Mitochondrial DNA Structure in North Africa Reveals a Genetic Discontinuity in the Nile Valley

Karima Fadhlaoui-Zid, Laura Rodríguez-Botigué, Nejib Naoui, Amel Benammar-Elgaaied, Francesc Calafell, and David Comas

1Laboratory of Genetics, Immunology, and Human Pathology, Faculty of Science of Tunis, University Tunis El Manar, 2092 Tunis, Tunisia
2Institut de Biologia Evolutiva (CSIC-UPF), Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, 08003 Barcelona, Spain
3Laboratory of Medical Biology, Tunis, Tunisia
4CIBERESP, 08003 Barcelona, Spain

ABSTRACT

Human population movements in North Africa have been mostly restricted to 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. 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, N1b, I, J) 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.

KEY WORDS uniparental marker; haplotype; gene flow

North Africa is a region characterized by a complex history of demographic events and the extent of its genetic effect on extant human populations is still far from being known. Despite being part of the African continent, its demographic history, mostly limited to an east-west axis due to the barriers imposed by the Mediterranean Sea and the Sahara Desert, has been completely different from the rest of the continent. According to archaeological records, the first modern humans established in North Africa around 160,000 years ago (ya) (Smith et al., 2007). Human settlements dated to be between 80,000 and 40,000 ya are associated with the Aterian industry (Garcea and Giraudi, 2006), and those between 22,000 and 9,000 ya with the Iberomaurusian culture (Newman, 1995). So far, no clear connections have been established between these first human industries and those that succeeded them. The Ibero-Maurusian culture was followed by the Capsian industry (10,000–4,700 ya) (Desanges, 1990) that persisted well after the adoption of farming and agriculture, which began around 5,500 years ago in the region. The persistence of a pre-Neolithic culture in Neolithic times might indicate cultural replacement with admixture, rather than a population replacement of the autochthonous pre-Neolithic people by Neolithic farmers originated in the Middle East. In general terms, the prehistoric cultural changes in North Africa were quite independent of the dynamics on the European shores of the Mediterranean. Historical records document trade routes across the Sahara Desert and contacts between both Mediterranean shores and the Middle East, such as Phoenicians, Romans, Vandals, and Byzantines. The first Arab invasion, initially confined to Egypt, started in A.D. 643 and may have involved only a few thousand individuals (McEvedy and Jones, 1980). The Arabs began to impose their religion and language over the Berber

Additional Supporting Information may be found in the online version of this article.

Karima Fadhlaoui-Zid and Laura Rodríguez-Botigué are the first authors and contributed equally to this work.

Grant sponsor: Dirección General de Investigación, Ministerio de Ciencia e Innovación, Spain; Grant number: CGL2010-14944/BOS; Grant sponsor: Programa de Cooperação Interuniversitária e Investigación Científica (AECI), Ministerio de Asuntos Exteriores y Cooperación, Spain; Grant numbers: A/5180/06, A/8394/07, B/018514/08; Grant sponsor: Direcció General de Recerca, Generalitat de Catalunya; Grant number: 2009SGR1101; Grant sponsor: Tunisian Ministry of Higher Education and Scientific Research.

*Correspondence to: David Comas, Institut de Biologia Evolutiva (CSIC-UPF), Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, 08003 Barcelona, Spain.

E-mail: david.comas@upf.edu

Received 1 October 2010; accepted 16 November 2010

DOI 10.1002/ajpa.21472
Published online in Wiley Online Library (wileyonlinelibrary.com).
autochthonous population, a process that culminated with the second and more numerous Arab wave in which the Bedouins reached the Maghreb (northwest Africa) in the 11th century. The later arrivals to northern Africa in colonial times include Europeans and Ottoman Turks, mainly in Egypt.

