



# OPEN The origin of modern North Africans as depicted by a massive survey of mitogenomes

Giulia Colombo<sup>1</sup>, Elisabetta Moroni<sup>1</sup>, Alessandro Raveane<sup>1</sup>, Nicola Rambaldi Migliore<sup>1</sup>, Vincenzo Agostini<sup>1</sup>, Rosalinda Di Gerlando<sup>1,2</sup>, Claudio Fiorini<sup>3</sup>, Leonardo Caporali<sup>3</sup>, Francesca Gandini<sup>4,5</sup>, Elena Raimondi<sup>1</sup>, Eugenia D'Atanasio<sup>6</sup>, Hovirag Lancioni<sup>7</sup>, Valerio Carelli<sup>3,8</sup>, Maria Pala<sup>9</sup>, Beniamino Trombetta<sup>10</sup>, Andrea Novelletto<sup>11</sup>, Jean-Michel Dugoujon<sup>12</sup>, Alessandro Achilli<sup>1</sup>, Antonio Torroni<sup>1</sup>, Martin B. Richards<sup>9</sup>, Fulvio Cruciani<sup>10</sup>, Ornella Semino<sup>1</sup> & Anna Olivieri<sup>1</sup>✉

North Africa occupies a unique geographical position in the Mediterranean basin. The genetic profile of modern North African populations is the result of ancient and recent movements of people and periods of isolation. In this paper, we provide a comprehensive mitogenome analysis based on unbiased population sampling from across North Africa including a dataset of 733 modern (238 of which new) and 43 ancient mitogenomes from Morocco, Algeria, Tunisia, Libya and Egypt, compared with modern mitogenomes from Eurasia and sub-Saharan Africa. Our analyses, based on a phylogeographic and demographic approach suggest that mitogenome variability in North African populations (i) can be traced back to three main origins - Eurasia, North Africa and sub-Saharan Africa - and (ii) has been defined during crucial key time periods - from the Palaeolithic to the last millennia. In particular, we identified a signal for population movements from sub-Saharan Africa to North Africa during the favourable climatic period of the Green Sahara, providing evidence from complete mitogenomes that the human dispersals associated with this period also involved females.

The genetic profile of modern North African populations is the result of both long periods of isolation and the variegated contribution of people who arrived in the region from different geographic sources at different times for more than 40,000 years.

The first securely dated evidence of the early phase of *H. sapiens* evolution in Africa is represented by the fossils discovered at the Jebel Irhoud site, in present-day Morocco dated to around  $315 \pm 34$  thousand years ago (kya)<sup>1</sup>. Several successive archaeological cultures have also appeared before the Neolithic (such as the Middle Palaeolithic Aterian, the Upper Palaeolithic Dabban in Cyrenaica and Iberomaurusian, and the Mesolithic Capsian or Oranian) across North Africa<sup>2–4</sup>. The Neolithic and subsequent periods saw the rise and spread of populations of the Maghreb, known as Berbers, Amazigh or Imazighen. These names refer to diverse cultures characterised by the same linguistic group, the Berber languages, which belong to the Afro-Asiatic family. Neolithic movements also involved the arrival of populations from both Europe and the Middle East<sup>5,6</sup> while historical times are associated with the advent of several populations from the Mediterranean basin, including Phoenicians (from ~ 1200 B.C.), Romans (from the IInd century B.C.), Vandals, and Byzantines (from ~ 400 A.D.). The arrival of the Arabs in the Maghrebi regions (in north-west Africa) from the VIIth century A.D. had had

<sup>1</sup>Department of Biology and Biotechnology "Lazzaro Spallanzani", University of Pavia, Pavia 27100, Italy.

<sup>2</sup>Molecular Biology and Transcriptomic Unit, IRCCS Mondino Foundation, Pavia, Italy. <sup>3</sup>IRCCS Istituto delle Scienze Neurologiche di Bologna, Programma di Neurogenetica, Bologna 40139, Italy. <sup>4</sup>B-Cell Neoplasia Unit and Strategic Research Program on CLL, IRCCS Ospedale San Raffaele, Milan, Italy. <sup>5</sup>Medical School, Università Vita-Salute San Raffaele, Milan, Italy. <sup>6</sup>Istituto di Biologia e Patologia Molecolari (IBPM), Consiglio Nazionale delle Ricerche (CNR), Rome 00185, Italy. <sup>7</sup>Department of Chemistry, Biology and Biotechnology, University of Perugia, Perugia 06123, Italy. <sup>8</sup>Department of Biomedical and Neuromotor Sciences, University of Bologna, Bologna 40139, Italy. <sup>9</sup>School of Applied Sciences, University of Huddersfield, HD1 3DH, Queensgate, Huddersfield, England, UK. <sup>10</sup>Dipartimento di Biologia e Biotecnologie "Charles Darwin", Sapienza Università di Roma, Rome 00185, Italy. <sup>11</sup>Department of Biology, University of Rome "Tor Vergata", Rome 00133, Italy. <sup>12</sup>Laboratoire d'Anthropologie Moléculaire et Imagerie de Synthèse (AMIS), UMR 5288, Université Paul Sabatier Toulouse III, Toulouse 31073, France. ✉email: anna.olivieri@unipv.it

the greatest impact on Berber cultures, bringing the consequent spread of the Arab language and Arabisation of Berber populations. Finally, in colonial times, the region saw the arrival of Europeans and Ottoman Turks<sup>7</sup>.

During the 1990s, analyses on classical markers<sup>8</sup> indicated that a major role in shaping the genetic landscape of North Africa had been played by its geographic position. The region is delimited by the Mediterranean to the north, the Atlantic in the west, the Nile in the east, and the Sahara Desert to the south. At different times these served as either point of access or physical barriers, with two consistently accessible waypoints to the outer regions represented by the Sinai towards the Levant, and the Strait of Gibraltar towards Europe. This likely contributed to the establishment of a broadly east-west gradient of genetic diversity<sup>9</sup>.

In addition, the isolation of North Africa from sub-Saharan Africa has varied over the millennia, due to strong periodic climatic fluctuations that affected the Sahara, the world's largest desert, alternating arid and humid periods<sup>10,11</sup>. The most recent African Humid Period, the so-called last Green Sahara that peaked 9 to 6 kya<sup>12</sup> made possible settlements and movements of human groups in and throughout the Saharan belt<sup>13</sup>. After 5 kya, rapid desertification established a geographic barrier, limiting human movements across this vast region until the domestication of the dromedary and consequent expansion of trans-Saharan trade, likely around the end of the second millennium BC<sup>14</sup>. This relative isolation of northern regions perpetuated the profound genetic difference from the rest of the African continent established during the Late Palaeolithic, so that North Africa is genomically much closer to West Eurasian than sub-Saharan populations<sup>15,16</sup>.

However, despite the paucity of genetic data from North African populations in comparison with the rest of the African continent and Eurasian populations<sup>17</sup> recent genomic and archaeogenomic data indicate that the picture is more complex than previously thought<sup>15–21</sup>. Modern genomics points to four major genetic ancestries to explain the genetic variability of North African populations with three of them, the sub-Saharan, Middle Eastern, and European components, representing genetic inputs from surrounding populations that at different times left their traces in North Africa, and the fourth one, also known as Maghrebi, typical of North Africa and ancestral<sup>22</sup>. This component is also supported by the mitochondrial DNA (mtDNA) haplogroups U6 and M1, the only two out-of-Africa haplogroups primarily introduced in the region<sup>23–25</sup> and is the legacy of an early “back-to-Africa” dispersal event from western Eurasia that dates back to pre-Holocene times.

This picture has recently been refined using ancient genomics, although North African climatic conditions render ancient DNA (aDNA) preservation particularly poor. The most ancient remains analysed so far are from the archaeological site of Taforalt in Morocco, dating to 13.9–15.1 kya<sup>18</sup>. They belong to the Iberomaurusian culture, which spanned the Last Glacial Maximum and the start of the Holocene (roughly equivalent to Epipalaeolithic times in the Near East) and are characterised entirely by the genomic Maghrebi component mentioned above U6 and M1 lineages. U6 is also seen at the only other pre-Neolithic sites analysed in Morocco and, more recently, Algeria and Tunisia<sup>26</sup>. Interestingly, when additional early and late Neolithic sites in Morocco were assessed, it emerged that early Neolithic North Africans were characterised by both Maghrebi and sub-Saharan genomic components, while late Neolithic individuals were mainly influenced by a Mediterranean-like ancestry, probably brought by the Neolithic expansion that occurred first from the Iberian Peninsula through the Gibraltar straits and later from the Near East<sup>15,26</sup>. Indeed, genome-wide sub-Saharan ancestry fell from around a quarter in the Palaeolithic and Early Neolithic to less than 10% by the Late Neolithic according to these datasets, although the Guanches, thought to have settled the Canary Islands from Morocco ~2.5 kya, retained about 15% sub-Saharan ancestry<sup>15</sup> (Fig. 2 and supplementary notes 6 and 7; cf. also<sup>27</sup> Fig.S3 corrected).

More recent individuals from the sites of Kerkouane in North Tunisia<sup>20</sup> and Kulubnarti in Nubia<sup>19</sup> dating back to the mid-Iron Age and late Iron Age, respectively, revealed that the genetic history of North African regions was strongly influenced by the increased rate of mobility across the Mediterranean, bringing to share ancestries within various populations of the entire basin.

The maternally inherited mtDNA has typically been used for one or other of two different kinds of analyses: (i) molecular dissection and characterisation of selected mtDNA lineages/haplogroups; (ii) population survey of unselected samples. Modern North African population analyses have been mostly limited to the mtDNA control-region<sup>28–31</sup> whereas complete mitogenomes can provide a much more detailed and precise picture<sup>16,32–40</sup>. To date, only one study surveyed a population-based dataset<sup>40</sup> but this was restricted to a single Berber population from north-east Algeria. Thus, a comprehensive mitogenome analysis based on unbiased population sampling from across North Africa is still a priority.

To this end, we obtained a dataset of 238 novel complete mitogenomes from modern North Africans, encompassing samples from Morocco, Algeria, Tunisia, and Libya, together with 135 new mitogenomes from the neighbouring countries of Burkina Faso, Niger, Nigeria, Cameroon, and Chad. The latter mitogenomes include samples from two nomadic populations, the Foulbe (from Cameroon, Niger and Nigeria) and the Berber-speaking Tuareg (from Niger, Supplementary Table 1). This new dataset was analysed in the context of 495 modern and 43 ancient previously published mitogenomes from the same geographic regions (Table 1).

