OPEN ACCESS Freely available online

PLOS | GENETICS

Early Back-to-Africa Migration into the Horn of Africa

Jason A. Hodgson[1], Connie J. Mulligan[2], Ali Al-Meeri[3], Ryan L. Raaum[4,5]*

1 Department of Life Sciences, Silwood Park Campus, Imperial College London, Ascot, Berkshire, United Kingdom, **2** Department of Anthropology and the Genetics Institute, University of Florida, Gainesville, Florida, United States of America, **3** Department of Biochemistry and Molecular Biology, Sana'a University, Sana'a, Yemen, **4** Department of Anthropology, Lehman College and The Graduate Center, The City University of New York, Bronx, New York, New York, United States of America, **5** The New York Consortium in Evolutionary Primatology (NYCEP), New York, New York, United States of America

Abstract

Genetic studies have identified substantial non-African admixture in the Horn of Africa (HOA). In the most recent genomic studies, this non-African ancestry has been attributed to admixture with Middle Eastern populations during the last few thousand years. However, mitochondrial and Y chromosome data are suggestive of earlier episodes of admixture. To investigate this further, we generated new genome-wide SNP data for a Yemeni population sample and merged these new data with published genome-wide genetic data from the HOA and a broad selection of surrounding populations. We used multidimensional scaling and ADMIXTURE methods in an exploratory data analysis to develop hypotheses on admixture and population structure in HOA populations. These analyses suggested that there might be distinct, differentiated African and non-African ancestries in the HOA. After partitioning the SNP data into African and non-African origin chromosome segments, we found support for a distinct African (Ethiopic) ancestry and a distinct non-African (Ethio-Somali) ancestry in HOA populations. The African Ethiopic ancestry is tightly restricted to HOA populations and likely represents an autochthonous HOA population. The non-African ancestry in the HOA, which is primarily attributed to a novel Ethio-Somali inferred ancestry component, is significantly differentiated from all neighboring non-African ancestries in North Africa, the Levant, and Arabia. The Ethio-Somali ancestry is found in all admixed HOA ethnic groups, shows little inter-individual variance within these ethnic groups, is estimated to have diverged from all other non-African ancestries by at least 23 ka, and does not carry the unique Arabian lactase persistence allele that arose about 4 ka. Taking into account published mitochondrial, Y chromosome, paleoclimate, and archaeological data, we find that the time of the Ethio-Somali back-to-Africa migration is most likely pre-agricultural.

Citation: Hodgson JA, Mulligan CJ, Al-Meeri A, Raaum RL (2014) Early Back-to-Africa Migration into the Horn of Africa. PLoS Genet 10(6): e1004393. doi:10.1371/journal.pgen.1004393

Editor: Scott M. Williams, Dartmouth College, United States of America

Received May 20, 2013; **Accepted** April 7, 2014; **Published** June 12, 2014

Funding: This work was supported by NSF BCS-0518530 to CJM and by research funds provided by the School of Natural and Social Sciences of Lehman College to RLR. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: ryan.raaum@lehman.cuny.edu

Introduction

The timing and extent of migration and admixture are questions that are central to the entire scope of human evolutionary history from the origin of our species to the present day. The most important event underlying human population structure is the origin of anatomically modern humans in Africa and their subsequent migration around the globe [1–3]. Following the initial out-of-Africa migration, the rate of migration between sub-Saharan Africa and the rest of the Old World was low throughout prehistory, but not absent; there is statistically significant evidence for a deep history of intercontinental migration [4–7]. Beginning around 11 ka (thousand years ago), the switch to reliance on domesticated plants and animals is associated with major population and language expansions from multiple centers of domestication around the world [8–10]. Finally, migration and admixture accelerated during the last few thousand years with increasing international trade, including the trade in slaves and the transplantation and shuffling of populations in the colonial era, culminating in the modern era of high international migration.

Populations in the Horn of Africa (HOA: Ethiopia, Eritrea, Djibouti, and Somalia) have substantial non-African ancestry [11–15]. The most recent genomic studies estimate 30–50% non-African ancestry in the Cushitic and Semitic speaking populations of the HOA resulting primarily from admixture around 3 ka [16,17]. This timeframe corresponds to the estimated time of origin of the Ethiosemitic languages [18] and there are some carved inscriptions in South Arabian scripts associated with temple ruins and ritual items in South Arabian styles dated to the early first millennium BCE in the north Ethiopian highlands [19–23]. These linguistic and archaeological connections have been cited in the recent population genomic studies to support a hypothesis of high levels of non-African migration into the HOA around 3 ka.

However, more recent archaeological research shows that non-African influences in the HOA were limited and transient. Of the early first millennium BCE inscriptions in non-African scripts complete enough to identify a language, only a small proportion are written in a non-African (South Arabian) language - the majority are written in indigenous proto-Ge'ez [24]. In the HOA, architecture with non-African (primarily South Arabian) elements is entirely monumental or ritual [25] and ritual items with exclusively non-African elements are rare [26]. There are few to no indications of non-African material culture in everyday objects: the ceramics and lithics found outside of the ritual context are almost entirely indigenous with clear local precedents [24,25,27].

Author Summary

The Horn of Africa (HOA) occupies a central place in our understanding of modern human origins. This region is the location of the earliest known modern human fossils, a possible source for the out-of-Africa migration, and one of the most genetically and linguistically diverse regions of the world. Numerous genetic studies over the last decades have identified substantial non-African ancestry in populations in this region. Because there is archaeological, historical, and linguistic evidence for contact with non-African populations beginning about 3,000 years ago, it has often been assumed that the non-African ancestry in HOA populations dates to this time. In this work, we find that the genetic composition of non-African ancestry in the HOA is distinct from the genetic composition of current populations in North Africa and the Middle East. With these data, we demonstrate that most non-African ancestry in the HOA cannot be the result of admixture within the last few thousand years, and that the majority of admixture probably occurred prior to the advent of agriculture. These results contribute to a growing body of work showing that prehistoric hunter-gatherer populations were much more dynamic than usually assumed.

While earlier scholarship conceived of a South Arabian origin D'MT polity with sovereignty over much of the northern HOA, it is now clear that this polity, if it ever existed at all as an integrated state [24], was geographically restricted to the regions around Yeha and Aksum in what is now the Tigray region of Ethiopia [25]. Artifacts with non-African features are effectively absent in the material culture (ritual or otherwise) of contemporaneous populations in the Eritrean highlands on the Asmara plateau (the "Ancient Ona") [25,28,29]. Prior to the first millennium BC, the archaeology of the HOA is less well studied, but what is available shows no substantial non-African material culture beyond trade relations [25]. Taken all together, the archaeological data could be consistent with limited non-African (primarily South Arabian) migration into the north Ethiopian highlands at the outset of the first millennium BCE, but cannot support large-scale population movements from any foreign population.

Archaeological data indicate trade between the HOA and Arabia by at least 8 ka [30,31] and genetic analyses of mitochondrial and Y chromosome data suggest much earlier migrations into the HOA. Mitochondrial data are suggestive of as many as three waves of prehistoric non-African migration into the HOA. First, HOA populations carry several unique M1 lineages of the otherwise South and East Asian mitochondrial haplogroup M [13,32–34]. Many of these HOA M1 lineages have deep roots, diverging from M1 representatives elsewhere 20–30 ka [34–36]. Second, representatives of N1a and N2a in the HOA diverged from their most closely related haplotypes in the Middle East and the Caucasus 15–20 ka [37]. Third, in the Eurasian mitochondrial HV1 and R0a lineages there are several sub-haplogroups (HV1a3, HV1b1, R0a2b, R0a2g) that are found in both the HOA and the Arabian Peninsula. Within these shared sub-haplogroup lineages, the HOA and Arabian haplotypes are distinct, suggesting that the migration that brought these lineages into the HOA happened soon after the sub-haplogroups began to diversify at 6–10 ka [38,39].

Y chromosome data are also suggestive of at least two episodes of non-African migration into the HOA prior to 3 ka. First, HOA populations carry E-M78 Y chromosomes at high frequencies [40,41]. E-M78 originated in northeastern Africa around 19 ka with a descendant lineage (E-V32) unique to the HOA that arrived by at least 6 ka [41]. Because northern African populations in this timeframe are inferred to have substantial non-African ancestry [42,43], the expansion south of E-M78 could have introduced non-African ancestry into the HOA prior to 6 ka. Second, some HOA populations carry moderate to high frequencies of T-M70 (previously K2-M70) Y chromosomes [44–46]. The T haplogroup originated in the area of the Levant approximately 21 ka and the T-M70 sub-haplogroup was present in northeast Africa by at least 14 ka, possibly arriving in the HOA as early as 5 ka [44,45,47].

In order to investigate the discrepancy among the archaeological, historical, mitochondrial, Y chromosome, and genome-wide data for recent vs. more ancient evidence of admixture in the HOA, we generated new genome-wide SNP data for a Yemeni sample and analyzed these new data with publicly available data [16,43,48–51]. Our objectives were to verify the presence of admixture in the HOA, determine the affinities of any HOA non-African ancestry, and evaluate the number of distinct admixture episodes and their timing.

Results and Discussion

For these analyses, we generated new genome-wide SNP data using the Illumina 370K array from 61 Yemenis, chosen to represent all geographic regions of the country. These new data were merged with published data from the HOA [16], the Middle East [48], North Africa [43], Qatar [50], southern Africa [51], west Africa [49], the HapMap3 project [52], and the Human Genome Diversity Project [53]. After reduction to SNPs shared across all source datasets and quality control, the main merged dataset included 2,194 individuals from 81 populations for 16,766 SNPs (Table S1).

Horn of Africa populations in the regional genetic landscape

We first investigated the position and dispersion of HOA populations in the genetic landscape in a multi-dimensional scaling (MDS) analysis of pairwise identity by state (IBS). Consistent with prior analyses of global genome-wide genetic variation [3,53,54], the first dimension of the IBS MDS analysis separates sub-Saharan Africans from non-Africans (Figure 1A). The HOA samples are broadly dispersed between the main sub-Saharan Africa cluster and the non-African populations and several sub-clusters of HOA samples are apparent. To see the specific distribution of all of the included HOA samples, we plotted the HOA samples in isolation (Figure 1B). While we include many more African and non-African population samples than prior analysis of these HOA data, our results in the MDS analysis for the HOA samples are not qualitatively different than those of Pagani et al. [16], who showed that the different HOA clusters correspond to linguistic groups: the Gumuz are Nilotic-speaking, the Ari and Wolayta are Omotic-speaking, and the rest speak Cushitic or Semitic languages. The dispersion of HOA samples between the sub-Saharan and non-African clusters is suggestive of admixture between African and non-African ancestors [55,56].

In order to better understand the genetic structure of HOA populations, we analyzed the SNP data using the model-based, maximum likelihood ancestry estimation procedure implemented in the ADMIXTURE software [57] for K values from 2 to 20 (Figure S1). For this analysis, we excluded SNPs in strong linkage disequilibrium, which reduced the main dataset to 16,420 SNPs. We used the cross-validation method encoded in the ADMIX-

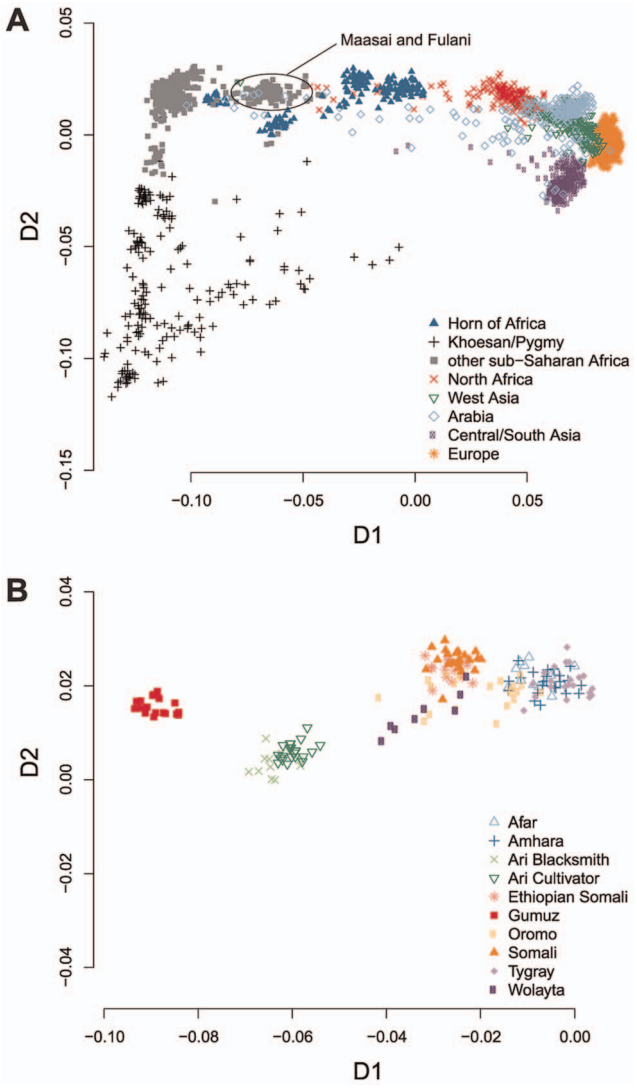

Figure 1. Multidimensional scaling analysis shows the great genetic diversity within the Horn of Africa. We plotted the first two dimensions of a multidimensional scaling analysis of pairwise identity by state across all study populations. (A) The HOA populations are broadly scattered between out-of-African populations and the bulk of sub-Saharan African populations along the first dimension. Some clusters of HOA individuals are much closer to the main sub-Saharan African cluster, while others are much closer to North African and Arabian clusters. (B) In this plot, we zoom in on the HOA samples and leave out all other populations. While the region as a whole covers a broad swath of the first MDS dimension, most individual populations are tightly clumped, with groups separated by language. The Nilo-Saharan speaking Gumuz are on the far left, the Omotic speaking Ari are in the center, and the Cushitic and Semitic speaking populations are on the right.
doi:10.1371/journal.pgen.1004393.g001

TURE software in an attempt to estimate the optimal number of inferred ancestral components (K) [58]. This cross validation procedure splits the genotype data into partitions and masks (marks as missing) each partition in turn, predicting the genotypes of the masked sites from the remaining unmasked data. For our data, the cross-validation error is minimized at K = 12, but there is little difference in error from K = 9 to K = 14 (Figure S2). For HOA populations, the ADMIXTURE estimates for K = 9–14 fall into three distinct patterns at K = 9–10, K = 11, and K = 12–14 (Figure S1). Here we focus on the ancestral component estimates for K values of 10, 11, and 12 as representative of these three patterns (Figure 2).

