African genetic diversity provides novel insights into evolutionary history and local adaptations

Ananyo Choudhury[1], Shaun Aron[1], Dhriti Sengupta[1], Scott Hazelhurst[1,2] and Michèle Ramsay[1,3]

[1]Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

[2]School of Electrical & Information Engineering, University of the Witwatersrand, Johannesburg, South Africa

[3]Division of Human Genetics, National Health Laboratory Service and School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

Abstract

Genetic variation and susceptibility to disease are shaped by human demographic history. We can now study the genomes of extant Africans and uncover traces of population migration, admixture, assimilation and selection by applying sophisticated computational algorithms. There are four major ethnolinguistic divisions among present day Africans: Hunter-gatherer populations in southern and central Africa; Nilo-Saharan speakers from north and northeast Africa; Afro-Asiatic speakers from east Africa; and Niger-Congo speakers who are the predominant ethnolinguistic group spread across most of sub-Saharan Africa. The enormous ethnolinguistic diversity in sub-Saharan African populations is largely paralleled by extensive genetic diversity and until a decade ago, little was known about the origins and divergence of these groups. Results from large-scale population genetic studies, and more recently whole genome sequence data, are unraveling the critical role of events like migration and admixture and environment factors including diet, infectious diseases and climatic conditions in shaping current population diversity. It is now possible to start providing quantitative estimates of divergence times, population size and dynamic processes that have affected populations and their genetic risk for disease. Finally, the availability of ancient genomes from Africa is providing historical insights of unprecedented depth. In this review, we highlight some key interpretations that have emerged from recent African genome studies.

Early population divergence in Africa

The genomic diversity in Africa has largely been shaped by deep population structure maintained by relatively long periods of geographic and cultural isolation. **Figure 1A** summarizes some of the ancient divergence events that have been critical to the origin of the major ethnolinguistic divisions. The oldest split in the human population has been suggested to be between Khoesan (KS) hunter-gatherers and other human populations. Although the estimated date of this event varies across studies, there is general consensus that the split occurred over 100 thousand years ago (kya) (1,2). The next major split was between the rain-forest foragers (RFF) and other human lineages, and is estimated to have occurred about 60-70 kya (2–5). Studies suggest that around 30 kya the KS population differentiated into the northern KS (e.g. Ju/'hoansi) and southern KS (e.g. ≠Khomani) groups (1,6). Similarly, it is estimated that around 20 kya the western RFF (e.g. Baka) diverged from the eastern RFF (e.g. Mbuti) (2,3,5). The separation of the Afro-Asiatic speakers currently living in east Africa from other sub-Saharan African (SSA) populations has been dated to about 50 kya (7). The final major separation is suggested to be between the Nilo-Saharan and Niger-Congo

speakers dated to ~28 kya (7). Despite being restricted to a rather small geographic area, genetic data suggest a significant differentiation within both Afro-Asiatic and Nilo-Saharans in east Africa (8–10). Similarly the Bantu and non-Bantu Niger-Congo speakers (from west Africa) also demonstrate observable genetic differentiation (11,12).

Notably, most of the divergence time estimates described above were based on uniparental markers or a limited set of autosomal markers. Computational analyses based on whole genome sequence (WGS) data from extant and ancient individuals are providing alternative timelines for some of these events. A recent ancient genome study (13) supports a much earlier date (>260 kya) for KS divergence from other human lineages as proposed by Scally and Durbin in 2012 (14). Similarly, the RFF split has been proposed to be up to 150 kya old (15). Both of these estimates could be about ~100 thousand years older than previous estimates hinting at deep population structure on the continent (**Figure 1B**). Coalescence analyses of WGS data are providing better estimates of demographic history, for example, population size estimates using Markovian coalescent simulations have suggested large population sizes of KS for the majority of human history (16,17). These populations appear not to have been affected by the events ~30-120 kya that significantly reduced the size of all other human populations (16). The observation of possible ancient KS admixture in western Africa (11) and widespread distribution of KS related ancestry in eastern and southern Africa (13,18) further indicate a much wider geographic spread of this group compared to their present distribution.

Recently, based on an analysis of 16 ancient African genomes, Skoglund and colleagues proposed a model that suggests an early separation of a west African branch from other human lineages that might predate the split of KS hunter-gatherers (18) (**Figure 1B**). Though this dating requires further investigation, the existence of at least two ancient and semi-independent populations, one in east and one in west Africa, in addition to the ancestors of the present day hunter-gatherer and forager populations seems a strong possibility. This could also help to explain the differentiation of the Bantu and non-Bantu genetic components that have been observed in several studies of west African populations (11,12). In addition, there is also strong evidence that these populations were not isolated, and that there was bidirectional gene flow between east and west Africa for a significant proportion of history. For example, a study based on Y chromosome markers in thousands of individuals showed considerable population movement and demographic expansion across the present day Sahara during the recent ”green Sahara” period that ended ~5 kya (19). Similarly, computational models also predict multiple ancient gene flow events between RFF and Niger-Congo speakers (7,15). Although the evidence for ancient introgression in Africa is still not adequate, it is highly plausible that such events have contributed to a fraction of the current genetic diversity (20,21).

Studies based on populations from a wide range of geographic locations are uncovering evidence of other ancient migration events. One such migration was of east African pastoralists to southern Africa, at a time that predates the Bantu-migration by more than a thousand years (18,22,23). This migration event is thought to be the source of lactose tolerance variants in southern Africa (24–26). Similarly, evidence for a relatively old Nilo-Saharan migration to western Africa, across the Sahel region, has been identified in some of the populations from this region (12). The availability of more deep-sequenced genomes, including ancient genomes, along with the development of improved algorithms would lead to more robust models of African demographic history, ancient migrations and gene flow.

The Bantu Expansion ~5000 years ago

One of the most recent events shaping the genomic landscape in Africa is the movement of Bantu-speaking populations from the central western region of Nigeria and Cameroon into the rest of SSA. Archaeological evidence supports the spread of Bantu languages together with the advent of agricultural practices from western central Africa to eastern, western and southern Africa between ~4 to 5 kya (27). The exact route and extent of the Bantu-migration across SSA remains an area of active research although several recent genetic studies have provided evidence of regional admixture events that support specific routes. It is likely that multiple waves of migration, coupled with exposure to new environments and interactions with indigenous hunter-gatherer populations, are obscuring some historical events (28).

There are currently two proposed routes to explain the migration of Bantu-speakers into SSA (12,29,30). The early-split model suggests that the western and eastern branches split early within the region of the Nigeria/Cameroon frontier around 3 to 2.5 kya and migrated directly east to the Great Lakes Region of east Africa. The late-split model suggests that populations first moved into the equatorial rainforest region near Gabon/Angola with a split to the east ~2 kya. Although both routes are supported by several genetic studies, more recently there is greater support for the late-split hypothesis (31) based on data showing that eastern Bantu-speakers and southeastern Bantu-speakers are genetically closer to western Bantu-speakers from the southern, rather than the northern region. Furthermore, haplotype analyses showed that eastern Bantu-speakers arose from two consecutive admixture events between western Bantu-speakers and an Afro-Asiatic speaking population from Ethiopia, the first occurring 1000 to 1500 years ago and the second a mere 150 to 400 years ago. The subsequent best-matched western Bantu-speaking parental population for both admixture events was identified from Angola. Furthermore, southeastern Bantu-speakers show a unique admixture event ~700 years ago between a parental Bantu-speaking population, most likely located in Angola, and the Jo/'hoansi San from Namibia. Together with further evidence of rain-forest hunter-gatherer admixture dated at ~800 years ago in western Bantu-speakers, this presents strong support that the Bantu migration first travelled south through the central rainforests before moving to the eastern and southern regions of the continent (31).

