Mitochondrial DNA haplogroup H structure in North Africa

Hajer Ennafaa†1, Vicente M Cabrera†2, Khaled K Abu-Amero3, Ana M González2, Mohamed B Amor1, Rym Bouhaha1, Nduna Dzimiri4, Amel B Elgaaïed1 and José M Larruga*2

Address: 1Laboratory of Genetics, Immunology and Human Pathology at the Faculty of Sciences of Tunis, Faculty of Sciences of Tunis, University El Manar I, 1060 Tunis, Tunisia, 2Department of Genetics, Faculty of Biology, University of La Laguna, Tenerife 38271, Spain, 3Molecular Genetics Laboratory, College of Medicine, King Saud University, Riyadh, Saudi Arabia and 4Department of Genetics, King Faisal Specialist Hospital, Riyadh, Saudi Arabia

Email: Hajer Ennafaa - ennafaahajer@yahoo.fr; Vicente M Cabrera - vcabrera@ull.es; Khaled K Abu-Amero - abuamero@gmail.com; Ana M González - amglez@ull.es; Mohamed B Amor - benamor_mohamed2000@yahoo.fr; Rym Bouhaha - rymab12@yahoo.fr; Nduna Dzimiri - dzimiri@kfshrc.edu.sa; Amel B Elgaaïed - a_benammarelgaaied@yahoo.fr; José M Larruga* - jlarruga@ull.es

* Corresponding author †Equal contributors

Abstract

Background: The Strait of Gibraltar separating the Iberian Peninsula from North Africa is thought to be a stronger barrier to gene flow for male than for female lineages. However, the recent subdivision of the haplogroup H at mitochondrial DNA (mtDNA) level has revealed greater genetic differentiation among geographic regions than previously detected. The dissection of the mtDNA haplogroup H in North Africa, and its comparison with the Iberian Peninsula and Near-East profiles would help clarify the relative affinities among these regions.

Results: Like the Iberian Peninsula, the dominant mtDNA haplogroup H subgroups in North Africa are H1 (42%) and H3 (13%). The similarity between these regions is stronger in the North-West edge affecting mainly Moroccan Arabs, West Saharans and Mauritians, and decreases eastwards probably due to gene flow from Near East as attested for the higher frequencies of H4, H5, H7, H8 and H11 subgroups. Moroccan Berbers show stronger affinities with Tunisian and Tunisian Berbers than with Moroccan Arabs. Coalescence ages for H1 (11 ± 2 ky) and H3 (11 ± 4 ky) in North Africa point to the possibility of a late Palaeolithic settlement for these lineages similar to those found for other mtDNA haplogroups. Total and partial mtDNA genomic sequencing unveiled stronger mtDNA differentiation among regions than previously found using HVSI mtDNA based analysis.

Conclusion: The subdivision of the mtDNA haplogroup H in North Africa has confirmed that the genetic differentiation found among Western and Eastern populations is mainly due to geographical rather than cultural barriers. It also shows that the historical Arabian role on the region had more a cultural than a demic effect. Whole mtDNA sequencing of identical H haplotypes based on HVSI and RFLP information has unveiled additional mtDNA differences between North African and Iberian Peninsula lineages, pointing to an older mtDNA genetic flow between regions than previously thought. Based on this new information, it seems that the Strait of Gibraltar barrier affected both male and female gene flow in a similar fashion.
Background

Bounded by the Mediterranean Sea in the North and the Sahara desert in the South, North Africa behaved as an anthropological island in the African Continent. Previously, only the Strait of Gibraltar in the West and the Suez Isthmus in the East connected this region to Europe and the Near East respectively. These two passageways were the most probable migratory routes followed by human intercontinental dispersals since prehistoric times. In addition, periodic wetter climatic conditions allowed contacts with sub-Saharan African peoples across the Sahara desert and seafaring achievement brought numerous Mediterranean and Atlantic cultures to the North African shores [1]. Archaeological information points to a modern human occupation of this area since 45,000 years ago (ya), as attested by the Aterian industry [2]. However, there is no unanimous assent about the degree of human continuity since that time, as some of the posterior Palaeolithic industries (Iberomarussian, Dabban) exhibit no clear cultural connections with the earlier Aterian form. Furthermore, there is also controversy about the demic impact of the Near East Neolithic on the Northwest African autochthonous Capsian Neolithic [1]. Even more tenuous are the putative connections of the Aterian with the Solutrean of Iberia, or those of the Capsian with the Mediterranean Neolithic [3]. Nonetheless, it is agreed that the historic penetration in the area of the Pharaonic and classical Mediterranean cultures, ending with the Islamic domination, imposed strong cultural influences with only a minor demic impact [4,5]. Population genetic studies using classical markers pointed to a sizeable Upper Palaeolithic component in Northwest African populations [6], whereas the Neolithic diffusion in that region was more a cultural than a demic process [7]. More recently, the haploid characteristics of the uniparental genetic markers allowed the successful application of phylogenetic and phylogeographic approaches to population genetics. Thus, mitochondrial DNA (mtDNA) phylogeographic analyses have enhanced the power of this maternal non-recombinating marker to detect human migrations on continental [8-10] and regional [11-14] scales. Focusing on North Africa, several mtDNA studies have shown that, in spite of an important Sub-Saharan African contribution, the majority of the lineages detected in this region belong to, or have common roots with, Eurasian haplogroups [15-23]. Some of these haplogroups, including the X1 [12], U6 [11,13] and M1 [13,14], although of West Asian origin, have Paleolithic coalescence ages in North Africa. Others seem to be of more recent acquisition as a result of European (U5, V [24-26]) or Middle Eastern influences (R0a, J1b, U3 [17,27-29]). In agreement with classical markers and mtDNA, in an early analysis of Northwest African populations using paternal Y-chromosome variation, it was proposed that the main haplogroups defined by the M78 and M81 binary markers could be the paternal counterparts of the classical and maternal Paleolithic components [30]. However, more recent studies in which those and other markers were further subdivided suggested a predominantly Neolithic origin for the Y-chromosomal DNA variation in North Africa [20,22-34].