## Materials and methods

### Sample

To survey mitogenome sequence variation across North African populations and in surrounding regions, DNA was obtained from 238 North African individuals sampled from four countries, i.e. Morocco, Algeria, Tunisia, and Libya, and 135 from neighbouring countries (Burkina Faso, Cameroon, Chad, Niger and Nigeria), for a total of 373 novel mitogenomes. Specifically, as for North Africa, we sequenced the mitogenomes of 95 Moroccans (35 of Arab origin and 60 from the general population), 49 Algerians, 34 Tunisians, and 60 Libyans, from the general population. The new mitogenomes from neighbouring countries are from 22 individuals from Burkina Faso's general population, 29 Foulbe from Cameroon, 65 individuals from Chad (including one Goulaye, two Madjingay, two Ngambai, and the remaining 60 from the general population), three Foulbe and seven Tuareg from Niger, and nine Foulbe from Nigeria.

Macroarea	Country	This study	Literature*	
		Modern (n = 373)	Modern (n = 495)	Ancient (n = 43)
North Africa	Morocco	95	7	26
	Algeria	49	302	–
	Tunisia	34	49	13
	Libya	60	2	2
	Egypt	–	135	2
Total North Africa		238	495	43
Sub-Saharan Africa	Burkina Faso	22	–	–
	Niger	10	–	–
	Nigeria	9	–	–
	Cameroon	29	–	–
	Chad	65	–	–
Total sub-Sahara		135	–	–

**Table 1.** Mitogenomes analysed in the present study. \*Population surveys only.

All the DNA samples analysed here were selected from our lab collections, assembled on a voluntary basis over the past decades after obtaining appropriate informed consent and according to all the regulations in effect at that time. The use of the “historical” collections in genomic studies was approved by the Ethical Committee Fondazione IRCCS Policlinico San Matteo (protocol number 0028298/22). DNA was extracted by a standard phenol-chloroform method from peripheral blood after a red blood cell lysing step.

### Sequencing of entire mitogenomes and analysis of the sequences obtained

The complete sequences of the 373 North African and sub-Saharan African mtDNAs were obtained by using next-generation sequencing (NGS) with an Illumina MiSeq at the Molecular Biology and Transcriptomic Unit of the IRCCS Mondino Foundation in Pavia (Italy) and at the IRCCS Institute of Neurological Sciences in Bologna (Italy). The workflow included the amplification of the entire mtDNA within two overlapping long range PCR fragments, PCR purification, Illumina NGS library preparation, pooling and sequencing reactions, following the protocols previously described in Brandini et al.<sup>41</sup>. The files obtained from the MiSeq system were demultiplexed using the software bcl2fastq. FASTQ files were aligned to the reference sequence (rCRS; NC\_12920.1)<sup>42</sup> using the software BWA<sup>43</sup> the obtained BAM files were then filtered and sorted using SAMtools<sup>44</sup>. Sequence variants (nucleotide substitutions and indels) were then called using HaplotypeCaller, a tool within the Genome Analysis Toolkit (GATK)<sup>45</sup>. BCFtools<sup>46</sup> was applied to further filter the final SNP dataset. The threshold used to detect heteroplasmies was 30% of mutated bases and heteroplasmies with coverage lower than 20 were filtered out. Lastly, fasta were produced using HaploGrep2<sup>47</sup>, which was also employed to predict haplogroup for every subject. The sequence data are available in GenBank with the accession numbers (PV621465 - PV621837) and can be accessed via the following link: <https://www.ncbi.nlm.nih.gov> (Here is the link: **373 mitogenomes - Colombo**). Any additional information concerning sequence and analytical data employed in this paper is available from the lead contact upon request.

### Phylogenetic and phylogeographic analyses

The 373 North African and sub-Saharan African mitogenomes were analysed together with 495 modern and 43 ancient complete mitogenomes publicly available (Supplementary Tables S1 and S2). The sequences were compared to the rCRS<sup>42</sup> using Sequencher 5.0 (Gene Codes Corporation). We built the phylogenetic tree by using a maximum parsimony approach. Haplogroup labels were assigned by following the nomenclature proposed by PhyloTree database build 17 (at <http://www.phylotree.org/>)<sup>48</sup>. Known problematic regions and mutations close to indels (from nucleotide positions (nps) 302 to 318 and from nps 3105 to 3110) were not considered in the phylogenetic tree<sup>49</sup>. Variant calls were further evaluated phylogenetically by assessing for each mitogenome the congruence with the expected haplogroup background<sup>48</sup>. By using this approach, potentially incorrect calls were identified, and eventually corrected.

The frequencies of macro-haplogroups and haplogroups were evaluated across the entire dataset of modern and ancient mitogenomes analysed in this study. Due to the small size of groups based on ethnic/linguistic affiliation, mitogenomes were grouped solely on geographic origin. A smaller phylogenetic tree, encompassing only North African-specific haplogroups (H1v, H1x, H1w and J2a2b1a\*) was also built. These three H1 sub-lineages were previously described as North African-specific<sup>50</sup> while J2a2b1a\* was considered as North African-specific as it fulfilled the requirements previously established by Olivieri et al.<sup>49</sup>: (i) it includes only mtDNAs of North African origin; (ii) it is defined by at least three mitogenomes, encompassing a minimum of two haplotypes (i.e. differing by at least one homoplasmic mutation); (iii) it is characterised by at least one stable mutation (i.e. not recurrent in the tree) at its root. A total of 40 subjects was employed for the construction of the tree, also including mitogenomes published in papers centred on the characterisation of specific lineages. Finally, in order to answer questions concerning a single clade, the phylogenetic tree of the L1b1a6 haplogroup was also built by including all publicly available mitogenomes. These datasets were also analysed using the software BEAST v. 2.5<sup>51</sup> with the same settings reported below.

The entire dataset was analysed with the software Bayesian Evolutionary Analysis Sampling Trees (BEAST) 2.5<sup>51,52</sup>. The program ran under the HKY substitution model (gamma-distributed rates plus invariant sites) with a fixed molecular clock for a minimum of 150,000,000 iterations, with samples drawn every 10,000 Markov chain Monte Carlo (MCMC) steps, after a discarded burn-in of 10,000,000 steps, as in Olivieri et al.<sup>49</sup>. All haplogroups and their major subclades were considered monophyletic in the analyses. Following the same approach as in Olivieri et al.<sup>49</sup> the two clock rates proposed by Soares et al.<sup>53</sup> and Posth et al.<sup>54</sup> ( $2.3 \pm 0.2 \times 10^{-8}$  and  $2.7 \pm 0.2 \times 10^{-8}$  base substitutions per nucleotide per year over the entire mitogenome, respectively) were entered as priors in separate runs. Radiocarbon dates of ancient specimens or the age estimate of the burial cultural context were used as priors in the analysis. The demographic analysis was performed by using Tracer v1.7.1<sup>55</sup>, the results were then plotted with the package *ggplot2*<sup>56</sup> employing a generation time of 25 years.

### Principal component analysis

In order to graphically display (and summarise) the relationships among different populations, Principal Component Analysis (PCA) was performed. The North African mitogenomes analysed in this study, were compared to a dataset of around 18,000 mitogenomes, selected from western Eurasian and sub-Saharan African populations. Only data from population-based studies and available on public repositories (Supplementary Table S3) was used for this step of the analysis. The PCA was performed on haplogroup frequency and computed with *prcomp*<sup>57</sup> command from the *stats* R package and plotted using *ggplot2*<sup>56</sup> and *factoextra*<sup>58</sup>.

### Results and discussion

We evaluated the mitogenome variation of 776 (733 modern and 43 ancient) North African subjects from Morocco, Algeria, Tunisia, Libya, and Egypt. The dataset, which includes 238 newly sequenced modern mitogenomes of North African origin together with 495 and 43 previously published mitogenomes from modern and ancient subjects, respectively (Table 1, Supplementary Tables S1 and S2), was employed to reconstruct a comprehensive phylogenetic tree of all available North African mitogenomes (Fig. 1, see Supplementary Figure S1 for details).

#### Mitogenome variation in modern North Africa

The analysis revealed that modern North African mitogenomes ( $N = 733$ ) cluster into 622 different haplotypes, showing a value of haplotype diversity of 0.9993, similarly to the results previously obtained on control-region sequences of Berber populations<sup>30</sup>. The high haplotype variability is also correlated with the identification of a high number of haplogroups and sub-haplogroups most of which (236 out of 368 identified clades) are members of typical western Eurasian branches (H, HV, I, J, K, N1, R0a, T, U, W, and X), observed at a frequency range between 62.6% (Tunisia) and 88.7% (Libya) (Fig. 2). A significant number of typical/autochthonous North African clades (45 out of 368, specifically H1v, H1x, M1, J2a2b1a\* and U6 and their subclades) were also detected with a frequency range between 15.4% (Algeria) and 4.4% (Egypt). Lastly, sub-Saharan haplogroups (within macro-haplogroup L) encompass the remaining clades (87 out of 368) with frequencies spanning from 26.5% (Tunisia) to 3.2% (Libya) (Fig. 2, Supplementary Figure S1, Supplementary Tables S1 and S4). The frequency data of modern mitogenomes analysed and published ancient North African mitogenomes were compared in Fig. 2. However, the latter dataset, for the sake of clarity, will be discussed later.

Overall, the typical/autochthonous component accounts for a minor portion of North African modern mitogenomes, while the Eurasian one plays a major role, albeit with qualitative and quantitative differences from population to population. For this reason, the mtDNA gene pool of single North African populations may be closer to that of western Eurasian populations. This is well illustrated in the graph of the first two principal components shown in Fig. 3 where North African groups are located at various distances between western Eurasian populations without forming a distinctive cluster. The Eurasian component is the largest contributor to the North African variability, comprising more than half of the lineages described in the region (quadrants I and II of the inset plot, see Table S3 for details).

#### The Eurasian component in North Africa

North African populations show high frequencies of haplogroups of Eurasian origin, that potentially diffused in this area at different times from three main sources: western Europe through the Gibraltar strait, the Middle East via the Sinai Peninsula, and the Mediterranean basin by sea. The five North African populations analysed in our dataset confirmed this trend, with at least 68% of the mtDNA variability represented by Eurasian clades (Fig. 2; Supplementary Table S4). Considering that multiple sources contributed to the diffusion of these lineages in North Africa, the Eurasian component in its totality appears almost equally frequent from west to east (ranging from 62.6 to 71.8%), with the only exception being represented by the Libyan population, with the highest value of 88.7% of Eurasian clades (Fig. 2). However, if we take into consideration one by one the haplogroups that make up the Eurasian component, the scenario changes and we can observe four different trends of distribution: (i) haplogroups with frequencies decreasing from west to east; (ii) haplogroups with frequencies decreasing from east to west; (iii) haplogroups with frequency peaks in central North Africa and decreasing clines at both sides; and (iv) haplogroups equally represented in the whole area.

