Reconstructing Prehistoric African Population Structure

Graphical Abstract

Highlights

- Genome-wide analysis of 16 African individuals who lived up to 8,100 years ago
- Forager populations related to southern African San were once widespread in eastern Africa
- Comparison of ancient and modern Africans reveal recent genomic adaptations
- Evidence for a divergent human lineage contributing to western Africans

Authors
Pontus Skoglund, Jessica C. Thompson, Mary E. Prendergast, ..., Ron Pinhasi, Johannes Krause, David Reich

Correspondence
skoglund@genetics.med.harvard.edu (P.S.), reich@genetics.med.harvard.edu (D.R.)

In Brief
The prehistory of African populations is explored by genome-wide analysis of 16 human remains providing insight into ancestral lineages, admixture, and genomic adaptations.
Reconstructing Prehistoric African Population Structure

Pontus Skoglund,1,* Jessica C. Thompson,2 Mary E. Prendergast,3 Alissa Mittnik,4,5,33 Kendra Sirak,2,6,33 Mateja Hajdinjak,7,33 Tasneem Salie,6,33 Nadine Rohland,1 Swapan Mallick,1,9 Alexander Peltzer,4,10 Anja Heinze,7 Inigo Olalde,1 Matthew Ferry,1,11 Eadaoin Harney,1,11,12 Megan Michel,1,11 Kristin Stewardson,1,11 Jessica I. Cerezo-Román,13 Chirissy Chiiumia,14 Alison Crowther,1,13 Kristin Stewardson,1,11
1Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
2Department of Anthropology, Emory University, Atlanta, GA 30322, USA
3Radcliffe Institute for Advanced Study, Harvard University, Cambridge, MA 02138, USA
4Max Planck Institute for the Science of Human History, Jena 07745, Germany
5Institute for Archeological Sciences, Eberhard-Karls-University, Tuebingen 72070, Germany
6School of Archaeology and Earth Institute, University College Dublin, Dublin 4, Ireland
7Max Planck Institute for Evolutionary Anthropology, Leipzig 04103, Germany
8Division of Human Genetics, Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town 7925, South Africa
9Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA
10Integrative Transcriptomics, Centre for Bioinformatics, University of Tuebingen, Tuebingen 72076, Germany
11Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA
12Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA
13Department of Geography and Anthropology, California State Polytechnic University, Pomona, Pomona, CA 91768, USA
14Malawi Department of Museums and Monuments, Lilongwe 3, Malawi
15School of Social Science, The University of Queensland, Brisbane, Queensland 4072, Australia
16National Museums of Tanzania, Dar es Salaam, Tanzania
17Department of Anthropology and Archaeology, University of Wisconsin - La Crosse, La Crosse, WI 54601, USA
18Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, UK
19Canterbury Archaeological Trust, Canterbury CT1 2LU, UK
20Department Archaeology and Anthropology, University of Bristol, Bristol BS8 1UU, UK
21Department of Archaeology, University of Cape Town, Cape Town 7700, South Africa

SUMMARY

We assembled genome-wide data from 16 prehistoric Africans. We show that the anciently divergent lineage that comprises the primary ancestry of the southern African San had a wider distribution in the past, contributing approximately two-thirds of the ancestry of Malawi hunter-gatherers ∼8,100–2,500 years ago and approximately one-third of the ancestry of Tanzanian hunter-gatherers ∼1,400 years ago. We document how the spread of farmers from western Africa involved complete replacement of local hunter-gatherers in some regions, and we track the spread of herders by showing that the population of a ∼3,100-year-old pastoralist from Tanzania contributed ancestry to people from north-eastern to southern Africa, including a ∼1,200-year-old southern African pastoralist. The deepest diversifications of African lineages were complex, involving either repeated gene flow among geographically disparate groups or a lineage more deeply diverging than that of the San contributing more to some western African populations than to others. We finally leverage ancient genomes to document episodes of natural selection in southern African populations.