Most of the African genetic diversity studies have been focused on the origin of our species and the first dispersals out of Africa (see for instance Tishkoff et al., 2009), processes in which North Africa had a marginal role, which made the region less attractive to human population geneticists. The analyses based on frequencies of classical genetic polymorphisms (blood groups, red cell enzymes, and serum proteins) have shown that the genetic landscape in northern Africa presents an east-west pattern of variation without differences between Arabs and Berbers, pointing to a sizeable Upper Paleolithic component in current northern African populations, whereas the Neolithic diffusion in the region was more a cultural than a demic process (Barbujani et al., 1994; Bosch et al., 1997). As for autosomal markers, only some STRs (Bosch et al., 2000; Cherni et al., 2005a; Coudray et al., 2007; Khodjet-El-Khil et al., 2008) and Alu polymorphisms (Comas et al., 2000; Flores et al., 2000; Gonzalez-Perez et al., 2003; Frigi et al., 2010) have been analyzed in a few northern African samples. Concerning massive analysis of genome-wide markers, only 30 Mozabite individuals (a Berber isolate from Algeria) have been analyzed for 650,000 SNPs (Li et al., 2008), showing various degrees of admixture between sub-Saharan, Middle Easterners, and Europeans. The analysis of Y chromosome lineages has shown a high frequency of two specific North African haplogroups (E1b1b1a and E1b1b1b), although their origins have been controversial since some analyses have suggested a Paleolithic component (Bosch et al., 2001), whereas others have pointed to a Neolithic origin (Arredi et al., 2004; Cruciani et al., 2004, 2007, 2010; Semino et al. 2004). The analysis of mitochondrial (mtDNA) lineages has shown that, in spite of an important sub-Saharan contribution, most haplogroups in North Africa are of Eurasian origin (Rando et al., 1998; Krings et al., 1999; Plaza et al., 2003; Fadhlaoui-Zid et al., 2004; Harich et al., 2010). Some can be traced to ancient Paleolithic times (such as haplogroups U6, M1b, which are almost specific of northern African populations); however, some maternal lineages have been recently acquired from Europe or the Middle East (such as haplogroups U5, V, R0a, J1b, U3) (Maca-Meyer et al., 2003; Olivieri et al., 2006; Gonzalez et al., 2007). Several studies suggest that at the end of the Last Glacial Maximum (LGM), around 13,000 ya, humans expanded from the Franco-Cantabrian refuge toward Europe and North Africa, spreading mtDNA haplogroups H1, H3, and V (Torroni et al., 1998, 2001; Achilli et al., 2004, 2005; Loogvai et al., 2004; Pereira et al., 2005; Cherni et al., 2005b; Ennafaa et al., 2009; Rhouda et al., 2009; Ottolini et al., 2010). However, a recent analysis of mtDNA diversity on Iberian populations points to the opposite conclusion: it suggests the absence of such an expansion (Garcia et al., 2010). In addition, a large degree of genetic heterogeneity has been shown in North African maternal lineages compared to other geographical regions such as Europe (Plaza et al., 2003; Fadhlaoui-Zid et al., 2004).

One of the main limitations of the genetic analyses of North African populations is the lack of representative and homogeneously distributed samples. For instance, most of the studies have focused on the north western samples and Egypt, being Libya a region with almost no genetic data with the exception of a Tuareg sample (Ottoni et al., 2009). The presence of gaps in the coverage of genetic studies across North Africa creates an artificial division between Eastern (Egypt) and Western populations (Morocco, Algeria, and Tunisia) and prevents explicitly geography-based analyses. Because of this lack of data, most of population genetic research in the area has a local scope rather than being comprehensive and covering the whole region.

Within the present analysis we aim to address several questions concerning the population history of North Africa. Is there any genetic structure that differentiates North African populations? What is the influence of the trade routes and natural corridors such as the Nile? Conversely, have the particularly inhospitable conditions of the Western Egyptian and Libyan deserts created a genetic barrier? These questions have not been successfully answered yet, partly because of the ~2,000 km gap in sampling between Egypt and Tunisia. In the present work we analyze for the first time the mtDNA sequences of a set of 269 Libyans representing the general population of the country that will allow us filling this gap and study the whole region.

