

# Working toward a synthesis of archaeological, linguistic, and genetic data for inferring African population history

Laura B. Scheinfeldt<sup>a</sup>, Sameer Soi<sup>b</sup>, and Sarah A. Tishkoff<sup>a,c,1</sup>

<sup>a</sup>Department of Genetics, <sup>b</sup>Genomics and Computational Biology Graduate Group, and <sup>c</sup>Department of Biology, University of Pennsylvania, Philadelphia, PA 19104

Although Africa is the origin of modern humans, the pattern and distribution of genetic variation and correlations with cultural and linguistic diversity in Africa have been understudied. Recent advances in genomic technology, however, have led to genome-wide studies of African samples. In this article, we discuss genetic variation in African populations contextualized with what is known about archaeological and linguistic variation. What emerges from this review is the importance of using independent lines of evidence in the interpretation of genetic and genomic data in the reconstruction of past population histories.

genetic variation | human evolution | mitochondrial DNA

**D**isentangling past population histories is a formidably complicated task that benefits from the synthesis of archaeological, linguistic, and genetic data. Archaeology permits insights into ancient technology and culture and provides a timetable for the emergence of innovations. Historical linguistic data complement the archaeological record by contributing an independent phylogenetic analysis of language relationships and providing clues about ancient population migration and admixture events. Similarly, genetic data provide an independent data source to understand the biological relationships among modern peoples and likely points of origin and expansion of their ancestors. Undoubtedly, the specific details of human demographic history are more complex than any synthesis can account for, but we are focusing here on the overlap among the archaeological, linguistic, and genetic data collected in Africa to make inferences about African demographic history.

## African Language Family Classification

Africa is home to almost a third of all modern languages, encompassing >2,000 ethno-linguistic groups (1) that have largely been classified into four language families: Niger-Kordofanian, Afroasiatic, Nilo-Saharan, and Khoesan. As displayed in Fig. 1, Niger-Kordofanian languages are spoken throughout western Africa, eastern Africa, central Africa, and southern Africa and include the common Bantu languages. The Afroasiatic language family includes languages spoken in northern, central, and eastern Africa such as Cushitic, Chadic, Semitic, and ancient Egyptian. The Nilo-Saharan language family is spoken predominantly in central and eastern Africa and includes the Sudanic and Nilotic languages. The Khoesan language family, which includes languages that contain click consonants and is spoken by hunter-gatherer populations in eastern (Hadza and Sandawe) and southern Africa [the San, referred to here as “southern African Khoesan” (“SAK”)], is the most contentious of the African language families because there is so much divergence among the Hadza, Sandawe, and SAK languages (2, 3).

## Modern Human Origins and Migration out of Africa

The earliest emergence of anatomically modern humans in the fossil record occurred in eastern Africa 200–150 thousand years ago (kya) (4–7). Although the earliest dated modern humans outside of Africa were identified in the Middle East ~90 kya (5, 8–

11), there was no continuous occupation of regions outside of Africa until ~60–40 kya; modern human remains are documented in Papua New Guinea 60–40 kya (12), southwest Asia ~35 kya, Europe ~40 kya, and mainland Asia ~35 kya (5). Therefore, over half of modern human history took place within Africa exclusively, and understanding patterns of variation within Africa is critical for the elucidation of modern human demographic history.

Genetic data from extant modern humans complement the fossil record in the reconstruction of modern human origins. The uniparentally inherited mitochondrial DNA (mtDNA) and non-recombinant portion of the Y chromosome (NRY) are two loci that have been extensively studied in human populations, in part because they represent the maternal and paternal population histories, respectively, in a population sample and in part because they do not undergo recombination and, therefore, lineages can be more easily traced back to a single common ancestor. Unfortunately, the mtDNA and NRY loci are single loci, which are susceptible to the effects of natural selection and genetic drift because they have smaller effective population sizes relative to the autosomes and because any selective pressure will impact the entire locus. Thus, combined mtDNA, NRY, and autosomal data are necessary for a thorough understanding of any population history.

MtDNA, NRY, and autosomal DNA studies demonstrate that the highest levels of genetic variation are present in African samples relative to non-Africans, consistent with a model of African ancestry for all modern humans (e.g., refs. 1, 13–19). Further, phylogenetic analysis of mtDNA and NRY variation reveals that the deepest phylogenetic clades are found exclusively in African samples and all non-African lineages derive from a subset of these African lineages (15, 16, 20–24). Consistent with the archaeological record, estimates of the time to the most recent ancestor (TMRCA) for the mtDNA lineages give an age range of ~200–100 kya (23–26) and similar results have been published for NRY lineages, ~200–65 kya (26–28). Therefore, the genetic data corroborate a model in which modern humans arose in African 200–100 kya and subsequently, one or more populations split off and migrated out of Africa. The migration out of Africa was accompanied by a population bottleneck, which resulted in a reduction in genetic diversity in non-African populations relative to Africans (29).

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, “In the Light of Evolution IV: The Human Condition,” held December 10–12, 2009, at the Arnold and Mabel Beckman Center of the National Academies of Sciences and Engineering in Irvine, CA. The complete program and audio files of most presentations are available on the NAS Web site at [www.nasonline.org/SACKLER\\_Human\\_Condition](http://www.nasonline.org/SACKLER_Human_Condition).

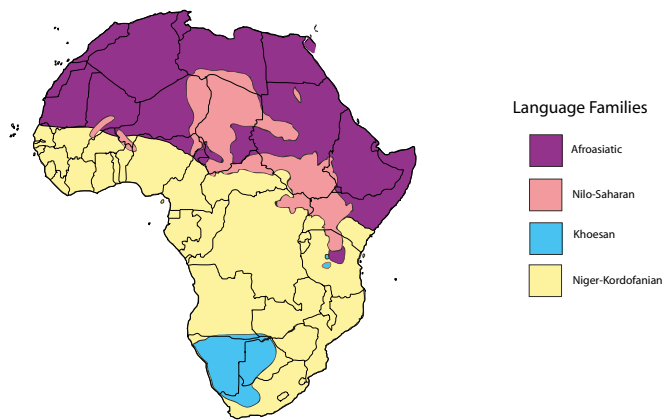
Author contributions: S.A.T. designed research; S.A.T. supervised analyses; L.B.S. and S.S. analyzed data; and L.B.S. and S.A.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

<sup>1</sup>To whom correspondence should be addressed. E-mail: [tishkoff@mail.med.upenn.edu](mailto:tishkoff@mail.med.upenn.edu).

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1002563107/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1002563107/-DCSupplemental).



**Fig. 1.** Map of Africa colored by the language family spoken in each region (adapted from ref. 29). The Afroasiatic language family is shown in purple, the Nilo-Saharan language family is shown in pink, the Khoesan language family is shown in blue, and the Niger-Kordofanian language family is shown in yellow.

