Evolutionary Genetics and Admixture in African Populations

Aaron Pfennig [1], Lindsay N. Petersen[2], Paidamoyo Kachambwa[2], and Joseph Lachance [1,*]

[1]School of Biological Sciences, Georgia Institute of Technology, Atlanta, Georgia
[2]Mediclinic Precise Southern Africa, Cape Town, South Africa

*Corresponding author: E-mail: joseph.lachance@biology.gatech.edu.

Accepted: 22 March 2023

Abstract

As the ancestral homeland of our species, Africa contains elevated levels of genetic diversity and substantial population structure. Importantly, African genomes are heterogeneous: They contain mixtures of multiple ancestries, each of which have experienced different evolutionary histories. In this review, we view population genetics through the lens of admixture, highlighting how multiple demographic events have shaped African genomes. Each of these historical vignettes paints a recurring picture of population divergence followed by secondary contact. First, we give a brief overview of genetic variation in Africa and examine deep population structure within Africa, including the evidence of ancient introgression from archaic "ghost" populations. Second, we describe the genetic legacies of admixture events that have occurred during the past 10,000 years. This includes gene flow between different click-speaking Khoe-San populations, the stepwise spread of pastoralism from eastern to southern Africa, multiple migrations of Bantu speakers across the continent, as well as admixture from the Middle East and Europe into the Sahel region and North Africa. Furthermore, the genomic signatures of more recent admixture can be found in the Cape Peninsula and throughout the African diaspora. Third, we highlight how natural selection has shaped patterns of genetic variation across the continent, noting that gene flow provides a potent source of adaptive variation and that selective pressures vary across Africa. Finally, we explore the biomedical implications of population structure in Africa on health and disease and call for more ethically conducted studies of genetic variation in Africa.

Key words: admixture, africa, demographic history, evolutionary genetics, population genetics, population structure.

Significance

Despite recent progress, African populations are still dramatically underrepresented in genetic studies, and more studies of African genetic variation and population structure are needed. Such studies may not only hold new insights about human origins but are also crucial for equitable biomedical research, with implications that possibly extend beyond Africa. In this review, we provide an overview of our current understanding of how admixture—mostly during the last 10,000 years—has shaped present-day population structure in Africa and how recent genetic studies complement linguistics and archeology in reconstructing the history of African populations.

Introduction

Africa exhibits vast cultural and linguistic diversity, including a wide range of subsistence strategies and ~2,000 spoken languages. In addition, African populations harbor the greatest genetic diversity, exhibit the lowest levels of **linkage disequilibrium** (LD), have the largest long-term **effective population sizes** (N_e), and show the deepest split times of all human lineages (Tishkoff et al. 2009; Auton et al. 2015; Mallick et al. 2016; Bergström et al. 2020). For these reasons, Africa is commonly accepted as

Glossary

- **Linkage disequilibrium (LD)**—The nonrandom association of two alleles at different loci.
- **Effective population size (N_e)**—The number of breeding individuals in an idealized randomly mating population. N_e determines the strength of genetic drift acting on a population.
- **Population structure**—Systematic differences in allele frequencies between subpopulations.
- **Admixture**—The interbreeding of individuals from two or more subpopulations that were isolated for a relatively short evolutionary time.
- **Out-of-Africa (OOA) model** —Hypothesis that anatomically modern humans evolved in Africa and subsequently peopled the rest of the world.
- **Serial founder effect**—The successive loss of genetic variation when populations are sequentially founded by a small number of individuals.
- **Haplotype**—A set of linked genetic variants that are coinherited.
- **Population bottleneck**—An event that drastically reduces the effective size of a population, leading to increased genetic drift.
- **Genetic cline**—A gradual change of allele frequencies over a specified geographic area.
- **Genetic ancestry**—The genealogical paths through which an individual inherits DNA from specific ancestors in a reference population. Individuals with shared genetic ancestry tend to be more genetically similar.
- **Principal components**—A set of uncorrelated variables derived from the original data set through linear transformations, which maximize the variance between samples and reduce the dimensionality of the data while preserving the most important information.
- **Gene flow**—The movement of individuals and their genetic material from one population to another population.
- **Introgression**—The interbreeding of individuals from two or more populations that were isolated for a long evolutionary time but are not yet reproductively isolated.
- **Holocene**—The current geological epoch that started after the Last Glacial Maximum ~12 kya.
- **Iron Age**—The period of time during human prehistory when people began making tools from iron and steel, extending from ~4 to 1.5 kya in Africa.
- **Neolithic (New Stone Age)**—The period of time when people began using more sophisticated stone tools, leading to the emergence of farming and herding, extending from ~12 kya to 6.5 kya in Africa.
- **Paleolithic (Old Stone Age)**—The period of time in human evolution when people initially started using stone tools, extending from ~3.3 million years ago (Mya) to 12 kya.
- **Uniparental markers**—Mitochondrial DNA and Y chromosomes, which are transmitted exclusively maternally or paternally without recombination.
- **Fine-mapping**—The processes of refining the location of trait-associated variants in the genomic region of interest to identify likely causal variants based on association statistics and linkage disequilibrium patterns.
- **Isolation-by-distance model**—A theoretical framework explaining how genetic differentiation between populations increases with geographic distance due to spatially limited gene flow, that is, decreasing migration rate with increasing distance.
- **F_{ST} (Fixation index)**—The extent of genetic differentiation of two populations. Higher values of F_{ST} are indicative of greater population structure.

the cradle of humankind (Henn et al. 2018), and African population history is of exceptional interest to human evolution.

Most of our knowledge about African population history is derived from archeological and linguistic studies, as Africa has long been neglected in genetic studies (Popejoy and Fullerton 2016; Martin et al. 2018; Sirugo et al. 2019; Fatumo et al. 2022). However, archeological and linguistic studies are largely unable to disentangle cultural diffusion from demic diffusion, that is, movements of people (Robertson and Bradley 2000; Diamond and Bellwood 2003). By contrast, genetic studies are uniquely equipped to identify large-scale demic movements (e.g., Tishkoff et al. 2009). In the last decade, the importance of studying genetic variation in Africa has become more appreciated, and a heap of genetic studies of contemporary and ancient individuals has revealed complex **population structure** and history in Africa, complementing archeological and linguistic studies (e.g., Tishkoff et al. 2009; Schlebusch et al. 2012; Choudhury et al. 2020; Lipson et al. 2022).

In this review, we focus on genetic studies that uncovered extensive archaic, prehistoric, and recent **gene flow** that has occurred in Africa. We start by putting genetic

variation in Africa into a global context and giving a brief overview of population structure in Africa inferred from ancient and extant genomes, focusing on hunter–gatherer groups and deep population structure in the continent. We then discuss how this population structure was shaped by archaic and recent **admixture**, moving from the deeper past to more recent times. Given the scope of this review paper, we cannot comprehensively review the evolutionary history of every population. Instead, we focus on representative examples of major migratory and admixture events. Finally, we briefly review the evidence for local adaption and discuss the biomedical implication of population structure in Africa. In light of this, we call for more (responsibly conducted) studies of genetic variation in Africa and research capacity building on the African continent. Note that we tried to refer to populations according to current naming conventions, and when we refer to admixture between specific populations, this does not necessarily imply the mixing of these exact populations, but rather the mixing of genetically similar populations.

Patterns of Genetic Variation in Africa

Compared with the rest of the world, each African genome harbors ~25% more polymorphisms than each non-African genome (Auton et al. 2015; Mallick et al. 2016; Bergström et al. 2020). Furthermore, variants that are rare on a global level (<1% frequency) are more frequently found to be common in African populations, that is, there is an excess of variants exclusively found in Africans (Auton et al. 2015). Greater numbers of private African alleles are consistent with the **out-of-Africa** (OOA) model, as substantial numbers of polymorphisms were lost due to **serial founder effects**. In fact, the genetic variation found outside of Africa is largely a subset of African genetic diversity (Tishkoff et al. 2009; Lachance et al. 2012). Subsets of African genetic variation found outside of Africa also vary by region, indicating that multiple OOA migrations may have occurred (Rasmussen et al. 2011; Pagani et al. 2015). Additionally, African populations exhibit a faster decay of LD, leading to shorter **haplotypes** (Auton et al. 2015). This is of biomedical relevance (see below), and it also enables improved **fine-mapping** of causal variants in genome-wide association studies (GWAS) because casual variants are tagged by fewer other variants (Auton et al. 2015).

In line with the OOA model, many human populations experienced a major decline in N_e coinciding with the OOA migration 70–50 thousand years ago (kya) (Bergström et al. 2020). Concurrently, African populations experienced declines in N_e while maintaining consistently larger N_e than non-African populations (Auton et al. 2015; Mallick et al. 2016; Bergström et al. 2020). Thus, the higher genetic diversity and lower LD in African populations reflect historically larger N_e.

Deep Population Structure in Africa

Hunting and gathering was the predominant subsistence strategy prior to the introduction of agriculture and pastoralism during the **Neolithic** (i.e., 12–6.5 kya in Africa) (Marshall and Hildebrand 2002). Today, only a few traditional hunter–gatherer groups remain that live in small communities. Generally, it is assumed that they have either merged into or were replaced by neighboring agropastoral groups, obscuring some of the ancestral genetic variation and structure (Pagani et al. 2012; Schlebusch and Jakobsson 2018; Gopalan et al. 2022). Nevertheless, when accounting for recent admixture, studying the genetics of the traditional hunter–gatherer groups in Africa can provide a snapshot of deep population structure due to their long-term population continuity. Attempts to illuminate the deep population structure in Africa have been further aided by the emergence of ancient DNA from unadmixed hunter–gatherer individuals (e.g., Skoglund et al. 2017 and Lipson et al. 2022).

The remaining traditional hunter–gatherer groups in Africa can be broadly grouped into three major groups: Khoe-San, eastern African hunter–gatherers (EAHG), and rainforest hunter–gatherers (RHG). Khoe-San collectively refers to Khoisan-speaking San hunter–gatherers and Khoekhoe herders, who historically inhabit arid regions in southern Africa. Possibly, Khoe-San were the only inhabitants of southern Africa for much of its prehistory (Schlebusch et al. 2017; Skoglund et al. 2017; Vicente, Jakobsson, et al. 2019). Khoekhoe herders have adopted a pastoralist lifestyle only recently, likely after admixture with eastern African pastoralists over the past 1,500 years (Breton et al. 2014; Macholdt et al. 2014; Schlebusch et al. 2017). Similarly, EAHG groups, for example, the click-speaking Hadza and Sandawe in Tanzania and the Chabu in Ethiopia, are traditional foragers, who have practiced a hunter–gatherer lifestyle until recently or are still practicing it (Bower 1991; Prendergast 2020). These EAHG groups are more closely related to each other than to other African hunter–gatherer groups (Scheinfeldt et al. 2019). RHG groups comprise genetically diverse populations in equatorial Africa, which are often further subdivided into western (e.g., the Baka) and eastern (e.g., the Mbuti) RHG groups (Patin and Quintana-Murci 2018).

Many hunter–gatherer groups experienced declines in N_e during the **Holocene** and have small census population sizes today (Patin et al. 2014; Scheinfeldt et al. 2019; Bergström et al. 2020; Gopalan et al. 2022). Nevertheless, African hunter–gatherers have the highest level of genetic diversity of extant populations and represent the most deeply branching human lineages even after accounting for recent admixture (Henn et al. 2011; Barbieri et al. 2014; Schlebusch et al. 2020). Specifically, the Khoe-San exhibit the highest genetic diversity of all human

