Report

Current Biology

Heterogeneity in Palaeolithic Population Continuity and Neolithic Expansion in North Africa

Highlights

- Paleolithic genetic continuity is found in extant human North African genomes
- There is a West-to-East genetic cline of the Paleolithic component in North Africa
- Neolithization had a larger demographic impact than Arabization
- Differential admixture and genetic drift have modeled North African genomes

Authors

Gerard Serra-Vidal, Marcel Lucas-Sanchez, Karima Fadhlaoui-Zid, Asmahan Bekada, Pierre Zalloua, David Comas

Correspondence

david.comas@upf.edu

In Brief

Serra-Vidal et al. report that human North African genomes exhibit traces of continuity from Paleolithic times in a decreasing pattern from West to East. Neolithic and posterior migrations did not erase the pre-existing Paleolithic substrate. The demographic impact of Neolithization was larger than posterior migrations, such as Arabization.

Serra-Vidal et al., 2019, Current Biology *29*, 1–7
November 18, 2019 © 2019 Elsevier Ltd.
https://doi.org/10.1016/j.cub.2019.09.050

CellPress

Please cite this article in press as: Serra-Vidal et al., Heterogeneity in Palaeolithic Population Continuity and Neolithic Expansion in North Africa, Current Biology (2019), https://doi.org/10.1016/j.cub.2019.09.050

Current Biology
Report

Heterogeneity in Palaeolithic Population Continuity and Neolithic Expansion in North Africa

Gerard Serra-Vidal,[1] Marcel Lucas-Sanchez,[1] Karima Fadhlaoui-Zid,[2,3] Asmahan Bekada,[4] Pierre Zalloua,[5] and David Comas[1,6,*]

[1]Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Departament de Ciencies Experimentals i de la Salut, Carrer del Doctor Aiguader 88, Barcelona 08003, Spain

[2]College of Science, Department of Biology, Taibah University, Street Janadah Bin Umayyah Road, Al Madinah Al Monawarah 42353, Saudi Arabia

[3]Laboratory of Genetics, Immunology, and Human Pathologies, Faculty of Science of Tunis, University Tunis El Manar, Street Tolede, Tunis 2092, Tunisia

[4]Département de Biotechnologie, Faculté des Sciences de la Nature et de la Vie, Université Oran 1 (Ahmad Ben Bella), 1524 El M'Naouer, Oran 31000, Algeria

[5]The Lebanese American University, Chouran, School of Medicine, Chouran Street, Beirut 1102-2801, Lebanon

[6]Lead Contact

*Correspondence: david.comas@upf.edu

https://doi.org/10.1016/j.cub.2019.09.050

SUMMARY

North Africa is located at the crossroads of the Mediterranean Sea, the Middle East, and the Sahara Desert. Extensive migrations and gene flow in the region have shaped many different cultures and ancestral genetic components through time [1–6]. DNA data from ancient Moroccan sites [7, 8] has recently shed some light to the population continuity-versus-replacement debate, i.e., the question of whether current North African populations descend from Palaeolithic groups or, on the contrary, subsequent migrations swept away all pre-existing genetic signal in the region. In the present study, we analyze 21 complete North African genomes and compare them with extant and ancient genome data in order to address the demographic continuity-versus-replacement debate, to assess whether these demographic events were homogeneous (including Berber and Arabic-speaking groups), and to explore the effect of Neolithization and posterior migration waves. The North African genetic pool is defined as a melting pot of genetic components, including an endemic North African Epipalaeolithic component at low frequency that forms a declining gradient from Western to Eastern North Africa. This scenario is consistent with Neolithization having shaped most of the current genetic variation in the region when compared to posterior back-to-North-Africa migration waves such as the Arabization. A common and distinct genetic history of the region is shown, with internal different proportions of genetic components owing to differential admixture with surrounding groups as well as to genetic drift due to isolation and endogamy in certain populations.

RESULTS AND DISCUSSION

Genetic Components and Population Structure in North Africans

Seventeen North African genomes were sequenced together with sixteen sub-Saharan and Eurasian individuals. In total, 10.86 million SNPs were called, using Illumina HiSeq 2000 at a mean coverage of 26×. All analyses were restricted to a high-confidence, 2-Gbp fraction of the genome obtained after applying quality control filtering (see STAR Methods). As a validation process, seven samples were independently genotyped through Affymetrix 6.0 SNP array (see STAR Methods). See Table S1 for extra information and summary statistics.

The initial dataset was merged with other available datasets (see STAR Methods), providing ten current and four ancient North African groups (Figure S1). A first exploration of the data was performed using principal component analysis (PCA). The first component (PC1, accounting for 3.5% of the variation) captures the genetic differentiation between sub-Saharan Africans and non-African populations, with ancient and current North Africans placed in an intermediate position (Figure 1A). PC2 (0.7% of the variation) splits Middle Easterns and Europeans, with North Africans closer to the former. Regarding ancient North Africans, while Canary Guanches ($\sim 5^{th}$ century BCE) cluster with current North Africans (in agreement with their putative Berber origin [9]), Moroccan Epipalaeolithic samples from Taforalt cluster independently, while Moroccan Early (IAM) and Late Neolithic (KEB) have intermediate positions in the PC.

Population structure and ancestry components were determined by ADMIXTURE (Figures 1B and S2). The lowest cross-validation errors were found in the range between $K = 4$ and $K = 7$, which depicts North African ancestry as a mosaic of components that are consistently conserved across different values of K (Figure 1C): (1) a sub-Saharan component derived from trans-Saharan gene flow (black); (2) European and Anatolian Neolithic component (white); (3) an ancient Middle Eastern component, prevalent in Natufian and Levant Neolithic and

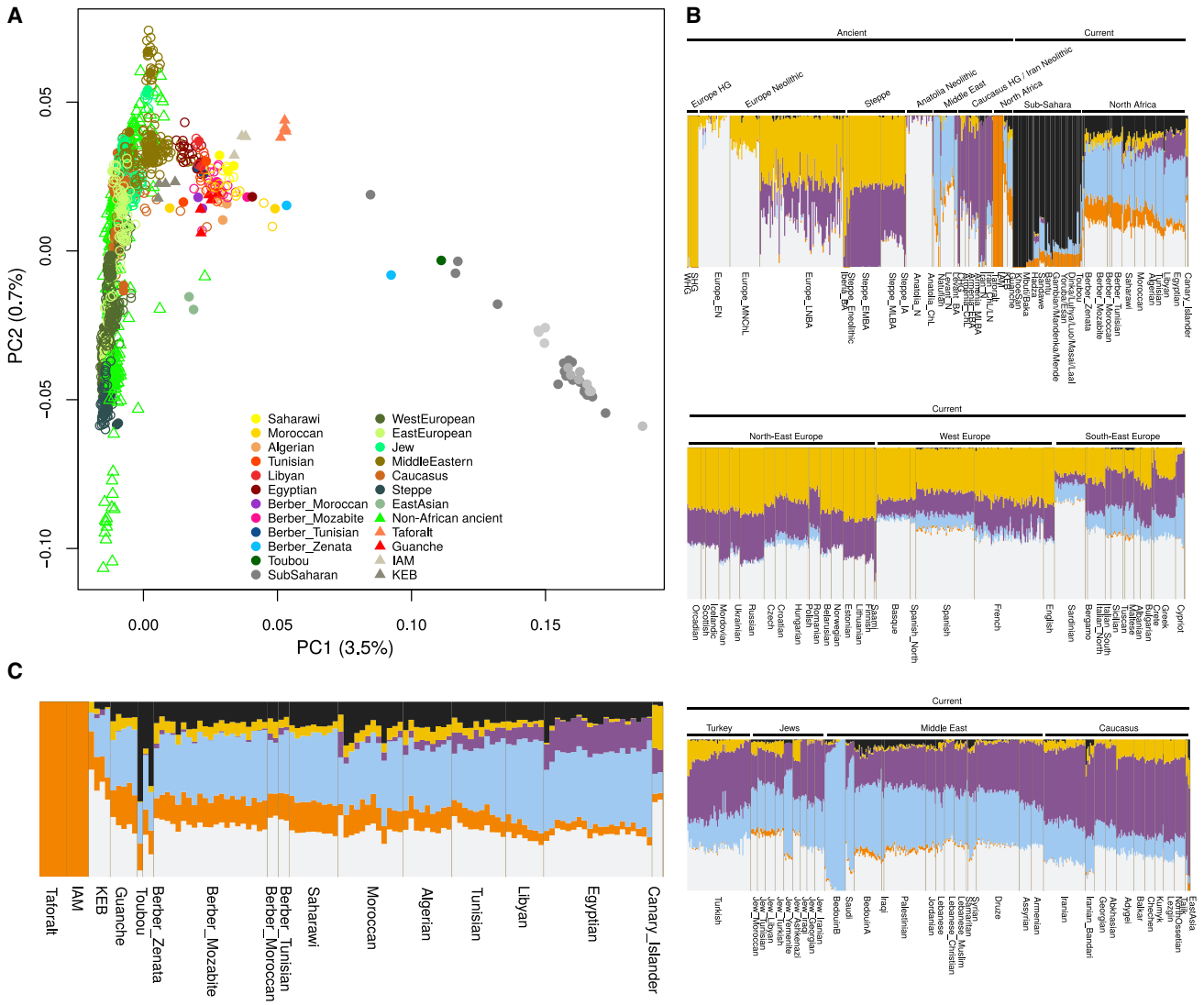

Figure 1. Principal Component Analysis and ADMIXTURE Analysis for *K* = 6

(A) Principal component analysis of North African samples together with a worldwide panel described in STAR Methods. Ancient samples (represented by triangles) are projected on top of current samples (circles). Filled circles and triangles correspond to sequence data, and empty shapes correspond to array data. (B and C) ADMIXTURE analysis (K = 6) of North African samples together with a broad panel of worldwide ancient and current populations from Africa, Europe, Middle East, and Caucasus (B) with a zoom-in on North African current and ancient populations (C).

