# Human Population History from an East African Perspective: the Forgotten Land

Hisham M. Eldai and Hiba Babiker

## Introduction

East Africa extends over 5.9 million square kilometres, with diverse geography and climates (Mather 2003), which have been host to the earliest examples of human existence (Coppens 1977; Leakey 1969; Liu et al. 2006; McDermott et al. 1996). Due to human interactions, migration and environmental cues over the long history of East Africa, its populations are largely multi-ethnic. However, the intricate genetic legacies precipitated in East African populations by these events remain unexplored.

Evidence supporting the origin of our common ancestors in East Africa and its place as a hub to Out-of-Africa (OoA) dispersals includes fossil remains, archaeological finds and genetic records (Reyes-Centeno et al. 2014; Groucutt et al. 2015; Tishkoff et al. 2009) (figure 1). The Singa (Sudan) and Omo (Ethiopia) hominid remains (figure 2), dated at 130 kya (Pearson 2013) and 200 kya (Wood & Leakey 2011) show that humans settled in East Africa long before the record of their organization into small monarchies and chiefdoms (which only began around 5000 BC) (Stringer 2016; Loeffler 2016; McDermott et al. 1996; McDougall et al. 2005). Remarkably, the Singa hominid is a rare specimen of a population that existed before the appearance of anatomically fully modern *Homo sapiens* (McDermott et al. 1996). Moreover, the dozens of 13-14 kya human skeletons recovered from Jebel Sahaba in northern Sudan "Upper Nubia" are the oldest evidence of early armed attacks ever known, making Jebel Sahaba the oldest battleground and one of the oldest cemeteries (Kelly 2005).

Archaeologically, distinct lithic industries found in Nubian complexes in Arabia, Egypt and the Levant indicate flowing cultural exchange and OoA dispersals along northern and eastern routes (Crassard & Hilbert 2013; Usik et al. 2013; Pagani et al. 2015). Genetic data infer that East Africa sustained a large effective population size (i.e. the actual number of successfully breeding individuals in a population) and had the greatest level of population substructure anywhere in the world. It is where deep branching of Y-chromosome and mtDNA haplogroups originated and dispersed to the rest of the world (Underhill et al. 2000; Semino et al. 2002; Gomes et al. 2010; Cruciani et al. 2007; Rosa & Brehem 2011; Soares et al. 2012; Gonder et al. 2007). However, the genomic record of this anthropologically-crucial region features sparse and low-coverage data for a small subset of geographically discontinuous populations and offers only limited insights for verifying the level of diversity implied by archaeological finds (reviewed in the next section).

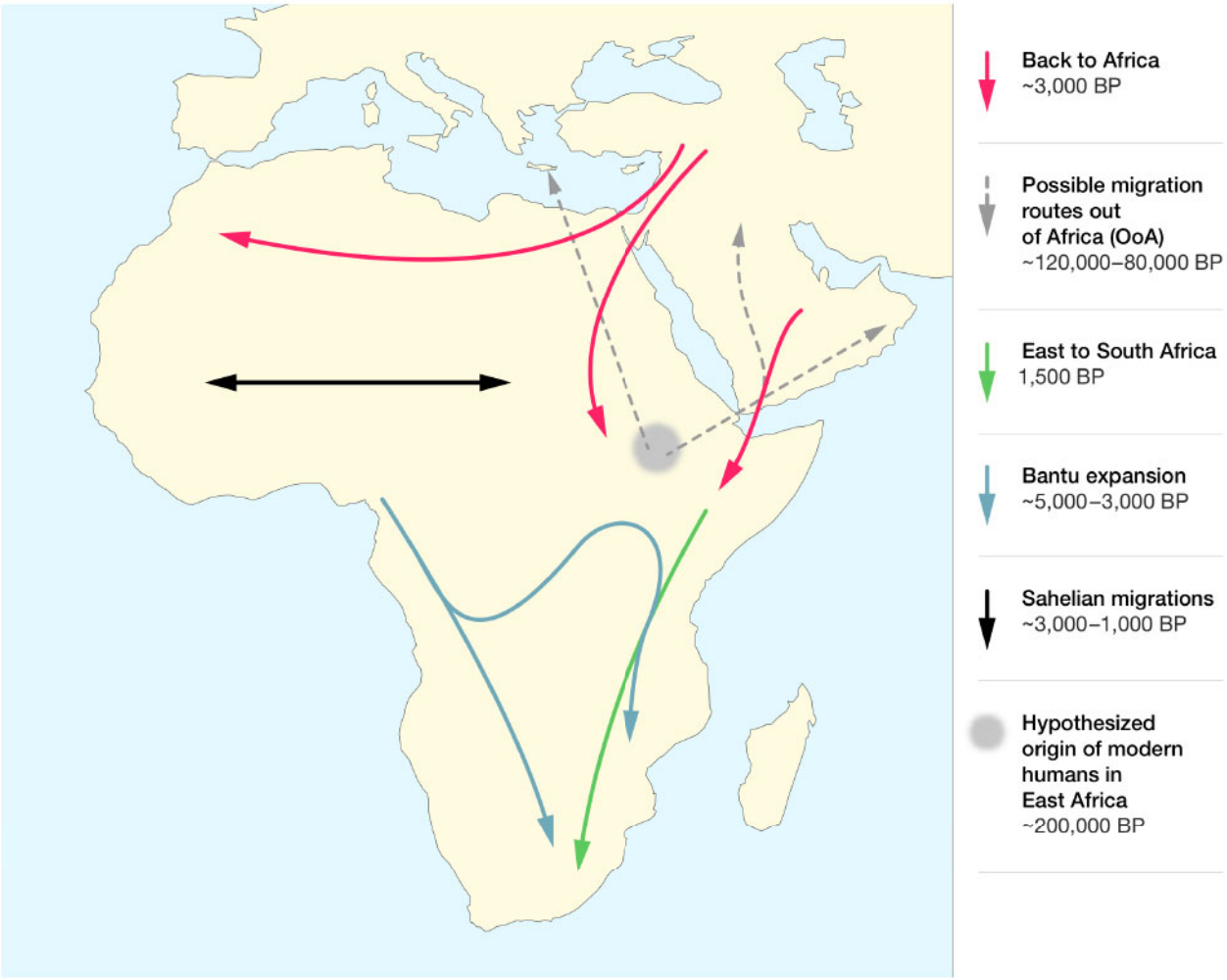



**Figure 1: map of Africa.** The map is indicative of the area in East Africa where anatomically modern humans are hypothesised to have originated; possible migration routes Out of Africa (OoA); the Bantu Expansion from Central West Africa to South and East Africa; the bidirectional migrations on the Sahel region and the migration back to Africa. Note that these events took place at different points of time (Rose et al. 2011; Walter et al. 2000; Luis et al. 2004; Underhill & Kivisild 2007; Musilová et al. 2011; Kopytoff 1987; Plaza et al. 2004; McDougall et al. 2005; Llorente et al. 2015; Schlebusch & Jakobsson 2018)

The strength of inferences made from a given population depends on the degree of its representation (sample size and genomic coverage) in a dataset, which in turn determines its utility in large-scale comparative population studies (Groucutt et al. 2015; Pemberton et al. 2013; Reyes-Centeno et al. 2016; Gonder et al. 2007). In the case of East Africa, the data limitation has considerable outcomes for our understanding of human history. First, as a result of low genomic coverage, the regional demographic relationships—as explained by simple models of gene flow—do not fully capture its population dynamics (i.e. migration, admixture, structure, and selection) (Pickrell & Reich 2014). Second, overlooking the impact of gene flow and concurrent population interactions from East Africa impedes a full understanding of the early events in human history and abates current inferences on relationships among the world's populations (Gonder et al. 2007; Groucutt et al. 2015; Reyes-Centeno et al. 2016). Therefore, the lack of systematic genomic studies on East African populations and the limitations in current data warrant establishing a representative East African dataset to enhance the region's data record (in sample size, historical and geographical continuity and genomic coverage) and

improve our current ability to evaluate the position of East Africa within the framework of human evolution and migration.

Uncovering African linguistic diversity and genetic relationships between populations of different language families can be advanced through interdisciplinary research. For example, integrating linguistic data with genetic data from modern and ancient populations will likely contribute to decoding long-standing questions in African linguistics, cultural and genetic diversity; and provide hints about ancient and recent population histories (Creanza et al. 2015; Pakendorf 2014; Günther et al. 2015).

This article revisits the available East African genetic records and discusses the importance of the region to our understanding of human history. We find that populations of East Africa are underrepresented in genetic studies, despite their potential to uncover the origin of our species in Africa and the possible routes of dispersal out of Africa. The evident need is for increasing the current data record by systematically sampling contemporary DNA and aDNA from the Sudan region given its archaeological, linguistic and anthropological significance. Having such data will maximise the role of simulation modelling and statistical approaches in providing precise inferences about population structure and divergence in Africa.

