

A weakly structured stem for human origins in Africa

Aaron P. Ragsdale¹, Timothy D. Weaver², Elizabeth G. Atkinson³, Eileen Hoal⁴, Marlo Möller⁵, Brenna M. Henn^{2,6,*}, and Simon Gravel^{7,**}

¹Department of Integrative Biology, University of Wisconsin, Madison, WI, USA

²Department of Anthropology, University of California, Davis, Davis, CA, USA

³FIXME: 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

⁴FIXME

⁵FIXME

⁶UC Davis Genome Center, University of California, Davis, Davis, CA, USA

⁷Department of Human Genetics, McGill University, Montreal, QC, Canada

*bmhenn@ucdavis.edu

**simon.gravel@mcgill.ca

January 5, 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 120-135ka, 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. **SG: IF we were not afraid of the racists, we could write something like: "Despite the genetic similarity between these populations, 4 to 10% of random genetic differences among [modern human populations] can be attributed to this early genetic drift.** In contrast to models with archaic introgression, we predict that fossil remains from coexisting ancestral populations should be morphologically similar. 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. 300ka-40ka) 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 Irhoud¹, Herto² and Klasies River³

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 300ka^{4,5,6}.

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 models⁷ or population coalescence and fragmentation⁶. 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^{8,9,10}. 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^{11,12,13,14,9,15,16,15}. 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.¹³). 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.¹⁷). 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 (MSL from Sierre Leone¹⁸), Gumuz (recent descendants of a hunter-gatherer group from Ethiopia^{19,20}), and Eastern African agriculturalists (Amhara and Oromo from Ethiopia¹⁹). The Amhara and Oromo populations, despite speaking distinct Afro-Asiatic languages, are highly genetically similar^{21,20} 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 in our demographic models as a representative source of back-to-Africa gene flow and recent colonial admixture in South Africa. Finally, we added a high-coverage ancient Neanderthal genome from Vindija Cave, Croatia²² to account for archaic gene flow from Neanderthals in non-Africans and gauge the relative time depth of divergence, assuming Neanderthals diverged 550ka 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^{9,23}. 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.

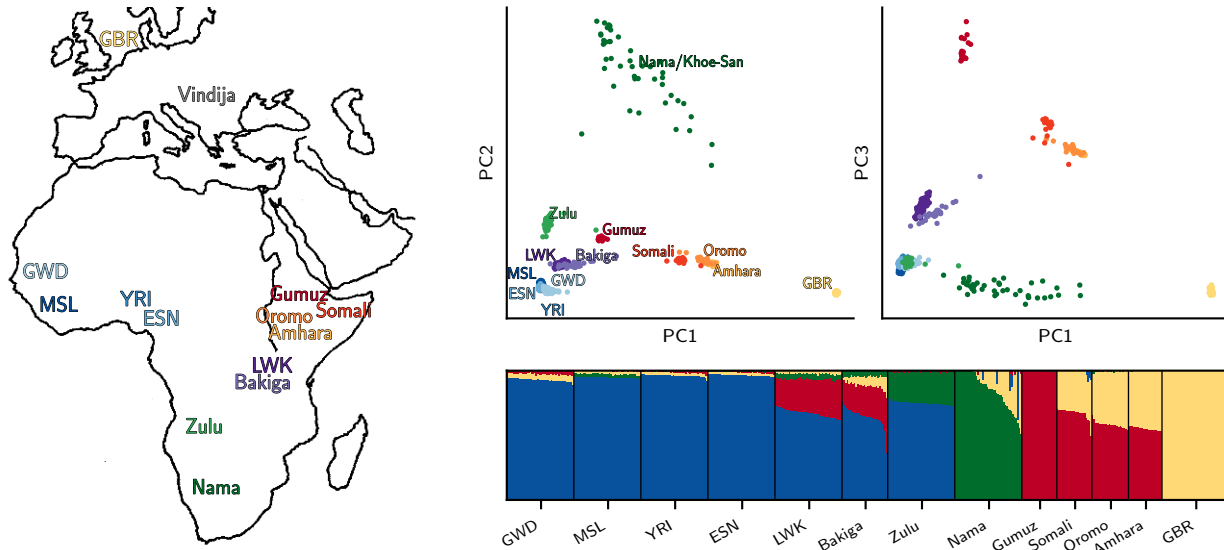


Figure 1: **Geography and genetic diversity within Africa.** APR: In the models presented below, we focus on the Nama, Mende (MSL), Gumuz, and Oromo and Amhara.

A Late Middle Stone Age common ancestry for contemporary humans

We began with a model of geographic expansion from a single ancestral source followed by migration between populations (Figure S2A), without allowing for contribution from an archaic African lineage. As expected⁹, this first model was a poor fit to the data qualitatively (Figure S4) and quantitatively (log-likelihood (LL) $\approx -189,000$, 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 recent human demographic history from two thousand APR: why two thousand, instead of just “up to 150ka”? to 150ka was remarkably robust. The earliest divergence among contemporary sampled human populations differentiates the southern African Nama from other African groups between 110–135ka, with low to moderate levels of subsequent gene flow APR: (Table 1). In none of the high-likelihood models which we explored did the divergence between Nama and other populations exceed ~ 150 ka. We conclude that, regardless of model specification, geographic patterns of *Homo sapiens* population structure date back to the Late Middle Stone Age in Africa, likely arising during 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 APR: (Figure 3).

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 60ka, just prior to the split of Eurasians and East Africans 50ka. Both models predict relatively high subsequent gene flow between the two regions ($m = 2 \times 10^{-4}$). 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 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^{24,25}, 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 [APR: cite](#) (Figure 1).

Deep population structure but not archaic admixture within Africa

Reconciling multiple lines of genetic evidence

Discussion

The Middle Stone Age in Africa

Contrasting archaic admixture and a weakly structured stem

Methods

A placeholder citation²⁶.

Figure 2: **The second main figure.** A placeholder - best fit model(s).

Figure 3: **The third main figure.** A placeholder - validation (Relate and cSFS).

Figure 4: **The fourth main figure.** A placeholder - predictions (FST and/or f4).

Acknowledgements

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] Jerome Kelleher, Alison M Etheridge, and Gilean McVean. Efficient coalescent simulation and genealogical analysis for large sample sizes. *PLoS Comput. Biol.*, 12(5):e1004842, May 2016.