See Figure S2 for ADMIXTURE results from K = 2 to K = 12.

also present in current Levantine populations, particularly in Bedouin groups (blue); (4) a component coming from Caucasus hunter-gatherers and Iran Neolithic (purple); and, (5) a North African autochthonous Epipalaeolithic component prevalent in the Moroccan Epipalaeolithic from Taforalt and Early Neolithic samples (orange), observed at low proportions in Moroccan Late Neolithic, Guanches, and in current Canary Islanders. The North African autochthonous component is absent in any other population outside North Africa from *K* = 7 onward, while its presence for lower *K* in sub-Saharan population values might be explained by the presence of a sub-Saharan ancestral component in North African Palaeolithic populations, as pointed out by [8].

Haplotype-based methods also point to the presence of a genetic component coming from Epipalaeolithic times in current North African populations. Taforalt samples cluster together with current North Africans in the fineSTRUCTURE haplotype-sharing-based tree (Figure S3), while the ChromoPainter coancestry matrix results point to higher levels of genomic tracts in current North Africans coming from their Epipalaeolithic ancestors than in any other extant population. Most North African current populations cluster together in a single cluster, although genetic groups within the North African cluster do not correspond to geographical or linguistic-based populations, in agreement with certain degree of genetic heterogeneity ([1] and others).

Even though a sudden genetic change comparing early and late Neolithic samples in Morocco has been recently proposed [7], our data show that there are still traces of the Epipalaeolithic ancestors in the genomes of extant North Africans, as shown by

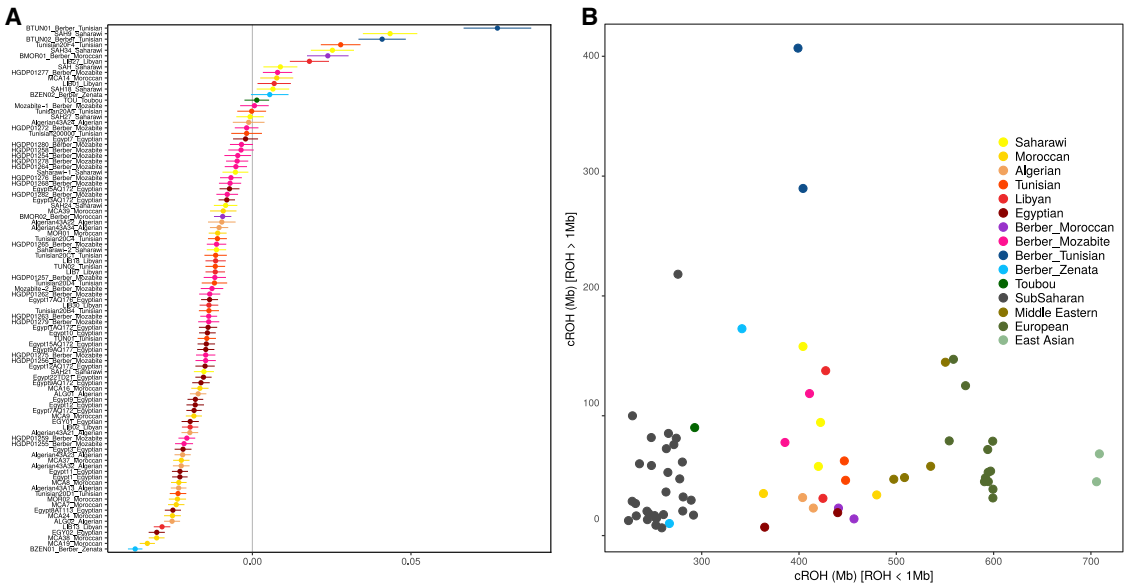

Figure 2. Admixture-*f3*(Yoruba-Mandenka, Basque-Iraqi; North African) and Runs of Homozygosity Analysis

(A) Admixture-*f3* statistic (*Z* score and standard error) for North African individuals computed as *f3*(Yoruba-Mandenka, Basque-Iraqi; North African).

(B) Scatterplot of the cumulative length of long (>1 Mbp) versus short (<1 Mbp) runs of homozygosity (ROH) per sample.

See Figure S4 for outgroup-*f3* results.

the admixture components and the amount of shared haplotypes between ancient and present North African genomes. Our results confirm that gene flow in the area, coming for surrounding regions such as Europe, the Middle East or sub-Saharan Africa did not completely erase the ancient background of autochthonous North Africans in the last 15,000 years. This Palaeolithic autochthonous genetic component might correlate with the Maghrebi component as a result of a back-to-Africa gene flow defined by [6], and its presence is not distributed following a uniform pattern in the area.

Genetic Heterogeneity within North Africa

Internal differences can be observed in both the ADMIXTURE and ChromoPainter results; in particular, some ancestral components present geographical gradients across the region. The gradients observed in the previous analyses were tested with f_3(X, Taforalt; Ju/'hoanNorth), f_3(X, Yoruba-Mandenka; Ju/'hoanNorth), and f_3(X, CHG-Iran_N; Ju/'hoanNorth) (Figure S4).

Saharawi and Berber groups show the highest outgroup-f_3 values for the Epipalaeolithic Taforalt component (0.218–0.222). The ancient samples IAM (0.322), KEB (0.235), and Guanche (0.227) show a higher North African Epipalaeolithic component than current populations (in agreement with the ADMIXTURE analysis), whose decrease is compatible with a dilution of the Taforalt component through time. This component is significantly more frequent in Western (Saharawi, Moroccan, Algerian) and Berber-speaking individuals (Saharawi, Mozabites, Moroccan, and Moroccan and Tunisian Berbers) (Mann-Whitney U test, p value = $4.52e^{-15}$), suggesting a continuity of this autochthonous North African component in Berber-speaking groups. Although no perfect correlation between culture (i.e., Arabic- and Berber-speaking groups) and genetics can be claimed ([1]), the consideration of the Berber-speaking groups as the autochthonous peoples of North Africa [10] is reinforced by the present results.

Zenata, Mozabite, Saharawi, and Moroccan groups show the highest proportions of sub-Saharan ancestry, which could be attributed to the slave trade routes carried out during Roman and Arab presence mainly in northwest Africa [11, 12]. The Taforalt and Moroccan Early Neolithic have a higher sub-Saharan affinity than most current North Africans (as stated by [8]), whereas the Moroccan Late Neolithic and the Guanches have a similar level of sub-Saharan affinity to most current groups analyzed in the present study.

Egyptian and Libyan show the highest proportion of the Caucasus-Iran component, in agreement with their geographical proximity to southwest Asia. A high f_3 value is also estimated for KEB (0.250), in contrast to the lower estimated values found for Taforalt and IAM (0.206, 0.216), suggesting that this component might have entered North Africa during the late Neolithic coming from Iran and may have been posteriorly diluted in western North Africa.

Admixture Models in North Africa

North African populations have been described as having a major influence from outside the African continent, together with a sub-Saharan component coming from trans-Saharan migrations [6]. In order to determine to which extent the North African genomes can be modeled as an admixture of sub-Saharans and out-of-African components, f_3 admixture test was performed as f_3(Sub-Saharan, West Eurasian; X), with Yoruba and Mandenka as sub-Saharan populations, Basques and Iraqis as West Eurasians, and X being each North African group analyzed. f_3 results are negative for most North African populations, fitting a model of admixture between sub-Saharans and Eurasians. However, some samples, such as some Algerian, Mozabite,

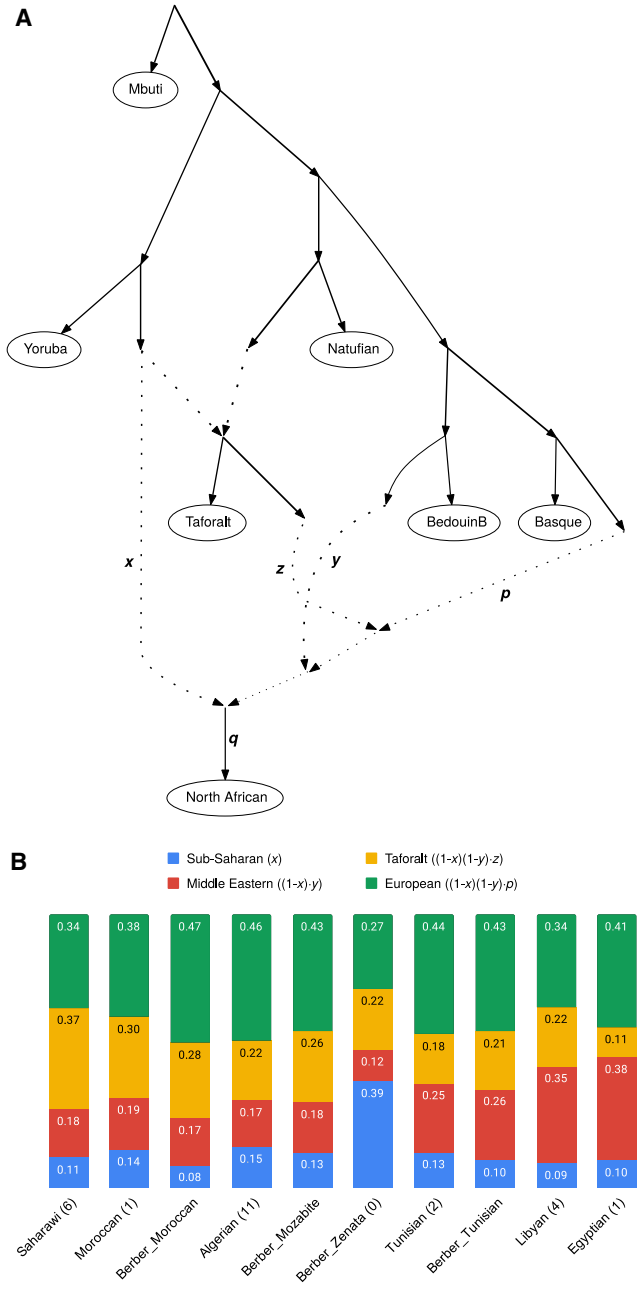

Figure 3. Tested qpGraph Model and Proportions of Ancestral Components in North African Populations

(A) qpGraph model fitting the data for the tested populations. The terminal node labeled as North African represents any North African group. x, y, z, and p stand for the proportion of admixture related to the sub-Saharan, Middle Eastern, Taforalt, and European components, respectively. q corresponds to the genetic drift parameter for each North African group. The tested models fitted the real data for all North African populations ($Z < 3$).

(B) Admixture proportions of ancestral components for each target North African population. The number after each population sample name corresponds to parameter q.

See Table S2 for MALDER results.