There are ten inferred ancestry components (IACs) that are consistent across all three focal K values (Figure 2) and are congruent with published analyses of African and Eurasian population structure [3,49,59,60]. Four IACs are found predominantly in sub-Saharan African populations: (1) one with high frequencies in the Mbuti and Biaka pygmies that is colored pink in the figure; (2) one with high frequencies in Khoesan speaking populations of southern Africa that is colored light purple; (3) one with high frequencies in Niger-Congo speaking populations throughout sub-Saharan Africa that is colored dark blue; and (4) one with high frequencies in Nilo-Saharan speaking populations that is colored light blue. Five IACs are found predominantly in Eurasia: (1) one with high frequencies in Central and South Asian populations that is colored red; (2) one with high frequencies in European populations that is colored light orange; (3) one with its highest frequencies in southern Europe, the Middle East, and Central Asia that is colored dark orange; (4) one with its highest frequencies in Arabian populations that is colored brown; and (5) one with its highest frequencies in Central Asian populations of known East Asian ancestry that is colored grey. The tenth shared IAC is colored light green and predominates in North African populations. This "Maghrebi" IAC has been recovered in previous studies of North African populations and is hypothesized to represent a late Pleistocene migration of non-African ancestors back into Africa [43,61].

From K = 10 to K = 12, the changes in ADMIXTURE results occur primarily in the HOA, where two new IACs appear at high frequencies (Figure 2). At K = 10 the African ancestry of HOA populations is dominated by the Nilo-Saharan IAC and the non-African ancestry is mostly split between Arabian and Maghrebi IACs. At K = 11 a new African IAC, colored dark purple in the figure, which we refer to as "Ethiopic", replaces much of the previously Nilo-Saharan attributed ancestry. The Ethiopic IAC reaches its highest frequencies in the Omotic speaking Ari and Wolayta populations, and is present at moderate frequencies in Semitic and Cushitic speaking populations. Pagani et al. [16] previously reported the presence of an equivalent Ethiopia-specific IAC (colored yellow in their Figure 1C). At K = 12 a second new IAC replaces almost all of the Maghrebi and much of the Arabian attributed non-African ancestry. This IAC is colored dark green on the figure and is referenced here as "Ethio-Somali". This Ethio-Somali IAC is found at its highest frequencies in Cushitic speaking Somali populations and at high frequencies in neighboring Cushitic and Semitic speaking Afar, Amhara, Oromo, and Tygray populations. This IAC was not identified in the source study for the HOA SNP data [16], but Tishkoff and colleagues [59], in an analysis of an independent autosomal microsatellite dataset, did recover an equivalent IAC (calling it "Cushitic"). While this Ethio-Somali IAC is found primarily in Africa, it has clear non-African affinities (Text S1).

Confident determination of the appropriate K value in an ADMIXTURE-like analysis in most human population genomic studies is problematic because the information required to set K *a priori* is unknown. In fact, there is no true K value in most cases because the simultaneous diversification model fit by ADMIXTURE is a poor reflection of human population history. Therefore, rather than take the ADMIXTURE IACs for one of K = 10,11,12 at face value, we used these estimates as hypotheses about the genetic structure of HOA populations and then evaluated these hypotheses in separate analysis.

First, for all focal K values, the ADMIXTURE analysis suggests that many HOA populations have admixture between African and

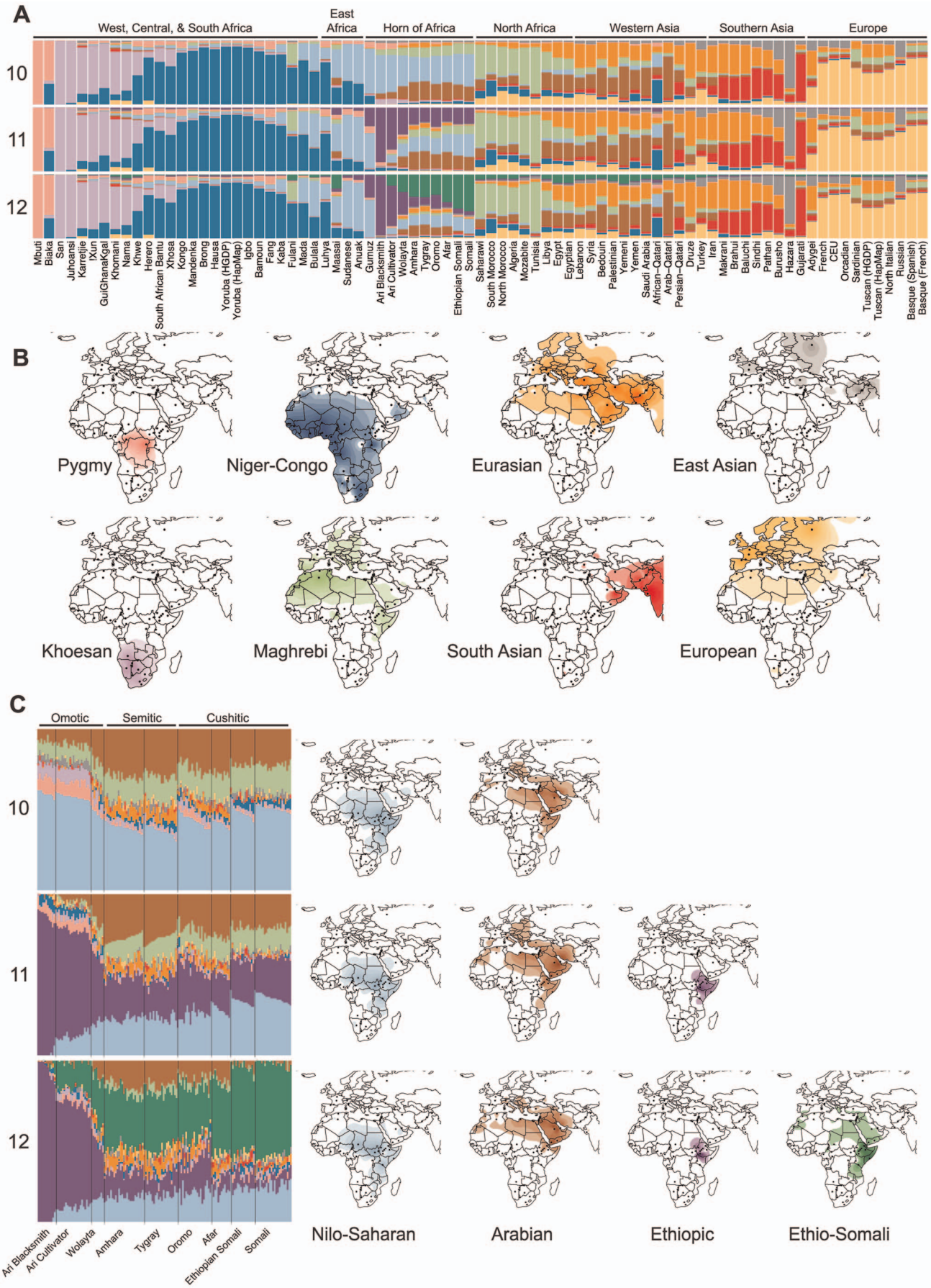
A
West, Central, & South Africa
East Africa
Horn of Africa
North Africa
Western Asia
Southern Asia
Europe
10
11
12
Mbuti
Biaka
San
Juhoansi
Karretjie
!Xun
GuiGhanaKgal
Khomani
Nama
Khwe
Herero
South African Bantu
Xhosa
Kongo
Mandenka
Brong
Hausa
Yoruba (HGDP)
Yoruba (HapMap)
Igbo
Bamoun
Fang
Kaba
Fulani
Mada
Bulala
Luhya
Maasai
Sudanese
Anuak
Gumuz
Ari Blacksmith
Ari Cultivator
Wolayta
Amhara
Tygray
Oromo
Afar
Ethiopian Somali
Somali
Saharawi
South Morocco
North Morocco
Algeria
Mozabite
Tunisia
Libya
Egypt
Egyptian
Lebanon
Syria
Bedouin
Palestinian
Yemeni
Saudi Arabia
African–Qatari
Arab–Qatari
Persian–Qatari
Druze
Turkey
Iran
Makrani
Brahui
Baluchi
Sindhi
Pathan
Burusho
Hazara
Gujarati
Adygei
French
CEU
Orcadian
Sardinian
Tuscan (HGDP)
Tuscan (HapMap)
North Italian
Russian
Basque (Spanish)
Basque (French)
B
Pygmy
Niger-Congo
Eurasian
East Asian
Khoesan
Maghrebi
South Asian
European
C
Omotic
Semitic
Cushitic
10
11
12
Ari Blacksmith
Ari Cultivator
Wolayta
Amhara
Tygray
Oromo
Afar
Ethiopian Somali
Somali
Nilo-Saharan
Arabian
Ethiopic
Ethio-Somali

Figure 2. Population structure of Horn of Africa populations in a broad context. ADMIXTURE analysis reveals both well-established and novel ancestry components in HOA populations. We used a cross-validation procedure to estimate the best value for the parameter for the number of assigned ancestral populations (K) and found that values from 9 to 14 had the lowest and similar cross-validation errors (Figure S2). (A) The differences in inferred ancestry from K = 9–14 are most pronounced in the HOA for K = 10–12, where two ancestry components that are largely restricted to the HOA appear (the dark purple and dark green components). (B) Surface interpolation of the geographic distribution of eight inferred ancestry components that are relatively unchanging and common to the ADMIXTURE results from K = 10–12. (C) Individual ancestry estimation for HOA populations (with language groups indicated) and surface plots of the changing distributions of the Nilo-Saharan (light blue) and Arabian (brown) ancestry components for K = 10–12. At K = 11, a new HOA-specific ancestry component that we call Ethiopic appears (dark purple) and at K = 12 a second new ancestry component that we call Ethio-Somali (dark green) appears with its highest frequencies in the HOA.
doi:10.1371/journal.pgen.1004393.g002

non-African ancestors in their history. To test this, we conducted three formal tests for admixture: the f_3-statistic test, the D-statistic test, and a weighted LD test [62,63]. We found that eight HOA populations (Afar, Amhara, Ari Cultivator, Oromo, Ethiopian Somali, Somali, Tygray, and Wolayta) had statistically significant signals of admixture with non-African populations for all three tests (Tables S2, S3, S4). With this strong support for a history of admixture between African and non-African ancestral populations, the differences among ADMIXTURE IACs across K = 10–12 suggest the following hypotheses for the African ancestry in the HOA:

1A. (K = 10) The HOA African ancestry is very similar to that found in neighboring Nilo-Saharan speaking populations.

1B. (K = 11,12) There is a distinct, differentiated African ancestry in HOA populations (the Ethiopic IAC).

And the following hypotheses for the non-African ancestry in the HOA:

2A. (K = 10,11) HOA populations experienced admixture with one or more non-African populations carrying high levels of the Arabian and Maghrebi IACs along with small amounts of the Eurasian IAC.

2B. (K = 12) There is a distinct non-African ancestry in the HOA that constitutes most of the non-African ancestry (the Ethio-Somali IAC).

To evaluate these ADMIXTURE-derived hypotheses, we used the CHROMOPAINTER software [64] to partition the chromosomes of HOA and neighboring populations into segments of African and non-African origin. We then sampled from the painted segments to create composite African and non-African ancestry chromosomes. To ensure that the African and non-African ancestry analyses would be directly comparable, we retained only those SNPs where samples could be generated from both the African and non-African painted segments across all populations, resulting in a dataset that includes 4,340 SNPs (the "4K partitioned" dataset). This dataset includes African and non-African partitioned samples from admixed HOA populations (Afar, Amhara, Ari [Blacksmith and Cultivator combined], Oromo, Somali [Ethiopian Somali and Somali combined], and Tygray) and from admixed Middle Eastern and North African (MENA) samples (Egypt, Mozabite, Palestinian, Yemen) as well as from relatively non-admixed African (Anuak, Gumuz, South Sudanese) and non-African (Bedouin, Druze, Saudi Arabia) populations. Further information on the population selection and ancestry painting methods is detailed in Materials and Methods. We used this 4K partitioned dataset to evaluate hypotheses of gene flow and population structure arising from the ADMIXTURE results.

African ancestry in the HOA

The hypothesis that African ancestry in the HOA is not distinct from that found in neighboring Nilo-Saharan speaking populations

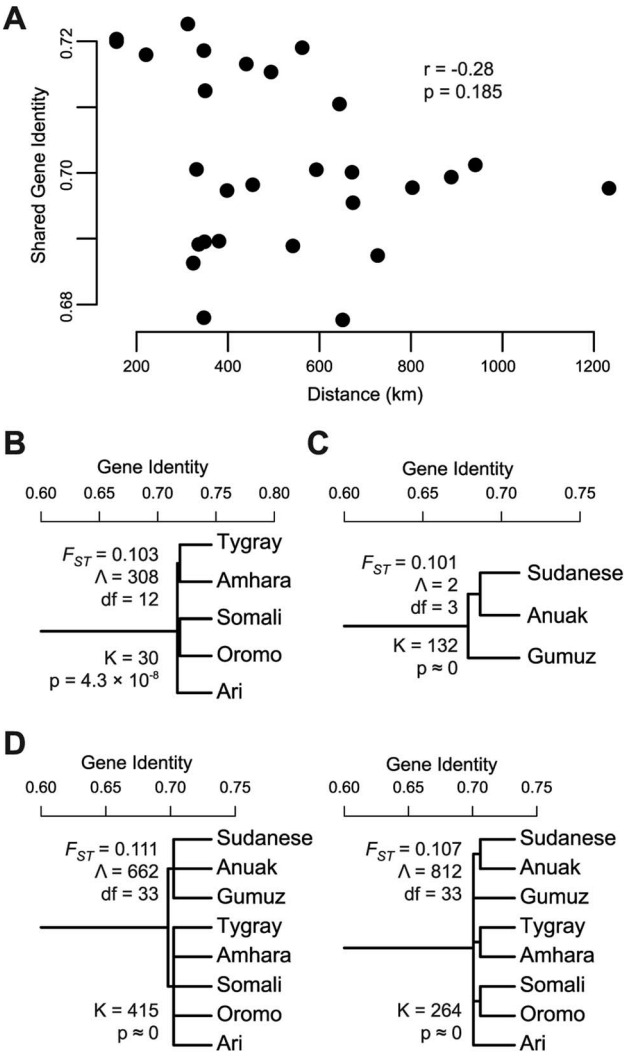

Figure 3. Tests of gene flow and population structure in the African ancestry of HOA populations. Using the African origin partition of the HOA data identified in a chromosome painting analysis, we evaluated the evidence for gene flow with neighboring populations and for population structure within and between the HOA and neighboring populations. (A) Shared gene identity plotted against distance for the HOA populations and the neighboring Anuak, Gumuz, and South Sudanese. (B) Linguistically structured population tree model within the African ancestry partition of HOA populations with the F_{ST} estimate from this tree model, the goodness-of-fit statistic Λ, and the likelihood ratio test statistic K for the improvement in model fit from the unstructured tree. (C) The linguistically structured population tree model for neighboring Nilo-Saharan language family populations. (D) Structured population tree models for the combined HOA and neighboring populations.
doi:10.1371/journal.pgen.1004393.g003

