scientific reports

Check for updates

OPEN

Whole mitogenomes reveal that NW Africa has acted both as a source and a destination for multiple human movements

Julen Aizpurua-Iraola[1], Amine Abdeli[2], Traki Benhassine[2], Francesc Calafell[1] & David Comas[1]✉

Despite being enclosed between the Mediterranean Sea and the Sahara Desert, North Africa has been the scenario of multiple human migrations that have shaped the genetic structure of its present-day populations. Despite its richness, North Africa remains underrepresented in genomic studies. To overcome this, we have sequenced and analyzed 264 mitogenomes from the Algerian Chaoui-speaking Imazighen (a.k.a. *Berbers*) living in the Aurès region. The maternal genetic composition of the Aurès is similar to Arab populations in the region, dominated by West Eurasian lineages with a moderate presence of M1/U6 North African and L sub-Saharan lineages. When focusing on the time and geographic origin of the North African specific clades within the non-autochthonous haplogroups, different geographical neighboring regions contributed to the North African maternal gene pool during time periods that could be attributed to previously suggested admixture events in the region, since Paleolithic times to recent historical movements such as the Arabization. We have also observed the role of North Africa as a source of geneflow mainly in Southern European regions since Neolithic times. Finally, the present work constitutes an effort to increase the representation of North African populations in genetic databases, which is key to understand their history.

Modern humans have inhabited North Africa at least since 300,000 years ago[1], and due to its geographic location, it has been home to many different populations and has seen multiple human movements throughout history[2]. However, knowledge about North African prehistory remains sketchy. From Morocco to Egypt and as far south as the Sahel, many and extremely diverse lithic industries dating to between 190 kya (thousand years ago) and 57 kya have been discovered and linked to the Aterian culture[3]. In Mesolithic times, two distinct lithic cultures are recognized in the Maghreb: the Iberomaurusian (~ 22–9 kya)[4] and the Capsian (~ 10–6 kya)[5], who are believed to have originated from the Paleolithic Aterian people[3]. Subsequently, the diffusion of the Neolithic (starting at ~ 5.5 kya) involved population growth, with diversified subsistence systems and an increase in sedentarism. These populations possibly gave rise to present day Amazigh (sing.)/Imazighen (pl.) populations, which are also known as *Berbers*[4,5].

Direct genetic analysis of prehistoric samples in North Africa has been hampered by the harsh climatic conditions that resulted in poor DNA preservation. Nevertheless, the analysis of the oldest aDNA samples from the entire African continent comes from the Taforalt Iberomaurusian settlement in Morocco from ~ 15 kya. Their study revealed the affinity of Epipaleolithic North Africans with Epipaleolithic Near Eastern populations[6]. Additionally, a sub-Saharan component was also detected, although none of the present day or ancient Holocene African groups were found to be a good proxy for the source of this component. Posterior ancient and modern samples revealed a certain degree of genetic continuity in the region form the Later Stone Age to Neolithic populations in the Maghreb and evidenced a remarkable impact of the Neolithic transition coming from Europe probably via the strait of Gibraltar[7,8]. After the Neolithic era, the major demographic movements in the region were: (i) the trans-Saharan gene flow caused mainly due to the slave-trade starting during the Roman empire rule (first century BCE) through the Arab conquest until the nineteenth century and (ii) the Arabization[9], starting in the seventh century CE and introducing gene flow from the Middle East and leading to an East to West

[1]Departament de Medicina i Ciències de la Vida, Institut de Biologia Evolutiva (CSIC-UPF), Universitat Pompeu Fabra, Barcelona, Spain. [2]Laboratorie de Biologie Cellulaire et Moléculaire, Faculté des Sciences Biologiques, Université des Sciences et de la Technologie Houari Boumediene, Alger, Algeria. ✉email: david.comas@upf.edu

genetic cline of the Middle Eastern component in North Africa[10]. Other possible external contributors were the Phoenicians, Romans, Vandals, Byzantines, Ottoman Turks and other Mediterranean populations.

Mitochondrial DNA (mtDNA) has been a widely used genetic marker in population genetics[11] and North Africa is no exception. Most of the studies conducted have focused on either control region lineage frequencies in different populations[12–14] or on particular lineages present in North African population[15–19]. Control region studies reported a high heterogeneity in the haplogroup distribution of North African populations[12,20], with some East–West clinal distributions for some lineages. Haplogroups H, HV0, L1b, L3b and U6 are more frequent in Western North Africa; while M1, L0a, R0a, N1b, I and J are more frequent in Eastern North Africa[16,18]. Overall, the North African maternal genetic landscape can be divided in (i) West Eurasian origin lineages, which generally make up for the majority of the maternal pool in North African populations; (ii) the U6 and M1 autochthonous North African haplogroups, which have been observed at least since the Epipaleolithic era in the region[6]; and (iii) sub-Saharan L lineages[13]. This lineage diversity evidences the importance of North Africa as a scenario of different human migrations, the study of which has been the objective of different lineage-specific analyses. For instance, an ancient link with Europeans was inferred since Iberian post-Glacial expansion lineages (namely U5b1b, H1, H3 and V) were detected in Amazigh populations[16] and the analysis of complete mitogenomes provided enough resolution to discover novel North African specific clades within these haplogroups with coalescence times of around 4 to 7 kya[15], whose origin could be attributed to the Neolithic expansion from the Iberian Peninsula. The expansion of some of these European lineages did not stop in North Africa and reached sub-Saharan African populations[21]. Other studies have expanded on the link with sub-Saharan populations (although mainly through the L haplogroup mtDNAs present in Europe) and discovered both ancient and recent origin mitochondrial lineages in Europe and North Africa[12,19]. In agreement with this, North African U6 and M1 lineages have also been observed in Europe[14,17]. All these results indicate that humans have permeated the Sahara Desert and Mediterranean Sea barriers enabling gene flow during different periods.

