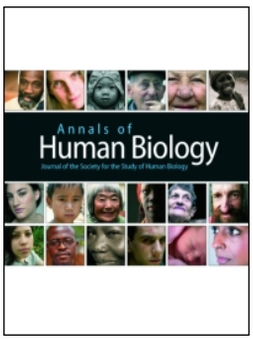

Annals of Human Biology

ISSN: 0301-4460 (Print) 1464-5033 (Online) Journal homepage: https://www.tandfonline.com/loi/iahb20

Searching for archaic contribution in Africa

Cindy Santander, Francesco Montinaro & Cristian Capelli

To cite this article: Cindy Santander, Francesco Montinaro & Cristian Capelli (2019): Searching for archaic contribution in Africa, Annals of Human Biology, DOI: 10.1080/03014460.2019.1624823

To link to this article: https://doi.org/10.1080/03014460.2019.1624823

Accepted author version posted online: 04 Jun 2019.

Searching for archaic contribution in Africa

Cindy Santander[1], Francesco Montinaro[1,2], Cristian Capelli[1]

1. *Department of Zoology, University of Oxford, Oxford, United Kingdom*

2. *Estonian Biocentre, University of Tartu, Tartu, Estonia*

Corresponding authors: cindy.santander@zoo.ox.ac.uk, cristian.capelli@zoo.ox.ac.uk

Acknowledgements

The authors would like to thank St. Hugh's College, the Leverhulme Trust, and Comisión Nacional de Investigación Científica y Tecnológica, Gobierno de Chile. We would also like to thank all the students, researchers, and collaborators who have contributed to the studies that elucidate on the genomic diversity across African populations and to the people who have donated their DNA to make those studies possible.

Searching for archaic contribution in Africa

Context: Africa's role in the narrative of human evolution is indisputably emphasised in the emergence Homo sapiens. However, once humans disperse beyond Africa, the history of those who stayed remains vastly understudied lacking the proper attention it deserves as the birthplace of both modern and archaic humans. The sequencing of Neanderthal and Denisovan genomes has elucidated evidence of admixture between archaic and modern humans outside of Africa but has not aided efforts in answering whether archaic admixture happened within Africa.

Objectives: We review here the state of research for archaic introgression in African populations and discuss recent insights into this topic.

Methods: Gathering published sources and recently released preprints, this review reports on the different methods developed for detecting archaic introgression. Particularly we discuss how relevant these are when implemented on African populations and what findings these studies have shown so far.

Results: Methods for detecting archaic introgression have been predominantly developed and implemented on non-African populations. Recent preprints present new methods considering African populations. While a number of studies using these methods suggest archaic introgression in Africa, without an African archaic genome to validate these results, such findings remain as putative archaic introgression.

Conclusion: In light of the caveats with implementing current archaic introgression detection methods in Africa, we recommend future studies to concentrate on unravelling the complicated demographic history of Africa through means of ancient DNA where possible and through more focused efforts to sequence modern DNA from more representative populations across the African continent

Keywords: Africa; archaic introgression; ancient admixture; human evolution

Current state-of-the-art for archaic introgression

The past decade has seen the industrial-scale sequencing and studying of ancient and modern genomes alike (Slatkin & Racimo 2016; Nielsen et al. 2017). Prior to this, studying the relationship between archaic hominins and anatomically modern humans was predominantly left to paleoanthropologists whom have proposed several evolutionary models since the first half of the 20th century. The most popular hypotheses, which have been in competition up until very recently, were the multiregional evolution and the recent out of Africa model (RAO). Multiregional evolution proposed that humans evolved across the Old World from local ancestors through an interbreeding network since the last 1.8 million years (Wolpoff et al. 1984), while RAO argued for an origin derived from Africa which then dispersed across the globe (Stringer & Andrews 1988). The turning point in this debate is often considered to be in the late 80s when Cann and colleagues conducted a study of mitochondrial DNA of present-day individuals where the results suggested a recent African origin for modern humans. This was followed by several other Y-chromosome and autosomal DNA studies suggesting similar results and therefore supporting the RAO model for many years (Jobling & Tyler-Smith 2003).

Since the sequencing of Neanderthal (Green et al. 2010) and Denisova genomes (Reich et al. 2010) we have been able to characterise the evolutionary relationship between archaic and modern humans using genetic data as well as explore the hypothesis of gene flow among them (Wall et al. 2013; Vernot & Akey 2014; Vernot et al. 2016; Prüfer et al. 2017; Browning et al. 2018; Hajdinjak et al. 2018)(note here that paleoanthropologists hold different positions on the significance of the degree of variation present within the genus *Homo*, and the classification for these different groups ranges from species to populations. The terms modern and archaic are used here for simplicity and are intended to refer to modern human populations usually defined as *H. sapiens* and broadly to all the other *Homo*

groups that co-existed, as for example Neanderthals and *H. floresiensis*, as well as poorly characterised forms as the Denisovans, and others still unknown). It is now understood that neither multiregionalism nor RAO can strictly explain the complex interactions between the hominin taxa (Galway-Witham & Stringer 2018). A number of intermediate models have been suggested where their main differences lie in how human genetic dispersal has taken place while mixing with other hominin groups outside of Africa. Intermediate models have included the "Leaky Replacement" model, which is fundamentally RAO with limited hybridization between modern and archaic populations. However, some of these models continue to assume an evolutionary history of humans diverging off a closely related core African population without accounting for the presence of multiple populations within Africa and their absorption through admixture before proximate extinctions of other *Homo* taxa prior the exodus from Africa. Others have made attempts to integrate genomic, archaeological, fossil and paleoenvironmental data to form models inclusive of African population structure (Lahr & Foley 1998; Harding & McVean 2004; Stringer 2016; Scerri et al. 2018; Henn et al. 2018).

As research continued and progressed in this field, the idea of admixture between archaic and modern humans has become an indispensable element to consider in any human evolution model. Evaluating admixture between non-Africans and archaic hominins has been comparatively more straightforward than in Africans given the availability of both Neanderthal and Denisova genomes (Green et al. 2010; Reich et al. 2010). The sequencing of non-African archaic hominins has led to irrefutable interest in wanting to understand what relationship they had with modern humans and moreover what possibly connects us to them. This yearning has manifested in a number of questions about our origin as a species, but it has also given rise to whether this phenomenon has occurred not just in Eurasia but also in

Africa (Garrigan et al. 2005; Wall et al. 2009; Hammer et al. 2011; Lachance et al. 2012; P H Hsieh et al. 2016; Skoglund et al. 2017).

In order to answer these queries, a number of tools have been developed to identify and characterise signatures of archaic admixture, predominantly depending on either variant distribution, linkage disequilibrium, or both. Patterson's *D* statistic, for example, measures excess sharing of derived alleles between two sister populations (ingroup) and an outgroup. If none of the ingroup populations received genetic material from the outgroup, they should share about the same number of derived alleles with the outgroup. While this method can detect whether there is an asymmetrical variant sharing between archaic and modern humans, it is unable to pinpoint segments in the genome that are of introgressed archaic origin (Patterson et al. 2012). Moreover, demographic scenarios such as ancient population structure can generate similar results (Eriksson & Manica 2012; Theunert & Slatkin 2017).

Mutation and recombination impact inherited DNA and therefore inevitably shape the segments of a putatively archaic origin. Consequently, these processes may be harnessed when developing algorithms to detect tracts of archaic introgression. Pedigree studies in humans would suggest that about 80 new mutations occur each generation leading to a mutation rate of approximately 0.5-1 x 10^{-9} per base pair per year (Scally & Durbin 2012; Besenbacher et al. 2015). Given that Neanderthals have diverged from the common ancestor with *Homo sapiens* about 520,000-630,000 years ago (Green et al. 2010; Prüfer et al. 2017; Hajdinjak et al. 2018), we can expect that the DNA of any two humans will be on average closer to each other when compared to a Neanderthal's DNA sequence.

Given that recombination is not evenly distributed across the human genome, several projects have built maps detailing recombination crossover rates for populations of European ancestry (Kong et al. 2002; Matise et al. 2007; Kong et al. 2010; The HapMap Consortium et al. 2010) and of West African ancestry (The HapMap Consortium et al. 2010; Hinch et al.

2011; Wegmann et al. 2011). From a study led by Hinch and colleagues, an African enrichment (AE) map, a map of hotspots unique to African ancestry, was deduced by comparing the Icelandic deCODE and African American (AA) recombination patterns (Kong et al. 2010; Hinch et al. 2011). Availability of resources, such as high-resolution genetic maps, which are relevant for the populations of interest are crucial for current methods that detect archaic introgression in modern genomes.