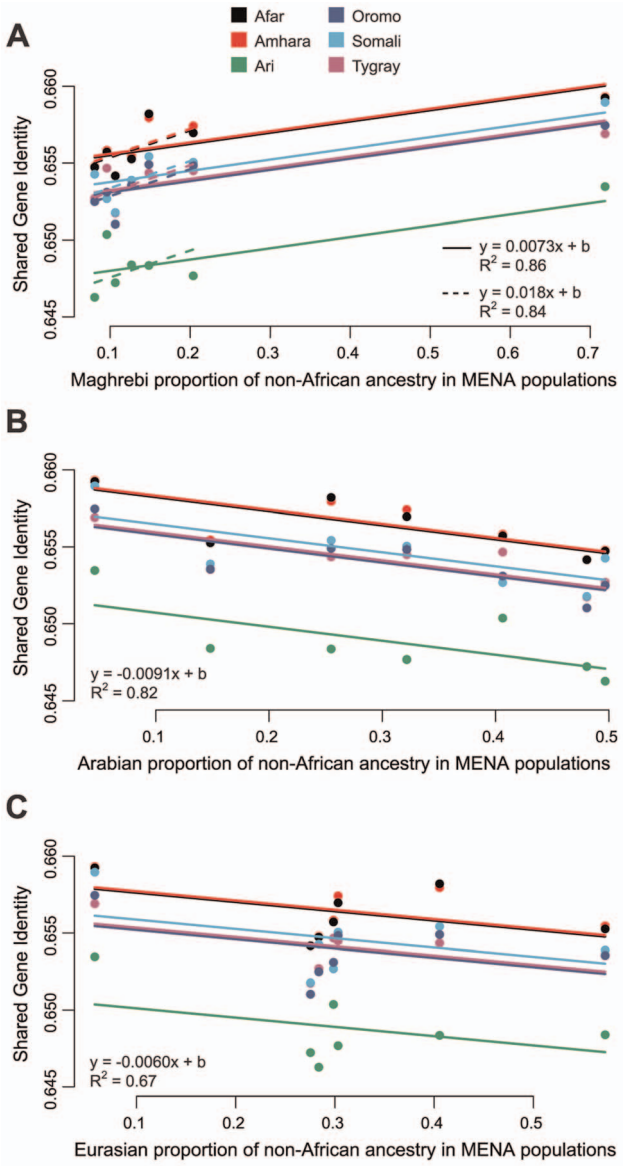

Figure 4. Relationship between non-African ADMIXTURE ancestry components and shared gene identity between HOA and MENA populations. ADMIXTURE results for K = 10,11 suggest that the non-African ancestry in HOA populations is indistinguishable from the "Maghrebi," "Arabian," and "Eurasian" ancestry components found in MENA populations. If this is a correct inference, then the shared gene identity of HOA populations should be higher with MENA populations with higher proportions of these ancestries. (A) There is significant positive relationship between shared gene identity and the proportion of Maghrebi ancestry in MENA populations. While there is variation across HOA populations in the overall shared gene identity (different intercepts for each population), the magnitude of the relationship is consistent (adding varying slopes did not significantly improve the model fit). This relationship holds whether or not the Mozabite (a high Maghrebi ancestry outlier) are included in the model. However, contrary to expectations, both the Arabian (B) and Eurasian (C) ancestry components showed a reduction in shared gene identity as the representation of these ancestry components in MENA populations increased.
doi:10.1371/journal.pgen.1004393.g004

(hypothesis 1A above) requires a history of homogenizing inter-population migration or relatively recent common origin. In the case of homogenizing gene flow, a correlation between genetic and geographic distance might be expected, with nearby populations more alike than distant populations. We calculated within and between population gene identity (the probability that two randomly drawn alleles are identical by state) for all the populations included in the 4K partitioned dataset. We then used the between population gene identity estimates among the predominantly Nilo-Saharan ancestry Anuak, Gumuz, South Sudanese, and the African ancestry partition of the Amhara, Ari, Oromo, Somali, and Tygray to test for a relationship between genetic and geographic distance. No significant relationship was recovered (Mantel test, $r = -0.28$, $p = 0.185$) (Figure 3A).

Since the pattern of genetic variation in the African ancestry of Sudanese and HOA populations is not a good fit to one model of ongoing gene flow, we tested the hypothesis that there is population substructure within and between HOA and Nilo-Saharan populations using AMOVA [65] and hierarchical population tree models [66,67]. First, within the HOA we used AMOVA to test for differentiation between linguistic groups – the Omotic speaking Ari, the Semitic speaking Amhara and Tygray, and the Cushitic speaking Oromo and Somali – and found a significant difference ($\Phi_{GT} = 0.013$, $p<0.001$). We also fit the HOA data to two population tree models, one without substructure and one with linguistically defined subgroups (Figure 3B), and found that the tree with the linguistic groups is a significantly better fit to the data ($K = 30$, $df = 1$, $p = 4.3\times10^{-8}$). Next, we tested for the presence of linguistically delineated subgroups within the Anuak, Gumuz, and South Sudanese. Most southern Sudanese populations speak languages in the Nilotic branch of the Nilo-Saharan language family and the Anuak language is also a Nilotic language [68]. The Gumuz language is either a highly divergent Nilo-Saharan language or a language isolate [69]. AMOVA reveals a statistically significant difference between these linguistic groups ($\Phi_{GT} = 0.024$, $p<0.001$) and the population tree with linguistically defined subgroups (Figure 3C) is a significantly better fit to the data than the tree without subgroups ($K = 132$, $df = 1$, $p\approx0$). Finally, putting all of the populations together in an AMOVA analysis, we find significant differences between linguistic subgroups at both a macro level (Nilo-Saharan vs Afro-Asiatic) ($\Phi_{GT} = 0.014$, $p<0.0001$) and a micro level (Nilotic, Gumuz, Omotic, Semitic, Cushitic) ($\Phi_{GT} = 0.022$, $p<0.0001$). Population tree models with these groupings are a significantly better fit to the data than a tree without subgroups ($K = 415$ and 264, $df = 1$, $p\approx0$) (Figure 3D). The tree with the larger subgroups (Nilo-Saharan vs Afro-Asiatic) is a slightly better fit to the data ($\Lambda = 662$) than the tree with the smaller subgroups ($\Lambda = 812$; smaller Λ values indicate better fit).

These results support the hypothesis from ADMIXTURE $K\geq 11$ of a distinct African ancestry with a long history in differentiated HOA populations (hypothesis 1B above) over the hypothesis from ADMIXTURE $K\leq10$ that African ancestry in the HOA is not substantially differentiated from that found in neighboring populations (hypothesis 1A). In fact, our results suggest a rather more complicated history for these regional populations. Studies of further population samples from ethnic groups in and near the western and southern edges of the Ethiopian escarpment are sure to be interesting.

Non-African ancestry in the HOA

The ADMIXTURE-derived hypothesis that non-African ancestry in the HOA derives from admixture with a population or populations with high levels of the Arabian and Maghrebi IACs and some of the Eurasian IAC (hypothesis 2A above) suggests that HOA populations should have higher levels of shared gene identity with populations with higher proportions

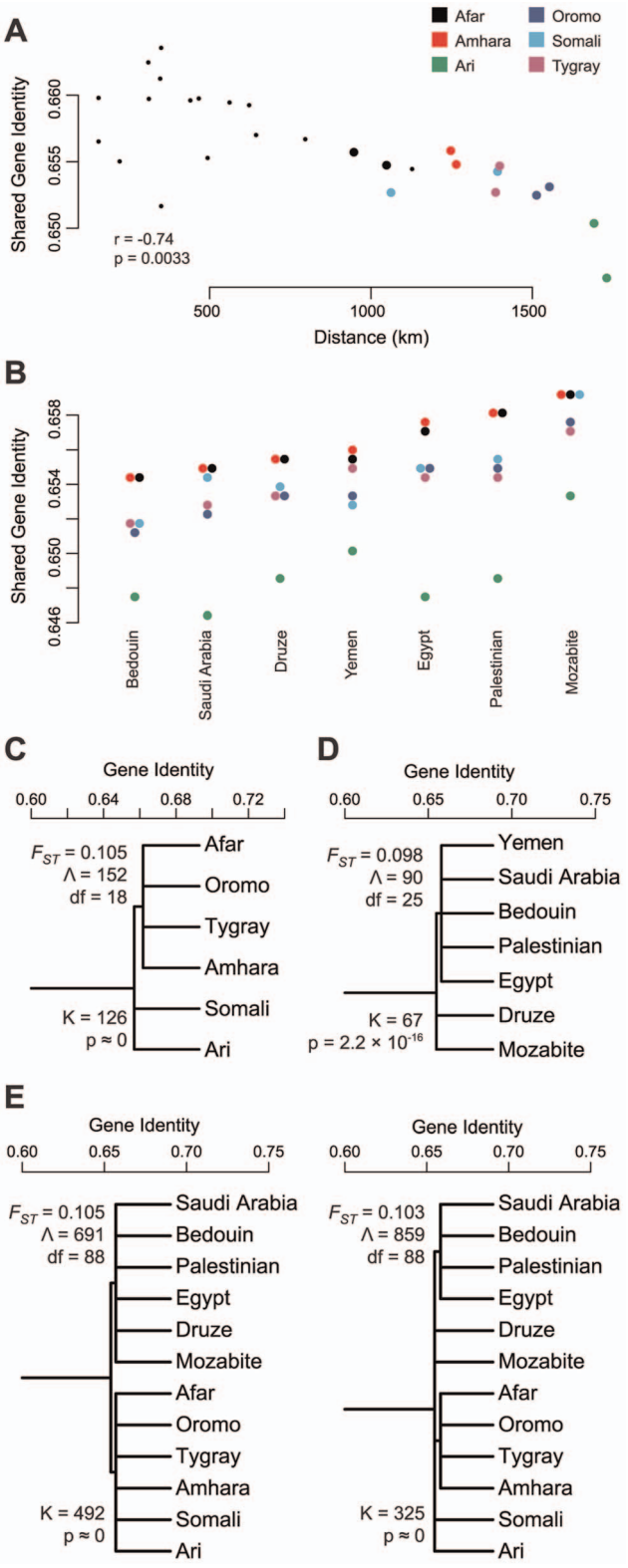

Figure 5. Tests of gene flow and population structure in the non-African ancestry of HOA populations. Using the non-African origin partition of the HOA data identified in a chromosome painting analysis, we evaluated the evidence for gene flow with MENA populations and for population structure within and between the HOA and MENA populations. (A) The only clear and statistically significant pattern of decreasing gene identity with geographic distance was between HOA populations and the Yemen and Saudi Arabia populations on the Arabian peninsula as evaluated by a Mantel test. This relationship only held for "as the crow flies" geographic distances; the relationship disappears using a waypoint through Egypt (Figure S3). (B) Shared gene identity between the non-African ancestry partition of HOA populations and MENA populations presented in increasing order. (C) Structured population tree model within the non-African ancestry partition of HOA populations with the F_{ST} estimate from this tree model, the goodness-of-fit statistic Λ, and the likelihood ratio test statistic K for the improvement in model fit from the unstructured tree. (D) Structured population tree model within the non-African ancestry partition of MENA populations. (E) Structured population tree models for the non-African ancestry partitions of both HOA and MENA populations. Both are significantly better fits to the data than the unstructured tree and the regional structure (HOA vs MENA) is a slightly better fit to the data as measured by the goodness-of-fit Λ statistic.
doi:10.1371/journal.pgen.1004393.g005

of those ancestries. To evaluate this prediction, we examined the relationship between shared gene identity and the ADMIXTURE-estimated proportion of the Arabian, Eurasian, and Maghrebi IACs in MENA population samples for each of the non-African ancestry partitions of the admixed HOA populations using varying intercepts linear models. Only the Maghrebi IAC analysis shows the expected relationship: shared gene identity between HOA and MENA populations increases as the proportion of Maghrebi ancestry increases (Figure 4A). Contrary to expectations, shared gene identity decreases between HOA populations and MENA populations as the proportion of the Arabian IAC (Figure 4B) and the Eurasian IAC (Figure 4C) increases.

Next, we looked for evidence for extended inter-population gene flow in the correlation of geographic distance and shared gene identity. We found no relationship between geographic and genetic distance within either HOA or MENA populations. We then examined this relationship for HOA populations to North African (Egypt, Mozabite), Levantine (Bedouin, Druze, Palestinian), and Arabian (Saudi Arabia, Yemen) populations (Figure S3). For North Africa and Arabia, we calculated both straight-line distances and distances involving a waypoint through Egypt. The only group for which there is a clear gradient of genetic similarity decreasing with geographic distance is for the straight-line distances with Arabian populations (Mantel test, $r = -0.74$, $p = 0.0033$) (Figure 5A). This relationship between genetic and geographic distance between HOA and Arabian populations might support a hypothesis of long-term equilibrium gene flow among these populations in an isolation-by-distance model. However, if this hypothesis were true, we would expect the highest levels of pairwise gene identity to be between HOA and Arabian populations, but this is not the case. The highest levels of shared gene identity are between HOA populations and the Levantine Palestinian and the North African Mozabite population samples (Figure 5B). Thus, it is more likely that the genetic-geographic HOA-Arabia distance gradient reflects secondary admixture of Arabian migrants into HOA populations already carrying substantial non-African ancestry or already admixed HOA populations sending migrants into Arabian populations.