INTRODUCTION

Africa harbors more genetic diversity than any other part of the world (Cann et al., 1987; Tishkoff et al., 2009). This is reflected both in a higher average number of differences among sub-Saharan African genomes than among non-African genomes (Cann et al., 1987; Ramachandran et al., 2005) and in the fact that the ancestry found outside of Africa is largely a subset of that within it (Tishkoff et al., 2009). Today, some of the earliest-branching African lineages are present only in populations with relatively small census sizes, including the southern African Khoe-San (see STAR Methods for terminology), central African rainforest hunter-gatherers, and Hadza of Tanzania (Gronau et al., 2011; Schlebusch et al., 2012; Veeramah et al., 2012). However, the population structure of Africa prior to the
expansion of food producers (pastoralists and agriculturalists) remains unknown (Busby et al., 2016; Gurdasani et al., 2015; Patin et al., 2017). Bantu-speaking agriculturalists originating in western Africa are thought to have brought farming to eastern Africa by −2,000 years BP (years before present, defined by convention as years before 1950 CE) and to southern Africa by −1,500 BP, thereby spreading the largest single ancestry component to African genomes today (Russell et al., 2014; Tishkoff et al., 2009). Earlier migration(s), which brought ancestry related to the ancient Near East (Lazaridis et al., 2016; Pagani et al., 2012; Pickrell et al., 2014), brought herding to eastern Africa by −4,000 BP (Marshall et al., 1984) and to southern Africa by −2,000 BP (Sadr, 2015).

RESULTS

To reconstruct African population structure prior to the spread of food production, we generated new genome-wide data from 15 ancient sub-Saharan Africans (Table 1; Tables S1 and S2; STAR Methods). For three individuals from the western Cape of South Africa (~2,300−1,300 BP), we carried out direct shotgun sequencing to 0.7- to 2.0-fold coverage. For 12 individuals from eastern and south-central Africa, we used in-solution enrichment of ~1.2 million single nucleotide polymorphisms (SNPs). These included four individuals from the coastal region of Kenya and Tanzania (~1,400–400 BP), one from interior Tanzania (~3,100 BP), and seven from Malawi (ranging over ~8,100–2,500 BP) (Figure S1). All individuals had postmortem degradation characteristic of ancient DNA (Table 1), and we confirmed that key results are unlikely to be artifacts of contamination by restricting analysis to sequences with postmortem degradation (Skoglund et al., 2012; Skoglund et al., 2014a) (Figure S2). We merged the new ancient DNA data with previously reported shotgun sequence data from a ~4,500 BP Ethiopian highland individual (Llorente et al., 2015) and with SNP genotypes from 584 present-day African individuals from 59 diverse populations (including new data from 34 Malawi individuals; STAR Methods) (Lazaridis et al., 2014; Patterson et al., 2012), as well as 300 high-coverage genomes from 142 worldwide populations (Mallick et al., 2016). We also determined mitochondrial DNA and Y chromosome haplogroups for all newly reported samples (Tables S3 and S4).

An Ancient Cline of Southern and Eastern African Hunter-Gatherers

We used principal component analysis (PCA) (Patterson et al., 2006) and automated clustering (Alexander et al., 2009) to relate the 16 ancient individuals to present-day sub-Saharan Africans (Figures 1 and S3). Whereas the two individuals buried in ~2,000 BP hunter-gatherer contexts in South Africa share ancestry with southern African Khoe-San populations in the PCA, 11 of the 12 ancient individuals who lived in eastern and south-central Africa between ~8,100 and ~400 BP form a gradient of relatedness to the eastern African Hadza on the one hand and southern African Khoe-San on the other (Figure 1A). The genetic cline correlates to geography, running along a north-south axis with ancient individuals from Ethiopia (~4,500 BP), Kenya (~400 BP), Tanzania (both ~1,400 BP), and Malawi (~8,100–2,500 BP), showing increasing affinity to southern Africans (both ancient individuals and present-day Khoe-San). The seven individuals from Malawi show no clear heterogeneity, indicating a long-standing and distinctive population in ancient Malawi that persisted for at least ~5,000 years (the minimum span of our radiocarbon dates) but which no longer exists today.