## The Genomic Record of East Africa and its Limitations

Variation analysis of genetic uniparental markers—mitochondrial DNA (mtDNA) and nonrecombining Y-chromosomal (NRY) markers—in contemporary populations informed early reconstruction of human ancestry. The analyses strongly support a recent African origin of *Homo sapiens*, in terms of Africa being the source of the deepest lineages and harbouring the greatest diversity (Grzybowski & Rogalla 2011; Patterson et al. 2012; Price et al. 2009; Pugach et al. 2011; Sankararaman et al. 2008). However, due to the low genomic coverage of uniparental markers, they cannot reveal complex details of population dynamics. They also suffer from inherent biases in sampling and in their capacity to illustrate gene flow. For example, uniparental markers can reveal ancestral patterns of gene flow, but not bidirectional gene flow relationships (Congiu et al. 2012; Veeramah & Hammer 2014). In contrast, high-resolution whole-genome sequencing (WGS) data would enable answering, more rigorously, questions in large-scale population dynamics studies (Malaspinas et al. 2016; Mallick et al. 2016; Pagani et al. 2016; Veeramah & Hammer 2014). For example, WGS data are amenable to investigating the bidirectional gene flow between diverging populations. Furthermore, aDNA analysis is promising in tracking human diversity and investigating population dynamics, as functions of time and historical events, through comparing ancient genomes to regionally modern ones (Pickrell & Reich 2014).

We surveyed the representation of East African populations in 31 publicly available genomic datasets (table S1). These studies had diverse objectives, but overall the current representation of East African populations does not reflect the complexity of East African population structure and the genomic record for the region is deficient in relation to that of the world. Moreover, notable issues arise in the few datasets that included East Africans. Examples of these are listed in table 1. The Allele Frequency Database (ALFRED) of anthropologically defined populations has 21, 451 and 724 populations from East Africa, continental Africa and the world (row 1, table 1) (Osier et al. 2002). But these alleles do not necessarily overlap, hence they are not comparable to one another since they were sourced from unrelated studies using disparate sets of genes. Another dataset that has been central to studying human diversity is the Human Genome Diversity Panel (HGDP) which genotyped 1,043 individuals from 52 populations. However, only one of the seven African populations in HGDP is from

East Africa (notably, 19 Bantu Kenyans) (Cavalli-Sforza 2005). Therefore, studies incorporating HGDP will either echo such underrepresentation (table 1, row 2); or in attempting to avoid it, they synthesise a dataset sourced from HGDP and several others by pooling only the overlapping markers common to all source datasets and discarding the non-overlapping ones; effectively reducing the genomic coverage of the attendant dataset (table 1, row 3). In the case of East Africa, one such source dataset is from a study by Tishkoff et al. (table 1, row 4). So, while the dataset in the Tishkoff et al. study reports 1327 genomic markers (Tishkoff et al. 2009), the synthesised dataset allows comparing only 645 microsatellites between populations.

**Table 1: sampling of East Africa in publicly available datasets.** Datasets were considered if they are actively maintained and their metadata lists a population's geographical location and/or ethnic affiliation. Supplementary table 1 provides further details for all studies surveyed in this review

| Row | Study | East Africa to Africa (%), East Africa to World (%) | Country Counts (Sudan, Ethiopia, Somalia, Uganda, Kenya, Tanzania, Djibouti, Eritrea) | Publicly available data |
|---|---|---|---|---|
| 1 | The Allele Frequency Database (ALFRED) (Osier et al. 2002) a. | 13, 1.4 | 17, 20, 5, 0, 10, 4, 1, 1 | Yes |
| 2 | Worldwide human relationships inferred from genome-wide patterns of variation(Osier et al. 2002; Li et al. 2008) | 14, 2 | 0, 0, 0, 0, 19, 0, 0, 0 | NA |
| 3 | Population structure in a comprehensive genomic data set on human microsatellite variation (Pemberton et al. 2013) b. | 43, 18 | 102, 55, 0, 0, 431, 499, 0, 0 | Yes |
| 4 | The genetic structure and history of Africans and African Americans (Tishkoff et al. 2009) bc. | 42, N.A | 102, 55, 0, 0, 431, 499, 0, 0 | Yes |
| 5 | hmtDB: A database of complete mtDNA genomes and coding regions (Rubino et al. 2011) d. | 9, 1 | 44, 118, 65, 3, 62, 49, 0, 0 | Yes |
| 6 | YHRD: Y Chromosome Haplotype Reference Database for multi-locus haplotypes and Y-SNP profile (Willuweit & Roewer 2015) d. | 57, 0.70 | 64, 278, 201, 118, 272, 162, 54, 161 | Yes |

| | | | | |
|---|---|---|---|---|
| 7 | An integrated map of genetic variation from 1,092 human genomes (1000 Genomes Project Consortium et al. 2012) e. | 49.1 | 0, 0, 0, 0, 98, 0, 0, 0 | Yes |
| 8 | Northeast African genomic variation shaped by the continuity of indigenous groups and Eurasian migrations (Hollfelder et al. 2017) c. | N.A | 221, 0, 0, 0, 0, 0, 0, 0 | Yes |
| 9 | The African Genome Variation Project shapes medical genetics in Africa (Gurdasani et al. 2015) e. | 68, N.A | 0, 120, 0, 100, 0, 0, 0, 0 | Yes |
| a=Allele frequency. b=Microsatellites. c=SNPs. d=Indels. e=Parental DNA (mtDNA, Y-Chromosome). f=Whole-Genome Sequencing. | | | | |

The region is also underrepresented in WGS and aDNA studies. Out of almost 300 million people living in East African countries, only 120 Ethiopian and 100 Bantu Ugandan genomes were sequenced in the African Genome Variation Project (AGVP) (Gurdasani et al. 2015) (table 1, row 9), another set of 125 Ethiopians was used in a study that aimed to resolve the OoA migration routes (Pagani et al. 2015). The Simons Genome Diversity Project sequenced 300 genomes from 142 diverse populations, including only three Dinka individuals from South Sudan and two individuals each of Maasai, Kenyan Bantu, Luo, Luhya and one Somali from Kenya (Mallick et al. 2016) (Supplementary table 1). In addition, 15 hunter-gatherers (including 10 Tanzanians) were sequenced to study evolutionary variation and adaptation in these communities (Lachance et al. 2012) (Supplementary table 1).

Globally, several large-scale WGS initiatives have sequenced hundreds of genomes (e.g., the >2500 genomes of the Icelandic population project, and the recently published Eurasian dataset) (Gudbjartsson et al. 2015; Kaiser 2015; Pagani et al. 2016). Moreover, Europe presented a more favourable place for harvesting aDNA from human remains than Africa, where DNA decay, accelerated by harsh climate, hindered the recovery of sufficient amounts for sequencing (Fu et al. 2016; Lazaridis et al. 2014; Richards et al. 2016; Günther et al. 2015; Haak et al. 2015). In addition, while interest in aDNA collection from Eurasia is gaining momentum (de Barros Damgaard et al. 2018), the trend does not reverberate for Africa in general and East Africa in particular.

### The Genetic Substructure in East Africa

Patterns of genetic diversity in East Africa were influenced by the complex histories of its populations, their geographical locations, and linguistic and cultural diversity (Tishkoff et al. 2009).

Analysis of contemporary populations from Africa based on uniparental and autosomal markers shows evidence of substructure between hunter-gatherers and agriculturalists (Quintana-Murci et al. 2008), populations within and between language families (Gomes et al. 2015; Pagani et al. 2012; Tishkoff et al. 2009) and populations with geographical boundaries (Hollfelder et al. 2017; Tishkoff et al. 2009).

### *Contemporary Studies in East Africa*

Pointing to population substructure and continuity; a recent study has shown that the Nilotic populations of North-East Africa, in particular, those from South Sudan, descended from ancestral populations that originated in East Africa and who were distinct from populations that lived in Ethiopia > 4000 years ago (Hollfelder et al. 2017). Further, the Nilo-Saharan genetic component observed in Western Sudanese populations of the Nuba Mountains suggests population contact in the region south of the Nile River (Dobon et al. 2015). The great Bantu Expansion (~5000-3000 BP, figure 1) that reached East and South Africa has influenced the patterns of genetic variations and linguistic diversity by introducing genetic diversity and replacing languages in the local populations (Patin et al. 2017; Grollemund et al. 2015). However, genetic traces of the Bantu Expansion were not evident in the Sahelian region of South Sudan despite evidence of Bantu languages being spoken in the region (Hammarström et al. n.d.; Hollfelder et al. 2017). These lines of evidence render the Bantu Expansion (routes and times of dispersals) open to deeper investigations.

Genetic research findings emerging from the region point to a complex and dynamic history and population structure across the Sahel and along the Nile Valley (Dobon et al. 2015; Busby et al. 2016; Pagani et al. 2012; Hollfelder et al. 2017). Physical traits such as skin pigmentation show high variability of skin colour in East Africans (Ethiopia, Tanzania, and Botswana) (Crawford et al. 2017). These findings highlight the need to cover more populations in East Africa to increase our understanding of the human population history and evolution of cultural and phenotypic traits.

### *Ancient DNA Era in East Africa: How Far are We?*

Data from whole-genome sequencing (WGS) studies over the past decade have progressively revealed partial aspects of human population history, yet big questions remain open and in need of answers. Lack of research funds and limited logistics in Africa contribute to slowing the pace of research in the region (Kumwenda et al. 2017) relative to Europe where an impressive increase in human population studies is seen for both modern and ancient populations remains from as far back as the pre-Neolithic periods (Hofmanová et al. 2016; Haak et al. 2010). Another challenge is socio-political, where the archaeological wealth of the region (including cemeteries and artefacts) is at risk of erosion under ambitious modernisation projects such as the Meroë dam in Sudan (Emberling 2009). This, in particular, implies that the historical continuity of archaeological sites can never be captured and fully understood despite best efforts by archaeologists to salvage and record them given the limits on excavation seasons; unless a concerted effort to sample aDNA from the cemeteries is carried out in conjunction with the archaeological salvation projects.