While these results suggest some history of gene flow between HOA populations and MENA populations, there is no simple pattern that emerges. In order to better understand the partitioning of genetic variation among these populations, we tested for population substructure within and between HOA and MENA populations using AMOVA and hierarchical population tree models. First, within the HOA, we tested for linguistically

Table 1. Estimates of admixture dates in Horn of Africa populations (ka).[1]

Population	African	non-African	ROLLOFF	ALDER[2]	ALDER[3]
Afar	Yoruba (YRI)	CEU	3.6±0.5	3.2±1.0†	4.4/0.1
	Sudanese	Sardinian	4.7±0.6	4.3±1.0†	
	Sudanese	Turkey	4.7±0.6	3.6±0.8*	
Amhara	Yoruba (YRI)	CEU	2.8±0.2	2.6±0.4*	3.0/0.3
	Sudanese	Sardinian	3.5±0.2	2.0±0.4*	3.4/0.4
	Juhoansi	Sardinian	2.8±0.3	2.0±0.5	3.3/0.5
Ari Cultivator	Yoruba (YRI)	CEU	3.2±0.4	4.1±0.5†	
	Juhoansi	Sardinian	3.4±0.5	3.2±0.6	
Oromo	Yoruba (YRI)	CEU	2.7±0.3	1.0±0.8†	2.6/0.1
	Juhoansi	Sardinian	3.1±0.4	1.7±0.5†	3.3/0.1
	Sudanese	Basque	2.9±0.3	2.0±0.4*	3.1/0.3
Ethiopian Somali	Yoruba (YRI)	CEU	3.7±0.5	3.8±1.4†	
	Sudanese	Sardinian	4.2±0.5	3.1±1.1†	
	Juhoansi	Tuscan (TSI)	3.7±0.4	4.0±0.6	
Somali	Yoruba (YRI)	CEU	2.9±0.4	2.6±0.4†	
	Sudanese	Sardinian	3.9±0.4	2.6±0.6†	
	Juhoansi	Basque	3.7±0.4	3.1±0.4*	
Tygray	Yoruba (YRI)	CEU	3.6±0.3	1.7±0.4†	
	Sudanese	Sardinian	3.6±0.4	2.5±0.6	
Wolayta	Yoruba (YRI)	CEU	3.1±0.8	1.3±0.4	
	Juhoansi	Sardinian	2.2±0.6	1.1±0.3†	

[1]Using 30 years per generation.
[2]Single admixture fit.
[3]Two admixtures fits that are significantly better than the single admixture fit.
†ALDER test for admixture not significant.
* ALDER reports inconsistent decay rates.
doi:10.1371/journal.pgen.1004393.t001

defined substructure between Cushitic, Semitic, and Omotic speaking populations, but found no significant differentiation ($\Phi_{GT} = 0.010$, $p = 0.066$). We then used the ADMIXTURE results to inform subgroup formation. At $K = 12$, the Amhara, Tygray, Oromo, and Afar all have similar proportions of non-African ancestries that differ from that seen in the Ari and Somali (Figure 2). This observation suggests a geographical structuring between the Amhara, Tygray, Oromo, and Afar in the Ethiopian highlands, the Somali in eastern Ethiopia and the Somalia lowlands, and the Ari in the southwestern Ethiopian Rift. AMOVA of these three population groups reveals significant between group differentiation ($\Phi_{GT} = 0.017$, $p<0.0001$). In addition, the population tree with these geographic subgroups (Figure 5C) is a significantly better fit to the data than the tree without subgroups ($K = 126$, $df = 1$, $p\approx0$). Within MENA populations, linguistic subgroups cannot be defined, so we tested several historic/geographic groupings. Between population differentiation was maximized in the AMOVA analysis with three subgroups: the northwest African Mozabite; the ethnic and religious isolate Druze; and the populations with histories entwined with the development and expansion of Islam - the Egyptians, Palestinians, Bedouin, Saudi Arabians, and Yemeni. For this set of subgroups, between population differentiation was statistically significant ($\Phi_{GT} = 0.011$, $p<0.0001$) and the population tree with these subgroups (Figure 5D) is a significantly better fit to the data than the tree without subgroups ($K = 67$, $df = 1$, $p = 3.3\times10^{-16}$). Finally, putting all of the populations together in an AMOVA analysis, we find significant differences between HOA and MENA subgroups at both a macro level (HOA vs MENA) ($\Phi_{GT} = 0.014$, $p<0.0001$) and a micro level (all of the individual subgroups identified above) ($\Phi_{GT} = 0.016$, $p<0.0001$). Population tree models for both the simple (HOA vs MENA) and more complex (all individual regional subpopulations) groups are a significantly better fit to the data than a tree without subgroups (Figure 5E). As measured by the goodness-of-fit statistic Λ of Long and Kittles [67], the simple HOA vs MENA structure is the better first level structure fit to the data than the more complex structure with six different subgroups.

Even though there is strong evidence for admixture between HOA and MENA populations, there is also clearly detectable substructure both within and between the HOA and the Middle East and North Africa. If the majority of the non-African ancestry in the HOA had entered during the last few thousand years (hypothesis 2A above), then population groups should be less differentiated within the HOA then within MENA samples. This expectation is reinforced by the smaller geographic area of sampling within the HOA when compared to the geographic spread of the MENA samples. However, what we observe is that population groups within the HOA are more differentiated ($\Phi_{GT} = 0.017$; $F_{ST} = 0.105$) than population groups across the MENA region ($\Phi_{GT} = 0.011$; $F_{ST} = 0.098$). All together, these results offer greater support to the hypothesis from ADMIXTURE $K\geq12$ that there is a distinct non-African ancestry in the HOA that is well-differentiated from the non-African ancestry in neighboring Middle Eastern and North African populations (hypothesis 2B).

Table 2. Estimated mean proportion of ancestry (greater than 5%) in Horn of Africa populations.[1]

	Nilo-Saharan speaking		Cushitic speaking				Semitic speaking		Omotic speaking		
	Anuak	Gumuz	Afar	Oromo	Ethiopian Somali	Somali	Amhara	Tygray	Wolayta	Ari Cultivator	Ari Blacksmith
Niger-Congo	0.11										
Nilo-Saharan	0.75	0.59	0.20	0.19	0.22	0.23	0.16	0.17	0.15	0.10	
Ethiopic	0.07	0.35	0.08	0.21	0.08	0.06	0.16	0.12	0.35	0.63	0.94
Ethio-Somali			0.43	0.32	0.53	0.57	0.35	0.35	0.27	0.17	
Arabian			0.12	0.12			0.16	0.16	0.08		
Eurasian							0.07	0.08			

[1]None of the Horn of Africa populations have 5% or greater ancestry from Khoesan, Pygmy, Maghrebi, European, South Asian, or East Asian ancestral populations, so these ancestries are not shown in the table.

doi:10.1371/journal.pgen.1004393.t002

Timing of non-African admixture in the HOA

We used two methods that model the pattern of linkage disequilibrium (LD) expected to result from admixture to estimate the date of admixture for all study populations in which we found statistically significant signals of admixture: ROLLOFF [62,70] and ALDER [63]. Our estimates are broadly compatible with the dates previously calculated for these same population samples [16,17]. Using the HapMap YRI (Yoruba) and CEU (Utah residents with Northern and Western European ancestry) as reference populations, Pagani et al. [16] calculated ROLLOFF admixture dates ranging between 2,000 and 3,000 years ago. Pickrell et al. [17] calculated ALDER admixture date estimates for these populations between about 2,500 and 3,500 years ago, with some experiencing secondary admixture between 100 and 300 years ago. Across the entire set of reference populations that we used, our ROLLOFF estimates range from 2,200 to 4,700 years ago, and our single-admixture ALDER estimates are somewhat younger, ranging from 1,000 to 4,300 years ago (Table 1). Following Pickrell et al. [17], we compared the fit of single and dual admixture histories from ALDER in HOA populations and found, in agreement with their results, strong evidence for two admixture events in the Amhara and Oromo (Table 1).

These relatively recent dates are not consistent with our results showing a distinct, differentiated non-African ancestry in the HOA. To be sure, the greater shared gene identity between HOA populations and MENA populations with higher proportions of the Maghrebi IAC (Figure 4A) and the observed genetic-geographic correlation with Arabian populations (Figure 4B) are supportive of some relatively recent admixture. However, the high population differentiation found in the AMOVA and population tree analyses (Figure 5) suggests that admixture within the last few thousand years is a poor explanation for the majority of non-African ancestry in the HOA.

In order to seriously entertain the hypothesis that most of the non-African ancestry in the HOA predates the last few thousand years, we must understand why the ROLLOFF and ALDER admixture date estimates are all relatively recent. To do so, we performed simulation tests of ROLLOFF and ALDER in episodic admixture scenarios. To start, we simulated two episodes of admixture, with the first (earliest) episode between 50 and 200 generations ago (1,500–6,000 years using a 30 year generation time) and the second (more recent) episode at either 10 or 30 generations ago (300 or 900 years). We found a strong bias towards the most recent admixture date for both ROLLOFF and single-episode ALDER (Figure S4). To investigate the importance of the relative contribution of the first and second admixtures, we simulated both equal (10% first, 10% second) and unequal (25% first, 10% second) admixture proportions. Variation in the relative contribution of the first and second admixture episodes had a much greater effect on the ROLLOFF estimates than on the ALDER estimates (Figure S4). The stronger bias towards the date of the most recent admixture in the ALDER results is actually desirable, as this tendency makes the ALDER results much more interpretable.

To get an intuition for how ROLLOFF and ALDER perform for truly ancient admixture followed by more recent admixture, we simulated 50% admixture between 1,500 and 35,000 years ago, followed by 10% admixture at 900 years ago (Figure S5). As in the first simulation results, the ALDER estimate is almost always a reasonable estimate of the most recent admixture. If these results hold more generally, then a ROLLOFF estimate many times greater than the ALDER estimate might indicate ancient admixture; however, we are wary of over-generalizing from these data because the relationship between ROLLOFF and ALDER

Table 3. Coefficients of variation for Ethio-Somali and non-African ancestry components present above 5%.

Population	Ethio-Somali	Arabian	Eurasian
Afar	0.46	1.16	-
Amhara	0.70	1.46	3.06
Ari Cultivator	1.04	-	-
Oromo	0.76	2.19	-
Ethiopian Somali	0.42	-	-
Somali	0.39	-	-
Tygray	0.62	1.02	2.55
Wolayta	0.34	1.66	-

doi:10.1371/journal.pgen.1004393.t003

estimates appears to be similar for a broad range of admixture scenarios. Unfortunately, from the ROLLOFF and ALDER estimates alone, it is not possible to say when admixture started, whether it was continuous or successive, if there were multiple sources, how long it lasted, or if there was variation in admixture proportions over time [62]. What is clear is that admixture date estimates from either ROLLOFF or ALDER within the last few thousand years do not preclude the possibility of earlier episodes of admixture.

In order to evaluate the hypothesis that there were two or more distinct episodes of non-African admixture in the HOA, with the Ethio-Somali admixture occurring during an earlier episode, we conducted four analyses. First, we looked at the distribution of IACs among HOA populations from the K = 12 ADMIXTURE results. If there have been successive episodes of admixture into a culturally diverse region, we expect that different populations will have different histories of admixture [60,71]. Over time, admixed ancestry will be transmitted throughout the region via intra-regional gene flow, including into populations that have no history of direct admixture. If admixture predates modern population divisions, contemporary populations may carry admixed ancestry from a common admixed ancestor. In the HOA, this suggests that the Ethiopic IAC has the deepest roots in the region, as it is present at appreciable frequencies in all populations (Figure 2, Table 2). Next, the Nilo-Saharan IAC is found in all but the Ari Blacksmiths. The Ethio-Somali IAC has the third broadest distribution, and is found in all Cushitic and Semitic speaking populations as well as the Omotic speaking Wolayta and Ari Cultivators, but not the Ari Blacksmiths. Arabian, Eurasian, and Niger-Congo IACs have successively narrower distributions in the HOA. Based on this distribution of IACs across HOA populations, the most parsimonious order of origin in or migration into the region is Ethiopic – Nilo-Saharan – Ethio-Somali – Arabian – Eurasian – Niger-Congo, with the Nilo-Saharan and Niger-Congo gene flow probably coming from the west/southwest and the Ethio-Somali, Arabian, and Eurasian IACs likely arriving from the east/north.

Second, at the time of admixture, there would be a great deal of variation among individuals in the amount of ancestry from introgressing populations. After admixture ends, there will be a decrease in variation among individuals in the amount of admixed ancestry over time [72]. We calculated the coefficient of variation for all non-African IACs present above 5% in admixed HOA populations (Table 3). While all significantly admixed HOA populations have at least 5% of the Ethio-Somali IAC, only in a few populations are the Eurasian and Arabian IACs present above 5% (Table 2). In all cases, the coefficient of variation for the Eurasian and Arabian IACs is 2–5 times greater than that for the Ethio-Somali IAC, suggesting that the Ethio-Somali admixture predates the Eurasian and Arabian admixture. The largest coefficient of variation found for the Ethio-Somali IAC is in the Ari Cultivators. As the Ari Blacksmiths have negligible Ethio-Somali ancestry, it seems most likely that the Ari Cultivators are the descendents of a more recent admixture between a population like the Ari Blacksmiths and some other HOA population (i.e. the Ethio-Somali ancestry in the Ari Cultivators is likely to substantially postdate the initial entry of this ancestry into the region).

Third, we estimated divergence times among the IACs (Table 4) using a simple model of the expected relationship among F_{ST}, effective population size, and divergence time [73], similar to analyses conducted in prior studies of North African and Levantine samples [43,60]. The most recent divergence date estimates for the Ethio-Somali ancestral population are with the Maghrebi and Arabian ancestral populations at 23 and 25 ka. Among the many assumptions made for this calculation is that the pairwise F_{ST} values for the ADMIXTURE IACs reflect post-divergence population isolation and could be used to construct a bifurcating population tree that is a good fit for the data. If the population tree model is generally valid, but there has been some post-divergence migration, then the fit of the data to the tree model will not be good and the true divergence date would have been earlier than what we estimate here. When we evaluate the fit of the non-African IACs to a population tree model (Figure S6) using the goodness-of-fit statistic Λ of Long and Kittles [67], we find that the data deviate significantly from a good fit ($\Lambda = 1064$, df = 15, $p \approx 0$) and therefore the dates calculated here assuming a population divergence model are likely to be underestimates.

An alternative interpretation of the F_{ST} estimates is also possible. Wright originally formulated F_{ST} as a measure of differentiation between populations that is the result of an equilibrium between the opposing forces of gene flow and genetic drift [73]. In this formulation, the measured populations have never been completely isolated from each other and it would be inappropriate to attempt to calculate a divergence date from the F_{ST} value. The extremely poor fit of the IAC data to a bifurcating population tree model suggests that a population history without population isolation is also a possibility. Overall, the pairwise F_{ST} estimates for the IACs suggest that either the ancestral Ethio-Somali population had begun to differentiate from other non-African populations by at least 23 ka or that the ancestral Ethio-Somali population has never been completely isolated from other non-African populations. In either case, these data do not indicate when this population arrived in the HOA. When complete genome sequences become available for HOA, North African, and Middle Eastern populations, it should be

Table 4. Minimum time of divergence of ADMIXTURE inferred ancestry components (ka).

