

# A weakly structured stem for human origins in Africa

Aaron P. Ragsdale<sup>1</sup>, Timothy D. Weaver<sup>2</sup>, Elizabeth G. Atkinson<sup>3</sup>, Eileen Hoal<sup>4</sup>, Marlo Möller<sup>4</sup>, Brenna M. Henn<sup>2,6,\*</sup>, and Simon Gravel<sup>7,\*\*</sup>

<sup>1</sup>Department of Integrative Biology, University of Wisconsin, Madison, WI, USA

<sup>2</sup>Department of Anthropology, University of California, Davis, Davis, CA, USA

<sup>3</sup>FIXME:

<sup>4</sup>DSI-NRF Centre of Excellence for Biomedical Tuberculosis Research; South African Medical Research Council Centre for Tuberculosis Research; Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

<sup>6</sup>UC Davis Genome Center, University of California, Davis, Davis, CA, USA

<sup>7</sup>Department of Human Genetics, McGill University, Montreal, QC, Canada

\*bmhenn@ucdavis.edu

\*\*simon.gravel@mcgill.ca

January 21, 2022

## Abstract

While a modern human origin within Africa is now broadly accepted, considerable uncertainty surrounds specific models of divergence and migration across the continent. Progress is hampered by a paucity of fossil and genomic data, as well as variability in dating. Here we use linkage disequilibrium and diversity-based statistics, optimized for rapid, complex demographic inference to discriminate among such models. We infer detailed demographic models for populations across Africa, including representatives from eastern and western groups, as well as 44 newly whole-genome sequenced individuals from the Nama (Kho-San). Despite the complexity of African population history, present-day population structure dates back to Marine Isotope Stage (MIS) 5. The earliest population divergence among contemporary populations occurs 120-135kya, between the Kho-San and other groups. Prior to the divergence of contemporary African groups, we infer long-lasting structure between two or more weakly differentiated ancestral *Homo* spp. populations connected by gene flow over hundreds of thousands of years (i.e. a weakly structured stem). We find that weakly-structured-stem models provide more likely explanations of polymorphism that had previously been attributed to contributions from archaic hominins in Africa. In contrast to models with archaic introgression, we predict that fossil remains from coexisting ancestral populations should be morphologically similar. Despite genetic similarity between these populations, an inferred **SG: 4 to 10%** of random genetic differences among [modern human populations] can be attributed to genetic drift between stem populations. We show that model misspecification explains variation in previous divergence time estimates and argue that studying a suite of models is key to robust inferences about deep history.

## Introduction

Archaeological sites from the Middle Stone Age (approx. 300kya-40kya) are widely distributed across Africa, and are particularly well represented in the northern, eastern and southern parts of the continent. Similarly, fossil crania such as those from the sites of Jebel Irhoud<sup>1</sup>, Herto<sup>2</sup> and Klasies River<sup>3</sup> demonstrate that anatomically derived *Homo sapiens* features were also present across the continent

during this period. It has been difficult to reconcile these lines of evidence with evidence from genomics, which have suggested a predominantly tree-like model of recent population divergence from a single ancestral population. It is unclear whether fossil specimens and archaeological sites represent populations which contributed to our ancestors as population precedents, or were local “dead-ends” from which contemporary *Homo sapiens* do not descend. Recently, synthetic attempts to reconcile genetic and paleoanthropological data include proposals for a Pan-African origin of *Homo sapiens* by which populations in many regions of the continent contributed to the formation of *Homo sapiens* beginning at least 300kya<sup>4,5,6</sup>.

Genetic models have been hampered in their contribution to this discussion because they primarily assume (or, at least, have been tested under) a tree-like model of isolation-with-migration. Alternative theoretical scenarios have been proposed, such as stepping stone models<sup>7</sup> or population coalescence and fragmentation<sup>6</sup>. These approaches are more challenging to interpret and fit to data. However, new population genetic inference tools now allow for inference involving tens to hundreds of genomes from multiple populations and greater complexity<sup>8,9,10</sup>. Inspired by evidence for Neanderthal admixture with modern humans in Eurasia, several recent articles have shown that introducing an archaic ghost population contributing to African populations in the period surrounding the Out-of-Africa migration event substantially improves the description of genetic data relative to single-origin models<sup>11,12,13,14,9,15,16,15</sup>. This has driven speculation about the geographic range of this ghost population, possible links to specific archaic remains, and the possibility of finding ancient DNA evidence (e.g., Hsieh et al.<sup>13</sup>). However, these prior articles share two weaknesses. First, they only contrast a single-origin model with an archaic admixture model, leaving out other plausible models (Figure S2 and Henn et al.<sup>17</sup>). Second, they focus on a small subset of African diversity, either because of small sample sizes (2-5 genomes) or because they rely on 1000 Genomes data which only recruited populations of recent West African ancestry. While ancient DNA from Eurasia has helped us understand early human history outside of Africa, there is no comparably ancient DNA to elucidate early history in Africa.

Here, we therefore aim to discriminate among a broader set of demographic models by interrogating the genomes of contemporary populations. We take as our starting point 4 classes of models (single population expansion, single population expansion with regional persistence, archaic admixture, and multi-regional evolution, Figure S2), using 290 genomes from southern, eastern, and western Africa as well as Eurasia. By including geographically and genetically diverse populations across Africa, we infer demographic models that explain more aspects of genetic diversity in more populations than previously reported. These analyses confirm the inadequacy of tree-like models and provide an opportunity to directly evaluate a wide range of alternative models.

## Results

We inferred detailed demographic histories using 4x-8x whole-genome sequencing data for four diverse African populations, comprising the Nama (Khoen-San from South Africa, newly presented here), Mende (from Sierra Leone, MSL from the Phase 3 1000 Genomes Project<sup>18</sup>), Gumuz (recent descendants of a hunter-gatherer group from Ethiopia<sup>19,20</sup>), and Eastern African agriculturalists (Amhara and Oromo from Ethiopia<sup>19</sup>). The Amhara and Oromo populations, despite speaking distinct Afro-Asiatic languages, are highly genetically similar<sup>21,20</sup> and thus the two groups were combined for a larger sample size (Figure 1). We also included the British (GBR) from the 1000 Genomes Project **SG: phase 3** in our demographic models as a representative source of back-to-Africa gene flow and recent colonial admixture in South Africa. Finally, we used a high-coverage ancient Neanderthal genome from Vindija Cave, Croatia<sup>22</sup> to account for archaic gene flow from Neanderthals in non-Africans and gauge the relative time depth of divergence, assuming Neanderthals diverged 550kya from a common stem. For each population, we computed low-order allele frequency and linkage disequilibrium (LD) statistics that are well suited for both low- and high-coverage genomes<sup>9,23</sup>. Using a maximum-likelihood inference framework, we then fit to these statistics a family of parameterized demographic models that involve population splits, size changes, continuous and variable migration rates, and punctuated admixture events, to learn about the nature of population structure over the past million years.