Here we delineate some of the most up-to-date methods used in detecting archaic introgression in both Africans and non-Africans. We consider the pitfalls of these methods bearing in mind the absence of a sequenced African archaic genome, the complex demographic histories yet to disentangle in African population, and the overall scarcity of modern African genomes which are comprehensively representative of the diverse ethnic-groups found on the continent.

Methods to infer archaic introgression

Linkage disequilibrium-based methods

We can consider recombination and mutations as the basis of understanding an expected length of an introgressed tract in relation to the time since the admixture event. Introgression is distinct from incomplete lineage sorting (ILS) in that it should leave behind longer tracts, as the former being more recent (Liang & Nielsen 2014). The difference in the size of the tracts should provide a way to tell whether a shared tract with an archaic hominin is in fact introgression or the genomic relic from an earlier common ancestral population.

Prior to the sequencing of the full Neanderthal and Denisova genomes, Plagnol and Wall (Plagnol & Wall 2006) sought to take advantage of the logic that putatively introgressed tracts would have had a limited number of generations (e.g. ~2000) to be broken down by recombination. Those tracts could be identified by linkage disequilibrium (LD) where

variants in an archaic segment should be strongly associated with other archaic variants in the genome, in other words introgressed archaic variants should be found in high LD. The authors came up with the *S** statistic (Figure 1), a summary statistic of LD, which extracts this particular information through a scoring scheme that searches for derived mutations that are in high LD. This method has become widely used in several studies which seek signals of archaic introgression in both non-African and African populations alike as it was originally designed to identify introgression without knowledge of the donor population (Plagnol & Wall 2006; Wall et al. 2009). However, the availability of archaic hominin sequences outside of Africa have provided a means of further corroborating haplotypes detected by *S** in non-African populations (Wall et al. 2013; Vernot & Akey 2014; Vernot et al. 2016; Browning et al. 2018). Such studies have been able to find more Neanderthal contribution in East Asians than in Europeans potentially because of a two admixture waves – one in Eurasian ancestors and another in Asians (Wall et al. 2013; Vernot & Akey 2014). However, these signatures can also be influenced by the different effective population sizes for Europeans and Asians as well as a dilution of Neanderthal signatures in Europeans due to gene flow from or into Africa, which can confound the true contribution from Archaics into non-Africans (Sankararaman et al. 2014; Petr et al. 2019).

A subsequent study was able to detect introgressed sequences which are uniquely Neanderthal or Denisovan or shared amongst populations that have received contribution from both archaic humans, such as in the case of the Melanesians and Asians (Vernot et al. 2016). A more recent study introduced a *S**-like method, Sprime, which uses a similar scoring scheme as its predecessor with the exception that it now performs detection on whole chromosomes instead of sliding windows and takes into account local mutation and recombination rates (Browning et al. 2018). The authors of this method report that Sprime helped them to discern that Asians today carried Denisovan introgression from two waves of

Denisovan admixture, one from a population closely related to the Altai Denisovan into East Asian and another more distantly related to the Altai Denisovan in Papuans and South Asians. A more recent study has elaborated on the complexity of archaic contact between Denisovans and modern humans (Jacobs et al. 2019).

Probabilistic machine learning methods

Another option to using LD-based methods in detecting introgressed sequences is to incorporate parametric assumptions into a probabilistic framework. Two methods that are utilised for this are hidden Markov models (HMM) (Baum & Eagon 1967) and conditional random field (CRF) (Lafferty et al. 2001). They are implemented with the concept that each single nucleotide polymorphism (SNP) across the genome is a hidden random variable with two states: either human or archaic (Figure 2). These methods integrate what we know about the biological processes such as human recombination and mutation and what we understand about the demographic events since the human exodus from Africa. In the case of looking for introgression in non-Africans, these methods are implemented with the assumption that we would not expect to find Neanderthal or Denisovan contributions in West Africans from archaic admixture. With the resulting parametric information one can calculate the probability that a given sequence is of archaic origin and if it is present in modern humans via admixture.

Two different studies made use of HMM methods but the main difference between the two was that, respectively, one provided *a priori* chosen parameters whereas the other gathered parameters from a reference dataset (Prüfer et al. 2014; Seguin-Orlando et al. 2014). Prüfer and colleagues (2014) estimated a 70-100% archaic enrichment in East Asians compared to Europeans. Seguin-Orlando and colleagues (2014), by looking at modern human aDNA, found the approximate time of admixture between modern humans and Neanderthals

occurred 16,600 years earlier than the ancient sample tested, in-line with what other studies approximate (37,000–86,000 years ago) (Green et al. 2010; Reich et al. 2010; Sankararaman et al. 2012; Fu et al. 2014).

Similar to HMM are CRF models which can incorporate other forms of information related to the data for example, LD, haplotype structure and allele configurations from multiple samples. The parameters are then calibrated by training them using simulations with particular demographic assumptions ranging from divergence dates to effective population size.

Sankararaman et al. (2014) used a CRF model in their study where the first emission function provided a high probability of being archaically introgressed if a variant was found in Non-Africans and in the Neanderthal reference genome but absent in Africa; akin to what the aforementioned studies had done with their HMM methods. Using this CRF model the authors detected in the 1000 Genomes Project 15% of introgressed sequence with 99% precision. As they detected more archaic sequence, 38%, their precision decreased to 98%. This study found 1.17-1.20% of the autosomes in Europeans have Neanderthal ancestry and that between 1.37-1.40% of East Asian autosomes have Neanderthal ancestry. Two years after developing this CRF-based method, Sankararaman and colleagues (2016) applied it to the Simons Genome Diversity Project (Mallick et al. 2016) which include 257 high coverage individuals from 142 worldwide populations. Here both the Neanderthal and Denisova genomes were used in the analysis and they identified 1.06% of autosomal DNA in Europeans of Neanderthal origin while in East Asians the proportion was slightly higher at 1.40%. Furthermore, they found a small although still significant amount of Denisova contribution in East Asians, estimated to be 0.06%.

While inarguably powerful methods, so far HMM and CRF have predominantly relied on two things: *a priori* chosen parameters which are based off of well-established demographic assumptions and a reference dataset. Recently an HMM method was released by Skov et al. (2018) which does not depend on an archaic reference but does require a specific phylogenetic arrangement that allows a large fraction of variation to be removed using an outgroup population to detect introgression. This model bases its logic on the observation that an archaic tract introgressed into a population should display a high density of variants not found in populations which have not experienced introgression. Like the previous methods reviewed here, Skov and colleagues (2018) focus on a scenario where admixture has occurred between a deeply divergent archaic population (e.g. Denisova) and a modern human population (ingroup - Papuans) but where admixture has only occurred with the ingroup and not with the outgroup (e.g. Yoruba). By removing variants that are shared with the outgroup, they inspect the density of the remaining variants in the ingroup which are, essentially, private SNPs. In their analysis they use Sub-Saharan Africans from the 1000 Genomes Project as an outgroup. The authors demonstrate, with 89 Papuan genomes, that they are able to recover more Denisovan introgressed segments (77 Mb out of 164.23 Mb of archaic sequence in Papuans) than previous methods (Vernot et al. 2016; Browning et al. 2018) possibly because this approach does not rely upon validation with the Altai Denisovan genome reference, ultimately demonstrating the advantage of a reference-free detection method. In addition to detecting introgressed fragments, this method can also infer admixture proportions as well as divergence time of human and archaic populations. Although this method does not require an archaic reference and does not require phased data, without using a suitable outgroup, detecting introgression in African population with just genome samples remains a challenge.

Detection of putative archaic introgression in Africans

So far, we have reviewed studies that have implemented the above methods outside Africa, but which have also relied on an archaic reference to filter out false positive candidates as well as recover false negatives. Without using an archaic reference, the authors of those studies have collectively shown that methods like *S** and Sprime can identify anywhere from 30-60% of all true positive introgressed sequences at low false discovery rates, and in the case of Sprime, with an accuracy of 93% as was shown on simulated data for different demographic histories (Vernot & Akey 2014; Vernot et al. 2016; Browning et al. 2018). Skov et al. (2018) have shown a low rate of false detection on both simulated and real data using an HMM method without the need of an archaic reference.