	Niger-Congo	Nilo-Saharan	Ethiopic	Ethio-Somali	Maghrebi	Arabian	Eurasian
Khoesan	34	39	40	**66**	66	73	75
Pygmy	26	31	34	**49**	59	67	68
Niger-Congo	-	15	24	**34**	43	52	54
Nilo-Saharan	15	-	22	**33**	43	51	53
Ethiopic	24	22	-	**31**	38	44	46
Ethio-Somali	34	33	31	**-**	23	26	31
Maghrebi	43	43	38	**23**	-	19	21
Arabian	52	51	44	**26**	19	-	18
Eurasian	54	53	46	**31**	21	18	-
European	58	57	50	**33**	19	19	14
South Asian	50	49	44	**35**	29	27	24
East Asian	52	51	46	**37**	31	30	26

doi:10.1371/journal.pgen.1004393.t004

possible to obtain better estimates using new methods being developed for unbiased sequence data, such as those based on the site frequency spectrum [74].

Fourth, a unique East African lactase persistence allele is found at its highest frequency in the Maasai [75] who have about 21% Ethio-Somali ancestry (Table S5). This lactase persistence allele is different from the alleles associated with lactase persistence in Europe [76,77] or Arabia [78,79], and likely arose during the last 7,000 years [75]. The Maasai do not have the Arabian lactase persistence allele, which is estimated to have originated about 4,000 years ago (95% CI: 250–27,575) and is present at high frequencies in Arabian populations (>50%) [78,79]. This Arabian allele is also almost absent in the Somali (1.6%) [79], which further supports our hypothesis that gene flow from Arabia within the last few thousand years cannot explain the non-African ancestry in HOA populations.

In summary, while LD-based methods estimate the time of non-African admixture in HOA populations to be within the last few thousand years, all sampled neighboring populations in North Africa or the Middle East are substantially differentiated from the non-African ancestry in the HOA. Based on this discrepancy, we undertook a closer examination of the properties of the LD-based ROLLOFF and ALDER admixture time estimation methods and found that earlier episodes of admixture are largely masked by more recent admixture events. Therefore, the admixture dates that are found within the last few thousand years do not falsify the hypothesis that the Ethio-Somali IAC arrived in the HOA during an earlier admixture event. (The ALDER/ROLLOFF simulation results do not directly support the hypothesis of earlier admixture, but they do show that earlier admixture cannot be excluded). The key lines of evidence that support a hypothesis of earlier admixture are that the Ethio-Somali IAC is broadly distributed across almost all ethnic groups in the HOA, consistent with an early entry into the region; that there is less inter-individual variance in Ethio-Somali IAC among individuals within ethnic groups than in Arabian or Eurasian IACs, suggesting that Ethio-Somali admixture predates the Arabian and Eurasian admixture; that the Ethio-Somali IAC is estimated to have diverged from all other non-African IACs by at least 23 ka; and that the Ethio-Somali IAC does not contain the unique Arabian lactase persistence allele that arose about 4 ka. In combination, these data suggest that the Ethio-Somali ancestors admixed with African-origin HOA ancestors sometime after 23 ka, but before the Middle Eastern admixture during the last few thousand years.

Non-genetic evidence for the timing of the Ethio-Somali back-to-Africa migration

Agriculture was established in the HOA by at least 7 ka [14,80], which suggests that local population densities were likely to have been relatively high from that time forwards. An external migration that occurred recently leading to 30–60% total genome-wide representation into pre-existing agricultural populations (Table S5) would require large or sustained population movements, which is not supported by either the historical or archaeological record [4,14]. The Ethio-Somali ancestry is more likely to have arrived during an earlier hunter-gatherer phase, when a smaller migration could make a significant contribution. As a point of reference, the slave trade into North Africa and the Middle East of over 11 million sub-Saharan Africans over the last 1,400 years [81] has led to a maximum of 30% total African ancestry in these populations.

Paleoclimate data offer some information on time ranges when human migration back-to-Africa would be most likely to succeed. During arid periods in North Africa and the Middle East, most

plausible routes into Africa experienced desertification, reducing the likelihood of successful migration. In our time frame of interest, there have been two major peaks of aridity in the region, the Last Glacial Maximum (LGM: ~21.5 ka) and the Younger Dryas (YD: ~12.5 ka), during which successful human migrations would not have been likely [82–86]. Since the end of the YD there have been fluctuations of arid and wet phases, but no arid periods as extreme or long lasting as these earlier two intervals [82]. Thus, if the Ethio-Somali ancestors diverged from all other non-African populations by 23 ka and were present in the HOA before the advent of HOA agriculture at around 7 ka [80], then there are three possible window of migration: post-YD, between the YD and the LGM, and pre-LGM. Because agriculturalist populations were expanding rapidly in the Middle East beginning about 12 ka and early agriculture in the HOA has an independent origin [80], the earlier YD-LGM and pre-LGM windows are favored.

There is abundant archaeological material in the HOA dating to between 5 and 30 ka, but most of the published literature is descriptions of surface surveys or test excavations [87,88]. More extensive investigations have focused on patterns of resource utilization [89–92], a key archaeological research goal, but less helpful for identifying cultural or biological affinities of early HOA populations. Contemporary HOA populations have occasionally been included in craniometric or dental studies of the biological affinities of ancient North Africans and Egyptians [93–95], but very little comparative analysis is available for the few prehistoric HOA skeletal collections [89,96]. One possible indication of ancient Ethio-Somali admixture might be found in studies of Late Pleistocene Nubians (~12 ka) from the Nile River Valley, who have been variously interpreted as sharing affinities with contemporaneous North African Iberomaurusians [97] and with sub-Saharan Africans [98]. Admixture of Ethio-Somali ancestors with African-origin populations in this region might explain these divergent interpretations of this Late Pleistocene Nubian population.

Relationship to the North African back-to-Africa migration

Like the Ethio-Somali, the Maghrebi IAC in North African populations derives from a early back-to-Africa migration [34,43,61,99–102]. Studies of North African populations reveal a complex layered history of admixture in North Africa, with an inferred pre-Last Glacial Maximum settlement of North Africa by a non-African population followed by gene flow from European, Middle Eastern, and sub-Saharan African populations dating from the end of the LGM to the recent past [43,103–105].

A single prehistoric migration of both the Maghrebi and the Ethio-Somali back into Africa is the most parsimonious hypothesis. That is, a common ancestral population migrated into northeast Africa through the Sinai and then split into two, with one branch continuing west across North Africa and the other heading south into the HOA. For the Ethio-Somali, the lowest F_{ST} value from the ADMIXTURE estimated ancestral allele frequencies is with the Maghrebi (Text S1), which is consistent with a common origin hypothesis. In contrast, the Maghrebi component has lower F_{ST} values with Arabian, European, and Eurasian ancestral populations than with the Ethio-Somali, which suggests that the Maghrebi diverged most recently from those populations, and might indicate separate back-to-Africa migrations for the Ethio-Somali and the Maghrebi. Unfortunately, the F_{ST} estimates alone are not robust enough to distinguish between single or separate back-to-Africa migrations. While the F_{ST} estimates for the ancestral populations are, in theory, free of confounding admixture, they derive from a simplified model of population history that is known to be inaccurate (simultaneous divergence) and are all assumed to be in Hardy-Weinberg equilibrium [57,106]. As a result, fine-scale differences in pairwise F_{ST} among ancestral populations should be interpreted with care.

Mitochondrial M1 and U6 lineages – sub-clades of mitochondrial haplogroups that are otherwise found only in Eurasian populations – are found both in North Africa and the HOA [34]. U6 has its highest frequencies and diversity in Northwest Africa and M1 has its highest frequencies and diversity in the HOA. The differing representation of deeply diverging M1 and U6 mitochondrial lineages in North Africa and the HOA shows that these regions have exchanged few female migrants since approximately 20 ka [36]. While these mitochondrial data further support our hypothesis that most of the non-African ancestry in the HOA has an ancient origin, we still cannot distinguish between single or separate migrations of the Maghrebi and Ethio-Somali back-to-Africa. If we could identify the geographical origins of both M1 and U6 and if these lineages originated in the same area, then a common migration hypothesis would seem more likely. The geographical origin of a mitochondrial clade is usually inferred from the presence of diverse early branching lineages within a region. To date, no region has been identified with a diversity of early branching lineages of either M1 or U6. Given the exclusively Eurasian distribution of the larger M and U haplogroups, it is generally inferred that M1 and U6 originated outside of Africa [34,35,100] but since all other early branches of M1 and U6 appear to have gone extinct, it is not possible to specify their location of origin. Most recently, Pennarun and colleagues [36] found that sub-lineages within U6 began diversifying in North Africa about 10,000 years before M1 sub-lineages began diversifying in the HOA (~30 ka vs. ~20 ka). This difference in coalescence times might be taken as evidence for separate migrations, but could also be explained by smaller population sizes in the HOA ancestors between 30 and 20 ka following a common migration.

Summary and implications

We find that most of the non-African ancestry in the HOA can be assigned to a distinct non-African origin Ethio-Somali ancestry component, which is found at its highest frequencies in Cushitic and Semitic speaking HOA populations (Table 2, Figure 2). In addition to verifying that most HOA populations have substantial non-African ancestry, which is not controversial [11–14,16], we argue that the non-African origin Ethio-Somali ancestry in the HOA is most likely pre-agricultural. In combination with the genomic evidence for a pre-agricultural back-to-Africa migration into North Africa [43,61] and inference of pre-agricultural migrations in and out-of-Africa from mitochondrial and Y chromosome data [13,32–37,47,99–102], these results contribute to a growing body of evidence for migrations of human populations in and out of Africa throughout prehistory [5–7] and suggests that human hunter-gatherer populations were much more dynamic than commonly assumed.

We close with a provisional linguistic hypothesis. The proto-Afro-Asiatic speakers are thought to have lived either in the area of the Levant or in east/northeast Africa [8,107,108]. Proponents of the Levantine origin of Afro-Asiatic tie the dispersal and differentiation of this language group to the development of agriculture in the Levant beginning around 12 ka [8,109,110]. In the African-origins model, the original diversification of the Afro-Asiatic languages is pre-agricultural, with the source population living in the central Nile valley, the African Red Sea hills, or the HOA [108,111]. In this model, later diversification and expansion within particular Afro-Asiatic language groups may be associated with agricultural expansions and transmissions, but the deep

diversification of the group is pre-agricultural. We hypothesize that a population with substantial Ethio-Somali ancestry could be the proto-Afro-Asiatic speakers. A later migration of a subset of this population back to the Levant before 6 ka would account for a Levantine origin of the Semitic languages [18] and the relatively even distribution of around 7% Ethio-Somali ancestry in all sampled Levantine populations (Table S6). Later migration from Arabia into the HOA beginning around 3 ka would explain the origin of the Ethiosemitic languages at this time [18], the presence of greater Arabian and Eurasian ancestry in the Semitic speaking populations of the HOA (Table 2, S6), and ROLLOFF/ALDER estimates of admixture in HOA populations between 1–5 ka (Table 1).

Materials and Methods

Ethics statement

Saliva samples were collected in Yemen in 2007 with informed consent under Western IRB approval, Olympia, WA. Subsequent analysis of anonymized SNP data was approved by the Lehman College IRB.

Genotyping of new Yemeni samples

Sixty-four Yemeni, chosen to represent all geographic regions of the country, were selected for SNP genotyping. Genomic DNA was extracted from saliva samples (DNA Genotek Oragene collectors) using the manufacturer's protocol. This DNA was genotyped using the Illumina 370k SNP chip by the University of Florida Interdisciplinary Center for Biotechnology Research Core Facility following the manufacturer's protocols. These new data are available from the Dryad Digital Repository (http://dx.doi.org/10.5061/dryad.d9s74) [112].

Data sets

We merged genome-wide SNP data from the HOA [16] with the new Yemeni data and other published data from the Middle East [48], North Africa [43], Qatar [50], southern Africa [51], west Africa [49], the HapMap3 project [52], and the Human Genome Diversity Project [53] using PLINK version 1.07 [113]. We excluded symmetric SNPs and SNPs and individuals with greater than 10% missing data. All known and inferred relatives were removed from the HapMap3 and HGDP data [114,115]. We then estimated kinship coefficients across all remaining individuals in all included populations using the "robust" algorithm, which is tolerant of population structure, in the KING software [116]. For all sets estimated to be second degree or closer relatives, we removed the individual(s) that would maximize the number of included individuals.

After pre-processing, the main dataset included 2,194 individuals from 81 populations for 16,766 SNPs (Table S1). We generated the linkage map for this dataset using the online map interpolator from the Rutgers second-generation combined linkage-physical map [117]. This dataset include some markers in strong linkage disequilibrium (LD), which is required for some of the analyses we conducted, but can bias other methods. For the methods that can be confounded by high levels of LD, we randomly excluded one of every pair of SNPs having pairwise genotypic correlation greater than 0.5 within a sliding 50 SNP window. After this exclusion, the "reduced-LD" dataset had 16,420 SNPs.

Many methods are known to perform better with more SNPs, especially those based on patterns of LD. To ensure that the estimates using these methods from our main dataset are reliable, we created two additional verification datasets with reduced population representation, which allows for greater overlap of mutually typed SNPs across studies. The "90K" dataset includes data for 91,101 SNPs from HOA, HapMap3, HGDP, and North Africa populations. The "260K" dataset includes data for 259,257 SNPs from the HOA, HapMap3, HGDP, southern Africa, and selected West Asian populations (see Table S1 for populations in the 90K and 260K datasets). All of the procedures described above for the main datasets were followed.

Population structure

Multidimensional scaling (MDS) was performed upon a genome wide matrix of identity by state (IBS) for all individual pairs in the reduced-LD dataset using PLINK [113]. For each increase in K from 2 to 5, there were substantial changes in reduced stress, but not for K greater than 5, so the IBS matrices were projected to 5-dimensional space. We inferred genetic structure and estimated admixture proportions in the reduced-LD dataset using ADMIXTURE [57]. Ancestry proportions were estimated for K values ranging from 2 to 20, and cross-validation error was calculated for each value of K. The geographic distribution of estimated admixture proportions were plotted using methods modified from Olivier François [118] using the MAPS, MAPTOOLS, and SPATIAL packages in R [119–122].