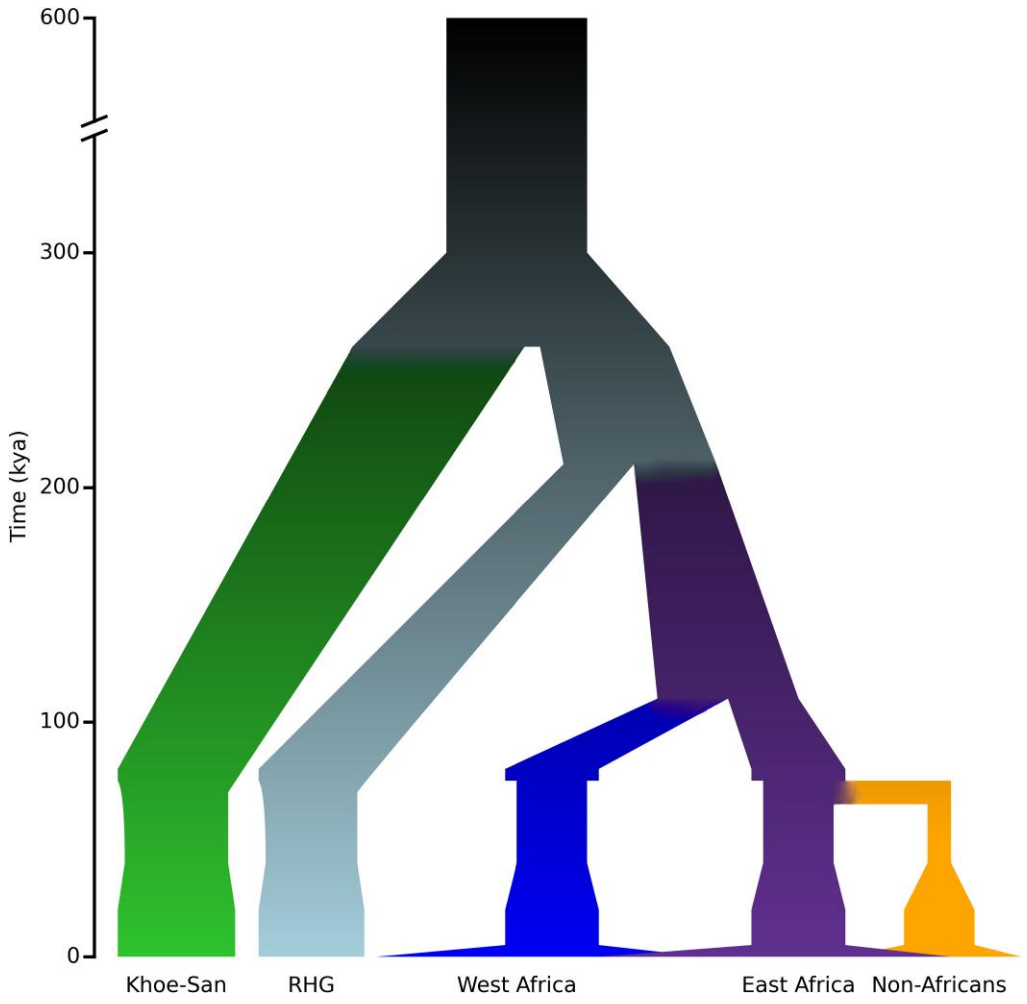

Fig. 1.—Human population tree. Khoe-San show the deepest split times, followed by RHG groups (e.g., Mbuti) and other extant populations. Divergence times are based on estimates from Schlebusch et al. (2020). Note that subsequent gene flow can confound these estimates. Changes in the widths of the arms represent qualitative changes in effective population sizes. The design of this figure was inspired by Schlebusch et al. (2017).

lineages, with a mean heterozygosity of 1.154×10^{-3} compared with 1.09×10^{-3} in the Mandenka (Schlebusch et al. 2020). As their genetic diversity is still significantly higher after accounting for recent admixture with non–Khoe-San groups, it reflects their historically larger N_e (Kim et al. 2014; Fan et al. 2019; Schlebusch et al. 2020). The lineage leading to the Khoe-San is basal to all other human lineages with an estimated divergence time of 300–200 kya (e.g., the Ju|'Hoan with the lowest level of recent admixture diverged ~270 ± 12 kya). Subsequently, the Mbuti (RHG) diverged ~220 ± 10 kya from all other human lineages, forming a second basal lineage (Schlebusch et al. 2020) (fig. 1). These estimates may reflect lower bounds as recent admixture reduces divergence time estimates. For these reasons, assuming regional population continuity, it has been argued for a southern African origin of modern humans (Henn et al. 2011), although models involving eastern Africa and/or multiple geographic regions are also debated (Henn et al. 2018).

The genetic relationship between these different hunter–gatherer groups can largely be modeled by an **isolation-by-distance model** (Skoglund et al. 2017; Vicente, Jakobsson, et al. 2019; Wang et al. 2020a; Lipson et al. 2022). There appears to be a **genetic cline** connecting the eastern African Hadza and southern African Khoe-San, as ancient hunter–gatherer genomes from eastern Africa show affinities to extant southern African San and EAHG (Pickrell et al. 2012; Skoglund et al. 2017; Wang et al. 2020a; Lipson et al. 2022; Fan et al. 2023). For instance, ancient hunter–gatherers genomes from Malawi (~8,100–2,500 BP) and Tanzania (~1,400 BP) exhibited two-third and one-third San-related **ancestry**, respectively, suggesting that the San previously occupied a larger geographic area extending into eastern

Africa (Skoglund et al. 2017) or admixture with a hunter–gatherer group who later gave rise to contemporary San (Wang et al. 2020a). An additional east–southwest cline was recently identified by the incorporation of six novel genomes of ancient hunter–gatherers from eastern and south–central Africa. Some of these individuals are located closer to ancient and present-day central African RHG in **principal component** space (Lipson et al. 2022). This could either suggest deep population structure with EAHG and southern hunter–gatherer groups tracing some of their ancestries to a basal central African RHG lineage (Lipson et al. 2020, 2022) or gene flow between southern African and central African foragers, as indicated by a distinct allele-sharing pattern between the !Xun/Ju|'Hoan and Mbuti (Scheinfeldt et al. 2019; Bergström et al. 2020; Schlebusch et al. 2020). However, in general, ancient genomes reveal deep divergence times of eastern, southern, and central African hunter–gatherer groups, indicating little historical gene flow (Fan et al. 2019; Scheinfeldt et al. 2019; Wang et al. 2020a; Lipson et al. 2022).

Although recent admixture with agriculturists and pastoralists partially obscures ancestral variation and population structure in traditionally foraging groups, their genomes may still provide exciting glimpses into the deep demographic history of modern humans (Bryc et al. 2010; Patin et al. 2014; Lipson et al. 2022). The sequencing of more ancient African genomes will likely reveal new complexities of human origins, although the tropical climate is complicating the analysis of ancient DNA in sub-Saharan Africa. Nevertheless, ancient DNA has recently been obtained of 18,000-year-old individuals (Lipson et al. 2022). Additional details about the deep population structure and the state of ancient DNA research in Africa can be found in reviews by Hollfelder et al. (2021) and Prendergast et al. (2022), respectively.

Evidence of Archaic (Ghost) Introgression in Africa

With the sequencing of genomes of archaic hominins, it has become evident that modern humans interbred with archaic hominins on multiple occasions in Eurasia (Green et al. 2010; Meyer et al. 2012). Although modern human–Neanderthal interbreeding most likely occurred in Eurasia after the OOA migration (possibly in the Levant) (Lazaridis et al. 2016), African populations also exhibit signals of Neanderthal admixture—especially northeastern African but also some West African populations (Gurdasani et al. 2015; Bergström et al. 2020; Chen et al. 2020). This signal of Neanderthal admixture observed in African genomes is most likely not the result of direct admixture but rather the result of admixture with back-migrating Europeans. This is because most Neanderthal haplotypes are shared with Europeans (Bergström et al. 2020; Chen et al. 2020). However, evidence supporting additional admixture events with unknown archaic hominins—the so-called archaic "ghost" populations—within Africa is also mounting (Lorente-Galdos et al. 2019; Wall et al. 2019; Durvasula and Sankararaman 2020; Schaefer et al. 2021).

The first evidence for archaic ghost **introgression** in Africa was obtained by applying S*—an approach that searches for highly divergent haplotypes—to African populations (Plagnol and Wall 2006). The time to the most recent common ancestor (TMRCA) of such identified putatively introgressed haplotypes was found to be significantly older than the deepest split of all modern human lineages and similar to the TMRCA of introgressed Neanderthal haplotypes found in Eurasian populations (Hammer et al. 2011; Lachance et al. 2012; Hsieh, Woerner, et al. 2016). This suggests that the introgressing archaic ghost lineage diverged approximately at the same time from the modern human lineage as Neanderthals (Lachance et al. 2012). Later studies fitting demographic models to the data (Skoglund et al. 2017; Hey et al. 2018; Lorente-Galdos et al. 2019; Lipson et al. 2020) or comparing empirical data to simulated data (Durvasula and Sankararaman 2020; Wang, Mathieson, et al. 2020) found that models which include archaic admixture in Africa consistently describe the data better than models that do not include archaic admixture. Reassuringly, the different approaches also inferred similar demographic scenarios, involving an archaic lineage that diverged around the same time as the Neanderthal lineage (~800–500 kya) and recurring, low-level admixture as recently as 30 kya (Hsieh, Woerner, et al. 2016; Lorente-Galdos et al. 2019; Lipson et al. 2020; Wang, Mathieson, et al. 2020). Lorente-Galdos et al. estimated that Khoe-San derive 3.8% (95% CI: 1.7–4.8%), Mbuti 3.9% (95% CI: 1.3–4.9%), and western African populations 5.8% (95% CI: 0.7–9.7%) of their ancestry from an archaic ghost lineage.

Furthermore, several candidate introgressed genes have been identified. Xu et al. (2017) concluded that a highly divergent haplotype of *MUC7* introgressed into modern West Africans from an archaic lineage. This salivary protein has previously been associated with being protective against asthma. However, Durvasula and Sankararaman (2020) did not find evidence for introgression at the *MUC7* locus when they applied a novel statistical method (ArchIE) that identifies introgressed segments based on multiple population genetics statistics to western African genomes. Using ArchIE, they identified a set of possibly adaptively introgressed genes that are at high frequencies in West Africans (99.9th percentile of putatively introgressed allele frequencies): *NF1*, *MTFR2*, *HSD17B2*, *KCN1P4*, and *TRPS1* (Durvasula and Sankararaman 2020). These examples underline the importance of potential archaic admixture for African genomic medicine (Pereira et al. 2021).

Despite the evidence for archaic admixture, it cannot be ruled out that deep population structure confounds the inference of archaic ghost introgression in Africa (Ragsdale et al. 2022; Fan et al. 2023). For instance, Ragsdale et al. (2022) recently found that a structured model with two stems, that is, two weakly differentiated *Homo* populations connected by gene flow over evolutionary time, can also explain the observed signals of archaic ghost introgression in Africa. However, the possibility of archaic ghost admixture is also supported by fossil records from across Africa, indicating that modern humans spatially and temporally overlapped with hominins exhibiting archaic features (Harvati et al. 2011). Thus, there were ample opportunities for admixture between modern humans and archaic hominins.

Pervasive Admixture in Africa during the Past 10,000 Years

In conjunction with archeological and linguistic studies, genetic studies of contemporary humans and ancient remains have painted a complex pattern of human history in Africa, as many African populations are connected by gene flow. Most contemporary African groups share some of their ancestries with groups from different geographic regions (fig. 2). Nevertheless, different genetic ancestries tend to cluster geographically (fig. 3*A*), with deserts and rainforests acting as major barriers to gene flow (fig. 3*B*). In the following subsections, we discuss major migration events that have shaped population structure in Africa during the past 10,000 years. We start with discussing admixture events in the deeper past and move to admixture events closer to the present day.

Fine-Scale Population Structure and Admixture of Khoe-San Populations

The Khoe-San are basal to all other human lineages with an estimated divergence time of 300–200 kya (Schlebusch et al. 2020; Fan et al. 2023). Although these populations are traditional foragers, some Khoe-San groups have recently adopted (agro-)pastoralist lifestyle. Initial studies leveraging autosomal genotyping data (Pickrell et al. 2012; Schlebusch et al. 2012) and mitochondrial DNA (mtDNA) (Barbieri et al. 2014) suggested differentiation between Khoe-San populations living north and south of the Kalahari Desert, an area that was dominated by lake Makgadikgadi during prehistoric times (i.e., >10 kya) (Goudie 2003). An additional central Khoe-San–related ancestry component has been identified in more recent studies that leveraged bigger and more diverse data sets (Uren et al. 2016; Montinaro et al. 2017; Vicente, Jakobsson, et al. 2019). Notably, these three ancestry components correlate with geography but not linguistics or present-day subsistence strategy. The Kx`a-speaking Ju|'Hoan and ! Xun and the Khoekhoe-speaking Hai||om are representative of the North Khoe-San ancestry component, the Khoekhoe-speaking Nama and Tuu-speaking ǂKhomani and Karretije are representative of the South Khoe-San ancestry component, and all remaining Khoe-San population are representative of the central Khoe-San ancestry component (Montinaro et al. 2017; Vicente, Jakobsson et al. 2019). Interestingly, the pairwise genetic divergences of these three components were found to be similar (i.e., similar fixation index [F_{ST}] values), and the divergence time was estimated to be ~25 kya (95% CI: 18–32 kya) (Montinaro et al. 2017).

Although most of the genetic variation among Khoe-San populations is explained under an isolation-by-distance model (Montinaro et al. 2017; Vicente, Jakobsson, et al. 2019), there is evidence of modest admixture between the three Khoe-San–related ancestry components. In formal admixture tests (f_3-analysis), the ǂKhomani (southern component) showed significant evidence of admixture with Taa populations (central), and the Ju|'Hoan (northern) showed significant signs of admixture with the !Xun (northern) and the Naro (central). Additionally, the Naro (central) showed evidence of admixture with the Ju|'Hoan (northern) and another population characterized by the Central Khoe-San component (e.g., Taa or |Gui). However, none of the populations characterized by the central Khoe-San component showed significant evidence of being a mixture between northern and southern Khoe-San groups (Montinaro et al. 2017). Using SpaceMix analyses, Vicente, Jakobsson et al. (2019) found additional evidence for gene flow from the Ju|'Hoan (northern) into the ǂHoan (central), from the |Gui/Xade San (central) into the Naro (central), and from an undefined Khoe-San population into the Nama (southern). Note that these tests do not definitively establish admixture between specific populations—the actual historical gene flow may have involved other related populations. It has been argued that this gene flow must have occurred within the last 10 ky after the prehistoric lake Makgadikgadi dried up (Barbieri et al. 2014). However, more studies of whole genome sequences are needed for exact dating. For further review of the history of Khoe-San populations, see Pakendorf and Stoneking (2021).