Earlier studies initially implemented *S** on African populations using genotype data for both coding and non-coding regions (Plagnol & Wall 2006; Wall et al. 2009; Hammer et al. 2011). In particular, a study looking at 61 noncoding autosomal regions was able to infer about 2% of genetic material introgressed into contemporary Mandenka, Biaka, and San around 35 kya from an archaic population that split from the ancestors of modern humans approximately around the same time as Neanderthal and Denisova, ≈700 kya (Hammer et al. 2011). The authors use an inferential approach to test whether the data they have for two hunter-gatherer populations (Biaka and San) and an agricultural population (Mandenka) fit a scenario of no admixture or low levels of admixture under two possible models—a two-population or a three-population model. Under a two-population model they test for archaic introgression by comparing *S** values from their data to values estimated using parameters under a no admixture scenario. In a complementary fashion, the authors also introduce three summary statistics for an approximate-likelihood method which consists in first identifying the two most divergent sequences for a locus and establishing two groups where the rest of

the sequences will cluster to correspondingly. The protocol then calls for estimating the distribution of the fraction of shared polymorphisms between the two groups (D_1), the ratio of the number of differences between them (D_2), and finally the size of the smaller of the two groups (D_3). Each one is meant to represent the time of introgression, the time of the archaic-split, and finally the proportion of admixture, respectively. This distribution of summary statistics is calculated by simulating several ancestral recombination graphs (ARGs) (Griffiths & Marjoram 1997). In using both S^* and these three summary statistics (D_1, D_2, D_3), the authors rejected the null hypothesis that an ancestral population with no admixture gave rise to anatomically modern humans. Instead both their inferential methods identify three exceptionally long haplotypes (Table 1) at low-frequency amongst African hunter-gatherers (central African rainforest hunter-gatherers and San), signalling these regions as putatively introgressed from an archaic population. Specifically, the authors suggest that central Africa may have been the place of origin for an extinct archaic hominin that admixed with modern humans in light of finding all three haplotypes amongst the Mbuti of Democratic Republic of Congo.

With the advances in sequencing technology whole-genome sequencing become accessible enough to sequence multiple individuals of several populations. Lachance and colleagues (2012) were the first to implement S^* on whole-genomes from African Hunter-Gatherer populations: rainforest hunter-gatherers from Cameroon, click-speaking Hadza, and Sandawe from Tanzania. Consistent with previous studies (Wall et al. 2009; Hammer et al. 2011), they find evidence of archaic admixture in all three populations with candidate loci corresponding to a time to recent common ancestor comparable to those observed in Europeans from Neanderthal introgression. The authors through coalescent simulations concluded that S^* was robust enough to detect admixture and differences in amounts of introgression.

An ensuing study by Hsieh and colleagues (2016) corroborated this evidence for archaic admixture in Africa by addressing the confounding effects of the demographic history of the population in question. *S* P*-value distributions were calculated for simulations based on two models inferred from a previous study (Hsieh et al. 2016) that incorporated both isolation and gene flow with neighbouring farming populations. These distributions were compared with *S* P*-value distributions from the observed data. This approach is an improvement to what Lachance and colleagues (2012) performed on their data by accounting for sequences that may have extreme *S** values but not being statistically significant if one considers the effects of demography and genomic processes such as mutation and recombination rates. Only the significant top 1% in this *P*-value distribution were chosen as candidate introgressed loci resulting in a total of 265 candidate loci, spanning ~20 Mb in length. They estimated a false discovery rate (FDR) between 19% and 68% in these top candidates. Surprisingly, from these putatively introgressed regions, Hsieh and colleagues (2016) using a variant of D_3 from Hammer et al. (2011) inferred at least one admixture event with low amounts of introgression around 9,000 years ago albeit a fine-scale understanding of the nature of these recurrent events in Africa still remains unresolved.

In a similar vein, a recent study that analysed 21 high coverage African genomes within an Approximate Bayesian computation (ABC) with Deep Learning framework (Mondal et al. 2019) also estimated that interbreeding occurred between modern humans in Africa (i.e Khoe-San, Mbuti, and West Africans) and an archaic ghost population that diverged from the basal human lineage around the same temporal scale seen between Neanderthal and Denisovans (Lorente-Galdos et al. 2019).

Finally, a genomic analysis comprising genomic material from ancient individuals from Southern and Eastern Africa, and therefore lacking confounding genomic fragments derived by recent demographic events, revealed that a model depicting Southern Africa

Khoe-San populations as basal of all the African populations is not fully supported, since the former show different relatedness to East and Western Africa (Skoglund et al. 2017). Furthermore, different West African populations show different relatedness to ancient Khoe-San, which is inconsistent with being derived from a homogeneous ancestral population which diverged from ancient southern Africans. This scenario would be compatible with either "archaic" admixture in Western Africans, affecting different populations heterogeneously, or with long term admixture which affected Western African groups with different magnitudes.

Although the results of recently released preprints are subject to change, we also report here some of the interesting and strongly relevant findings recently submitted to preprint archives. Durvasula and Sankararaman (2018) developed a machine learning method, ArchIE, which makes use of training datasets to calculate a set of features that are potentially informative of introgression. A prediction about archaic local ancestry could then be made for any given window by using a binary logistic regression model with the set of computed parameters. Results are then summarised for what is indicative of archaic admixture for each haplotype. The authors found that ArchIE weighs firstly the number of private SNPs followed by the skew of the distance vector in the underlying logistic regression model. After training their logistic regression predictor using the parameters from a dataset with confirmed Neanderthal introgression in non-African populations, the authors applied this method on Yoruban individuals from the 1000 Genomes Project. They implemented this under the assumption that the predictor is expected to be sensitive to introgression events from populations that shared ancestral population structure with Yorubans. Their results suggest that the archaic ancestry in Yoruba is best explained by admixture with an archaic ghost population more than the possibility of Neanderthal ancestry from back-migration or from admixture with an extant modern human population. In total, about 258 Mb of

introgressed sequences in the Yoruba were recovered using this method in several protein coding regions at high frequency in the population (Table 1). An update to this study earlier this year claimed a further recovery of 482 and 502 Mb of archaic ancestry in Yoruba and Mende populations, respectively and that sub-Saharan populations derive 2-19% of their genetic ancestry from an archaic population that diverged before the split between Neanderthals and modern humans (Durvasula & Sankararaman 2019).

In another recent preprint, Ragsdale and Gravel (2018) explore classic statistics, as well as less familiar tests, to measure introgression. These were used to infer a demographic model with archaic introgression within a likelihood framework in the absence of an archaic reference genome. By implementing this approach of joint statistics on intergenic data from the 1000 Genomes Project, the authors found that the Luhya in Webuye, Kenya and the Yoruba of Ibadan, Nigeria exhibited approximately 6-8 % archaic admixture, respectively. Moreover, this study shows that the commonly used model of human demographic history, derived from single-site allele frequency spectrum (AFS) and corroborated by LD decay curves, tends to fit the real data well but significantly underestimates the levels of LD among rare alleles. They show that by modelling archaic introgression worldwide, including African admixture with an archaic population that split off around 460-540 kya, this discrepancy in the levels of LD among rare alleles is resolved.

Speidel et al. (2019) in a most recent preprint have released a method, Relate, capable of inferring genome-wide genealogies for thousands of samples which they also implemented on African populations from the 1000 Genomes Project. Their results support separate ancient events unique to African populations, in particular an introgression event in the Yoruba with a hominin not closely related to Neanderthals (also diverging before the split between Neanderthals and modern humans). This method might be useful to implement on

other African populations, such as the San and the Mbuti, as more whole genomes for these populations become available.

Caveats with the state-of-the-art

Methods that do not require an archaic reference can be used to exploit modern whole-genome data but still have their caveats such as high false positive rates. There have been but a few studies focusing on Africa in attempts to detect archaic sequences and potentially introgressed regions (Garrigan et al. 2005; Wall et al. 2009; Hammer et al. 2011; Lachance et al. 2012; Mendez et al. 2013; Hsieh et al. 2016; Durvasula & Sankararaman 2018; Ragsdale & Gravel 2018). Previous studies focusing on archaic admixture outside of Africa have consistently used Africa as a proxy population that did not experience "archaic introgression" which has essentially meant "did not experience contact with Eurasian Archaics" such as Neanderthal and Denisova (Green et al. 2010; Reich et al. 2010; Sankararaman et al. 2014; Sankararaman et al. 2016; Prüfer et al. 2017; Durvasula & Sankararaman 2018). Those that have taken up the question of archaic introgression in Africa have had to do so in absence of an African archaic reference. Some of those studies have used LD-based methods such as *S** to search for signals and have equated the identified regions, which display a fairly old time to recent common ancestor (TMRCA) and long haplotype, as evidence of archaic introgression (Lachance et al. 2012). Although TMRCA does not always equate population divergence when considering selection and complex demographic models (Henn et al. 2018). Others have, similarly, run *S** on African hunter-gatherer populations and then chosen top candidates of archaic introgression by comparing against simulations with a set of demographic assumptions (Hsieh et al. 2016) to account for confounding effects on TMRCA due to selection. Nevertheless, both of these studies emphasise the importance to better characterise the nature of admixture in Africa and, more generally, the demographic history

of early African populations.

Early African demography: A complicated history

Our knowledge of what may have occurred since the divergence of modern humans and other archaic forms within Africa is limited both in the fossil record and in availability of modern genomes. Few genetic studies have addressed early human history in Africa despite the amount of fossil and archaeological findings elucidating on the emergence of anatomically modern humans (AMH) (Campana et al. 2013; Schlebusch & Jakobsson 2018).

