flexible, which made it possible for its users to adapt and exploit different environments including the desert [586]. The Lower Nile valley complex is considered as a continuation of industry practiced by earlier occupiers of the Nile Valley (from 90 kya), before subsequent transition to Upper Paleolithic industry around 40 kya [586]. Similarity between stone tool technologies dated to between 35-44 kya at Nazlet khater 4 in Egypt

and those used at the site in the Levant (Boker Tachtit) [691, 692] serves as evidence of a transition of modern humans from Africa to Eurasia. In fact, archeological findings from several 30 - 50 kyo sites in the Levant have been termed “transitional industries” between the Middle and Upper Paleolithic [29].

The Upper Paleolithic industries thought to be the work of modern humans seem to have appeared somewhat earlier in western Asia than in Europe [29]. Some historians argue that the 40 kya common tradition that marks the behavioral modernity sometimes attributed to Neanderthal (art, personal decoration, ritualized burials, formal bone tools and gift exchange), might represent the expansion of Upper Paleolithic anatomically modern human populations across Europe [393, 685, 693]. Despite the fact that Europe was settled by modern humans by the end of the Middle Paleolithic (by 30 kya) [404, 694], the early European modern human fossil evidence from Mladec (28 kya) [695] exhibit features that support substantial and relatively recent African ancestry [26]. Moreover, a study of comparison in body proportions of skeleton samples from the European Early Upper Paleolithic (30 -20 kya) shows that they cluster with recent African samples rather than European Late upper Paleolithic (19-10 kya) samples indicating that there was some gene flow and/or migration from Africa associated with the emergence of modern humans in Europe [696]. Therefore, the sequence of historical events and archaeological evidence above, indicates that the expansion from Africa into Southwest Asia might have taken place around 40 - 50 kya. This is further supported by anatomically modern human produced tools shared between North Africa and the Near east [697].

During two periods, the African faunal zone seems to have extended briefly into the Near East and allowed modern humans to expand their range out of Africa into southwestern Asia [698] before contracting back [699]. These two periods, about 100 kya and 50 kya, coincide with the initial unsuccessful and the second the later expansion out of Africa, respectively [29, 148, 684, 686]. Therefore, the contiguous area that constitutes part of north Africa, specifically the Nile valley and near East, might have also acted as a corridor of human range expansion from Africa and population contraction back to Africa from 40 kya up to the late Pleistocene (20 kya) [586].

The scenario described above fits the genetic evidence and time period for modern human dispersal from Africa through the northern route [17], mostly by individuals with R0 *mt*DNA lineages (sub-family of the N-clade). Based on principal components analysis (PCA) of a dataset of 940 individuals from 53 representative global populations typed at ~650,000 SNPs as part of the Human Genome Diversity Project [343], Reich *et al.,* [555] speculated that there was sub-Saharan African gene flow into Europe and the rest of Eurasia. Moreover, based on a novel PCA and clustering method which was used to determine the phylogeny of 1737 complete human *mt*DNA sequences, Alexe *et al.,* [191] argued that M and N *mt*DNA clades arose due to two different migration events that represent the previously described southern and northern routes respectively. They [191] further argue that the N carrying population that followed the northern route split along an East-West geographic division, resulting in a western "European R clade" containing the haplogroups H, V, H/V, J, T and U, and an eastern "Eurasian N subclade" containing haplogroups B, R5, F, A, N9, I, W and X. However, considering the distribution pattern of the 'Eurasian N-clades' in Southeast Asia, the

Pacific and the Americas, some of the N clades might have been present in individuals who followed the southern route. Interestingly, the R clades that are found in South/Southeast Asia and the Americas (A, B, F, N9, R5-R11, P and Y) seem to have split off from other R clades that are mainly found in the Near East and Europe about 50-70 kya [190]. Such a scenario may indicate that the N clade split within Africa before its expansion out of Africa. The M haplogroup, whose M1 haplotype is predominantly East African and whose other haplotypes are found in the Indian subcontinent and southeastern Asia, might also reflect a population split just before/or after the out-of Africa migration, with most of the M haplotype carrying populations expanding through the southern route. The TMRCA age estimates based on ~4600 sequences – N=3191, M=1416 (60 from this study and the rest from previously published data) of the N and M haplogroup lineages are 41 – 67 ky (Kivisild *et al.,* [220] 62.11±6.09 ky and Mishmar *et al.,* [178] 47.92±6.98 ky) and 41 – 62 (Kivisild *et al.,* [220] 55.76±4.36 ky and Mishmar *et al.,* [178] 45.11±4.53 ky), respectively **(Appendix 7b)**. These age estimates for the two haplogroups concurs with estimates done using a corrected time-dependent mutation rate based on the entire *mt*DNA genome using a maximum likelihood method which estimated the TMRCA of the N haplogroup to be 76.92±17.53 ky and the TMRCA of the M haplogroup to be 73.3±9.64 ky [190] **(Appendix 7b)**. It is still not yet clear whether M and N arose in Africa just before the exodus, or just after it (as indicated by the close relationship and similar ages for M and N (as estimated above), but it is highly unlikely that it happened further east in India as speculated elsewhere [598] based on high diversity of M [412, 600, 700, 701]. The ages of the haplogroups coupled with the distribution pattern of N and M haplotypes described above are consistent with the

hypothesis that they diverged prior to migration of modern humans out of Africa or just after it. This time period coincides perfectly with the return to warm, moist conditions in global climate after volcanic ash from the Mt. Toba eruption (which took place 73 kya in modern-day Sumatra) dissipated. The effects of the eruption on the tropics and sub tropics were reduced temperature, precipitation and increased aridity, and may have lasted until 60 kya [393]. It is hypothesized that these events led to a contraction of the human population, reducing genetic diversity and limiting the distribution of human populations to areas with climatically favorable conditions and ecologically stable environments [48, 110, 393]. The climatic conditions improved around 57 kya with increased insolation (solar radiation received) and precipitation in northern Africa [393]. During dry periods, environmental barriers associated with the severity of the Sahara desert could have made the northern route difficult, so it is likely that this route was more suitable during wetter climatic periods [702]; thus the expansion may have been more likely during the wet periods of 43-57 kya [393, 592]. Recent study's [703] findings of a crude age estimate (13.6 – 108.4 kya) and distribution pattern of 17q21 inversion (microtubular associated protein tau (MAPT) inversion), mainly across Europe, Central and southwestern Asia and Africa [703], also seem to conform to northern route out-of-Africa human and Neolithic expansion.

The expansion of modern human through the northern route is also supported by results from the simulation based global *Helicobacter pylori* (*Hp*) diversity studies [681, 682] that have shown that *Hp* accompanied anatomically modern humans during the initial expansions out of East Africa, estimated at 50-60 kya [39, 682]. Interestingly, the proportions of ancestry of Helicobacter pylori isolates that were found in NE Africa

(*hp*NEAfrica) formed a continuum with that from Europe (*hp*Europe) and not with that from Asia (*hp*Asia) [682] (meaning that the European strains shared ancestry with Northeastern African strains). The *hp*NEAfrica strain is found in populations as far away as Southeast Asia in China, Bangladesh and Thailand, albeit at low frequency [682]. Most of the strains isolated from Europeans, Middle Easterners and Central Asians belong to *hp*Europe [681, 682], and most isolates from East Asia belong to *hp*EastAsia [681, 682]. Moreover, *hp*Europe is found in Asia, while *hp*Asia2 is not found in European populations [682]. The *Hp*Europe strain has high diversity because it is a recombinant between AE1 and AE2 bacteria [681], with the proportion of AE1 nucleotides highest in northern Europe (Finland, Estonia), and Ladakh [681] and lowest levels in Sudan and Israel [681]. Therefore, the spatial distribution of ancestral nucleotides indicated that ancestral Europe2 (AE2) originated in East Africa, and that AE1 originated in Central Asia [682]. This finding is consistent with the hypothesis that current Europeans originated from Africa via the northern route and with some geneflow from East Asia (of the original population that migrated out-of-Africa through the southern route later expanding west into Europe).

Appendix 17: The historical movement of Livestock based on Archaeological and Paleoclimatic evidence

In Africa, domestic animals appear to have spread from north to south in several brief episodes separated by long periods of no expansion. This pattern, may be due to climatic changes [471, 472], though the spread from East Africa to southern Africa may not have been associated with any obvious environmental factors [133]. The chronological order of archeological evidence of pastoralism that starts in North Africa and ends in South Africa was therefore not direct or continuous **(Figure A17.1)**. There might have been a barrier that stunted rapid movement or caused detour in the path of human/livestock expansions. In the last 10 ky there is evidence that water levels in Lake Turkana have risen due to wetter conditions, especially in the Ethiopian highlands during three specific periods: 9.5-7.5 kya, 6.5 - 4 kya and 3.3 kya [704]. These changes in water levels and those also observed for Lakes Victoria and Albert, resulted in water runoff into southern Sudan and caused the "sudd" swamps that persisted up to 2 kya [704]. The time period (6.5 - 4 kya) when this swamp existed overlaps with the time when livestock moved from North Africa into Kenya and might have acted as a barrier to movement of pastoralism through southern Sudan [492]. In the last 10 kya there is evidence that water levels in Lake Turkana have risen due to wetter conditions, especially in the Ethiopian highlands during the three periods mentioned above [477]. This made expanding pastoralist populations use the narrow corridor that passes through lowland southwest Ethiopia and subsequent paths into Kenya followed by Tanzania and into Zambia [121, 492]. Based on archeological evidence of humped cattle from Upper Egypt in the form of

figurines and paintings in tombs and monuments dated to 3.5 – 4 kya [466], and a lack of archaeological evidence of humped cattle in the horn of Africa before 1.8-1.7 kya [494, 705], *Bos indicus* appear to have reached Africa via the Isthmus of Suez by 4 kya [706].

Bousman [707] reviewed early archaeological evidence of livestock in southern Africa and argues that south of Zambia there were two major routes of introduction of stock into South Africa; (i) first one by Khoisan herders that extended west through southern Angola then to southwest Cape via Namibia (ii) the second one involving both Khoisan herders and iron age pastoralists (read Bantu), through Zimbabwe to eastern part of South Africa and Mozambique. There seems to have been climatic changes to moist conditions that opened up both paths, for movement of herds south of Zambia ([707] and reference therein).

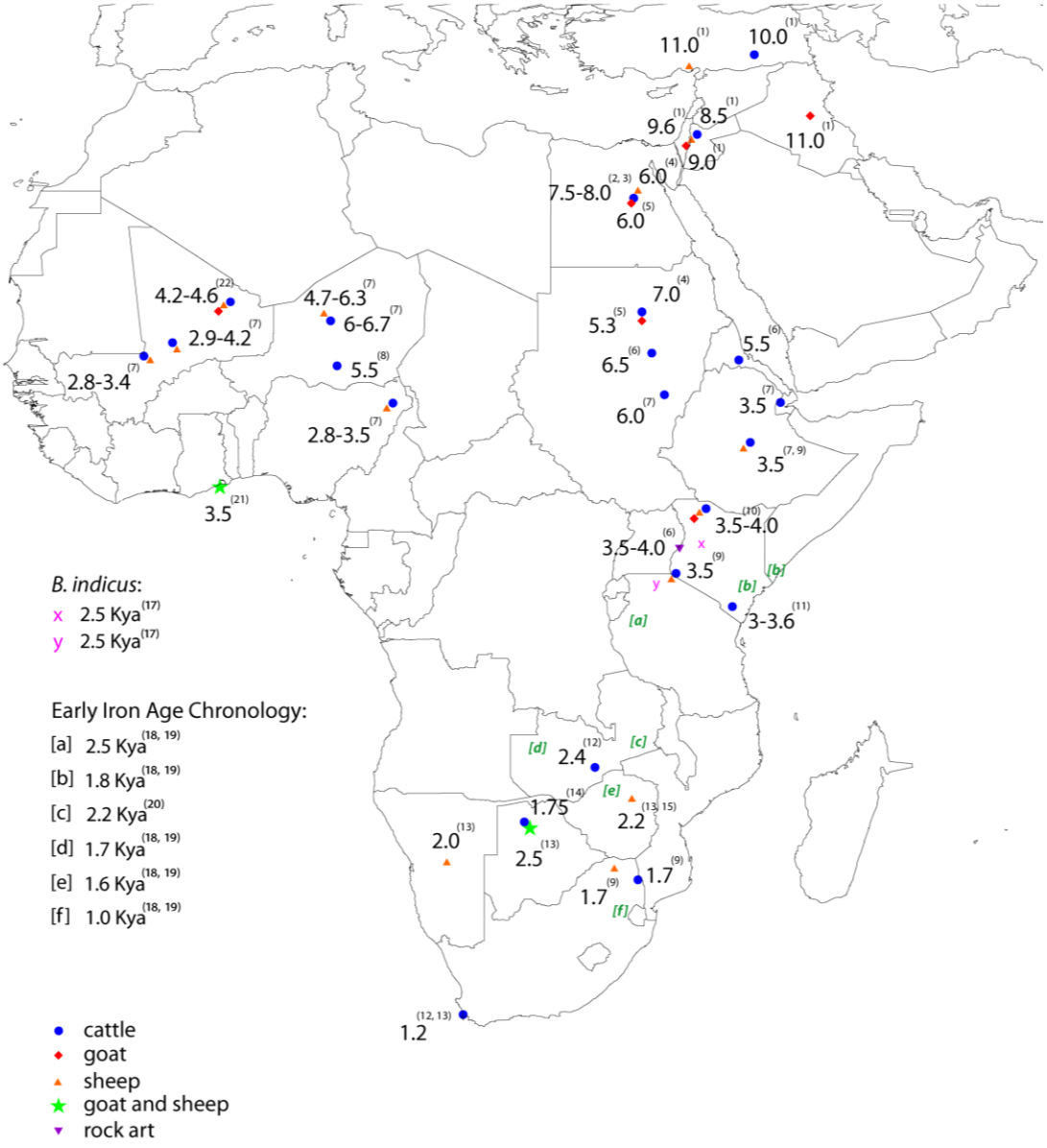
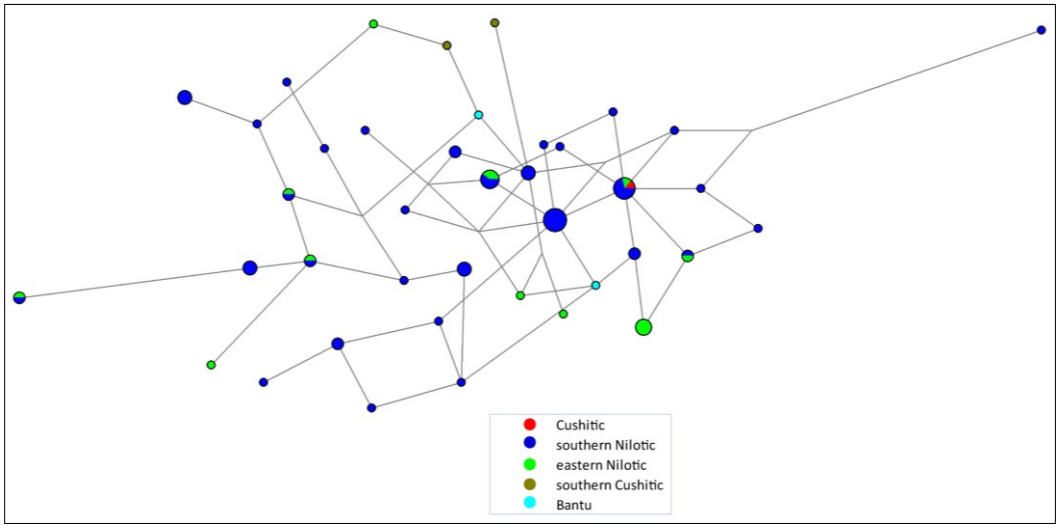


Figure A17.1: Archeological evidence of pastoralism in Africa. Dates were collated from the following literatures: [1][708]; [2][709]; [3][710]; [4][359]; [5][711]; [6][479]; [7][71]; [8][712]; [9][713]; [10][492]; [11][714]; [12][121]; [13][715]; [14][716]; [15][717]; [16][718]; [17][494]; [18][487]; [19][719]; [20][720]; [21][721]; and [22][722].

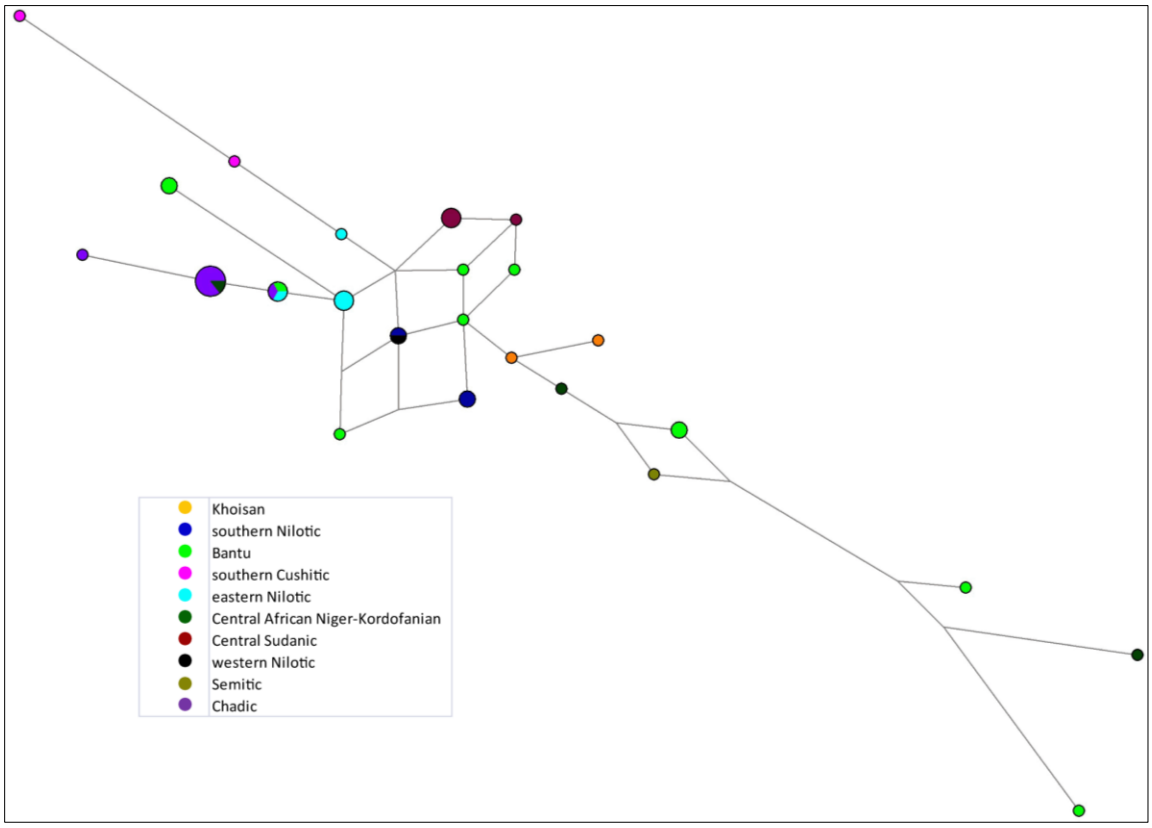
Appendix 18: Y chromosome networks for some selected lineages not shown in the results and Appendix 9

Figure A18.1: Median-joining network of 8-loci (DYS389I, DYS458, DYS437, DYS448, DYS391, DYS392, DYS393 and DYS635) Y-STR haplotypes for a) E2a b) E2b c) E3a7 d) E3a8 e) E3b1 positive individuals in East African populations. Network was generated using Network 4.1.1.1 [258]. Networks were processed first by the reduced-median method, and then by the median-joining method [182, 259] without weighting any of the STR loci. Areas of circles are proportional to absolute frequencies and colored according to which language family/branch an individual is from.