MATERIALS AND METHODS

Mitochondrial DNA sequencing and SNP genotyping in Libyan individuals

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

Four TaqMan® probes (Applied Biosystems) were used, following supplier’s recommendations, to genotype positions 3594, 10873, 12705, and 14783, diagnostic for major lineages L3/J, N, R, and M, respectively. After this first classification, haplogroup assignment for 115 subjects was further refined by genotyping eight SNPs in the mtDNA coding region (7028, 10400, 10873, 11251 11719, 12308, and 12705 diagnostic for haplogroups H, M, N, J/ T, R0, U, and R respectively) by means of a SNaphet™ Multiplex kit (Applied Biosystems), as described previously (Bosch et al., 2006). Two of them (10873 and 12705) were typed with both methods and were used as controls. Finally, a dissection of haplogroup H was carried out using an additional SNaphet™ multiplex reaction: positions 3010, 4793, 4336, 6776, and 14872 were typed to classify individuals into subhaplogroups H1, H7, H5a, H3, and H13, respectively. Sequences of primers used for PCR amplification and primers used for genotyping can be found in Supporting Information Table 1.

Samples were assigned to haplogroups with the joint information of the control region sequence and the SNPs in the coding region following the nomenclature previously described (Richards et al., 2000; Finnilä et al.,...
TABLE 1. Diversity measures within mtDNA HVSI in North African samples