The first group is the most frequent in our dataset and includes haplogroups mainly diffused in western Europe, such as H (represented by more than 20 sub-clades in our dataset). In particular, H1 and H3 are observed with frequency peaks in western North Africa (Morocco or Algeria) and lower frequencies in the eastern edge of Egypt, except for Libya that, despite being in the centre of North Africa, shows a high percentage of both clades (9.7% and 4.8%, respectively). A similar trend was observed for haplogroups HV0 and V (including V25) that are present in Northwest Africa and not observed in Tunisia (V and V25), Libya, or Egypt (HV0). Haplogroups H (including H1 and H3), HV0 and V reached North Africa mainly via the Strait of Gibraltar, as suggested by their geographical distribution across Eurasia and as already envisioned by previous studies<sup>59–61</sup>.



The second group, with a frequency cline from east towards west, includes haplogroups HV1, with a peak of 2.9% in Egypt and not detected in Morocco, T, with 19.3% in Egypt and then decreasing towards west, and N1 (including both N1a and N1b), which reaches 5.2% in Egypt. Both HV1 and N1 most likely originated in the Near East<sup>62–64</sup> as well as haplogroup T, together with its sister clade J<sup>65,66</sup>, and they all moved towards Europe, North Africa, and even sub-Saharan Africa (as for haplogroup N1<sup>39</sup>) during both pre- and Neolithic periods as well as in more recent times. Haplogroup T includes a sub-clade, namely T1a7, which was previously identified as typical of North African regions<sup>40</sup>. In our comprehensive phylogeny (Supplementary Figure S1), it encompasses eight mitogenomes (six from Egypt<sup>34,67</sup> and two from Algerian Berbers<sup>40</sup>) and represents a direct link between the Near East and North Africa, due to the detection of mitogenomes in the Arabian Peninsula, Israel, Iraq and Lebanon descending from the clade root. Haplogroup X, despite being quite rare in Europe, reaches significant frequencies in North Africa, especially in Egypt, where it is present in the 7.4% of the population, with a decline towards west, with 3.6% in Tunisia and 2.3% in Algeria, but it was not observed in Morocco and shows an intermediate frequency of 1.6% in Libya. Finally, the rare haplogroups I and W are thought to have a similar history to the much more frequent J and T<sup>68</sup> even if haplogroup I is particularly represented in Morocco (2.9%).

The third group includes haplogroup K, which shows its highest frequency in Tunisia (7.2%) and a decline at both eastern and western sides. A similar trend can be observed also for haplogroup U (including all its sub-clades from U1 to U8 with the exception of the North African U6), with a peak of 9.1% in Algeria and a slight decline in the other regions, again with the exception of Libya, showing the highest value of 12.9%, mainly due to an excess of the characteristic European haplogroup U5<sup>69</sup>.

The fourth group of haplogroups, identified by those equally represented in the whole North African region, is not well represented in our dataset, with only haplogroup J ranging from about 3–4.4% in western, central and eastern North Africa. Haplogroup R0a, originated in the Arabian Peninsula, is most frequent in Arabia and the Horn of Africa<sup>70</sup> and is also present in our dataset although it does not show a frequency cline from east to west, as previously reported<sup>70</sup>. Indeed, although it reaches a frequency peak of 6.4% in Libya, it maintains significant occurrences in all the regions (4.9% in Morocco, 3% in Egypt, 2.4% in Tunisia and 0.9% in Algeria).

### The North African component

The North African mtDNA component includes two different types of clades. The first group is represented by haplogroups M1 and U6, which did not originate in North Africa, but are markers of a pre-glacial migration from southwest Asia<sup>23–25,71</sup>. They are both considered to be typical of North Africa, although M1 reaches peaks of frequency in the Horn of Africa and they are also distributed, at lower frequencies, throughout the entire Mediterranean basin. As already mentioned, their Palaeolithic presence in this area is also supported by the classification of ancient mitogenomes from Morocco (dated to 14.2–14.6 kya;<sup>15,18,21</sup>) into these haplogroups (six out of seven are U6 and one is M1, Supplementary Table S2). A basal haplotype of clade U6 was also found in two ancient individuals from Romania, dated back to around 34–33 kya, representing the legacy of the spread of the clade from western Asia into Eastern Europe<sup>72,73</sup> probably the European counterpart of the parallel spread to North Africa. Among the modern North African mitogenomes analysed in this study, haplogroup U6 was observed at frequencies ranging from 1.6 to 8.4% (4.9% in Morocco, 7.1% in Algeria, 8.4% in Tunisia, 1.6% in Libya and 1.5% in Egypt), in full agreement with previous data<sup>24</sup>. As for haplogroup M1, it was observed as M1a in Morocco (2.9%), Algeria (2.3%), Libya (4.8%) and Egypt (3%), and as M1b in Morocco (5.9%) and Algeria (2.3%).

The second type of North African component consists of haplogroups that can be defined as North African-specific because they arose in situ and their distribution remained restricted to North African populations. This group includes three sub-clades of haplogroup H1, namely H1v, H1x and H1w<sup>50</sup> and J2a2b1a\*, a sub-clade within haplogroup J2<sup>40</sup>, all representing rare lineages. To shed light on the origin and dissection of these North African-specific clades, we built a phylogenetic tree encompassing all available mitogenomes (Fig. 4). The list of included mitogenomes is reported in Supplementary Table S6. Within H1 sub-clades, haplogroup H1v is the most frequent with 23 mitogenomes, including ten from Algeria, five from Libya and two from Tunisia. Together with these mitogenomes of confirmed North African origin, the clade also includes one mitogenome of unknown origin and five additional sequences that can be attributed to recent gene flow from North Africa, i.e. two mitogenomes from the Canary Islands, whose population originally moved from North Africa and is still characterised by a strong component from this source<sup>74</sup>, a Spanish subject, an Italian individual of Sephardic Jewish origin and one Basque individual. Haplogroup H1w includes seven mitogenomes, of which three from Libya, three from Italy and an additional sample of unknown origin, while haplogroup H1x encompasses five mitogenomes, three from Libya, one from Algeria and one from Spain. H1v and H1w show very similar coalescence ages (5.4 and 5.1, respectively), while H1x is slightly younger with a coalescence age of 3.9 kya, which largely overlap with the others when error is taken into account. This suggests that concomitant events of in situ differentiation occurred. As for the fourth North African-specific haplogroup, namely J2a2b1a\*, it appears to be even rarer than the other three, comprising only six mitogenomes (two from Algeria and four from the Canary Islands, Morocco, Tunisia and Libya), but older, being dated back to about 7.4 kya.

The geographic distribution of the four North African-specific clades, which covers virtually all North African regions from west to east, does not allow to hypothesise specific places of origin in North Africa, even if haplogroup J2a2b1a\* bears a mitogenome from Libya directly descending from its root with a unique haplotype. Additionally, because North African regions have experienced continuous gene flow from various Eurasian sources at different times, it is difficult to determine the timing and geographical distribution of nearby upstream nodes in the phylogeny and challenging to identify the origin and timing of the differentiation event that occurred in situ. With these considerations in mind, we evaluated the upstream node in the phylogeny, from which the closest external mitogenomes (out of North Africa) radiate. For the three H1 sub-haplogroups, the closest external mitogenomes are all those directly descending from the root of clade H1, thus counting a very

high number of sequences and covering the entire areal of distribution of H1, the most represented H sub-clade across western and Mediterranean Europe. These data suggest that the founder haplotypes of the three H1 sub-clades arrived from Western/Mediterranean Europe and were already present in North Africa by at least 4–5 kya.

As for J2a2b1a\*, the situation is different: only a single mitogenome belonging to an individual from southern Italy (Calabria region) is directly descended from the root of the upstream node defining sub-clade J2a2b1a, suggesting a direct link between the Mediterranean basin and this latter North African specific clade. However, for the reasons explained above, we cannot exclude that both the Italian mitogenome and the North African-specific clade J2a2b1a\* share a common ancestor that was already present in North Africa or the Middle East before their differentiation (at least 7 kya), since both these regions are the home of the mitogenomes descending from the more upstream nodes of J2a2b1. Furthermore, as a general observation, the four North African-specific sub-haplogroups, except for H1v, encompass a relatively small number of mitogenomes, and, in the case of H1w, they are often characterised by the same haplotype, so that their estimated ages should be treated with some caution.

### The sub-Saharan African component in North Africa

The sub-Saharan contribution, defined by the sub-Saharan L lineages, was observed in all North African regions (Fig. 2, Supplementary Table S4), at different frequencies, without a specific trend in a north-south or east-west direction. Indeed, macro-haplogroup L reaches frequencies of 26.5% in Tunisia, 23.7% in Egypt, 19.7% in Algeria, and 15.7% in Morocco. Macro-haplogroup L includes five clades in North African populations, i.e. haplogroups L0, L1, L2, L3 (excluding haplogroups M and N) and L4 (Fig. 2), with L2 and L3 being the most frequent (as previously reported by<sup>75</sup>). Haplogroup L2 is widely distributed across North Africa (12% in Tunisia, around 5–6% in Algeria and Egypt and around 3% in Libya and Morocco), as already observed<sup>30,76</sup>. Haplogroups L3 and L1 (present in Algeria, Egypt, Morocco, and Tunisia with frequency ranges of around 8–10% and 4–5%, respectively) are also common in Libya (as already observed in<sup>31</sup> although absent in our sample). As expected, due to their sub-Saharan origin, all the five lineages within macro-haplogroup L mentioned above (Fig. 1; Supplementary Figure S1) encompass mitogenomes not only from North Africa, but also from the sub-Saharan African countries included in this study (i.e. Burkina Faso, Cameroon, Chad, Niger, and Nigeria). Within haplogroup L1, L1b1a6 (Supplementary Figure S2, Table S7), defined by two transitions at nps 9755 and 14,110, encompasses ten mitogenomes in our phylogeny, most of which are from west-central Africa (two from Burkina Faso, one from Nigeria and four from Cameroon), two from Morocco and one from Tunisia. This clade has been considered an example of direct input from the African continent to Europe/Iberia via North Africa<sup>77,78</sup>. We re-evaluated this hypothesis in light of our new data, building a phylogenetic tree with all available L1b1a6 mitogenomes (Supplementary Figure S2 and Table S7). We confirmed that haplogroup L1b1a6, dated back to almost 14 Kya, is distributed in western Africa, including mitogenomes from western sub-Saharan countries (i.e. Burkina Faso, Cameroon, Guinea-Bissau, and Nigeria) and from Morocco and Tunisia in North Africa. The European mitogenomes contained in clade L1b1a6 do not form a separate subclade, but they all descend from a node dated back to around 13 Kya, thus setting this date as the upper limit to the initial migration event, similarly to what previously proposed by Hernandez et al.<sup>78</sup>.