The discrepancies in the uniparental marker results could be due to real differences in male and female demographic histories [35]. However, a lack of mtDNA haplogroup resolution could also be responsible. Recently, the accurate dissection of the most frequent Western Eurasian haplogroup H into several monophyletic subhaplogroups [36,10] [37-40] changed a rather uniform genetic landscape into one with several regional peaks and clinal variations. Thus, the frequencies of H1 and H3 subhaplogroups are highest in Western Europe, decreasing gradually to the East. In contrast, H2 occurs more frequently in Eastern than in Western Europe. Besides, there are subhaplogroups that characterize other regions, such as H6 and H8 in Central Asia, H13 in the Near East, H20 and H21 in the Caucasus, and H18 in the Arabian Peninsula. Haplogroup H is also the most frequent clade in North Africa. Global frequencies are highest in the North-west, representing 37% and 34% of their mtDNA lineages in Berber and Arab speaking Moroccans [15], 24% [41] and 32% [19] in Berber Algerian samples and 26% in Tunisian Berbers [20]. The frequencies drop slightly southwards, showing 24% in Saharans [15] and 23% in Mauritans [42], as well as eastwards displaying 21% [16] and 14% [21] in Egyptian samples. The aims of this paper are: 1) to subdivide the North African haplogroup H lineages into its known subhaplogroups, 2) to establish the phylogenetic and phylogeographic patterns of these subhaplogroups in the region, and 3) to compare them with those present in Europe and the Near East, in order to establish the strength of the human migrations from both continents into North Africa in spatial and temporal dimensions.

Results

As described previously, total frequencies for the haplogroup H decline toward both the East and the South (Table 1). The haplogroup H represents 44% of the mtDNA variation in the Iberian Peninsula, but only 22% in the Near East. Likewise, this distribution still reaches 25% in North Africa, but drops to only 9% in the Arabian Peninsula. Haplogroup H subclade distribution is also very different in the various regions. Subhaplogroups H1 and H3 are the dominant subgroups in the Iberian Peninsula (45% and 16%, respectively) and North Africa (42% and 13%, respectively) whereas unclassified H haplotypes (H*) account for 40–50% of the H diversity in the Arabian Peninsula and the Near East. Furthermore, while H1 (12%) is still the most frequent subgroup, followed by the H5 (8%) in the Near East, the modal subclades in the Ara-
Table 1: Distribution of subhaplogroup H frequencies (%) in the studied populations.

| Subhaplogroup | IP | Mau | Sah | Mor | MoB | TuB | Tun | NA | AP | NE |
|---------------|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| H*           | 18 | 18  | 14  | 11  | 30  | 21  | 48  | 26  | 43  | 51  |
| H1           | 45 | 64  | 68  | 63  | 30  | 24  | 24  | 42  | 6   | 12  |
| H1a          | 1  | -   | -   | -   | -   | -   | -   | -   | -   | -   |
| H1b          | 2  | -   | -   | -   | -   | -   | -   | -   | -   | -   |
| H2a          | 4  | -   | -   | -   | -   | -   | -   | -   | -   | 2   |
| H2a1a        | 1  | -   | -   | -   | -   | -   | -   | -   | 18  | -   |
| H3           | 16 | 9   | 7   | 17  | 4   | 21  | 15  | 13  | -   | -   |
| H4           | 3  | -   | -   | -   | 17  | 7   | 2   | 5   | 4   | 4   |
| H5           | 1  | -   | -   | 3   | 2   | 3   | 4   | 2   | 3   | 8   |
| H5a          | 5  | -   | -   | 3   | -   | -   | -   | -   | < 1 | 2   |
| H6           | < 1| -   | -   | -   | -   | -   | -   | -   | -   | -   |
| H6a          | 3  | -   | 4   | -   | 4   | -   | -   | -   | 1   | 3   |
| H6b          | -  | -   | 7   | -   | -   | -   | -   | 1   | 14  | 2   |
| H7           | 1  | 9   | -   | -   | 2   | 17  | 7   | 5   | 3   | 4   |
| H8           | -  | -   | -   | -   | -   | 3   | -   | < 1 | -   | -   |
| H11          | < 1| -   | -   | 3   | -   | 3   | -   | -   | -   | -   |
| H13a1        | 1  | -   | -   | -   | -   | -   | -   | -   | 1   | 4   |
| H14a         | -  | -   | -   | -   | -   | -   | -   | -   | -   | 1   |
| H20          | < 1| -   | -   | -   | -   | -   | -   | -   | 1   | 2   |

H classified 510 11 28 35 46 29 46 203 148 253

N<sub>H</sub> 593 18 30 42 49 29 46 224 149 265
N 1349 102 128 180 154 101 186 880 1685 1201

%H 44 18 23 23 32 29 25 9 22

hHG ± se 74 ± 2 60 ± 15 53 ± 10 58 ± 9 79 ± 3 85 ± 3 70 ± 5 74 ± 2 76 ± 3 72 ± 3

1. IP = Iberian Peninsula; Mau = Mauritania; Sah = Sahara; Mor = Morocco; MoB = Moroccan Berbers; TuB = Tunisian Berbers; Tun = Tunisia; NA = North Africa; AP = Arabian Peninsula; NE = Near East.
2. Haplogroup H (N<sub>H</sub>) and total (N) sample sizes.
3. Gene haplogroup diversity and standard error (hHG ± se), in percentage.
4. < 1 = subhaplogroup present in frequency of less than 1%.