<table>
<thead>
<tr>
<th>Population</th>
<th>n</th>
<th>K</th>
<th>Seq. diversity</th>
<th>n-s</th>
<th>PD</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moroccan Arabs</td>
<td>50</td>
<td>44</td>
<td>0.993 ± 0.007</td>
<td>0.0195 ± 0.0103</td>
<td>7.037 ± 3.356</td>
<td>Plaza et al., 2003</td>
</tr>
<tr>
<td>Moroccan Berbers</td>
<td>64</td>
<td>42</td>
<td>0.968 ± 0.013</td>
<td>0.0126 ± 0.0069</td>
<td>4.521 ± 2.251</td>
<td>Plaza et al., 2003</td>
</tr>
<tr>
<td>Figuig</td>
<td>94</td>
<td>29</td>
<td>0.937 ± 0.014</td>
<td>0.0171 ± 0.0091</td>
<td>6.173 ± 2.958</td>
<td>Coudray et al., 2009</td>
</tr>
<tr>
<td>Asni</td>
<td>53</td>
<td>36</td>
<td>0.963 ± 0.016</td>
<td>0.0151 ± 0.0082</td>
<td>5.424 ± 2.650</td>
<td>Coudray et al., 2009</td>
</tr>
<tr>
<td>Bouhria</td>
<td>70</td>
<td>38</td>
<td>0.964 ± 0.011</td>
<td>0.0157 ± 0.0084</td>
<td>5.661 ± 2.744</td>
<td>Coudray et al., 2009</td>
</tr>
<tr>
<td>Souss</td>
<td>50</td>
<td>34</td>
<td>0.961 ± 0.018</td>
<td>0.0128 ± 0.0071</td>
<td>4.604 ± 2.295</td>
<td>Brakez et al., 2001</td>
</tr>
<tr>
<td>West Saharanas</td>
<td>25</td>
<td>20</td>
<td>0.973 ± 0.022</td>
<td>0.0148 ± 0.0082</td>
<td>5.340 ± 2.668</td>
<td>Rando et al., 1998</td>
</tr>
<tr>
<td>Saharaoui</td>
<td>56</td>
<td>41</td>
<td>0.976 ± 0.012</td>
<td>0.0151 ± 0.0082</td>
<td>5.448 ± 2.659</td>
<td>Plaza et al., 2003</td>
</tr>
<tr>
<td>Mauritania</td>
<td>64</td>
<td>43</td>
<td>0.979 ± 0.008</td>
<td>0.0178 ± 0.0095</td>
<td>6.407 ± 3.071</td>
<td>Rando et al., 1998</td>
</tr>
<tr>
<td>Algerian</td>
<td>47</td>
<td>29</td>
<td>0.965 ± 0.012</td>
<td>0.0164 ± 0.0088</td>
<td>5.894 ± 2.861</td>
<td>Plaza et al., 2003</td>
</tr>
<tr>
<td>Mozabites</td>
<td>85</td>
<td>30</td>
<td>0.943 ± 0.010</td>
<td>0.0134 ± 0.0073</td>
<td>4.822 ± 2.375</td>
<td>Macaulay et al., 1999</td>
</tr>
<tr>
<td>Western tuareg</td>
<td>23</td>
<td>21</td>
<td>0.992 ± 0.015</td>
<td>0.0190 ± 0.0103</td>
<td>6.838 ± 3.330</td>
<td>Watson et al., 1997</td>
</tr>
<tr>
<td>Tunisian Urban</td>
<td>98</td>
<td>83</td>
<td>0.992 ± 0.004</td>
<td>0.0172 ± 0.0094</td>
<td>6.433 ± 3.070</td>
<td>Plaza et al., 2003, Cherni et al., 2009</td>
</tr>
<tr>
<td>Tunisian Berber Matmata</td>
<td>49</td>
<td>29</td>
<td>0.946 ± 0.021</td>
<td>0.0140 ± 0.0077</td>
<td>5.050 ± 2.490</td>
<td>Fadhloua-Zid et al., 2004</td>
</tr>
<tr>
<td>Tunisian Berber Sened</td>
<td>53</td>
<td>37</td>
<td>0.975 ± 0.011</td>
<td>0.0209 ± 0.0110</td>
<td>7.527 ± 3.565</td>
<td>Fadhloua-Zid et al., 2004</td>
</tr>
<tr>
<td>Tunisian Berber Chenini-Douiret</td>
<td>53</td>
<td>23</td>
<td>0.939 ± 0.017</td>
<td>0.0189 ± 0.0100</td>
<td>6.823 ± 3.259</td>
<td>Fadhloua-Zid et al., 2004</td>
</tr>
<tr>
<td>Zriba Arab</td>
<td>50</td>
<td>16</td>
<td>0.904 ± 0.022</td>
<td>0.0110 ± 0.0062</td>
<td>3.948 ± 2.908</td>
<td>Cherni et al., 2005</td>
</tr>
<tr>
<td>Kesra Berbers</td>
<td>47</td>
<td>20</td>
<td>0.931 ± 0.021</td>
<td>0.0174 ± 0.0093</td>
<td>6.264 ± 3.022</td>
<td>Cherni et al., 2005</td>
</tr>
<tr>
<td>Tunisian Andalusian</td>
<td>155</td>
<td>84</td>
<td>0.965 ± 0.010</td>
<td>0.0155 ± 0.0083</td>
<td>5.581 ± 2.693</td>
<td>Cherni et al., 2009</td>
</tr>
<tr>
<td>Skira Berbers</td>
<td>20</td>
<td>14</td>
<td>0.937 ± 0.043</td>
<td>0.0118 ± 0.0068</td>
<td>4.237 ± 2.185</td>
<td>Cherni et al., 2009</td>
</tr>
<tr>
<td>Djerba</td>
<td>59</td>
<td>43</td>
<td>0.977 ± 0.011</td>
<td>0.0153 ± 0.0083</td>
<td>5.517 ± 2.687</td>
<td>Loueslati et al., 2006</td>
</tr>
<tr>
<td>Eastern tuareg</td>
<td>129</td>
<td>20</td>
<td>0.677 ± 0.046</td>
<td>0.0115 ± 0.0064</td>
<td>4.131 ± 2.068</td>
<td>Ottoni et al., 2009</td>
</tr>
<tr>
<td>Libyan</td>
<td>269</td>
<td>162</td>
<td>0.988 ± 0.003</td>
<td>0.0189 ± 0.0099</td>
<td>6.746 ± 2.189</td>
<td>Present study</td>
</tr>
<tr>
<td>Egyptian</td>
<td>254</td>
<td>232</td>
<td>0.993 ± 0.002</td>
<td>0.0190 ± 0.0099</td>
<td>6.832 ± 2.224</td>
<td>Krings et al., 1999, Saunier et al., 2009</td>
</tr>
<tr>
<td>Upper Egypt</td>
<td>24</td>
<td>24</td>
<td>1.000 ± 0.012</td>
<td>0.0234 ± 0.0125</td>
<td>8.427 ± 4.028</td>
<td>Stevanovitch et al., 2004</td>
</tr>
<tr>
<td>Gurna, Egypt</td>
<td>34</td>
<td>29</td>
<td>0.989 ± 0.010</td>
<td>0.0231 ± 0.0122</td>
<td>8.331 ± 3.947</td>
<td>Stevanovitch et al., 2004</td>
</tr>
<tr>
<td>Siwa</td>
<td>78</td>
<td>72</td>
<td>0.914 ± 0.014</td>
<td>0.0151 ± 0.0081</td>
<td>5.436 ± 2.644</td>
<td>Coudray et al., 2009</td>
</tr>
<tr>
<td>Egyptian Nubian</td>
<td>80</td>
<td>53</td>
<td>0.977 ± 0.008</td>
<td>0.0228 ± 0.0118</td>
<td>8.203 ± 3.840</td>
<td>Krings et al., 1999</td>
</tr>
<tr>
<td>Sudanese Nubian</td>
<td>76</td>
<td>66</td>
<td>0.995 ± 0.003</td>
<td>0.0236 ± 0.0122</td>
<td>8.482 ± 3.963</td>
<td>Krings et al., 1999</td>
</tr>
</tbody>
</table>

