

Chad Genetic Diversity Reveals an African History Marked by Multiple Holocene Eurasian Migrations

Marc Haber,^{1,*} Massimo Mezzavilla,^{1,2} Anders Bergström,¹ Javier Prado-Martinez,¹ Pille Hallast,^{1,3} Riyadh Saif-Ali,⁴ Molham Al-Habori,⁴ George Dedoussis,⁵ Eleftheria Zeggini,¹ Jason Blue-Smith,^{6,10} R. Spencer Wells,⁷ Yali Xue,¹ Pierre A. Zalloua,^{8,9} and Chris Tyler-Smith^{1,*}

Understanding human genetic diversity in Africa is important for interpreting the evolution of all humans, yet vast regions in Africa, such as Chad, remain genetically poorly investigated. Here, we use genotype data from 480 samples from Chad, the Near East, and southern Europe, as well as whole-genome sequencing from 19 of them, to show that many populations today derive their genomes from ancient African-Eurasian admixtures. We found evidence of early Eurasian backflow to Africa in people speaking the unclassified isolate Laal language in southern Chad and estimate from linkage-disequilibrium decay that this occurred 4,750–7,200 years ago. It brought to Africa a Y chromosome lineage (R1b-V88) whose closest relatives are widespread in present-day Eurasia; we estimate from sequence data that the Chad R1b-V88 Y chromosomes coalesced 5,700–7,300 years ago. This migration could thus have originated among Near Eastern farmers during the African Humid Period. We also found that the previously documented Eurasian backflow into Africa, which occurred ~3,000 years ago and was thought to be mostly limited to East Africa, had a more westward impact affecting populations in northern Chad, such as the Toubou, who have 20%–30% Eurasian ancestry today. We observed a decline in heterozygosity in admixed Africans and found that the Eurasian admixture can bias inferences on their coalescent history and confound genetic signals from adaptation and archaic introgression.

Introduction

African genetic diversity is still incompletely understood, and vast regions in Africa remain genetically undocumented. Chad, for example, makes up ~5% of Africa's surface area, and its central location, connecting sub-Saharan Africa with North and East Africa, positions it to play an important role as a crossroad or barrier to human migrations. However, Chad has been little studied at a whole-genome level, and its position within African genetic diversity is not well known. With 200 ethnic groups and more than 120 indigenous languages and dialects, Chad has extensive ethnolinguistic diversity.¹ It has been suggested that this diversity can be attributed to Lake Chad, which has attracted human populations to its fertile surroundings since prehistoric times, especially after the progressive desiccation of the Sahara starting ~7,000 years ago (ya).^{2,3}

Important questions about Africa's ethnic diversity are the relationships among the different groups and the relationships between cultural groups and existing genetic structures. In the present study, we analyzed four Chadian populations with different ethnicities, languages, and modes of subsistence. Our samples are likely to capture recent genetic signals of migration and mixing and also have the potential to show ancestral genomic relationships that are shared among Chadians and other populations. An additional major question relates to the prehistoric

Eurasian migrations to Africa: what was the extent of these migrations, how have they affected African genetic diversity, and what present-day populations harbor genetic signals from the ancient migrating Eurasians? We have previously reported evidence of gene flow from the Near East to East Africa ~3,000 ya, as well as subsequent selection in Ethiopians on non-African-derived alleles related to light skin pigmentation.⁴ A recent attempt to quantify the extent of such backflow into Africa more generally, by using ancient DNA (aDNA), suggested that the impact of the Eurasian migration was mostly limited to East Africa.⁵ However, previous studies using mitochondrial DNA and the Y chromosome in populations from the Chad Basin found some with an East African⁶ or Mediterranean and Eurasian influence,^{7,8} and analysis based on genome-wide data⁹ found a non-African component (suggested to be from East Africa) in central Sahelian populations. Thus, studying diverse Chadian populations on a whole-genome level presents an opportunity to shed more light on the history of African-Eurasian mixtures, including whether or not selection after admixture is a widespread phenomenon in Africa and how the historical events in Chad are related to events that have occurred elsewhere in Africa and the Near East.

In this work, we present a genetic dataset of 480 Chadian, Near Eastern, and European individuals genotyped at 2.5 million SNPs, in addition to high-coverage

¹Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, UK; ²Institute for Maternal and Child Health, IRCCS Burlo Garofolo, University of Trieste, 34137 Trieste, Italy; ³Institute of Molecular and Cell Biology, University of Tartu, Tartu 51010, Estonia; ⁴Department of Biochemistry and Molecular Biology, Faculty of Medicine and Health Sciences, Sana'a University, Sana'a 19065, Yemen; ⁵Department of Nutrition and Dietetics, Harokopio University Athens, Athens 17671, Greece; ⁶National Geographic Society, Washington, DC 20036, USA; ⁷Department of Integrative Biology, University of Texas at Austin, Austin, TX 78712, USA; ⁸Lebanese American University, Chouran, Beirut 1102 2801, Lebanon; ⁹Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA

¹⁰Present address: Karius, Inc., 1505A Adams Drive, Menlo Park, CA 94025, USA

*Correspondence: mh25@sanger.ac.uk (M.H.), cts@sanger.ac.uk (C.T.-S.)

<http://dx.doi.org/10.1016/j.ajhg.2016.10.012>

© 2016 The Author(s). This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

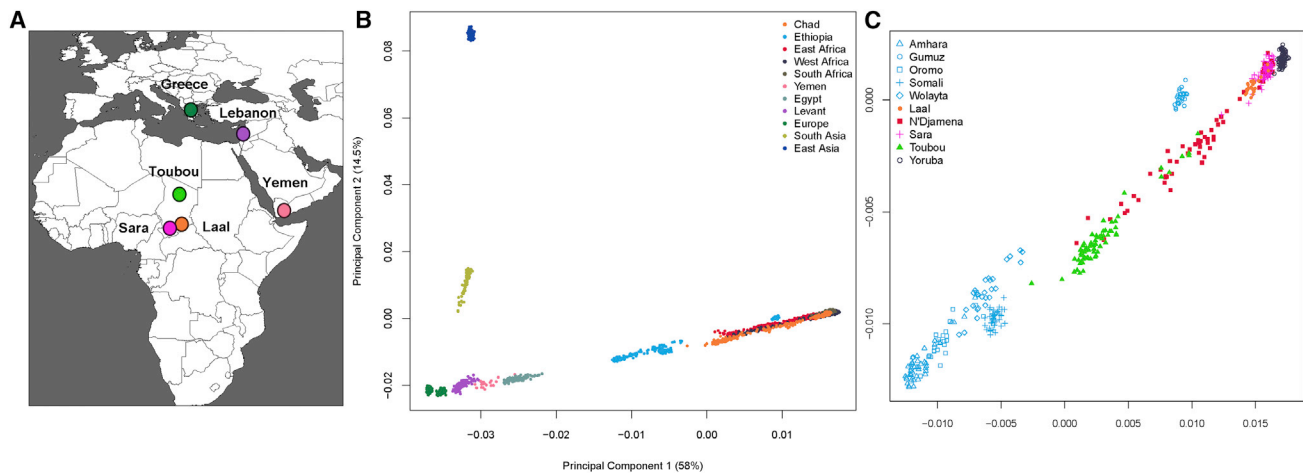


Figure 1. Population Locations and Genetic Structure

(A) The map shows the location of newly genotyped or sequenced populations.