Table 2: F<sub>ST</sub> (by 1,000) based on subhaplogroup, above the diagonal, and haplotype, below the diagonal, frequencies.

|                  | Tun | TuB | MoB | Mor | Sah | Mau | NA | IP | AP | NE |
|------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Tun              | -   | 33  | 30  | 200 | 227 | 129 | 86 | 6  | 20 |
| TuB              | 7   | -   | 22  | 109 | 152 | 56  | 36 | 1  | 100 |
| MoB              | 17  | 48  | -   | 115 | 121 | 55  | 38 | 83 | 61 |
| Mor              | 40  | 51  | 16  | -   | 0   | 0   | 14 | 334 | 308 |
| Sah              | 24  | 45  | 5   | 0   | -   | 0   | 31 | 334 | 316 |
| Mau              | 0   | 11  | 0   | 0   | 0   | -   | 0  | 260 | 231 |

NA 6 145 116

IP 10 24 13 13 4 0 3 - 192 170

AP 18 49 47 80 55 29 42 42 34

NE 6 41 28 64 40 14 28 30 13

* = p < 0.05, ** = p < 0.01, *** = p < 0.001
bian Peninsula are H2a1a (18%) and H6b (14%). Pairwise $F_{ST}$ distances based on sub-haplogroup frequencies display a high heterogeneity among the main regions (Table 2). However, the level of statistical significance between the Iberian Peninsula and North Africa ($p < 0.05$) is lower than that for any other pairwise comparison ($p < 0.001$). In addition, within North African populations, the Tunisians, Tunisian Berbers and Moroccan Berbers are different from the Saharan and Moroccan Arabs, while the last two are comparatively less different from the Iberian Peninsula. The relative proximity of the Iberian Peninsula to the westernmost North African populations is graphically reflected in Figure 1a. It is evident that Tunisians and Berbers are closest to the Near East and the Arabian Peninsula. A principal component analysis (PCA) points to sub-haplogroups H1 and H3 as being primarily responsible for the Iberian-Moroccan-Saharan connection, whereas H4, H5, H7, H8 and H11 testify the Near East influence (data not shown). Similarly, haplotypic based $F_{ST}$ distances show a strong influence of the Iberian Peninsula on the Western Moroccan and Saharan North African populations, and indicate that Tunisians are comparatively the most remarkably influenced by the Near East (Table 2 and Figure 1b). Globally, North Africa shares a similar number of haplotypes with the Iberian Peninsula compared with the Near East (Table 3). However, a detailed analysis of the ratios between haplotypic identities relating each North African population with the Iberian Peninsula or the Near East confirms that the Western populations, comprising Moroccan Arabs, Saharans and Mauritanians, are the most notably influenced by the Iberian Peninsula, whereas the Tunisian Berbers, Tunisians, and the Moroccan Berbers have received relatively more gene flow from the Near East (Table 3). At this point, it is noteworthy that all the Arabian Peninsula haplotypes shared with North Africa are a subset of those shared by the latter with the Near East, pointing to a minor direct input of the Arabian Peninsula on the North African populations. Haplogroup (Table 1) and haplotype (Table 3) genetic diversities demonstrate that the Northwestern African populations (Moroccan Arabs and Saharans) are genetically less diverse than the more central Tunisian and Berbers, a fact that could be explained by a stronger Near East influence on the later populations. Although global haplogroup and haplotypic diversities are not statistically different among regions (Table 1 and 3), the European subgroup H1 appears to be significantly more diverse in the Near East ($87 \pm 5$) than in the Iberian Peninsula ($75 \pm 3$) or North Africa ($67 \pm 6$). Moreover, the genetic diversity for the Western European subgroup H3, which is absent in the Near East, is also higher in North Africa ($74 \pm 9$) than in the Iberian Peninsula ($65 \pm 6$). Transformation of molecular genetic diversities in coalescence ages gives $18,345 \pm 4,051, 14,201 \pm 2,984$, and $11,366 \pm 2,354$ years for H1 in the Near East, Iberian Peninsula and North Africa, respectively. On the other hand, the coalescence ages for H3 in the Iberian Peninsula ($10,342 \pm 2,634$) and North Africa ($10,866 \pm 2,417$) are similar. However, only H1 ages in Near East and North Africa are statistically different from each other.

The relative affinities among regions are based on subhaplogroup frequencies, which do not take into account differences between haplotypes assorted in the same subgroup, or in haplotypic matches, whose identity is based only on partial HVSI sequences. In addition, it has to be taken into account that half of the H lineages detected in North Africa are not shared with other regions and that this percentage is even greater in the putative source regions of the Near East (70%) and the Iberian Peninsula (76%). These facts point to a higher differentiation among regions and between populations than those observed previously. Indeed, complete or nearly complete sequencing of some apparently identical samples indi-

### Table 3: Population and regional haplotypic composition.

| HT | HTU | %hHT | %HTU | %HTNA | %HTIP | %HTTNE | %HTC | IP | TNE | IP/TNE |
|----|-----|------|------|-------|-------|--------|------|----|-----|--------|
| IP | 161 | 122  | 93 ± 1| 76    |       |        |      |    |     |        |
| Mau| 9   | 1    | 95 ± 7| 11    | 22    | -      | 44   | 0.414 | 0.238 | 1.74  |
| Sah| 16  | 6    | 89 ± 5| 38    | -     | 31     | 6    | 0.349 | 0.167 | 2.09  |
| Mor| 20  | 9    | 86 ± 6| 45    | 15    | 20     | 5    | 0.217 | 0.107 | 2.03  |
| MoB| 20  | 5    | 90 ± 3| 25    | 15    | -      | 15   | 0.280 | 0.321 | 0.87  |
| TuB| 15  | 4    | 93 ± 3| 27    | -     | 20     | 13   | 0.373 | 0.285 | 1.31  |
| Tun| 30  | 12   | 97 ± 1| 40    | 7     | 13     | 17   | 0.228 | 0.214 | 1.06  |

| NA | 80  | 43   | 93 ± 1| 54    | 16    | 13     | 18   | 0.210 | 0.160 | 1.31  |
| AP | 58  | 36   | 96 ± 1| 62    |       |        |      |      |      |       |
| NE | 129 | 90   | 97 ± 1| 70    |       |        |      |      |      |       |