### Middle Stone Age in Africa

The Middle Stone Age, which took place ~250–40 kya (30), is a period in the archaeological record that indicates a significant change in culture and subsistence technology in Africa. Several sites in eastern, central, and southern Africa contain artifacts consistent with a shift in technology and population expansion ~75–55 kya, including hunting weapons, indications of increased plant utilization, signs of increased marine exploitation, and evidence of large-scale movement of red ochre (used for art), stone, and shell ornaments (30–33). It is tempting to speculate that these developments are tied to improvements in human communication; however, the reconstruction of proto-languages does not extend back this far in time; therefore, there is no empirical way to establish when or where human language emerged. Interestingly, an analysis of mtDNA data estimates a population expansion in Africa 70 kya (34), consistent with the archaeological evidence from the late Middle Stone Age. Furthermore, we would not expect to see the same signal of expansion in non-African populations given that the extreme bottleneck associated with the migration out of Africa most likely obscures more ancient demographic signals.

### Neolithic in Africa

The Neolithic period, beginning ~10 kya, included the development of agriculture and animal domestication in Africa, with concomitant changes in population demographics due to population growth and migration to new regions. Below we discuss several such movements including the spread of agriculture, the spread of pastoralism, and the dispersal of affiliated language groups and genetic lineages. It is important to note, however, that these associations among linguistic, archaeological, and genetic data are not presented here to paint a simple picture of migration or replacement, but rather to illustrate that large-scale movements of technology and culture have resulted in detectable amounts of gene flow among the involved peoples and that the interpretation of extant genetic patterns benefits from an understanding of the combined data.

**Neolithic in Northern Africa.** Approximately 14 kya, climatic changes associated with the end of the Last Glacial Maximum resulted in regions around the world becoming more favorable to human exploitation. Northern Africa is one such region, and ~13 kya, novel technologies (“Natufian”) thought to be the immediate precursor to agricultural technologies emerged and were associated with semisedentary subsistence and population expansions in

northeastern Africa (35). Moreover, before the emergence of the Natufian styled artifacts, the archaeological record includes two artifact styles, the “Geometric Kebaran” and the “Mushabian” associated with Middle Eastern and Northern African populations, respectively (35). The archaeological evidence suggests the peoples using these assemblages interacted for well over 1,000 years, and linguistic evidence suggests that the peoples using these assemblages may have spoken some form of proto-Afroasiatic (35, 36). Although the origins of the Afroasiatic language family remain contentious, linguistic data generally support a model in which the Afroasiatic language family arose in Northern Africa >10 kya (36). Moreover, analyses of the Cushitic branch of the Afroasiatic language family suggest that proto-Cushitic arose and diversified at least 7 kya, and this likely took place in Ethiopia (37).

Intriguingly, the origin and diversification of proto-Afroasiatic is consistent with the spread of intensive plant collection in the archaeological record, and some interpret this pattern to represent a model in which proto-Afroasiatic speakers developed the novel subsistence technology resulting in the expansion and spread of their Afroasiatic descendants in the region (37). Some examples of the relevant linguistic data include reconstructed Chadic root words for “porridge” and “sorghum” and the Cushitic root words for “grain” and “wheat” (37). Because these and other root words are present in many of the Chadic and Cushitic languages, it is assumed that they were present in the proto-Chadic and proto-Cushitic languages and therefore must be as old as those proto-languages (37).

The genetic data appear to be consistent with the archaeological and linguistic data indicative of extensive population interactions between North African and Middle Eastern populations. A recent NRY study explores the distribution of haplogroups in a sample of African, Middle Eastern, and European males (38). Whereas a subclade of haplogroup E (M35) appears to have arisen in eastern Africa over 20 kya and subsequently spread to the Middle East and Europe, haplogroup J (M267) appears to have arisen in the Middle East over 20 kya and subsequently spread into northern Africa (38). A recent study of genomewide autosomal microsatellite markers reports that Middle Eastern and African samples share the highest number of alleles that are also absent in other non-African samples, consistent with bidirectional gene flow (1). In addition, a recent study of domestic goat mtDNA and NRY variation reports similar findings as well as evidence of trade along the Strait of Gibraltar (39). The combined archaeological, linguistic, and genetic data, therefore, suggest bidirectional migration of peoples between northern Africa and the Levant for at least the past ~14 ky.

**Neolithic in Sahel.** There is increasing archaeological, linguistic, and genetic evidence that the Sahel has been an important region for bidirectional migration between western and eastern Africa (1, 40–42). Linguistic evidence indicates population interactions for ~20–10 kya between the Nilo-Saharan and Afroasiatic speakers in this region (43). The combined linguistic and archaeological data support a model in which the Nilo-Saharan language family arose in eastern Sudan >10 kya and Nilo-Saharan speakers subsequently migrated westward to Lake Chad and southward into southern Sudan (1, 44). Linguistic data also suggest that ~7 kya, proto-Chadic Afroasiatic speakers migrated from the Sahara into the Lake Chad Basin (45). This possibility is supported by an analysis of NRY variation that finds that the pattern and distribution of haplogroup R (V88) are consistent with the emergence of proto-Chadic ~7 kya and subsequent expansion of this linguistic group into the Lake Chad Basin (46). Whereas the inferred migration route is not consistent between NRY and mtDNA analyses, perhaps due to sex-biased migration, studies of mtDNA corroborate a model in which Sahel is a corridor for bidirectional migration between eastern and western Africa and, on the basis of the distribution

of haplogroup L3f3, the proto-Chadic speakers expanded from eastern Africa into the Lake Chad Basin (47, 48).

**The Spread of Pastoralism.** Archaeological data suggest that the emergence of animal husbandry in northeastern Africa took place as early as ~11 kya (49). Archaeological studies in Nabta Playa (in Egypt's Western Desert) reveal a spectrum of artifacts consistent with pastoralism and adaptation to the desert environment, including particular pottery styles (Khartoum tradition), evidence of well technology, and cattle burials (49, 50). By ~8 kya, evidence is present of imported (from the Middle East) sheep or goat remains in northeastern Africa (e.g., ref. 50). Some controversy persists in the archaeological community regarding whether cattle domestication was developed in northern Africa or imported from the Middle East; however, recent DNA analysis of extant indigenous African bovine taurine and zebu cattle (51) supports a model in which the earliest emergence of pastoralism involving taurine cattle took place in northeastern Africa and subsequently spread westward and southward (51). A recent analysis of NRY variation in 13 eastern and southern African population samples suggests that the spread of pastoralism from eastern Africa to southern Africa was accompanied by migration of pastoral peoples as well as pastoral technology as evidenced by the distribution of NRY haplogroup M293 (and the subclade E3b1f-M293) (22). Furthermore, the most likely source for this migration based on the samples included in Henn et al. (22) would have been the southern Nilotic speaking Datog (because the haplotype frequency and diversity of M293 is highest in the Datog) ~2 kya (22).

