

1  
2  
3 1 TITLE: Mitochondrial DNA structure in North Africa reveals a genetic discontinuity in the  
4 Nile Valley.  
5  
6 2

7  
8 3 AUTHORS: Karima Fadhlou-Zid <sup>1,\*</sup>, Laura Rodríguez-Botigué <sup>2,\*</sup>, Nejib Naoui <sup>3</sup>, Amel  
9 Benammar-Elgaaied <sup>1</sup>, Francesc Calafell <sup>2,4</sup>, David Comas <sup>2,4</sup>  
10  
11 4

12 5 \*Both authors contributed equally to this work  
13  
14 6

15  
16  
17 7 <sup>1</sup>Laboratory of Genetics, Immunology and Human Pathology, Faculty of Science of Tunis,  
18 University Tunis El Manar, 2092 Tunis, Tunisia.  
19  
20 8

21  
22 9 <sup>2</sup> Institut de Biologia Evolutiva (CSIC-UPF), Department de Ciències Experimentals i de la  
23 Salut, Universitat Pompeu Fabra, 08003 Barcelona, Spain.  
24  
25 10

26  
27 3  
28 11 Laboratory of Medical Biology. Tunis, Tunisia.  
29

30 4  
31 12 CIBERESP, 08003 Barcelona, Spain  
32  
33 13

34  
35 14 26 text pages, 3 figures, 3 tables (plus 2 tables and 1 figure as Supplementary material).  
36  
37 15

38  
39  
40 16 RUNNING TITLE: mtDNA structure in North Africa  
41  
42 17

43  
44  
45 18 KEYWORDS: uniparental marker, haplogroup, gene flow  
46

47 19 Corresponding author: DAVID COMAS  
48

49  
50 20 Institut de Biologia Evolutiva (CSIC-UPF), Department de Ciències Experimentals i de la  
51 Salut, Universitat Pompeu Fabra, 08003 Barcelona, Spain.  
52  
53 21

54 22 Tel : +34 93 3160843, Fax: +34 93 3160901  
55

56  
57 23 E-mail: [david.comas@upf.edu](mailto:david.comas@upf.edu)  
58  
59  
60

**Abstract**

Human population history in North Africa has been constrained in an east-west direction due to the geographical barriers imposed by the Sahara Desert and the Mediterranean Sea. Although these barriers have not completely impeded human migrations, genetic studies have shown that an east-west genetic gradient exists. However, the lack of genetic information of certain geographical areas and the focus of some studies in parts of the North African landscape have limited the global view of the genetic pool of North African populations. In order to provide a global view of the North African genetic landscape and population structure, we have analyzed ~2,300 North African mitochondrial DNA lineages (including 269 new sequences from Libya, in the first mtDNA study of the general Libyan population). Our results show a clinal distribution of certain haplogroups, some of them more frequent in Western (H, HV0, L1b, L3b, U6) or Eastern populations (L0a, R0a, Nib, I, J, M1) that might be the result of human migrations from the Middle East, sub-Saharan Africa, and Europe. Despite this clinal pattern, a genetic discontinuity is found in the Libyan/Egyptian border, suggesting a differential gene flow in the Nile River Valley. Finally, frequency of the post-LGM subclades H1 and H3 is predominant in Libya within the H sequences, highlighting the magnitude of the LGM expansion in North Africa.

1 North Africa is a region characterized by a complex history of demographic events  
2 and the extent of its genetic effect on extant human populations is still far from being known.  
3 Despite being part of the African continent, its demographic history, conditioned in an east-  
4 west direction by the barriers imposed by the Mediterranean Sea and the Sahara Desert, has  
5 been completely different from the rest of the continent. According to archaeological records,  
6 the first modern humans in North Africa produced the Aterian stone industry during the Early  
7 Upper Paleolithic, around 45,000 years ago (ya) (Garcea and Giraudi 2006). No clear  
8 connections have been established between this first human industry and subsequent cultures  
9 in the region, such as the Ibero-Maurusian industry (22,000 - 9,500 ya) (Newman 1995). The  
10 Ibero-Maurusian culture was followed by the Capsian industry (10,000 – 4,700 ya) (Desanges  
11 1990) that persisted well after the adoption of farming and agriculture, which began around  
12 5,500 years ago in the region. The persistence of a pre-Neolithic culture in Neolithic times  
13 might indicate cultural replacement with admixture, rather than a population replacement of  
14 the autochthonous pre-Neolithic people by Neolithic farmers originated in the Middle East. In  
15 general terms, the prehistoric cultural changes in North Africa were quite independent of the  
16 dynamics on the European shores of the Mediterranean. Historical records document trade  
17 routes across the Sahara Desert and contacts between both Mediterranean shores and the  
18 Middle East, such as Phoenicians, Romans, Vandals, and Byzantines. The first Arab invasion,  
19 initially confined to Egypt, started in A.D. 643 and may have involved only a few thousand  
20 individuals (McEvedy and Jones 1980). The Arabs began to impose their religion and  
21 language over the Berber autochthonous population, a process that culminated with the  
22 second and more numerous Arab wave in which the Bedouin reached the Maghreb (northwest  
23 Africa) in the 11th century. The later arrivals to northern Africa in colonial times include  
24 Europeans and Ottoman Turks, mainly in Egypt.

1           The genetic data available for North Africa is scarce, which limits the power to test  
2 population history hypotheses. Most of the African genetic diversity studies have been  
3 focused on the origin of our species and the first dispersals out of Africa (see for instance  
4 Tishkoff et al. (2009)), processes in which North Africa had a marginal role, which made the  
5 region less attractive to human population geneticists. The analyses based on frequencies of  
6 classical genetic polymorphisms (blood groups, red cell enzymes and serum proteins) have  
7 shown that the genetic landscape in northern Africa presents an east-west pattern of variation  
8 without differences between Arabs and Berbers, pointing to a sizeable Upper Paleolithic  
9 component in current northern African populations, whereas the Neolithic diffusion in the  
10 region was more a cultural than a demic process (Barbujani et al. 1994; Bosch et al. 1997). As  
11 for autosomal markers, only some STRs (Bosch et al. 2000; Khodjet-El-Khil et al. 2008) and  
12 *Alu* polymorphisms (Comas et al. 2000; Flores et al. 2000; Frigi et al. 2010; González-Pérez  
13 et al. 2003) have been analyzed in a few northern African samples. Concerning massive  
14 analysis of genome-wide markers, only 30 Mozabite individuals (a Berber isolate from  
15 Algeria) have been analyzed for 650,000 SNPs (Li et al. 2008), showing various degrees of  
16 admixture between sub-Saharan, Middle Easterners and Europeans. The analysis of Y  
17 chromosome lineages has shown a high frequency of two specific North African haplogroups  
18 (E1b1b1a and E1b1b1b), although their origins have been controversial since some analyses  
19 have suggested a Paleolithic component (Bosch et al. 2001), whereas others have pointed to a  
20 Neolithic origin (Arredi et al. 2004; Cruciani et al. 2004; Cruciani et al. 2007; Semino et al.  
21 2004). The analysis of mitochondrial (mtDNA) lineages has shown that, in spite of an  
22 important sub-Saharan contribution, most haplogroups in North Africa are of Eurasian origin  
23 (Fadhlaoui-Zid et al. 2004; Harich et al. 2010; Krings et al. 1999; Plaza et al. 2003; Rando et  
24 al. 1998). Some can be traced to ancient Paleolithic times (such as haplogroups U6, M1,  
25 which are almost specific of northern African populations); however, some maternal lineages

1  
2  
3 1 have been recently acquired from Europe or the Middle East (such as haplogroups U5, V,  
4  
5 2 R0a, J1b, U3) (Gonzalez et al. 2007; Maca-Meyer et al. 2003; Olivieri et al. 2006). Several  
6  
7 3 studies suggest that at the end of the Last Glacial Maximum (LGM), around 13,000 ya,  
8  
9 4 humans expanded from the Franco-Cantabrian refuge towards Europe and North Africa,  
10  
11 5 spreading mtDNA haplogroups H1, H3 and V (Achilli et al. 2004; Cherni et al. 2009; Ennafaa  
12  
13 6 et al. 2009; Loogvali et al. 2004; Pereira et al. 2005; Rhouda et al. 2009; Torroni et al. 1998;  
14  
15 7 Torroni et al. 2001). However, a recent analysis of mtDNA diversity on Iberian populations  
16  
17 8 points to the opposite conclusion: it suggests the absence of such an expansion (Garcia et al.  
18  
19 9 2010). In addition, a large degree of genetic heterogeneity has been shown in North African  
20  
21 10 maternal lineages compared to other geographical regions such as Europe (Fadhlaoui-Zid et  
22  
23 11 al. 2004; Plaza et al. 2003).