Number of different haplotypes (HT), unique haplotypes (HTU), haplotype diversity ($\pm$ standard error) ($%_{hHT}$), unique haplotype frequency ($%_{HTU}$). Frequency of North African haplotypes shared only with other North Africans ($%_{HTNA}$), IP ($%_{HTP}$), NE and/or AP ($%_{HTAP}$), and with both IP and TNE regions ($%_{HTC}$). North African haplotypic identity, by 100, with IP ($I_{HT}$) and with TNE ($I_{TNE}$) and ratio between them (IP/TNE).
cates that the real genetic heterogeneity among regions is greater than those estimated above (Figure 2). To begin with, the HVSI motif 16093-16189 that characterizes subgroup H1f was found in an individual (Mor 2047) from Morocco (Figure 2) also in an H1 background. This subgroup is particularly abundant and mainly restricted to Finland and the surrounding populations [36]. At first sight, this coincidence would seem to point to a new link between North European with North African populations like that found previously for U5b1b [26]. However, in this case, further analysis of the coding region in the North African sample revealed a lack of the three coding region mutations that additionally characterize the Finish H1f subgroup [38] (Figure 2). This lack of identity between haplotypes assorted in the same subgroup and sharing the same or similar HVSI motif can be extended to other cases. For instance, there is a group of H sequences that shares the 16145-16222 HVSI motif consistently found in Northwestern Africa, the Sahara and several Western Sahelian populations [15]. The complete sequencing of a Mauritanian sample (Mau 2027) allowed the assignation of this type to the subhaplogroup H1 (Figure 2). The direct connection of this motif with a German sequence was previously suggested [15]. However, the additional presence of transitions 16304 and 456 in the HVSI and HVSII regions respectively in that German haplotype [43] indicated that it should be classified as belonging to the H5 instead of the H1 subgroup, which does not support a direct link between these regions. In contrast, the two 16145-16222 haplotypes sporadically detected in the Iberian Peninsula [[44] and unpublished results] belonged to the North African subgroup as they shared the coding 10257 mutation, in addition to the H1 diagnostic transition 3010, with the totally sequenced Mauritanian sample (Figure 2). It seems that the 10257 transition defines a new subgroup within H1. This fact points to a possible, although not recent, North African demic influence on the Iberian genetic pool. Another interesting group of sequences belonging to the H1 subgroup in North Africa is that characterized by the 16172-16311 motif, which we [15] and others [19] have found mainly in Saharan samples. Haplotypes with, or including, this HVSI motif have also been detected in European [45,43,8][46-49] and in Asian [50-54] samples, but not in the Iberian Peninsula yet (see Additional file 1). However, the possibility of direct phylogenetic links among such distant regions is very weak, because all of those individuals further classified in both regions belong to the H5 subgroup or the HV haplogroup [48,49] in Europe, or to the HV or the R2 haplogroups [53,54] in the Middle East, which strongly points to yet another case of HVSI convergence in distinct backgrounds of coding regions. In addition to the CRS, the 16189 and the 16311 HVSI motifs are quite abundant in North Africa (see Additional file 1). However, when these samples were screened for the coding region positions observed in completely sequenced European or Middle East individuals that held the same HVSI motifs (Figure 2), none of these positions appeared in the North African samples. This lack of homogeneity again strongly points to their different monophyletic coding backgrounds, in spite of their HVSI matches, a fact repeatedly found in other studies [38]. Indeed, in this study, there are also instances of molecular convergence in the coding region. Sequences How 73H and Jor 843 share the 12236 transition, although they respectively belong to the H* and H5 subgroup (Figure 2). The 12358 transition also presents one such case that is shared by four sequences (Her 127, Ach 28, MM H2, and Mau 2027) belonging to different H subgroups (Figure 2).

**Discussion**

The dissection of mtDNA haplogroup H in North Africa has confirmed several genetic features of its populations. First, there is a significant genetic differentiation between

![Graphical relationships among the studied populations](http://www.biomedcentral.com/1471-2156/10/8)

**Figure 1**

Graphical relationships among the studied populations. Codes are as in Table 1. MDS plots based on $F_{ST}$ haplogroup (a) and haplotypic (b) frequency distances.
Northwestern, Central and Eastern populations already detected since the first genetic studies carried out in North Africa using classical genetic polymorphisms [6]. This differentiation has also been found by posterior molecular analyses using Y-chromosome markers [31-34] or X-chromosome SNPs [55]. Second, as Arab and Berber communities are present in both areas, geographic isolation, more than cultural barriers, seems to be the main cause of this genetic differentiation. This has been consistently reported in all previous studies using autosomal short tandem repeats [4], autosomal Alu insertion polymorphisms [56,57], high-resolution Y-chromosome analyses [30,58,59], and mtDNA polymorphisms [19,20,22,23]. As a consequence, it has been proposed that the North African gene pool has had Palaeolithic and Neolithic influences from the East, but that the impact of the historical invasions, such as the Arabic role, had more a cultural than a demic effect. The lack of exclusive haplotypic matches between North Africa and the Arabian Peninsula found here is in accordance with that hypothesis. Third, the southward clinal diminution of haplogroup H frequencies found at mitochondrial level is well explained as a counteracting effect of the northward clinal diminution of the Sub-Saharan maternal gene flow [15,5,19]. Fourth, the genetic heterogeneity detected between the North African and the Iberian Peninsula populations has been attributed to both the effect of the physical barrier imposed by the Strait of Gibraltar and strong cultural differences. However, some gene flow has been detected between areas and its strength depends mainly on the type of marker used. The strongest barrier effect has been detected in analyses based on Y-chromosome polymorphisms [30]. The levels of gene flow detected in autosomal studies have been of more diverse range [4,56] and, in some cases, seem to depend on the population samples used as is the case with, for instance, the CD4/Alu microsatellite haplotypes [60,61]. In contrast, a high female permeability has been deduced from several mitochondrial studies that pointed to the existence of an important maternal Iberian input on North Africa [15,19]. Although there is no archaeological evidence to justify such a demic flow from Iberia to North Africa, based on the phylogeographic range, comparative gene diversity and ages of several mitochondrial haplogroups such as V, H1, H3, and U5b1b [25,37,26], the presence of these haplogroups in North Africa is thought to be the result of a southward