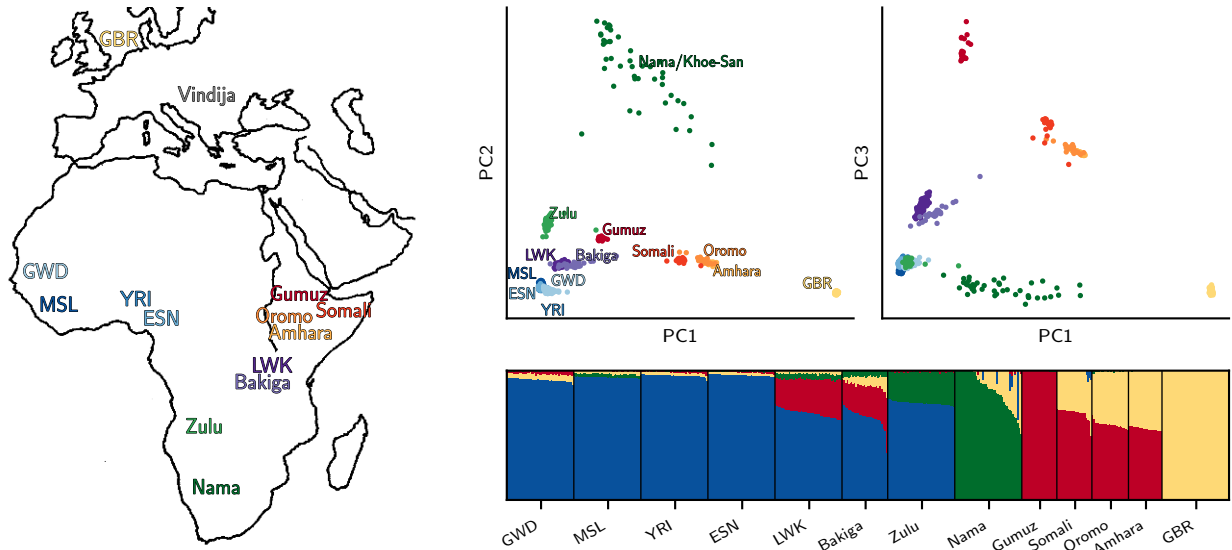


Figure 1: **Geography and genetic diversity within Africa.** The combined 1000 Genomes and AGRP datasets include populations from West, East, and South Africa. Both PCA (B, C) and ADMIXTURE (D) illustrate the well-characterized high levels of genetic diversity and signatures of recent migration and gene flow within Africa. We built parameterized models using the newly-sequenced Nama, Mende from Sierra Leone, Gumuz, Oromo and Amhara from Ethiopia, and the British and Vindija Neanderthal individual (bold fonts). **APR: work on this caption**

## A Late Middle Stone Age common ancestry for contemporary humans

We began with a model of geographic expansion from a single ancestral, unstructured source followed by migration between populations, without allowing for contribution from an archaic African lineage or population structure prior to the expansion (Figure S2A). As expected<sup>9</sup>, this first model was a poor fit to the data qualitatively (Figure S4) and quantitatively (log-likelihood ( $LL$ )  $\approx -189,400$ , Table S1). We next explored a suite of models in which population structure predates the differentiation of contemporary groups, including models allowing for ancestral reticulation (Figure S2B), archaic admixture (Figure S2C), and African multi-regionalism (Figure S2D).

Regardless of the model choice for early epochs, inference of human demographic history in the last 150kya was remarkably robust. The earliest divergence among contemporary human populations differentiates the southern African Nama from other African groups between 110–135kya, with low to moderate levels of subsequent gene flow (Table1). In none of the high-likelihood models which we explored did the divergence between Nama and other populations exceed  $\sim 140$ kya. We conclude that, regardless of model specification, geographic patterns of contemporary *Homo sapiens* population structure date back to the Late Middle Stone Age in Africa, likely arising during MIS 5. **SG: Alt formulation: Thus all models suggest that broad geographic patterns of contemporary population structure date back to MIS 5** Although we find evidence for earlier population structure in Africa (see below), present-day populations cannot be easily mapped onto these more ancient stem groups as only a small proportion of drift between modern populations can be attributed to drift between stems (Figures 3 and S10–S13).

Given this consistency in inferred recent history and the numerical challenge of optimizing a large number of parameters, we fixed several parameters related to recent population history so as to focus on more ancient events (Supp. Methods). The time of divergence of Western and Eastern African populations was set to 60kya, just prior to the split of Eurasians and East Africans 50kya. All models predict relatively high subsequent gene flow between the two regions ( $m \approx 2 \times 10^{-4}$ ). Admixture from Neanderthals to Europeans directly following the out-of-Africa migration was set to 1.5% at 45kya (Supp. Methods). Back-to-Africa gene flow at the beginning of the Holocene primarily **SG: Here it is not clear**

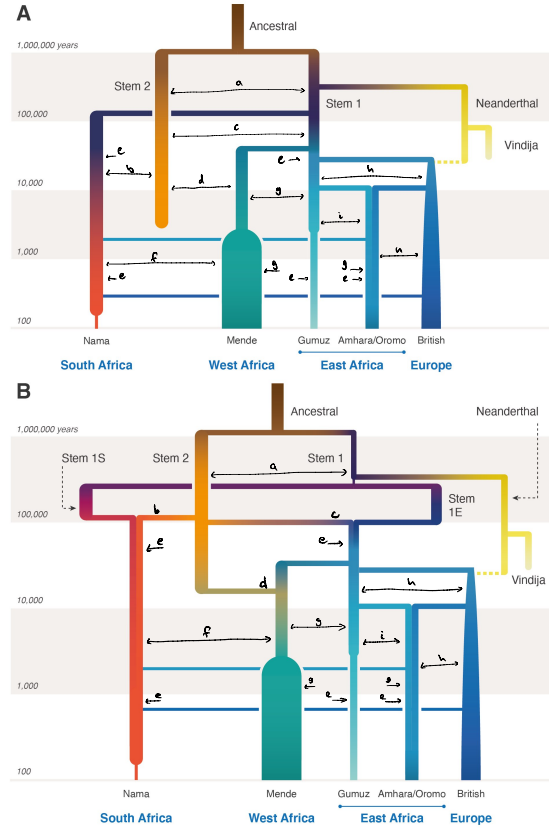


Figure 2: **Two best fit models with a weakly structured stem.** In both the the parameterizations of early population structure, continuous migration (A) and fragmentation with mergers (B), models that include ongoing migration between stem populations outperform those in which stem populations are isolated. Most populations are connected by continuous, reciprocal migration that are indicated by double-headed arrows (labels matched to migration rates and divergence times in Table 1), and which lasts for the duration of concurrence of contemporaneous populations. The merger-with-stem-migration model (B) outperformed the continuous-migration model (A),  $LL = -102,600$  vs  $-115,500$ , respectively, and both models provide a much better fit to diversity and LD statistics than a single-origin model of recent expansion.