Complex Spread of Pastoralism in Eastern and Southern Africa

Recent genetic studies paint a complex picture of population continuity and admixture in eastern Africa since the introduction of pastoralism in northeastern Africa some 8 kya (e.g., Haber et al. 2016; Skoglund et al. 2017; Prendergast et al. 2019; Naidoo et al. 2020; Wang et al. 2020a). Using DNA from ancient individuals from Kenya

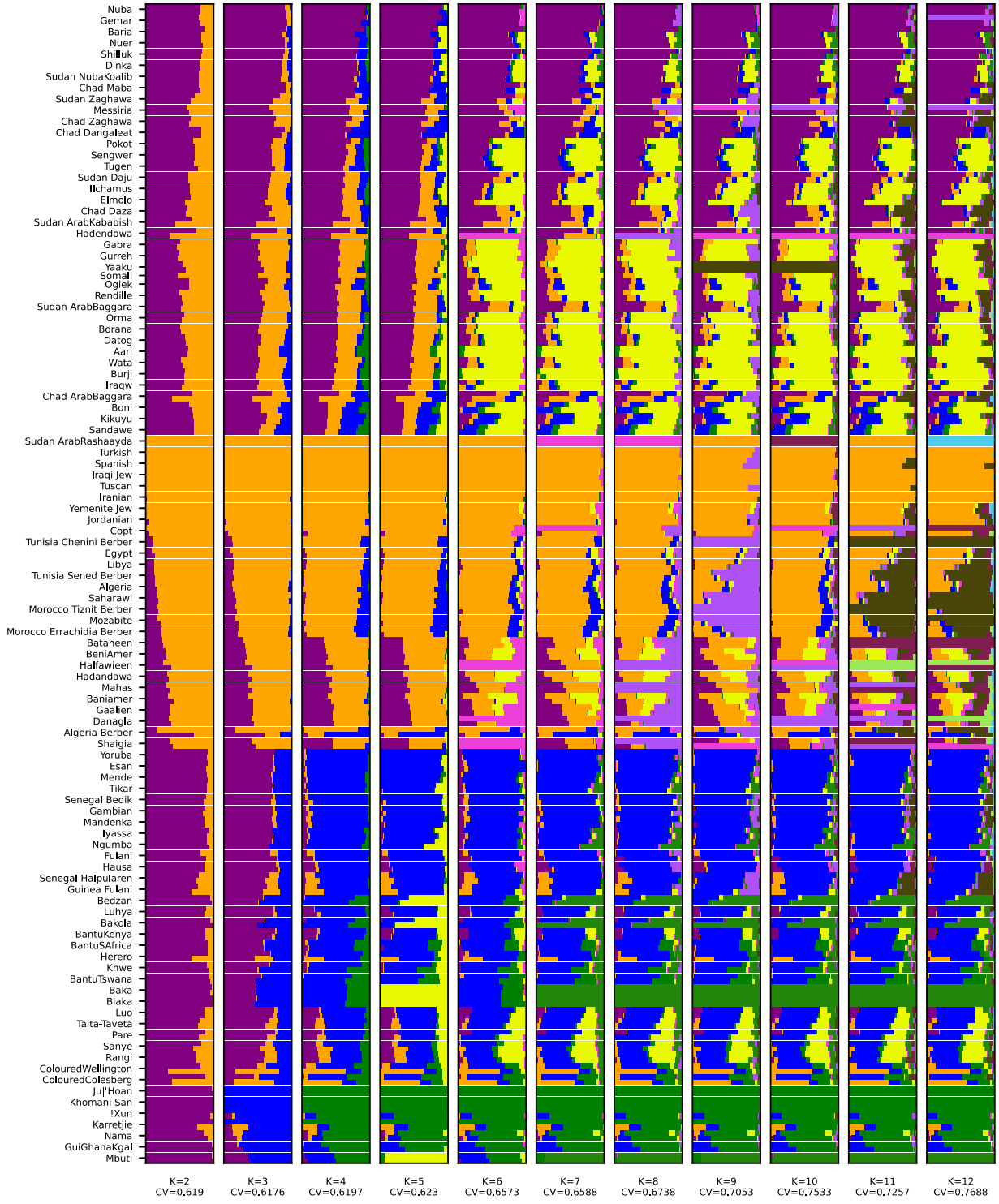

Fig. 2.—Population structure analysis of 97 African and 7 Eurasian populations. ADMIXTURE plots are shown for $K = 2$ to $K = 12$. At $K = 2$, African-like and European-like ancestry cluster separately, and at $K = 3$, a Khoe-San component appears. At $K = 4$, West and East African-like ancestry is distinguished. At higher K, additional fine-scale population structure is revealed. These analysis used harmonized and LD-pruned genotype data from Schlebusch et al. (2012), Mallick et al. (2016), Arauna et al. (2017), Crawford et al. (2017), Hollfelder et al. (2017), Scheinfeldt et al. (2019), and Fortes-Lima et al. (2022) (see supplementary methods and table S1, Supplementary Material online). Note that the results of ADMIXTURE analysis are contingent on which populations are included, as well as their sample sizes. Code used to generate this figure can be found at GitHub: https://github.com/LachanceLab/AfricanPopulationStructure.

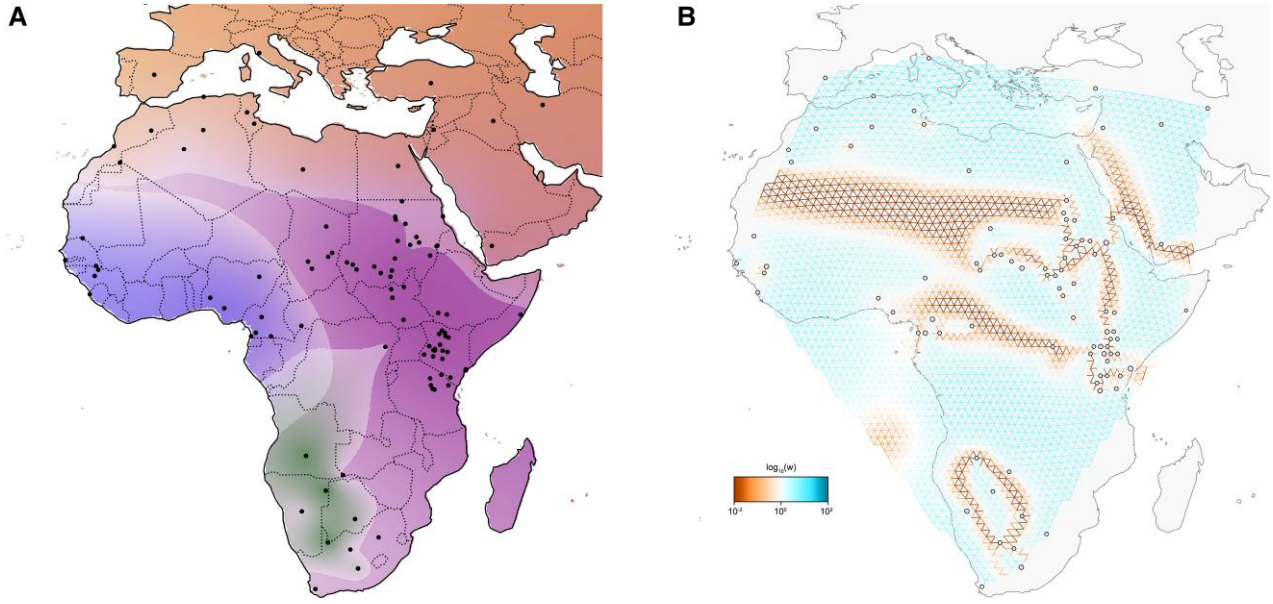

Fig. 3.—Spatial visualizations of admixture and migration in Africa. (*A*) Mapping of ADMIXTURE proportions at $K = 4$, that is, Eurasian-like, West African-like, Khoe-San-like (southern Africa), and East African-like ancestry, on a geographical map using the ordinary Kriging method. (*B*) Effective migration surfaces estimated using FEEMS (Marcus et al. 2021). Brown shading indicates lower effective migration rates, and blue shading indicates higher migration effective migration rates, with edge weights quantified by log10(*w*). As expected, the Sahara, Red Sea, central African rainforest, and the Kalahari Desert act as ecological barriers. These analyses used harmonized and LD-pruned genotype data from Schlebusch et al. (2012), Mallick et al. (2016), Arauna et al. (2017), Crawford et al. (2017), Hollfelder et al. (2017), Scheinfeldt et al. (2019), and Fortes-Lima et al. (2022) (see supplementary methods and table S1, Supplementary Material online). Code used to generate this figure can be found at GitHub: https://github.com/LachanceLab/AfricanPopulationStructure.

and Tanzania, it has been proposed that herding and farming spread in multiple steps into eastern Africa (Prendergast et al. 2019). First, in northeastern Africa, admixture between a population related to contemporary Nilo-Saharan speakers (e.g., the Dinka or Nuer) and a population related to modern groups from northern Africa or the Levant created a group of "early northeastern pastoralists." This group then migrated to eastern Africa and admixed with local foragers ~4 kya, receiving ~20% ancestry from a group related to a 4,500-year-old ancient individual from the Mota cave in Ethiopia that is genetically similar to the isolated, Afro-Asiatic–speaking Aari (Gallego Llorente et al. 2015) and present-day Afro-Asiatic speakers (fig. 4*A*). Given the high genetic affinity of a pastoralist individual who lived 4000 years ago in northern Sudan with ancient individuals from Kenya and Tanzania, it has been argued that this initial dispersal of northeastern pastoralists into East Africa occurred rapidly (Wang et al. 2022). Lastly, this group received another pulse of gene flow from a population related to Nilo-Saharan–speaking Dinka in Sudan ~2.2 kya, that is, during the **Iron Age** (fig. 4*A*; Prendergast et al. 2019; Fan et al. 2019). Based on varying amounts of Mota-related and Dinka-related ancestry in ancient individuals from the Democratic Republic of Congo, Uganda, and Botswana, it has been argued that a model with repeated, unidirectional gene flow from east African forager groups and Nilo-Saharan–speaking groups into the "early northeastern pastoralist" group provides a better fit (Wang et al. 2020a). However, with the currently available data, it is impossible to distinguish between multiple waves of migration and complex population structure.

In line with archeological studies, genetic studies of Khoe-San confirmed that pastoralism spread from East Africa to southern Africa by demic diffusion (Breton et al. 2014; Macholdt et al. 2014; Ranciaro et al. 2014; Schlebusch et al. 2017; Skoglund et al. 2017). Khoekhoe-speaking populations (e.g., the Nama), who currently practice a pastoralist lifestyle, have a high-frequency lactase persistence (LP) allele that is also found in East African populations (Schlebusch et al. 2012; Breton et al. 2014; Macholdt et al. 2014). This "East African" LP single nucleotide polymorphism (SNP) (14010G > C) is distinct from the "European" LP SNP (1391 °C > T) and is rare in southern African Bantu-speaking groups (Breton et al. 2014; Macholdt et al. 2014). Among Khoe-San groups, this "East African" LP SNP is found at the highest frequency in the Nama with a frequency of ~35%, which is much higher than expected given the ~13% East African admixture fraction in the Nama, suggesting positive selection (Breton et al. 2014; Macholdt et al. 2014). Interestingly, Prendergast et al. found that the "East African" LP allele is largely absent from ancient pastoralist individuals from Kenya and Tanzania, indicating that east African pastoralists were lactose intolerant as recently as 3–1 kya