---

**Figure 2**

Phylogenetic tree of complete (continuous branches) or nearly complete (discontinuous branches) haplogroup H mtDNA sequences. Numbers along links refer to nucleotide transitions. "A" and "T" indicate transversions; "d" deletions and "i" insertions. Recurrent mutations are underlined. The empty box represents a node from which other (not shown) sequences branch. Sequence references are: CRS [64,65]; How 73H, How 78H and How51H [66]; Bra HS [49]; Her 127 ([10], EF657262); Ach 28 and Ach 39 ([37], AY373967 and AY373979); Mis E6H [62], AY195757; MM H2 and MM H1 ([9], AF382002 and AF381992); Fra 27 and Fra 48 ([67], DOS23627 and GOS23648); Fin 413 ([36], AY339413); Jor 843, Mau 2027, Sah 7045, Sev 1179, Geo 2459, Mor 2047 (present study). Geographic origins are: How 73H, How 78H and How51H: Dutch, Jor 843: Jordan, Bra HS and Her 127: Europeans, Ach 28 and Ach 39: Italians, Mis E6H: Israeli, Mau 2027 and MM H1: Moroccans, MM H2 and Sev 1179: Spaniards, Sah 7045: West Saharan, Geo 2459: Georgian, Fra 27 and Fra 48: Sardinians, Mor 2047: Moroccan, Fin 413: Finlander.
expansion of Palaeolithic hunter-gatherers from the Franco-Cantabrian refuge after the Last Glacial Maximum. In fact, coalescence ages for H1 and H3 subclades estimated in this study are in good agreement with those previously published and are congruent with these expansions. Thus, our HVSI based coalescence ages for H1 (14.2 ± 3.0 ky) and H3 (10.3 ± 2.6 ky), in the Iberian Peninsula, are very close to those published by Pereira et al. [40] in the same area for H1 (14.0 ± 3.0 ky) and for all of Europe for H3 (11.0 ± 3.0 ky). Furthermore, striking similarities are observed when these ages are compared to those obtained from the coding region in similar geographic ranges, using the Mishmar et al. calibration [62]. Thus, H1 coalescence ages for Iberia (13.0 ± 6.0 ky; [40]) and Southwest Europe (12.8 ± 2.4 ky; [37]) are very similar between themselves, and not significantly different from those based on the HVSI. Likewise, H3 coding region based coalescence ages for whole Europe (9.0 ± 3.0 ky; [40]) and Southwest Europe (10.3 ± 2.4 ky; [37]) are also very similar to those based only on the HVSI. That Palaeolithic expansion would explain the notorious presence of H1 and H3 detected mainly in the most North-western populations of North Africa and the decrease in their frequency eastwards. However, if this hypothesis held, the comparatively high diversity of H1 and H3 in North Africa would point to an important Palaeolithic gene flow from the Iberian Peninsula to North Africa across the Strait of Gibraltar. On the contrary, a consensus exists regarding the Near East origin of the bulk of the Y-chromosome and mtDNA North African lineages. However, discrepancies still exist with respect to the time in which these settlements most probably occurred. In the first Y-chromosome pioneering studies of the region, a Palaeolithic settlement for the autochthonous E-M81 clade was hypothesized in accordance with the age proposed based on classical markers [30]. However, later studies have assigned this, and other subclades derived from E-M78, that are particularly abundant in North Africa, a Neolithic or even historic settlement age and a Near East or Northeast African source [63,31-34]. On the other hand, for those mtDNA haplogroups pre-eminent in North Africa, that have been analyzed at deep genomic and phylogeographic levels, such as U6 and M1, a Palaeolithic settlement and Middle East roots have been proposed [11,13,14]. From our data, it can be also deduced that the presence of the H1 and H3 subgroups in North Africa could have similar expansion times as in Europe and, therefore, a late Palaeolithic settlement in the region. Finally, it should be noted that the different levels of gene flow detected throughout the Strait of Gibraltar, with respect to Y-chromosome and mtDNA polymorphisms have been attributed to sexual migratory differences, with females showing more permeability than males due to paternal locality and polygyny [5,60,19], and to genetic drift differently affecting both sexes [22,59]. However, the first explanation is not in accordance with the demographic flows known to have occurred between Morocco and Iberia across the Strait of Gibraltar. Historically, the main human movement from Northwest Africa to the Iberian Peninsula was the Islamic Invasion. As a military enterprise, it is believed that this North African gene flow into Iberia was mainly a male contribution. If genetically important, it would homogenize the male lineages between Iberia and North Africa to a greater extent than the female lineages, in contradiction to the experimental results. Little is known about prehistoric contacts between these two areas, but human movements repeatedly crossing the Gibraltar Strait to establish patrilocality seems improbable. The lack of deep sequence identity for several mtDNA haplotypes assorted in the same H subgroup and considered haplotypic matches between North Africa and the Iberian Peninsula, clearly points to the existence of a higher mtDNA heterogeneity between these two regions than suggested in previous studies. If the greater level of differentiation established for H in the present study were extendable to other mitochondrial haplogroups, the female levels of gene flow between both areas would match approximately those of males. Further mtDNA studies at genomic level are necessary to test this hypothesis.

**Conclusion**

The subdivision of mtDNA haplogroup H in North Africa has confirmed that the genetic differentiation found among Western and Eastern populations is mainly due to geographical rather than cultural barriers. It also appears that the historical Arabian role on the region had more a cultural than a demic effect. Whole mtDNA sequencing of apparently identical H haplotypes, based on HVSI and RFLP information, has unveiled additional mtDNA differences between North Africa and the Iberian Peninsula, pointing to the Strait of Gibraltar barrier as affecting male and female gene flow in a similar fashion.

