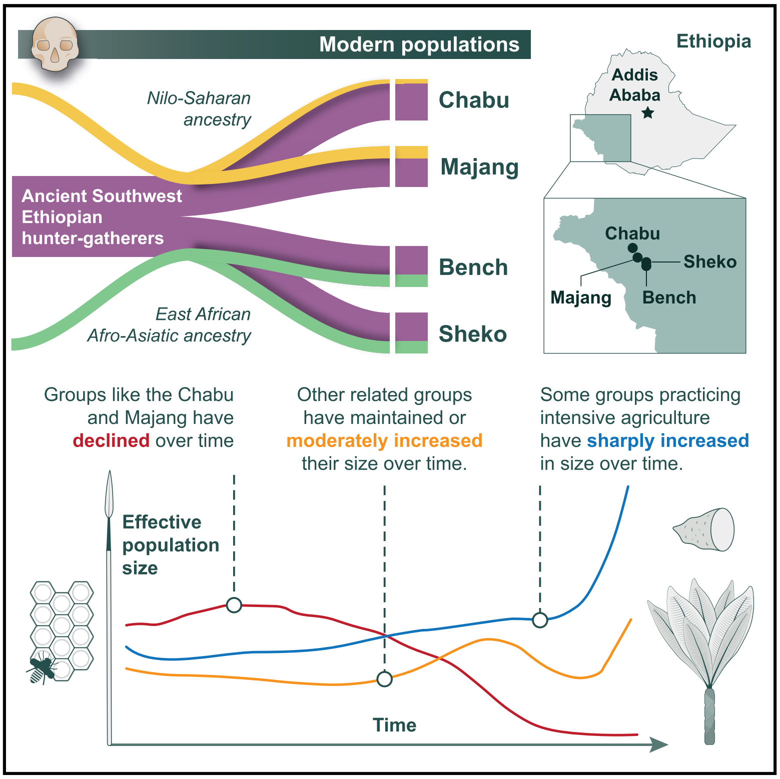
Current Biology

Report

Hunter-gatherer genomes reveal diverse demographic trajectories during the rise of farming in Eastern Africa

Graphical abstract



Highlights

- The Chabu people are related to ancient Southwest Ethiopian hunter-gatherers (HGs)
- Like other African HGs, Chabu population size has declined over the past 1,400 years
- However, other populations with Ethiopian HG ancestry have not experienced declines
- This heterogeneity may stem from variable HG responses to encroaching farmers

Authors

Shyamalika Gopalan,
Richard E.W. Berl, Justin W. Myrick, ...,
Marcus W. Feldman, Barry S. Hewlett,
Brenna M. Henn

Correspondence

hewlett@wsu.edu (B.S.H.),
bmhenn@ucdavis.edu (B.M.H.)

In brief

Using new genotype data, Gopalan et al. show that the Chabu people of Southwest Ethiopia are closely related to ancient people who lived in the region prior to the rise of farming. The Chabu population has declined sharply over the past 1,400 years. However, this trend is not universal among Ethiopian hunter-gatherer descendants.

Gopalan et al., 2022, Current Biology *32*, 1–9
April 25, 2022
https://doi.org/10.1016/j.cub.2022.02.050

Please cite this article in press as: Gopalan et al., Hunter-gatherer genomes reveal diverse demographic trajectories during the rise of farming in Eastern Africa, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.02.050

Current Biology

CellPress
OPEN ACCESS

Report

Hunter-gatherer genomes reveal diverse demographic trajectories during the rise of farming in Eastern Africa

Shyamalika Gopalan,[1,2,15,16] Richard E.W. Berl,[3,4,15,17] Justin W. Myrick,[5,6] Zachary H. Garfield,[7,8] Austin W. Reynolds,[5,9] Barnabas K. Bafens,[10] Gillian Belbin,[11] Mira Mastoras,[6] Cole Williams,[12] Michelle Daya,[12] Akmel N. Negash,[13] Marcus W. Feldman,[14] Barry S. Hewlett,[7,*] and Brenna M. Henn[5,6,18,*]

[1]Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY 11794, USA
[2]Center for Genetic Epidemiology, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA
[3]School of Biological Sciences, Washington State University, Pullman, WA 99164, USA
[4]Department of Human Dimensions of Natural Resources, Colorado State University, Fort Collins, CO 80523, USA
[5]Department of Anthropology, University of California, Davis, Davis, CA 95616, USA
[6]UC Davis Genome Center, University of California, Davis, Davis, CA 95616, USA
[7]Department of Anthropology, Washington State University, Vancouver, WA 98686, USA
[8]Institute for Advanced Study in Toulouse, Université Toulouse, Toulouse 31080, France
[9]Department of Anthropology, Baylor University, Waco, TX 76798, USA
[10]Diaspora and Protocol Affairs Office, Bench Sheko Zone Administration, Mizan, Ethiopia
[11]Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[12]Department of Medicine, University of Colorado, Anschutz Medical Campus, Aurora, CO 80045, USA
[13]Department of Anthropology, Hawassa University, Hawassa, SNNPR, Ethiopia
[14]Department of Biology, Stanford University, Stanford, CA 94305, USA
[15]These authors contributed equally
[16]Present address: Department of Evolutionary Anthropology, Duke University, Durham, NC 27708, USA
[17]Present address: U.S. Geological Survey, Eastern Ecological Science Center at the Patuxent Research Refuge, Laurel, MD 20708, USA
[18]Lead contact
*Correspondence: hewlett@wsu.edu (B.S.H.), bmhenn@ucdavis.edu (B.M.H.)

https://doi.org/10.1016/j.cub.2022.02.050

SUMMARY

The fate of hunting and gathering populations following the rise of agriculture and pastoralism remains a topic of debate in the study of human prehistory. Studies of ancient and modern genomes have found that autochthonous groups were largely replaced by expanding farmer populations with varying levels of gene flow, a characterization that is influenced by the almost universal focus on the European Neolithic.[1–5] We sought to understand the demographic impact of an ongoing cultural transition to farming in Southwest Ethiopia, one of the last regions in Africa to experience such shifts.[6] Importantly, Southwest Ethiopia is home to several of the world's remaining hunter-gatherer groups, including the Chabu people, who are currently transitioning away from their traditional mode of subsistence.[7] We generated genome-wide data from the Chabu and four neighboring populations, the Majang, Shekkacho, Bench, and Sheko, to characterize their genetic ancestry and estimate their effective population sizes over the last 60 generations. We show that the Chabu are a distinct population closely related to ancient people who occupied Southwest Ethiopia >4,500 years ago. Furthermore, the Chabu are undergoing a severe population bottleneck, which began approximately 1,400 years ago. By analyzing eleven Eastern African populations, we find evidence for divergent demographic trajectories among hunter-gatherer-descendant groups. Our results illustrate that although foragers respond to encroaching agriculture and pastoralism with multiple strategies, including cultural adoption of agropastoralism, gene flow, and economic specialization, they often face population decline.

RESULTS

In order to test hypotheses pertaining to the impact of foraging-to-farming transitions, we estimated genetic ancestry, relative genetic isolation, and the timing and magnitude of demographic fluctuations in Eastern African populations. Our investigation focuses on the Chabu (their preferred ethnonym,[7,8] but also referred to in the literature as the "Sabue," "Sabu," and "Shabo"[9–11]), a group of transitioning hunter-gatherers (HGs) who inhabit the Southwestern Ethiopian highland forests that straddle the borders between the Oromia Regional State, Gambella Regional State, and the Southern Nations, Nationalities, and Peoples' Region (SNNPR).[7] We generated genome-wide data from the Chabu (n = 83) and neighboring Majang,

Please cite this article in press as: Gopalan et al., Hunter-gatherer genomes reveal diverse demographic trajectories during the rise of farming in Eastern Africa, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.02.050

Shekkacho, Bench, and Sheko groups (n = 49, 45, 48, 50) at 1.7 million single nucleotide polymorphisms (SNPs). We combined this dataset with published genotypes from additional groups from across Eastern, Central, and Western Africa, as well as the Near East (Table S1). Importantly, we also include genomic data from Bayira, a 4,500-year-old individual found in Mota Cave in the nearby Gamo Highlands who lived well before any evidence of agriculture or pastoralism in the region,[12,13] as well as additional ancient genomes from Eastern Africa, the Levant, and Anatolia.

Inferring Chabu origins through patterns of genome-wide relatedness

To characterize the genetic relationships between the Chabu and other African and Near Eastern (NE) populations, we first estimated their global ancestry. We performed unsupervised clustering of autosomal SNPs (i.e., admixture; STAR Methods), varying "K," the hypothesized number of ancestral source populations, from 2 to 12 (Figure S1).[14] We focus on the pattern that arises at K = 7, where global ancestry patterns in Southwest Ethiopia are represented by genetic components that we identify here by the population or linguistic/geographic group that carries its highest frequencies (Figure 1). The Chabu and their near neighbors are primarily characterized by differences in their frequencies of five components: "Bayira-majority," "Chabu-majority," "Nilo-Saharan" (NS), "East African Afro-Asiatic" (EAAA), and "Near Eastern" (NE). Importantly, Bayira-majority and Chabu-majority components are genetically similar ($F_{st} = 0.05$) and jointly represent the ancestry of Southwest Ethiopian HGs.