Recent advances in optimised aDNA recovery techniques can increase the yield of aDNA extraction from skeletal remains (e.g. Petrous bone) (Pinhasi et al. 2015). Ancient DNA studies have been successful not only in sequencing hominin ancient remains (Neanderthal and Denisovans) (Prüfer et al. 2014; Meyer et al. 2012) but also naturally preserved human remains (Iceman, Saqqaq) (Keller et al. 2012; Rasmussen et al. 2010) and chemically treated ones (Egyptian Mummies) (Schuenemann et al. 2017). The aDNA studies from different geographical locations in Africa reflect how the aDNA technology is being adopted to study the most challenging DNA material (Schuenemann et al. 2017; Schlebusch et al. 2017; van de Loosdrecht et al. 2018; Fregel et al. 2018). These factors, together with the overall reduction in the cost of DNA sequencing, promise to realise the target of including more

underrepresented East African populations in sequencing projects and thus smoothing the geographical bias in sampling DNA from East Africa.

To demonstrate the sort of interesting questions that can arise from studying East African aDNA, consider these two examples. The first, aDNA genome from Ethiopia revealed population continuity in Ethiopia between ancient Mota's genome and contemporary Ethiopian highlanders known as the Ari, who also showed evidence of a Eurasian backflow into the region in the last 3 Kya (Llorente et al. 2015). The back-migration waves from Eurasia and the extent of their traces left in eastern Africa, and further west, remain incompletely surveyed (Llorente et al. 2015; Gurdasani et al. 2015; Haber et al. 2016).

A recent aDNA study of human remains found in eastern and southern Africa highlighted the deep and related ancestry of the East African Hadza to an ancient East African lineage and that they share less ancestry with Bantu populations occupying the same region (Skoglund et al. 2017). Interestingly, the study also pointed to population contact and gene-flow between East and South Africa predating the spread of farming and herding.

Understanding the complex demographic events and migrations in Africa will be a reference to studies of African populations and populations of African descent. In our opinion, aDNA skeletal remains from different sites in Sudan should be considered for new research avenues in Africa.

### Importance of East Africa in Human Settlement and Migration

The East African region encompasses Sudan, South Sudan, Eritrea, and Ethiopia. It is an outstanding place to study human diversity. The first humans transited the region in their exodus out of Africa, and many have settled there to form urban centres along the Nile River. The river is thought to have acted as a natural conduit/barrier of gene flow with different ethnicities on different sides of the Blue and White Niles and the several tributaries thereabout (Krings et al. 1999).

The area witnessed the rise and dwindling of ancient kingdoms spanning the East African region. The country of Sudan, specifically, has rich ancient cemeteries the study of which can widen our understanding of population history within and beyond Africa by providing new genetically-informed insights into the interplay between archaeology, anthropology, and linguistics in the region. As the region has been supporting human migration and settlement since antiquity, both modern and aDNA data are needed to better assign population relationships and trace migration turnovers to prehistoric depths.

Inarguably using accurate data to build demographic models of populations truly representative of East Africa will lead to more realistic and comprehensive conclusions about human evolution and migration. This is in contrast with estimating divergence times based on sampled populations that do not best represent those where historically important divergences occurred. For instance, estimating Yoruba-Eurasian divergence time (given the West African location of the Yoruba) may actually indicate an inter-African divergence and complicates estimation by divergence models (Gutenkunst et al. 2009).

On the other hand, the location of the Sudan to the East and at the nexus of migration routes made it a bottleneck and a corridor to settlement and OoA dispersal of anatomically modern humans, their crafts, crops, and animals (Winchell et al. 2017; Beldados et al. 2018; Krings et al. 1999; Gifford-Gonzalez 2017; Kimura et al. 2011). One hallmark of the diversity in this region is population continuity and turnover as evidenced by the succession of overlapping kingdoms that existed there (figure 2), each

leaving an abundance of cemeteries with burying styles reflective of their time periods (Chłodnicki et al. 2010; Chłodnicki et al. 2005; Chłodnicki & Żurawski 2004; Buzon 2014). At times, one location could even host mixed burial styles indicating the coexistence of populations or turnover of burial practice. A great example of this is in Sai Island where different styles of cemeteries can be seen within an area of only ~50 km$^2$, hailing from different periods of times (Neolithic, Napatan, Kerman, Meroitic, Kushitic, Christian, and Islamic) (Van Peer et al. 2003; Murail et al. 2004; Budka 2017; Eerkens et al. 2018; Francigny 2017). Figure 2 and table 2 list a few examples of such cemeteries to indicate that patterns of continuity are likely well spread over the Sudan region.

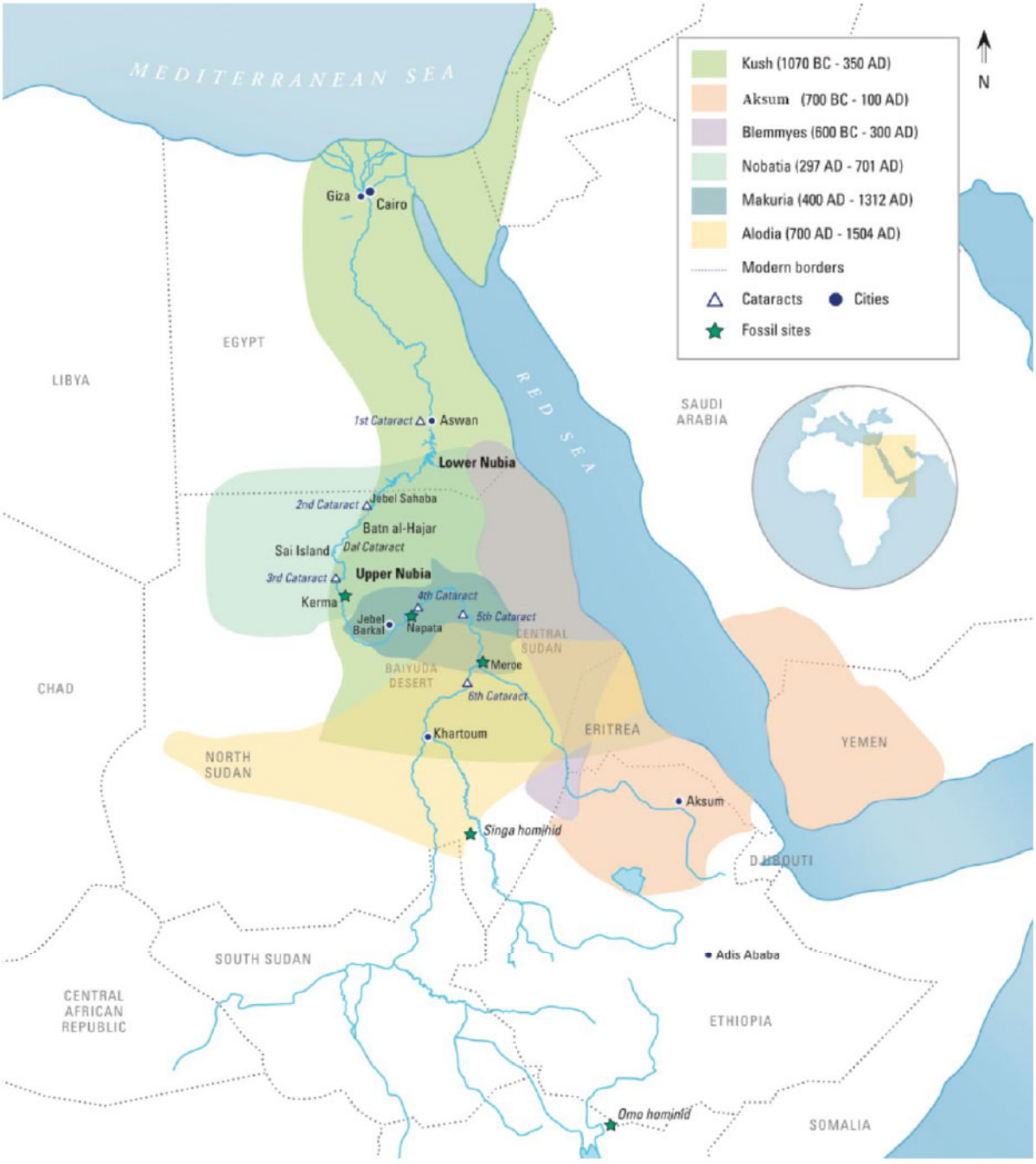



**Figure 2: Map of Ancient East African Kingdoms and sites of archaic fossil remains in Sudan and Ethiopia.** The region of East Africa witnessed a turnover of successive and contemporary kingdoms throughout history as evidenced by the richness of archaeological sites; making it an interesting place to study population structure and

continuity. The kingdoms of Kerma (2450 BC-1500 BC) and Napata (founded in 1400 BC) were riverine urban centres and powerful city-states. Together with Meroë, they were precursors to (and consecutive capitals of) the Kingdom of Kush. Kush expanded into Egypt, establishing the 25th Dynasty (760 BC–650 BC). It interacted commercially and militarily with the Assyrians in the Levant and the Phoenician coast (Kahn 2006). Aksum (founded in 400 BC) was a city-state and precursor to the Christian Aksum kingdom, which later infringed into Kush. Aksum and Kush were succeeded by Nobatia and Makuria, which eventually expanded over Nobatia

**Table 2: List of archaeological sites/cemeteries excavated in Sudan.** Sites are from different time periods spanning pre-Kerma, Kerma, Napata, Meroitic and medieval times. More details are shown in figures 2 & 3

| Archaeological Site/Cemetery | Cemeteries Spanning Periods | References |
|---|---|---|
| Jebel Sahaba | Pre-Kerma 14000–11000 | (Antoine et al. 2013) |
| Tombos | Napatan Nubian 750–660 BC | (Buzon 2014) |
| Sedeinga | Pre-Kerma, Meroitic and Napatan | (Rilly and Francigny 2011) |
| Berber Meroitic Cemetery | Meroitic Period (300 BC–350 AD) | (Rilly and Francigny 2011; Bashir and David 2015) |
| Hagar El-Beida 1 | Kerma, Napatan and Late/Post Meroitic (Christian) | (Chłodnicki et al. 2005) |
| ElSadda | Neolithic, Kerma, Late Meroitic, post-Meroitic, Old Kushitic, Christian | (Osypiński 2010) |

The region's inhabitants both generated and absorbed genetic and demographic flows. For example, the movement of ancient mtDNA lineages and linguistic diversity seen in other East African countries indicate southbound migration events originating in Sudan (Gonder et al. 2007). While some of these lineages are almost regionally exclusive, others (e.g. the M1 mtDNA haplogroup) have reached as far as Oceania (Gonder et al. 2007). This lineage was also suggested to have participated in shaping some of the West African mtDNA diversity coinciding with Kushitic migrations westward (Rosa et al. 2004). Unfortunately, mtDNA studies are too few to be considered comprehensive in representing the ancient and recent history of Sudan (Fox & Lalueza Fox 1997; Krings et al. 1999).

