

Article

Circum-Saharan Prehistory through the Lens of mtDNA Diversity

Mame Yoro Diallo ^{1,2} , Martina Čížková ¹, Iva Kulichová ^{2,3}, Eliška Podgorná ¹, Edita Priehodová ¹, Jana Nováčková ³, Veronica Fernandes ^{4,5} , Luísa Pereira ^{4,5}  and Viktor Černý ^{1,*} 

- ¹ Archaeogenetics Laboratory, Institute of Archaeology of the Academy of Sciences of the Czech Republic, Prague, Letenská 1, 118 01 Prague, Czech Republic; yorodiallo90@yahoo.fr (M.Y.D.); martina.ciz@email.cz (M.Č.); ell.6@email.cz (E.P.); priehodovaedita@seznam.cz (E.P.)
- ² Department of Anthropology and Human Genetics, Faculty of Science, Charles University in Prague, 128 01 Prague, Czech Republic; iwakulichova@gmail.com
- ³ Department of Forensic Genetics, Institute of Criminalistics, Police of the Czech Republic, 128 01 Prague, Czech Republic; novackova.janka@gmail.com
- ⁴ i3S—Instituto de Investigação e Inovação em Saúde, Universidade do Porto, 4200-135 Porto, Portugal; vfernandes@ipatimup.pt (V.F.); luisap@ipatimup.pt (L.P.)
- ⁵ IPATIMUP—Instituto de Patologia e Imunologia Molecular da Universidade do Porto, 4200-135 Porto, Portugal
- * Correspondence: cerny@arup.cas.cz

Abstract: African history has been significantly influenced by the Sahara, which has represented a barrier for migrations of all living beings, including humans. Major exceptions were the gene flow events that took place between North African and sub-Saharan populations during the so-called African Humid Periods, especially in the Early Holocene (11.5 to 5.5 thousand years ago), and more recently in connection with trans-Saharan commercial routes. In this study, we describe mitochondrial DNA (mtDNA) diversity of human populations from both sides of the Sahara Desert, i.e., both from North Africa and the Sahel/Savannah belt. The final dataset of 7213 mtDNA sequences from 134 African populations encompasses 470 newly collected and 6743 previously published samples, which were analyzed using descriptive methods and Bayesian statistics. We completely sequenced 26 mtDNAs from sub-Saharan samples belonging to the Eurasian haplogroup N1. Analyses of these N1 mitogenomes revealed their possible routes to the Sahel, mostly via Bab el-Mandab. Our results indicate that maternal gene flow must have been important in this circum-Saharan space, not only within North Africa and the Sahel/Savannah belt but also between these two regions.

Keywords: Sahel/Savannah belt; North Africa; mtDNA diversity; population history



Citation: Diallo, M.Y.; Čížková, M.; Kulichová, I.; Podgorná, E.; Priehodová, E.; Nováčková, J.; Fernandes, V.; Pereira, L.; Černý, V. Circum-Saharan Prehistory through the Lens of mtDNA Diversity. *Genes* **2022**, *13*, 533. <https://doi.org/10.3390/genes13030533>

Academic Editor: Marc Via

Received: 23 January 2022

Accepted: 18 February 2022

Published: 17 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The out-of-Africa event, during which a relatively small group of anatomically modern humans spread from East Africa into Eurasia [1], was a defining moment in the evolution of modern humans. Although paleoanthropology has detected several older waves of Middle Pleistocene migrations from Africa to Eurasia [2], genetic studies show that contemporary non-sub-Saharan populations are descendants of an ancestral population that spread from Africa only about 60 ka (thousands of years ago) [3]. Leaving aside the long-term isolation of Khoisan populations in southern Africa and of the Pygmies in the tropical rain forests of central Africa, the out-of-Africa event is nowadays considered to be the most significant restriction of gene flow between two groups of anatomically modern humans: the sub-Saharan and non-sub-Saharan populations had been separated by the Sahara Desert throughout most of prehistory. Differentiation between these groups is apparent in both mitochondrial (mtDNA) [4] and nuclear [5] DNA diversity. Due to this separation, we can detect different mtDNA haplogroups, which can be assigned either a sub-Saharan or Eurasian ancestry [6].

While the basis of the sub-Saharan mtDNA gene pool is classified as macro-haplogroup L, the rest of the world nowadays traces its maternal ancestry from haplogroup N or M [7].

That, however, does not mean that after the out-of-Africa event migration had stopped. The genetic structure of inhabitants of the Sahel/Savannah belt was analyzed by researchers with respect to their linguistic affiliation, subsistence structure, and geographic localization of local populations [8–10], and both population genetics and phylogeographic studies highlighted the significance of gene flow. The Sahel/Savannah belt has therefore been called a “bidirectional corridor of migrations” [11] and evidence of gene flow was also detected across the Sahara in populations inhabiting regions between the Sahel/Savannah belt and North Africa [12,13].

Interestingly, while some migrations may have had an ethnic association, others did not. For instance, the origin of the Chadic-speaking peoples living in the Lake Chad Basin was traced to East Africa based on linguistic evidence [14]. According to this theory, the ancestors of current Chadic-speaking peoples migrated, still as nomadic herders, from the Nile Valley through Wadi Howar to the Ennedi Mountains, and further through Wadi Hawash up to the Lake Chad Basin. Genetically, the Chadic-speaking peoples nowadays harbor mtDNA sequences belonging to the L3f haplogroup with East African ancestry, especially a private branch L3f3, which formed during their westward expansion of about 8 ka [15]. On the other hand, another mtDNA haplogroup, called L3e5, which was also detected in populations living today in the Lake Chad Basin but not only in Chadic-speaking populations, is also present in the Maghreb and its origin can be traced to an ancestral population that crossed the green Sahara during the Early Holocene approximately 10 ka [13].

Sahelian populations also carry Eurasian mtDNA haplogroups. They are found more frequently in the nomadic pastoralists than in sedentary farmers [16] and a surprising finding showed that some sub-Saharan Africans and even Northern Eurasians share some very recent maternal ancestry. For instance, it was shown that a Saami from Scandinavia and a Yakut from Siberia share with a Berber and a Fulani mtDNA sequences belonging to haplogroup U5b1b [17]. Given the enormous geographical distances between these populations, the most plausible explanation is that the most recent common ancestor (~8.6 ka) lived probably in southwestern Europe, from where the descendants spread both to northern Eurasia and sub-Saharan Africa. A more detailed study of originally Eurasian lineages beyond the Sahara has shown that not only U5b1b but also the H1 haplogroup (which both occur mainly in the Fulani pastoralists) came to form the new and younger sub-Saharan lineages called U5b1b1b and H1cb1 [18]. Their most recent common ancestor (~4 ka) dates to the time when, according to archaeology [19–21], the first herders settled in the western Sahel/Savannah belt.

It may therefore seem that the pastoralist food-production strategy did not spread to sub-Saharan Africa by demic diffusion from the Near Eastern domestication center via northeastern Africa, but through the ancestors of Berbers from the Maghreb. In this context, it should be noted that the genetic architecture of the circum-Mediterranean space had undergone substantial changes since the Neolithic. For instance, ancient Near Eastern farmers are genetically better represented by the current populations of central and western Mediterranean, such as the Sardinians and the Basques [22,23], than by the current populations of the Near East.

The importance of post-Neolithic gene flow from northwestern Africa to the western part of the Sahel/Savannah was also suggested by research on lactase persistence. In fact, the Fulani pastoralists from Burkina Faso share with Europeans the extended haplotype carrying Eurasian variant $-13,910 \times T$. It was suggested that their ancestors received this haplotype via admixture with the Eurasian population two times [24]. The first event is genetically dated to ~1828 years ago and the second one to ~302 years ago, whereby it seems that the admixture involved a group related to southwestern Europeans. Moreover, the geographical distribution of lactase persistence variants in the Sahel/Savannah belt shows clear differences between the pastoralists in the east (mostly Arabs harboring variant $-13,915 \times G$) and the west (mostly Fulani harboring variant $-13,910 \times T$) [25]. In fact, a

boundary between the western and eastern Sahelian genetic spaces lies somewhere near the Lake Chad Basin, as attested not only by lactase persistence but also by a genome-wide SNP study [26].

Last but not least, it was shown that Sahelian pastoralists tend to represent several mutually similar mtDNA haplotypes, which indicates either more recent origins of their diversity, isolation of their demes, lower gene flow, or lower effective size of the population [10]. Interestingly, thanks to coalescence analyses, it was possible to show there is an asymmetric gene flow between the pastoralists and the farmers in both parts of the Sahel/Savannah belt: while the western (Fulani) pastoralists are losing their mtDNA diversity, the eastern (Arabs) pastoralists are gaining it by admixture with local sub-Saharan agricultural populations [27]. This is further supported by the presence of various sub-Saharan mtDNA haplotypes in the gene pool of Arabic-speaking populations [9], mostly non-carriers of the lactase persistence $-13,915 \times G$ variant [28]. Interestingly, this genetic observation might correspond to a process of Arabization and/or language shift after the expansion of Arabs and their culture from North Africa into the Lake Chad Basin, from the 14th century AD onwards [29].

The above-mentioned studies show that inclusion of newly collected local populations, especially from the Sahel/Savannah belt, has significantly contributed to our knowledge of the peopling of Africa north of the equator by discoveries of not only new variants—which happens quite commonly when a new dataset of a sub-Saharan population is presented [30]—but even of entire new mtDNA haplogroups. In fact, since all new sub-Saharan population studies published so far revealed new genetic variants, one ought to admit we are so far aware of merely a fraction of the genetic diversity of sub-Saharan populations [31,32].

Because sub-Saharan Africa is still underrepresented in population genetic and genomic studies [30], we compiled a large mtDNA database composed of both newly collected and previously published mtDNA sequences and produced an updated survey of migration patterns in the circum-Saharan space. Additionally, we performed a complete mtDNA sequencing of the N1 haplogroup from sub-Saharan Africa, with most samples from the Sahel/Savannah belt but some also from East Africa. The N1 haplogroup's southwestern Asian ancestry is well known and goes as far as to ~ 60 ka [4] but its African phylogeny is still not well understood. We selected the N1 because this haplogroup was reintroduced back to the Sahel/Savannah belt by migration from southwestern Asia, possibly via North Africa, as became apparent when a related basal branch was recently discovered in a North-African skeleton (Takarkori rock shelter, Libya) dated to ~ 7 ka [33]. We can thus assume that phylogeny of this specific haplogroup could document an ancient gene flow back to Africa in the eastern circum-Saharan region.