At low levels of K = 3–5, the "Bayira-majority" and "Chabu-majority" ancestries are the same, distributed widely across Eastern Africa. At K = 7, Chabu fall out and are modeled as carrying over 90% of their own "Chabu-majority" ancestry. This component is also at significant frequencies in Bayira (9%), the neighboring Majang (38%), and nearby NS populations (5%–10%) (Figure 1A). The Bayira-majority component is found at highest frequency in the extant Aari Blacksmiths and Cultivators, who were previously found to be Bayira's closest relatives,[12] as well as the Bench and Sheko (Figure 1B). The Majang and Gumuz, a western Ethiopian population, also carry this component at substantial frequencies. More generally, other Ethiopian populations are distinguished by their relative proportions of EAAA and NE components (Figures 1E and 1F). The Chabu do not carry EAAA or NE ancestry.

Taken together, these results suggest that the Chabu are primarily descended from ancient Southwest Ethiopian HG groups; a hypothesis of secondary adoption of hunting and gathering is not supported. On the basis of these findings, we also consider the Majang, Gumuz, Aari Blacksmiths and Cultivators, Bench, and Sheko to be probable "HG descendants," even though they do not currently practice this subsistence strategy. The Majang and Gumuz are similar to each other and have genetic affinities with both NS and HG groups (Figures 1 and 2). While both groups are primarily small-scale farmers today, ethnographic studies show that they exhibit characteristic features of HG societies, such as high degrees of egalitarianism and reciprocity, and likely hunted and gathered regularly in the recent past.[15,16] The Aari Blacksmiths and Cultivators, Bench, and Sheko form another group of populations that show the strongest genetic affinities to Bayira (Figures 1A and 2B), suggesting that they are also direct descendants of the ancestral forager populations to which Bayira belonged.[12]

Investigating and dating episodes of gene flow

Based on our unsupervised clustering analysis, nearly all Ethiopian populations carry multiple distinct ancestries (Figure 1). Furthermore, some of these groups (e.g., the Wolayta) show substantial intra-population variance in their ancestry components and/or a relatively broad distribution in principal component (PC) space, suggesting recent gene flow (Figures 1A and 2A). Using F3 admixture statistics and a linkage disequilbrium (LD) decay-based method, we found additional evidence of gene flow in many Southwest Ethiopian populations and were able date some of these instances to within the last 125 generations (Table S2).[17,18]

Genetic signatures of isolation in the Chabu

We inferred spatial population structure by using genetic data to estimate the effective migration surface (EEMS) of humans across Eastern Africa.[19] This analysis reveals corridors of and barriers to gene flow that closely correspond to the geographic distribution of ancestral components (Figure 1G). Some of these barriers also correspond to major geographic features such as deserts, high elevation areas, and bodies of water. However, other features that might be expected to have been migration barriers, such as the Nubian Desert and northeastern Ethiopian Highlands, seem not to have impeded historical gene flow to the same extent (Figure S2). We also find that areas of low migration tend to lie along the boundaries between major African language families, while high migration corridors lie within them (Figure S2C). Together, these results emphasize the close association between geography and language in determining gene flow between groups.[20,21]

The Chabu lie directly in the center of a language contact area with negative effective migration rates, indicating their relative isolation from neighboring groups. We further quantified the effects of recent and historical genetic isolation by analyzing runs of homozygosity (RoH). Compared to their neighbors, the Chabu carry a much larger proportion of their genome in RoH (Figure 3A). As the Chabu practice clan exogamy and have no cultural tradition of close relative marriage, their elevated levels of homozygosity relative to their neighbors are likely driven by demographic pressure.[7] We compared the Chabu to the Batwa, Biaka, Mbuti, Hadza, and Sandawe. Only the Hadza of Tanzania, who were previously shown to carry the highest levels of RoH among Africans, exceeded the Chabu in cumulative RoH (Figure S3).[22]

Closely related populations experienced divergent demographic trajectories

Given these indications of recent demographic pressure on some HG groups, we sought to more precisely estimate their historical effective population sizes (N_e). We used a non-parametric method that leverages the distribution of segments that are shared identical-by-descent (IBD) across pairs of individuals.[23] This allowed us to estimate N_e at each generation from 4 to 60 generations ago (ga). We evaluated the robustness of IBDN_e to small sample sizes, gene flow, and SNP ascertainment by performing

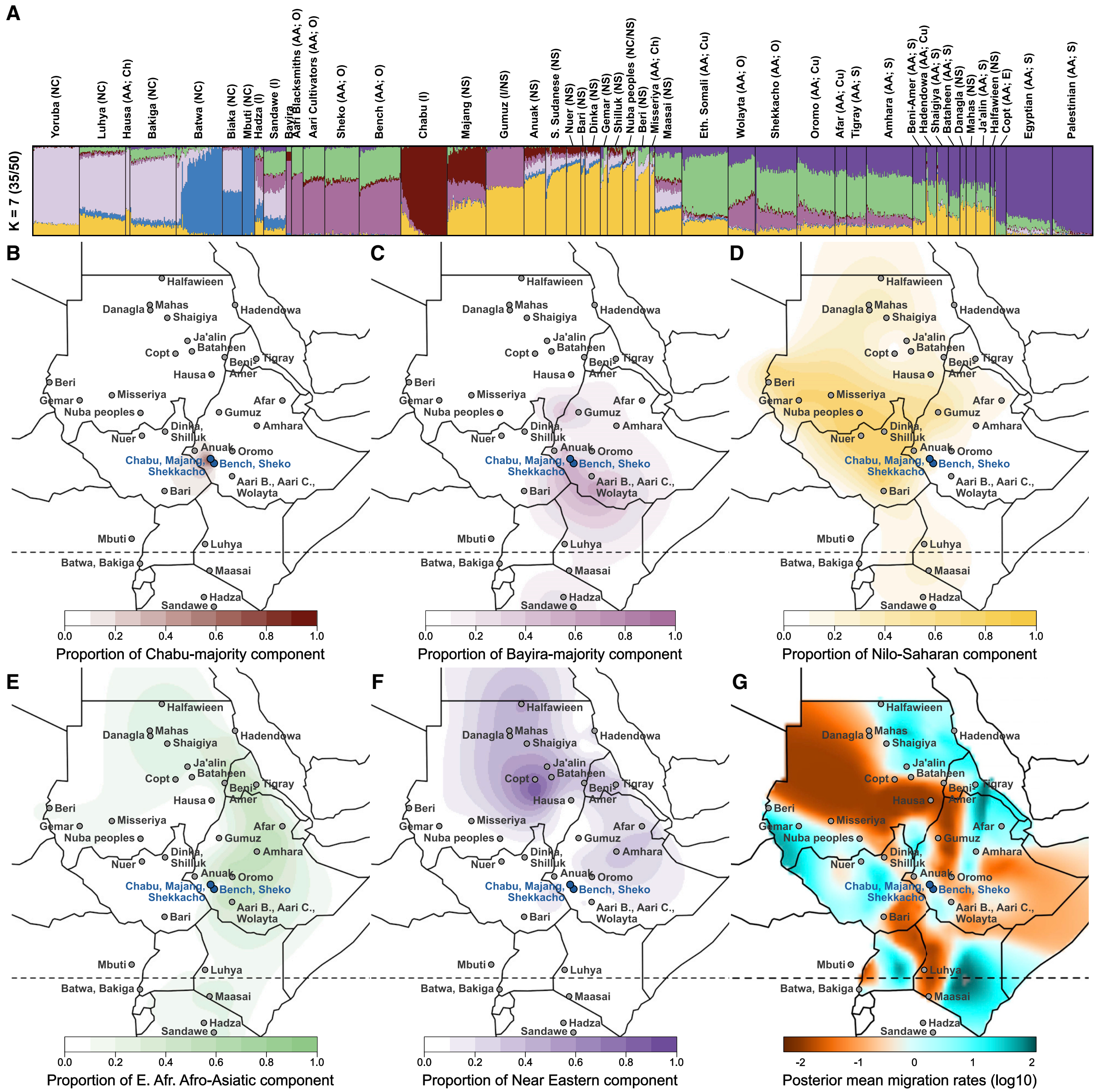


Figure 1. Global ancestry proportions of individuals inferred from unsupervised clustering of genotype data

(A) Each color corresponds to one of the K = 7 hypothesized genetic components, and each vertical bar represents one individual genome (Bayira bar is plotted five times wider to aid visualization). Population labels include linguistic codes in brackets: Afro-Asiatic (AA), Niger-Congo (NC), Nilo-Saharan (NS), and linguistic isolates (I). Within Afro-Asiatic speakers, we further differentiate between Chadic (Ch), Cushitic (Cu), Egyptian (E), Omotic (O), and Semitic (S) speakers.

(B–F) The geographic distributions for 5 of these ancestry components, with the intensity of the color corresponding to the mean population proportion of the respective ancestry. Each component is labeled below the map.

(G) The effective migration surface, inferred from the rate of decay of genetic similarity across geographic space. Cool colors correspond to effective migration corridors, while warm colors correspond to effective migration barriers.