Ehret (52) inferred the history of pastoralism in Africa from a linguistic analysis of shared cognates. His findings support a relatively ancient emergence of pastoralism in northeastern Africa corresponding to Eastern Sudanic, Central Sudanic, and possibly Southern Cushitic speakers, followed by the subsequent spread of cattle keeping to western and southern Africa (52). The relatively ancient emergence of pastoralism in the archeological record is supported by the reconstruction of proto-Cushitic languages. For example, there are at least two words for cattle that are thought to be relatively old, one in Northern Cushitic and the other in Central Cushitic. In proto-Cushitic, the word "hlee," which translates to "head of cattle," is related to the Southern Cushitic (Mbugu) word "hline," which translates to "heifer" (52), and so on. Furthermore, estimates of linguistic diversity of vocabulary related to cattle suggest that cattle keeping arose in northeastern Africa and subsequently spread to western and southern Africa (52).

Ehret (52) also argues that the spread of cattle milking was separate and more recent than the spread of cattle keeping. He discusses the assumption that the spread of cattle milking would require some discernible impact on the language used to discuss it (52). For example, the proto-Bantu word for milk is related to the proto-Bantu word for breast, but there are several root words for milk (many likely borrowed from other languages) among the Bantu languages. However, there is only one root word for milking (literally to squeeze). This observation, is interpreted to support a model in which a Bantu population in Tanzania borrowed the word (possibly from the southern Cushitic speakers) representing milking as well as the actual technology related to cattle milking and subsequently spread the technology to other Bantu speaking populations (52).

The shift from food gathering to food producing inferred from African archaeological and linguistic data also resulted in a detectable genetic signal. This relationship between subsistence, culture, and biology due to gene/culture coevolution is one that has been of special interest in human genetics studies. Models of Darwinian (i.e., positive) selection are consistent with subsistence being an environmental factor that can have a profound effect on patterns of genetic variation, and the emergence of agriculture and pastoralism is tied to increased population densities and dietary

changes. Thus, genetic variants that conveyed a selective advantage in this shift in diet from foraging to animal and plant products would have persisted and increased in frequency in agricultural and pastoralist communities.

Lactase persistence is one of the better studied examples of gene/culture coevolution (e.g., 53, 54). In most mammals, once an individual is weaned, it loses the ability to produce the enzyme lactase-phlorizin hydrolase (LPH), which is necessary to digest the sugar lactose present in milk without gastric distress (55). The majority of humans do not express this enzyme as adults (referred to as the "lactase nonpersistence" phenotype) (56). Several widespread mutations, however, result in the continued production of LPH into adulthood, a trait often referred to as lactase persistence (57). The distribution of the lactase persistence phenotype is intriguing given what is known about subsistence patterns worldwide (Fig. 2). Lactase persistence is present at high frequency in Northern European dairying and African pastoralist populations; at moderate frequency in southern European and Middle Eastern populations; and at low frequency in nonpastoral Asian, Pacific, American, and African populations (55). In Europeans, the most common mutation associated with lactase persistence is thought to be a regulatory mutation located upstream of the gene that encodes LPH (a T at position -13910), within intron 13 of the neighboring *MCM6* gene (56, 58). Further, this mutation is located within a large linkage disequilibrium block that is thought to have arisen ~20–2 kya, consistent with recent positive selection related to the emergence of cattle domestication and milk consumption ~10 kya in the Middle East (59, 60).

In African populations, the lactase persistence phenotype is generally highest in pastoral populations (55–57, 61, 62). However, with the exception of the Fulani and Hausa populations (62), other African pastoralist populations do not have the T-13910 mutation associated with the lactase persistence trait (57, 61). Recent studies have identified at least three additional and independent mutations that are associated with lactase persistence in East African pastoralist populations: C-14010, which is most common in Kenya and Tanzania (57); G-13907, which is present at low to moderate frequency in northeast Africa (57, 61); and G-13915, which is most common in the Middle East (60) and northeastern Africa (57, 61) and may be associated with camel domestication in the Middle East ~6 kya (60). Tishkoff et al. (57) demonstrated that all three variants result in significant increases in gene expression levels driven by the lactase promoter.

The most common variant within Africa associated with lactase persistence (C-14010) is also located within an extremely large linkage disequilibrium block (2 Mb) and is thought to have arisen ~6.8–2.7 kya in either the agropastoralist Afroasiatic populations that migrated into Kenya and Tanzania from Ethiopia within the past 5,000 years or the Nilo-Saharan pastoralist populations that migrated into the region from southern Sudan within the past 3,000 years, and the variant then subsequently spread throughout pastoral populations in eastern



**Fig. 2.** Global map showing the frequency of the lactase persistence trait for populations reported in Ingram et al. (55) and citations therein. Lactase persistence is shaded in black.

Africa relatively rapidly, consistent with the spread of pastoralism into sub-Saharan Africa ~4.5 kya (57). The estimates of the selection coefficients of the African mutations (0.035–0.097) are among the highest reported for modern humans, and intuitively this makes sense given not only the increased nutritional value of drinking milk as an adult but also the increased source of water in regions such as the Sahara where dehydration and diarrhea can cause death.

**Bantu Expansion.** In Sub-Saharan Africa, the long-range exchange networks of Neolithic technology and associated spread of Bantu languages (which we refer to here as the “Bantu expansion” for the sake of simplicity) have had a major influence on biological and cultural diversity in sub-Saharan Africa. On the basis of archaeological and linguistic data, the Bantu languages and associated agricultural and iron age technologies are thought to have originated in Nigeria or Cameroon (63) ~5,000 years ago (64, 65) and spread relatively rapidly across sub-Saharan Africa. The extent to which this was associated with the migration of populations vs. a diffusion of language and technology among populations has been debated.

The linguistic classification of the ~600 Bantu languages is interpreted to represent several dispersals throughout sub-Saharan Africa (e.g., ref. 66). Ehret (67) argues that proto-Bantu diverged into several daughter clades, all but one of which are spoken only in the northwestern region of the Bantu-speaking areas (i.e., western central Africa), and the other of which was a forest Savanna Bantu clade. Ehret (67) goes on to argue that the forest Savanna Bantu clade diverged into several daughter clades, including the Savanna Bantu clade, and this diversification is linked to the spread of Bantu languages into central and southern Africa. The Savanna Bantu clade includes most of the contemporary languages spoken in eastern Africa, southeastern Africa, southwestern Africa, and the southern Savanna belt. This reconstruction supports a model in which proto-Bantu emerged in western central Africa ~5,000 years ago and diversified and spread across the rainforest for ~2,000 years before the first archaeological evidence of eastern Bantu speakers in the Great Lakes region (67).

Archaeological evidence related to the Bantu expansion largely focuses on the distribution of Early and Late Iron Age sites in Africa. Phillipson argues that the Eastern Bantu languages likely arose in western central Africa around the time of the emergence of Early Iron Age artifacts consistent with cattle keeping, but that the spread of Eastern Bantu languages is associated with the distribution of “later Iron Age” sites in central and southern Africa (68).