We constructed a model where ancient and present-day African populations trace their ancestry to a putative set of nine ancestral populations. As proxies for these populations, three different ancient Near Eastern populations and six African populations were used that, according to our analyses, harbor substantial ancestry related to major lineages present in Africa today. The Mende from Sierra Leone are used in this model to represent a component of ancestry that exists in high proportions in western African populations, the ancient southern African genomes (South_Africa_2000BP) are used to represent the ancestry of southern Africa before agriculture, the Ethiopian individual (Ethiopia_4500BP) is used to represent northeastern African ancestry before agriculture, the Mbuti are used to represent central African rainforest hunter-gatherer ancestry, the individual from an eastern African pastoralist context (Tanzania_Luxmanda_31000BP) is used to represent an early pastoralist lineage from eastern Africa (see below; Figure 2), and the Dinka (from Sudan) are used to represent distinctive ancestry found in Nilotic speakers today. The ancient Near Eastern populations were representative of Anatolia, the Levant, and Iran,
<table>
<thead>
<tr>
<th>ID</th>
<th>Population label</th>
<th>Date (BP denotes calibrated years before present [defined as 1950], while &quot;uncalibrated BP&quot; denotes the radiocarbon age)</th>
<th>Location</th>
<th>Lat.</th>
<th>Long.</th>
<th>Y chromosome haplogroup</th>
<th>mtDNA haplogroup</th>
<th>Damage rate at 5’ CpG dinucleotides</th>
<th>SNPs hit on 1.2M autosomal targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>I9028</td>
<td>South_Africa_2000BP</td>
<td>2241-1965 BP (2330 ± 25 uncalibrated BP, UGAMS-7255)</td>
<td>St. Helena, South Africa</td>
<td>−32.8</td>
<td>18.0</td>
<td>A1b1b2a</td>
<td>L0d2c1$</td>
<td>28%</td>
<td>731,098 (1.1X shotgun)</td>
</tr>
<tr>
<td>I9133</td>
<td>South_Africa_2000BP</td>
<td>2017-1748 BP (2000 ± 50 uncalibrated BP, Pta-5283)</td>
<td>Faraoskop Rock Shelter, South Africa</td>
<td>−32.0</td>
<td>18.5</td>
<td>A1b1b2a</td>
<td>L0d1b2b1b</td>
<td>33%</td>
<td>1,028,904 (2.3X shotgun)</td>
</tr>
<tr>
<td>I9134</td>
<td>South_Africa_1200BP</td>
<td>1282-1069 BP (1310 ± 50 uncalibrated BP, Pta-4373)</td>
<td>Kasteelberg, South Africa</td>
<td>−32.8</td>
<td>17.9</td>
<td>Female</td>
<td>L0d1a1a</td>
<td>21%</td>
<td>641,971 (0.8X shotgun)</td>
</tr>
<tr>
<td>I4427</td>
<td>Malawi_Fingira_6100BP</td>
<td>6175-5913 BP (5270 ± 25 uncalibrated BP, UCIAMS-186346)</td>
<td>Fingira, Malawi</td>
<td>−10.8</td>
<td>33.8</td>
<td>BT</td>
<td>L0d1b2b</td>
<td>37%</td>
<td>99,341</td>
</tr>
<tr>
<td>I4466</td>
<td>Malawi_Fingira_6100BP</td>
<td>6177-5923 BP (5290 ± 25 uncalibrated BP, UCIAMS-186347)</td>