a) E2a

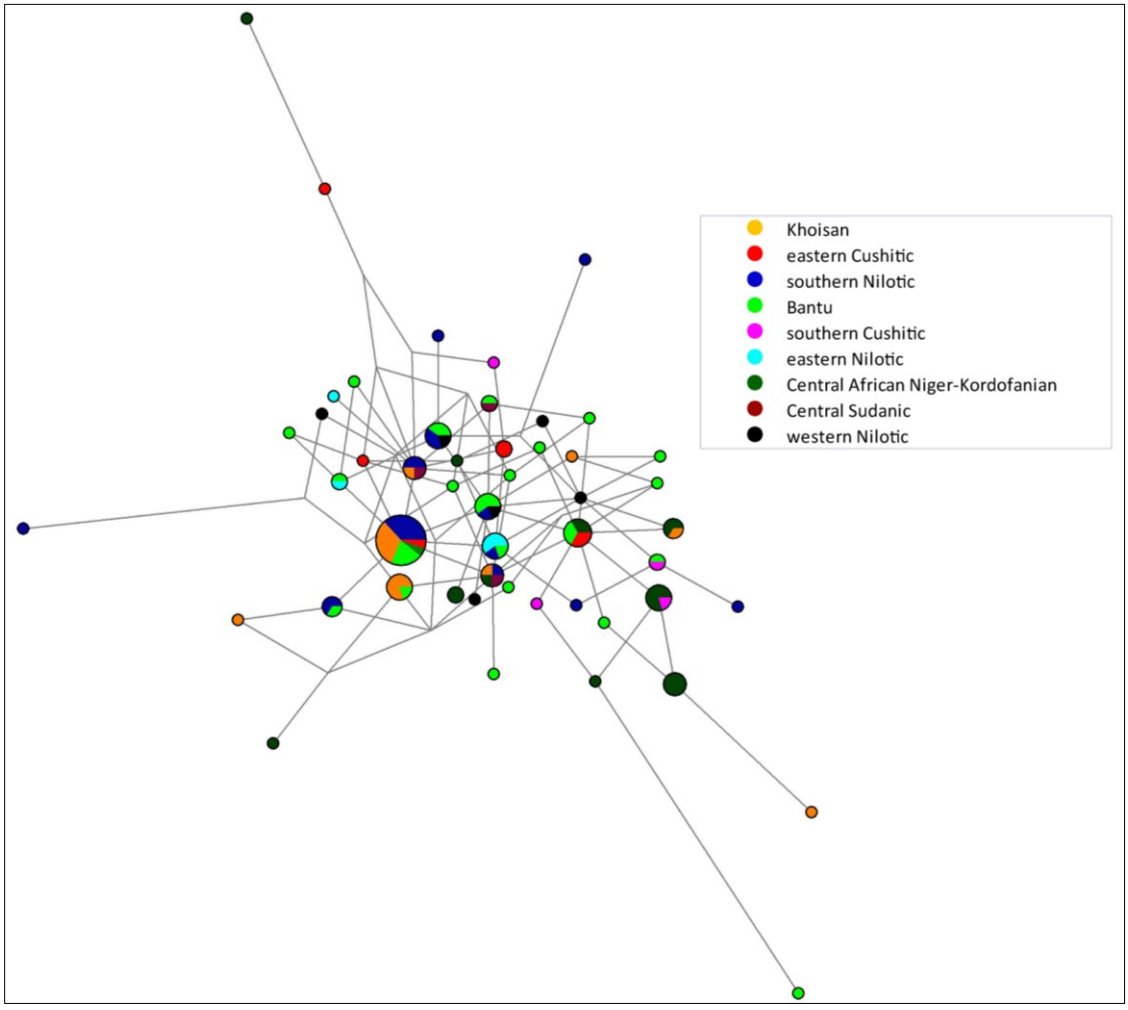


b) <u>E2b (plus DYS019)</u>

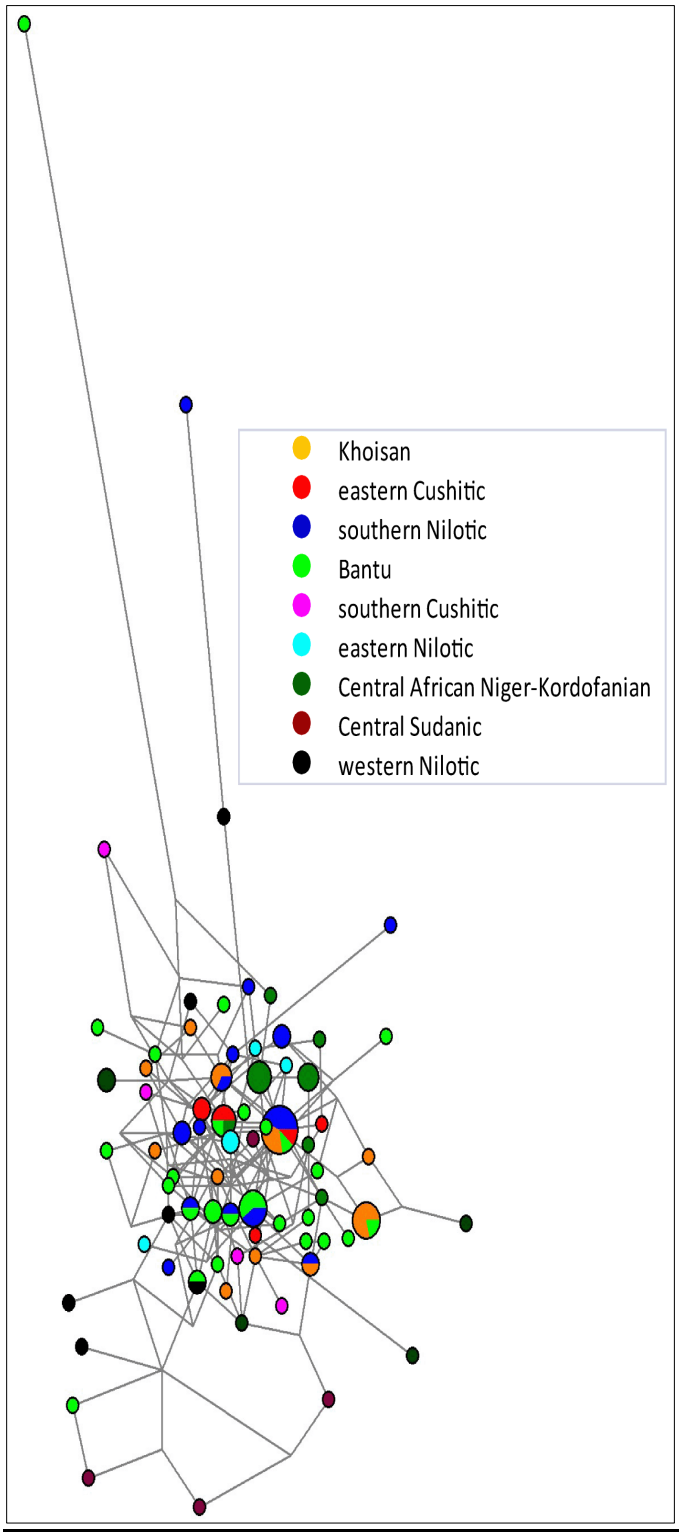


c) <u>E3a7</u>

(i) 8 primers

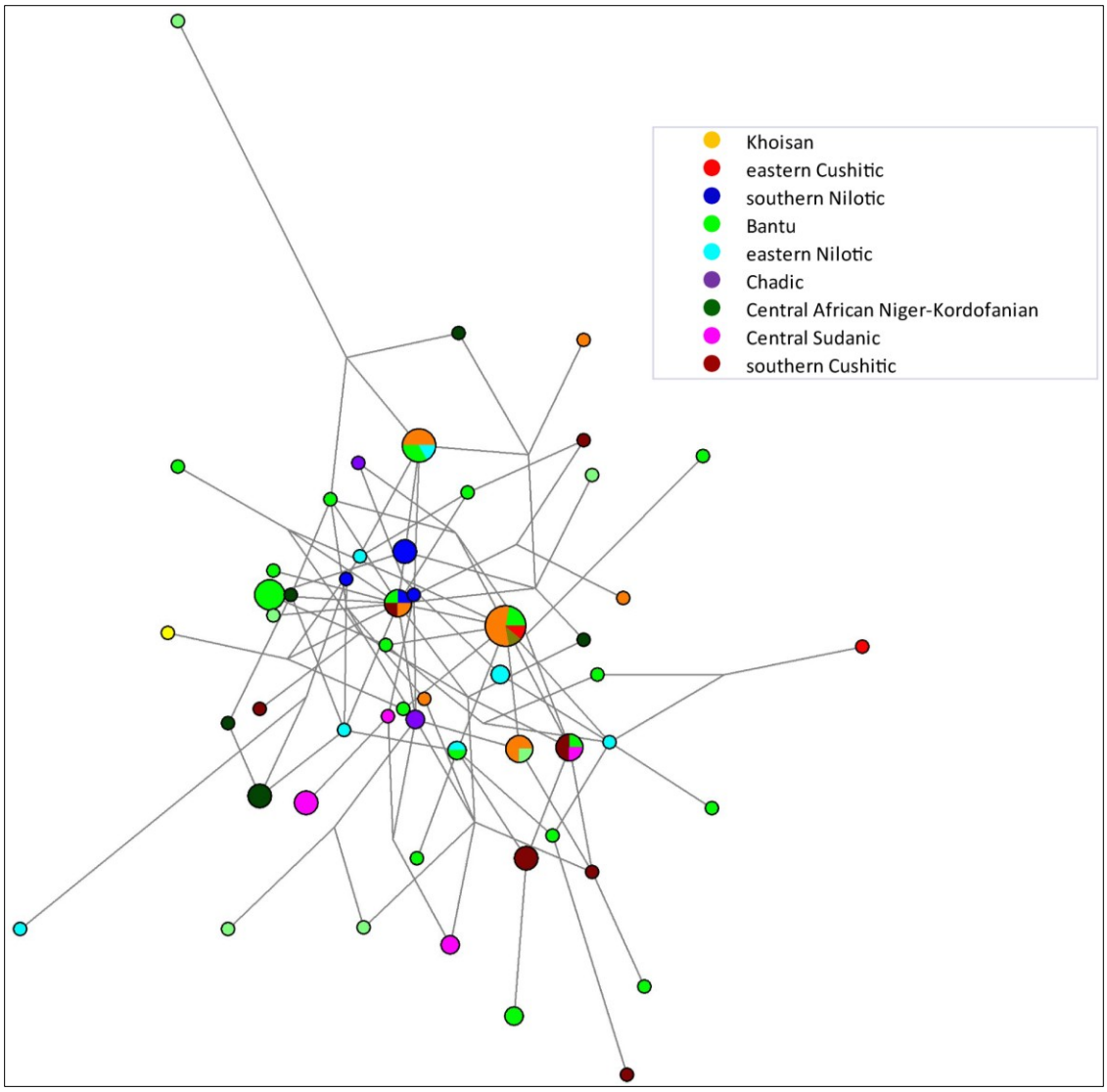


(ii) 9 primers (plus DYS385)

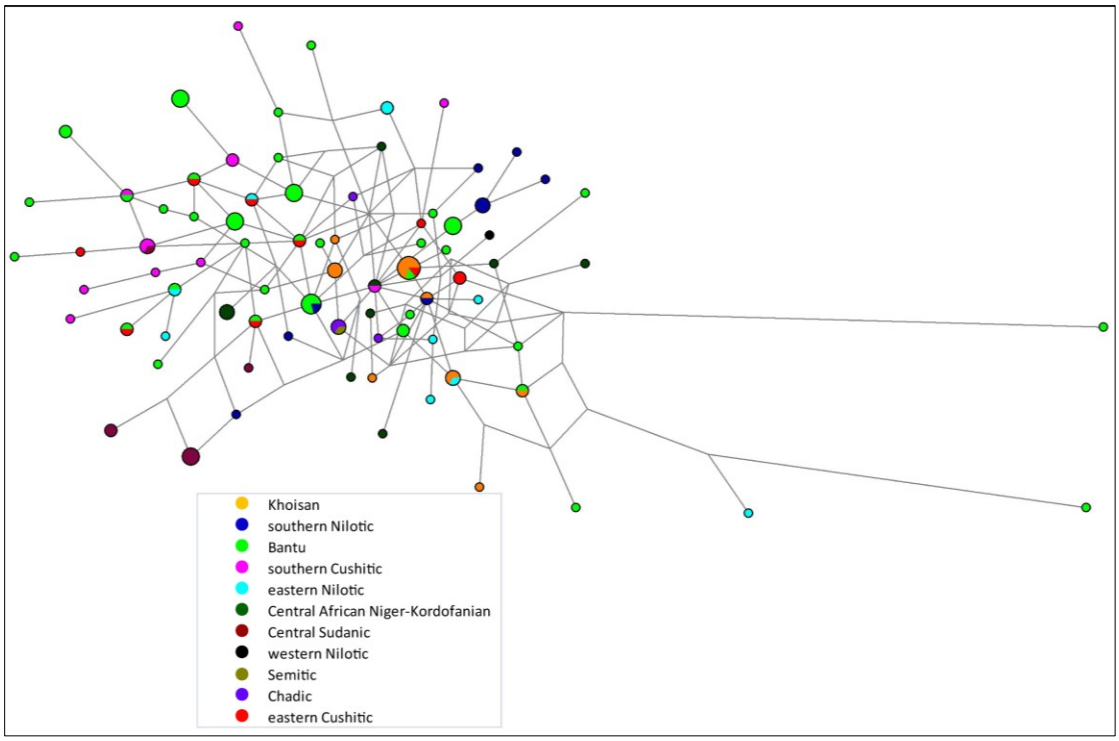


d) <u>E3a8</u>

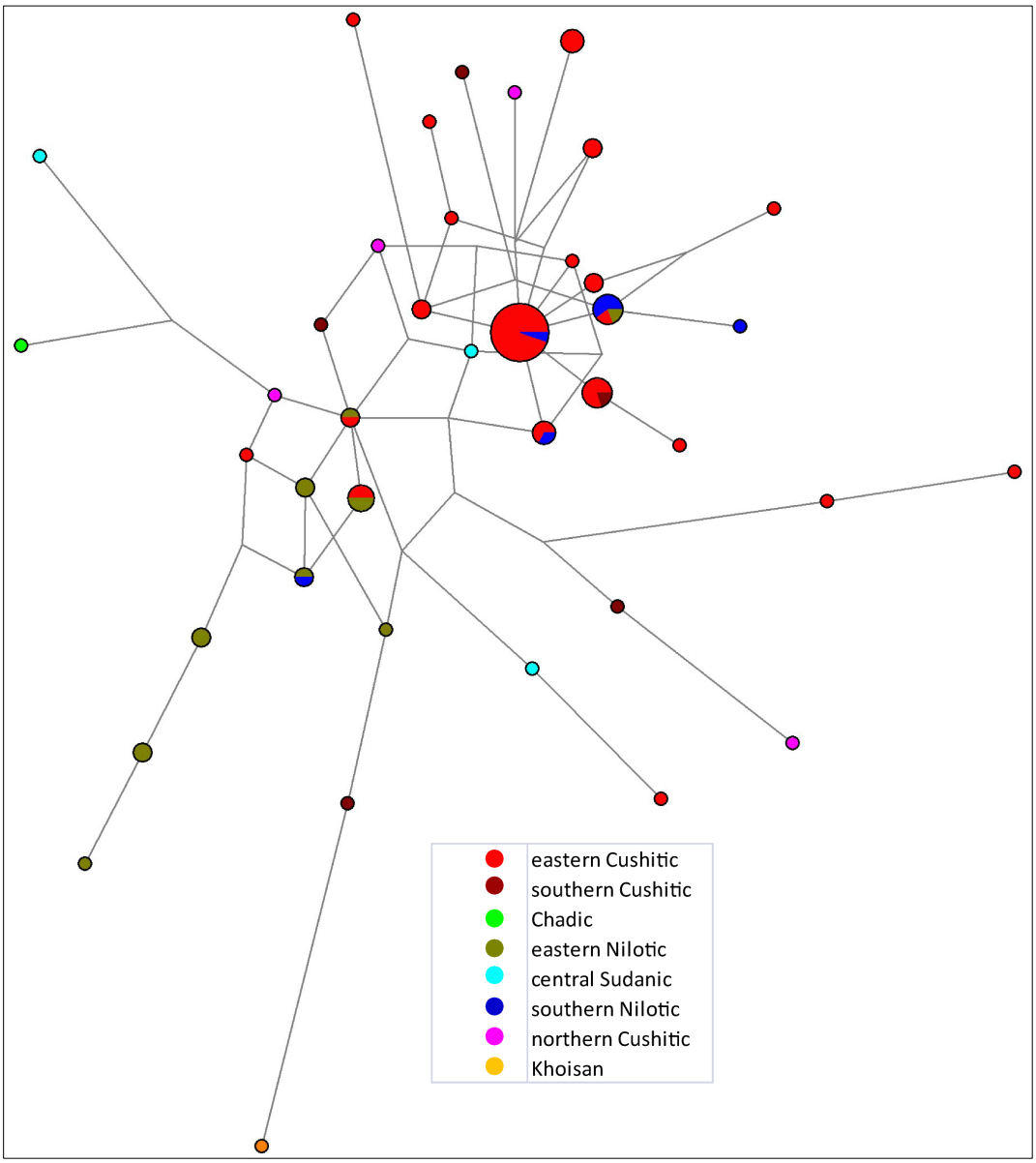
(i) 8 primers



(ii) 9 primers (plus DYS385)



e) **E3b1**


eastern Cushitic
southern Cushitic
Chadic
eastern Nilotic
central Sudanic
southern Nilotic
northern Cushitic
Khoisan

References:

1. Tishkoff, S.A. and S.M. Williams, *Genetic analysis of African populations: human evolution and complex disease.* Nat Rev Genet, 2002. **3**(8): p. 611-21.

2. Phillipson, D.W., *African archaeology*. 3rd ed. 2005, Cambridge: Cambridge University Press.

3. Rosenberg, N.A., et al., *Genetic structure of human populations.* Science, 2002. **298**(5602): p. 2381-5.

4. Tishkoff, S.A., et al., *The Genetic Structure and History of Africans and African Americans.* Science, 2009. **324**(5930): p. 1035-1044.

5. Macaulay, V., et al., *Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes.* Science, 2005. **308**(5724): p. 1034-6.

6. Thangaraj, K., et al., *Reconstructing the origin of Andaman Islanders.* Science, 2005. **308**(5724): p. 996.

7. Kumar, S., et al., *Reconstructing Indian-Australian phylogenetic link.* BMC Evol Biol, 2009. **9**: p. 173.

8. Prugnolle, F., A. Manica, and F. Balloux, *Geography predicts neutral genetic diversity of human populations.* Curr Biol, 2005. **15**(5): p. R159-60.

9. Ramachandran, S., et al., *Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa.* Proc Natl Acad Sci U S A, 2005. **102**(44): p. 15942-7.

10. Betti, L., et al., *Distance from Africa, not climate, explains within-population phenotypic diversity in humans.* Proc Biol Sci, 2008.

11. Cann, R.L., M. Stoneking, and A.C. Wilson, *Mitochondrial DNA and human evolution.* Nature, 1987. **325**(6099): p. 31-6.

12. Tishkoff, S.A., et al., *Global patterns of linkage disequilibrium at the CD4 locus and modern human origins.* Science, 1996. **271**(5254): p. 1380-7.

13. Hammer, M.F., et al., *Out of Africa and back again: nested cladistic analysis of human Y chromosome variation.* Mol Biol Evol, 1998. **15**(4): p. 427-41.

14. Underhill, P.A., et al., *Y chromosome sequence variation and the history of human populations.* Nat Genet, 2000. **26**(3): p. 358-61.

15. Quintana-Murci, L., et al., *Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa.* Nat Genet, 1999. **23**(4): p. 437-41.

16. Kivisild, T., et al., *Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears.* Am J Hum Genet, 2004. **75**(5): p. 752-70.

17. Lahr, M.M. and R.A. Foley, *Towards a theory of modern human origins: geography, demography, and diversity in recent human evolution.* Am J Phys Anthropol, 1998. **Suppl 27**: p. 137-76.

18. Relethford, J.H., *Ancient DNA and the origin of modern humans.* Proc Natl Acad Sci U S A, 2001. **98**(2): p. 390-1.

19. Stringer, C., *Modern human origins: progress and prospects.* Philos Trans R Soc Lond B Biol Sci, 2002. **357**(1420): p. 563-79.

20. Cavalli-Sforza, L.L., P. Menozzi, A. Piazza, *The History and Geography of Human Genes*. 1996, Princeton: Princeton University Press.

21. Ray, N., et al., *Recovering the geographic origin of early modern humans by realistic and spatially explicit simulations.* Genome Res, 2005. **15**(8): p. 1161-7.

22. McBrearty, S. and A.S. Brooks, *The revolution that wasn't: a new interpretation of the origin of modern human behavior.* J Hum Evol, 2000. **39**(5): p. 453-563.

23. Clark, J.D., et al., *Stratigraphic, chronological and behavioural contexts of Pleistocene Homo sapiens from Middle Awash, Ethiopia.* Nature, 2003. **423**(6941): p. 747-52.

24. Haile-Selassie, Y., B. Asfaw, and T.D. White, *Hominid cranial remains from upper Pleistocene deposits at Aduma, Middle Awash, Ethiopia.* Am J Phys Anthropol, 2004. **123**(1): p. 1-10.