See also Figures S1 and S2 and Table S1.

coalescent simulations with msprime (STAR Methods).[24] Briefly, we performed 10 simulation replicates for each demographic history, testing the effects of sample size, SNP density, and gene flow on N_e estimation. We found that, in the absence of gene flow, a true decline in N_e could be robustly inferred with as few as 20 samples (Figure S4). However, for the same sample size, constant and growing populations were often incorrectly estimated; constant N_e was estimated to be substantially increasing in 40% of

Please cite this article in press as: Gopalan et al., Hunter-gatherer genomes reveal diverse demographic trajectories during the rise of farming in Eastern Africa, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.02.050

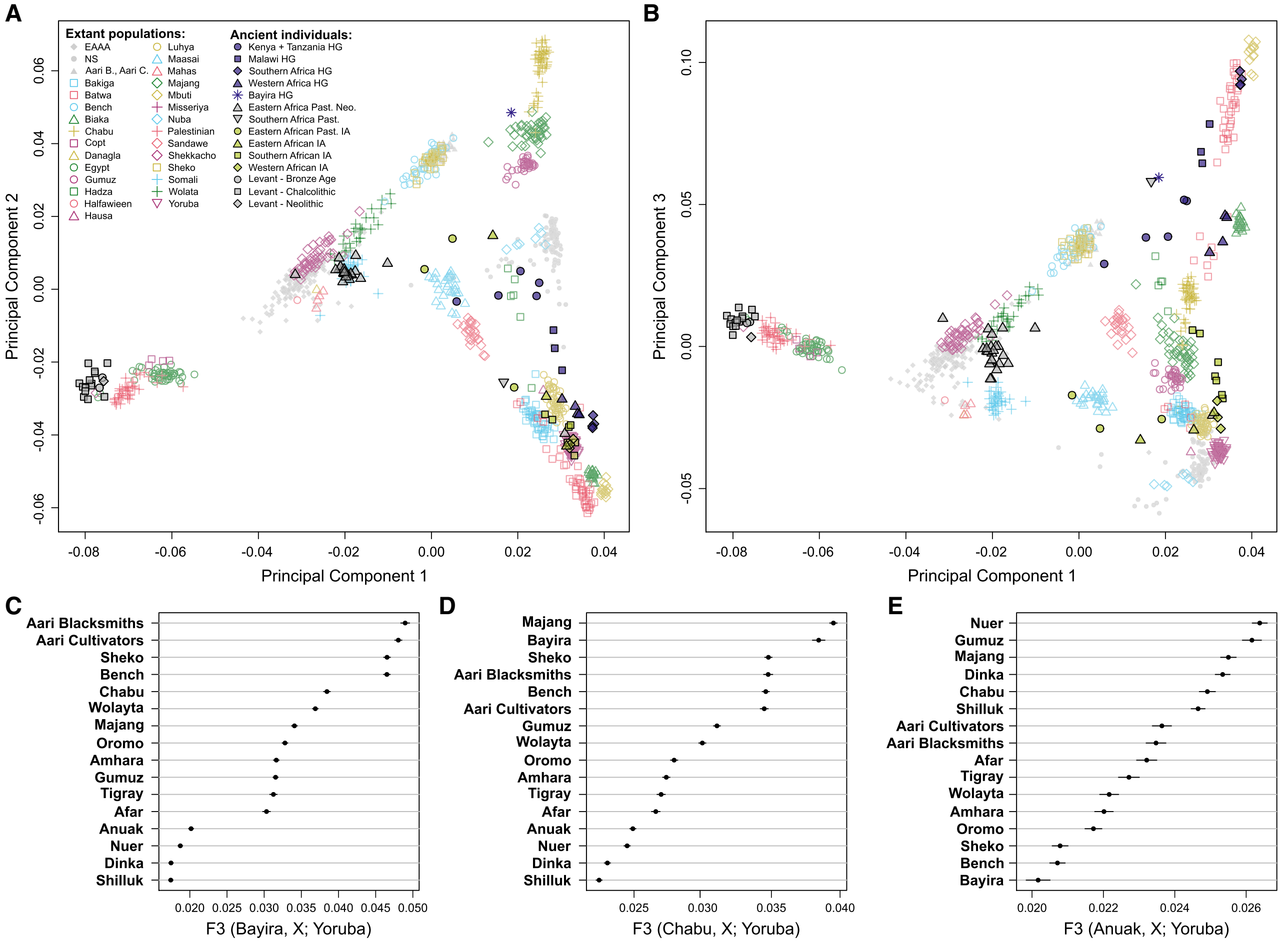


Figure 2. Population structure and F3 outgroup estimates for modern and ancient individuals

(A) A principal component analysis (PCA) of genotype data from modern populations from Eastern Africa and the Near East, with ancient DNA samples superimposed, for PC1 and PC2. Bayira falls close to the Chabu cluster, which anchors the second PC.

(B) As in (A), but for PC1 and PC3. Here, the Chabu lie near other modern and ancient hunter-gatherers.

(C–E) F3 outgroup tests for X, listed in each row, for shared drift with (C) ancient Bayira or (D) the Chabu relative to the Yoruba. (E) The Anuak, a Nilo-Saharan-speaking Ethiopian group have higher F3 outgroup statistics with the Majang and Gumuz than with the Dinka or Shilluk, despite having a high proportion of NS ancestry (Figure 1A). Overall, the F3 outgroup statistic indicates that, among extant populations, Bayira is most closely related to the Aari Blacksmiths, Aari Cultivators, Bench, and Sheko, followed by the Chabu; conversely, the Chabu carry the most shared drift with their Majang neighbors, followed by Bayira. The whiskers represent standard errors on the F3 estimates.

replicates, while increasing N_e was estimated to be holding steady or fluctuating in 10% and 20% of replicates, respectively. This discrepancy resolves when the sample size was increased to 50 (Figure S4).

Among Southwest Ethiopian populations, we found that the N_e of the Chabu, Majang, Bench, Sheko, and Aari Blacksmiths have all declined in the recent past, while the N_e of Aari Cultivators, Gumuz, Wolayta, and Shekkacho have increased (Figure 4). The decline in the Chabu and Aari Blacksmiths starting approximately 50 ga is consistent with the RoH results, but similar patterns of decline in the Majang, Bench, and Sheko were not suggested by our RoH analyses (Figure 3B). Two other HG-descendant groups in Tanzania, the Hadza and Sandawe, both have experienced net declines in N_e over the past 60 generations (Figure 4); the Hadza decline is consistent with RoH patterns (Figure S3). We caution that small sample sizes <50 are sensitive to gene flow (STAR Methods); while the Aari Blacksmiths lacked evidence of recent gene flow (Table S2), additional analysis of the Sandawe is warranted. Further details are contextualized below (see discussion).

DISCUSSION

Previous genetic research on the Neolithic transition has largely focused on Europe, especially with the advent of high-throughput ancient autosomal DNA.[5] However, ancient DNA studies that attempt to characterize the transition often represent HGs by single individuals or aggregate samples over millennia.[2,25,26] These studies are therefore limited in what they can infer regarding the processes of transition. Furthermore, patterns observed in Europe may not pertain to innovations and diffusions of agriculture and/or pastoralism in Africa.[27–30] Here, by studying extant

Please cite this article in press as: Gopalan et al., Hunter-gatherer genomes reveal diverse demographic trajectories during the rise of farming in Eastern Africa, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.02.050

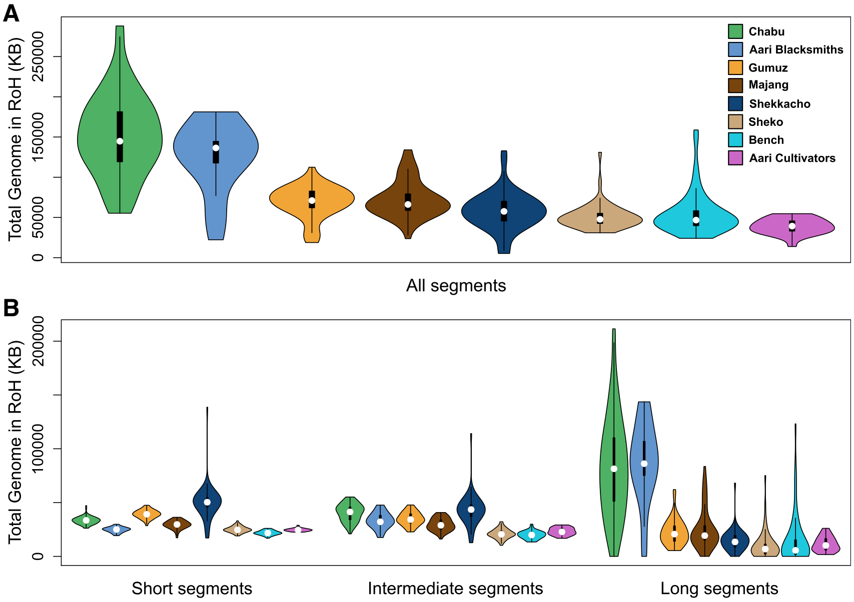


Figure 3. Distributions of the total amount of the genome in runs of homozygosity (RoH) in Southwest Ethiopian populations

RoH are represented by colored violins for (A) all RoH segments and (B) separate RoH size classes. The white point represents the median value of the distribution, and the black rectangle represents values between the lower and upper quartiles. The thin black "whiskers" extend to data points that lie within 1.5 times the interquantile range below or above the lower and upper quantiles, respectively. The Chabu and Aari Blacksmiths showed significantly elevated total RoH in only the longest class suggesting that these populations' genomic signatures of isolation and demographic decline are a result of relatively recent events. See also Figure S3.

agriculturalist and HG East African populations, including 276 new samples from 5 populations, we evaluate the mechanisms underlying the spread of farming in this part of the world.

As farmers expand into a geographic region, HG groups already living there ultimately become either replaced by farmers (i.e., local extirpation) or persist alongside them. Much work has highlighted the prevalence of the former outcome; however, we are interested in how HG populations and their genetic descendants persist in the midst of major cultural shifts. We interpret our results in the context of ethnographic and archeological evidence to consider the following mechanisms by which the present-day populations with HG genetic ancestry might have adjusted to encroachment by farmers: (1) reduce their geographic range, (2) move to an ecological region that is marginal for farming, (3) adopt different cultural (including subsistence) practices, and (4) enter into an economic-symbolic exchange relationship.[10,16,31–36] This list of responses is not exhaustive, nor are they mutually exclusive; the history of any particular group may have involved multiple different responses at various times.[37] Importantly, genetic data can give insights into the extent to which population size change and gene flow were associated with these different responses.