2. Materials and Methods

We built a large dataset of 7213 mtDNA hypervariable segment 1 (HVS-1) sequences from 134 African populations inhabiting the circum-Saharan space. It contains both published and newly acquired samples. All samples were collected with the help of African colleagues and appropriate informed consent was procured prior to sampling from all participants. Newly collected and here for the first time presented mtDNA ($n = 470$) sequences enriched the Sahelian dataset with four populations from Mali ($n = 135$), three from Mauritania ($n = 191$), and three from Sudan ($n = 144$). During the sampling, we focused on populations that have not been previously included in any genetic study, such as the Bella from Mali, Imraghen from Mauritania, Daju and Zaghawa from Sudan, but also on populations whose sampling coverage was insufficient due to their large geographic dispersals, such as the Fulani from Mali and Mauritania, Moors and Soninke from Mauritania, and Tuareg, Songhai, and Arabs from Mali.

The population dataset (Supplementary Table S1) was further divided according to three variables: from the perspective of subsistence strategy (lifestyle), each population was categorized as belonging either among pastoralists or farmers, from the perspective

of language affiliation, populations were classified as belonging to one of three language families (Niger-Congo, Nilo-Saharan, and Afro-Asiatic), and from the geographic perspective we divided populations in western Sahel/Savannah, western North African, eastern Sahel/Savannah, and eastern North African groups (Figure 1).

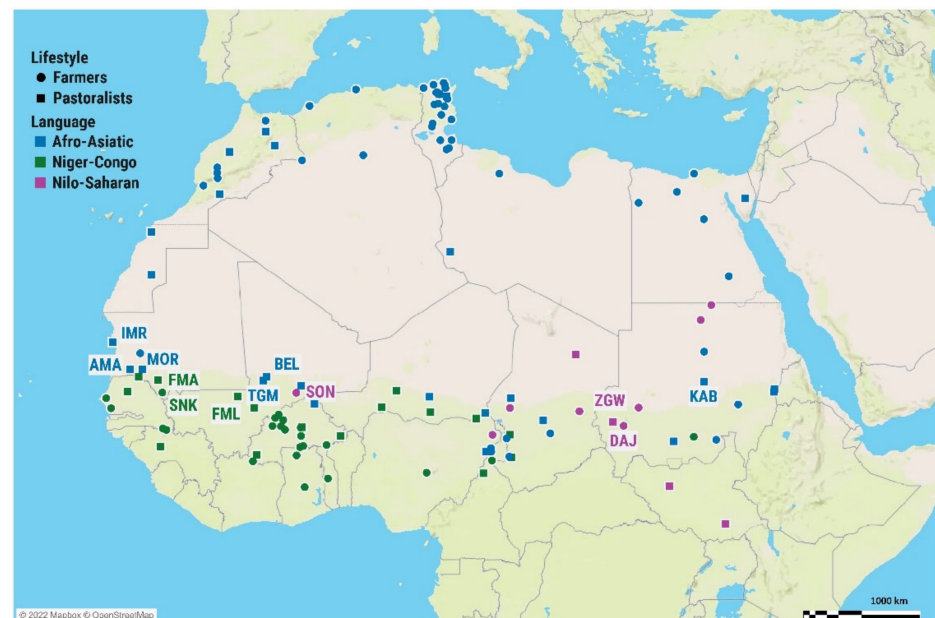


Figure 1. Geographic locations of 134 African populations included in this study. Newly obtained samples are marked with labels, and colors differentiate the linguistic affiliations and subsistence strategies.

Subsequently, we selected from our collections for complete mtDNA sequencing 26 DNA samples which, according to HVS-1 using the rCRS-oriented version of Build 17 on the PhyloTree website [34], belong (together with its daughter clade I) to haplogroup N1. For selected details of the N1 sample, see Supplementary Table S2. Subsequently, we compared these mitogenomes with worldwide samples ($n = 701$) belonging to the same N1 haplogroup. Our total N1 dataset thus includes 727 mitogenomes (Supplementary Table S3).

New samples published here for the first time were collected by the Oragene DNA Collection saliva kit (DNA Genotek Inc., Ottawa, ON, Canada). DNA was extracted according to supplier's protocol. PCR amplification of all newly sequenced samples ($n = 470$) was carried out using primer pairs previously published in Gonder et al. [35]. Due to differences in the lengths of mtDNA sequences published in available comparative studies, sequences analyzed in this study were restricted to a highly polymorphic segment of 339 bp (nps 16,032–16,370). The region between positions 16,184 and 16,194 was excluded from analyses because of uncertainty regarding its correct alignment which arose in consequence of heteroplasmy observed in several samples [36].

Samples selected for whole mtDNA sequencing were analyzed using the Massive Parallel Sequencing (MPS) method. Quantity of the extracted DNA was determined by Plexor HY (Promega) and the samples were diluted to the total DNA template of 15 ng per reaction. Whole-genome mtDNA MPS data were generated using commercially available QIAseq™ Targeted DNA Panel (Qiagen), which contains 222 overlapping primers, and then sequenced on MiSeq Fgx Sequencing system (Verogen) in the “Research Use Only” mode through the “Generate FASTQ” with “FASTQ Only” application and “Paired End Read”. MPS libraries and final sequencing reactions were prepared according to the manufacturer's protocol to a final loading concentration of 12 pM. Sequencing reaction was performed using MiSeq v2 Reagent Illumina kit (300 cycles per kit).

We computed both the standard and molecular diversity indices, such as haplotype diversity, nucleotide diversity, and the mean number of pairwise differences for each population using Arlequin software ver. 3.5.2.2 [37]. Differences in average haplotype and nucleotide diversities and in the mean number of pairwise differences among geographically specified subsistence modes were tested using Kruskal–Wallis one-way tests.

To investigate the demographic history of our samples, we calculated tests of selective neutrality such as Tajima's D [38], Fu's F_s [39], and R_2 [40], which are sensitive to deviations from a demographic equilibrium. Significant results could indicate not only natural selection but also demographic expansion or contraction of effective population size. Tajima's D and Fu's F_s tests were calculated in the same version of Arlequin using 10,000 iterations [37], while R_2 tests were computed by using DnaSP ver. 6.12.03 [41]. Statistical tests and confidence intervals for R_2 were based on parametric bootstrapping with coalescence simulations. It should be noted that in order to reject neutrality at 5% level of significance in Fu's F_s statistic, the p -value should be below 2% (because it is tested by a unilateral test). Additionally, we evaluated demographic history by Harpending's raggedness index (H_{ri}) and the sum of squared deviations (SSD) while considering a model of demographic expansion as implemented in Arlequin [42]. Statistical significance of values was assessed by a permutation test with 1000 replicates and, similarly, as in molecular diversity indices, variance between p -values was evaluated using Kruskal–Wallis one-way tests.

Interpopulation comparisons were assessed via pairwise Reynolds' genetic distances based on haplotype frequencies [43,44] using Arlequin, and significance was tested by 10,000 iterations. Evolutionary distances between haplotypes (Φ_{ST} indices) were used to weight Reynolds' genetic distances, to which purpose we used the 2-parameter Kimura model with a γ shape parameter of 0.4, transition/transversion ratio of 10/1, and indels not taken into consideration, as recommended in a previous study [45]. Coancestry coefficients of Reynolds' genetic distances were used for visualization via a multidimensional scaling analysis (MDS) with community ecology package *vegan* ver. 2.5-7 [46] implemented in RStudio [47].

Analyses of molecular variance (AMOVA) assessed the levels of genetic variation within and between different groups of populations structured, as noted above, by several factors, namely language affiliation, geographical location, and subsistence strategy. The settings were: language grouping (Afro-Asiatic vs. Nilo-Saharan vs. Niger-Congo); geographic grouping as Region1 (Sahel/Savannah belt vs. North Africa) and Region2 (eastern North Africa vs. western North Africa vs. eastern Sahel/Savannah belt vs. western Sahel/Savannah belt), subsistence grouping as Lifestyle1 (pastoralists vs. farmers) and Lifestyle2 (pastoralists of the Sahel/Savannah belt vs. pastoralists of North Africa vs. farmers of Sahel/Savannah belt vs. farmers of North Africa). By inspecting the fixation indices, we were able to describe the structure between groups (for Φ_{CT} index) and/or relationships between populations within groups (for Φ_{SC} index). The significance of fixation indices was tested using a hierarchical framework and 10,000 iterations of the random permutation procedure implemented in Arlequin [37].

Coalescent estimation of ancestral or contemporary migration patterns between populations was carried out by Migrate-n (version 3.6.11) using a Bayesian Markov chain Monte Carlo inference model to generate a posterior probability density distribution [48]. This method is capable of measuring complex models with asymmetric gene flow directly. The estimated parameters were Θ for population size (from which we could calculate effective population size via $\Theta = 2N\mu$ for haploid mtDNA data) and M (immigration rates), both scaled to mutation rate, with settings of one long chain; 1,000,000 genealogies with every 5000 recorded; and a burn-in per replicate of 10,000. To enhance our understanding of the complex patterns and trends in gene flows, we used a bilateral circular migration plot as an effective method of visualizing flow data, implemented in the R package *circsize* 0.4.13 [49].

In terms of treatment of whole mtDNA sequences, we analyzed the raw FASTQ files using NextGENe software (SoftGenetics, LLC, PA 16803, USA) and the resulting sequences were read with BioEdit version 7.0.9.1 [50]. Mutations were scored relative to the revised

reference sequence, rCRS [51] with numbers 1–16,569 referring to mutation's position in that sequence. The newly generated mitogenomes were compared to complete sequences available in GenBank. We found 701 sequences belonging to I and N1 haplogroups in a great majority of cases from non-African populations.

To estimate time to the most recent common ancestor (TMRCA) for specific clades in the phylogeny, we used the ρ statistic [52] and maximum likelihood (ML). For ρ , i.e., the mean sequence divergence from the inferred ancestral haplotype of the clade in question, we set the mutation rate estimate for the whole-mtDNA sequence corrected for purifying selection to one substitution in every 3624 years and for the synonymous mutation rate at one substitution in every 7884 years [53], while standard errors were estimated as previously described [54]. ML estimates of branch lengths were obtained using PAML 4.8a [55] while assuming an HKY85 mutation model with γ -distributed rates (approximated by a discrete distribution with 32 categories), partitions (the two hypervariable regions as a block vs. the remaining mtDNA genome), and a generation time of 25 years. We converted mutational distance in ML to time using the same whole-mtDNA genome clock of 3624 years.