25. Stringer, C.B. and P. Andrews, *Genetic and fossil evidence for the origin of modern humans.* Science, 1988. **239**(4845): p. 1263-8.

26. Trinkaus, E., *Early Modern Humans.* Annual Review of Anthropology, 2005. **34**: p. 207-30.

27. Bar-Yosef, O., *The role of western Asia in modern human origins.* Philos Trans R Soc Lond B Biol Sci, 1992. **337**(1280): p. 193-200.

28. Bar-Yosef, O., *The contribution of southwest Asia to the study of modern human origins*, in *Origins of Anotomically Modern Humans*, M.N.a.D. Nitecki, Editor. 1994, Plenum Press: New York. p. 22-36.

29. Bar-Yosef, O. *Modern humans, Neanderthals and the Middle/Upper Paleolithic transition in Western Asia*. in *The Lower and Middle Palaeolithic (Colloquia of the XIII International Congress of Prehistoric and Protohistoric Sciences*. 1996. Forlì, Italy.

30. O'Connell, J.F.a.F.J.A., *Dating the colonization of Sahul (Pleistocene Australia-New Guinea): A review of recent research.* Journal of Archaeological Science, 2004. **31**(835-853).

31. Petraglia, M., et al., *Middle Paleolithic assemblages from the Indian subcontinent before and after the Toba super-eruption.* Science, 2007. **317**(5834): p. 114-6.

32. Lahr, M.M. and R.A. Foley, *Multiple dispersals and modern human origins.* Evolutionary Anthropology, 1994. **3**(48-60).

33. Cavalli-Sforza, L.L., et al., *Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data*. Proc Natl Acad Sci U S A, 1988. **85**(16): p. 6002-6.

34. Chen, Y.S., et al., *Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups.* Am J Hum Genet, 1995. **57**(1): p. 133-49.

35. Barbujani, G. and R.R. Sokal, *Zones of sharp genetic change in Europe are also linguistic boundaries.* Proc Natl Acad Sci U S A, 1990. **87**(5): p. 1816-9.

36. Rosser, Z.H., et al., *Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language.* Am J Hum Genet, 2000. **67**(6): p. 1526-43.

37. Hunley, K., et al., *Genetic and linguistic coevolution in Northern Island Melanesia.* PLoS Genet, 2008. **4**(10): p. e1000239.

38. Wood, E.T., et al., *Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes.* Eur J Hum Genet, 2005. **13**(7): p. 867-76.

39. Liu, H., et al., *A geographically explicit genetic model of worldwide human-settlement history.* Am J Hum Genet, 2006. **79**(2): p. 230-7.

40. Manica, A., et al., *The effect of ancient population bottlenecks on human phenotypic variation.* Nature, 2007. **448**(7151): p. 346-8.

41. Degiorgio, M., M. Jakobsson, and N.A. Rosenberg, *Out of Africa: Modern Human Origins Special Feature: Explaining worldwide patterns of human genetic variation using a coalescent-based serial founder model of migration outward from Africa.* Proc Natl Acad Sci U S A, 2009.

42. Hofer, T., et al., *Large allele frequency differences between human continental groups are more likely to have occurred by drift during range expansions than by selection.* Ann Hum Genet, 2009. **73**(1): p. 95-108.

43. Relethford, J.H., *Mitochondrial DNA and ancient population growth.* Am J Phys Anthropol, 1998. **105**(1): p. 1-7.

44. Relethford, J.H. and H.C. Harpending, *Craniometric variation, genetic theory, and modern human origins.* Am J Phys Anthropol, 1994. **95**(3): p. 249-70.

45. Vigilant, L., et al., *African populations and the evolution of human mitochondrial DNA.* Science, 1991. **253**(5027): p. 1503-7.

46. Excoffier, L. and S. Schneider, *Why hunter-gatherer populations do not show signs of pleistocene demographic expansions.* Proc Natl Acad Sci U S A, 1999. **96**(19): p. 10597-602.

47. Zhivotovsky, L.A., et al., *Human population expansion and microsatellite variation.* Mol Biol Evol, 2000. **17**(5): p. 757-67.

48. Ambrose, S.H., *Late Pleistocene human population bottlenecks, volcanic winter, and differentiation of modern humans.* J Hum Evol, 1998. **34**(6): p. 623-51.

49. Armour, J.A., et al., *Minisatellite diversity supports a recent African origin for modern humans.* Nat Genet, 1996. **13**(2): p. 154-60.

50. Labuda, D., E. Zietkiewicz, and V. Yotova, *Archaic lineages in the history of modern humans.* Genetics, 2000. **156**(2): p. 799-808.

51. Marth, G.T., et al., *The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations.* Genetics, 2004. **166**(1): p. 351-72.

52. Rogers, A.R. and L.B. Jorde, *Genetic evidence on modern human origins.* Hum Biol, 1995. **67**(1): p. 1-36.

53. Tishkoff, S.A., et al., *A global haplotype analysis of the myotonic dystrophy locus: implications for the evolution of modern humans and for the origin of myotonic dystrophy mutations.* Am J Hum Genet, 1998. **62**(6): p. 1389-402.

54. Wainscoat, J.S., et al., *Evolutionary relationships of human populations from an analysis of nuclear DNA polymorphisms.* Nature, 1986. **319**(6053): p. 491-3.

55. Zietkiewicz, E., et al., *Genetic structure of the ancestral population of modern humans.* J Mol Evol, 1998. **47**(2): p. 146-55.

56. Mehlman, M.J., *Mumba-Hohle Revisited: The Relevance of a Forgotten Excavation to Some Current Issues in East African Prehistory.* World Archaeology, 1979. **11**(1): p. 80-94.

57. Ambrose, S.H., *Archaeology and Linguistic Reconstructions of History in East Africa*, in *The archaeological and linguistic reconstruction of Africa history*, C.a.P. Ehret, M., Editor. 1982, University of Carlifonia Press, Ltd: London, UK. p. 104-157.

58. Ehret, C., *Historical/Linguistic evidence for early food production*, in *From hunters to farmers: the causes and consequences of food production in Africa*, J.C.a.S. Brandt, Editor. 1984, University of Carlifornia Press: Berkeley. p. 26-35.

59. Ehret, C., *Cattle-Keeping and Milking in Eastern and Southern African History: The Linguistic Evidence.* The Journal of African History, 1967. **8**(1): p. 1-17.

60. Roland, O., et al., *Comments on Christopher Ehret, "Bantu History: Re-Envisioning the Evidence of Language".* The International Journal of African Historical Studies, 2001. **34**(1): p. 43-81.

61. Vansina, J., *Western Bantu Expansion.* The Journal of African History, 1984. **25**(2): p. 129-145.

62. Vansina, J., *New Linguistic Evidence and 'the Bantu Expansion'.* The Journal of African History, 1995. **36**(2): p. 173-195.

63. Chami, F.A., *A Response to Christopher Ehret's "Bantu Expansions".* The International Journal of African Historical Studies, 2001. **34**(3): p. 647-651.

64. Ehret, C., *Bantu Expansions: Re-Envisioning a Central Problem of Early African History.* The International Journal of African Historical Studies, 2001. **34**(1): p. 5-41.

65. Ehret, C., *Nilo-Saharans and the Saharo-Sudanese Neolithic*, in *The Archaeology of Africa: Food, Metals and Towns*, P.S. T. Shaw, B. Andah and A. Okpoko, Editor. 1993, Routledge Press: London. p. 104-125.

66. Ehret, C., *Language and History* in *African Languages: An introduction*, B.H.a.D. Nurse, Editor. 2000, Cambridge University Press: Cambridge. p. 272-97.

67. Greenberg, J.H., *The languages of Africa*. Publication of the Indiana University Research Center in Anthropology, Folklore, and Linguistics,. 1963, Bloomington,: Indiana University. vi, 171 p.

68. Greenberg, J.H., *Studies in African Linguistic Classification: I. The Niger-Congo Family.* Southwestern Journal of Anthropology, 1949. **5**(2): p. 79-100.

69. Greenberg, J.H., *Linguistic Evidence Regarding Bantu Origins.* The Journal of African History, 1972. **13**(2): p. 189-216.

70. Lewis, M.P.e., *Ethnologue: Languages of the World, Sixteenth edition.* 2009, SIL International, Dallas Texas.

71. Blench, R., *Archaeology, language, and the African past*. The African archaeology series 10. 2006, Lanham, MD: AltaMira Press. xxv, 361 p.

72. Blench, R., *restructuring Niger Congo*, in *Niger Congo Symposium*. 2004: Paris.

73. Roland, O., *Bantu Genesis: An Inquiry into Some Problems of Early Bantu History.* African Affairs, 1966. **65**(260): p. 245-258.

74. Nurse, D., *The Contributions of Linguistics to the Study of History in Africa.* The Journal of African History, 1997. **38**(3): p. 359-391.

75. Schoenbrun, D.L., *We Are What We Eat: Ancient Agriculture between the Great Lakes.* The Journal of African History, 1993. **34**(1): p. 1-31.

76. Twaddle, M., *Towards an Early History of the East African Interior.* History in Africa, 1975. **2**: p. 147-184.

77. Dundas, K.R., *76. Notes on the Origin and History of the Kikuyu and Dorobo Tribes.* Man, 1908. **8**: p. 136-139.

78. Bender, M.L., *Livestock and Linguistics in North and East African Ethnohistory.* Current Anthropology, 1982. **23**(3): p. 316-317.

79. Ehret, C., *A historical-comparative reconstruction of Nilo-Saharan.* 2001, Köln Rudiger Köppe.

80. Blench, R., *Recent developments in African language classification and their implications for prehistory*, in *The Archaeology of Africa: Food, Metals and Towns*, P.S. T. Shaw, B. Andah and A. Okpoko, Editor. 1993, Routledge Press: London. p. 71-103.

81. Blench, R. and K.C. MacDonald, *The origins and development of African livestock : archaeology, genetics, linguistics, and ethnography*. 2000, London ; New York, NY: UCL Press. xx, 546 p.

82. Bender, L., *Nilo-Saharan*, in *African Languages: An Introduction*, B.a.N. Heine, D., Editor. 2000, Cambridge University Press.

83. *Ehret, C.*, *Historical-Comparative Reconstruction of Nilo-Saharan*. 2001, Rüdiger Köppe Verlag: Cologne.

84. Sutton, J.E.G., *The Aquatic Civilization of Middle Africa.* The Journal of African History, 1974. **15**(4): p. 527-546.

85. Ehret, C., *The Civilizations of Africa: A History to 1800*. 2002, Charlottesville: University Press of Virginia

86. Lynch, B.M. and L.H. Robbins, *Namoratunga: The First Archeoastronomical Evidence in Sub-Saharan Africa.* Science, 1978. **200**(4343): p. 766-768.

87. Ehret, C., *Southern Nilotic history; linguistic approaches to the study of the past*. 1971, Evanston, Ill.,: Northwestern University Press. xiv, 200 p.

88. Ogot, B.A., *History of the southern Luo*. 1967, [Nairobi]: East African Pub. House. v.

89. Wrigley, C.C., *The Problem of the Lwo.* History in Africa, 1981. **8**: p. 219-246.

90. Ayodo, G., et al., *Combining evidence of natural selection with association analysis increases power to detect malaria-resistance variants.* Am J Hum Genet, 2007. **81**(2): p. 234-42.

91. Allison, A.C., E.W. Ikin, and A.E. Mourant, *Further Observations on Blood Groups in East African Tribes.* The Journal of the Royal Anthropological Institute of Great Britain and Ireland, 1954. **84**(1/2): p. 158-162.

92. Allison, A.C., et al., *Blood Groups in Some East African Tribes.* The Journal of the Royal Anthropological Institute of Great Britain and Ireland, 1952. **82**(1): p. 55-61.

93. Ssebabi, E.C.T. and D.F. Roberts, *East African Frequencies in Some Lesser Known Blood Group Systems.* Man, 1975. **10**(4): p. 524-529.

94. Roberts, D.F., *Serology and the History of the Northern Nilotes.* The Journal of African History, 1962. **3**(2): p. 301-305.

95. Roberts, D.F., E.W. Ikin, and A.E. Mourant, *Blood groups of the northern Nilotes.* Ann Hum Genet, 1955. **20**(2): p. 135-54.

96. Ehret, C., *Linguistic stratigraphies and Holocene history in Northeastern Africa*, in *Studies in African Archaeology*. 2006, Poznan Archaeological Museum: Poznan. p. 1019 - 1055.

97. Ehret, C., *Who were the Rock Artists? Identifying the Holocene populations of the Sahara*, in *International Rock Art Congress*. 1995, IFRAO: Turin.

98. Greenberg, J.H., *Studies in African Linguistic Classification: V. The Eastern Sudanic Family.* Southwestern Journal of Anthropology, 1950. **6**(2): p. 143-160.

99. Bender, L., *Livestock and Linguistics in North and East African ethnohistory.* Current Anthropology, 1982. **23**: p. 316-318.

100. Fleming, H., *Chadic external relations*, in *Studies in Chadic and Afroasiatic Linguistics*, E.W.a.H. Meyer-Bahlburg, Editor. 1983, Helmut Buske: Hamburg. p. 17-31.

101. Ehret, C., *Reconstructing Proto-Afroasiatic (Proto-Afrasian) : vowels, tone, consonants, and vocabulary*. University of California publications in linguistics v. 126. 1995, Berkeley: University of California Press. xvii, 557 p.

102. Fleming, H.C., *Ethiopians and East Africans.* The International Journal of African Historical Studies, 1978. **11**(2): p. 267-282.

103. Roland, O., *Review: Nilotic Ambiguities.* The Journal of African History, 1983. **24**(4): p. 539.

104. Schlee, G.n., *Inter-Ethnic Clan Identities, Ethnicity, Centrisms and Biases; A Response to Paul Spencer.* Africa: Journal of the International African Institute, 1993. **63**(4): p. 591-600.

105. Schlee, G.n., *Interethnic Clan Identities among Cushitic-Speaking Pastoralists.* Africa: Journal of the International African Institute, 1985. **55**(1): p. 17-38.

106. Lewis, H.S., *The Origins of the Galla and Somali.* The Journal of African History, 1966. **7**(1): p. 27-46.

107. Lynch, B.M. and L.H. Robbins, *Cushitic and Nilotic Prehistory: New Archaeological Evidence from North-West Kenya.* The Journal of African History, 1979. **20**(3): p. 319-328.

108. Distefano, J.A., *Hunters or Hunted? Towards a History of the Okiek of Kenya.* History in Africa, 1990. **17**: p. 41-57.

109. Ambrose, S.H., *historical and Linguistic reconstructions in East Africa; the archeological evidence*, in *The archeological and linguistic reconstruction of African History*, E.C.a.P. M, Editor. 1982, Boston University African Studies Center: Berkeley.

110. Ambrose, S.H., *East African Neolithic*, in *Encyclopedia of prehistory: Africa*, P.N.a.E. Peregrine, M., Editor. 2003, Kluwer academic/plenum publishers. p. 97-109.

111. Ehret, C. and Thomas Leiper Kane Collection (Library of Congress. Hebraic Section), *Ethiopians and East Africans : the problem of contacts*. Nairobi historical studies 3. 1974, Nairobi: East African Pub. House. 95 p.

112. Fleming, H.C., *Asa and Aramanik: Cushitic Hunters in Masai-Land.* Ethnology, 1969. **8**(1): p. 1-36.

113. Stiles, D., *Hunters of the Northern East African Coast: Origins and Historical Processes.* Africa: Journal of the International African Institute, 1981. **51**(4): p. 848-862.

114. Heine, B., *The sam Languages. A history of Rendille, Boni and Somali.* Afroasiatic linguistics, 1978. **6**: p. 25-115.

115. Ehret, C., *The historical reconstruction of Southern Cushitic phonology and vocabulary*. Kölner Beiträge zur Afrikanistik 5. Bd. 1980, Berlin: Reimer. 407 p.

116. Tucker, A.N., *Fringe Cushitic: An Experiment in Typological Comparison.* Bulletin of the School of Oriental and African Studies, University of London, 1967. **30**(3): p. 655-680.

117. **Haspelmath M., D.M.S., Gil D., Comrie B. eds.**, *The World Atlas of Language Structures Online*. 2008, ***Munich: Max Planck Digital Library.***

118. International, S. Ethnologue: Languages of the World 2005; 15:[

119. Ambrose, S., *Hunter-gatherer adaptations to non-marginal environments: An ecological and archaeological assessment of the Dorobo Model.* Sprache und Geschichte in Afrika (SUGIA), 1986. **7.2** p. 11-42.

120. Sutton, J., *The settlement of East Africa*, in *Zamani (New edition)*, B.A. Ogot, Editor. 1973: Nairobi. p. 70-97

121. Smith, A.B., *Origins and Spread of Pastoralism in Africa.* Annual Review of Anthropology, 1992. **21**: p. 125-141.

122. Kenny, M.G., *Mirror in the Forest: The Dorobo Hunter-Gatherers as an Image of the Other.* Africa: Journal of the International African Institute, 1981. **51**(1): p. 477-495.

123. Kassam, A., *The People of the Five ''Drums'': Gabra Ethnohistorical Origins Ethnohistory*, 2006. **53**(1): p. 173-193

124. Spielmann, K.A. and J.F. Eder, *Hunters and Farmers : Then and Now.* Annual Review of Anthropology, 1994. **23**: p. 303-323.

125. Huntingford, G.W.B., *Modern Hunters: Some Account of the Kamelilo-Kapchepkendi Dorobo (Okiek) of Kenya Colony.* The Journal of the Royal Anthropological Institute of Great Britain and Ireland, 1929. **59**: p. 333-378.

126. Ehret, C., *The civilizations of Africa : a history to 1800*. 2002, Charlottesville: University Press of Virginia. xii, 480 p.

127. Leakey, L.S.B., *The stone age cultures of Kenya Colony*. 1931, Cambridge, [Eng.]: The University Press. xiii, 287, [1] p. incl. plates. 2 fold. maps.

128. Bower, J.R.F. and C.M. Nelson, *Early Pottery and Pastoral Cultures of the Central Rift Valley, Kenya.* Man, 1978. **13**(4): p. 554-566.

129. Bower, J.R.F., *Seronera: excavations at a stone bowl site in the Serengeti National Park, Tanzania.* Azania, 1973. **8**: p. 71-104.

130. Barthelme, J.W., *Holocene sites north-east of Lake Turkana: preliminary report.* Azania, 1977. **12**: p. 33-41.

131. Butzer, K.W., Brown F. H., Thurber, D. L. , *Horizontal sedimentsof the lower Omo Basin: the Kibish Formation.* Quarternaria, 1969. **11**: p. 15-29.

132. Robbins, L.H., *Archeology in the Turkana District, Kenya.* Science, 1972. **176**(4033): p. 359-366.

133. Ambrose, S.H., *Chronology of the Later Stone Age and Food Production in East Africa Journal of Archaeological Science* 1998. **25**: p. 377-392.

134. Posnansky, M., *Pottery Types from Archaeological Sites in East Africa.* The Journal of African History, 1961. **2**(2): p. 177-198.

135. Posnansky, M., *168. Dimple-Based Pottery from Uganda.* Man, 1961. **61**: p. 141-142.

136. Soper, R.C., *Early Iron Age Pottery Types from East Africa: Comparative Analysis.* Azania, 1971. **6**: p. 39-52.

137. Soper, R.C., *New Radiocarbon Dates for Eastern and Southern Africa.* The Journal of African History, 1974. **15**(2): p. 175-192.

138. Sutton, J.E.G., *New Radiocarbon Dates for Eastern and Southern Africa.* The Journal of African History, 1972. **13**(1): p. 1-24.

139. Posnansky, M., *Bantu Genesis-Archaeological Reflexions.* The Journal of African History, 1968. **9**(1): p. 1-11.

140. Sutton, J.E.G., *101. Sirikwa Holes, Stone Houses and Their Makers in the Western Highlands of Kenya.* Man, 1965. **65**: p. 113-115.

141. Robertshaw, P., Collett, D. P., Gifford D. , Mbea, N., *Shell middens on the shores of Lake Victoria.* Azania, 1983. **18**: p. 1-43.

142. Morin, E., *Evidence for declines in human population densities during the early Upper Paleolithic in western Europe.* Proc Natl Acad Sci U S A, 2008. **105**(1): p. 48-53.

143. Gasse, F., *Hydrological changes in the African tropics since the Last Glacial Maximum.* Quaternary Science Reviews, 2000. **19**: p. 189-211.

144. Zhang, D.D., et al., *Global climate change, war, and population decline in recent human history.* Proc Natl Acad Sci U S A, 2007. **104**(49): p. 19214-9.

145. Ray, N. and L. Excoffier, *Inferring past demography using spatially explicit population genetic models.* Hum Biol, 2009. **81**(2-3): p. 141-57.

