# UNIVERSITÀ DEGLI STUDI DI ROMA "TOR VERGATA"

FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI

DOTTORATO DI RICERCA IN BIOLOGIA EVOLUZIONISTICA ED ECOLOGIA

XX CICLO

## HOLOCENE HUMAN PEOPLING OF LIBYAN SAHARA
Molecular analysis of maternal lineages in ancient and extant populations of Fezzan

CLAUDIO OTTONI

Docente Guida: Prof. Olga Rickards

Coordinatore: Prof. Patrizia Albertano

**ABSTRACT (English version)**

**Holocene human peopling of Libyan Sahara - Molecular analysis of maternal lineages in ancient and extant populations of Fezzan.**

The present work provides an important view of a region of Africa that is still almost unknown: the Central Sahara. The aim of the project as a whole, was to reconstruct from the maternal side, through the genetic analysis of mitochondrial DNA (mtDNA), the origins of a Pastoral nomad population in the Libyan Sahara, the Tuareg. The availability of both modern and ancient samples from the Fezzan (Libyan Sahara), collected in collaboration with the Italian Archaeological Mission in Libya directed by Prof. Savino Di Lernia, represented an important means of relating the mtDNA pool of extant Libyan Tuareg, with that of Pastoral people inhabiting the Central Sahara in prehistoric times, and with the Garamantes, the hypothetical ancestors of Libyan Tuareg. Nevertheless, molecular analysis carried out on the bones collected from the archaeological sites of the Acacus region, showed a very low state of preservation of the DNA, this probably due to the high temperatures that characterised burials over the centuries. Failure of the genetic analyses in the ancient individuals, necessarily limited the present work to the study of the extant Tuareg sample. Nevertheless, comparison with other genetic data collected so far in the modern African populations, and moreover the multidisciplinary integration with archaeological and ethnological data, helped to hypothetically reconstruct the origins of Libyan Tuareg, and their relationship with the ancient human migratory dynamics that occurred in Northern Africa during the Holocene.

A total of 129 individuals from two villages in the Acacus region, in Fezzan, were genetically analysed at the mtDNA level. The results here reported clearly show the low level of genetic diversity in the Libyan Tuareg sample, that is hypothetically due to high endogamy. Furthermore, phylogenetic analyses indicate that the mtDNA genetic pool of the Libyan Tuareg is characterized by a major “West-Eurasian” component, that is shared with many Berber groups and hypothetically comes from the Iberian Peninsula, and a minor “South-Saharan” component that shows some kind of relationship with Central and Eastern African populations.

**ABSTRACT (Italian version)**

**Popolamento umano del Sahara libico durante l'Olocene – Analisi molecolare delle linee materne in popolazioni antiche ed attuali del Fezzan.**

Il presente lavoro offre un quadro descrittivo, dal punto di vista genetico, di una regione che è ancora piuttosto sconosciuta: il Sahara Centrale. Lo scopo del progetto è quello di ricostruire attraverso l'analisi del DNA mitocondriale (DNAmt) le origini di una popolazione pastorale Tuareg insediata nel Sahara libico. L'area presa in esame è quella del Fezzan, una regione nella Libia Sud-Occidentale. Per poter risalire ad una eventuale continuità genetica tra i Tuareg che oggi abitano quella regione, e i gruppi umani nomadi che in epoca preistorica occuparono quella zona, sono state effettuate analisi genetiche sia su campioni moderni che antichi, raccolti in collaborazione con la Missione Italiana Archeologica in Libia diretta dal Prof. Savino Di Lernia. Le analisi molecolari condotte sul materiale archeologico a disposizione, hanno evidenziato però il basso stato di conservazione del DNA, probabilmente dovuto alle alte temperature cui le ossa sono state sottoposte all'interno delle sepolture almeno negli ultimi 2,000 anni. Per quanto riguarda il campione moderno, un totale di 129 individui provenienti da due villaggi situati nella zona dell'Acacus, nel Fezzan, sono stati analizzati a livello del DNAmt. Il confronto con i dati genetici di popolazioni Africane attuali presenti in letteratura, e l'integrazione con dati archeologici ed etnologici, ha permesso di ricostruire le origini dei Tuareg in Libia, e di capire in che modo questi sono stati coinvolti nelle migrazioni umane che caratterizzarono il Nord Africa durante l'Olocene.

I risultati riportati nel presente studio evidenziano un livello molto basso di diversità genetica nel campione Tuareg Libico, probabilmente associato a pratiche endogamiche. Inoltre, le analisi filogenetiche indicano che il *pool* genetico mitocondriale dei Tuareg in Libia è caratterizzato da una componente "Eurasiatica Occidentale" predominante, che è condivisa con alcuni gruppi Berberi Nord Africani e verosimilmente proviene dalla penisola Iberica, e una componente minore di tipo "Sud-Sahariano", che evidenzia una certa affinità con le popolazioni dell'Africa Centrale e Orientale.

**Key-words:** Tuareg**,** Phylogeny, Africa, Human mtDNA, aDNA, Holocene, Garamantes, Pastoral, Central Sahara.

# Contents

# Chapter 1. Introduction

## 1.1 GEOGRAPHICAL CHARACTERIZATION

### The African continent

Africa is the second largest continent in the world and is inhabited by about the 14% of human population. Its climate is variable: moist air is concentrated around the equator (latitude 5° North and South) and it leads to frequent precipitation, especially in the plains; here the range of temperatures throughout the year is generally low. All this area is covered by the tropical pluvial forest, which in the past was wider than it appears today: traces of its previous extent can be found even in the Gulf of Guinea. The reduction in size of the rainforest probably took place in the last 5,000 years.

Proceeding north and south of the equator, moisture decreases and the vegetation changes. Forests turn into savannas, dry and wet seasons alternate during the year. The more precipitations decreases, the spiner the shrubs and trees become, like baobab and acacia. In both directions, the sub-desert environment is found, and finally the desert. The two main African deserts are both a result the transformation in vegetation described above. The first desert is the Sahara desert, which extends North of the Equator for about a quarter of the whole continent surface. The second one is the Kalahari desert, located south of the Equator in Namibia and Botswana. The Sahara desert, in particular the Libyan Sahara, is the place which this study concentrates on.

### The Sahara desert

The Sahara is the largest desert in the world. It is almost as large as the United States, and is larger than Australia. At over 9,000,000 square kilometres, it covers most parts of Northern Africa, an area stretching from the Red Sea, including parts

of the Mediterranean coasts, to the outskirts of the Atlantic ocean. It is delimited in its northern border by the Mediterranean sea, and in the southern one by the so-called Sahel-Sudan belt, that extends from the Atlantic Ocean to the Red Sea.

About 10,000 years ago, the Sahara region differed considerably from the area we know today. Climatic changes brought heavy rains driven by south-western monsoons, transforming the desert into a verdant landscape. This marked the end of the arid Pleistocene period and attracted animals and humans from southern latitudes. As observed in the graph shown in figure 1.1, the first half of the Holocene was characterized by oscillations of arid and wet periods; these conditions lasted until around 5,000 years B.P.[1], when a dramatic, abrupt dry spell set in. This major arid phase of the Middle Holocene, also known as the Mahla event (Hassan 2002), forced human groups to adapt their food security and settlement systems. Since 1,500 years B.P. the Sahara has not changed in appearance from how it is today. The *wadis*[2] that are spread throughout the Sahara, represent remains of the rivers that flowed in the ancient Saharan verdant landscape. These *wadis* over time changed into deep canyons, turning over hundreds of years into the valleys we find today.

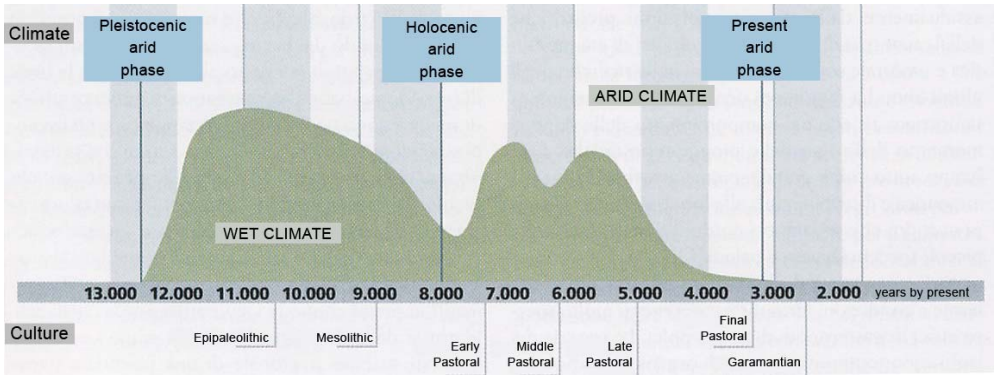


**Figure 1.1:** Climatic changes in the Central Sahara during the Holocene (Di Lernia mod.).

The climatic crisis of the second half of the Holocene is somehow attested by the reduction in size that characterized the lake Chad. Around 9,000 years ago, a large

[1] The quotation refers to uncalibrated radiocarbon years before present.
[2] *Wadi* is traditionally a valley. In some cases it can refer to a dry riverbed that contains water only during times of heavy rain.

part of the Chad basin was underwater, eventually giving rise to lake Megachad. It flowed through the Bahr el-Ghazal into the Bodélé plains of northern Chad. It is estimated that lake Megachad covered an area of 330,000 $km^2$. Today it is merely 20,000 $km^2$. Therefore the hypothesis that it is the remains of a much larger lake/sea is quite reasonable (Brunk and Gronenborn 2004, Cerny et al. 2007).

## The Fezzan: surveyed area

The Fezzan is situated in the South-Western part of Libya, in the Central Sahara (figure 1.2). It is a region wider than Southern Italy, and contains both homogenous landscapes and complex ones. Located in a remote region of Libya, at the border with Algeria and Niger, the Fezzan is characterized by the presence of two mountain massifs: the Acacus and the Messak Settafet. Both these massifs belong to a system of mountains crossing the Sahara from West to East.

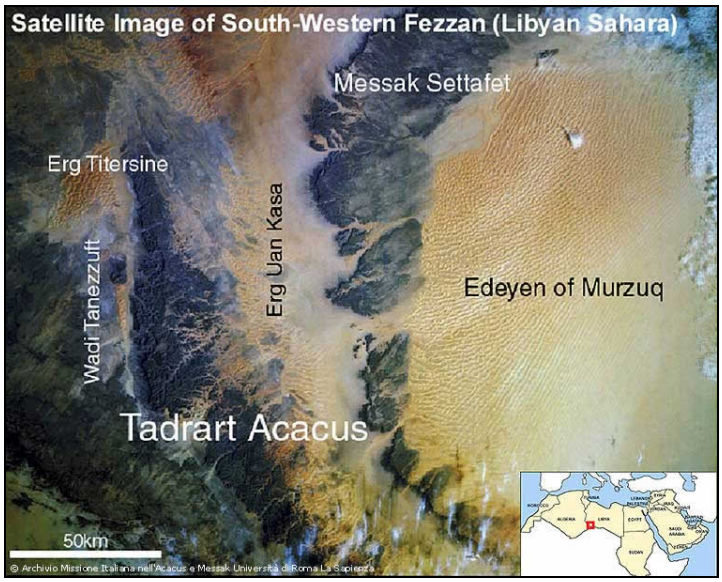


**Figure 1.2:** Satellite view of the surveyed area (Italian Archaeological Mission in Acacus and Messak Archive, University of Rome "La Sapienza").

The Acacus consists of a deeply dissected mountain range composed mainly of Permian and Carboniferous sandstone. East to the Acacus, the Messak Settafet represents a relict plateau, cut out of Cretaceous sandstone. Both the massifs are

ploughed by deep *wadis* and valleys, and preserve traces of the human occupation since the beginning of the Holocene. The Erg[3] Uan Kasa is situated between the Acacus and the Messak, while to the west of the Messak is located the Edeyen of Murzuq. They are sand-seas, characterized by dune-dominated landscapes, and fluvial and lacustrine deposits. The variety of environments that characterized the Fezzan during the Holocene, were able to provide man with many possibilities of adaptation, and so played a key role in the history of human populations that occupied the area (Cremaschi and Di Lernia 1999).

Nowadays, the geological features of the area tell us the history of the past climatic oscillations that characterized the whole Sahara during the Holocene. The main sources of evidence for paleoclimatic reconstruction in the area are the cave and shelter fills in the mountain ranges, and lacustrine or marsh deposits in the interdune corridors located in the *erg* areas. In the reconstruction of environmental changes, TL-, OSL-, U/Th-[4], and radiocarbon-dated geological and geoarchaeological evidences have been used (Cremaschi 1998, Cremaschi and Di Lernia 1999). In parallel, archaeological finds are found throughout the whole area: all these evidence makes it possible to reconstruct how the relationship between humans and the environment changed over time because of the climatic shifts.

Geological evidence (Martini et al. 1998) indicates severe desert conditions have existed for 90,000-70,000 years B.P.: these presumably correspond to the beginning of the so-called Ogolian desert encroachment (Cremaschi and Di Lernia 1999). The heavy rains which occurred at the end of the Pleistocene, that correspond to the beginning of the post-Ogolian wet conditions, changed the landscape of Fezzan dramatically: the Wadi Tanezzuft was a river that originated from the massifs of Tassilli and Acacus, flowed toward the North emptying into a sea located in the actual border between Algeria and Libya (Cremaschi 2004). In correspondence to the *erg,* i.e. the Erg Uan Kasa and the Erg Murzuq, the water collected by the dunes gave rise to large lakes, whose coasts where often inhabited by the pastoral groups.

[3] An *erg* (also *sand sea* or *dune sea*) is a large, relatively flat area of desert covered with wind-swept sand with little to no vegetation cover.

[4] TL: thermoluminescence; OSL: optically stimulated luminescence; U/Th: disequilibrium Uranium/Thorium.

In the 6th millennium the water level of the lakes reached its highest, but with the climatic crisis 5,000 years ago, they dried out irreversibely. Similarly, the Wadi Tanezzuft river gradually lost its capacity, its flow being limited to the Acacus region. Yet despite this, over a much longer period it continued to be an important hydrological source in the midst of an even-drier Sahara, and its fluvial activity is recorded as late as 3,800 years ago. The Wadi Tanezzuft fed an over 80 km wide oasis, which provided a refuge for man communities for millennia, even during the second half of the Holocene. The drying up of the river was certainly completed since 2,300 years ago. About 2,000 years ago, when the hyperarid conditions that characterize the Sahara nowadays set in, the contraction of the oasis led to the formation of three separated oases which exist up this very day: Barkat, Ghat and Fehwet (Cremaschi and Di Lernia 1999).

## 1.2 PREHISTORY IN THE ANCIENT CENTRAL SAHARA

The Saharan mountains of Southern Algeria, Libya and Northern Chad are famous for being of remarkable interest for prehistoric art (Lutz and Lutz 1995, Mori 1965), and for their rich archaeological records (Barich 1987, Lupacciolu 1992, Mori 1965). In fact this area preserves an incredible collection of paintings and engravings attesting the presence of humans in the Central Sahara since the Paleolithic. Rock art tell us many things about customs and way of living of the people inhabiting a region at any given period; in this sense it is an important instrument for reconstructing cultural transitions in a particular geographical context. Paintings depicting exotic and wild animals (e.g. elephants, giraffes, gazelles, rhinoceros) provide evidence of how the Sahara appeared in the first half of the Holocene, as well as the successive phases (figures 1.3a, b, c).

Apart from rock art, many important archaeological sites, from cave sites to open-air sites, can help to reconstruct the life-style of the inhabitants of the Central Sahara. The archaeological survey in the Acacus and Messak Settafet, carried out by the Italo-Libyan joint mission directed by Prof. Fabrizio Mori from 1990 to

1996, led to the identification of more than 450 archaeological sites (Cremaschi and Di Lernia 1999). The geoarchaeological survey of these sites shed light on the cultural dynamics of the human groups in relation to the environmental changes.

The Holocene sites with extractive economic basis have been classified as 'Early Acacus', and 'Late Acacus'; those with food-producing economy, as 'Early Pastoral', 'Middle Pastoral', and 'Late/Final Pastoral Neolithic'. This terminology has been proposed by Cremaschi and Di Lernia (1999), on the basis of local *facies*, referred to regional Holocene cultures. The transition from prehistorical to historical time is characterized by the presence of Garamantes, the last cultural phase being named after these people: the Garamantian period. The chronocultural attribution was based on many archaeological indicators: lithic assemblages, pottery samples, faunal remains and spatial configurations of archaeological structures (pits, fireplaces, etc.).

## Early Acacus Hunter-Gatherers (ca. 9,800-8,900 Years B.P.)

Geological evidence related to travertine deposits in the Acacus mountain range shows a clear increase in precipitation between 14,000 and 10,000 years B.P., as indicated by U/Th dating. Travertine deposits are the trace of the refilling of the hydrographic network in the mountains; the absence of subsequent travertine deposits indicates that probably this period represented the wettest phase of the late Quaternary. There is no evidence of human occupation in the area recorded at the beginning of this interval. This gap is consistent with the data from other North African contexts (Gabriel 1987, Close 1992, Petit-Maire 1993), and therefore it can be considered as of the time necessary for a significant biological recolonization of a region (Cremaschi and Di Lernia 1996). The oldest 'Early Acacus' site was recorded inside the Uan Afuda cave (9765+/–105; GX-20750) (Cremaschi and Di Lernia 1999).

The first Holocene occupation of the area does not appear to be related to any previous Pleistocene occupation, but rather to the arrival of human groups, who probably followed the Sahelian belt. In fact, the hypothesis of a Southern provenance for the 'Early Acacus' human groups cannot be discarded (Di Lernia

1997, 1999), above all when considering the evidence of a northward shift of the monsoon across the Sahara (Hassan 1996). Whatever the case, the absence of funerary remains belonging to this phase, and the general feature of the lithic industry showing no peculiar relationship with other complexes, provide no further support to this hypothesis.

Environmental conditions of this period provided an abundance of resources, making specialized forms of hunting possible. Most of the archaeological sites are distributed close to the lakes existing at that time inside the dune corridors, e.g. the Erg Uan Kasa and inside the Edeyen of Murzuq. These lowland sites appear to be small, and archaeological records, mostly based on lithic industry, seem to attest the specialised function of these sites. Other archaeological sites are located in the Acacus mountains, and the archaeological records make it possible to hypothesize the existence of multi-activity base camps in the mountains, characterized by stone structures, shallow hearts and a quantity of lithic artefacts struck from different raw material. Analysis on botanical remains in these sites indicates a narrow range of exploited plants (Castelletti et al. 1998), and up to now no seeds of wild cereals have been found. The almost exclusive presence of faunal remains of *Ammotragus lervia*, the Barbary sheep (Corridi 1998), indicates selective hunting practices, despite apparent evidence of fishing activity in lake sites. It has been hypothesized that the relationship between the multi-activity sites in the mountains and the specialized lowlands ones, is related to a different exploitation of the environment. It is tempting to suggest that human groups during the wet season (i.e. summer) dispersed into the lowlands, while coming together again during the dry season (i.e. winter) in the mountain sites, where the settlements appear larger, and where water sources were permanent (Cremaschi and Di Lernia 1999).

## Late Acacus Hunter-Gatherers (ca. 8,900-7,400 years B.P.)

At the end of the Early Acacus phase, geological evidence from the Uan Afuda cave indicates an increasing aridity. Most of the sites that belong to this chronological phase are located in the Acacus area, showing a continuity with the sites of previous

Early Acacus. By contrast, in the dune areas close to the lacustrine or marsh basin, the lowland sites significantly decrease. Mountain sites appear to be larger than the previous ones, and many indicators suggest that they were inhabited for longer periods during the year: organization of lithic technology, grinding equipment, and pottery seem to demonsrate an increase of sedentism. Moreover, the sites themselves were more complex, presenting an organization in separated specialized areas, e.g. fire, fodder stocking, and probably animal corralling. It is likely that human concentration in the mountain sites is related to the drying process, despite a more generic cultural choice cannot be ruled out (Cremaschi and Di Lernia 1999).

While in the Early Acacus the economy was specialized around hunting of Barbary sheep and the modest use of plants, in the Late Acacus it was more diversified, with hunting of both small and large mammals, fish, birds and, above all, with a flourishing exploitation of wild cereals (*Urochlea/Brachiaria* type) and the use of grinding equipment (Di Lernia and Cremaschi 1996, Castelletti et al. 1998). Furthermore, it is worth noting that in the site of Uan Afuda, the first traces of forced enclosure of *Ammotragus laervia* were recorded. It represented the first attempt of cultural control of these animals in the area, as documented by the accumulation of dung and forage in the more internal part of the cave. A sort of containment of the Barbary sheep and periodical slaughtering of the animals in particular periods has been hypothesized (Di Lernia and Cremaschi 1996, Castelletti et al. 1998).

The shift toward a broad spectrum of resources, and forms of planned exploitation of resources, are a typical response of hunter-gatherer groups to the situations of competition and stress that took place in critical periods, as when population increased, leading to an imbalance between people and resources (Binford 1983).

## Early Pastoral (ca 7,400-6,400 years B.P.)

Geoarchaeological evidence (Cremaschi and Di Lernia 1999) indicates that the second half of the Early Pastoral (around 6,900-6,400 years B.P.) was characterized by wet conditions, both in the mountains and the sand seas: it coincided with a dense occupation of both areas. Concerning the settlement system, a continuity seems to be documented between the Late Acacus phase and the Early Pastoral, with the latter system distributing preferentially in the interior areas of the mountains. The settlement pattern seems to indicate the existence of movements between the mountains and the lowlands, probably related to seasonal fluctuations. Movements from the mountain sites to lowlands can be seen as a first form of seasonal transhumance.

The archaeo-zoological records show the absence of continuity with the Late Acacus phase. This can be seen as an evidence of the external provenance of domestic ovicaprines, which were widespread in South-Western Asia while absent in Africa during all the Pleistocene. Bones of sheep and goats have in fact often been recovered between 7,300-7,000 years B.P. in a wide area, from the Red Sea to the Acacus itself (Di Lernia and Liverani 2004). Ovicaprines were probably introduced by Pastoral groups coming from East, who were forced to move westwards because of the increasing aridity in the North-Eastern regions of Africa, driven by the search of water and pastures. Probably the increasing cultural complexity of the Late Acacus hunter gatherers, and their attempt of containining Barbary sheep, favoured the introduction of domesticated animals, whose frequency in the faunal records grew quickly (Di Lernia 1998, 1999). Concerning cattle, whether its origin is autochthons or was introduced by Eastern people is still being debated (Hanotte et al. 2002)

## Middle Pastoral (ca. 6,100-5,000 years B.P.)

A short dry period may have occurred at the end of the 7$^{th}$ millennium B.P., as attested by pollen diagrams and cave deposits (Cremaschi and Di Lernia 1999). It corresponds to a drop in human occupation.

The Middle Pastoral appears as the most flourishing period in the Pastoral period. It corresponds to the climax of the Pastoral society, this favoured by the good weather conditions: lakes in the plains reached their highest level, encouraging an increase of human settlements that became larger and more complex than the previous ones. The mountain sites attest the continuous but intermittent presence of humans. The hypothesis of seasonal, probably winter occupations of such sites by shepherds has been made (Cremaschi 1996, Mercuri et al. 1998). A possible scenario is that Pastoral groups settled in the plains and bred bovine and ovicaprines for long periods of the year, while during the dry winter season small groups moved to the mountains, the Acacus and the Messak, with the animals, especially sheep and goats. The fact that bovine archaeological remains are frequent in the plain sites, while rare in the mountains, leads to the hypothesis that they were used mostly for secondary products, particularly milk and blood. The final picture is one of a mature semi-sedentary pastoralism, which resembles that of the extent African pastoralist groups. Evidence of the complexity and maturity of the Pastoral society in this period is related to the rock art. Engravings and paintings depict the physiognomy of the figures, these showing a variety of morphological traits. In fact, the Pastoral society appeared in this period as a multiethnic society: evidence of such heterogeneity comes from the analysis of funerary rituals; this reveals the existence of many traditions and cultures intermingling with each other: e.g., single and multiple burials, differences according to the age and sex of the dead, mummification. In particular, it is worth noting that in this phase, burials are always placed inside the human settlement, indicating a deep continuity between life and death (Di Lernia and Liverani 2004)

Even the physical anthropological analysis confirms the deep-rooted heterogeneity of Pastoral society (Ricci et al. 2008, in press).

## Late and Final Pastoral (ca. 5,000-2,700 years B.P.)

Around 5,000 years B.P., weather conditions dramatically changed: dry conditions that characterized this period drastically transformed the Saharan landscape, leading to the complete drying up of lakes. Fluvial activity and water availability are

attested at 3,800 years B.P. only along the Wadi Tanezzuft. Dessiccation was certainly completed by 2,300 years B.P.. High population density, overgrazing, and the uncontrolled use of resources contituted factors of deep stress and competition in the Pastoral society. Archaeological surveys show that the substantial homogeneity and unity of the Middle Pastoral is fragmented by the birth of regional entities, which resulted from humans adaptation to the climatic crisis. Both nomad and sedentary communities characterized this periond.

Nomad pastoralism. The new climatic conditions forced human groups to adapt their settlement system and their economy. Food security could no longer be provided by bovines, but only by more resistant animals, e.g. ovicaprines. Settlements consist of moderately-sized camps in the plains which were inhabited during the summer season, while hundreds of shelters were used as sites of settlement in the mountains. They were distributed over a wide area, both in the Acacus and in the Messak, and were occupied during the winter season. These small sites were mostly used as stables for goats (Cremaschi and Di Lernia 1999).
The exclusive use of ovicaprines, and the nomad pastoralism based on the annual mobility, are distinctive markers of this phase. This new kind of pastoralism attests the necessity for the man to continuously move to find new pastures, akin in this regard to the modern desert nomadic life-style.

Communities of the oases. In an environment ever more dominated by increasing aridity, fluvial valleys (e.g. the Wadi Tanezzuft valley), and their paleo-oases became a focal point of human concentration, which may have favoured some forms of sedentism, with intensive use of plant resources (Di Lernia and Livereani 2004). In these regions, control and access to resources acted as critical factors which led to the first forms of social stratification. This aspect is reflected in the funerary architecture too: for the first time megalithic structures appear in the pastoral culture. Tumuli, platforms and other rock monuments clearly represent evidence of hierarchical structure in the oases-communities (Di Lernia and Livereani 2004).

At the end of the Pastoral period, human communities appear to be completely adapted to the dry conditions. Little is known about the society of the Final Pastoral, although it appeared to be well structured and complex. Long-distance tradings with foreign people (e.g. sorghum in West Africa, and date palm in East Africa) clearly demonstrate the dynamic nature of these communities that crossed the Sahara desert looking for new resources. The relationship between these organized Pastoral communities and the Garamantes that followed are still open of debate.

## Garamantian (2,700-1,800 years B.P.)

Toward the mid of the 2nd millennium B.C. (Before Christ) the transformation process of Saharan communities was complete. Oases were reduced their present size (e.g. the oases of Ghat, Barkat, and Fewet in the Wadi Tanezzuft), and allowed the human groups to cultivate date palms, cereals, and leguminous plants.

In this period, the establishment in the Saharan border of trading centres for trans-Saharan exchanges, represents an important factor triggering the origin of the first state-entities based on sedentary settlements, like those in the Southern Saharan border along the Sahel-Sudan belt. At the same time, Greek (Cyrenaica) and Phoenician (Tunisia and Tripolitania) colonies were founded on the Mediterranean coast. Herodotus wrote about the opening of the first caravan route crossing the Sahara through Fezzan, from Lower Egypt to the middle valley of Niger. Through this trade system Garamantes brought salt from the Central Sahara to the Chad basin, exchanging gold and other exotic products. These were subsequently transported to Mediterranean, in return for oil and manufactured products (e.g. glasses, ceramic, jewels). The Herodotus's writings attest the birth in that period of new ethnic groups in the Central Sahara, such as Garamantes themselves in the Fezzan, Atlantes in the Hoggar, and Nasamoni south of Cyrenaica.

The first Garamantian settlements date back to the 6th century B.C.. The Garamantes had their most important centre in Germa, in the Wadi el-'Ajjal. Excavations conducted since 1997 by the Italo-Libyan joint mission, shed light on two important Garamantian sites situated in the southern region of the kingdom: the

citadel of Aghram Nadharif, near the oasis of Barkat, which provides a clear evidence of the trade activities of the Garamantes; and the village of Fewet, which attests the agricultural exploitation of the oasis. Since the 1st century B.C., contacts with the Roman empire, concurrently with the diffusion of metallurgy, led to the continuous growth of the Garamantian kingdom.

## Recent history

Around the 3rd-4th century A.C. (After Christ) the Garamantian reign reached its peak. Agricultural activities were improved through new techniques, i.e. the *foggara*, a system of underground irrigation channels. The dromedary was used as a beast of burden instead of donkey, as it afforded higher resistance in the desert for the trade systems. Settlements became more stable and complex, while many changes were also observed in the funerary culture.

In the mid of the 4th century, the crisis of the Roman Empire and consequently of the Mediterranean trades in the African provinces, led to the decline of the Garamantian kingdom itself. From 632 A.C. onwards saw a period of Arabic expansion, that in Africa at first limited to Egypt and the Mediterranean coast. During the Omayyadi Caliphate, in the 7th and 8th century, Arabs reached Maghreb (Whitehouse 1980), and later advanced southward, both to West and East. In the 9th and 10th century, Muslim colonizers reached Eastern Africa, while at the beginning of this millennium Bedouins repeatedly invaded Northern Africa. During this time, many Tuareg pastoralists were driven southwards and were displaced by Arab Bedouins who settled in Tripolitania and the Fezzan. Elsewhere in the Sahara, the Tuareg were able to defend most of their territories, but also to assimilate certain Arab tribes that had ventured southwards (Capot-Rey 1953, Nicolaisen 1963).

**Figure 1.3a:** Acacus phase. During this phase hunter-gatherers groups settled Sahara. Rock art is characterized by realistic representations of wild animals that human groups hunted. Animals depicted (e.g. gazelles, but even elephants, giraffes, rhinoceros) attest that the Sahara at the beginning of the Holocene appeared as a


**Figure 1.3b:** Pastoral phase. During the Pastoral period cattle, that represented the main source of human groups, became the main subject of the rock art. Differently from the Acacus phase, often humans were depicted too, and scenes of both social and work activities were represented.


**Figure 1.3c:** Garamantian phase. During this period the Sahara almost appeared as it is today. Human groups settled around the oases, that were represented in many paintings. Dromedaries, which were introduced during this phase being more resistant to the desert environment, were often depicted too.

## 1.3 THE CASE STUDY: THE TUAREG

The origin of the Tuareg is quite confused: the scarcity of written chronicles prevents us from reconstructing their history exactly. Most Arabian historians and geographers[5] attribute to the Tuareg a descent from Arabic or Semitic populations who reached Maghreb after various military campaigns, and who progressively entered the South of the region, intermingling with local Berber populations (Lhote 1955, Hama 1967). The origin and meaning of the name Tuareg (*Twareg*) has long been debated with various etymologies advanced, although it would appear that *Twārəg* is derived from the broken plural[6] of *Tārgi*, a name whose former meaning was 'inhabitant of *Targa*' (the Tuareg name of the Libyan region commonly known as Fezzan).

The Tuareg speak a Berber language: the *Tamajaq* (also called *Tamasheq* or *Tamahaq*, according to the regions where it is spoken), which appears to have several dialects among the different regions. The Berber language belongs to the Afroasiatic family, which encompasses about 240 languages spoken in Northern Africa from Egypt to Morocco (except for a region in Southern Libya where Nilo-Saharan languages are spoken), in Eastern Africa (Eritrea, Ethiopia, and Somalia) and in the Middle East and the Arabic peninsula. More particularly, the Berber branch is spoken by most of nomadic Northern African shepherds and is closely related to Pharaonic Egyptian and the Semitic languages (such as Arabic, Hebrew, and Amharic).

The Tamajaq writing system, *Tifinagh* (also called *Shifinagh* and *Tifinar*), descends directly from the original Berber script used by the Numidians in pre-Roman times (source: Smithsonian Institute, http://africa.si.edu/exhibits/tuareg/who.html).

[5] Among them: Ibn Kaldoun, Ibn Abd Al H'akam, Yakout, Ibn Hokal, and Ibn Batoutah (Tornieri 2002).

[6] In linguistics 'broken plurals' are a grammatical phenomenon typical in many Semitic languages of the Middle East and Ethiopia, in which a singular noun is "broken" to form a plural by having its root consonant embedded in a different 'frame', rather than by merely adding a prefix or suffix to the original singular noun.

Despite the common language and culture, the Tuareg population has always been divided into different groups. Each group appears as a mosaic of traditions and tribes often in conflict with each other. During the Middle Ages, the Tuareg were organized into two major groups: the *Ihaggaren*, the so called Northern Tuareg, which were located in the Ahaggar massif; and the *Tademaket*, known as Sudanese Tuareg, who were settled in the Adrar des Iforas in Mali. In figure 1.4, the precolonial geographic distribution of Tuareg in Africa is reported.

In the 16$^{th}$ – 17$^{th}$ century these two groups fragmented, probably because of internal struggles, into the *confederations* which characterised the pre-colonial time (see forward): *Kel Ahaggar* and *Kel Ajjer*, in the present-day Algeria, *Kel Adrar* in Mali, and *Kel Ioullimmidden* in the Western Niger.

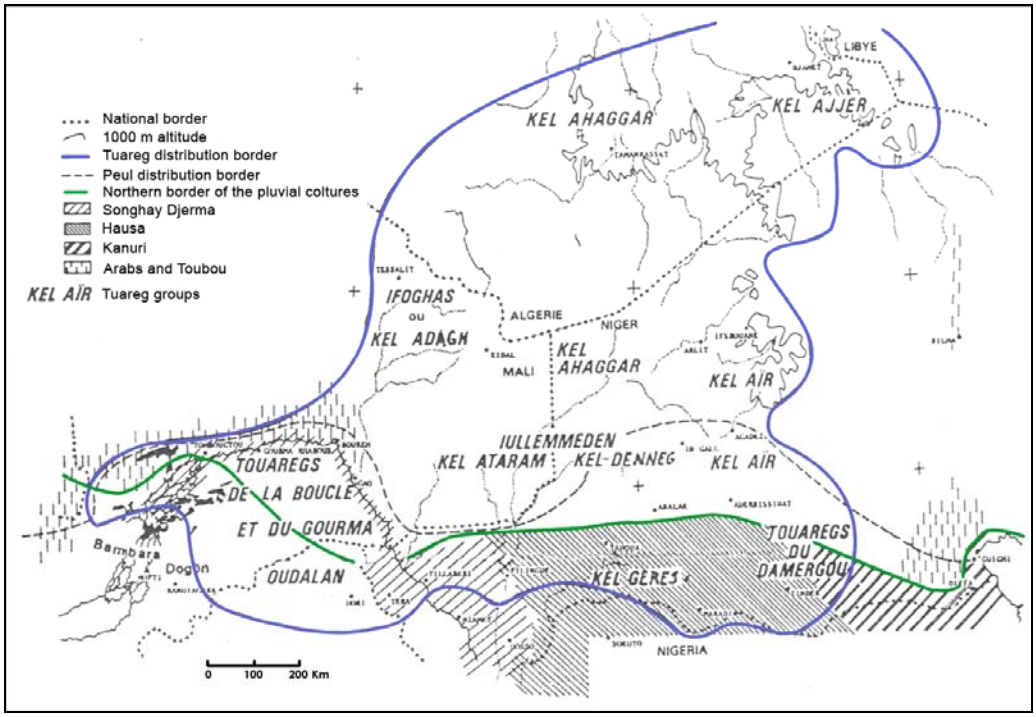


**Figure 1.4:** Precolonial geographic distribution of Tuareg in Africa (Wikipedia, Tornieri 2002).

Other isolated nomad tribes reached the Aïr region in the 10$^{th}$ century, mixing with local autochthonous populations. Here, the organization into a confederation was decided only after the arrival of some groups from *Kel Ahaggar*, who were attracted

by the rich pastures of the region. When the French army conquered the Sahara between the end of 19th and the beginning of the 20th century, the Tuareg were organized in the following confederations (Lhote 1955, Hama 1967, Tornieri 2002):

- *Ihaggaren* o *Kel Ahaggar*: placed in the Ahggar massif, in Algeria.
- *Kel Ajjer*: settled in the Tassili-n-Ajjer, near Ghadames and the Western part of Libyan Fezzan.
- *Kel Ioullimmidden*: located in the Boucle of Niger. They included the Kel Adrar, who after independence from this confederation were subjects of the Kel Ahaggar.
- *Tenguéréguif*, who inhabited the region of Timbuktu, in Mali.
- *Kel Aïr*, who include the Kel Geres, located in Damergou, and the *Kel Ewey* and *Kel Ferwan*, in the Aïr massif.

No stable political relationships were ever established between the confederations. Sometimes, neighbour tribes of different confederations stipulated military alliances against common enemies. Only during the French invasion were alliances between whole confederations established, but these never led to the Tuareg unification.

## Precolonial social structure of the Tuareg: a stratified society

Despite a certain cultural unity, it is not totally correct to speak about a Tuareg society: as previously described, many independent regional entities exist as a result of different social and geographic contexts. It is possible to provide a general description of the Tuareg and their relationship to territory, something common in all the Tuareg groups (Tornieri 2002).

Organization of precolonial Tuareg society is based on a rigid division in social classes, which reflects the separation into tribes. They are the result of long-term processes influenced by hegemonic classes, who try to maintain their privileges over generations. A social structure appeared in the precolonial times as follows.

*Imajeghan*. They represented the hegemonic class in the context of the *Ettebel*, i.e. the confederation. Only descent from such a tradition enabled one to be the leader of the *Ettebel*, who is called *Amenokal*.
They represented the upper class, possessing camels and lands near the oases which were generally cultivated by the slaves (*Iklan*). Wars represent their main activity and is what confers to *Imajeghan* the supremacy. Hooting and raids on the neighbouring camps of other tribes belonging to a different *ettebel*, or on the sedentary agricultural villages were often carried out in the dry season. Animals, slaves and precious objects were taken during these raids, and people were submitted to the control of the *Imajeghan*. Sometimes, political and commercial matters, or disputes concerning control of pasture, could cause wars involving tribes or even whole confederations. The lower classes were subjects of the *Imajeghan* as they were the only ones who could grant them protection against attacks from the *Imajeghan* of other tribes: access to means of warfare was in fact a peculiarity of this hegemonic class.

*Imghad*. They were shepherds who generally were allowed to possess only goats. Their submission to *Imajeghan* constituted a form of an annual tribute obliging them to provide the *Imajeghan* animals and other alimentary resources, and to participate in wars or raids.
They were politically organized in tribes, whose leader was autonomously elected among them. In fact, the *Imajeghan* generally did not interfere with the political issues of the *Imghad*. In the context of the *Ettebel*, the leaders of the *Imghad* tribes, could vote for the *Amenokal* in an assembly together with the Imajeghan.

*Ineslimen*. They represented the religious class. Their social importance varied according to the confederation: sometimes they were not organized in tribes, and could receive the same social status of the *Imajeghan*, so that marriages between these two classes were allowed. On the other hand they could be subjected to the noble class, and were organized into politically autonomous tribes based on a pastoral economy.

*Enaden*. They were artisans who worked with wood, leather, and metals. The *Enaden* generally lived together with the *Imajeghan* or the *Imghad* in the same camp, so they usually were not organized into separated tribes. Often, not more than two families lived in the same camp, and were never really integrated in the life of the camp as they were despised on account of being manual workers: they represented the lowest class among the 'free men'. Their marginal position was emphasized, linguistically, by the use of a peculiar dialect, the *tenet*, which was hardly understood by the other groups.
Mythical sources attribute to the *Enaden* a different origin in respect to the other Tuareg groups of free men: their descent probably derives from autochthonous South-Saharan populations external to the Tuareg.

*Iklan*. They were the slaves, who were generally captured by the Imajeghan during their raids, or bought from the Arab merchants of Tidikelt or Timbuktu (Mali). The *Iklan* took care of all the manual activities in the camps, from agriculture to breeding, as all manual work was despised by the Imajeghan. The *Imghad* and the *Enaden* could also possess slaves. They were completely estranged from social and political life, and were treated as private objects which could be exchanged in the social transactions.

## Precolonial political organization of the Tuareg

The political organization of the Tuareg society was based on two entities: the tribes (*Tawshit*) and the so-called confederations (*Ettebel*).

The confederation: *Ettebel*. As mentioned before, Tuareg were divided into confederations of tribes, the *Ettebel*. Each *Ettebel* was governed by the *Amenokal*, who was chosen among the tribes of *Imajeghan*. Composition of tribes corresponded to the different social categories, and could significantly vary from confederation to confederation. A tendency for slaves to be more numerous in the South was observed: from about 10% of the whole confederation in the Sahara, to even 70-80% in the Sahel-Sudan belt (Bernus 1981). The *Imajeghan* were always a minority, while the number of *Imghad* could vary.

Even external non-Tuareg groups could belong to the *Ettebel*, which in this sense was a very flexible political structure. Independently from their origin, foreign groups that could strengthen the *Ettebel* were allowed to participate in the socio-political life of the confederation, submitting themselves to its authority (e.g. the Arabs, who often were admitted into the *Ettebel* because of their guns) (Capot-Rey 1953, Nicolaisen 1963).

The tribes: *Tawshit*. Tribes were composed of individuals of the same social category who shared the same geographical area. Belonging to the tribe was defined by matrilineal descent: all the individuals descended by the common female ancestor. However, these blood ties were often weakened by adoption practices or protectorate relationships. The demographic size of the tribes could widely vary, both between different tribes, and within the tribe itself. In fact, wars, alliances, and even weather conditions (e.g. extreme dry seasons), could lead a faction of one tribe to move and join other ones. According to a study conducted by F. Nicolas in 1950, the confederation of *Kel Ferwan* was composed of 25 tribes, whose size varied from 25 to 635 individuals, with an average of 250 (Nicolas 1950, Nicolaisen 1962). It clearly suggests that the demographic profile of the confederations could be highly complex.

Tribes were strictly endogamic, especially the noble ones: in fact, marriages were allowed only between couples from the same tribe. Only exceptionally were marriages between people from different social categories authorized. Strict rules were generally followed after marriage: the new familiar nucleus settled in the camp of the bride's parents, up to the birth of the first son, when they had to move to the camp of the groom's parents. After the death of the man, the woman and her sons moved back to her mother's territory.

The camp: *Aghiwan*. In the context of each tribe, the exploitation of the resources, i.e. water and pasture, was carried out at the camp level (*Aghiwan*). This was the context in which the pastoral activities were carried out, and so, in contrast to the confederation and the tribe, it had a economic connotation. The nomad life-style of

the Tuareg led to the continuous movement of camps, whose time of stay in a particular place was generally associated with a series of pastoral cycles.

The practice of the pastoral activity was carried out at the family level. The camp was composed of a man and his wife, their sons with respective wives and sons, and their nubile daughters (or in case, divorced or widow daughters with their sons). Generally all the decisions about the pastoral activities were taken by the oldest man in the camp, who even decided when to move to another region and where to go.

## The pastoral circuit

The flock is generally composed by ovicaprines, camels and sometimes bovines and horses, especially in the Sahel area. Size and composition of flocks change according to the tribe they belong to (noble or tributary) and to the weather conditions (alternation of wet and dry periods). A precise organization concerning the choice of pastures is followed according to the species of animals, to their abundance, and to the availability of manual workers and resources, in particular water. It results in a very specialized pastoralism that aims to reduce competition among the animals and the pressure on pastures-land. The hardest moment is during the warm and dry season, when water demands increase while its availability is drastically reduced. It leads to the concentration of many shepherds near the same water source, i.e. the wells, this critical situation causing degradation of pasture and food shortage for the animals.

Movement of camps follows a cyclical trend and is strictly related to seasonal variation. Rainfall during the wet season leads to the growth of new pastures and marks the beginning of a new year. Frequency of movements during the year can vary in relation to the area and the availability of resources, abundance of rain in the last wet season, and the particular demands of the human groups. Camps can move as many as five or six times in a year (Bernus 1986).

During the wet season, that can last up to three months, camps generally settle in areas protected from the rainfall. Rainfall is sufficient to renew pastures with grass and water. During this period, high availability of resources allows the camps to

meet up and people from different groups to gather for rituals and sport competitions (*tendè*). Once the wet season ends, the landscape soon starts to dry. Temperatures begin to decrease and the camps disperse, looking for new locations in the valley. This period coincides with the dry and cold season. With the arrival of the hot season camps further disperse over a wider area. It coincides with the hardest period of the year: the grass is completely dry, and some animals die because of the lack of resources. Camps are forced to move with a higher frequency looking for new pastures to exploit. The arrival of the rains coincides with the progressive concentration of camps and so with the beginning of a new cycle.

## Relationship with other nomad and sedentary populations

Both nomad (Arabs, Bedouin, Toubou etc.) and sedentary populations (Songhay, Hausa, Djerma) shared the same territory with the Tuareg populations: their cohabitation led to various results. In fact, while some nomad groups have been integrated in the political, but not cultural life of the Tuareg (e.g. the Arabian group Eddès Deremshaka and Peul), often sedentary populations have found themselves subjected to military oppression, being the object of raids and ambushes (Bernus 1969, 1981, Tornieri 2002). Furthermore since the Middle Ages, caravan routes, strongly contributed the incentives for Tuaregs to establish contact with other foreign groups. Generally, trade was directed toward the Sahel-Sudan region where millet and animals were exchanged, or to the Saharan-Sudan region in order to exchange salt, dates and millet.

Two important commercial bases were at the centre of the trade system, one located in the Sudan region, the other in the Maghreb area, near the oasis of Ghat. A reconstruction of the caravan pathways in the Central Sahara is reported in figure 1.5.

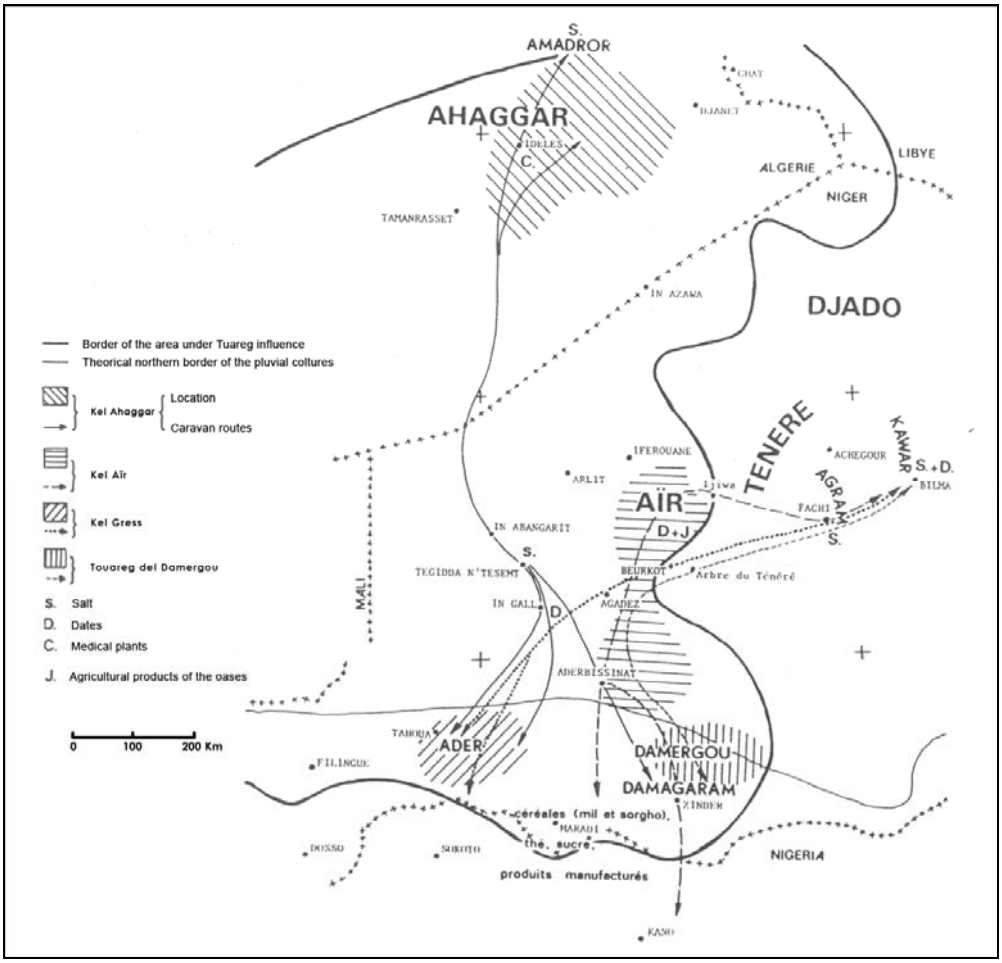


**Figure 1.5:** Caravan pathways in Central Sahara (Tornieri 2002).

## The French conquer and the dissolution of the Tuareg socio-political system

At the beginning of the 20th century, the French army finally conquered the Sahara, after many years of battle. The long period of war and the rebellion of Tuareg groups between 1916 and 1919, led to a severe weakening of the Tuareg socio-political system: the *Imajeghan*, who were directly involved in the battles, were decimated to a small minority, and lost control over the lower social groups. In this sense, French domination destroyed the traditional dependence/protectorate relationships at the basis of *Imajeghan*'s power and of their tributary system: in

fact, the colonial authority interposed between the nobles and the lower social classes, who had not to ask for protection anymore.