Our analyses demonstrate that the Chabu descend from a population with genetic affinities to Bayira, an individual who lived well before any evidence of intensive farming in the region. The Chabu say they are the original inhabitants of the forests they currently occupy, a claim their nearest neighbors generally support.[7] However, the lack of recognition of Chabu land claims and the migration of farmers from other parts of Ethiopia facing land shortages have resulted in the loss of traditional Chabu forests to development projects.[7] We hypothesize that this documented loss of Chabu land over the past two decades is a continuation of a centuries-long trend (Response 1, above). Specifically, by analyzing population-level genomic patterns, we estimate that the Chabu have experienced a precipitous decline from approximately N_e = 6,000 beginning ~40 ga to N_e = 200 four ga (Figure 4). Current estimates of the Chabu census size range between 1,700 and 2,500.[7] These findings are not at odds with limited gene flow between the Chabu and Nilotic groups 30–43 ga (Figures 1 and 2; Table S2) or previous findings of deep shared ancestry with the geographically distant Hadza and Sandawe.[11] Rather, they suggest that increased HG isolation and population decline is a relatively recent trend coinciding with the expansion of agriculturalists and pastoralists across Africa, which disrupted once-widespread HG networks.[11,29,38,39] Within just the last decade, ethnographic data show the Chabu are experiencing greater assimilation, with an increasing proportion of Chabu men preferring to take a Majang, Shekkacho, or Amhara spouse.[7]

A severe population bottleneck had been previously reported in the Hadza HGs of Tanzania, who speak a linguistic isolate.[22] Our analyses support this observation, and find a decline in Hadza N_e from 3,500 to 160, accelerating between 15 and 25 ga (Figure 4). Today, the Hadza live around Lake Eyasi, an area unsuitable for cultivation or pastoralism, which may explain their continued persistence as HGs (Responses 1 and 2).[40] The Sandawe, close neighbors of the Hadza who speak a distinct language isolate, also show an overall decline in N_e over the past 60 generations (Figure 4). This group of former HGs is known to have transitioned to agro-pastoralism in the last 500 years (Response 3).[41]

Interestingly, the Majang and Gumuz exhibit divergent N_e trajectories despite being highly genetically similar and both current practitioners of small-scale cultivation. The Majang have steadily declined starting 50 ga by about 85% from N_e = 5,000, while the Gumuz N_e has apparently nearly doubled from 3,500 over the same period (Figure 4). Like the Sandawe, who transitioned to cultivation well after their N_e had already declined significantly, we hypothesize that the Majang were "late adopters" of horticulture (Response 3). Historical and ethnographic accounts indicate that, beginning a century ago, the Gumuz were forced to migrate to increasingly inhospitable lands due to pressure from Afro-Asiatic-speaking farming neighbors (Response 2).[42] Prior to this, however, we find that the Gumuz N_e was robust, perhaps because of ecological differences or earlier adoption of cultivation relative to the Majang.

We also observe opposite demographic trends in the closely related Aari Blacksmiths and Aari Cultivators. Previous studies have shown that these two groups diverged within the last

Please cite this article in press as: Gopalan et al., Hunter-gatherer genomes reveal diverse demographic trajectories during the rise of farming in Eastern Africa, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.02.050

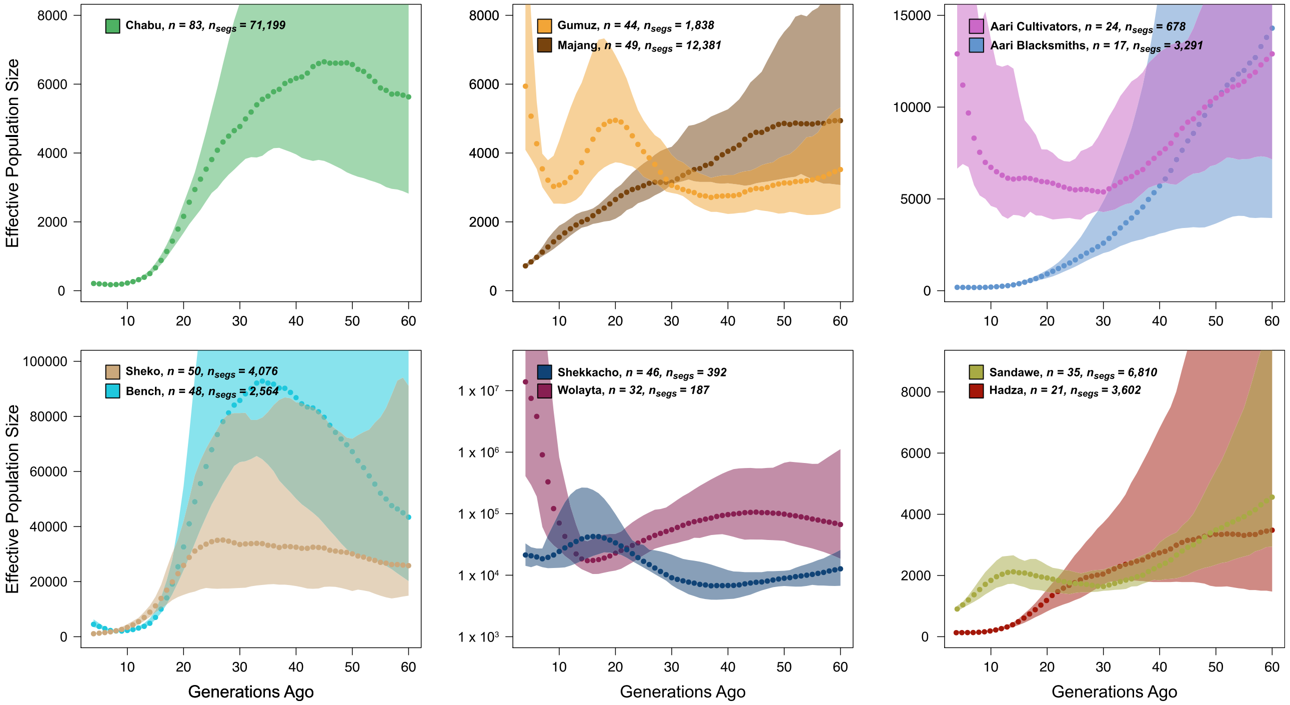


Figure 4. Divergent demographic trajectories for Eastern African populations over the past 2,000 years

Historical effective population sizes (N_e), from 4 to 60 generations ago, were inferred from distributions of identical-by-descent segments >4 cM among pairs of individuals. Shown from the upper left are the Chabu, the Majang and Gumuz, the Aari populations, the Bench and Sheko, the Shekkacho and Wolyata, and the Hadza and Sandawe hunter-gatherers of Tanzania. The number of samples and number of IBD segments used for each population are indicated by "n" and "n_{seg}," respectively. Filled circles represent the estimated N_e at a given generation. Colored ribbons indicate bootstrapped 95% confidence intervals around these estimates. Note that the y axis scale changes across panels and is on a log scale for the Shekkacho/Wolyata panel.

4,500 years, and are both probable descendants of a Bayira-like HG population.[12,43] Our results support earlier findings of a recent bottleneck in the Aari Blacksmiths, and more precisely estimate the timing and magnitude of this decline (Figure 4).[43] At the same time, we find that the N_e of the Aari Cultivators follows a "U-shape" decline and recovery (Figure 4). Today, the Aari Blacksmiths are a marginalized group of craftspeople who neighbor the Aari Cultivators and the Wolayta, with whom they engage in mutual economic exchange (Response 4).[43] Archeological evidence for blacksmithing, today considered a marginal occupational activity in southern Ethiopia (as is foraging and eating wild foods), appeared in nearby regions between 1,000 and 3,000 years ago.[44] There is also evidence from Southwest Ethiopia that the Manja HGs recently shifted to charcoal production, and that marginal occupational groups were lower castes associated with marriage prohibitions.[45,46] We hypothesize that the divergence of the two Aari populations within the last 4,500 years followed their differential adoption of novel cultural practices (i.e., blacksmithing versus farming; Response 3), and subsequent social marginalization of the Blacksmiths influenced their divergent patterns of N_e over the past 60 generations.

The Bench and Sheko are genetically indistinguishable from the Aari (Figures 1A and 2A), suggesting that the majority of their genetic ancestry also derives from Bayira-like HGs.[43] Both groups are currently farmers (Response 3), but unlike the Aari Cultivators, they appear to have experienced net declines in population size over the last few millennia. Our simulations suggest that the apparent extreme jumps in N_e to highs of 93,000 and 35,000 that precede the declines in the Bench and Sheko, respectively, may actually be an artifact of high gene flow (Figures 4 and S4). We find strong evidence for recent admixture with EAAA cultivator groups in both the Bench and Sheko (Table S2). By contrast, the F3 and LD-based tests indicate that if there was gene flow into the Aari from Afro-Asiatic cultivators, it occurred ~100 ga, which is unlikely to affect our IBD-based estimates of N_e (Table S2; Figure S4). Overall, the Aari Blacksmiths, Aari Cultivators, Bench, and Sheko exhibit evidence for Response 3 with qualitatively similar levels of gene flow from incoming EAAA groups (Figure 1A). Despite this, their demographic trajectories are heterogenous.