<td>Fingira, Malawi</td>
<td>−10.8</td>
<td>33.8</td>
<td>BT</td>
<td>L0d1c</td>
<td>49%</td>
<td>30,257</td>
</tr>
<tr>
<td>I4421</td>
<td>Malawi_Chengerere_5200BP</td>
<td>5400-4800 BP (radiocarbon dating was unsuccessful; the date is based on context of other materials from the same site)</td>
<td>Chencherere, Malawi</td>
<td>−14.4</td>
<td>33.8</td>
<td>Female</td>
<td>L0k2</td>
<td>28%</td>
<td>59,470</td>
</tr>
<tr>
<td>I4422</td>
<td>Malawi_Chengerere_5200BP</td>
<td>5293-4979 BP (4525 ± 25 uncalibrated BP, UCIAMS-186348)</td>
<td>Chencherere, Malawi</td>
<td>−14.4</td>
<td>33.8</td>
<td>Female</td>
<td>L0k1</td>
<td>42%</td>
<td>9,355</td>
</tr>
<tr>
<td>I2966</td>
<td>Malawi_Hora_8100BP</td>
<td>10000-5000 BP (radiocarbon dating was unsuccessful; the date is based on context of other materials from the same site)</td>
<td>Hora, Malawi</td>
<td>−11.7</td>
<td>33.6</td>
<td>BT</td>
<td>L0k2 (PMDS &gt; 3)</td>
<td>54%</td>
<td>610,605 (+0.26X shotgun)</td>
</tr>
<tr>
<td>I2967</td>
<td>Malawi_Hora_8100BP</td>
<td>8173-7957 BP (7230 ± 60 uncalibrated BP, PSUAMS-2535)</td>
<td>Hora, Malawi</td>
<td>−11.7</td>
<td>33.6</td>
<td>Female</td>
<td>L0a2</td>
<td>48%</td>
<td>65,686 (+0.03X shotgun)</td>
</tr>
<tr>
<td>I4426</td>
<td>Malawi_Fingira_2500BP</td>
<td>2676-2330 BP (2676-2343 BP (2425 ± 20 uncalibrated BP, PSUAMS-1734), 2483-2330 BP (2400 ± 20 uncalibrated BP, PSUAMS-1881))</td>
<td>Fingira, Malawi</td>
<td>−10.8</td>
<td>33.8</td>
<td>Female</td>
<td>L0f</td>
<td>39%</td>
<td>635,427</td>
</tr>
<tr>
<td>I3726</td>
<td>Tanzania_Luxmanda_3100BP</td>
<td>3141-2890 BP (2925 ± 20 uncalibrated BP, ISGS-A3806)</td>
<td>Luxmanda, Tanzania</td>
<td>−4.3</td>
<td>35.3</td>
<td>Female</td>
<td>L2a1</td>
<td>65%</td>
<td>845,016</td>
</tr>
<tr>
<td>I0589</td>
<td>Tanzania_Zanzibar_1400BP</td>
<td>1370-1303 BP (1479 ± 23 uncalibrated BP, OxA-31427)</td>
<td>Kuumbi Cave, Zanzibar Island, Tanzania</td>
<td>−6.4</td>
<td>39.5</td>
<td>Female</td>
<td>L4b2a2c</td>
<td>22%</td>
<td>752,917</td>
</tr>
<tr>
<td>I1048</td>
<td>Tanzania_Pemba_1400BP</td>
<td>1421-1307 BP (1520 ± 30 uncalibrated BP, Beta-434912)</td>
<td>Makangale Cave, Pemba Island, Tanzania</td>
<td>−4.9</td>
<td>39.6</td>
<td>Female</td>
<td>L0a</td>
<td>42%</td>
<td>168,117</td>
</tr>
<tr>
<td>I2298</td>
<td>Tanzania_Pemba_600BP</td>
<td>639-544 BP (623 ± 20 uncalibrated BP, Wk-43308)</td>
<td>Makangale Cave, Pemba Island, Tanzania</td>
<td>−4.9</td>
<td>39.6</td>
<td>Female</td>
<td>L2a1a2</td>
<td>29%</td>
<td>695,242</td>
</tr>
<tr>
<td>I0595</td>
<td>Kenya_400BP</td>
<td>496-322 BP (398 ± 27 uncalibrated BP, OxA-30803)</td>
<td>Panga ya Saidi, Kenya</td>
<td>−3.7</td>
<td>39.7</td>
<td>E1b1b1b2</td>
<td>L4b2a2</td>
<td>30%</td>
<td>150,383</td>
</tr>
</tbody>
</table>

