



OPEN

The origin of sheep settlement in Western Mediterranean

Asmae Kandoussi^{1,2}, Ismaïl Boujenane¹, Clément Auger², Bruno Serranito², Agnès Germot², Mohammed Piro³, Abderrahman Maftah², Bouabid Badaoui⁴ & Daniel Petit²

The arrival of Neolithic culture in North Africa, especially domestic animals has been essentially documented from archaeological records. As the data relative to sheep are scarce, we studied the genetic relationship between Moroccan sheep breeds and Mediterranean ones using the sequencing of 628 bp of the mitochondrial DNA control region in 193 Moroccan individuals, belonging to six breeds, and 652 sequences from other breeds in Europe and Middle East. Through Network analysis and an original phylogenetically derived method, the connection proportions of each Moroccan breed to foreign ones were estimated, highlighting the strong links between Moroccan and Iberian breeds. The first founders of Moroccan sheep population were issued at 79% from Iberia and 21% from a territory between Middle East and Africa. Their calculated expansion times were respectively 7,100 and 8,600 years B.P. This suggests that Neolithization was introduced by a double influence, from Iberia and from another route, maybe Oriental or Sub-Saharan. The consequence of the environmental changes encountered by founders from Iberia was tested using different neutrality tests. There are significant selection signatures at the level of Moroccan and European breeds settled in elevated altitudes, and an erosion of nucleotide diversity in Moroccan breeds living in arid areas.

The introduction of domesticated mammals in Northwest Africa was linked to a change in food production from hunting-gathering to farming¹. These changes were the result of cultural exchanges, followed by population movements well documented in Western Mediterranean². Recent researches conducted by several archaeological teams proposed, with a high accuracy, the different steps of the diffusion of Neolithic, since about 120 sites have been studied and dated in the range of 12,000 (Epipalaeolithic) to 4,000 years B.P. (late Neolithic)³. Indices of cereal and legume cultivation were attested at around 7,460 to 7,200 years B.P. and of sheep domestication at around 7,460 years B.P., at Kaf-That-el-Ghar in the Mediterranean side of the Tangier Peninsula⁴. It has been questioned whether the route of the migration was from south to north across the Gibraltar strait or the reverse^{2,5}. Beside the slight delay relatively to south Spain Neolithic culture recorded at around 7,700 years B.P.¹, the common pottery production in either sides of Gibraltar strait argue for the European source of this culture found in Northwest Africa^{6,7}. The progressive development of agriculture was associated with a local warming deduced from the measurement of $\delta^{18}\text{O}$ on shell margin of *Phorcus turbinatus* specimens collected and consumed by humans in Ifri Oudadane site⁸.

The origin of local sheep populations in Morocco is still obscure. The different scenarios reported in the literature⁹ are controversial, and many of them are not based on convincing evidence. The majority of these authors reported that except the Mountain population that has existed in the country for a very long time, the other local breeds were introduced into the country by Phoenicians, Romans, Arabs or brought from the sub-Saharan regions. The main breeds recognized by the public authorities with a well-defined standard are the Sardi, Timahdite, Boujaad, D'man, Beni-Guil, Blanche de Montagne and the Noire de Siroua, whereas the remaining breeds are composed by non-standardized populations. At the national level, characterization work on sheep breeds has been based on morphological characters¹⁰ and blood protein polymorphism¹¹. Recently, a study based on SNPs markers highlighted the differentiation between most of Moroccan and Algerian breeds, with a special focus on inter-population exchanges¹². However, the information about the link between Moroccan and European breeds is still lacking. The use of mitochondrial DNA (mtDNA) makes it possible to disentangle the

¹Department of Animal Production and Biotechnology, Institut Agronomique et Vétérinaire Hassan II, PO Box 6202 Rabat-Instituts, 10101, Rabat, Morocco. ²Glycosylation et différenciation cellulaire, EA 7500, Laboratoire Peirene, Université de Limoges, 123 av. A. Thomas, 87060, Limoges, Cedex, France. ³Department of Medicine, Surgery and Reproduction, Institut Agronomique et Vétérinaire Hassan II, PO Box 6202 Rabat-Instituts, 10101, Rabat, Morocco. ⁴Laboratory of Biodiversity, Ecology and Genome, Mohammed V University, 4 Avenue Ibn Battouta, B.P, 1014 RP, Rabat, Morocco. e-mail: daniel.petit@unilim.fr

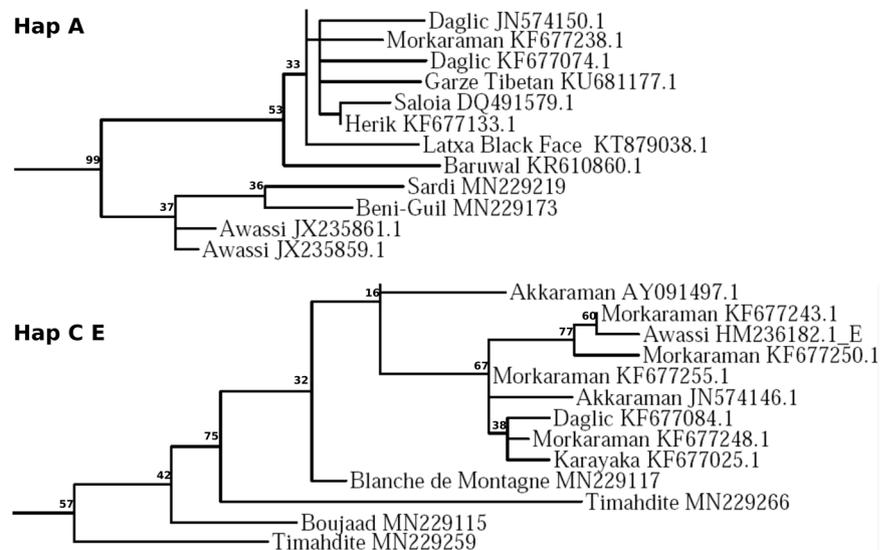


Figure 1. Position of the Moroccan breeds at the base of haplogroups A and C E. The bootstrap supports were calculated from 500 iterations.

origin, evolutionary history and phylogeny of many populations. So far, five maternal lineages, or haplogroups, have been identified in domestic sheep around the world: haplogroups A of Asian origin, B of European origin, C encountered in South Asia, and D and E, considered as rare, encountered in Eurasia^{13–17}. According to Tapio *et al.*¹⁶ and Sanna *et al.*¹⁸, the haplogroup B expanded around 6,400 years ago and reached Western Europe before the haplogroup A. As a result, many breeds represent an admixture of two to three haplogroups.

The aim of this work was to assess the relationships between Moroccan and Mediterranean breeds by using the control region (CR) of the mtDNA in order to reveal the history of sheep settlement in western Mediterranean, including their periods of expansion. Moreover, it was interesting to compare the genetic variability of such characterized breeds to foreign ones, and to highlight the selective pressure that might have affected the Moroccan breeds relatively to Mediterranean ones. Among the different sources of selection, we focused on the climatic constraints, taking advantage of the highly contrasted areas of Morocco described in a previous work¹⁹.