A completely new organization of tribes followed the Tuareg's defeat. French authority set most of the slaves free, and disassembled the confederations. Tribes were often divided or mixed each other, and were substituted by 'artificial groups' that were forced to settle in more accessible areas, so as to allow the French aouthority to have more control on them (Giazzi 1996). A significant decline of pastoralism is observed in this period: nomad tribes were confined to precise areas according to the new administrative rules, and the progressive contraction of the pastoral activities to small spaces was at the basis of frequent conflicts between nomads and agricultures, who were supported by the French administration. In this period it was also frequent that small tribes joined to richer group in order to survive.

Concurrently to the progressive decline of *Imajeghan*, a political and economical growth of *Imghad* and *Ineslimen*, and even an enrichment of the old slaves, was observed. In fact, once free from their noble owners, slaves had the access to private property and could keep their own flocks. Anyway, it is worth noting that only with the independence, in 1960, slavery was completely abolished.

In the recent history, it has to be noted that during Keddafi's Jamahiriya revolution, many Tuareg came to Libya from Chad, Algeria and Niger, and settled in the South of the country. Many of them underwent a process of sedentism near Ghat and Ubarj (Gaudio 1993). According to recent population size estimates, nowadays Tuareg are more than 1,300,000 in Northwest Africa. Recent census counted about 17,000 Tuareg in Libya (http://www.ethnologue.com), within which 5,000 are distributed in Fezzan


**Figure 1.6:** present distribution of Tuareg in Africa (Wikipedia).

and speak the dialect of Ghat, that is one of the most important inhabited area in Fezzan. In figure 1.6 the present distribution of Tuareg is reported.

## 1.4 THE MITOCHONDRIAL DNA

Human mitochondrial DNA (mtDNA) is a circular double-stranded molecule, 16,569 base pairs (bp) in length, that corresponds to the 0.0006% of the total haploid human genome (about $3x10^9$ bp). The two strands that compose the mtDNA are different in basis composition, so that they are named 'Heavy' (H, rich in Guanidine residues) and 'Light' (L, rich in Cytosine residues). No introns or repeated sequences are present in mtDNA, and it represents one of the most known eucaryotic genomes as its complete sequence has been studied in many organisms.
In humans, first complete sequencing of mtDNA was performed on an Englishman (Anderson et al. 1981). Since then, this sequence has been used as a reference for all human mtDNA sequencing studies, and it is termed Cambridge Reference Sequence (CRS). In 1999, CRS was reanalysed and revised (rCRS) (Andrews et al. 1999). Mitochondrial DNA codes for 13 subunits of the oxidative phosphorylation system, two ribosomal RNAs (rRNAs), and 22 transfer RNAs (tRNAs) (Anderson et al. 1981, Korhonen et al. 2004). It is present in hundreds to thousands of copies in each cell, not within the nucleus, but within the cell's energy-generating organelles, the mitochondria. In figure 1.7, a schematic map of mtDNA is shown. The mtDNA consists predominantly of coding DNA, with the exception of a 1100-bp long fragment that has mainly regulatory functions and is therefore termed the Control Region (CR). Within the CR, two hypervariable regions are present, i.e. the Hypervariable Sequence I (HVS-I) and the Hypervariable Sequence II (HVS-II). As mtDNA encodes essential components of the cellular energy production apparatus, lesions in mtDNA and mitochondrial dysfunction contribute to many metabolic human diseases (e.g. MELAS, MERFF, ataxias).

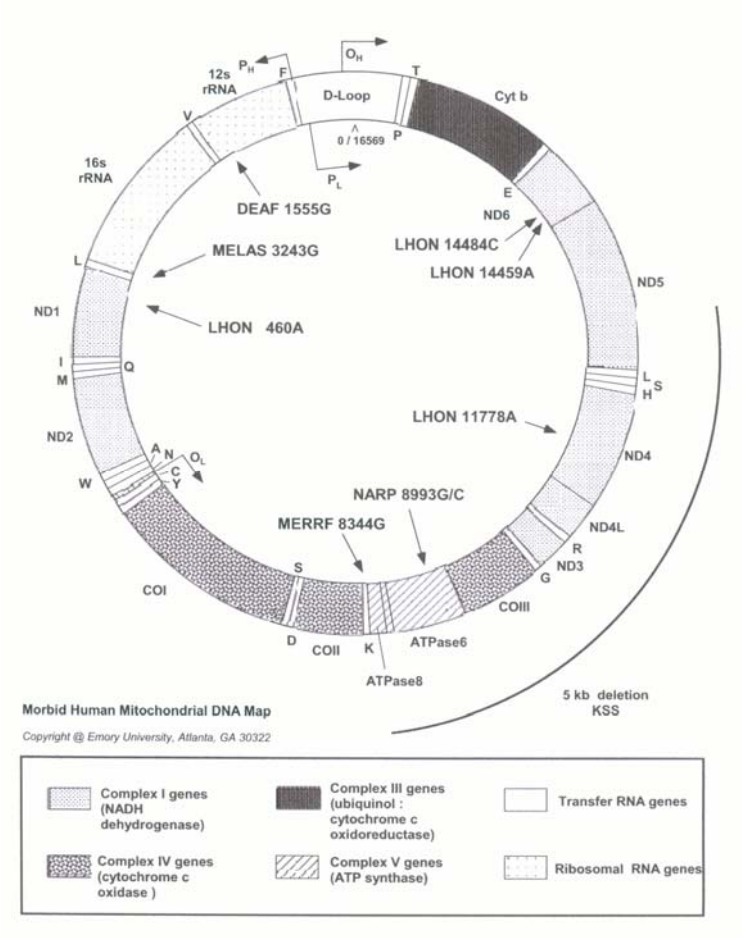


**Figure 1.7:** Mitochondrial DNA map. Nucleotide positions associated with mitochondrial human diseases are indicated.

Organization and inheritance of mtDNA remain poorly understood. Mitochondrial DNA is packaged in protein-DNA complexes that are associated to inner membrane of mitochondria (Stuart et al. 2005), and are called, by analogy to the bacterial chromosome, mitochondrial *nucleoids*. In table 1.1, an interspecies comparison of mitochondrial nucleoids is reported. Proteins involved in the replication (Korhonen et al. 2004), transcription (Gaspari et al. 2004), repair (Stuart et al. 2005), and perhaps recombination (D'Aurelio et al. 2004, Tsaousis et al. 2005) of mtDNA are among the nucleoid components. Recent findings indicate the existence of a dedicated apparatus that coordinates nucleoid segregation with mitochondrial dynamics and cell division (Azpiroz and Butow 1993, Nunnari et al. 1997,

Okamoto et al. 1998, Boldogh et al. 2003). In yeast, it has been observed that proteins Aco1, Abf2, and Ilv5, are directly involved in the metabolic remodelling of mitochondrial nucleoids (Chen and Butow 2005, Chen et al. 2005), as reported in Figure 1.8.

**TABLE 1.1**
An interspecies comparison of mitochondrial nucleoids.

| Species | Size | Number per cell | Number of mitochondrial genomes per cell | Size of mitochondrial genome |
|---|---|---|---|---|
| *Saccharomices cerevisiae* | ~0.2-0-4 μm in aerobic and ~0.6-0-9 μm in anaerobic cells (diameter) | ~40-60 in aerobic and ~7.6 in anaerobic cells | ~1-2 in aerobic and ~29 in anaerobic cells | 75-80 kb |
| *Physarum polycephatum* | Up to ~1.5 μm in length | ~15 | ~40-80 | 63 kb |
| *Chritidia fasciculate* | ~1.0 μm x ~0.35 μm | 1 | Several thousand mini circles and a few dozen maxi circles | 0.5-10 kb for mini circles and 20-40 kb for maxi circles |
| Humans | ~0.068 μm (diameter) | 466-806 in cell lines | ~2-10 | 16.5 kb |

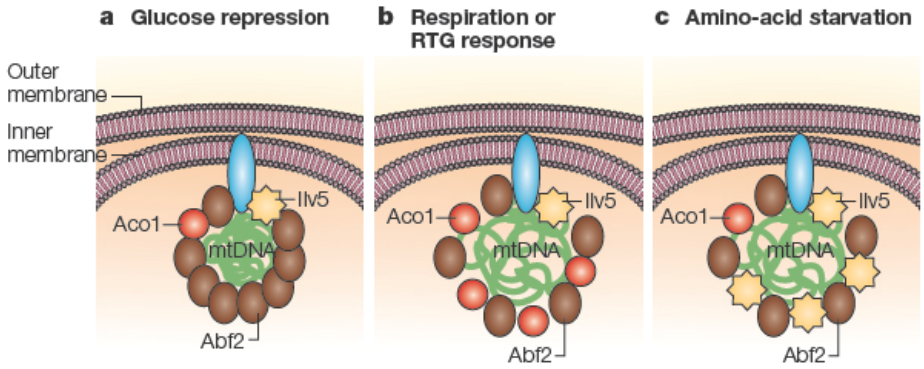


**Figure 1.8:** Hypothetical function of Aco1, Abf2, and Ilv5, in the metabolic remodelling of mitochondrial nucleoids in yeast (Chen and Butow 2005).

Since the first in-depth study of human mtDNA variation 25 years ago (Brown 1980), mtDNA has become widely used for studies of human evolution, migration, and population histories (Johnson et al. 1983, Cann et al. 1987, Vigilant et al. 1991,

Torroni et al. 1993, Krings et al. 1997, Melton et al. 1998, Reed and Stoneking 1999, Ingman et al. 2000, Tambets et al. 2004). In figures 1.9a, b, c, some schematic representations of human mtDNA phylogenies, and their use in the reconstruction of human migrations in the past, are reported. This widespread use is due to the unique features of mtDNA that make it particularly amenable to evolutionary studies. These features include the high copy number, the mathernal inheritance, and the lack of recombination.

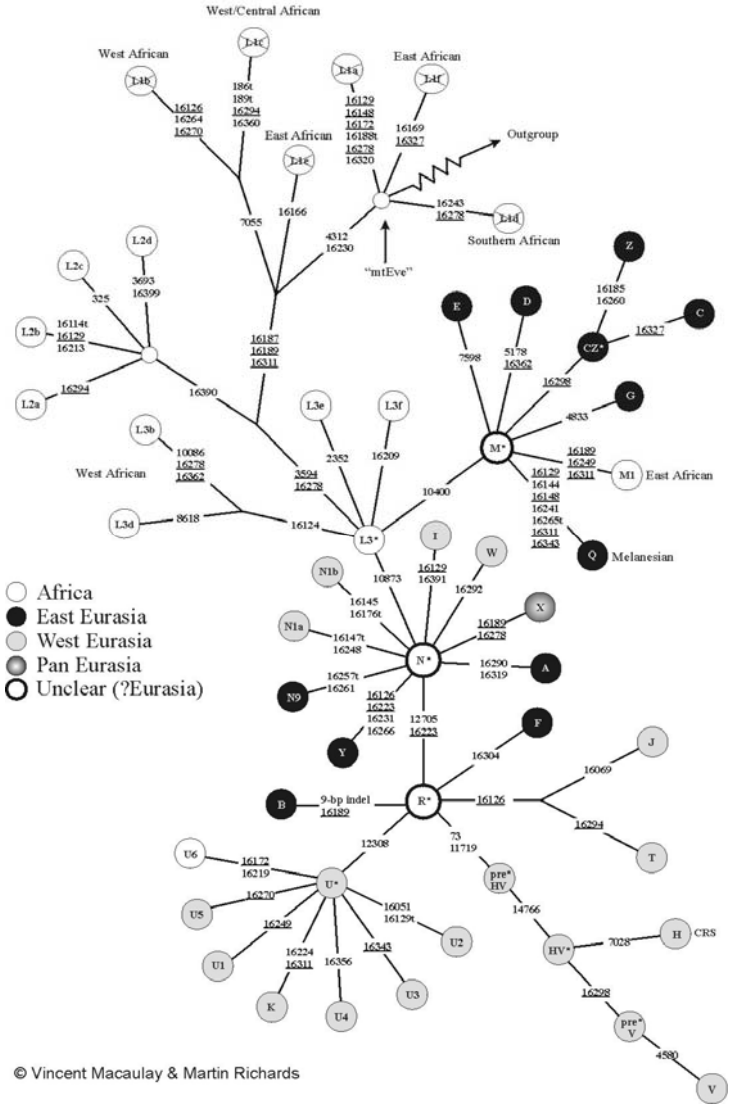


**Figure 1.9a:** Schematic representation of mitochondrial DNA evolutionary tree (2000)

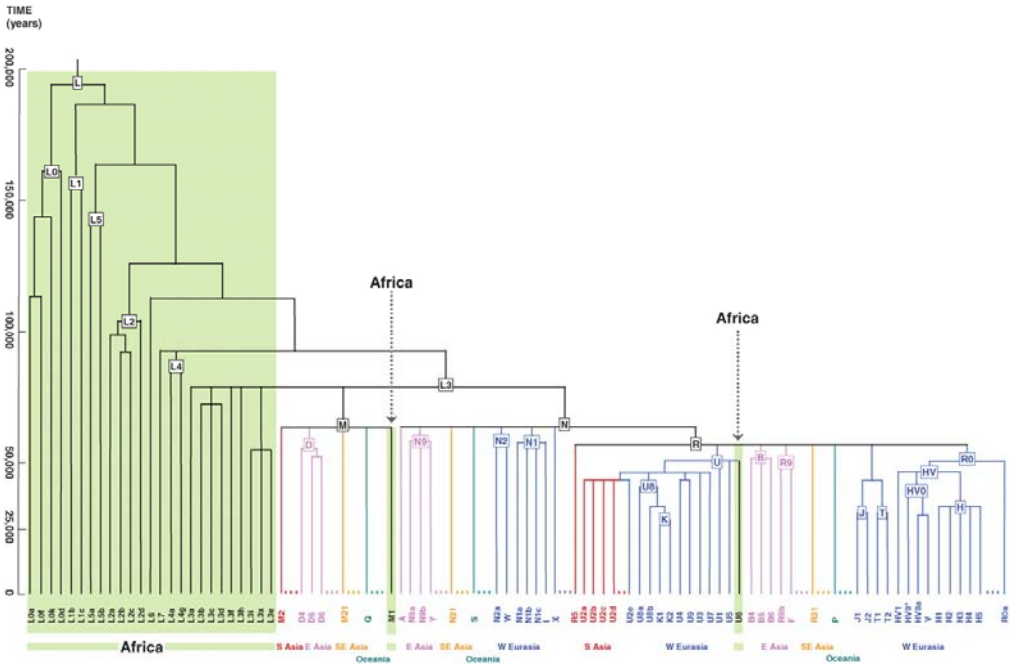


**Figure 1.9b:** One of the latest representation of mitochondrial DNA evolutionary tree (Olivieri et al. 2006).

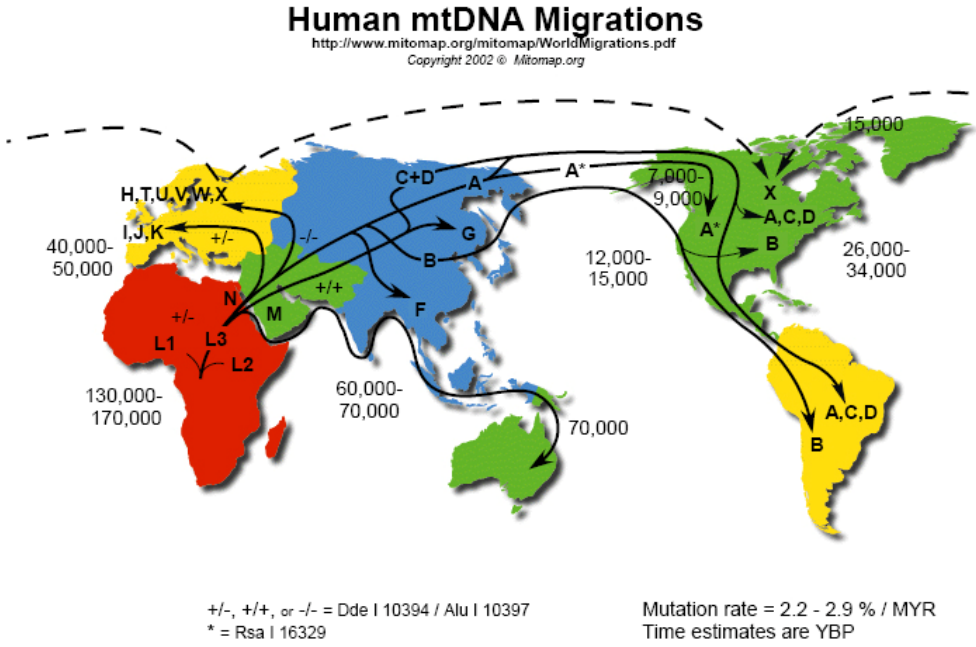


**Figure 1.9c:** Representation of human migrations on the basis of mtDNA phylogeny.

## Properties of human mtDNA

High copy number. The average somatic cell has just two copies of any given nuclear gene or DNA segment, but hundreds to thousands of copies of mtDNA (Robin and Wong 1988). More particularly, each nucleoid contains up to ten mtDNA molecules, and between 400-800 nucleoids are present in one single human cell (Chen and Butow 2005). The high copy number, along with the extranuclear, cytoplasmic location of mtDNA, makes it easier to obtain mtDNA for the genetic analysis, and also makes mtDNA the molecule of choice for analyzing ancient DNA and for certain forensic DNA applications. On the other hand, the multiple copies of the mtDNA genome within one individual need not to be identical, a topic that is called *heteropolasmy*. More precisely, heteropolasmy is the state in which a cell contain genomes that are genetically different (Chen and Butow 2005). Current best estimate of heteropolasmy is that about 14% of the population has a second mtDNA type present at a frequency of at least 1% (Tully et al. 2000), and indeed it is quite likely that all of us harbour more than one mtDNA type among the many mtDNA genomes in our bodies. Nevertheless, the overall homogeneity of mtDNA within individuals indicate that during the early stages of the oogenesis, a substantial bottleneck occurs (see next paragraph). Recent studies on segregation of mtDNA across generations hypothesized that only ten segregating units per individual exist (Bendall et al. 1996, Brown et al. 2001, Poulton and Machington 2002).

Mathernal inheritance. Human mtDNA is characterized by a strictly maternal inheritance (see Figure 1.10) (Stoneking 1993, Stoneking and Soodyall 1996, Wallace et al. 1999). The hypothetic process responsible for the maternal inheritance, is the selective destruction of sperm mitochondria in the oocytes after fertilisation (Manfredi et al. 1997, Shitara et al. 1998). In fact paternal mtDNA is marked for destruction through ubiquitination mechanisms. It is worth noting that recently, mtDNA from muscle tissue of a man affected by mitochondrial myopathy, showed a predominantly paternal origin (Schwartz and Vissing 2002). Nevertheless, this case remains an exceedingly rare phenomenon: in fact, subsequent investigations of more patients with the same mitochondrial dysfunction have

yielded no further cases of paternal inheritance (Filosto et al. 2003, Taylor et al. 2003, Schwartz and Vissing 2004). Therefore, at present maternal inheritance of mtDNA in humans can still be regarded as a rule (Schwartz and Vissing 2003). Up to now, exceptional paternal inheritance of mtDNA in mussels (Zouros et al. 1992), and inter- and intra-specific hybrids in *Drosophila*, mouse and birds (Gyllensten et al. 1991, Kaneda et al. 1995, Kondo et al. 1990, Kvist et al. 2003), have been observed.

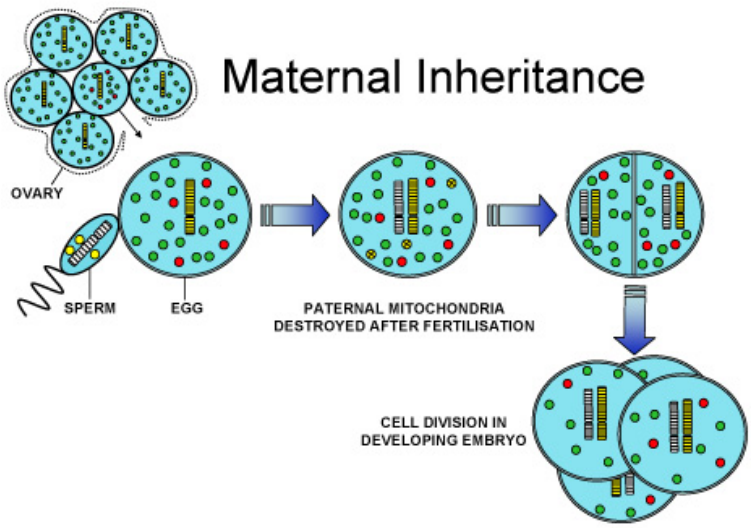


**Figure 1.10:** Representation of maternal inheritance in mtDNA (http://www.ncl.ac.uk/nnp/research/mrg/advice/inheritance.htm).

It should be noted that an embryo generally inherits from its mother identical copies of mtDNA, so homoplasmy is a matter of fact since the first hours of life. The first studies on mtDNA inheritance between generations were realized in Holstein cows pedigrees with a heteroplasmic mtDNA polymorphism (Hauswirth and Laipis 1982). A complete shift of mtDNA genotype between an animal and its descendant was shown to occur rapidly, over two generations. Heteroplasmy was then considered as a transient state of mtDNA returning to homoplasmy over generations. To account for the rapid segregation of mtDNA variants, a genetic bottleneck for mtDNA in the female germ line or in the early embryo was proposed (see Figure 1.11). Studies on mice showed that segregation of the mtDNA

genotypes probably occurs in oogonia, in which the number of mtDNA molecules per cell is reduced to 200 (Jenuth et al. 1996). During female germ line development, mitochondria and mtDNA molecules are both amplified, but the volume of the organelles and the amount of mitochondrial rRNA and mRNA increase ten times more than the mtDNA copy number, leading to a dramatic decrease of the number of mtDNA molecules per mitochondrion (Malka et al. 2006). During maturation of the primary oocyte, the number of mtDNA molecules increases one hundred fold. Further segregation takes place during cytoplasmic partition of the oocyte in the early embryogenesis, and in the following separation of the endoderm, mesoderm, and ectoderm, and migration and division of primordial cells in the undifferentiated gonad. This process explains the segregation of different mtDNA genotypes between tissues (Hauswirth and Laipis 1985, Macmillan et al. 1993, Chinnery et al. 1999). Alteration of the bottleneck effect due to defects in the organization and inheritance of mitochondrial nucleoids, could significantly change ratio of wild-type to mutant mtDNA molecules in the cells of a specific tissue, resulting in many mitochondrial disorders (Di Mauro and Schon 2001, Taylor and Turnbull 2005).

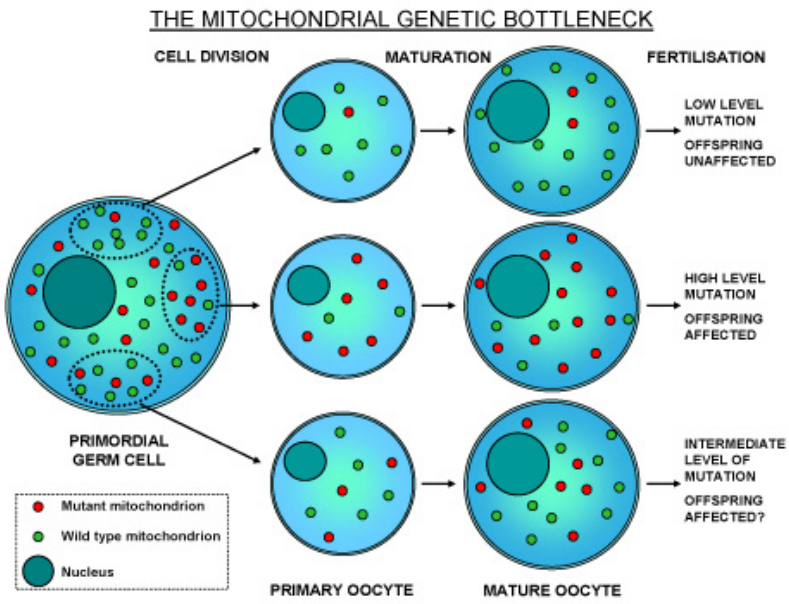


**Figure 1.11:** Model of genetic bottleneck in the female germ line proposed to account for the rapid segregation of mtDNA variants (http://www.ncl.ac.uk/nnp/research/mrg/advice/inheritance.htm).

The haploid inheritance of mtDNA leads to two significant consequences. The first one is that mitochondrial genome has a smaller 'effective population size' (Ne) than nuclear genes: this is estimated 1/4 of the nuclear genome. In fact, a couple of interbreeding individuals possess potentially four haploid nuclear genomes, but only one mtDNA genome to transmit to the next generation. This process is responsible of a higher rate of local differentiation through drift.
The second consequence of the uniparental inheritance, is the absence of segregation and recombination. It allows researchers to trace back over time related lineages, highlighting the maternal ancestry of a population without the confounding effects of biparental inheritance and recombination that characterize the nuclear DNA.

Lack of recombination. The absence of recombination in mtDNA has been demonstrated in many papers (Stoneking 1993, Stoneking and Soodyall 1996, Wallace et al. 1999, Pakendorf and Stoneking 2005). In 1999-2000 four paper claimed evidence for mtDNA recombination in humans (Awadalla et al. 1999, Awadalla et al. 2000, Eyre-Walker et al. 1999, Hagelberg et al. 1999), but all these studies were subsequently shown to be based on faulty and questionable statistical and phylogenetic methods, and new analysis of the same data gave no significant results. Subsequent studies (Ingman et al. 2000, Elson et al. 2001, Piganeau and Eyre-Walker 2004) could not find any evidence of recombination even in data sets from complete mtDNA sequences, although an excess of *homoplasic*[7] sites was detected, hypothetically due to heterogeneous mutation rates that characterize human mtDNA. It is worth noting that recently, recombination in human mtDNA was detected in the only individual so far observed with biparental inheritance: recombination between the maternal and parental contribute occurred in approximately 0.7 % of the total mtDNA in the patient's muscle tissue. Mitochondria possess a functional recombinase (Thyagarajan et al. 1996), but paternal leakage is an extremely rare phenomenon, and it is still unclear to what extent mitochondria within a cell fuse, and exchange their content (Enriquez et al.

[7] Homoplasy: the generation of the same state by independent means (convergent evolution) (Jobling et al. 2005).

2000, Ono et al. 2001Legros et al. 2002). So, in the absence of heteroplasmic DNA molecules, any recombination event would result in identical molecules.

High Mutation rate. The mutation rate of mtDNA is much higher than that of nuclear DNA. The estimated rate is $0.017 \times 10^{-6}$ substitutions per site per year for the whole genome excluding the CR (Ingman et al. 2000, Pakendorf and Stoneking 2005). In the HVR-I and HVR-II of the CR, the rate is even higher, although exactly how high is a matter of controversy: estimates based on inter-/intra-specific phylogenetic comparisons range between $0.075\text{-}0.165 \times 10^{-6}$ substitutions/site/year (Hasegawa et al. 1993, Stoneking et al. 1992, Tamura and Nei 1993, Ward et al. 1991). Nevertheless, significantly higher rates ($0.47 \times 10^{-6}$ substitutions/site/year) were estimated through direct observation of mutations in families or deep-rooting pedigrees. It is worth noting that within the CR the mutation rate is very heterogeneous, with some 'mutational hot spots' mutating even four or five time as fast as the average site (Excoffier 2002, Hasegawa et al. 1993, Heyer et al. 2001, Meyer et al. 1999, Stoneking and Soodyall 1996, Wakeley 1993). In this sense, the discrepancy between phylogenetic- and pedigree-based estimates could be explained by the fact that pedigree studies preferentially detect fast-evolving site, whose mutations have arisen recently and will not probably become fixed. Differently, phylogenetic studies, that encompass a larger number of transmissions, collect information from mutations that have reached an appreciable frequency in the population, which belong even to slowly evolving site (Heyer et al. 2001, Jazin et al. 1998, Macaulay et al. 1997, Pakendorf and Stoneking 2005, Pakendorf et al. 2003). In this sense, the phylogenetic rate may be preferable for studies of deep history, whereas the pedigree rate is preferentially used for studies of recent history (Macaulay et al. 1997, Pääbo 1996).

As the present work deals with human phylogeny, the phylogenetic rate of mutation will be used. Furthermore, it is worth noting that the fast mutation rate that characterize in general mtDNA, makes it a good genetic system for studying recent micro-evolutionary events like human origins.

## 1.5 MOVEMENTS OF PEOPLE IN NORTHERN AFRICA: GENETIC EVIDENCE

### The spreading of agriculture in Northern Africa and the Bantu expansion

Archaeologists still debate the origin of agricultural diffusion in Northern Africa: despite a reasonable consensus that it spread from Middle East into Egypt between 9,500 and 7,000 years ago, an earlier autochthonous origin in the Nile Valley and the Sahara itself has been hypothesized. In the Nile Valley, domestication of pigs and goats was firmly established by 5,500 years ago, while in the mountains of Central Sahara, evidence of cattle herding dates back to 8,000 years ago. As mentioned before, Sahara in that period was completely different from the hyper-arid landscape it is today: it was rich in antelope and wild oxen, *Bos primigenius*, the progenitor of modern cattle. It is still open of debate whether the cattle eventually tamed in North Africa were descended from local oxen or were brought from the Near East: mtDNA data support the hypothesis of an independent African domestication (Troy et al. 2001), while microsatellite data on nuclear DNA seem to show a close relationships with Eastern cattle, suggesting an early importation of cattle from South-Western Asia (Hanotte et al. 2002). As mentioned before, a sophisticated farming culture was already practiced by the Pastoral groups in the Sahara when the process of desertification progressively started to modify the territory: since 3,500 years ago, agriculture has been widespread throughout the savannah belt south of the Sahara, with farming communities using the shifting agriculture of woodland soils ('slash and burn').

The term 'Bantu expansion' refers to a complex series of population movements related to the rapid spread of farming economies through South-Saharan Africa. This phenomenon is thought to be associated with the introduction around 2,700 years ago of iron from Northern Africa, west of Egypt or from Egypt itself, (Jobling et al. 2004, Vogel 1994b) in South-Saharan Africa. As a result of this expansion, since 1,300 years ago, stock-breeding, iron working, and other forms of social organization has been widespread through Central and Southern Africa (Jobling et

al. 2004). The term 'Bantu' also clearly refers to the Bantu language, a sub-group of the Niger-Kordofonian family whose distribution is represented in figure 1.12.

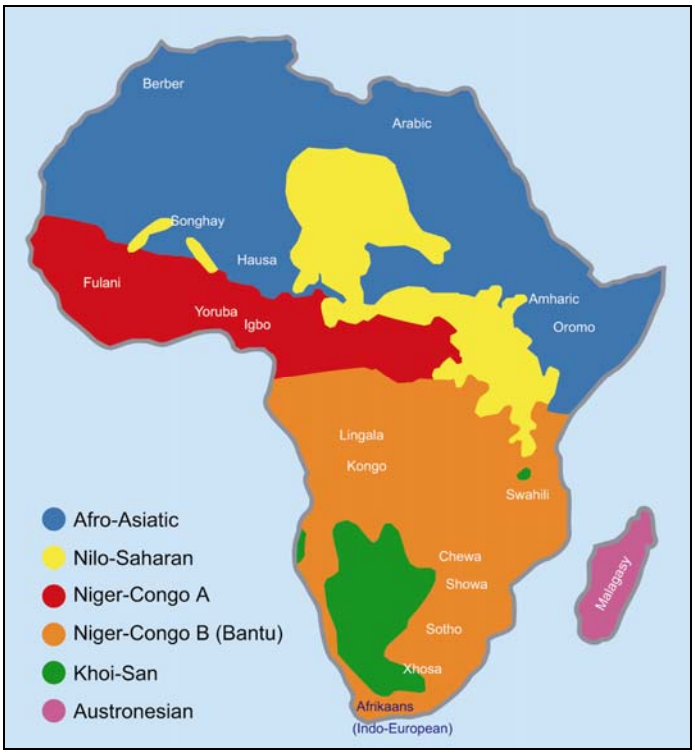


**Figure 1.12:** Map showing the distribution of African language families (according to Joseph H. Greenberg's classification) and some major African languages. Afro-Asiatic extends from the Sahel to Southwest Asia. Niger-Congo is divided to show the size of the Bantu sub-family.

The modern Bantu speakers today number over 200 million people (about 28% of Africans). The Bantu languages are thought to have been co-dispersed with agriculture over much of Central and Southern Africa from a homeland in Western Africa (Holden 2001). Linguistic evidence shows that the 'core area' of the Bantu expansion may be Eastern Nigeria or Western Cameroon. (Johnstone 1913, Greenberg 1972). Despite archaeologists being doubtful about this hypothesis, a consensus view accepts an origin in the Cross River Valley area of Western-Central Africa (Huffman 1982, Phillipson 1993, Vogel 1994a). Bantu languages fall into two main sub-groups, West and East, that probably resulted from distinct routes of dispersal (Vansina 1995): the 'eastern stream' moved eastward, north of the

rainforest, while the 'western stream' took a route southward, through the rainforest (see figure 1.13).

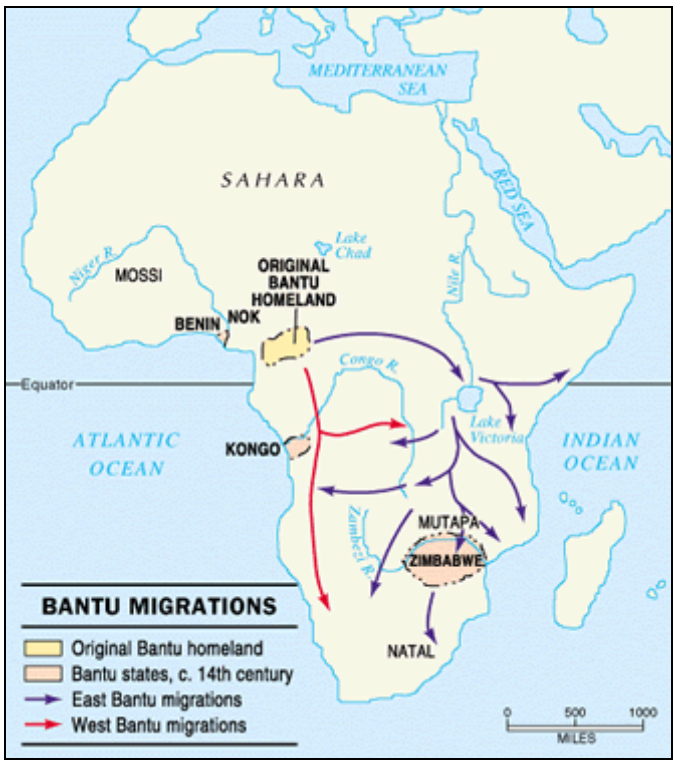


**Figure 1.13:** Hypothetical representation of the Bantu expansion.

The contribution of genetics to the debate about the Bantu dispersal has only recently taken shape. Genetic evidence by Cavalli Sforza (1994) from classical markers, highlights the high genetic affinity that characterize Bantu speakers from different parts of Africa, while some indications of the Bantu expansion are present on the second and fourth principal component of the Principal Component Analysis (PCA) which he performed. More recently, many genetic studies on mtDNA and the non-recombining region of the Y chromosome (NRY), have dealt with the issue of Bantu expansion (Bandelt et al. 1995, Chen et al. 1995, Soodyall et al. 1996, Scozzari et al. 1999, Pereira et al. 2001, Underhill et al. 2000, Underhill et al. 2001, Salas et al. 2002), trying to shed light on this still poorly understood process. In particular, in Salas et al. (2002), it has been shown that Bantu-speaking populations

from South-Eastern Africa are characterised by a high genetic homogeneity, compatible with a very recent common origin. Nonetheless, the high haplogroup diversity, points to a substantial number of people involved in the dispersal and who probably moved by means of a complex series of short-range migrations. Furthermore, it is worth noting that some typical Bantu mtDNA lineages, have even been observed in some areas north of the equator. This has led to the hypothesis that, as part of this Bantu expansion, some human groups moved northward reaching Ghana, Nigeria, Burkina Faso and Mauritania (Rosa et al. 2004).

## From Eastern Africa to the Chad basin.

The inhabitants of the Savannah/Sudan belt of Africa are well differentiated from the Pygmies and the Khoisan people to the South, and from the Berbers and the Arabs to the North (Excoffier et al. 1987, Cavalli-Sforza et al. 1994), but their mutual relationships are still unclear. Archaeological documents show that human movements in Central Africa are closely related to the changing environment of Lake Chad in the past (Connah 1981, Maley 1981, David and Mac Eachern 1988). In a recent paper by Cerny et al. (2004), four different populations from Cameroon were analyzed at the mtDNA level. The samples belong to Chadic-speaking populations, settled in the region of the lake Chad, in Northern Cameroon. They appear to be genetically homogenous and their pattern of mtDNA variability place them close to Eastern populations, in particular Kikuyu, Turkana, and Somali. An explanation for such results was attributable to the relatively prolonged isolation of the Western populations (e.g. Serer, Wolof, Mauritanians), whose contacts with central African people were limited, since the Early Holocene (10,000 years ago), by the almost impenetrable swamp of the Middle Delta of the Niger, and later in the Holocene by the Sahara desert. On the other hand, the lake Chad environment and the surrounding plains may have attracted different people from the East as well from the near West. A secondary explanation proposed by the authors, was the establishment of the Maghreb state from the 11$^{th}$ century onwards, so strenghtening the relationship between West Saharan and Berbers, limiting the southern contact with Central African populations.

Linguistic reconstruction of the Afro-Asiatic phylum shows that the Cushitic language is a *sister group* of the Chadic one. Their common ancestor probably inhabited the Middle Nile area around 6,000 years ago (Blench 1999). Subsequently, a split took place which led some populations to move southwards, contributing to the diversification of the Cushitic languages, while the second stream ventured westward, making the diversification of the Chadic languages in the Chad basin possible. Archaeological data seem to trace the westward route through Sudan along the Wadi Howar (Keding 1993, Blench 1999). All together these data suggest that probably a migration of Pastoral groups from the confluence of the White and Blue Nile through Wadi Howar and Wadi Hawach to the South Chad basin took place (see figure 1.14).

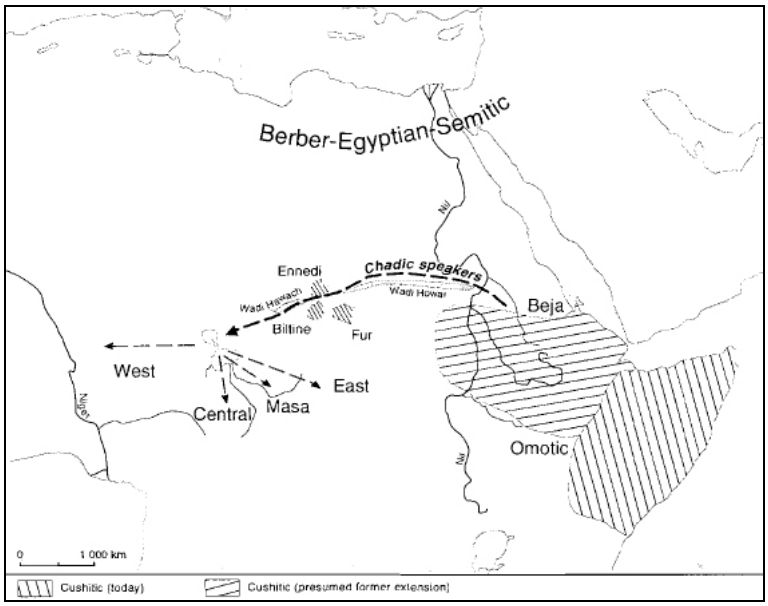


**Figure 1.14:** Hypothetical representation of the westward flow of Cushitic Pastoralists from Eastern Africa according to archaeological record (Blench 1999).

This westward migration is likely responsible for the genetic affinity observed between the Chadic speaking populations and the Eastern ones. Furthermore, the close genetic relationship of the Pastoralist Eastern Sudanese Beja to the Saharan Tuareg (Cavalli-Sforza et al. 1994), otherwise difficult to explain, may also be a relic of such a connection (Cerny et al. 2004). Data collected in 2004, were

integrated by Cerny et al. recently (2007) with other sequences from populations of the Sahel-Sudan region. Results showed how this area played a determining role as a corridor connecting West and East Africa in the first half of the Holocene.

## Movements of Pastoral groups: the Fulbe

The Fulani (also known as the Foulah, Peulh, Peul, Fulfulde, or Fulbe) are an ethnic group of South-Saharan Africa who live in 17 countries and number almost 30 million people (Cerny et al. 2006). They originally were a nomadic population, but today exist both as nomadic and settled populations, the latter cultivating sorghum and raising domesticated animals. Pastoral activity is at the basis of the Fulani nomads' cultural identity, their life-style finds its roots in the old pastoral groups inhabiting Northern Africa in the second half of the Holocene. According to archaeological documents, their origin is placed in the Central Sahara about 5,000 years ago (Dupuy 1999), or even earlier in the Neolithic (Ba and Dieterlen 1966). The modern nomadic Fulani live in the middle savannah belt, from Eastern Senegal to the Central African Republic, and represent the most numerous nomadic group in this area. Linguistically they belong to the Atlantic branch of the Niger-Congo family (Ruhlen 1987).

Although, in the 1930s, a Eastern African origin, even Middle Eastern, was proposed for Fulani (Tauxier 1937), the first molecular evidence about Rh system linked them to West African groups (Excoffier et al. 1987). Recently, in a study by Cerny et al. (2006), four Fulani nomad populations from the Chad basin and West African savannah were sampled and genetically analysed at the mtDNA level. Results show a clear and close genetic relationships among all the samples, irrespective of their geographic origin. Furthermore, they differ from practically all the neighbouring populations. Most of the haplotypes (79.6 %) belong to haplogroups of West African origin, while no close genetic relationship were observed with Nilotic populations. Admixture analysis, however, does not completely rule out the possible ancestral role for Eastern populations from the Nile area, as they present a high admixture coefficient.

The issue of Fulbe's origin appears be even more complicated if we look at some recent results of NRY and mtDNA (Cruciani et al. 2002, Coia et al. 2005). It was observed that populations from Northern Cameroon, including Fulbe, had high frequencies of a Y-chromosome haplogroup of hypothetical Asian origin, R1*-M173 (Cruciani et al. 2002), despite this lineage itself being less frequent from Northern to Equatorial Africa. Salas et al. (2002) hypothesized that R1*-M173 was introduced from the Middle East concurrently with the maternally inherited lineage U6, through movements of Berber populations between 20,000-50,000 years ago (Rando et al. 1998, Macaulay et al. 1999). Then, Fulbe, or also other Pastoralists groups, served to carry these lineages as far as Northern Cameroon by mixing with Berbers in Northern Africa. This hypothesis is based on the presence of R1*-M173 in the Fulbe from Northern Cameroon, and on the evidence that the Fulbe who settled in Nigeria show traces of gene flow of maternally transmitted markers from Berbers, namely sequences belonging to the haplogroups U6 and H (Watson et al. 1997). Nevertheless, no counterparts to R1*-M173 in terms of maternally transmitted lineages were observed in Fulbe from Cameroon (Coia et al. 2005), while haplotypes assigned to U5 haplogoup, close to Moroccan, Tunisian and Saharawi populations were observed, suggesting some kind of contact with these North-Western African people. These results show that even geographically neighbouring Fulbe populations (i.e. Nigeria and Cameroon), can be significantly heterogeneous, and that the dynamic of movements that characterize this nomadic population and other Pastoral groups is quite complex. Furthermore, it is worth noting that often demographic processes like founder effects, selective mating, and drift, can strongly affect the genetic composition of populations, especially nomad people, whose demography in the past is notoriously intricate.

## People from the Iberian peninsula

Recently, genetic characterization on both maternally (Rando et al. 1998, Torroni et al. 2001a Plaza et al. 2003, Achilli et al. 2004, Achilli et al. 2005) and paternally transmitted genetic markers (Bosch et al. 2001) of North-Western African populations, revealed some kind of relationship with South-Western European

people. The main demographic process that hypothetically is responsible of this genetic affinity, has been thought to be the Postglacial Recolonization of Europe from the Franco-Cantabrian refuge, about 15,000 years ago. During the Late Glacial Maximum (LGM), the two major refuges were in South-Western France/Cantabria, and Ukraine/Central Russia Plain, but other minor refuges could have existed (Dolukhanov 2000). Radiocarbon dating attests the successive recolonization of Northern Europe from glacial refugia (Housley et al. 1997), and genetic data from present day populations on both mtDNA (Torroni et al. 1998, Torroni et al. 2001a, Achilli et al. 2004, Achilli et al. 2005) and Y-chromosome (Malaspina et al. 2000, Semino et al. 2000) helped to shed light on this issue. More particularly, recent genetic data collected by sequencing of the whole mtDNA (Achilli et al. 2004, Achilli et al. 2005) clearly indicate that European hunter-gatherers also moved southward and, by crossing the Strait of Gibraltar, introduced U5b1b, H1, H3, and V mitochondrial lineages into North Africa. In fact, as reported in figure 1.15, most of these haplogroups display frequency peaks centered in Iberia and surrounding populations, including Berbers of Morocco.

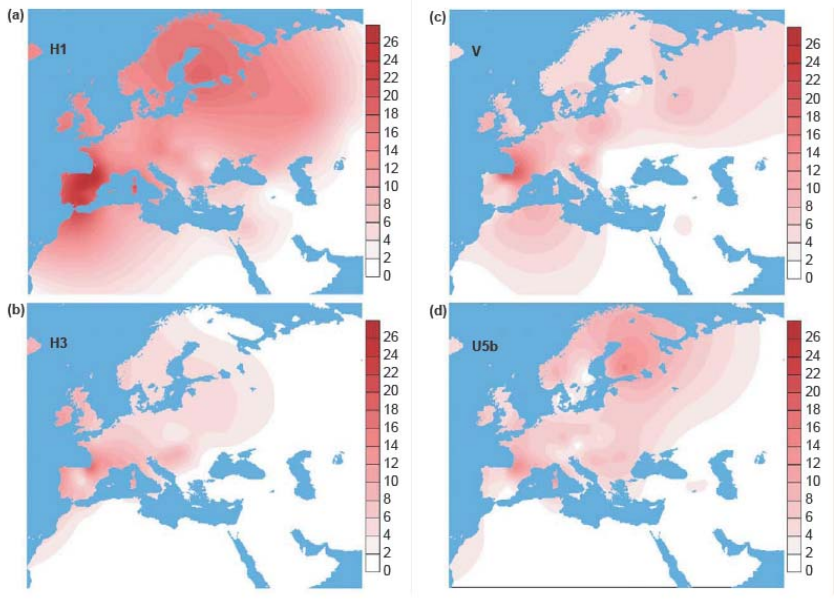


**Figure 1.15:** The spatial frequency distribution in Europe (excluding Saami because of their outlier haplogroup frequencies) and North Africa of haplogroups H1 (a), H3 (b), V (c), and U5b (d) (Torroni et al. 2006).

It is also worth noting that some U5b1b complete-mtDNA lineages, unequivocally link Berbers to Saami populations from Scandinavia; such a close genetic relationship between distantly-related populations in terms of geography, was explained as the result of the bidirectional expansion process of recolonization westward and southward from the South Western Europe refuge (Achilli et al. 2005). The coalescence ages of H1/H3 and U5b1b lineages, that date back respectively to 11,000 and 9,000 years ago, clearly support the scenario of such Post Glacial recolonization (Achilli et al. 2004, Achilli et al. 2005).

## General pattern of mtDNA variability in Northern Africa

Since the first genetic studies in North African populations, a clear contrast with the South-Saharan populations has been observed. Providing a general description of North African mtDNA genetic pool is somewhat daring, as from East to West, the genetic variation of populations has been shaped by many demographic events, both in prehistoric and historic times. The Sahara desert certainly represented a barrier to gene flow since the second half of the Holocene, and could catalyse isolation events, but at the same time the existence of many nomadic groups led different ethnic groups to be in contact each other and sometimes also to mix. In this sense demographic processes like founder effects, admixture events, or even drift in small isolates, could have shaped the genetic composition of a wide range of different local geographic entities, from populations to tribes.

