# Ancient west Eurasian ancestry in southern and eastern Africa

Joseph K. Pickrell[a,1,2], Nick Patterson[b], Po-Ru Loh[c], Mark Lipson[c], Bonnie Berger[b,c], Mark Stoneking[d], Brigitte Pakendorf[e,1], and David Reich[a,b,f,1]

[a]Department of Genetics and [f]Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115; [b]Broad Institute, Cambridge, MA 02142; [c]Department of Mathematics and Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139; [d]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany; and [e]Laboratoire Dynamique du Langage, Unité Mixte de Recherche 5596, Centre National de la Recherche Scientifique and Université Lyon Lumière 2, 69363 Lyon Cedex 07, France

The history of southern Africa involved interactions between indigenous hunter–gatherers and a range of populations that moved into the region. Here we use genome-wide genetic data to show that there are at least two admixture events in the history of Khoisan populations (southern African hunter–gatherers and pastoralists who speak non-Bantu languages with click consonants). One involved populations related to Niger–Congo-speaking African populations, and the other introduced ancestry most closely related to west Eurasian (European or Middle Eastern) populations. We date this latter admixture event to ~900–1,800 y ago and show that it had the largest demographic impact in Khoisan populations that speak Khoe–Kwadi languages. A similar signal of west Eurasian ancestry is present throughout eastern Africa. In particular, we also find evidence for two admixture events in the history of Kenyan, Tanzanian, and Ethiopian populations, the earlier of which involved populations related to west Eurasians and which we date to ~2,700–3,300 y ago. We reconstruct the allele frequencies of the putative west Eurasian population in eastern Africa and show that this population is a good proxy for the west Eurasian ancestry in southern Africa. The most parsimonious explanation for these findings is that west Eurasian ancestry entered southern Africa indirectly through eastern Africa.

prehistory | population genetics | migration

**H**unter–gatherer populations have inhabited southern Africa for tens of thousands of years (1). Within approximately the last 2,000 y, these populations were joined by food-producing groups (both pastoralists and agriculturalists), and a culturally diverse set of populations occupy the region today. Because written history was unavailable until recently in southern Africa, inferences about the migration patterns leading to the present distribution of populations have largely been informed by archaeology and linguistics.

Genetic data are an additional source of information about population history, but extracting this information remains challenging. Studies of diversity in southern Africa have highlighted the influence of precolonial population admixture on the genetic structure of populations in the region (2–4) but have come to different conclusions about the historical scenarios that led to this admixture. In particular, although there is agreement that the arrival of Bantu-speaking agriculturalist populations had a major demographic impact in many populations, the importance of population movements from other parts of Africa or the world is unclear. Schlebusch et al. (2) argued for eastern African ancestry specifically in the Nama, a pastoralist population, and Pickrell et al. (3) raised this possibility not just for the Nama but for several Khoe-speaking populations. Identifying the sources of non-Khoisan ancestry in southern Africa could shed light on the historical processes that led to the extensive linguistic and cultural diversity of the region.

Here, we develop techniques based on the extent of linkage disequilibrium to thoroughly examine the signal of admixture in the southern African Khoisan (defined here as indigenous populations speaking non-Bantu languages with click consonants, without implying cultural, linguistic, or genetic homogeneity of Khoisan

groups). First, we show that all Khoisan populations have some nonzero proportion of west Eurasian ancestry. (Throughout this paper, we will use geographic labels to refer to ancestry, with the caveat that the geographic labels are derived from modern populations. That is, when we refer to "west Eurasian ancestry" in "southern Africa" we are using this as a shorthand for the more cumbersome, but more accurate, phrasing of "ancestry most closely related to populations currently living in west Eurasia" in "populations currently living in southern Africa.") Second, we show that there are multiple waves of population mixture in the history of many southern and eastern African populations, and that west Eurasian ancestry entered eastern Africa on average 2,700–3,300 y ago and southern Africa 900–1,800 y ago. Third, we infer the allele frequencies of the ancestral west Eurasian population in eastern Africa and show that this population is a good proxy for the west Eurasian ancestry in southern Africa. We thus argue that the most plausible source of west Eurasian ancestry in southern Africa is indirect gene flow via eastern Africa.

## Results

We began with an analysis of population mixture in southern Africa, using the data from Pickrell et al. (3) supplemented with an additional 32 individuals from seven Khoisan populations genotyped on the Affymetrix Human Origins Array (*SI Appendix*, Table S1); note that the Damara are excluded from most of the subsequent analyses because they genetically resemble southern African Bantu-speaking groups (3). These southern African data were then combined with previously published worldwide data (5) (*SI Appendix*). After removing individuals who seemed to be genetic outliers with respect to others in their population (*SI Appendix*), we analyzed a final dataset consisting of 1,040

> ### Significance
>
> The hunter–gatherer and pastoralist populations of southern Africa are among the culturally, linguistically, and genetically most diverse human populations. However, little is known about their history. We show that all of these populations have some ancestry most closely related to Europeans and Middle Easterners and use this to reconstruct the history of population movements between Eurasia, eastern Africa, and southern Africa.

individuals from 75 worldwide populations, all genotyped on the Affymetrix Human Origins array at 565,259 SNPs. These data are available on request from the authors for use in analyses of population history.