The late split hypothesis further supports two main migration paths into southern Africa, one along the east of Africa and another along the west, giving rise to the southeastern Bantu-speakers (SEB) and southwestern Bantu-speakers (SWB), respectively (27). The Bantu migration into southern Africa resulted in interactions with several established groups already present in the region, including various KS groups **(Figure 2a)** and this contributed to the extensive genetic diversity observed in the current Bantu-speaking populations in South Africa (30,32).

Eurasian gene flow back to Africa

Genome-wide data have revealed significant Eurasian ancestry in African populations assimilated across a long timescale. Initial studies limited this Eurasian migration to a single event in east Africa (33), but subsequent studies revealed a more complex series of distinct admixture events occurring in western and eastern Africa **(Figure 2b)**. The African Genome Variation Project (AGVP) used genome data from 18 ethnolinguistic groups in 8 African countries to identify Eurasian admixture ranging from 0-50% in and across central, west and east African populations (11). This distinctive Eurasian admixture appears to have occurred over at least three time periods with ancient admixture in central west Africa (e.g. Yoruba

from Nigeria) occurring between ~7.5 and 10.5 kya (11), older admixture in east Africa (e.g. Ethiopia) occurring between ~2.4 and 3.2 kya (11,23,33) and more recent admixture between ~0.15 and 1.5 kya in some east African (e.g. Kenyan) populations (11).

Subsequent studies based on LD decay and haplotype-sharing in an extensive set of African and Eurasian populations confirmed the presence of Eurasian signatures in west, east and southern Africans. In the west, in addition to Niger-Congo speakers from The Gambia and Mali, the Mossi from Burkina Faso showed the oldest Eurasian admixture event ~7 kya (12). In the east, these analyses inferred Eurasian admixture within the last 4000 years in Kenya, with the Chonyi and Kauma showing possible south Asian admixture that may have been facilitated by Medieval trade across the Indian Ocean. Eurasian admixture in the Afro-Asiatic speaking populations of east Africa appears to have occurred in two waves, one ~2 kya and one in the last 200 years with ancestry best matched to the Tuscans from Italy. In the south, more recent direct Eurasian admixture from northern European populations has been observed in Khoesan-speaking populations in South Africa (**Figure 2B**). These include the Khomani and Karretjie, with events dating to ~225 years ago, with an historical link to the European colonial period in South Africa (12).

More recent gene flow and the African diaspora

The second millennium CE saw significant population exchanges between Africa and other continents. Well before the age of European exploration, there were well established trade routes between Africa and centres in Asia. Brucato et al. (34) recently surveyed these economic connections and compared them to finger prints of population exchange that can be seen in the genomes of current populations. Based on historical records, trade patterns are divided into four phases. In the first two phases, the trade routes where mainly land-based through the Sinai Peninsula or via the Bab al-Mandab, reaching deep into southern Africa. During the third phase, starting in the 11th Century CE, there was significant sea-borne trade from the Arabian Peninsula and Indonesia to east Africa and to Madagascar. Trade extended further into central Africa through to the Gulf of Guinea. Brucato et al. (34) show that this trade came with gene flow, and these events were highly correlated with peaks and troughs in trade routes.

The biggest drivers at a population level, out of and into Africa, were the slave trade and colonisation that led to millions of Africans forcibly leaving Africa for the Americas and Asia, and, to a lesser extent European and Asian populations establishing themselves in Africa, with some admixture into African populations. Slavery and forced migration of people is a human practice dating millennia. Over the last thousand years, millions of Africans were enslaved and sent to new homes thousands of kilometres away. Although there are many detailed records of where people boarded boats that took them to captivity, less is known about where they actually came from. Recent research, including genetic studies, on the African diaspora and groups who have oral traditions of African ancestry have shed new light on these histories.

The Arab slave trade from at least the 8th Century to the late 19th Century had two distinct routes. The first was the trade either into the Mahgreb and Egypt or through these places into Asia. There is no good method of accurately estimating the numbers, but several million people have conservatively been estimated to have been enslaved (35). A second Indian Ocean trade route from the east coast, particularly from the 16th Century onwards enslaved a further four million people (36). In the first phase mainly Nilotic or Afro-Asiatic speaking

people from the Horn of Africa were involved and from the 19th Century, mainly Bantu-speaking people, probably via regions controlled by the Omani Empire from their base in Zanzibar. At this scale, there is expected to be significant evidence of gene flow, but relatively little research has been done on the origins of these people or what became of them.

Laso-Jadart et al. (36) studied people from the Makrani group in Pakistan who have traditions suggesting descent from African slaves. Genotype data from 102 individuals showed an average 25% Africa ancestry, with significant diversity, and suggested that the likely source is southeastern Bantu-speaking people. GLOBEtrotter analysis suggests a single wave of admixture dating to the late 18th Century, although other models are possible.

The European slave trade from Africa took place over a shorter period of time, from the 16th-19th Centuries, but at a much more intense scale. Numbers are difficult to estimate, especially as many people who were enslaved died *en route*, but approximately 12 million people were enslaved and taken to the Americas from the 16th-19th centuries (37) of whom 7 million went to South America (38). Several recent studies have investigated their African origins (31,39,40). Patin et al. (31) used data from 2500 Africans and 5200 African-Americans to explore the range of African ancestry in African-Americans. They estimate that ~50% of the African ancestry of African-Americans originates from the Bight of Benin, ~30% from western central Africa (current day Angola and DRC), ~13% from Senegambia and ~7% from the Windward Coast (present day Cote d'Ivoire). Mathias et al. (40) detected similar patterns and showed considerable variability in admixture patterns. Sex-biased admixture patterns are evident, with significantly higher proportions of European male-biased ancestry (40,41) – contrasting with historical records showing that more African men were brought to the Americas than women (42). Complementing the broad-brushstroke papers, several recent studies have explored admixture in particular communities. These studies are bringing out the complexity of history, as well as the impact of admixture on health today (38,43–46).

Admixture into Africa in the last two millennia

The most significant admixture in Africa in the last two thousands years is likely due to invasions from Arabia after the rise of Islam. Recent work has shown very significant impact on north African and north-eastern populations (10,47–49). This impact is greatest in Egypt and Mahgreb, and strong signals can be found elsewhere. However, there is strong intriguing evidence for admixture down the east coast of Africa into southern Africa (34,50).