\[ a \] Number of different sequences.  
\[ b \] Nucleotide diversity.  
\[ c \] Mean number of pairwise differences.

Statistical and phylogenetic analyses

Previously published data of HVSI sequences ranging from positions 16024 to 16383 of 28 North African populations was used for population comparisons. Their names and references can be found in Table 1. For the purposes of the present analysis, and given our focus on Libya, we define eastern North Africa as Egypt and Northern Sudan (Nubia), and western North Africa as Morocco, Western Sahara, Mauritania, Algeria, and Tunisia. Tunisian samples were 13 out of the 21 original populations. The geographical pattern of haplogroup distribution was investigated by computing Pearson's correlation coefficients (using SPSS 15.0 software, SPSS, Chicago, IL) between the frequencies of each haplogroup and the geographical longitude of each population sample. Given the linear, East-West disposition of human settlement in

American Journal of Physical Anthropology
N. Africa, longitude captures most of the geographical distance between populations.

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

RESULTS
Mitochondrial DNA sequence diversity and haplogroup composition in Libya

A total of 164 different sequences were found in the analysis of both HVSI and HVSII segments in 269 Libyan individuals. Sequence diversity in Libyans is similar to other North African samples (Table 1), and when comparing only HVSI regions, 78 (29%) Libyan sequences were not found in the dataset used. Sequences and haplogroup frequencies of the Libyan population are shown in Supporting Information Table 2. Lineages of west Eurasian origin are the most frequent in Libyans (65%), followed by sub-Saharan L lineages (28%), and back to Africa haplogroups U6 and M1, which have an overall frequency of 7%.

Fig. 1. A. Map showing the location of the North African 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 Figueig Berber (FIG); 4 Asni Berber (ASN); 5 Bouthia Berber (BOU); 6 Sousi (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 Chemini-Douriet Berber; 17 Kesra Berber (KES); 18 Zriba Arab (ZRI); 19 Skira Berber (SKI); 20 Tunisian Andalusian (TUN_AND); 21 Djebel; 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 raw in the bottom of the Figure. When two groups are defined, the split is located in one of the barriers limiting two sections, and populations laying on the left represent the Western group and populations on the right represented the Eastern group. Percentage of variation between groups and within groups is shown at the left and right sides of the figure respectively. * Level of significance below 5%; ** Level of significance below 1%.
Haplogroups L2a1, L3f1b, L3b, and L1b are the most frequent (over 3%) sub-Saharan haplogroups in Libyans. Haplogroup L2a1 is common and apparently scattered throughout Africa (Salas et al., 2002), and therefore its geographical origin is difficult to assess. However, L1b, L3b, and L3f1b have more restricted locations in Africa. These haplogroups, together with other minor lineages in Libya such as L2b, L2c, L3d, and L3e have a typical Western Africa distribution (Salas et al., 2002; Harich et al., 2010). Nonetheless, other minor lineages present in Libya such as L0a, L3h, and L3x are more frequent in Eastern Africa. Such haplogroup frequency distribution suggests a predominantly Western origin of L lineages in Libya with some minor admixture of Eastern lineages. The back to Africa U6 lineage is mainly present in North Africa and shows an opposite frequency gradient respect to M1, being U6 significantly more frequent in the West, whereas M1 is more frequent in the East. Interestingly, these haplogroups display similar frequencies in the Libyan mtDNA pool (4.1% for U6 and 3.3% for M1).