**West Eurasian Ancestry in the Ju|'hoan_North.** We previously observed that the Ju|'hoan_North, although the least admixed of all Khoisan populations, show a clear signal of admixture when using a test based on the decay of admixture linkage disequilibrium (LD) (3). The theoretical and practical aspects of historical inference from admixture LD have since been examined in greater detail (6); we thus reevaluated this signal in the Ju|'hoan_North using the software ALDER v1.0 (6).

In particular, we were interested in identifying the source of the gene flow by comparing weighted LD curves computed using different reference populations. This is possible because theory predicts that the amplitude of these curves (i.e., the average level of weighted LD between sites separated by 0.5 centiMorgans) becomes larger as one uses reference populations that are closer to the true mixing populations. Loh et al. (6) additionally showed that this theory holds when using the admixed population itself as one of the reference populations. We thus computed weighted LD curves in the Ju|'hoan_North, using the Ju|'hoan_North themselves as one reference population and a range of 74 worldwide populations as the other, and examined the amplitudes of these curves (Fig. 1A). The largest amplitudes are obtained with European populations as references (Fig. 1A); taken literally, this would seem to implicate Europe as the source of admixture (although Middle Eastern populations are also among the best proxies). The estimated date for this gene flow is $43 \pm 2$ generations [$1,290 \pm 60$ y, assuming 30 y per generation (7)] before the present, consistent with our previously estimated date (3). This date is well before the historical arrival of European colonists to the region.

We next tested the robustness of this result. We confirmed that this observation is consistent across panels of SNPs with varied ascertainment (*SI Appendix*, Fig. S2). We then considered hunter–gatherer populations from other regions of Africa. In particular, we performed the same analysis on the Biaka (Fig. 1B) and Mbuti (*SI Appendix*, Fig. S3) from central Africa. As expected, the inferred source of admixture in these populations is a sub-Saharan African population (most closely related to the Yoruba, a Niger–Congo-speaking agriculturalist group from Nigeria).

A signal of west Eurasian ancestry in the Ju|'hoan_North should be identifiable by allele frequencies as well as by LD. We thus tested the population tree [Chimp,[Ju|'hoan_North, [Han, French]]] using an $f_4$ statistic (8, 9). This tree fails with a Z-score of 4.0 ($P = 3 \times 10^{-5}$). On smaller subsets of SNPs, the evidence is weaker, explaining why we had not noticed it previously (on the set of SNPs ascertained in a Ju|'hoan individual, $Z = 2.7$ [$P = 0.003$]; in a French individual, $Z = 0.6$ [$P = 0.27$]; in a Yoruba individual, $Z = 1.4$ [$P = 0.08$]). We thus conclude that there is a signal in both allele frequencies and linkage disequilibrium that the Ju|'hoan_North admixed with a population more closely related to western Eurasian (i.e., European or Middle Eastern) rather than eastern Eurasian populations, and that this signal is absent from hunter–gatherer populations in central Africa.
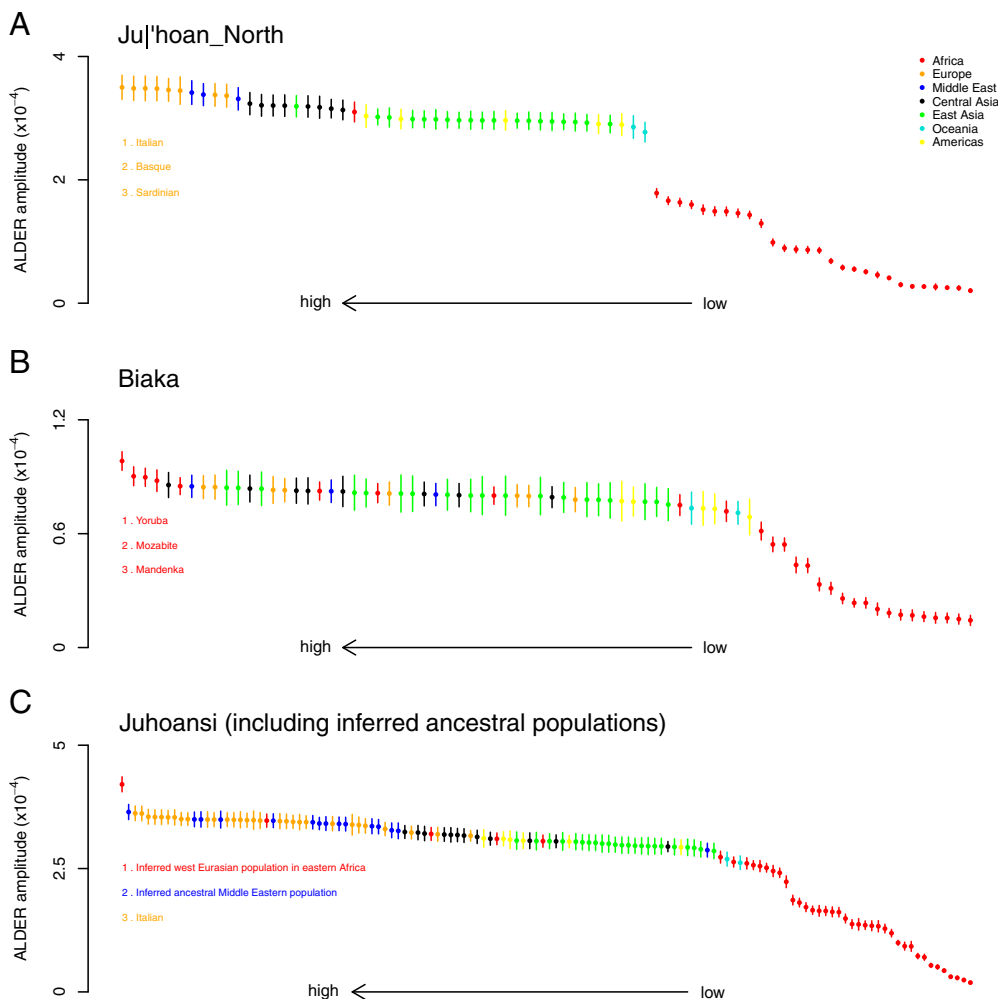