(B) PCA of worldwide populations shows that Near Easterners and East Africans are intermediate to Eurasians and sub-Saharan Africans on PC1. Chad populations are close to sub-Saharan Africans and have some samples drawn toward Ethiopians.

(C) Magnification of the African PCA shows different affinities of the Chad populations to other Africans: the Toubou cluster close to Ethiopians, whereas the Sara and Laal speakers are close to the Yoruba. The mixed samples from N'Djamena, the capital, are intermediate to the Toubou, Sara, and Laal speakers.

whole-genome sequences from 19 of these individuals. From Chad, we studied (1) the Toubou, who are nomads from northern Chad and speak a Nilo-Saharan language; (2) the Sara, who are a sedentary population from southern Chad and also speak a Nilo-Saharan language; (3) the Laal speakers, a population of just ~750 individuals who speak an unclassified language isolate and live in southern Chad; and (4) an urban population from the capital city of N'Djamena. In addition to the Chadians, we included Greek, Lebanese, and Yemen samples whose location and history suggest that they might be informative about early African-Eurasian migrations. We used this dataset to advance our understanding of human genetic diversity in Africa and neighboring regions by focusing on population migration and mixing and how the admixture process has shaped present-day genetic variation.

Subjects and Methods

Samples and Data

Samples were collected from Chad (238), Lebanon (126), Greece (96), and Yemen (20) (Figure 1A); details can be found in Table S1. All samples (except for those from Greece) were genotyped with the Illumina HumanOmni2.5-8 BeadChip, which covers ~2.5 million SNPs. Greek genotype information for the 2.5 million sites was extracted from sequence data (E.Z., unpublished data) and merged with array data from other populations. In addition, 19 samples (Chad [11], Greece [4], and Lebanon [4]) were whole-genome sequenced at >30× depth with Illumina HiSeq X Ten or HiSeq 2500 technology. Genotyping and sequencing were completed at the Wellcome Trust Sanger Institute. Informed consent was obtained from the studied subjects, and the use of the samples in genetic studies was approved by the Human Materials and Data Management Committee at the Wellcome Trust Sanger

Institute (approval numbers 09/056 and 14/072) and by the institutional review board (number SMPZ121307-02) of the Lebanese American University.

The genotyping data were merged with data from the African Genome Variation Project,¹⁰ the 1000 Genomes Project,¹¹ and Pagani et al.,¹² resulting in a combined dataset of ~1.1 million SNPs in 2,453 samples. Analyses including ancient genomes involved merging the panel described above with the Haak et al. dataset,¹³ resulting in ~90,000 SNPs in common. Comparative whole-genome sequences were obtained from Pagani et al.¹² and Complete Genomics.¹⁴

Genotype data were processed with PLINK.¹⁵ The SNP genotype success rate required was set to 99%, whereas SNPs with a minor allele frequency < 0.001 or Hardy-Weinberg p value < 0.000001 were removed. Genotypes from sequence data were called with SAMtools v.1.2¹⁶ and BCFtools v.1.2 with the command “samtools mpileup -q 20 -Q 20 -C 50 | bcftools call -c -V indels.” Concordance with array genotypes had a rate of 0.999. Phasing was carried out with SHAPEIT¹⁷ with 1000 Genomes Project phase 3 haplotypes¹⁸ as a reference panel.

Population Structure and Gene Flow

Principal components were computed with EIGENSOFT v.4.2.¹⁹ Effective population size and rates of gene flow were inferred by the multiple sequentially Markovian coalescent (MSMC) approach²⁰ with four high-coverage phased genomes from each population. We assumed a generation time of 30 years and a mutation rate of 1.25×10^{-8} mutations per nucleotide per generation. Admixture masks to identify African and Eurasian segments within mixed high-coverage genomes were generated with PCAdmix²¹ including two ancestral populations based on the 1000 Genomes Project YRI (Yoruba in Ibadan, Nigeria) and CEU (Utah residents with northern and western European ancestry from the CEPH collection) populations. 1 cM windows with a posterior probability of >0.9 for the most likely ancestral state were collected and used for creating African and Eurasian masks.

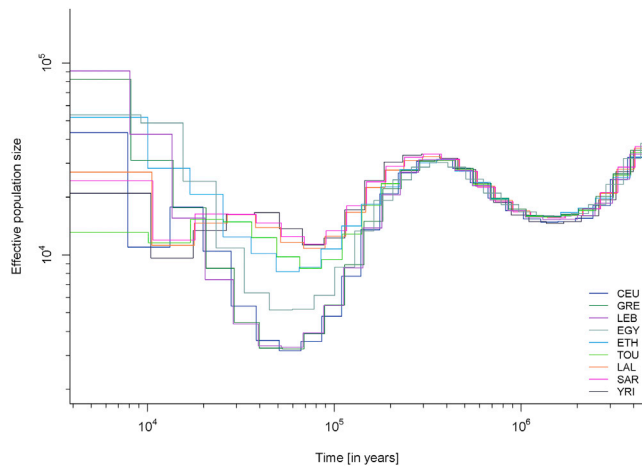


Figure 2. Population-Size Estimates from Whole-Genome Sequences

Population size was inferred by MSMC analysis with four haplotypes from each population. Eurasian populations had a distinctive bottleneck at the time of their exodus from Africa ~60,000 ya. Compared to other Africans, admixed Africans (from a Eurasian gene flow), such as Egyptians, Ethiopians, and the Toubou, also showed a decline in population size during the same period.

Phylogenetic analysis of whole Y chromosome sequences was carried out as described in Bergström et al.²² Internal node ages were estimated with the rho-statistic²³ and converted to units of years by application of a Y chromosome mutation rate of 0.76×10^{-9} (95% confidence interval [CI] = 0.67×10^{-9} to 0.86×10^{-9}) mutations per site per year.²⁴ Additionally, Y chromosome haplogroups were defined to the highest resolution possible with 636 SNPs from the array data that overlapped International Society of Genetic Genealogy (ISOGG) markers.