From a paternal lineage perspective, Y-chromosome studies point to interesting patterns of gene flow in Sudan and South Sudan that coincide with geographical affinities, languages, and history of the studied populations (Scheinfeldt et al. 2010). The discovery of E-M78* Y haplogroup in southeast Europe is an example of very ancient migratory events (~8500 BC) originating from Sudan (Battaglia et al. 2009; Cruciani et al. 2007). Other local examples include ancient population interactions between Nilo-Saharan and Afro-Asiatic speakers in the region and the possible origin of the Nilo-Saharan language family in eastern Sudan followed by westward and southward migrations to central Africa and southern Sudan (Cavalli-Sforza et al. 1994; Poloni et al. 2009). Studies on other Sahelian African populations reported contrasting patterns between Y-chromosome and mtDNA genome variations, which were shown to be a result of sex-biased migrations and gene flow (Wood et al. 2005; Barbieri et al. 2012; Cruciani et al. 2011). Hence extended investigation of Sahelian populations will uncover the population demographic history across the Sahel and will reveal the various patterns of migration and gene flow.

Evidence for under-sampling in eastern Africa, in particular, from Sudan has been revealed by a recent study where samples from different populations were examined using genome-wide data analysis pointing to population substructure and continuity in the region that had not been reported before (Hollfelder et al. 2017). Such findings highlight a much needed yet tardy progress of research on the

structure of regional indigenous populations such as the admixed modern Nubians and the relatively isolated Nilotes (Hollfelder et al. 2017).

Within the context of human population history, the population structure in the Sudan region lends itself to studying local and global population dynamics, both ancient and recent. The Sudanese populations can thus appropriately represent East Africa in large scale genomic studies. Therefore, an apparent approach to further enhance the current repertoire of genetic data is to systematically sample aDNA and modern DNA from Sudanese cemeteries and contemporary populations and to make that accessible to the scientific community.

### *The East African Linguistic Diversity Explained through Genetics: aDNA and Modern Data*

Linguistic diversity in Africa had been classically categorised into four language families: Afro-Asiatic, Niger-Congo, Nilo-Saharan, and Khoisan (Greenberg 1963). Despite linguists' recent criticism of the accuracy of this classification of African languages (Childs 2012; Güldemann 2016), it is possible to uncover population histories and substructure conforming to this linguistic classification (Pagani et al. 2012; Tishkoff et al. 2009). It is also worth noting that the co-evolution of genes and languages is a complex process that impacts the interpretation of phylogenies to reflect population histories (Pakendorf 2014).

The relationship between genes, phenotypic diversity, and languages is shaped by geography and the demographic histories of the populations (e.g. in East Africa: population size changes and admixture events with populations of both the Sahelian region and Eurasia (Pagani et al. 2012)). Within Africa, different correlations between languages, genetics, and geography are discernible between and within the language families (Krings et al. 1999; Tishkoff et al. 2009).

In the context of the Sudan region, this relationship presents an untapped resource of information pending systematic exploration. The evidence of the linguistic diversity in East Africa pinpoints to the location of Sudan in the overall continental linguistic diversity (Rosa et al. 2004; Poloni et al. 2009; Scheinfeldt et al. 2010; Barbieri et al. 2014; Gonder et al. 2007). Moreover, there are more than 137 languages spoken in Sudan (Hammarström et al. 2018), yet most populations speaking these languages have no representation in genetic datasets whatsoever since most of the previous studies were based on only a few numbers of markers and/or uniparental markers (Krings et al. 1999; Dobon et al. 2015).

It follows, therefore, that it remains inconclusive whether we will see a consistent matching between linguistic boundaries and genetic structure in the region pending integrating evidence from 01) aDNA (to study populations representing the ancient kingdoms in Sudan so as to set concrete linguistic evidence informing population continuity/displacement and past population contacts along the Nile Valley); 02) population-level WGS (to infer population divergences, assign population relationships and historical contacts, and date these events in a similar way to Indo-European studies (Haak et al. 2015); and 03) improved uniparental datasets that sample from more populations in the region. For this quest, Y-chromosomal variation can serve as a linguistic analysis tool provided that good quality chromosomal databases exist (St. Clair 2016).

***Systematic adoption of genomic anthropology to decode human population history***

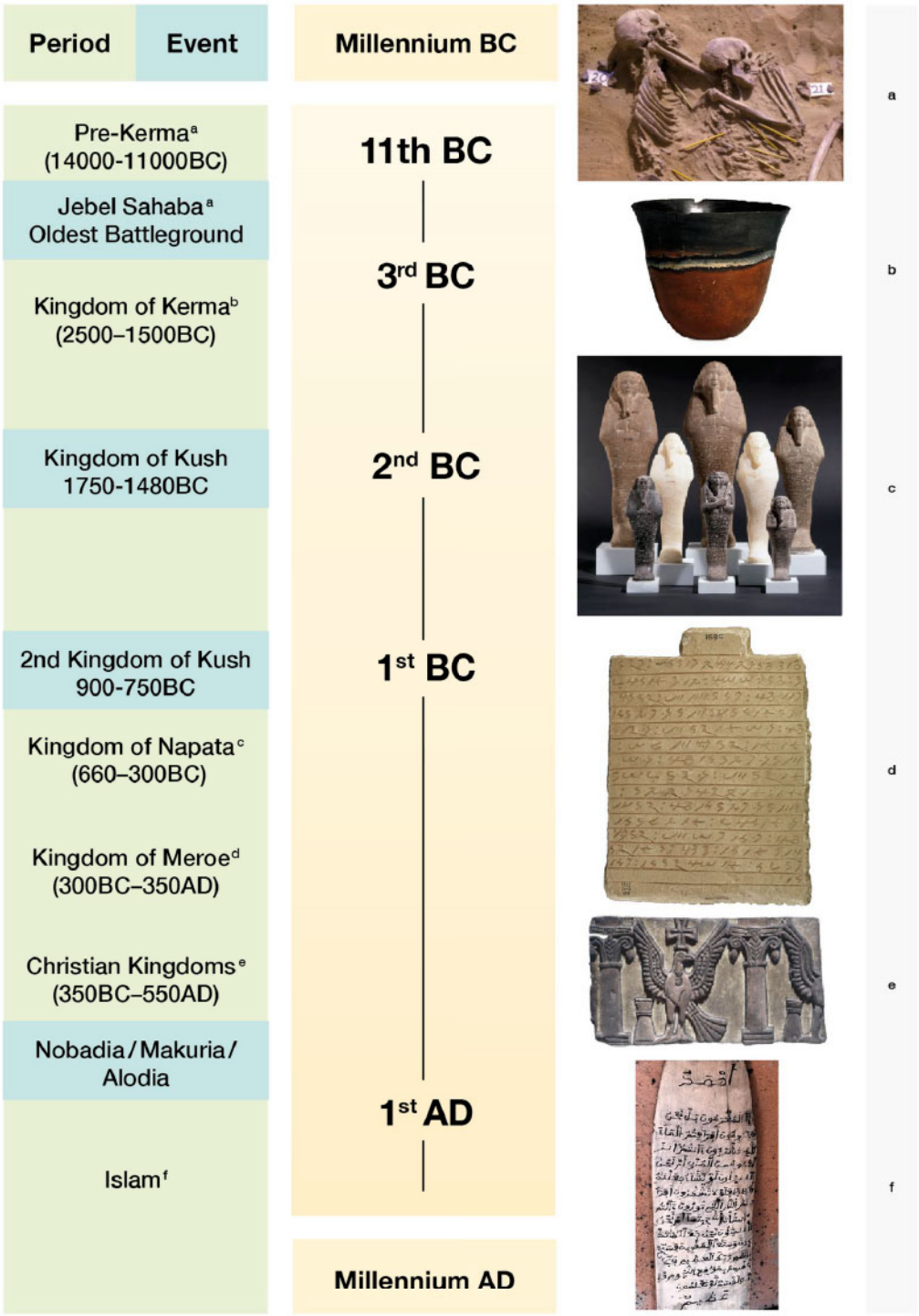
The inadequate inclusion of East African populations in population genomic studies minimises their contributions to the story of human history and to the field of genomic anthropology. Although there is a need for generating high quality East African whole genomes; the priority should be given to systematic sampling to avoid the shortcomings inherent in the current datasets. Specifically, a quantitative enhancement to this record, unless reflective of the diversity in the regional ethnic and linguistic repertoire and unless accounting for population continuity, is insufficient.