Zenata, Saharawi, and particularly Tunisian Berbers, show significantly positive f_3 results (Figure 2A) due to extensive genetic drift, as further proved by a TreeMix analysis (data not shown), which discards a different genetic history of these groups.

To determine the main Eurasian ancestry sources in North Africa during the back-to-Africa events and test whether these sources are homogeneous across North African groups, admixture-f_3 tests of the form f_3(OOA, SS; NA) were performed (OOA, out of Africa population; SS, Sub-Saharan population; NA, North African population). The sub-Saharan source was fixed to Yoruba (no significantly different results were found when choosing an East African group, such as Dinka, instead; data not shown). For all groups (with the exception of Tunisian Berbers, for which admixture was not detected through f_3, see above), the most significant results were found for Sardinians, Basques, and North Italians (data not shown). When the analyses were repeated adding ancient Eurasian populations [13] to the dataset, the higher significant values were found with the Neolithic farmers from Europe and Anatolia and, again, Sardinian and Basques, which is consistent with their high frequency of Neolithic component [13]. These results point to the main role of the Neolithization process when shaping the current North African genetic landscape, thus supporting the PCA, ADMIXTURE, and internal f_3 tests, where Eurasian gene flow after Neolithization (such as the Arabization process starting in the 7th century CE) seem to have had a lower impact. The arrival of the Neolithic to North Africa as a demic diffusion process from the Middle East with putative interactions with local groups is widely accepted [14, 15], although some hypotheses and recent genetic data also point to direct contacts with Iberia [7, 16]. Despite recent approximate Bayesian computational based analyses of genetic data suggesting that the synchronous Neolithic expansion from the Middle East through both Mediterranean shores (i.e., Europe and North Africa) had a similar demographic pattern [17], our results point to a larger demographic replacement in North African than in European populations. The genomes of current Europeans carry a larger amount of hunter-gatherer components (up to 50% in Northern Europeans, according to [18]) compared to those of North Africans, where the Palaeolithic component, although present in extant populations, it is found at much lower frequencies (from 18.1% in Western Sahara to 5.2% in Egypt, according to the crude estimates of ADMIXTURE). This might suggest that the Neolithic demographic imprint was lower in Europe than in North Africa, where fewer local hunter-gatherers were assimilated by Neolithic farmers [13].

Complex demographic scenarios were tested with qpGraph [19] on a model involving different ancestral components: sub-Saharan (Yoruba), European (Basque), Middle Eastern (BedouinB), and North African Epipalaeolithic (Taforalt) (Figure 3). All North African groups fit into the tested model, with varying proportions of the four ancestral components, yielding compatible results with the admixture-f_3 tests (Figure 2A). Admixture signals in contemporary North African populations were dated using MALDER. All population triplets of the form (Ref A, Ref B, Target) were tested, taking the same populations used in the qpGraph analysis (i.e. Yoruba, Basque, BedouinB, and Taforalt) as references. Tunisian Berbers were chosen instead of Taforalt when the latter lacked power to give significant results due to

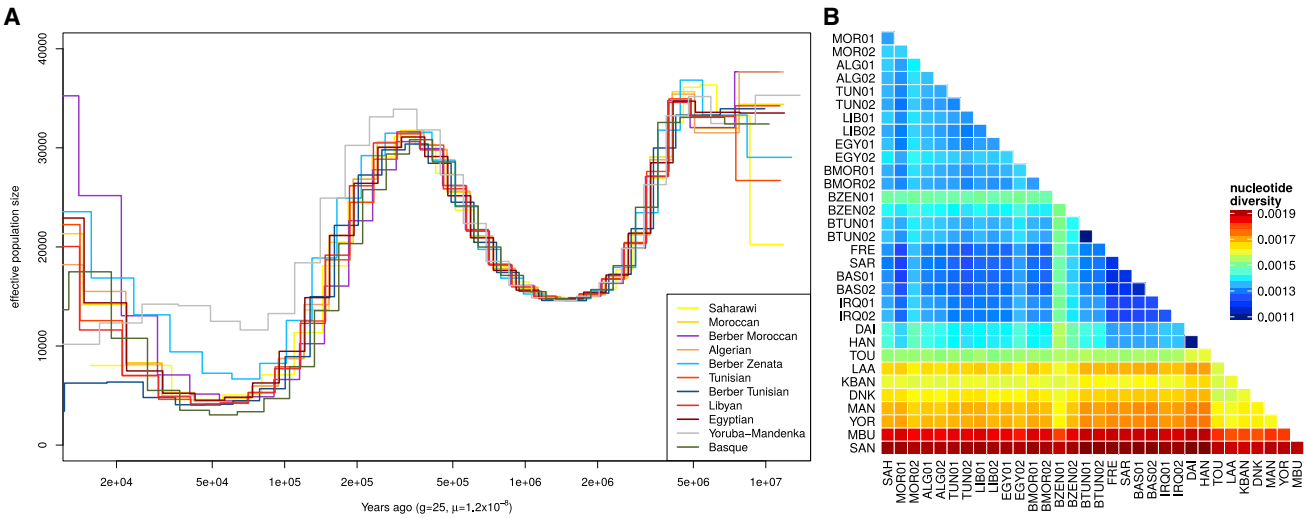

Figure 4. Effective Population Size throughout Time and Pairwise Genetic Differences

(A) Multiple sequentially Markovian coalescent (MSMC) model in North African samples. Sub-Saharan (Yoruba-Mandenka) and European (Basque) samples are included for comparison.

(B) Heatmap of pairwise nucleotide diversity per callable site.

their high missing data, since Tunisian Berbers have a high North African component, according to previous studies [1]. Interestingly, only tests involving the Yoruba population passed all the pre-test steps and yielded statistically significant results. Dates for these admixture times are shown in Table S2 for Saharawi and Egyptian, which account for both geographic and genetic diversity extremes in North Africa. These results (ranging from 1329 to 1643 AD) are compatible with previous estimates of the sub-Saharan introduction in North African populations in recent historical times ([6]).

Relatedness and Effective Population Sizes in North African Groups

Genetic drift detected in the admixture-f_3 analyses was further studied by computing the runs of homozygosity (ROH). Long runs of homozygosity are a signal of recent relatedness between individuals (i.e., interbreeding or consanguinity), which tend to occur in populations that are isolated or endogamous, while short runs of homozygosity show historically low effective population sizes. We compared the cumulative length of short ROHs (shorter than 1 Mb) and the length of long ROHs (longer than 1 Mb) (Figure 2B). North African samples show intermediate levels of length of short ROHs between sub-Saharan samples and non-African samples. Long ROHs, on the contrary, group North Africans with most sub-Saharan and Eurasian samples, with the exception of Tunisian Berber samples, which have significantly high values, pointing to high levels of interbreeding. Differences in the number and cumulative length of long and short ROHs were not statistically significant between Arabs and Berbers (Mann-Whitney U tests $p > 0.05$), pointing to no correlation between cultural groups and genetic heterogeneity. Genetic inbreeding might have been frequent and has been claimed in some North African populations as a result of isolation [20, 21]. Nonetheless, not all North African groups exhibit long ROH or signals of extensive drift, and these signals are not correlated by geography or culture, pointing to heterogeneity in the isolation patterns of North African populations [1, 20]. The simplistic idea of considering Berber-speaking groups as small, isolated, and inbred not only is misleading but also is not supported by the present data. Population dynamics in North Africa were further assessed using multiple sequentially Markovian coalescent (MSMC) analyses to estimate population changes in the effective population size (N_e) through time. North African individuals show similar effective population size patterns to Eurasians (Figure 4A), with a population decrease after the split with sub-Saharan Africans (~70 thousand years ago [kya] according to [22]), suggesting an N_e reduction after the out-of-Africa event. Almost all North African individuals show a very similar MSMC pattern, pointing to a relatively homogeneous population size evolution history (at least during the time period with MSMC credible interval values). The Zenata Berbers, however, show a minor N_e decay compared to the rest of North Africans and Eurasians analyzed. As pointed out by [22], an extensive admixture history may affect the coalescent history estimated by models that use the density of heterozygous sites, such as MSMC. This might explain why the decay in most North Africans is almost as sharp as the one observed in Basques (due to the high proportion of West Eurasian ancestry in North Africans), and it is moderate in the Zenata, which show a higher proportion of sub-Saharan ancestry (see for instance PCA, ADMIXTURE, f_3 analyses, and qpGraph estimates). To assess the influence of sub-Saharan admixture on MSMC results, we masked sub-Saharan tracts found in each North African sample and repeated MSMC analyses. Zenata individuals showed similar N_e decay as the rest of North Africans (data not shown). The nucleotide diversity analysis (π, as defined by [23]) (Figure 4B) shows a similar pattern, with higher values in the Zenata and reduced values in out-of-Africa populations and in North Africans with the highest proportions of non-African ancestry, which was affected by the out-of-Africa bottleneck.

Conclusions

The intricate genome landscape in North Africa is shaped by two factors. The first one is an amalgam of genetic components resulting of extensive gene flow coming from different geographical (sub-Saharan Africa, Europe, Middle East, Caucasus, and North Africa itself) and temporal sources (Palaeolithic migrations, Neolithization, Arabization, and recent migrations). The second factor is the result of internal admixture and genetic drift, which have produced an ample genetic heterogeneity within the region. These two factors of complexity should be considered in biomedical studies in which North African samples are included in order to avoid genetic biases and artifacts.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Samples and datasets
  - Sequencing, mapping, calling and annotation
  - SNP calling validation
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Principal component analysis
  - ADMIXTURE analysis
  - *f* -statistics
  - MALDER analysis
  - Runs of homozygosity
  - ChromoPainter and fineSTRUCTURE
  - Population size inferences
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cub.2019.09.050.

ACKNOWLEDGMENTS

We thank Lara R. Arauna and Alex Mas-Sandoval for helpful discussion and methodological support. We are grateful to all the volunteers who participated in the study. This work was supported by the Spanish Agencia Estatal de Investigación (MINEICO, AEI) and Fondo Europeo de Desarollo Regional (FEDER) grants CGL-2013-44351-P and CGL2016-75389-P, the "Unidad de Excelencia Maria de Maeztu" MDM2014-0370, and Agència de Gestió d'Ajuts Universitaris i de la Recerca (grant 2014SGR866). G.S.V. was supported by an FI (2017FI_B2 00010) scholarship awarded by the Generalitat de Catalunya.