In particular, East Africa is the most extensively excavated area in the continent, while Central and Western Africa are the least explored (Schlebusch & Jakobsson 2018). Some of the oldest fossils of the genus *Homo* can be found in southern Africa, with a fossil record that begins from about 2 million years ago with pronounced transitional forms ranging from 200-600 kya (Dusseldorp et al. 2013). Forms displaying transitional AMH features have appeared in the record of southern Africa around 100-300 kya and fully AMH emerge here approximately 120 kya. Strikingly, while the fossil record for North Africa has been limited, it is where the oldest fossils meeting the criteria of AMH can be found—particularly the ancient remains from Jebel Irhoud, dated to 300 kya (Hublin et al. 2017). Interestingly in a younger time period of 100-60 kya, North African remains still denote archaic morphological traits despite anatomical modernity (Rightmire 2009).

Fossils dating 160-180 kya in East Africa, Omo Kibish and Herto, have been observed to be fully AMH and have often been used to support East Africa as the birthplace of modern humans (White et al. 2003; McDougall et al. 2005). Although, an East African calvaria from Lukenya Hill in Kenya while regarded *Homo sapiens* demonstrates morphological features better represented in archaic hominins yet dated a recently as ~ 23 kya (Tryon et al. 2015).

Of the notably few and recent assemblages from western and central Africa, some remains display features that are more in line with archaic humans but which are fairly young such as the Ishango site in the eastern part of the Democratic Republic of Congo dated to 20-25 kya and Iwo Eleru skeletal remains from Nigeria dating back to ~13 kya (Harvati et al. 2011). Yet the human fossils from Shum Laka in Cameroon, which date 3-7 kya, are fully AMH (Lavachery 2001). Some have suggested, based on both morphology and the mosaic-like fashion that technological replacement took place from central to western Africa, that this could support a model where archaic admixture took place between surviving archaic populations and modern humans (Scerri 2017; Schlebusch & Jakobsson 2018). As consistent as some of these fossils may be with supporting archaic introgression in Africa, understanding how admixture impacts morphology to better understand the mosaic-like distribution for early *Homo sapiens* across different regions has been strongly overshadowed by the efforts to prove that modern humans originated from Africa (Henn et al. 2018).

Despite the several assemblages of fossils found across Africa, DNA preservation from these samples is often compromised—a great deal owed to climatic obstruction (Campana et al. 2013). Alternatively, recovering demographic history from modern genomes can be challenging with lack of large genomic datasets representing the whole of Africa but especially in the light of the continent's complex deep past. The movements, and subsequent gene flow, from the agricultural expansion and spread of pastoralism have cloaked the ancient variation pertaining to early modern humans (Tishkoff et al. 2009; Montinaro et al. 2017). Consequently, this has complicated demographic inferences that can be recovered regarding ancient African population structure. Although recent ancient samples have provided insight into the relatedness between populations prior to monumental demographic events (Skoglund et al. 2017), these sort of samples with endogenous aDNA remain rare and will most likely accumulate at a slower rate.

Simulating various demographic scenarios has corroborated our understanding of the demographic history of populations (Gravel 2012). In the case of African demographic history, modern genomes, albeit challenging, can still provide a window into the past. As such, there have been genetic studies that simulate important events in early migrations such as the Bantu expansion and pastoralist movements (Li et al. 2014; González-Santos et al. 2015; Marks et al. 2015). This can assist our understanding of African pre-history in order to make more appropriate demographic assumptions about the populations on the continent. Henn et al. (2018) layout distinct models of early modern human origins and corresponding support for them in the fields of morphology, archaeology and genetics. These models can provide guidance in model-testing and exploration of early demographic events. Nevertheless, without more archaeological data and more genomes, both ancient and modern, those explorations might overall be of limited scope and self-fulfilling.

Back to the basics: The landscape of recombination and genetic variation in African populations

In the absence of aDNA from an archaic hominin or ghost population in Africa, we must rely on modern genomes to uncover whether archaic introgression took place. Demographic history alone is not sufficient when using modern genomes as windows into the past. We must also consider biological processes in order to develop tools that aid us in answering questions about our genetic history. Consequently, it is crucial to be aware of the differences in the genomic landscape of different populations. For example, recombination maps have been built for populations with European-ancestry (deCODE, HapMapCEU) (Kong et al. 2010; The HapMap Consortium et al. 2010), West African-ancestry (HapMapYRI) (The HapMap Consortium et al. 2010) and mixed-ancestry (AAmap, AfAdm map) (Hinch et al. 2011; Wegmann et al. 2011). These studies have led to the conclusion that comparatively, West Africans have more recombination hotspots across the genome than Europeans and

therefore crossovers are more evenly distributed leading to shorter LD distances in West Africans (The HapMap Consortium et al. 2010; Hinch et al. 2011). Moreover, these studies have been able to show that recombination events appear to be concentrated at hotspots which correlate with a particular ancestry (The HapMap Consortium et al. 2010; Wegmann et al. 2011). Considering that African populations show the highest levels of genetic diversity in both between and within-population, it is essential to note that the recombination landscape across Africa may differ and therefore LD-based methods developed to search for introgression in non-African populations must be adjusted for these differences. For example, if we wanted to use *S** to explore archaic introgression in West Africans, we must account for the shorter extent of LD and therefore expect shorter introgressed haplotypes than what is seen in Eurasians even if admixture putatively took place at a similar date. Mutation rate too has been shown to differ between populations and still remains without a final consensus (Narasimhan et al. 2017; Ragsdale et al. 2018).

In a similar vein, it is vital to point out that the grand majority of tools that explore human history in high-throughput sequencing use SNPs to answer questions about evolution and variation. While SNPs are undoubtedly useful genetic markers, structural variants (SVs) such as insertions, deletions, duplications, and inversions make up most of the variation between individuals (Sudmant et al. 2015). Those with African-ancestry have been shown to exhibit significant difference in SV profile from other populations which is consistent with what has been observed with SNPs (Simons et al. 2014; Sudmant et al. 2015; Sherman et al. 2019). However, the complexity of SVs present on both the individual and population level can often go unappreciated first due to reference-bias (Sudmant et al. 2015; Sherman et al. 2019) but also because SVs and repetitive elements are not commonly considered when developing tools that measure genetic variation for non-clinical purposes (Santander et al. 2017). This reference bias can make it challenging to observe what other regions of the

genome are shared between certain populations and archaics either because of ancestral structure or introgression (Gardner et al. 2017; Günther & Nettelblad 2018).

When searching for archaic introgression in African populations, both Lachance et al. (2012) and Hsieh et al. (2016) found pronounced depletion of archaic sequence in genic regions suggesting that introgressed loci in hunter-gatherer populations are neutrally evolving remnants. However, a subsequent study (Xu et al. 2017) has found that archaic introgression in Africa contributes to the variation in the human salivary gene *MUC7* consequently affecting the composition of the oral microbiome in modern Africans today. This study also suggested that copy number variation in *MUC7* has rapidly evolved under adaptive forces potentially shaped by pathogenic pressures among primates (Xu et al. 2016).

Recent studies looking at African hunter-gatherer and farmer populations have explored how deleterious genetic variation amongst human populations is affected by changes in population size and gene flow (Lopez et al. 2018). Specifically, hunter-gatherers are efficient in purifying selection despite having a recent population collapse (Simons et al. 2014; Lopez et al. 2018). Long-term selection against archaic introgression in Africa has not been looked at although recent studies have shown that non-African populations demonstrate selection against introgression in regulatory regions more than in protein-coding regions (Petr et al. 2019). A study exploring the genetic relationship between sequenced archaic hominins and Africans showed substantial IBD sharing between Africans (East and West) and Denisovans best explained by interbreeding between the ancestor of humans and other archaic hominins (Povysil & Hochreiter 2016). Whether these short IBD fragments are skewed in their distribution across genic and non-genic regions remains to be confirmed as well if they are putatively introgressed haplotypes under any form of non-neutral selection. However, Durvasula and Sankararaman (2018) have detected a number of regions with archaic ancestry in Yoruba which are in genic regions and that may be under positive

selection (Table 1). Nevertheless, it is important to point out that without ancient African samples to confirm archaic introgression in present-day African populations, signs of adaptive introgression will remain putative.