Fig. 1. Identifying sources of admixture using LD. In each panel, we computed weighted LD curves with ALDER v1.0 using a test population as one reference and a panel of other populations as the second reference. We performed this analysis with different test populations: (A) the Ju|'hoan_North, (B) the Biaka, and (C) the Juhoansi [equivalent to the Ju|'hoan_North, but different samples genotyped on a different genotyping array (2)]. We fit an exponential decay curve to each LD curve, starting from 0.5 centiMorgans. Plotted are the fitted amplitudes for each curve; error bars indicate one SE. A larger amplitude indicates a closer relationship to one of the true admixing populations. Populations are ordered according to the amplitude and colored according to their continent of origin. The three populations with the largest amplitude (and thus the closest inferred relationship to the true mixing population) are listed. Note that the only populations from western Africa in these data are the Yoruba and Mandenka. In C, we include as references two inferred populations: the inferred west Eurasian population that entered Ethiopia (see main text for details) and an inferred Middle Eastern population before admixture with African populations (24).
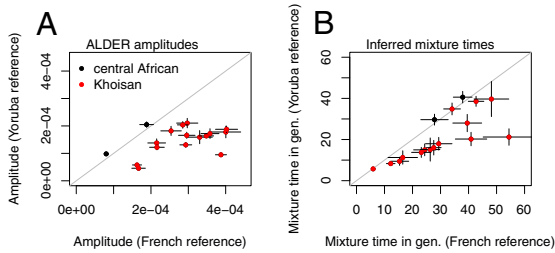
**Fig. 2.** Relationship with west Eurasia is shared by all Khoisan populations. We generated weighted LD decay curves in each Khoisan (or central African hunter–gatherer) population, using weights computed using the test population as one reference and either the French or the Yoruba as the other reference. We then fit an exponential decay model to each LD curve. Plotted are the inferred (A) amplitudes and (B) admixture times in each population. Larger amplitudes indicate a closer relationship to the true admixing population, and under a model of a single admixture event the admixture times do not depend on the reference populations used.

## Signal of West Eurasian Relatedness Is Shared Throughout Southern Africa.

We next examined whether this signal of relatedness to west Eurasia is present in other Khoisan populations. For each Khoisan population, we used ALDER to compute weighted LD decay curves using the test population as one reference and either the French or the Yoruba as the other reference. We included the central African Mbuti and Biaka populations as negative controls. In all Khoisan populations, the amplitude of the LD decay curve is larger when using the French as a reference than when using the Yoruba as a reference (Fig. 2A). In contrast, for the Mbuti and Biaka, the larger amplitude is seen when using the Yoruba as a reference (Fig. 2A).

A striking observation that emerges from this analysis is that in many of the southern African populations the inferred mixture times depend substantially on the second population used as a reference (Fig. 2B). Under a model of admixture from a single source population, the decay rate of the LD curve does not depend on the reference population used (6); this suggests that there are at least two separate non-Khoisan sources of ancestry in some of these Khoisan populations. In contrast, for the central African Mbuti and Biaka the inferred times do not depend on the reference used.

## Estimating Parameters of Multiple Admixture Events.

Motivated by the above observations, we designed a method to estimate dates of multiple admixture events in the history of a population (related ideas have been explored by Myers et al.*). We extended the population genetic theory of Loh et al. (6) to the case where a population has experienced multiple episodes of population admixture from different sources (*SI Appendix*). In this situation, the extent of admixture LD in the population is no longer a single exponential curve as a function of genetic distance, but instead is a mixture of exponential curves. Using a range of reference populations, we can thus formally test for the presence of multiple waves of mixture and estimate the dates of these mixture events (*SI Appendix*). We validated this approach using coalescent simulations of three pairs of mixture dates chosen to span the scenarios that our data suggest are relevant to southern and eastern Africa (*SI Appendix*). The simulations indicate that our method has reasonable but not perfect power; depending on the pair of dates we simulated, we successfully detected both events in between 50% and 90% of simulated cases.