In this context, Algeria, the largest country in Africa with an area of 2.4 million km^2, is home to different linguistic groups that include not only Arab speakers (who mainly inhabit the cosmopolitan cities), but also to many different Tamazight or Berber speaking groups like the Chaoui or Shawiya, Kabyle, Mozabite, Zenate, Chleuh, or Touareg. Previous studies have highlighted the heterogeneity between different Algerian groups, which appeared not to be correlated with linguistic or geography[20]. However, these studies analyzed control region data, which due to its limited phylogeographic resolution when comparing with complete mitogenomes, did not provide fine-grained phylogeographic information on the origin of the lineages. In addition, the informativeness of the mitochondrial phylogenies depends directly on the amount of available data from reference populations. In this sense, North Africa has been traditionally underrepresented in population genetic which is exemplified by the presence of just a single North African population (the Mozabite) in a global genomic database like the Human Genome Diversity Project[22] and the 4 samples (2 Mozabite and 2 from the Western Saharawi population) in the Simons Genome Diversity Project[23].

In the present study we aim to contribute to the representation of genetically neglected regions such as mainland North Africa, where, to the best of our knowledge, no population-based complete mitogenome studies have been conducted so far. To that effect, we have analyzed 264 mitogenomes of individuals from the Chaoui Amazigh population. The Chaoui Imazighen inhabit the Aurès mountainous region in North Eastern Algeria, and despite being one of the largest Amazigh populations in North Africa, just two genetic studies analyzing Y-chromosome and autosomal STRs have covered this population[24,25]. We focus on the mtDNA diversity observed in three different localities in the Aurès region and analyze the time and geographic origin of the maternal lineages in our dataset.

Materials and methods

Samples and sequencing. We amplified and sequenced the whole mtDNA from blood samples donated by 302 volunteers from three different localities in the Aurès region in Algeria: Batna (n = 136), Khenchela (n = 82) and Oum El Bouaghi (n = 84) (Fig. 1). All samples correspond to *Shawiya* or *Chaoui* speakers with all four grandparents from the Aurès region. A questionnaire and a written informed consent was obtained from the volunteers; the study complies with the ethical rules of all the institutions involved and has been approved by the *Laboratoire de Biologie Cellulaire et Moléculaire* at the *Faculté des Sciences Biologies* of Université de Sciences et Technologie Houari Boumedienne in Algiers and the CEIm-PSMAR IRB in Barcelona (2019/8900/I). All methods in this study were performed following the standard guidelines and regulations in accordance with the Declaration of Helsinki.

DNA was extracted using the QIAamp DNA blood mini kit (Qiagen GmbH, Hilden, Germany) following the manufacturer's recommendations and was quantified using Quantifiler™ Human DNA Kit on 7500 SDS Reak-Time PCR System (Applied Biosystems). PCR amplifications were performed in four different fragments as in Ref.[26]. Nextera XT libraries were prepared and sequenced in a Miseq sequencer (Illumina) following the Illumina mtDNA Genome Guidelines[27].

Sequence processing. Sequences were processed according to the GATK best practices protocol[28]. First, an initial quality check was performed using FastQC[29]. Then, the raw sequencing reads were mapped against the revised Cambridge Reference Sequence (rCRS)[30] using the BWA-MEM algorithm[31]. The PCR duplicates were removed with Picard tools[32], base quality scores were recalibrated with GATK's Base Quality Score Recalibration (BQSR)[33], and a final quality report was obtained with Qualimap2[34]. Finally, the sequence variants were called with GATK tools HaplotypeCaller and GenotypeGVCFs[33].

As a quality filter, we first set mean coverage values of above 15× for each of the amplified fragments and a Haplogrep quality score above 85%. Samples failing any these conditions were discarded unless they had at

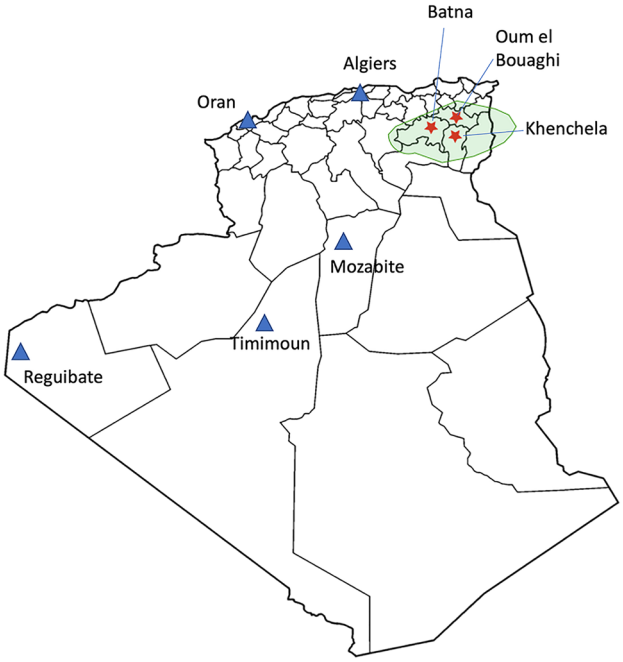

Figure 1. Map of Algeria with the three sampling sites in the Aurès marked with red stars. The green area represents the Aurès mountainous region. The map was created with the R package 'rnaturalearth'[59].

least a coverage of 5× across all the reference. 24 samples failed both the 15× coverage and the > 85% Haplogrep quality score, 7 samples failed the coverage filter and were discarded, and 8 samples failed just the Haplogrep score requirement, of which 4 of them were discarded since they contained regions with < 5× of coverage. Finally, three additional samples were discarded since we detected three pairs of third-degree relatives in our dataset after an autosomal genomewide kinship analysis with King software[35]. We ended up with a final number of 264 sequences (Batna n = 133, Khenchela n = 75, Oum El Bouaghi n = 56).