Eurasian haplogroups HV0, H1, and K are the most frequent in Libyans (7.4, 6.3, and 5.2%, respectively). To trace back the geographical origin of the H lineages in Libya, we dissected haplogroup H in several subclades (see Materials and Methods). Compared with previous published data (Ennafa et al., 2009), Libyan individuals exhibit an admixture of western and eastern H subclades (Table 2). As in North West African populations, H1 and H3 are the most frequent subclades, and account for 48% of the H lineages. However, these frequencies are lower than those found in Maghreb populations because of the relative high proportion of H5, H7, and H13 subgroups in Libya, which are more frequent in the Near East (Roostalu et al., 2006).

The age estimation of haplogroup H1 based on the HV0 lineage in Europe is 16.0 kya (Pereira et al., 2005), 11.7 kya in Tunisia (Cherni et al., 2009), and between 4.4 and 11.5 kya in Libyan Tuaregs (Ottoni et al., 2010). When H1 Libyan sequences are taken into account, coalescence age estimates in Libya (14.7 ± 4.4 kya) are compatible with those found in Tunisia and in Tuareg. The haplogroup H1 network can be found in Supporting Information Figure 1.

North African maternal lineage landscape

To have a general view of the maternal genetic landscape in North Africa, a correspondence analysis (CA) based on haplogroup frequencies was built (see Fig. 2). The first dimension separates southern and northern populations: Mandenka, Sudanese, southern Nubian, Mauritanian, and western Tuaregs lie on one edge, characterized by L haplogroups (except for some L3e subgroups), and Middle Easterners and Europeans are plotted on the opposite edge, characterized by most of the Eurasian lineages. The second dimension follows a longitudinal pattern, grouping the Saudi Arabian, most of the Egyptian and the Sudanese samples on one edge and the Moroccan and European ones at the opposite side. North African populations form a large cluster in the center of the chart, without a clear structure. Nevertheless, it is noticeable that Egyptian populations and Nubians are placed in one edge of the second dimension, whereas Maghreb and European populations are grouped in the opposite edge.

A series of Analyses of Molecular Variance (AMOVA) were performed to test the proportion of the genetic variance within and among samples in North Africa. When all North African populations were considered as a single group, 3.88% (P < 0.01) of the genetic variance was attributed to differences among samples. Then, we aimed to test how the apportionment of the maternal genetic variance was distributed across North Africa when two groups were considered in a west-east axis. To perform this test, we divided the whole region along four sections that were roughly limited by the actual geopolitical boundaries in 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 when the Eastern group was defined only by Egyptian and Sudanese populations. Values and levels of significance are shown in Figure 1B. Still, at least 3.39% of the variation is explained by differences among populations within groups, stressing the heterogeneity of the North African region.

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 toward 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

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 star-like phylogenies. Interestingly, both subclades bear high diversity (being the haplotype diversity estimates 0.76 ± 0.073 and 0.87 ± 0.029 for U6a and U6a1, respectively) 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 shown that minor subclades U6b and U6c are restricted to local areas (Maca-Meyer et al., 2003). The distribution of U6b was restricted to Morocco, Algeria, and Eastern Bedouins; however, it has been found in Libya and Saudi Arabia (Abu-Amero et al. 2008) as well, extending its presence to nearly the entire North African area.