**Methods**

A total of 5,115 mtDNA sequences were analyzed. Of these, 1,231 belonged to the haplogroup H, defined as -7025 AluI by RFLP screening. Four main geographic areas were covered by this study: the Iberian Peninsula (593 H individuals from a total of 1,349), North Africa (224 H individuals from a total of 880), the Near East (265 H individuals from a total of 1,201) and the Arabian Peninsula (149 H individuals from a total of 1,685). Detailed origin and geographic localization for all the samples are specified in Additional file 2. From the 1,231 individuals classified as H, 1,114 could be assorted into one of 19 different H subgroups by further screening for characteristic HVSI and/or HVSII sequence motifs, or diagnostic RFLPs (see Additional file 3). Analyzed individuals that could not be assorted into any of the known groups were considered as H* types. In addition, complete or nearly com-
complete mtDNA sequencing was carried out on 6 individuals with haplotypes found only in well-defined geographic areas or with HVSI haplotypic matches between very distant regions with the aim of accessing whether these matches also held for their coding regions (Figure 2, [64,65,36,9,10,66,62,37,67,49]). Furthermore, in order to find out additional subdivisions, those individuals presenting HVSI matches with already published complete haplogroup H sequences were screened for all coding region positions they hold (Figure 2). DNA extraction, primers, conditions used for PCR amplifications and total or partial sequencing have been published previously [9,29]. RFLP analyses and subhaplogroup H nomenclature (see Additional file 3) were as in Loogväli et al. [38] and Roostalu et al. [68]. Haplogroup and haplotype diversities (h) as well as molecular genetic diversities (π) were calculated according to Nei et al. [69]. Only HVSI positions from 16,024 to 16,365 were used for genetic comparisons of partial sequences with other published data. Phylogenetic relationships among HVSI and genomic mtDNA sequences were established using the reduced median network algorithm [70]. Ages of clades were estimated using the rho statistic [71], and a calibration of 1 transition within np 16090–16365 corresponds to 20,180 years [72] for HVSI sequences.

For population comparisons, Fst distances were calculated based on haplogroup and haplotype frequencies using Arlequin 2.0 [73]. In order to diminish the strong influence of the common haplotypes in Fst distances, an additional measure of haplotypic identity [HIT = (HTXY / (HTX + HTY))] was used, where HTXY is the number of shared haplotypes between populations X and Y, and HTX and HTY are the numbers of different haplotypes in the populations X and Y, respectively. Multidimensional scaling (MDS) plots were obtained from Fst distances and principal component analysis (PCA) from haplogroup frequencies using SPSS version 13.0 (SPSS Inc., Chicago, Illinois).

**Accession numbers**  
The six new complete mitochondrial DNA sequences are registered under GenBank accession numbers: FJ236978–FJ236983.

**Authors’ contributions**  
HE, VMC, AMG, ABEG and JML conceived the project, designed the experiments and wrote the paper; HE, MB, RB, ND, JML collected the samples; HE, MB, RB and VMC extracted the DNA; HE, VMC, AMG, KKAA and JML performed the experiments, sequencing aligning and haplogroups typing; HE, VMC, AMG and JML analyzed the data and performed the statistical analysis. All authors read, revised and approved the final manuscript.

**Additional material**

**Additional file 1**  
H haplotypes detected in the different populations analyzed. The data show the H-subgroup assignment, and HVSI I and HVSI II sequences of individuals analyzed in the following studied populations: Galicia, Castilla-Mancha, Castilla-León, Andalucía, Santander, León, Potes, Pasiegos, Maragatos, Tunisian Berbers, Tunisia, Libya, Algeria, Algeria Berbers, Moroccan Berbers, Morocco, Sahara, Mauritania, Saudi-Arabia, Jordan, Dead Sea, Palestine, Syria, Iran.

Click here for file  
[http://www.biomedcentral.com/content/supplementary/1471-2156-10-8-S1.xls]

**Additional file 2**  
Origin, localization, sample size, number of H, number of H1 classified in subhaplogroups, and percentage of H in different populations. The table list the populations and geographic areas used to search for H haplotypes assigned to an H subgroup. Origin, localization, sample size, number of H, number of H1 classified in subhaplogroups, percentage of H and references are shown.

Click here for file  
[http://www.biomedcentral.com/content/supplementary/1471-2156-10-8-S2.xls]

**Additional file 3**  
Diagnostic mt SNPs, RFLPs and primers used to detect the different H subhaplogroups. Table shows the diagnostic mt SNPs, RFLPs and primers used to detect the different H subhaplogroups in all or some of the samples.

Click here for file  
[http://www.biomedcentral.com/content/supplementary/1471-2156-10-8-S3.xls]

**Acknowledgements**  
This research was supported by grant BFU2006-04490 from Ministerio de Ciencia y Tecnología to J.M.L.

**References**

1. N. Newman JL: The peopling of Africa. A geographic interpretation New Haven and London: Yale University Press; 1995.
2. Garcea EA, Giraudi C: Late Quaternary human settlement patterning in the Jebel Gharbi. J Hum Evol 2006, 51:411-421.
3. Straus LG: Africa and Iberia in the Pleistocene. Quaternary International 2001, 75:91-102.
4. Bosch E, Calafell F, Perez-Lezaun A, Clarimon J, Comas D, Mateu E, Martinez-Arias R, Morera B, Bratek Z, Akhaya O, Safani A, Hariti G, Cambon-Thomsen A, Bertranpetit J: Genetic structure of northwest Africa revealed by STR analysis. Eur J Hum Genet 2000, 8:360-366.
5. Flores C, Hernandez M, Gonzalez AM, Cabrera VM: Genetic affinities among human populations inhabiting the Subsaharan area, northwest Africa, and the Iberian Peninsula. In Proceedings of Prehistoric Iberia: Genetics, Anthropology, and Linguistics Edited by: Arnaiz-Villena A. New York: Kluwer Academic/Plenum Press; 2000:33-50.
6. Bosch E, Calafell F, Perez-Lezaun A, Comas D, Mateu E, Bertranpetit J: Population history of north Africa: evidence from classical genetic markers. Hum Biol 1997, 69:295-311.
7. Barbuiani G, Pilastro A, De Domenico S, Renfrew C: Genetic variation in North Africa and Eurasia: Neolithic demic diffusion vs. Paleolithic Colonisation. Am J Phys Anthropol 1994, 95:137-154.
8. Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, Sellitto D, Cruciani F, Kivisild T, Villens R, Thomas M, Rychkov S,

(page number not for citation purposes)
Rychkov O, Rychkov Y, Golge M, Dimitrov D, Hill E, Bradley D, Romano V, Cali F, Vona G, Demaine A, Papila S, Triantaphyllids C, Stefancini G, Vina A, Belotti D, Di Rienzo A, Novellino M, Oefner PJ, Underhill PA, Bertranpetit J, Richards M, Torroni A, Bandelt H, Jähnig S, Trebicka K, Khusnutdinova E, Zeviani M, Briem E, Carelli V, Moral P, Dugoujon J-M, Roostalu U, Kivisild T, Bandelt HJ, Arredi B, Poloni ES, Paracchini S, Zerjal T, Fathallah DM, Makrelou M, Stefanescu G, Hatina J, Belledi M, Di Rienzo A, Novelletto A, Oppenheimer A, Newby S, Al-Zahery N, Santachiara-Benerecetti S, Scozzari R, Torroni A, Bandelt H: Tracking European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* 2000, 67:1251-1276.