Statistical analysis. Haplogroups were determined with Haplogrep v.2.4.0 using the forensic update of phylotree v.17[36]. Molecular diversity and summary statistics were calculated with the *pegas* package in R[37] and AMOVA and Φst distances were computed with *poppr*[38] and *ade4*[39].

Population analysis. Given the absence of whole mtDNA population reference datasets from North Africa, we downloaded control region data from different Algerian populations: Mozabite (n = 85)[40], Timimoun (n = 73)[20], Reguibate (n = 108)[20], Oran (n = 333)[20,41] and Algiers (n = 62)[20].

Phylogeographic analysis. We first selected the haplogroups with a count number of $n \geq 2$ and blasted the mitogenomes within those haplogroups in Genbank searching for the most similar publicly available sequences. We downloaded the 20 most similar sequences based on the identity percentage obtained and built phylogenetic networks looking for North African specific clades. We define North African specific clades as the groups of sequences found in North Africa that belong to the same haplogroup and share at least one mutation (not accounting for the highly recurrent mutations listed in Ref.[42]) that separates them from other sequences. For each North African clade, we dated the internal Time to the Most Recent Common Ancestor (TMRCA) in addition to the time to the closest non-North African individual using BEAST 1.10[43]. For the case of sequences belonging to M1 and U6 haplogroups, we inferred (i) the coalescent time for each of the non-North African clades found within the sequences obtained from GenBank, and (ii) the coalescence time between each non-North African sequence (not forming clades) and the closest North African sequence.

For the BEAST analysis, we used a prior mutation rate of 2.355×10^{-8} substitutions per nucleotide per year, taking into account purifying selection as in Ref.[44], the substitution model we used was the GTR as indicated by jModelTest2[45] and set a strict clock model assuming all tree branches evolve at the same rate. We used UPGMA as the starting trees for the analysis, which was conducted in five independent runs with 15,000,000 iterations, sampling the result every 15,000 iterations. The resulting log files were combined using LogCombiner and the combined log file was checked with Tracer 1.7[46] to ensure effective sample size (ESS) values over 200 for every parameter.

Results

mtDNA diversity in the Aurès region within the Algerian context. We generated a total of 264 complete mitogenomes with a mean coverage of 593.5× (Suppl. Figs. 1, 2) from three different localities in the Aurès mountainous region in North Eastern Algeria: Batna (n = 133), Khenchela (n = 75) and Oum El Bouaghi (n = 56) (Suppl. Table 1). We computed diversity statitics for the three localities (Suppl. Table 2A). Batna shows a slightly lower haplotype diversity in comparison to the other populations. We computed an AMOVA test to verify whether there was significant internal variation among the three localities, yielding a small but significant fraction of variance found among the localities (0.92%, p = 0.008), with Khenchela showing the highest φ_{st} values (Suppl. Table 2B).

Due to the lack of whole mtDNA population data from North African populations, we built a dataset of control region sequences from different parts of Algeria (see "Materials and methods" section). Within the Algerian context, we observe a clear heterogeneity in the haplogroup composition between the different populations as previously reported[20] (Fig. 2a, Suppl. Table 3). The mtDNA haplogroup composition from the Aurès region shows high frequencies of West Eurasian (WE) superhaplogroups (mainly H/HV) with 60% of WE haplogroups in Khenchela and 74.4% and 69.6% for Batna and Oum El Bouaghi respectively (Fig. 1). Overall, Khenchela is also differentiated because of higher proportions of U6 and specially M1 North African sequences, being the locality with the highest frequency of M1 lineages (12%) among all the Algerian populations we considered. Regarding sub-Saharan lineages, Khenchela also showed the highest proportion of L haplogroup sequences among the Aurès region but well below the Amazigh population in Timimoun or the Arab population in Algiers.

Both haplogroup frequency-based distances and φ_{st} distances show the populations from the Aurès region are closer to the population in Oran and the Reguibate population (Fig. 2b, Suppl. Fig. 3). Khenchela shows higher φ_{st} distances (Suppl. Table 4) and also appears slightly separated from Batna and Oum El Bouaghi the haplogroup frequency-based PCA, probably due to the high M1 proportions.

Origins of the Aurès mitogenome lineages. After exploring the phylogeographic origin of the lineages of non-North African origin in the Aurès sample, we found that 73 sequences from the Aurès region belonged to 26 different North African specific clades with origins outside Northern Africa (Fig. 3) (without accounting for the 18 clades formed by 31 identical sequences (Suppl. Fig. 26)). That is, the closest sequences to these clades had been sampled outside of North Africa, particularly in Europe, the Middle East, the Sahel, and West, Central, and East Africa.

Many European lineages that are present in the Aurès sample (i.e.: H1e1a, H1v, K1a3a, U5a1c, etc.) have internal coalescence ages of around 2000 years before present (ybp) and some are not only found in the Aurès region but also in other North African populations including the Canary Islands (Suppl. Figs. 4–15). Coalescence ages to the closest non-North African individuals, which serve as an upper limit of the origin of the lineage in North Africa, are more heterogeneous (ranging from ~ 3000 to ~ 11,000 ybp) as they are related to each haplogroup degree of representation in the literature. Besides, one lineage belonging to haplogroup U3a1c has an internal coalescence age of 5888 ybp (Fig. 3) suggesting European influence in the region dating as far back as ~ 6000 ybp.

Regarding Middle Eastern lineages, we found three North African specific clades also present in the Aurès region belonging to haplogroups T1a7, J2a2b1a* (sensu[47]) and J2a2d. The North African specific clade within T1a7 is just formed by two sequences from the Aurès and have a recent (967 ybp) coalescence age, while the two other clades are present in other North African populations (including the Canary Islands) and have prehistoric coalescence ages (5058 and 9179 ybp for J2a2b1a* and J2a2d respectively) (Fig. 3, Suppl. Figs. 16–18).