The peopling of Madagascar has been of great interest for decades. It has been long known from archaeological, linguistic, mt and Y DNA data that Madagascar was settled by people from Africa and Austronesia in the first millennium but the dates and sequence of events has been unclear. Madagascar comprised a set of independent kingdoms until being unified in the early 19th Century and subsequently colonised by the French in the late 19th Century. Recent genetic evidence demonstrates a complex pattern of settlement and interaction resulting in a heterogeneous population (51). The genetic evidence suggests that two major settlements occurred: Austronesian populations settling in the east of the country roughly 1.5 to 2.5 kya, and Bantu-speaking populations settling roughly around 1.5 kya, with admixture occurring over the last 1500 years. On average the relative contributions are ~37% Austronesian and ~59% African, although there is significant geographically clustered variation.

European colonisation brought many people from elsewhere into Africa, particularly in southern Africa. The most remarkable story is that of the Coloured community in South Africa. The Dutch brought slaves from India and Indonesia, who mixed with Europeans and Khoi, San, and Bantu-speaking people. This community came into being from the end of the 17th to the start of the 19th Century. The latest evidence from genotyping and sequencing studies illustrates this complex admixture (32,52,53).

Contributions from African whole genome sequence data

Population-level deep-sequencing whole genome data are enhancing our understanding of diversity and LD architecture in African populations (11,54,55). On average Africans have ~20-30% more SNVs compared to non-Africans (11,32,54,55) illustrating their tremendous potential for novel variant discovery. For example, the AGVP study of 320 African genomes identified ~9.5 million novel SNVs (11). Similarly, the Southern African Human Genome Programme study identified ~0.8 million novel SNVs from only 24 African genomes (32). These and other African WGS data have been used to develop genotyping platforms such at the Human Heredity and Health in Africa Consortium (H3Africa) SNP array that has better coverage of African genetic diversity and a more comprehensive representation of African LD blocks to enhance African genome-wide association studies (GWASs) (https://commonfund.nih.gov/globalhealth/h3aresources).

The statistical technique of imputation, providing genotypes for variants not present on a genotyping array, has emerged as a widely used technique to improve the effectiveness of GWASs (56,57). However, the quality and accuracy of imputation depends largely on the size of reference panels and the representation of haplotypes from the population to be studied (11,58,59). African haplotypes are still underrepresented in most reference panels (60,61) and therefore the inclusion of more African genomes into current reference panels is expected to significantly improve the quality of imputation in African studies. An African-centric reference panel of about 5000 genomes (about half of which are novel African genomes) is available at the Sanger Imputation server (https://imputation.sanger.ac.uk/) and others are under development. Better representation of African ethnolinguistic groups will provide a more comprehensive catalogue of genetic diversity and LD architecture in Africa and such studies are underway within the H3Africa Consortium (62,63).

Natural selection has also shaped African genetic diversity

Sub-Saharan Africa has a high burden of life-threatening non-communicable diseases, especially in regions around the tropical rain forests and such environments are suggested to have contributed to a number of local adaptations (11,39). Consistent with the current spread and prevalence of malaria in Africa, many of the first loci identified to be under strong selection in African populations have been associated with adaptation to malaria. The most studied example is the sickle cell mutation in the *HBB* gene (64–66), with haplotype analyses suggesting that the mutation arose independently at least four times in Africa (67). Allele frequencies are highest in west Africa and some east African populations where Plasmodium falciparum malaria is endemic. The *CD36* malaria associated allele is most common in Nigerian populations with lower frequencies elsewhere. Selection of variants in the *DARC* gene are known to be adaptive to Plasmodium vivax malaria, with some variants having reached fixation in selected west and east African populations **(Table 1)**. (31,68,69). Recent studies showed that copy number variants and complex structural rearrangements in the *GYPA, GYPB* and *GYPC* genes, which reduce the risk of severe malaria by up to 40%, have

also been selected in some east African populations (70). As another example, adaptive selection has been observed for *APOL1*, *LARGE* and *IL21* variants that protect against African trypanosomiasis and Lassa fever (11,71–74).

A major selective sweep around the *LCT* and *MCMC6* genes has been associated with dietary adaptation in pastoralist populations around the world. Variants in this genomic region confer a lactase persistence phenotype (the ability to digest fresh milk as an adult) and have been found to have higher allele frequencies in east and southern African populations (not well represented in the populations shown in Table 1) in parallel with the geographic spread of pastoralism (25,75,76). Similarly, variants in several genes, including *VAV3*, *ARNT2* and *THRB,* are suggested to be involved in high-altitude adaptation and were identified to be more common in populations native to the Ethiopian highlands (77). Selective sweeps have also been detected for variants in loci associated with short stature (*POU1F1, HESX1, DOCK3, CISH* and *STAT5*) in both east central and west central RFF populations (20,77). It is hypothesized that the short stature phenotype may confer some benefits in response to food scarcity, high humidity and heat or as a trade-off between cessation of growth and early reproduction (78).

With the availability of genome-scale data, there has been a surge in studies aimed at identifying signatures of selection in a wide range of African populations and this has led to the discovery of many new adaptive loci (18,31,79–81). Some studies have explored novel approaches to detect regions under selection, in addition to the traditionally used neutrality tests. Skoglund et al. 2017 (18) for example, used the comparison of ancient southern African genomes to genomes of modern-day San, and identified taste-receptor gene clusters to be under selection in the present day populations. Crawford et al. 2017 (79) employed a GWAS-based approach to identify loci associated with skin pigmentation and then studied selective sweeps around the signals (including *MFSD12, HERC2/OCA2, SLC24A5, TYR, DDB1/TMEM*). Sugden et al. 2018 (80) used a novel method to identify selection in Khoesan populations and detected selective sweeps around various metabolism-related genes.

Computational analyses are also generating insights into the evolutionary history of some of the positively selected variants. For example, some of the variants that show signatures of selection in present-day populations have likely been introduced by admixture (gene flow), a process referred to as adaptive introgression. The HLA loci in western Bantu-speakers and LCT loci in eastern Bantu-speakers were found to show an excess of local ancestry from rainforest hunter-gatherers and east Africans (Nilo-Saharan/Afro-Asiatic), respectively, hinting at possible adaptive introgression (31). Similarly, the positively selected variant (rs1426654) in the *SLC24A5* gene was estimated to have evolved roughly 30 thousand years ago in Eurasian lineages and is proposed to have been re-introduced in the east African populations in the last 3 to 9 thousand years (48,79,82). Interestingly, another study showed the *SLC24A5* derived allele to be more common in the lighter pigmented Khoesan populations hinting at either convergent evolution or Eurasian gene flow thousands of years ago (83). These studies suggest that the use of a wider variety of African populations along with a more comprehensive toolkit for detecting selective sweeps could generate a better understanding of the variation due to adaptive selection across the continent.

Conclusions and future perspective

Our inherent optimism in the power of scientific discovery has not diminished the challenges of understanding complex population relationships. These include the interpretation of genetic diversity to reconstruct human history and our ability to understand gene-to-gene, gene-to-environment interactions and inter-organismal relationships in shaping health and susceptibility to disease. Many unresolved questions related to the genetic variation observed in African populations remain. For example, when and where did key events during the Bantu migration take place? What processes during these events essentially shaped the emergence of the Bantu-speakers? How did cultural practices such as sex-biased behaviour influence admixture events during migrations?