There is also a genetic signature of past population movements thought to be associated with the Bantu expansion. The large majority of genetic analyses have focused on mtDNA and NRY data. Overall, both datasets tie particular mtDNA (e.g., L0a, L2a, L3b, and L3e) (25, 69–73) and NRY [e.g., E3a (M2/M180), E2 (M75), and B2a (M150)] (22, 65, 72) lineages to the Bantu expansion, because they are found in the highest frequencies in extant Bantu-speaking populations. Interestingly, comparative studies of mtDNA and NRY variation suggest different maternal and paternal population histories related to the Bantu expansion (71, 72). Specifically, NRY variation in regions affected by the Bantu expansion is low relative to mtDNA variation and consists almost exclusively of haplogroup lineages associated with the Bantu expansion (71). Conversely, the mtDNA haplogroup lineages in the same samples include lineages associated with the Bantu expansion as well as lineages that are thought to have been present in the region before the Bantu expansion (21). This discrepancy is largely attributed to sex-biased migration and gene flow due to the practice of patrilocality and/or polygyny (71, 74), both of which are common in present-day Bantu-speaking populations. Moreover, this pattern of sex-biased gene flow is docu-

mented independently in other regions of the world such as the Pacific Islands (75, 76). Both loci, however, are more susceptible to genetic drift than autosomal loci because of their relatively smaller effective population sizes; therefore, some of the differential male/female patterns may be attributed to chance. A recent analysis of genomewide autosomal data is consistent with a large genetic impact of the Bantu expansion on most of sub-Saharan Africa, as evidenced by the presence of Niger-Kordofanian ancestry in many central, eastern, and southern African populations (1). In addition, Tishkoff et al. (1) documented evidence from their analysis of genomewide autosomal loci of a distinct Bantu migration from eastern to southern Africa, which is consistent with the archaeological and linguistic evidence of dispersal of Bantu technology and languages from the Great Lakes region of East Africa (67).

### Contemporary African Genetic and Linguistic Variation

Scholars have studied language relationships within a cladistic framework since at least the early 19th century (77), and given the parallels in linguistic and genetic change over time, it is not unreasonable to use linguistic affiliations as a way of grouping individuals for genetic study. Several studies have demonstrated a correlation between linguistic and genetic variation, including cases in Europe (78, 79), Asia (80), the Pacific (75, 76, 81, 82), and the Americas (83–86). The main difficulty in these studies lies in the interpretation of linguistic similarities among populations. Whereas language sharing obviously results from some degree of contact among peoples, the horizontal transmission of language can occur with little to no genetic exchange. Likewise, there can be genetic exchange with little or no linguistic exchange. Therefore, the degree of correlation between genetic and linguistic variation varies depending on the populations being studied.

Studies of genetic variation within Africa, as mentioned above, have found extensive amounts of genetic variation relative to non-Africans owing to the fact that the “out of Africa” bottleneck significantly reduced genetic variation in non-Africans; however, most genetic studies of African populations are limited by the number of population samples included. More recent work has improved the understanding of genetic variation in Africa with a survey of genomewide genetic variation in geographically and ethnically diverse African samples (1). Tishkoff et al. (1) analyzed 1,327 genomewide autosomal microsatellite and insertion/deletion polymorphisms in 121 African population samples and a comparative sample of 1,394 non-Africans. The authors (1) studied population structure and relationships using the program STRUCTURE (87), among other phylogenetic analyses. The STRUCTURE program uses a model-based Bayesian clustering approach to identify genetic subpopulations and assign individuals probabilistically to these subpopulations on the basis of their genotypes, while simultaneously estimating ancestral population allele frequencies. The program STRUCTURE places individuals into  $K$  clusters, where  $K$  is chosen in advance and is varied across independent runs, and individuals can have membership in multiple clusters (87). Tishkoff et al. (1) inferred 14 ancestral population clusters globally as well as within Africa and found that the African samples cluster geographically as well as linguistically and ethnically (Table 1). In addition to the STRUCTURE analysis, the authors (1) constructed a neighbor-joining tree on the basis of pairwise population genetic distances that showed that the African samples clustered primarily by geographic region and to a lesser extent by linguistic affiliation with a few notable exceptions. The pygmies from central Africa, for example, clustered near the southern African San.

Several studies have looked at the relationship between genetic and linguistic variation in African samples (1, 21, 22, 40, 88–90). For example, an NRY study of Nilo-Saharan, Niger-Congo, and Afroasiatic speakers in Sudan revealed a strong correlation (Mantel test:  $r = 0.31$ ,  $P = 0.007$ ) between linguistic

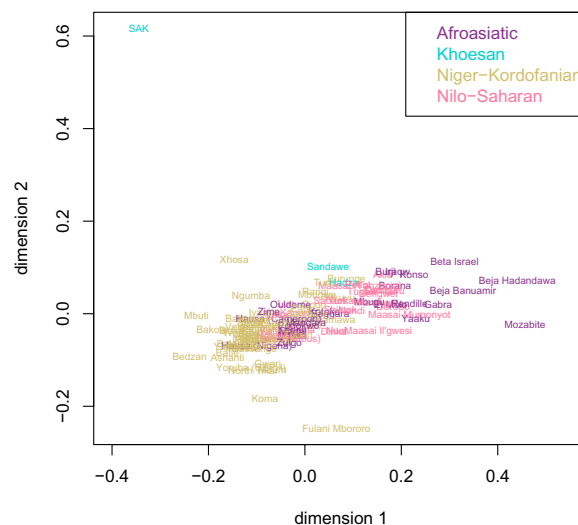
**Table 1. Inferred population clusters using the STRUCTURE analysis of autosomal microsatellite and insertion/deletion polymorphism data from global populations adapted from ref. 1**

K	Emerging clusters
2	African, non-African
3	East Asian, Oceanic, Native American
4	Eastern African
5	Hadza, Sandawe, SAK, Pygmy
6	Western Pygmy
7	Chadic, Nilo-Saharan
8	Indian, Oceanic
9	Oceanic
10	Native American
11	Mbuti Pygmy, SAK
12	Chadic/Nilo-Saharan speakers from northern Cameroon, Chad, and southern Sudan
13	Sandawe
14	Fulani

and NRY variation (40), and in this case the correlation between linguistic and genetic variation was stronger than the correlation between geographic and genetic distances (Mantel test:  $r = 0.29$ ,  $P = 0.025$ ). Similarly, a study of mtDNA and NRY variation in 40 African samples representing all four language families reports a significant correlation between genetic and linguistic distances (Mantel of NRY,  $r = 0.32$ ,  $P = 0.001$ ; Mantel of mtDNA,  $r = 0.23$ ,  $P = 0.016$ ) (71).