Admixture Analysis

Population-mixture signals and proportions were tested with qp3Pop, qpDstat, and qpF4Ratio from the ADMIXTOOLS package.²⁵ Admixture proportions were additionally estimated with ADMIXTURE v.1.3.0.²⁶ ALDER²⁷ and MALDER²⁸ were used to date the time of admixture with all pairs of African-Eurasian populations as references. Significant results with a p value < 0.01 were collected and plotted.

Measure of Heterozygosity and Simulations

Heterozygosity on a per-individual basis was estimated with VCFtools v.0.1.13²⁹ for ~2.17 Gb of the uniquely mappable genome.¹¹ Heterozygosity was also estimated after correction for recent inbreeding via the removal of long runs of homozygosity (>2 Mb). We investigated the effect of gene flow on the observed heterozygosity by using individual-based forward-time simulations implemented in SimuPOP v.1.1.7.³⁰

Selection after Admixture

Evidence for positive selection was tested with the population branch statistic (PBS)³¹ with correction for the long-term effective population size.³² We constructed a tree with the Toubou population branching from the Laal speakers and the Chinese Han as an outgroup. We collected values above the 95th percentile of the PBS distribution and looked for variants previously reported under putative selection in Europeans.

Results

Genetic Structure in Chad Indicates a Complex Admixture History

We performed an initial exploration of our dataset by using principal-component analysis (PCA).¹⁹ The first component (PC1) captured the genetic differentiation between Africans and Eurasians (Figure 1B). Populations such as the Near Easterners and North and East Africans fell between the Europeans and sub-Saharan Africans. The Chadian groups lay near the sub-Saharan Africans: the Sara and Laal speakers clustered tightly with sub-Saharan Africans, such as the Yoruba, whereas the Toubou were somewhat more distant and appeared drawn toward East Africans, such as the Ethiopians. Samples collected from the capital of Chad, N'Djamena, appeared in a central position between the Toubou cluster and the Sara and Laal cluster (Figure 1C). Many individuals from N'Djamena have not reported their ethnicity or have reported a mixed ethnic origin. Therefore, recent mixture could be responsible for their position on the PCA.

We further investigated the genetic variation in Chad by estimating changes in the effective populations size (N_e) over time via the MSMC approach.²⁰ Eurasians and Africans diverged around 60,000–80,000 ya and subsequently had different patterns of population-size changes: in particular, compared with Africans, the Eurasian population experienced a sharp decrease in size ~60,000 ya.²⁰ We observed this expected pattern in most populations in our dataset (Figure 2), but a few stood out: (1) Egyptians had a population bottleneck that was much more pronounced than that of other Africans but not as sharp as that of Eurasian populations; and (2) the Toubou and Ethiopians shared a very similar pattern during the bottleneck: they were close to other Africans but had a somewhat sharper decrease in population size (Figure 2). We would not expect such different fluctuations in population sizes at 60,000 ya in populations who shared a common origin during this period. For example, all Eurasians trace their origin to a population who exited Africa ~60,000 ya, and this is reflected in indistinguishable N_e patterns during this period,^{20,33} which we also observed in the CEU, Greeks, and Lebanese (Figure 2), as expected. A shared pattern of N_e in ancient times was also observed in the Sara, Laal speakers, and other Africans, such as the Yoruba. We suggest that the deviation from the expected N_e pattern in the Toubou is related to extensive admixture history with Eurasians, like the Eurasian admixture seen in Ethiopians, and we explore this possibility directly with admixture tests below.

Multiple Eurasian Admixtures in Africa after 6,000 ya

We have previously reported massive gene flow ~3,000 ya from Eurasians to Ethiopian populations.⁴ Here, we reassess the presence of Eurasian ancestry in Africa by using f_3 statistics²⁵ in the form of $f_3(X; \text{Eurasian}, \text{Yoruba})$, where

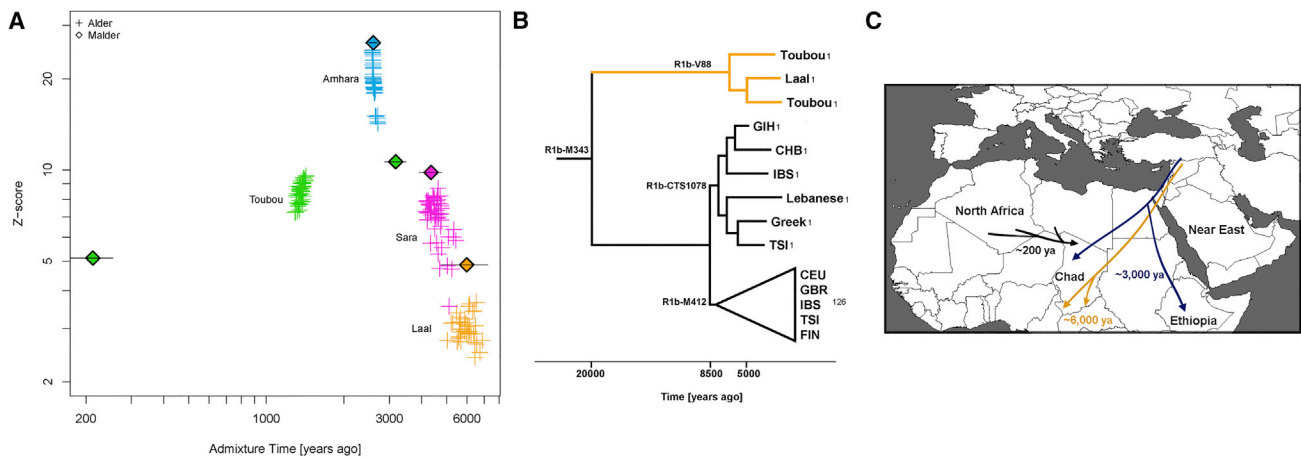


Figure 3. Timing of the Eurasian Admixture in Africa

(A) Crosses represent significant admixture events in the history of the Toubou, Amhara, Sara, and Laal speakers. Time of admixture is estimated from LD by ALDER with all pairs of African-Eurasian populations in our dataset as references. MALDER extends ALDER inference by detecting multiple mixture events, such as in the case of the Toubou population (shown here in green lozenges). (B) A maximum-likelihood tree shows the males belonging to haplogroup R1b in the 1000 Genomes Project and the R1b males in our dataset. The number of samples is shown on each branch tip. We estimate that the Chadian R1b emerged 5,700–7,300 ya, whereas most European R1b haplogroups emerged 7,300–9,400 ya. The African and Eurasian lineages coalesced 17,900–23,000 ya. (C) Putative sources and times of admixture of the Eurasian ancestry in Chad and East Africa.