CONCLUSION

In this work, we characterize nuanced and varied HG responses to recent changes associated with the spread of agriculture and pastoralism in Eastern Africa. While a shift to agricultural subsistence has been linked to increases in N_e,[47–49] we show that this is not a universal outcome. Furthermore, we observe declining N_e in populations that appear to resist cultural change. Continued ethnographic and genetic work in collaboration with the Chabu and other marginalized groups is likely to provide valuable insights into the interactions between farmers and HGs, the drivers of major cultural transitions over long periods of coexistence, and the reasons behind the divergence of demographic histories in genetically and culturally similar groups.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - ○ Lead contact
  - ○ Materials availability
  - ○ Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - ○ Sample collection
  - ○ Ethnographic interviews and return of results
- METHOD DETAILS
  - ○ Data generation and processing
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - ○ Unsupervised clustering analysis
  - ○ Principal component analysis
  - ○ Tests for genetic similarity (shared drift) and gene flow
  - ○ Estimating the date of gene flow
  - ○ Estimated effective migration surfaces
  - ○ Estimating the distributions of major language families
  - ○ Runs of homozygosity
  - ○ Genomic autozygosity regions likelihood-based inference (GARLIC)
  - ○ Historical effective population size inference
  - ○ Demographic simulations
  - ○ Political boundaries in maps

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2022.02.050.

ACKNOWLEDGMENTS

We thank all Chabu, Majang, Bench, Sheko, and Shekkacho participants for their generous contributions and involvement in this research. We also thank additional communities that have contributed data to prior studies used here, and thank Dr. Luca Pagani, Dr. Nina Hollfelder et al., and Dr. George Perry for making these datasets available to us. We acknowledge the sovereignty and rights of all of these groups to the governance, protection, and use of their own genetic data. We thank Yoell Eno and other local research assistants for their contributions. We thank the Stanford Center for Computational, Evolutionary, and Human Genomics and Dr. Carlos Bustamante for providing funding for data generation, and Alexandra Sockell for her assistance with sample preparation. We thank Dr. William Palmer for assisting with the curation of Y chromosomes. We thank Dr. Sharon Browning and Dr. Brian Browning for assistance with implementing the IBDNe algorithm. We thank Dr. Chris Gignoux for consultation on IBD inference and his support throughout the project. R.E.W.B. was supported by funding from the IGERT Program for Evolutionary Modeling at Washington State University and an Exploration Fund grant from The Explorers Club. Z.H.G. acknowledges IAST funding from the French National Research Agency (ANR) under the Investments for the Future (Investissements d'avenir) program, grant ANR-17-EURE-0010. This research was supported by NIH grants R35GM133531 (to B.M.H.) and 2R01HL104608 (to M.D.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

AUTHOR CONTRIBUTIONS

S.G. contributed to conceptualization, data curation, analysis, methodology, visualization, and writing. R.E.W.B. contributed to conceptualization, analysis, data collection, visualization, and writing. J.W.M., Z.H.G., and B.K.B. contributed to data collection. A.W.R. contributed to data analysis. G.B. contributed to analysis, methodology, and software. M.M. contributed to software. A.N.N. contributed to data collection and writing. C.W. contributed to data analysis. M.D. contributed to funding acquisition and data generation. M.W.F. contributed to funding acquisition, data generation, supervision, and writing. B.S.H. contributed to conceptualization, resources, project administration, funding acquisition, supervision, and writing. B.M.H. contributed to conceptualization, project administration, methodology, supervision, and writing.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We worked to ensure gender balance in the recruitment of human subjects. We worked to ensure ethnic or other types of diversity in the recruitment of human subjects. We worked to ensure that the study questionnaires were prepared in an inclusive way. One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in science. One or more of the authors of this paper self-identifies as living with a disability. One or more of the authors of this paper received support from a program designed to increase minority representation in science. The author list of this paper includes contributors from the location where the research was conducted who participated in the data collection, design, analysis, and/or interpretation of the work.

Received: September 25, 2019
Revised: May 12, 2021
Accepted: February 16, 2022
Published: March 9, 2022

REFERENCES

1. Hofmanová, Z., Kreutzer, S., Hellenthal, G., Sell, C., Diekmann, Y., Díez-del-Molino, D., van Dorp, L., López, S., Kousathanas, A., Link, V., et al. (2016). Early farmers from across Europe directly descended from Neolithic Aegeans. Proc. Natl. Acad. Sci. USA *113*, 6886–6891.
2. Brace, S., Diekmann, Y., Booth, T.J., van Dorp, L., Faltyskova, Z., Rohland, N., Mallick, S., Olalde, I., Ferry, M., Michel, M., et al. (2019). Ancient genomes indicate population replacement in Early Neolithic Britain. Nat. Ecol. Evol. *3*, 765–771.
3. Bramanti, B., Thomas, M.G., Haak, W., Unterlaender, M., Jores, P., Tambets, K., Antanaitis-Jacobs, I., Haidle, M.N., Jankauskas, R., Kind, C.-J., et al. (2009). Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. Science *326*, 137–140.
4. Skoglund, P., Malmström, H., Raghavan, M., Storå, J., Hall, P., Willerslev, E., Gilbert, M.T.P., Götherström, A., and Jakobsson, M. (2012). Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. Science *336*, 466–469.
5. Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P.H., Schraiber, J.G., Castellano, S., Lipson, M., et al. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature *513*, 409–413.
6. Hildebrand, E.A., Brandt, S.A., and Lesur-Gebremariam, J. (2010). The Holocene archaeology of Southwest Ethiopia: new insights from the Kafa Archaeological Project. Afr. Archaeol. Rev. *27*, 255–289.
7. Dira, S.J., and Hewlett, B.S. (2017). The Chabu hunter-gatherers of the highland forests of Southwestern Ethiopia. Hunter Gatherer Research *3*, 323–352.
8. Kibebe, T.T. (2015). Documentation and grammatical description of Chabu. PhD thesis (Addis Ababa University).
9. Schnoebelen, T. (2009). (Un)classifying Shabo: phylogenetic methods and results. In Proceedings of the Conference on Language Documentation & Linguistic Theory 2, *2*, P.K. Austin, O. Bond, M. Charette, D. Nathan, and P. Sells, eds., pp. 274–284.

10. González-Ruibal, A., Marín Suárez, C., Sánchez-Elipe, M., Lesur, J., and Martínez Barrio, C. (2014). Late hunters of Western Ethiopia: the sites of Ajilak (Gambela), c. AD 1000–1200. Azania Archaeol. Res. Afr. *49*, 64–101.

11. Scheinfeldt, L.B., Soi, S., Lambert, C., Ko, W.-Y., Coulibaly, A., Ranciaro, A., Thompson, S., Hirbo, J., Beggs, W., Ibrahim, M., et al. (2019). Genomic evidence for shared common ancestry of East African hunting-gathering populations and insights into local adaptation. Proc. Natl. Acad. Sci. USA *116*, 4166–4175.

12. Gallego Llorente, M., Jones, E.R., Eriksson, A., Siska, V., Arthur, K.W., Arthur, J.W., Curtis, M.C., Stock, J.T., Coltorti, M., Pieruccini, P., et al. (2015). Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. Science *350*, 820–822.

13. Arthur, J.W., Curtis, M.C., Arthur, K.J.W., Coltorti, M., Pieruccini, P., Lesur, J., Fuller, D., Lucas, L., Conyers, L., Stock, J., et al. (2019). The transition from hunting–gathering to food production in the Gamo Highlands of southern Ethiopia. Afr. Archaeol. Rev. *36*, 5–65.

14. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res *19*, 1655–1664.

15. Alemayehu, E.E. (2015). Mapping the socio-cultural landscape of the Gumuz Community of Metekel, Northwestern Ethiopia. Afr. J. Hist. Cult. *7*, 209–218.

16. Stauder, J. (1971). The Majangir: Ecology and Society of a Southwest Ethiopian People (Cambridge University Press).

17. Loh, P.-R., Lipson, M., Patterson, N., Moorjani, P., Pickrell, J.K., Reich, D., and Berger, B. (2013). Inferring admixture histories of human populations using linkage disequilibrium. Genetics *193*, 1233–1254.

18. Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. Genetics *192*, 1065–1093.

19. Petkova, D., Novembre, J., and Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. Nat. Genet. *48*, 94–100.

20. Creanza, N., Ruhlen, M., Pemberton, T.J., Rosenberg, N.A., Feldman, M.W., and Ramachandran, S. (2015). A comparison of worldwide phonemic and genetic variation in human populations. Proc. Natl. Acad. Sci. USA *112*, 1265–1272.

21. López, S., Tarekegn, A., Band, G., van Dorp, L., Bird, N., Morris, S., Oljira, T., Mekonnen, E., Bekele, E., Blench, R., et al. (2021). Evidence of the interplay of genetics and culture in Ethiopia. Nat. Commun. *12*, 3581.

22. Henn, B.M., Gignoux, C.R., Jobin, M., Granka, J.M., Macpherson, J.M., Kidd, J.M., Rodríguez-Botigué, L., Ramachandran, S., Hon, L., Brisbin, A., et al. (2011). Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. Proc. Natl. Acad. Sci. USA *108*, 5154–5162.

23. Browning, S.R., and Browning, B.L. (2015). Accurate non-parametric estimation of recent effective population size from segments of identity by descent. Am. J. Hum. Genet. *97*, 404–418.

24. Kelleher, J., Etheridge, A.M., and McVean, G. (2016). Efficient coalescent simulation and genealogical analysis for large sample sizes. PLoS Comput. Biol. *12*, e1004842.

25. Mathieson, I., Alpaslan-Roodenberg, S., Posth, C., Szécsényi-Nagy, A., Rohland, N., Mallick, S., Olalde, I., Broomandkhoshbacht, N., Candilio, F., Cheronet, O., et al. (2018). The genomic history of southeastern Europe. Nature *555*, 197–203.