24  
25 12 One of the main limitations of the genetic analyses of North African populations is the  
26  
27 13 lack of representative and homogeneously distributed samples. For instance, most of the  
28  
29 14 studies have focused in the north western samples and Egypt, being Libya a region with  
30  
31 15 almost no genetic data with the exception of a Tuareg sample (Ottoni et al. 2009). The  
32  
33 16 presence of gaps in the coverage of genetic studies across North Africa creates an artificial  
34  
35 17 division between Eastern (Egypt) and Western populations (Morocco, Algeria and Tunisia)  
36  
37 18 and prevents explicitly geography-based analyses. Due to this lack of data, most of population  
38  
39 19 genetic research in the area has a local scope rather than being comprehensive and covering  
40  
41 20 the whole region.

42  
43 21 Within the present analysis we aim to address several questions concerning the  
44  
45 22 population history of North Africa. Is there any genetic structure that differentiates North  
46  
47 23 African populations? What is the influence of the trade routes and natural corridors such as  
48  
49 24 the Nile? Conversely, have the particularly inhospitable conditions of the Western Egyptian  
50  
51 25 and Libyan deserts created a genetic barrier? These questions have not been successfully

1 answered yet, partly because of the ~2,000 Km gap in sampling between Egypt and Tunisia.  
2 In the present work we analyze for the first time the mtDNA sequences of a set of 269  
3 Libyans representing the general population of the country that will allow us filling this gap  
4 and study the whole region.

## 5 MATERIALS AND METHODS

### 6 Mitochondrial DNA sequencing and SNP genotyping in Libyan individuals

7 DNA was extracted from fresh blood from a total of 269 individuals from Libya using  
8 standard phenol-chloroform methods. The mtDNA control region was PCR amplified using  
9 primer pairs L15996 and H408, purified using the GFX PCR DNA and Gel Band purification  
10 Kit (GE Healthcare), and sequenced for both mtDNA hypervariable segments (HVS I and  
11 HVS II) as described previously (Plaza et al. 2004) using primer pairs L15996, H16401, L29,  
12 and H408 (Vigilant et al. 1989). Positions 16024 to 16391 for HVS I and positions 63 to 323  
13 for HVS II (Anderson et al. 1981; Andrews et al. 1999) were considered for further analysis.

14 Four TaqMan® probes (Applied Biosystems) were used, following supplier's  
15 recommendations, to genotype positions 3594, 10873, 12705, and 14783, diagnostic for major  
16 lineages L0-L2/L5-L6, L/M, R, and M, respectively. After this first classification, 105  
17 samples were subsequently refined by genotyping eight SNPs in the mtDNA coding region  
18 (7028, 10400, 10873, 11251 11719, 12308, and 12705 diagnostic for haplogroups H, M, L/M,  
19 J/T, R, U, and N and respectively) by means of a SNaPshot™ Multiplex kit (Applied  
20 Biosystems), as described previously (Bosch et al. 2006). Two of them (10873 and 12705)  
21 were typed with both methods and were used as controls. Finally, a dissection of haplogroup  
22 H was carried out using an additional SNaPshot™ multiplex reaction: positions 3010, 4793,  
23 4336, 6776, and 14872 were typed in order to classify individuals into subhaplogroups H1,  
24 H7, H5a, H3, and H13, respectively. Sequences of primers used for PCR amplification and  
25 primers used for genotyping can be found in Supplementary Table 1.

1 Samples were assigned to haplogroups with the joint information of the control region  
2 sequence and the SNPs in the coding region following the nomenclature previously described  
3 (Behar et al. 2008; Finnilä et al. 2001; Kivisild et al. 2004; Loogvali et al. 2004; Maca-Meyer  
4 et al. 2003; Maca-Meyer et al. 2001; Metspalu et al. 2004; Olivieri et al. 2006; Palanichamy et  
5 al. 2004; Richards et al. 2000).

### 6 **Statistical and phylogenetic analyses**

7 Previously published data of HVSI sequences ranging from positions 16024 to 16383  
8 of 28 North African populations was used for population comparisons. Their names and  
9 references can be found in Table 1. For the purposes of the present analysis, and given our  
10 focus on Libya, we define eastern North Africa as Egypt and Northern Sudan (Nubia), and  
11 western North Africa as Morocco, Western Sahara, Mauritania, Algeria, and Tunisia.  
12 Tunisian samples were 13 out of the 21 original samples in the database (62%); since Tunisia  
13 was clearly overrepresented (it contains ~6.5% of the population of the region), we sought to  
14 pool some of the Tunisian samples. The Tunisian samples that were not significantly different  
15 from each other and could be merged were identified with an AMOVA (Analysis of the  
16 Molecular Variance). In all subsequent analyses the so-called Tunisian “Andalousian”  
17 populations were pooled together in a single group and the Tunisian from Plaza et al. (2003)  
18 and the Urban Tunisian from Ottoni et al. (2009) were also merged into a single population.  
19 Sample locations are shown in Figure 1A. Intrapopulation diversity parameters (number of  
20 unique sequences, sequence diversity, nucleotide diversity and mean pairwise differences) of  
21 samples in the database were calculated using Arlequin 3.0 software (Excoffier et al. 2005).  
22 Analyses of the molecular variance (AMOVAs) were also performed with the Arlequin  
23 software.

24 A Correspondence Analysis (CA) was built using absolute haplogroup frequencies  
25 with the SPSS 15.0 software (SPSS Inc., Chicago, Illinois). Haplogroups with relative

1 frequencies below 1% were grouped with their superior clade. After using this criterion, a  
2 total of 23 haplogroups were included in the analysis. In addition, Andalusians, Macedonians,  
3 and Saudi Arabians were included as southern European and Middle Eastern representatives  
4 (Abu-Amero et al. 2008; Larruga et al. 2001; López-Soto and Sanz 2000; Plaza et al. 2003;  
5 Zimmermann et al. 2007), and South Sudanese and Mandenka as sub-Saharan representatives  
6 (Graven et al. 1995; Krings et al. 1999).

7 The geographical pattern of haplogroup distribution was investigated by computing  
8 Pearson's correlation coefficients (using SPSS 15.0 software, SPSS Inc., Chicago, Illinois)  
9 between the frequencies of each haplogroup and the geographical longitude of each  
10 population sample. Given the linear, East-West disposition of human settlement in N. Africa,  
11 longitude captures most of the geographical distance between populations.

12 Haplogroup specific median networks for haplogroups U6, M1 and H1 present in the  
13 data set and on Saudi Arabian populations were generated with the median joining algorithm  
14 (Bandelt et al. 1999) using the Network 4.5.1.0 program ([http://www.fluxus-](http://www.fluxus-engineering.com)  
15 [engineering.com](http://www.fluxus-engineering.com)). Networks were weighed taking into account the mutation rate of each  
16 position (Allard et al. 2002). Positions 16182 and 16183 were not considered because they  
17 mutate recurrently and therefore they are not phylogenetically informative. Time estimates  
18 were calculated using the rho statistic (Saillard et al. 2000) with one nucleotide substitution  
19 every 19,171 years for the HVSI sequences according to Soares et al. (2009).

## 20 RESULTS

### 21 Mitochondrial DNA sequence diversity and haplogroup composition in Libya

22 A total of 164 different sequences were found in the analysis of both hypervariable  
23 segments (HVSI and HVSII) in 269 Libyan individuals. Sequence diversity in Libyans is  
24 similar to other North African samples (Table 1), and when comparing only HVSI regions, 78  
25 (29%) Libyan sequences were not found in the dataset used. Sequences and haplogroup



1  
2  
3 1 frequencies of the Libyan population are shown in Supplementary Table 2. Lineages of west  
4  
5 2 Eurasian origin are the most frequent in Libyans (65%), followed by sub-Saharan L lineages  
6  
7  
8 3 (28%), and North African specific haplogroups U6 and M1, which have an overall frequency  
9  
10 4 of 7%.