African and non-African origin data partitions

After phasing the 260K dataset using the haplotypes inference algorithm implemented in version 2 of the SHAPEIT software [123], we partitioned the phased data from admixed HOA and MENA populations into African and non-African chromosome segments using the chromosome painting method implemented in the CHROMOPAINTER software [64]. This algorithm "paints" each target individual as a combination of segments from "donor" populations. As donors, we selected individuals from African and non-African populations without significant evidence for admixture: African populations used as donors were the Anuak, Ju/'hoansi, Mandenka, Mbuti, San, South Sudanese, and Yoruba; non-African ancestry populations used as donors were the Adygei, Basque, Bedouin, Brahui, Burusho, CEU, Druze, Gujarati (GIH), Hazara, Makrani, Orcadians, Pathan, Sardinians, and Saudi Arabians. For each admixed individual, each chromosome segment that was "painted" with 80% or greater confidence from African or non-African donor populations was assigned that origin. On average, 85% of each admixed individual's genome could be confidently partitioned. We then sampled from the painted segments to create 12 African ancestry and 12 non-African ancestry chromosomes for the admixed HOA population samples and the key neighboring admixed population samples of the Yemeni, Palestinians, Egyptians, and Mozabite (12 chromosomes was chosen as a compromise between maximizing sample size and maximizing the included populations). The Ari Blacksmith and Ari Cultivator samples were combined into a single Ari sample and the Ethiopian Somali and Somali samples were combined into a single Somali sample. The small original sample size of the Afar ($n = 12$) made it impossible to assemble enough African ancestry painted chromosome segments for this population and neither enough African nor non-African painted chromosome segments could be assembled for the Wolayta (original $n = 8$). To ensure that the African and non-African ancestry analyses would be directly comparable, we retained only those sites where 12 alleles could be selected from both the African and non-African painted segments across all populations; this reduced the starting 260K dataset to 4,340 SNPs (the "4K partitioned" dataset). Because we required a complete dataset with no missing data, the intersection across populations of available data considerably reduces the number of

available sites (even though 85% of each individual genome could be confidently partitioned into African and non-African origin ancestries). Because of this, we had to use the 260K dataset, which unfortunately has reduced population representation, missing in particular most of the North African populations.

Tests of gene flow and population structure in partitioned data

Using the 4K partitioned dataset, we evaluated the evidence for gene flow and population structure using Mantel tests, AMOVA, and population tree models. We tested for geographically mediated gene flow using Mantel tests of the correlation between genetic distance as measured by shared gene identity and geographic distance using the implementation in the R ADE4 package [124] with 10,000 random permutations of the data to estimate p-values. When appropriate, geographic distance was calculated both "as the crow flies" and through a northeastern African waypoint in Egypt. Population structure was assessed using AMOVA and population tree models. For AMOVA, we modeled structure at three levels, within populations, between populations within groups, and between groups, but focused on the tests for between group population structure. We used the AMOVA implementation in the R ADE4 package [124] with 10,000 random permutations of the data to estimate p-values. Population tree models were constructed following the method of Long and Kittles [67]. The fit of the data to the tree was assessed using their likelihood ratio statistic Λ. In most cases, the data deviate significantly from a perfect fit, which is not unexpected: Long and Kittles note that this statistic is likely to be very sensitive to any violation of the model assumptions. We assessed the improvement in fit from a less structured population tree to a more structured population tree using the K likelihood ratio statistic [67]. Both the Λ and K statistics are chi-squared distributed random variables.

Ancestral population divergence

Population divergence times of the ADMIXTURE-inferred ancestral populations were estimated using the relationship $1-F_{ST}=(1-1/2N_e)^t$ [73]. This estimate assumes that effective population sizes are known and have remained stable through time. We used a generation time of 30 years [125–127] and estimated minimum divergence times using an N_e of 5,000, which is on the lower end of the N_e values estimated for relevant HGDP populations [53]. Wright's original formulation of F_{ST} as a measure of differentiation resulting from the equilibrium between gene flow and genetic drift that is discussed in the main text is $F_{ST}=1/(4N_em+1)$ [73].

Admixture tests and proportions

We formally tested for the presence of admixture in all study populations using the f_3-statistic, the D-statistic, and a weighted LD statistic [62,63]. Because a significant result for any one of these tests may be produced by histories other than admixture, we only report support for an admixture hypothesis when we found support for admixture from all three tests. To test for admixture between a sub-Saharan African and a non-African population, the f_3 test requires a reference population for each, which need not be the actual admixture source. For sub-Saharan Africa reference populations, we used populations that showed very little admixture of ancestral population components in the ADMIXTURE analysis: Mbuti Pygmies, Ju/'hoansi, HapMap3 Yoruba, South Sudanese, and Ari Blacksmith. For non-African reference populations, we used the HapMap3 CEU, Gujarati, and Tuscan populations in addition to Basque, Turkey, and Sardinian. The f_3 test was run for all other study populations for all possible pairs of reference populations. A strict Bonferroni correction was applied to control for multiple testing, only Z-scores less than -4 for the most negative f_3 statistic for each test population were considered significant. For those populations with significant f_3 statistics, the bounds of the admixture proportion were then estimated with the addition of a chimpanzee outgroup. The f_3 tests on the 90K and 260K datasets have more power, but return almost exactly the same f_3 statistic values (Table S7).

The test for admixture based on the D-statistic requires three populations in addition to the test population [62]. D-statistics significantly different from zero indicate either admixture or ancestral population structure. As in the f_3 test, the reference population suspected to be the source of admixture need not be the true source. We chose our population sets such that only positive values would reflect the admixture of interest. For sub-Saharan African and HOA test populations, the unrooted tree tested was ((African reference, test population), (Papuan, Basque)), where the African reference populations are the same as for the f_3 test. Since there is no indication in the literature of any African admixture in the Papuan population, any significantly positive D-statistic was taken as support for admixture between the test population and (a population related to) the Basque. For North African, Middle Eastern, and Eurasian test populations, the unrooted tree tested was ((Papuan, African reference), (Basque, test population)), where the African reference populations are the same as before. Again, since there is no indication in the literature of any admixture between Papuans and Basque, any significantly positive D-statistic indicates admixture between the test population and an African reference population. A strict Bonferroni correction was applied to control for multiple testing, only Z-scores greater than 4 for the most positive D-statistic for each test population were considered significant. The D tests on the 90K and 260K datasets have more power but recover indistinguishable D statistic values (Table S8).

Like the f_3 test, the weighted LD test in the ALDER software requires two reference populations, which need not be the actual admixture sources [63], and we used the same sets of non-African and sub-Saharan African reference populations. The test procedure implemented in ALDER controls for multiple testing across all the pairs of populations for each test population, but we still controlled for multiple testing across the whole family of tests using a strict Bonferroni correction, with only Z-scores greater than 3.2 considered statistically significant. The ALDER tests for admixture on the 90K and 260K datasets have more power but return similar results (Table S9).

We used three methods to calculate non-African admixture proportions in significantly admixed populations. First, we estimate the lower and upper bounds of non-African admixture using the bounding procedure allied with the f_3 admixture test [62]. This method requires an outgroup to the three populations in the f_3 test, but does not require a large sample, or even polymorphism, for the chosen outgroup. Therefore, following the recommendation in the description of this method, we used chimpanzee as the outgroup. Second, we estimated admixture proportions using the f_4 ratio estimation method [62]. The required number of populations and relationships among those populations for this method are as described for the D statistic test for admixture above, with the addition of an outgroup. Again, we used chimpanzee as the outgroup. Finally, for our third measure of non-African admixture proportions, we summed the proportions attributed to non-African ancestries from our ADMIXTURE analysis at K = 12.

Admixture dating and simulations

We estimated the time of admixture for all populations identified as admixed using two LD-based methods: ROLLOFF [62,70] and ALDER [63]. Following Pickrell et al. [17], we also compared the fit of single and double admixture models for admixed HOA populations. For comparison with other published admixture dates, we used the HapMap3 CEU and Yoruba populations as references. We also used the reference populations that gave the top f_3 statistic in the f_3 test for admixture and the reference populations giving the strongest signal in the ALDER test for admixture (sometimes these were the same). To verify the admixture date estimates calculated from the main (~17K SNP) dataset are reliable, we ran ROLLOFF and ALDER on both the 90K and 260K datasets using the HapMap3 Yoruba and CEU as the reference populations. Using the main dataset, we estimate ROLLOFF admixture dates from 2.6–3.7 ka and ALDER admixture dates from 1.1–4.1 ka for admixed HOA population. The verification estimates are not meaningfully different from these, with ROLLOFF admixture dates from 2.6–3.7 ka and ALDER admixture dates from 1.2–3.3 ka for the 260K dataset (Table S10).

We simulated individuals of admixed ancestry following published protocols [70,128]. We extracted 20 CEU and 40 Yoruba (YRI) individuals from a 260K SNP combined HapMap3 and HDGP dataset and phased them using fastPHASE [129]. These phased chromosomes were combined in episodic admixture scenarios, with two instances of admixture. We started with 20 CEU individuals and selected 20 random Yoruba individuals, and simulated admixture at time λ_0 with admixture proportion α_0 deriving from the Yoruba and $1 - \alpha_0$ from the CEU. For each haploid admixed genome, we randomly selected one chromosome from each source population. We then created a vector of ancestry transition events along each chromosome by sampling with probability $1 - e^{-\lambda 0 g}$, where g is the genetic distance in Morgans. Using this vector of transition event locations, we selected ancestry from the Yoruba chromosome with probability α_0 at each transition. This procedure was repeated until we had 40 haploid admixed genomes. We then used these admixed chromosomes as a source population for the second episode of admixture at time λ_1 with admixture proportion from α_1 from the remaining 20 YRI individuals not selected for the first admixture. We randomly combined the 40 haploid admixed genomes into 20 diploid individuals. We chose to simulate 20 admixed individuals because the modal number of individuals in our admixed populations was about 20.

In our first set of simulations, we simulated admixture with λ_0 equal to 50, 100, 150, or 200 generations and λ_1 equal to 10 or 30 generations. Admixture proportion α_0 was either 0.10 or 0.25 and admixture proportion α_1 was 0.10. Three independent replicates were performed for each combination of parameters (48 simulations in total). The second set of simulations used λ_0 equal to 50, 100, 150, 300, 500, 650, 850, 1000, or 1150 generations and λ_1 equal to 30 generations. Admixture proportion α_0 was 0.50 and admixture proportion α_1 was 0.10. Again, three independent replicates were performed for each combination of parameters (27 simulations in total). Admixture dates were estimated for the simulation data using ROLLOFF and ALDER with the remaining unadmixed CEU and Yoruba individuals as the reference populations. In addition, we reduced the simulated data to the 16,766 SNPs present in the main dataset used to estimate admixture dates for the study populations and estimated admixture dates using ROLLOFF and ALDER for the same set of reference population pairs.

Supporting Information

Figure S1 ADMIXTURE-inferred population average ancestry components for K = 2–20. The Ethiopian-specific Ethiopic ancestry component (dark purple) first appears at K = 11 and is stable from K = 12–20. The back-to-Africa Ethio-Somali ancestry component (green) first appears at K = 12 and is stable from K = 13–20.
(EPS)

Figure S2 Cross-validation error for K = 2–20 from the ADMIXTURE analysis. Ancestry proportions were estimated for K values ranging from 2 to 20, and cross-validation error was calculated for each value of K. The cross-validation error was minimized at K = 12.
(TIFF)

Figure S3 Tests for geographically mediated gene flow within and between HOA and MENA populations. The relationship between shared gene identity and inter-population distance is shown within and between HOA and MENA populations for both straight line distances and distance calculated using a waypoint through Egypt (where appropriate).
(EPS)

Figure S4 ROLLOFF and ALDER estimates of admixture times for simulated episodic admixture. We simulated two instances of admixture between CEU and YRI source populations using the 260K SNP dataset and estimated the time of admixture using ROLLOFF and ALDER. The 260K SNP dataset was also pruned down to ~17K SNPs (to match the main dataset used in this study) and admixture dates were estimated from this reduced data as well. Loess lines were fit for each set of estimates: ROLLOFF 260K & 17K, ALDER 260K & 17K (thick solid lines). Overall, there is little difference in the estimates from the 260K and 17K simulated datasets. Previous simulation studies have characterized the ROLLOFF estimate of admixture time in episodic admixture scenarios as intermediate between the true dates of admixture; to better understand what this means, we plotted the weighted arithmetic, geometric, and harmonic means of the simulated admixture times (thin dotted lines). The ROLLOFF and ALDER admixture estimates are always heavily biased towards the more recent admixture. (A) The first admixture event was simulated between 50 and 200 generations ago, followed by a second admixture at 10 generations ago. During both episodes, 10% YRI ancestry was admixed. (B) Same as A, but with 25% YRI ancestry at the first (earlier) admixture. (C & D) Same as A and B, but with the second (more recent) admixture at 30 generations ago.
(EPS)

Figure S5 ROLLOFF and ALDER estimates for simulated ancient admixture followed by more recent admixture. same methods were followed as described for Figure S3, except a broader time range for the first (earlier) admixture was simulated, up to 1150 generations ago. In the first admixture, the YRI contributed 50% ancestry and in the second the YRI contributed 10% ancestry. Overall, there is little difference between the estimates based on the 260K SNP and the 17K SNP datasets. Both ROLLOFF and ALDER estimates are strongly biased towards the more recent admixture episode, with ALDER generally recovering values closer to the more recent admixture time.
(EPS)

Figure S6 Hierarchical population tree models for the non-African inferred ancestry components. Using the non-African IACs from the K = 12 ADMIXTURE analysis, a series of increasingly structured population tree models were constructed, starting from the unstructured tree (A), continuing through intermediately structured trees (B–D) to a final, fully bifurcating

tree model (E). The fit to the data improves significantly with each more structured tree. While the final fully bifurcating is a significantly better fit to the data than any less structured tree, it deviates significantly from a perfect fit to the data.
(EPS)

Table S1 Population samples included in the main 17K, 90K, and 260K SNP datasets.
(XLSX)

Table S2 Results of f3 test for admixture, with estimated bounds of admixture for significantly admixed populations.
(XLSX)

Table S3 Results of D statistic test for admixture.
(XLSX)

Table S4 Results of ALDER test for admixture.
(XLSX)

Table S5 Non-African ancestry proportion estimated for significantly admixed study populations.
(XLSX)

Table S6 Ancestry proportions for each study population from ADMIXTURE ancestry components.
(XLSX)

Table S7 Verification of f3 test results from main (17K SNP) dataset in verification 90K and 260K datasets.
(XLSX)

Table S8 Verification of D statistic test results from main (17K SNP) dataset in verification 90K and 260K datasets.
(XLSX)

Table S9 Verification of ALDER test results from main (17K SNP) dataset in verification 90K and 260K datasets.
(XLSX)

Table S10 Verification of ROLLOFF and ALDER admixture date estimates from main (17K SNP) dataset in verification 90K and 260K datasets.
(XLSX)

Text S1 Affinities of the Ethio-Somali ancestry component. Three different analyses demonstrate that the Ethio-Somali ancestry component that is found at high frequencies in many HOA populations has a non-African origin.
(PDF)

Acknowledgments

We are grateful to all project participants who provided samples to this study. Thanks are due to Kirstin Sterner, Nelson Ting, and Darcy Westphal for discussion and comments.