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

In a similar way to U6, M1 network shows that the basal lineages of M1 (including M1b) and M1a1 are the most common and have a star-like phylogeny (Fig. 3B) as well. Unlike U6, this clade is mostly represented in
TABLE 3. Pearson correlation indexes and significance observed for the correlation between the longitudinal coordinate and the haplogroup frequencies for North African samples

<table>
<thead>
<tr>
<th>Pearson correlation</th>
<th>Pearson correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>-0.508 (0.001)</td>
</tr>
<tr>
<td>HV1</td>
<td>0.075 (0.656)</td>
</tr>
<tr>
<td>HV0</td>
<td>-0.420 (0.009)</td>
</tr>
<tr>
<td>R0a</td>
<td>0.614 (&lt;0.001)</td>
</tr>
<tr>
<td>J</td>
<td>0.527 (0.001)</td>
</tr>
<tr>
<td>T</td>
<td>0.123 (0.464)</td>
</tr>
<tr>
<td>U3</td>
<td>0.181 (0.276)</td>
</tr>
<tr>
<td>U5</td>
<td>0.296 (0.071)</td>
</tr>
<tr>
<td>K</td>
<td>-0.040 (0.814)</td>
</tr>
<tr>
<td>N1b</td>
<td>0.345 (0.034)</td>
</tr>
<tr>
<td>I</td>
<td>0.328 (0.044)</td>
</tr>
<tr>
<td>X</td>
<td>-0.098 (0.557)</td>
</tr>
</tbody>
</table>

a Level of significance (two tailed) is 5% and is shown in brackets.
b Negative values represent higher frequencies in western samples, whereas positive values represent higher frequencies in Eastern samples.

mtDNA STRUCTURE IN NORTH AFRICA

routes from Segal (2002), northern Libya was directly connected to western Africa with the Chad basin and was also interconnected with Tunisia, Algeria, and Morocco, which were in turn connected with other western Africa locations. This would explain as well differences found in L haplogroups between our Libyan results and those found in Libyan Tuareg populations, where 18% of the L sequences are L0a1, a typical eastern African haplogroup (Ottoni et al., 2009). Libyan Tuaregs live in south-western Libya, along a trade route that interconnected this region with Egypt. Therefore, differences in sub-Saharan haplogroup distribution between these two Libyan samples could be due to gene flow either across the trade routes connecting North Africa and sub-Saharan Africa, or across North Africa itself. Indeed, the significant gradient of frequencies of haplogroups L1b and L3b shown with the correlation analysis agrees with this sub-Saharan genetic exchange within North Africa. However, a differentiation of mitochondrial DNA lineages in the Libyan Tuaregs could also be explained by drift given the level of isolation of this population.

The distribution of subsets of haplogroups U6 and M1 also suggests the presence of a discontinuity between Libya and Egypt, separating western North Africa from eastern North Africa. Even if both haplogroups are thought to have been carried by a back-to-Africa migration from the Near East, significant increased U6 frequencies have been detected in the West compared to the East. The network of all U6 sequences found in the database presents two nodes with star-like shape, U6a* and U6a1. In a similar way, M1a1 is the node with star-like topology in haplogroup M1, and the node where most of the eastern sequences are found. Time estimates of these nodes are 13.5 ± 3.7, 13.0 ± 5.7, and 13.1 ± 7.0 kya for haplogroups U6a*, U6a1, and M1a1 respectively. The most plausible explanation of the frequency distribution of M1a, U6, and M1b1 lineages, their coalescence age estimates, and the star-like shape would be an early split in the back to Africa migration followed by a period of stability and a period of expansion. The split would have produced two different migration waves, one westward, represented by U6 and possibly M1b1 in lower frequencies, and the other southward, represented by M1a. Each haplogroup would have increased its frequency by drift and subsequently accumulated diversity over time. Coalescent time estimates point to a possible second expansion of these haplogroups at the end of the LGM, simultaneously with some Eurasian haplogroups, as suggested by Olivieri et al. (2006). Moreover, all but one M1a1 sequence are found in eastern North Africa, which suggests that this subclade might have appeared in the East, and only after that have migrated westwards at this period.