11  
12 5 Haplogroups L2a1, L3b, L3f1b, and L1b are the most frequent (over 3%) sub-Saharan  
13  
14 6 haplogroups in Libyans. Haplogroup L2a1 is common and apparently scattered throughout  
15  
16 7 Africa (Salas et al. 2002), and therefore its geographical origin is difficult to assess. However,  
17  
18 8 L1b, L3b, and L3f1b have more restricted locations in Africa. These haplogroups, together  
19  
20 9 with other minor lineages in Libya such as L2b, L2c, L3d, and L3e have a typical Western  
21  
22 10 Africa distribution (Harich et al. 2010; Salas et al. 2002). Nonetheless, other minor lineages  
23  
24 11 present in Libya such as L0a, L3f, L3h, and L3x are more frequent in Eastern Africa. Such  
25  
26 12 haplogroup frequency distribution suggests a predominantly Western origin of L lineages in  
27  
28 13 Libya with some minor admixture of Eastern lineages.

29  
30 14 The back to Africa M1 and U6 lineages are mainly present in North Africa and show  
31  
32 15 opposite frequency gradients, being U6 significantly more frequent in the West, whereas M1  
33  
34 16 is more frequent in the East. Interestingly, these haplogroups display similar frequencies in  
35  
36 17 the Libyan mtDNA pool (4.1% for U6 and 3.3% for M1).

37  
38 18 Eurasian haplogroups HV0, H1, and K are the most frequent in Libyans (7.4%, 6.3%  
39  
40 19 and 5.2% respectively). In order to trace back the geographical origin of the H lineages in  
41  
42 20 Libya, we dissected haplogroup H in several subclades (see Material and Methods).  
43  
44 21 Compared with previous published data (Ennafaa et al. 2009), Libyan individuals exhibit an  
45  
46 22 admixture of western and eastern H subclades (Table 2). As in North West African  
47  
48 23 populations, H1 and H3 are the most frequent subclades, and account for 48% of the H  
49  
50 24 lineages. However, these frequencies are lower than those found in Maghreb populations  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1 because of the relative high proportion of H5, H7, and H13 subgroups in Libya, which are  
2 more frequent in the Near East (Roostalu et al. 2006).

3 The age estimation of haplogroup H1 based on the HVSI region in Europe is 16.0 kya  
4 (Pereira et al. 2005) and 11.7 kya in Tunisia (Cherni et al. 2009). When H1 Libyan sequences  
5 are taken into account, coalescence age estimates in Libya ( $14.7 \pm 4.4$  kya) are compatible  
6 with those found in Tunisia. The haplogroup H1 network can be found in supplementary  
7 Figure 1.

### 8 **North African maternal lineage landscape**

9 In order to have a general view of the maternal genetic landscape in North Africa, a  
10 Correspondence Analysis (CA) based on haplogroup frequencies was built (Fig. 2). The first  
11 dimension separates sub-Saharan and southern samples on one edge, characterized by L  
12 haplogroups (except for some L3e subgroups), and Saudi Arabians on the opposite edge,  
13 characterized by R0a and J lineages. The second dimension follows a longitudinal pattern,  
14 grouping the Saudi Arabian, the Egyptian and the Sudanese samples on one edge and the  
15 Moroccan and European ones at the opposite side. North African populations form a large  
16 cluster in the center of the chart, without a clear structure. Nevertheless, it is noticeable that  
17 Egyptian populations and Nubians are placed in one edge of the second dimension, whereas  
18 Maghreb and European populations are grouped in the opposite edge.

19 A series of Analyses of Molecular Variance (AMOVA) were performed in order to  
20 test the proportion of the genetic variance within and among samples in North Africa. When  
21 all North African populations were considered as a single group, 3.88% ( $p < 0.01$ ) of the  
22 genetic variance was attributed to differences among samples. Then, we aimed to test how the  
23 apportionment of the maternal genetic variance was distributed across North Africa when two  
24 groups were considered in a west-east axis. In order to perform this test, we divided the whole  
25 region along four sections that were roughly limited by the actual geopolitical boundaries in

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

the region (Fig. 1B). Next, we performed a series of AMOVA analyses: in each new analysis the border between the two groups was moved progressively eastwards. Results showed that the amount of genetic variation was maximal (1.09%,  $p$ -value < 0.01) when the Eastern group was defined only by Egyptian and Sudanese populations.

In order to assess which haplogroups might be responsible of the differences found in the AMOVAs between eastern North Africa (Egypt and Sudan) and western North Africa together with Libya, we performed a correlation analysis between haplogroup frequencies and the longitude coordinates of the populations in our dataset (Table 3). Some lineages have higher frequencies in the West and decrease significantly towards the East, such as Eurasian H and HV0 haplogroups, sub-Saharan L1b and L3b haplogroups, and the North African U6 haplogroup. On the contrary, some lineages are more frequent in eastern samples, such as L0a and Eurasian haplogroups R0a, N1b, I, and J lineages. Interestingly, M1 does not reach statistical significance ( $p=0.055$ ).

#### **Phylogeographic analysis of North African U6 and M1 lineages**

In order to deeply analyze the distribution and relationships of U6 and M1 lineages, a phylogeographic analysis was performed. Figure 3A shows the Median network of U6 haplotypes. As expected, the most represented groups are U6a\* and U6a1, both of which show starlike phylogenies. Interestingly, both subclades bear similar diversity and many of the derived nodes are unique lineages (found only once in the database). The Maghreb is largely represented in the U6a clade. Of note is that most of the non-Maghreb U6a sequences are indeed from Libya. Moreover, Libya bears many unique sequences placed in basal and intermediate nodes spread all over the network, showing a high level of variability, in contrast with more eastern samples that show little diversity since all their sequences but one bear the root motifs of haplogroups U6a and U6a1. This is consistent with the mtDNA pool from Libyans being genetically closer to the Maghreb than to the northeast. Previous studies have

1 shown that minor subclades U6b and U6c are restricted to local areas (Maca-Meyer et al.  
2 2003). The distribution of U6b was restricted to Morocco, Algeria and Eastern Bedouins;  
3 however, it has been found in Libya and Saudi Arabia (Abu-Amro et al. 2008) as well,  
4 extending its presence to nearly the entire North African area.

5 The coalescence time estimate for the U6 network (except for the U6c branch) is 44.0  
6  $\pm$  21.6 kya. Our coalescence age estimation based on the HVSI region for the haplogroup  
7 U6a1 is 13.0  $\pm$  5.7 kya, whereas for U6a\* is 13.5  $\pm$  3.7 kya.

8 In a similar way to U6, M1 network shows that the basal lineages of M1 and M1a1 are  
9 the most common and have a starlike phylogeny (Fig. 3B) as well. Unlike U6, this clade is  
10 mostly represented in Egypt, Sudan and Saudi Arabia, where most of the unique sequences  
11 are found. Moreover, it should be noted that all Maghreb sequences, with only one exception,  
12 belong to the M1 haplogroup. Haplogroup distribution, haplogroup diversity and the high  
13 number of terminal nodes suggest that M1a1 arose in North East Africa and subsequently  
14 spread westwards. Coalescence age estimation for M1a1 is 13.1  $\pm$  7.0 kya. When the entire  
15 M1 clade is considered, the time estimate increases to 23.1  $\pm$  9.2. Note that coalescence ages  
16 for all the nodes that show starlike shape -M1a1, U6a\*, and U6a1- are similar, which may be  
17 concordant with a common expansion.

## 18 DISCUSSION

19 The maternal lineage background in North Africa shows a moderate degree of East-  
20 West differentiation, with a genetic discontinuity between Libya and Egypt. This difference is  
21 summarized in the AMOVA that attributes 1.09% of the North African genetic variance to  
22 differences between Eastern and Western groups. Despite that other groupings within North  
23 Africa also yield significant differences in the AMOVAs, the differences found between  
24 Eastern and Western groups defined in the Libyan-Egyptian border are more than double

1 compared to the rest. Overall, the genetic structure within North Africa is the result of  
2 different haplogroup frequency distribution of L, U6, M1, and probably H lineages.