We also found an influence of populations from the Sahel region (mainly Fulani populations from Burkina Faso, Chad, Niger and Mali) as we found North African lineages within haplogroups H1cb, L3e5a and L3b1a9 with a putative origin in the Sahel region (Suppl. Figs. 19–21). The time origin of these lineages in North Africa appears to be prehistoric since all but one internal coalescence ages inferred are above 5000 ybp (Fig. 3). The branching pattern of the H1cb lineages shows the presence of a more basal sequence in Mauritania, while the

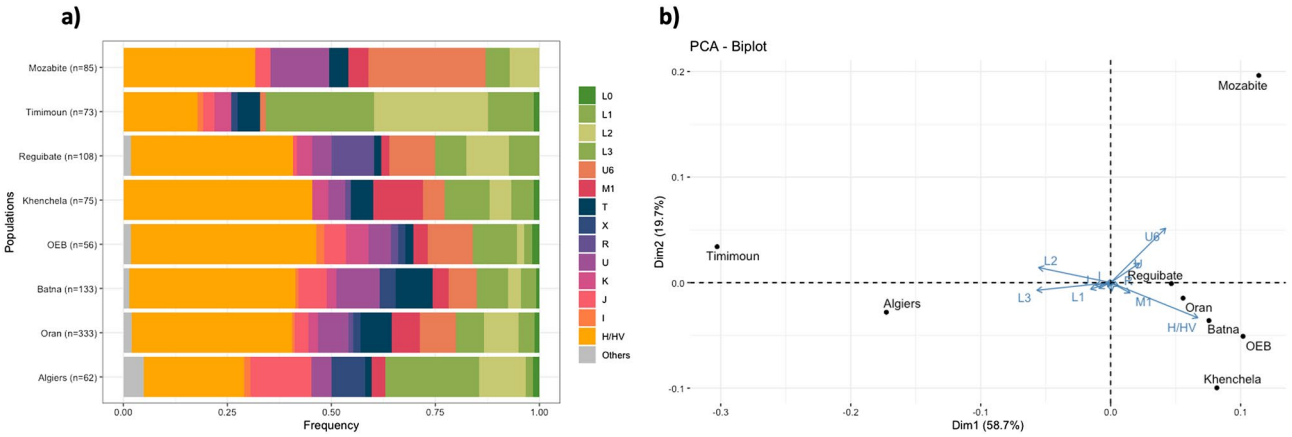

Figure 2. (**a**) The haplogroup composition for each of the Algerian populations based on HVS-I data and (**b**) principal component analysis (PCA) based on the haplogroup frequencies obtained from HVS-I data from Algerian populations.

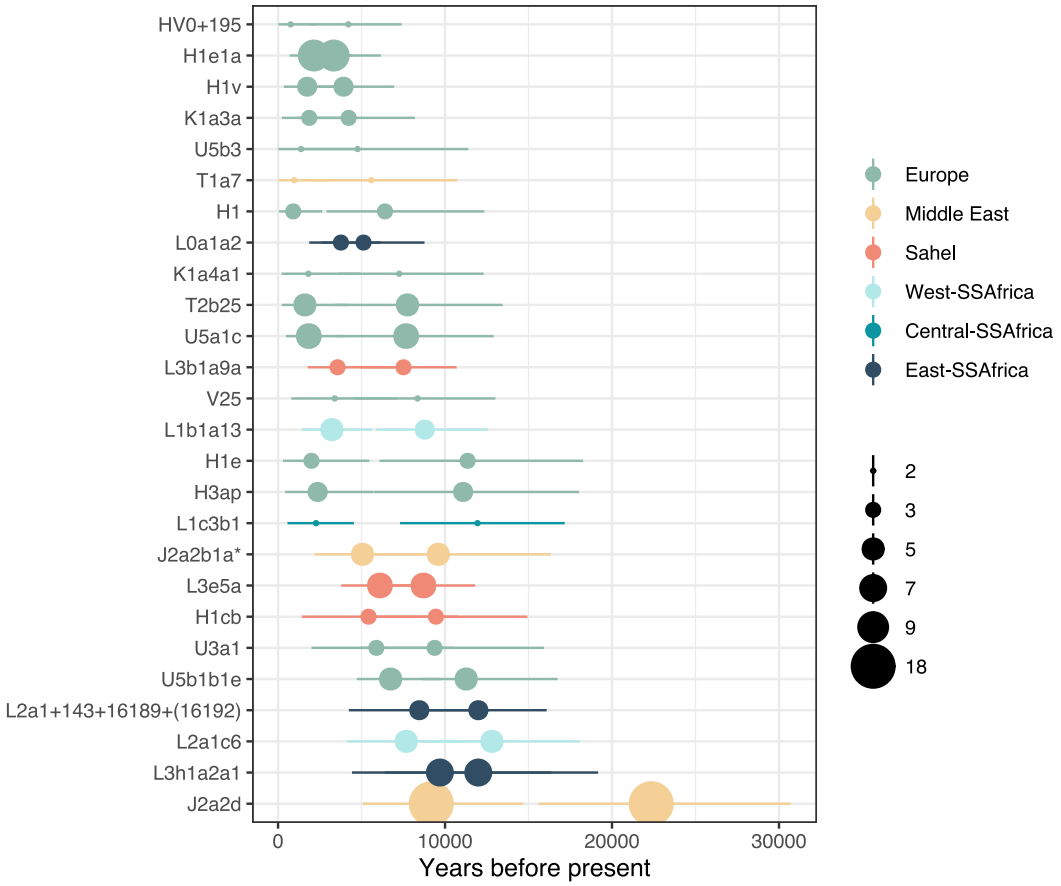

Figure 3. Inferred coalescent dates for all North African specific clades (left point) and inferred date to the closest non-North African sample (right point). Colour indicates the geographic origin and the point size refers to the number of sequences within the North African clades.