146. Scholz, C.A., et al., *East African megadroughts between 135 and 75 thousand years ago and bearing on early-modern human origins.* Proc Natl Acad Sci U S A, 2007. **104**(42): p. 16416-21.

147. Hassan, F.A., *Holocene Palaeoclimates of Africa.* African Archaeological Review, 1997. **14**(4): p. 213-230.

148. Mellars, P., *Why did modern human populations disperse from Africa ca. 60,000 years ago? A new model.* Proc Natl Acad Sci U S A, 2006. **103**(25): p. 9381-6.

149. Powell, A., S. Shennan, and M.G. Thomas, *Late Pleistocene demography and the appearance of modern human behavior.* Science, 2009. **324**(5932): p. 1298-301.

150. Passarino, G., et al., *Different genetic components in the Ethiopian population, identified by mtDNA and Y-chromosome polymorphisms.* Am J Hum Genet, 1998. **62**(2): p. 420-34.

151. Watson, E., et al., *mtDNA sequence diversity in Africa.* Am J Hum Genet, 1996. **59**(2): p. 437-44.

152. Watson, E., et al., *Mitochondrial footprints of human expansions in Africa.* Am J Hum Genet, 1997. **61**(3): p. 691-704.

153. Cruciani, F., et al., *Phylogeographic analysis of haplogroup E3b (E-M215) y chromosomes reveals multiple migratory events within and out of Africa.* Am J Hum Genet, 2004. **74**(5): p. 1014-22.

154. Krings, M., et al., *mtDNA analysis of Nile River Valley populations: A genetic corridor or a barrier to migration?* Am J Hum Genet, 1999. **64**(4): p. 1166-76.

155. Semino, O., et al., *Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny.* Am J Hum Genet, 2002. **70**(1): p. 265-8.

156. Luis, J.R., et al., *The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations.* Am J Hum Genet, 2004. **74**(3): p. 532-44.

157. Behar, D.M., et al., *The dawn of human matrilineal diversity.* Am J Hum Genet, 2008. **82**(5): p. 1130-40.

158. Hammer, M.F.a.Z., Z.L. , *THE HUMAN Y CHROMOSOME HAPLOGROUP TREE: Nomenclature and Phylogeography of Its Major Divisions.* Annual Review of Anthropology, 2002. **31**: p. 303-321.

159. Thomson, R., et al., *Recent common ancestry of human Y chromosomes: evidence from DNA sequence data.* Proc Natl Acad Sci U S A, 2000. **97**(13): p. 7360-5.

160. Scozzari, R., et al., *Combined use of biallelic and microsatellite Y-chromosome polymorphisms to infer affinities among African populations.* Am J Hum Genet, 1999. **65**(3): p. 829-46.

161. Seielstad, M.T., E. Minch, and L.L. Cavalli-Sforza, *Genetic evidence for a higher female migration rate in humans.* Nat Genet, 1998. **20**(3): p. 278-80.

162. Durbin et al., *A map of human genome variation from population-scale sequencing.* Nature. **467**(7319): p. 1061-1073.

163. Underhill, P.A., et al., *The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations.* Ann Hum Genet, 2001. **65**(Pt 1): p. 43-62.

164. Cruciani, F., et al., *A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes.* Am J Hum Genet, 2002. **70**(5): p. 1197-214.

165. Goncalves, R., et al., *Y-chromosome lineages in Cabo Verde Islands witness the diverse geographic origin of its first male settlers.* Hum Genet, 2003. **113**(6): p. 467-72.

166. Ingman, M. and U. Gyllensten, *Analysis of the complete human mtDNA genome: methodology and inferences for human evolution.* J Hered, 2001. **92**(6): p. 454-61.

167. Ingman, M., et al., *Mitochondrial genome variation and the origin of modern humans.* Nature, 2000. **408**(6813): p. 708-13.

168. Macaulay, V., M. Richards, and B. Sykes, *Mitochondrial DNA recombination-no need to panic.* Proc Biol Sci, 1999. **266**(1433): p. 2037-9; discussion 2041-2.

169. Brown, W.M., M. George, Jr., and A.C. Wilson, *Rapid evolution of animal mitochondrial DNA.* Proc Natl Acad Sci U S A, 1979. **76**(4): p. 1967-71.

170. Giles, R.E., et al., *Maternal inheritance of human mitochondrial DNA.* Proc Natl Acad Sci U S A, 1980. **77**(11): p. 6715-9.

171. Macaulay, V., et al., *The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs.* Am J Hum Genet, 1999. **64**(1): p. 232-49.

172. Richards, M.B., et al., *Phylogeography of mitochondrial DNA in western Europe.* Ann Hum Genet, 1998. **62 (Pt 3)**: p. 241-60.

173. Richards, M. and V. Macaulay, *The mitochondrial gene tree comes of age.* Am J Hum Genet, 2001. **68**(6): p. 1315-20.

174. Tamura, K. and M. Nei, *Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees.* Mol Biol Evol, 1993. **10**(3): p. 512-26.

175. Wakeley, J., *Substitution rate variation among sites in hypervariable region 1 of human mitochondrial DNA.* J Mol Evol, 1993. **37**(6): p. 613-23.

176. Herrnstadt, C., et al., *Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups.* Am J Hum Genet, 2002. **70**(5): p. 1152-71.

177. Maca-Meyer, N., et al., *Major genomic mitochondrial lineages delineate early human expansions.* BMC Genet, 2001. **2**(1): p. 13.

178. Mishmar, D., et al., *Natural selection shaped regional mtDNA variation in humans.* Proc Natl Acad Sci U S A, 2003. **100**(1): p. 171-6.

179. Wallace, D.C., M.D. Brown, and M.T. Lott, *Mitochondrial DNA variation in human evolution and disease.* Gene, 1999. **238**(1): p. 211-30.

180. Alves-Silva, J., et al., *The ancestry of Brazilian mtDNA lineages.* Am J Hum Genet, 2000. **67**(2): p. 444-61.

181. Bandelt, H.J., et al., *Phylogeography of the human mitochondrial haplogroup L3e: a snapshot of African prehistory and Atlantic slave trade.* Ann Hum Genet, 2001. **65**(Pt 6): p. 549-63.

182. Bandelt, H.J., et al., *Mitochondrial portraits of human populations using median networks.* Genetics, 1995. **141**(2): p. 743-53.

183. Chen, Y.S., et al., *mtDNA variation in the South African Kung and Khwe-and their genetic relationships to other African populations.* Am J Hum Genet, 2000. **66**(4): p. 1362-83.

184. Graven, L., et al., *Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample.* Mol Biol Evol, 1995. **12**(2): p. 334-45.

185. Salas, A., et al., *The making of the African mtDNA landscape.* Am J Hum Genet, 2002. **71**(5): p. 1082-111.

186. Soodyall, H., et al., *mtDNA control-region sequence variation suggests multiple independent origins of an "Asian-specific" 9-bp deletion in sub-Saharan Africans.* Am J Hum Genet, 1996. **58**(3): p. 595-608.

187. Torroni, A., et al., *Do the four clades of the mtDNA haplogroup L2 evolve at different rates?* Am J Hum Genet, 2001. **69**(6): p. 1348-56.

188. Huelsenbeck, J.P. and F. Ronquist, *MRBAYES: Bayesian inference of phylogenetic trees.* Bioinformatics, 2001. **17**(8): p. 754-5.

189. Huelsenbeck, J.P., et al., *Bayesian inference of phylogeny and its impact on evolutionary biology.* Science, 2001. **294**(5550): p. 2310-4.

190. Soares, P., et al., *Correcting for purifying selection: an improved human mitochondrial molecular clock.* Am J Hum Genet, 2009. **84**(6): p. 740-59.

191. Alexe, G., et al., *PCA and clustering reveal alternate mtDNA phylogeny of N and M clades.* J Mol Evol, 2008. **67**(5): p. 465-87.

192. Bandelt, H.J. and P. Forster, *The myth of bumpy hunter-gatherer mismatch distributions.* Am J Hum Genet, 1997. **61**(4): p. 980-3.

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

194. Weber, J.L., et al., *Human diallelic insertion/deletion polymorphisms.* Am J Hum Genet, 2002. **71**(4): p. 854-62.

195. Kimura, M. and T. Ohta, *Stepwise mutation model and distribution of allelic frequencies in a finite population.* Proc Natl Acad Sci U S A, 1978. **75**(6): p. 2868-72.

196. Di Rienzo, A., et al., *Mutational processes of simple-sequence repeat loci in human populations.* Proc Natl Acad Sci U S A, 1994. **91**(8): p. 3166-70.

197. Goldstein, D.B.S.t., C., *eds.* , *Microsatellites: Evolution and Applications*. 1999, Oxford: Oxford Univ. Press.

198. Kino, Y., et al., *Muscleblind protein, MBNL1/EXP, binds specifically to CHHG repeats.* Hum Mol Genet, 2004. **13**(5): p. 495-507.

199. Miller, J.W., et al., *Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy.* EMBO J, 2000. **19**(17): p. 4439-48.

200. Ho, T.H., et al., *Colocalization of muscleblind with RNA foci is separable from mis-regulation of alternative splicing in myotonic dystrophy.* J Cell Sci, 2005. **118**(Pt 13): p. 2923-33.

201. Stallings, R.L., et al., *Evolution and distribution of (GT)n repetitive sequences in mammalian genomes.* Genomics, 1991. **10**(3): p. 807-15.

202. Jorde, L.B., et al., *Microsatellite diversity and the demographic history of modern humans.* Proc Natl Acad Sci U S A, 1997. **94**(7): p. 3100-3.

203. Bowcock, A.M., et al., *High resolution of human evolutionary trees with polymorphic microsatellites.* Nature, 1994. **368**(6470): p. 455-7.

204. Coia, V., et al., *Brief communication: mtDNA variation in North Cameroon: Lack of asian lineages and implications for back migration from Asia to sub-Saharan Africa.* Am J Phys Anthropol, 2005.

205. Jorde, L.B., et al., *Origins and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data.* Am J Hum Genet, 1995. **57**(3): p. 523-38.

206. Mountain, J.L. and L.L. Cavalli-Sforza, *Inference of human evolution through cladistic analysis of nuclear DNA restriction polymorphisms.* Proc Natl Acad Sci U S A, 1994. **91**(14): p. 6515-9.

207. Wilder, J.A., Z. Mobasher, and M.F. Hammer, *Genetic evidence for unequal effective population sizes of human females and males.* Mol Biol Evol, 2004. **21**(11): p. 2047-57.

208. Birky, C.W., Jr., T. Maruyama, and P. Fuerst, *An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results.* Genetics, 1983. **103**(3): p. 513-27.

209. Pritchard, J.K., et al., *Population growth of human Y chromosomes: a study of Y chromosome microsatellites.* Mol Biol Evol, 1999. **16**(12): p. 1791-8.

210. Tang, H., et al., *Frequentist estimation of coalescence times from nucleotide sequence data using a tree-based partition.* Genetics, 2002. **161**(1): p. 447-59.

211. Hamilton, G., M. Stoneking, and L. Excoffier, *Molecular analysis reveals tighter social regulation of immigration in patrilocal populations than in matrilocal populations.* Proc Natl Acad Sci U S A, 2005. **102**(21): p. 7476-80.

212. Hammer, M.F., et al., *Human population structure and its effects on sampling Y chromosome sequence variation.* Genetics, 2003. **164**(4): p. 1495-509.

213. Fischer, M., *Ethnographic Atlas Codebook*

214. Miller, S.A., D.D. Dykes, and H.F. Polesky, *A simple salting out procedure for extracting DNA from human nucleated cells.* Nucleic Acids Res, 1988. **16**(3): p. 1215.

215. Rowold, D.J. and R.J. Herrera, *Alu elements and the human genome.* Genetica, 2000. **108**(1): p. 57-72.

216. Yconsortium, *A nomenclature system for the tree of human Y-chromosomal binary haplogroups.* Genome Res, 2002. **12**(2): p. 339-48.

217. Anderson, S., et al., *Sequence and organization of the human mitochondrial genome.* Nature, 1981. **290**(5806): p. 457-65.

218. Gonder, M.K., et al., *Whole-mtDNA genome sequence analysis of ancient African lineages.* Mol Biol Evol, 2007. **24**(3): p. 757-68.

219. Behar, D.M., et al., *Counting the founders: the matrilineal genetic ancestry of the Jewish Diaspora.* PLoS ONE, 2008. **3**(4): p. e2062.

220. Kivisild, T., et al., *The role of selection in the evolution of human mitochondrial genomes.* Genetics, 2006. **172**(1): p. 373-87.

221. Torroni, A., et al., *Harvesting the fruit of the human mtDNA tree.* Trends Genet, 2006. **22**(6): p. 339-45.

222. Olivieri, A., et al., *The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa.* Science, 2006. **314**(5806): p. 1767-70.

223. Derenko, M., et al., *Phylogeographic analysis of mitochondrial DNA in northern Asian populations.* Am J Hum Genet, 2007. **81**(5): p. 1025-41.

224. Gasparre, G., et al., *Disruptive mitochondrial DNA mutations in complex I subunits are markers of oncocytic phenotype in thyroid tumors.* Proc Natl Acad Sci U S A, 2007. **104**(21): p. 9001-6.

225. Achilli, A., et al., *The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool.* Am J Hum Genet, 2004. **75**(5): p. 910-8.

226. Just, R.S., et al., *Complete mitochondrial genome sequences for 265 African American and U.S. "Hispanic" individuals.* Forensic Sci Int Genet, 2008. **2**(3): p. e45-8.

227. Costa, M.D., Cherni, L., Fernandes,V., Freitas, F., el-Gaaied,A.B.A.and Pereira, L, *Data from complete mtDNA sequencing of Tunisian centenarians: testing haplogroup association and the' golden mean 'to longevity*. 2008.

228. Kujanova, M., Pereira, L., Fernandes,V., Pereira, J.B. and Cerny,V. , *Predominantly Neolithic contribution o the population history of the Egyptian Western Desert; mtDNA and Y chromosome analyses from The oasis of el-Hayez* 2008.

229. Malyarchuk, B., Derenko, M., Perkova, M., Grzybowski,T., Vanecek, T. and Lazur, J., *Identification of African mitochondrial DNA lineages in Slavs*. 2003.

230. Greenspan, B. 2008, Family Tree DNA - Genealogy Genetics, Ltd.. Houston, TX.

231. GeneCode, *Sequencher: software for DNA sequencing*. 2008, Gene Codes corporation: Ann Arbor, Michigan.

232. Andrews, R.M., et al., *Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA.* Nat Genet, 1999. **23**(2): p. 147.

233. Fu, Y.X., *Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection.* Genetics, 1997. **147**(2): p. 915-25.

234. Tajima, F., *Statistical method for testing the neutral mutation hypothesis by DNA polymorphism.* Genetics, 1989. **123**(3): p. 585-95.

235. Fu, Y.X. and W.H. Li, *Statistical tests of neutrality of mutations.* Genetics, 1993. **133**(3): p. 693-709.

236. Rozas, J., et al., *DnaSP, DNA polymorphism analyses by the coalescent and other methods.* Bioinformatics, 2003. **19**(18): p. 2496-7.

237. Watterson, G.A., *On the number of segregating sites in genetical models without recombination.* Theor Popul Biol, 1975. **7**(2): p. 256-76.

238. Tajima, F., *Evolutionary relationship of DNA sequences in finite populations.* Genetics, 1983. **105**(2): p. 437-60.

239. Innan, H. and W. Stephan, *The coalescent in an exponentially growing metapopulation and its application to Arabidopsis thaliana*. Genetics, 2000. **155**(4): p. 2015-9.

240. Tajima, F., *The effect of change in population size on DNA polymorphism.* Genetics, 1989. **123**(3): p. 597-601.

241. Aris-Brosou, S. and L. Excoffier, *The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism*. Mol Biol Evol, 1996. **13**(3): p. 494-504.

242. Bertorelle, G. and M. Slatkin, *The number of segregating sites in expanding human populations, with implications for estimates of demographic parameters.* Mol Biol Evol, 1995. **12**(5): p. 887-92.

243. Elson, J.L., D.M. Turnbull, and N. Howell, *Comparative genomics and the evolution of human mitochondrial DNA: assessing the effects of selection.* Am J Hum Genet, 2004. **74**(2): p. 229-38.

244. Eltsov, N., *New tool (mtPhyl) proposed for phylogenetic analysis of human complete mitochondrial genomes*. 2009: Novosibirsk, Russia.

245. Rand, D.M. and L.M. Kann, *Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from Drosophila, mice, and humans.* Mol Biol Evol, 1996. **13**(6): p. 735-48.

246. Nachman, M.W. and S.L. Crowell, *Estimate of the mutation rate per nucleotide in humans.* Genetics, 2000. **156**(1): p. 297-304.

247. Nachman, M.W., *Deleterious mutations in animal mitochondrial DNA*. Genetica, 1998. **102-103**(1-6): p. 61-9.

248. Weinreich, D.M. and D.M. Rand, *Contrasting patterns of nonneutral evolution in proteins encoded in nuclear and mitochondrial genomes.* Genetics, 2000. **156**(1): p. 385-99.

249. Rand, D.M., D.M. Weinreich, and B.O. Cezairliyan, *Neutrality tests of conservative-radical amino acid changes in nuclear- and mitochondrially-encoded proteins.* Gene, 2000. **261**(1): p. 115-25.

250. McDonald, J.H. and M. Kreitman, *Adaptive protein evolution at the Adh locus in Drosophila.* Nature, 1991. **351**(6328): p. 652-4.

251. Schneider, S., Roessli, D., Excoffier, L., *ARLEQUIN: A Software for Population Genetic Analysis*. 2000, Genetics and Biometry Laboratory, University of Geneva: Geneva, Switzerland.

252. Slatkin, M., *A measure of population subdivision based on microsatellite allele frequencies.* Genetics, 1995. **139**(1): p. 457-62.

253. Excoffier, L., P.E. Smouse, and J.M. Quattro, *Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data.* Genetics, 1992. **131**(2): p. 479-91.

254. Excoffier, L. and M. Slatkin, *Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population.* Mol Biol Evol, 1995. **12**(5): p. 921-7.

255. WRIGHT, S., *The genetical structure of populations.* Annals of Eugenics, 1951 **15**: p. 323-354.

256. Nei, M. and W.H. Li, *Mathematical model for studying genetic variation in terms of restriction endonucleases.* Proc Natl Acad Sci U S A, 1979. **76**(10): p. 5269-73.

257. Bandelt, H.J., V. Macaulay, and M. Richards, *Median networks: speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA.* Mol Phylogenet Evol, 2000. **16**(1): p. 8-28.

258. FluxusTechnology, *Network program*, http://www.fluxus-engineering.com: Clare, Suffolk, UK.

259. Bandelt, H.J., P. Forster, and A. Rohl, *Median-joining networks for inferring intraspecific phylogenies.* Mol Biol Evol, 1999. **16**(1): p. 37-48.

260. Forster, P., et al., *Origin and evolution of Native American mtDNA variation: a reappraisal.* Am J Hum Genet, 1996. **59**(4): p. 935-45.

261. Saillard, J., et al., *mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion.* Am J Hum Genet, 2000. **67**(3): p. 718-26.

262. Zhivotovsky, L.A., et al., *The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time.* Am J Hum Genet, 2004. **74**(1): p. 50-61.

263. Bianchi, N.O., et al., *Characterization of ancestral and derived Y-chromosome haplotypes of New World native populations.* Am J Hum Genet, 1998. **63**(6): p. 1862-71.

264. Heyer, E., et al., *Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees.* Hum Mol Genet, 1997. **6**(5): p. 799-803.

265. Kayser, M., et al., *Evaluation of Y-chromosomal STRs: a multicenter study.* Int J Legal Med, 1997. **110**(3): p. 125-33, 141-9.

266. Kayser, M., et al., *Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs.* Am J Hum Genet, 2000. **66**(5): p. 1580-8.

267. Strimmer, K. and O.G. Pybus, *Exploring the demographic history of DNA sequences using the generalized skyline plot.* Mol Biol Evol, 2001. **18**(12): p. 2298-305.

268. Pybus, O.G., A. Rambaut, and P.H. Harvey, *An integrated framework for the inference of viral population history from reconstructed genealogies.* Genetics, 2000. **155**(3): p. 1429-37.

269. Drummond, A., O.G. Pybus, and A. Rambaut, *Inference of viral evolutionary rates from molecular sequences.* Adv Parasitol, 2003. **54**: p. 331-58.

270. Drummond, A.J., et al., *Bayesian coalescent inference of past population dynamics from molecular sequences.* Mol Biol Evol, 2005. **22**(5): p. 1185-92.