African genome data remain scarce and many African populations are yet to be studied. It is now evident that hunter-gatherer groups historically populated large regions of SSA and left traces of their genomes in almost all the populations currently living on the continent. It is important that we engage with extant hunter-gatherer groups about participating in genome studies to ensure a better understanding of their demographic history and the high genetic diversity among their members. The San communities remain poorly studied and to address potential exploitation of this vulnerable group, The San Council of South Africa launched a code of ethics for research in 2017 where they document their views and terms for promoting equitable and fair research in their communities, with respect and honesty (http://trust-project.eu/san-council-launches-san-code-of-ethics/).

There is great opportunity in studying ancient African genomes. The race for accessing specimens is evident (84) and therefore this intensifies the need to share responsibly and to preserve specimens for a time when improved techniques could provide more insights.

Engrained in the genomes of Africans is the deep and rich history of our species and evidence of migrations that reach the furthest corners of our planet. Over the past decade novel approaches to interpreting DNA sequence divergence have generated hypotheses that have yet to stand the test of time. African populations continue to provide fascinating insights on how diversity was shaped by historical events triggered by changing environments, including climatic and dietary adaptation and encounters with infectious pathogens, and sometimes by chance.

Acknowledgements

MR is a South African Research Chair in Genomics and Bioinformatics of African populations hosted by the University of the Witwatersrand, funded by the Department of Science and Technology and administered by National Research Foundation of South Africa (NRF). AC and DS are funded by the National Institutes of Health (National Human Genome Research Institute (NHGRI)) AWI-Gen Collaborative Centre under award number U54HG006938, as part of the H3Africa Consortium. SH is partially funded and SA is funded by the National Institutes of Health, through the Pan-African Bioinformatics Network for H3Africa (H3ABioNet) award number U41HG006941.

Conflict of interest
None

References

1. Schlebusch,C.M., Skoglund,P., Sjödin,P., Gattepaille,L.M., Hernandez,D., Jay,F., Li,S., De Jongh,M., Singleton,A., Blum,M.G.B., *et al.* (2012) Genomic variation in seven Khoe-San groups reveals adaptation and complex African history. *Science (80-.).*, 10.1126/science.1227721.
2. Veeramah,K.R., Wegmann,D., Woerner,A., Mendez,F.L., Watkins,J.C., Destro-Bisol,G., Soodyall,H., Louie,L. and Hammer,M.F. (2012) An early divergence of KhoeSan ancestors from those of other modern humans is supported by an ABC-based analysis of autosomal resequencing data. *Mol. Biol. Evol.*, 10.1093/molbev/msr212.
3. Patin,E., Laval,G., Barreiro,L.B., Salas,A., Semino,O., Santachiara-Benerecetti,S., Kidd,K.K., Kidd,J.R., Der Veen,L. Van, Hombert,J.M., *et al.* (2009) Inferring the demographic history of African farmers and Pygmy hunter-gatherers using a multilocus resequencing data set. *PLoS Genet.*, 10.1371/journal.pgen.1000448.
4. Verdu,P., Austerlitz,F., Estoup,A., Vitalis,R., Georges,M., Théry,S., Froment,A., Le Bomin,S., Gessain,A., Hombert,J.M., *et al.* (2009) Origins and Genetic Diversity of Pygmy Hunter-Gatherers from Western Central Africa. *Curr. Biol.*, 10.1016/j.cub.2008.12.049.
5. Batini,C., Lopes,J., Behar,D.M., Calafell,F., Jorde,L.B., Van Der Veen,L., Quintana-Murci,L., Spedini,G., Destro-Bisol,G. and Comas,D. (2011) Insights into the demographic history of African pygmies from complete mitochondrial genomes. *Mol. Biol. Evol.*, 10.1093/molbev/msq294.
6. Pickrell,J.K., Patterson,N., Barbieri,C., Berthold,F., Gerlach,L., Güldemann,T., Kure,B., Mpoloka,S.W., Nakagawa,H., Naumann,C., *et al.* (2012) The genetic prehistory of southern Africa. *Nat. Commun.*, 10.1038/ncomms2140.
7. Shriner,D., Tekola-Ayele,F., Adeyemo,A. and Rotimi,C.N. (2014) Genome-wide genotype and sequence-based reconstruction of the 140,000 year history of modern human ancestry. *Sci. Rep.*, 10.1038/srep06055.
8. Gebremeskel,E.I. and Ibrahim,M.E. (2014) Y-chromosome E haplogroups: their distribution and implication to the origin of Afro-Asiatic languages and pastoralism. *Eur. J. Hum. Genet.*, 10.1038/ejhg.2014.41.
9. Dobon,B., Hassan,H.Y., Laayouni,H., Luisi,P., Ricaño-Ponce,I., Zhernakova,A., Wijmenga,C., Tahir,H., Comas,D., Netea,M.G., *et al.* (2015) The genetics of East African populations: A Nilo-Saharan component in the African genetic landscape. *Sci. Rep.*, 10.1038/srep09996.
10. Hollfelder,N., Schlebusch,C.M., Günther,T., Babiker,H., Hassan,H.Y. and Jakobsson,M. (2017) Northeast African genomic variation shaped by the continuity of indigenous groups and Eurasian migrations. *PLoS Genet.*, 10.1371/journal.pgen.1006976.
11. Gurdasani,D., Carstensen,T., Tekola-Ayele,F., Pagani,L., Tachmazidou,I., Hatzikotoulas,K., Karthikeyan,S., Iles,L., Pollard,M.O., Choudhury,A., *et al.* (2015) The African Genome Variation Project shapes medical genetics in Africa. *Nature*, 10.1038/nature13997.
12. Busby,G.B., Band,G., Si Le,Q., Jallow,M., Bougama,E., Mangano,V.D., Amenga-Etego,L.N., Enimil,A., Apinjoh,T., Ndila,C.M., *et al.* (2016) Admixture into and within sub-Saharan Africa. *Elife*, 10.7554/eLife.15266.
13. Schlebusch,C.M., Malmström,H., Günther,T., Sjödin,P., Coutinho,A., Edlund,H., Munters,A.R., Vicente,M., Steyn,M., Soodyall,H., *et al.* (2017) Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. *Science (80-.).*, 10.1126/science.aao6266.