AUTHOR CONTRIBUTIONS

D.C. and G.S.V. designed the study, analyses, and methods; G.S.V. performed most of the analyses; M.L.S. contributed to the analyses; K.F.Z., A.B., and P.Z. provided the samples and contributed to the discussion; D.C. and G.S.V. wrote the manuscript; all authors read and approved the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 11, 2018
Revised: August 2, 2019
Accepted: September 19, 2019
Published: October 31, 2019

REFERENCES

1. Arauna, L.R., Mendoza-Revilla, J., Mas-Sandoval, A., Izaabel, H., Bekada, A., Benhamamouch, S., Fadhlaoui-Zid, K., Zalloua, P., Hellenthal, G., and Comas, D. (2017). Recent historical migrations have shaped the gene pool of Arabs and Berbers in North Africa. Mol. Biol. Evol. *34*, 318–329.
2. Bekada, A., Fregel, R., Cabrera, V.M., Larruga, J.M., Pestano, J., Benhamamouch, S., and González, A.M. (2013). Introducing the Algerian mitochondrial DNA and Y-chromosome profiles into the North African landscape. PLoS ONE *8*, e56775.
3. Botigué, L.R., Henn, B.M., Gravel, S., Maples, B.K., Gignoux, C.R., Corona, E., Atzmon, G., Burns, E., Ostrer, H., Flores, C., et al. (2013). Gene flow from North Africa contributes to differential human genetic diversity in southern Europe. Proc. Natl. Acad. Sci. USA *110*, 11791–11796.
4. Elkamel, S., Boussetta, S., Khodjet-El-Khil, H., Benammar Elgaaied, A., and Cherni, L. (2018). Ancient and recent Middle Eastern maternal genetic contribution to North Africa as viewed by mtDNA diversity in Tunisian Arab populations. Am. J. Hum. Biol. *30*, e23100.
5. Fadhlaoui-Zid, K., Haber, M., Martínez-Cruz, B., Zalloua, P., Benammar Elgaaied, A., and Comas, D. (2013). Genome-wide and paternal diversity reveal a recent origin of human populations in North Africa. PLoS ONE *8*, e80293.
6. Henn, B.M., Botigué, L.R., Gravel, S., Wang, W., Brisbin, A., Byrnes, J.K., Fadhlaoui-Zid, K., Zalloua, P.A., Moreno-Estrada, A., Bertranpetit, J., et al. (2012). Genomic ancestry of North Africans supports back-to-Africa migrations. PLoS Genet. *8*, e1002397.
7. 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. (2018). 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.
8. 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. (2018). Pleistocene North African genomes link Near Eastern and sub-Saharan African human populations. Science *360*, 548–552.
9. Rodríguez-Varela, R., Günther, T., Krzewińska, M., Storå, J., Gillingwater, T.H., MacCallum, M., Arsuaga, J.L., Dobney, K., Valdiosera, C., Jakobsson, M., et al. (2017). Genomic analyses of pre-European conquest human remains from the Canary Islands reveal close affinity to modern North Africans. Curr. Biol. *27*, 3396–3402.e5.
10. Ilahiane, H. (2006). Historical Dictionary of the Berbers (Imazighen).
11. Harich, N., Costa, M.D., Fernandes, V., Kandil, M., Pereira, J.B., Silva, N.M., and Pereira, L. (2010). The trans-Saharan slave trade - clues from interpolation analyses and high-resolution characterization of mitochondrial DNA lineages. BMC Evol. Biol. *10*, 138.
12. Segal, R. (2001). Islam's Black Slaves: the Other Black Diaspora (Farrar, Straus and Giroux).
13. Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D.C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., et al. (2016). Genomic insights into the origin of farming in the ancient Near East. Nature *536*, 419–424.
14. Morales, J., Pérez-Jordà, G., Peña-Chocarro, L., Zapata, L., Ruíz-Alonso, M., López-Sáez, J.A., and Linstädter, J. (2013). The origins of agriculture in North- West Africa: Macro-botanical remains from Epipalaeolithic and Early Neolithic levels of Ifri Oudadane (Morocco). J. Archaeol. Sci. *40*, 2659–2669.
15. Mulazzani, S., Belhouchet, L., Salanova, L., Aouadi, N., Dridi, Y., Eddargach, W., Morales, J., Tombret, O., Zazzo, A., and Zoughlami, J. (2016). The emer- gence of the Neolithic in North Africa: a new model for the Eastern Maghreb. Quat. Int. *410*, 123–143.

16. Linstädter, J., and Kehl, M. (2012). The Holocene archaeological sequence and sedimentological processes at Ifri Oudadane, NE Morocco. J. Archaeol. Sci. *39*, 3306–3323.
17. Pimenta, J., Lopes, A.M., Comas, D., Amorim, A., and Arenas, M. (2017). Evaluating the neolithic expansion at both shores of the mediterranean sea. Mol. Biol. Evol. *34*, 3232–3242.
18. Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P.H., Schraiber, J.G., Castellano, S., Lipson, M., et al. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature *513*, 409–413.
19. Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. Genetics *192*, 1065–1093.
20. Bekada, A., Arauna, L.R., Deba, T., Calafell, F., Benhamamouch, S., and Comas, D. (2015). Genetic heterogeneity in Algerian human populations. PLoS ONE *10*, e0138453.
21. Fadhlaoui-Zid, K., Martinez-Cruz, B., Khodjet-el-khil, H., Mendizabal, I., Benammar-Elgaaied, A., and Comas, D. (2011). Genetic structure of Tunisian ethnic groups revealed by paternal lineages. Am. J. Phys. Anthropol. *146*, 271–280.
22. Haber, M., Mezzavilla, M., Bergström, A., Prado-Martinez, J., Hallast, P., Saif-Ali, R., Al-Habori, M., Dedoussis, G., Zeggini, E., Blue-Smith, J., et al. (2016). Chad genetic diversity reveals an African history marked by multiple Holocene Eurasian migrations. Am. J. Hum. Genet. *99*, 1316–1324.
23. Nei, M., and Li, W.H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA *76*, 5269–5273.
24. Lorente-Galdos, B., Lao, O., Serra-Vidal, G., Santpere, G., Kuderna, L.F.K., Arauna, L.R., Fadhlaoui-Zid, K., Pimenoff, V.N., Soodyall, H., Zalloua, P., et al. (2019). Whole-genome sequence analysis of a Pan African set of samples reveals archaic gene flow from an extinct basal population of modern humans into sub-Saharan populations. Genome Biol. *20*, 77.
25. Reich, D., Green, R.E., Kircher, M., Krause, J., Patterson, N., Durand, E.Y., Viola, B., Briggs, A.W., Stenzel, U., Johnson, P.L., et al. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature *468*, 1053–1060.
26. Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A., et al. (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature *538*, 201–206.
27. Lachance, J., Vernot, B., Elbers, C.C., Ferwerda, B., Froment, A., Bodo, J.M., Lema, G., Fu, W., Nyambo, T.B., Rebbeck, T.R., et al. (2012). Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse African hunter-gatherers. Cell *150*, 457–469.
28. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.
29. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.
30. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.
31. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.
32. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al.; 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158.
33. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience *4*, 7.
34. Korn, J.M., Kuruvilla, F.G., McCarroll, S.A., Wysoker, A., Nemesh, J., Cawley, S., Hubbell, E., Veitch, J., Collins, P.J., Darvishi, K., et al. (2008). Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. Nat. Genet. *40*, 1253–1260.
35. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.
36. Cingolani, P., Platts, A., Wang, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) *6*, 80–92.
37. Schubert, M., Lindgreen, S., and Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res. Notes *9*, 88.
38. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L., and Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics *29*, 1682–1684.
39. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664.
40. Behr, A.A., Liu, K.Z., Liu-Fang, G., Nakka, P., and Ramachandran, S. (2016). pong: fast analysis and visualization of latent clusters in population genetic data. Bioinformatics *32*, 2817–2823.
41. Loh, P.R., Lipson, M., Patterson, N., Moorjani, P., Pickrell, J.K., Reich, D., and Berger, B. (2013). Inferring admixture histories of human populations using linkage disequilibrium. Genetics *193*, 1233–1254.
42. Lawson, D.J., Hellenthal, G., Myers, S., and Falush, D. (2012). Inference of population structure using dense haplotype data. PLoS Genet. *8*, e1002453.
43. Leslie, S., Winney, B., Hellenthal, G., Davison, D., Boumertit, A., Day, T., Hutnik, K., Royrvik, E.C., Cunliffe, B., Lawson, D.J., et al.; Wellcome Trust Case Control Consortium 2; International Multiple Sclerosis Genetics Consortium (2015). The fine-scale genetic structure of the British population. Nature *519*, 309–314.
44. Schiffels, S., and Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. Nat. Genet. *46*, 919–925.
45. Maples, B.K., Gravel, S., Kenny, E.E., and Bustamante, C.D. (2013). RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. Am. J. Hum. Genet. *93*, 278–288.
46. O'Connell, J., Gurdasani, D., Delaneau, O., Pirastu, N., Ulivi, S., Cocca, M., Traglia, M., Huang, J., Huffman, J.E., Rudan, I., et al. (2014). A general approach for haplotype phasing across the full spectrum of relatedness. PLoS Genet. *10*, e1004234.
47. Campbell, C.D., Chong, J.X., Malig, M., Ko, A., Dumont, B.L., Han, L., Vives, L., O'Roak, B.J., Sudmant, P.H., Shendure, J., et al. (2012). Estimating the human mutation rate using autozygosity in a founder population. Nat. Genet. *44*, 1277–1281.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
21 human whole genome sequences	This study	Table S1
3 human whole-genome sequences	[24]	Table S1
9 human whole-genome sequences	[25]	Table S1
119 human whole-genome sequences from Simons Genome Diversity Project	[26]	http://simonsfoundation.s3.amazonaws.com/share/SCDA/datasets/10_24_2014_SGDP_metainformation_update.txt
15 human whole-genome sequences from Pygmy groups	[27]	Hadza (5), Sandawe (5), Baka (3), Bakola (1), and Bedzan (1).
5 ancient human genome sequences from Taforalt	[8]	TAF010, TAF011, TAF012, TAF013, TAF014
4 ancient human genome sequences from Ifri n'Amr or Moussa (IAM)	[7]	IAM3, IAM4, IAM6, IAM7
4 ancient human genome sequences from Kelif el Boroud (KEB)	[7]	KEB1, KEB4, KEB6, KEB8
5 ancient human genome sequences from Guanche population	[9]	gun002, gun005, gun008, gun011, gun012
963 current samples and 281 ancient samples genotyped with Affymetrix Human Origins Array from Human Origins dataset	[13]	https://reich.hms.harvard.edu/sites/reich.hms.harvard.edu/files/inline-files/NearEastPublic.tar.gz
Deposited Data		
Raw data	https://www.ebi.ac.uk/ena/data/view/PRJEB29142	ENA: PRJEB29142
Software and Algorithms		
fastqc	Babraham Bioinformatics	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
Burrows-Wheeler aligner (bwa) v0.7.7	[28]	http://bio-bwa.sourceforge.net/
Picard tools v2.8.3	Broad Institute	https://broadinstitute.github.io/picard/
Genome Analysis Toolkit (GATK) v4.0.12	[29]	https://software.broadinstitute.org/gatk/
Samtools v1.9	[30]	http://www.htslib.org/
BEDTools v2.21	[31]	https://bedtools.readthedocs.io/en/latest/
VCFtools v0.1.14	[32]	https://vcftools.github.io/index.html
PLINK v1.90	[33–35]	https://www.cog-genomics.org/plink/1.9/
tabix v0.2.6	[30]	http://www.htslib.org/doc/tabix.html
snpEff v4.3	[36]	http://snpeff.sourceforge.net/
AdapterRemoval	[37]	https://github.com/MikkelSchubert/adapterremoval
pileupCaller	[9]	https://github.com/stschiff/sequenceTools
mapDamage	[38]	https://ginolhac.github.io/mapDamage/
EIGENSOFT v6.0.1	[19]	https://github.com/DReichLab/EIG
ADMIXTURE v1.3	[39]	http://software.genetics.ucla.edu/admixture/
pong v1.4.7	[40]	https://github.com/ramachandran-lab/pong
AdmixTools v4.1	[19]	https://github.com/DReichLab/AdmixTools
MALDER v1	[41]	https://github.com/joepickrell/malder/tree/master/MALDER
ChromoPainter v2 and fineSTRUCTURE v2.1.0	[42, 43]	http://www.paintmychromosomes.com/
MSMC2	[44]	https://github.com/stschiff/msmc2
RFMix	[45]	https://github.com/slowkoni/rfmix