### Mitogenome variation in adjacent sub-Saharan regions

The high genetic variability of North Africans, measured in terms of number of haplogroups observed, is particularly evident when comparing North African mitogenomes with sub-Saharan sequences (from Burkina Faso, Niger, Nigeria, Cameroon and Chad) (Fig. 2, Supplementary Table S2). Indeed, most of the subjects from these countries belong to macro-haplogroup L, with few exceptions represented by one Nigerian mitogenome belonging to haplogroup U6b, two mitogenomes from Niger belonging to two distinct sub-haplogroups within U6a (U6a3b1 and U6a5a) and one U5b1b1b sample identified in a Fulbe from Cameroon. As mentioned above, haplogroup U6 is a typical North African clade, with frequency peaks in the north-western part of North Africa, while the U5b1b1b clade was already identified in Berber populations, as a marker of diffusion events from the Iberian Peninsula<sup>69</sup>.

### Mitogenome variation in ancient North Africa

Haplogroup distribution was also investigated for comparison in a dataset of 43 ancient published mitogenomes from North African populations (Fig. 2, Supplementary Tables S2 and S5). The ancient samples are from four countries (Morocco, Tunisia, Libya, and Egypt), spanning a time frame of about 13 ky.

The most ancient mitogenomes, represented by seven samples, are from the Iberomaurusian site of Taforalt in Morocco, which date back to 15 kya and constitute the oldest North African DNAs<sup>18</sup>. One mitogenome belongs to haplogroup M1b, while the remaining six classify into different U6a sub-clades, thus confirming the Palaeolithic origin of the two typical North African haplogroups M1 and U6 (Fig. 2). Nineteen additional unrelated ancient mitogenomes are available from Moroccan archaeological sites, from Epipalaeolithic and Neolithic cultures, with dates ranging from ~7 kya to ~4 kya<sup>15,21</sup>. Eight U6 (six of them U6a) and one M1a mitogenomes were detected in Epipalaeolithic, Early Neolithic and two (out of three) Middle Neolithic samples, while the remaining mitogenomes including all the Late Neolithic ones belong to both European and Near Eastern clades, specifically HV0, J1c, J2a, K1a, T2b3, U5b, and X2b. These overall data confirm the presence of an ancient North African genomic component with an ultimate source in Upper Palaeolithic Eurasia, as previously suggested by modern dataset<sup>23,24</sup> (albeit possibly via several waves<sup>71</sup>). This input was followed by Neolithic gene flow from western Europe and subsequent flow from the Middle East starting in the Middle Neolithic<sup>15,21</sup>. These details are supported more generally by genome-wide analyses<sup>15,21</sup>.

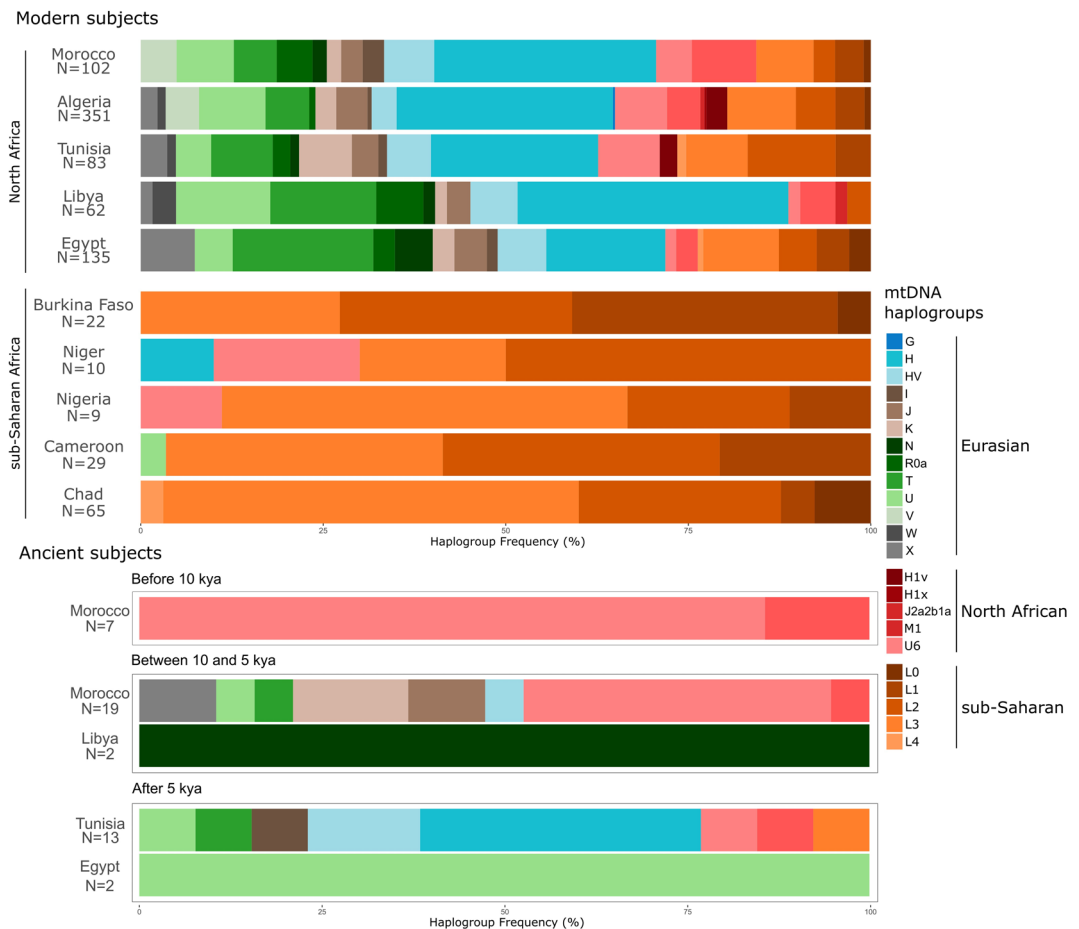
From a Neolithic site in Libya's central Sahara (dated to ~7 kya), two mitogenomes descending from haplogroup N were identified<sup>64</sup>. All the remaining 15 ancient individuals available from North African regions



**Fig. 1.** Schematic phylogenetic tree of modern ( $N=733$ ) and ancient ( $N=43$ ) mitogenomes from North Africa. Grey triangles correspond to different haplogroups and sub-haplogroups and their width is proportional to the number of included samples. Ancient subjects are indicated in bold italic and are flanked with a circular marker (see Supplementary Figure S1 for details).

date back to more recent times (from 4 kya to 1.3 kya), and were collected from Tunisia (from two Phoenician sites)<sup>20,79</sup> and Egypt<sup>80–82</sup>. Notably, among these complete ancient genomes, only three from the Phoenician site of Kerkouane belong to haplogroups M1, U6 and L3e<sup>20</sup> while all the others classify within haplogroups of Eurasian origin (H, HV, I5, T2, and U) (Fig. 2).

However, it is important to note that of the 43 available ancient samples in North Africa, 39 are from the western regions (26 from Morocco and 13 from Tunisia), with just two samples available from Libya and two from Egypt. As mentioned in the introduction, climatic conditions in Africa are generally not favourable for the preservation of ancient DNA, and these studies are generally affected by mosaic sampling. Expanding the sampling to eastern regions of North Africa would therefore be mandatory in order to obtain a broad picture of the ancient genomics of the region.



**Fig. 2.** Barplots illustrating macro-haplogroup/haplogroup frequencies (as reported in Supplementary Table S4) in North African (Morocco, Algeria, Tunisia, Libya, and Egypt from west to east) and sub-Saharan (Burkina Faso, Niger, Nigeria, Cameroon, and Chad from west to east) modern and ancient mitogenomes analysed in this study. Ancient subjects are clustered into three groups based on their age estimates (before 10 kya, between 10 and 5 kya and after 5 kya). Haplogroups on the right are grouped according to their geographic origin/distribution.

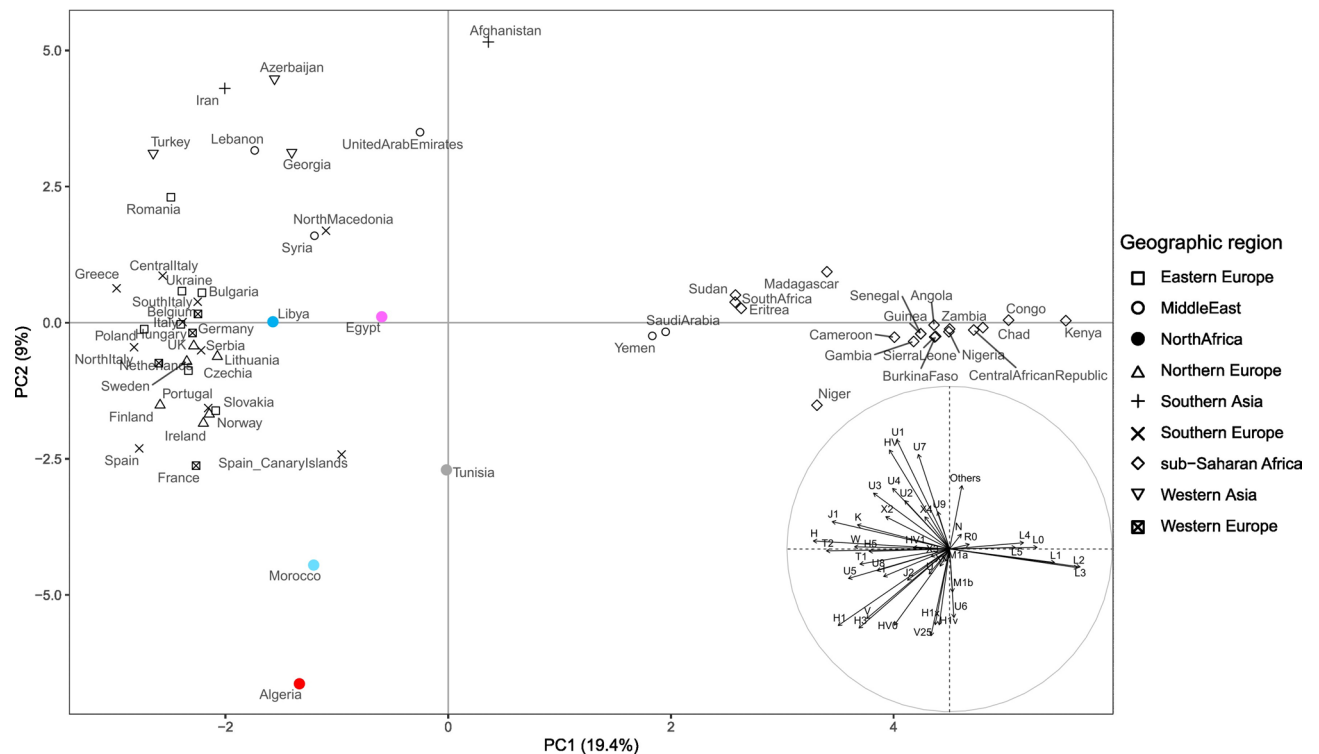
### The mitogenome links between North Africa and Sardinia

In 2017, a study of mitogenomes from Sardinia identified a large number of autochthonous sub-haplogroups that arose in situ, in some cases shortly after the first peopling event of the island and were defined as Sardinian-specific since their geographic distribution was apparently restricted to the Mediterranean island<sup>49</sup>. By comparing the phylogeny of North African regions with that based on Sardinian-specific haplogroups, we identified a connection between North Africa and Sardinia involving at least four sub-clades. Indeed, the haplogroups H1cn1a, H3aw, H3ay1, and J1c3h, previously defined as Sardinian-specific, now include not only Sardinian mitogenomes but also one or few mitogenomes from our North African dataset.