14. Scally,A. and Durbin,R. (2012) Revising the human mutation rate: implications for understanding human evolution. *Nat. Rev. Genet.*, 10.1038/nrg3295.
15. Hsieh,P.H., Veeramah,K.R., Lachance,J., Tishkoff,S.A., Wall,J.D., Hammer,M.F. and Gutenkunst,R.N. (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.*, 10.1101/gr.192971.115.
16. Kim,H.L., Ratan,A., Perry,G.H., Montenegro,A., Miller,W. and Schuster,S.C. (2014) Khoisan hunter-gatherers have been the largest population throughout most of modern-human demographic history. *Nat. Commun.*, 10.1038/ncomms6692.
17. Mallick,S., Li,H., Lipson,M., Mathieson,I., Gymrek,M., Racimo,F., Zhao,M., Chennagiri,N., Nordenfelt,S., Tandon,A., *et al.* (2016) The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature*, 10.1038/nature18964.
18. Skoglund,P., Thompson,J.C., Prendergast,M.E., Mittnik,A., Sirak,K., Hajdinjak,M., Salie,T., Rohland,N., Mallick,S., Peltzer,A., *et al.* (2017) Reconstructing Prehistoric African Population Structure. *Cell*, 10.1016/j.cell.2017.08.049.
19. D'Atanasio,E., Trombetta,B., Bonito,M., Finocchio,A., Di Vito,G., Seghizzi,M., Romano,R., Russo,G., Paganotti,G.M., Watson,E., *et al.* (2018) The peopling of the last Green Sahara revealed by high-coverage resequencing of trans-Saharan patrilineages. *Genome Biol.*, 10.1186/s13059-018-1393-5.
20. Lachance,J., Vernot,B., Elbers,C.C., Ferwerda,B., Froment,A., Bodo,J.M., Lema,G., Fu,W., Nyambo,T.B., Rebbeck,T.R., *et al.* (2012) Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse African hunter-gatherers. *Cell*, 10.1016/j.cell.2012.07.009.
21. Hsieh,P.H., Woerner,A.E., Wall,J.D., Lachance,J., Tishkoff,S.A., Gutenkunst,R.N. and Hammer,M.F. (2016) Model-based analyses of whole-genome data reveal a complex evolutionary history involving archaic introgression in Central African Pygmies. *Genome Res.*, 10.1101/gr.196634.115.
22. Gomez,F., Hirbo,J. and Tishkoff,S.A. (2014) Genetic variation and adaptation in Africa: Implications for human evolution and disease. *Cold Spring Harb. Perspect. Biol.*, 10.1101/cshperspect.a008524.
23. Pickrell,J.K., Patterson,N., Loh,P.-R., Lipson,M., Berger,B., Stoneking,M., Pakendorf,B. and Reich,D. (2014) Ancient west Eurasian ancestry in southern and eastern Africa. *Proc. Natl. Acad. Sci.*, **111**, 2632–2637.
24. Breton,G., Schlebusch,C.M., Lombard,M., Sjödin,P., Soodyall,H. and Jakobsson,M. (2014) Lactase persistence alleles reveal partial east African ancestry of southern African Khoe pastoralists. *Curr. Biol.*, 10.1016/j.cub.2014.02.041.
25. Macholdt,E., Slatkin,M., Pakendorf,B. and Stoneking,M. (2015) New insights into the history of the C-14010 lactase persistence variant in Eastern and Southern Africa. *Am. J. Phys. Anthropol.*, **156**, 661–4.
26. Macholdt,E., Lede,V., Barbieri,C., Mpoloka,S.W., Chen,H., Slatkin,M., Pakendorf,B. and Stoneking,M. (2014) Tracing pastoralist migrations to southern Africa with lactase persistence alleles. *Curr. Biol.*, 10.1016/j.cub.2014.03.027.
27. Phillipson, W,D. (2005) African Archaeology Third. Cambridge University Press, Cambridge.
28. Tishkoff,S.A., Reed,F.A., Friedlaender,F.R., Ehret,C., Ranciaro,A., Froment,A., Hirbo,J.B., Awomoyi,A.A., Bodo,J.M., Doumbo,O., *et al.* (2009) The genetic structure and history of Africans and African Americans. *Science (80-.).*, 10.1126/science.1172257.
29. de Filippo,C., Bostoen,K., Stoneking,M. and Pakendorf,B. (2012) Bringing together linguistic and genetic evidence to test the Bantu expansion. *Proc. R. Soc. B Biol. Sci.*,

10.1098/rspb.2012.0318.
30. Li,S., Schlebusch,C. and 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.*, 10.1098/rspb.2014.1448.
31. Patin,E., Lopez,M., Grollemund,R., Verdu,P., Harmant,C., Quach,H., Laval,G., Perry,G.H., Barreiro,L.B., Froment,A., *et al.* (2017) Dispersals and genetic adaptation of Bantu-speaking populations in Africa and North America. *Science (80-.).*, 10.1126/science.aal1988.
32. Choudhury,A., Ramsay,M., Hazelhurst,S., Aron,S., Bardien,S., Botha,G., Chimusa,E.R., Christoffels,A., Gamieldien,J., Sefid-Dashti,M.J., *et al.* (2017) Whole-genome sequencing for an enhanced understanding of genetic variation among South Africans. *Nat. Commun.*, 10.1038/s41467-017-00663-9.
33. Pagani,L., Kivisild,T., Tarekegn,A., Ekong,R., Plaster,C., Gallego Romero,I., Ayub,Q., Mehdi,S.Q., Thomas,M.G., Luiselli,D., *et al.* (2012) Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. *Am. J. Hum. Genet.*, **91**, 83–96.
34. Brucato,N., Kusuma,P., Beaujard,P., Sudoyo,H., Cox,M.P. and Ricaut,F.-X. (2017) Genomic admixture tracks pulses of economic activity over 2,000 years in the Indian Ocean trading network. 10.1038/s41598-017-03204-y.
35. Austen,R.A. (1992) The mediterranean islamic slave trade out of africa: A tentative census. *Slavery Abol.*, **13**, 214–248.
36. Laso-Jadart,R., Harmant,C., Quach,H., Zidane,N., Tyler-Smith,C., Mehdi,Q., Ayub,Q., Quintana-Murci,L. and Patin,E. (2017) The Genetic Legacy of the Indian Ocean Slave Trade: Recent Admixture and Post-admixture Selection in the Makranis of Pakistan. *Am. J. Hum. Genet.*, **101**, 977–984.
37. Schroeder,H., Ávila-Arcos,M.C., Malaspinas,A.-S., Poznik,G.D., Sandoval-Velasco,M., Carpenter,M.L., Moreno-Mayar,J.V., Sikora,M., Johnson,P.L.F., Allentoft,M.E., *et al.* (2015) Genome-wide ancestry of 17th-century enslaved Africans from the Caribbean. *Proc. Natl. Acad. Sci.*, **112**, 201421784.
38. Fortes-Lima,C., Gessain,A., Ruiz-Linares,A., Bortolini,M.C., Migot-Nabias,F., Bellis,G., Moreno-Mayar,J.V., Restrepo,B.N., Rojas,W., Avendaño-Tamayo,E., *et al.* (2017) Genome-wide Ancestry and Demographic History of African-Descendant Maroon Communities from French Guiana and Suriname. *Am. J. Hum. Genet.*, **101**, 725–736.
39. Rotimi,C.N., Tekola-Ayele,F., Baker,J.L. and Shriner,D. (2016) The African diaspora: history, adaptation and health. *Curr. Opin. Genet. Dev.*, 10.1016/j.gde.2016.08.005.
40. Mathias,R.A., Taub,M.A., Gignoux,C.R., Fu,W., Musharoff,S., Connor,T.D.O., Vergara,C., Torgerson,D.G., Pino-yanes,M., Shringarpure,S.S., *et al.* (2016) A continuum of admixture in the Western Hemisphere revealed by the African Diaspora genome. *Nat. Commun.*, **7**, 2522.
41. Adhikari,K., Mendoza-Revilla,J., Chacón-Duque,J.C., Fuentes-Guajardo,M. and Ruiz-Linares,A. (2016) Admixture in Latin America. *Curr. Opin. Genet. Dev.*, **41**, 106–114.
42. Eltis,D. and Engerman,S.L. (1993) Fluctuations in sex and age ratios in the transatlantic slave trade, 1663-1864. *Econ. Hist. Rev.*, **46**, 308–323.
43. Conley,A.B., Rishishwar,L., Norris,E.T., Valderrama-Aguirre,A., Mariño-Ramírez,L., Medina-Rivas,M.A. and Jordan,I.K. (2017) A Comparative Analysis of Genetic Ancestry and Admixture in the Colombian Populations of Chocó and Medellín. *G3Genes, Genomes, Genet.*, **7**, 3435–3447.
44. Fuller,H. and Torres,J.B. (2018) Investigating the "Taíno" ancestry of the Jamaican Maroons: a new genetic (DNA), historical, and multidisciplinary analysis and case study of the Accompong Town Maroons. *Can. J. Lat. Am. Caribb. Stud.*, **43**, 47–78.