To illustrate the intuition behind this method, in Fig. 3 we plot one of the weighted LD curves calculated in the G‖ana. Under a model with a single admixture event, the mean admixture date in the G‖ana is estimated as 14 ± 3 generations, identical to the

date obtained by Pickrell et al. (3). However, it is visually apparent that this model is a poor fit to the data (Fig. 3). Indeed, we find that adding a second mixture event significantly improves the fit (minimum Z-score on the two admixture times of 2.8; $P = 0.003$). The two inferred mean admixture times in the G‖ana are 4 ± 1 and 39 ± 6 generations ago.

This method additionally estimates amplitudes of the LD decay curves for each pair of populations on each mixture time, which are a function of the relationship between the reference populations and the true source populations. These amplitudes can be used to infer the references closest to the true mixing populations. However, if a source population is itself admixed, under some conditions this method will identify a population related to one of the ancestral components of the source population instead of the source population itself (*SI Appendix*). By examining these amplitudes, we conclude that the west Eurasian ancestry in the G‖ana
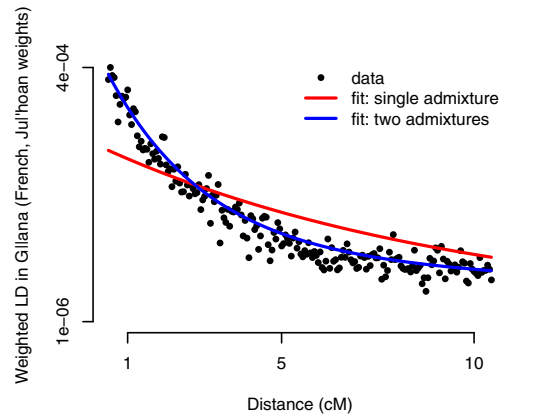


| Reference populations | Amplitude 1 (x10⁻⁴) | Amplitude 2 (x10⁻⁴) |
|---|---|---|
| **Italian;Taa_West** | **4.14** | 0.954 |
| **Sardinian;Taa_West** | **4.09** | 0.97 |
| **Orcadian;Taa_West** | **4.04** | 0.962 |
| **Druze;Taa_West** | **4.04** | 0.957 |
| **Russian;Taa_West** | **4.02** | 0.972 |
| **Glui;Yoruba** | 2.99 | **1.36** |
| **Jul'hoan_North;Yoruba** | 2.81 | **1.32** |
| **Jul'hoan_South;Yoruba** | 2.81 | **1.32** |
| **Glui;Mandenka** | 2.92 | **1.31** |
| **Taa_North;Yoruba** | 2.93 | **1.3** |

**Fig. 3.** LD evidence for multiple waves of mixture in the G‖ana. We computed 990 (45 choose 2) weighted LD curves in the G‖ana and fit two models: one with a single admixture event, and one with two admixture events. Shown is the LD curve computed using the French and Jul'hoan_North populations as references, along with the fitted curves from the two models (note that the decay rates in the fitted curves are shared across the data for all 990 pairs of populations, not only to the shown data). Below the plot, we show a schematic representation of the fitted model with two admixture events. In the table, we show the population pairs with the five largest estimated amplitudes on each admixture event (that is, the population pairs in dark blue are those with the largest weights on the dark blue curve, and those labeled in light blue are those with the largest weights on the light blue curve).

---

*Myers S, et al., LD patterns in dense variation data reveal information about the history of human populations worldwide. Presented at the 61st Annual Meeting of the American Society of Human Genetics, October 11–15, 2011, Montreal, QC, Canada.
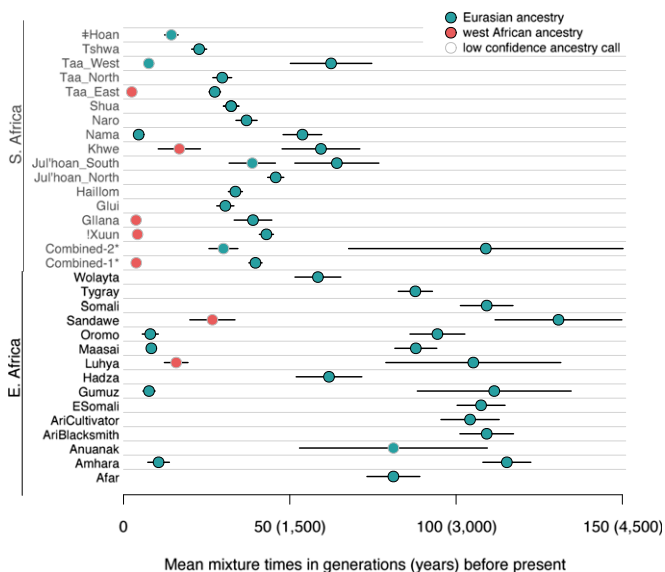
**Fig. 4.** Mean times of admixture in southern and eastern Africa. For each southern or eastern African population, we estimated the number of mixture events and their dates. Plotted are the estimated dates. Black lines show one SE on the estimates. Points are colored according to the populations inferred as proxies for the mixing populations (*SI Appendix*). *The combined-1 population is the Tshwa, Shua, Hai‖om, ‡Hoan, Naro, and Taa_North. The combined-2 population is the Ju|'hoan_North and G‖ui (see main text for details).

entered the population through the older admixture event (Fig. 3). Because of the caveat noted above, however, we cannot distinguish between two historical scenarios with this method: direct gene flow from a west Eurasian population and gene flow from a west Eurasian-admixed population.