Results

Database and conventional phylogenetic analysis. A 628 bp-long sequence of the control region of mtDNA were sequenced from 193 Moroccan individuals, belonging to 6 breeds (access numbers in GenBank: MN229085–MN229277). To enrich the database, the use of these sequences as seeds allowed retrieving from NCBI Blastn tool 652 foreign sequences, distributed in 9 Iberian, 7 Italian, and 10 Oriental breeds. There are two non-native breeds: the Lacaune of Italy and the German Merino. The number of individuals per breed, their origin and their access number are indicated in Suppl. S1. 1.

The phylogenetic analysis generated a tree encompassing a total of 845 individuals, rooted by individuals of C and E haplogroups according to Meadows *et al.*²⁰, and displayed in Suppl. S2. Among the 193 Moroccan individuals, 185 were found to belong to haplogroup B, i.e. 96% (Suppl. S1. 2). It is worth noting that at the base of the C and E haplogroup clade, there is a series of four embedded Moroccan individuals (GenBank access numbers: MN229117, MN229266, MN229115, MN229259). Similarly, there are two sequences of Moroccan breeds also at the base of the A haplogroup clade (Fig. 1) (GenBank access numbers: MN229219, MN229173). Conversely, at the root of the B haplogroup clade, there are three Moroccan individuals associated to Portuguese Churra Badana breed, two others to Italian Merinizzata Italiana breed, and the last one to the Turkish Herik breed.

Variation in the composition and connectivity among regions and breeds. The treatment of the topology information at the level 1, i.e. terminal branches, using Newick-Extra produced the table of affinities between breeds (Suppl. S1. 3), and then their proportions for each breed. As a first approach, the non-metric multidimensional scaling (N-MDS) analysis (Fig. 2a) revealed the isolated position of Oriental breeds, and the connections between Moroccan and Iberian breeds on one hand, and between Iberian and Italian breeds on the other hand. A more detailed view was given by the Cluster analysis (Fig. 2b), which highlighted three groups: (i) Oriental breeds, (ii) Moroccan breeds associated to four Portuguese breeds, and (iii) a mix of Iberian and Italian breeds.

In the same perspective of studying connectivity and kinship between populations, the genetic distances D_{xy} and D_a between breeds (Suppl. S1. 4) were visualized by NMDS. In both graphs (Suppl. S3. A–B), the Moroccan, Iberian and Italian breed ellipses were strongly superposed, while the Oriental breeds were scattered for D_{xy} genetic distance and relatively grouped for the D_a genetic distance. These results were confirmed by Cluster Analyses (Suppl. S3. C–D).

The use of Network software v.5.0.0.1 showed seven groups. On the left and the right sides, two groups gather individuals of haplogroups C–E and A respectively, dominated by animals of Oriental breeds (Fig. 3).

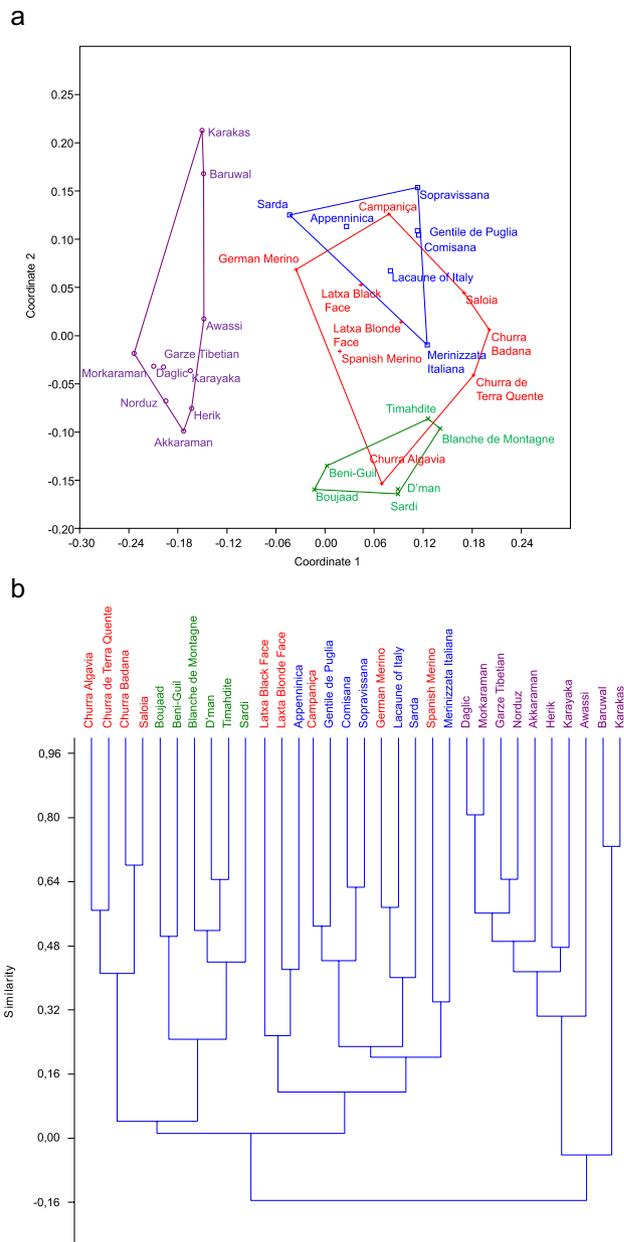


Figure 2. Relationship between breeds according to their geographic origin, from Newick-Extra program. (a) N-MDS analysis showing the intermediate position of Iberian breeds among western Mediterranean ones. (b) Cluster analysis highlighting the proximity between Moroccan and Portuguese breeds. Violet = Oriental breeds, Blue = Italian breeds, Red = Iberian breeds, Green = Moroccan breeds. The figure was drawn using PAST software v. 2.97, available at <https://folk.uio.no/ohammer/past/>.

The situation concerning haplogroup B is more complex as the individuals were organized in four groups: one Italian, one Iberian and two Moroccan ones, named 1 and 2, represented by 119 and 32 individuals respectively, making 62% and 17% of the actual Moroccan genetic background. The Iberian one was localized in the center and closer to the Moroccan group 1 than to the Italian one. The Moroccan group 2 was linked to the Italian one and the ensemble composed of individuals of haplogroups C and E. At a closer look, it appears that (i) the Italian and Iberian groups included 20 and 14 Moroccan individuals respectively, corresponding to c. 10% and 7% each of the total from Morocco; (ii) moreover, the Italian group also contained several Iberian (Spanish Merinos only) and Oriental individuals; (iii) the Iberian group itself hosted individuals of Oriental, Italian, and of the Moroccan group 1. As regard haplogroup D, composed of a few oriental individuals, its connections concern haplogroup A group and two ensembles belonging to haplogroup B, i.e. Iberian group and Moroccan group 1.