45. Pardo-Seco,J., Heinz,T., Taboada-Echalar,P., Martinón-Torres,F. and Salas,A. (2016) Mapping the genomic mosaic of two 'Afro-Bolivians' from the isolated Yungas valleys. *BMC Genomics*, **17**, 1–12.
46. Santangelo,R., González-Andrade,F., Børsting,C., Torroni,A., Pereira,V. and Morling,N. (2017) Analysis of ancestry informative markers in three main ethnic groups from Ecuador supports a trihybrid origin of Ecuadorians. *Forensic Sci. Int. Genet.*, **31**, 29–33.
47. Arauna,L.R., Mendoza-Revilla,J., Mas-Sandoval,A., Izaabel,H., Bekada,A., Benhamamouch,S., Fadhlaoui-Zid,K., Zalloua,P., Hellenthal,G. and Comas,D. (2017) Recent Historical Migrations Have Shaped the Gene Pool of Arabs and Berbers in North Africa. *Mol. Biol. Evol.*, **34**, 318–329.
48. Pagani,L., Schiffels,S., Gurdasani,D., Danecek,P., Scally,A., Chen,Y., Xue,Y., Haber,M., Ekong,R., Oljira,T., *et al.* (2015) Tracing the route of modern humans out of Africa by using 225 human genome sequences from Ethiopians and Egyptians. *Am. J. Hum. Genet.*, **96**, 986–91.
49. Font-Porterias,N., Solé-Morata,N., Serra-Vidal,G., Bekada,A., Fadhlaoui-Zid,K., Zalloua,P., Calafell,F. and Comas,D. (2018) The genetic landscape of Mediterranean North African populations through complete mtDNA sequences. *Ann. Hum. Biol.*, **45**, 98–104.
50. Soodyall,H. (2013) Lemba origins revisited : Tracing the ancestry of Y chromosomes in South African and Zimbabwean Lemba. *South African Med. J.*, **103**, 1009–1013.
51. Pierron,D., Heiske,M., Razafindrazaka,H., Rakoto,I., Rabetokotany,N., Ravololomanga,B., Rakotozafy,L.M.-A., Rakotomalala,M.M., Razafiarivony,M., Rasoarifetra,B., *et al.* (2017) Genomic landscape of human diversity across Madagascar. *Proc. Natl. Acad. Sci.*, **114**, E6498–E6506.
52. Chimusa,E.R., Daya,M., Möller,M., Ramesar,R., Henn,B.M., van Helden,P.D., Mulder,N.J. and Hoal,E.G. (2013) Determining ancestry proportions in complex admixture scenarios in South Africa using a novel proxy ancestry selection method. *PLoS One*, **8**, e73971.
53. Uren,C., M??ller,M., van Helden,P.D., Henn,B.M. and Hoal,E.G. (2017) Population structure and infectious disease risk in southern Africa. *Mol. Genet. Genomics*, **292**, 1–11.
54. Altshuler,D.M., Durbin,R.M., Abecasis,G.R., Bentley,D.R., Chakravarti,A., Clark,A.G., Donnelly,P., Eichler,E.E., Flicek,P., Gabriel,S.B., *et al.* (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*, 10.1038/nature11632.
55. Auton,A., Abecasis,G.R., Altshuler,D.M., Durbin,R.M., Bentley,D.R., Chakravarti,A., Clark,A.G., Donnelly,P., Eichler,E.E., Flicek,P., *et al.* (2015) A global reference for human genetic variation. *Nature*, 10.1038/nature15393.
56. Howie,B., Fuchsberger,C., Stephens,M., Marchini,J. and Abecasis,G.R. (2012) Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.*, 10.1038/ng.2354.
57. Visscher,P.M., Wray,N.R., Zhang,Q., Sklar,P., McCarthy,M.I., Brown,M.A. and Yang,J. (2017) 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am. J. Hum. Genet.*, 10.1016/j.ajhg.2017.06.005.
58. Zhang,P., Zhan,X., Rosenberg,N.A. and Z??llner,S. (2013) Genotype imputation reference panel selection using maximal phylogenetic diversity. *Genetics*, 10.1534/genetics.113.154591.
59. Roshyara,N.R., Horn,K., Kirsten,H., Ahnert,P. and Scholz,M. (2016) Comparing performance of modern genotype imputation methods in different ethnicities. *Sci. Rep.*, 10.1038/srep34386.
60. Chou,W.C., Zheng,H.F., Cheng,C.H., Yan,H., Wang,L., Han,F., Richards,J.B.,