3. Results

Gene diversity, nucleotide diversity, and the mean number of pairwise differences of each population are shown in Supplementary Table S4. Figure 2a shows the distribution of values of these indices within geographical regions, subsistence categories, and language families. These results clearly show that populations from the Sahel/Savannah belt tend to be more diverse than those living in North Africa and that the eastern parts of both of these regions are more diverse than their western parts. One can also observe that sedentary farmers are slightly more diverse than nomadic pastoralists are, especially in the eastern part of the investigated regions. Our results regarding the distribution of these indices within language families may be affected by unequal representation of individual populations in groups because the Afro-Asiatic family is much better represented than other families: for instance, the Nilo-Saharan family is represented in our dataset by just 14 populations. A Kruskal–Wallis one-way test (Table 1) indicated statistically significant differences from both the perspective of nucleotide diversity and the mean number of pairwise differences among all the groups (except for those based on lifestyle, which seems to significantly differ from the perspective of haplotype diversity). When lifestyle was combined with regional information, all molecular indices showed significant differences.

Table 1. Kruskal–Wallis one-way test (p -values) of differences in average diversity indices and demographic parameters among specified groups of populations.

	Kruskal–Wallis Test (p -Values)					
	Haplotype Diversity	Nucleotide Diversity	Mean Number of Pairwise Differences	Fu's F_s (p)	Hri (p)	SSD (p)
NA vs. SS	0.380	0.000	0.000	0.563	0.068	0.208
FM vs. PS	0.004	0.856	0.630	0.133	0.271	0.249
AA vs. NC vs. NS	0.293	0.011	0.001	0.682	0.033	0.147
eNA vs. eSS vs. wNA vs. wSS	0.453	0.000	0.000	0.940	0.266	0.231
NAFM vs. NAPS vs. SSFM vs. SSPS	0.000	0.000	0.000	0.507	0.482	0.874

NA: North Africa; SS: Sahel/Savannah belt; FM: farmers, PS: pastoralists; AA: Afro-Asiatic; NC: Niger-Congo; NS: Nilo-Saharan; eNA: eastern North Africa; eSS: eastern Sahel/Savannah; wNA: western North Africa; wSS: western Sahel/Savannah; NAFM: farmers North Africa; NAPS: pastoralists North Africa; SSFM: farmers Sahel/Savannah; SSPS: pastoralists Sahel/Savannah; p -values marked with bold indicate statistically significant differences between the groups.

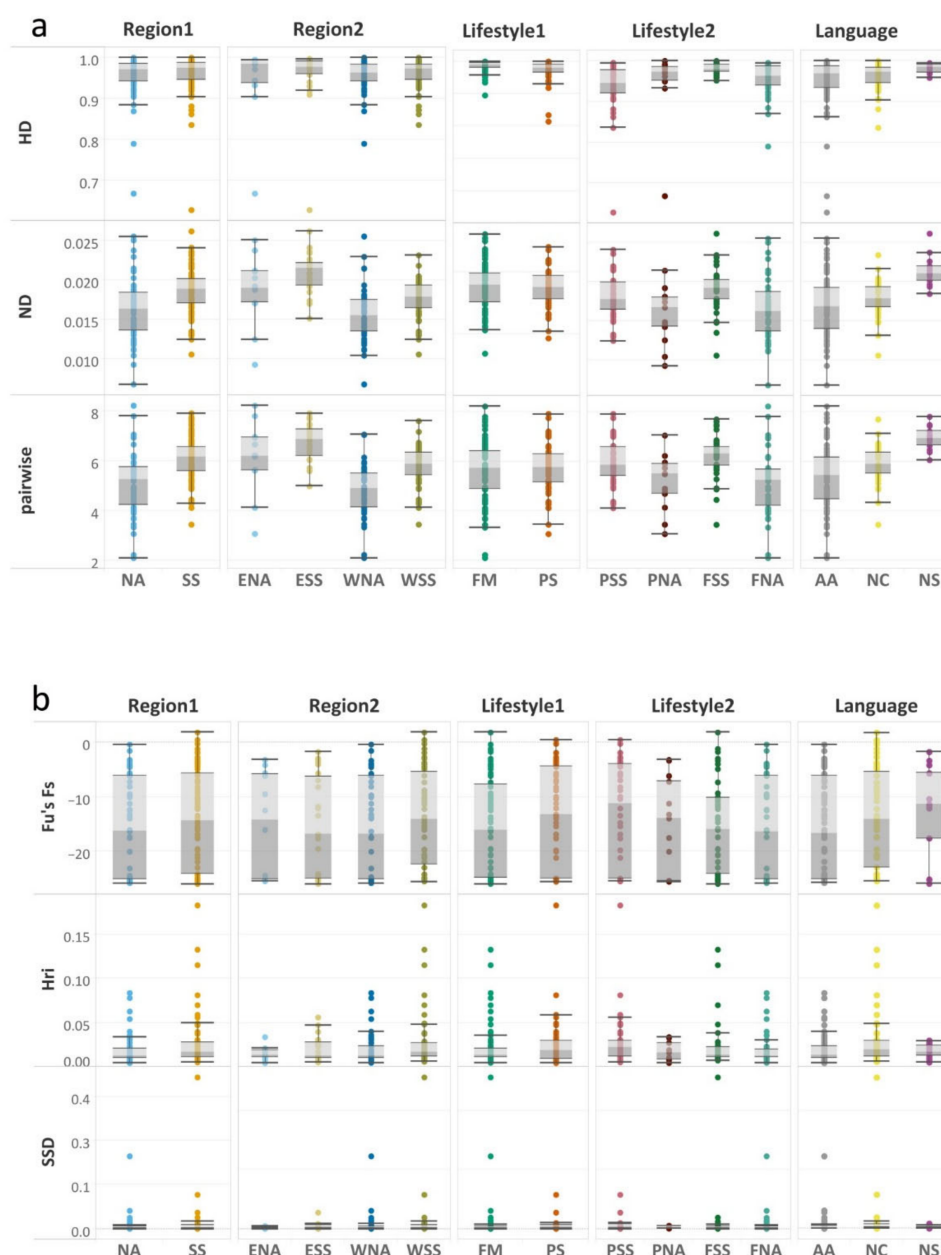


Figure 2. Boxplots representing diversity values of basic standard and molecular indices in five groups defined by subsistence strategy, language, and geographic region. (a) HD—haplotype diversity, ND—nucleotide diversity; pairwise—the mean number of pairwise differences; (b): Fs—Fu's; Hri—Harpending's raggedness index; SSD—the sum of squared deviations. Grouping label abbreviations: Region1: NA North Africa, SS Sahel/Savannah belt, Region2: ENA Eastern North Africa, ESS Eastern Sahel/Savannah belt, WNA Western North Africa, WSS Western Sahel/Savannah belt. Lifestyle1: PS Pastoralists, FM Farmers, Lifestyle2: PSS pastoralists of Sahel/Savannah belt, PNA pastoralists of North Africa, FSS farmers of Sahel/Savannah belt, FNA farmers of North Africa. Language: AA Afro-Asiatic, NC Niger-Congo, NS Nilo-Saharan. The color palette is set according to the defined groups. The line dividing boxplots into two parts represents the median of data for the variable concerned. The ends of the box indicate the upper (Q3) and lower (Q1) quartiles. The difference between quartiles 1 in light grey and 3 in grey indicates the interquartile range (IQR). The top and bottom lines show the maximum and minimum values. Dots (or other markers) beyond the extreme line indicate potential outliers.

Patterns of demographic expansions were investigated using selective neutrality tests. Figure 2b shows the distribution of values of these indices within geographical regions, subsistence categories, and languages families. A more sensitive analysis—indicated by the highly negative values of F_u 's F_s (running around -25) supported by non-significant p -values of SSD and Hri—revealed demographic expansion in several North African populations. As far as sedentary farmers are concerned, demographic contraction can be deduced for the Bedik and Mandenka, but also for some groups of Moroccan Berbers such as the Kesra and Zriba. No signals of demographic expansion (based on F_u 's F_s) were detected also in several nomadic Fulani groups from the western Sahel/Savannah belt. The exact values of all indices are specified in Supplementary Table S4. We also tested the variance of p -values with a Kruskal–Wallis one-way test and significant differences were detected only in the Hri index between linguistic groups (Table 1).

A matrix of coancestry coefficients for all analyzed populations is presented in Supplementary Table S5. MDS plots based on these values (calculated from genetic distances by $-\ln(1 - F_{ST})$) with a stress value of 0.112 imply differentiation of populations in a 2D scale. Four subplots show the populations colored according to geography (Figure 3a), lifestyle (Figure 3b,c), and language (Figure 3e). In general, the space covered by populations from the Sahel/Savannah belt is apparently larger than the space covered by the North African populations, which shows a higher level of differentiation (higher pairwise distances) in the southern part of this circum-Saharan area. Interestingly, the newly introduced populations collected in the Sahel/Savannah belt are distributed not only within their own genetic space: some, such as the Moors from Mauritania and Arabs from Mali, are located within the area covered by North African populations. This can best be explained by the recent immigration of these groups to the Sahel/Savannah belt. Figure 3d shows that not only geography but also language might be considered a good tool for the determination of population structure in the circum-Saharan space.

AMOVA was applied to evaluate genetic differentiation among populations grouped according to geography, language, and subsistence strategy. The results are presented in Supplementary Table S6 and Figure 4. Interestingly, definitions of the groups based on lifestyle do not seem to reflect the population structure either in North Africa or in the Sahel/Savannah belt: in all these groups, the values for between-group variation (Φ_{CT}) were lower than for variation between populations within these groups (Φ_{SC}). On the other hand, we were able to confirm the structure when considering languages or regionally divided populations (Region1 and Region2), or when the farmers/pastoralists in North Africa were contrasted with farmers in the Sahel/Savannah belt. Our results show rather balanced indices indicating no significant structure among pastoralists in North Africa and the Sahel/Savannah belt ($\Phi_{CT} = 0.0674$, $\Phi_{SC} = 0.0765$). All values were statistically highly significant ($p < 0.05$) except that for variance between the groups of farmers vs. pastoralists in the North African division.