Aurès specific clade stems from the Sahelian diversity, which could point to a "back to North Africa" migration of this lineage[21]. When considering the origin of the Sahelian L3 lineages in North Africa, we observe a number of different North African sequences outside the North African specific clades which might reflect independent migrations events throughout history.

Finally, we also found North African specific clades with Eastern, Western and Central sub-Saharan African origins. The time origin of these lineages is overall heterogenous with internal coalescent date inferences ranging from 2275 to 9696 ybp (Fig. 3). As in the case of the lineages with Sahelian origin, we observe many North African individuals with haplogroups from different regions of sub-Saharan Africa outside North African specific clades, which might also suggest independent migration events (Suppl. Figs. 22–25).

Although most North African clades were composed of non-clonal sequences separated by greater or less divergence time, we also observed North African specific clades belonging to non-North African haplogroups constituted by identical sequences, which could be a sign of very recent migration events into the region (Suppl. Fig. 26). The origin of most of these lineages seems to be European, although we find some Middle Eastern lineages like R0a2 which is present in Bedouins (Suppl. Fig. 27). Interestingly, two samples from the Aurès region share their haplotype with an ancient Phoenician sample collected in Sardinia belonging to the W5 haplogroup (Suppl. Fig. 28).

Influence of the Aurès and North Africa over other territories. Besides the phylogeographic structure of non-North African sequences in Northern Africa we also analyzed the M1 and U6 lineages present in the Aurès region outside North Africa. These haplogroups originated, or at least greatly expanded, in North Africa[10,17,18]. For this purpose, we inferred the coalescence age between the closest non-North African and North African individuals belonging to haplogroups present in the Aurès sample. The results show that the time divergences of M1 and U6 lineages between North African and non-North African individuals are roughly uninterrupted along time and that most of these lineages are found in Southern Europe (Suppl. Fig. 29).

We found some European specific U6 and M1 clades for which we inferred the divergence time and found that many of the TMRCA of these clades fall between the present time and 5,000 ybp. Besides, these lineages are also found in Southern Europe (Suppl. Fig. 30).

Discussion

Despite its extensive geographic range and the presence of numerous distinct populations, North Africa has generally received little attention in genetic studies. Algeria, the largest country in Africa, is home to several different linguistic groups including Arab speakers, who make up the majority in urban areas, and also several distinct Tamazight speaking groups, such as the Chaoui or Shawiya, Kabyle, Mozabite, Zenate, Chleuh, and Touareg. The genetic studies covering these populations, however, are still scarce. We present 264 new complete mitochondrial sequences from the Chaoui Amazigh group collected in three different localities in the Aurès region in North-eastern Algeria.

We observed small but significant variation among the three localities sampled in the Aurès region (0.92%, $p = 0.008$), with Khenchela being the locality contributing most to this local substructure. Looking at the linage composition of each of the sampling sites, Khenchela's contribution to the local maternal substructure observed is possibly caused by the high proportion of M1 lineages, the highest among the Algerian populations examined in this study.

In an Algerian context, we analysed the HVS-I control region of different Algerian groups together with our samples and observed the genetic heterogeneity previously described in the country as well as for other Imazighen[9,12,20]. The haplogroup composition of the Aurès sample is close to that of the city of Oran (Fig. 2), the second largest city in Algeria and with an Arab majority, which might suggest genetic similarities between the Aurès and inhabitants from cosmopolitan cities. Their haplogroup composition is mainly composed by H/HV lineages evidencing the European influence in North Africa, although other West Eurasian haplogroups can also be observed in most Algerian populations in lower frequencies. Besides, we also observe a moderate frequency of M1 and U6 North African autochthonous lineages that have been present in the region since Palaeolithic times[6], and finally sub-Saharan L lineages at frequencies ~ 20%.

Genetic and archaeological evidence have dated the first Western European influence in North Africa around 6000 ybp coming from the Iberian Peninsula and have linked it to the Neolithization process in North Africa[7,48]. In accordance with this, several European mitochondrial lineages discovered in the Fulani and other Sahelian populations (namely H1cb1 and U5b1b1b) have been linked to this same migration across the Gibraltar strait to North Africa from where it would have spread southwards to the Sahel[21,49,50]. After inspecting the time and geographic origin of the lineages present in the Aurès region and other parts of North Africa, we observed not only clades belonging to these specific haplogroups (H1cb1) but also other clades of West Eurasian origin with coalescent dates around the same period (U3a1c, ~ 6000 ybp and U5b1b1e, ~ 6700 ybp) that might also be linked to the migrations across the Gibraltar strait during the Early Neolithic period. The rest of European origin lineages found in the Aurès region and in other North African samples have more recent coalescent dates (between ~ 2000 and ~ 3000 ybp) indicating an entrance during historical times during the Phoenician, Roman or other contacts with Mediterranean populations. Nevertheless, given that some of these lineages are also present in the Canary Islands (H1e1a)[51], their coalescence age should be at least higher than ~ 2000 ybp (the estimated age for the colonization of the archipelago[52]).

Regarding Middle Eastern origin lineages in North Africa, we have observed the presence of a North African specific J2a2d haplogroup with an estimated coalescence age of ~ 9100 ybp. This date predates the Neolithization period in North Africa and, given the link between the Taforalt samples and the Epipaleolithic Middle Eastern populations (Natufians)[6], it could represent pre-agricultural contact connection between the Near East and North Africa, as it occurs with the Y-chromosome lineage E1b1b1a1 (M-78), which is very frequent in North Africa and is very closely related to the lineage E1b1b1b (M-123) found in Natufians and Pre-Pottery Neolithic Levantines[53]. The North African haplogroup J2a2b1a* (sensu[47]) also presents a prehistoric coalescent date around 5,000 ybp but its more recent origin could also be related to posterior migrations. The clade within T1a7 seems to be the only recent Middle Eastern lineage found in the Aurès region with a coalescence age of < 1000 ybp, and possibly reflects the incoming maternal lineages due to the Arabization or the Bedouin expansion[8,54].