whether we are still talking about parameters that we fixed. I think we are also missing a clear description of the best-fit model. That *\*could\** go after the description of the model uncertainty affected the ancestors of the Ethiopian agricultural populations, comprising over half of their genetic ancestry, estimated to be 64–65%. The past 5,000 years also saw major demographic changes, including strong population growth for Western Africans as they specialized in yam and oil palm agriculture (estimated 3-fold growth). We observe significant gene flow from the Amhara and Oromo into the Nama, a signal which is likely a proxy for the movement of Eastern African caprid and cattle pastoralists<sup>24,25</sup>, here estimated to constitute a 25% ancestry contribution 2,000 ya. Colonial period admixture from Europeans into the Nama was estimated at 15%, similar to proportions inferred by ADMIXTURE<sup>26</sup> (Figure 1).

## Deep but connected population structure within Africa

To account for population structure prior to 135kya, we allowed SG: alt: three of our four models allowed for two or more “stem” populations which could diverge either before or after the Neanderthal split. We considered models both with and without migration between these stem populations, and in both cases we tested two different types of gene exchange during the expansion phase. 1) One of the stem popula-

| Likelihood              | Label | Population Pair         | Divergence Time (kya) | Migration rate per generation | Migration duration (ky) |
|-------------------------|-------|-------------------------|-----------------------|-------------------------------|-------------------------|
| <b>Continuous Model</b> |       |                         |                       |                               |                         |
| LL= -115,500            | a     | Stem 1, Stem 2          | 1,163                 | 6.43e-5                       | 1,028                   |
|                         | b     | Stem 2, Nama            | NA                    | 5.82e-5                       | 130                     |
|                         | c, d  | Stem 2, Other Africans* | NA                    | 3.10e-5, <b>1.64e-4</b>       | 130, 55                 |
|                         | e, f  | Nama, Other Africans*   | 135                   | 4.1e-5, 9.8e-6                | 135, 60                 |
|                         | g     | Mende, East Africans    | 60                    | <b>2.14e-4</b>                | 60                      |
|                         | h     | East Africans, British  | 50                    | 4.17e-5                       | 50                      |
|                         | i     | Gumuz, Amhara/Oromo     | 12                    | <b>3.36e-4</b>                | 12                      |
| <b>Merger Model</b>     |       |                         |                       |                               |                         |
| LL= -102,600            | a     | Stem 1, Stem 2          | 1,442                 | <b>1.16e-4</b>                | 963                     |
|                         | –     | Stem 1S, Stem 1E        | 479                   | 0 (fixed)                     | –                       |
|                         | b     | Stem 2 to Nama          | 119                   | <b>0.71</b>                   | pulse                   |
|                         | c     | Stem 2 to Stem 1E       | 98                    | <b>0.50</b>                   | pulse                   |
|                         | d     | Stem 2 to Mende         | 25                    | <b>0.18</b>                   | pulse                   |
|                         | e, f  | Nama, Other Africans*   | 119                   | 4.4e-5, 7.1e-6                | 119, 60                 |
|                         | g     | Mende, East Africans*   | 60                    | <b>1.98e-4</b>                | 60                      |
|                         | h     | East Africans, British  | 50                    | 3.87e-5                       | 50                      |
|                         | i     | Gumuz, Amhara/Oromo     | 12                    | <b>3.59e-4</b>                | 12                      |

Table 1: **Migration and divergence parameters from best fit models.** Labeled migration rates correspond to symmetric continuous migration bands shown in Figure 2. Both the continuous-migration and merger models inferred a relatively deep split of human stem branches, though they were connected by ongoing migration that maintained their genetic similarity. \*: The ancestors of the Nama diverged from other African groups ~120–135kya, which later branch into West and East African groups 60kya. Migration rates and durations are shown for 1) East Africans and their ancestors, and 2) the Mende, respectively.

tion expands (splits into contemporary populations), and the other stem population(s) has continuous symmetric migration with those populations; or 2) one or more of the stem populations expands, with instantaneous pulse (or “merger”) events from the other stem population, so that diversifying branches are formed through the **SG: This sentence is not clear to me. Do we mean: In other words, the more recent populations are formed by mergers of multiple ancestral populations** coalescence of multiple ancestral populations. Depending on parameter values, this scenario encompasses archaic introgression and fragmentation-and-coalescence models.

Allowing for continuous migration between the stem populations substantially improved the fits relative to zero migration ( $LL \approx -102,600$  vs.  $-107,700$  in the merger scenario and  $LL \approx -115,500$  vs.  $-126,600$  with continuous migration). With continuous migration between stems, population structure extends back to 1.1–1.4Mya (Table 1). Migration between the stems in these models is moderate, with a fraction of migrant lineages each generation estimated as  $m = 6.43 \times 10^{-5} - 1.16 \times 10^{-4}$ . For comparison, this roughly compares to inferred migration rates between connected contemporary populations over the past 50kya (Table 1). This ongoing (or at least, periodic) gene flow qualitatively distinguishes these models from previously proposed archaic admixture models (Figure S2C) as the early branches remain closely related (Figure 3).

Bootstrap-based confidence intervals are relatively narrow within each specified model (Tables S1–S5), reflecting an informative statistical approach. As expected, model assumptions have a larger impact on parameter estimates (and thus, uncertainty). To convey model uncertainty, we highlight features of two inferred models with high likelihoods, the “multiple-merger” and the “continuous-migration” model that both allow migration between stem branches. These differ primarily in the timing of the early divergence of stem populations and their relative  $N_e$  (Figure 2). The two models also differ in the mode of divergence during the Middle Stone Age.