Maca-Meyer N, González AM, Larruga JM, Flores C, Cabrera VM: Major genomic mitochondrial lineages delineate early human expansions. *BMCC Genetics* 2001, 2:13.

Hernštädt C, Elson JF, Faly E, Preston G, Turnbull DM, Anderson C, Ghosh SS, Olefsky JM, Beal MF, Davis RE, Howell N: Reduced genomic- network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian and European haplogroups. *Am J Hum Genet* 2002, 70:1151-1171.

Maca-Meyer N, González AM, Pestaño J, Flores C, Larruga JM, Cabrera VM: Mitochondrial DNA transition between West Asia and Europe inferred from U6 phylogeography. *BMCC Genetics* 2003, 4:15.

Reidla M, Kivisild T, Vennemann TD, Parik J, Loogväl El, Derenzo M, Ma;larchuk B, Bermisheva M, Zhedanov S, Pennekinen E, Gubina M, Golubenko M, Damba L, Fedorova S, Gusar V, Grechina A, Girad VA, Moshitsky PP, Khusnutdinova E, Osipova L, Stepanov V, Voevoda M, Achilli A, Rengo C, Richards O, De Stefano GF, Papila S, Beckman L, Janicijevic B, Rudan P, Anagnou P, Malakidezi E, Koziol S, Usanga E, Geberhiwot T, Hernštädt C, Howell N, Torroni A, Villens R: Origin and diffusion of European mtDNA haplogroup X. *Am J Hum Genet* 2003, 73:1166-1176.

Olivieri A, Achilli A, Pala M, Battaglia V, Fadhlaoui-Zid K, Plaza S, Calafell F, Bertranpetit J: High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between northeastern and southwestern Africa. *Am J Hum Genet* 2001, 68:1019-1029.

Arredi B, Poloni ES, Paracchini S, Zerjal T, Fadhlaoui-Zid K, Macaulay V, Caprioli A, Vona G, Demaine A, Papiha S, Triantaphyllidis C, Marchini JL, Brindisi M, Benincasa T, Underhill PA, Santachiara-Benerecetti AS: Origin, diffusion and differentiation of Y-chromosome haplogroups E1b3a, F and G1 in Africa. *Am J Hum Genet* 2004, 75:1014-1022.

Semino O, Magni C, Benuzzi G, Liakou AA, Al-Zahery N, Battaglia V, Macaulay V, Santachiara-Benerecetti AS, Semino O, Bandelt HJ, Torroni A: The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa. *Science* 2004, 312:1767-1770.

González AM, Larruga JM, Abu-Amero KK, Shi Y, Pestaño J, Cabrera VM: Mitochondrial lineage M1 traces an early human back-flow to Africa. *BMCC Genetics* 2007, 8:223.

Rando JC, Pinto F, González AM, Hernández M, Larruga JM, Cabrera VM, Bandelt HJ: Mitochondrial DNA analysis of Northwest African populations reveals genetic exchanges with European, Near-Eastern, and sub-Saharan populations. *Ann Hum Genet* 1998, 62:531-550.

Kring RA, Hladik CM, Burke R, Glass KA, Chakraborty K, Gubina M, Golubenko M, Damba L, Fedorova S, Gusar V, Grechina A, Girad VA, Moshitsky PP, Khusnutdinova E, Osipova L, Stepanov V, Voevoda M, Achilli A, Rengo C, Richards O, De Stefano GF, Papila S, Beckman L, Janicijevic B, Rudan P, Anagnou P, Malakidezi E, Koziol S, Usanga E, Geberhiwot T, Hernštädt C, Howell N, Torroni A, Villens R: Origin and diffusion of European mtDNA haplogroup X. *Am J Hum Genet* 2003, 73:1166-1176.

Olivieri A, Achilli A, Pala M, Battaglia V, Fadhlaoui-Zid K, Plaza S, Calafell F, Bertranpetit J: High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between northeastern and southwestern Africa. *Am J Hum Genet* 2001, 68:1019-1029.

Arredi B, Poloni ES, Paracchini S, Zerjal T, Fadhlaoui-Zid K, Macaulay V, Caprioli A, Vona G, Demaine A, Papiha S, Triantaphyllidis C, Marchini JL, Brindisi M, Benincasa T, Underhill PA, Santachiara-Benerecetti AS: Origin, diffusion and differentiation of Y-chromosome haplogroups E1b3a, F and G1 in Africa. *Am J Hum Genet* 2004, 75:1014-1022.

Semino O, Magni C, Benuzzi G, Liakou AA, Al-Zahery N, Battaglia V, Macaulay V, Santachiara-Benerecetti AS, Semino O, Bandelt HJ, Torroni A, Scozzari R: Phylogeographic analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and outside of Africa. *Am J Hum Genet* 2004, 75:1023-1034.

Cruciani F, La Fratta R, Santolamazza P, Zeviani M, Briem E, Carelli V, Moral P, Watson E, Guida V, Colomb EB, Zaharova B, Lavinha J, Vona G, Aman R, Cali F, Akar N, Richards M, Torroni A, Novelletto A, Scozzari R: Phylogeographic analysis of haplogroup E6b (E-M215) Y chromosomes reveals multiple migratory events within and outside of Africa. *Am J Hum Genet* 2004, 75:1014-1022.

Wood ET, Stover DA, Ehret C, Destro-Bisol G, Spedini G, McLeod H, Louie L, Bamshad M, Strassmann BI, Soodyall H, Hammer MF: Contrasting patterns of Y-chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes. *Eur J Hum Genet* 2005, 13:867-876.