To test whether the genetic composition of the three western Mediterranean areas (Moroccan, Iberian, and Italian) was different, the AMOVA revealed significant molecular differences among breeds and within breeds for the 3 groups. Regardless of the compared groups, the main source of variation lay within populations, exceeding 96%. More interestingly, it appeared that there were significant differences between all the combinations of

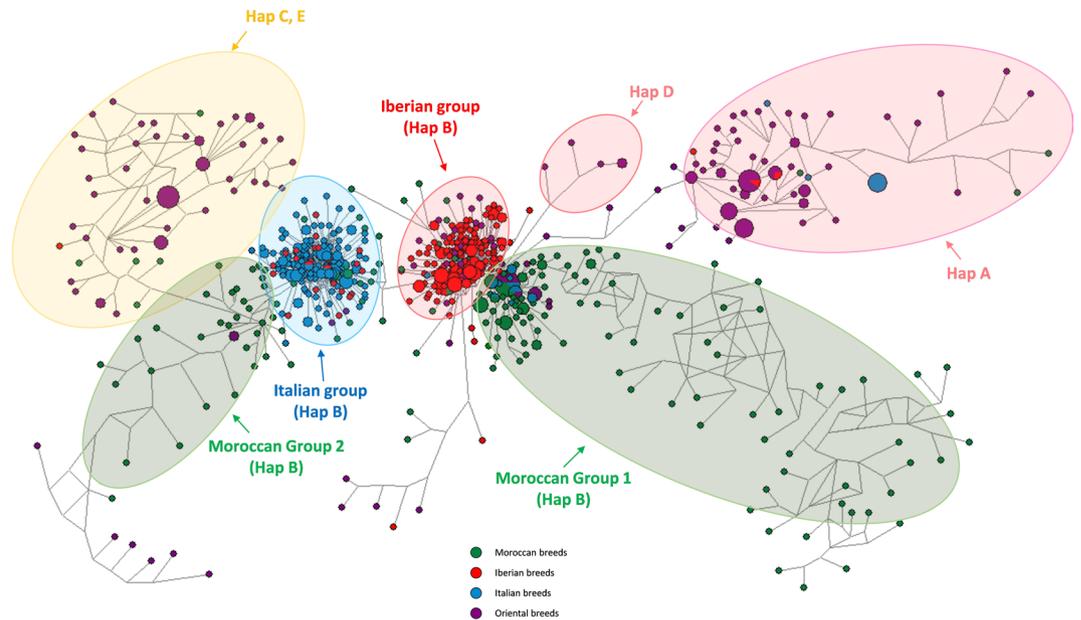


Figure 3. The individuals are organized in four haplogroup ensembles. The haplogroup B is dominant and comprises the Italian, the Iberian and the two Moroccan groups. The figure was drawn using Network software v.5.0.0.1, available on the website: <http://www.fluxus-engineering.com/>.

groups, except between Iberian and Italian ones. At a closer look, the difference between Moroccan and Iberian breeds was lower than between Moroccan and Italian ones (Suppl. S4. A).

In summary, the structuration of the western Mediterranean populations was relatively concordant with the results of procedures involving the Network, the AMOVA and Newick-Extra program, while the D_a and D_{xy} genetic distances provided a confusing view. To test specifically the level of compatibility between the Newick-Extra program and the Network methods, we generated 2 matrices that presented the percentages of connections between each Moroccan breed and the other breeds according to their geographical origin (Moroccan, Iberian, Italian or Oriental). In contrary to the Newick program for which this counting was straightforward, the count of connections from the network needed to be done manually. For each Moroccan breed, the numbers of individuals belonging to Italian, Iberian, Moroccan (1, 2 and 3) or Oriental (Haplogroups A, C, D, and E) were summed (Suppl. S1. 5). The correlation between the two matrices was highly significant ($N = 24$, $r = 0.93$, $p < 0.001$).

For each Moroccan breed, the percentages of affinity (Fig. 4) showed contrasted situations toward Italian, Iberian and Oriental breeds. Beni-Guil breed had the highest proportion of influence from foreign countries (about a third), especially from Iberia, while D'man breed affinity was specifically Moroccan. The remaining Moroccan breeds had about a fifth relationship with foreign breeds, the Timahdite being the only breed devoid of connection with oriental breeds. There were moderate connections between Moroccan breeds themselves, except for the D'man, which showed more than 20% of relationship with the Timahdite, the Blanche de Montagne and the Sardi breeds. The analysis of the phylogenetic tree (Suppl. S2) indicated that all these connections were situated in the terminal branches of haplogroup B clade, and thus corresponded to recent exchange events.

Climatic parameters. As it seems that the Italian, Iberian and Moroccan regions are centers of differentiation of sheep populations, it is useful to characterize the climatic constraints that exist in these three regions. The ANOVA conducted on the sum of localities recorded for each breed within each region lead to significant differences for most parameters (Suppl. S1. 6 and S3. E). As the altitudes recorded in the breeding areas in Morocco are higher than in Italy and Iberia, the winter temperatures expressed by m are colder. Given the south position of Morocco, the summer temperatures expressed by M are higher, leading to a decrease in rainfall (R) and vegetal productivity ($Q3$).

With regard to the breeds themselves, the mean values for climatic parameters were gathered in S1. 7.

Genetic variability parameters. Among the groups defined from haplogroup B, it appeared that the Moroccan group 1 had a significant lower haplotype diversity H_d than the Iberian and Italian ones, but a higher nucleotide diversity π in relation to numerous singletons. At the level of breeds, it was expected that the variations in genetic diversity between Moroccan breeds followed the degree of affinity to foreign ones showed in Fig. 4. Effectively, the haplotype diversity H_d had its highest and lowest values in Beni-Guil and D'man, respectively (Suppl. S4. B., Suppl S1. 2 for all breeds).

Neutrality tests and climatic constraints. The calculated neutrality tests F_u and $Li F^*$ and D^* among the four ensembles defined from haplogroup B, indicate lower significant values in Moroccan group 2 and Italian group than in the two other ones. These same tests applied to Moroccan breeds highlighted the Boujaad,