Data collected so far provide a general portrait of North Africa, in which on a West-Eurasian genetic substrate, introductions of South-Saharan lineages took place differentially from East to West. Providing such a description is somewhat reductive but it can help to furnish a general trend.

As regards Northeast African populations, Egyptians have been widely described. Data collected so far show that the West Eurasian component is represented by haplogroups H, T, J, K and X, at frequencies even superior to 30% (Krings et al. 1999, Stevanovitch et al. 2004). Haplogroup M1 is significantly represented in Ethiopians, but is widespread in Northeast Africa too. It is present in Ethiopia at frequencies of about 20% (Passarino et al. 1998), and declines north-westwards

(Nubians 10%, Egyptians 8%) (Krings et al. 1999), whereas its frequency in the Middle East is lower (3% in Jordan, 2% in Palestinians, 2% in Druze) (Macaulay et al. 1999). Furthermore, M1 frequencies significantly diminish westward (where it reaches frequencies of 3-4%) (Rando et al. 1998, Plaza et al. 2003), and also going South of sub-equatorial areas. Recently in Upper Egypt, high frequencies of M1 (17%) have been found, pointing to a close genetic relationship with Ethiopia (Stevanovitch et al. 2004).

A general tendency to a northwards decreasing cline distribution of South-Saharan lineages can be observed in Northeast Africa, in particular for what concerns L3 lineages: 37% in Sudan, 18% in Nubia and 9% in Egypt. This trend in frequencies may be due to the role of the Nile river Valley as a migration corridor that favoured gene flow between populations (Krings et al. 1999).

Like in the Northeast, Northwest African populations are characterized by a West Eurasian genetic substrate that reaches the highest frequencies in Northern countries (e.g. 82% in Moroccan Berbers, between 55-60% in Algeria, Tunisia, and Saharawi; Plaza 2003), decreasing southward, where South-Saharan lineages are more frequent (e.g. 27% in Algeria, 34% in the Saharawi, 45% in Mauritania; Gonzalez et al. 2006). Nonetheless, it might be thought that most of the sampled and genetically characterized populations in Northwest Africa are nomads (moreover Berbers and Fulbe), and that their particular history, cultures, and customs may have significantly affected locally their genetic composition. It is the reason why it would be quite difficult to provide a general view of this part of Africa, while focusing on single populations or even tribes would certainly furnish a better understanding of the relationships between neighbouring people.

A significant presence of lineages from the Iberian peninsula is observed, in particular in the furthermost coastal fringe of Northwest Africa (e.g. V at 10% in Moroccan Berbers from Souss, and 17% in Saharawi; Brakez et al. 2001, Plaza et al. 2003). As previously described, the introduction of these lineages, i.e. U5b, V, H1, and H3, is hypothetically related to the Post Glacial Recolonization from the Franco-Cantabrian refuge which took place about 10,000 years ago (Torroni et al.

2001a, Achilli et al. 2004, Achilli et al. 2005), and their distribution follows a southward decreasing cline up to Guinea (e.g. only U5b at 2.5%; Rosa et al. 2004).

In many Northwest African populations, a significant component is represented by haplogroup U6. It is largely distributed in the Mozabites Berbers (28.2%) (Macaulay et al. 1999) and Mauritanians (20%) (Rando et al. 1998, Gonzalez et al. 2006). Lower frequencies of U6 are present in Tunisians (4.2%) and Moroccans (8%) (Rando et al. 1998, Plaza et al. 2003). It is worth noting that U6 is mainly found in Northern Africa but was also observed in Eastern Africa (Ethiopia 3%) (Kivisild et al. 2004). Recently, complete mtDNA sequencing of U6 and M1 haplotypes allowed to shed light on the phylogeny of these two lineages (Olivieri et al. 2006, Gonzalez et al. 2007). Both of them are predominantly North African clades that originated in Southwest Asia and spread together to North Africa about 40,000 to 45,000 years ago. This demic diffusion took place concurrently to the peopling of Europe by modern humans, and was most likely related to the same change in climatic conditions that allowed humans to enter the Levant from Eastern Africa during the 'Out of Africa' dispersal, between 55,000 and 85,000 years ago (Forster and Matsumura 2005), opening the way to the colonization of both Europe and North Africa. The Early Upper Paleolithic populations carrying M1 and U6 lineages probably followed the Mediterranean coastal route, and then spread differentially into Eastern and Northern Africa.

The last and permanent Islamic expansion may have had a strong cultural influence and possibly even a genetic impact, especially on Berber groups. In fact, the Arab expansion largely submerged the original Berber language and customs, except for the tribes that were forced back to the mountains and some villages in Southern Tunisia. Furthermore, many Arab tribes themselves, such as Bedouin, have coexisted with Berbers since their arrival at the beginning of the 2$^{nd}$ millennium (Murdock 1959, Rando et al. 1998). Notwithstanding this, a recent mtDNA study on Tunisian Berbers (Fadhlaoui-Zid et al. 2004), suggests that Arabization of the area was mainly a cultural process, rather than a demographic replacement of Berber populations who inhabited the area where the Arabic expansion took place. It is worth noting that the Arabic presence in Northwest Africa increased the

introduction of South-Saharan L3 lineages, hypothetically as a result of the slave trade (Gonzalez et al. 2006). In figure 1.16 a recent representation of a African mtDNA evolutionary tree, and of the distribution of the main lineages is reported (Gonder et al. 2007).

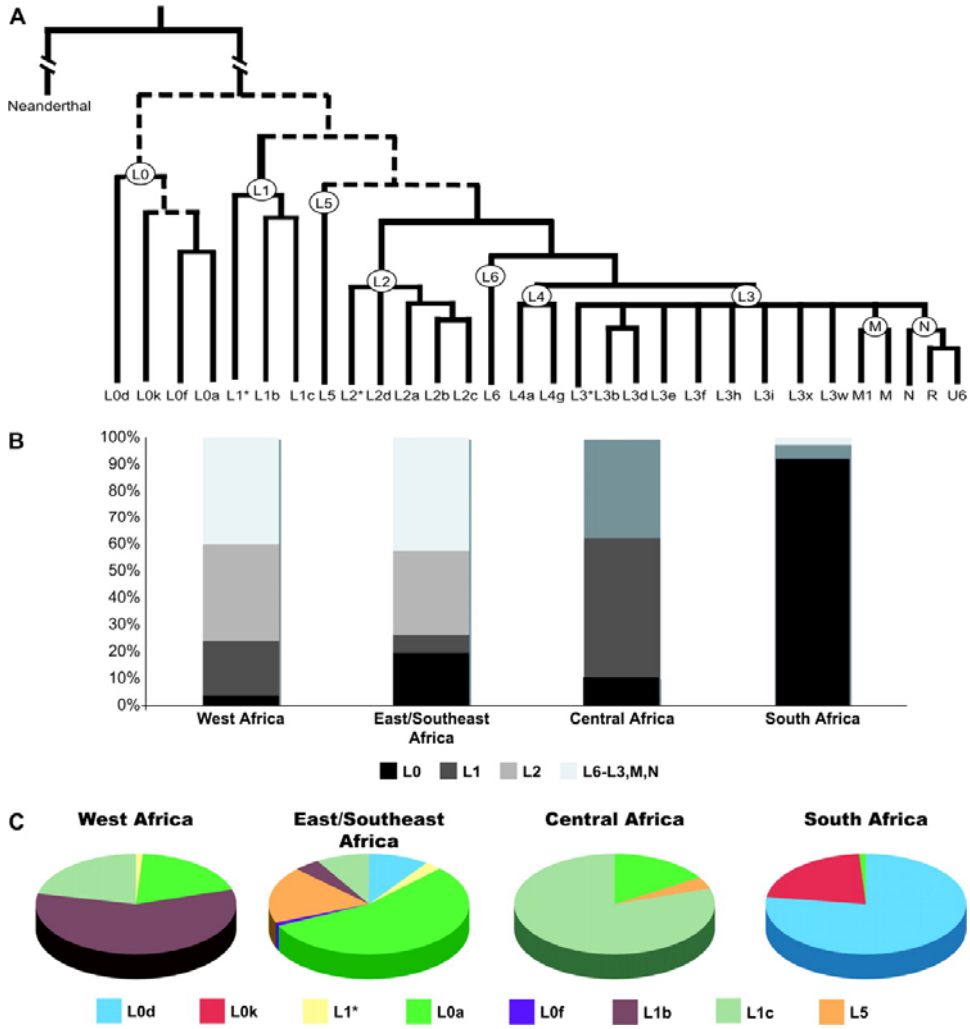


**Figure 1.16:** Evolutionary history of mtDNA haplogroup structure in African populations inferred from mtDNA D-Loop and RFLP analysis. (**A**) Relationships among different mtDNA haplogroup lineages inferred from mtDNA D-Loop sequences and mtDNA coding region SNPs from previous studies (Kivisild et al. 2006). Dashed lines indicate previously unresolved relationships. (**B**) Relative frequencies of haplogroups L0, L1, L5, L2, L3, M, and N in different regions of Africa from mtDNA D-loop and mtDNA coding region SNPs from previous studies. (**C**) Relative frequencies of haplogroups L0, L1, and L5 sub-haplogroups (excluding L2 and L3) in different regions of Africa from mtDNA d-loop and mtDNA coding region SNPs from previous studies (see Gonder et al. 2007).

## Tuareg: state of the art of genetic studies.

Genetic data collected so far on Tuareg are quite scarce. In his work on classical genetic makers, Cavalli Sforza (1994) highlighted a high genetic affinity between Tuareg and Eastern Africa populations from Ethiopia, in particular the Beja. They are a nomadic Pastoral population speaking a Cushitic language and who probably have inhabited for a long time (hypothetically even since 4,000 years ago) the area of Northern Sudan that they presently occupy, between the Nile and the Red Sea. As Cavalli Sforza noticed, the higher genetic relationship of Tuareg with Eastern Ethiopian people, rather than with Northern Berbers, hypothetically points to their common origin. In detail, genetic differentiation between Tuareg and Beja, would date back to 5,000 years ago, when Tuareg could have moved westward. In this sense, the linguistic differentiation represents a secondary acculturation factor, related to the influence that Berbers had on Tuareg once they arrived in Northern Africa. Both Tuareg and Beja genetic pool seems to be the result of an admixture between South-Saharan and West Eurasian genes. Cavalli Sforza hypothesized that the West Eurasian component is associated with repeated genetic contacts between Arabia and Eastern Africa and which would have taken place in the last 5,000-6,000 years, before the Tuareg-Beja split.

Even the hypothesis of a middle-eastern origin of Beja cannot be totally ruled out. Further introduction of South-Saharan lineages in the Tuareg might be the result of slavery practices. It is worth noting that Cavalli Sforza's analysis was carried out on samples including tribes distributed on a wide area. This fact, according to the author, would exclude the influence of drift on the data collected.

As regards mtDNA, very little is known about Tuareg. In fact, only few haplotypes have been collected so far. In 1996 Watson et al. analysed 23 individuals, coming from Mali, Niger and Nigeria. One haplotype more was collected by Gonzales et al. (2006) in Mali. Comparison with other Northwest African populations (Rando et al. 1998, Gonzalez et al. 2006) showed a high affinity of these Tuareg lineages with Senegal, in particular due to the high proportion of South-Saharan lineages (L1, L2, L3). Geographic proximity can hypothetically explain such a relationship. For this

reason, in the present study we refer to the Tuareg sample by Watson as 'Western Tuareg'.

## 1.6 AIM OF THE PRESENT STUDY

The present work aims to trace the origin of Libyan Tuareg inhabiting the Fezzan (Libya), through the analysis of mtDNA lineages in two samples from the region of Tahala, near the Acacus massif. In detail, individuals were sampled in two villages: Tahala and Al Awaynat, and the sample sizes are respectively 111 and 18. The particular topics of mtDNA, makes it one of the most effective instruments for phylogeographic analyses, and particularly for investigating human origins. In fact, analysis at the mtDNA level has been widely used for 25 years for studies of human evolution, migrations, and population histories. It is worth noting that genetic data on mtDNA collected so far on the Tuareg are almost scarce (Watson et al. 1996), and this study appears to be the first with such a high number of mitochondrial DNA sequences from Tuareg.

Most of the knowledge on the Tuareg population is based on ethnical studies and few historical chronicles. Furthermore, much archaeological information has been collected about history and prehistory in the Central Sahara, particularly concerning the Acacus region (Mori's studies and Italo-Libyan joint mission's surveys), but it is always hard to relate Tuareg origin with past human dynamics in the Central Sahara. Some oral and written historical sources hint at the origin of the Tuareg: tales that are transmitted through generations in many villages point to the Arabic peninsula as the place were Tuareg are from, but it should be noted that Islamic populations generally claim the provenance from Muhammad's land, and a direct relation with his lineage. Oral traditions in the villages of Fezzan, attribute to Tuareg a direct descent from Garamantes, whose presence in the Central Sahara is attested since 2,500 years ago. Nevertheless, the origin of Garamantes and their relationship with Pastoral people inhabiting Sahara during the second half of the Holocene are fairly unknown. In the present study, the genetic analysis of ancient

individuals dating back to Pastoral and Garamantian period is included. In fact, in the last two decades, ancient DNA analysis (aDNA) has played an important role for the reconstruction of many past human migratory dynamics, e.g. the peopling of Europe (Haak et al. 2005). Nevertheless, the molecular techniques carried out in the present study, on each occasion failed to recover DNA from the bones and teeth samples of human skeletons from the Acacus, probably because of the high temperatures of the environment that prevented DNA from being preserved at least for the last 2,000 years. For this reason, it is impossible to look at the genetic relationships between Pastoralists and Garamantes, and moreover to test the descent of Tuareg from Garamantes. Hypotheses about Tuareg origin will be inferred from genetic data of modern samples, and the integration with archaeological, ethnological, and historical data will tentatively help to shed light on the past of this population.

In order to find out the relationships between Tuareg and other populations from the African and the Euroasiatic continent, 5,064 HVS-I mtDNA sequences have been collected from literature, and comparisons at the population level with multivariate exploratory techniques (e.g. Multidimensional Scaling, and Correspondence Analysis), as well as at the haplotype level through phylogenetic methods (e.g. Network analysis) have been performed. This comparative analysis makes it possible to test on the maternal side the Cavalli Sforza's hypothesis about Tuareg origin, which he attributed, on the basis of nuclear markers analysis, to Eastern Africa about 5,000 years ago. More in detail, results of this study tentatively helps to find a connection between Tuareg and the movements of people that took place in the Sahel-Sudan corridor, and in the Sahara as well, throughout the whole Holocene. Particular attention is given to the genetic relationship with other North African nomad populations, as well as neighbour settled ones, so as to look at any degree of contact between them and the Tuareg.

As mentioned before, genetic data so far collected on the Tuareg are rather scarce (Cavalli-Sforza et al. 1994, Watson et al. 1996). Nonetheless, integration of these data as a whole, would tentatively help to characterize the Tuareg as an ethnic group, and to unravel their origin. What DNA tells us about Tuareg's genetic

composition and origin, represents an important instrument to test whether the Tuareg in Northwest Africa share a common origin, attributable to a particular geographic area, or they are the result of ancient migratory dynamics that took place in the changing Saharan environment in the second half of the Holocene, these hypothetically leading different nomadic groups to be in contact each other and to mix.

Finally, it should be noted that molecular data on the Tuareg sample collected in the present study, apart from representing the first genetic study strictly focused on Tuareg, open an important insight into a region of Africa that is almost unknown, at least from the genetic side: the Central Sahara. In fact, as depicted in Figure A1, no other samples exist in literature from this region, and even in the surrounding areas. Of course, it might be said that Sahara is almost uninhabited. Nonetheless we think that all the small local ethnic groups, both settled and nomadic, that nowadays occupy this region, represent an important source for collecting information about the dynamic of human migrations in the Sahara, and from a wider point of view, in all Northern Africa.

# Chapter 2 Material and methods

## 2.1 SAMPLES

In the present study both ancient and extant samples were analyzed. Geographic location of the sampled areas is reported in figure 2.1

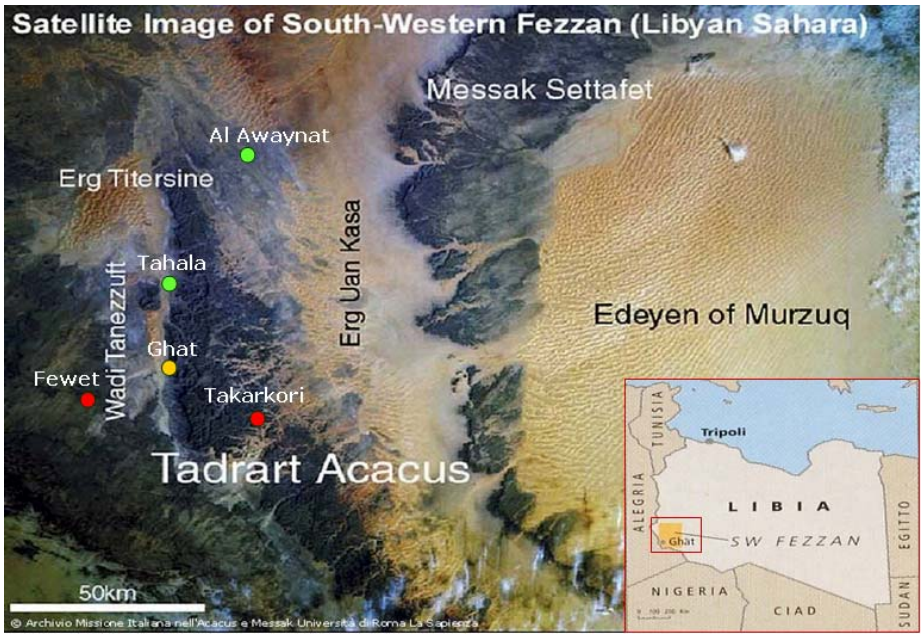


**Figure 2.1:** Geographic distribution of sampled areas. Red dots: ancient sites, Green dots: extant sites, Yellow dot: significant inhabited area. (Italian Archaeological Mission in the Acacus, University of Rome "La Sapienza").

### Ancient samples

Bone samples were taken from 18 skeletons recovered in two archaeological sites: Takarkori, that can be roughly characterized as a Pastoral site, and Fewet, a Garamantian site (Table 2.1). Excavations were conducted recently (between 1997 and 2004) by the 'Joint Italian-Libyan Archaeological Mission in the Acacus and Messak', directed by Prof. Savino Di Lernia, in collaboration with Prof. Giorgio

Manzi. Skeletons were collected in museum 'Sergi' at the University of Rome 'La Sapienza', and the samples used for the genetic analysis were not washed and chemically treated. These were carefully handled by only one Italian archaeologist who wore a disposable facemask and latex gloves. In this way, any source of DNA contamination due to handling of the remains was reduced. Working on freshly excavated bones, under strictly controlled post-excavations procedures, appears to be a key factor for reducing contaminations and degradation of aDNA (Pruvost et al. 2007). Long bones were the preferential material selected for sampling. All the samples were successively stored at –20°C in the laboratory of Molecular Anthropology at the University of Rome 'Tor Vergata'.

**TABLE 2.1**
List of samples analyzed in the present study. Uncalibrated radiocarbon dating is shown in years B.P. where available.

| ID Sample | C14 dating | Site | Cultural phasis |
|---|---|---|---|
| TK H1 | 6090+/–60 | Takarkori | Middle Pastoral |
| TK H3 | - | Takarkori | Pastoral |
| TK H4 | 6740+/–70 | Takarkori | Early Pastoral |
| TK H5 | 6540+/–70 | Takarkori | Early Pastoral |
| TK H6 | 6900+/–70 | Takarkori | Early Pastoral |
| TK H7 | 7130+/–70 | Takarkori | Late Acacus |
| TK H8 | 7973+/–45 | Takarkori | Late Acacus |
| TK H9 | 5600+/–70 | Takarkori | Middle Pastoral |
| TK H10 | - | Takarkori | Pastoral |
| TK H11 | - | Takarkori | Pastoral |
| TK H12 | 7155+/–65 | Takarkori | Late Acacus |
| TK H13 | 4291+/–50 | Takarkori | Late Pastoral |
| TK H14 | 7327+/–65 | Takarkori | Late Pastoral |
| TK H15 | - | Takarkori | Pastoral |
| FW T1197 H1 | - | Fewet | Garamantian |
| FW T399 H1 | - | Fewet | Garamantian |
| FW T715 H1 | - | Fewet | Garamantian |
| FW T1261 H1 | - | Fewet | Garamantian |

### Extant samples

Two samples of healthy and maternally unrelated individuals of ascertained Tuareg descent, were collected from Fezzan. A total of 129 individuals were genetically analyzed: 111 from the village of Al Awaynat, and 18 from the village of Tahala. Appropriate informed consent to anonymously use their data was obtained from all individuals. Sampling was carried out by Prof. Olga Rickards and Prof. Gianfranco Biondi in 2004.

## 2.2 PRECAUTIONS REGARDING CONTAMINATION DURING ANCIENT DNA ANALYSES

Generally, archaeological samples contain only small amounts of DNA that is often damaged. This makes undertaking a series of precautions necessary to minimize the risk of amplifying and typing modern contaminating DNA molecules.

Only Italian archaeologists and paleoanthropologists were in proximity of the human remains, and were genetically screened for their mtDNA haplotype in order to detect any source of contamination on the bone material. As previously mentioned, the samples used for the genetic analysis were not washed but were carefully handled by only one Italian archaeologist wearing a disposable facemask and latex gloves. To maximize the probability of extracting and sequencing authentic DNA from the ancient samples, the strictest available criteria (Cooper and Poinar 2000, Pääbo et al. 2004) were followed. Standard contamination precautions were employed for the ancient DNA analysis: pre-PCR work was carried out in a UV irradiated (254 nm) exclusively dedicated laboratory; DNA was extracted in a UV-irradiated laminar flow cabinet; workbenches were cleaned daily with 1N HCl, bleach and DNA*zap*™ (Ambion, Austin, TX); personnel wore clean-room overalls, disposable hats, shoes, face masks and latex gloves. All materials entering the room were extensively washed with bleach and then UV-irradiated. All items used for liquid manipulation, water used for PCR and non-organic buffers, were autoclaved. Entry in the clean room was not permitted if the PCR products had been handled in

the same day. A separate laboratory was used for the extraction of the modern samples. Post PCR procedures of both the ancient and the modern samples were carried out in a separated laboratory.

Other methodological precautions were taken: at least two independent aDNA extractions of each sample were performed, mock-extraction controls were carried out, and at least two independent amplifications for each investigated region were performed on each extract. The amplification products of all the mtDNA regions were cloned.

## 2.3 DNA EXTRACTION

### Ancient samples

The outer layer of the bone was scraped with a sterile razor blade in order to remove potential contamination from previous handling, and then each side was exposed to UV light (254 nm) for 30 minutes. The tissue was pulverized with a mortar and pestle. Bone powder (300-500 mg) was dimineralised with EDTA [0.5 M, pH8] for three days at room temperature. The EDTA was removed and stored at -20 °C. The residual fraction was incubated at 37 °C for three days in 7.5 ml of extraction buffer (0.5 M EDTA, pH 8, 1 M Tris-HCl, 10% SDS; 2 mg proteinase-K). The supernatant was removed, and a phenol/chloroform protocol was followed for DNA extraction. The EDTA fraction was submitted to phenol/chloroform extraction as well. Then, the EDTA fraction and the residual fraction were pooled, and DNA was concentrated using a Centricon-30 ultrafiltration device (Millipore, Billerica, MA), purified with silica (Qbiogene Inc, Carlsbad, CA) and suspended in a 20-50 µl volume of autoclaved dd$H_2O$. Extracted DNA was preserved at –20°C. At least two extractions were performed for each individual. When satisfactory reproducible results could not be obtained in two different extractions, a third or even a fourth extraction was performed.

### Extant samples

Total genome DNA was extracted from mouth swab samples. These were incubated over night at 56°C, in 300 µl of extraction buffer (10mM Tris-HCl pH 8, 100mM NaCl, 10mM EDTA pH 8, 2% SDS) and 7,5 µl of Proteinase K (20mg/ml) (Budowle et al. 2000). After incubation, the swab was removed and a phenol/chloroform protocol was followed for DNA extraction. DNA was concentrated using a Microcon-100 ultrafiltration device (Millipore, Billerica, MA), and resuspended in 100 µl of sterile water. Extracted DNA was preserved at –20°C.

## 2.4 ANALYSIS OF PHYLOGENETICALLY INFORMATIVE REGIONS WITHIN mtDNA

### Analysis of HVS-I and HVS-II in the ancient samples

Amplification through PCR of the two hypervariable segments of the mtDNA, the HVS-I, from np (nucleotide position) 15996 to 16401, and the HVS-II, from np 00029 to 00408, was carried out in a 50µl reaction volume (Rickards et al. 2001). For details about the reaction conditions see table A1a. The HVS-I and the HVS-II were amplified respectively with four and two overlapping primer pairs (see table A2a). Cycle conditions consisted of an initial denaturation at 94°C for 15 min, 35 cycles of 30 sec at 94°C, 30 sec at 58°C, 30 sec at 72°C, followed by a final extension at 72°C for 10 min. PCR products were visualized by gel electrophoresis on a 2% agarose gel stained with ethidium bromide. Whenever the first PCR did not provide positive amplification, a second amplification attempt was carried out by performing two rounds of nested PCR for each fragment. Twenty-five cycles were carried out for each round of PCR and the aforementioned cycle conditions were used. Whenever even the second attempt provided no amplification, the sample was not further analysed. Samples which gave negative results in both the extractions were considered to be too damaged for genetic analysis.

Positive amplification products were purified through the standard ethanol precipitation technique. Sequence data were obtained using fluorescent dye labelling and an ABI PRISM 3100 AVANT DNA sequencer (Applied Biosystem, Foster City, CA) following the recommended sequencing kit protocols. Dyed products were purified through the ethanol precipitation technique. Overlapping sequencing of light and heavy strands for each region was performed to make the results more reliable. At least two independent amplifications were performed on each DNA extract, and when necessary even three or four independent amplifications were carried out, in order to get as much reproducibility as possible, and to rule out sporadic contamination episodes. Reproducibility of results is at the basis of the authenticity of ancient DNA results, and results were considered reproducible when at least two independent amplifications in two independent extractions provided the same sequence.

## Analysis of HVS-I, HVS-II and phylogenetically informative coding regions in the modern samples

Amplification through PCR of the HVS-I, from np 15996 to 16401, and HVS-II, from np 00029 to 00408 (see table A2c), was carried out in a 25µl reaction volume. Cycle conditions consisted of an initial denaturation at 94°C for 15 min, 25 cycles of 30 sec at 94°C, 30 sec at 60°C, 30 sec at 72°C, followed by a final extension at 72°C for 10 min (see also table A1c). PCR products were visualized by gel electrophoresis on a 2% agarose gel stained with ethidium bromide. Positive amplification products were purified and sequenced as described for the ancient samples. As dye labelling was performed only with the forward primer (named primer 'L' in the mtDNA analysis), whenever the stretch of poly-C was observed in HVS-I, due to the transition at np 16189, dye labelling with the reverse primer (primer 'H') was accomplished in order to cover the length of the whole region up to np 16401.

When the classification could not be resolved through sequencing of HVS-I and HVS-II, properly selected diagnostic sites for mtDNA phylogeny were analyzed in the mtDNA coding region. List of primers used for each site, reaction conditions,

and cycle conditions are reported in table A2b, A1b and A1d. When possible, Restriction Fragment Length Polymorphism (RFLP) analysis was performed: amplifications products were digested and the resulting fragments were visualized in 12% polyacryilamide gels followed by ethidium bromide staining. All details about the restriction pattern of each enzyme are reported in table A2b. Polymorphic site at np 15904, that is diagnostic for haplogroup V, was screened through sequencing analysis, despite the RFLP analysis through enzyme *Mse*I being commonly used. Amplification products relative to diagnostic sites in the coding region that have been recently described in the mtDNA phylogeny (i.e. sites that characterize H sub-haplogroups; Achilli et al. 2004, Loogvali et al. 2004, Pereira et al. 2005) were submitted to direct sequencing, following the procedure previously described. Amplifications were carried out in a 25µl reaction volume, according to the chemical conditions reported in table A1a. List of primers used and the cycles conditions of the amplification reactions are reported in table A2c and A1c.

## Sequencing of the whole mtDNA in properly selected haplotypes of the modern samples

Some properly selected haplotypes were submitted to complete mitochondrial DNA sequencing. More particularly, a sub-sample of H1 haplotypes observed after the analysis of the HVS-I, the HVS-II and other informative coding regions, were submitted to complete mtDNA sequencing by the researchers of the laboratory of genetic in the Department of Genetic and Microbiology at the University of Pavia, directed by Prof. Antonio Torroni. This analysis allowed to resolve the genealogies of H1 lineages to a higher extent of phylogenetic resolution. In the present work results of the complete mtDNA sequencing on five individuals (three from Tahala and two from Al Awaynat) are reported.

## Ancient samples: cloning of PCR product

As mentioned before, archaeological samples contain only small amounts of DNA that is often damaged. In fact, bone tissues and ancient DNA as well, during depositional time, are subjected to many chemical and physical processes, that are

known as diagenetic damages. Some of these damages result in fragmentation or oxidation of DNA molecules, which both reduce the efficiency of amplifying DNA molecules through PCR. Other damages result in chemical modifications of nitrous bases (i.e. hydrolytic reactions). These latter kind of modifications are known as *miscoding lesions*, as they are responsible for the introduction of wrong nucleotides by the *Taq* polymerase during the amplification process, so that a different sequence from the original one is obtained. The presence of miscoding lesions in some endogenous molecules (i.e. molecules belonging to the bone sample), results in a heterogeneous pool of extracted ancient DNA molecules. Furthermore, it should be noted that even when undertaking all the precautions necessary to minimize the risk of contamination, a certain degree of exogenous contaminant molecules are introduced, so that heterogeneity of the extracted molecules increases. Such a heterogeneous pool of molecules obtained after the extraction process, when submitted to amplification and direct sequencing analyses, results in a electropherogram that is characterized by many heteroplasmies. In order to resolve heteroplasmies, cloning of PCR products was accomplished on every fragment from independent PCRs, and independent extractions as well. In fact, cloning allows to discriminate for any sequence resulting from contamination of exogenous human DNA, or from artifacts arising during the amplification reaction, the latter being due to miscoding lesions and also to *Taq* misincorporations.

Since neither primer dimers nor non–specific bands were visible by agarose gel electrophoresis, the amplification products were directly cloned using a TOPO TA cloning kit (Invitrogen, Carlsbad, CA) according to the supplier's protocol. Positive transformants were directly analyzed by PCR after selecting the colonies and suspending them individually in 20µl of the PCR product; 15µl of each positive PCR product were directly sequenced as reported above. At least 10 clones corresponding to each amplified region were sequenced so as to obtain a statically significant consensus sequence.

## 2.5 ANCIENT SAMPLES: ATTESTING PRESERVATION OF THE BONES

### Thermal Age analysis

Thermal age analysis was accomplished in collaboration with Dr. Colin Smith (Department of Archaeology at the University of Durham). This represents an index of DNA preservation and is based on a mathematical model which assumes that depurination is the principal pathway of DNA degradation in most environments (Lindahl and Nyberg 1972). Apurinic sites represent in fact preferential sites where *single strand breaks* occur, these leading to the fragmentation of the DNA molecule. In the calculation of the Thermal Age, both altitude and the integrated regional paleotemperature data are considered. The expected deterioration in DNA by depurination alone is expressed in terms of its 'Thermal Age', which is the number of years required, at a constant 10°C, to produce the degradation calculated from its thermal history. Bones with Thermal Ages higher than that of Feldhofer Neanderthal (around 20 $_{kyr@10°C}$), were observed to fail to yield DNA (Smith et al. 2001), so that this represents the current technical limit for DNA retrieval. Nonetheless, it should be noted that thermal age is not the only factor that preserves DNA quality for successful amplification, but there are many other parameters to be taken into account, e.g. humidity, soil pH, phosphorus content of the soil, average temperature in different earth layers, microbial mediated decay (Smith et al. 2001).

Thermal age estimates were made following the method described by Smith et al. (2003). Local weather station data for the Fezzan area is scarce, however three stations near the site were used to estimate a likely temperature range for the Holocene. Thermal ages for 'cave sites' are based on the mean annual temperature with no seasonal fluctuation, representing a scenario where there is little temperature fluctuation within caves and deep burial. Thermal ages for 'open sites' include full seasonal variation, representing a scenario where a fossil is exposed to the full extent of seasonal fluctuation, i.e. in exposed shallow burial. Deeper burials in open sites are likely to incur some dampening of the seasonal variation, but the full seasonal variation represents a 'worst case' scenario.

### Racemization analysis

Racemization analysis was performed in collaboration with Prof. Antonella Casoli, in the laboratory of the Department of Inorganic and General Chemistry, Analytical Chemistry and Physical Chemistry, at the University of Parma. This represents a biochemical assay of macromolecular preservation that serves to support the claim that a specimen is well enough preserved to allow preservation of DNA (Poinar et al. 1996, Krings et al. 1997). In fact, the preservation of other organic macromolecules, such as proteins, in ancient specimens may serve as useful proxies for predicting the survival of DNA. Using this rationale, the degree of racemization of aspartic acid (D/L Asp value) have been used to screen samples besides DNA extraction and amplification (Poinar et al. 1996, Poinar and Stankiewicz 1999). Poinar et al. (1996) showed that no DNA sequences could be retrieved from a set of bone samples, dating back from 50 to 40,000 years ago, where the D/L Asp value was higher than 0.08, whereas all samples with D/L values below 0.08 yielded DNA sequences. Nevertheless, it is worth noting that experimental results recently collected in the laboratory of Molecular Anthropology of Rome, about relationship between DNA preservation and aminoacid racemization, suggest that DNA most likely follows a different diagenetic trajectory to proteins (Ottoni et al. 2008, submitted).

Amino acid racemization analysis was carried out on the whole bone fraction of the youngest archaeological samples (i.e. those from Fewet site). D/L ratio values for three aminoacids were calculated (Alanin, Glutamic Acid, and Aspartic Acid).

## 2.6 STATYSTICAL ANALYSIS

### Haplogroup classification

Each mtDNA sequence was phylogenetically classified according to the available literature (Macaulay and Richards 2002, Salas et al. 2002, Kivisild et al. 2003, Achilli et al. 2004, Kivisild et al. 2004, Salas et al. 2004, Achilli et al. 2005, Olivieri et al. 2006).

## Database of HVS-I sequences

For comparative analysis, HVS-I sequences collected from the GENEBANK international database and literature were used. These reference haplotypes belong to samples of extant populations from Mediterranean Europe, the Middle East and Northern and South-Saharan Africa. In particular for the African continent a total of 5,064 HVS-I sequences from 80 population samples, were collected (TableA3, figA1). This comparative survey provided us with valuable phylogeographic information about the lineages characterizing the individuals of our samples. Samples collected from the African continent were divided in geographical groups, namely: Northern Africa, Eastern Africa, Western Africa, Central Africa, Equatorial Africa, Southern Africa[8]. For the multivariate and phylogenetic analyses carried out in this study, populations and haplotypes belonging to each geographical group were graphically represented with different colours. Details about the African samples used for the comparative analyses are presented in TableA3. The complete mtDNA H1 sequences analysed in the present study were compared with other complete mtDNA haplotypes analysed in the laboratory of genetic of the University of Pavia (Achilli et al. 2004).

Update of the database and transformation of the files in the appropriate format for the analyses (i.e. text, fasta, phylip) was accomplished through the program mtDNA 2.4ver, created by Dr. Enrico Fabrini (www.doppiovu3.it/mtdna).

## Analysis at the intra-population level: Standard diversity indices and Molecular indices

Standard Diversity indices and Molecular indices were calculated with the program Arlequin 2.000ver (Schneider et al. 2000). The HVS-I sequences collected in the database were used for the calculation of all the parameters in the pooled Libyan Tuareg sample (i.e. including sequences from both Tahala and Al Awaynat, see Results) and other African populations. The same calculations were repeated for the

[8] As populations south to the Equator appeared to be less informative for the comparative analyses, they were all enclosed in one single group, i.e. Southern Africa, despite their very different cultural and genetic origins (e.g. Bantu and Khoisan populations).

Libyan Tuareg when considering the HVS-I/HVS-II haplotypes. The following parameters were calculated:

K: number of haplotypes. Number of different HVS-I haplotypes observed in the sample.

S: number of polymorphic sites. Number of loci that present more than one allele per locus.

H: Gene Diversity. It is equivalent to the expected heterozigosity for diploid data. It is defined as the probability that two randomly chosen haplotypes are different in the sample. Gene Diversity and its sampling variance are estimated as:

$$\hat{H} = \frac{n}{n-1}\left(1-\sum_{i=1}^{K} p_i^2\right)$$

$$V(\hat{H}) = \frac{2}{n(n-1)}\left\{2(n-2)\left[\sum_{i=1}^{K} p_i^3-\left(\sum_{i=1}^{K} p_i^2\right)^2\right]+\sum_{i=1}^{K} p_i^2-\left(\sum_{i=1}^{K} p_i^2\right)^2\right\}$$

where n is the number of gene copies, K is the number of haplotypes, and $p_i$ is the samples frequency of the i-th haplotype (Nei 1987).

π: Mean Number of Pairwise differences. It is the mean number of differences between all pairs of haplotypes in the sample (Tajima 1983, 1993). It is given by:

$$\pi = \sum_{i=1}^{K} \sum_{j<i} p_i\, p_j\, d_{ij}$$

where $d_{ij}$ is an estimate of the number of mutations having occurred since the divergence of haplotypes i and j, K is the number of haplotypes, and $p_i$ is the frequency of the haplotype i. The total variance (over the stochastic and the sampling process), assuming no recombination between sites and selective neutrality, is obtained as:

$$V(\pi) = \frac{3n(n+1)\pi+2(n^2+n+3)\pi^2}{11(n^2-7n+6)}$$

Nucleotide diversity: or Average Gene Diversity over L loci. It is the probability that two randomly chosen homologous nucleotides are different. It is equivalent to the gene diversity at the nucleotide level (Nei 1987, Tajima 1983).

$$\pi_n = \frac{\sum_{i=1}^{K} \sum_{j<i} p_i p_j d_{ij}}{L}$$

$$V(\pi_n) = \frac{n+1}{3(n-1)L} \pi_n + \frac{2(n^2+n+3)}{9n(n-1)} \pi_n^2$$

## Calculation of 95% Credible Region

We estimated the posterior distribution of the proportion of haplogroups in the populations, given the two samples of Tahala and Al Awaynat, by using a binomial likelihood and a uniform prior on the populations proportion. From this posterior distribution we calculated a central 95% 'credible region' (Richards et al. 2000). Overlapping distributions of frequencies of the main lineages present in the samples, suggest that differences in the haplogroup frequencies observed between the samples are stochastic, and consequently the samples can be pooled in a single one. Bayesian 95% credible regions (CRs) for allele and haplotype frequencies were calculated with the computer program Sampling (kindly provided by V. Macaulay, Department of Statistics, University of Glasgow).

## Analysis at the inter-population level: $F_{ST}$ Genetic distances

$F_{ST}$ genetic distances were calculated between the extant pooled Tuareg sample and all the other African populations collected in the database. Computations of $F_{ST}$ was

accomplished with the software Arlequin 2.000ver (Schneider et al. 2000), and HVS-I sequences for each population were used as input data.

The pairwise $F_{ST}$ can be used as short term genetic distance between populations, with the application of a slight transformation to linearize the distance with population divergence time (Reynolds et al. 1983, Slatkin 1995). The pairwise $F_{ST}$ values are given in the form of a matrix. The null distribution of pairwise $F_{ST}$ values under the hypothesis of no difference between the populations is obtained by permuting 1,023 times the haplotypes between populations. The P-value of the test is the proportion of permutations leading to a Fst value larger or equal to the observed one. The P-value are also represented in matrix form. Slatkin's linearized Fst were used in the present work.

Slatkin's linearized $F_{ST}$ (Slatkin 1995): Slatkin considered a simple demographic model where two haploid populations of size N have diverged $\tau$ generations ago from a population of identical size. These two populations have remained isolated ever since, without exchanging any migrants. Under such conditions, $F_{ST}$ can be expressed in terms of the coalescence times $t_1$, which is the mean coalescence time of two genes drawn from two different populations, and $t_0$, which is the mean coalescence time of two genes drawn from the same population. Using the analysis of variance approach, the $F_{ST}$ values are expressed as (Slatkin 1991, 1995):

$$F_{ST} = \frac{t_1 - t_0}{t_1}$$

As $t_0$ is equal to N generations (Hudson 1990), and $t_1$ is equal to $\tau$ + N generations, the above expression reduces to:

$$F_{ST} = \frac{\tau}{\tau + N}$$

Therefore, the ratio $D=F_{ST}/(1-F_{ST})$ is equal to $\tau/N$, and is therefore proportional to the divergence time between the two populations.

## Multidimensional Scaling (MDS)

Multidimensional Scaling is a dimensional reduction technique, in which $\delta_{ij}$ observed distances between $n$ items are represented in a low-dimensional coordinate system (e.g. two-/three-dimensional), and distances $d_{ij}$ closely match the original distances $\delta_{ij}$ (Rencher 2002), that is:

$$d_{ij} \cong \delta_{ij} \quad \text{for all } i, j$$

The program Statistica 6.0ver was used for the MDS analysis. Tthe goal of the analysis is to detect meaningful underlying dimensions that allow the researcher to explain observed distances between the investigated objects. Therefore, MDS attempts to arrange 'objects' in a space with a particular number of dimensions in order to reproduce the observed distances. The orientation of axes in the 'reduced' representation is arbitrary, since distances between objects remain the same. Nevertheless the calculation is not such an exact procedure, as observed distances are 'rearranged' in the new low-dimensional representation, and the new configuration is only an approximation of the original observed distances. In fact, the program moves objects in the space defined by the required number of dimensions, and checks how well the distances between objects can be reproduced by a new low-dimensional configuration. In more technical terms, the program uses a minimization function algorithm that evaluates different configurations with the goal of maximizing the goodness-of-fit (or minimizing the 'lack of fit'). The most common measure that is used to evaluate how well (or poorly) a particular configuration reproduces the observed distance matrix is the *stress* measure. The raw *stress* value of a configuration is defined by:

$$stress = \sum [\, d_{ij} - f(\delta_{ij})]^2$$

Where $f(\delta_{ij})$ indicates a nonmetric, monotone transformation of the observed input distance $\delta_{ij}$. Thus, the program attempts to reproduce the general rank-ordering of distances between the objects in the analysis. The smaller the stress value, the better the reproduced distance matrix fits the observed distance matrix. Kruskal suggests the following verbal evaluations for various levels of *stress*:

| Stress | Goodness of fit |
|---|---|
| 0.40 | Poor |
| 0.20 | Fair |
| 0.10 | Good |
| 0.05 | Excellent |
| 0.00 | 'Perfect' |

In the present study, matrixes of Slatkin's linearized $F_{st}$s were used as input data, and a two-dimensional plot was used as final representation of distances. Population samples that appear to be *outgroups* in the two-dimensional representation were discarded in order to reduce the stress value.

## Correspondence Analysis (CA)

Like MDS, CA is a multivariate exploratory technique that is used to reduce dimensionality of data, and portrays relationship among observations and variables. It is a graphic technique used for representing the information in a two-way contingency table, which contains the counts (frequencies) of items for a cross-classification of two categorical variables. With CA, a plot is reconstructed showing the interaction of the two categorical variables along with the relationship of the rows to each other and the columns to each other (Rencher 2002). When dealing with more than two variables, Multiple CA (MCA) is performed.

In general, CA computations extract from the observed data new dimensions, so as to maximize the distances between the row or column points. The new dimensions (which are independent, or orthogonal, to each other) will 'explain' less and less the overall information contained in the observed data (i.e., the *inertia*, which is defined as the total Pearson *Chi*-square for the two-way table divided by the total sum). The final goal of CA, is that the first two or three coordinates can be represented in a two-/three-dimensional plot, and the *inertia* enclosed in the new dimensions, indicates how good is the representation of data.

In the present study CA was performed on available haplogroup frequencies of the African populations collected in the database. The program Statistica 6.0ver was used for computations, and data were represented in a two-dimensional plot. As data from literature contain different levels of phylogenetic classification (less or

more 'refined'), in order to standardize the input matrix of frequencies to the level of phylogenetic resolution performed in this study, some samples were discarded (e.g. data collected from old publications, where the extent of mitochondrial phylogenetic resolution was lower than that used in the present study), and some haplogroups were grouped in single variables (table A4).

## Network Analysis and Coalescence Ages

Network analysis is a phylogenetic method, so the analysis is shifted from the population level to the genealogical level. In detail, the Median Joining (MJ) Network identifies groups of closely related types, and introduces hypothesized ancestral types in order to unite the types into a parsimonious tree or network (Bandelt et al. 1999). The parameter ε (epsilon) governs the initial grouping of closely related types, and consequently the chance of recovering Most Parsimonious (MP) trees. The parameter ε specifies a weighted genetic distance to the known sequences in the data set, within which potential median vectors may be constructed. If ε is set less than the greatest weighted genetic distance within the data set, then there is a theoretical possibility that the MJ network will not contain all possible shortest trees. If ε is set equal to (or even greater than) the greatest weighted genetic distance, the MJ algorithm is guaranteed to yield a full median network.

MJ network analysis using Network 4.1 software (Fluxus Technology Ltd., Clare, Suffolk, UK) was carried out to locate most of the HVS-I haplotypes observed in the Tuareg sample in the context of other African lineages present in the database. All the networks were performed at the haplogroup level. The MJ networks were calculated by setting ε=0, in order to reduce run time of the analyses, and to avoid too complex networks difficult to interpret. Furthermore, different weights were assigned to the polymorphic sites according to the transition rate previously described (Meyer et al. 1999), and to the recurrence of mutations in the network, so as to minimize homoplasy. This allowed to resolve high-dimensional cubes and large cycles. In general, events that might be much less likely to happen need to have a higher weight, because they are significant when they happen. By contrast,

weights for events that might be much more likely to happen have to be decreased. A different weight was also given to transversions with respect to transitions (3:1), as suggested in the Network's user manual.

Once defined the networks for the main lineages observed in the Tuareg sample, the coalescence ages were calculated. Coalescence time estimation was carried out for some significant nodes observed in the networks, which hypothetically might be related to demographic processes (e.g. expansion events). The ρ (r*ho*) statistic is the average distance of the derived nodes to the roots (Morral et al. 1994), and it measures the age of an ancestral node in mutational units. This mutational age is then converted into years by multiplication with the mutation rate, that is for human mtDNA in the region between nps 16090-16365, 20,180 years. The standard deviation is defined by the parameter σ (*sigma*) (Saillard et al. 2000). The ρ is inherently unbiased (model-free): past demography will have influenced the shape of the evolutionary tree, but this influences only the error of the time estimate rather than systematically increasing or decreasing the time estimate. After calculating coalesce time, *star likeness* was calculated, in order to attest how rapid was the hypothesized demographic event (i.e. the population expansion). Star likeness is defined as:

$$Star\ Likeness = \frac{\rho}{n\sigma^2}$$

where *n* is the sample size. Star likeness values higher than 0.3, indicate a rapid expansion. Star likeness equal to 1 indicates a 'perfect star phylogeny', in which the tips represent single individuals and the central node harbours the remainder of the samples (Saillard et al. 2000).

# Chapter 3. Results

## 3.1 ANCIENT SAMPLES

### Genetic Analysis

Genetic analysis performed on the bones of all the ancient individuals did not give any reproducible result. The high state of degradation of DNA is hypothetically due to the high temperatures of the Saharan environment, to which the burials were exposed at least in the last 3,000 years. In most of the samples (13/18, i.e. 72.2%) multiple amplifications from independent extractions provided unsuccessful results. Even when the size of the amplicons was reduced, so as to favour the amplification of short fragments to which presumably aDNA is reduced, PCR gave negative results. It was observed that when attempting amplification with two rounds of nested PCR, contaminant DNA was sporadically amplified. In fact, exclusively sequences belonging to technicians working in the laboratory were observed after direct sequencing, and cloning of PCR products confirmed the presence of exclusively contaminant modern sequences (frequency=100% of the clones).