a negative value with a Z score < -4 indicates that X is a mixture of Africans and Eurasians. We found, as expected, that most Ethiopians are a mixture of Africans and Eurasians. An exception is the Gumuz population, where $f_3(\text{Gumuz}; \text{Eurasian}, \text{Yoruba})$ is always positive. The Gumuz language belongs to the Nilo-Saharan family, which could have isolated the Gumuz from the Afro-Asiatic-speaking Ethiopians. However, we found that the Toubou in Chad, who also speak a Nilo-Saharan language, are a mixture of Africans and Eurasians, making $f_3(\text{Toubou}; \text{Eurasian}, \text{Yoruba})$ always significantly negative. This suggests that the impact of Eurasian migrations today extends beyond East Africa and the Afro-Asiatic-speaking populations. We did not detect significant (Z score < -4) Eurasian admixture in the Sara (Nilo-Saharan language family) or the Laal speakers (unclassified language) with the use of f_3 statistics (lowest Z score for the Sara was > -2.9 ; for the Laal speakers, Z scores were all positive). However, this statistic loses sensitivity with small mixture proportions and post-admixture drift,²⁷ so positive values from the f_3 statistics do not necessarily reflect a complete absence of admixture. We thus further tested for admixture by using ALDER and MALDER, which assess admixture-induced linkage disequilibrium (LD) and can detect small mixture proportions from a substantially diverged reference possibly missed by the f_3 statistic. ALDER detected admixture in the Toubou, Sara, and Laal speakers (Table S2). MALDER, which has the potential to determine whether or not the admixture LD in the population is best represented as the result of one or multiple mixtures, showed that two mixture events had occurred in the Toubou (Figure 3A; Table S3). The first event occurred 2,850–3,500 ya (Z score = 11), a time close to the date of mixture in East Africans 2,500–2,700 ya (Z score = 26).

The second mixture event occurred much more recently at 170–260 ya (Z score = 5). In southern Chad, we detected mixture events that were more ancient than those in the north. Mixture occurred 3,900–4,800 ya (Z score = 10) in the Sara and 4,750–7,200 ya (Z score = 5) in the Laal speakers (Figure 3A). These time estimates overlap, and we interpret them as signals from the same admixture event, whose time in the distant past was estimated more reliably in the Laal speakers because they carry more Eurasian ancestry (1.25%–4.5%) than the Sara (0.3%–2%) (see estimates of admixture proportions below), even though the Sara have smaller standard errors because of their larger sample size. In particular, we suggest that the Eurasian mixture event in the Sara and Laal speakers is independent of the mixture event in East Africans and the Toubou for two reasons: (1) admixture LD showed that the events in southern Chad preceded the events in East Africa by 2,000–4,500 years, and (2) we found in Chad a Eurasian Y chromosome lineage (Y haplogroup R1b-V88) that had penetrated all Chadian populations examined but was absent or rare from the Ethiopians examined (Table S4; Figure S1). From whole Y chromosome sequences (Figure S2), we estimate that the Chadian R1b-V88 chromosomes sampled emerged 5,700–7,300 ya (Figure 3B), a time comparable to the Laal speaker admixture dates (4,750–7,200 ya) estimated from genome-wide LD-decay patterns.

The Sources of Eurasian Backflow into Chad and East Africa Are Correlated

Previous studies have suggested that the Eurasian backflow into East Africa came from a population related to early Neolithic farmers.⁵ We wanted to know whether the Eurasian ancestry we found in the Toubou, which we

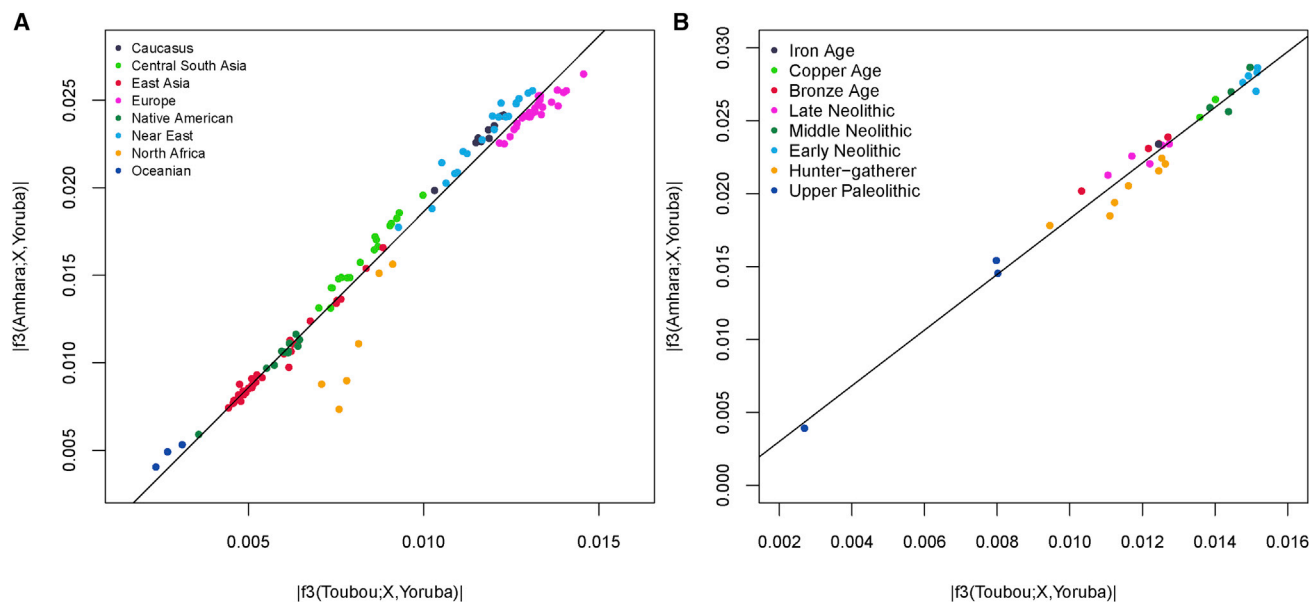


Figure 4. Sources of the Eurasian Ancestry in Chad and Ethiopia

The plot shows significant Eurasian sources for the Toubou and Amhara according to a three-population test (Z score < -4). An increase in the absolute value of the f_3 statistic implies an increase in the genetic affinity of the Eurasian population X to the Toubou and Amhara. (A) With the exception of North Africans, who showed increased affinity to the Toubou, present-day populations showed correlated affinity to both the Toubou and Amhara. Among modern populations, Sardinians showed the highest genetic affinity to both the Toubou and Amhara.