Karasik,D., Kiel,D.P. and Hsu,Y.H. (2016) A combined reference panel from the 1000 Genomes and UK10K projects improved rare variant imputation in European and Chinese samples. *Sci. Rep.*, 10.1038/srep39313.
61. McCarthy,S., Das,S., Kretzschmar,W., Delaneau,O., Wood,A.R., Teumer,A., Kang,H.M., Fuchsberger,C., Danecek,P., Sharp,K., *et al.* (2016) A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.*, **48**, 1279–83.
62. Hindorff,L.A., Bonham,V.L., Brody,L.C., Ginoza,M.E.C., Hutter,C.M., Manolio,T.A. and Green,E.D. (2018) Prioritizing diversity in human genomics research. *Nat. Rev. Genet.*, 10.1038/nrg.2017.89.
63. H3Africa Consortium,T.H., Rotimi,C., Abayomi,A., Abimiku,A., Adabayeri,V.M., Adebamowo,C., Adebiyi,E., Ademola,A.D., Adeyemo,A., Adu,D., *et al.* (2014) Research capacity. Enabling the genomic revolution in Africa. *Science*, **344**, 1346–8.
64. Allison,A.C. (1954) Protection afforded by sickle-cell trait against subtertian malareal infection. *Br. Med. J.*, **1**, 290–4.
65. Pasvol,G., Weatherall,D.J. and Wilson,R.J.M. (1978) Cellular mechanism for the protective effect of haemoglobin S against P. falciparum malaria. *Nature*, **274**, 701–703.
66. Aidoo,M., Terlouw,D.J., Kolczak,M.S., McElroy,P.D., ter Kuile,F.O., Kariuki,S., Nahlen,B.L., Lal,A.A. and Udhayakumar,V. (2002) Protective effects of the sickle cell gene against malaria morbidity and mortality. *Lancet*, **359**, 1311–1312.
67. Bitoungui,V.J.N., Pule,G.D., Hanchard,N., Ngogang,J. and Wonkam,A. (2015) Beta-globin gene haplotypes among cameroonians and review of the global distribution: is there a case for a single sickle mutation origin in Africa? *OMICS*, **19**, 171–9.
68. Hamblin,M.T. and Di Rienzo,A. (2000) Detection of the Signature of Natural Selection in Humans: Evidence from the Duffy Blood Group Locus. *Am. J. Hum. Genet.*, **66**, 1669–1679.
69. Howes,R.E., Patil,A.P., Piel,F.B., Nyangiri,O.A., Kabaria,C.W., Gething,P.W., Zimmerman,P.A., Barnadas,C., Beall,C.M., Gebremedhin,A., *et al.* (2011) The global distribution of the Duffy blood group. *Nat. Commun.*, **2**, 266.
70. Leffler,E.M., Band,G., Busby,G.B.J., Kivinen,K., Le,Q.S., Clarke,G.M., Bojang,K.A., Conway,D.J., Jallow,M., Sisay-Joof,F., *et al.* (2017) Resistance to malaria through structural variation of red blood cell invasion receptors. *Science*, **356**, eaam6393.
71. Ko,W.-Y., Rajan,P., Gomez,F., Scheinfeldt,L., An,P., Winkler,C.A., Froment,A., Nyambo,T.B., Omar,S.A., Wambebe,C., *et al.* (2013) Identifying Darwinian selection acting on different human APOL1 variants among diverse African populations. *Am. J. Hum. Genet.*, **93**, 54–66.
72. Grossman,S.R., Shylakhter,I., Karlsson,E.K., Byrne,E.H., Morales,S., Frieden,G., Hostetter,E., Angelino,E., Garber,M., Zuk,O., *et al.* (2010) A Composite of Multiple Signals Distinguishes Causal Variants in Regions of Positive Selection. *Science (80-.).*, **327**, 883–886.
73. Vitti,J.J., Grossman,S.R. and Sabeti,P.C. (2013) Detecting Natural Selection in Genomic Data. *Annu. Rev. Genet*, **47**, 97–120.
74. Andersen,K.G., Shylakhter,I., Tabrizi,S., Grossman,S.R., Happi,C.T. and Sabeti,P.C. (2012) Genome-wide scans provide evidence for positive selection of genes implicated in Lassa fever. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, **367**, 868–77.
75. Ranciaro,A., Campbell,M.C., Hirbo,J.B., Ko,W.Y., Froment,A., Anagnostou,P., Kotze,M.J., Ibrahim,M., Nyambo,T., Omar,S.A., *et al.* (2014) Genetic origins of lactase persistence and the spread of pastoralism in africa. *Am. J. Hum. Genet.*, 10.1016/j.ajhg.2014.02.009.
76. Liebert,A., López,S., Jones,B.L., Montalva,N., Gerbault,P., Lau,W., Thomas,M.G., Bradman,N., Maniatis,N. and Swallow,D.M. (2017) World-wide distributions of lactase

persistence alleles and the complex effects of recombination and selection. *Hum. Genet.*, 10.1007/s00439-017-1847-y.
77. Jarvis,J.P., Scheinfeldt,L.B., Soi,S., Lambert,C., Omberg,L., Ferwerda,B., Froment,A., Bodo,J.M., Beggs,W., Hoffman,G., *et al.* (2012) Patterns of ancestry, signatures of natural selection, and genetic association with stature in Western African pygmies. *PLoS Genet.*, 10.1371/journal.pgen.1002641.
78. Perry,G.H., Foll,M., Grenier,J.-C., Patin,E., Nédélec,Y., Pacis,A., Barakatt,M., Gravel,S., Zhou,X., Nsobya,S.L., *et al.* (2014) Adaptive, convergent origins of the pygmy phenotype in African rainforest hunter-gatherers. *Proc. Natl. Acad. Sci.*, **111**, E3596–E3603.
79. Crawford,N.G., Kelly,D.E., Hansen,M.E.B., Beltrame,M.H., Fan,S., Bowman,S.L., Jewett,E., Ranciaro,A., Thompson,S., Lo,Y., *et al.* (2017) Loci associated with skin pigmentation identified in African populations. *Science (80-.).*, 10.1126/science.aan8433.
80. Sugden,L.A., Atkinson,E.G., Fischer,A.P., Rong,S., Henn,B.M. and Ramachandran,S. (2018) Localization of adaptive variants in human genomes using averaged one-dependence estimation. *Nat. Commun. 2018 91*, **9**, 703.
81. Busby,G., Christ,R., Band,G., Leffler,E., Le,Q.S., Rockett,K., Kwiatkowski,D. and Spencer,C. (2017) Inferring adaptive gene-flow in recent African history. *bioRxiv*, 10.1101/205252.
82. Basu Mallick,C., Iliescu,F.M., Möls,M., Hill,S., Tamang,R., Chaubey,G., Goto,R., Ho,S.Y.W., Gallego Romero,I., Crivellaro,F., *et al.* (2013) The light skin allele of SLC24A5 in South Asians and Europeans shares identity by descent. *PLoS Genet.*, **9**, e1003912.
83. Martin,A.R., Lin,M., Granka,J.M., Myrick,J.W., Liu,X., Sockell,A., Atkinson,E.G., Werely,C.J., Möller,M., Sandhu,M.S., *et al.* (2017) An Unexpectedly Complex Architecture for Skin Pigmentation in Africans. *Cell*, 10.1016/j.cell.2017.11.015.
84. Morris,A.G. (2017) Ancient DNA comes of age, but still has some teenage problems. *S. Afr. J. Sci.*, 10.17159/sajs.2017/a0232.