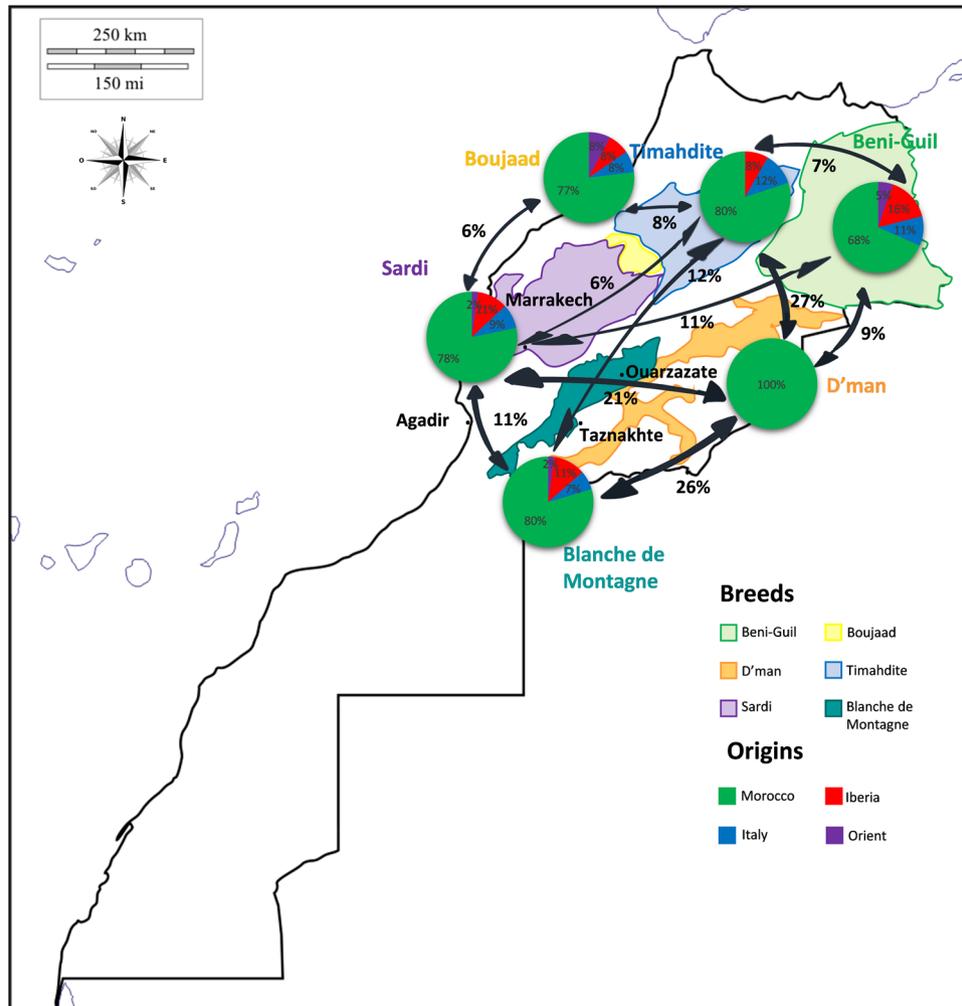


Figure 4. Map of connections between each Moroccan breed and (i) other Moroccan ones, and (ii) breeds of other Mediterranean areas.

the Blanche de Montagne and the Beni-Guil breeds, partially retrieved using Tajima D test. In contrast, Fu's F_s revealed that all the Moroccan breeds were significantly out of neutrality. As it is difficult to quantify a direct influence of human selection by neutrality tests, it seems easier to investigate the effects of environmental constraints, especially the climatic parameters. For this purpose, a synthetic database was constructed by including both climatic and genetic data (Suppl. S1. 7). The relationships among all the climatic and genetic variables were well represented on the PCA first plan (70.02% of variance) (Fig. 5). The pluviometric coefficient Q_3 and annual rainfall R varied in opposite to the temperature variables (M, T max, M - m) and altitude. In contrast, the variables m and T min seems to vary in the same direction as most genetic variables except H_d . With regard to breeds, the Moroccan ones and the Iberian Spanish Merino and Churra Badana breeds were located towards high temperature and low rainfall variables. Another set of breeds was clearly associated to high rainfall, e.g. Sarda, Sopravissana, Saloia, Latxa Blonde and Black face. There is another set of three breeds (Campaniça, Churra Algarvia and Comisana) sharing high values of neutrality test variables. The result of correlation indicated that the variables T min and m were positively and significantly associated to π , Fu and Li F* and D*, and to some extent to D of Tajima (Suppl. S1. 7, Suppl. S3. F-H). If the breeds for which the significant neutrality tests Fu and Li F* and D* were considered (Blanche de Montagne, Boujaad and Beni-Guil), it appears that they were located in the PCA plan towards higher altitudes, low minimal temperatures (m and T min) and higher maximal temperatures (M and T max).

Expansion time. To assess the prospective expansion events of western Mediterranean native breeds, the calculation of mismatch distribution revealed that the four groups of haplogroup B and breeds followed the coalescent simulation (p-value of SSD test >0.1, Suppl. S3. I,J), except the limit case of Comisana (p-value of SSD test = 0.08). The τ variations did not correlate to any climatic parameters (Suppl. S1. 7). From the first method of expansion dating, the yearly mutation rate could be estimated at 3.4×10^{-5} and 2.3×10^{-5} using 1.26 and 1.72 Myrs ancestors, respectively, resulting in expansion times at 140,000 and 207,000 years for the D'man breed for example. From the second method using archaeological records of Italy and Iberia, it appears that the Moroccan group 1 seems to have expanded later than Iberian one around 7,100 years B.P., whereas the Moroccan 2 group

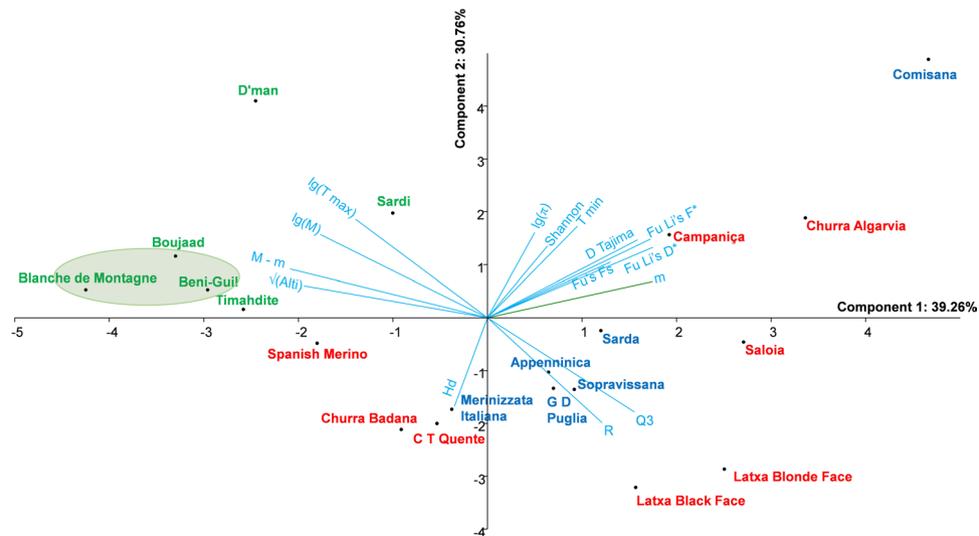


Figure 5. PCA of the climatic and genetic parameters. The ellipse gathers the breed significantly under selection (Fu and Li D* and F* parameters). Color code: see Fig. 2A. The figure was drawn using PAST software v. 2.97, available at <https://folk.uio.no/ohammer/past/>.

showed an earlier expansion than the Italian one itself around 8,600 years B.P. Due to the low population size of Moroccan individuals recorded in Italian and Iberian groups, no mismatch analysis could be done. With regard to the Moroccan breeds, the northern ones (Timahdite, Sardi, Beni-Guil and Boujaad) are in the range of c. 6,700–7,400 years B.P., while the southern ones are associated to 7,700 years B.P. for the Blanche de Montagne and 8,400 years B.P. for the D'man (Suppl. S1. 8).