A point of emphasis is thus: sampling needs to be anthropologically informed such that the generated datasets would serve to expand our ability to ask more concrete questions regarding, for examples, untangling the relationships between East African populations and the World; solving the gaps in our understanding of human evolutionary history; retracing the migration routes that traversed the region in the OoA exodus and testing population continuity through serial sampling from several periods (figure 3) (Silva et al. 2017).

Comparative population analysis contributes to our knowledge about human genetic diversity and enhances the record of relationships between world populations in the contexts of understanding human migration, settlement, admixture and divergence times. This will allow the re-evaluation of migration models of human dispersal after *plugging-in* the East African component and will reveal the region's position in human evolution and migration through defining the contribution of the region to the population structures of the World. Additionally, it would also help in explaining how the regional population dynamics shaped OoA migration and were shaped by the backflow into Africa. By considering the genomic data in different ancestry models, it will be possible to comprehensively and systematically construct and test various divergence hypotheses to reveal deeper insights into the complex human demography; enabling researchers to systematically answer questions based on well-designed research priors (Hey 2004).

Regional population genetic studies have the potential to resolve/characterise inter-population relationships; patterns of selection and ancestral relationships; the evolutionary triggers that led to the selection of genetic traits along migration routes and bio-medically significant population-specific traits. Specific open questions in this regard include: exploring haplogroup diversity; describing the dynamics of male and female-mediated gene flow; calculating changes in effective population size (Schiffels & Durbin 2014); inferring demographic events and complex population history from genome-wide data (Gutenkunst et al. 2009) and the contributions of geography in shaping these relationships.

Based on all of the above, the question thus arises as to which region could harbour promising leads and hence can realistically represent East Africa in addressing the research opportunities presented thus far (in archaeology, linguistics, and anthropology) through genomics. To answer this question, we suggest 01) establishing an East African "flagship" dataset to bypass the limitations of the currently available datasets in coverage and resolution, and which contains sequence information of both modern and ancient DNAs, to facilitate exploring the broad themes discussed in this review; 02) and that populations from Sudan are well-suited representatives of East Africa in this endeavour.




*a-e © Trustees of the British Museum / *f credit Hiba Babiker

**Figure 3: Ancient East African Kingdoms and some representative skeletal remains/artefacts from each, reflecting the cultures of each period.** (a) Skeletons excavated in Palaeolithic Cemetery 117 of Jebel Sahaba in Sudanese Nubia. (b) Classical ware pottery beaker from Kerma period. (c) Napatan Red quartzite shabti (figurines) of Taharqo. (d) Meroitic Sandstone stela in the shape of an offering-table; fourteen incised lines of Meroitic text. (e) Christian period and the overlap between the Christian kingdoms. (f) Wood tablet used to teach Quran reading and writing until today. Images of artefacts were obtained from the British Museum's printed material. Representative cemeteries from each period, which can be explored for aDNA studies are detailed in table 2

## Conclusion and Future Directions

We have highlighted in this review that the inclusion of the Sudan region in the world genomic repertoire could lead to reconstructing human population history. Key to linking eastern Africa to the global picture is systematic DNA sampling from contemporary and ancient populations within the contexts of archaeology, linguistics, and anthropology. Availability of a flagship East African genomic dataset will promote the design and development of robust simulation modelling that best utilises the data both in spatial and temporal windows, which will allow making precise inferences about population history.

Revealing the true position of the region in the human migration and diversity story is hampered, however, by the limitations in the already sparse genomic data on populations of the Sudan region. Unfortunately, the currently available East African datasets will only yield diminishing insights when compared to what can be gleaned by including data from the latest and more comprehensive high throughput approaches. Echoing datasets across studies will not improve the situation. Hence a challenging task is compounded by the potential for information exhaust out of the current East African datasets. Therefore, a shift toward adopting large genomic datasets, systematically, is warranted. A primary motivation for such a transition is that large genomic datasets can lend themselves to statistical and computational modelling when studying complex demographic histories. This has the potential to refine our understanding of human history and local ancestry through tracing admixture signals and dating these events (Medina et al. 2018; Price et al. 2009; Pugach et al. 2011; Sankararaman et al. 2008); facilitate studying population continuity; reveal OoA and Sahelian migration routes and furthering the depth of questions asked. Indeed, incorporation of more complex genetic models will be crucial to revealing ancient human migrations and population divergences; hence it is relevant to answering longstanding questions about our species' history, evolution and dispersal inside and outside Africa (Scerri et al. 2018).

The few ancient DNA studies from southern, eastern and northern Africa were successful in reconstructing parts of the complex African population structure (Llorente et al. 2015; Schlebusch et al. 2017; Skoglund et al. 2017; van de Loosdrecht et al. 2018). These studies should serve as motivating examples for both archaeological and genetic research. Ancient DNA data spanning the Neolithic and later periods from Europe and the Near East are becoming available for comparative analysis and for tracing back past migration routes. The inclusion of east African aDNA will further clarify the origin of the story of human migration by revealing the times and routes of human dispersals/divergences out of Africa and the extent and regional depth of the migration waves and backflow into Africa.

Equally interesting is that aDNA sampled from the remains of humans that lived in the Sudan region at different points of time will be useful in comparing the structure of early regional inhabitants with those of present populations. Integrating insights from modern and ancient genomes is likely to clarify the environmental impact on migration and adaptation and allow testing such hypotheses as to whether the ancient population was gradually assimilated or replaced at different times in history (i.e., population continuity) (Silva et al. 2017). Moreover, analysis of modern and ancient human genomes specifically from the region will reveal information about the peopling of the region, the genetic affinities underlying ancient—modern population relationships and the impact of ancient populations on modern regional diversity. Therefore, archaeological work in the Sudan region should accommodate for aDNA collection from the field and from contemporary populations in the area. Also, remains found in the region and stored in worldwide museums need to be considered for aDNA sequencing studies (Fletcher et al. 2014).

Interdisciplinary research will play a significant role in the growing interest in human evolutionary history. As an example, subsistence patterns during the last 5000-3200 years in East Africa have been reflected in findings from studies in disciplines including archaeobotany and archaeozoology (Gifford-Gonzalez 1998; Gifford-Gonzalez & Hanotte 2011; Gifford-Gonzalez 2017), confirming that migration routes, population movements and dispersals can be implied from the faunal distributions and diversity. Analysis of ancient material such as animals and plants has proven its merit in connecting East Africa to the Indian Ocean and South Asia and provided a deep look into transitions of life subsistence activities in the area (Crowther et al. 2017).

Interdisciplinary research will allow undertaking more rigorous analysis and integrating additional layers of evidence from linguistics, archaeology, and anthropology through genomics, leading to a better understanding of the cultural and linguistic variation in light of modern and aDNA data (Stoneking & Krause 2011). Success in interdisciplinary research of this nature, however, requires concerted efforts and open collaboration among scientists in the region and the international community. This is crucial for asking questions, transferring knowledge, sharing data and expertise and above all; positioning the geno-anthropological tesserae from the Sudan region in their respective places within the mosaic of human diversity.

## Author Contributions and Acknowledgements

H. E. and H. B. designed the survey, collected and analyzed the data and wrote the manuscript. We thank Abbie Hantgan (Langage, Langues et Cultures d'Afrique Noire, Centre National de la Recherche Scientifique, Paris, France) for suggestions on the first draft; Mohamed Saad Abd Allah (National Corporation for Antiquities and Museums, Khartoum, Sudan) for helpful suggestions on figures and we thank Carina Schlebusch (Department of Organismal Biology, Uppsala University, Sweden) for critical review of the manuscript. This work was supported through funds of the Max Planck Society to the Department of Linguistic and Cultural Evolution (H.B.).

## References

Antoine, D. M., A. Zazzo and R. Friedman. 2013. Revisiting Jebel Sahaba: New Apatite Radiocarbon Dates for one of the Nile Valley's Earliest Cemeteries. *American Journal of Physical Anthropology* 150: 68–68.

Barbieri, C. et al. 2012. Contrasting Maternal and Paternal Histories in the Linguistic Context of Burkina Faso. *Molecular Biology and Evolution* 29 (4): 1213–1223.

Barbieri, C. et al. 2014. Migration and Interaction in a Contact Zone: mtDNA Variation among Bantu-Speakers in Southern Africa. *PloS One* 9(6): e99117.

de Barros Damgaard, P. et al. 2018. The First Horse Herders and the Impact of Early Bronze Age Steppe Expansions into Asia. *Science*: http://dx.doi.org/10.1126/science.aar7711.

Bashir M. S. and R. David. 2015. The Meroitic Cemetery at Berber. Recent Fieldwork and Discussion on Internal Chronology. *Sudan & Nubia* 19: 97–106.

Battaglia, V. et al. 2009. Y-Chromosomal Evidence of the Cultural Diffusion of Agriculture in Southeast Europe. *European Journal of Human Genetics* 17 (6): 820–830.

Beldados, A. et al. 2018. Evidence of Sorghum Cultivation and Possible Pearl Millet in the Second Millennium BC at Kassala, Eastern Sudan. In *Plants and People in the African Past: Progress in African Archaeobotany*, eds. A. M. Mercuri, A. C. D'Andrea, R. Fornaciari and A. Höhn, 503–528. New York: Springer.

Budka, J. 2017. Pyramid Cemetery SAC5, Sai Island, Northern Sudan. An Update Based on Fieldwork from 2015–2017. *Ägypten und Levante* 27: 107–130.