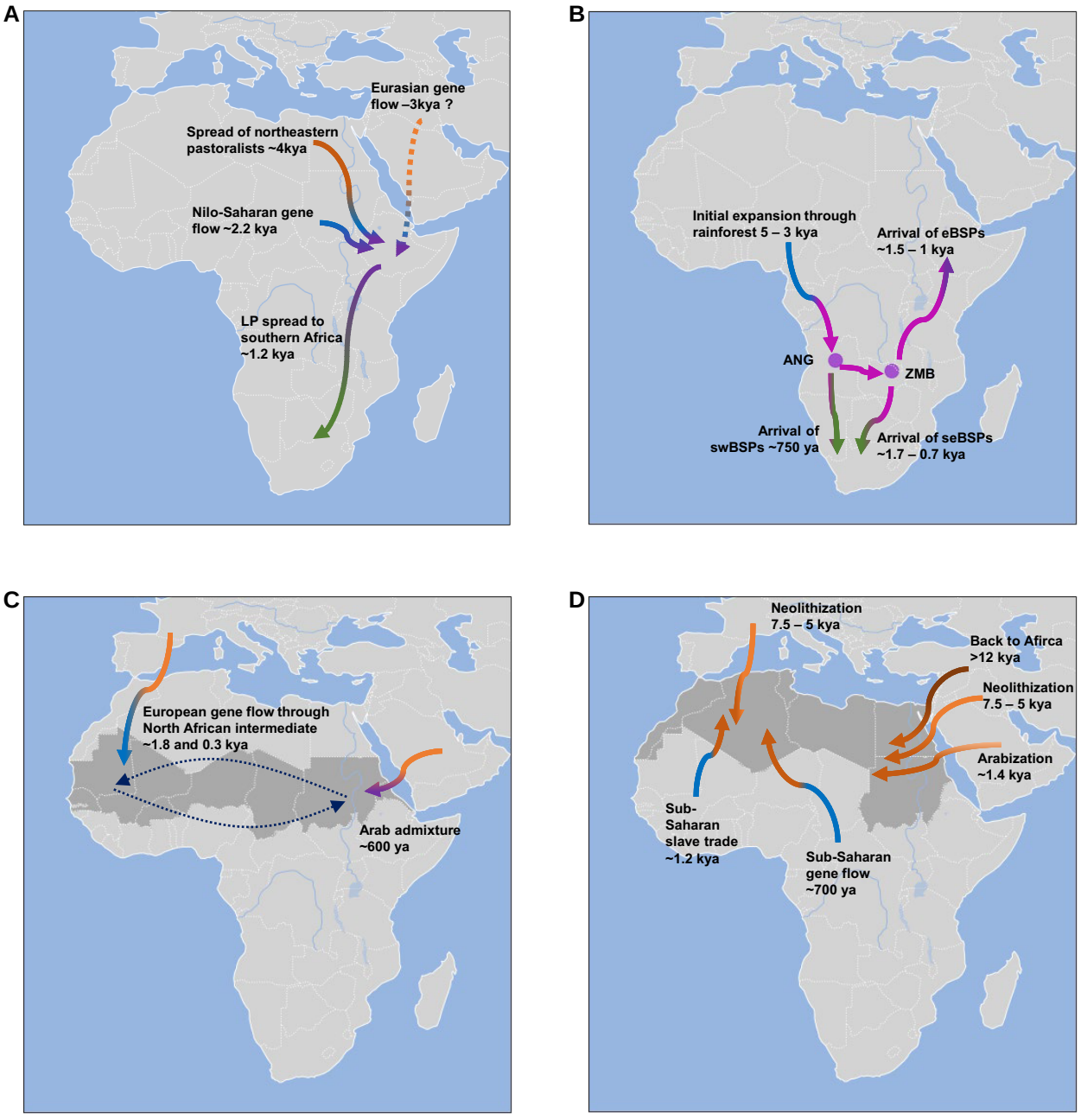

Fig. 4.—Visual summary of key admixture events in Africa. (*A*) The stepwise spread of lactose persistence from northeastern Africa into eastern Africa and subsequently into southern Africa. (*B*) Southward migration of Bantu-speaking people through the rainforest to modern-day Angola (ANG) and Zambia (ZMB) before splitting into eBSPs and seBSPs, in concordance with the late-split hypothesis. (*C*) Extensive admixture between Sahelian populations with European groups in the West and Middle Eastern groups in the East, but only limited gene flow among Sahelian populations. (*D*) Repetitive gene flow from the Middle East/Europe and sub-Saharan Africa into Northern African populations.

(Prendergast et al. 2019). However, it is also possible that this allele has not been detected in ancient samples due to a limited number of surveyed individuals.

A direct link between Afro-Asiatic–speaking eastern African (i.e., Amhara- or Oromo-related ancestry) and southern African pastoralists has been established by showing that a 1,200-year-old individual from southern Africa, who has genetic similarities with modern Khoekhoe-speaking pastoralist groups (e.g., the Nama), traces ~40% of their ancestry to a Eurasian admixed group related to a 3,100-year-old pastoralist individual from Luxmanda, Tanzania (Skoglund et al. 2017). Thus, this study indicates that admixture of Khoe-San groups with eastern African pastoralists occurred at least ~1.2 kya (fig. 4*A*). Concordantly, another study estimated

that all modern Khoe-San populations received 9–30% gene flow from an admixed East African/Eurasian pastoralist group 1.5–1.3 kya (Schlebusch et al. 2017). Furthermore, east African pastoralist contributions to Khoe-San groups are lower on X chromosomes than autosomes (Vicente et al. 2021), indicating that male-biased admixture occurred. Overall, these results suggest that eastern African pastoralists reached southern Africa prior to and independently of Bantu-speaking groups. For a detailed review of the spread of lactase persistence in Africa, see Campbell and Ranciaro (2021).

Multiple Migration Waves of Bantu Speakers

Genetic studies showed that the spread of Bantu languages, agricultural practices, and iron use 5–3 kya was accompanied by multiple migration waves of Bantu speakers from western Africa (i.e., current eastern Nigeria and western Cameroon) to other regions in sub-Saharan Africa (Tishkoff et al. 2009; Schlebusch et al. 2012; Li et al. 2014). Consequently, the Bantu expansion extensively contributed to population structure due to differential levels of admixture with and replacement of local hunter–gatherer groups over the past 3,500 years (Skoglund et al. 2017; Sengupta et al. 2021; González-Santos et al. 2022).

Two major migratory routes of Bantu-speaking populations (BSPs) have been hypothesized. The early-split hypothesis suggests that BSPs split at an early stage north of the rainforest, with one group then moving directly South through the rainforest, whereas the other migrated East, north of the rainforest, toward the Great African Lakes. In contrast, the late-split hypothesis states that BSPs first migrated South through the rainforest before splitting into two groups, with one moving further South and the other one migrating East toward the Great African Lakes. Similarly to phylolinguistics (e.g., Rexová et al. 2006), genetics are in favor of the late-split hypothesis, as eastern BSPs (eBSPs) and south-eastern BSPs (seBSPs) are genetically closer to western BSPs (wBSPs) south of the rainforest (i.e., Angola) than to wBSPs north of the rainforest (Patin et al. 2017). A subsequent study using samples from wider geographic and ethnolinguistic groups showed that eBSPs, seBSPs, and southwestern BSPs (swBSPs) are genetically closest to Bantu speakers from Zambia (Choudhury et al. 2020). Together, these findings suggest that Bantu speakers first migrated South through the rainforest to Angola and subsequently to Zambia before splitting into two groups (fig. 4*B*).

In western Africa, wBSPs asymmetrically mixed with resident RHG groups, with RHG groups receiving higher amounts of gene flow from wBSPs (Jarvis et al. 2012; Hsieh, Veeramah, et al. 2016; Patin et al. 2017; Lopez et al. 2018). wBSPs in Angola have small amounts of RHG-related ancestry from an admixture event that occurred after the split of BSPs ~800 ya (Patin et al. 2017), although a recent study inferred a more ancient admixture date of ~1.9 kya for Bantu speakers in Cabinda, Angola (Tallman et al. 2022). The amount of gene flow from wBSPs into individual RHG groups varied. Whereas the Mbuti and the Biaka have <6% wBSP-related ancestry, the Bezan and the Bongo trace as much as 38.5% and 47.5% to wBSPs, respectively (Patin et al. 2014). This gene flow from wBSPs into RHGs was inferred to have occurred ~7 kya using models of site-frequency spectra (Hsieh, Veeramah, et al. 2016; Lopez et al. 2018), whereas methods leveraging LD patterns yielded estimates <1 kya (Patin et al. 2014, 2017). Analyses of **uniparental markers** as well as autosomal and X chromosomal data also showed that this gene flow from wBSPs into RHGs was male-biased (Verdu et al. 2013; Patin et al. 2014). For an excellent review of the interactions between BSPs and RHGs, see Patin and Quintana-Murci (2018).

In eastern Africa, two admixture events 1.5–1 kya and 400–150 years ago have been inferred between wBSPs (~75% contribution) and an Afro-Asiatic–speaking population from Ethiopia (~10%) (Patin et al. 2017). These estimates are in slight disagreement with the estimates of Skoglund et al. (2017), who estimated that admixture between Bantu speakers and eastern African pastoralists occurred 800–400 years ago, but are in agreement with 71% Bantu-related ancestry in an ancient Iron Age individual dated to ~1,160 years ago from the Rift Valley in Kenya (Prendergast et al. 2019).

In South Africa, seBSPs received between 1.5% (e.g., the Tsonga) and 20% (e.g., the Tswana) gene flow from Khoe-San groups during independent admixture events (Sengupta et al. 2021). The exact admixture timings differ between populations (1.7 kya–700 ya), with northern groups showing older dates than southern groups (Sengupta et al. 2021; Tallman et al. 2022). However, Tallman et al. (2022) also reported an older Khoe-San admixture event in the Zulu ~3 kya. Furthermore, uniparental markers and X chromosomal and autosomal data suggest male-biased seBSPs contributions and female-biased Khoe-San contributions (Bajić et al. 2018; Sengupta et al. 2021). In contrast to the admixture in South Africa, seBSPs appeared to have replaced resident hunter–gatherer populations in Malawi and Mozambique with present-day individuals deriving ≥97% of their ancestry from the Bantu expansion (Skoglund et al. 2017; Tallman et al. 2022). Overall, this suggests multiple migration waves of Bantu speakers or that Khoe-San admixture did not occur immediately.

In contrast to seBSPs, swBSPs appear to have reached southern Africa more recently (~750 ya), as indicated by more recent admixture of a western African-related source in the Khoisan-speaking Khwe and !Xun from Angola (Busby et al. 2016). This suggests that swBSPs took a different

route directly south along the western coast and thus have different recent population histories than seBSPs (fig. 4*B*).

Genetic studies of uniparental and autosomal markers initially suggested that BSPs are largely genetically homogenous groups of people (i.e., $F_{ST} \leq 0.02$) (Coelho et al. 2009; Ansari-Pour et al. 2013; Gurdasani et al. 2015; Busby et al. 2016). Despite the modest F_{ST} values, fine-scale population structure of BSPs has recently been emphasized (Semo et al. 2020; Sengupta et al. 2021; Tallman et al. 2022). Semo et al. (2020) showed that eBSPs from Mozambique and Angola form a North–South genetic cline of relatedness along the coast from Kenya/Tanzania to South Africa. They also found that genetic homogeneity increases east- and southward, indicating serial founder effects and little admixture with local populations until Bantu speakers reached South Africa. Additionally, Sengupta et al. (2021) found that fine-scale genetic substructure among seBSPs in South Africa correlates well with geography and linguistics and persists even after accounting for differential levels of Khoe-San admixture.

Overall, recent genetic studies highlight the spatially and temporally complex dynamics of the Bantu expansion, with differential levels of admixture among sub-Saharan populations and multiple migration waves. Increasing sample sizes, as well as the number of sampled BSPs and additional ancient genomes, may allow for clarifying the exact migratory route, dating major events, and revealing further fine-scale population structure among BSPs. For a comprehensive review of the population history of Bantu speakers, see Schlebusch and Jakobsson (2018) as well as Choudhury et al. (2021).

Admixture of Pastoralists and Farmers in the Sahelian Populations

The Sahel/Savannah belt was formed with the aridification of the Sahara Desert ~5.5 kya (Manning and Timpson 2014), pushing human populations, among others, southward closer to the tropical rainforest, which demarcates the southern border of the belt. Nowadays, this region is inhabited by populations practicing one of two main subsistence strategies, tracing their origin to the Early Holocene (~10 kya) (Pereira et al. 2010). Nomadic pastoralists (i.e., the Fulani in the West and the Arabs in the East) maintain large numbers of cattle that require seasonal movements to pastures and water resources, whereas farming populations (e.g., the Hausa or Mandinka) are more sedentary.

Genetic analyses generally revealed weak population structure, with most of the variation found within groups rather than between groups (Čížková et al. 2017; Nováčková et al. 2020; Diallo et al. 2022; Fortes-Lima et al. 2022). Depending on subsistence strategy, different distributions of uniparental markers have been observed in the Sahel. Whereas sedentary farmers are stratified based on geography but not linguistics, the opposite is true for western Fulani pastoralists (Nováčková et al. 2020). Furthermore, Y chromosomal haplogroups are genetically more diverse in nomadic pastoralists groups, whereas mtDNA haplogroups are more diverse in sedentary farmers (Čížková et al. 2017; Nováčková et al. 2020; Diallo et al. 2022). It has been suggested that low levels of sex-biased gene flow with sedentary farmers caused the Fulani to lose mtDNA diversity (Čížková et al. 2017). However, this may also be the result of a strong **population bottleneck** (Fortes-Lima et al. 2022). In contrast to the Fulani, Arab pastoralists have a higher mtDNA diversity, suggesting variable levels of female admixture into pastoral populations (Čížková et al. 2017).

A recent study of genome-wide genotype data from 327 individuals comprising 14 ethnolinguistic groups highlighted fine-scale population structure and admixture in the Sahel region that is mostly correlated with the geographical distribution of populations. Arabic-speaking populations from Central and Eastern Sahel form an east-to-west genetic cline due to varying amounts of Middle Eastern–related and East African–related ancestry (Fortes-Lima et al. 2022). The mixing of Middle Eastern–related and African–related ancestry components has been dated to ~600 ya in Sudanese Arab populations (fig. 4*C*; Shriner and Rotimi 2018a). Middle Eastern–related ancestry was found to range from ~27.6% in the Baggara from Chad and Sudan to 95.1% in the Rashaayda from Sudan (Hollfelder et al. 2017; Fortes-Lima et al. 2022). The high proportion of Middle Eastern–related ancestry in the Rashaayda is consistent with high frequencies of Middle Eastern mtDNA haplogroups, that is, R0a2c and J1b (Čížková et al. 2017; Priehodová et al. 2017).