Only in three samples from Takarkori (TK H1, TK H5 and TK H9), and two from Fewet (FW T399, FW T1197), amplification of some HVS-I fragments after the first round of PCR was obtained in independent multiple experiments. Nevertheless, the pattern of mutation observed in the sequences was never clear and reproducible, even after multiple amplifications from three (in TK H5 and FW T1197), or four independent extractions (in TK H1, TK H9, and FW T399). Cloning of the PCR products showed the presence of different sequences, indicating a high heterogeneity of the amplification products. A brief summary of all the cloning experiments that were carried out on the aforementioned samples is reported as follows. The following sequences were found in the clones:

- Sporadic contaminations of laboratory technicians who had been working in the clean room in the same period in which the ancient Libyan samples were analysed. These kind of contaminant sequences were observed in all the cloning experiments, and their frequencies relative to the total number of clones sequenced in each single cloning experiment, ranged from 30% to 70%.
- Two *chimera* sequences were observed in two cases (TK H5, TK H1). It is likely that they were the results of *jumping PCR* (Pääbo et al. 1990, Hofreiter et al. 2001b), between hypothetically fragmented endogenous molecules and contaminant ones.
- Sequences of the Nuclear Mitochondrial Pseudogene (NUMT) on chromosome 11 (Zischler et al. 1995) were observed in two clones (frequency=6%) of the HVS-I fragment L16209-H16331, in the individual TK H5. In fact, this couple of primers appear to have a complete match with the NUMT on chromosome 11. It is worth noting that the clones sequenced in this experiment were mostly contaminant (frequency=80%). A more detailed analysis of the primers which are generally used for mtDNA at the HVS-I level, showed a 100% match between NUMT on chromosome 11 and the following primers:

| **Primers L (length)** | **Primers H (length)** |
|---|---|
| L16159 (19) | H16139 (19) |
| L16209 (19) | H16142 (19) |
| L16347 (19) | H16164 (18) |
| | H16331 (19) |
| | H16410 (20) |

Consequently, the possible combination between these couples of primers in PCR experiments of highly degraded samples, should be carefully used in aDNA analysis.
- Sporadic unknown sequences: they were observed at low frequencies (range 10-40%) in all the samples who provided successful amplifications after the first round of PCR. Hypothetically they were endogenous ancient (presumably damaged) sequences. Nevertheless, the same sequence was never observed in the clones of the independent experiments carried out for each sample (i.e. two extractions for

sample FW T1197, three extractions for samples TK H1, TK H9, FW T399, four extractions for sample TK H5, being the amplifications performed at least twice in each extract). Consequently, these sequences could not be considered as authentic.

## Thermal Age

Thermal Age calculation was accomplished in the ancient samples which were C14 dated. Thermal Age could not be directly calculated for the samples from the Fewet site, as C14 dating was not available. Nevertheless, C14 dating was available for two Garamantian individuals, from two sites located in the Wadi Tanezzuft Valley, not far from Fewet (i.e. In Aghelachem and Aghram Nadharif, data not shown). As the same cultural context and chronological phase was attested for these samples and those from Fewet (i.e. Garamantian), and the three archaeological sites hypothetically shared the same regional paleoclimate, Thermal Ages calculated in the samples from the Wadi Tanezzuft Valley were used as representative of the Garamantian sites, including Fewet.

In Table 3.1 approximate range of Thermal Age calculations of the archaeological sites is reported.

**TABLE 3.1**

Estimated Thermal Age ranges (reported as yr@10°C) of human bones from Garamantian, Pastoral and Late Acacus sites. Values are reported as ranges between the lowest and the highest C14 dating observed in the samples.

| Site | Cultural phase | Nominal Age range | Termal Age (open) | Termal Age (cave) |
|---|---|---|---|---|
| Takarkori | Late Acacus/Pastoral | 5,600-7,973 | 83,000-105,000 | 46,000-57,500 |
| Fewet[a] | Garamantian | 1,700-2,550 | 23,200-55,000 | 12,800-30,100 |

[a] Thermal Age calculations for the Fewet site, were calculated from C14 dating of two samples from a different Garamantian site: RT H1 (In Aghelachem site, 1,700+/–40) and T4 H1 (96/129 site, 2,550+/–190).

Range of values for the samples from Takarkori, are extremely higher than the value of Feldhofer Neanderthal (around 20 $_{kyr@10°C}$), and so they are not compatible with the retrieval of aDNA. Differently, the range of values in the samples from Fewet (under the assumption of analogous conditions with the Garamantian sites of

the Wadi Tanezzuft Valley) is approximately compatible with the retrieval of aDNA, at least under the scenario of 'cave sites'. This scenario can be assumed for the Fewet site, being the burials very well preserved, intact and sealed (Castelli et al. 2005). Burials were dug deeply in the ground and corpses were placed in a sort of chamber delimited by stone plates. Burial environment was dry, without standing water. Furthermore, human bodies were found in leather envelopes: despite being difficult to precisely state what kind of effect it had on the body preservation, it can be hypothesized that it further helped to isolate the corpse from the surrounding environment.

## Aminoacid Racemization

Aminoacid racemization was accomplished on all the Fewet samples, which macroscopically appeared to be well preserved, and whose Thermal Age calculation approximately yielded values comparable with that of Feldhofer Neanderthal (around 20 $_{kyr@10°C}$), at least when assuming the scenario of 'cave' sites. In Table 3.2, D/L ratio values of Ala, Asp, and Glu aminoacids for the whole bone fraction are reported.

**TABLE 3.2**

Results of the aminoacid racemization in the whole bone fraction in four samples from Fewet sites. Asp D/L values below 0.08 are compatible with DNA retrieval from archaeological samples (Poinar et al. 1996)

| | D/L ratio | | |
|---|---|---|---|
| **Sample** | **Ala** | **Asp** | **Glu** |
| FW T1197 | 0.22 | 0.47 | 0.13 |
| FW T715 | 0.119 | 0.50 | 0.119 |
| FW T1261 | 0.21 | 0.37 | 0.115 |
| FW T399 | 0.139 | 0.350 | 0.055 |

It is worth noting that in the present study only the whole bone fraction, which might be affected by contaminant amino acids/peptides, was submitted to the Racemization analysis. Further data need to be collected on the EDTA insoluble

fraction. This contains the bulk of the total bone protein, i.e. insoluble collagen, and therefore should provide a more precise measure of the degree of protein racemization, as soluble non-collagenous proteins and contaminant amino acids/peptides are absent.

## 3.2 MODERN SAMPLES: GENETIC ANALYSIS

It was possible to determine HVS-I and HVS-II sequences from all the 129 Tuareg individuals. Then, according to the pattern of mutation observed in the D-Loop region, analysis of the polymorphic sites in the coding region was accomplished through RFLP, in order to increase as much as possible the phylogenetic resolution. Exclusively for H haplotypes, sequencing of properly selected coding regions of mtDNA was carried out in order to screen some diagnostic SNPs which were previously described in literature (Achilli et al. 2004, Loogvali et al. 2004, Achilli et al. 2005, Pereira et al. 2005, Roostalu et al. 2007). This made possible to get a more refined classification for the haplogroup H, and to increase the resolution at the phylogeographic level. The list of polymorphic sites for each haplotype, and the haplogroup classification is reported in Table A5.

### Description of lineages found in the Tuareg from Fezzan

After the first step of HVS-I and HVS-II sequencing, 63 individuals from Al Awaynat, and nine individuals from Tahala, were characterised by the haplotype CRS-00263. The RFLP typing allowed to attribute all of them to the haplogroup H, which is characterised by the diagnostic markers –7025 *Alu*I, and –14766 *Mse*I. The same RFLP pattern was observed in the haplotypes harbouring HVS-I mutations 16037-16256 (four individuals in Al Awaynat, and two in Tahala), and 16183C-16189 (one individual in Al Awaynat). In particular the latter haplotype was characterised by the reversion at position 00073. Then, sequencing of informative coding regions of mtDNA allowed to screen H haplotypes for the presence of polymorphisms that were characterised as diagnostic for the H sub-haplogroups. All

the sequences submitted to this analysis, differed from the rCRS (that belongs to sub-haplogroup H2) for transitions A1438G, G3010A, and A4769G. Such pattern of mutations allowed to assign all the H haplotypes to sub-haplogroup H1.

Four haplotypes from Al Awaynat and one from Tahala, are characterized by the HVS-I transition 16298 and the HVS-II transversions 72C, that are two diagnostic mutations for haplogroup V (Torroni et al. 1998). Assignment to this haplogroup was confirmed by the mutation at np 15904, and by the marker –4577 *Alu*I (Richards et al. 1996, Torroni et al. 1996, Torroni et al. 1997).

The L0a1 lineage was observed exclusively in eight individuals from Al Awaynat. This is characterised by the HVS-I motif 16129-16148-16168-16172-16187-16188G-16189-16223-16230-16311-16320 (Salas et al. 2004). All the L0a1 haplotypes share the same pattern of mutation in the HVS-I and HVS-II and are characterised by aforementioned HVS-I except for the reversion at position 16223.

Two different L1b1 lineages were found in two individuals from Al Awaynat, and three from Tahala. The HVS-I motif that characterises the L1b lineages is: 16126-16187-16189-16223-16264-16270-16278-16311 (Salas et al. 2004). The transition at position 16293, which is diagnostic for haplogroup L1b1 (Salas et al. 2002), was observed in both the Tuareg lineages.

The HVS-I basal motif 16223-16278-16294-16390 (Salas et al. 2004) allowed to assign 11 individuals to haplogroup L2a. All these individuals share the same HVS-I and HVS-II haplotype, except for two of them who have one mutation more (i.e. 16368 in AL15, and 00198 in AL96). One individual from Tahala (TA14), apart from the basal L2a HVS-I basal motif, harbours the transition at position 16309, which is diagnostic for the L2a1 sub-haplogroup. Furthermore, transitions at positions 16189 and 16192, characterize this haplotype as L2a1β3 (Salas et al. 2002). In the Al Awaynat sample, one individual (AL64) presents the typical L2a1 HVS-I motif and the mutation 16286 which is diagnostic for the sub-haplogroup L2a1a.

Two individuals from Al Awaynat were classified as L2b. They share the same haplotype, and harbour all the mutations which characterise the L2b HVS-I basal motif (i.e. 16114A-16129-16213-16223-16278-16390) plus the transition 16218.

The RFLP marker +2349 *Mbo*I, which is diagnostic of haplogroup L3e, was observed in ten individual from Al Awaynat. The pattern of mutation in the HVS-I allowed to discriminate them in three different L3e sub-haplogroups. The HVS-I haplotype 16223-16327 is the basal motif of the sub-haplogroup L3e1*. It was observed in four individual from Al Awaynat and all of them share the same HVS-I and HVS-II haplotype. Two individuals harbour the mutation at np 16320, that is characteristic of sub-haplogroup L3e2. Furthermore, transitions at positions 16189 and 16172 allowed to classify the individual AL50 as L3e2b (Salas et al. 2004). Four identical HVS-I and HVS-II haplotypes from Al Awaynat were characterised by transversion 16265T. This mutation defines the sub-haplogroup L3e3.

Two identical haplotypes from Tahala harbour the HVS-I motif 16209-16223-16311, which is diagnostic for haplogroup L3f (Salas et al. 2004). Furthermore, they are characterised by the presence of transition 16292, which makes it possiple to classify them as L3f1.

Three individuals from Al Awaynat share the same HVS-I and HVS-II haplotype and were classified as L3w. This lineage has not been widely described in literature so far, nevertheless, the three Tuareg haplotypes share the HVS-I/HVS-II motif 16223-16260-00150-00152 with some individuals from Ethiopia, who were previously classified as L3w (Kivisild et al. 2004).

Finally, two individuals from Al Awaynat were characterized by the presence of the restriction site +10397 *Alu*I, which characterizes the haplogroup M. They share the same haplotype and harbour the HVS-I/HVS-II motif 16129-16189-16249-16311-00195, which allowed to insert them in the sub-haplogroup M1 (Olivieri et al. 2006). Nonetheless, it is worth noting the lack of mutation 16223 in the Tuareg haplotypes.

## Haplogroup frequencies in the Tuareg sample

Haplogroup frequencies in the Tuareg samples from Al Awaynat and Tahala, and the frequency of the pooled sample, are reported in Table 3.3.

**TABLE 3.3**

Haplogroup frequencies in the Tuareg samples from the villages of Al Awaynat and Tahala (Fezzan, Libya) analysed in the present study. Haplogroup frequencies of the pooled sample (Tuareg Libya) are reported too.

| | **Al Awaynat** | | **Tahala** | | **Tuareg Libya**[a] | |
|---|---|---|---|---|---|---|
| | **Fr (N)** | **Fr (%)** | **Fr (N)** | **Fr (%)** | **Fr (N)** | **Fr (%)** |
| **H1** | 68 | 61.26 | 11 | 61.11 | 79 | 61.24 |
| **V** | 4 | 3.60 | 1 | 5.56 | 5 | 3.88 |
| **L0a1** | 8 | 7.21 | - | - | 8 | 6.2 |
| **L1b1** | 2 | 1.80 | 3 | 16.67 | 5 | 3.88 |
| **L2a** | 11 | 9.91 | - | - | 11 | 8.53 |
| **L2a1** | - | - | 1 | 5.56 | 1 | 0.78 |
| **L2a1a** | 1 | 0.90 | - | - | 1 | 0.78 |
| **L2b** | 2 | 1.80 | - | - | 2 | 1.55 |
| **L3e1** | 4 | 3.60 | - | - | 4 | 3.1 |
| **L3e2** | 1 | 0.90 | - | - | 1 | 0.78 |
| **L3e2b** | 1 | 0.90 | - | - | 1 | 0.78 |
| **L3e3** | 4 | 3.60 | - | - | 4 | 3.1 |
| **L3f1** | - | - | 2 | 1.11 | 2 | 1.55 |
| **L3w** | 3 | 2.70 | - | - | 3 | 2.33 |
| **M1** | 2 | 1.80 | - | - | 2 | 1.55 |
| **total** | 111 | | 18 | | 129 | |

[a] Individuals from Al Awaynat and Tahala pooled in one sample.

Calculation of the 95% Credible Region showed that most of the lineages in the two samples are characterised by overlapping distributions of the haplogroup frequencies (dara not shown). In particular as regards the H1 lineage, which constitutes the main component in both the samples, frequencies are almost the same (see figures 3.1a, and 3.1b). The only exception is represented by the L1b1 lineage, that is unexpectedly higher in the sample from Tahala. Results of the 95% Credible Region analysis and the fact that the aim of this study is not focused on the

genetic relationship between the villages of Al Awaynat and Tahala, led to the assumption that the two samples belong to a unique population, and so they were pooled in one sample. In the successive statistical analyses, the haplogroup frequencies of the pooled samples ('Tuareg Libya') are used. In figure 3.1a and 3.1b, graphic diagrams of the haplogroup frequencies in the two Tuareg samples are shown.

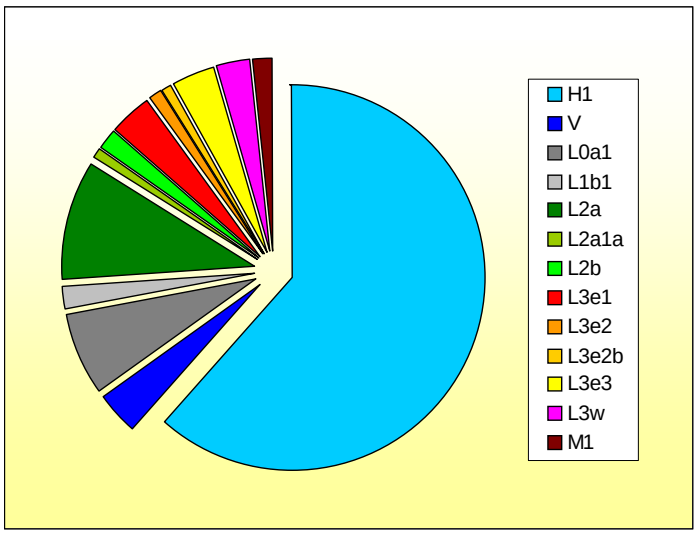


**Figure 3.1a:** Haplogroup frequencies in the sample from Al Awaynat.

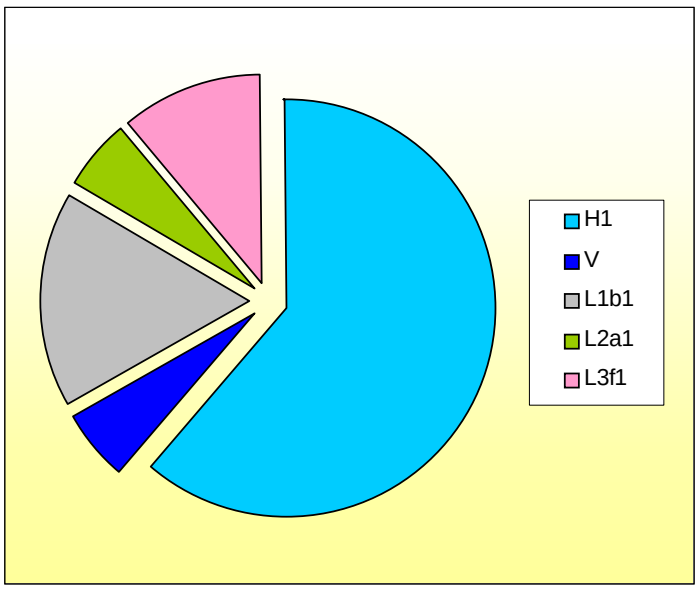


**Figure 3.1b:** Haplogroup frequencies in the sample from Tahala.

## Complete mtDNA sequencing

In figure 3.2, the most parsimonious tree of complete H1 mtDNA sequences is shown. The tree, rooted in haplogroup R, includes five Tuareg mtDNAs sequenced in this study, and other 12 H1 mtDNAs previously reported in Achilli et al. (2004). Phylogeny construction was performed by hand, following a parsimony approach, and was confirmed by use of the program Network as described in Achilli et al. 2004. The rCRS sequence, which belong to sub-haplogroup H2, is included in the tree.

It is worth noting that the Tuareg haplotypes are all different, and that two of them (i.e. TA16 and AL89) are characterised by the presence of the RFLP site 4313, which can be detected through the enzyme *Alu*I.

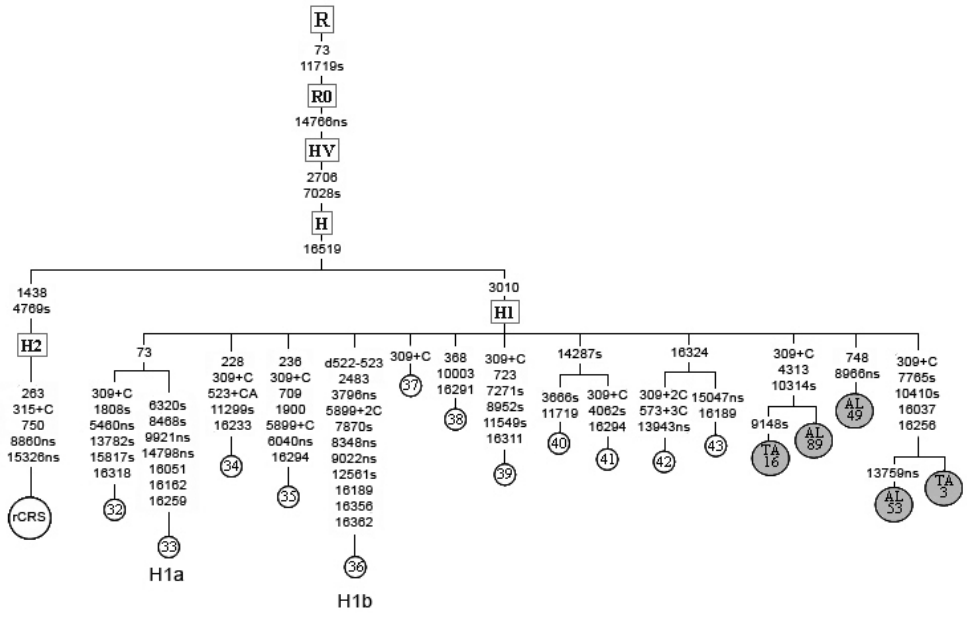


**Figure 3.2:** Most-parsimonious tree of complete H1 mtDNA sequences. Tuareg mtDNAs (ID sample is reported) are depicted in grey. The other sequences follow the code reported in Achilli et al. (2004), i.e. 32 and 33 are from Georgia, 34 is from a Berber subject, 35-37 and 39-43 are from Italy, 38 is from Iraq. Other codes as follow: d, deletion; s, sense; ns, non-sense; +, insertion.

## 3.3 STATYSTICAL ANALYSES

### Standard diversity indices and molecular indices

Standard diversity indices (i.e. N, K, and S) and Molecular indices (i.e. π and Nucleotide Diversity) were calculated in all the population collected in the database. Indices observed in some African populations are reported in Table 3.4.

It is worth noting that these results are limited to the HVS-I haplotypes (except for Libyan Tuareg, whose results of calculations from the HVS-I and HVS-II haplotypes are reported in the table). Despite its high size, which is one of the highest in the African samples available so far, the Libyan Tuareg sample shows the lowest value of K (i.e. Number of different haplotypes), after Mbuti and Mbenzele, which have a smaller sample size. High homogeneity at the haplotype level, is reflected by the lowest value of Gene Diversity (0.677 +/– 0.046). As regards the Mean Number of Pairwise Differences, Tuareg are characterised by the lowest value (4.398 +/– 2.186); comparable values are found in the Berber samples from Morocco (including those from Souss area). A similar result is observed in the Nucleotide Diversity values: Tuareg from Libya are characterised by a low value (0.01099 +/– 0.00605), which is comparable only with the Berber samples (Matmata and Moroccan Berbers). No similarities are observed between the Tuareg from Libya and the Western Tuareg (Watson et al. 1996), whose sample has an H value (0.993 +/– 0.014) comparable with that of the other African populations (which is generally higher than 0.95). In the Westen Tuareg, the values of π (6.989 +/– 3.402) and Nucleotide Diversity (0.01747 +/– 0.00949) fall in the range of Northwest African populations, especially some Berber and Arab groups.

**TABLE 3.4**

Standard Diversity indices and Molecular indices in some African populations[a]. All indices were calculated from HVS-I haplotypes of population collected in the database, apart from Tuareg Libya whose indices calculated on HVS-I/HVS-II are reported too. Abbreviations are as follows: N, size of the sample; K, number of haplotypes; S, number of polymorphic sites; π, Mean number of pairwise differences.

| POPULATION | N | K | S | H | π | Nuclotide Diversity |
|---|---|---|---|---|---|---|
| **Tuareg Libya** | 129 | 20 | 41 | 0.677 +/– 0.046 | 4.398 +/– 2.186 | 0.01099 +/– 0.00605 |
| **Tuareg Libya**[b] | 129 | 21 | 61 | 0.678 +/– 0.046 | 7.769 +/– 3.642 | 0.00997 +/– 0.00517 |
| **Western Tuareg** | 24 | 22 | 40 | 0.993 +/– 0.014 | 6.989 +/– 3.402 | 0.01747 +/– 0.00949 |
| **Oromo** | 84 | 66 | 94 | 0.993 +/– 0.003 | 9.100 +/– 4.229 | 0.02521 +/– 0.01298 |
| **Amhara** | 90 | 75 | 104 | 0.994 +/– 0.003 | 8.523 +/– 3.977 | 0.02442 +/– 0.01260 |
| **Fayum Egypt** | 69 | 63 | 82 | 0.997 +/– 0.003 | 8.549 +/– 4.000 | 0.02137 +/– 0.01109 |
| **Fon** | 49 | 39 | 48 | 0.990 +/– 0.006 | 7.439 +/– 3.537 | 0.01860 +/– 0.00981 |
| **Berba** | 46 | 42 | 67 | 0.996 +/– 0.005 | 8.911 +/– 4.182 | 0.02228 +/– 0.01161 |
| **Bariba** | 50 | 43 | 67 | 0.989 +/– 0.008 | 8.776 +/– 4.117 | 0.02194 +/– 0.01142 |
| **Kanembu** | 50 | 37 | 58 | 0.988 +/– 0.006 | 9.295 +/– 4.343 | 0.02324 +/– 0.01205 |
| **Fali** | 40 | 23 | 43 | 0.947 +/– 0.019 | 7.546 +/– 3.597 | 0.01886 +/– 0.00999 |
| **Chad Arabs** | 27 | 20 | 37 | 0.963 +/– 0.023 | 7.242 +/– 3.500 | 0.01810 +/– 0.00975 |
| **Tunisian Arabs** | 47 | 42 | 61 | 0.990 +/– 0.009 | 6.150 +/– 2.977 | 0.01709 +/– 0.00918 |
| **Algerian Arabs** | 47 | 27 | 50 | 0.956 +/– 0.014 | 5.681 +/– 2.772 | 0.01578 +/– 0.00855 |
| **Fulani Chad** | 185 | 57 | 61 | 0.938 +/– 0.008 | 7.215 +/– 3.396 | 0.01804 +/– 0.00940 |
| **Fulbe** | 65 | 37 | 44 | 0.945 +/– 0.018 | 6.781+/– 3.236 | 0.01885 +/– 0.00997 |
| **Kotoko** | 56 | 31 | 53 | 0.956 +/– 0.014 | 7.477 +/– 3.545 | 0.01870 +/– 0.00983 |
| **Mandenka** | 119 | 50 | 52 | 0.947 +/– 0.012 | 6.513 +/– 3.101 | 0.02055 +/– 0.01083 |
| **Chenini Berbers** | 53 | 23 | 41 | 0.939 +/– 0.017 | 6.822 +/– 3.264 | 0.01706 +/– 0.00905 |
| **Matmata** | 49 | 29 | 56 | 0.946 +/– 0.021 | 5.050 +/– 2.494 | 0.01262 +/–0.00692 |
| **Sened** | 53 | 37 | 64 | 0.975 +/– 0.011 | 7.526 +/– 3.570 | 0.01882 +/– 0.00990 |
| **Souss Berbers** | 50 | 34 | 38 | 0.961 +/– 0.018 | 4.604 +/– 2.298 | 0.01151 +/– 0.00638 |
| **Moroccan Berbers** | 64 | 42 | 51 | 0.968 +/– 0.013 | 4.497 +/– 2.243 | 0.01249 +/– 0.00691 |
| **Mozabiti** | 85 | 30 | 37 | 0.943 +/– 0.010 | 4.821 +/– 2.377 | 0.01597 +/– 0.00872 |
| **Mbuti** | 20 | 9 | 19 | 0.858 +/– 0.054 | 6.205 +/– 3.077 | 0.01799 +/– 0.00996 |
| **Mbenzele** | 57 | 12 | 22 | 0.805 +/– 0.037 | 4.986 +/– 2.460 | 0.01381 +/– 0.00756 |

[a] Because of the large amount of data, only a sub-sample of populations from those collected in the database is reported.
[b] Calculations from the HVS-I and HVS-II haplotypes.

## Comparison of CRS haplotype frequencies

The H1 haplogroup represents the main component in the Libyan Tuareg sample, with a frequency higher than 60%. Together with the haplogroup V, it constitutes an 'extra-African' component. Within the H1 haplogroup, 72 individuals on a total of 79 (91.14%) are characterised by a CRS HVS-I haplotype. The comparison of HVS-I H-CRS (the abbreviation H-CRS is used here) frequencies between the Libyan Tuareg sample and the other African samples collected in the database was carried out, and results are reported in Table 3.5 and Figure 3.3. Despite most of the data on African populations in literature are limited to the HVS-I sequencing, this making not possible to ascertain that the CRS haplotypes belong to H1 (e.g. data from Watson et al. 1996), we think that looking at the distribution of this haplotype in Africa can provide important information about the origin of the 'extra-African' component in the Libyan Tuareg sample, and about the relationships between the Libyan Tuareg and the neighbouring populations.

In order to reduce the size of the data and to present the most informative results, only a sub-sample of Central, Northwest and Northeast African population from the database are reported in the analysis. All the other African populations collected in the database which are not reported in Table 3.6, did not present any H-CRS haplotype, or had very low frequencies. The frequency of H-CRS in the Libyan Tuareg sample is significantly much higher than observed in all the other African populations. Nevertheless, distribution of frequencies shows that this HVS-I haplotype is particularly spread in Northwest Africa, more in detail in the Berber groups, which present frequencies higher then 15%. Differently, Eastern African (e.g. Egypt, 5.56%; Nubia, 1.25%) and typically South-Saharan populations, are characterised by very low frequencies of the H-CRS haplotype.

**TABLE 3.5**

H-CRS HVS-I haplotype frequencies in some African populations. Absolute frequencies (N), and relative frequencies in percentage are reported. For references see table A3.

| POPULATION | H-CRS (N) | Sample size | Freq (%) |
|---|---|---|---|
| **Libyan Tuareg** | 72 | 129 | 55.81 |
| **Mozabiti Berbers** | 7 | 85 | 8.24 |
| **Tunisians** | 5 | 47 | 10.64 |
| **Algerians** | 4 | 47 | 8.51 |
| **Western Tuareg** | 2 | 24 | 8.33 |
| **Chenini Douiret Berbers** | 3 | 53 | 5.66 |
| **Matmata Berbers** | 10 | 49 | 20.41 |
| **Sened Berbers** | 5 | 53 | 9.43 |
| **Moroccan Berbers** | 10 | 64 | 15.63 |
| **Souss Berbers** | 9 | 50 | 18.00 |
| **Moroccans** | 4 | 50 | 8.00 |
| **Saharawi** | 3 | 80 | 3.75 |
| **Mauritania** | 7 | 95 | 7.37 |
| **Fulbe** | 0 | 65 | 0.00 |
| **Fulani** | 1 | 185 | 0.54 |
| **Nubia** | 1 | 80 | 1.25 |
| **Sudan** | 0 | 60 | 0.00 |
| **Egypt** | 7 | 126 | 5.56 |
| **Hide** | 1 | 23 | 4.35 |
| **Mafa** | 0 | 32 | 0.00 |
| **Kotoko** | 0 | 56 | 0.00 |
| **Masa** | 0 | 31 | 0.00 |
| **Kanuri_a** | 1 | 31 | 3.23 |
| **Kanembu** | 0 | 50 | 0.00 |
| **Fali** | 0 | 40 | 0.00 |
| **Buduma** | 1 | 30 | 3.33 |
| **Arab Shuwa** | 0 | 38 | 0.00 |
| **Arab Chad** | 0 | 27 | 0.00 |

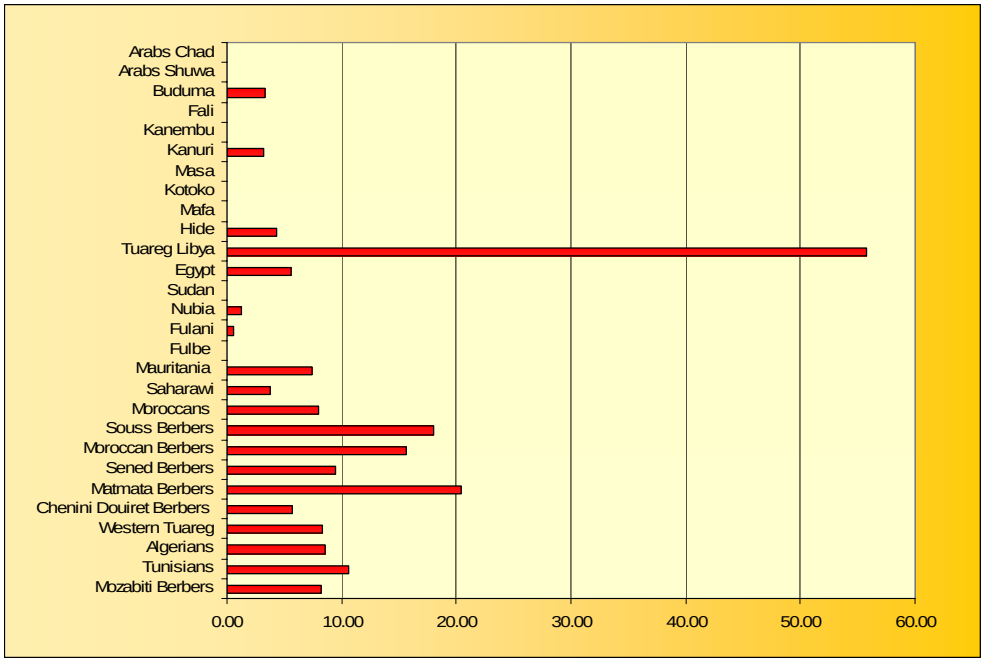


**Figure 3.3:** Graphical representation of HVS-I H-CRS frequencies in some African populations reported in table 3.5.

## Haplogroup V frequencies

Only in the populations which present the highest frequencies of the H-CRS haplotype, the frequency of haplogroup V was observed too (Table 3.6). The highest values were found in the Berber groups, particularly in those characterised by a high frequency of the haplotype H-CRS (Matmata, 16.33%; Souss Berbers, 10%; Mozabiti, 8.2%; Moroccan Berbers, 6.25%), with the exception of Sened. It is worth noting that non Berber groups, i.e. Tunisians and Algerians, despite the high frequencies of H-CRS, do not show any V lineage.

**TABLE 3.6**
Haplogroup V frequencies in the African population characterised by high frequencies of the HVS-I H-CRS haplotype.

| POPULATION | Freq H-CRS (%) | Freq V (%) |
|---|---|---|
| **Tuareg Libya** | 55.81 | 3.88 |
| **Mozabiti Berbers** | 8.24 | 8.2 |
| **Tunisians** | 10.64 | 0 |
| **Algerians** | 8.51 | 0 |
| **Chenini Douiret Berbers** | 5.66 | 0 |
| **Matmata Berbers** | 20.41 | 16.33 |
| **Sened Brebers** | 9.43 | 0 |
| **Moroccan Berbers** | 15.63 | 6.25 |
| **Souss Berbers** | 18 | 10 |
| **Western Tuareg** | 8.33 | 0 |

## Multidimensional Scaling

The matrix of Saltkin's linearized Fst was generated by the program Arlequin, and included genetic distances between all the African populations present in the database. Reduction of data to two coordinates was carried out through the program Statistica, and the two-dimensional representation of distances between all the African populations, which had a *stress* value of 0.11, clearly showed the outlier position of five populations (Mbenzele, Biaka, Mbutu, Kung, Herero), which consequently compressed all the others in one single group (data not shown). Once discarded the outliers, a 'good' resolution of the plot was obtained (see forward Figure 3.4a): the *stress* value is 0.12, and the relationships between populations are clearer than the first representation. Geographical criteria seem to explain the grouping of populations in the two-dimensional plot: almost all North African populations are located in the left side, while the typical South-Saharan ones are placed in the right side. More particularly, Berber groups and some isolates from North Africa occupy the most marginal position in the left side. In the context of the South-Saharan populations, it is possible to observe a further geographical partition, as Eastern populations are placed at the top, Western populations at the bottom, and

Central populations in the middle of the graph. African populations south of the Equator are located at the top-right fringe (e.g. Khwe, Mozambico, and Southeast Bantu).

After removing all the 'extra-African' haplotypes (i.e. haplotypes classified as H1 and V) from the Libyan Tuareg sample, an other matrix of Slatkin's Fst was calculated with all the populations of the database. After discarding the *outliers* the *stress* value is 0.12, and the two-dimensional representation shows the shift of the Libyan Tuareg to the upper right of the plot, near Kikuyu and Kenya (see Figure 3.4b).

## Correspondence Analysis

Correspondence Analysis was performed using the haplogroup frequencies reported in Table A4. As the H-CRS haplotypes represent the main component of the Libyan Tuareg mtDNA pool, we decided to consider their frequency as a separated variable from the other H haplotypes. It makes possible to better discriminate the populations in relation to the HVS-I CRS motif within the H lineages. Haplogroups which could be considered as *sister clade*, or strictly phylogenetically related, were grouped in single variables (e.g. N/N1/R, preHV/HV, L1*/L1b/L1c, L2b/L2c, L3b/L3d, U5/U5a, and U2-U4/U7-U9/K): it allowed us to reduce the number of variables, and furthermore to have a uniform dataset of frequencies in which even samples from literature whose classification has not the same degree of phylogenetic resolution of the other samples were included. In the particular case of U lineages, haplogroups U5b, U5/U5a, and U6 were kept separated from the others, as their distribution in North Africa is hypothetically related to different migration events (Achilli et al. 2005, Coia et al. 2005, Olivieri et al. 2006). Extra-African lineages that were found at low frequencies in the African continent were all included in the variable 'others'.

The two-dimensional plot is reported in Figure 3.5a. It represents 35.89 % of the total *inertia* (the first dimension 22.13 %, the second dimension 13.76 %). Relationship between populations and haplogroups in the dotted-square are shown at higher resolution in Figure 3.5b.

Most of the typically South-Saharan populations are clustered in the left side, while on the right side populations from Eastern and Northern Africa are placed. The particular distribution of the populations along the first dimension seems to reflect, from left to right, a distinction between typically South-Saharan populations, characterised by L lineages with an almost exclusive African distribution (mainly L1, L2, and L3e), and populations from East and North Africa, these presenting lineages with an 'extra-African' distribution which are mainly shared both with the Near East and the Mediterranean area.

On the second dimension, a hypothetical distinction between East Africa and North Africa is observed. In fact, most of the East African populations (mostly from Ethiopia) are located in the lower right of the plot, in strict correlation with some typical Eastern African lineages (e.g. L3w, L5, and L4), and even with some lineages from Near East (e.g. M1, preHV/HV). Two samples from Egypt (Fayum Egyptians and Gurna), two Berber samples (Chenini and Sened) and the Tunisian samples, form a cluster in the middle of the second dimension, in association with lineages from the Near East (preJT/JT, U5/U5a) and Northern/Eastern African lineages (L3*, U6). In the upper right of the plot, the Libyan Tuareg, Mozabite Berbers, and Matmata Berbers are located. Their position is strictly related to lineages H, H-CRS, and V. More particularly, the position of Mozabite is associated with haplogroup H, while the Libyan Tuareg sample, which appears to be an *outlier*, is related to the H-CRS lineages.

**Figure 3.4**

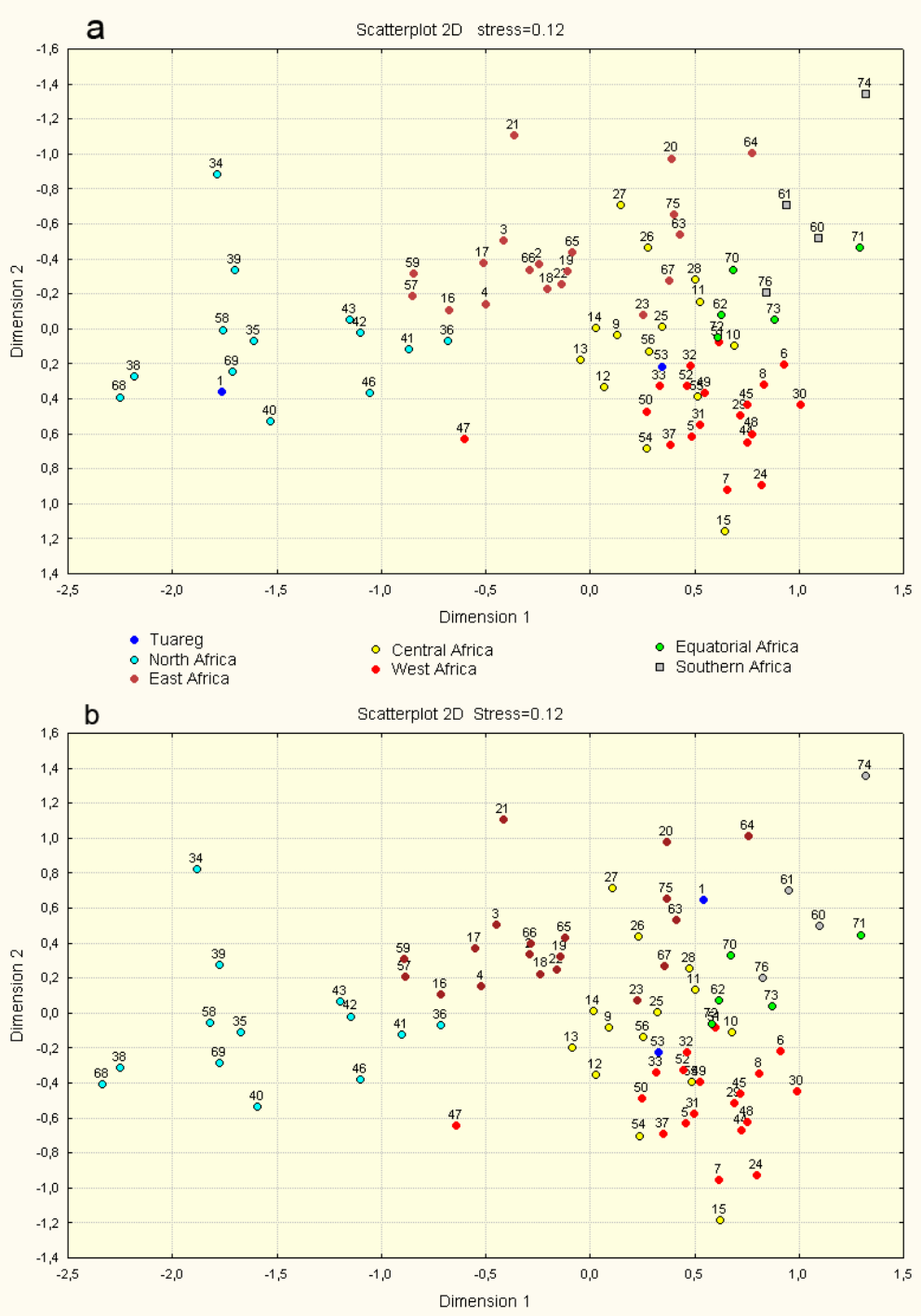
a
Scatterplot 2D stress=0.12
Dimension 2
Dimension 1
Tuareg
North Africa
East Africa
Central Africa
West Africa
Equatorial Africa
Southern Africa
b
Scatterplot 2D Stress=0.12
Dimension 2
Dimension 1

**Figure 3.4:** Multidimensional Scaling. **a,** two-dimensional plot of Slatkin's linearized Fst between African populations. **b,** two-dimensional plot of Slatkin's linearized Fst between African populations after excluding the 'West Eurasian' genetic component (i.e. the H1 and V lineages) from the Libyan Tuareg sample. Numeric code are as follows: 1 Tuareg Libya; 2 Oromo; 3 Amhara; 4 Fayum Egypt; 5 Fon; 6 Dendi; 7 Berba; 8 Bariba; 9 Kanuri; 10 Kanembu; 11 Fali; 12 Buduma; 13 Arabs Shuwa; 14 Arabs Chad; 15 Fulani; 16 Yemen; 17 Tigrai; 18 Oromo_b; 19 Gurage; 20 other Ethiopians; 21 Eritreans; 22 Amahra_b; 23 Afar; 24 Mandenka; 25 Hide; 26 Mafa; 27 Kotoko; 28 Masa; 29 Mende; 30 Loko; 31 Limba; 32 Temne; 33 Guineans; 34 Chenini-Douiret Berbers; 35 Matmata Berbers; 36 Sened Berbers; 37 Senegal; 38 Moroccan Berbers; 39 Mozabiti Berbers; 40 Souss Berbers; 41 Moroccans; 42 Tunisians; 43 Algerians; 44 Wolof; 45 Serer; 46 Saharawi; 47 Mauritania; 48 Mali; 49 Barbara; 50 Capo Verde; 51 Yoruba; 52 Songhai; 53 Western Tuareg; 54 Fulbe; 55 Hausa; 56 Kanuri_b; 57 Egyptians; 58 Canary Islands; 59 Libyans; 60 Mozambicans; 61 Southeast Bantu; 62 Sao Tomè/Bioko; 63 Kikuyu; 64 Turkana; 65 Somalia; 66 Nubia; 67 Sudan.

**Figure 3.5**

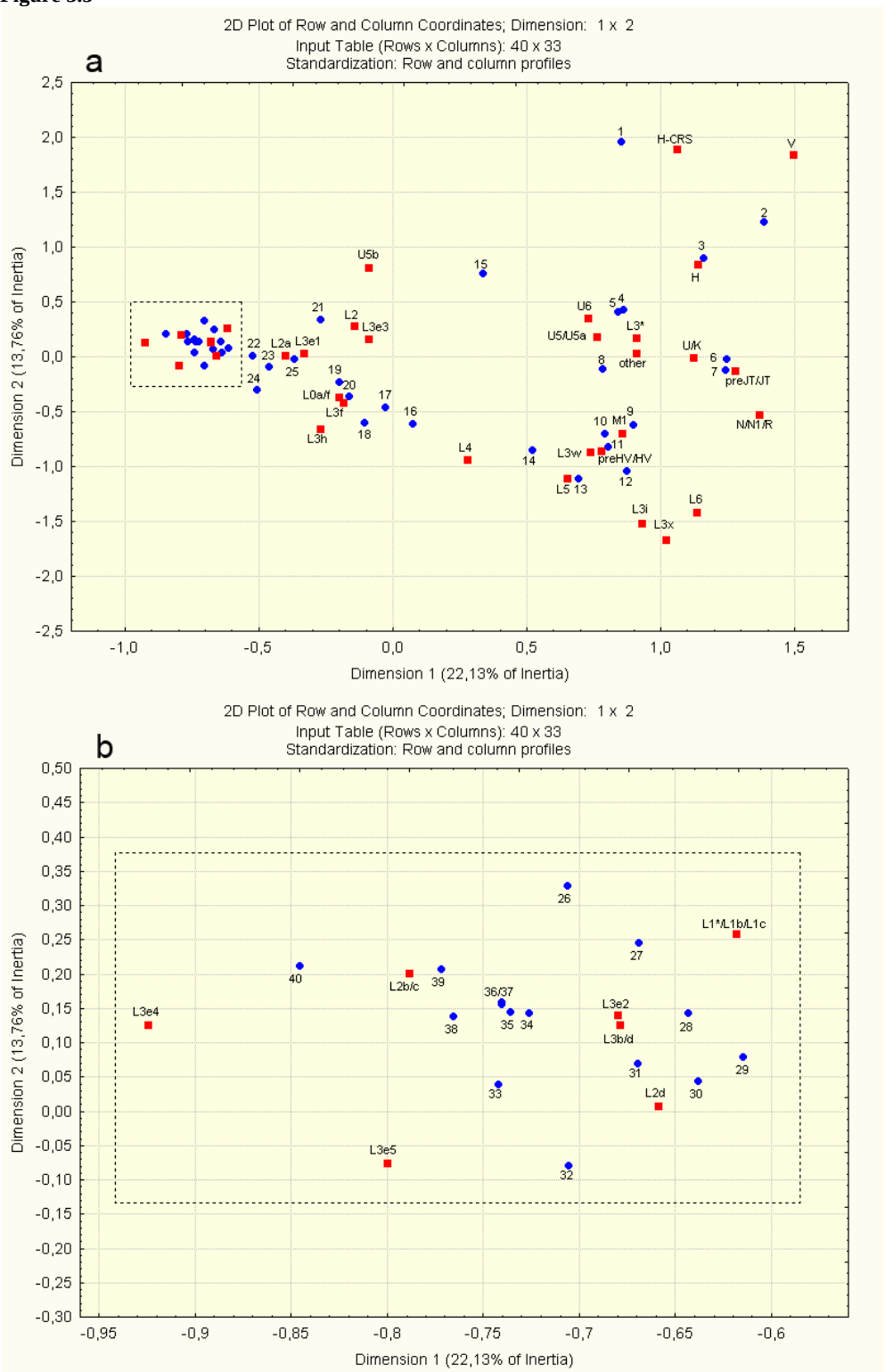
2D Plot of Row and Column Coordinates; Dimension: 1 x 2
Input Table (Rows x Columns): 40 x 33
Standardization: Row and column profiles
a
Dimension 2 (13,76% of Inertia)
Dimension 1 (22,13% of Inertia)
1
H-CRS
V
2
3
H
U5b
15
21
L2
L3e3
22
L2a
L3e1
23
25
24
19
20
L0a/f
L3f
L3h
17
16
18
L4
14
U6
5 4
L3*
U5/U5a
8
other
U/K
6
7
preJT/JT
10
M1
9
N/N1/R
11
L3w
preHV/HV
12
L5
13
L6
L3i
L3x
2D Plot of Row and Column Coordinates; Dimension: 1 x 2
Input Table (Rows x Columns): 40 x 33
Standardization: Row and column profiles
b
Dimension 2 (13,76% of Inertia)
Dimension 1 (22,13% of Inertia)
26
L1*/L1b/L1c
27
40
L2b/c
39
36/37
L3e4
38
35
34
L3e2
L3b/d
28
31
29
33
L2d
30
L3e5
32

**Figure 3.5:** Correspondence Analysis. **a**, two-dimensional representation of relationship between some African populations and their haplogroup frequency. Relationship between populations and haplogroups in the dotted-square are shown at higher resolution in figure **b.** Symbols are as follows: Red Square, haplogroups; Blue Circles, African populations numbered as in legend. Numeric codes are as follows: 1 Tuareg Libya; 2 Matmata Berbers; 3 Mozabiti Berbers; 4 Tunisians; 5 Sened Berbers; 6 Chenini Berbers; 7 Egyptians (Gurna); 8 Fayum Egypt; 9 Amhara; 10 Oromo; 11 Tigrai; 12 Gurage; 13 Oromo_b; 14 Amhara_b; 15 Mauritania; 16 Kikuyu; 17 Afar; 18 Kenya; 19 Buduma; 20 Arabs Chad; 21 Western Tuareg; 22 Kanuri; 23 Kanembu; 24 Kotoko; 25 Arabs Shwa; 26 Fulani; 27 Loko; 28 Mende; 29 Hide; 30 Mafa; 31 Berba; 32 Masa; 33 Fali; 34 Bariba; 35 Temne; 36 Guineans; 37 Limba; 38 Barbara; 39 Fon; 40 Malinke.