*Table S2 provides detailed information on the direct radiocarbon dating measurements.

$Consistent with previously published mtDNA sequence by Morris et al. (2014).
respectively (Lazaridis et al., 2016; Mathieson et al., 2015). We used qpAdm (Haak et al., 2015), a generalization of $f_2$ symmetry statistics, to successively test 1-source, 2-source, or 3-source models and admixture proportions for all other ancient and present-day African populations, with a set of 10 non-African populations as outgroups (STAR Methods).

We find that ancestry closely related to the ancient southern Africans was present much farther north and east in the past than is apparent today. This ancient southern African ancestry comprises up to 91% of the ancestry of Khoe-San groups today (Table S5), and also 31% ± 3% of the ancestry of Tanzania-Zanzibar_1400BP, 60% ± 6% of the ancestry of Malawi_Fingira_6100BP, and 65% ± 3% of the ancestry of Malawi_Fingira_2500BP (Figure 2A). Notably, the Khoe-San-related ancestry in ancient individuals from Malawi and Tanzania is symmetrically related to the two previously identified lineages present in the San ($Z < 2$; Figure S2), estimated to have diverged at least 20,000 years ago (Mallick et al., 2016; Pickrell et al., 2012; Schlebusch et al., 2012), implying that this was an ancient divergent branch of this group that lived in eastern Africa at least until 1,400 BP. However, it was not present in all eastern Africans, as we do not detect it in the ~400-year-old individual from coastal Kenya nor in the present-day Hadza.

**Displacement of Forager Populations in Eastern Africa**

Both unsupervised clustering (Figure 1B) and formal ancestry estimation (Figure 2B) suggest that individuals from the Hadza group in Tanzania can be modeled as deriving all their ancestry from a lineage related deeply to ancient eastern Africans such as the Ethiopia_4500BP individual (Figure 3A; Table S5). However, this lineage appears to have contributed little ancestry to present-day Bantu speakers in eastern Africa, who instead trace their ancestry to a lineage related to present-day western Africans, with additional components related to the Nilotic-speaking Dinka and to the Tanzania_Luxmanda_3100BP.

---

**Figure 1. Overview of Ancient Genomes and African Population Structure**

(A) Map of sampling locations in Africa and principal component analysis of all individuals. Present-day individuals are indicated with gray circles. (B) Automated clustering of key ancient and present-day populations (for K = 7 cluster components). Present-day populations are labeled in gray.
pastoralist (see below; Figure 2). The Sandawe, another population that like the Hadza uses click consonants in their spoken language, are modeled as having ancestry similar to the Hadza but also admixture related to that of neighboring populations (Figure 3A; Table S5) consistent with previous findings (Henn et al., 2011; Tishkoff et al., 2009). Population replacement by incoming food producers appears to have been nearly complete in Malawi, where we detect little if any ancestry from the ancient individuals who lived ~8,100–2,500 BP. Instead, present-day Malawian individuals are consistent with deriving all their ancestry from the Bantu expansion of ultimate western African origin (Figure 3).

Among the ancient individuals analyzed here, only a ~600 BP individual from the Zanzibar archipelago has a genetic profile similar to present-day Bantu speakers (Figure 1). Notably, this individual has even more western-African-related ancestry than the present-day Bantu speakers we analyzed from Kenya, who also derive some of their ancestry from lineages related to Dinka and Tanzania_Luxmanda_3100BP (Figure 1B). Using linkage disequilibrium, we estimate that this admixture between western- and eastern-African-related lineages occurred an average of 800–400 years ago (STAR Methods). This suggests a scenario of genetic isolation between early farmers and
previously established foragers during the initial phase of the Bantu expansion into eastern Africa (Crowther et al., 2017; Ribot et al., 2010), a barrier that broke down over time as mixture occurred. This parallels the patterns previously observed in genomic analyses of the Neolithic expansion into Europe (Haak et al., 2015; Skoglund et al., 2012) and the East Asian farming expansion into Remote Oceania (Skoglund et al., 2016). However, this process of delayed admixture did not always apply in Africa, as is evident in the absence of admixture from previously established hunter-gatherers in present-day Malawians.