3 Besides L2a1, which is widespread in Africa, most sub-Saharan mtDNA haplogroups  
4 found in North Africa exhibit a slight east-west cline. L1b, L3b, and L3f1b lineages, which  
5 have a mainly western African distribution (Harich et al. 2010; Salas et al. 2002) are more  
6 frequent in NW African samples and rare in NE African populations. Harich and collaborators  
7 (2010) proposed that the origin of most of the sub-Saharan sequences found in North Africa  
8 can be found in the impact of the trans-Saharan slave trade routes that were established during  
9 recent times. This hypothesis could well explain the results found in Libya. According to  
10 trans-Saharan slave trade routes from Segal (2002), northern Libya was directly connected to  
11 western Africa with the Chad basin and was also interconnected with Tunisia, Algeria and  
12 Morocco, which were in turn connected with other western Africa locations. This would  
13 explain as well differences found in L haplogroups between our Libyan results and those  
14 found in Libyan Tuareg populations, where 18% of the L sequences are L0a1, a typical  
15 eastern African haplogroup (Ottoni et al. 2009). Libyan Tuaregs live in south-western Libya,  
16 along a trade route that interconnects this region with Egypt. Therefore, differences in sub-  
17 Saharan haplogroup distribution between these two Libyan samples could be due to gene flow  
18 either across the trade routes connecting North Africa and sub-Saharan Africa, or across  
19 North Africa itself. Indeed, the significant gradient of frequencies of haplogroups L1b and  
20 L3b shown with the correlation analysis agrees with this sub-Saharan genetic exchange within  
21 North Africa.

22 The distribution of haplogroups U6 and M1 also suggests the presence of a  
23 discontinuity between Libya and Egypt, separating western North Africa from eastern North  
24 Africa. Even if both haplogroups are thought to have been carried by a back-to-Africa  
25 migration from the Near East, significant increased U6 frequencies have been detected in the

1 West compared to the East. The network of all U6 sequences found in the database presents  
2 two nodes with star-like shape, U6a\* and U6a1. In a similar way, M1a1 is the node with star-  
3 like topology in haplogroup M1. Time estimates of these nodes are  $13.5 \pm 3.7$ ,  $13.0 \pm 5.7$  and  
4  $13.1 \pm 7.0$  kya for haplogroups U6a\*, U6a1, and M1a1 respectively. The most plausible  
5 explanation of the frequency distribution of M1 and U6 lineages, the coalescence age  
6 estimates, and the starlike shape would be an early split in the back to Africa migration  
7 followed by a period of stability and a period of expansion. The split would have produced  
8 two different migration waves, one westward, represented by U6, and the other southwards,  
9 represented by M1. Each haplogroup would have increased its frequency by drift and  
10 subsequently accumulated diversity over time. Coalescent time estimates point to an  
11 expansion of these haplogroups at the end of the LGM, simultaneously with some Eurasian  
12 haplogroups, as suggested by Olivieri et al. (2006). Moreover, all but one M1a1 sequence are  
13 found in eastern North Africa, which suggests that this subclade might have appeared in the  
14 East, and only after that have migrated westwards.

15 A similar East-West structure has been found with haplogroups related to the post-  
16 LGM expansion in the European Franco-Cantabrian area. A declining gradient of frequencies  
17 from west to east is detected for haplogroups H1 and H3. Moreover, the estimate age of  
18 haplogroup H1 agrees with previous estimates in North Africa, being  $14.7 \pm 4.4$ ,  $11.7 \pm 3.6$ ,  
19 and  $11.3 \pm 2.3$  ya for Libya (present study), Tunisia (Cherni et al. 2009), and North Africa  
20 (Ennafaa et al. 2009) respectively. These dates set an upper limit for the presence of H1 in  
21 North Africa, which in any case is unlikely to have entered the region before the LGM. They  
22 are compatible with the posited post-LGM expansion from the Franco-Cantabrian glacial  
23 refuge area, although subsequent introductions cannot be ruled out. Unfortunately, no data is  
24 available for haplogroup H subclades in Egypt. The dissection of Egyptian H lineages would  
25 help to discern whether H1 is ubiquitous along North Africa or if a clear genetic barrier exists

1  
2  
3 1 between Libya and Egypt. Moreover, it would also be possible to discern whether a similar  
4  
5 2 pattern has taken place from a post-LGM expansion in the Near Eastern refuge, considering  
6  
7  
8 3 that Libya has an increased frequency of typically Near Eastern haplogroups as H5, H7 and  
9  
10 4 H13 compared to western North Africa.

11  
12 5 The most plausible explanation for the differences found between NW and NE Africa  
13  
14 6 is the presence of a demographic corridor along the Nile Valley. This corridor might have  
15  
16 7 allowed the contact between Egypt, East Africa, and the Near East; leaving the rest of NW  
17  
18 8 Africa apart from this Eastern contacts. Later, the Arab movements tied to the expansion of  
19  
20 9 Islam did not apparently bridge the gap, at least for the female-transmitted mtDNA. The arid  
21  
22 10 conditions west of the Nile may have conditioned population movement throughout much of  
23  
24 11 human (pre)history, to the point of partially isolating the genetic pool of Egypt from those of  
25  
26 12 countries to its west, including Libya.

### 31 32 13 **ACKNOWLEDGEMENTS**

33  
34 14 We would like to thank Roger Anglada, Stephanie Plaza, and Mònica Vallés (UPF,  
35  
36 15 Barcelona, Spain) for technical support. We also would like to thank all the volunteers that  
37  
38 16 donated their DNA, making this study possible. The present research was supported by the  
39  
40 17 Dirección General de Investigación, Ministerio de Ciencia e Innovación, Spain (CGL2010-  
41  
42 18 14944/BOS); Tunisian Ministry of High Education and Scientific Research ; Programa de  
43  
44 19 Cooperación Interuniversitaria e Investigación Científica (AEIC), Ministerio de Asuntos  
45  
46 20 Exteriores y Cooperación, Spain (A/5180/06, A/8394/07, B/018514/08); and Direcció General  
47  
48 21 de Recerca, Generalitat de Catalunya (2009SGR1101).

## LITERATURE CITED

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Abu-Amero KK, Larruga JM, Cabrera VM, and Gonzalez AM. 2008. Mitochondrial DNA structure in the Arabian Peninsula. *BMC Evolutionary Biology* 8(1):45.

Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, Scozzari R, Cruciani F, Zeviani M, Briem E, Carelli V and others. 2004. The Molecular Dissection of mtDNA Haplogroup H Confirms That the Franco-Cantabrian Glacial Refuge Was a Major Source for the European Gene Pool. *The American Journal of Human Genetics* 75(5):910-918.

Allard M, Miller K, Wilson M, Monson K, and Budowle B. 2002. Characterization of the Caucasian haplogroups present in the SWGDAM forensic mtDNA dataset for 1771 human control region sequences. Scientific Working group on DNA analysis methods. *Journal of Forensic Sciences* 47:1215-1223.

Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F and others. 1981. Sequence and organisation of the human mitochondrial genome. *Nature* 290:457 - 465.

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

Arredi B, Poloni ES, Paracchini S, Zerjal T, Fathallah DM, Makrelouf M, Pascali VL, Novelletto A, and Tyler-Smith C. 2004. A predominantly neolithic origin for Y-chromosomal DNA variation in North Africa. *American Journal of Human Genetics* 75(2):338-345.



1  
2  
3  
4  
5 Bandelt HJ, Forster P, and Rohl A. 1999. Median-joining networks for inferring intraspecific  
6 phylogenies. *Molecular Biology and Evolution* 16:37 - 48.  
7  
8

9  
10  
11  
12 Barbuji G, Pilastro A, De Domenico S, and Renfrew C. 1994. Genetic variation in North  
13 Africa and Eurasia: neolithic demic diffusion vs. Paleolithic colonisation. *American Journal of*  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Physical Anthropology 95(2):137-154.

Behar DM, Vilems R, Soodyall H, Blue-Smith J, Pereira L, Metspalu E, Scozzari R, Makkan  
H, Tzur S, Comas D and others. 2008. The Dawn of Human Matrilineal Diversity. *The*  
*American Journal of Human Genetics* 82(5):1130-1140.

Bosch E, Calafell F, González-Neira A, Flaiz C, Mateu E, Scheil HG, Huckenbeck W,  
Efremovska L, Mikerezi I, Xirotiris N and others. 2006. Paternal and maternal lineages in the  
Balkans show a homogeneous landscape over linguistic barriers, except for the isolated  
Aromuns. *Annals of Human Genetics* 70(4):459-487.