## Network analysis

The Network analysis was carried out as described in chapter two. It is worth noting that in order to resolve the structure of the networks, significantly different weights were assigned to the same nucleotide site in different haplogroups. It is hypothetically due to the fact that some L lineages, particularly L2a, show a violation of the clock-like evolution (Torroni et al. 2001b, Howell et al. 2004). In this sense the mutation rate of some variable sites, and consequently the their recurrence in the network, may change in different lineages.

In each network the root node is represented by the symbol '#', and all the mutated positions reported in the branches has to be read as +16,000 (e.g. 223, stands for 16223), as they belong to the HVS-I region. State of polymorphism is reported only when transversions occurred, while all the other mutations are transitions. Node sizes are proportional to the absolute haplotype frequencies, as reported in the legends. Relative frequencies of haplotypes from different geographic African areas in the same nodes, are represented as different coloured pies. Each colour corresponds to a geographic area of the African continent as reported in table A3 and in figure A1, apart from the Western Tuareg, who are reported in light blue (like Berbers and other North African populations) in order to distinguish them from the Libyan Tuareg (dark blue).

**L2a**. The MJ Network of the L2a HVS-I lineages is reported in figure 3.6. The wide distribution of L2a in South-Saharan Africa makes it difficult to resolve its phylogeny (Salas et al. 2002). In order to have a clear topology of the Network, L2a1 haplotypes defined by transition at np 16309 (TA14 and AL64) were removed. In fact, as previously described in literature (Salas et al. 2002, Kivisild et al. 2004), high homoplasy at np 16309, together with 16189 and 16192, originates a reticulation which cannot be resolved with certainty. Some haplotypes from literature which were located in long branches from the root were discarded from the analysis, because they appeared to be scarcely informative. A total of 158 HVS-I haplotypes are finally represented in the network. Weights at nps are listed below:

| Nucleotide position (np) | Weight |
|---|---|
| 16093 | 15 |
| 16145 | 15 |
| 16189 | 8 |
| 16192 | 8 |
| 16193 | 8 |
| 16213 | 15 |
| 16291 | 7 |
| 16292 | 8 |
| 16301 | 15 |
| 16311 | 5 |
| 16354 | 15 |
| 16355 | 8 |
| 16362 | 5 |

The default value of 10 was kept for all the other nps, and transversions were weighted three times more than transitions.

The basal L2a motif (16223-16278-16294-16390, indicated by '#') is mostly distributed in Central Africa, which is represented by 55.5% of the haplotypes in the node. Western Africa is represented by 33.3% of the haplotypes and the remaining fraction is from Equatorial Africa. Ten of the eleven L2a Tuareg haplotypes are located in the node departing from the root through transitions 16189, and 16192. The Libyan Tuareg represent the main fraction of this node (45.4%) and it is worth noting that an other Tuareg haplotype from literature (Western Tuareg) shares the same HVS-I pattern of mutations. Both Western (13.6%) and Eastern African individuals (18.2%) are present in this node, while the remaining fraction is mainly characterised by haplotypes from Southeast Africa. One Libyan Tuareg haplotype is located as a direct derivative from the 16189-16192 node through transition 16368. Most of the other derivate haplotypes are from Eastern Africa, and are located as single individuals 1-mt step (one mutational step) distant from the ancestor node: this results in a *star-like* structure of the branch 16189-16192.

Some Western Tuareg L2a lineages are spread in the network: one is present in the node departing from the root through mutation 16189, together with both Eastern, which represent the main fraction, and Western African haplotypes. An other Tuareg haplotype is located in the branch 16209-16301-16354 from the root, which

appears to be exclusively characterised by Central and Eastern African individuals. The last two Tuareg haplotypes are placed in a long Central African branch as two '1-mt step-distantly related' individuals.

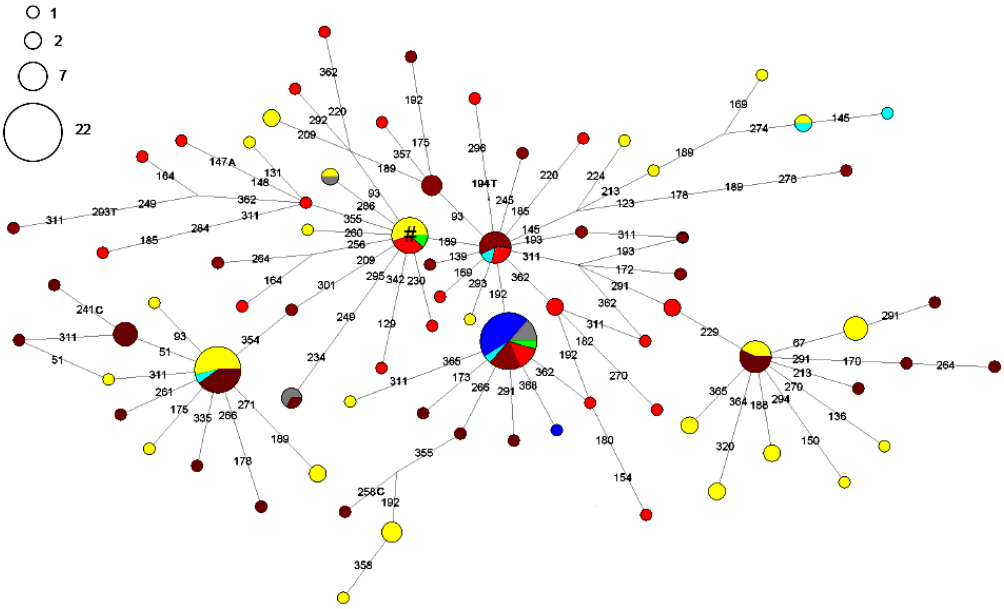


**Figure 3.6:** Network of African L2a lineages. Colours are as follows: Yellow, Central Africa; Red, Western Africa; Brown, Eastern Africa; Green, Equatorial Africa; Light Blue, Northern Africa and Western Tuareg; Grey, Southern Africa (SA); Dark Blue, Libyan Tuareg.

The coalescence times were calculated as described in chapter two. More particularly, calculations were focused on three nodes whose topology indicated some kind of population expansion: the L2a-16209-16301-16354, the L2a-16189-16229-16291-16311, and the L2a-16189-16192 branches. For the first and second node, calculations were carried out twice: when including all the derivatives, and when considering only the 1-mt step distantly related ones. Data are shown in table 3.7. It is worth noting that under the assumption of a high endogamy effect in the Libyan Tuareg population, coalescence times are affected by the high fraction of Tuareg lineages in the basal node. In order to take account of the hypothetical effect of endogamy, the TMRCA for the L2a-16189-16192 node was also calculated after

reducing the number of Tuareg haplotypes in the basal node to one. This resulted in a slightly higher value of the TMRCA.

**TABLE 3.7**

Coalescence ages in the the main L2a nodes. Two different calculations were carried out: the first includes all derivatives from the node, the second includes only the 1-mutational step distant derivatives (when available)

| node | TMRCA (years BP) (all derivatives) | TMRCA (years BP) (1-mt step derivatives) |
|---|---|---|
| L2a-16209-16301-16354 | 41,700 +/– 19,600 | 7,500 +/– 3,800 |
| L2a-16189-16229-16291-16311 | 21,000 +/– 6,600 | 11,300 +/– 6,300 |
| L2a-16189-16192 | 5,045 +/– 1,900[a] | - |
| L2a-16189-16192[b] | 7,400 +/– 2,800[a] | - |

[a] Only the one/two mutational-steps derivatives were considered.
[b] Number of Tuareg haplotypes reduced to one.

**L0a1.** The MJ Network of the L0a1 HVS-I lineages is shown in Figure 3.7. The default weight value of 10 was kept for all the nps apart from the following ones:

| Nucleotide position (np) | Weight |
|---|---|
| 16093 | 7 |
| 16148 | 6 |
| 16223 | 7 |

Transversions were weighted three times more than transitions. A total of 221 HVS-I haplotypes are represented in the network. All the eight HVS-I L0a1 Libyan Tuareg haplotypes are located in a node deriving from the basal root (HVS-I motif 16129-16148-16168-16172-16187-16188G-16189-16223-16230-16311-16320, indicated with '#') through reversion at np 16223. They appear to be the only haplotypes so far collected in literature with this topic. The L0a1 root is widely distributed in both Eastern (43.5% of the haplotypes) and Western Africa (34.8%), and a lower fraction is present in Central Africa too (17.4%). Many derived haplotypes are 1-mt step distant from the basal node, this resulting in a *star-like* structure of the node. These derivatives are mainly from Western and Central Africa.

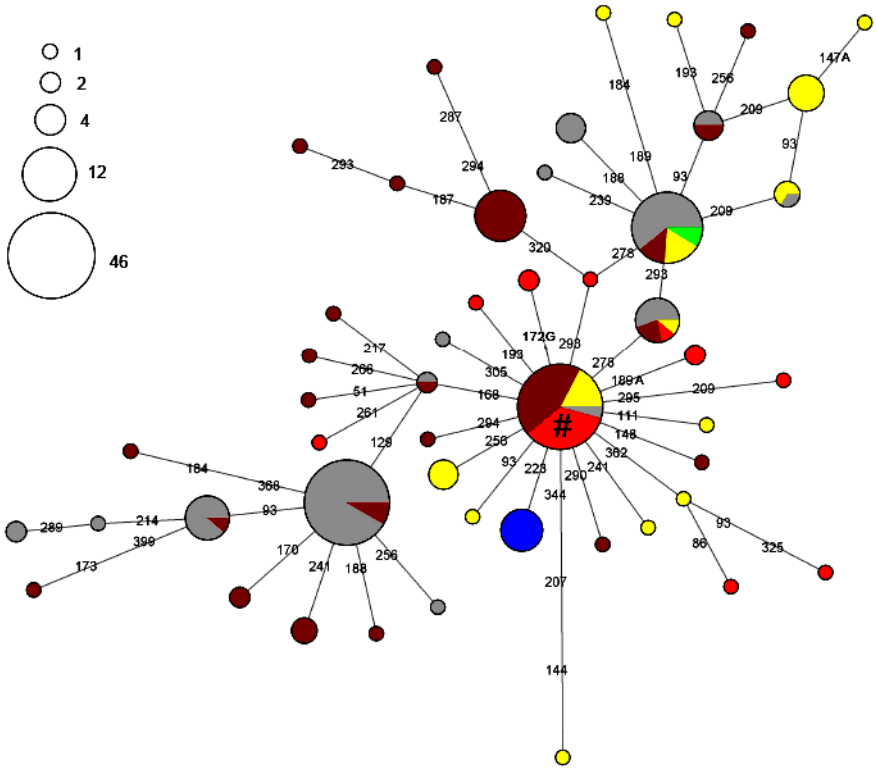


**Figure 3.7:** Network of African L0a1 lineages. Colours are as follows: Yellow, Central Africa; Red, Western Africa; Brown, Eastern Africa; Green, Equatorial Africa; Light Blue, Northern Africa and Western Tuareg; Grey, Southern Africa (SA); Dark Blue, Libyan Tuareg.

The coalescence age was calculated considering the basal L0a1 node, which presents a *star-like* structure, as the root of a population expansion event. When considering only the 1-mt step derivatives, the value of 9,400 +/– 3,300 years was observed.

**L1b.** The MJ Network of the L1b HVS-I lineages is shown in Figure 3.8. The default weight value of 10 was kept for all the nps apart from the following ones:

| Nucleotide position (np) | Weight |
|---|---|
| 16093 | 30 |
| 16223 | 6 |
| 16274 | 6 |
| 16293 | 30 |

Transversions were weighted three times more than transitions. A total of 205 HVS-I haplotypes are represented in the network.

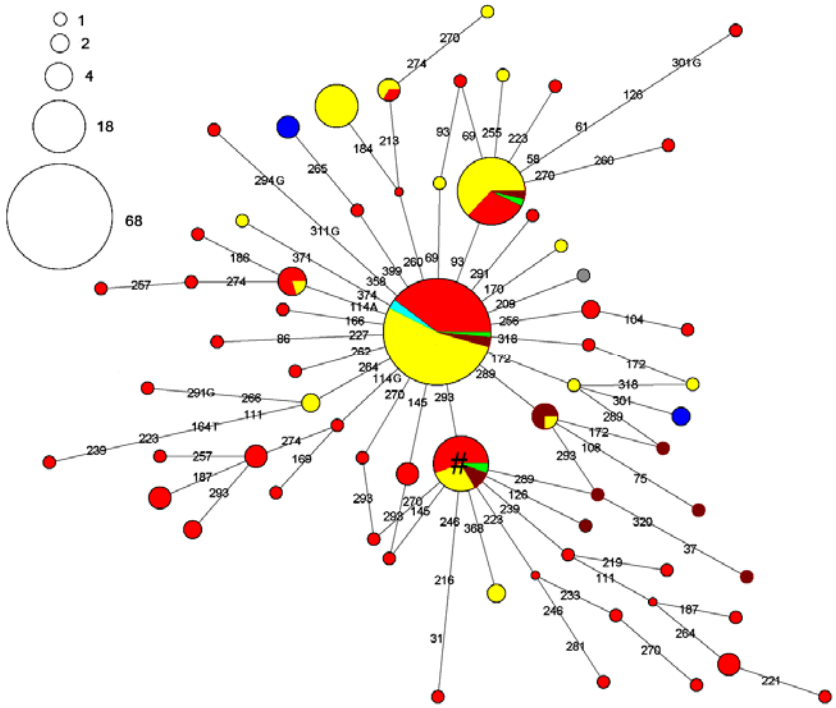


**Figure 3.8:** Network of African L1b lineages. Colours are as follows: Yellow, Central Africa; Red, Western Africa; Brown, Eastern Africa; Green, Equatorial Africa; Light Blue, Northern Africa and Western Tuareg; Grey, Southern Africa (SA); Dark Blue, Libyan Tuareg.

It is worth noting that some reticulations in the network were manually resolved by considering different phylogenetic pathways at np 16114. In fact this polymorphic site is characterised by the double transversion C16114A and C16114G, which were not discriminated by the program during calculations. The basal L1b motif (16126-16187-16189-16223-16264-16270-16278-16311) is mainly distributed in Western Africa (55.6% of the haplotypes), and then in Central Africa (27.8%). The L1b1 sub-haplogroup basal node, which is defined by the transition 16293, represents almost one-third of the entire L1b dataset (68 haplotypes on a total of 205 L1b haplotypes from the database, i.e. 33.2%). This is mainly represented by Central (52.9%) and Western African haplotypes (39.7%). Interestingly, two

Western Tuareg haplotype are placed in this node, and most of the Central African haplotypes belong to Fulani individuals (69.4%).

A total of five Libyan Tuareg haplotypes were classified as L1b1: three individuals from Tahala are located as derivatives of a haplotype from Mali (Western Africa), on a branch departing from the L1b1 root trough mutations 16265-16399. Similarly, the two Al Awaynat haplotypes are placed on a branch departing from the L1b1 root trough mutations 16172-16301, as derivatives of a haplotype from the Fali population (Central Africa).

Coalescent age was calculated in the L1b1 root, whose *star-like* topology strongly suggests an expansion event. When considering only the 1-mt step derivatives, the hypothetical expansion process dates back to 10,400 +/– 4,800 years ago. When including in the calculation the 2-mt step derivatives, the coalescence age increases to 17,040 +/– 4,900 years.

**L2b.** The MJ Network of the L2b HVS-I lineages is shown in Figure 3.9. Default values of the weights were used for all the sites (i.e. 10) and, as usual, transversions were weighted three times more than transitions. A total of 57 HVS-I haplotypes are represented in the network. The basal L2b motif 16114A-16129-16213-16223-16278-16390 is exclusively distributed in Western Africa and presents a *star-like* structure whose derivatives are mainly located in Western and Central Africa. The two identical Libyan Tuareg haplotypes are loated as derivatives departing from the root through transition 16218. Their position represents the ancestor node of two Central African haplotypes from the Kanembu and the Buduma populations.

It is worth noting that most of the haplotypes in the network are from Western Africa, and many Central African haplotypes are usually represented as derivates. Nontheless, transition 16145 seems to characterise a Eastern African branch. The coalescence age calculated in the L2b root including the 1-mt step derivatives is 17,190 +/– 5,740 years.

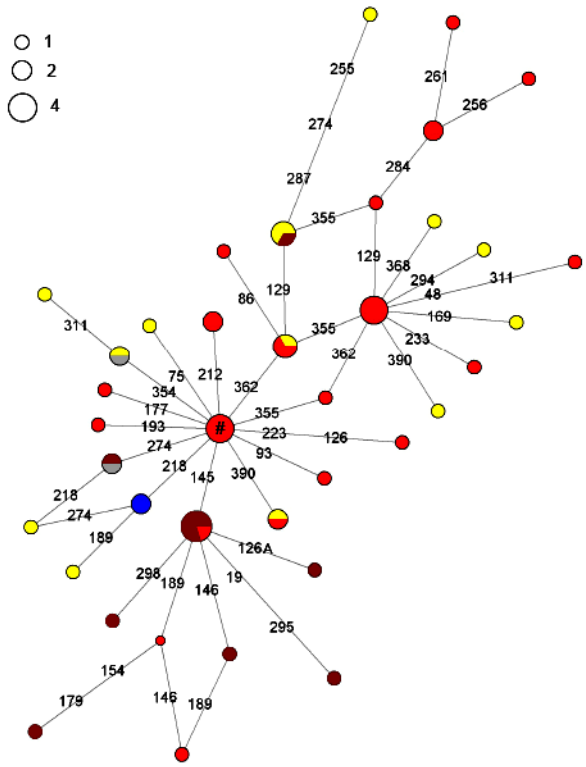


**Figure 3.9:** Network of African L2b lineages. Colours are as follows: Yellow, Central Africa; Red, Western Africa; Brown, Eastern Africa; Green, Equatorial Africa; Light Blue, Northern Africa and Western Tuareg; Grey, Southern Africa (SA); Dark Blue, Libyan Tuareg.

**L3e1.** The MJ Network of the L3e1 HVS-I lineages is shown in Figure 3.10. Default values of the weights were used (i.e. 10) except for np 16311, whose weight was set to six. Transversions were weighted three times more than transitions. A total of 69 HVS-I haplotypes are represented in the network.

The L3e1 root is characterised by the motif 16223-16327, and is geographically distributed as follows: 35.7% of the haplotypes is from Southeast Africa, 28.6% are Libyan Tuareg, and 21.4% are from Central Africa. The remaining fraction is represented by both Eastern and Western African haplotypes. All the four L3e1 Libyan Tuareg haplotypes are located in the root, and interestingly, a Western Tuareg individual departs from the root through the transition 16248. Differently from the previous network, L3e1 is characterised by the presence of many haplotypes from Southeast Africa, in particular as regards the branch with the

mutation 16185, and the one with the deletion at np 16325, which have been both already characterized as Bantu markers (Pereira et al. 2001, Salas et al. 2002). Some derivate haplotypes from Equatorial Africa have been also found. It is worth noting that despite the presence of some long branches, the L3e1 root appears to have a *star-like* structure. The coalescence ages were calculated: when the root and all its 1-mt step derivatives were considered in the calculation, a value of 14,000 +/– 5,900 years was found. When neglecting the two typically Southeast branches, which are assumed to be related to the Bantu expansion, the value of 11,400 +/– 4,400 years was observed.

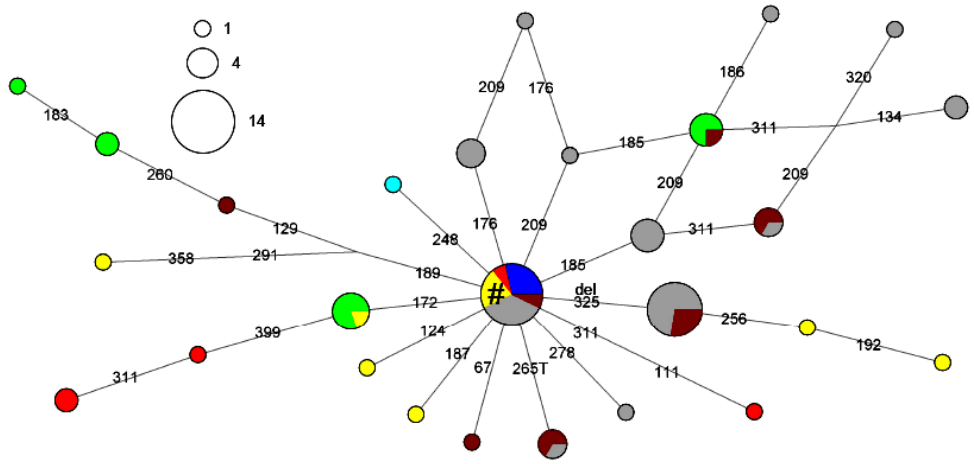


**Figure 3.10:** Network of African L3e1 lineages. Colours are as follows: Yellow, Central Africa; Red, Western Africa; Brown, Eastern Africa; Green, Equatorial Africa; Light Blue, Northern Africa and Western Tuareg; Grey, Southern Africa (SA); Dark Blue, Libyan Tuareg.

**L3e2.** The MJ Network of the L3e2 HVS-I lineages is shown in Figure 3.11. Default values of the weights were used (i.e. 10) except for nps 16172 and 16189, whose weights were set to six. A total of 98 HVS-I haplotypes are represented in the network. In the L3e2 network transversions were not over-weighted than transitions, because many sequences in the dataset harbour the transversion A16183C which is a homoplasic site.

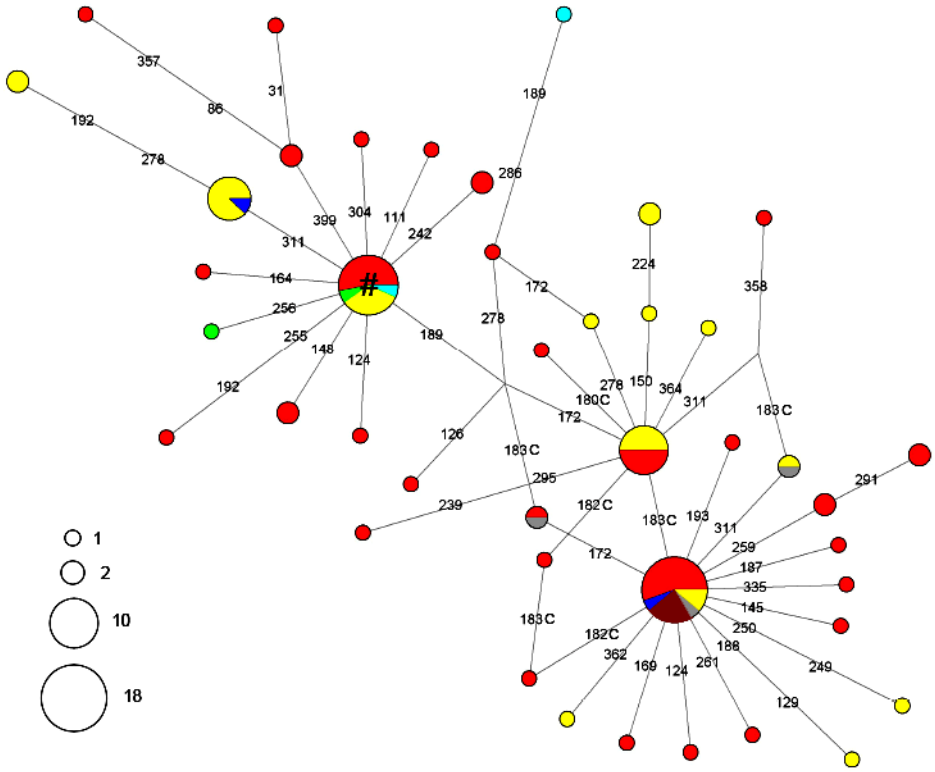


**Figure 3.11:** Network of African L3e2 lineages. Colours are as follows: Yellow, Central Africa; Red, Western Africa; Brown, Eastern Africa; Green, Equatorial Africa; Light Blue, Northern Africa and Western Tuareg; Grey, Southern Africa (SA); Dark Blue, Libyan Tuareg.

The root is defined by the HVS-I motif 16223-16320 and represents one of the three main nodes observed in the network. The other two significant nodes are the node departing from the root through mutation 16172-16189, which defines the sub-haplogroup L3e2b, and the one departing from the L3e2b basal node through the transversion A16183C. High homoplasy that characterises site 16183 generates some reticulations that were left unresolved as we think that more data need to be collected in order to ascertain the phylogeny of this haplogroup, in particular as regards lineages L3e2b (e.g. increasing the number of haplotypes, or complete mtDNA sequencing).

The L3e2 root is mainly distributed in Western Africa (53.3% of the lineages in the node) and Central Africa (33.3%). It is worth noting the presence of a Western Tuareg haplotype in the root. Most of the derivatives are 1-mt step distant from the root and are mainly from Western Africa, apart from the 16311-branch that appears

to be specific of Central Africa, and one Equatorial African haplotype harbouring mutation 16256. One Libyan Tuareg haplotype is present in the 16311 branch, together with other sequences from Central Africa, especially from the Fali population.

As previously mentioned, the L3e2b sub-haplogroup is represented by two main nodes separated by the transversion 16183C. The basal L3e2b node is equally distributed in Western and Central Africa, while the L3e2b-16183C node is widely spread in Western Africa (55.5% of the lineages in the node), but significantly high fractions are present even in Eastern Africa (22.2%) and Central Africa (11.1%). It is worth noting that most of the derivate haplotypes are mainly Western African single individuals 1-mt step distant from the L3e2b-16183C node, this resulting in a typical *star-like* structure. The coalescence ages were calculated when considering only the 1-mt step derivatives in the L3e2 root (11,300 +/– 5,300 years), and in the L3e2b-16183C node (8,400 +/– 2,700 years).

**L3e3.** The MJ Network of the L3e3 HVS-I lineages is shown in Figure 3.12. This is a small clade and only 30 haplotypes are represented in the network. Default weights were used for all the sites and transversions were weighted three times more than transitions.

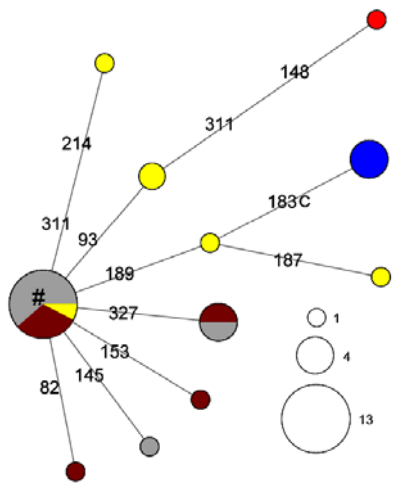


**Figure 3.12:** Network of African L3e3 lineages. Colours are as follows: Yellow, Central Africa; Red, Western Africa; Brown, Eastern Africa; Green, Equatorial Africa; Light Blue, Northern Africa and Western Tuareg; Grey, Southern Africa (SA); Dark Blue, Libyan Tuareg.

The basal node of L3e3 is defined by the HVS-I motif 16223-16265T. This is mainly distributed in Southeast Africa (61.5%), and Eastern Africa (30.8%), the remaining fraction being represented by only one Central African haplotype. Derivate haplotypes are both from Eastern/Southeast Africa and from Central Africa. The four L3e3 Libyan Tuareg haplotypes, which are characterised by the transversions 16183C, are placed in the 16189-branch, in which only Central African haplotypes are present.

**L3f1.** The MJ Network of the L3f1 HVS-I lineages is shown in Figure 3.13. It encloses a total of 75 HVS-I sequences. Transversions were weighted three times more than transitions and default weights were used for all the sites except for the following ones:

| Nucleotide position (np) | Weight |
|---|---|
| 16093 | 4 |
| 16126 | 6 |
| 16292 | 30 |

The network shows a typical *star-like* structure. The root is mainly distributed in Central Africa (46.9% of the haplotypes in the node), and Eastern Africa (43.7%). One haplotype from Western Tuareg and two from Western Africa are also present in the root. The two Libyan Tuareg haplotypes, both from the Tahala village, are located as derivatives departing from the root through the mutation 16218, and they appear to be the ancestors of a sequence from Mozambico (Southeast Africa), which is separated from them by the mutation 16256. Interestingly, the Western African haplotypes are mostly represented as derivatives from the root, and in particular the 16295-branch seems to be specific of West Africa. The coalescent age of L3f1 provided values of 8,000 +/– 2,100 years when considering only the 1-mt step derivatives from the root, and 11,800 +/– 2,300 years when including the 2-mt step derivatives.

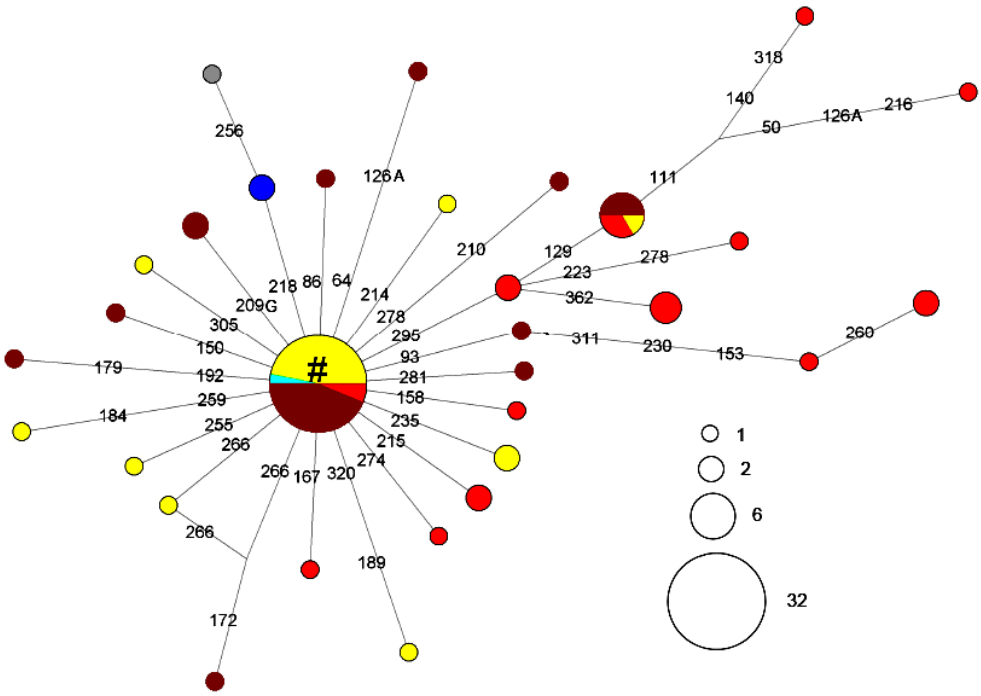


**Figure 3.13:** Network of African L3f1 lineages. Colours are as follows: Yellow, Central Africa; Red, Western Africa; Brown, Eastern Africa; Green, Equatorial Africa; Light Blue, Northern Africa and Western Tuareg; Grey, Southern Africa (SA); Dark Blue, Libyan Tuareg.

**L3w.** Haplogroup L3w has been only recently described. This appears to be rare and mainly distributed in Eastern Africa (Kivisild et al. 2004). The MJ Network of the L3w HVS-I lineages is shown in Figure 3.14.

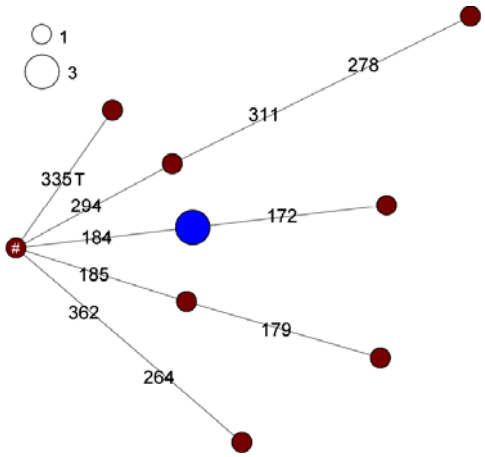


**Figure 3.14:** Network of African L3w lineages. Colours are as follows: Yellow, Central Africa; Red, Western Africa; Brown, Eastern Africa; Green, Equatorial Africa; Light Blue, Northern Africa and Western Tuareg; Grey, Southern Africa (SA); Dark Blue, Libyan Tuareg.

It includes only 11 HVS-I haplotypes, because few sequences have been so far collected in literature. Default weights were used for all the sites.

As expected, L3w clearly shows a Eastern African distribution. The L3w root is here represented by the motif 16223-16260-16311, and is represented by an Eritrean individual. The three Libyan Tuareg haplotypes depart directly from the root through the mutation 16184. One sequence from Somalia is a direct descendant from the Tuareg haplotype.

**M1.** The MJ Network of the M1 HVS-I lineages is shown in Figure 3.15. A total of 78 sequences are represented in the network. Default weights were used for all the sites and transversions were weighted three times more than transitions.

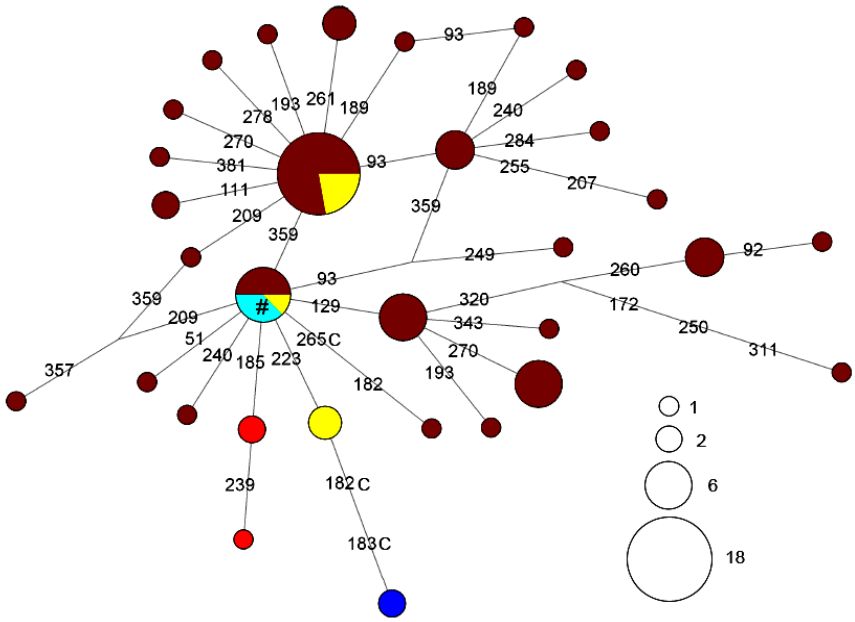


**Figure 3.15:** Network of African M1 lineages. Colours are as follows: Yellow, Central Africa; Red, Western Africa; Brown, Eastern Africa; Green, Equatorial Africa; Light Blue, Northern Africa and Western Tuareg; Grey, Southern Africa (SA); Dark Blue, Libyan Tuareg.

Geographic provenance of the haplotypes in the network clearly shows that M1 is mainly distributed in Eastern Africa, but it is worth noting the presence of some Northern Africa haplotypes from Berber populations in the root (37.5%). The root is defined by the HVS-I motif 16129-16189-16223-16249-16311. Three Central

African haplotypes depart directly from the root through the reversion at np 16223, and the two M1 Libyan Tuareg haplotypes, which harbour mutations 16182-16183C, are located in this branch as descendants of the Central African sequences.

## Star Likeness

Star Likeness was calculated as described in Chapter 2, and the values correspondent to the main coalescence events previously described in the Networks, are reported in Table 3.8, together with the TMRCA, ρ, σ, and n values.

It is worth noting that as regards the L2a lineages, values higher than 0.3 have been observed. Particularly the L2a-16189-16192 branch has a value equal to 1, this indicating a 'perfect star' phylogeny. High values of Star-Likeness characterize even L3f1 and the L3e2-16183 branch.

**TABLE 3.8**
Star Likeness values and the correspondent TMRCA for the main coalescence events described in the Networks of the African lineages observed in the Libyan Tuareg.

| NODE | TMRCA | ρ | σ | n | STAR LIKENESS |
|---|---|---|---|---|---|
| L2a-16209-16301-16354 | 7,500 +/– 3,800 | 0.375 | 0.191 | 24 | 0.43 |
| L2a-16189-16229-16291-16311 | 11,300 +/– 6,300 | 0.562 | 0.312 | 17 | 0.34 |
| L2a-16189-16192[a] | 7,400 +/– 2,800 | 0.25 | 0.094 | 28 | 1 |
| L0a-root | 9,400 +/– 3,300 | 0.46 | 0.16 | 84 | 0.21 |
| L1b-root (1-mt step) | 10,400 +/– 4,800 | 0.51 | 0.24 | 140 | 0.06 |
| L1b-root (2-mt step) | 17,040 +/– 4,900 | 0.84 | 0.25 | 183 | 0.08 |
| L2b-root | 17,190 +/– 5,740 | 0.85 | 0.28 | 27 | 0.39 |
| L3e1-root | 14,000 +/– 5,900 | 0.70 | 0.3 | 46 | 0.17 |
| L3e1-root | 11,400 +/– 4,400 | 0.55 | 0.23 | 31 | 0.35 |
| L3e2-root (without Bantu) | 11,300 +/– 5,300 | 0.56 | 0.26 | 34 | 0.23 |
| L3e2-16183C | 8,400 +/– 2,700 | 0.42 | 0.13 | 31 | 0.76 |
| L3f1 (1-mt step) | 8,000 +/– 2,100 | 0.4 | 0.1 | 54 | 0.67 |
| L3f1 (2-mt step) | 11,800 +/– 2,300 | 0.58 | 0.11 | 61 | 0.72 |

[a] The TMRCA calculated when maximally reducing the hypothetical effect of endogamy in the Libyan Tuareg is reported.

# Chapter 4. Discussion

## 4.1 ANCIENT DNA ANALYSIS AND HOT ENVIRONMENTS

The aDNA analysis has the potential to answer to paleonthological, archaeological, and anthropological matters, when the classical approach of these disciplines cannot do so. Water-soluble DNA has been shown to persist in fossil bones up to 130,000 years in temperate regions (Loreille et al. 2001). Nevertheless, DNA is at least partially degraded and chemically modified over the years, e.g.: oxidations (Hoss et al. 1996), enzymatic and non-enzymatic hydrolytic damages (Pääbo 1989, Lindahl 1993, Pääbo et al. 2004), intra-/inter-molecular cross-links (Poinar et al. 1998, Willerslev and Cooper 2005), and hydrolytic deaminations (Hofreiter et al. 2001a, Gilbert et al. 2003a, Gilbert et al. 2003b, Gilbert et al. 2006,). Consequently, despite some important achievements, the failure rate of paleogenetic investigations is high, because good DNA preservation is rare.

A trend in the success rate of DNA amplification and increasing average temperature has been experimentally observed: whereas 78% and 62% of permafrost samples were reported to be successfully amplified, samples from regions with moderate temperature were amplified with a 23-67% success rate, and from arid hot climates with a mere 2-4% (Pruvost et al. 2007). Temperature indeed represents a key factor in DNA preservation, despite it is not the only one: e.g., pH, oxidizing reducing Eh, irradiation, chemical composition of bone and soil, and hydrology, may play a role not yet well understood (Ovchinnikov et al. 2000, Pruvost et al. 2007, Smith et al. 2001).

In the past, many studies pointed to temperature as one of the main factor affecting the degradation of DNA, to the point that genetic analyses on samples from hot environment rarely resulted in positive results. A brief summary of the most important studies regarding the preservation of DNA in relation with temperature,

and the aDNA analyses carried on samples from hot environment so far, is reported in the following paragraph.

## Ancient DNA from hot environments: the story so far

1972. Lindhal and Nyberg produced a model of the DNA degradation rate by considering the *depurination rate* and consequently the breakdown of DNA strands. Depurination is a hydrolytic cleavage that affect the glycosidic bonds between the nitrous bases and the sugar backbone, resulting in abasic sites. The abasic site can undergo a chemical rearrangement that promotes occurrence of strand breakage. Generally, after putrefaction, DNA molecules undergoes chemical degradation, mainly based on depurination (Pruvost et al. 2007). The higher the temperature of burial conditions and the longer the time since the individuals were buried, the higher is the degree of DNA fragmentation.

1985. Pääbo reported that pieces of repetitive nuclear DNA could be cloned and sequenced from ancient human remains of an Egyptian 2,300 years old mummy (Pääbo 1985). A total of 23 mummies were analysed and only in one sample DNA was recovered. In detail, a fragment of 3.4 Kb was cloned. However, the amounts of DNA present in old tissues were so small that the isolation of bacterial clones carrying the same DNA sequence was not possible. So, the reproducibility of the results being lacking, authenticity of these data could not be achieved. Furthermore in 1998, Cooper and Wayne criticized this study, by stating that it was not possible to recover an aDNA fragment as long as the one found in the mummy, and consequently considered the results as artefacts (Cooper and Wayne 1998).

1991. Pääbo and Wilson (Sidow et al. 1991) use Lindahl's model on depurination rate to evaluate bacterial DNA survival in the Clarkia fossils. They observed that DNA degradation rate increases about threefold for every 10°C increase.

1994. Van der Kuyl et al. analyse 2,000-2,400 years old monkey's bones, collected in the museum of Saqqara (Egypt). Experimental results showed that, despite extractions controls were empty, high levels of contaminations characterised

the samples, as only human or birds sequences were found in the clones. Contaminations hypothetically occurred in the museum (van der Kuyl et al. 1994).

1996. Poinar et al. carried out a comparative analysis between aminoacid racemization and DNA preservation on samples from Egypt dating back to 2,300-4,500 years (respectively, one baboon, and five humans). No amplification of DNA was obtained in all the samples, and the racemization rate indicated that DNA hypothetically did not survive in hot environments (Poinar et al. 1996).

1996. Hoss et al. attested through gas chromatography/mass spectrometry (GC/MS), that DNA in the Egyptian monkey's samples dating back to the Tholemaic period (Poinar et al. 1996), was significantly modified, and consequently was refractory to amplification (Hoss et al. 1996).

1998. Crubezy et al. reported the recovery of DNA belonging to *Mycobacterium tuberculosis* in an Egyptian skeleton dating back to the Pre-Dynastic period, i.e. 5,400 years ago (Crubezy et al. 1998).

1999. Krings et al. carried out genetic analyses on skeletons and mummies dating back to the Pre-Dynastic and Early Pharaonic period. Rate of successful amplifications was very low (2/132) (Krings et al. 1999).

2000-2003. Zink et al. successfully amplified bacterial and mycobacterial DNA in Egyptian mummies, from the necropolis of Tebe West, and the Abydos site, which date back to 3,500-2,800 years ago, demonstrating that infection with *M. tuberculosis* was relatively frequent in Egypt. Furthermore, human aDNA regions, i.e. human β-actin and amelogenin, were amplified in 30 of 41 samples analysed (Zink et al. 2001). In their studies, the authors attested the feasibility of molecular studies on ancient Egyptian material (Zink and Nerlich 2003, Zink et al. 2005, Zink et al. 2003), nevertheless, authenticity of Zink's studies has been recently strongly criticised (Gilbert et al. 2005, Zink and Nerlich 2005).

2001. Graver et al. successfully recovered DNA from 34/50 individuals buried in a Roman cemetery (100-450 AD) in the oasis of Dakhleh (Western Egypt). The authors presented a preliminary report, but nonetheless provided clear evidences for the authenticity of their results (Graver et al. 2001).

2002. Marota et al. described a study on plant chloroplast DNA in modern and ancient papyri specimens in order to evaluate the half–life of aDNA in papyri, and in parallel they performed the racemization of Aspartic acid. The results were compared to the findings in animal and human remains from ancient Egypt in the literature. As a result, a strong and rapid destruction of aDNA was assumed. On the basis of the high temperatures in Egypt, an upper limit for DNA preservation of about 700-800 years was calculated, therefore suggesting that all the previous aDNA analyses on Egyptian remains were not authentic (Marota et al. 2002).

2003. Poinar et al. in 1998 observed that Pleistocene coprolites stemming from the extinct Shasta sloth (*Nothrotheriops shastensis*), contain mtDNA from the animal that produced them, as well as chloroplast DNA (cpDNA) from the ingested plants (Poinar et al. 1998). Nevertheless, the phylogeny of two families of extinct sloths was resolved only when three single-copy nuclear gene fragments were amplified from a 28,000 years old ground sloth coprolite from Gypsum cave, Nevada. It is worth noting that Thermal Age values of this sample were four times higher than the suggested theoretical limit for DNA retrieval. The author hypothesized that, despite the high temperature of the environment, depurination and deamination would also be retarded at very low levels of humidity. These results showed that aDNA can be recovered from warm, arid climates.

This short report of aDNA studies performed so far on samples from hot climate, suggests that the topic is rather controversial. As some of these studies provided successful results, high temperature seems not to always affect DNA preservation in the hot environment. Hypothetically, when particular chemical-physical conditions are present in the 'microenvironment' of a burial, effects of high temperatures might be reduced, as Poinar suggested in his work on sloth coprolite. Furthermore, it should be noted that most of the samples analysed in the aforementioned studies, were taken from museums, and in a recent paper (Pruvost et al. 2007), a significant increase in the DNA degradation rate due to storage at room temperature and to archaeological treatments, has been demonstrated. Nevertheless, still today, debate on DNA preservation in old bones is open, as very little is known about all the chemical and physical parameters that affect the degradation of DNA.

## State of preservation of the aDNA in the bones from the Libyan Sahara

As regards the results of the aDNA analyses accomplished in this study, the absence of amplification of authentic DNA in almost all the samples clearly indicates the very low state of preservation of DNA. Hypothetically, DNA is too fragmented to make the amplification at least of 124bp-long fragments possible (that corresponds to the shortest length of amplicons tested in the PCR experiments of the HVS-I_1 region, see Table A2a). Nevertheless, other kinds of damages that prevent DNA from amplifying (i.e. presence of hydantoins due to oxidation, or intra-/inter-molecular cross-links) cannot be ruled out. Very low amounts of endogenous DNA would hypothetically explain the results of cloning experiments in the samples which provided positive amplification reactions at the first round of PCR. Unknown sequences observed among the clones could be attributed to aDNA template molecules, but they were never reproduced in independent experiments, so that their authenticity and the presence of artefacts (e.g. TaqPol misincorporations or miscoding lesions) could not be tested. Furthermore, in one experiment, two *chimera* sequences were observed: it indicates the presence of extremely fragmented template molecules which undergo 'jumping PCR' events (Hofreiter et al. 2001a, Pääbo et al. 1990).

On the basis of the results of Thermal Age analysis, we would have expected to recover aDNA at least in the samples from the Fewet site. In fact, Thermal Age values for the samples from Fewet are not far from the accepted threshold, or even lower. This occurred when we considered the 'cave site' scenario, as envisaged for Fewet due to the fact that the corpses were recovered in intact burials dug at a reasonable depth in the soil, isolated from the surrounding environment (Castelli et al. 2005, see Chapter 3). By contrast, graves in Takarkori are laid in shallow pits just beneath the present ground surface, so that they were hypothetically more exposed to temperature fluctuations. Hypothetically, our assumption of the 'cave site' scenario for Fewet is wrong, but it is worth noting that the sample size is too small (i.e. four individuals) to completely rule out the possibility to recover aDNA in samples from this site. Racemization analysis confirms the very low state of

preservation of the bone samples, and shows that, at least in this case, DNA and proteins follow the same diagenetic trend.

It is worth noting that all the samples submitted to genetic analyses in the present study were taken from freshly excavated bones. It rules out the hypothesis of low DNA preservation due to treatments by archaeologists and storage in the museums (Pruvost et al. 2007), which could explain the unsuccessful results on many Egyptian samples in literature.