Busby, G. B. et al. 2016. Admixture into and within Sub-Saharan Africa. *eLife* 5: http://dx.doi.org/10.7554/eLife.15266.

Buzon, M.R. 2014. Tombos during the Napatan period (~750–660 BC): Exploring the Consequences of Sociopolitical Transitions in Ancient Nubia. *International Journal of Paleopathology* 7: 1–7.

Cavalli-Sforza, L. L. 2005. The Human Genome Diversity Project: Past, Present and Future. *Nature Reviews. Genetics* 6 (4): 333–340.

Cavalli-Sforza, L. L., P. Menozzi and A. Piazza. 1994. *The History and Geography of Human Genes*. Princeton: Princeton University Press.

Childs, T. 2012. Languages of Africa. In *Oxford Bibliographies in Linguistics*, ed. M. Aronoff, DOI: 10.1093/OBO/9780199772810-0092.

Chłodnicki, M. et al. 2005. Fourth Cataract Excavations at es-Sadda and Hagar el-Beida, Survey between es-Sadda and Shemkhiya: Two Seasons of Work in 2005. *Polish Archaeology in the Mediterranean* 17: 337–347.

Chłodnicki, M., A. Longa and O. Piotr. 2010. Fourth Cataract Archaeological Survey between El-Ar (Shemkhiya) and El-Gamamiya (November–December 2007). *Polish Archaeology in the Mediterranean* 19: 377–393.

Chłodnicki, M. and B. Żurawski. 2004. From Khor Umm Ghizlan to Shemkhiya: Archaeological Survey in the Fourth Cataract Region, December 2004. *Polish Archaeology in the Mediterranea* 16: 369–383.

Congiu, A. et al. 2012. Online Databases for mtDNA and Y Chromosome Polymorphisms in Human Populations. *Journal of Anthropological Sciences* 90: 201–215.

Coppens, Y. 1977. Hominid Remains from the Plio/Pleistocene Formations of the Omo Basin, Ethiopia. *Journal of Human Evolution* 6 (2): 169–173.

Crassard, R. and Y. H. Hilbert. 2013. A Nubian Complex Site from Central Arabia: Implications for Levallois Taxonomy and Human Dispersals during the Upper Pleistocene. *PloS One* 8 (7): e69221.

Crawford, N. G. et al. 2017. Loci Associated with Skin Pigmentation Identified in African Populations. *Science* 358 (6365): http://dx.doi.org/10.1126/science.aan8433.

Creanza, N. et al. 2015. A Comparison of Worldwide Phonemic and Genetic Variation in Human Populations. *Proceedings of the National Academy of Sciences of the United States of America* 112 (5): 1265–1272.

Crowther, A. et al. 2017. Subsistence Mosaics, Forager-Farmer Interactions, and the Transition to Food Production in Eastern Africa. *Quaternary International* 489: 101–120.

Cruciani, F. et al. 2011. A Revised Root for the Human Y Chromosomal Phylogenetic Tree: the Origin of Patrilineal Diversity in Africa. *American Journal of Human Genetics* 88 (6): 814–818.

Cruciani, F. et al. 2007. Tracing Past Human Male Movements in Northern/Eastern Africa and Western Eurasia: New Clues from Y-Chromosomal Haplogroups E-M78 and J-M12. *Molecular Biology and Evolution* 24 (6): 1300–1311.

Dobon, B. et al. 2015. The Genetics of East African Populations: a Nilo-Saharan Component in the African Genetic Landscape. *Scientific Reports (Nature)* 5: 9996.

Eerkens, J. W. et al. 2018. Early Childhood Diets on the Nile: δ13 C and δ15 N in Serial Samples of Permanent First Molars in an Elite Meroitic Population from Sai Island, Sudan. *International Journal of Osteoarchaeology* 28: 552–562.

Emberling, G. 2009. Letter from Sudan: The Gold of Kush. *Archaeology* 62 (6): 55–59.

Fletcher, A., D. Antoine J. D. Hill. 2014. *Regarding the Dead: Human Remains in the British Museum*. London: British Museum.

Fox, C. L. and C. Lalueza Fox. 1997. mtDNA Analysis in Ancient Nubians Supports the Existence of Gene Flow between Sub-Sahara and North Africa in the Nile Valley. *Annals of Human Biology* 24 (3): 217–227.

Francigny, V. 2017. The Elite Meroitic Necropolis of Sai Island Part I: Mortuary Interpretations. *The 81st Annual Meeting of the Society for American* Archaeology: https://core.tdar.org/document/429355/the-elite-meroitic-necropolis-of-sai-island-part-i-mortuary-interpretations [Accessed September 6, 2018].

Fregel, R. et al. 2018. Ancient Genomes from North Africa: Evidence Prehistoric Migrations to the Maghreb from both the Levant and Europe. *Proceedings of the National Academy of Sciences of the United States of America* 115 (26): 6774–6779.

Fu, Q. et al. 2016. The Genetic History of Ice Age Europe. *Nature* 534 (7606): 200–205.

Gifford-Gonzalez, D. 1998. Early Pastoralists in East Africa: Ecological and Social Dimensions. *Journal of Anthropological Archaeology* 17 (2): 166–200.

Gifford-Gonzalez, D. 2017. Pastoralism in Sub-Saharan Africa. In *The Oxford Handbook of Zooarchaeology*, eds. U. Albarella, M. Rizzetto, H. Russ, K. Vickers and S. Viner-Daniels. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199686476.013.27.

Gifford-Gonzalez, D. and O. Hanotte. 2011. Domesticating Animals in Africa: Implications of Genetic and Archaeological Findings. *Journal of World Prehistory* 24 (1): 1–23.

Gomes, V. et al. 2010. Digging Deeper into East African Human Y Chromosome Lineages. *Human Genetics* 127 (5): 603–613.

Gomes, V. et al. 2015. Mosaic Maternal Ancestry in the Great Lakes Region of East Africa. *Human Genetics* 134 (9): 1013–1027.

Gonder, M. K. et al. 2007. Whole-mtDNA Genome Sequence Analysis of Ancient African Lineages. *Molecular Biology and Evolution* 24 (3): 757–768.

Greenberg, J. 1963. Summering. *The Hudson Review* 16 (2): 203.

Grollemund, R. et al. 2015. Bantu Expansion Shows that Habitat Alters the Route and Pace of Human Dispersals. *Proceedings of the National Academy of Sciences of the United States of America* 112 (43): 13296–13301.

Groucutt, H. S. et al. 2015. Rethinking the Dispersal of *Homo sapiens* Out of Africa. *Evolutionary Anthropology* 24 (4): 149–164.

Grzybowski, T. and U. Rogalla. 2011. Mitochondria in Anthropology and Forensic Medicine. *Advances in Experimental Medicine and Biology* 942: 441–453.

Gudbjartsson, D. F. et al. 2015. Large-Scale Whole-Genome Sequencing of the Icelandic Population. *Nature Genetics* 47 (5): 435–444.

Güldemann, T. 2016. *The Languages and Linguistics of Africa: a Comprehensive Guide*. Berlin: De Gruyter.

Günther, T. et al. 2015. Ancient Genomes Link Early Farmers from Atapuerca in Spain to Modern-Day Basques. *Proceedings of the National Academy of Sciences of the United States of America* 112 (38): 11917–11922.

Gurdasani, D. et al. 2015. The African Genome Variation Project Shapes Medical Genetics in Africa. *Nature* 517 (7534): 327–332.

Gutenkunst, R. N. et al. 2009. Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. *PLoS Genetics* 5 (10): p.e1000695.

Haak, W. et al. 2010. Ancient DNA from European Early Neolithic Farmers Reveals their Near Eastern Affinities. *PLoS Biology* 8 (11): p.e1000536.

Haak, W. et al. 2015. Massive Migration from the Steppe was a Source for Indo-European Languages in Europe. *Nature* 522 (7555): 207–211.

Haber, M. et al. 2016. Chad Genetic Diversity Reveals an African History Marked by Multiple Holocene Eurasian Migrations. *American Journal of Human Genetics* 99 (6): 1316–1324.

Hammarström, H. et al. 2018. *Glottolog 3.2*. Available at: http://glottolog.org/ accessed [Accessed May 26, 2018].

Hey, J. 2004. Multilocus Methods for Estimating Population Sizes, Migration Rates and Divergence Time, with Applications to the Divergence of Drosophila pseudoobscura and D. persimilis. *Genetics* 167 (2): 747–760.

Hofmanová, Z. et al. 2016. Early Farmers from across Europe Directly Descended from Neolithic Aegeans. *Proceedings of the National Academy of Sciences of the United States of America* 113 (25): 6886–6891.

Hollfelder, N. et al. 2017. Northeast African Genomic Variation Shaped by the Continuity of Indigenous Groups and Eurasian Migrations. *PLoS Genetics* 13 (8): p.e1006976.

Kahn, D. 2006. The Assyrian Invasions of Egypt (673–663 B.C.) and the Final Expulsion of the Kushites. *Studien zur Altägyptischen Kultur* 34: 251–267.

Kaiser, J. 2015. Who Has Your DNA—Or Wants It. *Science* 349 (6255): 1475–1475.

Keller, A. et al. 2012. New Insights into the Tyrolean Iceman's Origin and Phenotype as Inferred by Whole-Genome Sequencing. *Nature Communications* 3: 698.

Kelly, R. C. 2005. The Evolution of Lethal Intergroup Violence. *Proceedings of the National Academy of Sciences* 102 (43): 15294–15298.