We applied this method to each Khoisan population in turn (with the exception of the Damara, who are genetically similar to non-Khoisan populations) using 45 other African and non-African populations as references (*SI Appendix*, Figs. S10–S23). In several populations there is evidence for two waves of population mixture (!Xuun, Taa_West, Taa_East, Nama, Khwe, G‖ana, and Ju|'hoan_South), whereas in others a single wave of population mixture fits the data (Fig. 4). For populations with two waves of mixture, west Eurasian ancestry entered through the earlier admixture event (Fig. 4 and *SI Appendix*, Table S3). In the Nama, both the early and more recent admixture events are predicted to involve populations with west Eurasian ancestry, consistent with known postcolonial European admixture in this population. The Taa_West also show two episodes of west Eurasian admixture, but the more recent one has low confidence.

It is important to mention a few caveats in interpreting results from this method. First, in cases in which the method detects two admixture events from the same source (as in the Nama and Taa_West above), simulations suggest an alternative interpretation is sustained population mixture over many generations (*SI Appendix*). Second, the numbers of admixture events inferred by this method are lower bounds; for example, this method fails to detect that the Naro are admixed between two distinct Khoisan groups (3), and we find evidence of west African ancestry in just four Khoisan populations (!Xuun, G‖ana, Khwe, and Taa_East) when treated individually (but see analyses of combined populations below). Finally, the method has low confidence when assigning an admixture event to a population with west African ancestry (Fig. 4); this reflects a relative lack of genetic drift specific to the west African reference populations (Yoruba and Mandenka), which makes it difficult to detect with high confidence (in contrast, there is considerable genetic drift in west Eurasian populations because of the out-of-Africa bottleneck, which allows admixture events to be more confidently assigned to this ancestry).

In most populations in which our method detects only a single admixture event, the fitted model visually seems inadequate to fully explain the data (e.g., *SI Appendix*, Figs. S11, S13, S20, and S21). Indeed, there is marginal statistical evidence for two admixture events in many of these populations (*SI Appendix*, Table S4). To increase our power to detect additional admixture events, we performed analyses of combined populations. In a combined set of populations (the Tshwa, Shua, Hai‖om, ‡Hoan, Naro, and Taa_North) that have marginal evidence for a second, more recent admixture event, we infer two dates of admixture: one $40 \pm 2$ generations ago and one $4 \pm 1$ generations ago (Z-score for the hypothesis that the admixture time is zero is 3.2, $P = 7 \times 10^{-4}$). In a combined set of two populations (the Ju|'hoan_North and G‖ui) that have marginal evidence for a second, more ancient admixture event, we also infer two dates of admixture (*SI Appendix*, Fig. S24), but with different dates from all other samples: one $30 \pm 4$ generations ago (Z-score of 6.9, $P = 2 \times 10^{-12}$) and one $109 \pm 41$ generations ago (Z-score of 2.6, $P = 0.005$). We interpret this as suggestive evidence that the population that introduced west Eurasian ancestry to southern Africa was itself admixed, and that this more ancient admixture happened around 110 generations ago (although the confidence intervals here are clearly large).

**Variation in West Eurasian Ancestry Proportions in the Khoisan.** We next asked whether there are systematic differences between Khoisan populations in their levels of west Eurasian ancestry. To test this, we constructed an $f_4$ ratio estimate to specifically measure west Eurasian ancestry. This ratio is $f_4$(Han, Orcadian; X, Druze)/$f_4$(Han, Orcadian; Yoruba, Druze), where X is any southern African population; this ratio takes advantage of the fact that the west Eurasian ancestry is more closely related to Middle Eastern than to northern European populations (*SI Appendix*). We applied this method to all Khoisan populations and included southern African Bantu speakers for comparison. The highest levels of west Eurasian ancestry are found in Khoe–Kwadi speakers (Table 1, southern Africa), particularly the Nama, where our estimate of west Eurasian ancestry reaches 14% (although note we cannot distinguish between the impact of recent colonialism and older west Eurasian ancestry in the Nama using this method). Other populations of note include the Khwe, Shua, and Hai‖om, whom we estimate to have ~5% west Eurasian ancestry. The apparent correlation between language group and west Eurasian ancestry may have implications for the origins of this ancestry in southern Africa; we return to this point in the discussion.