**SG: I don’t love this new structure because it breaks the model description in two parts separated by the bootstrap. Maybe we have to, but I would have preferred a unified description of the models. For example, we could start this section by the bootstrap discussion (“In all models, likelihood differences are large and bootstrap confidence interval narrow within each specified model ... We therefore highlight two models”). Then we describe model 1, describe model 2, and contrast the two. )** Under the continuous-migration model, one of the two stems diverges into lineages leading to contemporary populations in western, southern and eastern Africa, and the other (Stem 2) contributes variable ancestry to those populations. This migration from Stem 2 is highest with the Mende ( $m = 1.6 \times 10^{-4}$ ) compared to

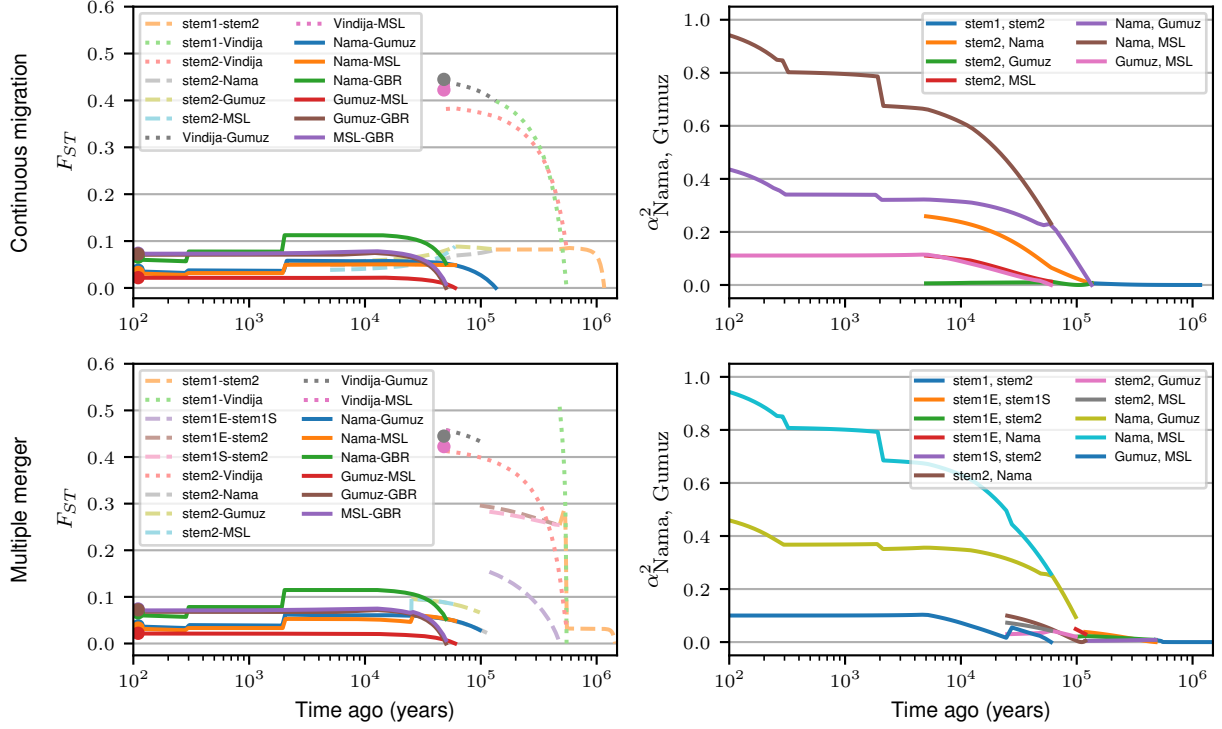


Figure 3: **Predicted population structure over time.** From the best fit models of our two parameterizations (top row: continuous migration, bottom row: multiple merger), we computed predictions of population structure between contemporaneous populations over time ( $F_{ST}$  SG: not sure I want  $F_{ST}$  here, vs  $\pi$ ) and the proportion of drift between pairs of sampled populations that aligns with drift between populations in the past ( $\alpha^2$ ). SG: point to supplement, since nonstandard

the Nama and East African populations ( $m = 5.8 \times 10^{-5}$  and  $3.1 \times 10^{-5}$ , respectively), with migration allowed to occur until the relatively recently demographic events 5kya. A sampled lineage from the Nama, Mende, and Gumuz have probabilities of being in Stem 2 at the time of Stem 1 expansion (135ka) of approximately 0.145, 0.2, and 0.13, respectively, though these probabilities change over time, precluding the notion of a fixed admixture proportion.

In contrast, under the multiple-merger model, stem populations merge with varying proportions to form the different contemporary groups. We observe a sharp bottleneck in Stem 1 down to  $N_e = 117$  after the split of the Neanderthal branch. This represents the lower bound allowed in our optimization (i.e. 1% of the ancestral  $N_e$ ), although the size of this bottleneck is poorly constrained ( $\sigma_{N_e} = 838$ ). After a long period of exchange with Stem 2, Stem 1 then fractures into “Stem 1E” and “Stem 1S” 479kya. The timing of this divergence was also poorly constrained ( $\sigma_T = 166$ kya). These populations evolve independently until approximately 119kya when Stem 1S and Stem 2 combine to form the ancestors of the Nama, with proportions 29%, 71% respectively. Similarly, Stem 1E and Stem 2 combine in equal proportions (50% each) to form the ancestors of the Western Africans and Eastern Africans (and thus also all individuals who later disperse during the Out of Africa event). Finally, the Mende receive a large additional pulse of gene flow from Stem 2, replacing 18% of their population 25kya. The later Stem 2 contribution to the Western African Mende resulted in better model fits (LL). This may indicate that an ancestral Stem 2 population occupied Western or Central Africa, broadly speaking. The differing proportions in the Nama and Eastern Africans may also indicate geographic separation of Stem 1S in southern Africa and Stem 1E in eastern Africa.

## Reconciling multiple lines of genetic evidence

Previous studies have found support for archaic admixture in Africa using two-locus statistics<sup>13,9</sup>, the conditional SFS (cSFS)<sup>15</sup>, and reconstruction of gene genealogies<sup>10</sup>. However, none of these studies considered a weakly structured stem. We validated our inferred models with additional independent approaches. We find that the observed cSFS (conditioned on the derived allele being carried in the Neanderthal sample) is very well-described by the merger model (Figures 4A-C and S14-S17), even though this statistic was not used in the fit. Both models outperform archaic models fit directly to the cSFS (for example, compare with Figure 1 in Durvasula and Sankararaman<sup>15</sup>).