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, David Comas (david.comas@upf.edu). This study did not generate new unique reagents.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All biological samples were collected with the appropriate informed consent and the approval of the IRB CEIC- Hospital del Mar 2013/5429/I. All the sampled individuals were male adults.

METHOD DETAILS

Samples and datasets

Twenty one complete genomes from North African individuals (n = 17), Basque (n = 2), and Iraqi (n = 2) were sequenced for the present study. Samples from North Africa belong to Arab- and Berber-speaking groups from Western Sahara (n = 1), Morocco (n = 4, including two Berbers from the Rif), Algeria (n = 4, including two Zenata Berbers), Tunisia (n = 4, including two Berbers from Chenini), Libya (n = 2), and Egypt (n = 2). Besides North African samples, eight Eurasian samples (Basques (n = 2), Iraqi (n = 2), French, Sardinian, Han, and Dai) and eight sub-Saharan samples (Toubou, Laal, Dinka, Mandenka, Yoruba, Kenya Bantu, Mbuti, and San) were included to the dataset. One sample per group was used unless otherwise stated. The Toubou, Laal, and Kenya Bantu samples were published by [24] (as Toubou- A408, Laal-A409, EastAfricanBantu-A401). All remaining non-North-African samples (French, Sardinian, Han, Dai, Dinka, Mandenka, Yoruba, Mbuti, San) were published by [25] (HGDP00521, HGDP00665, HGDP00778, HGDP01307, DNK02, HGDP01284, HGDP00927, HGDP00456, HGDP01029, respectively).

For some analyses, samples coming from five extra datasets were included:

1. Simons Genome Diversity Project (SGDP) [26]: 119 samples coming from 57 populations from sub-Saharan Africa (40), North Africa (4, two Mozabite Berbers and two Saharawi), Europe (41), the Middle East (15) and the Caucasus (19). Variant-only individual VCFs were obtained from https://sharehost.hms.harvard.edu/genetics/reich_lab/sgdp/vcf_variants/vcfs.variants.public_samples.279samples.tar. After excluding sites out of the defined callable fraction of the genome (see below), all samples were merged into a single VCF file, and then merged with the original dataset. The final dataset was filtered by minor allele frequency (5%) to avoid batch effect.
2. Hunter-gatherer complete genomes [27]: 15 genomes from five sub-Saharan hunter-gatherer groups: Hadza (5), Sandawe (5), Baka (3), Bakola (1), and Bedzan (1). Complete Genomics format was converted to VCF and posteriorly merged with the previous dataset.
3. Five Epipalaeolithic 15,000 year-old genomes from the Taforalt site in Morocco [8]: TAF010, TAF011, TAF012, TAF013 TAF014. Variants were called from BAM files containing the mapped reads with minimum mapping and base quality of 37 using SAMtools mpileup and pileupCaller as described in [9].
4. Five ancient Guanche genomes from Tenerife and Gran Canaria (Canary Islands) [9] from between the 7^{th} and 11^{th} centuries CE; gun002, gun005, gun008, gun011, gun012. Reads were processed using the same pipeline as in the previous dataset.
5. Eight North African Neolithic samples: four samples Early Neolithic from the site of Ifri n'Amr or Moussa (~7,000 ya) (IAM3, IAM4, IAM6, IAM7) and four Late Neolithic samples from the site of Kelif el Boroud (KEB) (~5, 000 ya) (KEB1, KEB4, KEB6, KEB8) [7]. The retrieved file format was FASTQ (raw non-mapped reads), which reads were processed with AdapterRemoval [37] (options –trimns –trimqualities –minlength 30 –mm 3 –minquality 30) and later aligned to the hg19 reference genome using the Burrows-Wheeler Aligner (BWA) [28] with options aln -l 1024 -n 0.01 -o 2 before variant calling. Duplicates were removed using SAMtools [30], and post-mortem DNA damage was assessed with mapDamage [38].
6. Human Origins dataset (Affymetrix Human Origins Array) [13]. A subset containing 281 ancient samples from 29 populations (European Palaeolithic and Neolithic, Middle Eastern Natufian and Neolithic, Anatolia, Caucasus Palaeolithic, Iranian Neolithic and Eurasian Steppe) and 963 current samples from 69 populations (North Africa, Europe, Middle East, and Caucasus) was chosen. Duplicated samples in the Human Origins dataset and the Simons Genome Diversity Project dataset were excluded for the analyses.

Two different final datasets were built; one of them including all datasets previously described, and another one excluding the Human Origins array data in order to avoid the reduction of the number of SNP due to the merging process with array data. The two datasets contain 1,359 and 177 samples, and 346, 211 and 4, 015, 113 SNPs, respectively.

Sequencing, mapping, calling and annotation

The sequencing process of the initial dataset was performed using Illumina HiSeq 2000 sequencing platform, producing paired-end reads. Read mapping and variant calling were performed according to the guidelines from [24]. Read quality was assessed using fastqc (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and they were subsequently mapped to the hg19 reference genome using the Burrows-Wheeler Aligner (BWA) [28].

The mapped reads were merged and PCR duplicates were removed using MarkDuplicates from Picard (https://broadinstitute.github.io/picard/), indels were realigned and base quality scores were recalibrated using Genome Analysis Toolkit (GATK) [29]. SNPs were called and their quality scores were recalibrated using the Unified Genotyper and the VQSR tools from GATK, respectively.

In order to get a confident set of variants, we defined a callable fraction of the genome taking into account the following criteria: *(i)* a minimum of 5 reads mapped for each locus; *(ii)* quality score threshold for the assertion of the alternative allele of the variant (minimum score of 20 in the QUAL field of the VCF file); *(iii)* genotyping quality threshold ('PASS' in the 'FILTER' field of the VCF file); *(iv)* exclusion of regions covered by structural variants (as in [24]): TandemRepeatMarker repeats of length > 80 bp (UCSC browser), RepeatMasker repeats of length > 80 bp and < 90% identity (UCSC browser), hg19 segmental duplications (UCSC browser), 1000 Genomes Project copy number variants (https://www.ncbi.nlm.nih.gov/dbvar/studies/, study variants_for_estd199.csv); *(v)* exclusion of regions adjacent to indels (6 bp flanking regions); *(vi)* exclusion of multiallelic variants.

A 2,059,756,821 bp-long 'callable genome' was obtained, containing a total high- confidence set of 10,855,685 SNPs.

The coverage of the mapped genomes was computed using the genomeCoverageBed tool from BEDTools v2.21 [31]. The length of the reads was obtained using SAMtools.

Several statistics were computed from the final set of variants in order to assess SNP calling quality and are shown in Table S1: the number of variants for each individual and the proportion of heterozygous and homozygous. Private variants (singletons and private doubletons) were obtained using VCFtools [32] -singletons switch.

Functional annotation was performed using snpEff [36]. In particular, the following parameters were computed for each sample: number of known variants (i.e., previously described in dbSNP v138), synonymous and non-synonymous SNPs, missense, nonsense and silent mutations, intron, exon and intergenic mutations and Ti/Tv ratio. All obtained values fell within the expected range of values.

SNP calling validation

In order to further validate the called variants, a comparison between the sequencing data and an independent genotype analysis performed in seven individuals with Affymetrix's Genome-Wide Human SNP array 6.0 data from [1] was performed. The North African validated samples are BTUN01, BTUN02, BZEN01, BZEN02, LIB02, MOR02 and SAH.

The SNP array datasets, which were filtered using population and frequencies criteria (missing data, minor allele frequency, and LD pruning), were converted to VCF using PLINK v1.90 beta [33–35], sorted with VCFtools and zipped and indexed with bgzip and tabix [30].

For each individual sample, only the genomic positions that passed GATK's filter, found to have been assayed in both the array and sequencing SNP calling, and that had a genotype assigned to them, were further considered for comparison. Data from sequencing and array was compared with both BEDTools and own scripts. All identity proportions were found to be high (> 99.30% in heterozygous calls, > 99.86% in homozygous calls, and > 99.74% globally).