**Origin of West Eurasian Ancestry in Southern Africa.** We next considered the origin of the west Eurasian ancestry in southern Africa. Direct interactions between Europe and southern Africa seem unlikely given the inferred admixture dates, especially because this ancestry is widespread throughout southern Africa. It has been reported that many populations in eastern Africa admixed with populations from the Levant (10) or the Arabian peninsula (11). Because there is suggestive genetic evidence of a migration from eastern Africa to southern Africa (2, 3, 12) as well as linguistic and archaeological indications (13), we hypothesized that indirect gene flow through eastern Africa might be a plausible source for the west Eurasian ancestry in southern Africa. This hypothesis makes two major predictions: First, that the west Eurasian ancestry in eastern Africa should have the same source as that in southern Africa, and second, that the mixture times in eastern Africa should be older than those in southern Africa, perhaps with a date of around 110 generations (corresponding to the oldest date identified in southern Africa).

To test these predictions, we assembled a dataset of individuals from southern Africa, eastern Africa, and west Eurasia typed on an Illumina platform by merging data from previous studies (10, 14–17). The eastern African populations in these combined data include populations from Ethiopia, Kenya, Tanzania, and Sudan [the majority of these populations were genotyped by Pagani et al. (10)]. We first confirmed using $f_3$ tests (8) that many eastern African

**Table 1. Estimates of the proportion of west Eurasian ancestry in southern and eastern African populations**

| Population | Language classification | West Eurasian ancestry, % |
|---|---|---|
| Southern Africa | | |
| Nama | Khoe–Kwadi | 14.0 |
| Shua | Khoe–Kwadi | 5.4 |
| Hai‖om | Khoe–Kwadi | 5.2 |
| Khwe | Khoe–Kwadi | 4.0 |
| Tshwa | Khoe–Kwadi | 3.0 |
| Naro | Khoe–Kwadi | 2.2 |
| G\|ui | Khoe–Kwadi | 2.0 |
| Taa_North | Tuu | 1.9 |
| G‖ana | Khoe–Kwadi | 1.6 |
| !Xuun | Kx'a | 1.2 |
| ǂHoan | Kx'a | 1.5 |
| Damara | Khoe–Kwadi | 1.3 |
| Kgalagadi | Bantu | 1.1 |
| Ju\|'hoan_North | Kx'a | 1.0 |
| Taa_East | Tuu | 0.4 |
| Mbukushu | Bantu | 0.5 |
| Taa_West | Tuu | 0.4 |
| Himba | Bantu | 0.1 |
| Ju\|'hoan_South | Kx'a | 0 |
| Tswana | Bantu | 0 |
| Wambo | Bantu | 0 |
| Eastern Africa | | |
| Tygray | Semitic | 50.4 |
| Amhara | Semitic | 49.2 |
| Afar | Cushitic | 46.0 |
| Oromo | Cushitic | 41.6 |
| Somali | Cushitic | 38.4 |
| Ethiopian Somali | Cushitic | 37.9 |
| Wolayta | Omotic | 34.1 |
| Maasai | Nilotic | 18.9 |
| Ari Cultivator | Omotic | 18.2 |
| Sandawe | isolate | 15.8 |
| Ari Blacksmith | Omotic | 15.7 |
| Hadza | isolate | 6.4 |
| Luhya | Bantu | 2.4 |
| Gumuz | isolate | 1.7 |
| Anuak | Nilotic | 0 |

We estimated the percentage of west Eurasian ancestry in each southern and eastern African population (*SI Appendix*). For each region, populations are sorted according to the estimated proportion of west Eurasian ancestry. SEs on all estimates ranged from 0.3 to 1.1%, with an average of 0.7%. We additionally estimated the Eurasian ancestry proportion in the Mandenka from western Africa using this method as 2.0%.

populations have statistically significant evidence for admixture with west Eurasian populations (*SI Appendix*, Table S5). The smallest $f_3$ statistics in nearly all eastern African populations involve a southern European (or Levantine) population as reference. We then evaluated the fraction of west Eurasian ancestry in each population, using the same $f_4$ ratio estimate as used in the Khoisan (Table 1, eastern Africa). The fraction of west Eurasian ancestry in eastern African populations is generally higher in eastern than in southern Africa; the highest levels of admixture (40–50%) are observed in some Ethiopian populations.

To test whether the west Eurasian ancestry in southern and eastern Africa is from the same source, we reconstructed the allele frequencies of the west Eurasian population involved in the admixture in eastern Africa (*SI Appendix*). We then tested whether this hypothetical population is a good proxy for the west Eurasian ancestry in southern Africa. Indeed, this reconstructed population is a better proxy than samples of modern Eurasians (Fig. 1C). In the Juhoansi (who correspond to the Ju\|'hoan_North), we obtain an ALDER amplitude in the one-reference test of

$4.2 \times 10^{-4} \pm 1.5 \times 10^{-5}$ when using this imputed population as a reference versus $3.6 \times 10^{-4} \pm 1.5 \times 10^{-5}$ when using Italians as a reference (one-sided *P* value for difference of 0.0015).