## 4.2 INFERRING PEOPLING OF LIBYAN SAHARA

Lack of the aDNA data from Fezzan, avoids to relate extant lineages found in the modern sample, to the people who occupied the same area in the past. In this sense, no hypotheses about continuity of genealogies in the Central Sahara can be done, and the reconstruction of peopling of the Acacus is consequently reduced to inferences based on data collected in the modern Tuareg sample, with the integration of information from other disciplines (e.g. archaeology, ethnology, and other genetic data from literature).

### Haplogroup composition and genetic diversity in the Libyan Tuareg

Results of the 95% Credible Region, showed that the two Tuareg samples from different villages are genetically similar, and consequently can approximately be considered as belonging to the same population. Significant differences were found in the frequencies of L1b1 lineages, but it could reflect defect of the sampling in the Tahala village. Nevertheless, increasing the size of this sample, would shed more light in the relationship between these two villages. In this study, we assume their genetic homogeneity, so that all results have been pooled in one single sample.

For phylogeographic purposes, it is important to characterize all the mitochondrial haplogroups found in the Tuareg sample, according to the data collected in literature so far:

**Haplogroup H1.** Haplogroup H is the most common haplogroup in all European populations and reaches its highest frequencies (40%-60%) in Western and Northern Europe. This haplogroup is also common in some population of the Near East and North Africa, and is observed in Northern India and among the Yakuts. In Europe haplogroups H is present at a frequency of 46%, while 25 % in the Middle East. Despite being more common in Europe than in the Near East, analysis of sequence divergence indicates that haplogroup H harbours a much higher diversity in the Near East than in Europe. Divergence values suggest an origin for haplogroup H between 25,000 and 30,000 years ago in the Middle East, and expanded into Europe before the Second Pleniglacial (15,000-20,000 years ago) (Otte 1990). The expansion of haplogroup H into Europe represents a second Paleolithic wave, that was contemporary with the diffusion of the Gravettian technology (20,000-25,000 years ago) (Torroni et al. 1998). Then, the cold and extremely dry conditions of the Second Pleniglacial caused an almost complete retreat of people from the central plains of Europe in two refuges: the Iberian peninsula in Western Europe, and Southern Ukraine in Eastern Europe (Otte 1990). After the LGM (Late Glacial Maximum), about 14,700-12,600 years ago, a warm period took place which allowed people to return to northern areas from the Franco-Cantabrian refuge (see Introduction for more details). As haplogroup H was one the most common haplogroup in the Iberian refuge, it marked together with haplogroup V the Postglacial Recolonization of Central-Northern Europe (Torroni et al. 1998). Recently, dissection of haplogroup H in sub-haplogroups through complete mtDNA sequencing, demonstrated that H1 and H3 were the main lineages involved in the Postglacial expansion toward Northeast and South. Their geographic distribution and coalescence ages (10,800 +/– 1,100 years, and 11,000 +/– 1,400 years, respectively for H1 and H3) confirm this hypothesis (Achilli et al. 2004).

**Haplogroup V.** Haplogroup V is closely related to haplogroup H: they are two *sister groups* that originated from a common ancestor (Torroni et al. 1998). Haplogroup V has a much more limited geographical distribution than haplogroup H, and is mainly observed in Northwest Europe and North Africa: it reaches high frequencies in some Iberian populations, and is also very common among the

Berbers of North Africa. The highest frequencies (40.9%) of V are observed among the Scandinavian Saami (Torroni et al. 1998). Nevertheless, the higher diversification of Iberian sequences suggests an origin in the Iberian peninsula, while the lack of variation observed in the Saami indicates a very recent founder event. The age of the haplogroup is estimated to be 11,200 +/– 2,700 years (Torroni et al. 2001a). The coalescence age and the geographic distribution of haplogroup V, depict a phylogeographic scenario analogous to that of H1, characterised by the Postglacial Recolonization from the Iberian Peninsula southward and north-eastward (Torroni et al. 1998).

**Haplogroup L0a1.** Haplogroup L0 is the earliest offshoot of the mtDNA tree in Africa. This appears as a *sister group* to the branch that holds all the other haplogroups found in the extant humans (see Figure 1.9b and 1.16). It includes four sub-clades (L0a, L0d, L0f, and L0k). A recent estimate of the coalescence age for L0 is 146,400 +/– 25,100years (Gonder et al. 2007). L0a is common in East, Central, and Southeast Africa (~20%-25%) and is almost absent in North, West, and Southern Africa (Salas et al. 2002). The TMRCA for L0a is 54,600 +/–5,700 years (Gonder et al. 2007). L0a1 is the main sub-clade, it has a quite star-like phylogeny and a predominantly East/Southeast African distribution. The root type for L0a1 is common in East Africa, and the coalescence age is estimated to be 33,350 +/–16,600 (Salas et al. 2002). L0a seems likely to have been brought to Southeast Africa into the Bantu community by the eastern stream of the Bantu expansion itself, having been picked up in East African non-Bantu speakers (Salas et al. 2002).

**Haplogroup L1b.** Haplogroup L1 is more frequent and diverse in West and Central Africa than in East Africa (Salas et al. 2002). Together with L0, L1 mtDNA haplotypes represent the most basal lineages of the human mtDNA gene tree. L1 appears to be slightly more recent then L0, as its TMRCA has been estimated as equal to 140,400 +/– 33,000 years (Gonder et al. 2007). Sub-haplogroup L1b is concentrated in West Africa, with some overflow in Central and North Africa, particularly in geographically adjacent areas connected by the West African coastal pathway. It is also common in Africa Americans, as a result of the Atlantic salve

trade. A Western origin of L1b, with significant diffusion into North and Central Africa, has been hypothesized (Salas et al. 2002). Nevertheless, it is worth noting that the coalescence time of L1b is 30,550 +/– 16,250, that is significantly lower than its sister clade L1c (i.e. 59,650 +/– 11,800) which likely originated in Central Africa. Consequently, the hypotheses of a more recent demographic scenario that shaped L1b phylogeny (e.g. bottleneck an re-expansion in West Africa), and an older Central African origin cannot be ruled out (Salas et al. 2002).

**Haplogroup L2a.** Haplogroup L2a is the most common and widely distributed South-Saharan African haplogroup and is also frequent in the Americas (~19%) (Salas et al. 2002). The wide distribution of L2a in Africa makes identifying geographical origins of lineages difficult. Moreover, recent analyses indicated the occurrence of marked homoplasy at multiple sites in the Control Region (e.g. 16189, 16192, 16309) (Howell et al. 2004), which confound the phylogeny of L2a. Furthermore, only few complete genomes have been sequenced so far (Ingman et al. 2000, Torroni et al. 2001b, Mishmar et al. 2003, Howell et al. 2004, Quintana-Murci et al. 2008), so that phylogeny of this haplogroup has not been clearly resolved yet. For this reason, in the present study nomenclature defined by Salas et al. (2002) is still followed. They defined two main clusters, one of them with the basal motif of L2a (i.e. HVS-I haplotype 16223-16278-16294-16390), and the other harbouring the transition at position 16309, that is diagnostic for haplogroup **L2a1**. Recently, the hypothesis of an ancestral L2a sequence carrying the transition at position 16309, and subsequent homoplasic events with reversion and additional forward mutations, has been even proposed (Howell et al. 2004). As regards the origin of L2a, Salas et al. (2002) calculated the TMRCA both in East (61,250 +/– 13,500) and West Africa (54,100 +/– 17,087), which appear to be almost similar. He proposed as possible solution an origin somewhere between East and West, followed by dispersals in both directions along the Sahel corridor. Sub-haplogroup L2a1 is further divided in two sub-clades: **L2a1a**, that is defined by transition at np 16286, and **L2a1b** characterised by transition at np 16290. Both of these appear to have a West African origin, and to have undergone a dramatic expansion either in Southeast Africa, or in a population ancestral to present-day Southeast Africans.

The L2a1a founder candidate dates back to 2,700 +/– 1,200 years (Salas et al. 2002). The very recent starbursts in L2a1a suggest a signature for the Bantu expansion (Pereira et al. 2001).

**Haplogroup L2b.** Haplogroup L2b appear to be largely confined, together with the L2c and L2d, to West and West-Central Africa, where they possibly originated (Salas et al. 2002). Some derived types are present in Southeast Africa too. In the context of L2 haplogroup, clusters L2b and L2c are the most recent, with an estimated TMRCA respectively of 31,600 +/– 11,200 and 27,500 +/– 7,250, while L2d is the oldest, with an estimated TMRCA of 121,900 +/– 34,200 (Salas et al. 2002).

**Haplogroup L3e.** Haplogroup L3e is the most widespread, frequent, and ancient of the African L3 clades. It hypothetically originated in Central Africa/Sudan region, about 45,000 years ago (Bandelt et al. 2001). Three of its four sub-clades (L3e1-L3e4), have been found in the Libyan Tuareg sample: L3e1, L3e2, and L3e3. Sub-haplogroup **L3e1** is distributed throughout South-Saharan Africa, but is especially common in Southeast Africa, among Bantu speakers, while is rare in West Africa. Salas et al. (2002) hypothesized a West-Central African origin for L3e1. It hypothetically spread from there to Kenya through the eastern Bantu stream, and successively reached the South-Eastern regions of Africa. Nevertheless, data are quite scarce, and a back migration from Southeast northward to Kenya cannot be ruled out (Salas et al. 2002). The TMRCA estimation for L3e1 is 32,150 +/– 11,450. Sub-haplogroup **L3e2,** is mostly distributed in Central Africa and West Africa. Its sub-clade derivative, L3e2b, is found primarily in West Africa, indicating an hypothetical westward expansion from Central Africa about 9,000 years ago (Salas et al. 2002). Coalescent Age has been calculated for L3e2 to be 37,400 +/– 18,350 years, while for its sub-clade L3e2b is 9,150 +/– 3,100 (Salas et al. 2002). Sub-haplogroup **L3e3,** is a small clade that is mainly distributed in Central, Eastern and Southeast Africa. Its root type is spread at elevated frequencies in Southeast Africa, together with some derivatives. A possible connection with the

eastern Bantu stream has been hypothesized. The TMRCA has been estimated to be 14,150 +/– 4,500 years (Salas et al. 2002).

**Haplogroup L3f.** L3f is quite rare and is spread mainly in East Africa. Sub-clade **L3f1**, defined by transition at position 16293, hypothetically spread into West Africa at an early date, and is correspondingly well-represented in African Americans. The TMRCA for L3f1 has been estimated to be 28,650 +/– 8,650 (Salas et al. 2002).

**Haplogroup L3w.** Haplogroup L3w has been recently defined by Kivisild et al. (2004). It is characterised by the HVS-I motif 16223-16260, that it shares with haplogroup L4a. The transition 15388, makes it possible to distinguish L3w from L4a (Kivisild et al. 2004). Despite the site 15388 being not screened in the present study, three Tuareg sequences harbouring the motif 16223-16260, were classified as L3w because they lack the mutation 16362, which appears to be diagnostic for haplogroup L4a at least in the Ethiopian sequences by Kivisild et al. (2004). Nevertheless, HVS-I motif 16223-16260 seems to be restricted to East and Northeast Africa, where it can be detected at low frequencies (Krings et al. 1999, Kivisild et al. 2004). No estimations of the coalescence age are yet available in literature for L3w.

**Haplogroup M1.** Haplogroup M is almost absent in West Eurasia, while it is dominant in Asia. Nevertheless, its sub-clade M1 is present at high frequencies in the Horn of Africa and appears to be predominantly African specific. Complete mtDNA sequencing demonstrated that the presence of this lineage in Eastern Africa is the result of a back migration from Southwest Asia, where it arose, suggesting a return to Africa of populations carrying M1 and U6 from the Mediterranean area about 40,000-45,000 years ago. The coalescence age for M1 lineages dates back to 36,800 +/– 7,100 years (Olivieri et al. 2006). Kivisild et al. (2004) observed that M1 lineages constitute 17% of the Ethiopian mtDNA sequences. It is worth noting that M1 is rare or absent in North Africans (Corte-Real et al. 1996, Rando et al. 1998, Brakez et al. 2001, Plaza et al. 2003), but its sub-clade M1c, defined by mutation at position 16185, covers most of haplogroup M1 variation in Northwest

Africa, the Canary islands, and the Near East, while it has not yet been found in Ethiopians (Kivisild et al. 2004). High resolution phylogenetic analysis of M1 lineages, indicates that presumably Northwest African haplogroup M1 variation cannot be interpreted as a derivative of the East African mtDNA pool (Maca-Meyer et al. 2001, Kivisild et al. 2004).

The Tuareg sample as a whole appears to be extremely homogenous, as the estimate of the Gene Diversity indicates. Its value is the lowest ever observed (0.677 +/– 0.046) in the African populations analysed so far, and only 20 HVS-I different haplotypes (K value) have been found in a total of 129 Tuareg individuals. The reason for this low genetic diversity, at least at the mtDNA level, can hypothetically be explained by the high endogamy that characterises the Tuareg. In fact, as described in the Introduction, tribal organization that characterized Tuareg at least up to the colonial period, permit marriages only between couples from the same tribe, especially when they are nobles. Furthermore, it is worth noting that the strict rules that had to be followed after marriage, led women to come back in their mother's territory after the death of the groom: in this sense, the same matrilinear descent is always maintained in the strict local context of the camps, and mtDNA lineages appear to be characterised by high matrilocality.

Like Gene Diversity, the values of $\pi$ and Nucleotide Diversity in the Libyan Tuareg are the lowest observed in the African samples of the database. This reflects the high fraction of HVS-I CRS sequences present in the Tuareg sample (55.81 %). Comparable values of $\pi$ and Nucleotide Diversity are observed exclusively in Berber groups (Matmata and Moroccan Berbers), where a high incidence of H-CRS haplotypes has been observed as well (see figure 3.3 and table 3.5). Such affinity strongly suggests the provenance of this lineage from Berber groups, and the hypothesis of a foundation event cannot be ruled out. Interestingly the other Tuareg sample present in literature (that is called here 'Western Tuareg'), has both Gene Diversity and other indices similar to those of the other African populations. Such genetic heterogeneity contrasts with that observed in the Libyan sample, and could hypothetically result either from the fact that the Western Tuareg sample include

individuals from different areas, or from the dissolution of the Tuareg socio-political system during the colonial period, when many confederations were disassembled and people from different tribes mixed each other. Nevertheless, the scarce information that we have about this sample, does not make possible to infer a precise hypothesis.

The high genetic homogeneity of the Libyan Tuareg sample is confirmed when calculation of Standard Diversity Indices and Molecular Indices is extended to the HVS-I/HVS-II haplotypes. In fact, 21 different haplotypes were observed, and Gene Diversity value is almost the same (0.678 +/– 0.046). As the molecular diversity contained in the HVS-II is now considered, the Mean Number of Pairwise Differences ($\pi$) significantly increases (7.769 +/– 3.642 vs 4.398 +/– 2.186), while the value of Nuclotide Diversity is slightly lower (0.00997 +/– 0.00517 vs 0.01099 +/– 0.00605), that is quite reasonable as the HVS-II is less variable than HVS-I.

Analyses of some polymorphic sites in the coding regions in the Libyan Tuareg sample, made it possible to classify all the HVS-I CRS sequences as H1. Haplogroup H1 appears to be strongly predominant in the Tuareg sample, with frequencies higher than 60%. The H1 and V lineages represent the 'West Eurasian' genetic component, and together they constitute almost two-thirds of the mtDNA genetic pool of the Libyan Tuareg. The geographical pattern of distribution of these lineages, suggests that they spread into Northwest Africa after crossing the Strait of Gibraltar during the expansion from the Franco-Cantabrian glacial refuge, at the end of the Pleistocene (Torroni et al. 1998, Achilli et al. 2004). It is interesting to note that the V lineages are present at significantly high frequencies in the same Berber groups where a high incidence of HVS-I H-CRS haplotypes has been observed (up to 16.33% in the Matmata, see table 3.6). It suggests that the H-CRS component, despite not being defined at the sub-haplogroup level in many North African populations, can be likely considered as a H1 marker at least in the Berber groups; this would support the hypothesis that Libyan Tuareg are somehow genetically related to Berbers, at least as the 'West Eurasian' component is concerned.

Beside the 'West Eurasian' component, the typical South-Saharan one is present. Despite its being more heterogeneous, at least as regards the haplogroup composition, a certain intra-haplogroup homogeneity of lineages is observed. As lineages from both West, Central, and East Africa are observed, it can be hypothesized that they were introduced into the Central Sahara at different times, through different demographic events. While for the most represented lineages (i.e. L2a and L0a1) some kind of significant migratory shift can be hypothesized, the others, whose frequencies are much lower, could have been introduced through sporadic events, like 'erratic' haplotypes.

## Analysis at the population level: comparison with other African samples

The relationship between Libyan Tuareg and the other African populations collected in the database was investigated through MDS and CA.

**Multidimensional Scaling**. In the MDS a geographic partition of population has been observed (figure 3.4a). Berbers from North Africa are grouped on the left side, this position being likely due to their common 'West-Eurasian' component. The Libyan Tuareg sample is placed on the left side of the two-dimensional plot, together with Berber populations. More in detail, its close relationship to Berber groups (i.e. Matmata and South Moroccan Berbers), confirms their high genetic affinity, at least at the mtDNA level, this likely related to their common H-CRS component. The other Tuareg sample (i.e. 'Western Tuareg'), appears to be significantly different from the Libyan one, as already observed in the Diversity Indices. In fact, in the two-dimensional plot the Western Tuareg sample is found in the border region between Western and Central African population, in close relationship with Kanuri from Nigeria, and Western Guineans. The particular position of Western Tuareg suggests that they were affected by an influx of lineages from the surrounding populations.

When discarding the 'West Eurasian' component (i.e. H1 and V haplotypes) from the Libyan Tuareg sample, it shifts in the two-dimensional plot to the upper right, together with Eastern African populations, near the Kikuyu and Kenya (figure

3.4b). This kind of representation suggests a Eastern provenance of the South-Saharan component in the Libyan Tuareg. In detail, it is worth noting that two Central Africa samples (i.e. Kotoko and Mafa) are located in the same region of the plot. It could suggest that they share the same Eastern mtDNA component. The Kotoko and Mafa from Northern Cameroon were recently analysed by Cerny et al. (2004, 2007), who attributed their genetic proximity to Eastern Africa populations, to the significant fraction of L0a, and L3f lineages, which we found in the Libyan Tuareg sample as well. The authors proposed that the Chad basin was the epicentre of a bidirectional 'genetic' corridor between East and West Africa. In particular at the beginning of the Holocene, Lake Megachad and its surrounding plains, could have attracted humans from East and West, who found there fertile regions where to settle.

**Correspondence Analysis.** In the CA the relationships between African populations are represented in terms of their haplogroup composition. The *outlier* position of the Tuareg sample is strongly affected by their high frequency of H-CRS lineages (figure 3.5a). These lineages were all classified as H1 in the Libyan Tuareg, and represent together with haplogroup V, the West Eurasian component that hypothetically come from the Iberian peninsula (Torroni et al. 1998, Achilli et al. 2004). As previously described in the Results section, a trend from East African populations (at the bottom of the graph) to North African ones (at the top of the graph) can be observed along the second dimension of the plot. Nevertheless, in the context of North African populations, it is possible to distinguish between those where lineages from West Eurasian are observed (i.e. Libyan Tuareg and in part Matmata Berbers, in relation with H-CRS and V lineages), and those in which Near Eastern lineages are found (i.e. the Egyptian and the other Berber samples, in relation with preJT/JT, U5/U5a, U/K haplogroups). Such a distinction reflects the two different 'extra-African' components that affected North African populations in different ways and times: the Near Eastern and the West Eurasian ones. The former, might be hypothetically related to both prehistoric (e.g. the incoming of Pastoral groups from Near East) and historic migration events (e.g. the Arabic Diaspora since the 7$^{th}$ century A.C.). Beside this, the typically Northern Africa haplogroup

U6 is observed, that arose in Southwest Asia and moved to Africa about 40,000 to 45,000 years ago (Olivieri et al. 2006). In contrast, the Western Eurasian component, as previously explained, is associated with the Post-Glacial re-expansion from the Franco-Cantabrian refuge (Achilli et al. 2004, Torroni et al. 1998).

Interestingly in the CA that we performed, H sub-haplogroups were not considered because of the scarce data so far available in literature (Achilli et al. 2004). For this reason, it is important to note that, in both the variable H-CRS and H that we considered, no distinction between 'West Eurasian' H lineages (i.e. H1 and H3), and Near Eastern H lineages can be done for all the samples. Nevertheless, the H-CRS haplotypes in North Africa can be likely considered of West Eurasian origin, as their frequency in Northwest Europe is higher than in the Middle East and even in North Africa a decreasing frequency cline from West to East can be observed (Richards et al. 2000). Under this assumption, results of the CA suggest once more that the strong genetic affinity previously observed between the Libyan Tuareg and some Berber groups (particularly the Matmata), is strictly related to their common mtDNA H1 component. No significant relationship between the Libyan Tuareg and the other Tuareg sample present in literature (Western Tuareg) is observed, despite the fact that it presents some H-CRS lineages (8.3%). Western Tuareg are placed in the left side of the two-dimensional plot, indicating that it is characterised by a typical South-Saharan haplogroup composition, which reflects that of the surrounding populations.

## Gene genealogies: coalescence of the Tuareg lineages

Coalescent theory focuses on genealogical descriptions of the ancestry of a number of copies of the same haplotype (Kingman 1982, Hudson 1990). This set of ancestral relationships is known as the *gene genealogy*. As we move backwards in time through the generations, we will start to encounter haplotypes that are ancestral to two existing haplotypes. The merging of lineages as we go backwards in time is at the basis of the coalescence process (Jobling et al. 2004). In the present study gene genealogies were investigated for most of the Libyan Tuareg haplotypes. More

particularly, as regards the West Eurasian lineages, a high resolution phylogenetic analysis was carried out through the sequencing of the whole mtDNA in five H1 individuals. The whole sequencing of mtDNA made possiple to investigate the ancestral relationships within the Tuareg haplotypes as well as with other complete mtDNAs in literature, this being impossible when limiting the analysis to the HVS-I CRS haplotypes. As regards the African lineages, the Network analysis on the HVS-I haplotypes was accomplished (see Chapter 2 and 3 for details).

**Western Eurasian lineages: H1 and V**. Complete mtDNA sequencing carried out on five H1 Tuareg individuals, indicates a certain degree of diversification of H1 lineages in Northern Africa. In fact, they are all different and tend to cluster in Northern African specific sub-haplogroups. It supports the hypothesis of an early arrival at the beginning of the Holocene of the H1 lineages in Northern Africa, this being a consequence of the Post-Glacial expansion from the Iberian Peninsula. Furthermore, the topology of the tree, that is characterised by many haplotypes departing directly from the root, clearly reflects the star-like structure of the haplogroup H1. This is an evidence of the expansion event that involved the H1 lineages at the beginning of the Holocene, as previously reported in literature (Pereira et al. 2001, Achilli et al. 2004, Roostalu et al. 2006).
As regards the V lineages, no genealogical was carried out. Nevertheless it is worth noting that the HVS-I haplotype 16189-16298 is shared with six sequences from Europe, three of which are from a Basque sample (Richards et al. 2000). We think that it is a further significant proof of the linkage between the West-Eurasian lineages in Northern Africa and the Post-Glacial expansion from the Iberian refuge.

**African lineages.** The network analysis was carried out on the African lineages (all the L lineages and the M1) found in the Libyan Tuareg samples. In this analysis the distribution of a particular lineage, with its pattern of mitochondrial variable sites, is observed over a wide range of populations within the entire continent. This phylogeographic approach makes it possible to reconstruct the direct ancestor-descendant relationship that stands between the haplotypes, and so provides information about the likely geographic provenance of a given lineage. Furthermore, the particular topology of a network, can tell us more about

demographic events (e.g. expansion events) and the period to which they date back, through calculation of the ρ statistic (i.e. coalescence ages).

**Genealogy of L2a lineages.** L2a lineages constitute the main component within the African lineages of the Tuareg sample, with a frequency of 8.53 %. As previously observed, L2a is widely distributed in all South-Saharan Africa, and it makes difficult to precisely resolve its phylogeny, at least at the HVS-I haplotypes level. In fact, haplotypes from Western, Central, and Eastern Africa are concurrently present at significantly high frequencies in the network. Nevertheless, their particular distribution in the nodes could give some important indications, in particular as regards the origin of the Libyan Tuareg lineages. A Central African origin of L2a could be hypothesized on the basis of the high fraction of HVS-I haplotypes in the root. Then, it likely spread to West, and East Africa as well: in fact, many derived Western haplotypes depart from the root, and they even constitute a significant fraction of it (33.3%). Nonetheless, despite lacking Eastern haplotypes in the root, an Eastern African origin cannot be totally rejected too: a high fraction of Eastern African haplotypes is in fact present in the derived 16189-node, and it should be considered that many factors could affect distribution of the haplotypes in the nodes: migrations, local population expansions, or even sampling effects. This is particularly true for such a old and widely distributed haplogroup as L2a, especially when limiting the analysis the HVS-I.

As reported in figure 3.6, two branches show a exclusive Central/Eastern African distribution: the 16209-16301-16354 branch from the root, and the 16189-16229-16291-16311 one. Geographic distribution of the haplotypes (data not reported) in the two basal nodes of these branches, strongly suggests a close genetic relationship between the Sudan area and the Chad basin. It could result from the migration westward of Eastern African people that was hypothesized by Cerny et al. (2004, 2007), which was even attested by linguistic and archaeological data (see Introduction).

The Libyan Tuareg haplotypes represent the main fraction (45.4%) of the 16189-16192 node. This particular distribution is likely biased by the effect of endogamy in the Tuareg sample, that contributed to increase the frequency of this haplotype in

the Tuareg in comparison to the other African populations. Nonetheless, a significant fraction is also represented by Eastern haplotypes (18.2%, half of them from the Sudan region), and Western Africa as well (13.6%). In trying to reconstruct a possible scenario for the origin of the Libyan Tuareg L2a lineages, it is worth noting that the ancestor node of the L2a-16189-16192 haplotype (i.e L2a-16189) is mainly distributed in Eastern Africa, particularly in the Sudan region. This opens the possibility that the L2a Libyan Tuareg lineages are from Eastern Africa. The same result was found by Cavalli-Sforza too (1997), who observed a strong genetic relationship at the nuclear DNA level between Tuareg and the Beja populations, whose separation dates back to about 6,000 years ago. Furthermore, we know from multidisciplinary data collected so far, that a westward flow from Eastern to Central Africa, is documented by archaeological, genetic, and linguistic data: this hypothetically originated in the Sudan area, where the Beja population settled about 6,000 years ago. It is likely that the L2a Tuareg lineages are somehow related to this movement: whether they separated from the main flux that led people to Chad basin (that would explain the absence of Central African haplotypes both in the Tuareg node and its ancestor L2a-16189 node), or whether they are from a small group of people which directed northward once it reached the Chad area, is impossible to ascertain. More data needs to be collected, both at the level of population samples, and at the level of high-phylogenetic resolution (i.e. complete mtDNA sequencing).

It is worth noting that the particular position in the network of the other Tuareg haplotypes described in literature, the Western Tuareg, confirms the relationship with Central/Eastern Africa, and suggests some kind of genetic affinity between Tuareg from different areas, at least as regards this lineage.

The coalescence age calculated from the L2a-16189-16192 root and all its derivatives, provides values (i.e. 5,045 +/– 1,900) compatible with the scenario of a spread from East Africa in the first half of the Holocene. It is worth noting that the coalescence value could be affected by the fraction of Tuareg haplotypes in the basal node, that is assumed to be due to the high endogamy. As reported in table 3.7 and 3.8, TMRCA values without the hypothetical effect of endogamy have been

calculated, and higher values of the coalescence times are observed (i.e. 7,400 +/– 1,800). This result suggests that when assuming the endogamy effect in the Libyan Tuareg population, this underestimates the coalescence age. Nevertheless, despite dating shifting back in time, the hypothetical westward human migration from East is supported as well. Star-Likeness value was found to be equal to 1; this indicates a 'perfect star' phylogeny, and the very rapid expansion event which characterized this lineages. As regards the two Central/Eastern African specific nodes, significantly high coalescence times were observed. Nevertheless, under the assumption that these data could be affected by the presence of more ancient lineages present since the Pleistocene, the same calculation was repeated when considering only the 1-mt step derivatives; the results appear to be more compatible with our scenario of a diffusion from the East in the first half of the Holocene: 7,500 +/– 3,800 years, and 11,300 +/– 6,300 years. Both Star-Likeness values correspondent to these TMRCAs, appear to be higher than 0.3, indicating a rapid expansion event.

**Genealogy of L0a1 lineages.** L0a1 lineages constitute the second main African component within the Tuareg mtDNA genetic pool, with a frequency of 6.2 %. As previously mentioned, L0a1 has been described in literature by Salas et al. (2002) as mainly distributed in East/Southeast Africa. In fact, topology of the network shows the effects of the Bantu expansion, in particular concerning the L0a-16129-16168 and the L0a-16278-16293 nodes, in which many Southeast African haplotypes are located (represented in grey).

In comparison to the work of Salas (2002), the network analysis performed in this study include the latest published mtDNA sequences, enclosing many Western and Central African samples recently characterised (Rosa et al. 2004, Cerny et al. 2007). Topology of the network (figure 3.7), shows that Western and Central African haplotypes represent a significant fraction in the root (West 34.8%, and Central 17.4%), but the Eastern ones constitutes the main component (43.5%). It is worth noting that most of the Western haplotypes are mostly from one single country (i.e. 75% from Guinea, Rosa et al. 2004), and the Eastern African haplotypes as well (Upper/Lower Egypt, 60%). The Central African haplotypes are all from the Lake

Chad area (mainly Arabs from Chad, Kotoko, and Masa). The *star-like* structure of the L0a1 root suggests a population expansion event in the past, which led people to spread in different directions, including the Central Sahara. Although the spread distribution of the haplotypes in Western, Central, and Eastern Africa, makes it difficult to find the origin of this hypothetical expansion, integration with data from other studies can help to find the most likely scenario. Two hypothesis can be done:

Hypothetical expansion from Northeast Africa: archaeological records collected from the Italo-Libyan joint mission attest the arrival of Pastoral groups in the Acacus (Central Sahara), in the middle of the first half of the Holocene. This hypothesis is supported by the presence of ovicaprines remains in the Acacus mountains at least 7,200 years ago (see Introduction). It coincide with the beginning of the Pastoral period in the Fezzan, as described from paleonthological surveys (see Introduction). This hypothesis is compatible with the presence of Northeast African haplotypes in the root, these being mainly from Upper/Lower Egypt (more in detail, Nubia), since it can be hypothesized that Pastoral groups from Middle East, before moving to Central Sahara, stopped in the fertile areas of the Nile area, in Egypt. This hypothesis indicates some kind of genetic relation between the Pastoral groups inhabiting the Fezzan since the first half of the Holocene and the extant Tuareg.

Hypothetical expansion from Central Africa: both archaeological and genetic data support the hypothesis of an expansion from the Lake Chad. Archaeological records (Haour 2003) attest that people from the Chad basin spread northward between the end of the Pleistocene and the beginning of the Holocene. Furthermore, mtDNA data recently collected from populations of the Chad basin (Cerny et al. 2004), support this hypothesis, in particular as regards the lineage L3e5, which hypothetically originated in Central Africa, and spread through a population expansion about 7,100 +/– 3,800 years ago to Northern Africa, where it has been found in Berber groups (Fadhlaoui-Zid et al. 2004, Cerny et al. 2007). Hypothetically, the Tuareg L0a1 lineages derive from this northward shift of human groups from the Chad basin; it is likely that this migration is related to a population

expansion associated with the new foraging opportunities (e.g. fishing and hunting ) that characterised the Chad basin during the wet period, in the early Holocene.

Both these scenarios are chronologically placed between the end of the Pleistocene and the beginning of the Holocene, and consequently are compatible with the coalescence age that was calculated for the L0a1 node and its 1-mt step derivatives (9,400 +/– 3,300). Relationship of the L0a1 Tuareg haplotypes with a hypothetical expansion from Western Africa (e.g. from Guinea) is less likely, as at present no further data from other disciplines or other genetic studies support a movement of people from West Africa to the Central Sahara. Nevertheless, only more refined phylogenetic analyses such as the complete mtDNA sequencing, could establish more precisely the exact provenance of the L0a1 Tuareg lineages, and the geographical origin of a hypothetical expansion event in the context of L0a1 lineages. Concerning the Star-Likeness calculated for the L0a1 lineages, it should be noted its low value is an indication of a slow expansion event. It contrasts with the high Star-Likeness values observed for the L2a lineages, which have been hypothetically linked to a rapid expansion event from Central Africa. In this sense, the hypothesis of a Eastern African origin of L0a1 lineages appears to be more likely, despite the existence of two different expansion events (a rapid and a slow one) that separately involved L0a1 and L2a lineages cannot be ruled out under the scenario of a Central African origin.

**Other networks.** It should be noted the other African lineages all represent a minor component in the Libyan Tuareg mtDNA genetic pool. When dealing with such low distributed lineages, it is possible to relate them to significant demographic events; nonetheless we should take into consideration the hypothesis that they were sporadically introduced as 'erratic lineages', even recently.

**L3e1, L3e3.** Both these haplogroups represent 3.1 % of the African lineages in the Tuareg sample. As regards L3e1, the topology of the network is affected by the recent Bantu expansion, which represented a very significant demographic event in the context of this lineages: in fact, a significant fraction of root-haplotypes are from Southeast Africa, and two typically Bantu branches are observed: the L3e1-16325del, and L3e1-16185, that can be both considered as Bantu-markers (Pereira

et al. 2001). However, the presence of Central African haplotypes in the root (21.4%), and of Equatorial individuals as well, strongly supports the hypothesis made by Salas (2002) of a West-Central African origin of L3e1. Many of these lineages may have taken part about 3,000 years ago in the Bantu expansion from West-Central Africa, which drove them to East/Southeast Africa. The position of the Libyan Tuareg haplotypes in the root clearly suggests a relationship with Central Africa. This kind of genetic affinity is likely related to a demographic expansion that took place in the Chad basin between the end of the Pleistocene and the beginning of the Holocene, as the coalescence ages indicate: 14,900 +/– 5,900 when considering all the 1-mt step derivatives, and 11,400 +/– 4,400 when discarding the two typically Bantu branches, which likely originated later, hypothetically in the context of the Bantu expansion. These results would support the hypothesis of a demographic expansion in the Chad basin during the wet period, between the end of the Pleistocene and the beginning of the Holocene. This likely led people to move northward, and to reach the Central Sahara, which at that time appeared as a verdant flourishing region. The Star-Likeness value equal to 0.35 for the L3e1 lineages (excluding the typical Bantu branches), suggests that it was a quite rapid expansion.

Significantly the network of L3e3 lineages, that represent 3.13% of the African haplotypes in the Tuareg sample, strongly supports the close genetic affinity between the Tuareg sequences and Central Africa as well. In fact, the Libyan Tuareg haplotypes are placed in a typical Central African branch, as derivatives of a Hausa sequence. The high distribution of Eastern/Southeast African haplotypes in the network, indicates once again the strong Bantu component. Similarly to L3e1, we can hypothesize a West/Central origin of this haplogroup, although an Eastern origin could be likely as well. High resolution phylogeny is needed to resolve this matter.

**L1b1, L2b, L3e2.** These lineages represent, like L3e1 and L3e3, a minor component in the African mtDNA pool of the Tuareg sample (respectively 3.88 %, 1.55 %, and 1.55 %). Topology of these networks as a whole, suggests a derivation of these lineages from both Western and Central Africa. More particularly, it is of

note that one L3e2 haplotype shares the same pattern of mutations of individuals from the Chad basin, in particular from the Fali population. The L3e2b Tuareg haplotype is located in the 16172-16183C-16189 node, which is mainly distributed in West Africa. The particular *star-like* structure and the coalescence ages of the L3e2 root, and of the L3e2b-16183C node, indicate an expansion event that hypothetically took place in Western Africa between the end of the Pleistocene, and the beginning the Holocene. Nonetheless, the significantly different Star-Likeness values calculated in the L3e2 root, and in the L3e2b-16183C node, hypothetically indicate that two different expansion events occurred: the first was quite slow, and was followed by a more rapid one. The coalescence ages, and the Star-Likeness observed in the L2b network, that appear to be unambiguously of West African origin, support the hypothesis of a rapid expansion in West Africa. Whether the presence of L3e2 and L2b lineages in the Central Sahara could be related to such West African expansions is not possible to ascertain: as previously mentioned, no other data exist about such a hypothetical migration eastward from West, and furthermore the hypothesis that the Libyan Tuareg lineages were introduced through sporadic events, like slavery, is very likely. In fact, historical documents attest that slavery practices led Tuareg up to Timbuktu (Mali), which was one of the main centres for the slave trade in Northern Africa (Tornieri 2002).

The same interpretation could explain the presence of the L1b lineages in the Libyan Tuareg. Distribution of the haplotypes in the root of the L1b network indicates a Western origin, although, as previously mentioned, a Central African origin cannot be ruled out (Salas et al. 2002). Significantly, most of the Central African haplotypes in the L1b1 node are from nomad Fulani groups, hypothetically as a result of their mating pattern which led to a reduction in genetic diversity and to high frequencies of this lineage in the population (Cerny et al. 2006). The coalescence age, and the Star-Likeness of the L1b1 node indicates a slow expansion event which took place at the end of the Pleistocene hypothetically in Western Africa, as already observed for the L3e2 lineages. The results collected here on L1b, L2b, and L3e2 lineages, clearly show their origin in Western Africa, but

demographic human dynamics that shape their phylogenies need to be further investigated.

**L3f1, L3w, M1.** They respectively represent 1.55 %, 2.33 %, and 1.55 % of the mtDNA pool in the Libyan Tuareg sample. The Tuareg L3w lineages appear to be unambiguously of Eastern African origin. They could have been hypothetically introduced through the same movements of people that led in the Central Sahara the L0a1 lineages, though a more recent sporadic introduction cannot be excluded. Concerning L3f1, of particular interest is its high distribution in Eastern and Central Africa. This supports the hypothesis of a westward movement of human groups from East Africa to the Chad basin at the beginning of the Holocene, as previously observed in the L0a1 network. The coalescence ages of L3f1 are compatible with such a scenario, while the Star-Likeness value indicates a rapid expansion event.

The network of M1 lineages clearly shows the typical East African distribution of this haplogroup. In the root, three haplotypes from Northern Africa have been found, two of which belong to Berbers from Tunisia. The particular position of the Libyan Tuareg haplotypes suggests a derivation from Central African sequences (mainly form Buduma population). More interestingly, it should be noted that the reversion at np 16223, which characterises the branch to which the Libyan Tuareg and the Central African haplotypes belong, has been recently found in a similar Berber haplotype (HVS-I haplotype: 16129-16189-16249-16311-16374C) (Olivieri et al. 2006). On the other hand, a haplotype harbouring the 16182 mutation has been found in Egypt (HVS-I haplotype 16129-16182-16189-16223-16249-16265C-16311). It suggests that the Libyan Tuareg lineages, as well as the Central African ones present in the same branch, are likely from Northern rather than Eastern Africa.

# Chapter 5. Conclusions

The present work provides an important view of a region of Africa that is still almost unknown: the Central Sahara. The aim of the project as a whole, was to reconstruct from the maternal side the origins of a Pastoral nomad population in the Libyan Sahara, the Tuareg. In fact, the availability of both modern and ancient samples from the Fezzan (Libyan Sahara), collected in collaboration with the Italian Archaeological Mission in Libya, represented an important means of relating the mtDNA pool of extant Libyan Tuareg, with that of Pastoral people inhabiting the Central Sahara in prehistoric times, and with the Garamantes, the hypothetical ancestors of Libyan Tuareg. Nevertheless, molecular analysis carried out on the bones collected from the archaeological sites of the Acacus region, showed a very low state of preservation of the DNA, this probably due to the high temperatures that characterised burials over the centuries. In fact, no aDNA was recovered in a total of 18 ancient individuals, dating back to the Garamantian time (Fewet site, 4 individuals) up to the Pastoral time (Takarkori site, 14 individuals). Hypothetically, preservation of DNA was compromised by the high temperatures of the sites. Concerning the Takarkori archaeological site, the old age of the burials (C14 dating range: from 4,200 up to 7,300 years ago), and the particular burial context (shallow pits few centimetres from the surface), were determining factors leading to the complete degradation of biomolecules in the bones. Nevertheless, some aspects would appear to tentatively increase the size of the Garamantian sample, in particular the good state of preservation of the Garamantian burials, and their typology, which hypothetically could have protected DNA from the high thermal shifts of the Saharan environment. By assuming such a scenario, results of the Thermal Age calculation confirm this hypothesis.

Failure of the genetic analyses in the ancient individuals, necessarily limited the present work to the study of the extant Tuareg sample. Nevertheless comparison

with other genetic data collected so far in the modern African populations, and moreover the multidisciplinary integration with archaeological and ethnological data helped to hypothetically reconstruct the origins of Libyan Tuareg, and their relationship with the ancient human migratory dynamics that occurred in Northern Africa during the Holocene. Nonetheless, as first consequence of the lack of aDNA data, we cannot shed light on the relationship between Garamantes and the extant Libyan Tuareg.

The genetic analysis at the mtDNA level of 129 Tuareg individuals from two neighbour villages in the Fezzan (Al Awaynat and Tahala), clearly shows the very low level of genetic diversity in the Libyan Tuareg, that is hypothetically due to endogamy practices. It should be noted that Tuareg since the pre-colonial time, were characterized by the society's matriarchal structure. Furthermore, at least as long as organization in tribes existed, it was normal praxis that women after marriage followed her husband, but after his death, she moved back to her mother's territory, together with all her sons. We think that this topic contributed to increase the matrilocality of Tuareg, leading to a further reduction of genetic diversity at the regional level.

Analysis of the mtDNA lineages made it possible to characterise Tuareg as a mixed group, in which two main components were identified. The first component is characterised by 'Western Eurasian' lineages and it represents almost two-thirds (65.12 %) of the whole sample. The Western-Eurasian component appears to be very homogeneous: only five different haplotypes have been observed when considering all the mtDNA regions investigated (both the D-Loop and some coding regions). More in detail, three haplotypes belong to haplogroup H1, and constitute the main fraction of the West-Eurasian component (94 %). The other two haplotypes belong to haplogroup V. Within the H1 haplotypes, the strong predominance of the lineage characterised by the HVS-I CRS motif is observed (72/79 haplotypes, 91.14 %). The resulting high homogeneity that characterizes the West Eurasian component, hypothetically originate due to a founder effect. It should be noted, that the HVS-I H-CRS motif distribution in Northern Africa, shows close genetic relationship between North African Berbers, and the Libyan

Tuareg. A scenario can be hypothesized in which the continuously changing Saharan environment, in particular in the second half of the Holocene with the beginning of the arid phase, was responsible for human migratory dynamics that led different groups to mix and also to separate from each other. In this sense, the hypothesis of a Berber origin of the Libyan Tuareg as a consequence of a founder effect seems likely. This hypothesis is supported by linguistic data, which characterise the Tuareg language as a proto-Berber language.

The presence of West-Eurasian lineages in Northern Africa, is hypothetically related to the spread of hunter-gatherers human groups, who left the Franco-Cantabrian refuge at the end of the Pleistocene, when climatic conditions began to improve (about 15,000 years ago). From the Iberian Peninsula, they gradually repopulated much of Central and Northern Europe (Achilli et al. 2004, Torroni et al. 1998), and likely headed southward as well. After crossing the Strait of Gibraltar, they likely contributed their lineages (i.e. H1, H3, V, and U5b1b) to modern North Africans (Achilli et al. 2005).

The second component observed in the Libyan Tuareg mtDNA pool, is characterised by typical South-Saharan lineages. Despite being less represented in the sample (34.88 %), it appears to be more heterogeneous than the West-Eurasian component. L0a1 (6.2 %) and L2a (8.53 %) are the most frequent lineages within the South-Saharan component. Genealogical analysis accomplished through Network analysis, showed that these lineages are related to the spread of Cushitic Pastoral groups from Sudan to the Chad basin. This westward migratory shift hypothetically took place in the first half of the Holocene, when the wet climatic conditions transformed the Chad basin into a flourishing area where to settle. Both linguistic and archaeological data seem to support the westward movement of people from East (Blench 1999, Keding 1993). Then, it is likely that the same good environmental conditions in Central Africa, could have catalysed some kind of human expansion that led people to shift northward, so explaining the presence of Central African lineages in the Central Sahara and more generally North to the Chad basin (Cerny et al. 2007). Archaeological data seem to support this hypothesis (Haour 2003). Nevertheless, we think that more data need to be collected in order to

shed light on the human migratory dynamics that took place in Central and Northern Africa during the Holocene: sampling in still unknown areas, and high resolution phylogenetic methods (i.e. complete mtDNA sequencing), would help to further ascertain the human migratory shifts involved.
As far as the other South-Saharan lineages are concerned, they are all represented at very low frequencies, ranging from 1.55 % to 3.1 %. Despite genealogical analysis linking them with a significant demographic event (e.g. human expansions in both Central and Western Africa), the hypothesis of a sporadic introduction into the Libyan Tuareg population cannot be ruled out. Slavery practices for example, could have introduced 'South-Saharan' lineages in the Tuareg mtDNA genetic pool, and this hypothesis is confirmed by ethnological studies, which describes the old slaves as characterised by typically South-Saharan morphological traits.

As previously mentioned, the absence of aDNA data makes it difficult to investigate the relationship between modern Tuareg and the ancient human groups of Pastoralists inhabiting the Central Sahara. On the basis of the archaeological data so far collected, it is hypothesized that Pastoral groups came from the Middle East in the first half of the Holocene. Their arrival in the Central Sahara would be associated with the appearance of ovicaprines remains, a Middle Eastern species, in the archaeological records of the Acacus sites, at least 7,200 years ago. Furthermore, both the rock art and the morphological data collected from the physical anthropologists (Ricci et al. 2008, in press), depict the Pastoral society as a multiethnic society, indicating some kind of intermixture with people from different areas, hypothetically the hunter-gatherer groups who occupied Central Sahara at the beginning of the Holocene. Concerning our genetic data, the alternative hypothesis about the provenance of the L0a1 lineages from Nile Valley has been done. Such a scenario, would open the possibility of a kind of genetic relationship between the extant Tuareg and the old Pastoralists from East, under the assumption that the Nile Valley acted as the original place for the diffusion of Pastoralism in Northern Africa. Apart from this hypothesis, that still needs to find further proof, no traces of a hypothetical 'Middle Eastern' component has been found in the extant Libyan

Tuareg mtDNA pool. Nevertheless, without aDNA data that could characterize the ancient Pastoral groups in the Central Sahara, no reliable hypotheses can be made concerning the dynamics of the diffusion of Pastoralism in Northern Africa, nor can the relationship between Libyan Tuareg and ancient Pastoral people can be investigated.
It is worth noting that no significant relationships have been observed between the Libyan Tuareg and the Fulbe, which represents one of the most important Pastoral groups today, and which in the second half of the Holocene hypothetically crossed the Sahara.

Finally, it should be noted that our data are limited to a small Tuareg group in the Central Sahara. According to recent estimates, the Tuareg number almost 1,300,000, and are mainly distributed in North-West Africa. It should be remembered that scarce genetic data have been collected so far about this population, so no extensive comparison between samples from different areas can be made in order to investigate the origins of the Tuareg in the Northern Africa. A close genetic relationship between Tuareg and Eastern people from the Sudan area, was already observed by Cavalli Sforza (1994), on the basis of nuclear DNA genetic data. Such a result is concordant with the presence in the Libyan Tuareg mtDNA pool of lineages from East Africa, who hypothetically reached the Central Sahara passing through the Chad basin. Nevertheless, it is worth noting that nuclear genetic data by Cavalli-Sforza, showed a closer genetic relationship between Tuareg and Beja, than between Tuareg and Berbers. This result contrasts with the high affinity of the Libyan Tuareg with Berber groups, but it can be hypothesized that demographic factors (isolation, founder effects, drifts, and gene flow), could have shaped differentially the mtDNA genetic pool of Tuareg groups from distant areas.
Concerning mtDNA data, a few sequences from some Tuareg settled in Western Africa were previously described by Watson et al. (1996). Despite some lineages being shared with the Libyan Tuareg ones (see Networks), this sample appears to be closely related with the surrounding populations, showing a typical West Africa mtDNA pool.

Actually, it is necessary to say that data available so far do not allow us to make further hypotheses about the Tuareg origins in Northern Africa: more studies need to be conducted in order to compare data of different samples from a wide geographic area, so as to eventually find a shared root that would suggest a common origin of all the Tuareg.