QUANTIFICATION AND STATISTICAL ANALYSIS

Principal component analysis

Principal component analysis (PCA) was carried out using the smartpca tool from the EIGENSOFT software package (v6.0.1) [19]. The data were pruned for linkage disequilibrium between markers using PLINK v1.90 [33–35] and the parameters–indep-pairwise 200 25 0.4. Principal components were computed using current samples and the ancient ones were projected on top of them (using the poplistname and lsqproject options).

ADMIXTURE analysis

ADMIXTURE v1.3 [39] was applied on the whole dataset, which was previously pruned for linkage disequilibrium between markers using PLINK v1.9, after assigning an r^2 threshold of 0.4 in every continuous window of 200 SNPs with a step of 25 SNPs (i.e., –indep-pairwise 200 25 0.4). ADMIXTURE in unsupervised mode was run assuming a number of ancestral clusters ranging from K = 2 to K = 12 with 10 independent runs for each K using different randomly generated seeds for each run.

The cross-validation error was assessed for each run, with K = 3 to K = 6 giving the minimum error. pong [40] in greedy mode was run in order to identify common modes among different runs for each K to align clusters across different values of K.

To further assess and validate the nature of the ancestral clusters found with ADMIXTURE in the North African samples, we performed a supervised analysis (– supervised) for K = 6 taking as reference populations the proxies for the major com- ponents detected in North Africans, i.e., Yoruba/Esan/Mandenka/Mende; Natufian; Europe_EN/Anatolia_N; CHG/Iran_N; Taforalt; and WHG/SHG, since 6 it is the lowest K value showing the five components that repeatedly appear in North Africans through different values of K. The result of the supervised analysis is virtually identical to the non-supervised analysis, implying that the defined clusters do correspond to the previously identified ancestral components.

f -statistics

Admixture-f_3, outgroup-f_3, and F_4-ratio statistics were computed using the qp3Pop and qpF4ratio tools from AdmixTools v4.1 [19]. Standard errors were assessed using weighted block jackknife with block size of 5 Mb.

MALDER analysis

MALDER v1 [41], a modification of ALDER v1.3, was used with the default parameters and the "multiple admixture tests" mode in order to date admixture events.

Runs of homozygosity

Runs of homozygosity (ROH) were computed with PLINK using the whole genome sequences dataset. Runs were identified with PLINK using sliding windows of 50 SNPs across the genome and were defined as chunks containing at least 50 SNPs and 100 kbp, with no more than 2000 kbp between consecutive SNPs and with no more than one heterozygous site and 5 missing sites.

Differences between Arabs and Berbers were computed with a Mann-Whitney U test using several ROH counts: total number of ROH, cumulative length of ROH, average length of ROH, number of short ROH, cumulative length of short ROH, number of long ROH, and cumulative length of long ROH.

ChromoPainter and fineSTRUCTURE

ChromoPainter v2 and fineSTRUCTURE v2.1.0 [42, 43] were used in order to study haplotype sharing between pairs of samples.

The dataset was phased using SHAPEIT v2.12 [46] against the phased reference panel of the 1000 Genomes Project Phase 3 (build 37) downloaded from https://mathgen.stats.ox.ac.uk/impute/1000GP_Phase3.tgz and using the genetic map from the 1000 Genomes Project Phase 3 as well.

Chromopainter was run in two runs for computational efficiency reasons. The first run estimated global mutation probability and average switch rate parameter using an expectation-maximization algorithm across chromosomes 1, 4, 17 and 20. These two parameters were then fixed in the second run, in which ChromoPainter inferred the coancestry matrix, composed by the number of shared chunks (chunk counts) and their length (chunk lengths) between all pairs of individuals. Chromo- Combine was used to combine all chromosomes and individuals into a single final matrix.

This coancestry matrix was used by fineSTRUCTURE to classify the individuals into genetically homogeneous clusters using Markov Chain Monte Carlo (MCMC) with 1,000,000 burn-in iterations, 2,000,000 sample iterations, from which 1 in every 10,000 was recorded, and 100,000 additional hill-climbing moves to reach the final inferred tree.

Population size inferences

MSMC2 [44] was applied to the phased whole genome sequences dataset, four haplotypes at a time (two haplotypes per individual and two individuals per population) to estimate the evolution of effective population size using 32 time segments (parameters "-p 1*2+25*1+1*2+1*3"), a mutation rate of $1.2\ 10^{-8}$ [47] and an average generation time of 35 years. MSMC analysis provided estimates for cross- population and within-population coalescence rates. The relative cross-coalescence rate was computed by dividing the cross-population coalescence rate by the average of within-population coalescence rates. In order to assess the effect of admixture on the MSMC results, MSMC was performed after sub-Saharan tracts were masked from each North African sequence. Tracts were inferred using RFMix [45], considering 1,007 samples from sub-Saharan Africa (YRI, LWK, GWD, ESN, MSL) and European (CEU, GBR, IBS, TSI, FIN) populations from the 1000 Genomes Project Phase 3 panel as references.

DATA AND CODE AVAILABILITY

The accession number for the newly generated whole-genome data reported in this paper is ENA: PRJEB29142 (https://www.ebi.ac.uk/ena/data/view/PRJEB29142).

Current Biology, Volume *29*

Supplemental Information

Heterogeneity in Palaeolithic Population Continuity and Neolithic Expansion in North Africa

Gerard Serra-Vidal, Marcel Lucas-Sanchez, Karima Fadhlaoui-Zid, Asmahan Bekada, Pierre Zalloua, and David Comas

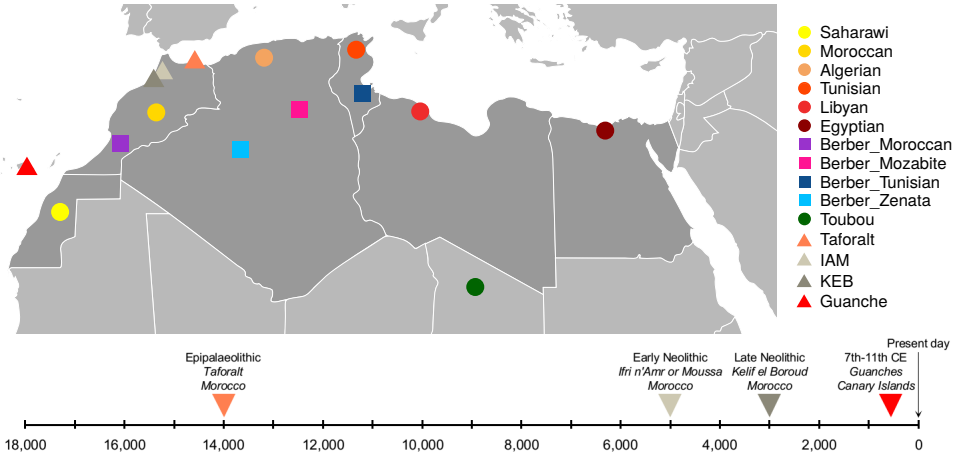

Figure S1: North African samples used for this study. Related to STAR Methods.

North African samples used for this study. Berber groups (■): Moroccan Berbers (BMOR), Algerian Mozabite Berbers (BMOZ), Algerian Zenata Berbers (BZEN) and Tunisian Berbers (BTUN). Arab groups (●): Western Sahara (SAH), Morocco (MOR), Algeria (ALG), Tunisia (TUN), Libya (LIB) and Egypt (EGY). Ancient groups (▲): Guanche (Canary Islands, 7^{th}-11^{th} centuries CE), KEB (Morocco, 5,000 ya), IAM (Morocco, 7,000 ya), and Taforalt (Morocco, 15,000 ya).

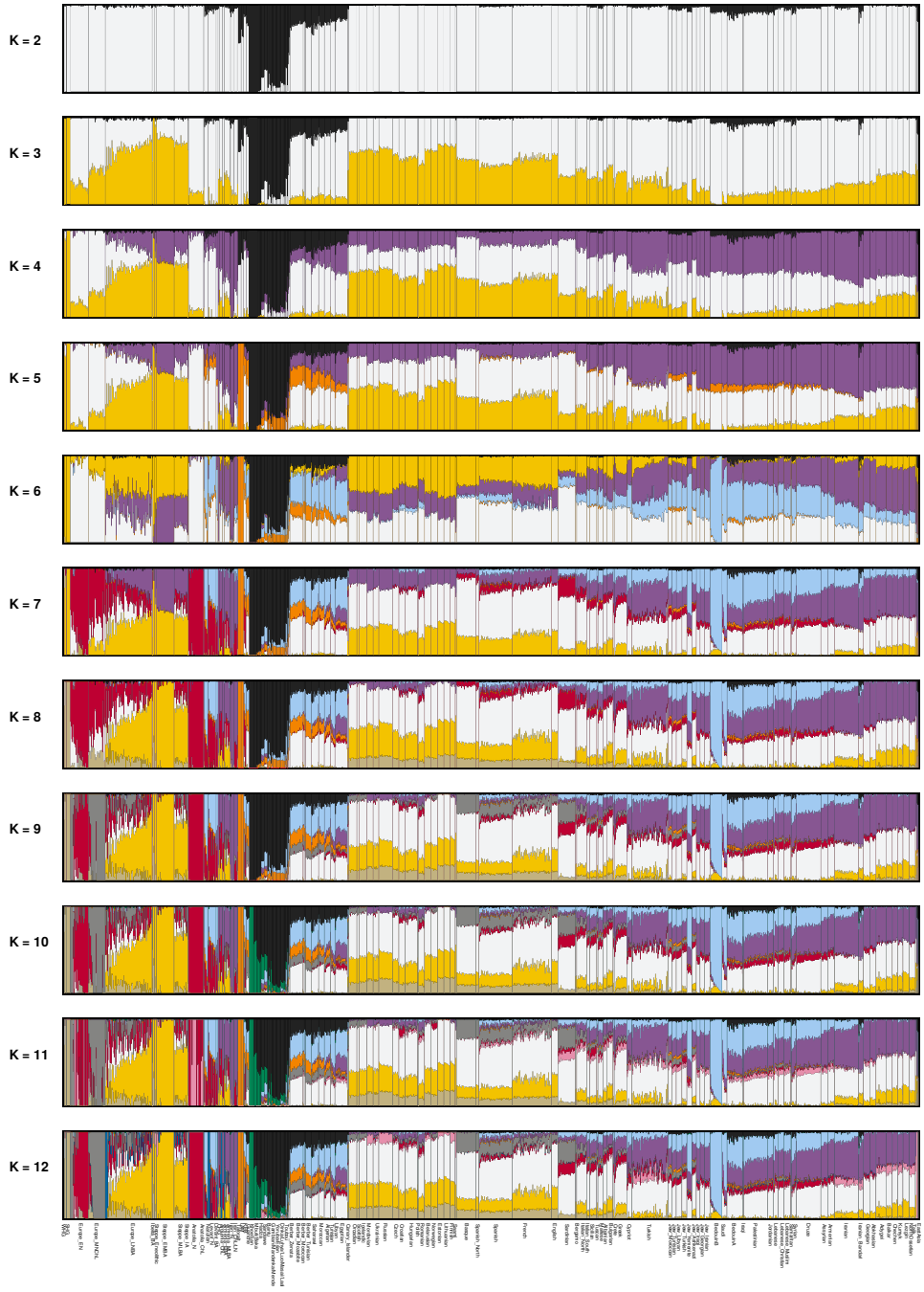

Figure S2: ADMIXTURE analysis of North African samples and worldwide ancient and current populations. Related to Figure 1.