**Early Levantine Farmer-Related Admixture in an ~3,100-Year-Old Pastoralist from Tanzania**

Western-Eurasian-related ancestry is pervasive in eastern Africa today (Pagani et al., 2012; Tishkoff et al., 2009), and the timing of this admixture has been estimated to be ~3,000 BP on average (Pickrell et al., 2014). We found that the ~3,100 BP individual (Tanzania_Luxmanda_3100BP), associated with a Savanna Pastoral Neolithic archeological tradition, could be modeled as having 38% ± 1% of her ancestry related to the nearly 10,000-year-old pre-pottery farmers of the Levant (Lazaridis et al., 2016), and we can exclude source populations related to early farmer populations in Iran and Anatolia. These results could be explained by migration into Africa from descendants of pre-pottery Levantine farmers or alternatively by a scenario in which both pre-pottery Levantine farmers and Tanzania_Luxmanda_3100BP descend from a common ancestral population that lived thousands of years earlier in Africa or the Near East. We fit the remaining approximately two-thirds of Tanzania_Luxmanda_3100BP as most closely related to the Ethiopia_4500BP (p = 0.029) or, allowing for three-way mixture, also from a source closely related to the Dinka (p = 0.18; the Levantine-related ancestry in this case was 39% ± 1%) (Table S4).
While these findings show that a Levant-Neolithic-related population made a critical contribution to the ancestry of present-day eastern Africans (Lazaridis et al., 2016), present-day Cushitic speakers such as the Somali cannot be fit simply as having Tanzania_Luxmanda_3100BP ancestry. The best fitting model for the Somali includes Tanzania_Luxmanda_3100BP ancestry, Dinka-related ancestry, and 16% ± 3% Iranian-Neolithic-related ancestry (p = 0.015). This suggests that ancestry related to the Iranian Neolithic appeared in eastern Africa after earlier gene flow related to Levant Neolithic populations, a scenario that is made more plausible by the genetic evidence of admixture of Iranian-Neolithic-related ancestry throughout the Levant by the time of the Bronze Age (Lazaridis et al., 2016) and in ancient Egypt by the Iron Age (Schuenemann et al., 2017).

Direct Evidence of Migration Bringing Pastoralism to Eastern and Southern Africa

In contrast to the Malawi and Zanzibar individuals, all three ancient southern Africans show affinities to the ancestry predominant in present-day Tuu speakers in the southern Kalahari more than to present-day Ju’hoan speakers in the northern Kalahari (Figures S2B and S2C). However, the ~1,200 BP sample from the western Cape that is found in a pastoralist context has a specific similarity in clustering analyses to present-day Khoe-Khoe-speaking pastoralist populations such as the Nama (Figure 1B), and like them it has affinity to three groups: Khoe-San, western Eurasians, and eastern Africans. This supports the hypothesis that a non-Bantu-related population carried eastern African and Levantine ancestry to southern Africa by at least around 1,200 BP, providing direct evidence for claims previously made based on analysis of present-day populations (Pickrell et al., 2014).

We used our modeling framework to show that the South_Africa_1200BP pastoralist individual from the western Cape is consistent with being a mixture of just two streams of ancestry relative to non-southern African populations, with 40.3% ± 2.3% ancestry related to the Tanzania_Luxmanda_3100BP individual (54% ± 7% when restricting analysis to sequences with postmortem damage) and the remaining being related to the South_Africa_2000BP hunter-gatherers (Table S5). This supports the hypothesis that the Savanna Pastoral Neolithic archaeological tradition in eastern Africa is a plausible source for the spread of herding to southern Africa. Even the Ju’hoan San group with the least genetic affinity to eastern Africans (Ju’hoan_North), have 9% ± 1% of their ancestry most closely related to Tanzania_Luxmanda_3100BP, consistent with previous findings that the ancestries of all present-day San and Khoe were affected by agropastoralist migrations in the last two millennia (Pickrell et al., 2014).