(B) Ancient Eurasians also showed correlated affinity to both the Toubou and Amhara; the early Neolithic LBK (Linearbandkeramik, or Linear Pottery) population ($\sim 5,000$ BCE) had the highest affinity.

attribute to a mixture close in time to the date of mixture in East Africans, can be traced to the same source populations that influenced Ethiopia. We performed the tests $f_3(\text{Toubou}; \text{Yoruba}, X)$ and $f_3(\text{Amhara}; \text{Yoruba}, X)$, where X is a present-day non-sub-Saharan African population in our dataset and is related to one that contributed ancestry to the Toubou and Amhara (Z score < -4) (Table S5). We then looked at the correlation of the f_3 statistic values between the two tests (Figure 4A). We found that the Eurasian source populations for the Amhara and Toubou were highly correlated ($r = 0.98$; 95% CI = 0.98–0.99; p value $< 2.2 \times 10^{-16}$) and that the most significant result was for present-day Sardinians. Exceptions to this correlation were the North African populations (Tunisians, Mozabite, Algerians, and Saharawi), who appeared to have contributed more ancestry to the Toubou than to the Amhara. We repeated the tests by using published ancient genomes (Table S6) and also found a high correlation of the Eurasian sources for the Amhara and Toubou ($r = 0.98$; 95% CI = 0.97–0.99; p value $< 2.2 \times 10^{-16}$); early Neolithic farmers were the most significant contributors, as reported previously⁵ (Figure 4B). When we substituted the Amhara with other Ethiopians (Wolayta and Oromo), we found similar results (data not shown). In a parallel comparison, we checked whether the sources of the African ancestry in different Near Eastern populations were also correlated. We tested $f_3(\text{Lebanese}; \text{British}, X)$ and $f_3(\text{Yemeni}; \text{British}, X)$ and found a lower correlation of the f_3 values ($r = 0.62$; 95% CI = 0.32–0.80), suggesting a

more complicated history of gene flow from genetically different Africans to different populations in the Near East.

We next quantified the proportion of African-Eurasian mixture in the study populations by using two methods: (1) ADMIXTURE²⁶ supervised with $K = 2$ and the British and Yoruba as ancestral populations and (2) the f_4 ratio $\alpha = f_4(\text{British}, \text{chimp}; X, \text{Yoruba})/f_4(\text{British}, \text{chimp}; \text{early Neolithic farmer}, \text{Yoruba})$, where X is one of the populations in our dataset (Figure S3). The results from the two tests were highly correlated ($r = 0.998$; 95% CI = 0.996–0.999; p value $< 2.2 \times 10^{-16}$). Eurasian ancestry was estimated at 26%–30% in the Toubou, 0.3%–2% in the Sara, and 1.2%–4.5% in the Laal speakers. Eurasian ancestry in Ethiopians ranged from 11%–12% in the Gumuz to 53%–57% in the Amhara. African ancestry in the Near East ranged from 7%–14% (Yemen) to 0.7%–5% (Lebanese Christians).

Eurasian Gene Flow Shaped the Genomes of Admixed Africans

Our results from the PCA and MSMC analysis showed a deviation of the admixed populations from the patterns observed in unadmixed (or less admixed) populations in the same geographical region. The MSMC analysis, in particular, showed that admixed Africans had patterns indicative of a decline in heterozygosity (increased bottleneck $\sim 60,000$ ya), somewhat similarly to Eurasians. We tested whole-genome heterozygosity in these populations and found that it decreased in admixed Africans according

to their Eurasian ancestry (Figure S4A). This decrease was not related to recent inbreeding, given that removing segments with long runs of homozygosity did not change the overall pattern. Our simulations suggest that decay in heterozygosity is expected after gene flow from a population with diversity comparable to that of Eurasians (Figures S4B and S4C). We further investigated heterozygosity in admixed Africans by assessing heterozygosity of the different ancestral segments in the Toubou genome. We found that admixed African-Eurasian segments had more heterozygosity (1.23 hets/kb) than segments of the genome where African-African haplotypes were present (1.19 hets/kb) (Figure S5). However, the Toubou genome segments with complete Eurasian ancestry (Eurasian-Eurasian) had considerably lower heterozygosity (~0.96 hets/kb; Figure S5), leading to the genome-wide pattern of decay in heterozygosity observed in Africans with Eurasian ancestry (Figure S4).

We wanted to understand the consequence of admixture on the models that use the density of heterozygous sites to infer the demographic history of populations. We first tested whether the coalescent history estimated by MSMC was affected by a small proportion of mixture, such as the African mixture found in Greeks and Lebanese (ranging from 0% to 5%). We tested the Greek, Lebanese, CEU, and CHB (Han Chinese in Beijing, China) split times from the Yoruba and found that all populations split from the Yoruba ~70,000–80,000 ya, implying that the low proportions of African admixture in the Greeks and Lebanese did not detectably affect the estimates of relative cross-coalescence rate (Figure S6A). We next tested the Toubou, who have ~30% Eurasian ancestry. The Toubou appeared to split from Eurasians ~30,000–40,000 ya, a time more recent than expected considering the African-Eurasian split 60,000–80,000 ya²⁰ (Figure S6B). We tested other Africans in our dataset and found that the Sara, Laal speakers, and Yoruba split from Eurasians, as expected, ~70,000–80,000 ya (Figure S6B). We then tested directly whether the Eurasian ancestry affected the relative cross-coalescence rate between the Toubou and Eurasians by masking some of the Eurasian ancestry in the Toubou. We used PCAdmix²¹ to estimate the ancestry along each chromosome and then used the identified Eurasian segments as a negative mask in our analysis. The split times between the Toubou and Lebanese, for example, increased by ~15,000 years (Figure S6B), shifting the split date toward the expected African-Eurasian split time.

We found that, in addition to influencing the relative cross-coalescence rate, admixture can also inflate putative signals of positive selection. For example, using the PBS³¹ to detect recent positive selection that occurred in the Toubou after their divergence from the Yoruba, we found signals of selection on *MCM6* (MIM: 601806) rs4988235, a variant associated with the lactase-persistence phenotype. This SNP was previously found to be under strong positive selection in Europeans, where it was probably advantageous to individuals living in pastoralist societies.³⁴

The frequency of this variant in the Toubou is 2%, and it is absent from the sub-Saharan African and other Chadic samples (the Sara and Laal speakers) examined here. Although this SNP appears to be a candidate for selection, we suggest that it has probably drifted neutrally in the Toubou after the Eurasian gene flow: the Toubou have ~30% Eurasian ancestry from a population similar to the Greeks, who have 13% derived alleles at rs4988235, suggesting an expectation of ~3.9% of the derived allele simply from admixture. We similarly found in the Toubou signals at *HERC2* (MIM: 605837) rs1129038 a major contributor to blue eye color in Europeans³⁵ (Toubou derived allele frequency [DAF] = 0.014; Greek DAF = 0.33; Yoruba, Sara, and Laal DAF = 0), as well as a signal at *SLC24A5* (MIM: 609802) rs1834640, a major contributor to pigmentation³⁶ (Toubou DAF = 0.19; Greek DAF = 0.99; Yoruba, Sara, and Laal DAF = 0–0.04).