Cluster-based analysis ($K = 2$ to $K = 12$) using ADMIXTURE of North African samples and worldwide ancient and current populations from Africa, Europe, Middle East and Caucasus.

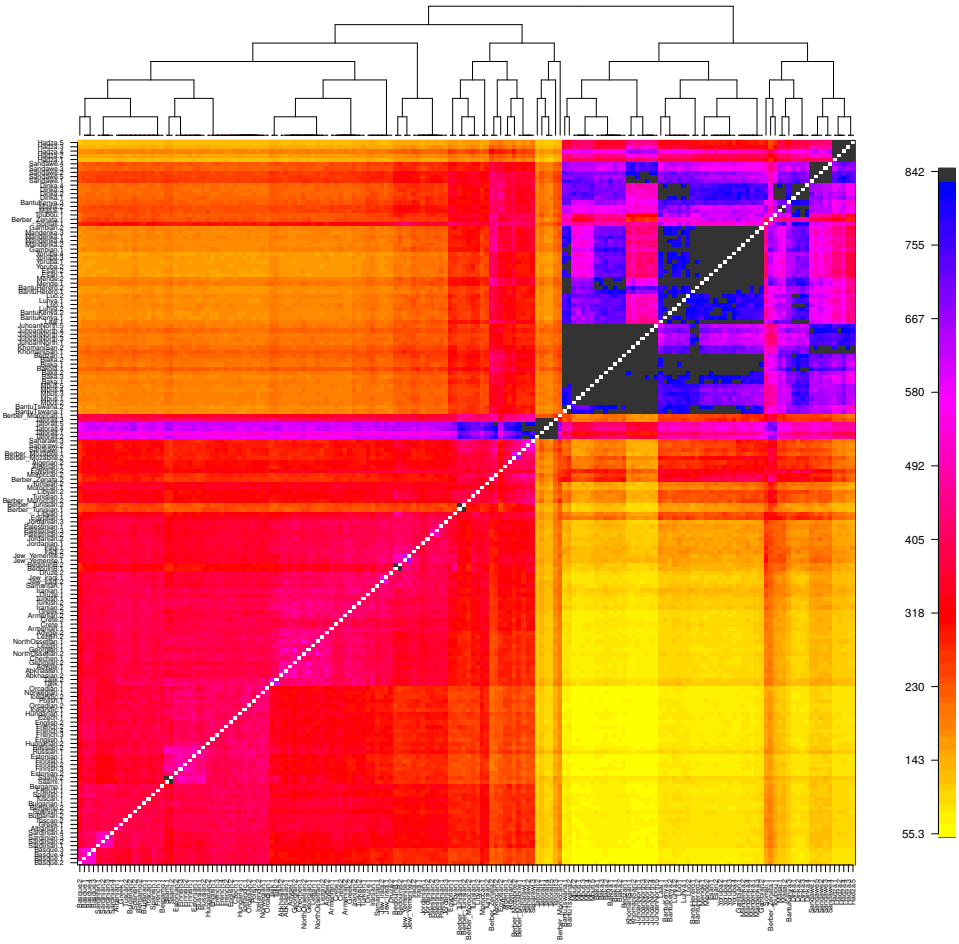

Figure S3: ChromoPainter chunk counts coancestry matrix. Related to STAR Methods.

ChromoPainter chunk counts coancestry matrix. Donor samples are displayed in the left column and recepients are at the bottom. The dendogram of the ChromoPainter chunk counts coancestry matrix inferred by fineSTRUCTURE is show at the top.

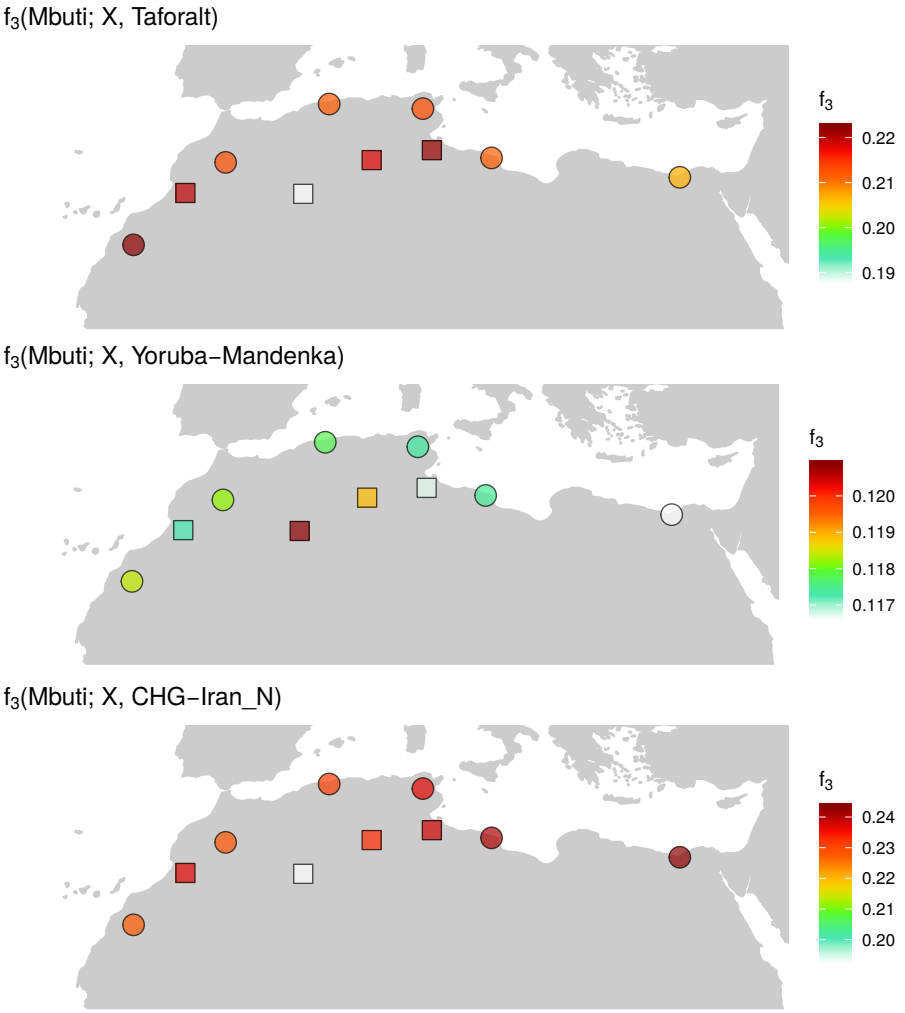

Figure S4: Outgroup-f_3 results for North African groups. Related to Figure 2.

(A) Results for f_3(X, Taforalt; Ju/'hoanNorth), **(B)** f_3(X, Yoruba-Mandenka; Ju/'hoanNorth), and **(C)** f_3(X, CHG-Iran_N; Ju/'hoanNorth)