The Earliest Divergences among Modern Human Populations

Previous studies have suggested that the primary ancestry in the San is from a lineage that separated from all other lineages represented in modern humans today, before the latter separated from each other (Gronau et al., 2011; Veeramah et al., 2012). Such a model emerges when we automatically fit a tree without admixture to the data (Figure 3A), but we also find that a tree-like representation is a poor fit (Figure S4A), in the sense that ancient southern Africans who lived ~2,000 BP were not strictly an outgroup to extant lineages in other parts of sub-Saharan Africa. In particular, we find that ancient southern Africans, who have none of the eastern African admixture that is ubiquitous today, share significantly more alleles with present-day and ancient eastern Africans (including Dinka, Hadza, and Ethiopia_4500BP) than they do with present-day western Africans (Figure 3B; Table S6). Even within present-day western Africans, the genetic differences between Yoruba from Nigeria and the Mende from Sierra Leone are inconsistent with descent from a homogeneous ancestral population isolated from ancient southern Africans. The asymmetry between Yoruba and Mende is also observed with non-Africans but is no stronger than in eastern Africans (the most closely related Africans to the ancestral out-of-Africa population), and thus these signals are not driven by admixture from outside Africa and instead likely reflect demographic events entirely within Africa (Figure 3C; Table S6).

We carried out admixture graph modeling of the allele frequency correlations and found two parsimonious models that fit the data. The first posits that present-day western Africans harbor ancestry from a basal African lineage that contributed more to the Mende than it did to the Yoruba, with the other source of western African ancestry being related to eastern Africans and non-Africans (Figures 3D, S4, and S5; Table S7). The second model posits that long-range and long-standing gene flow has connected southern and eastern Africa to some groups in western Africa (e.g., the ancestors of the Yoruba) to a greater extent than to other groups in western Africa (e.g., the ancestors of the Mende) (Figure 3E) (Pleurdeau et al., 2012). The possible basal western African population lineage would represent the earliest known divergence of a modern human lineage that contributed a major proportion of ancestry to present-day humans. Such a lineage must have separated before the divergence of San ancestors, which is estimated to have begun on the order of 200–300 thousand years ago (Scally and Durbin, 2012). Such a model of basal western African ancestry might support the hypothesis that there has been ancient structure in the ancestry of present-day Africans, using a line of evidence independent from previous findings based on long haplotypes with deep divergences from other human haplotypes (Hammer et al., 2011; Lachance et al., 2012; Plagnol and Wall, 2006). One scenario consistent with this result could involve ancestry related to eastern Africans (and the out-of-Africa population) expanding into western Africa and mixing there with more basal lineages. Our genetic data do not support the theory that this putative basal lineage diverged prior to the ancestors of Neanderthals, since the African populations we analyze here are approximately symmetrically related to Neanderthals (Mallick et al., 2016; Prüfer et al., 2014).

A Selective Sweep Targeting a Taste Receptor Locus in Southern Africa

The availability of ancient African genomes provides an opportunity to search for genomic footprints of natural selection, manifested as regions of greater allele frequency differentiation between ancient and present-day populations than predicted by the genome-wide background. We compared the two ancient
southern African ~2,000 BP shotgun sequence genomes to six present-day high-coverage San genomes with minimal recent mixture. The small number of ancient individuals does not permit inference of changing allele frequencies at single loci, so we performed a scan for high allele frequency differentiation in 500 kb windows with a step size of 10 kb. Using ~500 windows spaced at least 5 million base pairs apart as a null distribution, we found that the most differentiated locus was 15 standard deviations from the observed genome-wide mean and overlapped a cluster of eight taste-receptor genes on chromosome 12 (Figure 4A; Table 2). Taste-receptor genes have previously been identified as targets of natural selection in humans, as they modulate the ability to detect poisonous compounds in plants (Campbell et al., 2012).

Figure 4. Ancient Genomes Provide Evidence of Natural Selection in Present-Day Southern African San Populations
We computed branch-specific allele frequency differentiation in six present-day high-coverage San genomes compared to a pool of two ~2,000 BP South African genomes as an outgroup using two approaches.
(A) We computed the statistic in windows of 500 kb separated by 10 kb. We also estimated genome-wide average and standard deviation of the statistic using windows separated by at least 5 Mb, and we transformed the genome-wide distribution of the sliding windows to be approximately normal (right). We observed outliers 15 standard deviations from the mean in a taste-receptor gene cluster on chromosome 12 and a secondary peak in the Keratin Associated Protein 4 gene cluster. The outgroup used was 4 Central African Mbuti genomes. See Table 2 for details on all major outlying regions.
(B) Illustration of the branch-specific allele frequency differentiation approach.
(C) y axis: We computed the statistic and block jackknife standard errors for 208 gene ontology categories with at least 50 genes each. The outgroup used was western Africans. x axis: As a control to confirm that outlier categories do not show larger magnitudes of allele frequency differentiation across populations, we replaced the present-day San with the central African Mbuti.
DISCUSSION