In addition to introducing to African populations genes that were positively selected in Europe, the recent African-Eurasian admixture carried Neanderthal alleles to Central and East Africa. Neanderthals are closer to the Amhara than to the Yoruba: $D(\text{Neanderthal, chimp, Amhara, Yoruba}) = 0.0094$. Neanderthals are also closer to the Toubou than to the Yoruba: $D(\text{Neanderthal, chimp, Toubou, Yoruba}) = 0.0041$. On the other hand, we found that Neanderthals are closer to Europeans than to Near Easterners: $D(\text{Neanderthal, chimp; French, Yemen}) = 0.0056$ and $D(\text{Neanderthal, chimp; French, Palestinian}) = 0.0040$.

We estimated the archaic ancestry proportions by using the ratio

$$\alpha = \frac{f_4(\text{Altai, chimp; X, Yoruba})}{f_4(\text{Altai, chimp; Vindija, Yoruba})}$$

and found that Neanderthal ancestry was ~0.5% in the Toubou and ~1% in the Amhara. We then computed the correlation between the Neanderthal ancestry proportions and the Eurasian and African ancestry proportions we identified. Neanderthal ancestry in admixed Africans and Near Easterners was highly correlated with their Eurasian ancestry ($r = 0.98$; p value $< 2.2 \times 10^{-16}$) and inversely correlated with their African ancestry (Figure 5).

Discussion

We have generated an extensive set of genotyping and high-coverage whole-genome sequencing data to study the genetic history of Chad and neighboring populations. We found substantial genetic differences between the ethnic groups inhabiting Chad today and suggest that multiple ancient Eurasian migrations played a major role in shaping the genetic diversity of the region (Figure 3C). Here, we discuss these migrations and how the mixed ancestry can confound proper interpretation of the evolutionary processes that occurred in their history and therefore needs to be thoroughly accounted for in the study of African genetic diversity.

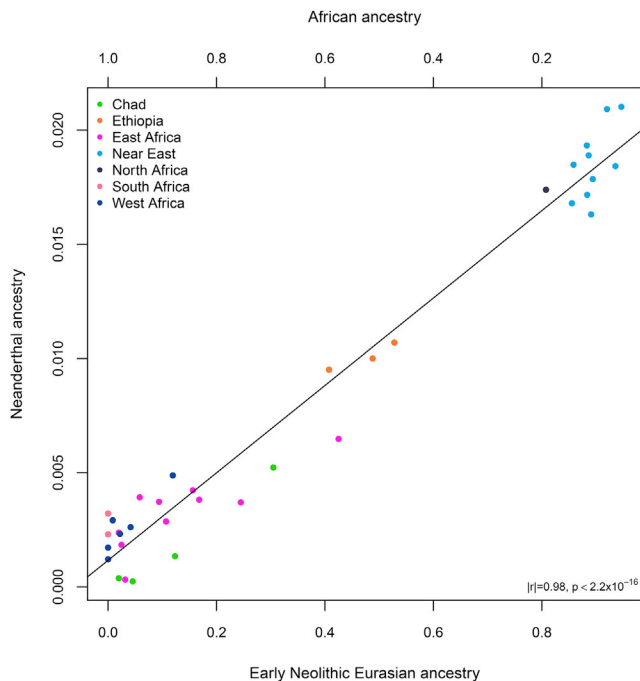


Figure 5. Neanderthal Ancestry Correlation with the African-Eurasian Admixture

Neanderthal ancestry is not expected in Africa, yet today many Africans carry Neanderthal-derived alleles. The plot shows that the Neanderthal ancestry proportion in Africans is correlated with gene flow from Eurasians. For example, knowing that today Eurasians carry ~2% of Neanderthal ancestry, we observed that East Africans (Ethiopians) had ~1% Neanderthal ancestry and ~50% Eurasian ancestry. Correspondingly, Near Easterners showed a decline in Neanderthal ancestry proportional to their levels of African ancestry.

We detected the earliest Eurasian migrations to Africa in the Laal-speaking people, an isolated language group of fewer than 800 speakers who inhabit southern Chad. We estimate that mixture occurred 4,750–7,200 ya, thus after the Neolithic transition in the Near East, a period characterized by exponential growth in human population size. Environmental changes during this period (which possibly triggered the Neolithic transition) also facilitated human migrations. The African Humid Period, for example, was a humid phase across North Africa that peaked 6,000–9,000 ya³⁷ and biogeographically connected Africa to Eurasia, facilitating human movement across these regions.³⁸ In Chad, we found a Y chromosome lineage (R1b-V88) that we estimate emerged during the same period 5,700–7,300 ya (Figure 3B). The closest related Y chromosome groups today are widespread in Eurasia and have been previously associated with human expansions to Europe.^{39,40} We estimate that the Eurasian R1b lineages initially diverged 7,300–9,400 ya, at the time of the Neolithic expansions. However, we found that the African and Eurasian R1b lineages diverged 17,900–23,000 ya, suggesting that genetic structure was already established between the groups who expanded to Europe and Africa. R1b-V88 was previously found in Central and West Africa

and was associated with a mid-Holocene migration of Afro-asiatic speakers through the central Sahara into the Lake Chad Basin.⁸ In the populations we examined, we found R1b in the Toubou and Sara, who speak Nilo-Saharan languages, and also in the Laal people, who speak an unclassified language. This suggests that R1b penetrated Africa independently of the Afro-asiatic language spread or passed to other groups through admixture.