Discussion

The central issue of the study was to understand the extension of Neolithic culture associated with sheep domestication in western Mediterranean. The results obtained provide a new vision of the subject, and exclude the hypotheses emitted by historians who did not think to movements of sheep populations earlier than a few hundred years B.C.E.⁹

The different approaches conducted in the present study brought complementary information on the origin of Moroccan sheep breeds. If the Cluster Analysis generated by phylogeny and Newick-Extra method were considered, all the Iberian, Italian and Moroccan breeds are clearly separated from Oriental ones. This suggests that the western Mediterranean populations have undergone a long period of evolution during which numerous mutations have been accumulated. Moreover, it appears that the Moroccan sheep breeds are closer to Portuguese than to Spanish ones and even more so the Italian ones. This view is supported by the network results, where Iberian group is situated between the Italian and Moroccan groups. However, there are two Moroccan groups, a majority group 1 representing around 62% of the genetic background and linked to Iberian populations, and a minority group 2 around 17% having relationships with Italian and Oriental breeds. The individuals situated at the base of A and CE haplogroups, related to Oriental breeds, could be assigned to the Moroccan group 2 in reason of their link and ancient origin, making this group up to c. 20% of the total genetic background. The question raised is to determine the significance of these 2 groups in the light of archaeological and expansion time evidences. For this purpose, the method based on mutation rate of control region lead to unrealistic dates, an issue previously discussed by Guo *et al.*¹⁴ who found a mutation rate of the same order of ours. It can be hypothesized a low accumulation of mutations in the genus *Ovis* (about a tenth) or an over-estimation for the dates of the two calibration ancestors. Using the second method based on the dating of early sheep remains in Italy and Iberia, the expansion date of Moroccan group 1 for example at 7,100 years B.P. is close to archaeological records at c. 7,400. As a result, the data provided by this second method were retained.

There is a progressive consensus among archaeologists for an Iberian origin of most features of Neolithic culture in Northwest Africa²¹, although Sánchez *et al.*²² argued for a reverse way for some aspects. This view was globally supported by the molecular dating and network approach of the present study, since the Moroccan group 1 expanded later than Iberian populations around 7,100 years B.P. This estimation is close to the indices of domestic sheep remains mentioned at around 7,460 years B.P., at Kaf-That-el-Ghar near Tangier Peninsula⁴. Concerning human populations, a work dealing with Ifri n'Amr or Moussa (at around 7,000 years B.P., early Neolithic) and Kelif el Boroud (at around 5,000 years B.P., upper Neolithic) near the Atlantic coast, showed that the early Neolithic populations were isolated, in the continuity of the upper Paleolithic ones in relation to Berber settlement, while the upper Neolithic individuals were strongly influenced by Iberian populations²¹.

About the origin of Moroccan group 2, the possibility of an ancient settlement of population from a route different from the Gibraltar strait was explored as for goat populations²³. The absence of any relationship with Iberian breeds of haplogroup B strongly suggests that these individuals came to Morocco from Middle East via Africa, maybe from a south Italian – Tunisia route, or from an entirely African route. The connection between

Italian and Moroccan individuals of group 2 support the 1st idea. In contrast, the existence of several Moroccan individuals at the root of A and C E haplogroup clades argue for the second solution. Moreover, the dating of Moroccan group 2 expansion around 8,600 years B.P., more ancient than for Italian group, seems to favor this second solution. Unfortunately, the archaeological works are useless until now as the earliest sheep remains out of Morocco are dated around 7,000–7,500 years B.P. in Sinai Peninsula²⁴. If our calculations were not overestimated due to calibration bias, we think that new investigations should be conducted in the south of Morocco, and of course in other areas of Africa, in order to detect very early Neolithic sheep remains. This idea does not appear to be unrealistic, since there is a lack of archaeological sites for the period around 8,000 years B.P., especially in the south of Atlas region³.

It is interesting to draw a parallel with the genetics of ancient humans. Van de Loosdrecht *et al.*²⁵ established that between 15,000 and 14,000 years B.P. (upper Paleolithic), the Taforalt's population, in the Mediterranean coast of Morocco, presented affinities with the population of Levant and to sub-Saharan Africans. However, the studies on the genetics of Sahelian sheep are still scarce^{26–28}.

Besides the early founders of the two Moroccan groups, there were 20 individuals (around 10%) and 14 (around 7%) showing very similar sequences to Italian and Iberian groups, respectively, suggesting late introduction events southward. Conversely, the Moroccan group 1 contained 9 Italian individuals that are interpreted as a late event of importation from Morocco to Italy, maybe at the Roman's epoch: these individuals had stayed a long time in the Moroccan territory, acquiring their genetic identity. Similarly, the Italian group included all the individuals of Spanish Merinos, a view supported by the text of 1st century Roman Columella in chapter 7²⁹, who indicated that the selection of fine wool sheep was conducted using individuals from Tarento region in South East Italy. Moreover, the 13 Oriental individuals found in the Moroccan group 1 could be due to recent trade events toward Turkey and beyond. Following the successive population movements of the founders on Moroccan territory, the animals have progressively differentiated into local populations and then breeds. As shown previously, the four breeds of the North, between 6,700 and 7,400 years B.P. have expanded later than the two breeds settled on south of Atlas, between 7,700 years B.P. (Blanche de Montagne) and 8,400 years B.P. (D'man). This dating suggests that some D'man ancestors would have reach Morocco via another route than Gibraltar strait, maybe from Africa. However, the origin of this last breed remains still obscure and new data on Sahelian breed could enlighten the situation. The four northern ones from Sardi in the West to Beni-Guil in the East were well genetically differentiated with less than 9% of connection. In contrast, the two southern ones, Blanche de Montagne and D'man had strong and recent links with Sardi and Timahdite breeds. The strong proportion of connections between the D'man and the other breeds were also found from the study based on SNPs over Northwest African sheep breeds¹². It can be assumed that the D'man and the Blanche de Montagne, being small sized breeds¹⁰, were crossed with the biggest Sardi and Timahdite individuals to increase their meat production. Otherwise, the D'man being well known for its high prolificacy³⁰, was crossed to Sardi, Timahdite and Blanche de Montagne breeds to improve their reproductive characteristics.

The European genetic component of Moroccan breeds, mostly Iberian, is actually highly dominant compared to the native "Berber" one. Given the climatic contrast between the two areas from each side of Gibraltar strait, it was expected that the drought and the high temperatures observed in Morocco would induce a selection pressure on the sheep flocks. Effectively, the shortage of rainfall and low vegetal productivity expressed by Q₃ decreased the nucleotide diversity at the level of Moroccan breeds. At the level of European and Moroccan breeds, this correlation is not true, but the low temperatures (T min and m) induced a selection signature (Fu and Li F* and D* statistics) for the breeds settled in high lands above 650 m A.S.L (Blanche de Montagne, Boujaad, Beni-Guil, and Timahdite to some extent). The influence of these altitudes on low temperatures had also suggested a decrease of nucleotide diversity.

Conclusion

Much effort has been made in the recent years to document the transition between the Paleolithic and Neolithic cultures in Morocco. The results presented here evidenced two successive waves of domestic sheep arrival on the Moroccan territory, corresponding to this transition: the most ancient one could be associated to the first Berbers around 8,600 years B.P., still representing 20% for the Moroccan genetic background, maybe as a response to the aridification of Northern Sahara at the end of the "African humid period". The second one was the most important with 62% of this same background, and reached Morocco around 7100 years B.P. from Iberia. Two further independent importations of individuals from Iberia and Italy were recorded, less or equal to 10% of the genetic diversity each, but hardly datable due to their weak numbers. Archaeologists have shown an increase in exchanges with Iberia since 7,500 years B.P., in good agreement with the dating conducted on the expansion of most Moroccan breeds. However, for the special case of the D'man breed, the calculated expansion exceeded the period of human movements between Morocco and European continent, suggesting non-European route for the introduction of this breed ancestors. The human genetic researches argued for a continuity of local populations from broadly 15,000 to 7,000 years B.P., more or less related to Middle East, South and East African ones, and a sudden change of population since 5,000 years B.P., corresponding to a transition from hunting-gathering to agricultural practices.