The single-locus studies of genetic and linguistic correlation are consistent with the regression analysis reported by Tishkoff et al. (1) that documents significant correlations between linguistic and genetic distances within the Niger-Kordofanian and Nilo-Saharan language families after correction for geographic distances. To further explore the relationship among genetic and linguistic variation in Africa, we used the published dataset of genomewide data from Tishkoff et al. (1) that includes 103 population samples ( $n \geq 10$ ) that speak languages representing all four African language families. We first performed a Mantel test to determine to what extent genetic and linguistic distances are correlated within language families. Not surprisingly, all three tests showed that linguistic and genetic distances were significantly correlated (with 100,000 permutations): Niger-Kordofanian,  $r = 0.32$ ,  $P = 9.99 \times 10^{-6}$ ; Nilo-Saharan,  $r = 0.29$ ,  $P = 9.99 \times 10^{-6}$ ; and Afroasiatic,  $r = 0.27$ ,  $P = 9.99 \times 10^{-6}$  (the linguistic relationships among the Khoesan speakers are not clearly understood and therefore did not permit the construction of a linguistic distance matrix needed to perform a Mantel test); and the correlation coefficient is >25% in all three tests.

Because we and others (1) have established a significant correlation between linguistic affiliation and genetic variation within three of the African language families, we wanted to explore to what degree samples plotted by genetic distance cluster by language family. We used multidimensional scaling (MDS) to construct a two-dimensional plot of a pairwise genetic distance matrix taken from the above-mentioned 103 population samples (1). Consistent with the mtDNA and NRY studies discussed above (40, 71), our genomewide analysis of microsatellite data shows that populations generally cluster on the basis of both geographic region and linguistic classification. Fig. 3 demonstrates that populations generally separate by linguistic affiliation along dimension 1. Dimension 2 separates the SAK speakers from all other Africans including the eastern Khoesan speakers, the Hadza and Sandawe, that cluster closely with other eastern Africans. Another interesting pattern that emerges in the MDS plot that is consistent with previous work (1) is the clustering of



**Fig. 3. Multidimensional scaling (MDS) analysis of autosomal microsatellite data from Tishkoff et al. (1).** A pairwise genetic distance matrix using  $(\delta\mu)^2$  (as described in ref. 1) was constructed for populations with a sample size of  $n \geq 10$  and used for MDS analysis. Populations are colored on the basis of linguistic affiliation. The Afroasiatic speakers are shown in purple, the Nilo-Saharan speakers are shown in pink, the Khoesan speakers are shown in blue, and the Niger-Kordofanian speakers are shown in yellow. The x axis represents dimension 1 and the y axis represents dimension 2.

the Afroasiatic Chadic speakers with the Nilo-Saharan speakers, which may reflect a past language shift (1).

Because the distribution of language families in Africa roughly follows a geographic distribution (Fig. 1), we also performed MDS within geographic regions that include at least three language families. In central Africa (Fig. S1), the samples cluster by language family with a few notable exceptions. For example, the Fulani who are nomadic pastoralists that speak a Niger-Kordofanian language and reside across central and western Africa do not cluster with other Niger-Kordofanian-speaking populations. Moreover, the Fulani are distinguished from other African samples at  $K = 14$  in Tishkoff et al.'s (1) STRUCTURE analysis. Morphological analyses of the Fulani have been interpreted to suggest a Middle Eastern origin for the Fulani (91), and there has been some speculation based on linguistic data that the Fulani migrated to central Africa from northern Africa or the Middle East (91). In addition, there is evidence of shared recent ancestry among the Fulani and European/Middle Eastern samples from studies of mtDNA (92), NRY (40), and autosomal microsatellites (1) and from the presence in this population of the mutation associated with lactose tolerance in Europeans (T-13910) (62).

Whereas previous work on mtDNA (92) is consistent with a West African origin for the Fulani (consistent with other Bantu speakers), the NRY data reveal that the Fulani share recent ancestry with Nilo-Saharan and Afroasiatic speaking populations (40). As in other cases where the maternal and paternal patterns of population history are not in agreement, this result could reflect differential patterns of Fulani male and female migration and gene flow, or it could reflect the influence of genetic drift or some combination of the two. A more recent analysis of genomewide autosomal data shows that the Fulani cluster most closely with the Chadic- and Central Sudanic-speaking populations (1). This result is consistent with our MDS analysis in which both Fulani cluster most closely with the Chadic- and Central Sudanic-speaking populations, as well as with the Baggara (Semitic). The clustering of the Baggara near the Fulani is also consistent with Tishkoff et al. (1), who report that the Baggara share ancestry with the Fulani and with the Chadic speakers.

To a lesser extent, the Hausa from Nigeria and Cameroon cluster more closely with the Niger-Kordofanian speakers along dimension 2 (Fig. S1). This result is consistent with previous genetic analysis (1) and with linguistic analysis of the Hausa that suggests extensive interaction between the Hausa (who speak an Afroasiatic Chadic language) and Niger-Kordofanian speakers as evidenced by an analysis of loanwords (93).

In eastern Africa (Fig. S2), dimension 1 separates the Afroasiatic and Niger-Kordofanian samples, and dimension 2 separates the Nilo-Saharan samples. As in Fig. 3, the Hadza and Sandawe do not separate from the eastern African samples along either dimension to any large extent, although they do cluster closely to each other (Fig. S2), and this pattern is consistent with extensive regional gene flow with neighboring populations. The other noteworthy pattern in this plot is the Luo sample (Fig. S2), who speak a Western Nilotic language but cluster separately from other Nilo-Saharan speakers along dimension 1, together with Bantu-speaking populations. This clustering is consistent with previous findings that the Luo show predominately Bantu ancestry (1) and may reflect high levels of admixture among the Luo and geographically nearby Bantu populations (94).

**History of Hunter–Gatherer Populations.** As mentioned previously, the classification of languages within the Khoesan language family is contentious given the high diversity within each subclade and extreme divergence among them (95, 96), particularly for the Sandawe and Hadza. A common classification, therefore, groups the three languages spoken in South Africa into a separate branch (SAK) from the more divergent Sandawe and Hadza (97). One interpretation of this extreme linguistic diversity is that the last common ancestor of the language family must be extremely ancient, and Ehret (95) estimates the TMRCA to be at least 20 kya (which approaches the limit in timescale to linguistic reconstruction). The Sandawe and SAK are more similar to each other linguistically than either one is to the Hadza. Geographically, however, the Sandawe and Hadza are extremely close to each other (150 km apart in Tanzania), and both are geographically distant from the SAK populations residing in southern Africa.