These North African mitogenomes most likely reveal temporally distinct events of female-mediated gene flow from Sardinia to different regions of North Africa. In detail, the Sardinian-specific haplogroup H1cn1a includes also one mitogenome from Morocco that differs from other Sardinian mitogenomes by only one mutation (np 16162) in the control region. This most likely represents a very recent connection between the two regions, dated back to only a few hundred years, given the low molecular divergence accumulated within the clade. The H3aw haplogroup, a subclade of H3 defined by the transition at np 15,289, includes three identical mitogenomes from Algerian Berbers harbouring the control-region transition at np 194 from the H3aw root. This connection could possibly be further back in time, as the haplotype found in the three Berbers departs from the root of H3aw that had been dated to around 3000 years ago<sup>49</sup>. The Sardinian-specific haplogroup H3ay1 shows a mitogenome from Libya with a haplotype corresponding to the clade root itself and identical to a sample from southern Sardinia. This too might be a rather ancient link, as the clade root was dated to around 3400 years ago. However, in both cases, recent events of gene flow cannot be excluded. Finally, an Algerian Berber has a mitogenome descending from the J1c3h haplogroup root, dated 5400 years ago<sup>49</sup> with three mutations, the reversion of a transition in the control region and two transitions in the coding region.

Another interesting case concerns the H3u haplogroup, which was not classified as Sardinian-specific, but contains only Sardinian and North African samples, including three mitogenomes in our phylogeny (from





**Fig. 3.** PCA plot showing the genetic variation of North Africa in the context of neighbouring sub-Saharan African and Western Eurasian landscapes. Data from sub-Saharan Africans are from the present study, while the Eurasian dataset includes over 18 thousand sequences available in public databases (Supplementary Table S3). The plot is based on modern mtDNA haplogroup frequencies from both this study and the literature. Reference populations are indicated by the country name, followed by a symbol to represent their macro-area. North African populations analysed in this study are highlighted in different colours: red (Algeria); pink (Egypt); blue (Libya); light blue (Morocco) and grey (Tunisia). The inset shows haplogroup contributions to PC1 and PC2 (accounting for 19.4% and 9.0% of the total variance, respectively).

Algeria, Libya and Morocco). It is worth noting that an ancient Sardinian sample<sup>49</sup> dated back to the Final Bronze Age (4000 years ago) and from the south of the island, belongs to the same clade and shares the mutation at np 16,240 present in all three North African mitogenomes of this study, again indicating a link between the two regions that most likely goes back several millennia. Finally, as previously noted<sup>40</sup> two mitogenomes from Algeria share the same W5 haplotype with a Punic-era Sardinian individual from the late 5th century<sup>83</sup>.

On the other hand, if we consider the possible connections between North African regions, Sardinia and one or more Mediterranean regions, the examples are too numerous to list. To give just one example among many, haplogroup H1e1a6 includes two mitogenomes of Moroccan origin, produced in this work, together with an ancient Sardinian sample from the Punic era<sup>83</sup> and modern Spanish mitogenomes<sup>84,85</sup>.

### The demography of North African populations over time and a signal from the last green Sahara

In order to outline the complexity of mtDNA sequence variability in North African populations, observed in both ancient and modern times, we performed a Bayesian analysis. The Bayesian Skyline Plot (BSP) in Fig. 5 shows changes in the effective population size over time for mtDNA lineages identified in North African populations. The same analysis was performed by using two different clock rates<sup>53,54</sup> and the results are virtually overlapping. As discussed previously, North African maternal genetic components can be divided into three groups of distinct geographic origin, i.e. sub-Saharan, North African typical/autochthonous and Eurasian (as illustrated in Supplementary Tables S4 and S5). These three components trace their origins to different places and different times. For this reason, we performed coalescent Bayesian skyline analyses employing the entire dataset of modern ( $N=733$ ) and ancient ( $N=43$ ) mitogenomes from North African populations, partitioned according to these three components (Fig. 5).

Each component reveals a distinct major population growth event. The expansion of the Eurasian component (blue line) begins in pre-Neolithic times, at ~17.5 kya, corresponding to the retreat of the glaciation in Europe. This result indicates that along with the Neolithic impact that brought Eurasian lineages into North Africa, as also attested by the available genomic and ancient DNA data<sup>4,15,16,86</sup> a pre-Neolithic expansion also left traces in the Eurasian mtDNA variability in North African populations, although this expansion most likely took place in Europe or western Asia before the arrival in North Africa. Considering the age and geographical distribution of the haplogroups, it is likely that, for example, U5b1b arrived from the Strait of Gibraltar, while others may

have arrived from the Sinai or the Mediterranean. The autochthonous/typical North African component (red line), by contrast, indicates an expansion beginning around 35 kya, from populations in the Middle East, as also previously hypothesised by mitochondrial and nuclear ancient and modern DNA data<sup>18,24,87,88</sup>. The discovery of putatively haplogroup pre-N lineages in early Holocene southwest Libya<sup>64,89</sup> which lack the N-defining 8701 A variant similar to Oase-1 from Palaeolithic Romania, but carry the N-defining 9540T that Oase-1 also lacks<sup>73</sup>, may point to the same process. On the other hand, the 60% Natufian-related ancestry in the Libyan samples may suggest a more recent wave of Middle Eastern ancestry.

The sub-Saharan component (orange line), by contrast with both, shows a primary event of expansion between 15 and 10 kya. This time span overlaps with the last Green Saharan period, which occurred 15 to 5 kya and was characterised by human settlements throughout the Sahara. Y-chromosome and genomic data supported the scenario of trans-Saharan movements towards North Africa over a time span of ~ 12 – 5 kya mediated by a ghost population they called the Green Sahara Population<sup>90,91</sup>. As for the female-mediated mtDNA, a signal of gene flow across the Sahel belt during the Holocene was proposed from founder analyses<sup>92</sup> on the HVS-I variability in African populations. In particular, these previous data indicated a signal of expansion from sub-Saharan Africa towards North Africa with a peak at around 6.5 Kya, attributable mostly to haplogroups L1b (L1b1a in particular), L3e5, and L0a1<sup>92</sup>, albeit potentially biased by primarily haplogroup-based sampling at the time.

Here, BSP analysis with a much larger dataset, based on both modern and ancient mitogenomes from North Africa provides further evidence that the Green Sahara may have played an important role in shaping the genetic landscape of North African populations also from a female-mediated perspective.

This signal is present in sub-haplogroups of all sub-Saharan lineages (L0 to L5), suggesting that the sources from which these movements originated may be multiple. It is worth mentioning that the limited amount of genomic data available from both modern and ancient populations from sub-Saharan Africa makes it difficult to identify and date the specific lineages that might be responsible for the observed signal. However, we can hypothesize a few possible examples. Haplogroup L2a1 is observed in modern subjects from both sides of the Sahara Desert and in ancient Sudanese<sup>19</sup> but also in ancient samples from Andalusia dated 3.5 kya<sup>93</sup> and in two samples from Sardinia (2.5–1.7 kya)<sup>94</sup> thus suggesting that this lineage arrived in Europe prior to 3.5 kya and may have crossed the Mediterranean basin at a later date. The distribution of the L2a1 clade in modern North Africans is widespread throughout the region, suggesting that multiple events of gene flow occurred across the Sahara during the period of favourable climatic conditions in the Green Sahara. Similarly, as previously discussed (Supplementary Figure S2, Supplementary Table S7) haplogroup L1b1a6 includes sub-Saharan and North African, as well as European, mitogenomes descended from the same node with an upper limit set by the age of the L1b1a6 node at almost 14 thousand years ago.

Future analyses may increase the number of proxy populations from sub-Saharan Africa, providing a more accurate estimate of the times and routes of contact that shaped the formation of modern North African population groups.

## Conclusion

Since the 1990s, uniparental markers, i.e. the maternally-inherited mtDNA and the male-specific non recombining portion of the Y chromosome (MSY) have provided informative data for population geneticists to trace ancient human migrations at both macro- and micro-geographic levels<sup>32,35,95,96</sup>. In particular, mitochondrial DNA, with its small molecular size and high evolutionary rate, has long been a key genetic tool in population studies. It is important to emphasise that, compared to nuclear genes, mtDNA does not reflect the full complexity of past demographic processes. It is a single, maternally-inherited locus that is particularly susceptible to genetic drift due to its reduced effective population size. Nevertheless, this non-recombining genetic system can provide high phylogenetic and phylogeographic resolution<sup>32</sup>, making the mtDNA data an informative complement to nuclear genome data in the genomic era. In particular, large-scale surveys of complete mitogenomes allowed unprecedented resolution of the mtDNA phylogeny (in terms of haplogroup and sub-haplogroup definitions). Phylogenetic and phylogeographic approaches have allowed accurate dating and calibration of the mtDNA molecular clock<sup>53,97</sup>, useful also as a stand-point for dating demographic events detected through genomic data<sup>98</sup>.