Material and methods

Sample collection, DNA isolation and sequencing. Blood sampling was performed by veterinarians of the ANOC (Association Nationale Ovine et Caprine) during routine medical care or follow up, such that no ethical authorization was required. Breeders and ANOC consented to the blood samplings and their utilization for scientific purpose.

A total of 193 blood samples were taken from sheep belonging to six Moroccan sheep breeds: Sardi (n = 33), Timahdite (n = 37), Boujaad (n = 31), Beni-Guil (n = 31), D'man (n = 29), and Blanche de Montagne (n = 32).

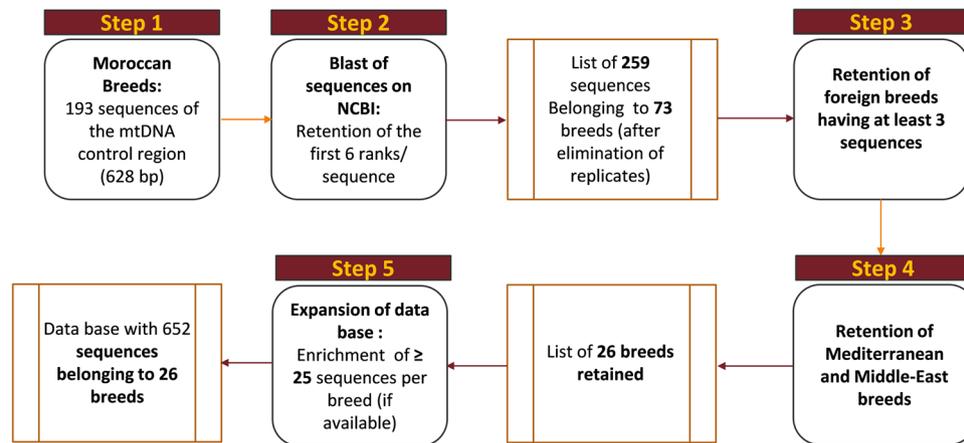


Figure 6. Workflow for data base construction.

The blood samples were taken by an authorized person. The individuals were collected from farm breeders in the breeding areas of different breeds and supervised by the “Association Nationale Ovine et Caprine” ANOC, in order to ensure that they correspond to native breeds without introgression with foreign commercial or other local breeds.

Specimens were chosen according to their genealogy so that they are unrelated. Blood samples were taken from the jugular vein and collected in ethylene diamine tetra acetic (EDTA) acid tubes and frozen at -20°C until extraction. Total genomic DNA was extracted from blood samples using an alkaline lysis method³¹. Details about PCR amplification and sequencing are reported in Suppl. S4. C. The sheep breeds sequences were recorded in GenBank (accession numbers in Suppl. S1. 1).

Mitochondrial sequence analysis. *Construction of database.* We built a database of control region (CR) sequences including 193 from Moroccan and 652 from non-Moroccan breeds. This database was constructed by exploring the NCBI gene bank using BLASTn algorithm. Taking each Moroccan sequence as seed, we consistently selected at least the top six ranks per blasted sequence. Then, we looked at the following ranks to possibly retain those who had the same score as the sixth. Among all the breeds obtained, we retained those that showed more than three individuals, belonging to Mediterranean, African, and Near East regions. Then, we expanded the database by including 25 sequences or more per breed if available. The sequences selected were then recorded with their access number and the reference to their breed (Fig. 6). Furthermore, the data matrix included 5 mtDNA control region sequences from sheep representing the five mitochondrial haplogroups (A, B, C, D, E), in order to correctly assign each sequence to its proper haplogroup.

Analytical methodologies. The sequences were corrected and aligned using MEGA software v.10.0³², resulting in 628 pb available for further treatments.

In order to determine the genetic variation within populations, several diversity parameters were calculated with the DNAsp v5 program³³: the number of haplotypes (H), the haplotype diversity (H_d), the nucleotide diversity (π)³⁴, the number of segregation sites (S), the number of singletons (S_g) and the number of parsimoniously informative sites (P). In addition, four neutrality tests were performed: Fu and Li F*, Fu and Li D*³⁵, Fu Fs³⁶ and D of Tajima³⁷ using DNAsp for each Moroccan, Iberian and Italian breed. The number of segregation sites (S), the singletons (S_g) as well as the parsimoniously informative sites (P) were calculated for each breed by the “Polymorphic sites” option, while the remaining diversity indices were calculated by the “DNA Polymorphism” option. To calculate these previous parameters for each of the four groups within haplogroup B, the sequence numbers of Iberian and Italian ones were adjusted by a random method to the number recorded in Moroccan group 1, i.e. 119, to avoid a sampling size effect. The haplogroup frequencies were deduced from the phylogenetic analysis, and the haplogroup diversity from Shannon index calculated using PAST software v.2.97³⁸.

To estimate the divergence between populations (breeds), the genetic distances D_a (the corrected average pair-wise differences) and D_{xy} (the average number of pair-wise differences)³⁴ were calculated by DNAsp v5³³. The obtained results were highlighted with the help of “Non-Metric Multidimensional Scaling Plot” and Cluster Analysis with User distance using PAST software v.2.97³⁸. Connectivities as well as evolutionary relationships between different individuals were also analyzed through the construction of a phylogenetic tree by the Neighbor-Joining method, using the NETWORK software v.5.0.0.1 (available on the website: <http://www.fluxus-engineering.com/>). The Median-Joining Network³⁹ is an haplotype network of all individuals. To reduce the large data set, the “Star Contraction” option was used. In order to simplify the graphical representation of the constructed network, the optional “MP calculation” procedure according to Neighbor-Joining method was used to remove unnecessary vectors and median links and to avoid reticulations. In addition, the length of the branches of some individuals were also modified manually, when necessary, to improve the visualization of results.

A phylogenetic analysis was carried out in order to reveal the affinities between Moroccan breeds and themselves and also to other breeds in the world. The phylogenetic tree was constructed using the Maximum Likelihood (ML) method implemented in MEGA software v.10.0³². The best model according to the lowest

Bayesian Information Criterion corresponded to Hasegawa–Kishino–Yano model⁴⁰, including Gamma distribution (5 discrete categories) and Invariant sites (G + I). The gap treatment was done by eliminating segments that had less than 95% presence (Partial Deletion). The robustness of branches was tested using 500 bootstrap replicates. The tree was redrawn using Seaview v.4⁴¹.

As expected, the phylogenetic tree produced several sets of individuals corresponding to each haplogroup, preventing to evidence the relationships between breeds. The use of the conventional genetic distance indices produced the same effects. To overcome this problem, we developed a new phylogeny-derived method called Newick-Extra (Suppl. S4. D).