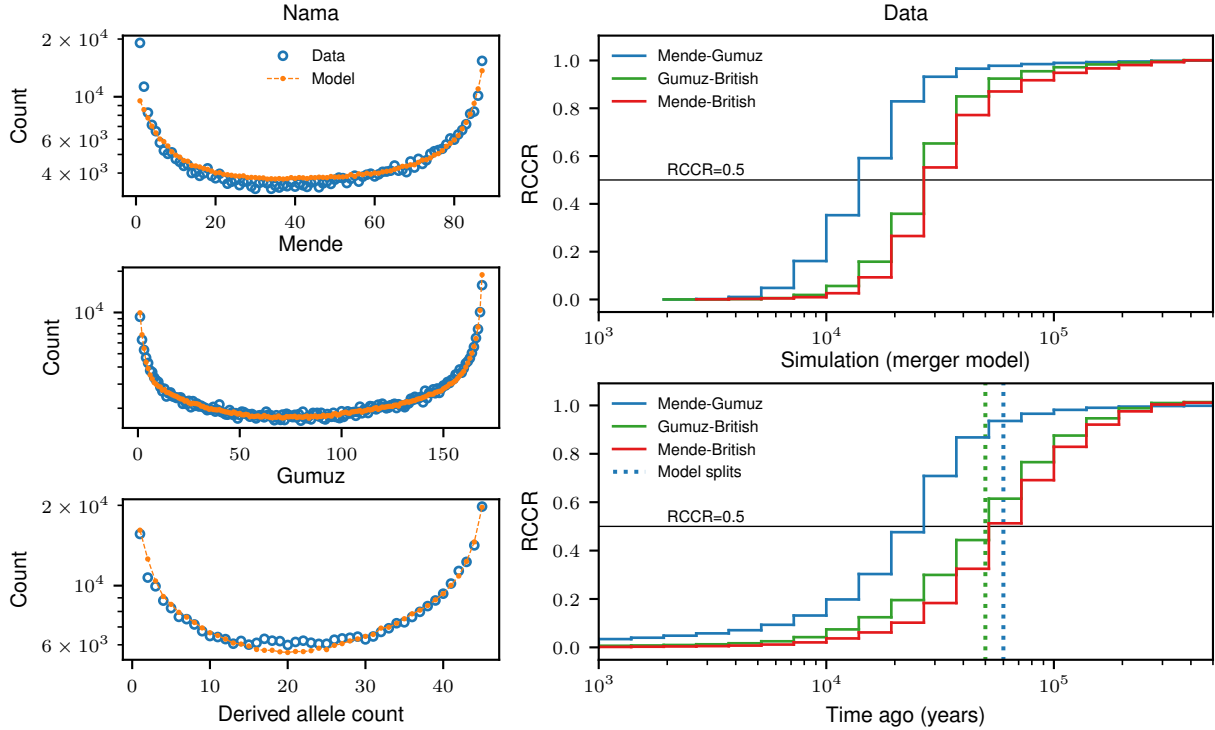


Figure 4: **Model validation using independent statistics.** (A–C) Using our best fit models to the LD and pairwise diversity statistics, we simulated expected conditional site-frequency-spectra (cSFS) and compared to the observed cSFS from the data. Our inferred models provide a good fit to the data, even though these were not used in our inference. Across the three populations, ancestral state misidentification was consistently inferred to be 1.5–1.7% for intergenic loci (Supp. Methods). (D, E) We used Relate<sup>10</sup> to reconstruct genome-wide gene genealogies, which we used to estimate coalescence rate trajectories and and cross-coalescence rates between pairs of populations. While coalescence rate distributions are informative statistics about past evolutionary processes, interpretation can be hindered by migration and population structure, and translating relative cross-coalescence rate curves (RCCR) into population divergence times is especially prone to misinterpretation. For example, the Mende-Gumuz comparison shows a more recent increased RCCR than either population with the British, a pattern that is recapitulated under our best-fit model, even though the Mende-Gumuz split occurs prior to the Gumuz-British split.

We used Relate<sup>10</sup> to infer the distribution of coalescence rates over time, which are commonly interpreted as effective population size histories. Trajectories of  $N_e$  from prior studies strongly support a bottleneck in Out of Africa populations. African populations also dip in effective size around 100–200kya, although this decrease is more modest than the OOA event<sup>27</sup>. One possible explanation is that ancestral population structure during the Middle Stone Age inflated  $N_e$  and subsequent gene flow among



ancestral Africans resulted in a more recent decreased  $N_e$ <sup>28</sup>. We simulated genomic data under four inferred demographic models to compare with observed coalescence rate histories. All models, including the single-origin model, generate multi-modal curves indicating that the ancestral increase in  $N_e$  between 100kya-1Mya is a general feature of the coalescent rate in humans, although none of the models perfectly match the exact timing of the early fluctuations in  $N_e$  observed in the data. Each simulated model also underestimated the level of population growth during the Holocene, indicating additional free parameters are necessary to refine these growth rates.

Relative cross-coalescence rates (rCCR) have recently been used to estimate divergence between a pair of populations, as measured by the rate of coalescence between two groups divided by the mean within population coalescence. Simulations of rCCR accuracy, however, focus on a ‘clean split’ between populations whereby groups diverge without subsequent gene flow. Published estimates of the earliest human divergences with rCCR, which range from 150kya-100kya<sup>29</sup>, may be significantly biased when compared to more complex models with gene flow as inferred here. We find that midpoint estimates of rCCR are poor estimates for population divergence, often underestimating divergence time by 50% or greater (e.g., Mende vs. Gumuz ~15kya compared to a true divergence of 60kya), and recent migration can lead to the misordering of divergence events (Figure 4E). We suggest that rCCR analyses which do not fit multiple parameters including gene flow should be interpreted with caution.

## Discussion

Any attempt at building detailed models of human history is subject to model misspecification. This is true of earlier studies, which often assumed that data inconsistent with a single origin model must be explained by archaic admixture. This is also true of this study. While it remains prohibitive to fully explore the space of plausible models of early human population structure, we sought to capture model uncertainty by exploring multiple parameterizations of early history. **BMH: we use this phrase a lot. human or hominin? Tim?** The best-fit models presented here include reticulation and migration between early human populations rather than archaic admixture from long-isolated branches. We cannot rule out that more complex models involving additional stems, or hybrid models including both weak structure and archaic admixture may better explain the data. Because parameters related to the split time, migration rates, and relative sizes of the early stems were variable across models, reflecting a degree of confounding among these parameters, we refrained from introducing additional branches (and therefore additional parameters) during that period. **SG: Some statement about hitting the boundary** Rather than interpreting the two stems as representing well-defined and stable populations over hundreds of thousands of years, we interpret the weakly structured stem as consistent, e.g., with a coalescence and fragmentation model<sup>6</sup>. Models including additional diversity within Africa, and eventual ancient DNA samples from Africa, could further distinguish the archaic admixture model from the weakly-structured-stem model.