When evaluating signals of gene flow obtained from the Bayesian coalescent approach in Migrate-n software (see Supplementary Tables S7 and S8 and Figure 5), we detected the highest population size parameters in farmer populations from eastern North Africa and western Sahel/Savannah belt, which implies larger N_e in these groups. Moreover, we calculated the immigration rates and the numbers of immigrants into specific groups. Surprising results are seen in North African populations (in particular eastern pastoralists and western farmers), who seem to have received the least number of immigrants from other regions. On the other hand, we ought to consider the differences in sample sizes of these groups: for eastern North African pastoralists, we had just 195 samples, while for western North African farmers we had 2343. The results also suggest a higher gene flow into farmer populations in general, with the exception of the eastern Sahel/Savannah belt where it was even lower than in other pastoralist groups. Figure 5 represents the summary of population contacts (gene flow) in a circular plot. In general, this specific analysis showed relatively high migration activities not only along the east–west but also the north–south axis.

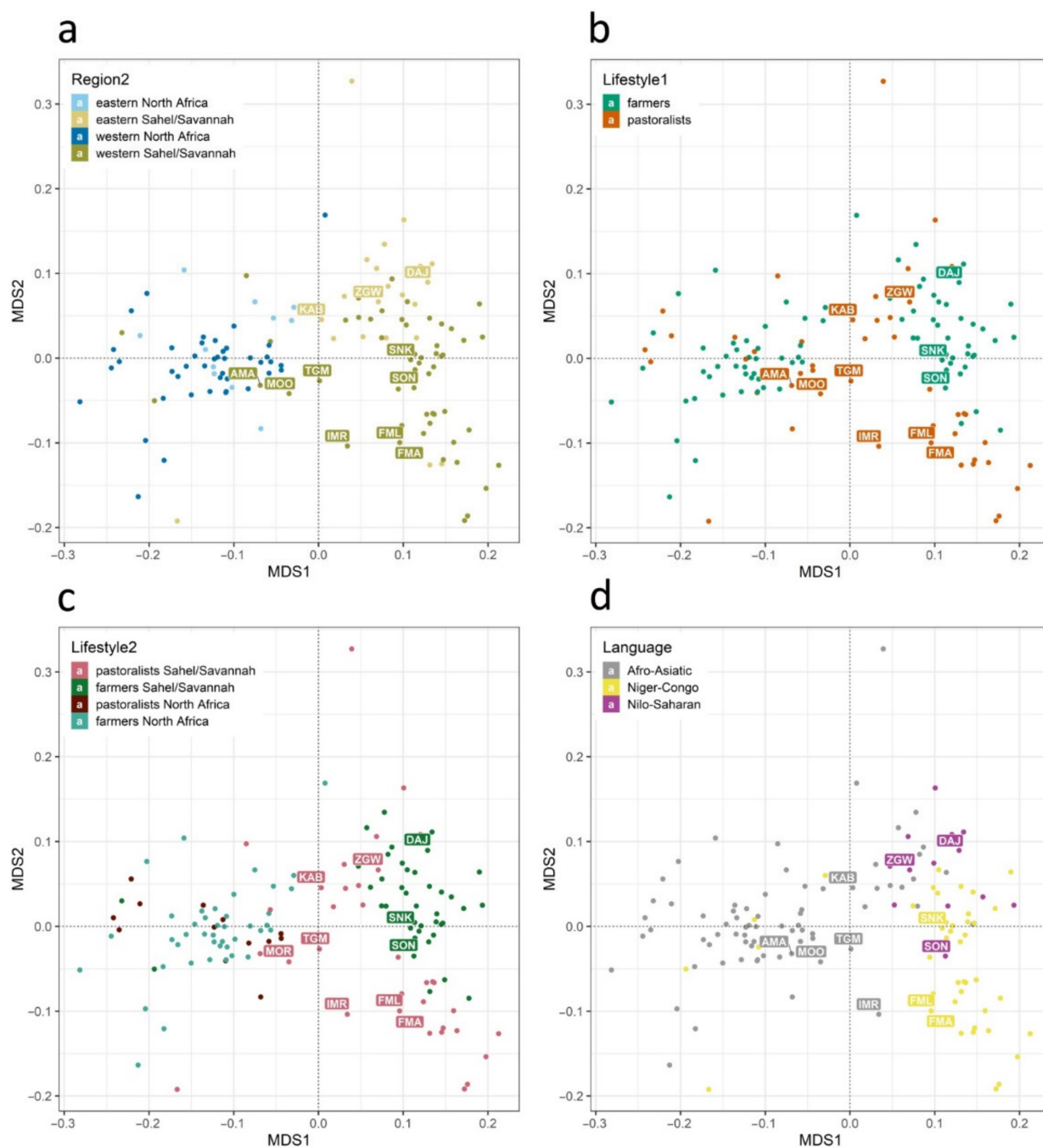


Figure 3. MDS plot (stress value 0.112) of pairwise mtDNA genetic distances between 134 populations colored according to geography as Region2 (a), subsistence as Lifestyle1 (b) and Lifestyle2 (c), and Language (d). Only newly collected populations are highlighted with labels.

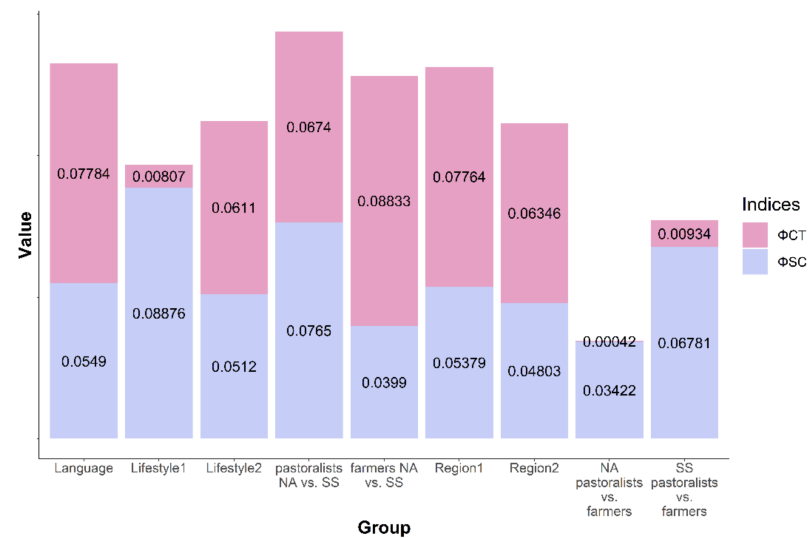


Figure 4. Values of two variation indices Φ_{CT} in pink (for variation between groups) and Φ_{SC} in violet (for variation among populations within the groups) from analysis of molecular variance (AMOVA) made for several groups defined by language, lifestyle, and geographic region (groups specified on x-axis).

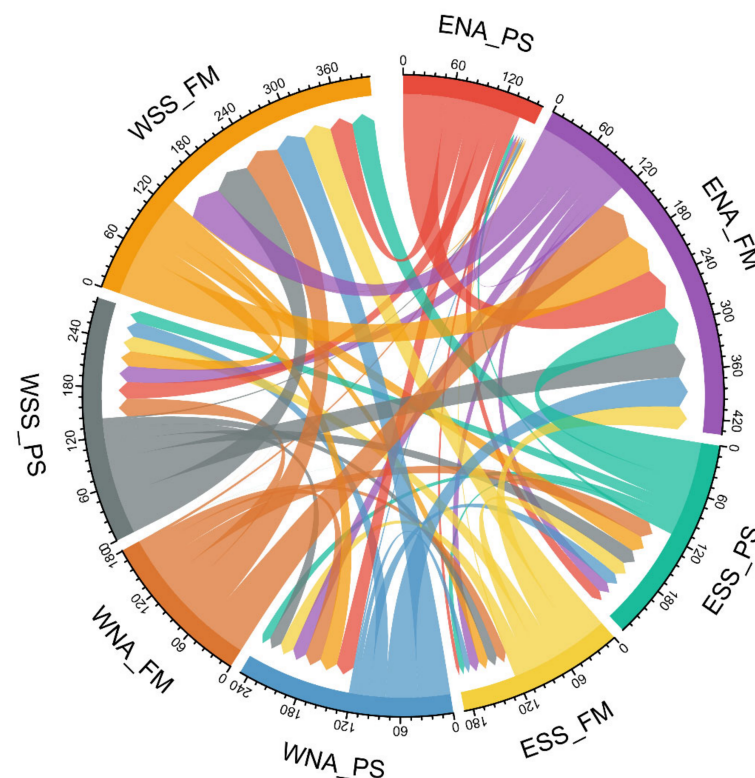


Figure 5. Circular plot of migration/gene flows resulting from the Bayesian coalescent approach in Migrate-n. Numbers around the perimeter illustrate values of $2Nm = \theta \times M$ (the number of immigrants). The arrow signs and their width represent the direction and extent of gene flows between specified groups (ENA_PS: Eastern North African pastoralists; ENA_FM: Eastern North African farmers; WNA_PS: Western North African pastoralists; WNA_FM: Western North African farmers; ESS_PS: Eastern Sahel/Savannah pastoralists; ESS_FM: Eastern Sahel/Savannah farmers; WSS_PS: Western Sahel/Savannah pastoralists; WSS_FM: Western Sahel/Savannah farmers).

The phylogeny of haplogroup N1, together with its derived clade I, is presented in Figure 6 and Supplementary Figure S1. These figures show that a large proportion of N1 lineages from the Sahel/Savannah belt and from eastern Africa (Somalia, Sudan, Ethiopia, Kenya, and Tanzania), which are affiliated in N1a1a and N1b2 sub-branches, have the closest relatives in the Arabian Peninsula and are the oldest lineages in those sub-branches, while most I-Sahel/Savannah belt lineages are basal in the branches together with other European and Near Eastern sequences. When we look at the age of the N1a1a and N1b2 clusters (Table 2 and Supplementary Table S9), we can see age estimates at most around 20 ka, whereby these should be interpreted as the upper limit age of introgression of these lineages into the Sahel.