Overcoming the oversimplification of Africa: Future directions

Detection of archaic introgression with current state-of-the-art methods relies on at least one of three things: (1) an archaic reference, (2) known demographic history, and (3) large sample size. While acquiring an African archaic reference will take time considering the limitations of the current technology and the integrity of available samples, this is a possibility that shows signs of becoming more feasible (Skoglund et al. 2017). In the meantime, efforts should be strengthened to enrich the geographic coverage for genetic data across Africa. More modern African genomes will elucidate the history of different populations across the continent and how they relate to each other in terms of ancient gene flow and structure in light of more recent migration. Simulations without more insight into demographic history, which are used to calibrate whether archaic introgression is present in absence of an archaic reference, can only be as accurate as the demographic assumptions they are modelled upon (Henn et al. 2018). Moreover, caution should be taken when implementing methods where certain biological assumptions are better in line with non-African populations and where extrapolating what has been done in non-African populations to African populations can potentially lead to both false positives and false negatives alike.

In absence of both an archaic reference and clear demographic inferences for Africa, new (Speidel et al. 2019) or less explored (Ragsdale & Gravel 2018) methods might be better suited in answering questions surrounding the nature of archaic introgression in Africa once more modern African genomes become available. Several questions still remain such as: did admixture take place between archaic hominins and modern humans or with the ancestors of

modern humans? Or did ancient admixture take place between existing populations in Africa today and ancient ghost populations? How often and for how long did admixture potentially take place? Lastly, more needs to be investigated regarding putative adaptive introgression in Africans (Xu et al. 2017; Durvasula & Sankararaman 2018) and, in general, evidence of selection for or against archaic ancestry in modern African genomes (Durvasula & Sankararaman 2018). New ancient and modern genetic data will allow us to explore and infer better demographic models so that we may move on to reliably answer these outstanding questions which are crucial to understanding our modern human origins.

Conflict of interest statement

The authors declare no competing interests.

References

Baum LB, Eagon JA. 1967. An inequality with applications to statistical estimation for probabilistic functions of markov processes and to a model for ecology. Bull Am Math Soc.

Besenbacher S, Liu S, Izarzugaza JMG, Grove J, Belling K, Bork-Jensen J, Huang S, Als TD, Li S, Yadav R, et al. 2015. Novel variation and de novo mutation rates in population-wide de novo assembled Danish trios. Nat Commun.

Browning SR, Browning BL, Zhou Y, Tucci S, Akey JM. 2018. Analysis of Human Sequence Data Reveals Two Pulses of Archaic Denisovan Admixture. Cell. 173.

Campana MG, Bower MA, Crabtree PJ. 2013. Ancient DNA for the Archaeologist: The Future of African Research. [place unknown]; [cited 2018 Dec 7]. Available from: https://www.jstor.org/stable/pdf/42641807.pdf

Durvasula A, Sankararaman S. 2018. Recovering signals of ghost archaic admixture in the genomes of present-day Africans. bioRxiv [Internet]. [cited 2018 Oct 1]:285734. Available from: https://www.biorxiv.org/content/early/2018/03/21/285734

Durvasula A, Sankararaman S. 2019. Recovering signals of ghost archaic introgression in African populations. bioRxiv [Internet]. [cited 2019 Apr 30]:285734. Available from: https://www.biorxiv.org/content/10.1101/285734v2

Dusseldorp G, Lombard M, Wurz S. 2013. Pleistocene homo and the updated stone age sequence of South Africa. S Afr J Sci.

Eriksson A, Manica A. 2012. Effect of ancient population structure on the degree of polymorphism shared between modern human populations and ancient hominins. Proc Natl Acad Sci [Internet]. [cited 2017 Oct 23]; 109:13956–13960. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22893688

Fu Q, Li H, Moorjani P, Jay F, Slepchenko SM, Bondarev AA, Johnson PLF, Aximu-Petri A,

Prüfer K, De Filippo C, et al. 2014. Genome sequence of a 45,000-year-old modern human from western Siberia. Nature.

Galway-Witham J, Stringer C. 2018. How did Homo sapiens evolve? Science [Internet]. [cited 2018 Aug 1]; 360:1296–1298. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29930123

Gardner EJ, Lam VK, Harris DN, Chuang NT, Scott EC, Stephen Pittard W, Mills RE, Devine SE. 2017. The mobile element locator tool (MELT): Population-scale mobile element discovery and biology. Genome Res.

Garrigan D, Mobasher Z, Kingan SB, Wilder JA, Hammer MF. 2005. Deep Haplotype Divergence and Long-Range Linkage Disequilibrium at Xp21.1 Provide Evidence That Humans Descend From a Structured Ancestral Population. Genetics [Internet]. [cited 2018 Dec 7]; 170:1849–1856. Available from: http://www.genetics.org/content/170/4/1849

González-Santos MG, Montinaro F, Oosthuizen O, Oosthuizen E, Busby GBJ, Anagnostou P, Destro-Bisol G, Pascali V, Capelli C. 2015. Genome-Wide snp analysis of southern african populations provides new insights into the dispersal of bantu-Speaking groups. Genome Biol Evol.

Gravel S. 2012. Population genetics models of local ancestry. Genetics.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MHY, et al. 2010. A draft sequence of the neandertal genome. Science (80-).

Griffiths RC, Marjoram P. 1997. An ancestral recombination graph. Prog Popul Genet Hum Evol (Minneapolis, MN, 1994).

Günther T, Nettelblad C. 2018. The presence and impact of reference bias on population genomic studies of prehistoric human populations. bioRxiv [Internet]. [cited 2018 Dec 10]:487983. Available from: https://www.biorxiv.org/content/early/2018/12/06/487983

Hajdinjak M, Fu Q, Hübner A, Petr M, Mafessoni F, Grote S, Skoglund P, Narasimham V, Rougier H, Crevecoeur I, et al. 2018. Reconstructing the genetic history of late Neanderthals. Nature. 555:652–656.

Hammer MF, Woerner AE, Mendez FL, Watkins JC, Wall JD. 2011. Genetic evidence for archaic admixture in Africa. Proc Natl Acad Sci.

Harding RM, McVean G. 2004. A structured ancestral population for the evolution of modern humans. Curr Opin Genet Dev.

Harvati K, Stringer C, Grün R, Aubert M, Allsworth-Jones P, Folorunso CA. 2011. The Later Stone Age Calvaria from Iwo Eleru, Nigeria: Morphology and Chronology.Relethford JH, editor. PLoS One [Internet]. [cited 2019 Jan 18]; 6:e24024. Available from: https://dx.plos.org/10.1371/journal.pone.0024024

Henn BM, Steele TE, Weaver TD. 2018. Clarifying distinct models of modern human origins in Africa. Curr Opin Genet Dev [Internet]. [cited 2018 Nov 13]; 53:148–156. Available from: https://www.sciencedirect.com/science/article/pii/S0959437X1730182X?openDownloadIssueModal=true

Hinch AG, Tandon A, Patterson N, Song Y, Rohland N, Palmer CD, Chen GK, Wang K, Buxbaum SG, Akylbekova EL, et al. 2011. The landscape of recombination in African Americans. Nature [Internet]. [cited 2018 Dec 10]; 476:170–175. Available from: http://www.nature.com/articles/nature10336

Hsieh Ping Hsun, Veeramah KR, Lachance J, Tishkoff SA, Wall JD, Hammer MF, Gutenkunst RN. 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.

Hsieh P H, Woerner AE, Wall JD, Lachance J, Tishkoff SA, Gutenkunst RN, Hammer MF.

2016. Model-based analyses of whole-genome data reveal a complex evolutionary history involving archaic introgression in Central African Pygmies (vol 26, pg 291, 2016). Genome Res. 26:717.

Hublin J-J, Ben-Ncer A, Bailey SE, Freidline SE, Neubauer S, Skinner MM, Bergmann I, Le Cabec A, Benazzi S, Harvati K, Gunz P. 2017. New fossils from Jebel Irhoud, Morocco and the pan-African origin of Homo sapiens. Nature.

Jacobs GS, Hudjashov G, Saag L, Kusuma P, Darusallam CC, Lawson DJ, Mondal M, Pagani L, Ricaut F-X, Stoneking M, et al. 2019. Multiple Deeply Divergent Denisovan Ancestries in Papuans. Cell [Internet]. [cited 2019 May 8]; 177:1010-1021.e32. Available from: https://www.sciencedirect.com/science/article/pii/S0092867419302181?via%3Dihub#sec3

Jobling MA, Tyler-Smith C. 2003. The human Y chromosome: An evolutionary marker comes of age. Nat Rev Genet.

Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, et al. 2002. A high-resolution recombination map of the human genome. Nat Genet. 31:241–247.

Kong A, Thorleifsson G, Gudbjartsson DF, Masson G, Sigurdsson A, Jonasdottir AA, Walters GB, Jonasdottir AA, Gylfason A, Kristinsson KT, et al. 2010. Fine-scale recombination rate differences between sexes, populations and individuals. Nature. 467:1099–1103.