In contrast to eastern Arabic-speaking populations, western Fulani groups are the closest to western Africans but also show significant fractions of European-related and East African–related ancestry (Henn et al. 2012; Triska et al. 2015; Busby et al. 2016; Vicente, Priehodová, et al. 2019; Fortes-Lima et al. 2022). Two admixture events involving a West African group and two different European groups dating to ~1.8 kya and ~300 years ago have been identified. During the first admixture event ~1.8 kya, the European component is best resembled by present-day northwestern Europeans, whereas during the second pulse ~300 years ago, the European component is more closely related to contemporary southwestern Europeans (Vicente, Priehodová, et al. 2019). This European-related ancestry was most likely indirectly introduced into the Fulani via admixture with a northern African population (e.g., a Mozabite-like population; fig. 4*C*) (Triska et al. 2015; Vicente, Priehodová, et al. 2019). Through this admixture event, the Fulani likely received a "European" LP variant −13910*T (rs4988235) in the *LCT* gene region that was then positively selected, reaching frequencies between 18% and 60% in Fulani groups (Lokki et al. 2011;

Ranciaro et al. 2014; Vicente, Priehodová, et al. 2019; Priehodová et al. 2020).

Lastly, small amounts of admixture among Sahelian groups have been inferred from genome-wide markers (Fortes-Lima et al. 2022) as well as mtDNA and Y haplogroups (Čížková et al. 2017; Nováčková et al. 2020) (fig. 4*C*). Limited sex-biased gene flow between the Fulani (and/or other sub-Saharan populations) and Arab nomadic pastoralists has been suggested, as more mtDNA than Y chromosomal haplogroup sharing was observed between the two groups, with most shared haplogroups being of sub-Saharan origin (Čížková et al. 2017; Shriner and Rotimi 2018a; Nováčková et al. 2020). In this section, we focused on the population history of two nomadic populations in the Sahel as they experienced the most admixture. For a comprehensive review of Sahelian populations' demographic history, including Niger-Congo–speaking populations, we refer to Černý et al. (2021).

Continuous, Multi-faceted Admixture in North Africa

Many studies of African genetics have historically focused on sub-Saharan populations, as northern African populations grouped separately from sub-Saharan populations and closer to non-African populations in studies of classical genetic markers (Cavalli-Sforza and Piazza 1993). However, studies of uniparental markers revealed 1) genetic heterogeneity among North African populations with a west-to-east cline of mtDNA and Y chromosomal haplogroup frequencies, 2) a lack of differentiation between Arabs and Imazighen (Berbers), 3) preliminary evidence for extensive admixture of populations with European-related, Middle Eastern–related, and sub-Saharan African–related ancestry, and 4) an autochthonous North African component (Haak et al. 2010; Fadhlaoui-Zid et al. 2011; Pennarun et al. 2012; Hervella et al. 2016; Solé-Morata et al. 2017; Vai et al. 2019).

Studies of genome-wide data largely confirmed the North African population structure inferred from uniparental markers while emphasizing fine-scale population structure (Henn et al. 2012; Arauna et al. 2017; Serra-Vidal et al. 2019; Wohlers et al. 2020; Lucas-Sánchez, Font-Porterias, et al. 2021). Henn et al. (2012) initially reported a clear genetic differentiation between Arabs and Imazighen. This study found that Tunisian Imazighen trace all their ancestry to an autochthonous North African—the so-called Maghrebi—ancestral component, whereas all Arab populations also have European-related, Middle Eastern–related, and/or sub-Saharan–related ancestry (Henn et al. 2012). However, the Tunisian Imazighen were the only Imazighen population in this study and were subsequently found to be an outlier in terms of ancestry composition, low genetic diversity, and high amount of runs of homozygosity (Arauna et al. 2017; Serra-Vidal et al. 2019; Lucas-Sánchez, Font-Porterias, et al. 2021). Subsequent studies of genome-wide data that included more Imazighen populations confirmed that most Arab and Imazighen populations are weakly genetically differentiated (Arauna et al. 2017; Serra-Vidal et al. 2019; Anagnostou et al. 2020).

The Maghrebi component is represented by 15,000-year-old **Paleolithic** individuals from Taforalt, Morocco, whose ancestry is best modeled as a mix of an early Holocene Middle Eastern (63.5%), that is, Levantine Natufians, and a sub-Saharan component (Van De Loosdrecht et al. 2018). Consistent with the age of the Taforalt individuals, it was estimated that the Maghrebi component diverged from the Middle Eastern ancestral component 38–18 kya, indicating back-to-Africa gene flow prior to the Holocene (>12 kya; fig. 4*D*) (Henn et al. 2012). This estimate is broadly consistent with the estimated coalescence times of North African-specific mtDNA lineages (44 ± 21.6 kya for the U6 lineage, 13.0 ± 5.7 kya for the U6a1 lineage, and 13.5 ± 3.7 kya for the U6a* lineage) and Y chromosome haplogroups (~15–12 kya for E-M78 in most populations and 44–30 kya in Tunisian Imazighen) (Fadhlaoui-Zid et al. 2011, 2013; D'Atanasio et al. 2018; Van De Loosdrecht et al. 2018). Furthermore, consistent with the west-to-east clines observed in uniparental markers, the autochthonous Maghrebi component decreases eastward (Henn et al. 2012; Arauna et al. 2017; Serra-Vidal et al. 2019). Approximately 5,000-year-old Early Neolithic individuals from Ifri n'Amr or Moussa, Morocco, show high genetic affinity to the Taforalt individuals, suggesting population continuity between the Paleolithic and Early Neolithic (Fregel et al. 2018).

Neolithization, Arabization, and sub-Saharan gene flow led to the dilution of this Maghrebi component in North African populations (fig. 4*D*). During the Neolithization, North African populations admixed with European Neolithic groups. This admixture is evident from ~3,000-year-old Late Neolithic individuals from Kelif el Boroud, Morocco, who are best modeled as a mixture of Ifri n'Amr or Moussa and European Neolithic groups (Fregel et al. 2018; Serra-Vidal et al. 2019). Subsequently, the Arabization introduced recent Middle Eastern–related ancestry ~1.4 kya, which decreases westward on a genetic cline (Henn et al. 2012; Arauna et al. 2017; Serra-Vidal et al. 2019). It has been hypothesized that the Arab expansion might also have introduced some sub-Saharan–related ancestry through the slave trade (Newman 1995), which is supported by sub-Saharan ancestry in North African populations that could be traced to an admixture event ~1.2 kya with a West African population (Henn et al. 2012). However, most sub-Saharan gene flow was inferred to have occurred more recently during the last 700 years, leading to varying degrees of sub-Saharan ancestry in contemporary North African populations (1–55%) (Henn et al.

2012; Arauna et al. 2017). Furthermore, consistent with patterns observed in the Americas (Micheletti et al. 2020), sub-Saharan gene flow was also likely sex-biased with female-biased sub-Saharan and male-biased Middle Eastern contributions (Arauna et al. 2017; D'Atanasio et al. 2018). Altogether, this suggests that North Africa has a deep history of continuous human migration and admixture. North African population history was also recently reviewed by Lucas-Sánchez, Serradell et al. (2021).

Complex Patterns of Admixture in the Western Cape

The Western Cape, at the southernmost part of South Africa, harbors one of the most diverse admixed populations, namely, the South African Coloured (SAC) population, which is the largest ethnic group in this region and has its origins slightly >350 ya (de Wit et al. 2010; Petersen et al. 2013; Choudhury et al. 2017). The SAC population represents >49% of the estimated 7 million inhabitants in this province, with the vast majority being historically Afrikaans speakers (a unique South African language ancestrally linked to Dutch), although this is more recently changing (Patterson et al. 2010; Republic of South Africa 2021). Their complex origin of admixture is attributed to significant historical events that occurred within the last few millennia, starting ~1.7 kya with the arrival of Bantu-speaking agro-pastoralists in South Africa (Sengupta et al. 2021). During the last few centuries, European colonization of the Cape by the Dutch, Germans, and French, later followed by British seizure and rule, contributed to the complex admixture patterns at the Western Cape. During this time, slaves were trade by the Dutch East India Company from East Africa, Madagascar and surrounding islands, India, and Indonesia, leading to settler–slave admixture, including indigenous Khoe-San people (de Wit et al. 2010; Patterson et al. 2010; Montinaro and Capelli 2018). Genomic studies of the SAC population revealed that these historic events correlate with the complex five-way admixture observed in this population, with ancestral contributions occurring predominantly from the indigenous Khoe-San, the Bantu-speaking Africans, European-descent groups, and Southeast Asian and South Asian populations (de Wit et al. 2010; Daya et al. 2013; Petersen et al. 2013; Chimusa et al. 2014; Choudhury et al. 2017; Swart et al. 2020). In addition, cultural and religious practices contributed to the high degree of heterogeneity in ancestral contributions among SAC individuals sampled from different regions of South Africa (de Wit et al. 2010; Daya et al. 2014; Choudhury et al. 2017).

Several studies have revealed a sex-biased gene flow in SAC that supports the historical records indicating that almost all mixed marriages were between a male settler and either a free Black female (where the man bought the slave their freedom) or an indigenous Khoe-San female (Patterson et al. 2010; Petersen et al. 2013; Choudhury et al. 2017). Petersen et al. (2013) used uniparental markers to ascertain likely ancestral contributions using unique population-specific mtDNA and Y chromosomal haplogroup identifiers. Khoe-San derived maternal lineage L0d had a 68% representation in the SAC group studied, while the M/N Eurasian mtDNA lineages were only represented at low frequencies. In contrast, there was a significant Eurasian paternal contribution (71.4%) defined by haplogroups R/I/G/N/O/J in the same group, and the Western European R1b haplogroup was prevalent at 44.4%. Similar distributions of mtDNA and Y haplogroups were observed from whole genome sequencing data of a small group of SAC males from the Western Cape region (Choudhury et al. 2017). Overall, these findings demonstrate that recent admixture involved sex-biased gene flow.

Admixture in the African Diaspora following the Transatlantic Slave Trade

As a consequence of the transatlantic slave trade, >12.5 million people were forcefully displaced from Africa to the Americas between the sixteenth and nineteenth centuries, creating the largest present-day African diaspora (Eltis 2007). Subsequent admixture with European-like ancestry and Native American-like ancestry populations was spatially and temporally complex, leading to varying amounts of recent African-like ancestry in admixed populations in the Americas (Bryc et al. 2010; Ongaro et al. 2019; Gouveia et al. 2020; Micheletti et al. 2020). African-related ancestry is the highest in the British Caribbean (~75%) and the United States (~71%) and the lowest in South America (~11–12%) and Central America (~8%, including Mexico) (Micheletti et al. 2020). The remaining ancestry can be predominantly assigned as European-like, with minor contributions from Native American groups in some populations (Micheletti et al. 2020). Additionally, despite more males being deported to the Americas, it has been shown that African contributions to gene pools in the Americas were likely female-biased, whereas European contributions were likely male-biased (Mathias et al. 2016; Ongaro et al. 2019; Micheletti et al. 2020). However, the magnitude of the sex bias is difficult to pinpoint from X chromosomal and autosomal ancestry proportions due to potential confounding from complex demographic histories, among others (Pfennig and Lachance 2023).

Broadly in agreement with historical records of the transatlantic slave trade, genetic studies of admixed populations from the Americas showed that most of the African ancestry can be traced to West–Central Africa, for example, similar to the Yoruba or Esan from Nigeria, with a smaller fraction being similar to south-eastern African ancestry, for example, Mbukushu-like from Botswana and/or

Luhya-like from Kenya (Patin et al. 2014; Gouveia et al. 2020; Micheletti et al. 2020). However, the distribution of these African ancestries varies between different populations in the Americas, with western/central African-related ancestry being more common in the northern parts, for example, the United States, and southeastern African-related ancestry being more common in the southern parts, for example, Brazil (Gouveia et al. 2020). The different sources of African-like ancestry and the different timing of admixture for different African source populations in the Americas may be attributed to geography and changing geopolitics at the time, influencing the voyage routes (Ongaro et al. 2019; Gouveia et al. 2020). Interestingly, there is less differentiation between the African ancestries found in admixed genomes in the Americas (as quantified by F_{ST} statistics) compared with what is seen between each of the contributing ancestries in Africa (Gouveia et al. 2020). For a more granular review of the demographic histories in light of the transatlantic slave trade of admixed population in the Americas, see Fortes-Lima and Verdu (2021).

Evidence of Local Adaptation in African Genomes

Environmental conditions vary over time and space. Because of this, African populations have experienced a heterogeneous mix of selection pressures. Nonetheless, African populations are connected via gene flow, which can serve as a potent source of adaptive variation. Although the specific genes implicated in African scans of selection vary by the method used and population studied, some common themes arise. Regulatory DNA appears to be a frequent target of adaptation in African genomes (Quiver and Lachance 2022). Furthermore, many noteworthy instances of selection in Africa are associated with physiology, diet, or pathogen pressure.