## The Middle Stone Age in Africa

By contrast, our inferred models paint a more consistent picture of the late Middle Stone Age as a critical period of change, assuming that estimates from the recombination clock accurately relate to geological chronologies (Supporting Information). During the Middle Stone Age, the Multiple Merger indicates three major stem lineages in Africa, tentatively assigned to southern (Stem 1S), eastern (Stem 1E) and western/central Africa (Stem 2). While the length of isolation among the stems is variable across model fits, models with this period of divergence, isolation and then a merger event (i.e. a “reticulation”) out-performed models with bifurcating divergence and continuous gene flow.

A population reticulation involves multiple stems contributing genetically to the formation of a group. One way in which this can happen is through the geographic expansion of one or both stems. For example, if during MIS 5, either Stem 1S from southern Africa moved northward thereby encountering the Stem 2, or Stem 2 moved from central/western Africa southward into Stem 1S – then we could observe disproportionate ancestry contributions from different stems in modern groups. We observed two merger events. The first, between Stem 1S and Stem 2, results in the formation of an ancestral Khoe-San population 120kya. The second 100kya between Stem 1E and Stem 2, results in the formation of the ancestors of East/West Africans as well as later “Out of Africans”. The rapid rise in sea levels and increased precipitation during MIS 5e, following a glacial period of aridity across Africa<sup>30</sup>, might have triggered migration inland away from the coasts, as has been suggested for the Paleo-Agulhas plain<sup>31</sup>.



Following these merger events, the stems subsequently fracture into subpopulations which then appear to persist over the past 120kya. These subpopulations can be linked to modern-day groups despite subsequent gene flow across the continent; for example, a genetic lineage sampled in the Nama has an XX probability of being in the ancestral ‘southern’ subpopulation 50kya versus XX probability of being in the ‘eastern’ subpopulation (Figure XX). We also find that Stem 2 continued to contribute to western Africans during the Last Glacial Period, indicative that this gene flow likely occurred in western/central Africa (Table 1).

## Contrasting archaic admixture and a weakly structured stem

Evidence for archaic admixture in Eurasia has bolstered the plausibility of archaic admixture having also occurred in Africa. For this reason, previous work has focused on archaic admixture to explain patterns of polymorphism inconsistent with a single origin model. Here, we have shown that weakly-structured-stem models better capture these patterns. They also help explain an ecological riddle posed by the archaic admixture model. Neanderthal populations were separated from early *Homo sapiens* by thousands of kilometers and continental geographic barriers. By contrast, an archaic hominin population in Africa would need to have stayed in relative reproductive isolation from the human lineage over hundreds of thousands of years despite closer geographic proximity and reproductive compatibility. The weakly-structured-stem model resolves this ecological riddle by accommodating continuous or recurrent contact between two or more groups present in Africa.

There is evidence for both deleterious and adaptive archaic-derived alleles in modern genomes in the form of a depletion of Neanderthal ancestry in regulatory regions<sup>32</sup> or an increased frequency of archaic-related haplotypes such as at *EPAS1* among Tibetans (e.g., Zhang et al.<sup>33</sup>). **SG: Probably not the best reference** Under an archaic African admixture model, the estimated 8–10% introgression rate is much higher than Neanderthal gene flow, and would have plausibly been fertile ground for dramatic selection for or against archaic-derived haplotypes over a short period<sup>34</sup>. By contrast, adaptation under a weakly structured stem would have occurred continuously over much longer periods. Patterns of polymorphism that are inconsistent with the single-stem model predictions have been used to infer putative archaic admixed segments<sup>11,13,34,15</sup>, negative selection against such segments<sup>34</sup>, and pervasive positive selection<sup>35</sup>. However, such approaches are subject to high false positives in the presence of population structure with migration<sup>32</sup>, and their interpretation should be re-examined in light of a weakly-structured-stem model in Africa.

Multiple studies have shown a correspondence between phenotypic differentiation, usually assessed with measurements of the cranium, and genetic differentiation among human populations and between humans and Neanderthals<sup>36,37,38</sup>. If we assume that this correspondence also holds for early human ancestral populations in Africa, the weakly-structured-stem model would predict that the fossils deriving from each of the two stems should be morphologically quite similar (supplementary text X) **APR: Will we still include this?**, even if they are from different parts of Africa, with differences across stems closer to those across contemporary human populations than to human-Neanderthal differences. Differences within stems would be comparable to differences within recent human populations. Identifying the morphological signature of ancient structure in the fossil record will require careful analysis of many broadly similar remains. Our model further predicts that morphologically distinct crania such as XX and YY are unlikely representatives of branches that contributed appreciably to modern human ancestries. **SG: Here I am being somewhat forceful. We could use conditional to emphasize that these conclusion are under the weakly structured stem model.**

## Methods

Methods are detailed in the Supporting Information.

## Acknowledgements

We are grateful for the DNA contribution from each participant which enabled this study; in particular we wish to highlight the generous participation of the Richtersveld Nama community in South Africa and help from local research assistants Willem DeKlerk and Hendrik Kaimann. Additional assistance

and community engagement was conducted by Justin Myrick, Chris Gignoux, Caitlen Uren and Cedric Werely. We thank the African Genome Diversity Project for data generation, including Tommy Carensen, Deepti Gurdasani, and Manj Sandhu. We also thank Sriram Sankararam, . . . for helpful discussion. This research was supported by an NIH grant R35GM133531 (to BMH) and XXX. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## References