271. Atkinson, Q.D., R.D. Gray, and A.J. Drummond, *Bayesian coalescent inference of major human mitochondrial DNA haplogroup expansions in Africa.* Proc Biol Sci, 2008.

272. Henn, B.M., et al., *Characterizing the time dependency of human mitochondrial DNA mutation rate estimates.* Mol Biol Evol, 2009. **26**(1): p. 217-30.

273. Atkinson, Q.D., R.D. Gray, and A.J. Drummond, *Bayesian coalescent inference of major human mitochondrial DNA haplogroup expansions in Africa.* Proc Biol Sci, 2009. **276**(1655): p. 367-73.

274. Endicott, P. and S.Y. Ho, *A Bayesian evaluation of human mitochondrial substitution rates.* Am J Hum Genet, 2008. **82**(4): p. 895-902.

275. Drummond, A.J. and A. Rambaut, *BEAST: Bayesian evolutionary analysis by sampling trees.* BMC Evol Biol, 2007. **7**: p. 214.

276. Fisher, R.A., *On the Interpretation of Ï‡2 from Contingency Tables, and the Calculation of P.* Journal of the Royal Statistical Society, 1922. **85**(1): p. 87-94.

277. Fisher, R.A., *Statistical methods for research workers*. 12th ed. Biological monographs and manuals, no. 5. 1954, Edinburgh,: Oliver and Boyd. 356p.

278. Agresti, A., *A Survey of Exact Inference for Contingency Tables.* Statistical Science, 1992. **7**(1): p. 131-153.

279. Marascuilo, L.A., *Large-sample multiple comparisons.* Psychol Bull, 1966. **65**(5): p. 280-90.

280. Team, R.D.C., *R: A Language and Environment for Statistical Computing*. 2010, R Foundation for Statistical Computing: Vienna, Austria.

281. Kruskal, J.B., *Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis.* Pyschometrika, 1964. **29**: p. 1– 27.

282. Mantel, N., *The detection of disease clustering and a generalized regression approach.* Cancer Res, 1967. **27**(2): p. 209-20.

283. Pritchard, J.K., M. Stephens, and P. Donnelly, *Inference of population structure using multilocus genotype data.* Genetics, 2000. **155**(2): p. 945-59.

284. Bamshad, M.J., et al., *Human population genetic structure and inference of group membership.* Am J Hum Genet, 2003. **72**(3): p. 578-89.

285. Evanno, G., S. Regnaut, and J. Goudet, *Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study*. Mol Ecol, 2005. **14**(8): p. 2611-20.

286. Falush, D., M. Stephens, and J.K. Pritchard, *Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies*. Genetics, 2003. **164**(4): p. 1567-87.

287. Pritchard, J.K., Wen, W. , *Documentation for STRUCTURE software: Version 2*. 2003, University of Chicago: Chicago.

288. Rosenberg, N.A., *distruct: a program for the graphical display of population structure.* Molecular Ecology Notes, 2004. **4**(1): p. 137-138.

289. Yang, M.C., et al., *Population subdivision of the tri-spine horseshoe crab, Tachypleus tridentatus, in Taiwan Strait.* Zoolog Sci, 2007. **24**(3): p. 219-24.

290. Hassan, H., Underhill PA, Cavalli-Sforza LL, Ibrahim ME, *Y-chromosome variation among Sudanese: restricted gene flow, concordance with language, geography, and history.* American Journal of Physical Anthropologist, 2008. **137**(3): p. 8.

291. Moran, C.N., et al., *Y chromosome haplogroups of elite Ethiopian endurance runners.* Hum Genet, 2004. **115**(6): p. 492-7.

292. Beleza, S., et al., *The genetic legacy of western Bantu migrations.* Hum Genet, 2005. **117**(4): p. 366-75.

293. King, T.E., et al., *Africans in Yorkshire? The deepest-rooting clade of the Y phylogeny within an English genealogy.* Eur J Hum Genet, 2007. **15**(3): p. 288-93.

294. Berniell-Lee, G., et al., *Genetic and demographic implications of the Bantu expansion: insights from human paternal lineages.* Mol Biol Evol, 2009.

295. Quintana-Murci, L., et al., *Strong maternal Khoisan contribution to the South African coloured population: a case of gender-biased admixture.* Am J Hum Genet. **86**(4): p. 611-20.

296. Vansina, J., *Bantu in the Crystal Ball, II.* History in Africa, 1980. **7**: p. 293-325.

297. Coelho, M., et al., *On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genetic variation in southwestern Angola.* BMC Evol Biol, 2009. **9**(1): p. 80.

298. Pereira, L., et al., *Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade.* Ann Hum Genet, 2001. **65**(Pt 5): p. 439-58.

299. Stefflova, K., et al., *Evaluation of group genetic ancestry of populations from Philadelphia and dakar in the context of sex-biased admixture in the americas.* PLoS ONE, 2009. **4**(11): p. e7842.

300. Silva, W.A., et al., *MtDNA haplogroup analysis of black Brazilian and sub-Saharan populations: implications for the Atlantic slave trade.* Hum Biol, 2006. **78**(1): p. 29-41.

301. Mendizabal, I., et al., *Genetic origin, admixture, and asymmetry in maternal and paternal human lineages in Cuba.* BMC Evol Biol, 2008. **8**: p. 213.

302. Destro-Bisol, G., et al., *The analysis of variation of mtDNA hypervariable region 1 suggests that Eastern and Western Pygmies diverged before the Bantu expansion.* Am Nat, 2004. **163**(2): p. 212-26.

303. Clinton, B., *Dating the Arrival of the Bedouin Tribes in Sinai and the Negev.* Journal of the Economic and Social History of the Orient, 1985. **28**(1): p. 20-49.

304. Cerny, V., et al., *mtDNA sequences of Chadic-speaking populations from northern Cameroon suggest their affinities with eastern Africa.* Ann Hum Biol, 2004. **31**(5): p. 554-69.

305. Saunier, J.L., et al., *Mitochondrial control region sequences from an Egyptian population sample.* Forensic Sci Int Genet, 2009. **3**(3): p. e97-103.

306. Gonzalez, A.M., et al., *Mitochondrial DNA variation in Mauritania and Mali and their genetic relationship to other Western Africa populations.* Ann Hum Genet, 2006. **70**(Pt 5): p. 631-57.

307. Rosa, A., et al., *MtDNA profile of West Africa Guineans: towards a better understanding of the Senegambia region.* Ann Hum Genet, 2004. **68**(Pt 4): p. 340-52.

308. Loueslati, B.Y., et al., *Islands inside an island: reproductive isolates on Jerba island.* Am J Hum Biol, 2006. **18**(1): p. 149-53.

309. Turchi, C., et al., *Polymorphisms of mtDNA control region in Tunisian and Moroccan populations: an enrichment of forensic mtDNA databases with Northern Africa data.* Forensic Sci Int Genet, 2009. **3**(3): p. 166-72.

310. Cherni, L., et al., *Post-last glacial maximum expansion from Iberia to North Africa revealed by fine characterization of mtDNA H haplogroup in Tunisia.* Am J Phys Anthropol, 2009. **139**(2): p. 253-60.

311. Fadhlaoui-Zid, K., et al., *Genetic diversity in Tunisia: a study based on the GM polymorphism of human immunoglobulins.* Hum Biol, 2004. **76**(4): p. 559-67.

312. Cerny, V., et al., *Regional differences in the distribution of the sub-Saharan, West Eurasian, and South Asian mtDNA lineages in Yemen.* Am J Phys Anthropol, 2008. **136**(2): p. 128-37.

313. Abu-Amero, K.K., et al., *Mitochondrial DNA structure in the Arabian Peninsula.* BMC Evol Biol, 2008. **8**: p. 45.

314. Richards, M., et al., *Tracing European founder lineages in the Near Eastern mtDNA pool.* Am J Hum Genet, 2000. **67**(5): p. 1251-76.

315. Alvarez, J.C., et al., *Characterization of human control region sequences for Spanish individuals in a forensic mtDNA data set.* Leg Med (Tokyo), 2007. **9**(6): p. 293-304.

316. Alvarez-Iglesias, V., et al., *New population and phylogenetic features of the internal variation within mitochondrial DNA macro-haplogroup R0.* PLoS ONE, 2009. **4**(4): p. e5112.

317. Jackson, B.A., et al., *Mitochondrial DNA genetic diversity among four ethnic groups in Sierra Leone.* Am J Phys Anthropol, 2005. **128**(1): p. 156-63.

318. Brehm, A., et al., *Mitochondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of Atlantic slave trade.* Ann Hum Genet, 2002. **66**(Pt 1): p. 49-60.

319. Alshamali, F., et al., *Mitochondrial DNA control region variation in Dubai, United Arab Emirates.* Forensic Sci Int Genet, 2008. **2**(1): p. e9-10.

320. Metspalu, M., et al., *Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans.* BMC Genet, 2004. **5**(1): p. 26.

321. Chandrasekar, A., et al., *Updating phylogeny of mitochondrial DNA macrohaplogroup m in India: dispersal of modern human in South Asian corridor.* PLoS ONE, 2009. **4**(10): p. e7447.

322. Hill, C., et al., *A mitochondrial stratigraphy for island southeast Asia.* Am J Hum Genet, 2007. **80**(1): p. 29-43.

323. Hill, C., et al., *Phylogeography and ethnogenesis of aboriginal Southeast Asians.* Mol Biol Evol, 2006. **23**(12): p. 2480-91.

324. Merriwether, D.A., et al., *Ancient mitochondrial M haplogroups identified in the Southwest Pacific.* Proc Natl Acad Sci U S A, 2005. **102**(37): p. 13034-9.

325. Rootsi, S., et al., *A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe.* Eur J Hum Genet, 2007. **15**(2): p. 204-11.

326. Martinez, L., et al., *Paleolithic Y-haplogroup heritage predominates in a Cretan highland plateau.* Eur J Hum Genet, 2007. **15**(4): p. 485-93.

327. Contu, D., et al., *Y-chromosome based evidence for pre-neolithic origin of the genetically homogeneous but diverse Sardinian population: inference for association scans.* PLoS ONE, 2008. **3**(1): p. e1430.

328. Nasidze, I., et al., *MtDNA and Y-chromosome variation in Kurdish groups.* Ann Hum Genet, 2005. **69**(Pt 4): p. 401-12.

329. Nasidze, I., et al., *Genetic evidence for the Mongolian ancestry of Kalmyks.* Am J Phys Anthropol, 2005. **128**(4): p. 846-54.

330. Malyarchuk, B., et al., *Differentiation of mitochondrial DNA and Y chromosomes in Russian populations.* Hum Biol, 2004. **76**(6): p. 877-900.

331. Nasidze, I., et al., *Mitochondrial DNA and Y-chromosome variation in the caucasus.* Ann Hum Genet, 2004. **68**(Pt 3): p. 205-21.

332. Wilson, J.F., et al., *Population genetic structure of variable drug response.* Nat Genet, 2001. **29**(3): p. 265-9.

333. Keinan, A. and D. Reich, *Can a sex-biased human demography account for the reduced effective population size of chromosome X in non-Africans?* Mol Biol Evol. **27**(10): p. 2312-21.

334. Keinan, A., et al., *Accelerated genetic drift on chromosome X during the human dispersal out of Africa.* Nat Genet, 2009. **41**(1): p. 66-70.

335. Amato, R., et al., *Genome-wide scan for signatures of human population differentiation and their relationship with natural selection, functional pathways and diseases.* PLoS ONE, 2009. **4**(11): p. e7927.

336. Kayser, M., et al., *Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea.* Am J Hum Genet, 2003. **72**(2): p. 281-302.

337. Oota, H., et al., *Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence.* Nat Genet, 2001. **29**(1): p. 20-1.

338. Hammer, M.F., et al., *Sex-biased evolutionary forces shape genomic patterns of human diversity.* PLoS Genet, 2008. **4**(9): p. e1000202.

339. Segurel, L., et al., *Sex-specific genetic structure and social organization in Central Asia: insights from a multi-locus study.* PLoS Genet, 2008. **4**(9): p. e1000200.

340. Destro-Bisol, G., et al., *Variation of female and male lineages in sub-Saharan populations: the importance of sociocultural factors.* Mol Biol Evol, 2004. **21**(9): p. 1673-82.

341. Wilkins, J.F. and F.W. Marlowe, *Sex-biased migration in humans: what should we expect from genetic data?* Bioessays, 2006. **28**(3): p. 290-300.

342. Wilkins, J.F., *Unraveling male and female histories from human genetic data.* Curr Opin Genet Dev, 2006. **16**(6): p. 611-7.

343. Li, J.Z., et al., *Worldwide human relationships inferred from genome-wide patterns of variation.* Science, 2008. **319**(5866): p. 1100-4.

344. Jakobsson, M., et al., *Genotype, haplotype and copy-number variation in worldwide human populations.* Nature, 2008. **451**(7181): p. 998-1003.

345. Prins, A.H.J., *The Didemic Diarchic Boni.* The Journal of the Royal Anthropological Institute of Great Britain and Ireland, 1963. **93**(2): p. 174-185.

346. Gomes, V., et al., *Digging deeper into East African human Y chromosome lineages.* Hum Genet. **127**(5): p. 603-13.

347. Vossen, R., *The eastern Nilotes : linguistic and historical reconstructions*. 1982, D. Reimer, Universität zu Köln.: Berlin. p. 512 p.

348. Bagshawe, F.J., *The Peoples of the Happy Valley (East Africa) The Aboriginal Races of Kondoa Irangi. Part III: The Sandawi.* Journal of the Royal African Society, 1925. **24**(95): p. 219-227.

349. Castri, L., et al., *mtDNA variability in two Bantu-speaking populations (Shona and Hutu) from Eastern Africa: Implications for peopling and migration patterns in sub-Saharan Africa.* Am J Phys Anthropol, 2009.

350. Oota, H., et al., *Recent origin and cultural reversion of a hunter-gatherer group.* PLoS Biol, 2005. **3**(3): p. e71.

351. Quintana-Murci, L., et al., *Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers.* Proc Natl Acad Sci U S A, 2008. **105**(5): p. 1596-601.

352. Verdu, P., et al., *Origins and genetic diversity of pygmy hunter-gatherers from Western Central Africa.* Curr Biol, 2009. **19**(4): p. 312-8.

353. Xu, S., et al., *Genetic evidence supports linguistic affinity of Mlabri--a hunter-gatherer group in Thailand.* BMC Genet. **11**: p. 18.

354. Bagshawe, F.J., *The Peoples of the Happy Valley (East Africa): The Aboriginal Races of Kondoa Irangi. Part IV: The Goroa and Their Kindred Hamitic Tribes, the Wambulu, or Erokh, the Wasi, or Alawa, and the Burungi.* Journal of the Royal African Society, 1925. **25**(97): p. 59-74.

355. Henn, B.M., et al., *Y-chromosomal evidence of a pastoralist migration through Tanzania to southern Africa.* Proc Natl Acad Sci U S A, 2008. **105**(31): p. 10693-8.

356. Tucker, A.N. and M.A. Bryan, *The 'Mbugu' Anomaly.* Bulletin of the School of Oriental and African Studies, University of London, 1974. **37**(1): p. 188-207.

357. Bagshawe, F.J., *The Peoples of the Happy Valley (East Africa): The Aboriginal Races of Kondoa Irangi. Part III (Continued).* Journal of the Royal African Society, 1925. **24**(96): p. 328-347.

358. Yellen, J.E., *Barbed Bone Points: Tradition and Continuity in Saharan and Sub-Saharan Africa* African Archaeological Review, 1998. **15**(3): p. 173-198.

359. Edwards, D.N., *The Archaeology of Sudan and Nubia.* Annual Review of Anthropology, 2007. **36**: p. 211-28.

360. Mohammed-Ali, A.S. and A.-R.M. Khabir, *The Wavy Line and the Dotted Wavy Line Pottery in the Prehistory of the Central Nile and the Sahara-Sahel Belt.* The African Archaeological Review, 2003. **20**(1): p. 25-58.

361. Irish, J.D., *Who were the ancient Egyptians? Dental affinities among Neolithic through postdynastic peoples.* Am J Phys Anthropol, 2006. **129**(4): p. 529-43.

362. Zakrzewski, S.R., *Population continuity or population change: formation of the ancient Egyptian state.* Am J Phys Anthropol, 2007. **132**(4): p. 501-9.

363. Schillaci, M.A., J.D. Irish, and C.C. Wood, *Further analysis of the population history of ancient Egyptians.* Am J Phys Anthropol, 2009.

364. Rilly, C., *The Linguistic position of Meroitic.* Sudan Electronic Journal of Archaeology and Anthropology, 2004.

365. Godde, K., *Lower Nubia. Who were the Meroites? A Biological Investigation into the Nubian Post-hiatus Group* International Journal of Osteoarchaeology, 2009. **DOI: 10.1002/oa**.

366. Brandstatter, A., et al., *Mitochondrial DNA control region sequences from Nairobi (Kenya): inferring phylogenetic parameters for the establishment of a forensic database.* Int J Legal Med, 2004. **118**(5): p. 294-306.

367. Di Rienzo, A. and A.C. Wilson, *Branching pattern in the evolutionary tree for human mitochondrial DNA.* Proc Natl Acad Sci U S A, 1991. **88**(5): p. 1597-601.

368. Veeramah, K.R., et al., *Little genetic differentiation as assessed by uniparental markers in the presence of substantial language variation in peoples of the Cross River region of Nigeria.* BMC Evol Biol. **10**: p. 92.

369. Stevanovitch, A., et al., *Mitochondrial DNA sequence diversity in a sedentary population from Egypt.* Ann Hum Genet, 2004. **68**(Pt 1): p. 23-39.

370. Hassan, H.Y., et al., *Y-chromosome variation among Sudanese: restricted gene flow, concordance with language, geography, and history.* Am J Phys Anthropol, 2008. **137**(3): p. 316-23.

371. Satzinger, H., *The Old Coptic Schmidt Papyrus.* Journal of the American Research Center in Egypt, 1975. **12**: p. 37-50.

372. Carleton, T.H., *An Egyptian Etymology: Egypto-Coptic mÈ□Ä□.* Anthropological Linguistics, 1997. **39**(2): p. 196-219.

373. Lindgren, B.r., *The Internal Dynamics of Ethnicity: Clan Names, Origins and Castes in Southern Zimbabwe.* Africa: Journal of the International African Institute, 2004. **74**(2): p. 173-193.

374. Ambrose, S., *Excavation at Deloraine, Rongai, 1978.* Azania, 1984. **19**(79-104).

375. Cohen, D.W. and E.S.A. Odhiambo, *Ayany, Malo, and Ogot: Historians in Search of a Luo Nation (Ayany, Malo et Ogot: des historiens Ã la recherche d'une nation luo).* Cahiers d'Ã‰tudes Africaines, 1987. **27**(107/108): p. 269-286.

376. Binagi, L.A., *Basuba, Baganda and Abhakuria.* Africa: Journal of the International African Institute, 1979. **49**(4): p. 418-419.

377. Kenny, M.G., *Letter to the Editor.* Africa: Journal of the International African Institute, 1980. **50**(3): p. 311-312.

378. Newman, P., *The Classification of Chadic within Afroasiatic*. 1980, Leiden: Leiden Universitaire Press.

379. Carleton, T.H., *Afroasiatic: The Horizon and beyond.* The Jewish Quarterly Review, 1983. **74**(2): p. 137-158.

380. Ehret, C., *An African Classical Age: Eastern and Southern Africa in World History, 1000 B.C. to A.D. 400*. 1998, Charlottesville: University Press of Virginia.

381. Munson, P.J., *Africa's Prehistoric Past*. Africa, ed. P.M.M.P. O'Meara. 1977, Bloomington, IN: University of Indiana press.

382. Blench, R. *The westward wanderings of Cushitic pastoralists.* in *L'homme et l'animal dans le basin du lac Tchad. Actes du colloque du Réseau Méga-Tchad Orleans*. 1999. New Orleans.

383. Lancaster, A., *Y haplogroups, archaeological cultures and language families: a review of the possibility of multidisciplinary comparisons using the case of E-M35.* Journal of Genetic Genealogy, 2009. **5**(1): p. 35-65.

384. Bereir, R.E., et al., *Co-introgression of Y-chromosome haplogroups and the sickle cell gene across Africa's Sahel.* Eur J Hum Genet, 2007. **15**(11): p. 1183-5.

385. Voigt, R. *Verbal conjugation in proto-chadic*. in *Symposium on chadic linguistics*. 1987. Boulder, Colorado.

386. Cruciani, F., et al., *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, 2006. **27**(8): p. 831-2.

387. Sr, E.M.v.Z.B., *Presidential Address: The Origin of Crop Cultivation with Special Reference to Africa.* The South African Archaeological Bulletin, 1980. **35**(132): p. 67-72.