26. González-Fortes, G., Jones, E.R., Lightfoot, E., Bonsall, C., Lazar, C., Grandal-d'Anglade, A., Garralda, M.D., Drak, L., Siska, V., Simalcsik, A., et al. (2017). Paleogenomic evidence for multi-generational mixing between Neolithic farmers and Mesolithic hunter-gatherers in the Lower Danube basin. Curr. Biol. *27*, 1801–1810.e10.

27. Skoglund, P., Thompson, J.C., Prendergast, M.E., Mittnik, A., Sirak, K., Hajdinjak, M., Salie, T., Rohland, N., Mallick, S., Peltzer, A., et al. (2017). Reconstructing prehistoric African population structure. Cell *171*, 59–71.e21.

28. Prendergast, M.E., Lipson, M., Sawchuk, E.A., Olalde, I., Ogola, C.A., Rohland, N., Sirak, K.A., Adamski, N., Bernardos, R., Broomandkhoshbacht, N., et al. (2019). Ancient DNA reveals a multistep spread of the first herders into sub-Saharan Africa. Science *365*, eaaw6275.

29. Wang, K., Goldstein, S., Bleasdale, M., Clist, B., Bostoen, K., Bakwa-Lufu, P., Buck, L.T., Crowther, A., Dème, A., McIntosh, R.J., et al. (2020). Ancient genomes reveal complex patterns of population movement, interaction, and replacement in sub-Saharan Africa. Sci. Adv. *6*, eaaz0183.

30. Pickrell, J.K., Patterson, N., Loh, P.-R., Lipson, M., Berger, B., Stoneking, M., Pakendorf, B., and Reich, D. (2014). Ancient west Eurasian ancestry in southern and eastern Africa. Proc. Natl. Acad. Sci. USA *111*, 2632–2637.

31. Beauclerk, J. (1993). Hunters and Gatherers in Central Africa: on the Margins of Development (Oxfam).

32. Köhler, A., and Lewis, J. (2002). Putting hunter-gatherer and farmer relations in perspective. A commentary from central Africa. In Ethnicity, Hunter-Gatherers, and the "Other": Association or Assimilation in Southern Africa? (Smithsonian Institute), pp. 276–305.

33. Fentaw, A. (2007). A history of the Shekacho (1898–1974). PhD thesis (Addis Ababa University).

34. Malmström, H., Gilbert, M.T.P., Thomas, M.G., Brandström, M., Storå, J., Molnar, P., Andersen, P.K., Bendixen, C., Holmlund, G., Götherström, A., and Willerslev, E. (2009). Ancient DNA reveals lack of continuity between Neolithic hunter-gatherers and contemporary Scandinavians. Curr. Biol. *19*, 1758–1762.

35. de Filippo, C., Heyn, P., Barham, L., Stoneking, M., and Pakendorf, B. (2010). Genetic perspectives on forager-farmer interaction in the Luangwa valley of Zambia. Am. J. Phys. Anthropol. *141*, 382–394.

36. Patin, E., Siddle, K.J., Laval, G., Quach, H., Harmant, C., Becker, N., Froment, A., Régnault, B., Lemée, L., Gravel, S., et al. (2014). The impact of agricultural emergence on the genetic history of African rainforest hunter-gatherers and agriculturalists. Nat. Commun. *5*, 3163.

37. Page, A.E., and French, J.C. (2020). Reconstructing prehistoric demography: what role for extant hunter-gatherers? Evol. Anthropol. *29*, 332–345.

38. Kim, H.L., Ratan, A., Perry, G.H., Montenegro, A., Miller, W., and Schuster, S.C. (2014). Khoisan hunter-gatherers have been the largest population throughout most of modern-human demographic history. Nat. Commun. *5*, 5692.

39. Verdu, P., Austerlitz, F., Estoup, A., Vitalis, R., Georges, M., Théry, S., Froment, A., Le Bomin, S., Gessain, A., Hombert, J.-M., et al. (2009). Origins and genetic diversity of pygmy hunter-gatherers from Western Central Africa. Curr. Biol. *19*, 312–318.

40. Blurton Jones, N. (2016). Demography and Evolutionary Ecology of Hadza Hunter-Gatherers (Cambridge University Press).

41. Newman, J.L. (1970). The Ecological Basis for Subsistence Change among the Sandawe of Tanzania (National Academy of Sciences).

42. Ahmad, A.H. (1995). The Gumuz of the lowlands of Western Gojjam: the frontier in history 1900–1935. Africa *50*, 53–67.

43. van Dorp, L., Balding, D., Myers, S., Pagani, L., Tyler-Smith, C., Bekele, E., Tarekegn, A., Thomas, M.G., Bradman, N., and Hellenthal, G. (2015). Evidence for a common origin of Blacksmiths and Cultivators in the Ethiopian Ari within the last 4500 years: lessons for clustering-based inference. PLoS Genet. *11*, e1005397.

44. Phillipson, D.W. (2005). African Archaeology (Cambridge University Press).

45. Hailu, G.K. (2016). Social stratification and marginalization in the Southern Nations Nationalities and People Region of Ethiopia: the case of Manja minority groups. Glob. J Human-Social Sci. *16* https://globaljournals.org/GJHSS_Volume16/6-Social-Stratification-and-Marginalization.pdf.

46. Yimer, N.A. (2020). The social challenges of potters and tanners among the Yem people, Southwest Ethiopia. Soc. Ment. Res. Thinkers J. *6*, 919–926.

47. Gignoux, C.R., Henn, B.M., and Mountain, J.L. (2011). Rapid, global demographic expansions after the origins of agriculture. Proc. Natl. Acad. Sci. USA *108*, 6044–6049.

48. Aimé, C., Laval, G., Patin, E., Verdu, P., Ségurel, L., Chaix, R., Hegay, T., Quintana-Murci, L., Heyer, E., and Austerlitz, F. (2013). Human genetic

Please cite this article in press as: Gopalan et al., Hunter-gatherer genomes reveal diverse demographic trajectories during the rise of farming in Eastern Africa, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.02.050

data reveal contrasting demographic patterns between sedentary and nomadic populations that predate the emergence of farming. Mol. Biol. Evol. *30*, 2629–2644.