Bosch E, Calafell F, Pérez-Lezaun A, Clarimón J, Comas D, Mateu E, Martínez-Arias R,  
Morera B, Brakez Z, Akhayat O and others. 2000. Genetic structure of north-west Africa  
revealed by STR analysis. *European Journal of Human Genetics* 8(5):360-366.

Bosch E, Calafell F, Pérez-Lezaun A, Comas D, Mateu E, and Bertranpetit J. 1997.  
Population history of north Africa: evidence from classical genetic markers. *Human Biology*  
69(3):295-311.

1  
2  
3 Bosch E, Clarimón J, Pérez-Lezaun A, and Calafell F. 2001. STR data for 21 loci in  
4  
5  
6 northwestern Africa. *Forensic Science International* 116(1):41-51.  
7

8  
9  
10 Comas D, Calafell F, Benchemsi N, Helal A, Lefranc G, Stoneking M, Batzer MA,  
11  
12 Bertranpetit J, and Sajantila A. 2000. Alu insertion polymorphisms in NW Africa and the  
13  
14 Iberian Peninsula: evidence for a strong genetic boundary through the Gibraltar Straits.  
15  
16  
17 *Human Genetics* 107(4):312-319.  
18

19  
20  
21  
22 Cruciani F, La Fratta R, Santolamazza P, Sellitto D, Pascone R, Moral P, Watson E, Guida V,  
23  
24 Colomb EB, Zaharova B and others. 2004. Phylogeographic analysis of haplogroup E3b (E-  
25  
26 M215) y chromosomes reveals multiple migratory events within and out of Africa. *American*  
27  
28 *Journal of Human Genetics* 74(5):1014-1022.  
29

30  
31  
32  
33  
34 Cruciani F, La Fratta R, Trombetta B, Santolamazza P, Sellitto D, Colomb EB, Dugoujon J,  
35  
36 Crivellaro F, Benincasa T, Pascone R and others. 2007. Tracing past human male movements  
37  
38 in northern/eastern Africa and western Eurasia: new clues from Y-chromosomal haplogroups  
39  
40 E-M78 and J-M12. *Molecular Biology and Evolution* 24(6):1300-1311.  
41  
42

43  
44  
45  
46 Cherni L, Fernandes V, Pereira JB, Costa MD, Goios A, Frigi S, Yacoubi-Loueslati B, Amor  
47  
48 MB, Slama A, Amorim A and others. 2009. Post-last glacial maximum expansion from Iberia  
49  
50 to North Africa revealed by fine characterization of mtDNA H haplogroup in Tunisia. *The*  
51  
52 *American Journal of Human Anthropology* 139(2):253-260.  
53  
54

55  
56  
57 Desanges J. 1990. *The proto-Berbers*. Mokhtar G, editor. Paris: Unesco.  
58  
59  
60

- 1  
2  
3 Ennafaa H, Cabrera V, Abu-Amero K, Gonzalez A, Amor M, Bouhaha R, Dzimiri N,  
4  
5 Elgaaied A, and Larruga J. 2009. Mitochondrial DNA haplogroup H structure in North  
6  
7 Africa. *BMC Genetics* 10(1):8.  
8  
9  
10  
11  
12 Excoffier L, Laval G, and Schneider S. 2005. Arlequin (version 3.0): An integrated software  
13  
14 package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1:47-50.  
15  
16  
17  
18  
19 Fadhlaoui-Zid K, Plaza S, Calafell F, Amor MB, Comas D, Bennamar A, and El g. 2004.  
20  
21 Mitochondrial DNA Heterogeneity in Tunisian Berbers. *Annals of Human Genetics*  
22  
23 68(3):222-233.  
24  
25  
26  
27  
28  
29 Finnilä S, Lehtonen MS, and Majamaa K. 2001. Phylogenetic network for European mtDNA.  
30  
31 *The American Journal of Human Genetics* 68:1475 - 1484.  
32  
33  
34  
35  
36 Flores C, Maca-Meyer N, González AM, and Cabrera VM. 2000. Northwest African  
37  
38 distribution of the CD4/Alu microsatellite haplotypes. *Annals of Human Genetics* 64(Pt  
39  
40 4):321-327.  
41  
42  
43  
44  
45 Frigi S, Ennafaa H, Amor MB, Cherni L, and Ammar-Elgaaied BA. 2010. Assessing human  
46  
47 genetic diversity in Tunisian Berber populations by Alu insertion polymorphisms. *Annals of*  
48  
49 *Human Biology*(epub).  
50  
51  
52  
53  
54  
55 Garcea EA, and Giraudi C. 2006. Late Quaternary human settlement patterning in the Jebel  
56  
57 Gharbi. *Journal of Human Evolution* 51:411-421.  
58  
59  
60

1  
2  
3 Garcia O, Fregel R, Larruga JM, Alvarez V, Yurrebaso I, Cabrera VM, and Gonzalez AM.  
4  
5 2010. Using mitochondrial DNA to test the hypothesis of a European post-glacial human  
6  
7 recolonization from the Franco-Cantabrian refuge. *Heredity*(epub).

8  
9  
10  
11  
12 González-Pérez E, Via M, Esteban E, López-Alomar A, Mazieres S, Harich N, Kandil M,  
13  
14 Dugoujon JM, and Moral P. 2003. Alu insertions in the Iberian Peninsula and north west  
15  
16 Africa -- genetic boundaries or melting pot? *Collegium Antropologicum* 27(2):491-500.

17  
18  
19  
20  
21  
22 Gonzalez AM, Larruga JM, Abu-Amero KK, Shi Y, Pestano J, and Cabrera VM. 2007.  
23  
24 Mitochondrial lineage M1 traces an early human backflow to Africa. *BMC Genomics* 8:223.

25  
26  
27  
28  
29 Graven L, Passarino G, Semino O, Boursot P, Santachiara-Benerecetti AS, Langaney A, and  
30  
31 Excoffier L. 1995. Evolutionary correlation between control region sequence and restriction  
32  
33 polymorphisms in the mitochondrial genome of large Senegalese Mandenka sample.  
34  
35  
36 *Molecular Biology and Evolution* 12(2):334-345.

37  
38  
39  
40  
41 Harich N, Costa MD, Fernandes V, Kandil M, Pereira JB, Silva NM, and Pereira L. 2010. The  
42  
43 trans-Saharan slave trade - clues from interpolation analyses and high resolution  
44  
45 characterization of mitochondrial DNA lineages. *BMC Evolutionary Biology* 10:138-156.

46  
47  
48  
49  
50 Khodjet-El-Khil H, Fadhlouzi-Zid K, Gusmão L, Alves C, Benammar-Elgaaied A, and  
51  
52 Amorim A. 2008. Substructure of a Tunisian Berber population as inferred from 15 autosomal  
53  
54 short tandem repeat loci. *Human Biology* 80(4):435-448.  
55  
56  
57  
58  
59  
60

1  
2  
3 Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, Pennarun E, Parik J, Geberhiwot T,  
4  
5 Usanga E, and VILLEMS R. 2004. Ethiopian Mitochondrial DNA Heritage: Tracking Gene  
6  
7 Flow Across and Around the Gate of Tears. *The American Journal of Human Genetics* 75:752  
8  
9 - 770.

10  
11  
12  
13  
14  
15 Krings M, Salem A-eH, Bauer K, Geisert H, Malek AK, Chaix L, Simon C, Welsby D, Di  
16  
17 Rienzo A, Utermann G and others. 1999. mtDNA Analysis of Nile River Valley Populations:  
18  
19 A Genetic Corridor or a Barrier to Migration? *The American Journal of Human Genetics*  
20  
21 64(4):1166-1176.  
22  
23  
24

25  
26  
27 Larruga JM, Diez F, Pinto F, Flores C, and Gonzalez AM. 2001. Mitochondrial DNA  
28  
29 characterization of European isolates: the Maragatos from Spain. *European Journal of Human*  
30  
31 *Genetics* 9(9):708-716.  
32  
33

34  
35  
36 Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh  
37  
38 GS, Feldman M, Cavalli-Sforza LL and others. 2008. Worldwide human relationships  
39  
40 inferred from genome-wide patterns of variation. *Science* 319(5866):1100-1104.  
41  
42  
43

44  
45  
46 Loogvali EL, Roostalu U, Malyarchuk BA, Derenko MV, Kivisild T, Metspalu E, Tambets K,  
47  
48 Reidla M, Tolk HV, Parik J and others. 2004. Disuniting uniformity: a pied cladistic canvas  
49  
50 of mtDNA haplogroup H in Eurasia. *Molecular Biology and Evolution* 21:2012 - 2021.  
51  
52

53  
54  
55 López-Soto M, and Sanz P. 2000. Polimorfismos de ADN Mitochondrial en individuos  
56  
57 residentes en Andalucía y Extremadura. *Cuadernos de Medicina Forense* 20:17-29.  
58  
59  
60

1  
2  
3 Maca-Meyer N, Gonzalez A, Pestano J, Flores C, Larruga J, and Cabrera V. 2003.