- [1] Jean-Jacques Hublin, Abdelouahed Ben-Ncer, Shara E Bailey, Sarah E Freidline, Simon Neubauer, Matthew M Skinner, Inga Bergmann, Adeline Le Cabec, Stefano Benazzi, Katerina Harvati, and Philipp Gunz. New fossils from jebel irhoud, morocco and the pan-african origin of homo sapiens. *Nature*, 546(7657):289–292, June 2017.
- [2] Tim D White, Berhane Asfaw, David DeGusta, Henry Gilbert, Gary D Richards, Gen Suwa, and F Clark Howell. Pleistocene homo sapiens from middle awash, ethiopia. *Nature*, 423(6941):742–747, June 2003.
- [3] H J Deacon. Two late Pleistocene-Holocene archaeological depositories from the southern cape, south africa. *The South African Archaeological Bulletin*, 50(162):121–131, 1995.
- [4] Chris Stringer. The origin and evolution of homo sapiens. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 371(1698), July 2016.
- [5] Eleanor M L Scerri, Mark G Thomas, Andrea Manica, Philipp Gunz, Jay T Stock, Chris Stringer, Matt Grove, Huw S Groucutt, Axel Timmermann, G Philip Rightmire, Francesco d’Errico, Christian A Tryon, Nick A Drake, Alison S Brooks, Robin W Dennell, Richard Durbin, Brenna M Henn, Julia Lee-Thorp, Peter deMenocal, Michael D Petraglia, Jessica C Thompson, Aylwyn Scally, and Lounès Chikhi. Did our species evolve in subdivided populations across africa, and why does it matter? *Trends Ecol. Evol.*, 33(8):582–594, August 2018.
- [6] Eleanor M L Scerri, Lounès Chikhi, and Mark G Thomas. Beyond multiregional and simple out-of-africa models of human evolution. *Nat Ecol Evol*, 3(10):1370–1372, October 2019.
- [7] Armando Arredondo, Beatriz Mourato, Khoa Nguyen, Simon Boitard, Willy Rodríguez, Camille Noûs, Olivier Mazet, and Lounès Chikhi. Inferring number of populations and changes in connectivity under the n-island model. *Heredity*, 126(6):896–912, June 2021.
- [8] Jack Kamm, Jonathan Terhorst, Richard Durbin, and Yun S Song. Efficiently inferring the demographic history of many populations with allele count data. *J. Am. Stat. Assoc.*, 115(531):1472–1487, 2020.
- [9] Aaron P Ragsdale and Simon Gravel. Models of archaic admixture and recent history from two-locus statistics. *PLoS Genet.*, 15(6):e1008204, June 2019.
- [10] Leo Speidel, Marie Forest, Sinan Shi, and Simon R Myers. A method for genome-wide genealogy estimation for thousands of samples. *Nat. Genet.*, 51(9):1321–1329, September 2019.
- [11] Vincent Plagnol and Jeffrey D Wall. Possible ancestral structure in human populations. *PLoS Genet.*, 2(7):e105, July 2006.
- [12] Michael F Hammer, August E Woerner, Fernando L Mendez, Joseph C Watkins, and Jeffrey D Wall. Genetic evidence for archaic admixture in africa. *Proc. Natl. Acad. Sci. U. S. A.*, 108(37):15123–15128, September 2011.
- [13] Pinghsun Hsieh, August E Woerner, Jeffrey D Wall, Joseph Lachance, Sarah A Tishkoff, Ryan N Gutenkunst, and Michael F Hammer. Model-based analyses of whole-genome data reveal a complex evolutionary history involving archaic introgression in central african pygmies. *Genome Res.*, 26(3):291–300, March 2016.
- [14] Jody Hey, Yujin Chung, Arun Sethuraman, Joseph Lachance, Sarah Tishkoff, Vitor C Sousa, and Yong Wang. Phylogeny estimation by integration over isolation with migration models. *Mol. Biol. Evol.*, 35(11):2805–2818, November 2018.
- [15] Arun Durvasula and Sriram Sankararaman. Recovering signals of ghost archaic introgression in african populations. *Sci Adv*, 6(7):eaax5097, February 2020.

- [16] Belen Lorente-Galdos, Oscar Lao, Gerard Serra-Vidal, Gabriel Santpere, Lukas F K Kuderna, Lara R Arauna, Karima Fadhlou-Zid, Ville N Pimenoff, Himla Soodyall, Pierre Zalloua, Tomas Marques-Bonet, and David Comas. Whole-genome sequence analysis of a pan african set of samples reveals archaic gene flow from an extinct basal population of modern humans into sub-saharan populations. *Genome Biol.*, 20(1):77, April 2019.
- [17] Brenna M Henn, Teresa E Steele, and Timothy D Weaver. Clarifying distinct models of modern human origins in africa. *Curr. Opin. Genet. Dev.*, 53:148–156, December 2018.
- [18] 1000 Genomes Project Consortium, Adam Auton, Lisa D Brooks, Richard M Durbin, Erik P Garrison, Hyun Min Kang, Jan O Korbel, Jonathan L Marchini, Shane McCarthy, Gil A McVean, and Gonalo R Abecasis. A global reference for human genetic variation. *Nature*, 526(7571):68–74, October 2015.
- [19] Deepti Gurdasani, Tommy Carstensen, Fasil Tekola-Ayele, Luca Pagani, Ioanna Tachmazidou, Konstantinos Hatzikotoulas, Savita Karthikeyan, Louise Iles, Martin O Pollard, Ananyo Choudhury, Graham R S Ritchie, Yali Xue, Jennifer Asimit, Rebecca N Nsubuga, Elizabeth H Young, Cristina Pomilla, Katja Kivinen, Kirk Rockett, Anatoli Kamali, Ayo P Doumatey, Gershim Asiki, Janet Seeley, Fatoumatta Sisay-Joof, Muminatou Jallow, Stephen Tollman, Ephrem Mekonnen, Rosemary Ekong, Tamiru Oljira, Neil Bradman, Kalifa Bojang, Michele Ramsay, Adebawale Adeyemo, Endashaw Bekele, Ayesha Motala, Shane A Norris, Fraser Pirie, Pontiano Kaleebu, Dominic Kwiatkowski, Chris Tyler-Smith, Charles Rotimi, Eleftheria Zeggini, and Manjinder S Sandhu. The african genome variation project shapes medical genetics in africa. *Nature*, 517(7534):327–332, January 2015.
- [20] S Gopalan, R E W Berl, G Belbin, C R Gignoux, and others. Hunter-gatherer genomes reveal diverse demographic trajectories following the rise of farming in east africa. *Biorxiv*, 2019.
- [21] Luca Pagani, Stephan Schiffels, Deepti Gurdasani, Petr Danecek, Aylwyn Scally, Yuan Chen, Yali Xue, Marc Haber, Rosemary Ekong, Tamiru Oljira, Ephrem Mekonnen, Donata Luiselli, Neil Bradman, Endashaw Bekele, Pierre Zalloua, Richard Durbin, Toomas Kivisild, and Chris Tyler-Smith. Tracing the route of modern humans out of africa by using 225 human genome sequences from ethiopians and egyptians. *Am. J. Hum. Genet.*, 96(6):986–991, June 2015.
- [22] Kay Prufer, Cesare de Filippo, Steffi Grote, Fabrizio Mafessoni, Petra Korlevi, Mateja Hajdinjak, Benjamin Vernot, Laurits Skov, Pingsun Hsieh, Stephane Peyregne, David Reher, Charlotte Hopfe, Sarah Nagel, Tomislav Maricic, Qiaomei Fu, Christoph Theunert, Rebekah Rogers, Pontus Skoglund, Manjusha Chintalapati, Michael Dannemann, Bradley J Nelson, Felix M Key, Pavao Rudan, eljko Kucan, Ivan Gui, Liubov V Golovanova, Vladimir B Doronichev, Nick Patterson, David Reich, Evan E Eichler, Montgomery Slatkin, Mikkel H Schierup, Aida M Andres, Janet Kelso, Matthias Meyer, and Svante Pabo. A high-coverage neandertal genome from vindija cave in croatia. *Science*, 358(6363):655–658, November 2017.
- [23] Aaron P Ragsdale and Simon Gravel. Unbiased estimation of linkage disequilibrium from unphased data. *Mol. Biol. Evol.*, 37(3):923–932, March 2020.
- [24] Brenna M Henn, Christopher Gignoux, Alice A Lin, Peter J Oefner, Peidong Shen, Rosaria Scozzari, Fulvio Cruciani, Sarah A Tishkoff, Joanna L Mountain, and Peter A Underhill. Y-chromosomal evidence of a pastoralist migration through tanzania to southern africa. *Proc. Natl. Acad. Sci. U. S. A.*, 105(31):10693–10698, August 2008.
- [25] Gwenna Breton, Carina M Schlebusch, Marlize Lombard, Per Sjodin, Himla Soodyall, and Mattias Jakobsson. Lactase persistence alleles reveal partial east african ancestry of southern african khoe pastoralists. *Curr. Biol.*, 24(8):852–858, April 2014.
- [26] David H Alexander, John Novembre, and Kenneth Lange. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.*, 19(9):1655–1664, September 2009.