Lachance J, Vernot B, Elbers CC, Ferwerda B, Froment A, Bodo JM, Lema G, Fu WQ, Nyambo TB, Rebbeck TR, et al. 2012. Evolutionary History and Adaptation from High-Coverage Whole-Genome Sequences of Diverse African Hunter-Gatherers. Cell. 150:457–469.

Lafferty JD, McCallum A, Pereira FCN. 2001. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In: Proc Eighteenth Int Conf Mach Learn. [place unknown].

Lahr MM, Foley RA. 1998. Towards a theory of modern human origins: geography, demography, and diversity in recent human evolution. Am J Phys Anthropol. 41:137–176.

Lavachery P. 2001. The holocene archaeological sequence of Shum Laka rock shelter (Grassfields, western Cameroon). African Archaeol Rev.

Li S, Schlebusch C, Jakobsson M. 2014. Genetic variation reveals large-scale population expansion and migration during the expansion of Bantu-speaking peoples. Proc R Soc B Biol Sci.

Liang M, Nielsen R. 2014. The lengths of admixture tracts. Genetics.

Lopez M, Kousathanas A, Quach H, Harmant C, Mouguiama-Daouda P, Hombert JM, Froment A, Perry GH, Barreiro LB, Verdu P, et al. 2018. The demographic history and mutational load of African hunter-gatherers and farmers. Nat Ecol Evol.

Lorente-Galdos B, Lao O, Serra-Vidal G, Santpere G, Kuderna LFK, Arauna LR, Fadhlaoui-Zid K, Pimenoff VN, Soodyall H, Zalloua P, et al. 2019. Whole-genome sequence analysis of a Pan African set of samples reveals archaic gene flow from an extinct basal population of modern humans into sub-Saharan populations. Genome Biol [Internet]. [cited 2019 Apr 30]; 20:77. Available from: https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1684-5

Marks SJ, Montinaro F, Levy H, Brisighelli F, Ferri G, Bertoncini S, Batini C, Busby GBJ, Arthur C, Mitchell P, et al. 2015. Static and Moving Frontiers: The Genetic Landscape of Southern African Bantu-Speaking Populations. Mol Biol Evol.

Matise TC, Chen F, Chen W, De La Vega FM, Hansen M, He C, Hyland FCL, Kennedy GC,

Kong X, Murray SS, et al. 2007. A second-generation combined linkage-physical map of the human genome. Genome Res.

McDougall I, Brown FH, Fleagle JG. 2005. Stratigraphic placement and age of modern humans from Kibish, Ethiopia. Nature.

Mendez FL, Krahn T, Schrack B, Krahn AM, Veeramah KR, Woerner AE, Fomine FLM, Bradman N, Thomas MG, Karafet TM, Hammer MF. 2013. An African American paternal lineage adds an extremely ancient root to the human y chromosome phylogenetic tree. Am J Hum Genet.

Mondal M, Bertranpetit J, Lao O. 2019. Approximate Bayesian computation with deep learning supports a third archaic introgression in Asia and Oceania. Nat Commun [Internet]. [cited 2019 May 8]; 10:246. Available from: http://www.nature.com/articles/s41467-018-08089-7

Montinaro F, Busby GBJ, Gonzalez-Santos M, Oosthuitzen O, Oosthuitzen E, Anagnostou P, Destro-Bisol G, Pascali VL, Capelli C. 2017. Complex Ancient Genetic Structure and Cultural Transitions in Southern African Populations. Genetics [Internet]. [cited 2019 May 8]; 205:303–316. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27838627

Narasimhan VM, Rahbari R, Scally A, Wuster A, Mason D, Xue Y, Wright J, Trembath RC, Maher ER, van Heel DA, et al. 2017. Estimating the human mutation rate from autozygous segments reveals population differences in human mutational processes. Nat Commun [Internet]. [cited 2019 Jan 22]; 8:303. Available from: http://www.nature.com/articles/s41467-017-00323-y

Nielsen R, Akey JM, Jakobsson M, Pritchard JK, Tishkoff S, Willerslev E. 2017. Tracing the peopling of the world through genomics. Nature.

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T,

Reich D. 2012. Ancient admixture in human history. Genetics.

Petr M, Pääbo S, Kelso J, Vernot B. 2019. Limits of long-term selection against Neandertal introgression. Proc Natl Acad Sci [Internet]. [cited 2019 Jan 16]; 116:1639–1644. Available from: https://www.pnas.org/content/early/2019/01/14/1814338116

Plagnol V, Wall JD. 2006. Possible Ancestral Structure in Human Populations. PLoS Genet [Internet]. [cited 2017 Oct 9]; 2:e105. Available from: http://dx.plos.org/10.1371/journal.pgen.0020105

Povysil G, Hochreiter S. 2016. IBD sharing between Africans, Neandertals, and Denisovans. Genome Biol Evol.

Prüfer K, Filippo C de, Grote S, Mafessoni F, Korlević P, Hajdinjak M, Vernot B, Skov L, Hsieh P, Peyrégne S, et al. 2017. A high-coverage Neandertal genome from Vindija Cave in Croatia. Science (80-) [Internet]. [cited 2017 Oct 5]:eaao1887. Available from: http://science.sciencemag.org/content/early/2017/10/04/science.aao1887.full

Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G, Sudmant PH, De Filippo C, et al. 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature. 505:43–49.

Ragsdale AP, Gravel S. 2018. Models of archaic admixture and recent history from two-locus statistics. bioRxiv [Internet]. [cited 2018 Dec 10]:489401. Available from: https://www.biorxiv.org/content/early/2018/12/07/489401

Ragsdale AP, Moreau C, Gravel S. 2018. Genomic inference using diffusion models and the allele frequency spectrum. Curr Opin Genet Dev [Internet]. [cited 2019 Jan 22]; 53:140–147. Available from: https://www.sciencedirect.com/science/article/pii/S0959437X18300819

Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs AW, Stenzel U, Johnson PLFF, et al. 2010. Genetic history of an archaic hominin group from

Denisova cave in Siberia. Nature [Internet]. [cited 2017 Oct 23]; 468:1053–1060. Available from: http://www.nature.com/doifinder/10.1038/nature09710

Rightmire GP. 2009. Out of Africa: modern human origins special feature: middle and later Pleistocene hominins in Africa and Southwest Asia. Proc Natl Acad Sci U S A.

Sankararaman S, Mallick S, Dannemann M, Prufer K, Kelso J, Paabo S, Patterson N, Reich D. 2014. The genomic landscape of Neanderthal ancestry in present-day humans. Nature [Internet]. 507:354–357. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24476815

Sankararaman S, Mallick S, Patterson N, Reich D. 2016. The Combined Landscape of Denisovan and Neanderthal Ancestry in Present-Day Humans. Curr Biol [Internet]. 26:1241–1247. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27032491

Sankararaman S, Patterson N, Li H, Pääbo S, Reich D. 2012. The Date of Interbreeding between Neandertals and Modern Humans.Akey JM, editor. PLoS Genet [Internet]. [cited 2018 Oct 19]; 8:e1002947. Available from: https://dx.doi.org/10.1371/journal.pgen.1002947

Santander CG, Gambron P, Marchi E, Karamitros T, Katzourakis A, Magiorkinis G. 2017. STEAK: A specific tool for transposable elements and retrovirus detection in high-throughput sequencing data. Virus Evol [Internet]. 3:1–12. Available from: http://orcid.org/0000-0003-0841-9159

Scally A, Durbin R. 2012. Revising the human mutation rate: implications for understanding human evolution. Nat Rev Genet.

Scerri E. 2017. The Stone Age Archaeology of West Africa. Oxford Res Encycl African Hist [Internet]. [cited 2019 Jan 18]; 1. Available from: http://africanhistory.oxfordre.com/view/10.1093/acrefore/9780190277734.001.0001/acrefore-9780190277734-e-137

Scerri EML, Thomas MG, Manica A, Gunz P, Stock JT, Stringer C, Grove M, Groucutt HS,

Timmermann A, Rightmire GP, et al. 2018. Did Our Species Evolve in Subdivided Populations across Africa, and Why Does It Matter? Trends Ecol Evol [Internet]. [cited 2018 Jul 18]; 0. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30007846

Schlebusch CM, Jakobsson M. 2018. Tales of Human Migration, Admixture, and Selection in Africa. Annu Rev Genomics Hum Genet [Internet]. 19:405–428. Available from: https://www.annualreviews.org/doi/10.1146/annurev-genom-083117-021759

Seguin-Orlando A, Korneliussen TS, Sikora M, Malaspinas AS, Manica A, Moltke I, Albrechtsen A, Ko A, Margaryan A, Moiseyev V, et al. 2014. Genomic structure in Europeans dating back at least 36,200 years. Science (80-).