Sample	Population	Region	Mean Coverage (X)	Read length	SNPs	Unique SNPs	Singletons	Private Doubletons	Ref. homozygous	Heterozygous	Alt. Homozygous	Heterozygous ratio	Known SNPs	Known SNPs ratio
DAI	Dai	East Asia	23.87	100,101,94,95	2,411,885	107,761	102,741	5,020	8,443,800	1,366,964	1,044,921	0.5668	2,348,444	0.9737
HAN	Han	East Asia	25.56	100,101,94,95	2,423,071	108,622	103,621	5,001	8,432,614	1,364,465	1,058,606	0.5631	2,363,550	0.9754
BAS01	Basque	Europe	26.23	101	2,394,574	58,375	56,675	1,700	8,461,111	1,423,402	971,172	0.5944	2,350,015	0.9814
BAS02	Basque	Europe	25.48	101	2,394,909	55,684	53,547	2,137	8,460,776	1,431,911	962,996	0.5979	2,349,854	0.9812
FRE	French	Europe	24.47	100,101,94,95	2,404,916	62,184	60,771	1,413	8,450,769	1,439,695	965,221	0.5986	2,357,844	0.9804
SAR	Sardinian	Europe	22.58	100,101,94,95	2,402,878	58,212	56,830	1,382	8,452,807	1,434,186	968,692	0.5969	2,354,076	0.9797
IRQ01	Iraqi	Middle East	26.37	101	2,473,674	74,454	72,256	2,198	8,382,011	1,519,753	953,921	0.6144	2,402,735	0.9713
IRQ02	Iraqi	Middle East	26.66	101	2,482,819	77,271	75,651	1,620	8,372,866	1,525,795	957,024	0.6145	2,413,004	0.9719
SAH	Western Saharawi	North Africa	20.03	100	2,529,332	74,141	70,902	3,239	8,326,353	1,549,056	980,276	0.6124	2,452,474	0.9696
MOR01	Moroccan	North Africa	28.24	101	2,500,538	73,170	71,463	1,707	8,355,147	1,554,181	946,357	0.6215	2,434,480	0.9736
MOR02	Moroccan	North Africa	27.47	101	2,634,439	95,653	93,087	2,566	8,221,246	1,700,631	933,808	0.6455	2,547,897	0.9671
ALG01	Algerian	North Africa	27.05	101	2,573,951	85,140	83,169	1,971	8,281,734	1,631,028	942,923	0.6337	2,496,968	0.9701
ALG02	Algerian	North Africa	26.52	101	2,579,482	85,582	84,023	1,559	8,276,203	1,652,080	927,402	0.6405	2,502,214	0.9700
TUN01	Tunisian	North Africa	26.66	101	2,543,016	76,389	74,677	1,712	8,312,669	1,597,082	945,934	0.6280	2,467,994	0.9705
TUN02	Tunisian	North Africa	26.30	101	2,537,091	72,988	72,341	647	8,318,594	1,598,154	938,937	0.6299	2,466,893	0.9723
LIB01	Libyan	North Africa	26.02	101	2,521,242	79,089	76,110	2,979	8,334,443	1,546,163	975,079	0.6133	2,445,293	0.9699
LIB02	Libyan	North Africa	20.99	100	2,543,844	76,764	75,106	1,658	8,311,841	1,612,654	931,190	0.6339	2,468,282	0.9703
EGY01	Egyptian	North Africa	26.62	101	2,558,156	87,301	85,643	1,658	8,297,529	1,622,030	936,126	0.6341	2,477,162	0.9683
EGY02	Egyptian	North Africa	26.05	101	2,632,906	105,006	103,656	1,350	8,222,779	1,707,079	925,827	0.6484	2,539,558	0.9645
BMOR01	Moroccan Berber	North Africa	23.11	101	2,474,791	50,435	46,609	3,826	8,380,894	1,479,580	995,205	0.5979	2,415,681	0.9761
BMOR02	Moroccan Berber	North Africa	26.38	101	2,558,247	74,053	73,435	618	8,297,438	1,618,507	939,740	0.6327	2,485,679	0.9716
BZEN02	Zenata Berber	North Africa	26.75	101	2,597,358	98,549	93,278	5,271	8,258,327	1,608,931	988,426	0.6194	2,512,842	0.9675
BZEN01	Zenata Berber	North Africa	26.58	101	2,782,951	141,256	137,454	3,802	8,072,734	1,853,390	929,561	0.6660	2,680,433	0.9632
BTUN01	Tunisian Berber	North Africa	27.22	101	2,419,114	50,274	46,270	4,004	8,436,571	1,349,271	1,069,843	0.5578	2,351,058	0.9719
BTUN02	Tunisian Berber	North Africa	26.40	101	2,463,450	52,163	49,495	2,668	8,392,235	1,427,940	1,035,510	0.5797	2,393,851	0.9717
TOU	Toubou	Sub-Saharan Africa	19.50	101	2,756,120	139,659	133,454	6,205	8,099,565	1,756,456	999,664	0.6373	2,638,003	0.9571
LAA	Laal	Sub-Saharan Africa	16.23	101	2,913,363	202,714	195,685	7,029	7,942,322	1,897,329	1,016,034	0.6513	2,770,671	0.9510
DNK	Dinka	Sub-Saharan Africa	22.89	100,101,94,95	2,879,818	197,905	191,149	6,756	7,975,867	1,855,877	1,023,941	0.6444	2,714,725	0.9427
KBAN	Kenya Bantu	Sub-Saharan Africa	16.72	100	2,891,879	205,139	198,694	6,445	7,963,806	1,915,632	976,247	0.6624	2,736,631	0.9463
MAN	Mandenka	Sub-Saharan Africa	22.53	100,101,94,95	2,931,978	226,853	217,933	8,920	7,923,707	1,912,826	1,019,152	0.6524	2,791,760	0.9522
YOR	Yoruba	Sub-Saharan Africa	29.33	100,101,94,95	2,914,217	228,745	218,230	10,515	7,941,468	1,881,916	1,032,301	0.6458	2,796,174	0.9595
MBU	Mbuti Pygmy	Sub-Saharan Africa	22.13	100,101,94,95	3,071,156	412,306	375,438	36,868	7,784,529	1,888,456	1,182,700	0.6149	2,721,851	0.8863
SAN	Khoisan	Sub-Saharan Africa	29.38	100,101,94,95	3,155,830	562,522	513,936	48,586	7,699,855	1,954,991	1,200,839	0.6195	2,747,483	0.8706

Sample	Synoymous	Non-synonymous	Non-synonymous - synoymous ratio	Missense	Nonsense	Silent	Missense-Silent ratio	Intron	Intergenic	Exon	Ti/Tv
DAI	19,370	15,779	0.8146	15,838	96	19,886	0.7964	5,181,669	1,142,868	95,334	2.0870
HAN	20,050	15,989	0.7975	16,053	107	20,577	0.7801	5,218,342	1,145,468	96,998	2.0873
BAS01	19,340	15,838	0.8189	15,909	96	19,915	0.7988	5,138,773	1,131,460	95,617	2.0939
BAS02	19,209	15,671	0.8158	15,731	87	19,776	0.7955	5,162,132	1,133,067	94,335	2.0969
FRE	19,486	16,193	0.8310	16,250	99	20,026	0.8114	5,206,961	1,131,848	96,384	2.0958
SAR	19,033	15,999	0.8406	16,063	113	19,539	0.8221	5,209,040	1,131,292	95,219	2.1006
IRQ01	20,017	16,471	0.8229	16,550	93	20,612	0.8029	5,370,818	1,166,751	99,027	2.0963
IRQ02	19,760	16,631	0.8416	16,694	109	20,323	0.8214	5,341,036	1,176,031	99,538	2.0952
ALG01	20,993	16,659	0.7936	16,725	106	21,521	0.7771	5,561,913	1,217,002	102,826	2.0967
ALG02	20,509	17,003	0.8291	17,069	87	21,087	0.8095	5,587,792	1,213,869	102,859	2.0969
EGY01	20,259	16,936	0.8360	17,006	103	20,818	0.8169	5,520,109	1,207,653	100,782	2.1003
EGY02	21,126	17,225	0.8153	17,315	127	21,744	0.7963	5,704,192	1,241,615	104,913	2.0947
LIB01	20,244	16,686	0.8242	16,751	93	20,858	0.8031	5,458,554	1,190,045	100,522	2.0972
LIB02	21,150	16,972	0.8025	17,042	111	21,696	0.7855	5,505,842	1,197,321	103,459	2.0965
MOR01	20,223	16,330	0.8075	16,392	99	20,818	0.7874	5,446,802	1,175,225	99,640	2.0953
MOR02	21,041	17,434	0.8286	17,516	114	21,621	0.8101	5,700,833	1,243,515	105,433	2.0945
BMOR01	19,566	16,063	0.8210	16,122	116	20,138	0.8006	5,409,242	1,161,865	97,982	2.1130
BMOR02	20,878	16,553	0.7928	16,623	103	21,475	0.7741	5,509,923	1,207,241	101,414	2.0986
TUN01	20,251	16,774	0.8283	16,840	98	20,830	0.8084	5,491,598	1,200,098	100,968	2.0945
TUN02	20,042	16,670	0.8318	16,738	108	20,599	0.8126	5,439,064	1,203,051	99,939	2.0943
BTUN01	19,263	16,115	0.8366	16,178	86	19,839	0.8155	5,215,156	1,142,775	96,475	2.0959
BTUN02	19,595	16,397	0.8368	16,472	113	20,135	0.8181	5,340,170	1,158,602	98,130	2.0939
SAH	20,592	16,736	0.8127	16,816	101	21,141	0.7954	5,493,726	1,189,767	102,782	2.0968
BZEN02	21,185	17,083	0.8064	17,156	86	21,894	0.7836	5,661,078	1,220,893	105,051	2.0958
BZEN01	22,518	17,733	0.7875	17,813	119	23,120	0.7705	6,014,726	1,313,923	110,281	2.0915
DNK	23,357	18,928	0.8104	19,010	113	24,087	0.7892	6,234,587	1,361,009	114,945	2.0895
KBAN	23,073	18,929	0.8204	18,992	104	23,801	0.7980	6,284,168	1,362,441	114,961	2.0903
SAN	25,811	20,069	0.7775	20,148	136	26,651	0.7560	6,891,268	1,480,134	125,490	2.0833
LAA	22,988	18,898	0.8221	18,982	109	23,645	0.8028	6,311,457	1,370,939	114,773	2.0894
MAN	23,441	18,946	0.8082	19,028	106	24,194	0.7865	6,404,481	1,378,152	117,026	2.0879
MBU	24,983	20,193	0.8083	20,286	121	25,712	0.7890	6,681,061	1,444,168	123,504	2.0868
TOU	21,802	17,485	0.8020	17,559	111	22,451	0.7821	5,955,782	1,300,258	108,916	2.0925
YOR	23,741	19,122	0.8054	19,194	137	24,399	0.7867	6,322,794	1,370,000	117,027	2.0885

Table S1: SNP calling and functional annotation statistics by sample. Related to STAR Methods.
SNP calling and functional annotation statistics by sample. The number of SNPs refers to the number of sites with a non reference homozygous genotype (i.e., heterozygous + alternative homozygous).

test_pop	ref_A	ref_B	p-value	Z-score	Mean admixture time (generations)	95% CI (generations)	Mean admixture time (year)	95% CI lower bound (year)	95% CI upper bound (year)
Egyptian	Yoruba	BedouinB	0.000064	4	12.94	3.24	1643.74	1549.78	1737.7
Egyptian	Yoruba	Basque	0.0000023	4.72	17.98	3.81	1497.58	1387.09	1608.07
Egyptian	Yoruba	Berber_Tunisian	0.00000087	4.92	17.65	3.59	1507.15	1403.04	1611.26
Saharawi	Yoruba	BedouinB	7.60E-09	5.78	13.96	2.42	1614.16	1543.98	1684.34
Saharawi	Yoruba	Basque	1.50E-07	5.25	14.94	2.85	1585.74	1503.09	1668.39
Saharawi	Yoruba	Berber_Tunisian	1.90E-05	4.27	23.79	5.56	1329.09	1167.85	1490.33

Table S2: MALDER results. Related to Figure 3.
MALDER test results for Saharawi and Egyptian as test populations and Yoruba, BedouinB, Basque, Berber_Tunisian and Taforalt as reference populations. Only tests with statistically significant results are shown.