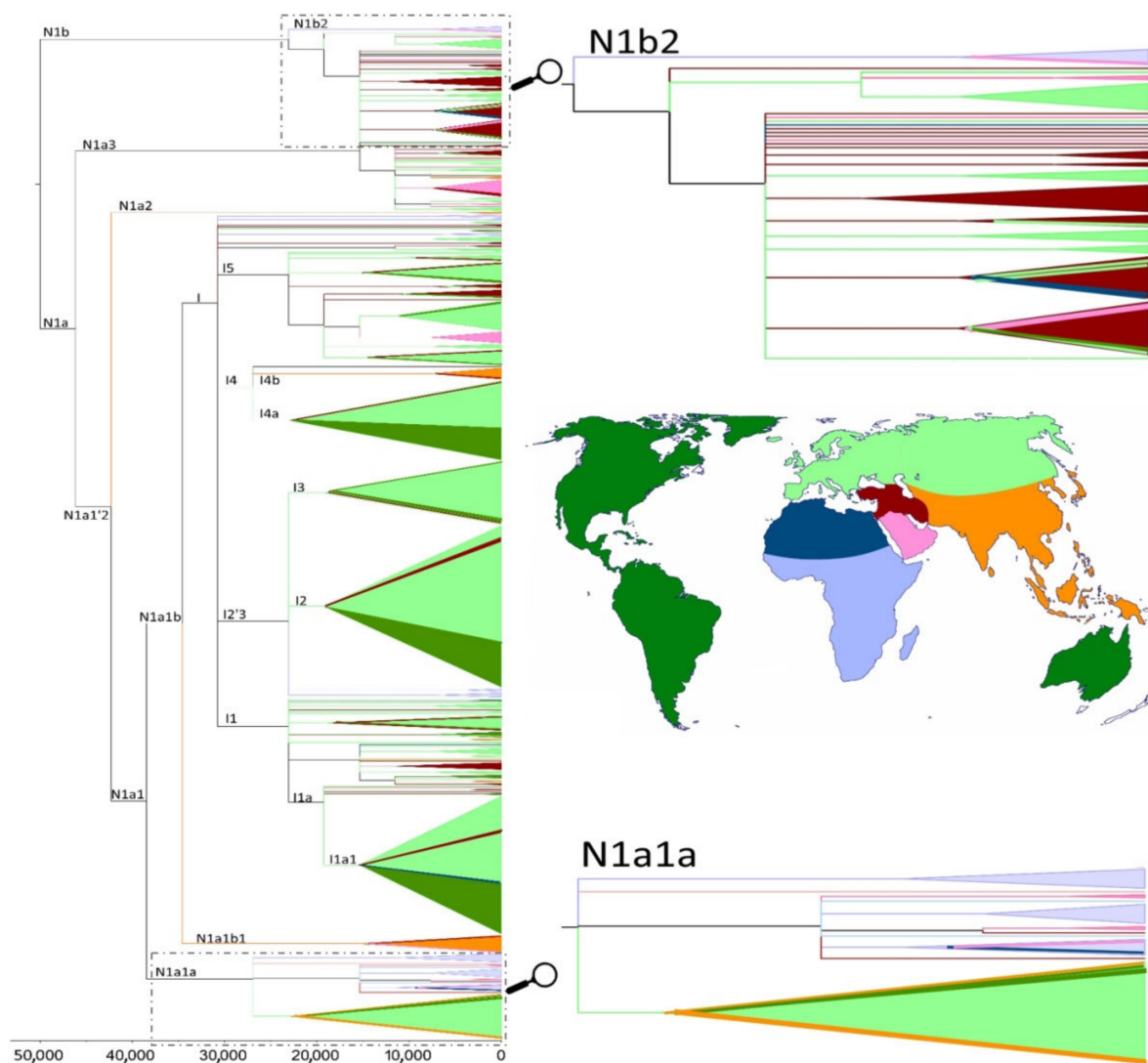


Figure 6. Phylogenetic reconstruction of mtDNA haplogroup N1, with time scale for the most recent common ancestors of the branches. Colors used in the tree correspond to the geographical regions illustrated in the map. Zooms of the N1a1a and N1b2 parts of the tree allow us to observe a closer relationship between the Sahel/Savannah belt and Arabian sequences.

Table 2. Age estimates and standard deviations (in years) for N1a haplogroups in Africa.

Haplogroup	Maximum Parsimony				Maximum Likelihood	
	Total Clock		Synonymous Clock		All SNPs	
	Age	95% CI	Age	95% CI	Age	95% CI
N1a1a	24,005	16,482–31,778	20,727	10,873–30,580	21,754	21,304–22,182
N1a1a3	13,900	10,071–17,808	14,266	7442–21,090	17,948	17,676–18,211
N1a3a+195C!	3188	1266–5135	1819	–838–4477	3610	3577–3643
N1b2+9325C	6816	553–13,327	14,191	–3 562–31,944	6908	6762–7048

4. Discussion

An mtDNA dataset containing 7213 mtDNA sequences in 134 African populations—which is much more than used in a previous studies [9,56]—and covering the entire circum-Saharan space had significantly contributed to our understanding of African population history north of the equator. First of all, we were able to show that North African populations have lower values of nucleotide diversity, especially in the western part of the region. This can be attributed to their lower effective population sizes, as attested also by the Bayesian coalescent approach employed in this study. When contrasting pastoralists with farmers, we found similar distributions of diversity values, which supports our previous finding of no significant structure associated with the subsistence strategy in the Sahel/Savannah belt [10].

The lower level of differentiation among the populations of North Africa than among the populations of the Sahel/Sudan belt supports the idea of higher migration activity homogenizing the gene pool and eroding the population structure in the southern Mediterranean space. It is well possible that there was a long-range influx of population(s) from the Near East to the Maghreb already in preagricultural times [50]. In fact, this finds support in recent aDNA analyses of Iberomaurusian skeletons [57]. Interestingly, both Natufian and Iberomaurusian specimens show a high level of Basal Eurasian ancestry, which was a population isolated > 50 ka in a Late Pleistocene refugium of the Arabo-Persian Gulf without contacts with the Neanderthals [58]. Further immigration to North Africa took place in the Neolithic and in later times both from the Near East [59–61] and from Europe via the Strait of Gibraltar [62–65]. According to our results, it seems that this expansion through populations of farmers reached all the way to the western Sahel/Savannah belt. It should also be noted that the general genetic homogeneity of North African populations is reflected in the linguistic homogeneity of Afro-Asian languages.

Our results which suggest that the eastern part of North African populations received immigrants from the Sahel/Savannah belt do not correspond with research on autosomal SNP variants, which had shown that the populations of Egypt and Libya are composed predominantly of a Near Eastern genetic component with very low input from sub-Saharan Africa [50]. However, that may be due to the fact that the last-mentioned study worked with limited sub-Saharan (and not really Sahelian) samples as the putative sources of migration to their North African datasets. In fact, a subsequent study revealed in some North African populations (e.g., in Algeria) a higher gene flow from the sub-Saharan space, especially in maternal lineages [66]. In Egypt, the importance of a migration corridor via the Nile Valley was described a number of times both in archaeology [67] and in genetics [68,69]. Moreover, many eastern Sahel/Savannah populations also have an admixture of sub-Saharan and Eurasian ancestries; especially the Arab groups have an important Eurasian component. This is thus why the sub-Saharan input in North Africa is larger in the west than in the east. Another point is that the sub-Saharan influence in North Africa was mainly via maternal lines: higher in mtDNA, almost nonexistent in the Y-chromosome, and intermediate in the autosomal DNA [70].

Migrations in the Sahel/Savannah belt were probably less important than in North Africa. When we look at the continent-wide African mtDNA diversity, the Sahel/Savannah belt can be viewed as a corridor between the Sahara and tropical rainforests which con-

nects eastern and western Africa but also has—especially in the Lake Chad Basin—some distinctive genetic features [11]. Food-producing strategies came to play an important role in demographic expansions in this region later than in North Africa, especially in the Holocene. The first expansion may have been related to pastoralism, which is a strategy perfectly well-adapted to Sahelian dryland ecosystems [71]. It spread through the Sahel/Savannah belt from northeastern Africa during the Holocene [72–74]. An expansion of herders started ~8 ka in northeastern Africa but was relatively slow because it reached the western Sahel/Savannah belt much later, at about 3 ka [75].

The cultivation of cereals and tubers, which is autochthonous in the Sahel/Savannah belt (especially in and around the Middle Niger Delta [76–79] and in the Middle Nile Valley [80]) was somewhat delayed because the first fully domesticated plants were consumed in Africa only about 4.5 ka. At present, the majority of Sahelian economics is based on mixed agro-pastoralism [81] but in many Sahelian countries, we still find purely nomadic pastoralists. We can thus see that a somewhat delayed spread of a particular food-production subsistence strategy may result in a lower migration activity and higher population differentiation. Despite some morphological differences detected between present-day full-time nomadic pastoralists and sedentary farmers [82], our genetic analyses show that lifestyle cannot be considered a determinative parameter of population sub-structuring in Africa, at least when considering the entire circum-Saharan region and not just the Sahel/Savannah belt where the biological separation of lifestyles may have played a more important role.

It seems that periodically, both in a long-term and short-term view, an increase in the size of the shallow Lake Chad in the middle of the east-western Sahelian corridor presented an obstacle to gene flow, forming a cul-de-sac. This can be documented by the spread of two different populations of nomadic pastoralists: in the west the Fulani, such as Wodaabe, and in the east the Arabs, such as Baggara or Shuwa. While the Arabs are of Eurasian (Arabian) ancestry and received gene flow from sub-Saharan Africans [28], the Fulani are of western African ancestry and their ancestors acquired some Eurasian ancestry by admixture with a northern African population possibly related to the Berbers [24,27]. Since the level of this admixture is relatively high (analyses show around 20% of a Eurasian component), it might be responsible for the noticeable differences between local Fulani populations and the surrounding relatively homogeneous Sahel/Savannah gene pool. On the other hand, the genetic diversity of sedentary farmers suggests that they lived in reproductively more isolated groups, which led to genetic drift and isolation by distance [10]. Interestingly, this can be associated also with the genetic diversity of their pearl millet landraces grown by different ethnolinguistic groups, especially in the western part of the Lake Chad Basin [83].