388. Wendorf, F., Schild, R., Close, A., Hillman, G., Gautier, A., Van Neer, W., Donahue, D.J., Jull, A. and Linick, T.W., *New radiocarbon dates and Late Palaeolithic diet at Wadi Kubbaniya, Egypt.* Antiquity, 1988. **62**: p. 279-83.

389. Close, A.E., *Current Research and Recent Radiocarbon, Dates from Northern Africa, III.* The Journal of African History, 1988. **29**(2): p. 145-176.

390. McCorriston, J. and F. Hole, *The Ecology of Seasonal Stress and the Origins of Agriculture in the Near East.* American Anthropologist, 1991. **93**(1): p. 46-69.

391. Wright, K.I., *Ground-Stone Tools and Hunter-Gatherer Subsistence in Southwest Asia: Implications for the Transition to Farming.* American Antiquity, 1994. **59**(2): p. 238-263.

392. Bar-Yosef, O., *Pleistocene Connexions between Africa and Southwest Asia: An Archaeological Perspective.* The African Archaeological Review, 1987. **5**: p. 29-38.

393. Barham, L. and P. Mitchell, *The first Africans : African archaeology from the earliest tool makers to most recent foragers*. Cambridge world archaeology. 2008, Cambridge ; New York: Cambridge University Press. xvii, 601 p.

394. Olszewski, D.I., *Subsistence Ecology in the Mediterranean Forest: Implications for the Origins of Cultivation in the Epipaleolithic Southern Levant.* American Anthropologist, 1993. **95**(2): p. 420-435.

395. Ehret, C., *On the Antiquity of Agriculture in Ethiopia.* The Journal of African History, 1979. **20**(2): p. 161-177.

396. Gonzalez, A.M., et al., *Mitochondrial lineage M1 traces an early human backflow to Africa.* BMC Genomics, 2007. **8**: p. 223.

397. Non, A.L., et al., *Mitochondrial DNA reveals distinct evolutionary histories for Jewish populations in Yemen and Ethiopia.* Am J Phys Anthropol.

398. Plaza, S., et al., *Joining the pillars of Hercules: mtDNA sequences show multidirectional gene flow in the western Mediterranean.* Ann Hum Genet, 2003. **67**(Pt 4): p. 312-28.

399. Corte-Real, H.B., et al., *Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis.* Ann Hum Genet, 1996. **60**(Pt 4): p. 331-50.

400. Cherni, L., et al., *Female gene pools of Berber and Arab neighboring communities in central Tunisia: microstructure of mtDNA variation in North Africa.* Hum Biol, 2005. **77**(1): p. 61-70.

401. Rando, J.C., et al., *Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern, and sub-Saharan populations.* Ann Hum Genet, 1998. **62**(Pt 6): p. 531-50.

402. Coudray, C., et al., *The Complex and Diversified Mitochondrial Gene Pool of Berber Populations.* Ann Hum Genet, 2008.

403. Atkinson, Q.D., R.D. Gray, and A.J. Drummond, *mtDNA variation predicts population size in humans and reveals a major Southern Asian chapter in human prehistory.* Mol Biol Evol, 2008. **25**(2): p. 468-74.

404. Haak, W., et al., *Ancient DNA from the first European farmers in 7500-year-old Neolithic sites.* Science, 2005. **310**(5750): p. 1016-8.

405. Ricaut, F.X., et al., *Genetic analysis of a Scytho-Siberian skeleton and its implications for ancient Central Asian migrations.* Hum Biol, 2004. **76**(1): p. 109-25.

406. Bramanti, B., et al., *Genetic Discontinuity Between Local Hunter-Gatherers and Central Europe's First Farmers.* Science, 2009: p. 1176869.

407. Xiao, F.X., et al., *Human X-chromosomal lineages in Europe reveal Middle Eastern and Asiatic contacts.* Eur J Hum Genet, 2004. **12**(4): p. 301-11.

408. Xu, S., W. Jin, and L. Jin, *Haplotype-sharing analysis showing Uyghurs are unlikely genetic donors.* Mol Biol Evol, 2009. **26**(10): p. 2197-206.

409. Li, H., et al., *Genetic Landscape of Eurasia and "Admixture" in Uyghurs.* The American Journal of Human Genetics, 2009. **85**(6): p. 934-937.

410. Reich, D., et al., *Reconstructing Indian population history.* Nature, 2009. **461**(7263): p. 489-94.

411. The HUGO Pan-Asian SNP Consortium, *Mapping Human Genetic Diversity in Asia.* Science, 2009. **326**(5959): p. 1541-1545.

412. Eaaswarkhanth, M., et al., *Traces of sub-Saharan and Middle Eastern lineages in Indian Muslim populations.* Eur J Hum Genet, 2009.

413. Comas, D., et al., *Admixture, migrations, and dispersals in Central Asia: evidence from maternal DNA lineages.* Eur J Hum Genet, 2004. **12**(6): p. 495-504.

414. Chaix, R., et al., *From social to genetic structures in central Asia.* Curr Biol, 2007. **17**(1): p. 43-8.

415. Metspalu, M., et al., *Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans.* BMC Genet, 2004. **5**: p. 26.

416. Quintana-Murci, L., et al., *Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor.* Am J Hum Genet, 2004. **74**(5): p. 827-45.

417. Gutala, R., et al., *A shared Y-chromosomal heritage between Muslims and Hindus in India.* Hum Genet, 2006. **120**(4): p. 543-51.

418. Cordaux, R., et al., *The northeast Indian passageway: a barrier or corridor for human migrations?* Mol Biol Evol, 2004. **21**(8): p. 1525-33.

419. Firasat, S., et al., *Y-chromosomal evidence for a limited Greek contribution to the Pathan population of Pakistan.* Eur J Hum Genet, 2007. **15**(1): p. 121-6.

420. Zerjal, T., et al., *A genetic landscape reshaped by recent events: Y-chromosomal insights into central Asia.* Am J Hum Genet, 2002. **71**(3): p. 466-82.

421. Enattah, N.S., et al., *Identification of a variant associated with adult-type hypolactasia.* Nat Genet, 2002. **30**(2): p. 233-7.

422. Babu, J., et al., *Frequency of lactose malabsorption among healthy southern and northern Indian populations by genetic analysis and lactose hydrogen breath and tolerance tests.* Am J Clin Nutr, 2009.

423. Enattah, N.S., et al., *Independent introduction of two lactase-persistence alleles into human populations reflects different history of adaptation to milk culture.* Am J Hum Genet, 2008. **82**(1): p. 57-72.

424. Bersaglieri, T., et al., *Genetic signatures of strong recent positive selection at the lactase gene.* Am J Hum Genet, 2004. **74**(6): p. 1111-20.

425. Enattah, N.S., et al., *Evidence of still-ongoing convergence evolution of the lactase persistence T-13910 alleles in humans.* Am J Hum Genet, 2007. **81**(3): p. 615-25.

426. Sapir, E., *Perspective in Aboriginal American Culture, a Study in Method.* Geological Survey Memoir 90. Vol. 13. 1916, Ottawa: Canada Department of Mines.

427. Dyen, I., *Language Distribution and Migration Theory.* Language, 1956. **32**(4): p. 611-626.

428. Diebold, A.R., Jr., *Determining the Centers of Dispersal of Language Groups.* International Journal of American Linguistics, 1960. **26**(1): p. 1-10.

429. Myres, N.M., et al., *A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe.* Eur J Hum Genet.

430. Cruciani, F., et al., *Strong intra- and inter-continental differentiation revealed by Y chromosome SNPs M269, U106 and U152.* Forensic Sci Int Genet.

431. Balaresque, P., et al., *A predominantly neolithic origin for European paternal lineages.* PLoS Biol. **8**(1): p. e1000285.

432. Oppenheimer, S., *Out of Eden : the peopling of the world*. 2003, London: Constable & Robinson. xxi, 440 p., [16] p. of plates.

433. Oppenheimer, S., *The real Eve : modern man's journey out of Africa*. 2003, New York, N.Y.: Carroll & Graf. xxi, 440 p., [16] p. of plates.

434. Semino, O., et al., *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, 2004. **74**(5): p. 1023-34.

435. Richard, C., et al., *An mtDNA perspective of French genetic variation.* Ann Hum Biol, 2007. **34**(1): p. 68-79.

436. Pereira, L., et al., *African female heritage in Iberia: a reassessment of mtDNA lineage distribution in present times.* Hum Biol, 2005. **77**(2): p. 213-29.

437. Pereira, L., M.J. Prata, and A. Amorim, *Diversity of mtDNA lineages in Portugal: not a genetic edge of European variation.* Ann Hum Genet, 2000. **64**(Pt 6): p. 491-506.

438. Shlush, L.I., et al., *The Druze: a population genetic refugium of the Near East.* PLoS ONE, 2008. **3**(5): p. e2105.

439. Cerny, V., et al., *Out of Arabia-The settlement of Island Soqotra as revealed by mitochondrial and Y chromosome genetic diversity.* Am J Phys Anthropol, 2008.

440. Brehm, A., et al., *Mitochondrial portraits of the Madeira and Acores archipelagos witness different genetic pools of its settlers.* Hum Genet, 2003. **114**(1): p. 77-86.

441. Santos, C., et al., *Genetic structure and origin of peopling in the Azores islands (Portugal): the view from mtDNA.* Ann Hum Genet, 2003. **67**(Pt 5): p. 433-56.

442. Gonzalez, A.M., et al., *Mitochondrial DNA variation in Jordanians and their genetic relationship to other Middle East populations.* Ann Hum Biol, 2008. **35**(2): p. 212-31.

443. Dubut, V., et al., *mtDNA polymorphisms in five French groups: importance of regional sampling.* Eur J Hum Genet, 2004. **12**(4): p. 293-300.

444. Tillmar, A.O., et al., *Homogeneity in mitochondrial DNA control region sequences in Swedish subpopulations.* Int J Legal Med, 2009.

445. Malyarchuk, B.A., et al., *Mitochondrial DNA variability in Bosnians and Slovenians.* Ann Hum Genet, 2003. **67**(Pt 5): p. 412-25.

446. Robinson, S.A., Black, S. , Sellwood, B. W., Valdes, P. J., *A review of palaeoclimates and palaeoenvironments in the Levant and Eastern Mediterranean from 25,000 to 5000 years BP: setting the environmental background for the evolution of human civilization.* Quaternary Science Reviews 2006. **25**: p. 1517-1541.

447. deMenocal, P., Ortiz, J., Guilderson, T., Adkins, J., Sarnthein, M., Baker, L., Yarusinsky, M., *Abrupt onset and termination of the African Humid Period: rapid climate responses to gradual insolation forcing.* Quaternary Science Reviews, 2000. **19**: p. 347-361.

448. Beek, G.W.v., *Monuments of Axum in the Light of South Arabian Archeology.* Journal of the American Oriental Society, 1967. **87**(2): p. 113-122.

449. Phillipson, L., *Aksumite Lithic Industries African Archaeological Review*, 2000. **17**(2): p. *49-63*.

450. Dahl, O., *The Early History of Arabia.* The Old and New Testament Student, 1890. **10**(3): p. 153-159.

451. Butzer, K.W., *Rise and Fall of Axum, Ethiopia: A Geo-Archaeological Interpretation.* American Antiquity, 1981. **46**(3): p. 471-495.

452. Phillipson, L., *Lithic Artefacts as a Source of Cultural, Social and Economic Information: the evidence from Aksum, Ethiopia.* African Archaeological Review, 2009. **26**: p. 45-58.

453. Kobishchanow, Y.M., *On the Problem of Sea Voyages of Ancient Africans in the Indian Ocean.* The Journal of African History, 1965. **6**(2): p. 137-141.

454. Mayerson, P., *A Confusion of Indias: Asian india and African India in the Byzantine Sources.* Journal of the American Oriental Society, 1993. **113**(2): p. 169-174.

455. Kitchen, A., et al., *Bayesian phylogenetic analysis of Semitic languages identifies an Early Bronze Age origin of Semitic in the Near East.* Proc Biol Sci, 2009. **276**(1668): p. 2703-10.

456. Thomason, S.G.a.K., T., *Language Contact, Creolization, and Genetic Linguistics*. 1988, Berkeley and Los Angeles: University of California Press.

457. Leslau, W., *The Influence of Cushitic on the Semitic Languages of Ethiopia: A Problem of Substratum.* WORD, 1945. **1**: p. 59-82.

458. Leslau, W., *The Influence of Sidamo on the Ethiopic Languages of Gurage.* Language, 1952. **28**(1): p. 63-81.

459. Ehret, C., *Cushitic Prehistory*, in *In The Non-Semitic Languages of Ethiopia*, M.L. Bender, Editor. 1976, African Studies Center, Michigan State University: East Lansing, Michigan. p. 85-96.

460. Tosco, M., *Is There an "Ethiopian Language Area"?* Anthropological Linguistics, 2000. **42**(3): p. 329-365.

461. Tofanelli, S., et al., *J1-M267 Y lineage marks climate-driven pre-historical human displacements.* Eur J Hum Genet, 2009.

462. Chiaroni, J., et al., *The emergence of Y-chromosome haplogroup J1e among Arabic-speaking populations.* Eur J Hum Genet, 2009.

463. Alexander, J., *Islam, Archaeology and Slavery in Africa.* World Archaeology, 2001. **33**(1): p. 44-60.

464. Richards, M., et al., *Extensive female-mediated gene flow from sub-Saharan Africa into near eastern Arab populations.* Am J Hum Genet, 2003. **72**(4): p. 1058-64.

465. Godber, M.J., et al., *The Hereditary Blood Factors of the Yemenite and Kurdish Jews.* Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 1973. **266**(876): p. 169-184.

466. Epstein, H. and I.L. Mason, *The origin of the domestic animals of Africa*. 1971, [Leipzig]: Edition Leipzig. 2 v.

467. Ingram, C.J., et al., *A novel polymorphism associated with lactose tolerance in Africa: multiple causes for lactase persistence?* Hum Genet, 2007. **120**(6): p. 779-88.

468. Tishkoff, S.A., et al., *Convergent adaptation of human lactase persistence in Africa and Europe.* Nat Genet, 2007. **39**(1): p. 31-40.

469. Ranciaro, A., *History of pastoralism in Africa*. 2009.

470. Kidd, K.K., Cavalli-Sforza, L.L., *ROLE OF GENETIC DRIFT IN DIFFERENTIATION OF ICELANDIC AND NORWEGIAN CATTLE.* Evolution, 1974. **28**(3): p. 381-395.

471. Gifford-Gonzalez, D., *The Fauna from Ele Bor: Evidence for the Persistence of Foragers into the Later Holocene of Arid North Kenya.* African Archaeological Review, 2003. **20**(2): p. 81-119.

472. Smith, A.B., *Environmental limitations on prehistoric pastoralism in Africa.* African Archaeological Review, 1984. **2**: p. 99–111.

473. Ehret, C., *Sheep and Central Sudanic Peoples in Southern Africa.* The Journal of African History, 1968. **9**(2): p. 213-221.

474. Blench, R. and M. Spriggs, *Archaeology and language*. One world archaeology 27, 29, 34-35. 1997, London ; New York: Routledge.

475. Jeffreys, M.D.W., *240. Pastoral Customs.* Man, 1962. **62**: p. 154.

476. Blench, R.M., *Ethnographic and linguistic evidence for the prehistory of African ruminant livestock, horses and ponies*, in *The Archaeology of Africa. Food, Metals and Towns*, T. Shaw, Sinclair, P., Andah, B. and Okpoko, A. , Editor. 1993, Routledge: London, UK. p. 71-103.

477. Hanotte, O., et al., *African pastoralism: genetic imprints of origins and migrations.* Science, 2002. **296**(5566): p. 336-9.

478. Freeman, A.R., et al., *Assessing the relative ages of admixture in the bovine hybrid zones of Africa and the Near East using X chromosome haplotype mosaicism.* Genetics, 2006. **173**(3): p. 1503-10.

479. Bower, J., *The Pastoral Neolithic of East Africa.* Journal of world prehistory, 1991. **5**(1): p. 49-82.

480. Robbins, L.H., *On the prehistory of Livestock in North and East Africa. Current Anthropology*, 1983. **24**(12): p. 407.

481. Stynder, D.D., *Craniometric evidence for South African Later Stone Age herders and hunter–gatherers being a single biological population.* Journal of Archaeological Science, 2009. **36**(3): p. 798-806.

482. Stynder, D.D., R.R. Ackermann, and J.C. Sealy, *Craniofacial variation and population continuity during the South African Holocene.* Am J Phys Anthropol, 2007. **134**(4): p. 489-500.

483. Carvalho, E.C.D., *"Traditional" and "Modern" Patterns of Cattle Raising in Southwestern Angola: A Critical Evaluation of Change from Pastoralism to Ranching.* The Journal of Developing Areas, 1974. **8**(2): p. 199-226.

484. Roland, O., *The Nilotic Contribution to Bantu Africa.* The Journal of African History, 1982. **23**(4): p. 433-442.

485. Holden, C.J. and R. Mace, *Spread of Cattle Led to the Loss of Matrilineal Descent in Africa: A Coevolutionary Analysis.* Proceedings: Biological Sciences, 2003. **270**(1532): p. 2425-2433.

486. Phillipson, D.W., *The later prehistory of eastern and southern Africa*. 1977, New York: Africana Pub. Co. xii, 323 p.

487. Phillipson, D.W., *Archaeology and Bantu Linguistics.* World Archaeology, 1976. **8**(1): p. 65-82.

488. Sikora, M., et al., *A genomic analysis identifies a novel component in the genetic structure of sub-Saharan African populations.* Eur J Hum Genet.

489. Blench, R. *Was there an interchange between Cushitic pastoralists and Khoisan speakers in the prehistory of Southern Africa and how can this be detected*. in *Migration, Settlement and Cultural History on the Basis of Linguistic Sources*. 2007. Königswinter, Germany: *Sprache und Geschichte in Afrika*.

490. Ambrose, S.H., *Chronology of the Later Stone Age and Food Production in East Africa* Journal of Archaeological Science, 1998. **25**: p. 377-392.

491. Huffman, T.N., *The Early Iron Age and the Spread of the Bantu.* The South African Archaeological Bulletin, 1970. **25**(97): p. 3-21.

492. Gifford-Gonzalez, D., *Animal Disease Challenges to the Emergence of Pastoralism in Sub-Saharan Africa African Archaeological Review*, 2000. **17**(3): p. *95-139*.

493. Badenhorst, S., *Goats (Capra hircus), the Khoekhoen and Pastoralism:*

Current Evidence from Southern Africa African Archaeological Reviews, 2006. **23**: p. 45-53.

494. Marshall, F., *Rethinking the Role of Bos Indicus in Sub-Saharan Africa.* Current Anthropology, 1989. **30**: p. 235-39.

495. Garrigan, D., et al., *Deep haplotype divergence and long-range linkage disequilibrium at xp21.1 provide evidence that humans descend from a structured ancestral population.* Genetics, 2005. **170**(4): p. 1849-56.

496. Garrigan, D., et al., *Evidence for archaic Asian ancestry on the human X chromosome.* Mol Biol Evol, 2005. **22**(2): p. 189-92.

497. Yotova, V., et al., *Tracing genetic history of modern humans using X-chromosome lineages.* Hum Genet, 2007. **122**(5): p. 431-43.

498. Plagnol, V. and J.D. Wall, *Possible ancestral structure in human populations.* PLoS Genet, 2006. **2**(7): p. e105.

499. Wall, J.D., K.E. Lohmueller, and V. Plagnol, *Detecting ancient admixture and estimating demographic parameters in multiple human populations.* Mol Biol Evol, 2009.

500. Hey, J. and R. Nielsen, *Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of Drosophila pseudoobscura and D. persimilis.* Genetics, 2004. **167**(2): p. 747-60.

501. Nielsen, R. and J. Wakeley, *Distinguishing migration from isolation: a Markov chain Monte Carlo approach.* Genetics, 2001. **158**(2): p. 885-96.

502. Cornuet, J.M., et al., *Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation.* Bioinformatics, 2008. **24**(23): p. 2713-9.

503. Beaumont, M.A., W. Zhang, and D.J. Balding, *Approximate Bayesian computation in population genetics.* Genetics, 2002. **162**(4): p. 2025-35.

504. Sanchez, J.J., et al., *High frequencies of Y chromosome lineages characterized by E3b1, DYS19-11, DYS392-12 in Somali males.* Eur J Hum Genet, 2005. **13**(7): p. 856-66.

505. Shen, P., et al., *Reconstruction of patrilineages and matrilineages of Samaritans and other Israeli populations from Y-chromosome and mitochondrial DNA sequence variation.* Hum Mutat, 2004. **24**(3): p. 248-60.