A recent study of mtDNA and NRY variation investigates the genetic relationship among the Hadza, Sandawe, and SAK (21). The authors find that in general, the Hadza and Sandawe are more genetically similar to each other than either one is to the SAK. However, the Sandawe and SAK share ancient mtDNA lineages, which may suggest an ancient common ancestry. For example, mtDNA haplogroup L0d is present at high frequency in the SAK and at low frequency in the Sandawe, but is not present in the Hadza samples (21), and the TMRCA estimate of the SAK and Sandawe L0d lineages is ancient (~60 kya) (21). Similarly, the SAK and Sandawe share NRY haplogroup A (M91), which is not present in the Hadza samples (21). On the other hand, haplogroup L4g is common in both the Sandawe and the Hadza and absent from the SAK samples, and the TMRCA for the Sandawe and Hadza L4g is more recent (~25 kya) (21). And all three samples share NRY haplogroup B2b (M112) (21). The authors (21) discuss more than one interpretation of these results. The absence of mtDNA haplogroup L0d and NRY haplogroup A (M91) from the Hadza could reflect loss due to genetic drift because there is evidence of a recent bottleneck in the Hadza (98). Alternatively, the pattern of haplogroup variation could reflect an ancient linguistic and genetic divergence of the Hadza from the SAK. Moreover, the authors (21) performed a likelihood analysis to estimate the time of divergence among the populations and found that the divergence between the Hadza and the Sandawe was >20 kya and the divergence between the Hadza/Sandawe and the SAK was >40 kya. Additional studies of mtDNA and NRY variation have identified ancient shared lineages among the SAK and the Hadza as well as several other eastern African populations (28, 38, 99–101). Consistent with the mtDNA and NRY

data, our MDS analysis shows that the Hadza and Sandawe cluster closely together with each other and with other eastern African populations (Fig. 3). Additionally, the Hadza are slightly farther from the SAK than the Sandawe along both dimensions (Fig. 3).

Tishkoff et al. (1) provide evidence for an ancient common ancestry of Khoesan and Pygmy populations, suggesting the possibility of a proto-Khoesan hunter–gatherer population in eastern Africa that diverged >30 kya. STRUCTURE analysis revealed that the pygmies cluster together with other hunter–gatherer samples, including the SAK, Hadza, and Sandawe at low  $K$  values ( $K = 3$ ), and then differentiate at higher  $K$  values ( $K = 5$ ) (Table 1). The analysis also shows that the Mbuti pygmies cluster with the SAK at higher  $K$  values ( $K = 7$ ), which could be due to either common ancestry or more recent gene flow. In addition, recent work on mtDNA, NRY, and autosomal data estimated the TMRCA of the pygmy and agricultural populations to be approximately 70–60 kya and the TMRCA of western and eastern pygmies to be approximately 20 kya (73, 102, 103). The findings of Tishkoff et al. (1) raise the possibility that the pygmy populations, who have lost their indigenous language, once spoke some form of proto-Khoesan with click consonants. Interestingly, linguistic analysis of the SAK suggests that they originated in eastern Africa and possibly as far north as Ethiopia before migrating into southern Africa, consistent with the identification of rock art in the Sandawe homeland and in southern Africa that is thought to be related to Khoesan speakers (104). There is further evidence that, although there has not been recent gene flow among these populations, there has been recent admixture between the Sandawe and neighboring populations as well as between the pygmies and neighboring populations, and this recent admixture may be obscuring the more ancient relationships among the hunter–gatherer populations (1). Future analyses that incorporate data from across the genome together with full-likelihood or approximate Bayesian computation methods will be necessary to more fully understand these complex population histories.

## Conclusions

We have presented here a synthesis of the archaeological, linguistic, and genetic data used to infer African population history. The general picture that emerges is that genetic variation in Africa is structured geographically and to a lesser extent linguistically. This is consistent with the fact that populations in close geographic proximity to each other as well as populations that speak linguistically similar languages are more likely to exchange genes. The pattern of genetic variation in Africa is also consistent with geographic barriers limiting gene flow as exemplified by the geographic/genetic distinction between northern African and sub-Saharan African populations. When we focus, however, on particular exceptions to these broad patterns, we are able to more fully appreciate the complex population histories that have contributed to extant patterns of genetic variation. The development of sequencing and genotyping technologies is advancing at an unprecedented rate and is allowing for the genotyping of millions of single-nucleotide polymorphisms and the sequencing of millions of nucleotides across populations. These data, coupled with computational methods for inferring demographic parameters and testing demographic models (e.g., maximum likelihood and approximate Bayesian computation), are well powered to refine our understanding of African past population histories. The incorporation of archaeological and linguistic data will be important for establishing testable hypotheses and elucidating the evolutionary processes (or forces) that have shaped the genomic landscape in Africa.

**ACKNOWLEDGMENTS.** We thank Scott Williams, Françoise Friedlaender, William Beggs, Charla Lambert, Joseph Jarvis, Michael Campbell, and the Tishkoff Lab members for support and helpful discussions. This work was funded by National Science Foundation Grants BCS-0196183 and BCS-0827436, National Institutes of Health Grants R01GM076637 and 1R01GM083606-01, and National Institutes of Health Pioneer 1-DP1-OD-006445-01 awards (to S.A.T.).