Sherman RM, Forman J, Antonescu V, Puiu D, Daya M, Rafaels N, Boorgula MP, Chavan S, Vergara C, Ortega VE, et al. 2019. Assembly of a pan-genome from deep sequencing of 910 humans of African descent. Nat Genet. 51:30–35.

Simons YB, Turchin MC, Pritchard JK, Sella G. 2014. The deleterious mutation load is insensitive to recent population history. Nat Genet. 46:220–224.

Skoglund P, Thompson JC, Prendergast ME, Mittnik A, Sirak K, Hajdinjak M, Salie T, Rohland N, Mallick S, Peltzer A, et al. 2017. Reconstructing Prehistoric African Population Structure. Cell [Internet]. [cited 2017 Sep 21]; 171:59-71.e21. Available from: http://www.cell.com/cell/fulltext/S0092-8674(17)31008-5

Skov L, Hui R, Shchur V, Hobolth A, Scally A, Schierup MH, Durbin R. 2018. Detecting archaic introgression using an unadmixed outgroup.Racimo F, editor. PLOS Genet [Internet]. [cited 2018 Oct 1]; 14:e1007641. Available from: http://dx.plos.org/10.1371/journal.pgen.1007641

Slatkin M, Racimo F. 2016. Ancient DNA and human history. Proc Natl Acad Sci.

Speidel L, Forest M, Shi S, Myers S. 2019. A method for genome-wide genealogy estimation

for thousands of samples. bioRxiv [Internet]. [cited 2019 Apr 23]:550558. Available from: http://dx.doi.org/10.1101/550558

Stringer C. 2016. The origin and evolution of homo sapiens. Philos Trans R Soc B Biol Sci. 371.

Stringer CB, Andrews P. 1988. Genetic and fossil evidence for the origin of modern humans. Science (80-).

Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, Zhang Y, Ye K, Jun G, Fritz MHY, et al. 2015. An integrated map of structural variation in 2,504 human genomes. Nature [Internet]. 526. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26432246

The HapMap Consortium, Altshuler D, Gibbs R a, Peltonen L, Dermitzakis E, Schaffner S, Yu F, Bonnen PE, de Bakker PIW, Deloukas P, et al. 2010. Integrating common and rare genetic variation in diverse human populations. Nature.

Theunert C, Slatkin M. 2017. Distinguishing Recent Admixture from Ancestral Population Structure. Genome Biol Evol [Internet]. [cited 2018 Sep 28]; 9:427–437. Available from: https://academic.oup.com/gbe/article/2982377/Distinguishing

Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, Hirbo JB, Awomoyi AA, Bodo JM, Doumbo O, et al. 2009. The genetic structure and history of Africans and African Americans. Science (80-) [Internet]. [cited 2017 Sep 18]; 324:1035–1044. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19407144

Tryon CA, Crevecoeur I, Faith JT, Ekshtain R, Nivens J, Patterson D, Mbua EN, Spoor F. 2015. Late Pleistocene age and archaeological context for the hominin calvaria from GvJm-22 (Lukenya Hill, Kenya). Proc Natl Acad Sci [Internet]. [cited 2019 Jan 18]; 112:2682–2687. Available from: www.pnas.org/cgi/doi/10.1073/pnas.1417909112

Vernot B, Akey JM. 2014. Resurrecting surviving Neandertal lineages from modern human genomes. Science (80-) [Internet]. [cited 2017 Oct 23]; 343:1017–1021. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24476670

Vernot B, Tucci S, Kelso J, Schraiber JG, Wolf AB, Gittelman RM, Dannemann M, Grote S, McCoy RC, Norton H, et al. 2016. Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. Science (80-).

Wall JD, Lohmueller KE, Plagnol V. 2009. Detecting ancient admixture and estimating demographic parameters in multiple human populations. Mol Biol Evol.

Wall JD, Yang MA, Jay F, Kim SK, Durand EY, Stevison LS, Gignoux C, Woerner A, Hammer MF, Slatkin M. 2013. Higher levels of Neanderthal ancestry in east Asians than in Europeans. Genetics.

Wegmann D, Kessner DE, Veeramah KR, Mathias RA, Nicolae DL, Yanek LR, Sun Y V, Torgerson DG, Rafaels N, Mosley T, et al. 2011. Recombination rates in admixed individuals identified by ancestry-based inference. Nat Genet [Internet]. [cited 2018 Dec 10]; 43:847–853. Available from: http://www.nature.com/articles/ng.894

White TD, Asfaw B, DeGusta D, Gilbert H, Richards GD, Suwa G, Howell FC. 2003. Pleistocene Homo sapiens from Middle Awash, Ethiopia. Nature.

Wolpoff MH, Wu XZ, Thorne AG. 1984. Modern Homo sapiens origins: a general theory of hominid evolution involving the fossil evidence from East Asia. In: Orig Mod Humans A World Surv Foss Evid. [place unknown].

Xu D, Pavlidis P, Taskent RO, Alachiotis N, Flanagan C, Degiorgio M, Blekhman R, Ruhl S, Gokcumen O. 2017. Archaic Hominin Introgression in Africa Contributes to Functional Salivary MUC7 Genetic Variation. Mol Biol Evol [Internet]. [cited 2018 Dec 30]; 34:2704–2715. Available from: https://academic.oup.com/mbe/article/34/10/2704/3988100

Xu D, Pavlidis P, Thamadilok S, Redwood E, Fox S, Blekhman R, Ruhl S, Gokcumen O. 2016. Recent evolution of the salivary mucin MUC7. Sci Rep.

Figure and Table Legends

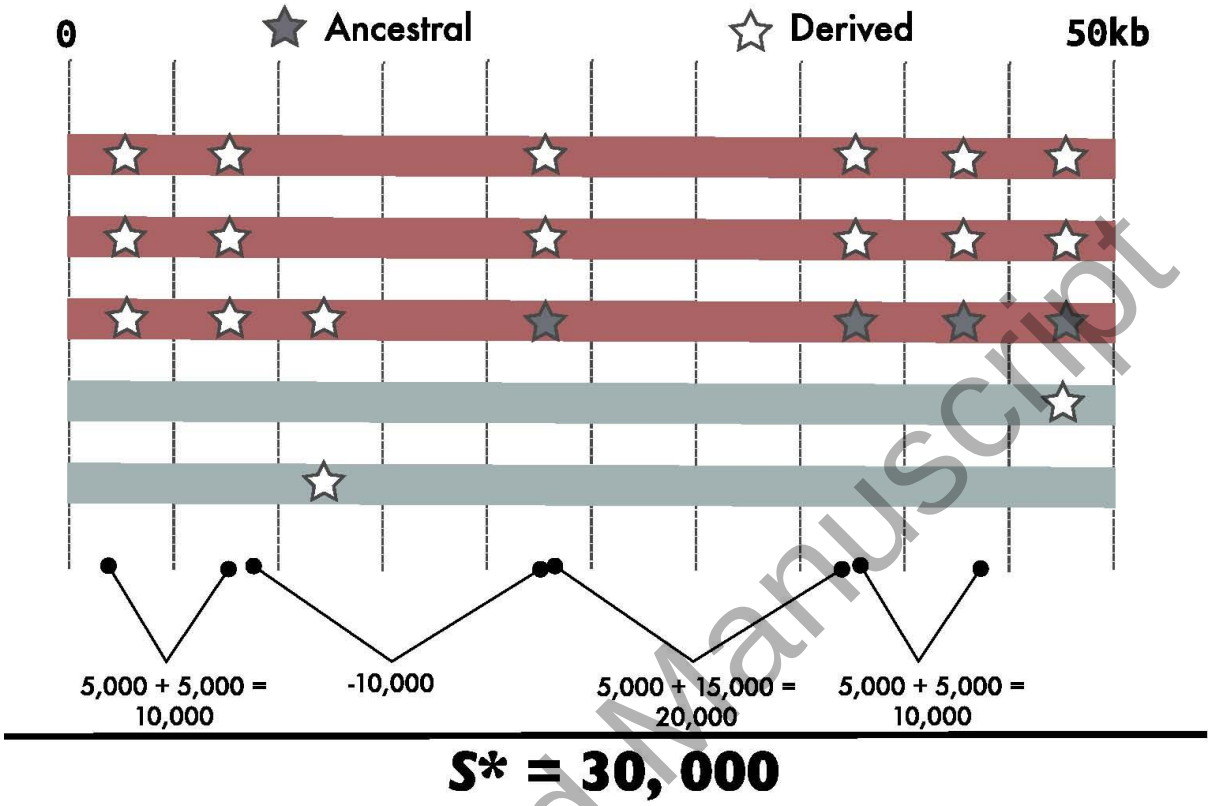

Figure 1. Schematic for *S** scoring. Here we see phased chromosomes and a test population that is represented in red and the outgroup in grey. *S** attempts to search for the most optimal sums of scores for an overall subset of SNPs at a given locus. It rewards fully linked pairs of sites, in other words where two successive SNP positions do not differ, and the increase in that reward is proportional to the distance between the positions. Consequently, *S** values increase with increasing LD within a window. The SNP positions that provide the optimal score are taken note of and can then be used to provide the delimited region that gives that optimal score. The lower bottom of the figure depicts an example of how *S** goes about calculating the most optimal sums of scores. Seven positions within a window of 50,000 bp and only SNPs not present in the outgroup are considered. Sites within the haplotype are labelled as a white star (i.e. ancestral) or a grey star (i.e. derived), those highlighted in red (target) are calculated to give the most optimal solution and that no other set of SNPs gives a higher score. The first 10,000 bp displays higher linkage than the following 10,000 bp where

there is a change in genotype leading to a penalty of -10,000. We can see that in the calculation that follows the positions from about 30,000 bp onwards are rewarded for being linked and outweigh any penalty from before. The scoring function works in rewarding linked sites with a score of 5,000 plus the distance between sites, penalties are given -10,000 for up to 5 mismatches in the above example. Regions with more than 5 mismatches or no linked sites are -∞, or essentially no score. The total *S** score for this example is 30,000.