Author Contributions

Conceived and designed the experiments: JAH RLR. Performed the experiments: JAH CJM RLR. Analyzed the data: JAH RLR. Contributed reagents/materials/analysis tools: CJM AAM. Wrote the paper: JAH RLR.

References

1. Howells WW (1976) Explaining modern man: Evolutionists Versus Migrationists. J Hum Evol 5: 477–495. doi:10.1016/0047-2484(76)90088-9.
2. Stringer CB, Andrews P (1988) Genetic and fossil evidence for the origin of modern humans. Science 239: 1263–1268. doi:10.1126/science.3125610.
3. Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, et al. (2002) Genetic structure of human populations. Science 298: 2381–2385. doi:10.1126/science.1078311.
4. Mitchell P (2005) African Connections: Archaeological Perspectives on Africa and the Wider World. Walnut Creek, CA: AltaMira Press.
5. Templeton AR (2005) Haplotype trees and modern human origins. Yrbk Phys Anthropol 48: 33–59. doi:10.1002/ajpa.20351.
6. Templeton AR (2007) Population biology and population genetics of Pleistocene Hominins. In: Henke W, Tattersall I, editors. Handbook of Paleoanthropology. Berlin: Springer-Verlag, Vol. 3. pp. 1825–1859.
7. Templeton AR (2013) Biological races in humans. Stud Hist Phil Biol Biomed Sci 44: 262–271. doi:10.1016/j.shpsc.2013.04.010.
8. Diamond J, Bellwood P (2003) Farmers and their languages: the first expansions. Science 300: 597–603. doi:10.1126/science.1078208.
9. Bellwood P, Oxenham M (2008) The Expansions of Farming Societies and the Role of the Neolithic Demographic Transition. In: Bocquet-Appel J-P, Bar-Yosef O, editors. The Neolithic Demographic Transition and its Consequences. New York: Springer. pp. 13–34.
10. Price TD, Bar-Yosef O (2011) The Origins of Agriculture: New Data, New Ideas. Curr Anthropol 52: S163–S174. doi:10.1086/659964.
11. Levine DM (2000) Greater Ethiopia: The Evolution of a Multiethnic Society. Second Edition. Chicago: University of Chicago Press.
12. Cavalli-Sforza LLL, Menozzi P, Piazza A (1994) History And Geography Of Human Genes. Princeton: Princeton University Press.
13. Passarino G, Semino O, Quintana-Murci L, Excoffier L, Hammer MF, et al. (1998) Different genetic components in the Ethiopian population, identified by mtDNA and Y-chromosome polymorphisms. Am J Hum Genet 62: 420–434. doi:10.1086/301702.
14. Ehret C (2002) The Civilizations of Africa: A History to 1800. Charlottesville: University of Virginia Press.
15. Non AL, Al-Meeri A, Raaum RL, Sanchez LF, Mulligan CJ (2011) Mitochondrial DNA reveals distinct evolutionary histories for Jewish populations in Yemen and Ethiopia. Am J Phys Anthropol 144: 1–10. doi:10.1002/ajpa.21360.
16. Pagani L, Kivisild T, Tarekegn A, Ekong R, Plaster C, et al. (2012) Ethiopian Genetic Diversity Reveals Linguistic Stratification and Complex Influences on the Ethiopian Gene Pool. Am J Hum Genet 91: 83–96. doi:10.1016/j.ajhg.2012.05.015.
17. Pickrell JK, Patterson N, Loh P-R, Lipson M, Berger B, et al. (2014) Ancient west Eurasian ancestry in southern and eastern Africa. Proc Natl Acad Sci USA 111: 2632–2637. doi:10.1073/pnas.1313787111.
18. Kitchen A, Ehret C, Assefa S, Mulligan CJ (2009) Bayesian phylogenetic analysis of Semitic languages identifies an Early Bronze Age origin of Semitic in the Near East. Proc Roy Soc B Biol Sci 276: 2703–2710. doi:10.1098/rspb.2009.0408.
19. Cerulli E (1960) Punti di vista sulla storia dell'Etiopia. Atti del Convegno Internazionale di Studi Etiopici, Roma 1959. Rome: Accademia Nazionale dei Lincei. pp. 5–27.
20. Bent T (1893) The Sacred City of the Ethiopians. London: Longmans, Green.
21. Conti Rossini C (1928) Storia D'Etiopia. Bergamo: Istituto Italiano d'Arti Grafiche.
22. Ullendorff E (1973) The Ethiopians. London: Oxford University Press.
23. Robin C, de Maigret A (1998) Le grand temple de Yèha (Tigray, Éthiopie) après la première campagne de fouilles de la mission française (1998). CR Acad Inscr Belle 142: 737–798.
24. Phillipson DW (2009) The First Millennium BC in the Highlands of Northern Ethiopia and South–Central Eritrea: A Reassessment of Cultural and Political Development. Afr Archaeol Rev 26: 257–274. doi:10.1007/s10437-009-9064-2.
25. Fattovich R (2010) The Development of Ancient States in the Northern Horn of Africa, c. 3000 BC–AD 1000: An Archaeological Outline. J World Prehist 23: 145–175. doi:10.1007/s10963-010-9035-1.
26. Manzo A (2009) *Capra nubiana* in *Berbere* Sauce? Afr Archaeol Rev 26: 291–303. doi:10.1007/s10437-009-9066-0.
27. Fattovich R (2009) Reconsidering Yeha, c. 800–400 BC. Afr Archaeol Rev 26: 275–290. doi:10.1007/s10437-009-9063-3.
28. Curtis MC (2009) Relating the Ancient Ona Culture to the Wider Northern Horn: Discerning Patterns and Problems in the Archaeology of the First Millennium BC. Afr Archaeol Rev 26: 327–350. doi:10.1007/s10437-009-9062-4.
29. Schmidt PR (2009) Variability in Eritrea and the Archaeology of the Northern Horn During the First Millennium BC: Subsistence, Ritual, and Gold Production. Afr Archaeol Rev 26: 305–325. doi:10.1007/s10437-009-9061-5.
30. Boivin N, Blench R, Fuller DQ (2010) Archaeological, Linguistic and Historical Sources on Ancient Seafaring: A Multidisciplinary Approach to the Study of Early Maritime Contact and Exchange in the Arabian Peninsula. In: Petraglia MD, Rose JI, editors. The Evolution of Human Populations in Arabia. New York: Springer. pp. 251–278.
31. Khalidi L (2010) Holocene Obsidian Exchange in the Red Sea Region. In: Petraglia MD, Rose JI, editors. The Evolution of Human Populations in Arabia. New York: Springer. pp. 279–291.
32. Quintana-Murci L, Semino O, Bandelt H-J, Passarino G, McElreavey K, et al. (1999) Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. Nat Genet 23: 437–441. doi:10.1038/70550.
33. Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, et al. (2004) Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. Am J Hum Genet 75: 752–770. doi:10.1086/425161.

34. Olivieri A, Achilli A, Pala M, Battaglia V, Fornarino S, et al. (2006) The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa. Science 314: 1767–1770. doi:10.1126/science.1135566.
35. González AM, Larruga JM, Abu-Amero KK, Shi Y, Pestano J, et al. (2007) Mitochondrial lineage M1 traces an early human backflow to Africa. BMC Genomics 8: 223. doi:10.1186/1471-2164-8-223.
36. Pennarun E, Kivisild T, Metspalu E, Metspalu M, Reisberg T, et al. (2012) Divorcing the Late Upper Palaeolithic demographic histories of mtDNA haplogroups M1 and U6 in Africa. BMC Evol Biol 12: 234. doi:10.1186/1471-2148-12-234.
37. Fernandes V, Alshamali F, Alves M, Costa MD, Pereira JB, et al. (2012) The Arabian Cradle: Mitochondrial Relicts of the First Steps along the Southern Route out of Africa. Am J Hum Genet 90: 347–355. doi:10.1016/j.ajhg.2011.12.010.
38. Černý V, Mulligan CJ, Fernandes V, Silva NM, Alshamali F, et al. (2011) Internal Diversification of Mitochondrial Haplogroup R0a Reveals Post-Last Glacial Maximum Demographic Expansions in South Arabia. Mol Biol Evol 28: 71–78. doi:10.1093/molbev/msq178.
39. Musilová E, Fernandes V, Silva NM, Soares P, Alshamali F, et al. (2011) 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 145: 592–598. doi:10.1002/ajpa.21522.
40. Cruciani F, Fratta RL, Santolamazza P, Sellitto D, Pascone R, et al. (2004) Phylogeographic analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and out of Africa. Am J Hum Genet 74: 1014–1022. doi:10.1086/386294.
41. Cruciani F, Fratta RL, Trombetta B, Santolamazza P, Sellitto D, et al. (2007) Tracing past human male movements in northern/eastern Africa and western Eurasia: new clues from Y-chromosomal haplogroups E-M78 and J-M12. Mol Biol Evol 24: 1300–1311. doi:10.1093/molbev/msm049.
42. Arredi B, Poloni ES, Paracchini S, Zerjal T, Fathallah DM, et al. (2004) A predominantly neolithic origin for Y-chromosomal DNA variation in North Africa. Am J Hum Genet 75: 338–345. doi:10.1086/423147.
43. Henn BM, Botigué LR, Gravel S, Wang W, Brisbin A, et al. (2012) Genomic Ancestry of North Africans Supports Back-to-Africa Migrations. PLoS Genet 8: e1002397. doi:10.1371/journal.pgen.1002397.
44. Sanchez JJ, Hallenberg C, Børsting C, Hernandez A, Morling N (2005) High frequencies of Y chromosome lineages characterized by E3b1, DYS19-11, DYS392-12 in Somali males. Eur J Hum Genet 13: 856–866. doi:10.1038/sj.ejhg.5201390.
45. Mendez FL, Karafet TM, Krahn T, Ostrer H, Soodyall H, et al. (2011) Increased Resolution of Y Chromosome Haplogroup T Defines Relationships among Populations of the Near East, Europe, and Africa. Hum Biol 83: 39–53. doi:10.3378/027.083.0103.
46. Plaster CA (2011) Variation in Y chromosome, mitochondrial DNA and labels of identity on Ethiopia [Doctoral]. UCL (University College London). Available: http://discovery.ucl.ac.uk/1331901/. Accessed 4 September 2013.
47. Luis JR, Rowold DJ, Regueiro M, Caeiro B, Cinnioğlu C, et al. (2004) The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. Am J Hum Genet 74: 532–544. doi:10.1086/382286.
48. Behar DM, Yunusbayev B, Metspalu M, Metspalu E, Rosset S, et al. (2010) The genome-wide structure of the Jewish people. Nature 466: 238–242. doi:10.1038/nature09103.
49. Bryc K, Auton A, Nelson MR, Oksenberg JR, Hauser SL, et al. (2010) Genome-wide patterns of population structure and admixture in West Africans and African Americans. Proc Natl Acad Sci USA 107: 786–791. doi:10.1073/pnas.0909559107.
50. Hunter-Zinck H, Musharoff S, Salit J, Al-Ali KA, Chouchane L, et al. (2010) Population Genetic Structure of the People of Qatar. Am J Hum Genet 87: 17–25. doi:10.1016/j.ajhg.2010.05.018.
51. Schlebusch CM, Skoglund P, Sjödin P, Gattepaille LM, Hernandez D, et al. (2012) Genomic Variation in Seven Khoe-San Groups Reveals Adaptation and Complex African History. Science 338: 374–379. doi:10.1126/science.1227721.
52. Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, et al. (2010) Integrating common and rare genetic variation in diverse human populations. Nature 467: 52–58. doi:10.1038/nature09298.
53. Li J, Absher D, Tang H, Southwick A, Casto A, et al. (2008) Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation. Science 319: 1100–1104. doi:10.1126/science.1153717.
54. Rosenberg NA, Mahajan S, Ramachandran S, Zhao C, Pritchard JK, et al. (2005) Clines, clusters, and the effect of study design on the inference of human population structure. PLoS Genet 1: e70. doi:10.1371/journal.pgen.0010070.
55. Patterson NJ, Price AL, Reich D (2006) Population structure and eigenanalysis. PLoS Genet 2: e190. doi:10.1371/journal.pgen.0020190.
56. Ma J, Amos CI (2012) Principal Components Analysis of Population Admixture. PLoS ONE 7: e40115. doi:10.1371/journal.pone.0040115.
57. Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19: 1655–1664. doi:10.1101/gr.094052.109.
58. Alexander D, Lange K (2011) Enhancements to the ADMIXTURE Algorithm for Individual Ancestry Estimation. BMC Bioinformatics 12: 246. doi:10.1186/1471-2105-12-246.
59. Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, et al. (2009) The Genetic Structure and History of Africans and African Americans. Science 324: 1035–1044. doi:10.1126/science.1172257.
60. Haber M, Gauguier D, Youhanna S, Patterson N, Moorjani P, et al. (2013) Genome-Wide Diversity in the Levant Reveals Recent Structuring by Culture. PLoS Genet 9: e1003316. doi:10.1371/journal.pgen.1003316.
61. Sánchez-Quinto F, Botigué LR, Civit S, Arenas C, Ávila-Arcos MC, et al. (2012) North African Populations Carry the Signature of Admixture with Neandertals. PLoS ONE 7: e47765. doi:10.1371/journal.pone.0047765.
62. Patterson NJ, Moorjani P, Luo Y, Mallick S, Rohland N, et al. (2012) Ancient Admixture in Human History. Genetics 192: 1065–1093. doi:10.1534/genetics.112.145037.
63. Loh P-R, Lipson M, Patterson N, Moorjani P, Pickrell JK, et al. (2013) Inferring Admixture Histories of Human Populations Using Linkage Disequilibrium. Genetics 193: 1233–1254. doi:10.1534/genetics.112.147330.
64. Lawson DJ, Hellenthal G, Myers S, Falush D (2012) Inference of Population Structure using Dense Haplotype Data. PLoS Genet 8: e1002453. doi:10.1371/journal.pgen.1002453.
65. Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131: 479–491.
66. Urbanek M, Goldman D, Long JC (1996) The apportionment of dinucleotide repeat diversity in Native Americans and Europeans: a new approach to measuring gene identity reveals asymmetric patterns of divergence. Mol Biol Evol 13: 943–953.
67. Long JC, Kittles RA (2003) Human genetic diversity and the nonexistence of biological races. Hum Biol 75: 449–471. doi:10.3378/027.081.0621.
68. Lewis MP, Simons GF, Fennig CD (2013) Ethnologue. Ethnologue: Languages of the World, 17th Edition. Available: http://www.ethnologue.com/. Accessed 17 December 2013.
69. Dimmendaal GJ (2011) Historical Linguistics and the Comparative Study of African Languages. Philadelphia, PA: John Benjamins Publishing.
70. Moorjani P, Patterson N, Hirschhorn JN, Keinan A, Hao L, et al. (2011) The History of African Gene Flow into Southern Europeans, Levantines, and Jews. PLoS Genet 7: e1001373. doi:10.1371/journal.pgen.1001373.
71. Petersen DC, Libiger O, Tindall EA, Hardie R-A, Hannick LI, et al. (2013) Complex Patterns of Genomic Admixture within Southern Africa. PLoS Genet 9: e1003309. doi:10.1371/journal.pgen.1003309.
72. Verdu P, Rosenberg NA (2011) A General Mechanistic Model for Admixture Histories of Hybrid Populations. Genetics 189: 1413–1426. doi:10.1534/genetics.111.132787.
73. Holsinger KE, Weir BS (2009) Genetics in geographically structured populations: defining, estimating and interpreting F_{ST}. Nat Rev Genet 10: 639–650. doi:10.1038/nrg2611.
74. Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. PLoS Genet 5: e1000695. doi:10.1371/journal.pgen.1000695.
75. Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, et al. (2007) Convergent adaptation of human lactase persistence in Africa and Europe. Nat Genet 39: 31–40. doi:10.1038/ng1946.
76. Enattah NS, Sahi T, Savilahti E, Terwilliger JD, Peltonen L, et al. (2002) Identification of a variant associated with adult-type hypolactasia. Nat Genet 30: 233–237. doi:10.1038/ng826.
77. Swallow DM (2003) Genetics of Lactase Persistence and Lactose Intolerance. Ann Rev Genet 37: 197–219. doi:10.1146/annurev.genet.37.110801.143820.
78. Imtiaz F, Savilahti E, Sarnesto A, Trabzuni D, Al-Kahtani K, et al. (2007) The T/G 13915 variant upstream of the lactase gene (LCT) is the founder allele of lactase persistence in an urban Saudi population. J Med Genet 44: e89. doi:10.1136/jmg.2007.051631.
79. Enattah NS, Jensen T, Nielsen M, Lewinski R, Kuokkanen M, et al. (2008) Independent Introduction of Two Lactase-Persistence Alleles into Human Populations Reflects Different History of Adaptation to Milk Culture. Am J Hum Genet 82: 57–72. doi:10.1016/j.ajhg.2007.09.012.
80. Ehret C (1979) On the Antiquity of Agriculture in Ethiopia. J Afr Hist 20: 161–177. doi:10.1017/S002185370001700X.
81. Lovejoy PE (2000) Transformations in Slavery: A History of Slavery in Africa. 2nd ed. Cambridge: Cambridge University Press.
82. Hoelzmann P, Gasse F, Dupont LM, Salzmann U, Staubwasser M, et al. (2004) Palaeoenvironmental changes in the arid and sub arid belt (Sahara-Sahel-Arabian Peninsula) from 150 kyr to present. In: Battarbee RW, Gasse F, Stickley CE, editors. Past Climate Variability through Europe and Africa. Developments in Paleoenvironmental Research. Springer Netherlands. pp. 219–256.
83. Hetherington R, Wiebe E, Weaver AJ, Carto SL, Eby M, et al. (2008) Climate, African and Beringian subaerial continental shelves, and migration of early peoples. Quatern Int 183: 83–101. doi:10.1016/j.quaint.2007.06.033.
84. Williams MAJ (2009) Late Pleistocene and Holocene environments in the Nile basin. Global Planet Change 69: 1–15. doi:10.1016/j.gloplacha.2009.07.005.
85. Parker AG (2010) Pleistocene Climate Change in Arabia: Developing a Framework for Hominin Dispersal over the Last 350 ka. In: Petraglia MD, Rose JI, editors. The Evolution of Human Populations in Arabia. New York: Springer. pp. 39–49.