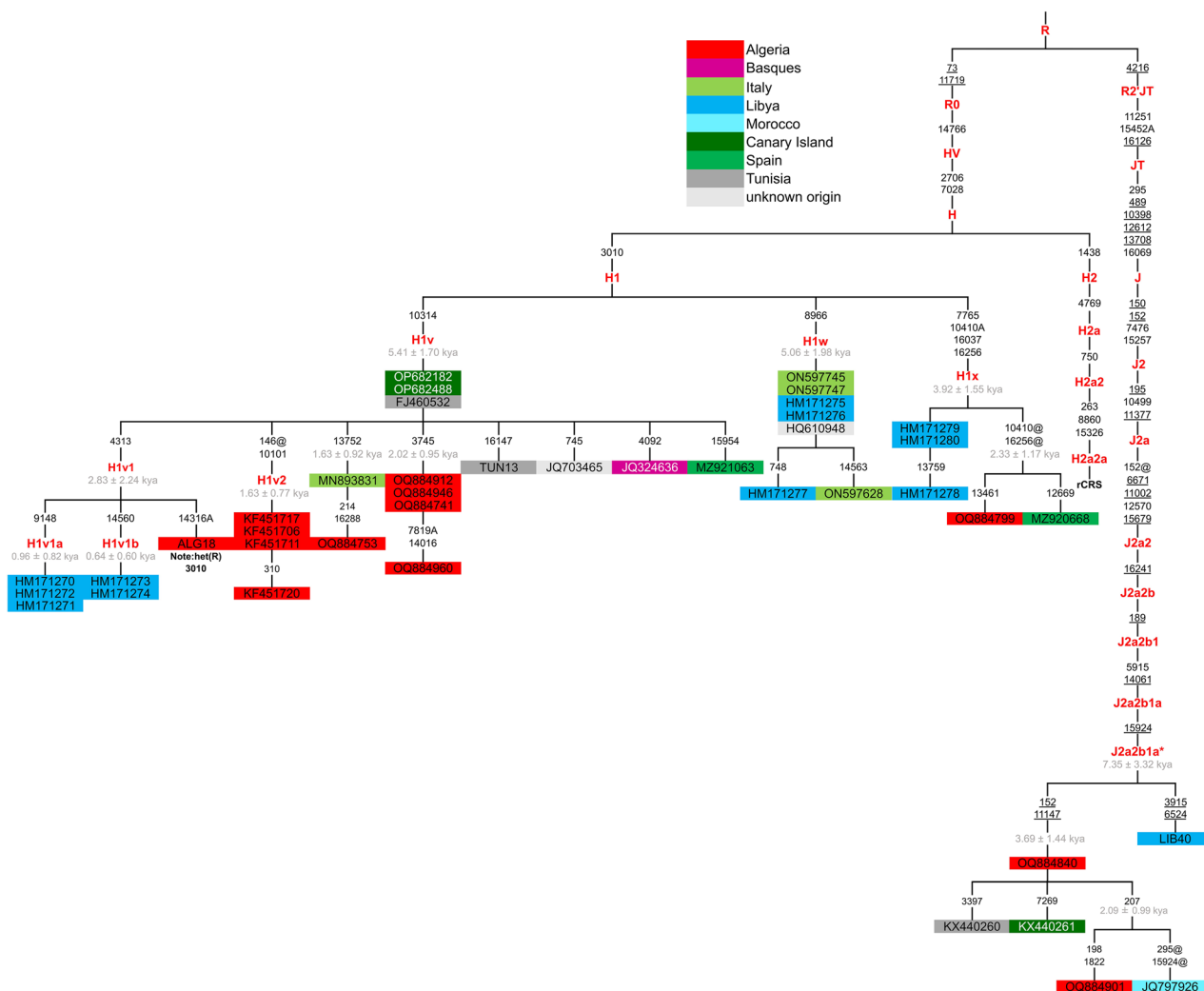
After an initial phase focusing on the mitogenomes (or portion of mtDNA) variability of specific phylogenetic lineages and assessing it on their entire distribution range, population mitogenomics moved towards the analysis of populations living in restricted geographical areas, through fine-scale and massive surveys of mitogenomes, in order to identify all the specific genetic components and lineages characteristic of that population<sup>49</sup>. One geographical region whose mitogenomic variability had never been studied using a population-based approach is North Africa. North Africa occupies a unique geographical position in the southern coast of the Mediterranean basin, has another geographical barrier to the south in the form of the Sahara Desert, and has two main routes of communication with the Middle East to the east via the Sinai Peninsula and with Europe to the west via the Straits of Gibraltar. For this reason, studying the genetic variability of this geographical region means looking for signs of admixture with multiple peoples who have contributed to its gene pool over the millennia.

In this paper we have analysed and discussed a dataset of 733 modern (238 of which new) and 43 ancient mitogenomes from Morocco, Algeria, Tunisia, Libya and Egypt, in comparison with 17,274 modern mitogenomes, 11,438 from Eurasia and 5,836 from sub-Saharan African countries (135 of which new). Our analyses, based on a phylogeographic approach, comparisons with population (and non-population) data on both modern and ancient mitogenomes, and demographic investigation using a Bayesian approach, in the light also of the available genomic data, suggest that the mitogenome variability in North African populations (i) can be traced back to three main origins - Eurasia, North Africa and Sub-Saharan Africa - and (ii) has been defined during four key time periods. The three components mainly include haplogroups that arrived in North Africa through movements and migrations, from different sources and at different times, as well as a small proportion of haplogroups that differentiated in situ, also at different times. As shown in Fig. 6, the first key events defining the

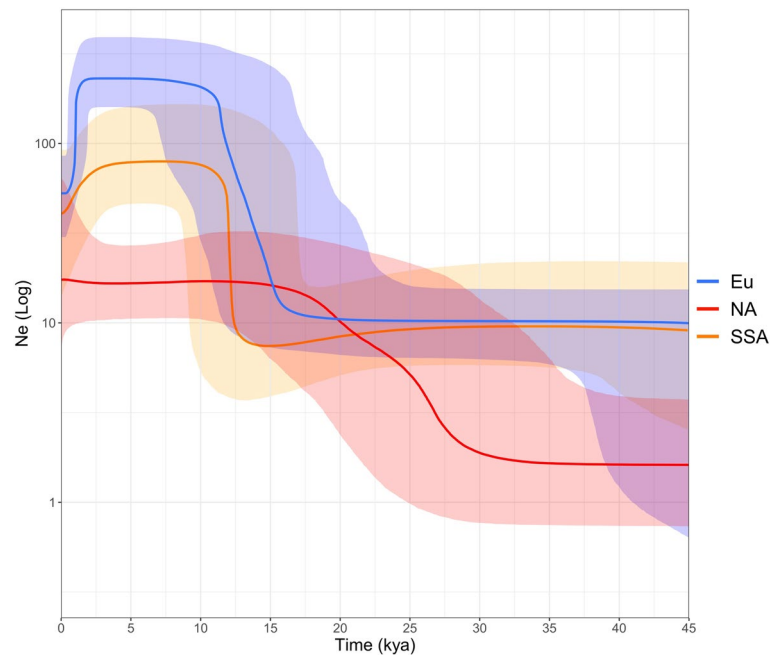
North African genetic pool occurred in the Palaeolithic, with the arrival from western Eurasia of a characteristic North African component characterised by haplogroups M1 and U6.

Two major routes, through the Strait of Gibraltar and Sinai, were followed during the emergence of Neolithic culture, which brought numerous subclades from both west and east as supported by genomic data from both modern and ancient samples<sup>4,15,16,21,86,91</sup>. The typically sub-Saharan African component (L sub-clades) found in North African populations seems to have come from the south mainly during the favourable climatic period of the Green Sahara. This corridor of human movements, involving a very large area and probably several population sources, has also been identified by modern genomic and Y-chromosome data<sup>90,92</sup>; the data presented here emphasises that it appears to have been not only a male-mediated event, but also left traces in the maternally inherited mtDNA. Once they reached North Africa, some sub-lineages of haplogroup L, such as L1b1a6, entered Europe via the Iberian Peninsula (as suggested by<sup>78</sup>).

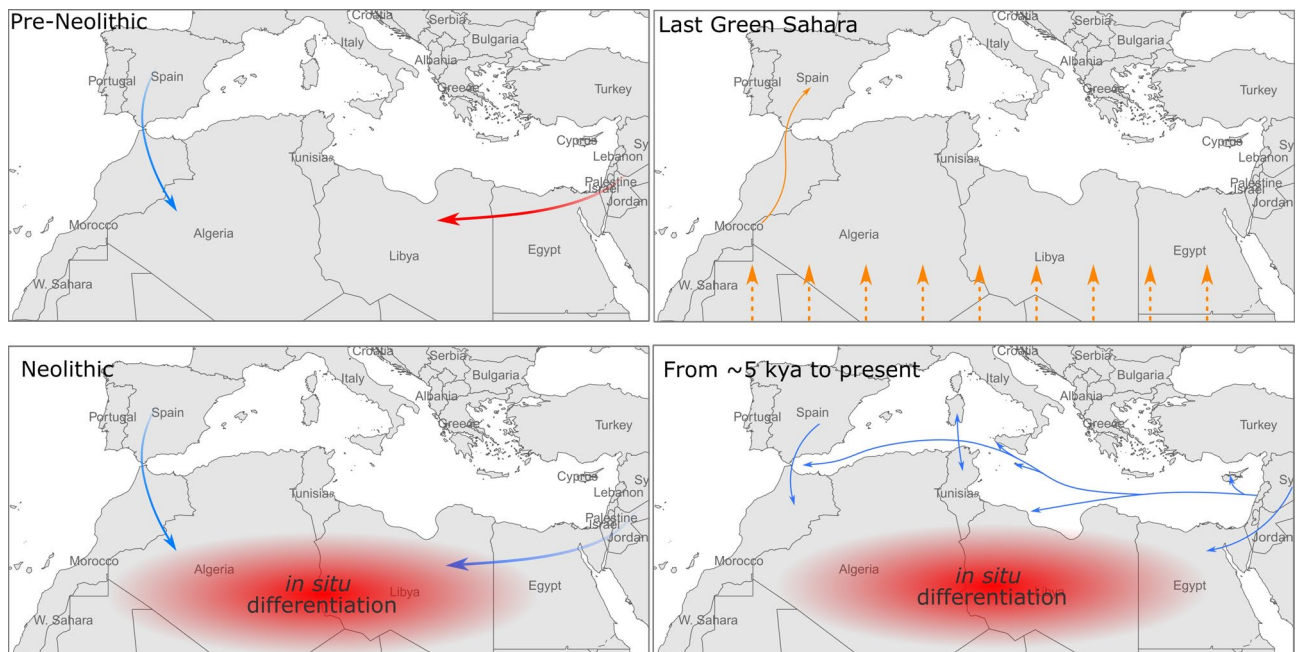
In more recent times, North Africa has been involved in a myriad of exchanges with peoples from neighbouring regions and the wider Mediterranean basin. Figure 6 shows just a few of the many possible exchange routes and indicates a period from about 7.5 kya, corresponding to the coalescence ages of J2a2b1a\*, which differentiated in situ in North Africa<sup>40</sup>. Other in situ differentiation events occurred in more recent times, starting from about 5 kya (Fig. 4), giving rise to the North African autochthonous haplogroups H1v, H1w and H1x<sup>50</sup>. From the Bronze Age until a few centuries ago, trade and exchange between peoples strongly contributed to North African mtDNA variability. Our data show matrilineal links between Sardinia and North Africa, and between these two regions and others, such as Spain, in the Mediterranean basin. Trade between the Iberian Peninsula and North Africa has also been constant over time<sup>99</sup>. Furthermore, previous studies<sup>63,83</sup> have reported how trade and/or invasions by peoples such as the Phoenicians and the Vandals left extensive genetic traces, supporting the hypothesis of extensive female movements throughout the Mediterranean, although it is not yet clear what social dynamics were involved in this process. Other routes, such as the Arab invasion, particularly from the 7th century AD onwards, have been suggested from the spread of Y-chromosome haplogroups<sup>100</sup> and are likely to



**Fig. 4.** Phylogenetic tree of the North African-specific haplogroups H1v, H1x, H1w, and J2a2b1a\* (included mitogenomes are listed in Supplementary Table S6).