4  
5 Mitochondrial DNA transit between West Asia and North Africa inferred from U6  
6  
7  
8 phylogeography. *BMC Genetics* 4(1):15.  
9

10  
11  
12 Maca-Meyer N, Gonzalez AM, Larruga JM, Flores C, and Cabrera VM. 2001. Major genomic  
13  
14 mitochondrial lineages delineate early human expansions. *BMC Genetics* 2:13.  
15  
16  
17

18  
19  
20 McEvedy C, and Jones R. 1980. *Atlas of the world population history*. Harmondsworth, UK.  
21  
22  
23

24  
25 Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, Serk P, Karmin M,  
26  
27 Behar D, Gilbert MT and others. 2004. Most of the extant mtDNA boundaries in South and  
28  
29 Southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically  
30  
31 modern humans. *BMC Genetics* 5(1):26.  
32  
33  
34

35  
36 Newman JL. 1995. *The peopling of Africa: a geographic interpretation*. Press YU, editor:  
37  
38 New Haven and London.  
39  
40  
41

42  
43 Olivieri A, Achilli A, Pala M, Battaglia V, Fornarino S Al-Zahery N, Scozzari R, Cruciani F,  
44  
45 Behar DM, Dugoujon JM, Coudray C and others. 2006. The mtDNA legacy of the Levantine  
46  
47 early Upper Palaeolithic in Africa. *Science* 314:1767 - 1770.  
48  
49  
50

51  
52  
53 Ottoni C, Martínez-Labarga C, Loogvali EL, Pennarun E, Achilli A, De Angelis F, Trucchi E,  
54  
55 Contini I, Biondi G, and Rickards O. 2009. First genetic insight into Libyan Tuaregs: a  
56  
57 maternal perspective. *Annals of Human Genetics* 73(4):438-448.  
58  
59  
60

1  
2  
3 Palanichamy MG, Sun C, Agrawal S, Bandelt HJ, Kong QP, Khan F, Wang CY, Chaudhuri  
4  
5 TK, Palla V, and Zhang YP. 2004. Phylogeny of mitochondrial DNA macrohaplogroup N in  
6  
7 India, based on complete sequencing: implications for the peopling of South Asia. *The*  
8  
9 *American Journal of Human Genetics* 75:966 - 978.

10  
11  
12  
13  
14  
15 Pereira L, Richards M, Gozos A, Alonso A, Albarran C, Garcia O, Behar DM, Golge M,  
16  
17 Hatina J, Al-Gazali L and others. 2005. High-resolution mtDNA evidence for the late-glacial  
18  
19 resettlement of Europe from an Iberian refugium. *Genome Research* 15:19 - 24.

20  
21  
22  
23  
24 Plaza S, Calafell F, Helal A, Bouzerna N, Lefranc G, Bertranpetit J, and Comas D. 2003.  
25  
26  
27 Joining the Pillars of Hercules: mtDNA Sequences Show Multidirectional Gene Flow in the  
28  
29 Western Mediterranean. *Annals of Human Genetics* 67(4):312-328.

30  
31  
32  
33  
34 Plaza S, Salas A, Calafell F, Corte-Real F, Bertranpetit J, Carracedo Á, and Comas D. 2004.  
35  
36  
37 Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola. *Human*  
38  
39 *Genetics* 115(5):439-447.

40  
41  
42  
43 Rando JC, Pinto F, Gonzalez AM, Hernandez M, Larruga JM, Cabrera VM, and Bandelt HJ.  
44  
45  
46 1998. Mitochondrial DNA analysis of Northwest African populations reveals genetic  
47  
48 exchanges with European, Near-Eastern, and sub-Saharan populations. *Annals of Human*  
49  
50 *Genetics* 62:531 - 550.

51  
52  
53  
54  
55 Rhouda T, Martinez-Redondo D, Gomez-Duran A, Noureddine E, Idaomar M, Diez-Sanchez  
56  
57  
58 C, Montoya J, Lopez-Perez MJ, and Ruiz-Pesini E. 2009. Moroccan mitochondrial genetic  
59  
60

1  
2  
3 background suggests prehistoric human migrations across the Gibraltar Strait. Mitochondrion  
4  
5 9:402-407.  
6  
7

8  
9  
10 Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, Sellitto D,  
11  
12 Cruciani F, Kivisild T and others. 2000. Tracing European founder lineages in the Near  
13  
14 Eastern mtDNA pool. The American Journal of Human Genetics 67:1251 - 1276.  
15  
16  
17

18  
19  
20 Roostalu U, Kutuev I, Loogvali EL, Metspalu E, Tambets K, Reidla M, Khusnutdinova EK,  
21  
22 Usanga E, Kivisild T, and Villems R. 2006. Origin and expansion of haplogroup H, the  
23  
24 dominant human mitochondrial DNA lineage in West Eurasia: the Near Eastern and  
25  
26 Caucasian perspective. Molecular Biology and Evolution 24:436 - 448.  
27  
28  
29

30  
31 Saillard J, Forster P, Lynnerup N, Bandelt HJ, and Norby S. 2000. mtDNA variation among  
32  
33 Greenland eskimos: the edge of the Beringian expansion. The American Journal of Human  
34  
35 Genetics 67:718 - 726.  
36  
37  
38

39  
40  
41 Salas A, Richards M, De la Fe T, Lareu MV, Sobrino B, Sanchez-Diz P, Macaulay V, and  
42  
43 Carracedo A. 2002. The making of the African mtDNA landscape. The American Journal of  
44  
45 Human Genetics 71:1082 - 1111.  
46  
47  
48

49  
50  
51 Segal R. 2002. Islam's black slaves: a history of Afrca's other black diaspora. Books A, editor.  
52  
53 London.  
54  
55

56  
57  
58 Semino O, Magri C, Benuzzi G, Lin AA, Al-Zaheri N, Battaglia V, Maccioni L,  
59  
60 Triantaphyllidis C, Shen P, Oefner P and others. 2004. Origin, diffusion, and differentiation



1  
2  
3 of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later  
4 migratory events in the Mediterranean area. *American Journal of Human Genetics*  
5  
6 74(5):1023-1034.  
7  
8  
9

10  
11  
12 Soares P, Ermini L, Thomson N, Mormina M, Rito T, Röhl A, Salas A, Oppenheimer S,  
13 Macaulay V, and Richards MB. 2009. Correcting for Purifying Selection: An Improved  
14  
15 Human Mitochondrial Molecular Clock. *The American Journal of Human Genetics*  
16  
17 84(6):740-759.  
18  
19  
20  
21

22  
23  
24 Tishkoff SA, Reed FA, Friedlaender F, Ehret C, Ranciaro A, Froment A, Hirbo JB, Awomoyi  
25 AA, Bodo JM, Doumbo O and others. 2009. The genetic structure and history of Africans and  
26  
27 African Americans. *Science* 324(5930):1035-1044.  
28  
29  
30  
31

32  
33  
34 Torroni A, Bandelt HJ, D'Urbano L, Lahermo P, Moral P, Sellitto D, Rengo C, Forster P,  
35 Savontaus ML, Bonne-Tamir B and others. 1998. mtDNA analysis reveals a major late  
36  
37 Paleolithic population expansion from southwestern to northeastern Europe. *The American*  
38  
39 *Journal of Human Genetics* 62:1137 - 1152.  
40  
41  
42  
43

44  
45  
46 Torroni A, Bandelt HJ, Macaulay V, Richards M, Cruciani F, Rengo C, Martinez-Cabrera V,  
47  
48 VILLEMS R, Kivisild T, Metspalu E and others. 2001. A signal, from human mtDNA, of  
49  
50 postglacial recolonization in Europe. *The American Journal of Human Genetics* 69:844 - 852.  
51  
52  
53