Kimura, B. et al. 2011. Ancient DNA from Nubian and Somali Wild Ass Provides Insights into Donkey Ancestry and Domestication. *Proceedings of the Royal Society B: Biological Sciences* 278 (1702): 50.

Krings, M. et al. 1999. mtDNA Analysis of Nile River Valley Populations: a Genetic Corridor or a Barrier to Migration? *American Journal of Human Genetics* 64 (4): 1166–1176.

Kumwenda, S. et al. 2017. Challenges Facing Young African Scientists in their Research Careers: a Qualitative Exploratory Study. *Malawi Medical Journal* 29 (1): 1–4.

Lachance, J. et al. 2012. Evolutionary History and Adaptation from High-Coverage Whole-Genome Sequences of Diverse African Hunter-Gatherers. *Cell* 150 (3): 457–469.

Lazaridis, I. et al. 2014. Ancient Human Genomes Suggest Three Ancestral Populations for Present-Day Europeans. *Nature* 513 (7518): 409–413.

Leakey, R. E. F. 1969. Early *Homo sapiens* Remains from the Omo River Region of South-West Ethiopia: Faunal Remains from the Omo Valley. *Nature* 222 (5199): 1132–1133.

Liu, H. et al. 2006. A Geographically Explicit Genetic Model of Worldwide Human-Settlement History. *American Journal of Human Genetics* 79 (2): 230–237.

Llorente, M. G. et al. 2015. Ancient Ethiopian Genome Reveals Extensive Eurasian Admixture in Eastern Africa. *Science* 350 (6262): 820–822.

Loeffler, H. 2016. *Introductory Guide to Ancient Civilizations*. Scotts Valley, CA: Createspace Independent Publishing Platform.

van de Loosdrecht, M. et al. 2018. Pleistocene North African Genomes Link Near Eastern and Sub-Saharan African Human Populations. *Science* 360 (6388): 548–552.

Luis, J. R. et al. 2004. The Levant Versus the Horn of Africa: Evidence for Bidirectional Corridors of Human Migrations. *American Journal of Human Genetics* 74 (3): 532–544.

Malaspinas, A.-S. et al. 2016. A Genomic History of Aboriginal Australia. *Nature* 538 (7624): 207–214.

Mallick, S. et al. 2016. The Simons Genome Diversity Project: 300 Genomes from 142 Diverse Populations. *Nature* 538 (7624): 201–206.

Mather, A. 2003. Review of Global Forest Resources Assessment 2000. *Land Use Policy* 20 (2): 195.

McDermott, F. et al. 1996. New Late-Pleistocene Uranium–Thorium and ESR Dates for the Singa Hominid (Sudan). *Journal of Human Evolution* 31 (6): 507–516.

McDougall, I., F. H. Brown and J. G. Fleagle. 2005. Stratigraphic Placement and Age of Modern Humans from Kibish, Ethiopia. *Nature* 433 (7027): 733–736.

Medina, P. et al. 2018. Estimating the Timing of Multiple Admixture Pulses During Local Ancestry Inference. *Genetics* 210 (3): 1089–1107.

Meyer, M. et al. 2012. A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science* 338 (6104): 222–226.

Kopytoff, I. 1987. *The African Frontier: the Reproduction of Traditional African Societies*. Bloomington: Indiana University Press.

Murail, P. et al. 2004. An Infant Cemetery of the Classic Kerma Period (1750–1500 BC, Island of Sai, Sudan). *Antiquity* 78 (300): 267–277.

Musilová, E. et al. 2011. Population History of the Red Sea: Genetic Exchanges Between the Arabian Peninsula and East Africa Signaled in the Mitochondrial DNA HV1 Haplogroup. *American Journal of Physical Anthropology* 145 (4): 592–598.

Osier, M. V. et al. 2002. ALFRED: An Allele Frequency Database for Anthropology. *American Journal of Physical Anthropology* 119 (1): 77–83.

Osypiński, P. 2010. El-Sadda. Excavations on the Polish concession (Hamadab Dam Rescue Project), January–February 2007. *Polish Archaeology in the Mediterranean* 19: 435–447.

Pagani, L. et al. 2012. Ethiopian Genetic Diversity Reveals Linguistic Stratification and Complex Influences on the Ethiopian Gene Pool. *American Journal of Human Genetics* 91 (1): 83–96.

Pagani, L. et al. 2016. Genomic Analyses Inform on Migration Events During the Peopling of Eurasia. *Nature* 538 (7624): 238–242.

Pagani, L. et al. 2015. Tracing the Route of Modern Humans Out of Africa by Using 225 Human Genome Sequences from Ethiopians and Egyptians. *American Journal of Human Genetics* 96 (6): 986–991.

Pakendorf, B. 2014. Coevolution of Languages and Genes. *Current Opinion in Genetics & Development* 29: 39–44.

Patin, E. et al. 2017. Dispersals and Genetic Adaptation of Bantu-Speaking Populations in Africa and North America. *Science* 356 (6337): 543–546.

Patterson, N. et al. 2012. Ancient Admixture in Human History. *Genetics* 192 (3): 1065–1093.

Pearson, O. M. 2013. Africa. In *The Origins of Modern Humans: Biology Reconsidered*, eds. F. H. Smith and J. C. M. Ahern, 1–43. Hoboken, NJ: Wiley.

Pemberton, T. J., M. DeGiorgio and N. A. Rosenberg. 2013. Population Structure in a Comprehensive Genomic Data Set on Human Microsatellite Variation. *G3* 3 (5): 891–907.

Pickrell, J. K. and D. Reich. 2014. Toward a New History and Geography of Human Genes Informed by Ancient DNA. *Trends in Genetics* 30 (9): 377–389.

Pinhasi, R. et al. 2015. Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous Bone. *PloS One* 10 (6): e0129102.

Plaza, S. et al. 2004. Insights into the Western Bantu Dispersal: mtDNA Lineage Analysis in Angola. *Human Genetics* 115 (5): 439–447.

Poloni, E. S. et al. 2009. Genetic Evidence for Complexity in Ethnic Differentiation and History in East Africa. *Annals of Human Genetics* 73 (6): 582–600.

Price, A. L. et al. 2009. Sensitive Detection of Chromosomal Segments of Distinct Ancestry in Admixed Populations. *PLoS Genetics* 5 (6): p.e1000519.

Prüfer, K. et al. 2014. The Complete Genome Sequence of a Neanderthal from the Altai Mountains. *Nature* 505 (7481): 43–49.

Pugach, I. et al. 2011. Dating the Age of Admixture Via Wavelet Transform Analysis of Genome-Wide Data. *Genome Biology* 12 (2): R19.

Quintana-Murci, L. et al. 2008. Maternal Traces of Deep Common Ancestry and Asymmetric Gene Flow Between Pygmy Hunter-Gatherers and Bantu-Speaking Farmers. *Proceedings of the National Academy of Sciences* 105 (5): 1596–1601.

Rasmussen, M. et al. 2010. Ancient Human Genome Sequence of an Extinct Palaeo-Eskimo. *Nature* 463 (7282): 757–762.

Reyes-Centeno, H. et al. 2014. Genomic and Cranial Phenotype Data Support Multiple Modern Human Dispersals from Africa and a Southern Route into Asia. *Proceedings of the National Academy of Sciences* 111 (20): 7248–7253.

Reyes-Centeno, H., K. Harvati and G. Jäger. 2016. Tracking Modern Human Population History from Linguistic and Cranial Phenotype. *Scientific Reports (Nature)* 6 (1): http://dx.doi.org/10.1038/srep36645.

Richards, M. B., P. Soares and A. Torroni. 2016. Palaeogenomics: Mitogenomes and Migrations in Europe's Past. *Current Biology* 26 (6): R243–246.

Rilly C. and V. Francigny. 2011. The Late Meroitic Cemetery at Sedeinga. Campaign 2010. *Sudan & Nubia* 15: 72–79.

Rosa, A. et al. 2004. MtDNA Profile of West Africa Guineans: Towards a Better Understanding of the Senegambia Region. *Annals of Human Genetics* 68 (4): 340–352.

Rosa, A. and A. Brehem. 2011. African Human mtDNA Phylogeography at-a-glance. *Journal of Anthropological Sciences* 89: 25–58.

Rose, J. I. et al. 2011. The Nubian Complex of Dhofar, Oman: an African Middle Stone Age Industry in Southern Arabia. *PloS One* 6 (11): e28239.

Sankararaman, S. et al. 2008. Estimating Local Ancestry in Admixed Populations. *American Journal of Human Genetics* 82 (2): 290–303.

Scerri, E. M. L. et al. 2018. Did Our Species Evolve in Subdivided Populations Across Africa, and Why Does It Matter? *Trends in Ecology & Evolution* 33 (8): 582–594.

Scheinfeldt, L. B., S. Soi and S. A. Tishkoff. 2010. Working Toward a Synthesis of Archaeological, Linguistic, and Genetic Data for Inferring African Population History. *Proceedings of the National Academy of Sciences* 107 (Supplement2): 8931–8938.

Schiffels, S. and R. Durbin. 2014. Inferring Human Population Size and Separation History from Multiple Genome Sequences. *Nature Genetics* 46 (8): 919–925.

Schlebusch, C. M. et al. 2017. Southern African Ancient Genomes Estimate Modern Human Divergence to 350,000 to 260,000 Years Ago. *Science* 358 (6363): 652–655.

Schlebusch, C. M. and M. Jakobsson. 2018. Tales of Human Migration, Admixture, and Selection in Africa. Annual Review of Genomics and Human Genetics 19: 405–428.