**Fig. 5.** Bayesian Skyline Plot showing the effective population size trends ( $N_e$ ) of haplogroups found in the dataset of 733 modern and 43 ancient mitogenomes from North Africa. Haplogroups were divided into three groups, according to the three genetic components (as reported in Supplementary Tables S4 and S5) contributing to the North African gene pool, Eu (Eurasian), NA (North African) and SSA (sub-Saharan African). Solid lines are the median estimates obtained by employing the mutation rates proposed by Soares et al.<sup>53</sup>; the shadings show the highest posterior density limits.



**Fig. 6.** Representation of the major movements of the three maternal genetic components reported in this study. The colours of the arrows are in agreement with the three genetic components identified (Eurasian -blue-, sub-Saharan -orange- and North African -red-). The underlying map was built using the R package *rnatuarearth* v.1.1.0 (<https://docs.ropensci.org/rnatuarearth/101>).



have significantly reshuffled and modified the genetic variability of North African populations, although a larger dataset, possibly including more Berber populations, should be analysed to provide evidence from a matrilineal perspective. For a clearer picture, it will be necessary to further increase the number of both modern and ancient genomes from North Africa, which is still poorly represented in terms of sequenced genomes in comparison with other parts of the world.

### Data availability

The sequence data of the paper “The origin of modern North Africans as depicted by a massive survey of mitogenomes” are available in GenBank with the accession numbers (PV621465 - PV621837) and can be accessed via the following link: <https://www.ncbi.nlm.nih.gov>.

Received: 13 May 2025; Accepted: 15 July 2025

Published online: 25 July 2025

### References

- Hublin, J. J. et al. New fossils from Jebel Irhoud, Morocco and the pan-African origin of Homo sapiens. *Nature* **546**, 289–292 (2017).
- Hunt, C. et al. Site formation processes in caves: the holocene sediments of the Haua Fteah, Cyrenaica, Libya. *J. Archaeol. Sci.* **37**, 1600–1611 (2010).
- Barton, R. N. E. et al. Origins of the Iberomaurusian in NW Africa: new AMS radiocarbon dating of the middle and later stone age deposits at Taforalt cave, Morocco. *J. Hum. Evol.* **65**, 266–281 (2013).
- Arauna, L. R. et al. Recent historical migrations have shaped the gene pool of Arabs and Berbers in North Africa. *Mol. Biol. Evol.* **34**, 318–329 (2017).
- Cavalli-Sforza, L. L., Cavalli-Sforza, L., Menozzi, P. & Piazza, A. *The History and Geography of Human Genes* (Princeton University Press, 1994).
- Fregel, R. Paleogenomics of the Neolithic transition in North Africa. In *Africa, the Cradle of Human Diversity: Cultural and Biological Approaches To Uncover African Diversity* Brill, Leiden, 213–235 (2022).
- Newman, J. L. *The Peopling of Africa: A Geographic Interpretation* (Yale University Press, 1995).
- Cavalli-Sforza, L. L. & Piazza, A. Human genomic diversity in Europe: a summary of recent research and prospects for the future. *Eur. J. Hum. Genet.* **1**, 3–18 (1993).
- Bosch, E. et al. Population history of North Africa: evidence from classical genetic markers. *Hum. Biol.* **69**, 295–311 (1997).
- Brooks, N. et al. The climate-environment-society nexus in the Sahara from prehistoric times to the present day. *J. North. Afr. Stud.* **10**, 253–292 (2005).
- Drake, N. A., Blench, R. M., Armitage, S. J., Bristow, C. S. & White, K. H. Ancient watercourses and biogeography of the Sahara explain the peopling of the desert. *Proc. Natl. Acad. Sci.* **108**, 458–462 (2011).
- Tierney, J. E. & deMenocal, P. B. Abrupt shifts in Horn of Africa hydroclimate since the last glacial maximum. *Science* **342**, 843–846 (2013).
- Sereno, P. C. et al. Lakeside cemeteries in the Sahara: 5000 years of holocene population and environmental change. *PLoS One* **3**, e2995 (2008).
- Almathen, F. et al. Ancient and modern DNA reveal dynamics of domestication and cross-continental dispersal of the dromedary. *Proc. Natl. Acad. Sci.* **113**, 6707–6712 (2016).
- Fregel, R. et al. Ancient genomes from North Africa evidence prehistoric migrations to the Maghreb from both the Levant and Europe. *Proc. Natl. Acad. Sci.* **115**, 6774–6779 (2018).
- Serra-Vidal, G. et al. Heterogeneity in palaeolithic population continuity and neolithic expansion in North Africa. *Curr. Biol.* **29**, 3953–3959.e4 (2019).
- Lucas-Sánchez, M., Serradell, J. M. & Comas, D. Population history of North Africa based on modern and ancient genomes. *Hum. Mol. Genet.* **30**, R17–R23 (2021).
- van de Loosdrecht, M. et al. Pleistocene North African genomes link near Eastern and sub-Saharan African human populations. *Science* **360**, 548–552 (2018).
- Sirak, K. A. et al. Social stratification without genetic differentiation at the site of Kulubnarti in Christian period Nubia. *Nat. Commun.* **12**, 7283 (2021).
- Moots, H. M. et al. A genetic history of continuity and mobility in the Iron age central Mediterranean. *Nat. Ecol. Evol.* **7**, 1515–1524 (2023).
- Simões, L. G. et al. Northwest African Neolithic initiated by migrants from Iberia and Levant. *Nature* **618**, 550–556 (2023).
- Henn, B. M. et al. Genomic ancestry of North Africans supports back-to-Africa migrations. *PLoS Genet.* **8**, e1002397 (2012).
- Macaulay, V. et al. The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am. J. Hum. Genet.* **64**, 232–249 (1999).
- Olivieri, A. et al. The mtDNA legacy of the Levantine Early Upper Palaeolithic in Africa. *Science* **314**, 1767–1770 (2006).
- González, A. M. et al. Mitochondrial lineage M1 traces an early human backflow to Africa. *BMC Genom.* **8**, 223 (2007).
- Lipson, M. et al. High continuity of forager ancestry in the Neolithic period of the Eastern Maghreb. *Nature* **641**, 925–931 (2025).
- Rodríguez-Varela, R. et al. Genomic analyses of pre-European conquest human remains from the Canary Islands reveal close affinity to modern North Africans. *Curr. Biol.* **27**, 3396–3402.e5 (2017).
- Brakez, Z. et al. Human mitochondrial DNA sequence variation in the Moroccan population of the Souss area. *Ann. Hum. Biol.* **28**, 295–307 (2001).
- Fadhlaoui-Zid, K. et al. Mitochondrial DNA heterogeneity in Tunisian Berbers. *Ann. Hum. Genet.* **68**, 222–233 (2004).
- Coudray, C. et al. The complex and diversified mitochondrial gene pool of Berber populations. *Ann. Hum. Genet.* **73**, 196–214 (2009).
- Fadhlaoui-Zid, K. et al. Mitochondrial DNA structure in North Africa reveals a genetic discontinuity in the Nile Valley. *Am. J. Phys. Anthropol.* **145**, 107–117 (2011).
- Torroni, A., Achilli, A., Macaulay, V., Richards, M. & Bandelt, H. J. Harvesting the fruit of the human mtDNA tree. *Trends Genet.* **22**, 339–345 (2006).
- Costa, M. D. et al. Data from complete mtDNA sequencing of Tunisian centenarians: testing haplogroup association and the golden mean to longevity. *Mech. Ageing Dev.* **130**, 222–226 (2009).
- Kujanová, M., Pereira, L., Fernandes, V., Pereira, J. B. & Černý, V. Near Eastern Neolithic genetic input in a small Oasis of the Egyptian Western desert. *Am. J. Phys. Anthropol.* **140**, 336–346 (2009).
- Behar, D. M. et al. A Copernican reassessment of the human mitochondrial DNA tree from its root. *Am. J. Hum. Genet.* **90**, 675–684 (2012).