49. Lopez, M., Kousathanas, A., Quach, H., Harmant, C., Mouguiama-Daouda, P., Hombert, J.-M., Froment, A., Perry, G.H., Barreiro, L.B., Verdu, P., et al. (2018). The demographic history and mutational load of African hunter-gatherers and farmers. Nat. Ecol. Evol. *2*, 721–730.
50. Hollfelder, N., Schlebusch, C.M., Günther, T., Babiker, H., Hassan, H.Y., and Jakobsson, M. (2017). Northeast African genomic variation shaped by the continuity of indigenous groups and Eurasian migrations. PLoS Genet *13*, e1006976.
51. Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A., et al. (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature *538*, 201–206.
52. Pagani, L., Kivisild, T., Tarekegn, A., Ekong, R., Plaster, C., Gallego Romero, I., Ayub, Q., Mehdi, S.Q., Thomas, M.G., Luiselli, D., et al. (2012). Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. Am. J. Hum. Genet. *91*, 83–96.
53. Perry, G.H., Foll, M., Grenier, J.-C., Patin, E., Nédélec, Y., Pacis, A., Barakatt, M., Gravel, S., Zhou, X., Nsobya, S.L., et al. (2014). Adaptive, convergent origins of the pygmy phenotype in African rainforest hunter-gatherers. Proc. Natl. Acad. Sci. USA *111*, E3596–E3603.
54. Getz, W.M., Fortmann-Roe, S., Cross, P.C., Lyons, A.J., Ryan, S.J., and Wilmers, C.C. (2007). LoCoH: nonparameteric kernel methods for constructing home ranges and utilization distributions. PLoS One *2*, e207.
55. Szpiech, Z.A., Blant, A., and Pemberton, T.J. (2017). GARLIC: genomic Autozygosity regions likelihood-based inference and classification. Bioinformatics *33*, 2059–2062.
56. Zhou, Y., Browning, S.R., and Browning, B.L. (2020). A fast and simple method for detecting identity-by-descent segments in large-scale data. Am. J. Hum. Genet. *106*, 426–437.
57. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience *4*, 7.
58. Williams, C.M., Scelza, B., Gignoux, C.R., and Henn, B.M. (2020). A rapid, accurate approach to inferring pedigrees in endogamous populations. Preprint at bioRxiv. https://doi.org/10.1101/2020.02.25.965376.
59. Delaneau, O., Zagury, J.-F., and Marchini, J. (2013). Improved whole-chromosome phasing for disease and population genetic studies. Nat. Methods *10*, 5–6.
60. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet. *2*, e190.
61. Goldstein, J.I., Crenshaw, A., Carey, J., Grant, G.B., Maguire, J., Fromer, M., O'Dushlaine, C., Moran, J.L., Chambert, K., Stevens, C., et al. (2012). zCall: a rare variant caller for array-based genotyping: genetics and population analysis. Bioinformatics *28*, 2543–2545.
62. 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. Nature *526*, 68–74.
63. Pagani, L., Schiffels, S., Gurdasani, D., Danecek, P., Scally, A., Chen, Y., Xue, Y., Haber, M., Ekong, R., Oljira, T., et al. (2015). Tracing the route of modern humans out of Africa by using 225 human genome sequences from Ethiopians and Egyptians. Am. J. Hum. Genet. *96*, 986–991.
64. Bergström, A., McCarthy, S.A., Hui, R., Almarri, M.A., Ayub, Q., Danecek, P., Chen, Y., Felkel, S., Hallast, P., Kamm, J., et al. (2020). Insights into human genetic variation and population history from 929 diverse genomes. Science *367*, eaay5012.
65. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.
66. Purcell, S., and Chang, C. (2021). PLINK1.9. https://www.cog-genomics.org/plink/.
67. Behr, A.A., Liu, K.Z., Liu-Fang, G., Nakka, P., and Ramachandran, S. (2016). pong: fast analysis and visualization of latent clusters in population genetic data. Bioinformatics *32*, 2817–2823.
68. Uren, C., Kim, M., Martin, A.R., Bobo, D., Gignoux, C.R., van Helden, P.D., Möller, M., Hoal, E.G., and Henn, B.M. (2016). Fine-scale human population structure in Southern Africa reflects ecogeographic boundaries. Genetics *204*, 303–314.
69. Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D.C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., et al. (2016). Genomic insights into the origin of farming in the ancient Near East. Nature *536*, 419–424.
70. Harney, É., May, H., Shalem, D., Rohland, N., Mallick, S., Lazaridis, I., Sarig, R., Stewardson, K., Nordenfelt, S., Patterson, N., et al. (2018). Ancient DNA from Chalcolithic Israel reveals the role of population mixture in cultural transformation. Nat. Commun. *9*, 3336.
71. Lipson, M., Ribot, I., Mallick, S., Rohland, N., Olalde, I., Adamski, N., Broomandkhoshbacht, N., Lawson, A.M., López, S., Oppenheimer, J., et al. (2020). Ancient West African foragers in the context of African population history. Nature *577*, 665–670.
72. Schlebusch, C.M., Malmström, H., Günther, T., Sjödin, P., Coutinho, A., Edlund, H., Munters, A.R., Vicente, M., Steyn, M., Soodyall, H., et al. (2017). Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. Science *358*, 652–655.
73. Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., et al. (2020). vegan: community ecology package. https://cran.r-project.org/web/packages/vegan/index.html.
74. Raghavan, M., Skoglund, P., Graf, K.E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., Stafford, T.W., Jr., Orlando, L., Metspalu, E., et al. (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature *505*, 87–91.
75. Reich, D., Thangaraj, K., Patterson, N., Price, A.L., and Singh, L. (2009). Reconstructing Indian population history. Nature *461*, 489–494.
76. Hammarström, H., Forkel, R., and Haspelmath, M. (2018). Glottolog 3.3. Jena: Max Planck Institute for the Science of Human History. http://glottolog.org.
77. Benson, S., and Duffield, M. (1979). Women's work and economic change: the Hausa in Sudan and in Nigeria. IDS Bull. *10*, 13–19.
78. Blench, R. (2017). African language isolates. In Language Isolates, L. Campbell, ed., pp. 176–206.
79. Pemberton, T.J., Absher, D.M., Feldman, M.W., Myers, R.M., Rosenberg, N.A., and Li, J.Z. (2012). Genomic patterns of homozygosity in worldwide human populations. Am. J. Hum. Genet. *91*, 275–292.
80. Browning, B.L., and Browning, S.R. (2013). Improving the accuracy and efficiency of identity-by-descent detection in population data. Genetics *194*, 459–471.
81. Ségurel, L., Wyman, M.J., and Przeworski, M. (2014). Determinants of mutation rate variation in the human germline. Annu. Rev. Genomics Hum. Genet. *15*, 47–70.
82. Hinch, A.G., Tandon, A., Patterson, N., Song, Y., Rohland, N., Palmer, C.D., Chen, G.K., Wang, K., Buxbaum, S.G., Akylbekova, E.L., et al. (2011). The landscape of recombination in African Americans. Nature *476*, 170–175.
83. Nelson, D., Kelleher, J., Ragsdale, A.P., Moreau, C., McVean, G., and Gravel, S. (2020). Accounting for long-range correlations in genome-wide simulations of large cohorts. PLoS Genet. *16*, e1008619.

Please cite this article in press as: Gopalan et al., Hunter-gatherer genomes reveal diverse demographic trajectories during the rise of farming in Eastern Africa, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.02.050

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical commercial assays		
Infinium Multi-Ethnic Global-8 Kit	Illumina	WG-316
Deposited data		
Genetic data from Bayira (Mota)	Gallego Llorente et al.[12]	https://doi.org/10.1126/science.aad2879
Genetic data from Sandawe and Hadza people	Henn et al.[22]	https://doi.org/10.1073/pnas.1017511108
Genetic data from Bari, Bataheen, Beni-Amer, Beri, Copts, Danagla, Dinka, Gemar, Hadendowa, Halfawieen, Hausa, Ja'alin, Mahas, Misseriya, Nuba, Nuer, Shaigiya, and Shilluk people	Hollfelder et al.[50]	https://doi.org/10.1371/journal.pgen.1006976
Genetic data from Mbuti people	Mallick et al.[51]	https://doi.org/10.1038/nature18964
Genetic data from Aari Blacksmith, Aari Cultivator, Afar, Anuak, Tigray, Amhara, Ethiopian Somali, Gumuz, Oromo, and Wolayta people	Hollfelder et al.[50]	https://doi.org/10.1016/j.ajhg.2012.05.015
Genetic data from Amhara, Egyptian, Ethiopian Somali, Gumuz, Oromo, and Wolayta people	Pagani et al.[52]	https://doi.org/10.1016/j.ajhg.2015.04.019
Genetic data from Bakiga and Batwa people	Perry et al.[53]	https://doi.org/10.1073/pnas.1402875111
Genetic data from ancient African individuals	Allen Ancient DNA Resource (v. 44.3)	https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data
Genetic data from Bench, Chabu, Majang, Shekkacho, and Sheko people	This study, deposited on dbGaP	phs001123.v2.p2
Software and algorithms		
Custom scripts	This study, deposited on Zenodo	https://doi.org/10.5281/zenodo.5911732
ADMIXTOOLS	Patterson et al.[18]	https://github.com/DReichLab/AdmixTools
ADMIXTURE	Alexander et al.[14]	https://dalexander.github.io/admixture/download.html
ALDER	Loh et al.[17]	https://github.com/joepickrell/malder/tree/master/MALDER
a-LoCoH, implemented in adehabitatHR (R package)	Getz et al.[54]	https://cran.r-project.org/web/packages/adehabitatHR/index.html
EEMS	Petkova et al.[19]	https://github.com/dipetkov/eems
GARLIC	Szpiech et al.[55]	https://github.com/szpiech/garlic
GenomeStudio v2.0.3	N/A	https://support.illumina.com/array/array_software/genomestudio/downloads.html
hap-ibd v1.0	Zhou et al.[56]	https://github.com/browning-lab/hap-ibd
IBDNe (ibdne.04Sep15.e78)	Browning and Browning[23]	https://faculty.washington.edu/browning/ibdne.html
maptools (R package)	The Comprehensive R Archive Network	https://cran.r-project.org/web/packages/maptools/index.html
msprime	Kelleher et al.[24]	https://github.com/tskit-dev/msprime
PLINK 1.9	Chang et al.[57]	https://www.cog-genomics.org/plink/

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
PONDEROSA	Williams et al.[58]	https://github.com/williamscole/PONDEROSA
R	The R Project for Statistical Computing	https://www.r-project.org/
RColorBrewer (R package)	The Comprehensive R Archive Network	https://cran.r-project.org/web/packages/RColorBrewer/index.html
SHAPEIT2	Delaneau et al.[59]	https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html
smartpca	Patterson et al.[60]	https://github.com/DReichLab/EIG
Spatial ancestry plotting functions	Ryan Raaum[52]	https://doi.org/10.1534/genetics.116.187369
vegan (R package)	The Comprehensive R Archive Network	https://cran.r-project.org/web/packages/vegan/index.html
zCall	Goldstein et al.[61]	https://github.com/jigold/zCall

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Brenna Henn (bmhenn@ucdavis.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Genotype data generated for this study are deposited at dbGaP: phs001123.v2.p2. This paper also analyzes existing, publicly available data. The accession numbers or DOIs for these datasets are listed in the key resources table. Additional plots and original code have been deposited at Zenodo and are publicly available as of the date of publication. The DOI is listed in the key resources table. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Sample collection

Samples from the Chabu and the Majang and Shekkacho were collected by REWB in May 2013, using Oragene•DISCOVER (OGR-500) kits for the Chabu and generic 5 ml tubes with Norgen preservation solution for the other two groups. Additional Chabu individuals, as well as Bench and Sheko individuals, were collected in October 2019 with OGR-500 kits. Ethiopian samples were collected after months of ethnographic research by Samuel Dira and BSH, as part of a larger formal collaborative research and capacity-building relationship between the Departments of Anthropology at Hawassa University, Ethiopia (HU) and Washington State University (WSU). The collaboration involves training several HU faculty in the WSU PhD program and cooperative participation in research projects in Southwestern Ethiopia. Prior approvals for the project were obtained from the leadership of each group being sampled, from the School of Behavioral Sciences at Hawassa University (#BS/502/05), and from the Majang Zone Council of the Gambella Regional State (#901/Majang Zone 1). Ethical approval for human subjects research was obtained from the Institutional Review Board of Washington State University under proposals #12972 and #13134. IRB approval was obtained from UC Davis #1445036-1 (July 2019) for additional sampling.

We aimed to sample 50 individuals per population per field season. Samples were obtained opportunistically within each group from the general population present in public or semi-public spaces such as village centers and municipal buildings. No participants were excluded from sampling *a priori* except those under 18 years of age. In recruiting participants, we relied on local informants and community leaders and experts for their aid. Informed consent was obtained from each participant after reading or hearing the approved text translated into their local language and providing their signature, or a fingerprint in lieu of a signature for non-literate participants.