Relationships between the studied breeds (asymmetric matrix generated by Newick-Extra) were visualized through the Non-Metric Dimensional Scaling Plot (NMDS) and Cluster analysis using Spearman's Rho as distance metrics, using PAST software v.2.97³⁸.

The Analysis of MOlecular VAriance (AMOVA) was computed to test significant differences between geographical breed groups, using ARLEQUIN v.3.5⁴². Sources of variation were considered in a single model: among groups, among breeds – within groups – and within breeds. Mismatch distribution was used for each European and Moroccan breed in order to address eventual demographic expansion events, by computing the τ parameter. This last one has shown to be useful to date historical expansion of human^{43,44}, wild⁴⁵, and domestic animals, e.g. goats⁴⁶. To test the validity of an expansion hypothesis, the observed mismatch distribution was compared to the simulation values generated by coalescent algorithm using the probability associated to the sum of square deviation (SSD)⁴⁴. In order to calculate the expansion time of breeds, two strategies were conducted. The first one used the formula $\tau = 2\mu t$, where μ is the mutation rate for the whole considered segment and t is time in generations⁴⁷. To calculate the mutation rate (μ) of the 628 pb segment, a data set was built, including two individuals of *Ovis aries*, *O. ammon*, *O. vignei*, per haplogroup and one for *O. orientalis*, *O. dalli* and *O. nivicola*. *nivicola* (access numbers in Suppl. S1. 1). After manual alignment, the topology given by ML with the same options as previously was saved as a user tree to run DNAPars implemented in PHYLIP software v.3.695⁴⁸, giving the possibility to get an exact count of mutations in the different branches. The ancestor of *O. vignei*, *O. orientalis*, and *O. aries* was calibrated at 1.26 Myrs, and the ancestor of this one with *O. ammon* at 1.72 Myrs⁴⁹. The second method was based on a calibration using τ values for Iberian and Italian groups, defined in the network analysis, and archaeological records. According to Rowley-Conwy *et al.*⁵⁰ and Sánchez *et al.*²², the earliest domestic sheep remains were found in South-East Italy around 8000 years B.P. and near Malaga in South East Spain around 7500 years B.P., respectively. The regression equation deduced was used to calculate the dates corresponding to τ values with 95% credible intervals for each group and breed.

Climatologic data. The aim was to define the optimal climatic conditions for the breeds in order to detect eventual sources of natural selection on genetic variations. The methods developed by Petit and Boujenane¹⁹ was extended to the Moroccan Blanche de Montagne, Iberian and Italian studied breeds. The Oriental animals were not considered because they were distantly related to Moroccan breeds. Briefly, the geographic perimeter of the breeding area of a given breed was obtained through bibliographic data: anoc.ma, agraria.org, sprega.com and confelac.org for Moroccan, Italian, Portuguese and Spanish breeds, respectively. Concerning the Spanish Merino breed, information was taken from Lancioni *et al.*⁵¹. The following climatic parameters of at least 10 localities distributed in the breeding area were recorded from climate-data.org: Alti: altitude (m), m: minima mean of the coldest month (°C), M: maxima mean of the hottest month (°C), R: annual rainfall (mm), T max: annual maxima mean (°C), T min: annual minima mean (°C) for the period 1984 to 2014, and $Q_3 = \frac{3.43R}{M-m}$ or pluviometric coefficient⁵². For each breed, the means of these parameters recorded for the different localities were calculated.

To test whether the regions showed significant differences in the climatic parameters, a one-way ANOVA conducted using SYSTAT vers. 12⁵³ was considered with three modalities (Iberia, Italy and Morocco). The non-autochthon breeds (German Merino and Lacaune) were removed from the lists of Iberian and Italian breeds, respectively.

The relationships between genetic and climatic data were studied using principal component analysis and correlation analysis implemented in PAST software v.2.97³⁸. Prior to the calculations, non-normal distributions were transformed (square root for Alti, Log for T max, M, and π). However, D of Tajima and the Shannon index were non-transformable.

Data availability

Newly-reported sequences in GenBank: MN229085–MN229277.

The program written in R Newick-extra is available from BS on request.

All data generated or analyzed during this study are included in this published article (and its Supplementary Information files).

Received: 8 October 2019; Accepted: 21 May 2020;

Published online: 23 June 2020

References

- Zeder, M. A. Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proc. Natl. Acad. Sci. USA* **105**, 11597–11604 (2008).
- Zilhão, J. Early prehistoric navigation in the Western Mediterranean: Implications for the Neolithic transition in Iberia and the Maghreb. *Euras. Prehist.* **11**, 185–200 (2014).
- Linstädter, J. Climate induced mobility and the missing Middle Neolithic of Morocco in: *Palaeoenvironment and the Development of Early Settlements. Menschen–Kulturen–Traditionen* (ForschungsCluster 1), M. Reindel, K. Bartl, F. Lüth, N. Benecke, Eds. (Rahden, Westf., Leidorf) 63–80 (2016).