We then applied our method for dating multiple admixture events to the eastern African populations in these data (*SI Appendix*, Figs. S25–S39). Pagani et al. (10) previously dated the earliest admixture events in Ethiopia to around 3,000 y ago, but with considerable variation between populations. We find evidence for multiple episodes of population mixture in eastern Africa; most populations have evidence for an early admixture event that we date to around 80–110 generations (2,400–3,300 y ago) (Fig. 4). As in southern Africa, the west Eurasian ancestry is present in the early admixture event (*SI Appendix*, Table S6). The earliest dates of population mixture that we estimate in eastern Africa are almost uniformly older than those we estimate in southern Africa (Fig. 4). One potential concern regarding this conclusion is that the southern and eastern African populations displayed in Fig. 4 were genotyped on different genotyping arrays; however, this pattern remains when using only populations typed on the same array (*SI Appendix*, Fig. S40). We conclude that the west Eurasian ancestry in southern Africa was likely brought by a migration of an already-admixed population from eastern Africa.

**Estimating the Proportion of Eastern African Ancestry in Southern Africa.** If west Eurasian ancestry indeed entered southern Africa via eastern Africa, then the relative proportions of west Eurasian ancestry in different southern African populations can be interpreted as reflecting different levels of eastern African gene flow. We thus attempted to split the ancestry of all southern African populations into three components: Khoisan ancestry, putative eastern African ancestry, and ancestry from Bantu-speaking immigrants to southern Africa. To do this, we make the following assumptions: First, that all eastern African ancestry in southern Africa came from a single source with a fixed level of west Eurasian admixture, and second, that all non-Khoisan ancestry in southern Africa is either from this putative eastern African population or from a Bantu-speaking population. Because these assumptions are largely unverifiable, the following should be viewed as more qualitative than quantitative.

We first attempted to estimate the proportion of west Eurasian ancestry in the putative eastern African population that entered southern Africa. Using ALDER, we estimate the lower bound on the proportion of non-Khoisan ancestry in the Ju\|'hoan_North as 4%. If ~1% of this is west Eurasian ancestry (Table 1) and the Ju\|'hoan_North have no Bantu-related ancestry, then this gives an admixture proportion of ~25% west Eurasian ancestry in the putative eastern African source population. Using this value, we then estimated the proportions of Khoisan, putative eastern African, and Bantu-related ancestry of all populations using a linear model (18) (*SI Appendix*). In *SI Appendix*, Table S8, we show our estimates of these three components (excluding in this case the Nama, who have recent European ancestry that confounds this analysis).

## Discussion

In this paper, we have examined the history of southern and eastern African populations using patterns of admixture LD. The most striking inference from this analysis is the presence of west Eurasian ancestry in southern Africa that we date to 900–1,800 y ago. Several lines of evidence suggest that the population that brought this ancestry to southern Africa was an already-admixed population from eastern Africa.

**Back-to-Africa Gene Flow in Eastern Africa.** A major open question concerns the initial source of the west Eurasian ancestry in eastern Africa. The estimated mean time of gene flow in eastern Africa is around 3,000 y ago, and the amount of gene flow must have been quite extensive, because the west Eurasian ancestry proportions reach 40–50% in some Ethiopian populations (Table 1 and ref. 10). Archaeological records from this region are sparse, so Pagani et al. (10) speculate that this admixture is related to the Biblical account of

the Kingdom of Sheba. However, archaeological evidence is not completely absent. During this time period, architecture in the Ethiopian culture of D'mt has an "unmistakable South Arabian appearance in many details" (19), although there is some debate as to whether these patterns can be attributed to large movements of people versus elite-driven cultural practices (19, 20). Additionally, linguistic evidence suggests that this time period was when Ethiosemitic languages were introduced to Africa, presumably from southern Arabia (21). It is perhaps not a coincidence that the highest levels of west Eurasian ancestry in eastern Africa are found in the Amhara and Tygray, who speak Ethiosemitic languages and live in what was previously the territory of D'mt and the later kingdom of Aksum.

**West Eurasian Ancestry in Southern Africa.** A second question is, Which population or populations introduced west Eurasian ancestry into southern Africa? The best genetic proxy for this ancestry that we have found is the west Eurasian ancestry in eastern Africa (Fig. 1C), and although we do not identify modern east African populations as the best source population, this is likely due to the lack of genetic drift specific to eastern Africa (*SI Appendix*, section 1.2.3). The most parsimonious explanation for this observation is that west Eurasian ancestry entered southern Africa indirectly via eastern Africa (the alternative scenario of direct contact with an unsampled west Eurasian population cannot formally be excluded; however, there is no archaeological, historical, or linguistic evidence of such contact). The relevant eastern African population may no longer exist. However, such a migration has been suggested based on shared Y chromosome haplotypes (12, 22) and shared alleles/haplotypes associated with lactase persistence (2, 23) between the two regions. Furthermore, based on a synthesis of archaeological, genetic, climatological, and linguistic data Güldemann (13) hypothesized that the ancestor of the Khoe–Kwadi languages in southern Africa was brought to the region by immigrating pastoralists from eastern Africa. Our observation of elevated west Eurasian ancestry in Khoe–Kwadi groups in general (Table 1) is consistent with this hypothesis.