Finally, the genetic influence of sub-Saharan population in North Africa can be dated back to Epipaleolithic times, as the Taforalt individuals present a sub-Saharan component making up to approximately a third of their genomes[6]. Besides, prehistoric contacts between both sides of the Sahara have been mainly evidenced by mtDNA studies, since most recent genome wide studies have mostly been able to detect recent sub-Saharan admixture in North Africa driven by the extensive slave trade[55]. European haplogroups related to the Neolithic expansion (H1cb1 and U5b1b1b) have been observed in the Sahelian populations[21,49,50] and at the same time, sub-Saharan lineages with high frequencies in Lake Chad Basin populations (L3e5) have been reported in North African populations and are believed to have a prehistorical origin[12,56]. We have detected the presence of these lineages in North Africa—in addition to another lineage with a presumed Sahelian origin (L3b1a9a)—and the coalescent times of these lineages agree with those inferred in previous studies pointing out to a trans-Saharan during the Green Sahara period (~ 10,000–5000 ybp)[56,57]. Besides, we have observed North African specific clades with even older coalescence ages (ranging from ~ 7600 to ~ 9700 ybp), whose origins seem less restricted to a particular region, and these could have also arrived at North Africa during the Green Sahara period. Finally, it is worth noting that the great majority of sub-Saharan sequences in North Africa do not form phylogenetic clades, and therefore, show no signals of having evolved in North Africa. This observation is compatible with a recent arrival of these sequences to the region as suggested by recent genomewide evidence[9,55].

Our results evidence that North Africa has not only acted as a sink of different human migrations, but also as a source and as a bidirectional corridor. North Africa's role as a corridor is evidenced by the European lineages mentioned earlier that crossed the Gibraltar strait and the Sahara and are nowadays present in both North African and sub-Saharan populations. Our results, in agreement with previous studies[19], also indicate that some lineages took the opposite direction. In fact, most of the North African clades with sub-Saharan origin contain sequences sampled in Europe (Suppl. Figs. 21–24). Finally, North Africa's role as a source is evidenced by the

presence of the autochthonous M1 and U6 sequences outside North Africa. Our results, despite focusing just on lineages present in the Aurès region, point to a roughly uninterrupted geneflow from North Africa mainly towards Southern Europe, since at least ~ 5000 ybp as suggested by the coalescence ages inferred from most of the European specific U6 and M1 clades observed. This agrees with previous results that point out higher historic spreads of North African lineages possibly during Phoenician rule, the Roman empire or the Arab expansion[19].

This present work, despite the multiple studies focused on specific haplogroups within North Africa[15,17,18,21,58], constitutes the first population-based whole mitogenome study of a mainland North African population. This, although it represents a positive effort to increase the representation of undersampled populations in public databases and in the literature, can also give rise to potential limitations. Due to the limited number of publicly available mitochondrial genomes from sub-Saharan Africa and North Africa in comparison with Europe, we need to be cautious when interpreting these results, since African populations remain largely uncharacterized from a population genetics standpoint and some of the phylogeographic inferences might change with the addition of more genetic data.

Data availability

The mitochondrial genomes produced in this study are available on GenBank with the accession numbers: OQ884717–OQ884980.