In addition to the early Eurasian migration to Africa ~6,000 ya, a second migration ~3,000 ya affected the Toubou population in northern Chad but had no detectable genetic impact on other Chadian populations. This migration appears to be associated with the previously reported Eurasian backflow into East Africa, given that the source populations and dates of mixture are similar. Occurring at the start of the Iron Age, these migrations could have been facilitated by advances in warfare and transportation technology in the Near East. It is uncertain why the impact of this migration in Chad affected only the Toubou. The African ancestral component in the Toubou is best represented by the Laal-speaking population, suggesting that the African-Eurasian mixture probably occurred in Chad. However, ethnolinguistic barriers could have already been established at this time between the Chad groups, preventing a widespread dissemination of the Eurasian ancestry. The Toubou, despite their Islamic faith, do not show the genetic admixture detected in many Near Eastern and North African populations around 1,100 ya,⁴¹ suggesting conversion without population mixing at this time. They did, however, receive additional Eurasian ancestry in the past 200 years from a source represented by North African populations such as Tunisians, Mozabite, Algerians, and Sahrawi (Figure 3C). This recent interaction could have been promoted by the nomadic lifestyle of the present-day Toubou and a shared Muslim religion with North Africans. Unsurprisingly, we also detected a likely mixing of Chad populations in the sample from the capital, which could be even more recent.

Eurasian backflow into Africa thus appears to have been a recurrent event in the history of many Africans, given its considerable impact on their genomes. Although population mixture in general is a process that increases genetic diversity, we observed a decrease in heterozygosity in the admixed Africans. Our simulations showed that these results are expected after mixture at these proportions with the Eurasians who suffered a significant bottleneck at the time of their exodus from Africa ~60,000 ya. Consequently, we found that mixture can complicate interpretation of the coalescent history inferred from models that use the density of heterozygous sites in their implementations. In addition, we detected in admixed Africans an inflation of positive-selection signals on alleles associated with adult lactose tolerance and pigmentation in Europeans, but we suggest that these alleles have drifted neutrally in Africans after admixture. Furthermore, we detected Neanderthal ancestry in admixed Africans and found it to be proportional to their Eurasian ancestry.

Similarly, in admixed Near Easterners, we found a decrease in Neanderthal ancestry proportional to the gene flow they have received from Africans. Although a higher genetic affinity of Neanderthals to Europeans than to Near Easterners was previously interpreted as additional Neanderthal admixture in the history of Europeans,⁴² we propose that a more parsimonious explanation for these observations is that African-Eurasian mixtures both introduced Neanderthal ancestry to Africa and “diluted” the Neanderthal ancestry in the Near East.

It is important to note that in this work we inevitably invoke Occam’s razor to support the simplest model consistent with our data; the history of the populations studied here, including the time and sources of the Eurasian admixture in Africa, could be more complex. aDNA from Chad and neighboring regions remains a challenge given the poor DNA preservation in hot climates, but future successful efforts in aDNA research could provide additional insights and reveal additional complexities not considered by the modern-DNA-based models favored here.⁴³

Our study has shown that human genetic diversity in Africa is still incompletely understood and that ancient admixture adds to its complexity. This work highlights the importance of exploring underrepresented populations, such as those from Chad, in genetic studies to improve our understanding of the demographic processes that shaped genetic variation in Africa and globally.

Accession Numbers

Whole-genome sequencing and SNP genotyping data are available through the European Genome-phenome Archive (EGA) under accession numbers EGA: EGAD00001002742 (sequences from Chad and Lebanon), EGAD00001001440 (sequences from Greece), and EGAS00001001231 (SNP data from Chad, Lebanon, and Yemen).

Supplemental Data

Supplemental Data include six figures and six tables and can be found with this article online at <http://dx.doi.org/10.1016/j.ajhg.2016.10.012>.

Acknowledgments

We thank all sample donors for making this work possible and the Wellcome Trust Sanger Institute pipelines for generating genotype and sequence data. We also thank Andrea Massaia for comments on analyzing Y chromosome haplogroups from array data and David Soria-Hernanz for comments on the ethnic groups and languages in Chad. M.H., M.M., A.B., J.P.-M., E.Z., Y.X., and C.T.-S. were supported by the Wellcome Trust (098051). P.H. was supported by Estonian Research Council grant PUT1036. J.B.-S. is a full-time employee of Karius, Inc., and R.S.W. is founder and CEO of Insiteome.

Received: August 15, 2016

Accepted: October 24, 2016

Published: November 23, 2016

Web Resources

European Genome-phenome Archive (EGA), <https://www.ebi.ac.uk/ega/home>

ISOGG Y-DNA Haplotype Tree v.11.201, <http://www.isogg.org/tree/>

References

1. Central Intelligence Agency. (2014). The World Factbook 2014. <https://www.cia.gov/library/publications/the-world-factbook/index.html>.
2. Kröpelin, S., Verschuren, D., Lézine, A.M., Eggermont, H., Cocquyt, C., Francus, P., Cazet, J.P., Fagot, M., Rumes, B., Russell, J.M., et al. (2008). Climate-driven ecosystem succession in the Sahara: the past 6000 years. *Science* 320, 765–768.
3. Löhr, D. (2009). Lake Chad and the migratory routes to Borno: A linguistic trail. In *Migrations and Spatial Mobility in the Lake Chad Basin*, XIIIth Mega-Chad Conference, H. Tourneaux, ed. (Editions de l’IRD), pp. 665–681.
4. Pagani, L., Kivisild, T., Tarekegn, A., Ekong, R., Plaster, C., Gallego Romero, I., Ayub, Q., Mehdi, S.Q., Thomas, M.G., Luiselli, D., et al. (2012). Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. *Am. J. Hum. Genet.* 91, 83–96.
5. Gallego Llorente, M., Jones, E.R., Eriksson, A., Siska, V., Arthur, K.W., Arthur, J.W., Curtis, M.C., Stock, J.T., Coltorti, M., Pieruccini, P., et al. (2015). Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa. *Science* 350, 820–822.
6. Cerný, V., Fernandes, V., Costa, M.D., Hájek, M., Mulligan, C.J., and Pereira, L. (2009). Migration of Chadic speaking pastoralists within Africa based on population structure of Chad Basin and phylogeography of mitochondrial L3f haplogroup. *BMC Evol. Biol.* 9, 63.
7. Cerezo, M., Cerný, V., Carracedo, Á., and Salas, A. (2011). New insights into the Lake Chad Basin population structure revealed by high-throughput genotyping of mitochondrial DNA coding SNPs. *PLoS ONE* 6, e18682.
8. Cruciani, F., Trombetta, B., Sellitto, D., Massaia, A., Destro-Bisoli, G., Watson, E., Beraud Colomb, E., Dugoujon, J.M., Moral, P., and Scozzari, R. (2010). Human Y chromosome haplogroup R-V88: a paternal genetic record of early mid Holocene trans-Saharan connections and the spread of Chadic languages. *Eur. J. Hum. Genet.* 18, 800–807.
9. Triska, P., Soares, P., Patin, E., Fernandes, V., Cerny, V., and Pereira, L. (2015). Extensive Admixture and Selective Pressure Across the Sahel Belt. *Genome Biol. Evol.* 7, 3484–3495.
10. Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., Karthikeyan, S., Iles, L., Pollard, M.O., Choudhury, A., et al. (2015). The African Genome Variation Project shapes medical genetics in Africa. *Nature* 517, 327–332.
11. Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., McVean, G.A.; and 1000 Genomes Project Consortium (2010). A map of human genome variation from population-scale sequencing. *Nature* 467, 1061–1073.
12. Pagani, L., Schiffels, S., Gurdasani, D., Danecek, P., Scally, A., Chen, Y., Xue, Y., Haber, M., Ekong, R., Oljira, T., et al. (2015). Tracing the route of modern humans out of Africa by using 225 human genome sequences from Ethiopians and Egyptians. *Am. J. Hum. Genet.* 96, 986–991.

13. Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., et al. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522, 207–211.
14. Drmanac, R., Sparks, A.B., Callow, M.J., Halpern, A.L., Burns, N.L., Kermani, B.G., Carnevali, P., Nazarenko, I., Nilsen, G.B., Yeung, G., et al. (2010). Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* 327, 78–81.
15. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575.
16. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R.; and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.
17. Delaneau, O., Zagury, J.F., and Marchini, J. (2013). Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* 10, 5–6.
18. Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., Abecasis, G.R.; and 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. *Nature* 526, 68–74.
19. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genet.* 2, e190.
20. Schiffels, S., and Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* 46, 919–925.
21. Brisbin, A., Bryc, K., Byrnes, J., Zakharia, F., Omberg, L., Degenhardt, J., Reynolds, A., Ostrer, H., Mezey, J.G., and Bustamante, C.D. (2012). PCAdmix: principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. *Hum. Biol.* 84, 343–364.
22. Bergström, A., Nagle, N., Chen, Y., McCarthy, S., Pollard, M.O., Ayub, Q., Wilcox, S., Wilcox, L., van Oorschot, R.A., McAllister, P., et al. (2016). Deep roots for Aboriginal Australian Y chromosomes. *Curr. Biol.* 26, 809–813.
23. Forster, P., Harding, R., Torroni, A., and Bandelt, H.J. (1996). Origin and evolution of Native American mtDNA variation: a reappraisal. *Am. J. Hum. Genet.* 59, 935–945.
24. Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S.M., Bondarev, A.A., Johnson, P.L., Aximu-Petri, A., Prüfer, K., de Filippo, C., et al. (2014). Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* 514, 445–449.
25. Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. *Genetics* 192, 1065–1093.
26. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664.
27. Loh, P.R., Lipson, M., Patterson, N., Moorjani, P., Pickrell, J.K., Reich, D., and Berger, B. (2013). Inferring admixture histories of human populations using linkage disequilibrium. *Genetics* 193, 1233–1254.
28. Pickrell, J.K., Patterson, N., Loh, P.R., Lipson, M., Berger, B., Stoneking, M., Pakendorf, B., and Reich, D. (2014). Ancient west Eurasian ancestry in southern and eastern Africa. *Proc. Natl. Acad. Sci. USA* 111, 2632–2637.
29. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al.; 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158.
30. Peng, B., and Kimmel, M. (2005). simuPOP: a forward-time population genetics simulation environment. *Bioinformatics* 21, 3686–3687.
31. Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z.X., Pool, J.E., Xu, X., Jiang, H., Vinckenbosch, N., Korneliussen, T.S., et al. (2010). Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* 329, 75–78.
32. Ayub, Q., Mezzavilla, M., Pagani, L., Haber, M., Mohyuddin, A., Khaliq, S., Mehdi, S.Q., and Tyler-Smith, C. (2015). The Kalash genetic isolate: ancient divergence, drift, and selection. *Am. J. Hum. Genet.* 96, 775–783.
33. Li, H., and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature* 475, 493–496.
34. Enattah, N.S., Sahi, T., Savilahti, E., Terwilliger, J.D., Peltonen, L., and Järvelä, I. (2002). Identification of a variant associated with adult-type hypolactasia. *Nat. Genet.* 30, 233–237.
35. Sturm, R.A., Duffy, D.L., Zhao, Z.Z., Leite, F.P., Stark, M.S., Hayward, N.K., Martin, N.G., and Montgomery, G.W. (2008). A single SNP in an evolutionary conserved region within intron 86 of the *HERC2* gene determines human blue-brown eye color. *Am. J. Hum. Genet.* 82, 424–431.
36. Stokowski, R.P., Pant, P.V., Dadd, T., Fereday, A., Hinds, D.A., Jarman, C., Filsell, W., Ginger, R.S., Green, M.R., van der Ouderaa, F.J., and Cox, D.R. (2007). A genomewide association study of skin pigmentation in a South Asian population. *Am. J. Hum. Genet.* 81, 1119–1132.
37. deMenocal, P.B., and Tierney, J.E. (2012). Green Sahara: African Humid Periods paced by Earth's orbital changes. *Nature Education Knowledge* 3, 12.
38. Larrasoana, J.C., Roberts, A.P., and Rohling, E.J. (2013). Dynamics of green Sahara periods and their role in hominin evolution. *PLoS ONE* 8, e76514.
39. Balaresque, P., Bowden, G.R., Adams, S.M., Leung, H.Y., King, T.E., Rosser, Z.H., Goodwin, J., Moisan, J.P., Richard, C., Millward, A., et al. (2010). A predominantly neolithic origin for European paternal lineages. *PLoS Biol.* 8, e1000285.
40. Batini, C., Hallast, P., Zadik, D., Delsler, P.M., Benazzo, A., Ghirrotto, S., Arroyo-Pardo, E., Cavalleri, G.L., de Knijff, P., Dupuy, B.M., et al. (2015). Large-scale recent expansion of European patrilineages shown by population resequencing. *Nat. Commun.* 6, 7152.
41. Haber, M., Gauguier, D., Youhanna, S., Patterson, N., Moorjani, P., Botigué, L.R., Platt, D.E., Matisoo-Smith, E., Soria-Hernanz, D.F., Wells, R.S., et al. (2013). Genome-wide diversity in the levant reveals recent structuring by culture. *PLoS Genet.* 9, e1003316.
42. Rodríguez-Flores, J.L., Fakhro, K., Agosto-Perez, F., Ramstetter, M.D., Arbiza, L., Vincent, T.L., Robay, A., Malek, J.A., Suhre, K., Chouchane, L., et al. (2016). Indigenous Arabs are descendants of the earliest split from ancient Eurasian populations. *Genome Res.* 26, 151–162.
43. Haber, M., Mezzavilla, M., Xue, Y., and Tyler-Smith, C. (2016). Ancient DNA and the rewriting of human history: be sparing with Occam's razor. *Genome Biol.* 17, 1.