This study, which multiplies by 16-fold the number of individuals with genome-wide ancient DNA data from sub-Saharan Africa, highlights the power of ancient African genomes to provide insights into prehistoric events that are difficult to discern based solely on analysis of present-day genomes. We reveal the presence of a hitherto unknown cline of geographically structured hunter-gatherer populations stretching from Ethiopia to South Africa, which we show existed prior to the great population transformations that occurred in the last few thousand years in association with the spread of herders and farmers. We also document deeper structure in western Africa, possibly predating the divergence of the ancestors of southern African hunter-gatherers from other population lineages. We finally provide case examples of how populations in eastern and southern Africa were transformed by the spread of food producers and show how the process gave rise to interactions with the previously established hunter-gatherers, with the outcomes ranging from no detectable mixture in present-day populations to substantial mixture. Our documentation of a radically different landscape of human populations before and after the spread of food producers highlights the difficulty of reconstructing the African past based solely on analysis of present-day populations and the importance of using ancient DNA to study deep African population history in an era in which technological improvements have now made this feasible. It is clear that ancient DNA studies with larger sample sizes and covering a broader geographic range have the potential to make major progress in improving our understanding of African prehistory.

STAR METHODS

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

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Termination
  - Panga ya Saidi Cave, Kilifi County, Kenya (n = 1)
  - Makangale Cave, Pemba, Tanzania (n = 2)
  - Kuumbi Cave, Zanzibar, Tanzania (n = 1)
  - Luxmanda, Babati District, Tanzania (n = 1)
  - Hora, Malawi (n = 2)
  - Fingira, Malawi (n = 3)
  - Chencereere II, Malawi (n = 2)
  - St. Helena Bay, South Africa (n = 1)
  - Faraoskop, South Africa (n = 1)
  - Kasteelberg, South Africa (n = 1)
ACKNOWLEDGMENTS

Permission to analyze the remains from Kenya and Tanzania was granted by the National Museums of Kenya; the Antiquities Division of the Ministry of Natural Resources and Tourism, Tanzania; and the Zanzibar Department of Museums and Antiquities. Permission to analyze the remains from Malawi was granted by the Malawi Department of Museums and Monuments and the University of Cape Town. We thank I. Lazaridis, M. Lipson, I. Mathiesen, and S. Tishkoff for discussions and I. Kucukkalipci and K. Majander for laboratory support. P.S. was supported by the Wenner-Gren Foundation and the Swedish Research Council (VR grant 2014-453). J.C.T was supported by a grant from the Program for the Enhancement of Research at Emory University. J.K. and A.M. were supported by the DFG grant KR 4015/1-1 and the Max Planck Society. K.Si. was supported by NSF grant BCS-1613577. M.Ha., A.H., M.Me., and S.P. were supported by the Max Planck Society. A.G.M. and J.P. were supported by the National Research Foundation of South Africa. R.R. was supported by the South African Medical Research Council. N.B. was supported by ERC starting grant SEALINKS (206148), and R.P. was supported by ERC starting grant ADNABIOARC (263441). M.G.T. was supported by Wellcome Trust Senior Investigator Award (grant number 100719/Z/12/Z). D.R. was supported by NIH grant GM100233 and by NSF HOMINID BCS-1032255 and is a Howard Hughes Medical Institute investigator. The laboratory at Penn State was supported by the NSF Archaeology program (BCS-1460369, D.J.K.).

Received: December 12, 2016
Revised: July 1, 2017
Accepted: August 29, 2017
Published: September 21, 2017

REFERENCES

Chami, F. (2009). Zanzibar and the Swahili coast from c. 30,000 years ago (E&D Vision Pub.).