Legends to figures

Figure 1: Major early African population splits showing our understanding prior to and after the availability of whole genome sequence data and novel analysis approaches. The events of the past ~5000 years, prior to the Bantu expansion are not shown and therefore the African regions (South, Central, East and West) reflect the groups that predominated in these regions at ~5000 thousand years ago (kya). Both trees are routed to the most common recent ancestor (MCRA) and the estimated major splits are shown in kya. (A) shows our understanding prior to ~2016 when the MRCA was estimated to be ~150 kya and (B) following further analyses that place the MCRA at ~300 kya, with revised estimates of major splits shown in blue. The dotted line shows the recently proposed deep split of a western African ancestry population ~250 kya. ([#1]Skoglund et al. 2017; [#2]Schlebusch et al 2017; [#3]Hseih et al. 2016)

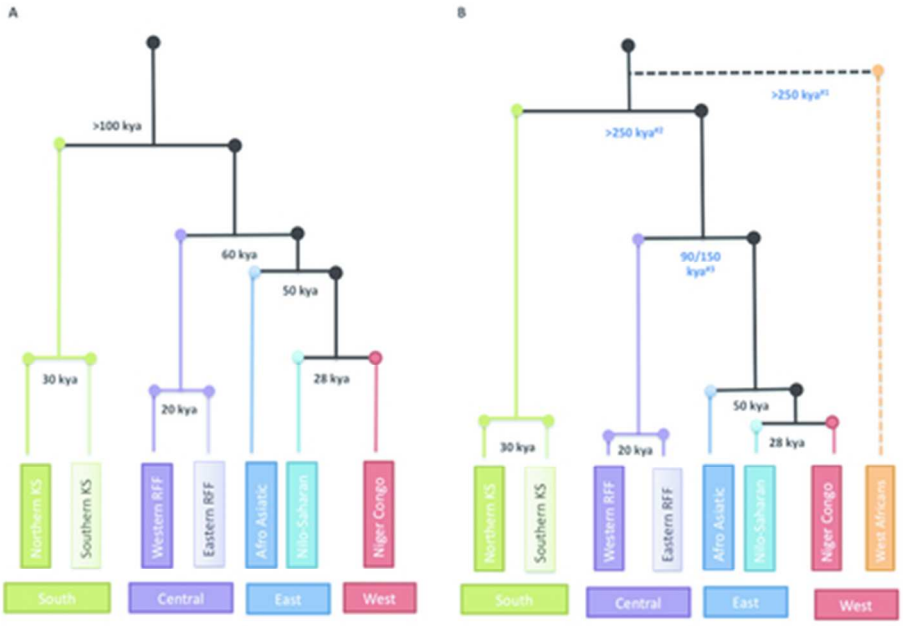
Figure 2: Hunter-gatherer (HG) and Eurasian admixture in African populations showing possible sources and timing of events. (A) shows HG admixture events while (B) shows Eurasian admixture events. Blue dotted arrows indicate the route of the Bantu expansion (reflecting the late split), while the colours of the stars and crosses correspond to the suggested timing of admixture events (kya) as shown on the scale at the bottom of the diagram. Populations are labelled as identified in the literature.

Table 1. Allele frequencies in AGVP and KGP African populations of variants previously detected to be under selection (west African in blue, southern African in green, east African in pink and non-African in yellow)

Gene	SNP ID	DA	GWD	MSL	ESN	YRI	ZUL	LWK	BAG	ETH	CEU	CHB
						Lactose tolerance						
MCM6	rs145946881	G	0.000	0.000	0.000	0.000	0.070	0.086	0.075	0.000	0.000	0.000
MCM6	rs41380347	C	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.000	0.000
MCM6	rs41525747	C	NA	NA	NA	NA	0.000	NA	0.000	0.050	NA	NA
MCM6	rs4954490	A	0.128	0.088	0.116	0.056	0.105	0.126	0.110	0.258	0.798	0.340
MCM6	rs4988235	A	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.058	0.737	0.000
MCM6	rs869051967	C	NA	NA	NA	NA	0.000	NA	0.000	0.008	NA	NA
						Malaria						
CD36	rs3211938	G	0.009	0.006	0.242	0.292	0.005	0.076	0.050	0.000	0.000	0.000
DARC	rs12075	G	0.000	0.000	0.000	0.000	0.070	0.000	0.000	0.113	0.429	0.932
DARC	rs2814778	C	1.000	1.000	1.000	0.995	0.800	1.000	1.000	0.738	0.000	0.000
DARC	rs3027011	G	0.080	0.123	0.081	0.102	0.185	0.121	0.185	0.058	0.000	0.000
DARC	rs55872368	T	0.925	0.882	0.929	0.898	0.730	0.879	0.825	0.700	0.131	0.000
DARC	rs7550207	C	0.080	0.123	0.081	0.097	0.190	0.121	0.185	0.125	0.197	0.000
DARC	rs863004	T	0.925	0.882	0.929	0.903	0.760	0.874	0.825	0.771	0.404	0.034
G6PD	rs1050828	T	0.040	0.070	0.160	0.210	NA	0.180	NA	NA	0.000	0.000
G6PD	rs1050829	C	0.360	0.280	0.350	0.380	NA	0.340	NA	NA	0.000	0.000
HBB (HbS)	rs334	A	0.115	0.123	0.121	0.139	0.005	0.101	0.070	0.000	0.000	0.000
HBB (HbC)	rs33930165	T	0.004	0.006	0.000	0.028	0.005	0.000	0.020	0.000	0.000	0.000
						Skin Pigmentation						
DDB1	rs11230664	T	0.208	0.229	0.455	0.310	0.195	0.253	0.255	0.296	0.995	0.971
HERC2	rs4932620	T	0.071	0.076	0.096	0.083	0.030	0.131	0.100	0.242	0.010	0.019
HERC2	rs6497271	G	0.549	0.559	0.722	0.699	0.475	0.712	0.410	0.467	0.015	0.029

KITLG	rs12821256	C	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.004	0.131	0.000
MFSD12	rs10424065	T	0.389	0.394	0.273	0.269	0.130	0.253	0.235	0.304	0.005	0.000
MFSD12	rs6510760	A	0.907	0.853	0.849	0.898	0.675	0.818	0.815	0.529	0.051	0.078
OCA2	rs1800404	T	0.150	0.082	0.136	0.069	0.145	0.086	0.045	0.258	0.823	0.393
OCA2	rs1800417	G	0.027	0.059	0.086	0.042	0.160	0.076	0.040	0.013	0.000	0.000
SLC24A5	rs1426654	A	0.075	0.088	0.025	0.014	0.075	0.076	0.025	0.371	1.000	0.029
SLC45A2	rs16891982	G	0.004	0.012	0.000	0.000	0.000	0.005	0.005	0.021	0.980	0.015
TMEM138	rs7948623	T	0.270	0.241	0.177	0.264	0.230	0.348	0.295	0.425	0.000	0.000
TYR	rs1042602	A	0.000	0.006	0.000	0.000	0.000	0.000	0.000	0.013	0.399	0.000
			Trypanosomiasis									
APOL1	rs60910145	G	0.243	0.123	0.495	0.375	0.120	0.056	0.055	0.000	0.000	0.000
APOL1	rs73885319	G	0.243	0.123	0.495	0.375	0.125	0.056	0.055	0.000	0.000	0.000
APOL1	rs143830837	-	0.195	0.182	0.116	0.079	NA	0.091	NA	NA	0.000	0.000

DA - derived allele
NA – missing data
Frequency of 0.000 – not observed in that population
Population codes: Esan in Nigeria (ESN); Yoruba in Ibadan, Nigeria (YRI); Gambian in Western Divisions in the Gambia (GWD); Mende in Sierra Leone (MSL); Luhya in Webuye, Kenya (LWK); Han Chinese in Bejing, China (CHB), Utah Residents with Northern and Western European Ancestry (CEU) from KGP Phase 3 and Baganda (BAG), Zulu (ZUL), Ethiopian (ETH) populations from AGVP. Both KGP Phase 3 and AGVP dataset have ~100 individuals per population.



40x30mm (300 x 300 DPI)

30x16mm (300 x 300 DPI)