- [27] Swapan Mallick, Heng Li, Mark Lipson, Iain Mathieson, Melissa Gymrek, Fernando Racimo, Mengyao Zhao, Niru Chennagiri, Susanne Nordenfelt, Arti Tandon, Pontus Skoglund, Iosif Lazaridis, Sriram Sankararaman, Qiaomei Fu, Nadin Rohland, Gabriel Renaud, Yaniv Erlich, Thomas Willems, Carla Gallo, Jeffrey P Spence, Yun S Song, Giovanni Poletti, Francois Balloux, George van Driem, Peter de Knijff, Irene Gallego Romero, Aashish R Jha, Doron M Behar, Claudio M Bravi, Cristian Capelli, Tor Hervig, Andres Moreno-Estrada, Olga L Posukh, Elena Balanovska, Oleg Balanovsky, Sena Karachanak-Yankova, Hovhannes Sahakyan, Draga Toncheva, Levon Yepiskoposyan, Chris Tyler-Smith, Yali Xue, M Syafiq Abdullah, Andres Ruiz-Linares, Cynthia M Beall, Anna Di Rienzo, Choongwon Jeong, Elena B Starikovskaya, Ene Metspalu, Jüri Parik, Richard Villems, Brenna M Henn, Ugur Hodoglugil, Robert Mahley, Antti Sajantila, George Stamatoyannopoulos, Joseph T S Wee, Rita Khusainova, Elza Khusnutdinova, Sergey Litvinov, George Ayodo, David Comas, Michael F Hammer, Toomas Kivisild, William Klitz, Cheryl A Winkler, Damian Labuda, Michael Bamshad, Lynn B Jorde, Sarah A Tishkoff, W Scott Watkins, Mait Metspalu, Stanislav Dryomov, Rem Sukernik, Lalji Singh, Kumarasamy Thangaraj, Svante Pääbo, Janet Kelso, Nick Patterson, and David Reich. The simons genome diversity project: 300 genomes from 142 diverse populations. *Nature*, 538(7624):201–206, October 2016.
- [28] O Mazet, W Rodríguez, S Grusea, S Boitard, and L Chikhi. On the importance of being structured: instantaneous coalescence rates and human evolution—lessons for ancestral population size inference? *Heredity*, 2016.
- [29] Anders Bergström, Chris Stringer, Mateja Hajdinjak, Eleanor M L Scerri, and Pontus Skoglund. Origins of modern human ancestry. *Nature*, 590(7845):229–237, February 2021.
- [30] Margaret Whiting Blome, Andrew S Cohen, Christian A Tryon, Alison S Brooks, and Joellen Russell. The environmental context for the origins of modern human diversity: A synthesis of regional variability in african climate 150,000–30,000 years ago. *J. Hum. Evol.*, 62(5):563–592, May 2012.
- [31] Curtis W Marean, Hayley C Cawthra, Richard M Cowling, Karen J Esler, Erich Fisher, Antoni Milewski, Alastair J Potts, Elzanne Singels, and Jan De Vynck. Stone age people in a changing south african greater cape floristic region. In *Fynbos*. Oxford University Press, October 2014.
- [32] Martin Petr, Svante Pääbo, Janet Kelso, and Benjamin Vernot. Limits of long-term selection against neandertal introgression. *Proc. Natl. Acad. Sci. U. S. A.*, 116(5):1639–1644, January 2019.
- [33] Xinjun Zhang, Kelsey E Witt, Mayra M Bañuelos, Amy Ko, Kai Yuan, Shuhua Xu, Rasmus Nielsen, and Emilia Huerta-Sanchez. The history and evolution of the Denisovan-EPAS1 haplotype in tibetans. *Proc. Natl. Acad. Sci. U. S. A.*, 118(22), June 2021.
- [34] Jeffrey D Wall, Aakrosh Ratan, Eric Stawiski, and GenomeAsia 100K Consortium. Identification of African-Specific admixture between modern and archaic humans. *Am. J. Hum. Genet.*, 105(6):1254–1261, December 2019.
- [35] Daniel R Schrider and Andrew D Kern. Soft sweeps are the dominant mode of adaptation in the human genome. *Mol. Biol. Evol.*, 34(8):1863–1877, August 2017.
- [36] J H Relethford. Craniometric variation among modern human populations. *Am. J. Phys. Anthropol.*, 95(1):53–62, September 1994.
- [37] Timothy D Weaver, Charles C Roseman, and Chris B Stringer. Close correspondence between quantitative- and molecular-genetic divergence times for neandertals and modern humans. *Proc. Natl. Acad. Sci. U. S. A.*, 105(12):4645–4649, March 2008.
- [38] Noreen von Cramon-Taubadel. Congruence of individual cranial bone morphology and neutral molecular affinity patterns in modern humans. *Am. J. Phys. Anthropol.*, 140(2):205–215, October 2009.