Received: 26 April 2023; Accepted: 23 June 2023
Published online: 27 June 2023

References

1. Callaway, E. Oldest Homo sapiens fossil claim rewrites our species' history. *Nature*. https://doi.org/10.1038/NATURE.2017.22114 (2017).
2. Lucas-Sánchez, M., Serradell, J. M. & Comas, D. Population history of North Africa based on modern and ancient genomes. *Hum. Mol. Genet.* **30**(R1), R17–R23. https://doi.org/10.1093/HMG/DDAA261 (2021).
3. Scerri, E. M. L. The Aterian and its place in the North African Middle Stone Age. *Quatern. Int.* **300**, 111–130. https://doi.org/10.1016/J.QUAINT.2012.09.008 (2013).
4. Irish, J. D. The Iberomaurusian enigma: North African progenitor or dead end? *J. Hum. Evol.* **39**(4), 393–410. https://doi.org/10.1006/JHEV.2000.0430 (2000).
5. Rahmani, N. Technological and cultural change among the last hunter-gatherers of the Maghreb: The Capsian (10,000–6000 B.P.). *J. World Prehist.* **18**(1), 57–105 (2004).
6. Van De Loosdrecht, M. *et al.* Pleistocene north African genomes link near eastern and sub-saharan African human populations. *Science* **360**(6388), 548–552. https://doi.org/10.1126/SCIENCE.AAR8380/SUPPL_FILE/AAR8380_VANDELOOSDRECHT_SM.PDF (2018).
7. 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. U.S.A.* **115**(26), 6774–6779. https://doi.org/10.1073/PNAS.1800851115 (2018).
8. Serra-Vidal, G. *et al.* Heterogeneity in palaeolithic population continuity and neolithic expansion in North Africa. *Curr. Biol.* **29**(22), 3953–3959. https://doi.org/10.1016/J.CUB.2019.09.050 (2019).
9. Arauna, L. R. *et al.* Recent historical migrations have shaped the gene pool of arabs and berbers in North Africa. *Mol. Biol. Evol.* **34**(2), 318–329. https://doi.org/10.1093/MOLBEV/MSW218 (2017).
10. Henn, B. M. *et al.* Genomic ancestry of North Africans supports back-to-Africa migrations. *PLoS Genet.* **8**(1), e1002397. https://doi.org/10.1371/JOURNAL.PGEN.1002397 (2012).
11. Kivisild, T. Maternal ancestry and population history from whole mitochondrial genomes. *Investig. Genet.* **6**(1), 1–10. https://doi.org/10.1186/S13323-015-0022-2/FIGURES/2 (2015).
12. Fadhlaoui-Zid, K. *et al.* Mitochondrial DNA heterogeneity in Tunisian berbers. *Ann. Hum. Genet.* **68**(3), 222–233. https://doi.org/10.1046/J.1529-8817.2004.00096.X (2004).
13. Salas, A. *et al.* The making of the African mtDNA landscape. *Am. J. Hum. Genet.* **71**(5), 1082. https://doi.org/10.1086/344348 (2002).
14. Plaza, S. *et al.* Joining the pillars of hercules: mtDNA sequences show multidirectional gene flow in the Western Mediterranean. *Ann. Hum. Genet.* **67**(4), 312–328. https://doi.org/10.1046/J.1469-1809.2003.00039.X (2003).
15. Ottoni, C. *et al.* Mitochondrial Haplogroup H1 in North Africa: An early holocene arrival from Iberia. *PLoS ONE* **5**(10), 13378. https://doi.org/10.1371/JOURNAL.PONE.0013378 (2010).
16. Achilli, A. *et al.* Saami and berbers—An unexpected mitochondrial DNA link. *Am. J. Hum. Genet.* **76**(5), 883. https://doi.org/10.1086/430073 (2005).
17. Secher, B. *et al.* The history of the North African mitochondrial DNA haplogroup U6 gene flow into the African, Eurasian and American continents. *BMC Evol. Biol.* **14**(1), 109. https://doi.org/10.1186/1471-2148-14-109 (2014).
18. Pennarun, E. *et al.* Divorcing the Late Upper Palaeolithic demographic histories of mtDNA haplogroups M1 and U6 in Africa. *BMC Evol. Biol.* **12**(1), 1–12. https://doi.org/10.1186/1471-2148-12-234/TABLES/3 (2012).
19. Cerezo, M. *et al.* Reconstructing ancient mitochondrial DNA links between Africa and Europe. *Genome Res.* **22**(5), 821–826. https://doi.org/10.1101/GR.134452.111 (2012).
20. Bekada, A. *et al.* Genetic heterogeneity in Algerian human populations. *PLoS ONE* **10**(9), e0138453. https://doi.org/10.1371/JOURNAL.PONE.0138453 (2015).
21. Kulichová, I. *et al.* Internal diversification of non-Sub-Saharan haplogroups in Sahelian populations and the spread of pastoralism beyond the Sahara. *Am. J. Phys. Anthropol.* **164**(2), 424–434. https://doi.org/10.1002/AJPA.23285 (2017).
22. Bergström, A. *et al.* Insights into human genetic variation and population history from 929 diverse genomes. *Science* **367**, 6484. https://doi.org/10.1126/SCIENCE.AAY5012/SUPPL_FILE/AAY5012-BERGSTROM-SM.PDF (2020).
23. Mallick, S. *et al.* The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**(7624), 201–206. https://doi.org/10.1038/nature18964 (2016).
24. Abdeli, A. & Benhassine, T. Paternal lineage of the Berbers from Aurès in Algeria: Estimate of their genetic variation. *Ann. Hum. Biol* **46**(2), 160–168. https://doi.org/10.1080/03014460.2019.1602166 (2019).
25. Abdeli, A. & Benhassine, T. Genetic diversity of 15 autosomal STRs in a sample of Berbers from Aurès region in the Northeast of Algeria and genetic relationships with other neighbouring samples. *Ann. Hum. Biol.* **47**(3), 284–293. https://doi.org/10.1080/03014460.2020.1736628 (2020).