54  
55 Vigilant L, Pennington R, Harpending H, Kocher T, and Wilson AC. 1989. Mitochondrial  
56  
57 DNA sequences in single hairs from a southern African population. *Proceedings of the*  
58  
59 *National Academy of Sciences of the United States of America* 86(23):9350-9354.  
60

1  
2  
3  
4  
5 Zimmermann B, Brandstätter A, Duftner N, Niederwieser D, Spiroski M, Arsov T, and  
6

7  
8 Parson W. 2007. Mitochondrial DNA control region population data from Macedonia.  
9

10 Forensic Science International: Genetics 1(3-4):e4-e  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8

## FIGURE LEGENDS

9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48

Fig.1. A. Map showing the location of the populations used in the present study. Boxes with numbers show the limits between sections used to divide the region. Populations are: **1** Moroccan Arab (MAR); **2** Moroccan Berber (MBE); **3** Figuig Berber (FIG); **4** Asni Berber (ASN); **5** Bouhria Berber (BOU); **6** Souss (SOU); **7** West Saharan (WSH); **8** Saharawi (SAH); **9** Mauritanian (MAU); **10** Algerian (ALG); **11** Mozabites (MZA); **12** Western Tuareg (WTUA); **13** Tunisian Urban (TUN\_URB); **14** Matmata Berber (TMA); **15** Sened Berber (TSE); **16** Chenini-Douriet Berber; **17** Kesra Berber (KES); **18** Zriba Arab (ZRI); **19** Skira Berber (SKI); **20** Tunisian Andalusian (TUN\_AND); **21** (DJE) Djerba; **22** Eastern Tuareg (ETUA); **23** Egyptian (EGY); **24** Upper Egypt (UPE); **25** Gurna (GUR); **26** Siwa (SIW); **27** Northern Nubian (NNUB); **28** Southern Nubian (SNUB); **33** Libya (LIB). B. Series of AMOVA results between and within groups including North African populations. Sample locations are represented in the map by dots. Five transects have been defined by the numbered white lines. Each analysis is represented by a row in the bottom of the Figure. When two groups are defined, the split is located in one of the barriers limiting two transects, and populations laying on the left represent the Western group and populations on the right represented the Eastern group.

49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Fig.2. Results of the Correspondence Analysis performed with the dataset used and three additional populations, the Sub-Saharan Mandenka and the Europeans Macedonian and Andalusians. Haplogroups are represented by dots and they have been colored according to its prevalence. Populations are represented by squares and have been colored according to the Transect where they belong. Population codes are the same as in Figure 1.

1  
2  
3 Fig.3. Median joining network of sequences present in the dataset that belong to A)  
4 haplogroup U6, B) haplogroup M1. In both images, each haplotype is represented by a circle  
5 and its dimension is proportional to the number of individuals that bear that haplotype.  
6 Haplogroups are located beside their most probable “root” haplotype and numbers separating  
7 haplotypes correspond to the positions of the HVSII region that change from one haplotype to  
8 the other (positions are under the form “position – 16000”). Small red dots represent  
9 reticulation.  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

TABLE 1. Diversity measures within mtDNA HVSI in North African samples.

Population	n	k	Seq. diversity	$\pi$	PD	Reference
Moroccan Arabs	50	44	$0.993 \pm 0.007$	$0.0195 \pm 0.0103$	$7.037 \pm 3.356$	Plaza et al. 2003
Moroccan Berbers	64	42	$0.968 \pm 0.013$	$0.0126 \pm 0.0069$	$4.521 \pm 2.251$	Plaza et al. 2003
Figuig	94	29	$0.937 \pm 0.014$	$0.0171 \pm 0.0091$	$6.173 \pm 2.958$	Coudray et al. 2009
Asni	53	36	$0.963 \pm 0.016$	$0.0151 \pm 0.0082$	$5.424 \pm 2.650$	Coudray et al. 2009
Bouhria	70	38	$0.964 \pm 0.011$	$0.0157 \pm 0.0084$	$5.661 \pm 2.744$	Coudray et al. 2009
Souss	50	34	$0.961 \pm 0.018$	$0.0128 \pm 0.0071$	$4.604 \pm 2.295$	Brakez et al. 2001
West Saharans	25	20	$0.973 \pm 0.022$	$0.0148 \pm 0.0082$	$5.340 \pm 2.658$	Rando et al. 1998
Saharai	56	41	$0.976 \pm 0.012$	$0.0151 \pm 0.0082$	$5.448 \pm 2.659$	Plaza et al. 2003
Mauritanian	64	43	$0.979 \pm 0.008$	$0.0178 \pm 0.0095$	$6.407 \pm 3.071$	Rando et al. 1998
Algerian	47	29	$0.965 \pm 0.012$	$0.0164 \pm 0.0088$	$5.894 \pm 2.861$	Plaza et al. 2003
Mozabites	85	30	$0.943 \pm 0.010$	$0.0134 \pm 0.0073$	$4.822 \pm 2.375$	Macaulay et al. 1999
Western tuareg	23	21	$0.992 \pm 0.015$	$0.0190 \pm 0.0103$	$6.838 \pm 3.330$	Watson et al. 1997
Tunisian Urban	98	83	$0.992 \pm 0.004$	$0.0172 \pm 0.0094$	$6.433 \pm 3.070$	Plaza et al. 2003, Cherni et al. 2009

1							
2							
3	Tunisian Berber Matmata	49	29	$0.946 \pm 0.021$	$0.0140 \pm 0.0077$	$5.050 \pm 2.490$	Fadahlaoui-Zid et al. 2004
4							
5	Tunisian Berber Sened	53	37	$0.975 \pm 0.011$	$0.0209 \pm 0.0110$	$7.527 \pm 3.565$	Fadahlaoui-Zid et al. 2004
6							
7	Tunisian Berber Chenini-Douiret	53	23	$0.939 \pm 0.017$	$0.0189 \pm 0.0100$	$6.823 \pm 3.259$	Fadahlaoui-Zid et al. 2004
8							
9	Zriba Arab	50	16	$0.904 \pm 0.022$	$0.0110 \pm 0.0062$	$3.948 \pm 2.008$	Cherni et al. 2005
10							
11	Kesra Berbers	47	20	$0.931 \pm 0.021$	$0.0174 \pm 0.0093$	$6.264 \pm 3.022$	Cherni et al. 2005
12							
13	Tunisian Andalusian	155	84	$0.965 \pm 0.010$	$0.0155 \pm 0.0083$	$5.581 \pm 2.693$	Cherni et al. 2009
14							
15	Skira Berbers	20	14	$0.937 \pm 0.043$	$0.0118 \pm 0.0068$	$4.237 \pm 2.185$	Cherni et al. 2009
16							
17	Djerba	59	43	$0.977 \pm 0.011$	$0.0153 \pm 0.0083$	$5.517 \pm 2.687$	Loueslati et al. 2006
18							
19	Eastern tuareg	129	20	$0.677 \pm 0.046$	$0.0115 \pm 0.0064$	$4.131 \pm 2.068$	Otoni et al. 2009
20							
21	Libyan	269	163	$0.988 \pm 0.003$	$0.0189 \pm 0.0099$	$6.746 \pm 3.189$	Present study
22							
23	Egyptian	344	232	$0.993 \pm 0.002$	$0.0190 \pm 0.0099$	$6.832 \pm 3.224$	Krings et al. 1999, Saunier et al. 2009
24							
25	Upper Egypt	24	24	$1.000 \pm 0.012$	$0.0234 \pm 0.0125$	$8.427 \pm 4.028$	Stevanovitch et al. 2004
26							
27	Gurna, Egypt	34	29	$0.989 \pm 0.010$	$0.0231 \pm 0.0122$	$8.331 \pm 3.947$	Stevanovitch et al. 2004
28							
29	Siwa	78	22	$0.914 \pm 0.014$	$0.0151 \pm 0.0081$	$5.436 \pm 2.644$	Coudray et al. 2009
30							
31	Egyptian Nubian	80	53	$0.977 \pm 0.008$	$0.0228 \pm 0.0118$	$8.203 \pm 3.840$	Krings et al. 1999
32							
33							
34							
35							
36							
37							
38							
39							
40							
41							
42							
43							
44							
45							
46							
47							

1  
2  
3 Sudanese Nubian 76 66  $0.995 \pm 0.003$   $0.0236 \pm 0.0122$   $8.482 \pm 3.963$  Krings et al. 1999  
4  
5

---

6 k stands for number of different sequences;  
7

8  $\pi$  stands for nucleotide diversity;  
9

10 PD stands for mean Number of pairwise differences.  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47

TABLE 2. Frequencies (%) of haplogroup H subclades within haplogroup H observed along the southern shore of the Mediterranean.