1. Tishkoff SA, et al. (2009) The genetic structure and history of Africans and African Americans. *Science* 324:1035–1044.
2. Nurse D (1997) The contributions of linguistics to the study of history in Africa. *J Afr Hist* 38:359–391.
3. Sands B (1998) *The Linguistic Relationship between Hadza and Khoisan* (Rudiger Kupper, Köln, Germany).
4. McDougall I, Brown FH, Fleagle JG (2005) Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature* 433:733–736.
5. Trinkaus E (2005) Early modern humans. *Annu Rev Anthropol* 34:207–230.
6. McDermott F, et al. (1996) New Late-Pleistocene uranium–thorium and ESR dates for the Singa hominid (Sudan). *J Hum Evol* 31:507–516.
7. Clark JD, et al. (2003) Stratigraphic, chronological and behavioural contexts of Pleistocene Homo sapiens from Middle Awash, Ethiopia. *Nature* 423:747–752.
8. Stringer CB, Grun R, Schwarcz HP, Goldberg P (1989) ESR dates for the hominid burial site of Es Skhul in Israel. *Nature* 338:756–758.
9. Schwarcz HP, et al. (1988) ESR dates for the hominid burial site of Qafzeh in Israel. *J Hum Evol* 17:733–737.
10. McDermott F, Grun R, Stringer CB, Hawkesworth CJ (1993) Mass-spectrometric U-series dates for Israeli Neanderthal/early modern hominid sites. *Nature* 363:252–255.
11. Mercier NVH, Bar-Yosef O, Vandermeersch B, Stringer C, Joron JL (1993) Thermoluminescence date for the Mousterian burial site of Es-Skhul, Mt. Carmel. *J Archaeol Sci* 20:169–174.
12. Groube L, Chappell J, Muke J, Price D (1986) A 40,000 year-old human occupation site at Huon Peninsula, Papua New Guinea. *Nature* 324:453–455.
13. International HapMap Consortium (2003) The International HapMap Project. *Nature* 426:789–796.
14. Akey JM, et al. (2004) Population history and natural selection shape patterns of genetic variation in 132 genes. *PLoS Biol* 2:e286.
15. Underhill PA, et al. (2001) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann Hum Genet* 65:43–62.
16. Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. *Nature* 325:31–36.
17. Frazer KA, et al. (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449:851–861.
18. Garrigan D, et al. (2007) Inferring human population sizes, divergence times and rates of gene flow from mitochondrial, X and Y chromosome resequencing data. *Genetics* 177:2195–2207.
19. Li JZ, et al. (2008) Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319:1100–1104.
20. Gonder MK, Mortensen HM, Reed FA, de Sousa A, Tishkoff SA (2007) Whole-mtDNA genome sequence analysis of ancient African lineages. *Mol Biol Evol* 24:757–768.
21. Tishkoff SA, et al. (2007) History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation. *Mol Biol Evol* 24:2180–2195.
22. Henn BM, et al. (2008) Y-chromosomal evidence of a pastoralist migration through Tanzania to southern Africa. *Proc Natl Acad Sci USA* 105:10693–10698.
23. Behar DM, et al. (2008) The dawn of human matrilineal diversity. *Am J Hum Genet* 82:1130–1140.
24. Ingman M, Kaessmann H, Paabo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. *Nature* 408:708–713.
25. Salas A, et al. (2002) The making of the African mtDNA landscape. *Am J Hum Genet* 71:1082–1111.
26. Tang H, Siegmund DO, Shen P, Oefner PJ, Feldman MW (2002) Frequentist estimation of coalescence times from nucleotide sequence data using a tree-based partition. *Genetics* 161:447–459.
27. Scozzari R, et al. (1999) Combined use of biallelic and microsatellite Y-chromosome polymorphisms to infer affinities among African populations. *Am J Hum Genet* 65:829–846.
28. Underhill PA, et al. (2000) Y chromosome sequence variation and the history of human populations. *Nat Genet* 26:358–361.
29. Campbell MC, Tishkoff SA (2008) African genetic diversity: Implications for human demographic history, modern human origins, and complex disease mapping. *Annu Rev Genomics Hum Genet* 9:403–433.
30. Henshilwood CS, et al. (2002) Emergence of modern human behavior: Middle Stone Age engravings from South Africa. *Science* 295:1278–1280.
31. Mellars P (2006) Why did modern human populations disperse from Africa ca. 60,000 years ago? A new model. *Proc Natl Acad Sci USA* 103:9381–9386.
32. Henshilwood CS, d'Errico F, Marean CW, Milo RG, Yates R (2001) An early bone tool industry from the Middle Stone Age at Blombos Cave, South Africa: Implications for the origins of modern human behaviour, symbolism and language. *J Hum Evol* 41:631–678.
33. McBrearty S, Brooks AS (2000) The revolution that wasn't: A new interpretation of the origin of modern human behavior. *J Hum Evol* 39:453–563.
34. Excoffier L, Schneider S (1999) Why hunter-gatherer populations do not show signs of Pleistocene demographic expansions. *Proc Natl Acad Sci USA* 96:10597–10602.
35. Bar-Yosef O (1987) Pleistocene connections between Africa and Southwest Asia: An archaeological perspective. *Afr Archaeol Rev* 5:29–38.
36. Ehret C, Keita SO, Newman P (2004) The origins of Afroasiatic. *Science*, 306:1680, and author reply (2004) 306:1680.
37. Ehret C (1979) On the antiquity of agriculture in Ethiopia. *J Afr Hist* 20:161–177.
38. Semino O, et al. (2004) Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: Inferences on the neolithization of Europe and later migratory events in the Mediterranean area. *Am J Hum Genet* 74:1023–1034.
39. Pereira F, et al. (2009) Tracing the history of goat pastoralism: New clues from mitochondrial and Y chromosome DNA in North Africa. *Mol Biol Evol* 26:2765–2773.
40. Hassan HY, Underhill PA, Cavalli-Sforza LL, Ibrahim ME (2008) Y-chromosome variation among Sudanese: Restricted gene flow, concordance with language, geography, and history. *Am J Phys Anthropol* 137:316–323.
41. Bereir RE, et al. (2007) Co-introgression of Y-chromosome haplogroups and the sickle cell gene across Africa's Sahel. *Eur J Hum Genet* 15:1183–1185.
42. Cerny V, Salas A, Hajek M, Zaloudkova M, Brdicka R (2007) A bidirectional corridor in the Sahel-Sudan belt and the distinctive features of the Chad Basin populations: A history revealed by the mitochondrial DNA genome. *Ann Hum Genet* 71:433–452.
43. Cavalli-Sforza LL, Menozzi P (1994) *History and Geography of Human Genes* (Princeton Univ Press, Princeton).
44. Ehret C (1983) *Population Movement and Culture Contact in the Southern Sudan*, eds Mack J, Robertshaw P (British Institute in Eastern Africa, Nairobi), pp 19–48.
45. Newman J (1997) *The Peopling of Africa* (Yale Univ Press, New Haven, London).
46. Cruciani F, et al. (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*, 10.1038/ejhg.2009.231.
47. Cerny V, et al. (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.
48. Cerny V, Salas A, Hajek M, Zaloudkova M, Brdicka R (2007) A bidirectional corridor in the Sahel-Sudan belt and the distinctive features of the Chad Basin populations: A history revealed by the mitochondrial DNA genome. *Ann Hum Genet* 71:433–452.
49. Wendorf F, Schild R (1998) Nabta Playa and its role in northeastern African prehistory. *J Anthropol Archaeol* 17:97–123.
50. McDonald M (1998) Early African pastoralism: View from Dakhleh Oasis (south central Egypt). *J Anthropol Archaeol* 17:124–142.
51. Hanotte O, et al. (2002) African pastoralism: Genetic imprints of origins and migrations. *Science* 296:336–339.
52. Ehret C (1967) Cattle-keeping and milking in eastern and southern African history: The linguistic evidence. *J Afr Hist* 8:1–17.
53. Durham W (1991) *Coevolution: Genes, Culture, and Human Diversity* (Stanford Univ Press, Palo Alto, CA).
54. Hollox E, Swallow D (2002) *Lactase Deficiency: Biological and Medical Aspects of the Adult Human Lactase Polymorphism* (Oxford Univ Press, Oxford).
55. Ingram CJ, Mulcare CA, Itan Y, Thomas MG, Swallow DM (2009) Lactose digestion and the evolutionary genetics of lactase persistence. *Hum Genet* 124:579–591.
56. Swallow DM (2003) Genetics of lactase persistence and lactose intolerance. *Annu Rev Genet* 37:197–219.
57. Tishkoff SA, et al. (2007) Convergent adaptation of human lactase persistence in Africa and Europe. *Nat Genet* 39:31–40.
58. Enattah NS, et al. (2002) Identification of a variant associated with adult-type hypolactasia. *Nat Genet* 30:233–237.
59. Edwards CJ, et al. (2007) Mitochondrial DNA analysis shows a Near Eastern Neolithic origin for domestic cattle and no indication of domestication of European aurochs. *Proc Biol Sci* 274:1377–1385.
60. Enattah NS, et al. (2008) Independent introduction of two lactase-persistence alleles into human populations reflects different history of adaptation to milk culture. *Am J Hum Genet* 82:57–72.
61. Ingram CJ, et al. (2007) A novel polymorphism associated with lactose tolerance in Africa: Multiple causes for lactase persistence? *Hum Genet* 120:779–788.
62. Mulcare CA, et al. (2004) The T allele of a single-nucleotide polymorphism 13.9 kb upstream of the lactase gene (LCT) (C-13.9kbT) does not predict or cause the lactase-persistence phenotype in Africans. *Am J Hum Genet* 74:1102–1110.
63. Greenberg JH (1972) Linguistic evidence regarding Bantu origins. *J Afr Hist* 13:189–216.
64. Phillipson D (1975) The chronology of the Iron Age in Bantu Africa. *J Afr Hist* 16:321–342.
65. Berniell-Lee G, et al. (2009) Genetic and demographic implications of the Bantu expansion: Insights from human paternal lineages. *Mol Biol Evol* 26:1581–1589.
66. Vansina J (1995) New linguistic evidence and 'the Bantu expansion'. *J Afr Hist* 36:173–195.
67. Ehret C (2001) Re-envisioning a central problem of early African history. *Int J Afr Hist Stud* 34:5–41.
68. Phillipson D (1976) Archaeology and Bantu linguistics. *World Archaeol* 8:65–82.
69. Pereira L, et al. (2001) Prehistoric and historic traces in the mtDNA of Mozambique: Insights into the Bantu expansions and the slave trade. *Ann Hum Genet* 65:439–458.
70. Plaza S, et al. (2004) Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola. *Hum Genet* 115:439–447.
71. Wood ET, et al. (2005) Contrasting patterns of Y chromosome and mtDNA variation in Africa: Evidence for sex-biased demographic processes. *Eur J Hum Genet* 13:867–876.
72. Beleza S, Gusmao L, Amorim A, Carracedo A, Salas A (2005) The genetic legacy of western Bantu migrations. *Hum Genet* 117:366–375.
73. Quintana-Murci L, et al. (2008) Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers. *Proc Natl Acad Sci USA* 105:1596–1601.
74. Pilkington MM, et al. (2008) Contrasting signatures of population growth for mitochondrial DNA and Y chromosomes among human populations in Africa. *Mol Biol Evol* 25:517–525.
75. Friedlaender JS, et al. (2008) The genetic structure of Pacific Islanders. *PLoS Genet* 4:e19.
76. Scheinfeldt L, et al. (2006) Unexpected NRY chromosome variation in Northern Island Melanesia. *Mol Biol Evol* 23:1628–1641.