# Literature cited

Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, et al. 2004. 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. 75:910-8

Achilli A, Rengo C, Battaglia V, Pala M, Olivieri A, et al. 2005. Saami and Berbers - An unexpected Mitochondrial DNA link. Am J Hum Genet. 76:883-6

Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, et al. 1981. Sequence and organization of the human mitochondrial genome. Nature. 290:457-65

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

Awadalla P, Eyre-Walker A, Maynard-Smith J. 1999. Linkage disequilibrium and recombination in hominid mitochondrial DNA. Science. 286:2524-5

Awadalla P, Eyre-Walker A, Maynard-Smith J. 2000. Questioning evidence for recombination in human mitochondrial DNA. Science. 288:1931a

Azpiroz R, Butow RA. 1993. Patterns of mitochondrial sorting in yeast zygotes. Mol Biol Cell. 4:21-36

Ba A, Dieterlen G. 1966. Les fresques d'époque bovidienne du Tassili-n-Ajjer et les traditions des Peuls. J Soc Afr. 36:141-57

Bandelt HJ, Forster P, Sykes BC, Richards MB. 1995. Mitochondrial portraits of human populations using median networks. Genetics. 141:743-53

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

Bandelt HJ, Alves-Silva J, Guimaraes PE, Santos MS, Brehm A, et al. 2001. Phylogeography of the human mitochondrial haplogroup L3e: a snapshot of African prehistory and Atlantic slave trade. Ann Hum Genet. 65:549-63

Barich BE. 1987. Archaelogy and Environment in the Libyan Sahara. The excavations in the Tadrart Acacus. B.A.R. Internat. Ser. 368:1978-83

Bendall KE, Macaulay V, Baker JR, Sykes B. 1996. Heteroplasmic point mutations in the human mtDNA control region. Am J Hum Genet. 59:1276-87

Bernus E. 1969. Particularismes éthniques en milieu urbain: l'exemple de Niamey. Paris: Institut d'ethnologie, Musée de l'homme.

Bernus E. 1981. Touaregs nigériens. Unité culturelle et diversité régionale d'un peuple pasteur. Paris: L'Harmattan.

Bernus E. 1986. Mobilité et flessibilité pastorales fâce à la sécheresse. In *Bulletin de liaison* pp. 137-43. Paris: Office de la Recherche Scientifique et Technique Outre-mer (ORSTOM)

Binford LR. 1983. In *Pursuit of the Past*. London: Thames and Hudson.

Blench R. 1999. The westward wandering of Cushitic pastoralists. Exploration in the prehistory of Central Africa. In *L' Homme et l' animal dans le basin du lac Chad* eds. C Baroin, J Boutrais, pp. 39-80. Paris: Editions IRD

Boldogh IR, Nowakowski DW, Yang HC, Chung H, Karmon S, et al. 2003. A protein complex containing Mdm10p, Mdm12p, and Mmm1p links mitochondrial membranes and DNA to the cytoskeleton based segregation machinery. Mol Biol Cell. 14

Bosch E, Calafell F, Comas D, Oefner PJ, Underhill PA, Bertranpetit J. 2001. High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between northwestern Africa and the Iberian Peninsula. Am J Hum Genet. 68:1019-29

Brakez Z, Bosch E, Izaabel H, Akhayat O, Comas D, et al. 2001. Human mitochondrial DNA sequence variation in the Moroccan population of the Souss area. Ann Hum Biol. 28:295-307

Brown WM. 1980. Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. Proc Natl Acad Sci U S A. 77:3605-9

Brown DT, Samuels DC, Michael EM, Chinnery PF. 2001. Random genetic drift determines the level of mutant mtDNA in human primary oocytes. Am J Hum Genet. 68:533-6

Brunk K, Gronenborn D. 2004. Floods, droughts, and migrations - The effects of Late Holocene lake level oscillations and climate fluctuations on the settlement and political history in the Chad basin. In *Living with the lake: perspectives on culture, economy and history of Lake Chad.*, ed. M Krings, E Platte, pp. 101-32. Köln: Rüdiger Köppe

Budowle B, Moretti T, Smith J, Dizinno J. 2000. DNA typing protocols: molecular biology and forensic analysis. Natick, MA: Eaton Publishing.

Cann RL, Stoneking M, Wilson AC. 1987. Mitochondrial DNA and human evolution. Nature. 325:31-6

Capot-Rey R. 1953. Le Sahara Francais: Presses Universitaires de France.

Castelletti L, Cottini M, Rottoli M. 1998. Early Holocene plant remains from Uan Afuda cave (Libyan Sahara). In *Before Food Production in North Africa. Questions and Tools Dealing with Resource Exploitation and Population Dynamics at 12,000-7000 bp*, ed. S Di Lernia, G Manzi, Workshop of the XIII Congress, pp. 91-102. Forlì: ABACO

Castelli R, Cremaschi M, Gatto M, Liverani M, Mori L. 2005. A preliminary report of excavations in Fewet, Libyan Sahara. Journal of African Archaeology. 3:69-102

Cavalli-Sforza L, Menozzi P, Piazza A. 1994. The history and geography of human genes: Princeton University Press.

Cerny V, Hajek M, Cmejla R, Bruzek J, Brdicka R. 2004. mtDNA sequences of Chadic-speaking populations from northern Cameroon suggest their affinities with eastern Africa. Ann Hum Biol. 31:554-69

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

Cerny V, Salas A, Hajek M, Zaloudkova M, Brdlcka R. 2007. 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. 71:433-52

Chen XJ, Butow RA. 2005. The organization and inheritance of the mitochondrial genome. Genetics. 6:815-25

Chen XJ, Wang X, Kaufman BA, Butow RA. 2005. Aconitase couples metabolic regulation to mitochondrial DNA maintenance. Science. 307:714-7

Chen YS, Torroni A, Excoffier L, Santachiara-Benerecetti AS, Wallace DC. 1995. Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. Am J Hum Genet. 57:133-49

Chinnery PF, Zwijnenburg M, Walker N, Howell RW, Taylor R, et al. 1999. Nonrandom tissue distribution of mutant mtDNA. Am J Med Genet. 85:498-501

Close AE. 1992. Holocene occupation of the Eastern Sahara. In *New Light on the Northeast African Past*, ed. F Klees, R Kuper, pp. 155-83. Köln, Heinrich Barth Institute.

Coia V, Destro-Bisol G, Verginelli F, Battaggia C, Boschi I, et al. 2005. 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. 128:678-81

Connah G. 1981. Three Thousand Years in Africa, Man and his environment in Lake Chad Regions of Nigeria. Cambridge: Cambridge University Press.

Cooper A, Wayne R.. 1998. New uses for old DNA. Curr Opin Biotechnol. 9:49-53

Cooper A, Poinar HN. 2000. Ancient DNA: do it right or not at all. Science. 289:1139

Corridi C. 1998. Faunal Remains from Holocene Archaeological sites of the Tadrart Acacus and Surroundings (Libyan Sahara). In *Wadi Teshuinat. Palaeoenvironment and Prehistory in South Western Fezzan (Libyan Sahara)*, ed. M Cremaschi, S Di Lernia, pp. 89-94. Milano

Corte-Real HB, Macaulay VA, Richards MB, Hariti G, Issad MS, et al. 1996. Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. Ann Hum Genet. 60:331-50

Cremaschi M. 1996. The rock varnish in the Messak Settafet (Fezzan, Libyan Sahara), age, archaeological context and paleoenvironmental implication. Geoarchaeology. 11:393-421

Cremaschi M. 1998. Late Quaternary geological evidence for encivronmental changes in South-Western Fezzan (Central Sahara, Libya). In *Wadi Teshuinat. Palaeoenvironment and Prehistory in south-western Fezzan (Libyan Sahara)*, ed. M Cremaschi, S Di Lernia, pp. 13-48. Milano

Cremaschi M, Di Lernia S. 1996. Current research on the prehistory of the Tadrart Acacus (Libyan Sahara). Survey and excavations 1991-1995. Nyame Akuma. 45:50-9

Cremaschi M. 2004. Diecimila anni di cambiamenti climatici e ambientali del Fezzan libico. *Darwin* 2:66-71

Cremaschi M, Di Lernia S. 1999. Holocene climatic changes and cultural dynamics in the Lybian Sahara. African Archaeological Review. 16:211-38

Crubezy E, Ludes B, Poveda JD, Clayton J, Crouau-Roy B, Montagnon D. 1998. Identification of Mycobacterium DNA in an Egyptian Pott's disease of 5,400 years old. C R Acad Sci III. 321:941-51

Cruciani F, Santolamazza P, Shen P, Macaulay V, Moral P, et al. 2002. A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. Am J Hum Genet. 70:1197-214

D'Aurelio M, Gajewski CD, Lin MT, Mauck WM, Shao LZ, et al. 2004. Heterologous mitochondrial DNA recombination in human cells. Hum. Mol. Genet. 13:3171-9

David N, Mac Eachern S. 1988. The Mandara Archaeological Project: preliminary results of the 1984 season. In *Le millieu et le hommes, recherches comparatives et historiques dans le bassin du Lav Tchad*, ed. D Barreteau, H Tourneaux, pp. 51-80. Paris: Editions ORSTOM

Di Lernia S. 1997. *Condizioni culturali e forme di adattamento prima della produzione del cibo: i massicci centrali sahariani nell' antico Olocene (10000-7500 bp)*. Rome: Università degli Studi di Roma 'La Sapienza'.

Di Lernia S. 1998. Cultural control over wild animals during the early Holocene: The case of Barbary sheep in central Sahara. In *Before Food Production in North Africa. Questions and Tools Dealing with Resource Exploitation and Population Dynamics at 12,000-7000 bp*, ed. S Di Lernia, G Manzi,Workshop of the XIII Congress, pp. 113-26. Forlì: ABACO

Di Lernia S, ed. 1999. *The Uan Afuda Cave Hunter-Gatherer Societies of Central Sahara*, Arid Zone Archaeology (AZA). Firenze: All'Insegna del Giglio

Di Lernia S, Cremaschi M. 1996. Taming barbary sheep: Wild animal management by Early Holocene hunter-gatherers at Uan Afuda (Libyan Sahara). Nyame Akuma. 46:43-54

Di Lernia S, Liverani M. 2004. Cacciatori, pastori e caravanserragli. Viaggio nelle antiche civiltà del Sahara. *Darwin* 2:72-9

Di Mauro S, Schon EA. 2001. Mitochondrial DNA mutations in human disease. Am J Med Genet. 106:18-26

Dolukhanov P. 2000. 'Prehistory revolutions' and languages in Europe. In *The roots of peoples and languages of northern Eurasia II and III*, ed. A Kunnap, pp. 71-8. Tartu: University of Tartu

Dupuy C. 1999. Les apports de l'achéologie et de l'ethnologie a la connaisance de l'histoire ancienne des Peuls. In *Figures Peules*, ed. R Botte, J Boutrais, J Schmitz. Paris: Karthala

Elson JL, Andrews RM, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. 2001. Analysis of European mtDNAs for recombination. Am J Hum Genet. 68:145-53

Enriquez JA, Cabezas-Herrera J, Bayona-Bafaluy MP, Attardi G. 2000. Very rare complementation between mitochondria carrying different mitochondrial DNA mutations points to intrinsic genetic autonomy of the organelles in cultured human cells. J Biol Chem. 275:11207-15

Excoffier L. 2002. Human demographic history: refining the recent African origin model. Curr Opin Genet Dev. 12:675-82

Excoffier L, Pellegrini P, Sanchez-Mazas A, Simon C, Langaney A. 1987. Genetics and history of Sub-Saharan Africa. Y Phys Anthrop 30:151-94

Eyre-Walker A, Smith NH, Smith JM. 1999. How clonal are human mitochondria? Proc R Soc London B Biol Sci. 266:477-83

Fadhlaoui-Zid K, Plaza S, Calafell F, Ben Amor M, Comas D, Bennamar El gaaied A. 2004. Mitochondrial DNA heterogeneity in Tunisian Berbers. Ann Hum Genet. 68:222-33

Filosto M, Mancuso M, Vives-Bauza C, Vila M, Shanske S, et al. 2003. Lack of paternal inheritance of muscle mitochondrial DNA in sporadic mitochondrial myopathies. Ann Neurol. 54:524-6

Forster P, Matsumura S. 2005. Evolution. Did early humans go north or south? Science. 308:965-6

Gabriel B. 1987. Palaecological evidence from neolithic fireplaces in the Sahara. African Archaeological Review. 5:93-103

Gaspari M, Larsson NG, Gustafsson CM. 2004. The transcription machinery in mammalian mitochondria. Acta. 1659:148-52

Gaudio A. 1993. Uomini Blu : il dramma dei Tuareg tra storia e futuro. Firenze: ECP.

Giazzi F. 1996. Etude initiale. La réserve naturelle nationale de l'Aïr et du Ténéré (Niger), MH/E, WWF et UICN, Gland, Suisse

Gilbert MT, Hansen AJ, Willerslev E, Rudbeck L, Barnes I, et al. 2003a. Characterization of genetic miscoding lesions caused by post-mortem damage. Am J Hum Genet. 72:48-61

Gilbert MT, Willerslev E, Hansen AJ, Barnes I, Rudbeck L, et al. 2003b. Distribution patterns of postmortem damage in human mitochondrial DNA. Am J Hum Genet. 72:32-47

Gilbert MT, Barnes I, Collins MJ, Smith C, Eklund J, et al. 2005. Long-term survival of ancient DNA in Egypt: response to Zink and Nerlich (2003). Am J Phys Anthropol. 128:110-4; discussion 5-8

Gilbert MT, Binladen J, Miller W, Wiuf C, Willerslev E, et al. 2006. Recharacterization of ancient DNA miscoding lesions: insights in the era of sequencing-by-synthesis. Nucleic Acids Res. 35:1-10

Gonder M, Mortensen H, Reed F, de Sousa A, Tishkoff S. 2007. Whole-mtDNA genome sequence analysis of ancient African lineages. Mol Biol Evol. 24:757-68

Gonzalez AM, Cabrera VM, Larruga JM, Tounkara A, Noumsi G, et al. 2006. Mitochondrial DNA variation in Mauritania and Mali and their genetic relationship to other Western Africa populations. Ann Hum Genet. 70:631-57

Gonzalez AM, Larruga JM, Abu-Amero KK, Shi Y, Pestano J, Cabrera VM. 2007. Mitochondrial lineage M1 traces an early human backflow to Africa. BMC Genomics. 8:223

Graver A, Molto J, Parr R, Walters S, Praymak R, Maki J. 2001. Mitochondrial DNA research in the Dakhleh oasis, Egypt: a preliminary report. Ancient Biomol. 3:239-53

Greenberg J. 1972. Linguistinc evidence concerning Bantu origins. J Afr Hist 13:189-216

Gyllensten U, Wharton D, Josefsson A, Wilson AC. 1991. Paternal inheritance of mitochondrial DNA in mice. Nature. 352:255-7

Haak W, Forster P, Bramanti B, Matsumura S, Brandt G, et al. 2005. Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. Science. 310:1016-8

Hagelberg E, Goldman N, Lio P, Wheloan S, Schiefenhövel W, et al. 1999. Evidence for mitochondrial DNA recombination in a human population of island Melanesia. Proc R Soc London B Biol Sci. 266:485-92

Hama B. 1967. Recherches sur l'Histoire des Touaregs Sahariens et Soudanais. Paris: Présence Africane.

Hanotte O, Bradley D, Ochieng JW, Verjee Y, Hill EW, Rege JEO. 2002. African pastoralism: genetic imprints of origins and migrations. Science.336-9

Haour A. 2003. One hundred years of Archaeology in Niger. Journal of World Prehistory. 17:181-234

Hasegawa M, Di Rienzo A, Kocher TD, Wilson AC. 1993. Toward a more accurate time scale for the human mitochondrial DNA tree. J Mol Evol. 37:347-54

Hassan F. 1996. Abrupt Holocene climatic events in Africa, pp. 83-9. Harare

Hassan F. 2002. Droughts, food and culture. New York: Kluwer Academic/Plenum Publishers.

Hauswirth WW, Laipis PJ. 1982. Mitochondrial DNA polymorphism in a maternal lineage of Holstein cows. Proc Natl Acad Sci U S A. 79:4686-90

Hauswirth WW, Laipis PJ. 1985. Transmission genetics of mammalian mitochondria: a molecular model and experimental evidence. In *Achievements and Perspectives of Mitochondrial Research*, ed. E Quagliarello, pp. 49-59. Amsterdam: Elsevier

Heyer E, Zietkiewicz E, Rochowski A, Yotova V, Puymirat J, Labuda D. 2001. Phylogenetic and familial estimates of mitochondrial substitution rates: study of control region mutations in deep-rooting pedigrees. Am J Hum Genet. 69:1113-26

Hofreiter M, Jaenicke V, Serre D, Haeseler Av A, Pääbo S. 2001a. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. Nucleic Acids Res. 29:4793-9

Hofreiter M, Serre D, Poinar HN, Kuch M, Pääbo S. 2001b. Ancient DNA. Nat Rev Genet. 2:353-9

Holden C. 2001. Bantu language trees reflect the spread of farming across sub-Saharan Africa: a maximum parsimony analysis. Proc R Soc London B. 269:793-9

Hoss M, Jaruga P, Zastawny TH, Dizdaroglu M, Pääbo S. 1996. DNA damage and DNA sequence retrieval from ancient tissues. Nucleic Acids Res. 24:1304-7

Housley R, Gamble C, Street M, Pettitt P. 1997. Radiocarbon evidence for the late glacial human recolonisation of northern Europe. Proc Prehist Soc. 63:25-54

Howell N, Elson JL, Turnbull DM, Herrnstadt C. 2004. African Haplogroup L mtDNA sequences show violations of clock-like evolution. Mol Biol Evol. 21:1843-54

Hudson RR. 1990. Gene genealogies and the coalescent process. In *Oxford surveys in Evolutionary Biology*, ed. D Futuyma, JD Antonovics, pp. 1-44. New York: Oxford University Press

Huffman T. 1982. Archaeology and the ethnohistory of the African Iron Age. Annu Rev Anthropol. 11:133-50

Ingman M, Kaessmann H, Pääbo S, Gyllensten U. 2000. Mitochondrial genome variation and the origin of modern humans. Nature. 408:708-13

Jazin E, Soodyall H, Jalonen P, Lindholm E, Stoneking M, Gyllensten U. 1998. Mitochondrial mutation rate revisited: hot spots and polymorphism. Nat Genet. 18:109-10

Jenuth JP, Peterson AC, Fu K, Shoubridge EA. 1996. Random genetic drift in the female germline explains the rapid segregation of mammalian mitochondrial DNA. Nat Genet. 14:146-51

Jobling MA, Hurles ME, Tyler-Smith C. 2004. Human evolutionary genetics. Origins, people and disease. New York: Garland Science.

Johnson MJ, Wallace DC, Ferris SD, Rattazzi MC, Cavalli-Sforza L. 1983. Radiation of human mitochondria DNA types analyzed by restriction endonuclease cleavage patterns. J. Mol. Evol. 19:255-71

Johnstone H. 1913. A survey of the ethnography of Africa anf the former racial and tribal migrations in that continent. J R Anthropol Inst. 43:345-421

Kaneda H, Hayashi JI, Takahama S, Taya C, Lindahl KF, Yonekawa H. 1995. Elimination of paternal mitochondrial DNA in intraspecific crosses during early mouse embryogenesis. Proc Natl Acad Sci U S A. 92:4542-6

Keding B. 1993. Leiterband sites in the Wadi Howar, North Sudan. In *Environmental change and human culture in the Nile basin and Northern Africa until the Second Millennium BC*, ed. L Krzyzaniak, L Kobusiewicz, J Alexander, pp. 371-80. Poznan: Poznan Archaeological Museum

Kingman JFC. 1982. On the genealogy of large populations. J Appl Prob. 19A: 27-43

Kivisild T, Rootsi S, Metspalu M, Mastana S, Kaldma K, et al. 2003. The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. Am J Hum Genet. 72:313-32

Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, et al. 2004. Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. Am J Hum Genet. 75:752-70

Kivisild T, Metspalu E, Bandelt H, Richards M, Villems R. 2006. The world mtDNA phylogeny. In *Human mitochondrial DNA and the evolution of Homo sapiens*, ed. H Bandelt, V Macaulay, M Richards, pp. 149-79. Berlin: Springer-Verlag

Kondo R, Satta Y, Matsuura ET, Ishiwa H, Takahata N, Chigusa SI. 1990. Incomplete maternal transmission of mitochondrial DNA in Drosophila. Genetics. 126:657-63

Korhonen JA, Pham XH, Pellegrini M, Falkenberg M. 2004. Reconstitution of a minimal mtDNA replisome in vitro. EMBO J. 23:2423-9

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

Krings M, Salem AE, Bauer K, Geisert H, Malek AK, et al. 1999. mtDNA analysis of Nile River Valley populations: A genetic corridor or a barrier to migration? Am J Hum Genet. 64:1166-76

Kvist L, Martens J, Nazarenko AA, Orell M. 2003. Paternal leakage of mitochondrial DNA in the great tit (Parus major). Mol Biol Evol. 20:243-7

Legros F, Lombes A, Frachon P, Rojo M. 2002. Mitochondrial fusion in human cells is efficient, requires the inner membrane potential, and is mediated by mitofusins. Mol Biol Cell. 13:4343-54

Lhote H. 1955. Les Touaregs du Hoggar. Paris: Payot.

Lindahl T, Nyberg B. 1972. Rate of depurination of native deoxyribonucleic acid. Biochemistry. 11:3610

Lindahl T. 1993. Instability and decay of the primary structure of DNA. Nature. 362:709-15

Loogvali EL, Roostalu U, Malyarchuk BA, Derenko MV, Kivisild T, et al. 2004. Disuniting uniformity: A pied cladistic canvas of mtDNA Haplogroup H in Eurasia. Mol Biol Evol. 21:2012-21

Loreille O, Orlando L, Patou-Mathis M, Philippe M, Taberlet P, Hanni C. 2001. Ancient DNA analysis reveals divergence of the cave bear, Ursus spelaeus, and brown bear, Ursus arctos, lineages. Curr Biol. 11:200-3

Lupacciolu M. 1992. Arte e culture del Sahara preistorico. Roma: Quasar.

Lutz R, Lutz G. 1995. The secret of the desert. The rock art of Messak Settafet and Messak Mellet, Libya. Innsbruck: Golf Verlag.

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

Macaulay VA, Richards MB, Forster P, Bendall KE, Watson E, et al. 1997. mtDNA mutation rates--no need to panic. Am J Hum Genet. 61:983-90

Macaulay V, Richards M, Hickey E, Vega E, Cruciani F, et al. 1999. The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. Am J Hum Genet. 64:232-49

Macaulay V, Richards M. 2002. In *http://www.stats.gla.ac.uk/~vincent/images/skeleton02.jpg*

Macmillan C, Lach B, Shoubridge EA. 1993. Variable distribution of mutant mitochondrial DNAs (tRNA(Leu[3243])) in tissues of symptomatic relatives with MELAS: the role of mitotic segregation. Neurology. 43:1586-90 *07-08-*

Malaspina P, Cruciani F, Santolamazza P, Torroni A, Pangrazio A, et al. 2000. Patterns of male-specific inter-population divergence in Europe, West Asia and North Africa. Ann Hum Genet. 64:395-412

Maley J. 1981. Etudes palynologiques dans le bassin du Tchad et paléoclimatologie de l'Afrique nord-tropicale de 30.000 and a l'epoque actuelle. Paris: Editions ORSTOM.

Malka F, Lombes A, Rojo M. 2006. Organization, dynamics and transimission of mitochondrial DNA: Focus on vertebrate nucleoids. Biochimica et Biophysica Acta. 1753:463-72

Manfredi G, Thyagarajan D, Papadopoulou LC, Pallotti F, Schon EA. 1997. The fate of human sperm-derived mtDNA in somatic cells. Am J Hum Genet. 61:953-60

Marota I, Basile C, Ubaldi M, Rollo F. 2002. DNA decay rate in papyri and human remains from Egyptian archaeological sites. Am J Phys Anthropol. 117:310-8

Martini M, Sibilia E, Zelaschi C, Troja SO, Forzese R, et al. 1998. TL and OSL dating of fossil dune sand in the Uan Afuda and Uan Tabu sites (Tadrart Acacus, Libyan Sahara). In *Wadi Teshuinat. Palaeoenvironment and Prehistory in south-western Fezzan (Libyan Sahara)*, ed. M Cremaschi, S Di Lernia, pp. 67-72. Milano

Melton T, Clifford S, Martinson J, Batzer M, Stoneking M. 1998. Genetic evidence for the proto-Austronesian homeland in Asia: mtDNA and nuclear DNA variation in Taiwanese aboriginal tribes. Am J Hum Genet. 63:1807-23

Mercuri AM, Trevisan, Grandi G, Mariotti, Lippi M, Cremaschi M. 1998. New pollen data from the Uan Muhuggiag rockshelter (Libyan Sahara, VII-VI millennia bp). In *Wadi Teshuinat. Palaeoenvironment and prehistory in south-western Fezzan (Libyan Sahara)*, ed. M Cremaschi, S Di Lernia, pp. 107-24. Milano

Meyer S, Weiss G, von Haeseler A. 1999. Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. Genetics. 152:1103-10

Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, et al. 2003. Natural selection shaped regional mtDNA variation in humans. Proc Natl Acad Sci U S A. 100:171-6

Mori F. 1965. Tadrart Acacus. Arte rupestre e culture del Sahara preistorico. Torino: Einaudi.

Morral N, Bertranpetit J, Estivill X, Nunes V, Casals T, et al. 1994. The origin of the major cystic fibrosis mutation (delta F508) in European populations. Nat Genet. 7:169-75

Murdock G. 1959. Africa: its peoples and their cultural history. New York: McGraw-Hill.

Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.

Nicolaisen J. 1962. Structures politiques et sociales des Touaregs de l'Aïr et de l'Ahaggar. Niamey: IFAN.

Nicolaisen J. 1963. The Pastoral Tuareg: ecology, culture and society. Copenaghen, National museum

Nicolas F. 1950. Contribution à l'étude des Twâreg de l'Aïr. In *Contribution à l'étude de l'Aïr*, pp. 459-80. Paris: Librairie Larose

Nunnari J, Marshall WF, Straight A, Murray A, Sedat JW, Walter P. 1997. Mitochondrial transmission during mating in Saccharomyces cerevisiae is determined by mitochondrial fusion and fission and the intramitochondrial segregation of mitochondrial DNA. Mol Biol Cell. 8:1233-42

Okamoto K, Perlman PS, Butow RA. 1998. The sorting of mitochondrial DNA and mitochondrial proteins in zygotes: preferential transmission of mitochondrial DNA to the medial bud. J Cell Biol. 142:613-23

Olivieri A, Achilli A, Pala M, Battaglia V, Fornarino S, et al. 2006. The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa. Science. 314:1767-70

Ono T, Isobe K, Nakada K, Hayashi JI. 2001. Human cells are protected from mitochondrial dysfunction by complementation of DNA products in fused mitochondria. Nat Genet. 28:272-5

Otte M. 1990. The northwestern European plain around 18,000 BP. In *The world at 18,000 BP*, ed. O Soffer, C Gamble, pp. 54-68. London: Unwin Hyman

Ottoni C, Koon H, Collins MJ, Penkman K, Rickards O, Craig O. 2008. Preservation of ancient DNA in thermally degraded archaeological bone. Submitted.

Ovchinnikov IV, Gotherstrom A, Romanova GP, Kharitonov VM, Liden K, Goodwin W. 2000. Molecular analysis of Neanderthal DNA from the northern Caucasus. Nature. 404:490-3

Pääbo S. 1985. Molecular cloning of Ancient Egyptian mummy DNA. Nature. 314:644-5

Pääbo S. 1989. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. Proc Natl Acad Sci U S A. 86:1939-43

Pääbo S. 1996. Mutational hot spots in the mitochondrial microcosmos. Am J Hum Genet. 59:493-6

Pääbo S, Irwin DM, Wilson AC. 1990. DNA damage promotes jumping between templates during enzymatic amplification. J Biol Chem. 265:4718-21

Pääbo S, Poinar H, Serre D, Jaenicke-Despres V, Hebler J, et al. 2004. Genetic analyses from ancient DNA. Annu Rev Genet. 38:645-79

Pakendorf B, Wiebe V, Tarskaia LA, Spitsyn VA, Soodyall H, et al. 2003. Mitochondrial DNA evidence for admixed origins of central Siberian populations. Am J Phys Anthropol. 120:211-224

Pakendorf B, Stoneking M. 2005. Mitochondrial DNA and human evolution. Annu Rev Genomics Genet. 6:165-83

Passarino G, Semino O, Quintana-Murci L, Excoffier L, Hammer M, Santachiara-Benerecetti AS. 1998. Different genetic components in the Ethiopian population, identified by mtDNA and Y-chromosome polymorphisms. Am J Hum Genet. 62:420-34

Pereira L, Macaulay V, Torroni A, Scozzari R, Prata MJ, Amorim A. 2001. Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. Ann Hum Genet. 65:439-58

Pereira L, Richards M, Goios A, Alonso A, Albarran C, et al. 2005. High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium. Genome Res. 15:19-24

Petit-Maire N. 1993. Recent Quaternary climatic change and man in Sahara. In *Vane e l'Ambiente del Sahara preistorico, dati ed interpretazioni.*, ed. G Calegari, pp. 411-6. Milano

Phillipson D. 1993. African archaeology. Cambridge: Cambridge University Press.

Piganeau G, Eyre-Walker A. 2004. A reanalysis of the indirect evidence for recombination in human mitochondrial DNA. Heredity. 92:282-8

Plaza S, Calafell F, Helal A, Bouzerna N, Lefranc G, et al. 2003. Joining the pillars of Hercules: mtDNA sequences show multidirectional gene flow in the western Mediterranean. Ann Hum Genet. 67:312-28

Poinar HN, Hoss M, Bada JL, Pääbo S. 1996. Amino acid racemization and the preservation of ancient DNA. Science. 272:864-6

Poinar HN, Hofreiter M, Spaulding WG, Martin PS, Stankiewicz BA, et al. 1998. Molecular coproscopy: dung and diet of the extinct ground sloth Nothrotheriops shastensis. Science. 281:402-6

Poinar HN, Stankiewicz AB. 1999. Protein preservation and DNA retrieval from ancient tissues. Proc Natl Acad Sci U S A. 96:8426-31

Poulton J, Machington DR. 2002. Segregation of mitochondrial DNA (mtDNA) in human oocytes and in animal models of mtDNA disease: clinical implications. Reproduction. 123:751-5

Pruvost M, Schwarz R, Correia VB, Champlot S, Braguier S, et al. 2007. Freshly excavated fossil bones are best for amplification of ancient DNA. Proc Natl Acad Sci U S A. 104:739-44

Quintana-Murci L, Quach H, Harmant C, Luca F, Massonnet B, et al. 2008. 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. 105:1596-601

Rando JC, Pinto F, Gonzalez AM, Hernandez M, Larruga JM, et al. 1998. Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern, and sub-Saharan populations. Ann Hum Genet. 62:531-50

Reed AJ, Stoneking M. 1999. Peopling of Sahul: mtDNA variation in aboriginal Australian and Papua New Guinean populations. Am J Hum Genet. 65:808-28

Rencher AC. 2002. Methods of Multivariate Analysis (2nd edition). New York, NY: Wiley-Interscience.

Reynolds J, Weir BS, Cockerham CC. 1983. Estimation of the Coancestry Coefficient: Basis for a Short-Term Genetic Distance. Genetics. 105:767-79

Ricci F, Fornai C, Tiesler Blos V, Rickards O, di Lernia S, Manzi G. 2008. Evidence of Artificial Cranial Deformation From the Later Prehistory of the Acacus Mts. (South-Western Libya, Central Sahara). International Journal of Osteoarchaeology. In press

Richards M, Corte-Real H, Forster P, Macaulay V, Wilkinson-Herbots H, et al. 1996. Paleolithic and neolithic lineages in the European mitochondrial gene pool. Am J Hum Genet. 59:185-203

Richards M, Macaulay V, Hickey E, Vega E, Sykes B, et al. 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. Am J Hum Genet. 67:1251-76

Rickards O, Martinez-Labarga C, Favaro M, Frezza D, Mallegni F. 2001. DNA analyses of the remains of the Prince Branciforte Barresi family. Int J Legal Med. 114:141-6

Robin ED, Wong R. 1988. Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. J Cell Physiol. 136:507-13

Roostalu U, Kutuev I, Loogvali EL, Metspalu E, Tambets K, et al. 2007. Origin and expansion of haplogroup H, the dominant human mitochondrial DNA lineage in West Eurasia: the Near Eastern and Caucasian perspective. Mol Biol Evol. 24:436-48

Rosa A, Brehm A, Kivisild T, Metspalu E, Villems R. 2004. MtDNA profile of West Africa Guineans: towards a better understanding of the Senegambia region. Ann Hum Genet. 68:340-52

Ruhlen M. 1987. A guide to world's languages. Stanford, CA: Stanford Univesity Press.

Saillard J, Forster P, Lynnerup N, Bandelt HJ, Norby S. 2000. mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. Am J Hum Genet. 67:718-26

Salas A, Richards M, De la Fe T, Lareu MV, Sobrino B, et al. 2002. The making of the African mtDNA landscape. Am J Hum Genet. 71:1082-111

Salas A, Richards M, Lareu MV, Scozzari R, Coppa A, et al. 2004. The African diaspora: mitochondrial DNA and the Atlantic slave trade. Am J Hum Genet. 74:454-65

Schneider S, Roessli D, Excoffier L. 2000. ARLEQUIN ver.2000. A software for population genetic data analysis. University of Geneva, Switzerland: Genetics and Biometry Laboratory Dept. of Anthropology and Ecology

Schwartz M, Vissing J. 2002. Paternal inheritance of mitochondrial DNA. N Engl J Med. 347:576-80

Schwartz M, Vissing J. 2003. New patterns of inheritance in mitochondrial disease. Biochem Biophys Res Commun. 310:247-51

Schwartz M, Vissing J. 2004. No evidence for paternal inheritance of mtDNA in patients with sporadic mtDNA mutations. J Neurol Sci. 218:99-101

Scozzari R, Cruciani F, Santolamazza P, Malaspina P, Torroni A, et al. 1999. Combined use of biallelic and microsatellite Y-chromosome polymorphisms to infer affinities among African populations. Am J Hum Genet. 65:829-46

Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, et al. 2000. The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. Science. 290:1155-9

Shitara H, Hayashi JI, Takahama S, Kaneda H, Yonekawa H. 1998. Maternal inheritance of mouse mtDNA in interspecific hybrids: segregation of the leaked paternal mtDNA followed by the prevention of subsequent paternal leakage. Genetics. 148:851-7

Sidow A, Wilson AC, Pääbo S. 1991. Bacterial DNA in Clarkia fossils. Philos Trans R Soc Lond B Biol Sci. 333:429-32; discussion 32-3

Slatkin M. 1991. Inbreeding coefficients and coalescence times. Genet Res. 58:167-75

Slatkin M. 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics. 139:457-62

Smith CI, Chamberlain AT, Riley MS, Cooper A, Stringer CB, Collins MJ. 2001. Neanderthal DNA. Not just old but old and cold? Nature. 410:771-2

Smith CI, Chamberlain AT, Riley MS, Stringer C, Collins MJ. 2003. The thermal history of human fossils and the likelihood of successful DNA amplification. J Hum Evol. 45:203-17

Soodyall H, Vigilant L, Hill AV, Stoneking M, Jenkins T. 1996. mtDNA control-region sequence variation suggests multiple independent origins of an 'Asian-specific' 9-bp deletion in sub-Saharan Africans. Am J Hum Genet. 58:595-608

Stevanovitch A, Gilles A, Bouzaid E, Kefi R, Paris F, et al. 2004. Mitochondrial DNA sequence diversity in a sedentary population from Egypt. Ann Hum Genet. 68:23-39

Stoneking M. 1993. DNA and recent human evolution. Evol Anthropol. 2:60-73

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

Stoneking M, Sherry ST, Redd AJ, Vigilant L. 1992. New approaches to dating suggest a recent age for the human mtDNA ancestor. Philos Trans R Soc London B Biol Sci. 337:167-75

Stuart JA, Mayard S, Hashiguchi K, Souza-Pinto NC, Bohr VA. 2005. Localization of mitochondrial DNA base excision repair to an inner membrane-associated particulate fraction. Nucleic Acids Res. 33:3722-32

Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics. 105:437-60

Tajima F. 1993. Measurement of DNA polymorphism. In *Mechanism of molecular evolution. Introduction to Molecular Paleopopulation biology.*, ed. N Takahata, AG Clark, pp. 37-59. Tokyo: Japan Scientific Society Press, Sinauer Associates, Inc.

Tambets K, Rootsi S, Kivisild T, Help H, Serk P, et al. 2004. The western and eastern roots of the Saami—the story of genetic 'outliers' told by mitochondrial DNA and Y chromosomes. Am J Hum Genet. 74:661-82

Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol. 10:512-26

Tauxier L. 1937. Moeurs et histoire des Peuls. Paris: Payot.

Taylor RW, Turnbull DM. 2005. Mitochondrial DNA mutations in human disease. Nature Rev Genet. 6:389-402

Taylor RW, McDonnel MT, Blakely EL, Chinnery PF, Taylor GA, et al. 2003. Genotypes from patients indicate no paternal mitochondrial DNA contribution. Ann Neurol. 54:521-4

Thyagarajan B, Padua RA, Campbell C. 1996. Mammalian mitochondria possess homologous DNA recombination activity. J Biol Chem. 271:27536-43

Tornieri G. 2002. *Società e spazio: l'evoluzione della territorialità Touareg*, Trento

Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, et al. 1993. Asian affinities and continental radiation of the four founding Native American mtDNAs. Am J Hum Genet. 53:563-590

Torroni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, et al. 1996. Classification of European mtDNAs from an analysis of three European populations. Genetics. 144:1835-50

Torroni A, Petrozzi M, D'Urbano L, Sellitto D, Zeviani M, et al. 1997. Haplotype and phylogenetic analyses suggest that one European-specific mtDNA background plays a role in the expression of Leber hereditary optic neuropathy by increasing the penetrance of the primary mutations 11778 and 14484. Am J Hum Genet. 60:1107-21

Torroni A, Bandelt HJ, D'Urbano L, Lahermo P, Moral P, et al. 1998. mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. Am J Hum Genet. 62:1137-52

Torroni A, Bandelt HJ, Macaulay V, Richards M, Cruciani F, et al. 2001a. A signal, from human mtDNA, of postglacial recolonization in Europe. Am J Hum Genet. 69:844-52

Torroni A, Rengo C, Guida V, Cruciani F, Sellitto D, et al. 2001b. Do the four clades of the mtDNA haplogroup L2 evolve at different rates? Am J Hum Genet. 69:1348-56

Troy CS, MacHugh DE, Bailey JF, Magee DA, Loftus RT, et al. 2001. Genetic evidence for Near-Eastern origins of European cattle. Nature. 410:1088-91

Tsaousis AD, Martin DP, Ladoukakis ED, Posada D, Zouros E. 2005. Widespread recombination in published animal mtDNA sequences. Mol Biol Evol. 22:925-33

Tully LA, Parsons TJ, Steighner RJ, Holland MM, Marino MA, Prenger VL. 2000. A sensitive denaturing gradient-gel electrophoresis assay reveals a high frequency of heteroplasmy in hypervariable region 1 of the human mtDNA control region. Am J Hum Genet. 67:432-43

Underhill PA, Shen P, Lin AA, Jin L, Passarino G, et al. 2000. Y chromosome sequence variation and the history of human populations. Nat Genet. 26:358-61

Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, et al. 2001. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann Hum Genet. 65:43-62

Van der Kuyl A, Dekker J, Attia M, Iskander N, Perizonius W, Goudsmit J. 1994. DNA from ancient Egyptian monkey bone. Ancient DNA Newslett. 2:19-21

Vansina J. 1995. New linguistic evidence and the 'Bantu expansion' J Afr Hist. 36:173-95

Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC. 1991. African populations and the evolution of human mitochondrial DNA. Science. 253:1503-7

Vogel J. 1994a. Bantu expansion. In *Encyclopedia of precolonial Africa*, ed. J Vogel, pp. 435-8. Walnut Creek: AltaMira Press

Vogel J. 1994b. Eastern and South-central Iron Age. In *Encyclopedia of precolonial Africa*, ed. J Vogel, pp. 439-44. Walnut Creek: AltaMira Press

Wakeley J. 1993. Substitution rate variation among sites in hypervariable region I of human mitochondrial DNA. J Mol Evol. 37:613-23

Wallace DC, Brown MD, Lott MT. 1999. Mitochondrial DNA variation in human evolution and disease. Gene. 228:211-30

Ward RH, Frazier BL, Dew-Jager K, Pääbo S. 1991. Extensive mitochondrial diversity within a single Amerindian tribe. Proc Natl Acad Sci U S A. 88:8720-4

Watson E, Bauer K, Aman R, Weiss G, von Haeseler A, Pääbo S. 1996. mtDNA sequence diversity in Africa. Am J Hum Genet. 59:437-44

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

Willerslev E, Cooper A. 2005. Ancient DNA. Proc Biol Sci. 272:3-16

Zink A, Nerlich AG. 2003. Molecular analyses of the 'Pharaos:' Feasibility of molecular studies in ancient Egyptian material. Am J Phys Anthropol. 121:109-11

Zink A, Nerlich AG. 2005. Long term survival of Ancient DNA in Egypt: reply to Gilbert et al. Am J Phys Anthropol. 128:115-8

Zink A, Haas CJ, Reischl U, Szeimies U, Nerlich AG. 2001. Molecular analysis of skeletal tuberculosis in an ancient Egyptian population. J Med Microbiol. 50:355-66

Zink AR, Sola C, Reischl U, Grabner W, Rastogi N, et al. 2003. Characterization of Mycobacterium tuberculosis complex DNAs from Egyptian mummies by spoligotyping. J Clin Microbiol. 41:359-67

Zink AR, Grabner W, Nerlich AG. 2005. Molecular identification of human tuberculosis in recent and historic bone tissue samples: The role of molecular techniques for the study of historic tuberculosis. Am J Phys Anthropol. 126:32-47

Zischler H, Geisert H, von Haeseler A, Pääbo S. 1995. A nuclear 'fossil' of the mitochondrial D-loop and the origin of modern humans. Nature. 378:489-92

Zouros E, Freeman AR, Ball AO, Pogson GH. 1992. Direct evidence for extensive paternal mitochondrial DNA inheritance in the marine mussel Mytilus. Nature. 359:412-4

# Appendix

**TABLE A1a**

Reaction conditions for the amplification of the HVS-I, HVS-II and coding region of the mtDNA in ancient and extant samples. All the PCR products were successively submitted to sequencing analysis.

| Reagent | Concentration in the reaction |
|---|---|
| PCR Buffer<br>KCl 500 mM<br>Tris-HCl pH8.3, 100 mM<br>$MgCl_2$ 15 mM<br>Gelatine 1 mg/ml | 1:10 React. Vol. |
| dNTPs 100 mM each (Invitrogen, Boeringer) | 1 mM each |
| Primers 100 µM each (Primm, MWG-Biotech) | 100 nM each |
| Taq Pol (Applied Biosystems) | 1U |
| ddH2O[a]<br>Ancient protocol<br>Modern protocol | <br>up to 50 µl<br>up to 25 µl |
| Template<br>Ancient DNA<br>Modern DNA<br>PCR product[b] | <br>1-8 µl<br>1-2 µl<br>1-3 µl |

[a]Autoclaved double-distilled water
[b] PCR product from the first round of PCR in the nested PCR experiments (see Material and methods)

**TABLE A1b**

Reaction conditions for the amplification of coding regions in the mtDNA of extant Tuareg. All the PCR products were successively submitted to RFLP analysis.

| Reagent | Concentration in the reaction |
|---|---|
| PCR Buffer II 10X (Applied Biosystems) | 1X |
| $MgCl_2$ 25mM (Applied Biosystems) | 1.5 mM |
| dNTPs 100 mM each (Boeringer, Invitrogen) | 0.2 mM each |
| Primers 100 µM (Primm, MWG-Biotech) | See table APP1d |
| Taq Pol (Applied Biosystems) | 1U |
| ddH2O | up to 25 µl |
| Template | 1-2 µl |

**TABLE A1c**

Cycles conditions used for the amplification of the HVS-I, HVS-II and some coding regions in the mtDNA of the extant Tuareg. All the amplification products were submitted to sequencing reaction. The 9700/2700 thermocyclers by Applied Biosystems were used for the amplification reactions.

| | Denaturation | Cycles | | Final extension |
|---|---|---|---|---|
| **HVS-I and HVS-II** | | | | |
| 15996-16401<br>29-408 | 94°C-15’ | 94°C-30’’<br>60°C-30’’<br>72°C-30’’ | x25 | 72°C-10’ |
| **Coding Regions** | | | | |
| 1361-1451<br>2978-3361<br>6740-7051 | 95°C-5’ | 95°C-10’’<br>58°C-30’’<br>72°C-30’’ | x35 | 72°C-10’ |
| 3641-4102<br>4180-4861 | 95°C-5’ | 95°C-10’’<br>54°C-30’’<br>72°C-30’’ | x35 | 72°C-10’ |
| 15761-16139 | 94°C-15’ | 94°C-30’’<br>60°C-30’’<br>72°C-30’’ | x25 | 72°C-10’ |

**TABLE A1d**

Concentrations of each primer (Forward and Reverse), and PCR conditions used for the amplification of coding regions in the mtDNA of the extant Tuareg. All the amplification products were submitted to RFLP analysis. Primers' sequences are listed in Table APP2b. For the amplification reactions the 9700/2700 Thermocyclers by Applied Biosystems were used.

| Primer concentration (each) | Restriction enzyme/Nucl. site | Denaturation | Cycles | | Final extension |
|---|---|---|---|---|---|
| 400 nM<br>400 nM<br>320 nM<br>400 nM<br>400 nM<br>400 nM<br>400 nM | *Mbo*I 2349<br>*Mbo*I 8616<br>*Hae*III 13957<br>*Mnl*I 10871<br>*Hae*III 13803<br>*Hae*III 1718<br>*Tsp*I 11251 | 94°C-5' | 94°C-30''<br>60°C-30''<br>72°C-30'' | x25 | 72°C-10' |
| 400 nM<br>200 nM<br>400 nM | *Taq*I 10084<br>*Hinf*I 12308<br>*Mbo*I 3693 | 94°C-5' | 94°C-15''<br>55°C-5''<br>72°C-10'' | x40 | 72°C-5' |
| 280 nM<br>280 nM<br>400 nM | *Dde*I 10394<br>*Alu*I 10397<br>*Alu*I 7055 | 94°C-5' | 94°C-15''<br>62°C-5''<br>72°C-10'' | x40 | 72°C-7' |
| 240 nM<br>320 nM<br>400 nM<br>400 nM | *Acc*I 14465<br>*Mse*I 14766<br>*Dde*I 1715<br>*Nla*III 4577 | 94°C-5' | 94°C-15''<br>51°C-5''<br>72°C-10'' | x30 | 72°C-7' |
| 240 nM | *Hae*III 8994 | 94°C-5' | 94°C-15''<br>65°C-8''<br>72°C-10'' | x40 | 72°C-7' |
| 400 nM<br>280 nM<br>200 nM<br>200 nM<br>320 nM<br>320 nM | *Nla*III 4216<br>*Alu*I 10032<br>*Hpa*I 3592<br>*Hae*III 9052<br>*Alu*I 7025<br>*Alu*I 4157 | 94°C-5' | 94°C-15''<br>46°C-5''<br>72°C-10'' | x30 | 72°C-7' |

**TABLE A2a**

Primers used for the amplification reactions of the HVS-I and HVS-II in the ancient samples. For each region, primers were used in different combination in order to tentatively increase the success of reaction (by decreasing the size of the PCR product), and to obtain reproducibility of results from different primer pairs.