Finally, we found that the N1 haplogroup, which diversified in southwestern Asia some 55 ka [4] and its younger lineages expanded across all Eurasia, is present also in eastern Africa and the Sahel/Savannah belt. On the other hand, the age estimates of N1 mitogenomes we detected in our Sahelian populations are much younger and thus congruent with population contacts via Ba el-Mandeb or the Red Sea [84], not via the southern Mediterranean space. Moreover, two ancient samples from Takarkori rock shelter in Libya dated to ~7 ka [33] are highly distinct from our N1 Sahelian and eastern African samples because they branched off before all the current N lineages (far away from N1). We did not find any traces of local expansion of mtDNA N1a lineages into the Sahel/Savannah belt, as described for example for the Y chromosome R1b-V88 haplogroup [85,86] or for L3f3 mtDNA haplogroup [15]. This kind of (maternal) Eurasian N1 impact is visible mainly in eastern Africa, especially in Sudan, Somalia, Ethiopia, Kenya, and Tanzania, and not further in the central or even western part of the Sahel/Savannah belt close to the Lake Chad Basin. In fact, migration associated with this N1 African diversity might be associated with the Late Pleistocene/Holocene expansions in Arabia and the neighboring region and more recently also with the spread of Ethiosemitic languages into Ethiopia at ~3 ka (for example N1a3a+195C!), which continued further to South Africa [87] and not to the west.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes13030533/s1>, Supplementary Table S1: General information on studied population samples. Supplementary Table S2: Samples used for mitogenome sequencing. Supplementary Table S3: Samples used for phylogenetic analyses of the N1 haplogroup. Supplementary Table S4: The results of intrapopulation analyzes. Supplementary Table S5: Analyses of molecular variance (AMOVA). Supplementary Table S6: Matrix of coancestry coefficients colored according to genetic distance between populations in green < 0.09; yellow < 0.19; red > 0.2. Supplementary Table S7: Estimation of population size parameters scaled to mutation rate based on coalescent Bayesian approach Migrate-n. Supplementary Table S8: Estimation of immigration rate (A) and calculation of the relative number of immigrants (B) based on the results of coalescent Bayesian approach Migrate-n. Supplementary Table S9: Age estimates and standard deviations (in years) for the TMRCA of all N1a haplogroups analyzed.

Author Contributions: Conceptualization, V.Č. and L.P.; methodology, M.Y.D., M.Č. and V.F.; software, M.Y.D., M.Č. and V.F.; validation, J.N. and I.K.; formal analysis, M.Y.D., M.Č. and V.F.; investigation, V.Č., M.Y.D., E.P. and M.Č.; resources, E.P. (Eliška Podgorná) and E.P. (Edita Priehodová); data curation, J.N., I.K. and E.P. (Edita Priehodová); writing—original draft preparation, V.Č., M.Y.D., M.Č. and L.P.; writing—review and editing, V.Č., M.Y.D. and M.Č.; visualization, M.Y.D., M.Č. and V.F.; supervision, V.Č. and L.P.; project administration, V.Č.; funding acquisition, V.Č. and J.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Czech Science Foundation grant no. 19-09352S-P505. I.K. and J.N. were funded by project no. VI20162020015 of the Ministry of the Interior of the Czech Republic.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of Charles University, Faculty of Sciences (protocol code 2019/12, 28 March 2019).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The sequences were submitted to GenBank and are available under accession numbers OM952925-OM953385 for HVS-1 sequences and OM990709-OM990737 for mitogenomes.

Acknowledgments: The authors would like to express their sincere gratitude to Youba Ould Cheikhna, Mohammed Bechiry, Yousef Deddi, and Tabara Ba from Institut Mauritanien de Recherche et de Formation en Matière du Patrimoine et de la Culture in Nouakchott in Mauritania and to Mohammed G. Mokhtar from Arabic Department, Faculty of Arts, University of Kordofan, Al-Ubayyid in Sudan for cooperation in all administrative matters in their countries and assistance in the collection of the samples. The authors would like to thank two anonymous reviewers for helpful suggestions on improving the manuscript, and Anna Pilatová for proofreading the text.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bae, C.J.; Douka, K.; Petraglia, M.D. On the Origin of Modern Humans: Asian Perspectives. *Science* **2017**, *358*, eaai9067. [[CrossRef](#)] [[PubMed](#)]
2. Harvati, K.; Röding, C.; Bosman, A.M.; Karakostis, F.A.; Grün, R.; Stringer, C.; Karkanas, P.; Thompson, N.C.; Koutoulidis, V.; Mouloupoulos, L.A.; et al. Apidima Cave Fossils Provide Earliest Evidence of Homo Sapiens in Eurasia. *Nature* **2019**, *571*, 500–504. [[CrossRef](#)] [[PubMed](#)]
3. Malaspinas, A.-S.; Westaway, M.C.; Muller, C.; Sousa, V.C.; Lao, O.; Alves, I.; Bergström, A.; Athanasiadis, G.; Cheng, J.Y.; Crawford, J.E.; et al. A Genomic History of Aboriginal Australia. *Nature* **2016**, *538*, 207–214. [[CrossRef](#)]
4. Fernandes, V.; Alshamali, F.; Alves, M.; Costa, M.D.; Pereira, J.B.; Silva, N.M.; Cherni, L.; Harich, N.; Cerny, V.; Soares, P.; et al. The Arabian Cradle: Mitochondrial Relicts of the First Steps along the Southern Route out of Africa. *Am. J. Hum. Genet.* **2012**, *90*, 347–355. [[CrossRef](#)] [[PubMed](#)]
5. Auton, A.; Abecasis, G.R.; Altshuler, D.M.; Durbin, R.M.; Abecasis, G.R.; Bentley, D.R.; Chakravarti, A.; Clark, A.G.; Donnelly, P.; Eichler, E.E.; et al. A Global Reference for Human Genetic Variation. *Nature* **2015**, *526*, 68–74. [[CrossRef](#)]
6. Behar, D.M.; Metspalu, E.; Kivisild, T.; Rosset, S.; Tzur, S.; Hadid, Y.; Yudkovsky, G.; Rosengarten, D.; Pereira, L.; Amorim, A.; et al. Counting the Founders: The Matrilineal Genetic Ancestry of the Jewish Diaspora. *PLoS ONE* **2008**, *3*, e2062. [[CrossRef](#)] [[PubMed](#)]
7. Kivisild, T. Maternal Ancestry and Population History from Whole Mitochondrial Genomes. *Investig. Genet.* **2015**, *6*, 3. [[CrossRef](#)]

8. Černý, V.; Pereira, L.; Musilová, E.; Kujanová, M.; Vašíková, A.; Blasi, P.; Garofalo, L.; Soares, P.; Diallo, I.; Brdička, R.; et al. Genetic Structure of Pastoral and Farmer Populations in the African Sahel. *Mol. Biol. Evol.* **2011**, *28*, 2491–2500. [[CrossRef](#)]
9. Čížková, M.; Munclinger, P.; Diallo, M.Y.; Kulichová, I.; Mokhtar, M.G.; Dème, A.; Pereira, L.; Černý, V. Genetic Structure of the Western and Eastern African Sahel/Savannah Belt and the Role of Nomadic Pastoralists as Inferred from the Variation of D-Loop Mitochondrial DNA Sequences. *Hum. Biol.* **2017**, *89*, 281–302. [[CrossRef](#)] [[PubMed](#)]
10. Nováčková, J.; Čížková, M.; Mokhtar, M.G.; Duda, P.; Stenzl, V.; Triska, P.; Hofmanová, Z.; Černý, V. Subsistence Strategy Was the Main Factor Driving Population Differentiation in the Bidirectional Corridor of the African Sahel. *Am. J. Phys. Anthr.* **2020**, *171*, 496–508. [[CrossRef](#)]
11. Černý, V.; Salas, A.; Hájek, M.; Zaloudková, M.; Brdička, R. A Bidirectional Corridor in the Sahel-Sudan Belt and the Distinctive Features of the Chad Basin Populations: A History Revealed by the Mitochondrial DNA Genome. *Ann. Hum. Genet.* **2007**, *71*, 433–452. [[CrossRef](#)] [[PubMed](#)]
12. Pereira, L.; Černý, V.; Cerezo, M.; Silva, N.M.; Hájek, M.; Vašíková, A.; Kujanová, M.; Brdička, R.; Salas, A. Linking the Sub-Saharan and West Eurasian Gene Pools: Maternal and Paternal Heritage of the Tuareg Nomads from the African Sahel. *Eur. J. Hum. Genet.* **2010**, *18*, 915–923. [[CrossRef](#)] [[PubMed](#)]
13. 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.* **2013**, *77*, 513–523. [[CrossRef](#)] [[PubMed](#)]
14. Blench, R. *The Westward Wanderings of Cushitic Pastoralists: Explorations in the Prehistory of Central Africa*. In *L'Homme et l'animal dans le bassin du lac Tchad*; Editions IRD: Paris, France, 1999; pp. 39–80.
15. Černý, V.; Fernandes, V.; Costa, M.D.; Hájek, M.; Mulligan, C.J.; Pereira, L. Migration of Chadid Speaking Pastoralists within Africa Based on Population Structure of Chad Basin and Phylogeography of Mitochondrial L3f Haplogroup. *BMC Evol. Biol.* **2009**, *9*, 63. [[CrossRef](#)] [[PubMed](#)]
16. Cerezo, M.; Černý, V.; Carracedo, Á.; Salas, A. New Insights into the Lake Chad Basin Population Structure Revealed by High-Throughput Genotyping of Mitochondrial DNA Coding SNPs. *PLoS ONE* **2011**, *6*, e18682. [[CrossRef](#)] [[PubMed](#)]
17. Achilli, A.; Rengo, C.; Battaglia, V.; Pala, M.; Olivieri, A.; Fornarino, S.; Magri, C.; Scozzari, R.; Babudri, N.; Santachiara-Benerecetti, A.S.; et al. Saami and Berbers—an Unexpected Mitochondrial DNA Link. *Am. J. Hum. Genet.* **2005**, *76*, 883–886. [[CrossRef](#)] [[PubMed](#)]
18. Kulichová, I.; Fernandes, V.; Deme, A.; Nováčková, J.; Stenzl, V.; Novelletto, A.; Pereira, L.; Černý, V. Internal Diversification of Non-Sub-Saharan Haplogroups in Sahelian Populations and the Spread of Pastoralism beyond the Sahara. *Am. J. Phys. Anthropol.* **2017**, *164*, 424–434. [[CrossRef](#)]
19. Gifford-Gonzalez, D.; Hanotte, O. Domesticating Animals in Africa: Implications of Genetic and Archaeological Findings. *J. World Prehistory* **2011**, *24*, 1–23. [[CrossRef](#)]
20. Ozainne, S. *Un Néolithique Ouest-Africain: Cadre Chrono-Culturel, Économique et Environnemental de l'Holocène Récent En Pays Dogon (Mali)*; Africa Magna Verlag: Frankfurt am Main, Germany, 2013; Volume 8, ISBN 3-937248-33-1.
21. Ozainne, S.; Lespez, L.; Garnier, A.; Ballouche, A.; Neumann, K.; Pays, O.; Huysecom, E. A Question of Timing: Spatio-Temporal Structure and Mechanisms of Early Agriculture Expansion in West Africa. *J. Archaeol. Sci.* **2014**, *50*, 359–368. [[CrossRef](#)]
22. Sikora, M.; Carpenter, M.L.; Moreno-Estrada, A.; Henn, B.M.; Underhill, P.A.; Sánchez-Quinto, F.; Zara, I.; Pitzalis, M.; Sidore, C.; Busonero, F.; et al. Population Genomic Analysis of Ancient and Modern Genomes Yields New Insights into the Genetic Ancestry of the Tyrolean Iceman and the Genetic Structure of Europe. *PLoS Genet.* **2014**, *10*, e1004353. [[CrossRef](#)]
23. Chiang, C.W.K.; Marcus, J.H.; Sidore, C.; Biddanda, A.; Al-Asadi, H.; Zoledziwska, M.; Pitzalis, M.; Busonero, F.; Maschio, A.; Pistis, G.; et al. Genomic History of the Sardinian Population. *Nat. Genet.* **2018**, *50*, 1426–1434. [[CrossRef](#)] [[PubMed](#)]
24. Vicente, M.; Priehodová, E.; Diallo, I.; Podgorná, E.; Poloni, E.S.; Černý, V.; Schlebusch, C.M. Population History and Genetic Adaptation of the Fulani Nomads: Inferences from Genome-Wide Data and the Lactase Persistence Trait. *BMC Genom.* **2019**, *20*, 915. [[CrossRef](#)] [[PubMed](#)]
25. Priehodová, E.; Austerlitz, F.; Čížková, M.; Nováčková, J.; Ricaut, F.-X.; Hofmanová, Z.; Schlebusch, C.M.; Černý, V. Sahelian Pastoralism from the Perspective of Variants Associated with Lactase Persistence. *Am. J. Phys. Anthr.* **2020**, *173*, 423–436. [[CrossRef](#)] [[PubMed](#)]
26. Černý, V.; Fortes-Lima, C.; Triska, P. Demographic History and Admixture Dynamics in African Sahelian Populations. *Hum. Mol. Genet.* **2021**, *30*, R29–R36. [[CrossRef](#)] [[PubMed](#)]
27. Triska, P.; Soares, P.; Patin, E.; Fernandes, V.; Černý, V.; Pereira, L. Extensive Admixture and Selective Pressure Across the Sahel Belt. *Genome Biol. Evol.* **2015**, *7*, 3484–3495. [[CrossRef](#)] [[PubMed](#)]
28. Priehodová, E.; Austerlitz, F.; Čížková, M.; Mokhtar, M.G.; Poloni, E.S.; Černý, V. The Historical Spread of Arabian Pastoralists to the Eastern African Sahel Evidenced by the Lactase Persistence –13,915*G Allele and Mitochondrial DNA. *Am. J. Hum. Biol.* **2017**, *29*, e22950. [[CrossRef](#)] [[PubMed](#)]
29. Levy, T.E.; Holl, A.F.C. Migrations, Ethnogenesis, and Settlement Dynamics: Israelites in Iron Age Canaan and Shuwa-Arabs in the Chad Basin. *J. Anthropol. Archaeol.* **2002**, *21*, 83–118. [[CrossRef](#)]
30. Pereira, L.; Mutesa, L.; Tindana, P.; Ramsay, M. African Genetic Diversity and Adaptation Inform a Precision Medicine Agenda. *Nat. Rev. Genet.* **2021**, *22*, 284–306. [[CrossRef](#)]