One key evolutionary challenge involves physiological responses to extreme conditions, including high-altitude desert environments. In Africa, the Ethiopian Highlands are 1,500 meters above sea level, with summits as high as 4,550 meters above sea level. For example, the Amahara people have adapted to low barometric pressure and hypoxia in the Ethiopian Highlands over the past 5,000 years. Interestingly, the specific adaptive mutations seen in the Ethiopian Highlands differ from what has been observed in the Tibetan Plateau and the Andean Altiplano. Selection scans comparing Amhara individuals living at high altitude to individuals living in lowland areas have implicated a number of adaptive loci, including rs10803083, a SNP that is associated with hemoglobin levels (Alkorta-Aranburu et al. 2012); *BHLHE41*, a gene that is involved in hypoxia response and circadian rhythm (Huerta-Sánchez et al. 2013); and *EGNL1*, a gene that plays a central role in mammalian oxygen homeostasis (Scheinfeldt et al. 2012). Intriguingly, *EGLN1* has also been implicated in selection scans of the click-speaking Sandawe people, who are traditional foragers from Tanzania (Lachance et al. 2012). This suggests that the benefits of adaptive *EGLN1* haplotypes may extend beyond high-altitude conditions.

Arid desert environments also present an evolutionary challenge in Africa. For instance, despite frequent droughts, the ǂKhomani San have lived in the Kalahari Desert for thousands of years. Using SWIF(r), an approach that combines multiple statistics to generate posterior probabilities of sweeps, researchers have identified multiple genes associated with adiponectin, body mass index (BMI), and metabolism as potential targets of selection in the ǂKhomani San (Sugden et al. 2018).

Another example of adaptation to extreme conditions are RHG groups, who evolved a short stature (mean adult height <160 cm). Several candidate loci under selection have been identified that are likely implicated in the short stature of RHG groups as they overlap with genes associated with bone synthesis (e.g., *EHB1* and *PRDM5*), muscular development (e.g., *OBSCN* and *COX10*), and growth hormone synthesis and secretion in the pituitary gland (e.g., *HESX1* and *ASB14*) (Jarvis et al. 2012; Lachance et al. 2012; Hsieh, Veeramah, et al. 2016; Fan et al. 2023), suggesting that the short stature of RHG evolved through positive selection on several loci. An intriguing example is *EPHB1*, an ephrin receptor at sites of osteogenesis. RHG groups and farmer groups are nearly fixed for different haplotypes, suggesting an incomplete selective sweep. The selective pressure at this locus appears to be of regulatory nature as no nonsynonymous variant was found (Hsieh, Veeramah, et al. 2016). Additionally, putatively selected regions also included genes unrelated to height, such as genes associated with reproduction, thyroid function, and immune traits, among others (Jarvis et al. 2012; Lachance et al. 2012; Hsieh, Veeramah, et al. 2016; Fan et al. 2023).

Diets vary across Africa, and these differences have provided ample opportunities for natural selection, be it via changes in metabolism, detoxifying harmful xenobiotics, or shifts in olfactory and taste perception. A textbook example of dietary adaptation and convergent evolution involves lactase persistence, and studies of African pastoralists have identified adaptive regulatory variants near the *LCT* and *MCM6* genes (Segurel and Bon 2017). However, the specific mutations conferring LP in Kenya (G-14010, rs145946881) and Sudan (G-13907, rs41525747) differ from LP mutations found in Northern Europe (T-13910, rs4988235) and the Middle East (G-13915, rs41380347) (Ranciaro et al. 2014). The timing of selection for LP appears to differ in Africa as well: Strong selection for LP in Maasai herders appears to have occurred more recently than selection for LP in Europe (Schlebusch et al. 2013). The ability to break down starchy

foods also appears to have been a target of selection. Substantial copy number variation of salivary amylase genes exists in African and non-African populations, with most humans having between 2 and 15 copies. Interestingly, the Hadza of Tanzania who have a diet rich in tubers tend to have higher copy numbers of amylase genes than populations with low-starch diets (Perry et al. 2007). Dietary differences have also probably contributed to the accelerated evolution of olfactory receptor and taste-perception genes in African populations (Sjöstrand et al. 2021).

Some of the strongest selection pressures on African populations involve pathogens and immune response, and few diseases have impacted human genomes as much as malaria. Each year, this tropical disease contributes to over 500,000 deaths in Africa, many of which involve children younger than 5 years of age (World Health Organization 2021). In sub-Saharan Africa, strong selection for malaria resistance has contributed to the near fixation of the Duffy blood group, elevated rates of *G6PD* deficiency, and sickle cell disease (Kariuki and Williams 2020). This selection has largely occurred during the Holocene, making it a relatively recent phenomenon from an evolutionary perspective. Indeed, the major sickle cell haplotype in central Africa appears to predate the Bantu expansion, with ancestral recombination graphs dating this mutation (rs334) to ~7300 years ago (Shriner and Rotimi 2018b). Furthermore, there is also genetic evidence that admixture has facilitated adaptation to Malaria in Cabo Verde during the past 20 generations, with selection coefficients acting on the *DARC* locus as high as $s = 0.08$ (Hamid et al. 2021). Additional infectious diseases that have been major targets of selection in Africa include HIV-1, trypanosomiasis (i.e., African sleeping sickness), smallpox, and tuberculosis (Karlsson et al. 2014; Rees et al. 2020; Pereira et al. 2021). However, many genes that are associated with immune response are highly pleiotropic, for example, major histocompatibility complex (MHC), human leukocyte antigen (HLA) genes, and apolipoprotein L1 (*APOL1*), complicating attempts to pin down the primary cause of recent adaptations. For instance, two *APOL1* haplotypes (G1 and G2) are protective against trypanosomiasis infection but are also associated with increased risk of kidney disease in African ancestry populations (Pereira et al. 2021). Finally, we note that natural selection on immune-related genes has also extended across the African diaspora. Specifically, Latin American genomes are enriched for African MHC/HLA haplotypes (Zhou et al. 2016; Norris et al. 2020), a pattern that is consistent with the evolutionary benefits of gene flow.

Biomedical Implications of Population Structure in Africa

A central premise of precision medicine is that ancestral variation plays a key role in disease processes. Because of this, the biomedical field benefits from an in-depth understanding of genomic variation in diverse populations (Rotimi and Jorde 2010). Investigating genetic variation in African populations is particularly promising due to their high genetic diversity and low levels of LD, increasing the pool of relevant causal variants (Auton et al. 2015) and allowing to narrow down causal variants (Jallow et al. 2009), respectively. For these reasons, studying more granular population structure in Africa, including potentially adapted genes, may increase our understanding of the genetics of complex traits (Chaichoompu et al. 2020). Such studies have produced large amounts of insightful data which have revealed medically relevant genetic loci and aided the interpretation of the pathogenicity of genetic variants, advancing precision medicine for all populations (Choudhury et al. 2020; Matjuda et al. 2021). This understanding together with knowledge of its interactions with sociocultural factors that influence disease risk or treatment response can improve clinical care by improving the accuracy of genetic testing and/or assessment of therapeutic response (Hindorff et al. 2018).

A greater understanding of the genetic architecture can help explain differences in disease risk between populations. One key example of this involves tuberculosis, a disease that has particularly severe infections in the SAC population (Chimusa et al. 2014; Swart et al. 2021). Leveraging local ancestry and population-specific high-density genotype data, a novel SNP (rs28647531) on chromosome 4q22 was associated with tuberculosis susceptibility in the SAC population. This displayed an SNP minor allelic effect while correcting for local ancestry for Bantu-speaking African ancestry (Swart et al. 2021). This example illustrates that admixed African populations are a promising opportunity to better understand ancestry-specific disease risk compared with homogeneous populations (Patterson et al. 2010). However, the choice of reference populations for multiway admixed populations may be sensitive and critical in biomedical research (Chimusa et al. 2013). For example, population-specific recombination maps have the potential to advance the detection of genotype–phenotype correlations in admixed populations and further the field of precision medicine relevant to all populations (Choudhury et al. 2017; Swart et al. 2020).

Additionally, multiple studies have also shown the significance of including ancestry to effectively direct the outcomes of treatment. These studies have shown that a patient's demographic medical and genetic information can be used for clinical decision-making or genetic counseling (Batai et al. 2021). For example, a patient's comedications, age, genetic variation, and ancestry are commonly used in inferring the dosage of the anticoagulant warfarin. It has been shown that genetic and ancestry-related information plays a significant role in accurately determining appropriate

dosage (Bress et al. 2012; Perera et al. 2013; Johnson et al. 2017; Batai et al. 2021). However, an FDA-approved test to inform the dosage of the anticoagulant warfarin surveys genetic variants that are not as relevant to Africans. For instance, one of the variants interrogated, rs1799853 in the *CYP2C9* gene, is rare in Africa and thus has limited pharmacogenetic utility in the continent (Dandara et al. 2011; Ndadza et al. 2021). Altogether, these examples illustrate that the heterogeneous admixture histories of African populations are important considerations in biomedical studies and an in-depth understanding of population genetics allows for improved functional annotation that may inform treatment options.

The Need for More Diversity in Genomic Research

Despite what is described here, we have only provided an overview of admixture events in the course of major migratory events, for example, the expansion of Bantu speakers. However, many more interesting admixture events are likely to have occurred along these migratory corridors.

As African population genetics research is still in its early stages compared with its European counterpart (Popejoy and Fullerton 2016; Martin et al. 2018; Sirugo et al. 2019; Fatumo et al. 2022), it is imperative that ongoing efforts to sequence diverse populations on the African continent need to be expanded. Only then we will be able to accrue a holistic picture of human genetic variation and fine-scale population structure on the African continent. It will then be important to understand the biomedical implications of this yet undiscovered genetic variation and population structure in Africa, to reduce health inequities between populations of African and European ancestry as genetics finds its way into clinical applications. For example, a recent study of 180 African hunter–gatherer genomes from 12 populations discovered ~5.3 million novel genetic variants of which ~78% are population-specific and of which many are predicted to be functionally relevant (Fan et al. 2023). Understanding how this population-specific genetic variation influences complex traits is particularly important in the context of polygenic scores. The high genetic diversity contributes to the poor generalizability of polygenic scores in Africa (Majara et al. 2023) because their accuracy decreases with distance to the study cohort (Privé et al. 2022). Furthermore, ~29% (44/154) of the "Likely Pathogenic" ClinVar variants in the data set by Fan et al. (2023) were common in their African data set (i.e., frequency >0.05) but rare outside of Africa (i.e., frequency <0.01). Considered that low frequency is a feature used for determining pathogenicity, this suggests that current classifications of variant pathogenicity are confounded by a lack of diversity in study cohorts. Overall, these examples underline the importance of surveying of genetic variation and population structure in Africa for clinical applications.

At the same time, however, it must be ensured that ethical guidelines and standards are obeyed to avoid unintended group harm. This requires meaningful engagement of community stakeholders on ethical, legal, and social issues as well as the communication of results, to guarantee that the benefits outweigh the risks (Lemke et al. 2022). Lastly, it is also imperative that the same ethical rigor applied to studying living participants needs to be extended to ancient DNA (Gibbon 2020).

The lack of diversity in study cohorts also extends to genomic scientists. Training more diverse scientists and building research capacities on the African continent not only leads to better research but may also help to address the lack of diversity in study cohorts (Hindorff et al. 2018). Altogether, if the current underrepresentation of marginalized groups in genomic research is not corrected, existing inequities are likely to be exacerbated.

Supplementary Material

Supplementary methods are available online at *Genome Biology and Evolution* online.

Acknowledgments

We thank Mimi Holness for her helpful comments and suggestions. We also thank Matt Hansen, Carina Schlebusch, Cesar Fortes-Lima, and Lara R. Arauna for their assistance in accessing African genomic data sets. This work was supported by an NIGMS MIRA grant to J.L. (R35GM133727).

Data Availability

Genotype data from previously published studies were used to generate ADMIXTURE and FEEMS plots (Schlebusch et al. 2012; Mallick et al. 2016; Arauna et al. 2017; Crawford et al. 2017; Hollfelder et al. 2017; Scheinfeldt et al. 2019; Fortes-Lima et al. 2022). Links to all data sources can be found in supplementary table S1, Supplementary Material online. A Snakemake (Mölder et al. 2021) workflow implementing data harmonization and preprocessing as well as ADMIXTURE and FEEMS analyses, including corresponding figures (figs. 2 and 3), is available at GitHub: https://github.com/LachanceLab/AfricanPopulationStructure.

Literature Cited

Alkorta-Aranburu G, et al. 2012. The genetic architecture of adaptations to high altitude in Ethiopia. PLOS Genet. 8:e1003110.

Anagnostou P, et al. 2020. Berbers and Arabs: tracing the genetic diversity and history of Southern Tunisia through genome wide analysis. Am J Phys Anthropol. 173:697–708.