| Fragment | Primer Forward (L) | | Primer Reverse (H) | | PCR product length (bp range) |
|---|---|---|---|---|---|
| HVS-I_1 | **L15996** | CCGAAGCTTCTCCACCATTAGCACCCAAAG | **H16139** | TACTACAGGTGGTCAAGTAT | 124-196 |
| | **L16055** | GAAGCAGATTTGGGTACCAC | **H16142** | ATGTACTACAGGTGGTCAAG | |
| HVS-I_2 | **L16131** | CACCATGAATATTGTACGGT | **H16236** | CTTTGGAGTTGCAGTTGATG | 117-180 |
| | **L16159** | TACTTGACCACCTGTAGTAC | **H16271** | GGTTTTGATGTGGATTGGGT | |
| HVS-I_3 | **L16209** | CCCCATGCTTACAAGCAAGT | **H16331** | TTGACTGTAATGTGCTATGT | 162-210 |
| | | | **H16356** | GTCATCCATGGGGACGAGAA | |
| | | | **H16379** | CAAGGGACCCCTATCTGAGG | |
| HVS-I_4 | **L16287** | CACTAGGATACCAACAAACC | **H16391** | GAGGATGGTGGTCAAGGGAC | 144-163 |
| | | | **H16401** | GGCCTCGAGTGATTTCACGGAGGATGGTG | |
| | | | **H16410** | GCGGGATATTGATTTCACGG | |
| HVS-II_1 | **L29** | CCGAAGCTTGGTCTATCACCCTATTAACCAC | **H285** | GGTTTGGTGGAAATTTTTTGT | 288-307 |
| | **L48** | CTCACGGGAGCTCTCCATGC | | | |
| HVS-II_2 | **L172** | CTGGCCACAGCACTTAAACAC | **H389** | CTGGTTAGGCTGGTGTTAGG | 258-288 |
| | | | **H 408** | GGCCTCGAGCTGTTAAAAGTGCATACCGCCA | |

**TABLE A2b**

Primers used for the amplification reactions of the coding regions of mtDNA submitted to RFLP analysis in the extant Tuareg samples.

| SNP | Primer Forward (L) | | Primer Reverse (H) | | Restriction Enzyme/Site | PCR product length | Restriction fragment size |
|---|---|---|---|---|---|---|---|
| **10398** | **L10359** | GTGTGGCCTATGAGTGACTAC | **H10466** | ATTTATGTAAATGAGGGGCATTTGG | *Dde*I 10394 | 131 | 96+35 |
| **10400** | **L10359** | GTGTGGCCTATGAGTGACTAC | **H10466** | ATTTATGTAAATGAGGGGCATTTGG | *Alu*I 10397 | 131 | 99+32 |
| **10873** | **L10810** | TCCAAAAAACACATAATTTG | **H10900** | GGGGAACAGCTAAATAGGTT | *Mnl*I 10871 | 90 | 59+31 |
| **3594** | **L3556** | CTCACCATCGCTCTTC | **H3630** | AACGGCTAGGCTAGAG | *Hpa*I 3592 | 108 | 56+52 |
| **12308** | **L12216** | CACAAGAACTGCTAACTCATGC | **H12309** | ATTACTTTTATTTGGAGTTGCACCAAGATT | *Hinf*I 12308 | 123 | 93+30 |
| **8618** | **L8571** | AGGCCTACCCGCCGCAGTAC | **H8660** | GTTTGATTAGTCATTGTTGG | *Mbo*I 8616 | 110 | 45+65 |
| **2349** | **L2281** | ACCCTATAGAAGAACTAATG | **H2371** | TGTAGATATTGGGCTGTTAA | *Mbo*I 2349 | 110 | 68+42 |
| **10086** | **L9976** | TTGATGAGGGTCTTACTC | **H10091** | TAGTAGTAAGGCTAGGAGG | *Taq*I 10084 | 136 | 108+28 |
| **14766** | **L14729** | TCAACTACCAGAACACCAATGACC | **H14832** | AGTGAGCCGAAGTTTCATCATG | *Mse*I 14766 | 70+55 | 70+17+38 |
| **7028** | **L6948** | ACCGTAGGTGGCCTGACTGG | **H7041** | GGCAAATACAGCTCCTATTGATAGGAC | *Alu*I 7025 | 120 | 79+41 |
| **4580** | **L4519** | CACTCATCACAGCGCTAAGC | **H4621** | TGGCAGCTTCTGTGGAAC | *Nla*III 4577 | 120 | 62+58 |
| **13803** | **L13761** | ATCCCCCTTCCAAACAACAA | **H13824** | CTAGGGCTGTTAGAAGTCCT | *Hae*III 13803 | 83 | 43+40 |
| **4158** | **L4091** | AGACCCTACTTCTAACCTCC | **H4181** | TAGGGTGAGTGGTAGGAAGT | *Alu*I 4157 | 110 | 67+43 |
| **13958** | **L13921** | TACCCTAGCATCACACACCG | **H14011** | GTTTTAGGTAATAGCTTTTC | *Hae*III 13957 | 110 | 73+37 |
| **7055** | **L6948** | ACCGTAGGTGGCCTGACTGG | **H7115** | GGCAAATACAGCTCCTATTGATAGGAC | *Alu*I 7055 | 187 | 107+80 |

**TABLE A2c**

Primers used for the amplification reactions of the D-Loop and some coding regions in the extant Tuareg samples. The resulting PCR products were submitted to sequencing analysis. SNPs are diagnostic for the following haplogroups: 15904-V, 1438/4769-H2, 3010-H1; 6776-H3 (Achilli et al. 2004); 3992/4024-H4 (Pereira et al. 2005).

| **Region of Interest/SNPs** | **Primer Froward (L)** | | **Primer Reverse (H)** | |
|---|---|---|---|---|
| **HVS-I (D-Loop)** | **L 15996** | CCGAAGCTTCTCCACCATTAGCACCCAAAG | **H16401** | GGCCTCGAGTGATTTCACGGAGGATGGTG |
| **HVS-II (D-Loop)** | **L29** | CCGAAGCTTGGTCTATCACCCTATTAACCAC | **H408** | GGCCTCGAGCTGTTAAAAGTGCATACCGCCA |
| **15904** | **L15761** | GGAGGACAACCAGTAAGCTA | **H16139** | TACTACAGGTGGTCAAGTAT |
| **1438** | **L1361** | GGCTACATTTTCTACCCCAG | **H1451** | TCAGGGCCCTGTTCAACTAA |
| **3010** | **L 2978** | GTCCATATCAACAATAGGGT | **H3361** | CGTTCGGTAAGCATTAGGAA |
| **3992, 4024** | **L3641** | CCGTTTACTCAATCCTCTGA | **H4102** | GAAGTAGGGTCTTGGTGAC |
| **4769** | **L4180** | AACTTCCTACCACTCACC | **H4861** | GGCTAGTTTTTGTCATGTGA |
| **6776** | **L6740** | TGGTCTGAGCTATGATATCA | **H7051** | GATGGCAAATACAGCTCCTA |

**TABLE A3**

Complete list of African samples used in the present study for comparative analyses. SA stands for Southern Africa: these populations are not reported in the geographical map, and whenever they appear in the graphical representations (i.e. MDS, Network) they are represented in grey.

| POPULATION | code | PLACE/COUNTRY | SIZE | REFERENCES | GEOGRAPHIC GROUP |
|---|---|---|---|---|---|
| Al Awaynat Tuareg | 1 | Libya | 111 | Present study | Central Sahara |
| Tahala Tuareg | 2 | Libya | 18 | Present study | Central Sahara |
| Oromo | 3 | Ethiopia | 84 | Our unpublished data | Eastern Africa |
| Amhara | 4 | Ethiopia | 90 | Our unpublished data | Eastern Africa |
| Fayum Egyptians | 5 | Egypt | 69 | Our unpublished data | Eastern Africa |
| Fon | 6 | Benin | 49 | Our unpublished data | Western Africa |
| Dendi | 7 | Benin | 48 | Our unpublished data | Western Africa |
| Berba | 8 | Benin | 46 | Our unpublished data | Western Africa |
| Bariba | 9 | Benin | 50 | Our unpublished data | Western Africa |
| Libyans | 10 | Libya | 13 | Our unpublished data | Northern Africa |
| Souss Berbers | 11 | Morocco | 50 | Brakez et al. 2001 | Northern Africa |
| Nairobi | 12 | Kenya | 100 | Brandstatter et al. 2004 | Eastern Africa |
| Capoverde | 13 | Capoverde | 122 | Brehm et al. 2002 | Western Africa |
| Açores islanders | 14 | Açores archipelago | 179 | Brehm et al. 2003 | Western Africa |
| Madeira islanders | 15 | Madeira archipelago | 155 | Brehm et al. 2003 | Western Africa |
| Hide | 16 | Cameroon | 23 | Cerny et al. 2004 | Central Africa |
| Mafa | 17 | Cameroon | 32 | Cerny et al. 2004 | Central Africa |
| Kotoko | 18 | Cameroon | 56 | Cerny et al. 2004 | Central Africa |
| Masa | 19 | Cameroon | 31 | Cerny et al. 2004 | Central Africa |
| Fulani[a] | 20, 21 | Chad/Cameroon | 185 | Cerny et al. 2006, 2007 | Central Africa |
| Kanuri | 22 | Nigeria | 31 | Cerny et al. 2007 | Central Africa |
| Kanembu | 23 | Chad | 50 | Cerny et al. 2007 | Central Africa |
| Fali | 24 | Cameroon | 40 | Cerny et al. 2007 | Central Africa |

**TABLE A3** continued

| POPULATION | code | PLACE/COUNTRY | SIZE | REFERENCES | GEOGRAPHIC GROUP |
|---|---|---|---|---|---|
| Buduma | 25 | Niger | 30 | Cerny et al. 2007 | Central Africa |
| Arabs Shuwa | 26 | Nigeria | 38 | Cerny et al. 2007 | Central Africa |
| Arabs Chad | 27 | Chad | 27 | Cerny et al. 2007 | Central Africa |
| Khwe | SA | South Africa | 31 | Chen et al. 2000 | Southern Africa |
| Mozabites Berbers | 28 | Algeria | 85 | Corte Real et al. 1996 | Northern Africa |
| Bamileke | 29 | Cameroon | 48 | Coia et al. 2005 | Equatorial Africa |
| Ewondo | 30 | Cameroon | 53 | Coia et al. 2005 | Equatorial Africa |
| Mbenzele | 31 | Central African Republic | 57 | Destro Bisol et al. 2004 | Equatorial Africa |
| Chenini-Douiret Berbers | 32 | Tunisia | 53 | Fadhaloi Zid et al. 2004 | Northern Africa |
| Matmata Berbers | 33 | Tunisia | 49 | Fadhaloi Zid et al. 2004 | Northern Africa |
| Sened Berbers | 34 | Tunisia | 53 | Fadhaloi Zid et al. 2004 | Northern Africa |
| Mandenka | 35 | Senegal | 119 | Graven et al. 1995 | Western Africa |
| Malinke | 36 | Mali | 31 | Gonzales et al. 2006 | Western Africa |
| Bambara | 37 | Mali | 52 | Gonzales et al. 2006 | Western Africa |
| Mende | 38 | Sierra Leone | 55 | Jackson et al. 2004 | Western Africa |
| Loko | 39 | Sierra Leone | 32 | Jackson et al. 2004 | Western Africa |
| Limba | 40 | Sierra Leone | 68 | Jackson et al. 2004 | Western Africa |
| Temne | 41 | Sierra Leone | 120 | Jackson et al. 2004 | Western Africa |
| Yemenite[b] | 42 | Yemen | 115 | Kivisild et al.2004 | Eastern Africa |
| Tigrai | 43 | Ethiopia | 45 | Kivisild et al.2004 | Eastern Africa |
| Oromo_b | 44 | Ethiopia | 33 | Kivisild et al.2004 | Eastern Africa |
| Gurage | 45 | Ethiopia | 21 | Kivisild et al.2004 | Eastern Africa |
| other Ethiopians | 46 | Ethiopia | 27 | Kivisild et al.2004 | Eastern Africa |
| Eritreans | 47 | Eritrea | 8 | Kivisild et al.2004 | Eastern Africa |
| Amhara_b | 48 | Ethiopia | 120 | Kivisild et al.2004 | Eastern Africa |
| Afar | 49 | Ethiopia | 16 | Kivisild et al.2004 | Eastern Africa |

**TABLE A3** continued

| POPULATION | code | PLACE/COUNTRY | SIZE | REFERENCES | GEOGRAPHIC GROUP |
|---|---|---|---|---|---|
| Nubia | 50 | Egypt/Sudan | 80 | Krings et al. 1999 | Eastern Africa |
| Sudanese | 51 | Sudan | 60 | Krings et al. 1999 | Eastern Africa |
| Egyptians | 52 | Egypt | 126 | Krings et al. 1998, Stevanovitch et al. 2003 | Eastern Africa |
| São Tomé-Bioko islanders | 53 | São Tomé/Bioko islands (Guinean Gulf) | 95 | Mateu et al. 1997 | Equatorial Africa |
| Mozambicans Bantu | SA | Mozambico | 68 | Pereira et al. 2001 | Southern Africa |
| Moroccan Berbers | 54 | Morocco | 64 | Pinto et al. 1996, Rando et al. 1998, Plaza et al. 2003 | Northern Africa |
| Canary Island | 55 | Canary Island | 55 | Pinto et al. 1996 | Northern Africa |
| Tunisians | 56 | Tunisia | 47 | Plaza et al. 2003 | Northern Africa |
| Algerians | 57 | Algeria | 47 | Plaza et al. 2003 | Northern Africa |
| Angolans | SA | Angola | 44 | Plaza et al. 2004 | Southern Africa |
| Senegalese | 58 | Senegal | 50 | Rando et al. 1998 | Western Africa |
| Moroccans | 59 | Morocco | 50 | Rando et al. 1998, Plaza et al. 2003 | Northern Africa |
| Wolof | 60 | Senegal | 48 | Rando et al. 1998 | Western Africa |
| Serer | 61 | Senegal | 23 | Rando et al. 1998 | Western Africa |
| Saharawi | 62 | Western Sahara | 80 | Rando et al. 1998, Plaza et al. 2003 | Northern Africa |
| Mauritanians | 63 | Mauritania | 95 | Rando et al. 1998, Gonzales et al. 2006 | Western Africa |
| Guineans | 64 | Guiné Bissau | 372 | Rosa et al. 2004 | Western Africa |
| Southeast Bantu | SA | Mozambico | 307 | Salas et al. 2002 | Southern Africa |
| Angolares | 52 | São Tomé/Príncipe islands (Guinean Gulf) | 30 | Trovoada et al. 2003 | Equatorial Africa |
| Forros | 52 | São Tomé/Príncipe islands (Guinean Gulf) | 38 | Trovoada et al. 2003 | Equatorial Africa |
| Tonga | 52 | São Tomé/Príncipe islands (Guinean Gulf) | 35 | Trovoada et al. 2003 | Equatorial Africa |
| Yoruba | 65 | Nigeria | 21 | Vigilant et al. 1991 | Western Africa |
| Herero | SA | Botswana | 27 | Vigiant et al. 1991 | Southern Africa |
| Biaka | 66 | Central African Republic | 17 | Vigiant et al. 1991 | Equatorial Africa |
| Mbuti | 67 | Democratic Republic of Congo | 20 | Vigiant et al. 1991 | Equatorial Africa |
| !Kung | SA | Botswana, South Africa | 67 | Vigilant et al. 1991, Chen et al. 2000 | Southern Africa |

**TABLE A3** continued

| POPULATION | code | PLACE/COUNTRY | SIZE | REFERENCES | GEOGRAPHIC GROUP |
|---|---|---|---|---|---|
| Songhai | 68 | Nigeria, Niger, Mali | 9 | Watson et al. 1996 | Western Africa |
| Western Tuareg | 69 | Nigeria, Niger, Mali | 24 | Watson et al. 1996 | Western Africa |
| Fulbe[c] | 70 | Nigeria, Niger, Benin, Cameroon, Burkina Faso | 61 | Watson et al. 1996 | Central Africa |
| Hausa | 71 | Nigeria, Niger | 20 | Watson et al. 1996 | Central Africa |
| Kanuri_b | 72 | Nigeria, Niger | 14 | Watson et al. 1996 | Central Africa |
| Kikuyu | 73 | Kenya | 25 | Watson et al. 1996 | Eastern Africa |
| Turkana | 74 | Kenya | 37 | Watson et al. 1996 | Eastern Africa |
| Somalians | 75 | Somalia | 15 | Watson et al. 1996 | Eastern Africa |

[a] Includes two samples: Bongor (code 20), and Tcheboua (code 21).
[b] Despite belonging to the Asian continent, Yemen was included in the Eastern Africa group as it shares many lineages with Ethiopia and other countries of the Horn of Africa.
[c] Fulbe is an other name used for the Fulani.

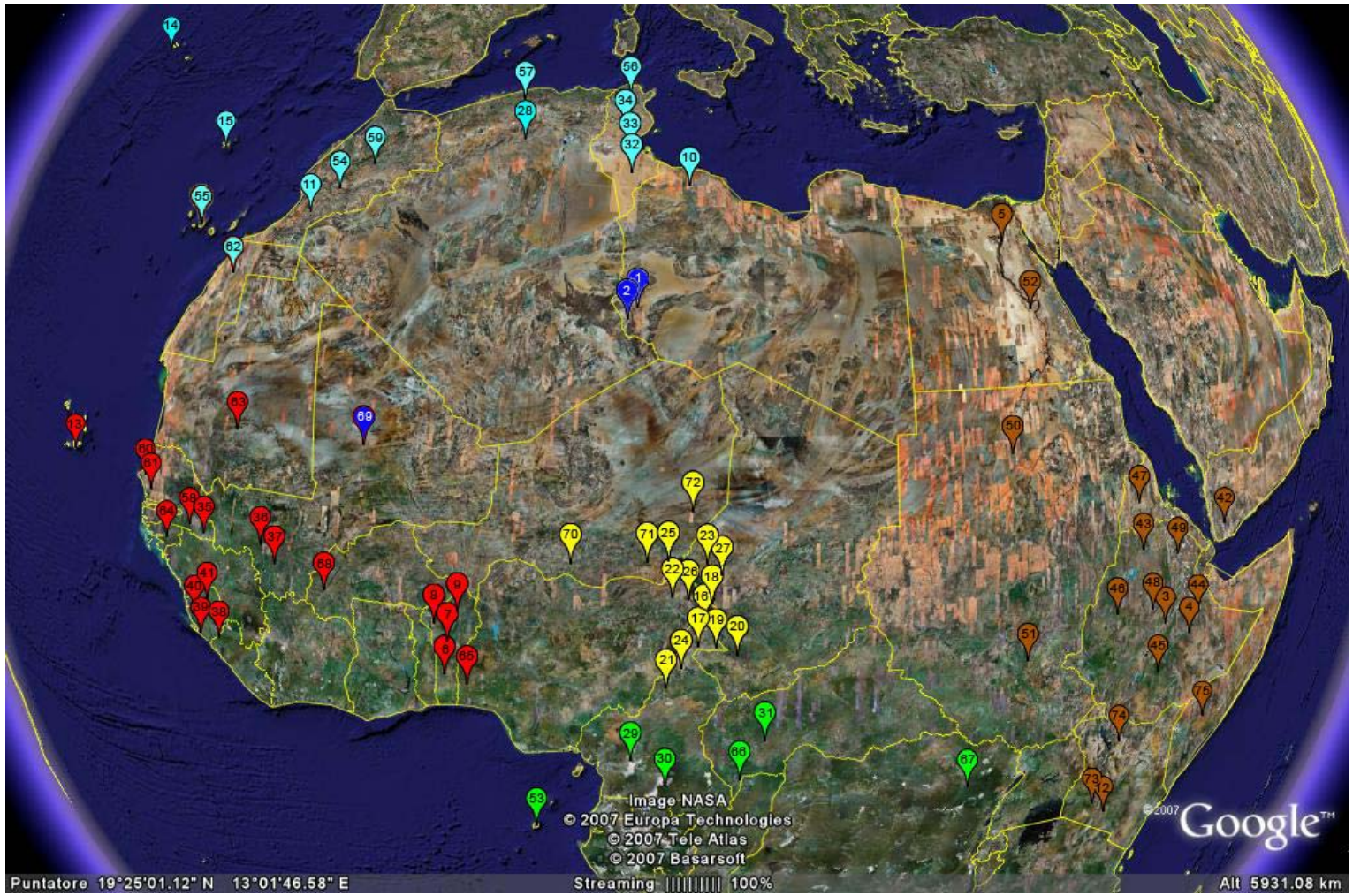
Image NASA
© 2007 Europa Technologies
© 2007 Tele Atlas
© 2007 Basarsoft
©2007 Google™
Puntatore 19°25'01.12" N 13°01'46.58" E
Streaming 100%
Alt 5931.08 km

**Figure A1**: Extant samples from populations approximately north of the Equator used in this study for comparative analyses. Numeric codes as in Table A3. Colours code are as follows: Blue dots, Tuareg samples; other colours indicate geographic groups: Red dots, Western Africa; yellow dots, Central Africa; light blue dots, North Africa; brown dots, Eastern Africa; green dots, Equatorial Africa. Populations south of the Equator (indicated as SA, Southern Africa, in Table A3) are not represented in the map being less significant for comparative analyses; whenever these populations (or their haplotypes in the case of Networks) appear in the graphic representations of statistical analyses (e.g. MDS, or Network), they are depicted in grey.

**TABLE A4**

Haplogroup frequencies in some African populations described in literature. In some cases, *sister groups* or phylogenetically strictly related haplogroups were grouped in a single variable in order to uniform data from samples with different level of classification (more/less "refined").

| Population | n | N/N1/R | M1 | U/K[a] | H-CRS | H | preHV/HV | U5/U5a | U5b | V | preJT/JT | L0a/f | L1*/L1b/L1c | L2a | L2b/c | L2d | L3e1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Tuareg Libya[b] | 129 | 0.0 | 1.6 | 0.0 | 55.8 | 5.4 | 0.0 | 0.0 | 0.0 | 3.9 | 0.0 | 6.2 | 3.9 | 10.1 | 1.6 | 0.0 | 3.1 |
| Hide | 23 | 0.0 | 0.0 | 0.0 | 4.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 8.7 | 13.0 | 17.4 | 4.3 | 4.3 | 4.3 |
| Mafa | 32 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 12.5 | 3.1 | 9.4 | 18.8 | 0.0 | 3.1 |
| Kotoko | 56 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 10.7 | 1.8 | 19.6 | 1.8 | 0.0 | 3.6 |
| Masa | 32 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 9.4 | 6.3 | 18.8 | 9.4 | 0.0 | 0.0 |
| Buduma | 30 | 0.0 | 13.3 | 3.3 | 3.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 3.3 | 0.0 | 23.3 | 6.7 | 16.7 | 0.0 |
| Arabs Chad | 27 | 0.0 | 3.7 | 3.7 | 0.0 | 0.0 | 18.5 | 0.0 | 0.0 | 0.0 | 0.0 | 11.1 | 0.0 | 33.3 | 0.0 | 0.0 | 0.0 |
| Arabs Shuwa | 38 | 0.0 | 10.5 | 5.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 2.6 | 21.1 | 0.0 | 0.0 | 2.6 |
| Fali | 40 | 0.0 | 0.0 | 0.0 | 0.0 | 2.5 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 10.0 | 12.5 | 12.5 | 0.0 | 2.5 | 0.0 |
| Kanembu | 50 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 2.0 | 0.0 | 0.0 | 0.0 | 6.0 | 16.0 | 38.0 | 4.0 | 0.0 | 0.0 |
| Kanuri | 31 | 0.0 | 0.0 | 0.0 | 3.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 6.5 | 3.2 | 12.9 | 6.5 | 0.0 | 3.2 |
| Kenia | 84 | 0.0 | 4.8 | 0.0 | 0.0 | 0.0 | 1.2 | 0.0 | 0.0 | 0.0 | 1.2 | 27.4 | 4.8 | 9.5 | 0.0 | 1.2 | 6.0 |
| Afar | 16 | 0.0 | 6.3 | 6.3 | 0.0 | 0.0 | 12.5 | 0.0 | 0.0 | 0.0 | 0.0 | 6.3 | 18.8 | 18.8 | 0.0 | 0.0 | 0.0 |
| Oromo_K | 33 | 0.0 | 18.2 | 6.1 | 0.0 | 0.0 | 18.2 | 0.0 | 0.0 | 0.0 | 3.0 | 6.1 | 0.0 | 12.1 | 0.0 | 0.0 | 0.0 |
| Amhara_K | 120 | 3.3 | 15.8 | 1.7 | 0.0 | 1.7 | 11.7 | 0.0 | 0.0 | 0.0 | 5.0 | 10.0 | 0.8 | 15.0 | 3.3 | 0.0 | 0.0 |
| Gurage | 21 | 9.5 | 14.3 | 9.5 | 0.0 | 0.0 | 14.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 4.8 | 9.5 | 0.0 | 0.0 | 0.0 |
| Tigrai | 45 | 2.2 | 11.1 | 6.7 | 0.0 | 0.0 | 17.8 | 0.0 | 0.0 | 0.0 | 13.3 | 6.7 | 0.0 | 6.7 | 2.2 | 0.0 | 0.0 |
| Mende | 55 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1.8 | 29.1 | 20.0 | 7.3 | 5.5 | 0.0 |
| Loko | 32 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 3.1 | 34.4 | 25.0 | 6.3 | 3.1 | 0.0 |
| Limba | 68 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1.5 | 10.3 | 22.1 | 22.1 | 1.5 | 2.9 |

**TABLE A4** continued

| Population | L3e2 | L3e3 | L3e4 | L3e5 | L3w | L3f | L3b/d | L4 | U6 | L3* | L5 | L2 | L3i | L3x | L3h | L6 | Other[c] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Tuareg Libya[b] | 1.6 | 3.1 | 0.0 | 0.0 | 2.3 | 1.6 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Hide | 0.0 | 0.0 | 0.0 | 17.4 | 0.0 | 4.3 | 17.4 | 4.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Mafa | 0.0 | 3.1 | 0.0 | 9.4 | 0.0 | 6.3 | 28.1 | 0.0 | 3.1 | 3.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Kotoko | 3.6 | 0.0 | 0.0 | 19.6 | 0.0 | 26.8 | 3.6 | 0.0 | 0.0 | 3.6 | 0.0 | 0.0 | 0.0 | 0.0 | 5.4 | 0.0 | 0.0 |
| Masa | 0.0 | 0.0 | 15.6 | 0.0 | 0.0 | 21.9 | 18.8 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Buduma | 3.3 | 0.0 | 0.0 | 0.0 | 0.0 | 20.0 | 3.3 | 3.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Arabs Chad | 3.7 | 0.0 | 0.0 | 0.0 | 0.0 | 3.7 | 22.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Arabs Shuwa | 13.2 | 0.0 | 0.0 | 7.9 | 0.0 | 2.6 | 28.9 | 0.0 | 2.6 | 2.6 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Fali | 17.5 | 0.0 | 0.0 | 22.5 | 0.0 | 5.0 | 15.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Kanembu | 0.0 | 0.0 | 0.0 | 6.0 | 0.0 | 10.0 | 8.0 | 4.0 | 0.0 | 4.0 | 2.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Kanuri | 6.5 | 3.2 | 0.0 | 6.5 | 0.0 | 12.9 | 25.8 | 3.2 | 3.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 3.2 | 0.0 | 0.0 |
| Kenia | 2.4 | 3.6 | 0.0 | 0.0 | 0.0 | 6.0 | 7.1 | 19.0 | 1.2 | 0.0 | 3.6 | 1.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Afar | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 6.3 | 12.5 | 6.3 | 0.0 | 0.0 | 6.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Oromo_K | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 9.1 | 3.0 | 3.0 | 3.0 | 0.0 | 3.0 | 0.0 | 3.0 | 12.1 | 0.0 | 0.0 | 0.0 |
| Amhara_K | 0.0 | 0.0 | 0.0 | 0.0 | 3.3 | 5.0 | 2.5 | 6.7 | 1.7 | 0.0 | 3.3 | 0.0 | 2.5 | 2.5 | 1.7 | 0.8 | 1.7 |
| Gurage | 0.0 | 0.0 | 0.0 | 0.0 | 4.8 | 4.8 | 0.0 | 4.8 | 4.8 | 0.0 | 9.5 | 0.0 | 0.0 | 4.8 | 0.0 | 4.8 | 0.0 |
| Tigrai | 0.0 | 0.0 | 0.0 | 0.0 | 2.2 | 6.7 | 0.0 | 13.3 | 4.4 | 0.0 | 0.0 | 0.0 | 0.0 | 2.2 | 0.0 | 0.0 | 4.4 |
| Mende | 5.5 | 0.0 | 0.0 | 0.0 | 0.0 | 5.5 | 18.2 | 1.8 | 0.0 | 3.6 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1.8 |
| Loko | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 21.9 | 0.0 | 6.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Limba | 1.5 | 0.0 | 7.4 | 0.0 | 0.0 | 4.4 | 22.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 4.4 |

**TABLE A4** continued

| Population | n | N/N1/R | M1 | U/K[a] | H-CRS | H | preHV/HV | U5/U5a | U5b | V | preJT/JT | L0a/f | L1*/L1b/L1c | L2a | L2b/c | L2d | L3e1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Fulani | 186 | 0.0 | 0.0 | 0.0 | 0.5 | 0.0 | 0.0 | 0.0 | 3.2 | 0.5 | 3.8 | 0.0 | 29.6 | 3.8 | 17.2 | 2.2 | 3.2 |
| Temne | 120 | 0.0 | 0.0 | 0.0 | 0.0 | 0.8 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 5.0 | 18.3 | 17.5 | 15.8 | 5.8 | 0.0 |
| Guineans | 372 | 0.0 | 1.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 2.7 | 0.0 | 0.0 | 5.1 | 15.3 | 16.4 | 24.2 | 1.9 | 0.0 |
| Mauritania | 64 | 0.0 | 0.0 | 4.7 | 7.8 | 10.9 | 3.1 | 1.6 | 4.7 | 3.1 | 3.1 | 0.0 | 21.9 | 12.5 | 3.1 | 0.0 | 0.0 |
| Malinke | 31 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 3.2 | 16.1 | 25.8 | 29.0 | 0.0 | 0.0 |
| Bambara | 52 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1.9 | 0.0 | 0.0 | 0.0 | 0.0 | 3.8 | 15.4 | 21.2 | 13.5 | 1.9 | 1.9 |
| Chenini Berbers | 53 | 1.9 | 0.0 | 20.8 | 5.7 | 7.5 | 15.1 | 0.0 | 0.0 | 0.0 | 35.8 | 0.0 | 3.8 | 3.8 | 0.0 | 0.0 | 0.0 |
| Sened Berbers | 53 | 0.0 | 11.3 | 5.7 | 9.4 | 15.1 | 5.7 | 3.8 | 0.0 | 0.0 | 7.5 | 0.0 | 5.7 | 9.4 | 0.0 | 0.0 | 0.0 |
| Matmata Berbers | 49 | 8.2 | 2.0 | 10.2 | 20.4 | 6.1 | 0.0 | 0.0 | 0.0 | 16.4 | 8.2 | 0.0 | 2.0 | 4.1 | 0.0 | 0.0 | 0.0 |
| Tunisians | 47 | 0.0 | 4.3 | 10.6 | 10.6 | 12.8 | 6.4 | 0.0 | 2.1 | 0.0 | 10.6 | 0.0 | 0.0 | 8.5 | 2.1 | 0.0 | 2.1 |
| Western Tuareg | 24 | 0.0 | 0.0 | 0.0 | 8.3 | 4.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 4.2 | 12.5 | 37.5 | 4.2 | 0.0 | 4.2 |
| Kikuyu | 24 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 16.7 | 0.0 | 12.5 | 0.0 | 0.0 | 8.3 |
| Egyptians (Gurna)[d] | 32 | 9.4 | 18.8 | 9.4 | 6.3 | 9.4 | 0.0 | 0.0 | 0.0 | 0.0 | 12.5 | 6.3 | 0.0 | 3.1 | 0.0 | 0.0 | 0.0 |
| Mozabiti Berbers | 85 | 0.0 | 4.7 | 10.6 | 8.2 | 16.5 | 0.0 | 0.0 | 0.0 | 8.2 | 8.2 | 0.0 | 0.0 | 5.9 | 0.0 | 0.0 | 0.0 |
| Fayum Egypt | 69 | 0.0 | 15.9 | 7.2 | 0.0 | 7.2 | 5.8 | 4.3 | 0.0 | 2.9 | 14.5 | 4.3 | 2.9 | 10.1 | 2.9 | 1.4 | 0.0 |
| Amhara | 90 | 2.2 | 17.8 | 1.1 | 0.0 | 2.2 | 1.1 | 1.1 | 0.0 | 0.0 | 16.7 | 3.3 | 2.2 | 3.3 | 1.1 | 0.0 | 0.0 |
| Oromo | 84 | 10.7 | 14.3 | 2.4 | 0.0 | 0.0 | 1.2 | 0.0 | 0.0 | 0.0 | 9.5 | 6.0 | 1.2 | 7.1 | 2.4 | 0.0 | 0.0 |
| Bariba | 50 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 4.0 | 20.0 | 34.0 | 2.0 | 4.0 | 0.0 |
| Dendi | 47 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 8.5 | 17.0 | 38.3 | 12.8 | 0.0 | 0.0 |
| Fon | 49 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 18.4 | 34.7 | 6.1 | 0.0 | 0.0 |

**TABLE A4** continued

| Population | L3e2 | L3e3 | L3e4 | L3e5 | L3w | L3f | L3b/d | L4 | U6 | L3* | L5 | L2 | L3i | L3x | L3h | L6 | Other[c] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Fulani | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1.6 | 33.3 | 0.0 | 0.0 | 1.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Temne | 8.3 | 0.0 | 0.0 | 0.0 | 0.0 | 5.8 | 20.8 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1.7 |
| Guineans | 4.3 | 0.0 | 3.0 | 0.0 | 0.0 | 2.4 | 18.0 | 0.0 | 2.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 3.5 | 0.0 | 0.0 |
| Mauritania | 0.0 | 0.0 | 1.6 | 0.0 | 0.0 | 0.0 | 4.7 | 0.0 | 17.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Malinke | 3.2 | 0.0 | 3.2 | 0.0 | 0.0 | 0.0 | 19.4 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Bambara | 7.7 | 0.0 | 0.0 | 0.0 | 0.0 | 1.9 | 30.8 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Chenini Berbers | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1.9 | 3.8 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Sened Berbers | 1.9 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 5.7 | 0.0 | 7.5 | 1.9 | 0.0 | 1.9 | 0.0 | 0.0 | 0.0 | 0.0 | 7.5 |
| Matmata Berbers | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 2.0 | 0.0 | 2.0 | 16.4 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 2.0 |
| Tunisians | 4.3 | 0.0 | 0.0 | 0.0 | 0.0 | 4.3 | 2.1 | 0.0 | 4.3 | 2.1 | 0.0 | 2.1 | 0.0 | 0.0 | 0.0 | 0.0 | 10.6 |
| Western Tuareg | 4.2 | 0.0 | 0.0 | 0.0 | 0.0 | 4.2 | 8.3 | 4.2 | 4.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Kikuyu | 4.2 | 4.2 | 0.0 | 0.0 | 8.3 | 4.2 | 0.0 | 8.3 | 8.3 | 4.2 | 8.3 | 0.0 | 0.0 | 0.0 | 12.5 | 0.0 | 0.0 |
| Egyptians (Gurna)[d] | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 12.5 | 6.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 6.3 |
| Mozabiti Berbers | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 4.7 | 0.0 | 30.6 | 2.4 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Fayum Egypt | 1.4 | 0.0 | 0.0 | 0.0 | 0.0 | 4.3 | 2.9 | 0.0 | 0.0 | 2.9 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 8.7 |
| Amhara | 0.0 | 0.0 | 0.0 | 0.0 | 2.2 | 13.3 | 0.0 | 5.6 | 4.4 | 8.9 | 1.1 | 0.0 | 0.0 | 3.3 | 0.0 | 3.3 | 5.6 |
| Oromo | 0.0 | 0.0 | 0.0 | 0.0 | 1.2 | 9.5 | 2.4 | 6.0 | 7.1 | 8.3 | 2.4 | 0.0 | 3.6 | 1.2 | 0.0 | 3.6 | 0.0 |
| Bariba | 8.0 | 0.0 | 0.0 | 0.0 | 0.0 | 2.0 | 24.0 | 0.0 | 0.0 | 0.0 | 0.0 | 2.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Dendi | 8.5 | 0.0 | 0.0 | 0.0 | 0.0 | 6.4 | 6.4 | 0.0 | 0.0 | 0.0 | 0.0 | 2.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Fon | 20.4 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 18.4 | 0.0 | 0.0 | 0.0 | 0.0 | 2.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |

[a] This variable includes phylogenetically realted haplogroups U2, U3, U4, U7, U8, U9, K, which likely derive from the same ancestor (Achilli et al. 2005), and U1. Haplogroups U5a, U5b, and U6 are keep separated as they are significantly spread in Africa, but their diffusion is hypothetically associated with different migration events (Olivieri et al. 2006, Achilli et al. 2005, Coia et al. 2005).

[b] Includes Al Awaynat (111) and Tahala (18), see Results.

[c] This variable includes haplogroups found at low frequencies in the African continent: I, A2, X, and W

[d] Here the Egyptian sample is represented only by the Gurnas, whose frequency data are available in Stevanovitch et al. 2004, Table 3.

**TABLE A5**

Polimorphic sites observed in the HVS-I, HVS-II, and some coding regions submitted to sequencing and RFLP analysis, in 129 Tuareg from two villages of Fezzan (Libya): Al Awaynat (AL) and Tahala (TA). In the RFLP typing columns, status of polymorphism is represented as presence (+) or absence (–) of the restriction site. Enzyme codes are as follows: a=*Alu*I, b=*Mse*I, c=*Nla*III, d=*Hinf*I, e=*Hpa*I, f=*Mbo*I, g=*Taq*I, h=*Hae*III, i=*Mnl*I, l=*Dde*I. Nucleotide changes are specified only for transversions. Mutations are numbered relative to revCRS (Andrews et al. 1999).

| ID sample | HVS-I (16---) | HVS-II | Sequenced coding region | 7025 (a) | 14766 (b) | 4577 (c) | 12308 (d) | 7055 (a) | 3592 (e) | 2349 (f) | 8616 (f) | 10084 (g) | 10397 (a) | 13803 (h) | 4157 (a) | 10871 (i) | 13957 (h) | 10394 (l) | Haplogroup |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AL1 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL3 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL5 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL10 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL11 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL12 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL13 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL14 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL16 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL17 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL20 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL23 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL25 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL26 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL27 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL30 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL31 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL33 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL35 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL36 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |

**TABLE A5** continued

| ID sample | HVS-I (16---) | HVS-II | Sequenced coding region | 7025 (a) | 14766 (b) | 4577 (c) | 12308 (d) | 7055 (a) | 3592 (e) | 2349 (f) | 8616 (f) | 10084 (g) | 10397 (a) | 13803 (h) | 4157 (a) | 10871 (i) | 13957 (h) | 10394 (l) | Haplogroup |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AL39 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL40 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL41 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL43 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL46 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL47 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL48 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL49 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL52 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL54 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL55 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL59 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL61 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL63 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL65 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL66 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL70 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL71 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL76 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL79 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL81 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL82 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL83 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL84 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |

**TABLE A5** continued

| ID sample | HVS-I (16---) | HVS-II | Sequenced coding region | 7025 (a) | 14766 (b) | 4577 (c) | 12308 (d) | 7055 (a) | 3592 (e) | 2349 (f) | 8616 (f) | 10084 (g) | 10397 (a) | 13803 (h) | 4157 (a) | 10871 (i) | 13957 (h) | 10394 (l) | Haplogroup |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AL86 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL87 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL88 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL89 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL90 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL93 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL94 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL95 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL97 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL98 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL99 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL100 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL103 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL104 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL105 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL106 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL108 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL109 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL110 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| TA2 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| TA4 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| TA9 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| TA10 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| TA11 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |

**TABLE A5** continued

| ID sample | HVS-I (16---) | HVS-II | Sequenced coding region | 7025 (a) | 14766 (b) | 4577 (c) | 12308 (d) | 7055 (a) | 3592 (e) | 2349 (f) | 8616 (f) | 10084 (g) | 10397 (a) | 13803 (h) | 4157 (a) | 10871 (i) | 13957 (h) | 10394 (l) | Haplogroup |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TA15 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| TA16 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| TA17 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| TA18 | CRS | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL32 | 37 256 | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL53 | 37 256 | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL68 | 37 256 | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL78 | 37 256 | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| TA3 | 37 256 | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| TA13 | 37 256 | 263 | 1438 3010 4769 | – | – | | | | | | | | | | | | | | H1 |
| AL111 | 183C 189 | 73 263 | 1438 3010 4769 | – | – | | – | | – | | | – | – | | | | | | H1 |
| AL2 | 129 148 168 172 187 188G 189 230 311 320 | 64 93 146 150 152 185 189 200 236 247 263 | | | | | | | | | | | | | | | | | L0a1 |
| AL8 | 129 148 168 172 187 188G 189 230 311 320 | 64 93 146 150 152 185 189 200 236 247 263 | | | | | | | | | | | | | | | | | L0a1 |
| AL19 | 129 148 168 172 187 188G 189 230 311 320 | 64 93 146 150 152 185 189 200 236 247 263 | | | | | | | | | | | | | | | | | L0a1 |
| AL22 | 129 148 168 172 187 188G 189 230 311 320 | 64 93 146 150 152 185 189 200 236 247 263 | | | | | | | | | | | | | | | | | L0a1 |
| AL29 | 129 148 168 172 187 188G 189 230 311 320 | 64 93 146 150 152 185 189 200 236 247 263 | | | | | | | | | | | | | | | | | L0a1 |
| AL44 | 129 148 168 172 187 188G 189 230 311 320 | 64 93 146 150 152 185 189 200 236 247 263 | | | | | | | | | | | | | | | | | L0a1 |
| AL51 | 129 148 168 172 187 188G 189 230 311 320 | 64 93 146 150 152 185 189 200 236 247 263 | | | | | | | | | | | | | | | | | L0a1 |

**TABLE A5** continued

| ID sample | HVS-I (16---) | HVS-II | Sequenced coding region | 7025 (a) | 14766 (b) | 4577 (c) | 12308 (d) | 7055 (a) | 3592 (e) | 2349 (f) | 8616 (f) | 10084 (g) | 10397 (a) | 13803 (h) | 4157 (a) | 10871 (i) | 13957 (h) | 10394 (l) | Haplogroup |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AL77 | 129 148 168 172 187 188G 189 230 311 320 | 64 93 146 150 152 185 189 200 236 247 263 | | | | | | | | | | | | | | | | | L0a1 |
| AL9 | 126 172 187 189 223 264 270 278 293 301 311 | 73 146 152 182 185 195 247 263 357 | | | | | | – | + | | | | | | | | | | L1b1 |
| AL38 | 126 172 187 189 223 264 270 278 293 301 311 | 73 146 152 182 185 195 247 263 357 | | | | | | – | + | | | | | | | | | | L1b1 |
| TA 5 | 126 187 189 223 264 265 270 278 293 311 399 | 73 182 185 247 263 357 | | | | | | – | + | | | | | | | | | | L1b1 |
| TA 6 | 126 187 189 223 264 265 270 278 293 311 399 | 73 182 185 247 263 357 | | | | | | – | + | | | | | | | | | | L1b1 |
| TA 7 | 126 187 189 223 264 265 270 278 293 311 399 | 73 182 185 247 263 357 | | | | | | – | + | | | | | | | | | | L1b1 |
| AL7 | 189 192 223 278 294 390 | 73 143 146 152 195 263 | | | | | | | + | | | | | | | | | | L2a |
| AL15 | 189 192 223 278 294 368 390 | 73 143 146 152 195 263 | | | | | | | + | | | | | | | | | | L2a |
| AL18 | 189 192 223 278 294 390 | 73 143 146 152 195 263 | | | | | | | + | | | | | | | | | | L2a |
| AL24 | 189 192 223 278 294 390 | 73 143 146 152 195 263 | | | | | | | + | | | | | | | | | | L2a |
| AL37 | 189 192 223 278 294 390 | 73 143 146 152 195 263 | | | | | | | + | | | | | | | | | | L2a |
| AL42 | 189 192 223 278 294 390 | 73 143 146 152 195 263 | | | | | | | + | | | | | | | | | | L2a |
| AL72 | 189 192 223 278 294 390 | 73 143 146 152 195 263 | | | | | | | + | | | | | | | | | | L2a |
| AL73 | 189 192 223 278 294 390 | 73 143 146 152 195 263 | | | | | | | + | | | | | | | | | | L2a |
| AL74 | 189 192 223 278 294 390 | 73 143 146 152 195 263 | | | | | | | + | | | | | | | | | | L2a |
| AL75 | 189 192 223 278 294 390 | 73 143 146 152 195 263 | | | | | | | + | | | | | | | | | | L2a |
| AL96 | 189 192 223 278 294 390 | 73 143 146 152 195 198 263 | | | | | | | + | | | | | | | | | | L2a |
| TA 14 | 189 192 223 263 278 294 309 390 | 73 143 146 152 195 263 | | | | | | | + | | | | | + | | | | | L2a1 |

**TABLE A5** continued

| ID sample | HVS-I (16---) | HVS-II | Sequenced coding region | 7025 (a) | 14766 (b) | 4577 (c) | 12308 (d) | 7055 (a) | 3592 (e) | 2349 (f) | 8616 (f) | 10084 (g) | 10397 (a) | 13803 (h) | 4157 (a) | 10871 (i) | 13957 (h) | 10394 (l) | Haplogroup |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AL64 | 223 278 286 294 309 390 399 | 73 146 152 195 263 | | | | | | | + | | | | | | | | | | L2a1a |
| AL6 | 114A 129 213 218 223 278 390 | 73 146 150 152 182 183 195 198 204 207 263 | | | | | | | + | | | | | | + | | | | L2b |
| AL67 | 114A 129 213 218 223 278 390 | 73 146 150 152 182 183 195 198 204 207 263 | | | | | | | + | | | | | | + | | | | L2b |
| AL34 | 184 223 260 311 | 73 150 152 263 | | | | | | | – | – | + | – | – | | | | | – | L3w |
| AL80 | 184 223 260 311 | 73 150 152 263 | | | | | | | – | – | + | – | – | | | | | – | L3w |
| AL102 | 184 223 260 311 | 73 150 152 263 | | | | | | | – | – | + | – | – | | | | | + | L3w |
| AL4 | 223 327 | 73 150 189 195 200 263 | | | | | | | | + | | | | | | | | | L3e1 |
| AL21 | 223 327 | 73 150 189 195 200 263 | | | | | | | | + | | | | | | | | | L3e1 |
| AL45 | 223 327 | 73 150 189 195 200 263 | | | | | | | | + | | | | | | | | | L3e1 |
| AL62 | 223 327 | 73 150 189 195 200 263 | | | | | | | | + | | | | | | | | | L3e1 |
| AL56 | 223 311 320 | 73 150 195 198 263 | | | | | | | | + | | | | | | | | | L3e2 |
| AL50 | 172 183C 189 223 320 | 73 150 195 263 | | | | | | | | + | | | | | | | | | L3e2b |
| AL58 | 183C 189 223 265T | 73 150 263 | | | | | | | | + | | | | | | | | | L3e3 |
| AL69 | 183C 189 223 265T | 73 150 263 | | | | | | | | + | | | | | | | | | L3e3 |
| AL85 | 183C 189 223 265T | 73 150 263 | | | | | | | | + | | | | | | | | | L3e3 |
| AL92 | 183C 189 223 265T | 73 150 263 | | | | | | | | + | | | | | | | | | L3e3 |
| TA 8 | 209 218 223 292 311 | 73 189 200 263 | | | | | | | | | | | | | | – | | | L3f1 |
| TA 12 | 209 218 223 292 311 | 73 189 200 263 | | | | | | | | | | | | | | – | | | L3f1 |
| AL60 | 189 298 | 72C 263 | | + | – | | | | | | | | | | | + | | | V |
| AL91 | 189 298 | 72C 263 | 15904 | + | – | – | | | | | | | | | | + | | | V |
| AL101 | 189 298 | 72C 263 | | + | – | | | | | | | | | | | + | | | V |

**TABLE A5** continued

| ID sample | HVS-I (16---) | HVS-II | Sequenced coding region | 7025 (a) | 14766 (b) | 4577 (c) | 12308 (d) | 7055 (a) | 3592 (e) | 2349 (f) | 8616 (f) | 10084 (g) | 10397 (a) | 13803 (h) | 4157 (a) | 10871 (i) | 13957 (h) | 10394 (l) | Haplogroup |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AL57 | 234 298 | 72C 263 | 15904 | + | – | – | | | | | | | | | | + | | | V |
| TA 1 | 234 298 | 72C 263 | | + | – | | | | | | | | | | | + | | | V |
| AL28 | 129 182 183C 189 249 311 | 73 195 263 | | | | | | | – | – | | – | | | | – | + | | M1 |
| AL107 | 129 182 183C 189 249 311 | 73 195 263 | | | | | | | – | – | | – | + | | | – | + | | M1 |