506. Msaidie, S., et al., *Genetic diversity on the Comoros Islands shows early seafaring as major determinant of human biocultural evolution in the Western Indian Ocean.* Eur J Hum Genet.

507. Trovoada, M.J., et al., *Dissecting the genetic history of Sao Tome e Principe: a new window from Y-chromosome biallelic markers.* Ann Hum Genet, 2007. **71**(Pt 1): p. 77-85.

508. Scozzari, R., et al., *Differential structuring of human populations for homologous X and Y microsatellite loci.* Am J Hum Genet, 1997. **61**(3): p. 719-33.

509. Scozzari, R., et al., *Human Y-chromosome variation in the western Mediterranean area: implications for the peopling of the region.* Hum Immunol, 2001. **62**(9): p. 871-84.

510. Arredi, B., et al., *A predominantly neolithic origin for Y-chromosomal DNA variation in North Africa.* Am J Hum Genet, 2004. **75**(2): p. 338-45.

511. Robino, C., et al., *Analysis of Y-chromosomal SNP haplogroups and STR haplotypes in an Algerian population sample.* Int J Legal Med, 2008. **122**(3): p. 251-5.

512. Al-Zahery, N., et al., *Y-chromosome and mtDNA polymorphisms in Iraq, a crossroad of the early human dispersal and of post-Neolithic migrations.* Mol Phylogenet Evol, 2003. **28**(3): p. 458-72.

513. Cadenas, A.M., et al., *Y-chromosome diversity characterizes the Gulf of Oman.* Eur J Hum Genet, 2008. **16**(3): p. 374-86.

514. Abu-Amero, K.K., et al., *Saudi Arabian Y-Chromosome diversity and its relationship with nearby regions.* BMC Genet, 2009. **10**(1): p. 59.

515. Zalloua, P.A., et al., *Y-chromosomal diversity in Lebanon is structured by recent historical events.* Am J Hum Genet, 2008. **82**(4): p. 873-82.

516. Flores, C., et al., *Isolates in a corridor of migrations: a high-resolution analysis of Y-chromosome variation in Jordan.* J Hum Genet, 2005. **50**(9): p. 435-41.

517. Zalloua, P.A., et al., *Identifying genetic traces of historical expansions: Phoenician footprints in the Mediterranean.* Am J Hum Genet, 2008. **83**(5): p. 633-42.

518. El-Sibai, M., et al., *Geographical Structure of the Y-chromosomal Genetic Landscape of the Levant: A coastal-inland contrast.* Ann Hum Genet, 2009.

519. Hammer, M.F., et al., *Extended Y chromosome haplotypes resolve multiple and unique lineages of the Jewish priesthood.* Hum Genet, 2009.

520. Regueiro, M., et al., *Iran: tricontinental nexus for Y-chromosome driven migration.* Hum Hered, 2006. **61**(3): p. 132-43.

521. Cinnioglu, C., et al., *Excavating Y-chromosome haplotype strata in Anatolia.* Hum Genet, 2004. **114**(2): p. 127-48.

522. Battaglia, V., et al., *Y-chromosomal evidence of the cultural diffusion of agriculture in Southeast Europe.* Eur J Hum Genet, 2009. **17**(6): p. 820-30.

523. Di Gaetano, C., et al., *Differential Greek and northern African migrations to Sicily are supported by genetic evidence from the Y chromosome.* Eur J Hum Genet, 2009. **17**(1): p. 91-9.

524. Adams, S.M., et al., *The genetic legacy of religious diversity and intolerance: paternal lineages of Christians, Jews, and Muslims in the Iberian Peninsula.* Am J Hum Genet, 2008. **83**(6): p. 725-36.

525. Beleza, S., et al., *Micro-phylogeographic and demographic history of Portuguese male lineages.* Ann Hum Genet, 2006. **70**(Pt 2): p. 181-94.

526. Poloni, E.S., et al., *Genetic Evidence for Complexity in Ethnic Differentiation and History in East Africa.* Ann Hum Genet, 2009.

527. Cerny, V., et al., *A bidirectional corridor in the Sahel-Sudan belt and the distinctive features of the Chad Basin populations: a history revealed by the mitochondrial DNA genome.* Ann Hum Genet, 2007. **71**(Pt 4): p. 433-52.

528. Cerny, V., et al., *MtDNA of Fulani nomads and their genetic relationships to neighboring sedentary populations.* Hum Biol, 2006. **78**(1): p. 9-27.

529. Mateu, E., et al., *A tale of two islands: population history and mitochondrial DNA sequence variation of Bioko and Sao Tome, Gulf of Guinea*. Ann Hum Genet, 1997. **61**(Pt 6): p. 507-18.

530. Trovoada, M.J., et al., *Pattern of mtDNA variation in three populations from Sao Tome e Principe.* Ann Hum Genet, 2004. **68**(Pt 1): p. 40-54.

531. Rando, J.C., et al., *Phylogeographic patterns of mtDNA reflecting the colonization of the Canary Islands.* Ann Hum Genet, 1999. **63**(Pt 5): p. 413-28.

532. Hurles, M.E., et al., *The dual origin of the Malagasy in Island Southeast Asia and East Africa: evidence from maternal and paternal lineages.* Am J Hum Genet, 2005. **76**(5): p. 894-901.

533. Tofanelli, S., et al., *On the origins and admixture of Malagasy: new evidence from high resolution analyses of paternal and maternal lineages.* Mol Biol Evol, 2009.

534. Ottoni, C., et al., *First genetic insight into Libyan Tuaregs: a maternal perspective.* Ann Hum Genet, 2009. **73**(Pt 4): p. 438-48.

535. Fadhlaoui-Zid, K., et al., *Mitochondrial DNA heterogeneity in Tunisian Berbers.* Ann Hum Genet, 2004. **68**(Pt 3): p. 222-33.

536. Harich, N., et al., *The trans-Saharan slave trade - clues from interpolation analyses and high-resolution characterization of mitochondrial DNA lineages.* BMC Evol Biol. **10**(1): p. 138.

537. Brakez, Z., et al., *Human mitochondrial DNA sequence variation in the Moroccan population of the Souss area.* Ann Hum Biol, 2001. **28**(3): p. 295-307.

538. Richards, M.B., et al., *Phylogeography of mitochondrial DNA in western Europe.* Ann Hum Genet, 1998. **62**(Pt 3): p. 241-60.

539. Torroni, A., et al., *mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe.* Am J Hum Genet, 1998. **62**(5): p. 1137-52.

540. Bini, C., et al., *Different informativeness of the three hypervariable mitochondrial DNA regions in the population of Bologna (Italy).* Forensic Sci Int, 2003. **135**(1): p. 48-52.

541. Turchi, C., et al., *Italian mitochondrial DNA database: results of a collaborative exercise and proficiency testing.* Int J Legal Med, 2008. **122**(3): p. 199-204.

542. Nei, M., *Molecular evolutionary genetics*. 1987, New York: Columbia University Press.

543. Horai, S., et al., *Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs.* Proc Natl Acad Sci U S A, 1995. **92**(2): p. 532-6.

544. McDougall, I., F.H. Brown, and J.G. Fleagle, *Stratigraphic placement and age of modern humans from Kibish, Ethiopia.* Nature, 2005. **433**(7027): p. 733-6.

545. Ehret, C., *Linguistic Evidence and Its Correlation with Archaeology.* World Archaeology, 1976. **8**(1): p. 5-18.

546. Sahlins, M.D., *The Segmentary Lineage: An Organization of Predatory Expansion.* American Anthropologist, 1961. **63**(2): p. 322-345.

547. Newcomer, P.J., *The Nuer Are Dinka: An Essay on Origins and Environmental Determinism.* Man, 1972. **7**(1): p. 5-11.

548. Premo, L.S. and J.J. Hublin, *Culture, population structure, and low genetic diversity in Pleistocene hominins.* Proc Natl Acad Sci U S A, 2009. **106**(1): p. 33-7.

549. Barbujani, G., R.R. Sokal, and N.L. Oden, *Indo-European origins: a computer-simulation test of five hypotheses.* Am J Phys Anthropol, 1995. **96**(2): p. 109-32.

550. Ammerman, A.J., and L. L. Cavalli-Sforza, *Measuring the rate of spread of early farming in Europe.* Man, 1971. **6**(4): p. 674-688.

551. Ammerman, A.J., and L. L. Cavalli-Sforza, *A population model for the diffusion of early farming in Europe*, in *The Explanation of Culture Change*, C. Renfrew, Editor. 1973, Duckworth: London. p. 343-357.

552. Ammerman, A.J. and L.L. Cavalli-Sforza, *The neolithic transition and the genetics of populations in Europe*. 1984, Princeton, N.J.: Princeton University Press. xv, 176 p.

553. Sokal, R.R., N.L. Oden, and C. Wilson, *Genetic evidence for the spread of agriculture in Europe by demic diffusion.* Nature, 1991. **351**(6322): p. 143-5.

554. Novembre, J. and M. Stephens, *Interpreting principal component analyses of spatial population genetic variation.* Nat Genet, 2008. **40**(5): p. 646-9.

555. Reich, D., A.L. Price, and N. Patterson, *Principal component analysis of genetic data.* Nat Genet, 2008. **40**(5): p. 491-2.

556. Lasker, G., and Mascie-Taylor, C., *The framework of migration studies*, in *Biological Aspects of Human Migration*, C. Mascie Taylor, and Lasker, G. , Editor. 1988, Cambridge University Press: Cambridge, UK. p. 1-13.

557. Dingle, H., *Migration : the biology of life on the move*. 1996, New York: Oxford University Press. vi, 474 p.

558. Edmonds, C.A., A.S. Lillie, and L.L. Cavalli-Sforza, *Mutations arising in the wave front of an expanding population.* Proc Natl Acad Sci U S A, 2004. **101**(4): p. 975-9.

559. Klopfstein, S., M. Currat, and L. Excoffier, *The fate of mutations surfing on the wave of a range expansion.* Mol Biol Evol, 2006. **23**(3): p. 482-90.

560. Excoffier, L. and N. Ray, *Surfing during population expansions promotes genetic revolutions and structuration.* Trends Ecol Evol, 2008. **23**(7): p. 347-51.

561. Capelli, C., et al., *Population structure in the Mediterranean basin: a Y chromosome perspective.* Ann Hum Genet, 2006. **70**(Pt 2): p. 207-25.

562. Francalacci, P., et al., *Peopling of three Mediterranean islands (Corsica, Sardinia, and Sicily) inferred by Y-chromosome biallelic variability.* Am J Phys Anthropol, 2003. **121**(3): p. 270-9.

563. Sims, L.M., D. Garvey, and J. Ballantyne, *Sub-populations within the major European and African derived haplogroups R1b3 and E3a are differentiated by previously phylogenetically undefined Y-SNPs.* Hum Mutat, 2007. **28**(1): p. 97.

564. Bosch, E., et al., *High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between northwestern Africa and the Iberian Peninsula.* Am J Hum Genet, 2001. **68**(4): p. 1019-29.

565. Hammer, M.F., et al., *Jewish and Middle Eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes.* Proc Natl Acad Sci U S A, 2000. **97**(12): p. 6769-74.

566. Quirin, J., *Caste and Class in Historical North-West Ethiopia: The Beta Israel (Falasha) and Kemant, 1300-1900.* The Journal of African History, 1998. **39**(2): p. 195-220.

567. Tamrat, T., *Process of Ethnic Interaction and Integration in Ethiopian History: The Case of the Agaw.* The Journal of African History, 1988. **29**(1): p. 5-18.

568. Shelemay, K.K., *"Historical Ethnomusicology": Reconstructing Falasha Liturgical History.* Ethnomusicology, 1980. **24**(2): p. 233-258.

569. Fleming, H.C., *Ethiopic Language History: Testing Linguistic Hypotheses in an Archaeological and Documentary Context.* Ethnohistory, 1968. **15**(4): p. 353-388.

570. Varzari, A., et al., *Searching for the origin of Gagauzes: inferences from Y-chromosome analysis.* Am J Hum Biol, 2009. **21**(3): p. 326-36.

571. Di Giacomo, F., et al., *Y chromosomal haplogroup J as a signature of the post-neolithic colonization of Europe.* Hum Genet, 2004. **115**(5): p. 357-71.

572. Karafet, T.M., et al., *Balinese Y-chromosome perspective on the peopling of Indonesia: genetic contributions from pre-neolithic hunter-gatherers, Austronesian farmers, and Indian traders.* Hum Biol, 2005. **77**(1): p. 93-114.

573. Cruciani, F., et al., *Human Y chromosome haplogroup R-V88: a paternal genetic record of early mid Holocene trans-Saharan connections and the spread of Chadic languages.* Eur J Hum Genet.

574. Kivisild, T., et al., *The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations.* Am J Hum Genet, 2003. **72**(2): p. 313-32.

575. Underhill, P.A., et al., *Maori origins, Y-chromosome haplotypes and implications for human history in the Pacific.* Hum Mutat, 2001. **17**(4): p. 271-80.

576. Underhill, P.A., et al., *Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a.* Eur J Hum Genet, 2009.

577. Lamphear, J., *The people of the Grey Bull: The origin and expansion of the Turkana.* Journal of African History, 1988. **29**: p. 27-39.

578. Herbert, R.K., *The Sociohistory of Clicks in Southern Bantu.* Anthropological Linguistics, 1990. **32**(3/4): p. 295-315.

579. Nurse, G.T., J.S. Weiner, and T. Jenkins, *The peoples of southern Africa and their affinities*. Research monographs on human population biology no. 3. 1985, Oxford [Oxfordshire]

New York: Clarendon Press ;

Oxford. xvi, 409 p.

580. Diegoli, T.M., Irwin, J. A., Just, R., Saunier, J. L., O'Callaghan, J. E., Parsons, T. J. , *Mitochondrial control region sequences from an African American population sample.* Journal of Forensic Science Genetics, 2009. **48**(2003): p. 908-911.

581. Monsalve, M.V. and E. Hagelberg, *Mitochondrial DNA polymorphisms in Carib people of Belize.* Proc Biol Sci, 1997. **264**(1385): p. 1217-24.

582. Batini, C., et al., *Phylogeography of the human mitochondrial L1c haplogroup: genetic signatures of the prehistory of Central Africa*. Mol Phylogenet Evol, 2007. **43**(2): p. 635-44.

583. Levanoni, A., *The Mamluks' Ascent to Power in Egypt.* Studia Islamica, 1990(72): p. 121-144.

584. Lev, Y., *Army, Regime, and Society in Fatimid Egypt, 358-487/968-1094.* International Journal of Middle East Studies, 1987. **19**(3): p. 337-365.

585. Walker, B.J., *Militarization to Nomadization: The Middle and Late Islamic Periods.* Near Eastern Archaeology, 1999. **62**(4): p. 202-232.

586. Peer, P.V., et al., *The Nile Corridor and the Out-of-Africa Model: An Examination of the Archaeological Record [and Comments and Reply]*. Current Anthropology, 1998. **39**(2): p. S115-S140.

587. Cremaschi, M., S.D. Lernia, and E.A.A. Garcea, *Some Insights on the Aterian in the Libyan Sahara: Chronology, Environment, and Archaeology.* The African Archaeological Review, 1998. **15**(4): p. 261-286.

588. Garcea, E.A.A., *Crossing Deserts and Avoiding Seas: Aterian North African-European Relations.* Journal of Anthropological Research, 2004. **60**(1): p. 27-53.

589. Wendorf, F., A.E. Close, and R. Schild, *Recent Work on the Middle Palaeolithic of the Eastern Sahara.* The African Archaeological Review, 1987. **5**: p. 49-63.

590. Tillet, T., *The Palaeolithic and Its Environment in the Northern Part of the Chad Basin.* The African Archaeological Review, 1985. **3**: p. 163-177.

591. Hublin, J.-J., *Recent Human Evolution in Northwestern Africa.* Philosophical Transactions: Biological Sciences, 1992. **337**(1280): p. 185-191.

592. Castaneda, I.S., et al., *Wet phases in the Sahara/Sahel region and human migration patterns in North Africa.* Proc Natl Acad Sci U S A, 2009.

593. Ottoni, C., et al., *Human mitochondrial DNA variation in Southern Italy.* Ann Hum Biol, 2009. **36**(6): p. 785-811.

594. Benn Torres, J., R.A. Kittles, and A.C. Stone, *Mitochondrial and Y chromosome diversity in the English-speaking Caribbean.* Ann Hum Genet, 2007. **71**(Pt 6): p. 782-90.

595. Salas, A., et al., *The mtDNA ancestry of admixed Colombian populations.* Am J Hum Biol, 2008. **20**(5): p. 584-91.

596. Thomas, M.G., et al., *Founding mothers of Jewish communities: geographically separated Jewish groups were independently founded by very few female ancestors.* Am J Hum Genet, 2002. **70**(6): p. 1411-20.

597. Cerny, V., et al., *Migration of Chadic speaking pastoralists within Africa based on population structure of Chad Basin and phylogeography of mitochondrial L3f haplogroup.* BMC Evol Biol, 2009. **9**(1): p. 63.

598. Forster, P., *Ice Ages and the mitochondrial DNA chronology of human dispersals: a review.* Philos Trans R Soc Lond B Biol Sci, 2004. **359**(1442): p. 255-64; discussion 264.

599. Maca-Meyer, N., et al., *Major genomic mitochondrial lineages delineate early human expansions.* BMC Genet, 2001. **2**: p. 13.

600. Rajkumar, R., et al., *Phylogeny and antiquity of M macrohaplogroup inferred from complete mt DNA sequence of Indian specific lineages.* BMC Evol Biol, 2005. **5**(1): p. 26.

601. Larruga, J.M., et al., *Mitochondrial DNA characterisation of European isolates: the Maragatos from Spain.* Eur J Hum Genet, 2001. **9**(9): p. 708-16.

602. Gonzalez, A.M., et al., *Mitochondrial DNA affinities at the Atlantic fringe of Europe.* Am J Phys Anthropol, 2003. **120**(4): p. 391-404.

603. Maca-Meyer, N., et al., *Mitochondrial DNA transit between West Asia and North Africa inferred from U6 phylogeography.* BMC Genet, 2003. **4**: p. 15.

604. Crespillo, M., et al., *Mitochondrial DNA sequences for 118 individuals from northeastern Spain.* Int J Legal Med, 2000. **114**(1-2): p. 130-2.

605. Salas, A., et al., *mtDNA analysis of the Galician population: a genetic edge of European variation.* Eur J Hum Genet, 1998. **6**(4): p. 365-75.

606. Francalacci, P., et al., *Sequence diversity of the control region of mitochondrial DNA in Tuscany and its implications for the peopling of Europe.* Am J Phys Anthropol, 1996. **100**(4): p. 443-60.

607. Shevchuk, N.A. and M.W. Allard, *Sources of incongruence among mammalian mitochondrial sequences: COII, COIII, and ND6 genes are main contributors.* Mol Phylogenet Evol, 2001. **21**(1): p. 43-54.

608. Stewart, J.B., et al., *Strong purifying selection in transmission of mammalian mitochondrial DNA.* PLoS Biol, 2008. **6**(1): p. e10.

609. Subramanian, S., *Temporal trails of natural selection in human mitogenomes.* Mol Biol Evol, 2009. **26**(4): p. 715-7.

610. Dowling, D.K., U. Friberg, and J. Lindell, *Evolutionary implications of non-neutral mitochondrial genetic variation.* Trends Ecol Evol, 2008. **23**(10): p. 546-54.

611. Ho, S.Y. and P. Endicott, *The crucial role of calibration in molecular date estimates for the peopling of the Americas.* Am J Hum Genet, 2008. **83**(1): p. 142-6; author reply 146-7.

612. Tiranti, V., et al., *A novel frameshift mutation of the mtDNA COIII gene leads to impaired assembly of cytochrome c oxidase in a patient affected by Leigh-like syndrome.* Hum Mol Genet, 2000. **9**(18): p. 2733-42.

613. Petros, J.A., et al., *mtDNA mutations increase tumorigenicity in prostate cancer.* Proc Natl Acad Sci U S A, 2005. **102**(3): p. 719-24.

614. Tessa, E.P., D. Cassandrini, T. Rizza, I. Wittig, F. Fattori, M. Meschini, E. Bertini, L. Santarelli, C. Bruno, *A novel mtDNA mutation in COIII impairs assembly of cytochrome c oxidase in a MELAS patient Neuromuscular Disorders* 2007. **17**(9): p. 768-769

615. Abu-Amero, K.K. and T.M. Bosley, *Mitochondrial abnormalities in patients with LHON-like optic neuropathies.* Invest Ophthalmol Vis Sci, 2006. **47**(10): p. 4211-20.