31. Mulindwa, J.; Noyes, H.; Ilboudo, H.; Pagani, L.; Nyangiri, O.; Kimuda, M.P.; Ahouty, B.; Asina, O.F.; Ofon, E.; Kamoto, K.; et al. High Levels of Genetic Diversity within Nilo-Saharan Populations: Implications for Human Adaptation. *Am. J. Hum. Genet.* **2020**, *107*, 473–486. [\[CrossRef\]](#)
32. Choudhury, A.; Aron, S.; Botigué, L.R.; Sengupta, D.; Botha, G.; Bensellak, T.; Wells, G.; Kumuthini, J.; Shriner, D.; Fakim, Y.J.; et al. High-Depth African Genomes Inform Human Migration and Health. *Nature* **2020**, *586*, 741–748. [\[CrossRef\]](#)
33. Vai, S.; Sarno, S.; Lari, M.; Luiselli, D.; Manzi, G.; Gallinaro, M.; Mataich, S.; Hübner, A.; Modi, A.; Pilli, E.; et al. Ancestral Mitochondrial N Lineage from the Neolithic ‘Green’ Sahara. *Sci. Rep.* **2019**, *9*, 3530. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Van Oven, M.; Kayser, M. Updated Comprehensive Phylogenetic Tree of Global Human Mitochondrial DNA Variation. *Hum. Mutat.* **2009**, *30*, E386–E394. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Gonder, M.K.; Mortensen, H.M.; Reed, F.A.; de Sousa, A.; Tishkoff, S.A. Whole-MtDNA Genome Sequence Analysis of Ancient African Lineages. *Mol. Biol. Evol.* **2007**, *24*, 757–768. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Bendall, K.E.; Sykes, B.C. Length Heteroplasmy in the First Hypervariable Segment of the Human MtDNA Control Region. *Am. J. Hum. Genet.* **1995**, *57*, 248.
37. Excoffier, L.; Lischer, H.E.L. Arlequin Suite Ver 3.5: A New Series of Programs to Perform Population Genetics Analyses under Linux and Windows. *Mol. Ecol. Resour.* **2010**, *10*, 564–567. [\[CrossRef\]](#)
38. Tajima, F. Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics* **1989**, *123*, 585–595. [\[CrossRef\]](#)
39. Fu, Y.X. Statistical Tests of Neutrality of Mutations against Population Growth, Hitchhiking and Background Selection. *Genetics* **1997**, *147*, 915–925. [\[CrossRef\]](#)
40. Ramos-Onsins, S.E.; Rozas, J. Statistical Properties of New Neutrality Tests against Population Growth. *Mol. Biol. Evol.* **2002**, *19*, 2092–2100. [\[CrossRef\]](#)
41. Rozas, J.; Ferrer-Mata, A.; Sánchez-DelBarrio, J.C.; Guirao-Rico, S.; Librado, P.; Ramos-Onsins, S.E.; Sánchez-Gracia, A. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Mol. Biol. Evol.* **2017**, *34*, 3299–3302. [\[CrossRef\]](#)
42. Rogers, A.R.; Harpending, H. Population Growth Makes Waves in the Distribution of Pairwise Genetic Differences. *Mol. Biol. Evol.* **1992**, *9*, 552–569. [\[CrossRef\]](#)
43. Reynolds, J.; Weir, B.S.; Cockerham, C.C. Estimation of the Coancestry Coefficient: Basis for a Short-Term Genetic Distance. *Genetics* **1983**, *105*, 767–779. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Slatkin, M. A Correction to the Exact Test Based on the Ewens Sampling Distribution. *Genet. Res.* **1996**, *68*, 259–260. [\[CrossRef\]](#)
45. Poloni, E.S.; Naciri, Y.; Bucho, R.; Niba, R.; Kervaire, B.; Excoffier, L.; Langaney, A.; Sanchez-Mazas, A. Genetic Evidence for Complexity in Ethnic Differentiation and History in East Africa. *Ann. Hum. Genet.* **2009**, *73*, 582–600. [\[CrossRef\]](#)
46. Jari Oksanen, F.; Blanchet, G.; Friendly, M.; Kindt, R.; Legendre, P.; McGlinn, D.; Peter, R.; Minchin, R.B.; O’Hara, G.; Simpson, L.; et al. *Vegan: Community Ecology Package. R Package Version*; R Package: Vienna, Austria, 2020; pp. 2–5.
47. RStudio Team. *RStudio: Integrated Development for R*; RStudio, Inc.: Boston, MA, USA, 2016.
48. Beerli, P. How to Use MIGRATE or Why Are Markov Chain Monte Carlo Programs Difficult to Use? In *Population Genetics for Animal Conservation*; Conservation Biology; Rizzoli, A., Vernesi, C., Bertorelle, G., Haufler, H.C., Bruford, M.W., Eds.; Cambridge University Press: Cambridge, UK, 2009; pp. 42–79, ISBN 978-0-521-86630-9.
49. Gu, Z.; Gu, L.; Eils, R.; Schlesner, M.; Brors, B. Circlize Implements and Enhances Circular Visualization in R. *Bioinformatics* **2014**, *30*, 2811–2812. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Henn, B.M.; Botigué, L.R.; Gravel, S.; Wang, W.; Brisbin, A.; Byrnes, J.K.; Fadhlouzi-Zid, K.; Zalloua, P.A.; Moreno-Estrada, A.; Bertranpetit, J.; et al. Genomic Ancestry of North Africans Supports Back-to-Africa Migrations. *PLoS Genet.* **2012**, *8*, e1002397. [\[CrossRef\]](#) [\[PubMed\]](#)
51. Andrews, R.M.; Kubacka, I.; Chinnery, P.F.; Lightowlers, R.N.; Turnbull, D.M.; Howell, N. Reanalysis and Revision of the Cambridge Reference Sequence for Human Mitochondrial DNA. *Nat. Genet.* **1999**, *23*, 147. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Forster, P.; Harding, R.; Torroni, A.; Bandelt, H.J. Origin and Evolution of Native American MtDNA Variation: A Reappraisal. *Am. J. Hum. Genet.* **1996**, *59*, 935–945.
53. Soares, P.; Ermini, L.; Thomson, N.; Mormina, M.; Rito, T.; Röhl, A.; Salas, A.; Oppenheimer, S.; Macaulay, V.; Richards, M.B. Correcting for Purifying Selection: An Improved Human Mitochondrial Molecular Clock. *Am. J. Hum. Genet.* **2009**, *84*, 740–759. [\[CrossRef\]](#)
54. Saillard, J.; Forster, P.; Lynnerup, N.; Bandelt, H.-J.; Nørby, S. MtDNA Variation among Greenland Eskimos: The Edge of the Beringian Expansion. *Am. J. Hum. Genet.* **2000**, *67*, 718–726. [\[CrossRef\]](#)
55. Yang, Z. PAML: A Program Package for Phylogenetic Analysis by Maximum Likelihood. *Comput. Appl. Biosci.* **1997**, *13*, 555–556. [\[CrossRef\]](#)
56. Černý, V.; Hájek, M.; Bromová, M.; Čmejla, R.; Diallo, I.; Brdička, R. MtDNA of Fulani nomads and their genetic relationships to neighboring sedentary populations. *Hum. Biol.* **2006**, *78*, 9–27. [\[CrossRef\]](#)
57. van de Loosdrecht, M.; Bouzouggar, A.; Humphrey, L.; Posth, C.; Barton, N.; Aximu-Petri, A.; Nickel, B.; Nagel, S.; Talbi, E.H.; El Hajraoui, M.A.; et al. Pleistocene North African Genomes Link Near Eastern and Sub-Saharan African Human Populations. *Science* **2018**, *360*, 548–552. [\[CrossRef\]](#) [\[PubMed\]](#)