Schuenemann, V. J. et al. 2017. Ancient Egyptian Mummy Genomes Suggest an Increase of Sub-Saharan African Ancestry in Post-Roman Periods. *Nature Communications* 8: 15694.

Semino, O. et al., 2002. Ethiopians and Khoisan Share the Deepest Clades of the Human Y-Chromosome Phylogeny. *American Journal of Human Genetics* 70 (1): 265–268.

Silva, N. M., J. Rio and M. Currat. 2017. Investigating Population Continuity with Ancient DNA under a Spatially Explicit Simulation Framework. *BMC Genetics* 18 (1): http://dx.doi.org/10.1186/s12863-017-0575-6.

Skoglund, P. et al. 2017. Reconstructing Prehistoric African Population Structure. *Cell* 171 (1): 59–71.e21.

Soares, P. et al., 2012. The Expansion of mtDNA Haplogroup L3 within and out of Africa. *Molecular Biology and Evolution* 29 (3): 915–927.

St. Clair, M. 2016. Harnessing Global Y-Chromosome Variation as a Tool for Linguistic Research. *The Genetic-Linguistic Interface*. Available at: http://www.genlinginterface.com/resource-guide/y-chromosome-variation-and-linguistic-research/.

Stoneking, M. & Krause, J. 2011. Learning about Human Population History from Ancient and Modern Genomes. *Nature Reviews. Genetics* 12 (9): 603–614.

Stringer, C. 2016. The Origin and Evolution of *Homo sapiens*. *Philosophical Transactions of the Royal Society of London*. Series B, Biological Sciences 371 (1698): http://dx.doi.org/10.1098/rstb.2015.0237.

Tishkoff, S. A. et al. 2009. The Genetic Structure and History of Africans and African Americans. *Science* 324 (5930): 1035–1044.

Underhill, P. A. et al. 2000. Y Chromosome Sequence Variation and the History of Human Populations. *Nature Genetics* 26 (3): 358–361.

Underhill, P. A. and T. Kivisild. 2007. Use of Y Chromosome and Mitochondrial DNA Population Structure in Tracing Human Migrations. *Annual Review of Genetics* 41: 539–564.

Usai, D. et al. 2014. The Al Khiday Cemetery in Central Sudan and its "Classic/Late Meroitic" Period Graves. *Journal of African Archaeology* 12 (2): 183–204.

Usik, V. I. et al. 2013. Nubian Complex Reduction Strategies in Dhofar, Southern Oman. *Quaternary International* 300: 244–266.

Van Peer, P. et al. 2003. The Early to Middle Stone Age Transition and the Emergence of Modern Human Behaviour at site 8-B-11, Sai Island, Sudan. *Journal of Human Evolution* 45 (2): 187–193.

Veeramah, K. R. and M. F. Hammer. 2014. The Impact of Whole-Genome Sequencing on the Reconstruction of Human Population History. *Nature Reviews. Genetics* 15 (3): 149–162.

Walter, R. C. et al. 2000. Early Human Occupation of The Red Sea Coast of Eritrea during the Last Interglacial. *Nature* 405 (6782): 65–69.

Winchell, F. et al. 2017. Evidence for Sorghum Domestication in Fourth Millennium BC Eastern Sudan: Spikelet Morphology from Ceramic Impressions of the Butana Group. *Current Anthropology* 58 (5): 673–683.

Wood, B. and M. Leakey. 2011. The Omo-Turkana Basin Fossil Hominins and their Contribution to Our Understanding of Human Evolution in Africa. *Evolutionary Anthropology* 20 (6): 264–292.

Wood, E. T. et al. 2005. Contrasting Patterns of Y Chromosome and mtDNA Variation in Africa: Evidence for Sex-Biased Demographic Processes. *European Journal of Human Genetics* 13 (7): 867–876.

## Appendix

The table below presents extended data for the sampling of East Africa in publicly available datasets. Datasets were considered if they are actively maintained and their metadata lists a population geographical location and/or ethnic affiliation.

| Study | Source of genetic data | East Africa to World (%) | Country Counts (Sudan, Ethiopia, Somalia, Uganda, Kenya, Tanzania, Djibouti, Eritrea) | Reference | Publicly available data |
|---|---|---|---|---|---|
| Ancient human genomes suggest three ancestral populations for present-day Europeans | Genome-wide SNPs, mtDNA | 22.8 | 7, 11, 13, 0, 38, 47, 0, 0 | Lazaridis et al. 2014 | Yes |
| Y-Chromosome Variation Among Sudanese: Restricted Gene Flow, Concordance With Languagc, Geography, and History | Y-chromosome | NA | 445, 0, 0, 0, 0, 0, 0, 0 | Hassan et al. 2008 | Yes |
| mtDNA analysis of Nile River Valley populations: A genetic corridor or a barrier to migration? | mtDNA | NA | 156, 0, 0, 0, 0, 0, 0, 0 | Krings M et al. 1999 | Yes |
| Ethiopian Genetic Diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool | Genome-wide SNPs | NA | 24, 188, 23, 0, 0, 0, 0, 0 | Pagani et al. 2012 | NA |
| Y-chromosome E haplogroups: their distribution and implication to the origin of Afro-Asiatic languages and pastoralism | Y-chromosome | NA | 0, 0, 0, 0, 0, 0, 0, 39 | Gebremeskel & Ibrahim 2014 | NA |

| | | | | | |
|---|---|---|---|---|---|
| Admixture into and within sub-Saharan Africa | Genome-wide SNPs | NA | 0, 0, 0, 0, 232, 148, 0, 0 | Busby et al. 2016 | Yes |
| Tracing the Route of Modern Humans out of Africa by Using 225 Human Genome Sequences from Ethiopians and Egyptians | Whole-Genome Sequences | NA | 0, 125, 0, 0, 0, 0, 0, 0 | Pagani et al. 2015 | Yes |
| The Genetics of East African Populations: A Nilo-Saharan Component in the African Genetic Landscape | 195,806 SNPs and 718 small INDELs | NA | 408, 39, 0, 0, 0, 0, 0, 0 | Dobon et al. 2016 | Yes |
| Mitochondrial footprints of human expansions in Africa | mtDNA | NA | 0, 0, 27, 0, 61 (Kikuyu 24, Turkana 37), 0, 0, 0 | Watson et al. 1997 | Yes |
| mtDNA Variation in East Africa Unravels the History of Afro-Asiatic Groups | mtDNA | NA | 0, 167, 0, 0, 285, 0, 0, 0, 0 | Boattini et al 2013 | NA |
| The Expansion of mtDNA Haplogroup L3 within and out of Africa | HVS (and whole-mtDNA sequencing) | NA | 102 (21), 77 (and 16), 148 (and 20), 0, 0, 0, 0, 0 | Soares 2011 | Yes |
| Ethiopian Mitochondrial DNA Heritage: Tracking Gene Flow Across and Around the Gate of Tears | HVS mtDNA | NA | 0, 271, 0, 0, 0, 0, 0, 0 | Kivisild 2004 | Yes |

| | | | | | |
|---|---|---|---|---|---|
| Genetic Evidence for Complexity in Ethnic Differentiation and History in East Africa | mtDNA HVS-I and HVS-II, 4 SNPS related to haplogroups L1, L2, L3, M and N | NA | 0, 161, 0, 0, 47, 0, 0, 0 | Poloni 2009 | Yes |
| Mitochondrial DNA control region sequences from Nairobi (Kenya): inferring phylogenetic parameters for the establishment of a forensic database | mtDNA | NA | 0, 0, 0, 0, 100, 0, 0, 0 | Brandstätter 2004 | Yes |
| African Y Chromosome and mtDNA Divergence Provides Insight into the History of Click Languages | mtDNA, Y-Chromosome | NA | 0, 0, 0, 0, 0, 100 (mtDNA) and 69 (Y-chromosome), 0, 0 | knight 2003 | Yes |
| History of Click-Speaking Populations of Africa Inferred from mtDNA and Y Chromosome Genetic Variation | mtDNA, Y-Chromosome | NA | 0, 0, 0, 0, 0, 316 (mtDNA) and 219 (Y-Chromosome), 0, 0 | Tishkoff et al. 2007 | Yes |
| Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa | aDNA Whole-Genome Sequences | NA | 0, 1, 0, 0, 0, 0, 0, 0 | Llorente et al. 2015 | Yes |
| Reconstructing prehistoric African population structure | aDNA Whole-Genome Sequences | NA | 0, 0, 0, 0, 1, 4, 0, 0 | Skoglund et al. 2017 | Yes |

| | | | | | |
|---|---|---|---|---|---|
| mtDNA analysis in ancient Nubians supports the existence of gene flow between sub-Sahara and North Africa in the Nile valley | mtDNA marker | NA | 29, 0, 0, 0, 0, 0, 0, 0 | Fox CL. 1997 | NA |
| Loci associated with skin pigmentation identifiers in African populations | Genome-wide SNPs and Whole-Genome Sequencing | NA | 0, 571, 0, 0, 0, 0, 476, 0, 0 | Crawford et al. 2017 | Yes |
| Evolutionary History and Adaptation from High-Coverage Whole-Genome Sequences of Diverse African Hunter-Gatherers | Whole-Genome Sequencing | NA | 0, 0, 0, 0, 0, 0, 10, 0, 0 | Lachance et al. 2012 | Yes |
| The Simons Genome Diversity Project: 300 genomes from 142 diverse populations | Whole-Genome Sequencing | 4 | 3, 0, 0, 0, 9, 0, 0, 0 | Mallick et al. 2016 | Yes |