	Haplogroup H subclades								
	H*	H1	H2a	H3	H5	H6	H7	H8	H13
Mauritania									
Saharawi	18	64	-	9	-	-	9	-	-
Moroccan Arabs	14	68	-	7	.	11	-	-	-
Moroccan Berbers	14	63	-	17	6	-	-	-	-
Tunisian Berbers	47	34	4	4	2	4	2	-	-
Tunisian	28	24	-	21	3	-	17	3	-
Libyan	50	24	-	15	4	-	7	-	-
North Africa	24	37	2	11	9	2	7	-	9
Arabian Peninsula	32	43	1	13	2	2	5	<1	-
Near East	49	7	20	-	3	17	3	-	1
	61	14	1	-	10	4	4	1	4

<sup>1</sup> Data from Ennafaa et al. 2009

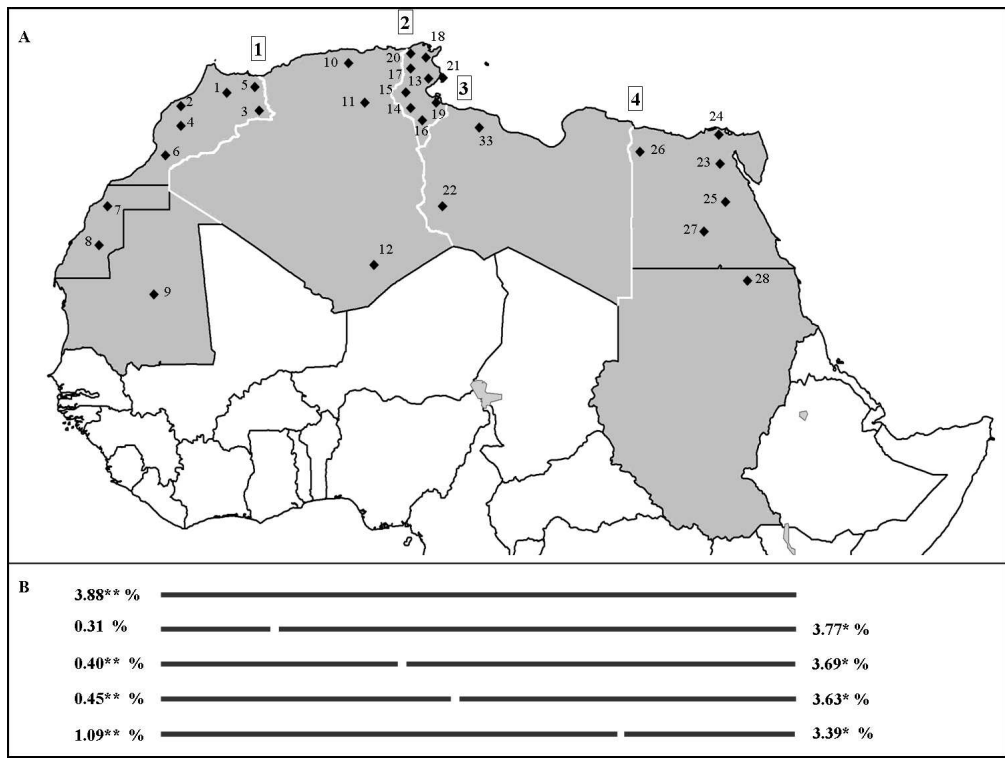


TABLE 3. Pearson correlation indexes and significance observed for the correlation between the longitudinal coordinate and the haplogroup frequencies for North African samples. Negative values represent higher frequencies in western samples, whereas positive values represent higher frequencies in Eastern samples.

	Pearson correlation		Pearson correlation
H	-0.508 (0.001)	L0a1	0.323 (0.048)
HV1	0.075 (0.656)	L1b	-0.569 (<0.001)
HV0	-0.420 (0.009)	L2a	-0.059 (0.727)
R0a	0.614 (<0.001)	L3b	-0.364 (0.025)
J	0.527 (0.001)	L3d	-0.097 (0.561)
T	0.123 (0.464)	L3e1	0.068 (0.687)
U3	0.181 (0.276)	L3e2	-0.193 (0.247)
U5	-0.296 (0.071)	L3e5	-0.271 (0.100)
K	-0.040 (0.814)	L3f	0.038 (0.820)
N1b	0.345 (0.034)	U6	-0.515 (0.001)
I	0.328 (0.044)	M1	0.314 (0.055)
X	-0.098 (0.557)		

Significance (2 tailed) is shown in brackets

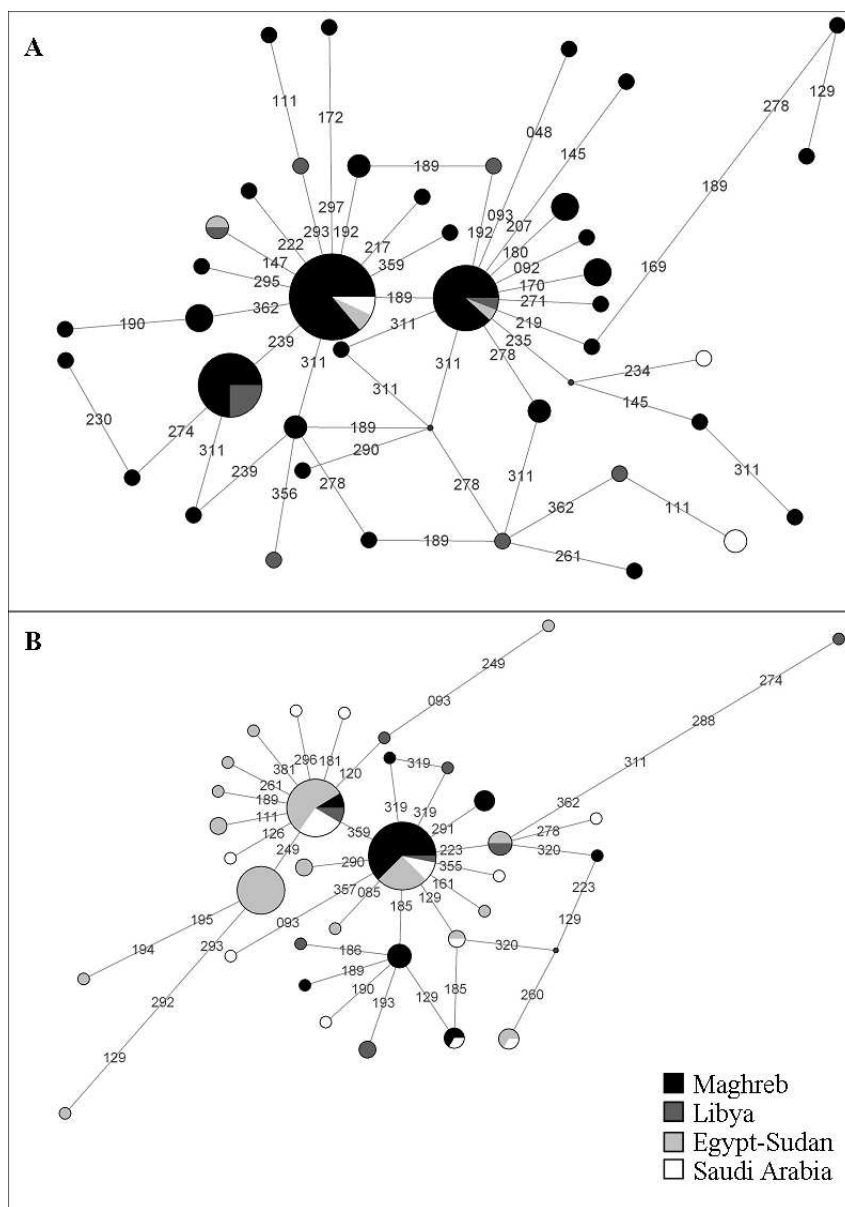
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



254x191mm (150 x 150 DPI)



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



134x189mm (150 x 150 DPI)