26. Aizpurua-Iraola, J., Giménez, A., Carballo-Mesa, A., Calafell, F. & Comas, D. Founder lineages in the Iberian Roma mitogenomes recapitulate the Roma diaspora and show the effects of demographic bottlenecks. *Sci. Rep.* **12**(1), 1–10. https://doi.org/10.1038/s41598-022-23349-9 (2022).
27. Illumina. *Human mtDNA Genome Guide 15037958*. https://emea.support.illumina.com/downloads/human_mtdna_genome_guide_15037958.html (Accessed April 2020) (2016).
28. van der Auwera, G. A. *et al.* From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinform.* **43**, 43. https://doi.org/10.1002/0471250953.bi1110s43 (2013).
29. Andrews, S. FastQC: A quality control tool for high throughput sequence data. *Babraham Bioinform*. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (Accessed 26 January 2022) (2010).
30. Andrews, R. M. *et al.* Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23**(2), 147. https://doi.org/10.1038/13779 (1999).
31. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. http://arxiv.org/abs/1303.3997 (2013).
32. Picard Toolkit. *Picard Toolkit. Broad Institute, GitHub Repository*. http://broadinstitute.github.io/picard/ (2019).
33. McKenna, A. *et al.* The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**(9), 1297. https://doi.org/10.1101/GR.107524.110 (2010).
34. Okonechnikov, K., Conesa, A. & García-Alcalde, F. Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **32**(2), 292–294. https://doi.org/10.1093/bioinformatics/btv566 (2016).
35. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**(22), 2867. https://doi.org/10.1093/BIOINFORMATICS/BTQ559 (2010).
36. Weissensteiner, H. *et al.* HaploGrep 2: Mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **44**(W1), W58–W63. https://doi.org/10.1093/nar/gkw233 (2016).
37. Paradis, E. & Barrett, J. pegas: An R package for population genetics with an integrated–modular approach. *Bioinformatics* **26**(3), 419–420. https://doi.org/10.1093/BIOINFORMATICS/BTP696 (2010).
38. Kamvar, Z. N., Brooks, J. C. & Grünwald, N. J. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Front. Genet.* **6**, 208. https://doi.org/10.3389/FGENE.2015.00208/BIBTEX (2015).
39. Bougeard, S. & Dray, S. Supervised multiblock analysis in R with the ade4 Package. *J. Stat. Softw.* **86**, 1–17. https://doi.org/10.18637/JSS.V086.I01 (2018).
40. Côrte-Real, H. B. S. M. *et al.* Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. *Ann. Hum. Genet.* **60**(4), 331–350. https://doi.org/10.1111/J.1469-1809.1996.TB01196.X (1996).
41. Bekada, A. *et al.* Introducing the Algerian mitochondrial DNA and Y-chromosome profiles into the north African Landscape. *PLoS ONE* **8**(2), e56775. https://doi.org/10.1371/JOURNAL.PONE.0056775 (2013).
42. Soares, P. *et al.* Correcting for purifying selection: An improved human mitochondrial molecular clock. *Am. J. Hum. Genet.* **84**(6), 740. https://doi.org/10.1016/J.AJHG.2009.05.001 (2009).
43. Suchard, M. A. *et al.* Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* **4**(1), 16. https://doi.org/10.1093/VE/VEY016 (2018).
44. Sá, L. *et al.* Phylogeography of sub-saharan mitochondrial lineages outside Africa highlights the roles of the holocene climate changes and the atlantic slave trade. *Int. J. Mol. Sci.* **23**(16), 9219. https://doi.org/10.3390/IJMS23169219 (2022).
45. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: More models, new heuristics and high-performance computing. *Nat. Methods* **9**(8), 772. https://doi.org/10.1038/NMETH.2109 (2012).
46. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Syst. Biol.* **67**(5), 901–904. https://doi.org/10.1093/SYSBIO/SYY032 (2018).
47. Dür, A., Huber, N. & Parson, W. Fine-tuning phylogenetic alignment and haplogrouping of mtDNA sequences. *Int. J. Mol. Sci.* **22**(11), 5747. https://doi.org/10.3390/IJMS22115747 (2021).
48. Vicente, M. *et al.* Population history and genetic adaptation of the Fulani nomads: Inferences from genome-wide data and the lactase persistence trait. *BMC Genom.* **20**(1), 1–12. https://doi.org/10.1186/S12864-019-6296-7/FIGURES/3 (2019).
49. Diallo, M. Y. *et al.* Circum-Saharan prehistory through the lens of mtDNA diversity. *Genes (Basel)* **13**(3), 533. https://doi.org/10.3390/GENES13030533 (2022).
50. Pala, M. *et al.* Mitochondrial Haplogroup U5b3: A distant echo of the epipaleolithic in Italy and the legacy of the early Sardinians. *Am. J. Hum. Genet.* **84**(6), 814–821. https://doi.org/10.1016/j.ajhg.2009.05.004 (2009).
51. García-Olivares, V. *et al.* Digging into the admixture strata of current-day Canary Islanders based on mitogenomes. *iScience* **26**(1), 105907. https://doi.org/10.1016/j.isci.2022.105907 (2023).
52. Fregel, R. *et al.* Mitogenomes illuminate the origin and migration patterns of the indigenous people of the Canary Islands. *PLoS ONE* **14**(3), e0209125. https://doi.org/10.1371/JOURNAL.PONE.0209125 (2019).
53. Lazaridis, I. *et al.* Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**(7617), 419–424. https://doi.org/10.1038/nature19310 (2016).
54. Elkamel, S. *et al.* The orientalisation of North Africa: New hints from the study of autosomal STRs in an Arab population. *Ann. Hum. Biol.* **44**(2), 180–190. https://doi.org/10.1080/03014460.2016.1205135 (2016).
55. Lucas-Sánchez, M., Fadhlaoui-Zid, K. & Comas, D. The genomic analysis of current-day North African populations reveals the existence of trans-Saharan migrations with different origins and dates. *Hum. Genet.* **142**(2), 305. https://doi.org/10.1007/S00439-022-02503-3 (2023).
56. Podgorná, E., Soares, P., Pereira, L. & Černý, V. The genetic impact of the lake chad basin population in North Africa as documented by mitochondrial diversity and internal variation of the L3e5 haplogroup. *Ann. Hum. Genet.* **77**(6), 513–523. https://doi.org/10.1111/AHG.12040 (2013).
57. Brooks, N. *et al.* The climate-environment-society nexus in the Sahara from prehistoric times to the present day. *J. N. Afr. Stud.* **10**(3–4), 253–292. https://doi.org/10.1080/13629380500336680 (2007).
58. Ern, V. *et al.* Migration of Chadic speaking pastoralists within Africa based on population structure of Chad Basin and phylogeography of mitochondrial L3f haplogroup. *BMC Evol. Biol.* **9**(1), 63. https://doi.org/10.1186/1471-2148-9-63 (2009).
59. Massicotte, P. & South, A. *rnaturalearth: World Map Data from Natural Earth*. https://docs.ropensci.org/rnaturalearth/ (2023).

Acknowledgements

The authors would like to thank all the volunteers involved in this study. This work was supported by the Spanish Ministry of Science and Innovation (Grant Number PID2019-106485GB-I00) and "Unidad María de Maeztu" (CEX2018-000792-M) funded by the MCIN and the AEI (https://doi.org/10.13039/501100011033).

Author contributions

J.A.-I., D.C. and F.C. designed the study. J.A.-I. conducted the analysis. J.A.-I., D.C. and F.C. contributed to the interpretation of the data. J.A.-I. wrote the manuscript with help of D.C. and F.C., and A.A. and T.B. contributed with the sampling and helped contextualizing the results. All authors revised and approved the manuscript.

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/s41598-023-37549-4.

Correspondence and requests for materials should be addressed to D.C.

Reprints and permissions information is available at www.nature.com/reprints.

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