58. Ferreira, J.C.; Alshamali, F.; Montinaro, F.; Cavadas, B.; Torroni, A.; Pereira, L.; Raveane, A.; Fernandes, V. Projecting Ancient Ancestry in Modern-Day Arabians and Iranians: A Key Role of the Past Exposed Arabo-Persian Gulf on Human Migrations. *Genome Biol. Evol.* **2021**, *13*, evab194. [\[CrossRef\]](#)
59. Arredi, B.; Poloni, E.S.; Paracchini, S.; Zerjal, T.; Fathallah, D.M.; Makrelouf, M.; Pascali, V.L.; Novelletto, A.; Tyler-Smith, C. A Predominantly Neolithic Origin for Y-Chromosomal DNA Variation in North Africa. *Am. J. Hum. Genet.* **2004**, *75*, 338–345. [\[CrossRef\]](#) [\[PubMed\]](#)
60. Kujanová, M.; Pereira, L.; Fernandes, V.; Pereira, J.B.; Cerný, V. Near Eastern Neolithic Genetic Input in a Small Oasis of the Egyptian Western Desert. *Am. J. Phys. Anthr.* **2009**, *140*, 336–346. [\[CrossRef\]](#) [\[PubMed\]](#)
61. Černý, V.; Čížková, M.; Poloni, E.S.; Al-Meer, A.; Mulligan, C.J. Comprehensive View of the Population History of Arabia as Inferred by MtDNA Variation. *Am. J. Phys. Anthr.* **2016**, *159*, 607–616. [\[CrossRef\]](#) [\[PubMed\]](#)
62. Currat, M.; Excoffier, L.; Penny, D. Modern Humans Did Not Admix with Neanderthals during Their Range Expansion into Europe. *PLoS Biol.* **2004**, *2*, e421. [\[CrossRef\]](#) [\[PubMed\]](#)
63. Pimenta, J.; Lopes, A.M.; Comas, D.; Amorim, A.; Arenas, M. Evaluating the Neolithic Expansion at Both Shores of the Mediterranean Sea. *Mol. Biol. Evol.* **2017**, *34*, 3232–3242. [\[CrossRef\]](#) [\[PubMed\]](#)
64. Fregel, R.; Méndez, F.L.; Bokbot, Y.; Martín-Socas, D.; Camalich-Massieu, M.D.; Santana, J.; Morales, J.; Ávila-Arcos, M.C.; Underhill, P.A.; Shapiro, B.; et al. Ancient Genomes from North Africa Evidence Prehistoric Migrations to the Maghreb from Both the Levant and Europe. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 6774–6779. [\[CrossRef\]](#)
65. Hernández, C.L.; Pita, G.; Cavadas, B.; López, S.; Sánchez-Martínez, L.J.; Dugoujon, J.-M.; Novelletto, A.; Cuesta, P.; Pereira, L.; Calderón, R. Human Genomic Diversity Where the Mediterranean Joins the Atlantic. *Mol. Biol. Evol.* **2020**, *37*, 1041–1055. [\[CrossRef\]](#)
66. Bekada, A. *Caractérisation Anthropogénétique d'un Échantillon de la Population Algérienne: Analyse des Marqueurs Parentaux*; Université d'Oran 1: Es Senia, Algeria, 2015.
67. Adams, W.Y. *Nubia: Corridor to Africa*; Princeton University Press: Princeton, NJ, USA, 1977; ISBN 978-0-691-09370-3.
68. Fox, C.L. MtDNA Analysis in Ancient Nubians Supports the Existence of Gene Flow between Sub-Sahara and North Africa in the Nile Valley. *Ann. Hum. Biol.* **1997**, *24*, 217–227. [\[CrossRef\]](#)
69. Krings, M.; Salem, A.E.; Bauer, K.; Geisert, H.; Malek, A.K.; Chaix, L.; Simon, C.; Welsby, D.; Di Rienzo, A.; Utermann, G.; et al. MtDNA Analysis of Nile River Valley Populations: A Genetic Corridor or a Barrier to Migration? *Am. J. Hum. Genet.* **1999**, *64*, 1166–1176. [\[CrossRef\]](#)
70. Segal, R. *Islam's Black Slaves: The History of Africa's Other Black Diaspora*; Atlantic Books: London, UK, 2002; ISBN 978-1-903809-81-5.
71. Turner, M.D.; Schlecht, E. Livestock Mobility in Sub-Saharan Africa: A Critical Review. *Pastoralism* **2019**, *9*, 13. [\[CrossRef\]](#)
72. Jesse, F.; Keding, B.; Lenssen-Erz, T.; Pöllath, N. I Hope Your Cattle Are Well': Archaeological Evidence for Early Cattle-Centred Behaviour in the Eastern Sahara of Sudan and Chad. In *Pastoralism in Africa: Past, Present and Future*; Berghahn Books: New York, NY, USA, 2013; pp. 66–103.
73. Linseele, V. From First Stock Keepers to Specialised Pastoralists in the West African Savannah. In *Pastoralism in Africa: Past, Present and Future*; Bollig, M., Schnegg, M., Wotzka, H.-P., Eds.; Berghahn Books: New York, NY, USA, 2013; pp. 145–170.
74. Ozainne, S.; Mayor, A.; Huysecom, E. Chronology of Human Occupation during the Holocene in West Africa. The Dogon country radiocarbon record. In *Winds of change*; Rudolf Habelt Verlag: Bonn, Germany, 2017; pp. 211–227.
75. Wright, D.K. Humans as Agents in the Termination of the African Humid Period. *Front. Earth Sci.* **2017**, *5*, 4. [\[CrossRef\]](#)
76. Manning, K.; Pelling, R.; Higham, T.; Schwenniger, J.-L.; Fuller, D.Q. 4500-Year Old Domesticated Pearl Millet (*Pennisetum glaucum*) from the Tilemsi Valley, Mali: New Insights into an Alternative Cereal Domestication Pathway. *J. Archaeol. Sci.* **2011**, *38*, 312–322. [\[CrossRef\]](#)
77. Burgarella, C.; Cubry, P.; Kane, N.A.; Varshney, R.K.; Mariac, C.; Liu, X.; Shi, C.; Thudi, M.; Couderc, M.; Xu, X.; et al. A Western Sahara Centre of Domestication Inferred from Pearl Millet Genomes. *Nat. Ecol. Evol.* **2018**, *2*, 1377–1380. [\[CrossRef\]](#) [\[PubMed\]](#)
78. Cubry, P.; Tranchant-Dubreuil, C.; Thuillet, A.-C.; Monat, C.; Ndjondjop, M.-N.; Labadie, K.; Cruaud, C.; Engelen, S.; Scarcelli, N.; Rhoné, B.; et al. The Rise and Fall of African Rice Cultivation Revealed by Analysis of 246 New Genomes. *Curr. Biol.* **2018**, *28*, 2274–2282.e6. [\[CrossRef\]](#)
79. Scarcelli, N.; Cubry, P.; Akakpo, R.; Thuillet, A.-C.; Obidiegwu, J.; Baco, M.N.; Otoo, E.; Sonké, B.; Dansi, A.; Djedatin, G.; et al. Yam Genomics Supports West Africa as a Major Cradle of Crop Domestication. *Sci. Adv.* **2019**, *5*, eaaw1947. [\[CrossRef\]](#)
80. Winchell, F.; Stevens, C.J.; Murphy, C.; Champion, L.; Fuller, D.Q. Evidence for Sorghum Domestication in Fourth Millennium BC Eastern Sudan: Spikelet Morphology from Ceramic Impressions of the Butana Group. *Curr. Anthropol.* **2017**, *58*, 673–683. [\[CrossRef\]](#)
81. Cervigni, R.; Morris, M. *Confronting Drought in Africa's Drylands: Opportunities for Enhancing Resilience*; World Bank Publications: Herndon, VA, USA, 2016; ISBN 978-1-4648-0818-0.
82. Kleisner, K.; Pokorný, Š.; Čížková, M.; Froment, A.; Černý, V. Nomadic Pastoralists and Sedentary Farmers of the Sahel/Savannah Belt of Africa in the Light of Geometric Morphometrics Based on Facial Portraits. *Am. J. Phys. Anthropol.* **2019**, *169*, 632–645. [\[CrossRef\]](#)
83. Naino Jika, A.K.; Dussert, Y.; Raimond, C.; Garine, E.; Luxereau, A.; Takvorian, N.; Djermakoye, R.S.; Adam, T.; Robert, T. Unexpected Pattern of Pearl Millet Genetic Diversity among Ethno-Linguistic Groups in the Lake Chad Basin. *Heredity* **2017**, *118*, 491–502. [\[CrossRef\]](#) [\[PubMed\]](#)

-
84. Musilová, E.; Fernandes, V.; Silva, N.M.; Soares, P.; Alshamali, F.; Harich, N.; Cherni, L.; Gaaied, A.B.; Al-Meer, A.; Pereira, L.; et al. Population history of the Red Sea—Genetic exchanges between the Arabian Peninsula and East Africa signaled in the mitochondrial DNA HV1 haplogroup. *Am. J. Phys. Anthropol.* **2011**, *145*, 592–598. [[CrossRef](#)] [[PubMed](#)]
 85. Cruciani, F.; Trombetta, B.; Sellitto, D.; Massaia, A.; Destro-Bisol, G.; Watson, E.; Beraud Colomb, E.; Dugoujon, J.-M.; Moral, P.; Scozzari, R. Human Y Chromosome Haplogroup R-V88: A Paternal Genetic Record of Early Mid Holocene Trans-Saharan Connections and the Spread of Chadic Languages. *Eur. J. Hum. Genet.* **2010**, *18*, 800–807. [[CrossRef](#)] [[PubMed](#)]
 86. Shriner, D.; Rotimi, C.N. Genetic History of Chad. *Am. J. Phys. Anthropol.* **2018**, *167*, 804–812. [[CrossRef](#)] [[PubMed](#)]
 87. Pickrell, J.K.; Patterson, N.; Loh, P.-R.; Lipson, M.; Berger, B.; Stoneking, M.; Pakendorf, B.; Reich, D. Ancient West Eurasian Ancestry in Southern and Eastern Africa. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 2632–2637. [[CrossRef](#)] [[PubMed](#)]