A similar East-West structure has been found with haplogroups related to the post-LGM expansion in the European Franco-Cantabrian area. A declining gradient of frequencies from west to east is detected for haplogroups H1 and H3. Moreover, the estimate age of haplogroup H1 agrees with previous estimates in North Africa, being 14.7 ± 4.4, 11.7 ± 3.6, and 11.3 ± 2.3 Kya for Libya (present study), Tunisia (Cherni et al., 2009), and North Africa (Ennaaia et al., 2009), respectively. A recent work published by Ottoni et al. (2010) based on a Tuareg Libyan sample from Fezzan estimates the age of H1 between 4.4 and 11.5 kya. This younger age estimate could be due to the fact that the Fezzan population is extremely homogeneous and has a very low haplotype

DISCUSSION

The maternal lineage background in North Africa shows a moderate degree of East-West differentiation, with a genetic discontinuity between Libya and Egypt. This difference is summarized in the AMOVA that attributes 1.09% of the North African genetic variance to differences between Eastern and Western groups. Despite that other groupings within North Africa also yield significant differences in the AMOVAs, the differences found between Eastern and Western groups defined in the Libyan-Egyptian border are more than double compared to the rest. Overall, the genetic structure within North Africa is the result of different haplogroup frequency distribution of L, U6, and probably H lineages.

Besides L2a1, which is widespread in Africa, most sub-Saharan mtDNA haplogroups found in North Africa exhibit a slight east-westcline. The L1b, L3b, and L3f lineages, which have a mainly western African distribution (Salas et al., 2002; Harich et al., 2010) are more frequent in NW African samples and rare in NE African populations. Harich et al. (2010) proposed that the origin of most of the sub-Saharan sequences found in North Africa can be found in the impact of the trans-Saharan slave trade routes that were established during recent times. This hypothesis could well explain the results found in Libya. According to trans-Saharan slave trade

American Journal of Physical Anthropology
diversity (Ottoni et al., 2009) These dates set an upper limit for the presence of H1 in North Africa, which in any case is unlikely to have entered the region before the LGM. They are compatible with the posited post-LGM expansion from the Franco-Cantabrian glacial refuge area, although subsequent introductions cannot be ruled out. Unfortunately, no data is available for haplogroup H subclades in Egypt. The dissection of Egyptian H lineages would help to discern whether H1 is ubiquitous along North Africa or if a clear genetic barrier exists between Libya and Egypt. Moreover, it would also be possible to discern whether a similar pattern has

Fig. 3. Median joining network of sequences present in the dataset that belong to A) haplogroup U6, B) haplogroup M1. In both images, each haplotype is represented by a circle and its dimension is proportional to the number of individuals that bear that haplotype. Haplogroups are located beside their most probable “root” haplotype and numbers separating haplotypes correspond to the positions of the HVSI region that change from one haplotype to the other (positions are under the form “position-16000”). Small dots represent reticulation.
taken place from a post-LGM expansion in the Near Eastern refuge, considering that Libya has an increased frequency of typically Near Eastern haplogroups as H5, H7, and H13 compared to western North Africa.

The most plausible explanation for the differences found between NW and NE Africa is the presence of a demographic corridor along the Nile Valley. This corridor might have allowed the contact between Egypt, East Africa, and the Near East; influencing only slightly the rest of NW Africa. Later, the Arab movements tied to the expansion of Islam did not apparently bridge the gap, at least for the female-transmitted mtDNA. In a similar way, the post-LGM expansion originated in the Egyptian Peninsula has contributed to genetically differentiate North African populations, displaying a gradient of frequencies of the LGM-associated haplotypes, though its influence in the point of differentiating the genetic pool of Egypt from those of countries to its west, including Libya.

ACKNOWLEDGMENTS

The authors thank Roger Anglada, Stephanie Plaza, and Mónica Valles (UPF, Barcelona, Spain) for technical support. They also thank all the volunteers that donated their DNA, making this study possible.

LITERATURE CITED


American Journal of Physical Anthropology