Finnila S, Lehtonen MS, Majamaa K: Phylogenetic network for European mtDNA haplogroup F. *Ann Hum Genet* 2001, 65:1475-1486.

Cruciani F, La Fratta R, Santolamazza P, Zeviani M, Briem E, Carelli V, Moral P, Watson E, Melegh B, Barbujani G, Fuselli S, Vona G, Zagradsinik B, Assum G, Kozlov AI, Efremov GD, Coppa G, Scozzari R, Zeviani M: Tracing past human male movements in northern/eastern Africa and western Eurasia: new clues from Y-chromosome haplogroups E-M78 and J-M12. *Mol Biol Evol* 2007, 24:1300-1311.

Wood ET, Stover DA, Ehret C, Destro-Bisol G, Spedini G, McLeod H, Louie L, Bamshad M, Strassmann BI, Soodyall H, Hammer MF: Contrasting patterns of Y-chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes. *Eur J Hum Genet* 2005, 13:867-876.

Finnila S, Lehtonen MS, Majamaa K: Phylogenetic network for European mtDNA haplogroup F. *Ann Hum Genet* 2001, 65:1475-1486.
vas of mtDNA haplogroup H in Eurasia. Mol Biol Evol 2004, 21:202-207.

39. Quintaúas B, Alvarez-Iglesias V, Salas A, Phillips C, Lareu MV, Car- racedo A: Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNAPSHOT minisequencing. Forensic Sci Int 2004, 140:251-257.

40. Pereira L, Richards M, Gozos A, Alonso A, Albarrán C, García O, Behar DM, Goebel T, Hatzina J, Al-Gazali L, Bradley DG, Macaulay V, Amorim A: High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium. Genome Res 2005, 15:9-24.

41. Corte-Real HB, Macaulay VA, Richards MB, Hariti G, Issad MS, Cambrion Thomas BM, Papatheo S, Bertranpetit J, Sykes BC: Genetic diversity in the Iberian Peninsula determined from mitochondrial DNA sequence. Ann Hum Genet 1996, 60:331-350.

42. González AM, Cabrera VM, Larruga JM, Noumsi G, Tho- bon-Thomsen A, Papiha S, Bertranpetit J, Sykes BC: Mitochondrial DNA control region in the Slovenian population. Forensic Sci Int 2005, 148:211-218.

43. Flores C, Maca-Meyer N, Gonzalez AM, Cabrera VM: mtDNA variation of Arab and Berber groups. Ann Hum Biol 2006, 33:634-640.

44. Flores C, Maca-Meyer N, Pérez JA, Hernández M, Cabrera VM: Y-chromosome differentiation in Northwest Africa. Hum Biol 2001, 73:513-524.

45. Ennafaa H, Amor MB, Yacoubi-Loueslati B, Khodjet el-khiil H, Gonzalez-Perez E, Moral P, Maca-Meyer N, Elgaied A: Alu polymorphisms in Jerba Island population (Tunisia): comparative study in Arab and Berber groups. Ann Hum Biol 2006, 33:634-640.

46. Brandstaetter A, Zimmermann B, Wagner J, Goebel T, Roeck AW, Brandt-Casadevall C, Hernández A, Papiha S, Bertranpetit J, Sykes BC: Mitochondrial DNA characterization of European isolates: the Marates from Spain. Eur J Hum Genet 2001, 9:708-716.

47. Pult I, Sajantila A, Simanainen J, Georgiev O, Schaffner W, Paabo S: Mitochondrial DNA sequences from Switzerland reveal sharing of mtDNA haplotypes with Northwest African populations. Biomed. Bioprop. Syst. 2005, 375(12):837-840.

48. Tagliafierri A, Turchi C, Buscemi L, Sassaroli C: Polymorphism of the mitochondrial DNA control region in Italians. Int J Legal Med 2001, 114:224-228.

49. Pastor C, Zjunic Pajic I, Balazic J, Kornel R: Sequence polymorphism of the mtDNA control region in the Slovenian population. Int J Legal Med 2004, 118:1-4.

50. brandstaetter A, Zimmermann B, Wagner J, Goebel T, Roeck AW, Salas A, Capraro A, Papatheo S, Caputo W: Timing and deci- pherment of mitochondrial DNA macro-haplogroup R0 variability in Central Europe and Middle East. BMC Evol Biol 2008, 8:191.

51. Voevodova M, Avskentsuyuk AV, Ivanova AV, Astakhova TI, Babenko VN, Kurilovitch SA, Duff LF, Segal B, Shields GF: Molecular genetic studies of the population of native inhabitants of Chukchee Peninsula. Analysis of polymorphism of mitochondrial DNA and of genes controlling alcohol metabolizing enzymes. Siberi- ski Ekolog Z 1994, 1:139-151.

52. Komlos J, Samboughini N, Berhingham E: Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. Genetics 1996, 142:1321-1334.

53. Comas D, Calafell F, Mateu E, Perez-Lezaun A, Bosch E, Martinez- Arias R, Clarimon J, Facchini F, Fiori G, Luissi D, Petcher D, Bertranpet- tit J: Treading genes along the silk road: mtDNA sequences and the origin of central Asian populations. Ann Hum Genet 1998, 62:1824-1838.

54. Al-Zahery N, Semino O, Benazzi M, Magri C, Passarino G, Torroni A, Santachiara-Benerecetti AS: Y-chromosome and mtDNA polymorphisms in Iraq, a crossroad of the early human dispersal and Neolithic migrations. Mol Phylogenet Evol 2004, 28:458-472.

55. Metspalu M, Kivisild T, Metspalu E, Valsangra L, Komul E, Loogvall E, Vilms M, Kivisild T: Extent and geographical distribution of Mitochondrial DNA haplogroup H in Eurasia. Ann Hum Genet 2000, 64:180-191.

56. Comas D, Calafell F, Benchemsi N, Helal A, Lefranc G, Stoneking M, Batzer MA, Bertranpetit J, Sajantilla A: Alu insertion polymor-