4. Martínez Sánchez, R.M., *et al.* Beyond Adornment: Cowry Use as Potter's Tool in the First Impressed Wares of the Southwestern Mediterranean Coast (Northern Morocco). *J. Isl. Coast Arch.* **13**: 420–437 (2018a).
5. García Borja, P. J., Tortosa, A., Emili, J., Jordá Pardo, J. F. & Salazar García, D. C. La cerámica neolítica de la Cueva de Nerja (Málaga, España): salas del Vestíbulo y la Mina. *Arch. Prehist. Levant.* **30**, 81–131 (2014).
6. Martínez Sánchez R.M., *et al.* The Middle Neolithic of Morocco's North-Western Atlantic Strip: New Evidence from the El-Khili Caves (Tangier). *Afr. Archaeol. Rev.* **35**: 417–442 (2018b).
7. Martínez-Sánchez, R.M., Vera-Rodríguez, J.C., Pérez-Jordá, G., Peña-Chocarro, L., & Bokbot, Y. The beginning of the Neolithic in northwestern Morocco. *Quat. Int.* **470**: 485–496 (2018c).
8. Yanes, Y., Hutterer, R. & Linstädter, J. On the transition from hunting-gathering to food production in NE Morocco as inferred from archeological *Phorcus turbinatus* shells. *Holocene.* **28**, 1301–1312 (2018).
9. Boujenane, I. *Les Ressources Génétiques Ovinnes au Maroc.* (Actes Editions (1999).
10. Boujenane, I. & Petit, D. Between-and within-breed morphological variability in Moroccan sheep breeds. *Anim. Genet. Resour.* **58**, 91–100 (2016).
11. Boujenane, I. *et al.* Variation at post-albumin, transferrin and haemoglobin proteins in Moroccan local sheep. *Small Rumin. Res.* **79**, 113–117 (2008).
12. Belabdi, I. *et al.* Genetic homogenization of indigenous sheep breeds in Northwest Africa. *Sci. Rep.* **9**, 7920 (2019).
13. Hiendleder, S., Mainz, K., Plante, Y. & Lewalski, H. Analysis of mitochondrial DNA indicates that domestic sheep are derived from two different ancestral maternal sources: no evidence for contributions from urial and argali sheep. *J. Hered.* **89**, 113–120 (1998).
14. Guo, J. *et al.* A novel maternal lineage revealed in sheep (*Ovis aries*). *Anim. Genet.* **36**, 331–336 (2005).
15. Pedrosa, S. *et al.* Evidence of three maternal lineages in Near Eastern sheep supporting multiple domestication events. *Proc. R. Soc. Lond. [Biol.]* **272**, 2211–2217 (2005).
16. Tapio, M. *et al.* Sheep mitochondrial DNA variation in European, Caucasian, and Central Asian areas. *Mol. Biol. Evol.* **23**, 1776–1783 (2006).
17. Meadows, J. R., Cemal, I., Karaca, O., Gootwine, E. & Kijas, J. W. Five ovine mitochondrial lineages identified from sheep breeds of the near East. *Genetics.* **175**, 1371–1379 (2007).
18. Sanna, D. *et al.* The first mitogenome of the Cyprus mouflon (*Ovis gmelini ophion*): new insights into the phylogeny of the genus *Ovis*. *PLoS one.* **10**, e0144257 (2015).
19. Petit, D. & Boujenane, I. Importance of determining the climatic domains of sheep breeds. *Animal.* **12**, 1501–1507 (2018).
20. Meadows, J. R., Hiendleder, S. & Kijas, J. W. Haplogroup relationships between domestic and wild sheep resolved using a mitogenome panel. *Heredity.* **106**, 700–706 (2011).
21. 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. USA* **115**, 6774–6779 (2018).
22. Sánchez, M. C. *et al.* The Mesolithic–Neolithic transition in southern Iberia. *Quat. Res.* **77**(2), 221–234 (2012).
23. Pereira, F. *et al.* Tracing the history of goat pastoralism: new clues from mitochondrial and Y chromosome DNA in North Africa. *Mol. Biol. Evol.* **26**, 2765–2773 (2009).
24. Muigai, A. W. & Hanotte, O. The origin of African sheep: archaeological and genetic perspectives. *Afr. Arch. Rev.* **30**(1), 39–50 (2013).
25. van de Loosdrecht, M. *et al.* Pleistocene North African genomes link near Eastern and sub-Saharan African human populations. *Science.* **360**, 548–552 (2018).
26. Álvarez, I. *et al.* Mitochondrial analysis sheds light on the origin of hair sheep. *Anim. Genet.* **44**(3), 344–347 (2013).
27. Brahi, O. H. D., Xiang, H., Chen, X., Farougou, S. & Zhao, X. Mitogenome revealed multiple postdomestication genetic mixtures of West African sheep. *J. Anim. Breed. Genet.* **132**(5), 399–405 (2015).
28. Resende, A., Gonçalves, J., Muigai, A. W. & Pereira, F. Mitochondrial DNA variation of domestic sheep (*Ovis aries*) in Kenya. *Anim. Genet.* **47**(3), 377–381 (2016).
29. Columella, translated by Du Bois L. *L'Économie Rurale, Bibliothèque Latine-Française.* (Panckoucke (1844).
30. Ben Jemaa, S. *et al.* The high prolificacy of D'man sheep is associated with the segregation of the FecLL mutation in the B4GALNT2 gene. *Reprod. Domest. Anim.* **54**, 531–53 (2019).
31. Miller, S. A., Dykes, D. D. & Polesky, H. F. R. N. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16**, 1215 (1988).
32. Kumar, S., Stecher, G., Li, M., Nnyaz, C. & Tamura, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
33. Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics.* **25**, 1451–1452 (2009).
34. Nei, M. *Molecular Evolutionary Genetics.* New-York: Columbia University Press (1987).
35. Fu, Y. X. & Li, W. H. Statistical tests of neutrality of mutations. *Genetics.* **133**, 693–709 (1993).
36. Fu, Y. X. New statistical tests of neutrality for DNA samples from a population. *Genetics.* **143**, 557–570 (1996).
37. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics.* **123**, 585–595 (1989).
38. Hammer, Ø., Harper, D. A. & Ryan, P. D. PAST: PAleontological STatistics software package for education and data analysis. *Palaeontol. Electron.* **4**, 9 (2001).
39. Bandelt, H. J., Forster, P. & Röhl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**, 37–48 (1999).
40. Hasegawa, M., Kishino, H. & Yano, T. A. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**, 160–174 (1985).
41. Gouy, M., Guindon, S. & Gascuel, O. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27**, 221–224 (2009).
42. Excoffier, L. & Lischer, H. E. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10**, 564–567 (2010).
43. Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. & Wilson, A. C. African populations and the evolution of human mitochondrial DNA. *Science.* **253**, 1503–1507 (1991).
44. Schneider, S. & Excoffier, L. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics.* **152**, 1079–1089 (1999).
45. Okello, J. B. A., Nyakaana, S., Masembe, C., Siegmund, H. R. & Arctander, P. Mitochondrial DNA variation of the common hippopotamus: evidence for a recent population expansion. *Heredity.* **95**, 206–215 (2005).
46. Naderi, S. *et al.* Large-scale mitochondrial DNA analysis of the domestic goat reveals six haplogroups with high diversity. *PLoS one.* **2**, e1012 (2007).
47. Harpending, H. C. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum. Biol.* **66**, 591–600 (1994).
48. Felsenstein, J. PHYLIP (phylogeny inference package), version 3.5 c. Joseph Felsenstein (1993).
49. Rezaei, H. R. *et al.* Evolution and taxonomy of the wild species of the genus *Ovis* (Mammalia, Artiodactyla, Bovidae). *Mol. Phyl. Evol.* **54**, 315–326 (2010).

50. Rowley-Conwy, P., Gourichon, L., Helmer, D. & Vigne, J.D. Early domestic animals in Italy, Istria, the Tyrrhenian islands and southern France. In *The Origins And Spread Of Domestic Animals In Southwest Asia And Europe* 161–194, (Routledge (2016).
51. Lancioni, H. *et al.* Phylogenetic relationships of three Italian merino-derived sheep breeds evaluated through a complete mitogenome analysis. *PLoS one*. **8**, e73712 (2013).
52. Sauvage, C. Le quotient pluviothermique d'Emberger, son utilisation et la représentation géographique de ses variations au Maroc. *Ann. Phys. Globe Météorol. Inst. Sci. Chérif*. **20**, 11–23 (1963).
53. SPSS Inc.. *Systat Software Vers 12*. (Chicago, Illinois, (2008).

Acknowledgements

This work was supported by grants from the University of Limoges (AOI 2016–2019 PHYLOSHEEP Project). We thank the ANOC (Association Nationale Ovine et Caprine) team and farmers for their precious help for the sampling of blood, and A. Bouffier, R. Marcillou and M. Safouani for their technical supports.

Author contributions

A.K., I.B., A.G., A.M., and D.P. designed research; A.K. and C.A. performed research; I.B and A.K., performed the blood sampling; A.K., D.P. and B.S. analyzed data; A.K. and M.P. extracted DNA; A.K. and D.P. wrote the paper and made the illustrations; I.B., A.G., A.M., B.B. and B.S. improved the successive drafts.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-67246-5>.

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

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.



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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020