36. Lippold, S. et al. Human paternal and maternal demographic histories: insights from high-resolution Y chromosome and MtDNA sequences. *Investig. Genet.* **5**, 13 (2014).
37. Fernandes, V. et al. Genetic stratigraphy of key demographic events in Arabia. *PLOS ONE*. **10**, e0118625 (2015).
38. Font-Porterias, N. et al. The genetic landscape of mediterranean North African populations through complete MtDNA sequences. *Ann. Hum. Biol.* **45**, 98–104 (2018).
39. Diallo, M. Y. et al. Circum-Saharan prehistory through the lens of MtDNA diversity. *Genes* **13**, 533 (2022).
40. Aizpurua-Iraola, J., Abdeli, A., Benhassine, T., Calafell, F. & Comas, D. Whole mitogenomes reveal that NW Africa has acted both as a source and a destination for multiple human movements. *Sci. Rep.* **13**, 10395 (2023).
41. Brandini, S. et al. The Paleo-Indian entry into South America according to mitogenomes. *Mol. Biol. Evol.* **35**, 299–311 (2018).
42. Andrews, R. M. et al. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23**, 147–147 (1999).
43. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2010).
44. Li, H. et al. The sequence alignment/map format and samtools. *Bioinformatics* **25**, 2078–2079 (2009).
45. McKenna, A. et al. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
46. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).
47. Weissensteiner, H. et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **44**, W58–W63 (2016).
48. van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30**, E386–E394 (2009).
49. Olivieri, A. et al. Mitogenome diversity in Sardinians: a genetic window onto an island's past. *Mol. Biol. Evol.* **34**, 1230–1239 (2017).
50. Ottoni, C. et al. Mitochondrial haplogroup H1 in North africa: an Early Holocene arrival from Iberia. *PLoS One*. **5**, e13378 (2010).
51. Drummond, A. J. & Rambaut, A. BEAST: bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).
52. Bouckaert, R. et al. BEAST 2.5: an advanced software platform for bayesian evolutionary analysis. *PLoS Comput. Biol.* **15**, e1006650 (2019).
53. Soares, P. et al. Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am. J. Hum. Genet.* **84**, 740–759 (2009).
54. Posth, C. et al. Pleistocene mitochondrial genomes suggest a single major dispersal of non-Africans and a late glacial population turnover in Europe. *Curr. Biol.* **26**, 827–833 (2016).
55. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior summarization in bayesian phylogenetics using tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).
56. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016).
57. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (2024).
58. Kassambara, A. & Mundt, F. Factoextra: extract and visualize the results of multivariate data analyses. R package version 1.0.7. (2020).
59. Torroni, A. et al. MtDNA analysis reveals a major late paleolithic population expansion from Southwestern to Northeastern Europe. *Am. J. Hum. Genet.* **62**, 1137–1152 (1998).
60. Torroni, A. et al. A signal, from human mtdna, of postglacial recolonization in Europe. *Am. J. Hum. Genet.* **69**, 844–852 (2001).
61. Achilli, A. et al. The molecular dissection of MtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Am. J. Hum. Genet.* **75**, 910–918 (2004).
62. Fernandes, V. et al. The Arabian cradle: mitochondrial relicts of the first steps along the Southern route out of Africa. *Am. J. Hum. Genet.* **90**, 347–355 (2012).
63. Shamooun-Pour, M., Li, M. & Merriwether, D. A. Rare human mitochondrial HV lineages spread from the near East and Caucasus during post-LGM and neolithic expansions. *Sci. Rep.* **9**, 14751 (2019).
64. Vai, S. et al. Ancestral mitochondrial N lineage from the neolithic 'green' Sahara. *Sci. Rep.* **9**, 3530 (2019).
65. Pala, M. et al. Mitochondrial DNA signals of late glacial recolonization of Europe from near Eastern refugia. *Am. J. Hum. Genet.* **90**, 915–924 (2012).
66. Pereira, J. B. et al. Reconciling evidence from ancient and contemporary genomes: a major source for the European Neolithic within Mediterranean Europe. *Proc. Biol. Sci.* **284**, 20161976 (2017).
67. Pagani, L. et al. Tracing the route of modern humans out of Africa by using 225 human genome sequences from Ethiopians and Egyptians. *Am. J. Hum. Genet.* **96**, 986–991 (2015).
68. Olivieri, A. et al. Mitogenomes from two uncommon haplogroups mark late glacial/postglacial expansions from the near East and neolithic dispersals within Europe. *PLoS One*. **8**, e70492 (2013).
69. Achilli, A. et al. Saami and Berbers—An unexpected mitochondrial DNA link. *Am. J. Hum. Genet.* **76**, 883–886 (2005).
70. Gandini, F. et al. Mapping human dispersals into the Horn of Africa from Arabian ice age refugia using mitogenomes. *Sci. Rep.* **6**, 25472 (2016).
71. Pennarun, E. et al. Divorcing the late upper palaeolithic demographic histories of MtDNA haplogroups M1 and U6 in Africa. *BMC Evol. Biol.* **12**, 234 (2012).
72. Hervella, M. et al. Ancient DNA from south-east Europe reveals different events during early and middle neolithic influencing the European genetic heritage. *PLoS One*. **10**, e0128810 (2015).
73. Fu, Q. et al. The genetic history of ice age Europe. *Nature* **534**, 200–205 (2016).
74. Serrano, J. G. et al. The genomic history of the Indigenous people of the Canary Islands. *Nat. Commun.* **14**, 4641 (2023).
75. Silva, M. et al. 60,000 years of interactions between central and Eastern Africa documented by major African mitochondrial haplogroup L2. *Sci. Rep.* **5**, 12526 (2015).
76. Ottoni, C. et al. First genetic insight into Libyan tuaregs: a maternal perspective. *Ann. Hum. Genet.* **73**, 438–448 (2009).
77. Cerezo, M. et al. Reconstructing ancient mitochondrial DNA links between Africa and Europe. *Genome Res.* **22**, 821–826 (2012).
78. Hernández, C. L. et al. Early holocenic and historic MtDNA African signatures in the Iberian peninsula: the Andalusian region as a paradigm. *PLoS One*. **10**, e0139784 (2015).
79. Matisoo-Smith, E. A. et al. A European mitochondrial haplotype identified in ancient phoenician remains from Carthage, North Africa. *PLoS One*. **11**, e0155046 (2016).
80. Molto, J. E. et al. Complete mitochondrial genome sequencing of a burial from a Romano–Christian cemetery in the Dakhleh oasis, egypt: preliminary indications. *Genes* **8**, 262 (2017).
81. Loreille, O. et al. Biological sexing of a 4000-Year-old Egyptian mummy head to assess the potential of nuclear DNA recovery from the most damaged and limited forensic specimens. *Genes* **9**, 135 (2018).
82. Drosou, K., Collin, T. C., Freeman, P. J., Loynes, R. & Freemont, T. The first reported case of the rare mitochondrial haplotype H4a1 in ancient Egypt. *Sci. Rep.* **10**, 17037 (2020).
83. Matisoo-Smith, E. et al. Ancient mitogenomes of Phoenicians from Sardinia and Lebanon: a story of settlement, integration, and female mobility. *PLoS One*. **13**, e0190169 (2018).
84. García, Ó., Alonso, S., Huber, N., Bodner, M. & Parson, W. Forensically relevant phylogeographic evaluation of mitogenome variation in the Basque country. *Forens. Sci. Int. Genet.* **46**, 102260 (2020).

85. Silva, M. et al. Biomolecular insights into North African-related ancestry, mobility and diet in eleventh-century Al-Andalus. *Sci. Rep.* **11**, 18121 (2021).
86. Botigué, L. R. et al. Gene flow from North Africa contributes to differential human genetic diversity in Southern Europe. *Proc. Natl. Acad. Sci.* **110**, 11791–11796 (2013).
87. Pimenta, J., Lopes, A. M., Comas, D., Amorim, A. & Arenas, M. Evaluating the neolithic expansion at both Shores of the Mediterranean sea. *Mol. Biol. Evol.* **34**, 3232–3242 (2017).
88. Serradell, J. M., Lorenzo-Salazar, J. M., Flores, C., Lao, O. & Comas, D. Modelling the demographic history of human North African genomes points to a recent soft split divergence between populations. *Genome Biol.* **25**, 201 (2024).
89. Salem, N. et al. Ancient DNA from the green Sahara reveals ancestral North African lineage. *Nature* **641**, 144–150 (2025).
90. D'Atanasio, E. et al. The peopling of the last green Sahara revealed by high-coverage resequencing of trans-Saharan patrilineages. *Genome Biol.* **19**, 20 (2018).
91. D'Atanasio, E. et al. The genomic echoes of the last green Sahara on the Fulani and Sahelian people. *Curr. Biol.* **33**, 5495–5504e4 (2023).
92. Soares, P., Rito, T., Pereira, L. & Richards, M. B. A genetic perspective on African prehistory. In *Africa from MIS 6–2: Population Dynamics and Paleoenvironments* (eds Jones, S. C. & Stewart, B. A.) Springer, 383–405 [https://doi.org/10.1007/978-94-017-7520-5\\_18](https://doi.org/10.1007/978-94-017-7520-5_18). (2016).
93. González-Fortes, G. et al. A western route of prehistoric human migration from Africa into the Iberian Peninsula. *Proc. R. Soc. B Biol. Sci.* **286**, 20182288 (2019).
94. Marcus, J. H. et al. Genetic history from the middle neolithic to present on the Mediterranean Island of Sardinia. *Nat. Commun.* **11**, 939 (2020).
95. Underhill, P. A. et al. Y chromosome sequence variation and the history of human populations. *Nat. Genet.* **26**, 358–361 (2000).
96. Jobling, M. A. & Tyler-Smith, C. The human Y chromosome: an evolutionary marker comes of age. *Nat. Rev. Genet.* **4**, 598–612 (2003).
97. Posth, C. et al. Deeply divergent archaic mitochondrial genome provides lower time boundary for African gene flow into Neanderthals. *Nat. Commun.* **8**, 16046 (2017).
98. Capodiferro, M. R. et al. Archaeogenomic distinctiveness of the Isthmo-Colombian area. *Cell* **184**, 1706–1723 (2021).
99. Olalde, I. et al. The genomic history of the Iberian Peninsula over the past 8000 years. *Science* **363**, 1230–1234 (2019).
100. Semino, O. et al. Semino, O. et al. Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. *Am. J. Hum. Genet.* **74**, 1023–1034 (2004).
101. Massicotte, P. & South, A. *rnaturalearth: World Map Data from Natural Earth*. (2025).

## Acknowledgements

We are grateful to all the volunteers who generously participated in this study and made this research possible. This study received support from the Italian Ministry of University and Research (MUR): the project PRIN2022 2022BZYFBE (to FC and OS), PRIN2022 2022Y8BSAL (to AA), PRIN2022 2022SNEBJY (to AT), PRIN2017 2017CWHLHY (to AT), PRIN2022 20224PBY5K (to AO); and from Fondazione Cariplo—Bando Giovani Ricercatori 2023, rif: 2023–1373 (to NRM).

## Author contributions

OS and AO conceived and supervised the project; HL, BT, AN, JMD, FC, and OS performed the sample collection; GC, EM, VA, and FG conducted the experiments, with guidance from AA, OS and AO; RDG, CF, and LC performed sequencing runs of the newly obtained mitogenomes, under the supervision of VC; GC was responsible for data submission to public databases; GC, AR, NRM, and EDA performed the analyses and interpreted the data with assistance from ER, AA, AT, OS, and AO; AA, AT, FC, OS, and AO contributed reagents/materials/analytical tools; GC, OS and AO wrote the manuscript with inputs from ER, MP, AT and MBR. All authors reviewed and approved the manuscript.

## Declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-12209-x>.

**Correspondence** and requests for materials should be addressed to A.O.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025