86. Williams MAJ, Williams FM, Duller GAT, Munro RN, El Tom OAM, et al. (2010) Late Quaternary floods and droughts in the Nile valley, Sudan: new evidence from optically stimulated luminescence and AMS radiocarbon dating. Quatern Sci Rev 29: 1116–1137. doi:10.1016/j.quascirev.2010.02.018.
87. Brandt SA (1986) The Upper Pleistocene and early Holocene prehistory of the Horn of Africa. Afr Archaeol Rev 4: 41–82. doi:10.1007/BF01117035.
88. Brandt SA (1997) Horn of Africa: History of Archaeology. In: Vogel JO, editor. Encyclopedia of Precolonial Africa. Walnut Creek, CA: AltaMira Press. pp. 69–75.
89. Brandt SA (1988) Early Holocene Mortuary Practices and Hunter-Gatherer Adaptations in Southern Somalia. World Archaeol 20: 40–56. doi:10.2307/124524.
90. Mayer DEB-Y, Beyin A (2009) Late Stone Age Shell Middens on the Red Sea Coast of Eritrea. J Island Coastal Archaeol 4: 108–124. doi:10.1080/15564890802662171.
91. Beyin A (2010) Use-wear analysis of obsidian artifacts from Later Stone Age shell midden sites on the Red Sea Coast of Eritrea, with experimental results. J Archaeol Sci 37: 1543–1556. doi:10.1016/j.jas.2010.01.015.
92. Beyin (2011) Early to Middle Holocene human adaptations on the Buri Peninsula and Gulf of Zula, coastal lowlands of Eritrea. Azania 46: 123–140. doi:10.1080/0067270X.2011.580139.
93. Brace CL, Seguchi N, Quintyn CB, Fox SC, Nelson AR, et al. (2006) The questionable contribution of the Neolithic and the Bronze Age to European craniofacial form. Proc Natl Acad Sci USA 103: 242–247. doi:10.1073/pnas.0509801102.
94. Irish JD, Konigsberg L (2007) The Ancient Inhabitants of Jebel Moya Redux: Measures of Population Affinity Based on Dental Morphology. Int J Osteoarchaeol 17: 138–156. doi:10.1002/oa.868.
95. Ricaut FX, Waelkens M (2008) Cranial Discrete Traits in a Byzantine Population and Eastern Mediterranean Population Movements. Hum Biol 80: 535–564. doi:10.3378/1534-6617-80.5.535.
96. Sellers TA (2008) The Influence of Subsistence Shift on Dental Reductions: A Comparison of Prehistoric and Modern Nubian and Somalian Dental Samples [M.A. Thesis]. Cincinatti, OH: University of Cincinnati.
97. Ferembach D (1985) On the origin of the iberomaurusians (Upper palaeolithic: North Africa). A new hypothesis. J Hum Evol 14: 393–397. doi:10.1016/S0047-2484(85)80047-6.
98. Irish JD (2000) The Iberomaurusian enigma: North African progenitor or dead end? J Hum Evol 39: 393–410. doi:10.1006/jhev.2000.0430.
99. Rando JC, Pinto F, González AM, Hernández M, Larruga JM, et al. (1998) Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern, and sub-Saharan populations. Ann Hum Genet 62: 531–550. doi:10.1046/j.1469-1809.1998.6260531.x.
100. Maca-Meyer N, González AM, Larruga JM, Flores C, Cabrera VM (2001) Major genomic mitochondrial lineages delineate early human expansions. BMC Genet 2: 13. doi:10.1186/1471-2156-2-13.
101. Maca-Meyer N, González AM, Pestano J, Flores C, Larruga JM, et al. (2003) Mitochondrial DNA transit between West Asia and North Africa inferred from U6 phylogeography. BMC Genet 4: 15. doi:10.1186/1471-2156-4-15.
102. Plaza S, Calafell F, Helal A, Bouzerna N, Lefranc G, et al. (2003) Joining the Pillars of Hercules: mtDNA Sequences Show Multidirectional Gene Flow in the Western Mediterranean. Ann Hum Genet 67: 312–328. doi:10.1046/j.1469-1809.2003.00039.x.
103. Cherni L, Fernandes V, Pereira JB, Costa MD, Goios A, et al. (2009) Post-last glacial maximum expansion from Iberia to North Africa revealed by fine characterization of mtDNA H haplogroup in Tunisia. Am J Phys Anthropol 139: 253–260. doi:10.1002/ajpa.20979.
104. Ennafaa H, Cabrera VM, Abu-Amero KK, González AM, Amor MB, et al. (2009) Mitochondrial DNA haplogroup H structure in North Africa. BMC Genet 10: 8. doi:10.1186/1471-2156-10-8.
105. Fadhlaoui-Zid K, Rodríguez-Botigué L, Naoui N, Benammar-Elgaaied A, Calafell F, et al. (2011) Mitochondrial DNA structure in North Africa reveals a genetic discontinuity in the Nile Valley. Am J Phys Anthropol 145: 107–117. doi:10.1002/ajpa.21472.
106. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945–959.
107. Greenberg JH (1971) African languages. In: Dil A, editor. Language, Culture, and Communication. Stanford: Stanford University Press. pp. 126–136.
108. Ehret C, Keita SOY, Newman P (2004) The Origins of Afroasiatic. Science 306: 1680–1680. doi:10.1126/science.306.5702.1680c.
109. Militarev A (2000) Towards the chronology of Afrasian (Afroasiatic) and its daughter families. In: Renfrew C, McMahon A, Trask L, editors. Time Depth in Historical Linguistics. Cambridge: The McDonald Institute for Archaeological Research, Vol. 1. pp. 267–307.
110. Militarev A (2002) The Prehistory of a Dispersal: the Proto-Afrasian (Afroasiatic) Farming Lexicon. In: Bellwood P, Renfrew C, editors. Examining the Farming/Language Dispersal Hypothesis. Cambridge: The McDonald Institute for Archaeological Research. pp. 135–150.
111. McCall DF (1998) The Afroasiatic Language Phylum: African in Origin, or Asian? Curr Anthropol 39: 139–144. doi:10.1086/204702.
112. Hodgson JA, Mulligan CJ, Al-Meeri A, Raaum RL (2014) Data from: Early Back-to-African Migration into the Horn of Africa. Dryad Digital Repository. doi:10.5061/dryad.d9s74.
113. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, et al. (2007) PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am J Hum Genet 81: 559–575. doi:10.1086/519795.
114. Rosenberg NA (2006) Standardized Subsets of the HGDP-CEPH Human Genome Diversity Cell Line Panel, Accounting for Atypical and Duplicated Samples and Pairs of Close Relatives. Ann Hum Genet 70: 841–847. doi:10.1111/j.1469-1809.2006.00285.x.
115. Pemberton TJ, Wang C, Li JZ, Rosenberg NA (2010) Inference of Unexpected Genetic Relatedness among Individuals in HapMap Phase III. Am J Hum Genet 87: 457–464. doi:10.1016/j.ajhg.2010.08.014.
116. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, et al. (2010) Robust relationship inference in genome-wide association studies. Bioinformatics 26: 2867–2873. doi:10.1093/bioinformatics/btq559.
117. Matise TC, Chen F, Chen W, Vega FMDL, Hansen M, et al. (2007) A second-generation combined linkage–physical map of the human genome. Genome Res 17: 1783–1786. doi:10.1101/gr.7156307.
118. François O (n.d.) How to display admixture coefficients (Q matrix) spatially? Available: http://membres-timc.imag.fr/Olivier.Francois/admix_display.html. Accessed 29 January 2013.
119. Venables WN, Ripley BD (2002) Modern applied statistics with S. Fourth Edition. New York: Springer.
120. Becker RA, Wilks AR, Brownrigg R, Minka TP (2012) maps: Draw Geographical Maps. Available: http://CRAN.R-project.org/package = maps.
121. Lewin-Koh NJ, Bivand R, Pebesma EJ, Archer E, Baddeley A, et al. (2012) maptools: Tools for reading and handling spatial objects. Available: http://CRAN.R-project.org/package = maptools.
122. R Development Core Team (2013) R: A Language and Environment for Statistical Computing. Vienna, Austria. Available: http://www.R-project.org.
123. Delaneau O, Zagury J-F, Marchini J (2013) Improved whole-chromosome phasing for disease and population genetic studies. Nat Meth 10: 5–6. doi:10.1038/nmeth.2307.
124. Dray S, Dufour A-B (2007) The ade4 Package: Implementing the Duality Diagram for Ecologists. J Stat Softw 22: 1–20.
125. Tremblay M, Vézina H (2000) New Estimates of Intergenerational Time Intervals for the Calculation of Age and Origins of Mutations. Am J Hum Genet 66: 651–658. doi:10.1086/302770.
126. Helgason A, Hrafnkelsson B, Gulcher JR, Ward R, Stefánsson K (2003) A Populationwide Coalescent Analysis of Icelandic Matrilineal and Patrilineal Genealogies: Evidence for a Faster Evolutionary Rate of mtDNA Lineages than Y Chromosomes. Am J Hum Genet 72: 1370–1388. doi:10.1086/375453.
127. Matsumura S, Forster P (2008) Generation time and effective population size in Polar Eskimos. Proc Roy Soc B Biol Sci 275: 1501–1508. doi:10.1098/rspb.2007.1724.
128. Price AL, Tandon A, Patterson N, Barnes KC, Rafaels N, et al. (2009) Sensitive Detection of Chromosomal Segments of Distinct Ancestry in Admixed Populations. PLoS Genet 5: e1000519. doi:10.1371/journal.pgen.1000519.
129. Scheet P, Stephens M (2006) A Fast and Flexible Statistical Model for Large-Scale Population Genotype Data: Applications to Inferring Missing Genotypes and Haplotypic Phase. Am J Hum Genet 78: 629–644. doi:10.1086/502802.