Ansari-Pour N, Plaster CA, Bradman N. 2013. Evidence from Y-chromosome analysis for a late exclusively eastern expansion of the Bantu-speaking people. Eur J Hum Genet. 21:423–429.
Arauna LR, et al. 2017. Recent historical migrations have shaped the gene pool of Arabs and Berbers in North Africa. Mol Biol Evol. 34:318–329.
Auton A, et al. 2015. A global reference for human genetic variation. Nature 526:68–74.
Bajić V, et al. 2018. Genetic structure and sex-biased gene flow in the history of southern African populations. Am J Phys Anthropol. 167: 656–671.
Barbieri C, et al. 2014. Unraveling the complex maternal history of Southern African Khoisan populations. Am J Phys Anthropol. 153:435–448.
Batai K, Hooker S, Kittles RA. 2021. Leveraging genetic ancestry to study health disparities. Am J Phys Anthropol. 175:363–375.
Bergström A, et al. 2020. Insights into human genetic variation and population history from 929 diverse genomes. Science 367:eaay5012.
Bower J. 1991. The pastoral Neolithic of East Africa. J World Prehistory 5:49–82.
Bress A, et al. 2012. Effect of NQO1 and CYP4F2 genotypes on warfarin dose requirements in Hispanic–Americans and African–Americans. Pharmacogenomics 13:1925–1935.
Breton G, et al. 2014. Lactase persistence alleles reveal partial East African ancestry of southern African Khoe pastoralists. Curr Biol. 24:852–858.
Bryc K, et al. 2010. Genome-wide patterns of population structure and admixture in West Africans and African Americans. Proc Natl Acad Sci. 107:786–791.
Busby GB, et al. 2016. Admixture into and within sub-Saharan Africa Pickrell, JK, editor. eLife 5:e15266.
Campbell MC, Ranciaro A. 2021. Human adaptation, demography and cattle domestication: an overview of the complexity of lactase persistence in Africa. Hum Mol Genet. 30:R98–R109.
Cavalli-Sforza LL, Piazza A. 1993. Human genomic diversity in Europe: a summary of human genomic diversity in Europe: a summary of recent research and prospects for the future. Eur J Hum Genet. 1:3–18.
Černý V, Fortes-Lima C, Tříska P. 2021. Demographic history and admixture dynamics in African Sahelian populations. Hum Mol Genet. 30:R29–R36.
Chaichoompu K, et al. 2020. A different view on fine-scale population structure in Western African populations. Hum Genet. 139:45–59.
Chen L, Wolf AB, Fu W, Li L, Akey JM. 2020. Identifying and interpreting apparent neanderthal ancestry in African individuals. Cell 180: 677–687.e16.
Chimusa ER, et al. 2013. Determining ancestry proportions in complex admixture scenarios in South Africa using a novel proxy ancestry selection method. PLoS One 8:e73971.
Chimusa ER, et al. 2014. Genome-wide association study of ancestry-specific TB risk in the South African Coloured population. Hum Mol Genet. 23:796–809.
Choudhury A, et al. 2017. Whole-genome sequencing for an enhanced understanding of genetic variation among South Africans. Nat Commun. 8:2062.
Choudhury A, et al. 2020. High-depth African genomes inform human migration and health. Nature 586:741–748.
Choudhury A, Sengupta D, Ramsay M, Schlebusch C. 2021. Bantu-speaker migration and admixture in Southern Africa. Hum Mol Genet. 30:R56–R63.
Čížková M, et al. 2017. Genetic structure of the western and Eastern African Sahel/Savannah belt and the role of nomadic pastoralists as inferred from the variation of D-loop mitochondrial DNA sequences. Hum Biol. 89:281–302.
Coelho M, Sequeira F, Luiselli D, Beleza S, Rocha J. 2009. On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genetic variation in southwestern Angola. BMC Evol Biol. 9:80.
Crawford NG, et al. 2017. Loci associated with skin pigmentation identified in African populations. Science 358:eaan8433.
Dandara C, et al. 2011. Genetic variants in CYP (-1A2, -2C9, -2C19, -3A4 and -3A5), VKORC1 and ABCB1 genes in a black South African population: a window into diversity. Pharmacogenomics 12:1663–1670.
D'Atanasio E, et al. 2018. The peopling of the last Green Sahara revealed by high-coverage resequencing of trans-Saharan patrilineages. Genome Biol. 19:20.
Daya M, et al. 2013. A panel of ancestry informative markers for the complex five-way admixed South African Coloured population. PLoS One 8:e82224.
Daya M, et al. 2014. Using multi-way admixture mapping to elucidate TB susceptibility in the South African Coloured population. BMC Genomics 15:1021.
de Wit E, et al. 2010. Genome-wide analysis of the structure of the South African Coloured population in the Western Cape. Hum Genet. 128:145–153.
Diallo MY, et al. 2022. Circum-Saharan prehistory through the lens of mtDNA diversity. Genes (Basel). 13:533.
Diamond J, Bellwood P. 2003. Farmers and their languages: the first expansions. Science 300:597–603.
Durvasula A, Sankararaman S. 2020. Recovering signals of ghost archaic introgression in African populations. Sci Adv. 6:eaax5097.
Eltis D. 2007. A brief overview of the trans-atlantic slave trade,' slave voyages: the Trans-Atlantic Slave Trade Database. https://www.slavevoyages.org/ (Accessed July 6, 2021).
Fadhlaoui-Zid K, et al. 2011. Mitochondrial DNA structure in North Africa reveals a genetic discontinuity in the Nile Valley. Am J Phys Anthropol. 145:107–117.
Fadhlaoui-Zid K, et al. 2013. Genome-wide and paternal diversity reveal a recent origin of human populations in North Africa. PLoS One 8:e80293.
Fan S, et al. 2019. African evolutionary history inferred from whole genome sequence data of 44 indigenous African populations. Genome Biol. 20:1–14.
Fan S, et al. 2023. Whole-genome sequencing reveals a complex African population demographic history and signatures of local adaptation. Cell 186:923–939.e14.
Fatumo S, et al. 2022. A roadmap to increase diversity in genomic studies. Nat Med. 28:243–250.
Fortes-Lima C, et al. 2022. Demographic and selection histories of populations across the Sahel/Savannah belt. Mol Biol Evol. 39: msac209.
Fortes-Lima C, Verdu P. 2021. Anthropological genetics perspectives on the transatlantic slave trade. Hum Mol Genet. 30:R79–R87.
Fregel R, et al. 2018. Ancient genomes from North Africa evidence prehistoric migrations to the Maghreb from both the Levant and Europe. Proc Natl Acad Sci. 115:6774–6779.
Gallego Llorente M, et al. 2015. Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa. Science 350:820–822.
Gibbon VE. 2020. African ancient DNA research requires robust ethics and permission protocols. Nat Rev Genet. 21:645–647.
González-Santos M, et al. 2022. Exploring the relationships between genetic, linguistic and geographic distances in Bantu-speaking populations. Am J Biol Anthropol. 179:104–117.
Gopalan S, et al. 2022. Hunter–gatherer genomes reveal diverse demographic trajectories during the rise of farming in Eastern Africa. Curr Biol. 32:1852–1860.e5.
Goudie AS. 2003. Great warm deserts of the world: landscapes and evolution. Oxford, New York: Oxford University Press.

Gouveia MH, et al. 2020. Origins, admixture dynamics, and homogenization of the African gene pool in the Americas. Mol Biol Evol. 37: 1647–1656.
Green RE, et al. 2010. A draft sequence of the neandertal genome. Science 328:710–722.
Gurdasani D, et al. 2015. The African Genome Variation Project shapes medical genetics in Africa. Nature 517:327–332.
Haak W, et al. 2010. Ancient DNA from European Early Neolithic farmers reveals their Near Eastern affinities. PLOS Biol. 8:e1000536.
Haber M, et al. 2016. Chad genetic diversity reveals an African history marked by multiple Holocene Eurasian migrations. Am J Hum Genet. 99:1316–1324.
Hamid I, Korunes KL, Beleza S, Goldberg A. 2021. Rapid adaptation to malaria facilitated by admixture in the human population of Cabo Verde przeworski, M, kana, BD, racimo, F, & busby, G, editors. eLife 10:e63177.
Hammer MF, Woerner AE, Mendez FL, Watkins JC, Wall JD. 2011. Genetic evidence for archaic admixture in Africa. Proc Natl Acad Sci. 108:15123–15128.
Harvati K, et al. 2011. The later stone age Calvaria from Iwo Eleru, Nigeria: morphology and chronology. PLoS One 6:e24024.
Henn BM, et al. 2011. Hunter–gatherer genomic diversity suggests a Southern African origin for modern humans. Proc Natl Acad Sci. 108:5154–5162.
Henn BM, et al. 2012. Genomic ancestry of North Africans supports back-to-Africa migrations. PLOS Genet. 8:e1002397.
Henn BM, Steele TE, Weaver TD. 2018. Clarifying distinct models of modern human origins in Africa. Curr Opin Genet Dev. 53:148–156.
Hervella M, et al. 2016. The mitogenome of a 35,000-year-old homo sapiens from Europe supports a Palaeolithic back-migration to Africa. Sci Rep. 6:25501.
Hey J, et al. 2018. Phylogeny estimation by integration over isolation with migration models. Mol Biol Evol. 35:2805–2818.
Hindorff LA, et al. 2018. Prioritizing diversity in human genomics research. Nat Rev Genet. 19:175–185.
Hollfelder N, et al. 2017. Northeast African genomic variation shaped by the continuity of indigenous groups and Eurasian migrations. PLOS Genet. 13:e1006976.
Hollfelder N, Breton G, Sjödin P, Jakobsson M. 2021. The deep population history in Africa. Hum Mol Genet. 30:R2–R10.
Hsieh P, Veeramah KR, et al. 2016. Whole-genome sequence analyses of Western Central African pygmy hunter–gatherers reveal a complex demographic history and identify candidate genes under positive natural selection. Genome Res. 26:279–290.
Hsieh P, Woerner AE, et al. 2016. Model-based analyses of whole-genome data reveal a complex evolutionary history involving archaic introgression in Central African pygmies. Genome Res 26(3):291–300.
Huerta-Sánchez E, et al. 2013. Genetic signatures reveal high-altitude adaptation in a set of Ethiopian populations. Mol Biol Evol. 30: 1877–1888.
Jallow M, et al. 2009. Genome-wide and fine-resolution association analysis of malaria in West Africa. Nat Genet. 41:657–665.
Jarvis JP, et al. 2012. Patterns of ancestry, signatures of natural selection, and genetic association with stature in Western African pygmies. PLOS Genet. 8:e1002641.
Johnson J, et al. 2017. Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline for pharmacogenetics-guided warfarin dosing: 2017 update. Clin Pharmacol Ther. 102:397–404.
Kariuki SN, Williams TN. 2020. Human genetics and malaria resistance. Hum Genet. 139:801–811.
Karlsson EK, Kwiatkowski DP, Sabeti PC. 2014. Natural selection and infectious disease in human populations. Nat Rev Genet. 15: 379–393.
Kim HL, et al. 2014. Khoisan hunter–gatherers have been the largest population throughout most of modern-human demographic history. Nat Commun. 5:5692.
Lachance J, et al. 2012. Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse African hunter–gatherers. Cell 150:457–469.
Lazaridis I, et al. 2016. Genomic insights into the origin of farming in the ancient near east. Nature 536:419–424.
Lemke AA, et al. 2022. Addressing underrepresentation in genomics research through community engagement. Am J Hum Genet. 109:1563–1571.
Li S, Schlebusch C, Jakobsson M. 2014. Genetic variation reveals large-scale population expansion and migration during the expansion of Bantu-speaking peoples. Proc R Soc B Biol Sci. 281: 20141448.
Lipson M, et al. 2020. Ancient West African foragers in the context of African population history. Nature 577:665–670.
Lipson M, et al. 2022. Ancient DNA and deep population structure in sub-Saharan African foragers. Nat 2022:1–7.
Lokki AI, et al. 2011. Lactase persistence genotypes and malaria susceptibility in Fulani of Mali. Malar J. 10:9.
Lopez M, et al. 2018. The demographic history and mutational load of African hunter–gatherers and farmers. Nat Ecol Evol. 2:721–730.
Lorente-Galdos B, et al. 2019. Whole-genome sequence analysis of a Pan African set of samples reveals archaic gene flow from an extinct basal population of modern humans into sub-Saharan populations. Genome Biol. 20:77.
Lucas-Sánchez M, Font-Porterias N, Calafell F, Fadhlaoui-Zid K, Comas D. 2021. Whole-exome analysis in Tunisian Imazighen and Arabs shows the impact of demography in functional variation. Sci Rep. 11:21125.
Lucas-Sánchez M, Serradell JM, Comas D. 2021. Population history of North Africa based on modern and ancient genomes. Hum Mol Genet. 30:R17–R23.
Macholdt E, et al. 2014. Tracing pastoralist migrations to Southern Africa with lactase persistence alleles. Curr Biol. 24:875–879.
Majara L, et al. 2023. Low and differential polygenic score generalizability among African populations due largely to genetic diversity. Hum Genet Genomics Adv. 4:100184.
Mallick S, et al. 2016. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature 538:201–206.
Manning K, Timpson A. 2014. The demographic response to Holocene climate change in the Sahara. Quat Sci Rev. 101:28–35.
Marcus J, Ha W, Barber RF, Novembre J. 2021. Fast and flexible estimation of effective migration surfaces perry, GH, alves, I, & tansey, W, editors. eLife 10:e61927.
Marshall F, Hildebrand E. 2002. Cattle before crops: the beginnings of food production in Africa. J World Prehistory 16:99–143.
Martin AR, Teferra S, Möller M, Hoal EG, Daly MJ. 2018. The critical needs and challenges for genetic architecture studies in Africa. Curr Opin Genet Dev. 53:113–120.
Mathias RA, et al. 2016. A continuum of admixture in the western hemisphere revealed by the African diaspora genome. Nat Commun. 7:12522.
Matjuda EN, Engwa GA, Anye SNC, Nkeh-Chungag BN, Goswami N. 2021. Cardiovascular risk factors and their relationship with vascular dysfunction in South African children of African ancestry. J Clin Med. 10:354.
Meyer M, et al. 2012. A high-coverage genome sequence from an archaic Denisovan individual. Science 338:222–226.
Micheletti SJ, et al. 2020. Genetic consequences of the transatlantic slave trade in the Americas. Am J Hum Genet. 107:265–277.
Mölder F, et al. 2021. Sustainable data analysis with Snakemake. F1000Res 10:33.