616. Abu-Amero, K.K., T.M. Bosley, and J. Morales, *Analysis of nuclear and mitochondrial genes in patients with pseudoexfoliation glaucoma.* Mol Vis, 2008. **14**: p. 29-36.

617. Bosley, T.M., et al., *Mitochondrial changes in leukocytes of patients with optic neuritis.* Mol Vis, 2007. **13**: p. 1516-28.

618. Johns, D.R. and M.J. Neufeld, *Cytochrome c oxidase mutations in Leber hereditary optic neuropathy.* Biochem Biophys Res Commun, 1993. **196**(2): p. 810-5.

619. Johns, D.R., M.J. Neufeld, and T.R. Hedges, 3rd, *Mitochondrial DNA mutations in Cuban optic and peripheral neuropathy.* J Neuroophthalmol, 1994. **14**(3): p. 135-40.

620. Cai, W., et al., *Mitochondrial variants may influence the phenotypic manifestation of Leber's hereditary optic neuropathy-associated ND4 G11778A mutation.* J Genet Genomics, 2008. **35**(11): p. 649-55.

621. Jaksch, M., et al., *A systematic mutation screen of 10 nuclear and 25 mitochondrial candidate genes in 21 patients with cytochrome c oxidase (COX) deficiency shows tRNA(Ser)(UCN) mutations in a subgroup with syndromal encephalopathy.* J Med Genet, 1998. **35**(11): p. 895-900.

622. Oostra, R.J., et al., *Leber's hereditary optic neuropathy: correlations between mitochondrial genotype and visual outcome.* J Med Genet, 1994. **31**(4): p. 280-6.

623. Ingman, M. and U. Gyllensten, *Rate variation between mitochondrial domains and adaptive evolution in humans.* Hum Mol Genet, 2007. **16**(19): p. 2281-7.

624. Howell, N., C. Howell, and J.L. Elson, *Time dependency of molecular rate estimates for mtDNA: this is not the time for wishful thinking.* Heredity, 2008. **101**(2): p. 107-8.

625. Laval, G., et al., *Formulating a historical and demographic model of recent human evolution based on resequencing data from noncoding regions*. PLoS ONE. **5**(4): p. e10284.

626. Ruiz-Pesini, E., et al., *Effects of purifying and adaptive selection on regional variation in human mtDNA.* Science, 2004. **303**(5655): p. 223-6.

627. Templeton, A.R., *Contingency tests of neutrality using intra/interspecific gene trees: the rejection of neutrality for the evolution of the mitochondrial cytochrome oxidase II gene in the hominoid primates.* Genetics, 1996. **144**(3): p. 1263-70.

628. Abu-Amero, K.K., et al., *The role of mitochondrial haplogroups in glaucoma: a study in an Arab population.* Mol Vis, 2008. **14**: p. 518-22.

629. Saitou, N. and M. Nei, *The neighbor-joining method: a new method for reconstructing phylogenetic trees.* Mol Biol Evol, 1987. **4**(4): p. 406-25.

630. Felsenstein, J., *Evolutionary trees from DNA sequences: a maximum likelihood approach.* J Mol Evol, 1981. **17**(6): p. 368-76.

631. Pereira, L., et al., *Linking the sub-Saharan and West Eurasian gene pools: maternal and paternal heritage of the Tuareg nomads from the African Sahel.* Eur J Hum Genet. **18**(8): p. 915-23.

632. Brandstatter, A., et al., *Timing and deciphering mitochondrial DNA macro-haplogroup R0 variability in Central Europe and Middle East.* BMC Evol Biol, 2008. **8**: p. 191.

633. Yao, Y.G., et al., *Different matrilineal contributions to genetic structure of ethnic groups in the silk road region in china.* Mol Biol Evol, 2004. **21**(12): p. 2265-80.

634. Parson, W., et al., *Population data for 101 Austrian Caucasian mitochondrial DNA d-loop sequences: application of mtDNA sequence analysis to a forensic case.* Int J Legal Med, 1998. **111**(3): p. 124-32.

635. Richards, M., et al., *Paleolithic and neolithic lineages in the European mitochondrial gene pool.* Am J Hum Genet, 1996. **59**(1): p. 185-203.

636. Belledi, M., et al., *Maternal and paternal lineages in Albania and the genetic structure of Indo-European populations.* Eur J Hum Genet, 2000. **8**(7): p. 480-6.

637. Hofmann, S., et al., *Analysis of the mitochondrial DNA from patients with Wolfram (DIDMOAD) syndrome.* Mol Cell Biochem, 1997. **174**(1-2): p. 209-13.

638. Tetzlaff, S., et al., *Mitochondrial DNA population data of HVS-I and HVS-II sequences from a northeast German sample.* Forensic Sci Int, 2007. **172**(2-3): p. 218-24.

639. Piercy, R., et al., *The application of mitochondrial DNA typing to the study of white Caucasian genetic identification.* Int J Legal Med, 1993. **106**(2): p. 85-90.

640. Helgason, A., et al., *Estimating Scandinavian and Gaelic ancestry in the male settlers of Iceland.* Am J Hum Genet, 2000. **67**(3): p. 697-717.

641. Sajantila, A., et al., *Genes and languages in Europe: an analysis of mitochondrial lineages.* Genome Res, 1995. **5**(1): p. 42-52.

642. Sajantila, A., et al., *Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population.* Proc Natl Acad Sci U S A, 1996. **93**(21): p. 12035-9.

643. Hedman, M., et al., *Finnish mitochondrial DNA HVS-I and HVS-II population data.* Forensic Sci Int, 2007. **172**(2-3): p. 171-8.

644. Grzybowski, T., et al., *Complex interactions of the Eastern and Western Slavic populations with other European groups as revealed by mitochondrial DNA analysis.* Forensic Sci Int Genet, 2007. **1**(2): p. 141-7.

645. Lappalainen, T., et al., *Migration waves to the Baltic Sea region.* Ann Hum Genet, 2008. **72**(Pt 3): p. 337-48.

646. Pliss, L., et al., *Mitochondrial DNA portrait of Latvians: towards the understanding of the genetic structure of Baltic-speaking populations.* Ann Hum Genet, 2006. **70**(Pt 4): p. 439-58.

647. Calafell, F., et al., *Population history of Corsica: a linguistic and genetic analysis.* Ann Hum Biol, 1996. **23**(3): p. 237-51.

648. Irwin, J., et al., *Hungarian mtDNA population databases from Budapest and the Baranya county Roma.* Int J Legal Med, 2007. **121**(5): p. 377-83.

649. Brandstatter, A., et al., *Migration rates and genetic structure of two Hungarian ethnic groups in Transylvania, Romania.* Ann Hum Genet, 2007. **71**(Pt 6): p. 791-803.

650. Opdal, S.H., et al., *Increased number of substitutions in the D-loop of mitochondrial DNA in the sudden infant death syndrome.* Acta Paediatr, 1998. **87**(10): p. 1039-44.

651. Zimmermann, B., et al., *Mitochondrial DNA control region population data from Macedonia.* Forensic Sci Int Genet, 2007. **1**(3-4): p. e4-9.

652. Lehocky, I., et al., *A database of mitochondrial DNA hypervariable regions I and II sequences of individuals from Slovakia.* Forensic Sci Int Genet, 2008. **2**(4): p. e53-9.

653. Mergen, H., Oner, R. and Oner, C., *Mitochondrial DNA sequence variation in the Anatolian Peninsula (Turkey).* Journal of Genetics, 2004. **83**(1): p. 39-47.

654. Comas, D., et al., *Geographic variation in human mitochondrial DNA control region sequence: the population history of Turkey and its relationship to the European populations.* Mol Biol Evol, 1996. **13**(8): p. 1067-77.

655. Calafell, F., et al., *From Asia to Europe: mitochondrial DNA sequence variability in Bulgarians and Turks.* Ann Hum Genet, 1996. **60**(Pt 1): p. 35-49.

656. Irwin, J., et al., *Mitochondrial control region sequences from northern Greece and Greek Cypriots.* Int J Legal Med, 2008. **122**(1): p. 87-9.

657. Bertranpetit, J., et al., *Human mitochondrial DNA variation and the origin of Basques.* Ann Hum Genet, 1995. **59**(Pt 1): p. 63-81.

658. Hofmann, S., et al., *Population genetics and disease susceptibility: characterization of central European haplogroups by mtDNA gene mutations, correlation with D loop variants and association with disease.* Hum Mol Genet, 1997. **6**(11): p. 1835-46.

659. Vanecek, T., F. Vorel, and M. Sip, *Mitochondrial DNA D-loop hypervariable regions: Czech population data.* Int J Legal Med, 2004. **118**(1): p. 14-8.

660. Helgason, A., et al., *mtDna and the islands of the North Atlantic: estimating the proportions of Norse and Gaelic ancestry.* Am J Hum Genet, 2001. **68**(3): p. 723-37.

661. Malmstrom, H., et al., *Ancient DNA Reveals Lack of Continuity between Neolithic Hunter-Gatherers and Contemporary Scandinavians.* Curr Biol, 2009.

662. Rogers, A.R. and H. Harpending, *Population growth makes waves in the distribution of pairwise genetic differences.* Mol Biol Evol, 1992. **9**(3): p. 552-69.

663. Harpending, H.C., et al., *The Genetic Structure of Ancient Human Populations.* Current Anthropology, 1993. **34**(4): p. 483-496.

664. Cox, M.P., et al., *Autosomal resequence data reveal Late Stone Age signals of population expansion in sub-Saharan African foraging and farming populations.* PLoS ONE, 2009. **4**(7): p. e6366.

665. Fagundes, N.J., R. Kanitz, and S.L. Bonatto, *A reevaluation of the Native American mtDNA genome diversity and its bearing on the models of early colonization of Beringia.* PLoS ONE, 2008. **3**(9): p. e3157.

666. Amos, W. and J.I. Hoffman, *Evidence that two main bottleneck events shaped modern human genetic diversity.* Proc Biol Sci, 2009.

667. Pilkington, M.M., et al., *Contrasting signatures of population growth for mitochondrial DNA and Y chromosomes among human populations in Africa.* Mol Biol Evol, 2008. **25**(3): p. 517-25.

668. Behrensmeyer, A.K., *Atmosphere. Climate change and human evolution.* Science, 2006. **311**(5760): p. 476-8.

669. Gasse, F., Chalie, F., Vincens, A., Williams, M. A.J., Williamson, D., *Climatic patterns in equatorial and southern Africa from 30,000 to 10,000 years ago reconstructed from terrestrial and near-shore proxy data* Quaternary Science Reviews, 2008. **27**: p. 2316-2340

670. Kiage, L.M.a.L., K., *Late Quaternary paleoenvironmental changes in East Africa: a review of multiproxy evidence from palynology, lake sediments, and associated records* Progress in Physical Geography, 2006. **30**: p. 633.

671. Laskar, J., Robutel, P., Joutel, F., Gastineau, M., Correia, A.C.M., Levrard, B., *A long-term numerical solution for the insolation quantities of the Earth.* Astronomy and Astrophysics 2004. **428**: p. 261-285.

672. Jousse, H., *What is the impact of Holocene climatic changes on human societies? Analysis of West African Neolithic populations dietary customs.* Quaternary International, 2006. **151**: p. 63-73.

673. Schefuss, E., S. Schouten, and R.R. Schneider, *Climatic controls on central African hydrology during the past 20,000 years.* Nature, 2005. **437**(7061): p. 1003-6.

674. Castaneda, I.S., Werne, J., and Johnson, T.C., *Wet and arid phases in the southeeast African tropics since the Last Glacial Maximum.* Geology, 2007. **35**: p. 823-826.

675. Shackelford, L.L., *Regional variation in the postcranial robusticity of late Upper Paleolithic humans.* Am J Phys Anthropol, 2007. **133**(1): p. 655-68.

676. Berger, A.a.L., M.F., *Insolation values for the climate of the last 10 million years.* Quaternary Sciences Review, 1991. **10**(4): p. 297-317.

677. Vaks, A., Bar-Matthews, M., Ayalon, A., Matthews, A., Halicz, L. and Frumkin, A. , *Desert speleothems reveal climatic window for African exodus of early modern humans.* Geology, 2007. **35**(9): p. 831-834.

678. Kuper, R. and S. Kropelin, *Climate-controlled Holocene occupation in the Sahara: motor of Africa's evolution.* Science, 2006. **313**(5788): p. 803-7.

679. Martinson, D.G., Pisias,N.G. Hays, J.D. Imbrie, J., Moore, T.C., Jr., and Shackleton, N.J., *Age dating and the orbital theory of the Ice Ages: development of a high-resolution 0 - 300,000 year chronostratigraphy.* Quaternary Research, 1987. **27**: p. 1-29.

680. Gunz, P., et al., *Early modern human diversity suggests subdivided population structure and a complex out-of-Africa scenario.* Proc Natl Acad Sci U S A, 2009.

681. Falush, D., et al., *Traces of human migrations in Helicobacter pylori populations.* Science, 2003. **299**(5612): p. 1582-5.

682. Linz, B., et al., *An African origin for the intimate association between humans and Helicobacter pylori.* Nature, 2007. **445**(7130): p. 915-8.

683. Zeitoun, V., *The taxonomical position of the skull of Zuttiyeh*, in *Comptes Rendus de l'Académie des Sciences, série IIa: Sciences de la Terre et des Planètes* 2001. p. 521-525.

684. Vandermeersch, B., *The evolution of modern humans: Recent evidence from Southwest Asia*, in *The human revolution: Behavioural and biological perspectives on the origins of modern humans*, P.M.a.C. Stringer, Editor. 1989, Princeton University Press: Princeton. p. 155-164.

685. Mellars, P.A., *Why did modern human populations disperse from Africa ca. 60,000 years ago? A new model.* Proceedings of the National Academy of Sciences (USA) 2006. **103** p. 9381-9386.

686. Shea, J.J., *Transitions or turnovers? Climatically-forced extinctions of Homo sapiens and Neanderthals in the east Mediterranean Levant.* Quaternary Science Reviews, 2008. **27**: p. 2253-2270.

687. Vermeersch, P.M., Gijselings, G. and Paulissen, E., *Discovery of Nazlet Khater man, Upper Egypt.* Journal of Human Evolution, 1984. **13**: p. 281-286.

688. Vermeersch, P.M., Paulissen, E., Stokes, S., Charlier, C., van Peer, P., Stringer, C. B. and Lindsay, W. , *A middle Palaeolithic burial of a modern human at Taramsa Hill, Egypt.* Antiquity, 1998. **72**: p. 475-482.

689. Schild, R.a.W., F. , *The late Pleistoicene Nile in Wadi Kubbaniya*, in *Prehistory of Wadi Kubbaniya* R.S.a.A.E.C. F. Wendorf, Editor. 1989, Southern Methodist University Press Dallas. p. 15-100.

690. Stokes, S.a.B., R. , *OSL-dating of Nazlet Safaha and Nazlet Khater 4*, in *Palaeolithic quarrying sites in Upper and Middle Egypt*, P.M. Vermeersch, Editor. 2002, Leuven University Press: Leuven. p. 349-350.

691. Marks, A.E., *The Middle to upper Palaeolithic transitions in the Levant.* Advances in World Archaelogy, 1983. **2** p. 51-98.

692. Bar-Yosef, O., *The Upper Paleolithic Revolution* Annual Review of Anthropology, 2002. **31**: p. 363-393.

693. Mellars, P.A., *The impossible coincidence. A single-species model for the origins of modern human behavior.* Evolutionary anthropology, 2005. **14**: p. 12-27.

694. Finlayson, C., *Biogeography and evolution of the genus Homo.* Trends Ecol Evol, 2005. **20**(8): p. 457-63.

695. Wild, E.M., et al., *Direct dating of Early Upper Palaeolithic human remains from Mladec.* Nature, 2005. **435**(7040): p. 332-5.

696. Holliday, T.W., *Body proportions in Late Pleistocene in Europe and Modern Human Origins.* Journal of Human of Human Evolution, 1997. **32** p. 423-447.

697. Finlayson, C. and J.S. Carrion, *Rapid ecological turnover and its impact on Neanderthal and other human populations.* Trends Ecol Evol, 2007. **22**(4): p. 213-22.

698. Klein, R.G., *Out of Africa and the evolution of human behavior.* Evolutionary Anthropology, 2008. **17**(6): p. 267-281.

699. Hodgson, J.A., C.M. Bergey, and T.R. Disotell, *Neandertal genome: the ins and outs of African genetic diversity.* Curr Biol. **20**(12): p. R517-9.

700. Thangaraj, K., et al., *Deep rooting in-situ expansion of mtDNA Haplogroup R8 in South Asia.* PLoS ONE, 2009. **4**(8): p. e6545.

701. Eaaswarkhanth, M., et al., *Diverse genetic origin of Indian Muslims: evidence from autosomal STR loci.* J Hum Genet, 2009. **54**(6): p. 340-8.

702. Beyin, A., *The Bab al Mandab vs the Nile-Levant: An Appraisal of the Two Dispersal Routes for Early Modern Humans Out of Africa.* African Archaeological Reviews, 2006. **23**: p. 5-30.

703. Donnelly, M.P., et al., *The distribution and most recent common ancestor of the 17q21 inversion in humans.* Am J Hum Genet. **86**(2): p. 161-71.

704. Harvey, C.P.D. and A.T. Grove, *A Prehistoric Source of the Nile.* The Geographical Journal, 1982. **148**(3): p. 327-336.

705. Epstein, H., Mason, I. L. , *Cattle*, in *Evolution of domesticated animals* I.L. Mason, Editor. 1984, Longman: London and New York. p. 6-27.

706. Magnavita, C., *Ancient Humped Cattle in Africa: A View from the Chad Basin.* African Archaeological Reviews, 2006. **23**: p. 55-84.

707. Bousman, C.B., *The Chronological Evidence for the Introduction of Domestic Stock into Southern Africa.* The African Archaeological Review, 1998. **15**(2): p. 133-150.

708. Zeder, M.A., *Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact.* Proc Natl Acad Sci U S A, 2008. **105**(33): p. 11597-604.

709. Gautier, A., *archaeozology of the Bir Kiseiba Region, Eastern Sahara*, in *Cattle keepers of the eastern Sahara*, F. Wendorf, Schild, R. and Close, A., Editor. 1984, Department of Anthropology, Southern Methodist University: Dallas. p. 49-72.

710. Gautier, A., *Prehistoric men and cattle in North Africa: A dearth of data and a surfeit of Models*, in *Prehistory of Arid North Africa*, A. Close, Editor. 1987, Southern Methodist University Press: Dallas. p. 163-88.

711. Reed, C.A., *Animal Domestication in the Prehistoric Near East.* Science, 1959. **130**(3389): p. 1629-1639.

712. Sereno, P.C., et al., *Lakeside cemeteries in the Sahara: 5000 years of holocene population and environmental change.* PLoS ONE, 2008. **3**(8): p. e2995.

713. Robertshaw, P., *Archaeology in Eastern Africa: Recent Developments and More Dates.* The Journal of African History, 1984. **25**(4): p. 369-393.

714. Wright, D., *New Perspectives on Early Regional Interaction Networks of East African Trade: A View from Tsavo National Park, Kenya.* African Archaeological Review, 2005. **22**(3): p. 111-140.

715. Denbow, J., *A New Look at the Later Prehistory of the Kalahari.* The Journal of African History, 1986. **27**(1): p. 3-28.

716. Robbins, L.H., et al., *The Advent of Herding in Southern Africa: Early AMS Dates on Domestic Livestock from the Kalahari Desert.* Current Anthropology, 2005. **46**(4): p. 671-677.

717. Walker, N.J., *The Significance of an Early Date for Pottery and Sheep in Zimbabwe.* The South African Archaeological Bulletin, 1983. **38**(138): p. 88-92.

718. Smith, A.B., *Prehistoric Pastoralism in the Southwestern Cape, South Africa.* World Archaeology, 1983. **15**(1): p. 79-89.

719. Phillipson, D.W., *The Chronology of the Iron Age in Bantu Africa.* The Journal of African History, 1975. **16**(3): p. 321-342.

720. Schmidt, P.R. and D.H. Avery, *More Evidence for an Advanced Prehistoric Iron Technology in Africa.* Journal of Field Archaeology, 1983. **10**(4): p. 421-434.

721. Marshall, F., Hildebrand, E., *Cattle Before Crops: The Beginnings of Food Production in Africa. Journal of World Prehistory*, 2002. **16**(2): p. 99-143.

722. Finucane, B., Manning, K., Toure, M. , *Late Stone Age subsistence in the Tilemsi Valley, Mali: Stable isotope analysis of human and animal remains from the site of Karkarichinkat Nord (KN05) and Karkarichinkat Sud (KS05).* Journal of Anthropological Archaeology, 2008. **27**: p. 82-92.