**Alternative Historical Scenarios.** We note that we have interpreted admixture signals in terms of large-scale movements of people. An alternative frame for interpreting these results might instead propose an isolation-by-distance model in which populations primarily remain in a single location but individuals choose mates from within some relatively small radius. In principle, this sort of model could introduce west Eurasian ancestry into southern Africa via a "diffusion-like" process. Two observations argue against this possibility. First, the gene flow we observe is asymmetric: Whereas some eastern African populations have up to 50% west Eurasian ancestry, levels of sub-Saharan African ancestry in the Middle East and Europe are considerably lower than this [maximum of 15% (24)] and do not seem to consist of ancestry related to the Khoisan. Second, the signal of west Eurasian ancestry is present in southern Africa but absent from central Africa, despite the fact that central Africa is geographically closer to the putative source of the ancestry. These geographically specific and asymmetric dispersal patterns are most parsimoniously explained by migration from west Eurasia into eastern Africa, and then from eastern to southern Africa.

## Conclusions

Based on these analyses, we can propose a model for the spread of west Eurasian ancestry in southern and eastern Africa as follows. First, a large-scale movement of people from west Eurasia into Ethiopia around 3,000 y ago (perhaps from southern Arabia and associated with the D'mt kingdom and the arrival of Ethiosemitic languages) resulted in the dispersal of west Eurasian ancestry throughout eastern Africa. This was then followed by a migration of an admixed population (perhaps pastoralists related to speakers of Khoe–Kwadi languages) from eastern Africa to southern Africa, with admixture occurring ~1,500 y ago. Advances in genotyping DNA from archaeological samples may allow aspects of this model to be directly tested.

## Materials and Methods

We genotyped 32 individuals on the Affymetrix Human Origins Array (5). Analysis of admixture linkage disequilibrium was done using ALDER v1.0 (6) and an extension allowing for multiple episodes of population mixture. Estimates of admixture proportions using $f_4$ ratios were performed using qpF4ratio version 300 in the ADMIXTOOLS package (5). *SI Appendix* gives details.

1. Phillipson D (2005) *African Archaeology* (Cambridge Univ Press, Cambridge, UK).
2. Schlebusch CM, et al. (2012) Genomic variation in seven Khoe-San groups reveals adaptation and complex African history. *Science* 338(6105):374–379.
3. Pickrell JK, et al. (2012) The genetic prehistory of southern Africa. *Nat Commun* 3:1143.
4. Petersen DC, et al. (2013) Complex patterns of genomic admixture within southern Africa. *PLoS Genet* 9(3):e1003309.
5. Patterson N, et al. (2012) Ancient admixture in human history. *Genetics* 192(3):1065–1093.
6. Loh P-R, et al. (2013) Inferring admixture histories of human populations using linkage disequilibrium. *Genetics* 193(4):1233–1254.
7. Fenner JN (2005) Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am J Phys Anthropol* 128(2):415–423.
8. Reich D, Thangaraj K, Patterson N, Price AL, Singh L (2009) Reconstructing Indian population history. *Nature* 461(7263):489–494.
9. Keinan A, Mullikin JC, Patterson N, Reich D (2007) Measurement of the human allele frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans. *Nat Genet* 39(10):1251–1255.
10. Pagani L, et al. (2012) Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. *Am J Hum Genet* 91(1):83–96.
11. Haber M, et al. (2013) Genome-wide diversity in the Levant reveals recent structuring by culture. *PLoS Genet* 9(2):e1003316.
12. Henn BM, et al. (2008) Y-chromosomal evidence of a pastoralist migration through Tanzania to southern Africa. *Proc Natl Acad Sci USA* 105(31):10693–10698.
13. Güldemann T (2008) A linguist's view: Khoe-Kwadi speakers as the earliest food-producers of southern Africa. *Southern African Humanities* 20(1):93–132.
14. Altshuler DM, et al.; International HapMap 3 Consortium (2010) Integrating common and rare genetic variation in diverse human populations. *Nature* 467(7311):52–58.
15. Behar DM, et al. (2010) The genome-wide structure of the Jewish people. *Nature* 466(7303):238–242.
16. Henn BM, et al. (2011) Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. *Proc Natl Acad Sci USA* 108(13):5154–5162.
17. Li JZ, et al. (2008) Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319(5866):1100–1104.
18. Patterson N, et al. (2010) Genetic structure of a unique admixed population: implications for medical research. *Hum Mol Genet* 19(3):411–419.
19. Munro-Hay SC (1991) *Aksum: An African Civilisation of Late Antiquity* (Edinburgh Univ Press, Edinburgh).
20. Mitchell P (2005) *African Connections: An Archaeological Perspective on Africa and the Wider World* (AltaMira, Lanham, MD), Vol 7.
21. 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 Biol Sci* 276(1668):2703–2710.
22. Cruciani F, et al. (2002) A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. *Am J Hum Genet* 70(5):1197–1214.
23. Coelho M, Sequeira F, Luiselli D, Beleza S, Rocha J (2009) On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genetic variation in southwestern Angola. *BMC Evol Biol* 9(1):80.
24. Moorjani P, et al. (2011) The history of African gene flow into Southern Europeans, Levantines, and Jews. *PLoS Genet* 7(4):e1001373.

GENETICS