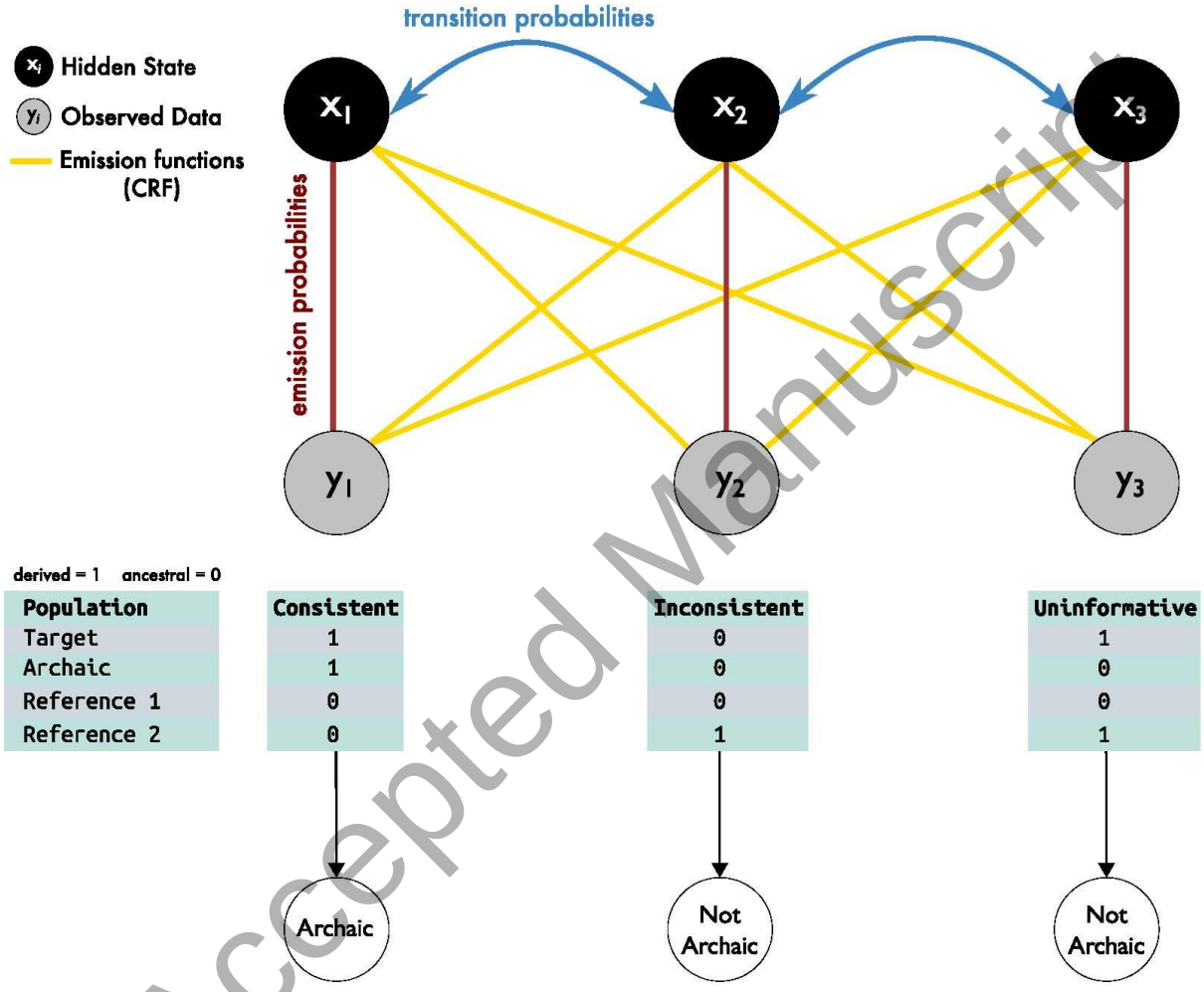

Figure 2. Overview of Hidden Markov model and conditional random field framework. Here is a depiction of how probabilistic models, such as HMM and CRF, can estimate the ancestry (x_i) of a SNP in a genome sequence (i). The possible hidden states for x_i are either introgressed (archaic) or not introgressed (not archaic). In the example, y_i is a matrix composed of individuals from a **target** population, an **archaic** population, and two outgroup populations that act as **references** for modern human variation. These are considered at three

SNP positions (below model). Sites which are consistent with introgression (x_1) are determined by the derived allele of the target population also present in the archaic population but not in the reference. In the case of x_2, the derived allele is only in one of the reference populations suggesting that the derived allele is modern human variation and therefore inconsistent with introgression. In x_3, both target and reference populations share the derived allele but absent in the archaic population which is uninformative. Hidden states (ancestry) are connected to the observed data (y_i) through emission probabilities (red) for HMM or emission functions (vertical red and diagonal yellow) for CRF. These emission functions in a CRF model can be used to evaluate whether a site is consistent or inconsistent with introgression, like in HMM. Additionally, they can also score whether a haplotype overall is closer to the archaic sequence than to the reference haplotypes. Horizontal connections between x_i and x_{i+1} denote transition probabilities (HMM) or transition functions (CRF) which model linkage between ancestral states along the genome. These parameters depend on the recombination, admixture proportion, and time of admixture.

Table 1. Archaic introgressed candidate loci in African populations.

Putatively introgressed regions	**Type**	**Population**	**Study**
RP11-286M16 (chr 1)	lincRNA	Yoruba	(Durvasula & Sankararaman 2018)
RN7SKP160 (chr 1)	Pseudogene	Yoruba	(Durvasula & Sankararaman 2018)
4qMB179	Non-coding	Biaka	(Hammer et al. 2011)
KCNIP4 (chr 4)	Protein coding	Yoruba	(Durvasula & Sankararaman 2018)
XRCC4 (chr 5)	Protein coding	Yoruba	(Plagnol & Wall 2006; Wall et al. 2009)
MTFR2 (chr 6)	Protein coding	Yoruba	(Durvasula & Sankararaman 2018)
TRPS1 (chr 8)	Protein coding	Yoruba	(Durvasula & Sankararaman 2018)
13qMB107	Non-coding	San	(Hammer et al. 2011)
TJP1 (chr 15)	Protein coding	Yoruba	(Plagnol & Wall 2006; Wall et al. 2009)
DUT (chr 15)	Protein coding	Yoruba	(Plagnol & Wall 2006; Wall

			et al. 2009)
HSD17B2 (chr 16)	Protein coding	Yoruba	(Durvasula & Sankararaman 2018)
KRT18P61 (chr 17)	Pseudogene	Yoruba	(Durvasula & Sankararaman 2018)
NF1 (chr 17)	Protein coding	Yoruba	(Durvasula & Sankararaman 2018)
RP1115E18 (chr 17)	Pseudogene	Yoruba	(Durvasula & Sankararaman 2018)
18qMB60	Non-coding	Biaka	(Hammer et al. 2011)
MIR125B2 (chr 21)	miRNA	Yoruba	(Durvasula & Sankararaman 2018)
Xp21.1	Non-coding	Mbuti	(Garrigan et al. 2005)
A00 (chr Y)	Sex chromosome	African-American & Mbo	(Mendez et al. 2013)
Top 350 candidates[†] (Across chr 1-22)	Unknown (genic depleted)	Sandawe, Western Pygmy, Hadza	(Lachance et al. 2012)
Distinct 265 candidates* (Across chr 1-22)	Genic & non-genic	Biaka and Baka	(Hsieh et al. 2016)