Montinaro F, et al. 2017. Complex ancient genetic structure and cultural transitions in Southern African populations. Genetics 205: 303–316.
Montinaro F, Capelli C. 2018. The evolutionary history of Southern Africa. Curr Opin Genet Dev. 53:157–164.
Naidoo T, et al. 2020. Y-chromosome variation in Southern African Khoe-San populations based on whole-genome sequences. Genome Biol Evol. 12:1031–1039.
Ndadza A, et al. 2021. Profiling of warfarin pharmacokinetics-associated genetic variants: Black Africans portray unique genetic markers important for an African specific warfarin pharmacogenetics-dosing algorithm. J Thromb Haemost. 19:2957–2973.
Newman J. 1995. The peopling of Africa: a geographic interpretation. New Haven: Yale University Press.
Norris ET, et al. 2020. Admixture-enabled selection for rapid adaptive evolution in the Americas. Genome Biol. 21:29.
Nováčková J, et al. 2020. Subsistence strategy was the main factor driving population differentiation in the bidirectional corridor of the African Sahel. Am J Phys Anthropol. 171:496–508.
Ongaro L, et al. 2019. The genomic impact of European colonization of the Americas. Curr Biol. 29:3974–3986.e4.
Pagani L, et al. 2012. Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. Am J Hum Genet. 91:83–96.
Pagani L, et al. 2015. Tracing the route of modern humans out of Africa by using 225 human genome sequences from Ethiopians and Egyptians. Am J Hum Genet. 96:986–991.
Pakendorf B, Stoneking M. 2021. The genomic prehistory of peoples speaking Khoisan languages. Hum Mol Genet. 30:R49–R55.
Patin E, et al. 2014. The impact of agricultural emergence on the genetic history of African rainforest hunter–gatherers and agriculturalists. Nat Commun. 5:3163.
Patin E, et al. 2017. Dispersals and genetic adaptation of Bantu-speaking populations in Africa and North America. Science 356:543 LP–543546.
Patin E, Quintana-Murci L. 2018. The demographic and adaptive history of Central African hunter–gatherers and farmers. Curr Opin Genet Dev. 53:90–97.
Patterson N, et al. 2010. Genetic structure of a unique admixed population: implications for medical research. Hum Mol Genet. 19: 411–419.
Pennarun E, et al. 2012. Divorcing the Late Upper Palaeolithic demographic histories of mtDNA haplogroups M1 and U6 in Africa. BMC Evol Biol. 12:234.
Pereira L, et al. 2010. Linking the sub-Saharan and West Eurasian gene pools: maternal and paternal heritage of the Tuareg nomads from the African Sahel. Eur J Hum Genet. 18:915–923.
Pereira L, Mutesa L, Tindana P, Ramsay M. 2021. African genetic diversity and adaptation inform a precision medicine agenda. Nat Rev Genet. 22:284–306.
Perera MA, et al. 2013. Genetic variants associated with warfarin dose in African–American individuals: a genome-wide association study. Lancet. 382:790–796.
Perry GH, et al. 2007. Diet and the evolution of human amylase gene copy number variation. Nat Genet. 39:1256–1260.
Petersen DC, et al. 2013. Complex patterns of genomic admixture within Southern Africa. PLOS Genet. 9:e1003309.
Pfennig A, Lachance J. 2023. Challenges of accurately estimating sex-biased admixture from X chromosomal and autosomal ancestry proportions. Am J Hum Genet. 110:359–367.
Pickrell JK, et al. 2012. The genetic prehistory of Southern Africa. Nat Commun. 3:1143.
Plagnol V, Wall JD. 2006. Possible ancestral structure in human populations. PLOS Genet. 2:e105.
Popejoy AB, Fullerton SM. 2016. Genomics is failing on diversity. Nat 5387624 538:161–164.
Prendergast ME, et al. 2019. Ancient DNA reveals a multistep spread of the first herders into sub-Saharan Africa. Science 365:eaaw6275.
Prendergast ME. 2020. The history of Eastern African foragers. Oxf Res Encycl Afr Hist. https://doi.org/10.1093/acrefore/9780190277734.013.405
Prendergast ME, Sawchuk EA, Sirak KA. 2022. Genetics and the African past. Oxf Res Encycl Afr Hist. https://doi.org/10.1093/acrefore/9780190277734.013.143
Priehodová E, et al. 2017. The historical spread of Arabian pastoralists to the Eastern African Sahel evidenced by the lactase persistence −13,915*G allele and mitochondrial DNA. Am J Hum Biol. 29: e22950.
Priehodová E, et al. 2020. Sahelian pastoralism from the perspective of variants associated with lactase persistence. Am J Phys Anthropol. 173:423–436.
Privé F, et al. 2022. Portability of 245 polygenic scores when derived from the UK Biobank and applied to 9 ancestry groups from the same cohort. Am J Hum Genet. 109:12–23.
Quiver MH, Lachance J. 2022. Adaptive eQTLs reveal the evolutionary impacts of pleiotropy and tissue-specificity while contributing to health and disease. Hum Genet Genomics Adv. 3:100083.
Ragsdale AP et al. 2022. A weakly structured stem for human origins in Africa. bioRxiv. https://doi.org/10.1101/2022.03.23.485528
Ranciaro A, et al. 2014. Genetic origins of lactase persistence and the spread of pastoralism in Africa. Am J Hum Genet. 94:496–510.
Rasmussen M, et al. 2011. An aboriginal Australian genome reveals separate human dispersals into Asia. Science 334:94–98.
Rees JS, Castellano S, Andrés AM. 2020. The genomics of human local adaptation. Trends Genet. 36:415–428.
Republic of South Africa SD. 2021. Statistical release (P0302): mid-year population estimates 2021. http://www.statssa.gov.za/publications/P0302/P03022021.pdf
Rexová K, Bastin Y, Frynta D. 2006. Cladistic analysis of Bantu languages: a new tree based on combined lexical and grammatical data. Naturwissenschaften 93:189–194.
Robertson JH, Bradley R. 2000. A new paradigm: the African early iron age without Bantu migrations. Hist Afr. 27:287–323.
Rotimi CN, Jorde LB. 2010. Ancestry and disease in the age of genomic medicine. N Engl J Med. 363:1551–1558.
Schaefer NK, Shapiro B, Green RE. 2021. An ancestral recombination graph of human, neanderthal, and Denisovan genomes. Sci Adv. 7:1–17.
Scheinfeldt LB, et al. 2012. Genetic adaptation to high altitude in the Ethiopian highlands. Genome Biol. 13:R1.
Scheinfeldt LB, et al. 2019. Genomic evidence for shared common ancestry of East African hunting–gathering populations and insights into local adaptation. Proc Natl Acad Sci. 116:4166–4175.
Schlebusch CM, et al. 2012. Genomic variation in seven Khoe-San groups reveals adaptation and complex African history. Science 338:374–379.
Schlebusch CM, et al. 2017. Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. Science 358:652–655.
Schlebusch CM, et al. 2020. Khoe-San genomes reveal unique variation and confirm the deepest population divergence in homo sapiens. Mol Biol Evol. 37:2944–2954.
Schlebusch CM, Jakobsson M. 2018. Tales of human migration, admixture, and selection in Africa. Annu Rev Genomics Hum Genet. 19:405–428.
Schlebusch CM, Sjödin P, Skoglund P, Jakobsson M. 2013. Stronger signal of recent selection for lactase persistence in Maasai than in Europeans. Eur J Hum Genet. 21:550–553.

Segurel L, Bon C. 2017. On the evolution of lactase persistence in humans. Annu Rev Genomics Hum Genet. 18:297–319.
Semo A, et al. 2020. Along the Indian ocean coast: genomic variation in Mozambique provides new insights into the Bantu expansion. Mol Biol Evol. 37:406–416.
Sengupta D, et al. 2021. Genetic substructure and complex demographic history of South African Bantu speakers. Nat Commun. 121 12:1–13.
Serra-Vidal G, et al. 2019. Heterogeneity in Palaeolithic population continuity and Neolithic expansion in North Africa. Curr Bioldm 29:3953–3959.e4.
Shriner D, Rotimi CN. 2018a. Genetic history of Chad. Am J Phys Anthropol. 167:804–812.
Shriner D, Rotimi CN. 2018b. Whole-genome-sequence-based haplotypes reveal single origin of the sickle allele during the Holocene wet phase. Am J Hum Genet. 102:547–556.
Sirugo G, Williams SM, Tishkoff SA. 2019. The missing diversity in human genetic studies. Cell 177:1080.
Sjöstrand AE, et al. 2021. Taste perception and lifestyle: insights from phenotype and genome data among Africans and Asians. Eur J Hum Genet. 29:325–337.
Skoglund P, et al. 2017. Reconstructing prehistoric African population structure. Cell 171:59–71.e21.
Solé-Morata N, et al. 2017. Whole Y-chromosome sequences reveal an extremely recent origin of the most common North African paternal lineage E-M183 (M81). Sci Rep. 7:15941.
Sugden LA, et al. 2018. Localization of adaptive variants in human genomes using averaged one-dependence estimation. Nat Commun. 9:703.
Swart Y, Uren C, van Helden PD, Hoal EG, Möller M. 2021. Local ancestry adjusted allelic association analysis robustly captures tuberculosis susceptibility loci. Front Genet. 12:716558.
Swart Y, van Eeden G, Sparks A, Uren C, Möller M. 2020. Prospective avenues for human population genomics and disease mapping in Southern Africa. Mol Genet Genomics 295:1079–1089.
Tallman S, Sungo M das D, Saranga S, Beleza S. 2022. Whole-genome sequencing of Bantu-speakers from Angola and Mozambique reveals complex dispersal patterns and interactions throughout sub-Saharan Africa. bioRxiv. https://doi.org/10.1101/2022.02.07.478793
Tishkoff SA, et al. 2009. The genetic structure and history of Africans and African Americans. Science 324:1035–1044.
Triska P, et al. 2015. Extensive admixture and selective pressure across the Sahel belt. Genome Biol Evol. 7:3484–3495.
Uren C, et al. 2016. Fine-scale human population structure in Southern Africa reflects ecogeographic boundaries. Genetics 204:303–314.
Vai S, et al. 2019. Ancestral mitochondrial N lineage from the Neolithic 'green' Sahara. Sci Rep. 9:3530.
Van De Loosdrecht M, et al. 2018. Pleistocene North African genomes link Near Eastern and sub-Saharan African human populations. Science 360:548–552.
Verdu P, et al. 2013. Sociocultural behavior, sex-biased admixture, and effective population sizes in Central African pygmies and non-pygmies. Mol Biol Evol. 30:918–937.
Vicente M, et al. 2021. Male-biased migration from East Africa introduced pastoralism into Southern Africa. BMC Biol. 19:259.
Vicente M, Jakobsson M, Ebbesen P, Schlebusch CM. 2019. Genetic affinities among Southern Africa hunter–gatherers and the impact of admixing farmer and herder populations. Mol Biol Evol. 36:1849–1861.
Vicente M, Priehodová E, et al. 2019. Population history and genetic adaptation of the Fulani nomads: inferences from genome-wide data and the lactase persistence trait. BMC Genomics 20:915.
Wall JD, et al. 2019. Identification of African-specific admixture between modern and archaic humans. Am J Hum Genet. 105:1254–1261.
Wang K, et al. 2020a. Ancient genomes reveal complex patterns of population movement, interaction, and replacement in sub-Saharan Africa. Sci Adv. 6:eaaz0183.
Wang K, et al. 2022. 4000-Year-old hair from the Middle Nile highlights unusual ancient DNA degradation pattern and a potential source of early Eastern Africa pastoralists. Sci Rep. 12:20939.
Wang K, Mathieson I, O'Connell J, Schiffels S. 2020. Tracking human population structure through time from whole genome sequences. PLOS Genet. 16:e1008552.
Wohlers I, et al. 2020. An integrated personal and population-based Egyptian genome reference. Nat Commun. 11:4719.
World Health Organization. 2021. *World malaria report* 2021. Geneva.
Xu D, et al. 2017. Archaic hominin introgression in Africa contributes to functional salivary MUC7 genetic variation. Mol Biol Evol. 34: 2704–2715.
Zhou Q, Zhao L, Guan Y. 2016. Strong selection at MHC in Mexicans since admixture. PLOS Genet. 12:e1005847.

Associate editor: Adam Eyre-Walker