77. Atkinson QD, Gray RD (2005) Curious parallels and curious connections—phylogenetic thinking in biology and historical linguistics. *Syst Biol* 54:513–526.
78. Piazza A, et al. (1995) Genetics and the origin of European languages. *Proc Natl Acad Sci USA* 92:5836–5840.
79. Cavalli-Sforza L, Feldman M (1981) *Cultural Transmission and Evolution* (Princeton Univ Press, Princeton).
80. Karafet T, et al. (2001) Paternal population history of East Asia: Sources, patterns, and microevolutionary processes. *Am J Hum Genet* 69:615–628.
81. Merriwether DA, et al. (1999) Mitochondrial DNA variation is an indicator of austronesian influence in Island Melanesia. *Am J Phys Anthropol* 110:243–270.
82. Robledo R, Scheinfeldt L, Merriwether DA, Thompson F, Friedlaender J (2003) A 9.1-kb insertion/deletion polymorphism suggests a common pattern of genetic diversity in Island Melanesia. *Hum Biol* 75:941–949.
83. Eshleman JA, et al. (2004) Mitochondrial DNA and prehistoric settlements: Native migrations on the western edge of North America. *Hum Biol* 76:55–75.
84. Malhi RS, Schults BA, Smith DG (2001) Distribution of mitochondrial DNA lineages among Native American tribes of Northeastern North America. *Hum Biol* 73:17–55.
85. Smith DG, et al. (2000) Implications of the distribution of Albumin Naskapi and Albumin Mexico for new world prehistory. *Am J Phys Anthropol* 111:557–572.
86. Wang S, et al. (2007) Genetic variation and population structure in native Americans. *PLoS Genet* 3:e185.
87. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
88. Lane AB, et al. (2002) Genetic substructure in South African Bantu-speakers: Evidence from autosomal DNA and Y-chromosome studies. *Am J Phys Anthropol* 119:175–185.
89. Sanchez-Mazas A (2001) African diversity from the HLA point of view: Influence of genetic drift, geography, linguistics, and natural selection. *Hum Immunol* 62:937–948.
90. Bryc K, et al. (2010) Genome-wide patterns of population structure and admixture in West Africans and African Americans. *Proc Natl Acad Sci USA* 107:786–791.
91. Ehret C (2008) The early livestock raisers of southern Africa. *South Afr Humanit* 30: 7–35.
92. Cerny V, et al. (2006) MtDNA of Fulani nomads and their genetic relationships to neighboring sedentary populations. *Hum Biol* 78:9–27.
93. Ehret C, ed (2006) *The Nilo-Saharan Background of Chad* (Ohio State University, Columbus, OH).
94. Bennett P (1983) Patterns in linguistic geography and the Bantu origins controversy. *Hist Afr* 10:35–51.
95. Ehret C, ed (2000) *Language and History* (Cambridge Univ Press, Cambridge, UK).
96. Sands B, ed (1998) *The Linguistic Relationship Between Hadza and Khoisan* (Rudiger Kupper, Köln, Germany).
97. Heine B, Nurse D (2000) *African Languages: An Introduction* (Cambridge Univ Press, Cambridge, UK).
98. Blurton Jones NG, Smith LC, O'Connell JF, Hawkes K, Kamuzora CL (1992) Demography of the Hadza, an increasing and high density population of Savanna foragers. *Am J Phys Anthropol* 89:159–181.
99. Knight A, et al. (2003) African Y chromosome and mtDNA divergence provides insight into the history of click languages. *Curr Biol* 13:464–473.
100. Semino O, Santachiara-Benerecetti AS, Falaschi F, Cavalli-Sforza LL, Underhill PA (2002) Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny. *Am J Hum Genet* 70:265–268.
101. 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* 70:1197–1214.
102. Patin E, et al. (2009) Inferring the demographic history of African farmers and pygmy hunter-gatherers using a multilocus resequencing data set. *PLoS Genet* 5:e1000448.
103. Destro-Bisol G, et al. (2004) The analysis of variation of mtDNA hypervariable region 1 suggests that Eastern and Western Pygmies diverged before the Bantu expansion. *Am Nat* 163:212–226.
104. Lim L (1992) *A Site-Oriented Approach to Rock Art: A Study from Usandawe, Central Tanzania*. PhD thesis (Brown University, Providence, RI).