Ethnographic interviews and return of results

Among the Chabu, interviews were largely conducted in the Chabu language. One author (ZHG) has moderate Chabu language skills and conducted many field interviews in Chabu (for simple demographics). A translator who spoke Chabu, Majang, and English assisted. In some cases, interviews were conducted in Majang. For all other ethnicities, interviews were conducted through a translator in the local language. For the purpose of analyses, individuals were classified as Chabu if they self-reported as Chabu and said they had at least one Chabu parent. 6 Chabu that we interviewed reported a non-Chabu (Majang) parent (7% of parents) and 10 reported a non-Chabu (Majang) grandparent (2% of grandparents). The majority of Chabu individuals spoke the Chabu language. Payment was provided to all participants whether or not they identified as Chabu. Chabu data collection was performed in two locations where, as far as the authors are aware, most Chabu reside.[7] One is a predominantly Chabu village deep in the forest and the other is a multi-ethnic frontier town with government presence.

Results presented here, including general population genetics results for the Chabu in the context of neighboring groups, Ethiopia, and globally, were returned and discussed with members of the Chabu community in October 2019. In an effort to increase community attendance, we arranged in advance with community members in the surrounding area to travel for the presentation of results. Research results were presented in a community gathering followed by a discussion; those who could not attend the presentation and whom we encountered opportunistically during our visit were presented the results individually or by household. Many Chabu members in the audience were pleased to hear our research results matched cultural models of their history. They expressed concern about issues regarding development and education related to the need to teach the Chabu language to their children in school.

METHOD DETAILS

Data generation and processing

50 individuals each of the Bench, Sheko, Majang, and Shekkacho, and 88 Chabu, were genotyped using the Illumina Infinium MultiEthnic Global Array, which assays over 1.7 million genetic markers. Genotypes were initially called using Illumina GenomeStudio v2.0.3 software and exported using the human genome build GRCh37. We removed samples that had a call rate below 90%. Calls for rare variants, defined as those having a minor allele frequency (MAF) $<$ 5%, were then replaced by using zCall following their published procedure.[61] Variants with more than 15% missing data, an observed heterozygosity greater than or equal to 80%, or with cluster separation less than or equal to 2% were removed from the dataset. In preparation for merging with additional datasets, all variants were converted to the Illumina top strand and oriented to match the 1000 Genomes reference. We renamed SNPs to match dbSNP version 144 and removed all indels and A/T or C/G transversion variants, leaving over 1.3 million SNPs in the final dataset.

QUANTIFICATION AND STATISTICAL ANALYSIS

Unsupervised clustering analysis

We merged our Ethiopian SNP data with previously published or publicly available genotype data from other Eastern Africans, as well as the Yoruba and Palestinians (Table S1).[12,22,50–53,62–64] We identified relatives in our dataset using PONDEROSA, which is able to accurately identify kinship categories across populations with differential levels of genetic diversity (i.e. due to inbreeding, bottlenecks etc).[58] We set PONDEROSA's parameters to join any segments separated by less than 1 cM and with fewer than 1 discordant homozygous site, and excluded any pairs of individuals exhibiting a 2nd degree relationship or closer.

We removed SNPs that had a missingness rate of over 5% or a minor allele frequency (MAF) less than 1% in the merged dataset, or were out of Hardy-Weinberg equilibrium (HWE) ($p < 0.001$) in any population.[57,65,66] Of the remaining individuals, we randomly discarded a set such that no population, as defined by their population labels, had more than 50 individuals. We then filtered for linkage disequilibrium in the merged dataset (using the PLINK command '–indep-pairwise 50 5 0.3) to the HWE, MAF and missingness filtered data, and removed individuals missing genotype data at more than 15% of sites, leaving 112,322 SNPs from 1,124 individuals across 45 populations plus Bayira.[57,65,66]

We ran the ADMIXTURE algorithm for Ks between 2 and 12 with 50 replicates each, and used pong to visualize concordance between different runs and to identify the most frequent mode per K among all replicates.[14,67] The lowest cross validation error was achieved for a mode that occurred when K=9. However, we focus on the pattern that arises at K=7 as Ks beyond this tended to identify population-specific components that were less informative about inter-population relationships (Figure S1). For extant Eastern African populations with known sampling or ethnographic coordinates, we also plotted the population averages of each ancestry component geographically, interpolating between data points across the landscape as in Uren et al.[68]

Principal component analysis

We performed principal component analysis (PCA) using smartpca (v. 16000) to visualize relationships between modern and ancient groups.[60] We took advantage of Procrustes transformation to include additional ancient samples from Africa and the Near East which had poor SNP overlap with the rest of the dataset.[27,29,69–72] We accessed the genotype data for these individuals through the harmonized dataset from the Allen Ancient DNA Resource (v. 44.3, accessed at https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data, February 19 2019). We retained ancient samples from sub-Saharan Africa, Israel, and Jordan that had a 'PASS' in the assessment column of the Allen Ancient DNA Resource dataset. We then merged our missingness, MAF, and HWE filtered dataset of the 1,124 individuals used for ADMIXTURE (the 'main' dataset) with the ancient individuals (the 'drop in' dataset).

We made all individuals in this combined dataset pseudohaploid by randomly retaining only one of an individual's alleles at any heterozygous sites. We then performed a series of PCAs that included all 'main' individuals plus one 'drop in' individual. We retained

SNPs that overlapped between the main dataset and the dropped in individual, filtered these SNPs for MAF and LD as described above, and computed the top 6 PCs using smartpca.[60] We inspected each of these 'drop in' PCAs individually, plotting two PCs against each other at a time and checking that the relative positions of the main dataset individuals were qualitatively similar across runs. In particular, biplots involving the top 3 PCs produced very consistent patterns. At this point, we chose to exclude samples for which fewer than 10,000 SNPs were used to calculate the PCs. Ultimately, we retained 74 ancient individuals. Using the R package 'vegan,'[73] we performed Procrustes transformation by designating the ancient sample with the highest number of SNPs as the 'baseline' PCA. By comparing the coordinates of only the main dataset individuals across the baseline PCA and each remaining PCA in turn, Procrustes transformation calculates the optimal translation, rotation, and scaling factors needed minimize the overall sum of squared differences between datasets. We then applied these factors to the dropped in individual to calculate its new coordinates in the Procrustes transformed PCA.

Tests for genetic similarity (shared drift) and gene flow

We calculated F3 outgroup statistics of the form F_3(Bayira; P_X, Yoruba) and F_3(Chabu; P_X, Yoruba) in order to estimate the degree of shared drift between various Eastern African populations (P_X) and Bayira and the Chabu, respectively (Figure S2).[74] We also calculated F3 admixture statistics of the form F_3(P_X; P_1, P_2) to test for the possibility that a target population (P_X) is the result of admixture between two diverged source groups (P_1, P_2).[75] For all tests, we used a merged dataset of individuals genotyped on Omni1M, Omni2.5M, and MEGA platforms, plus Bayira, containing 3,563,795 SNPs and 1,515 individuals, which had been filtered for close relatives within populations and for MAF and HWE, but not for LD or missingness. We calculated the F3 statistics using the qp3pop program from ADMIXTOOLS, which automatically retained only SNPs that overlapped across all three groups (P_X, P_1, and P_2).[18]

Estimating the date of gene flow

We tested for possible gene flow between populations, and additionally estimated the dates (in generations) of these gene flow events, using ALDER, a LD-based method, with the 'checkmap: NO' and 'mindis: 0.005' options.[17] We used the same dataset as for calculating F3 statistics, described above.

Estimated effective migration surfaces

We used a method of estimating effective migration surfaces (EEMS) to visualize variation in migration rates across East Africa (Figure S2).[19] The algorithm takes geo-referenced genotype data as input and simulates migration across a grid under a stepping-stone model, returning a spatial depiction of estimated historical rates of gene flow. We prepared the SNP dataset by following the same procedure as for ADMIXTURE, but excluded some samples prior to relatedness and LD filtering. The samples we removed were Bayira, an ancient sample, the Somali and Sudanese populations from Pagani et al. 2012,[52] due to lack of specific geographical information, and the Batwa, Bakiga, Biaka, Mbuti, Hadza, Sandawe, Palestinians, and Yoruba, due to the fact that these groups live outside the bounds of our geographic region of interest. We also applied a stricter individual missingness filter of 5%, which excluded all 5 Bari individuals. This left a dataset of 116,447 SNPs and 658 individuals across 35 populations. We primarily used the coordinates in the original publications with some adjustments; as EEMS is a spatial analysis based on historical population locations, the coordinates for Tigray individuals were changed from their sampling location near Addis Ababa to their traditional homeland using the Glottolog coordinates for Tigrinya.[76] For similar reasons, we also excluded the 3 Hausa individuals given the population's recent migration from outside the region of interest within the last 100 years.[77] However, we did not find that either of these changes led to major qualitative differences in the results. We performed 12 total runs under a range of starting parameters (number of demes specified as 200, 300, 400, and 500, each under three different starting seed values) and averaged the results to mitigate the possible bias of any single run. Each run was allowed to proceed for 30 million MCMC iterations to ensure convergence, with the first 15 million discarded as burn-in and the remaining 15 million thinned to retain 1 out of every 15,000 data points. Proposal variances were tuned so that proposals were accepted between 20% and 30% of the time for all runs.

Estimating the distributions of major language families

In order to determine the correspondence between EEMS-inferred migration barriers and corridors and linguistic boundaries, we calculated kernel estimates of language family distributions using the adaptive radius local convex hull (a-LoCoH) method.[54] Language centroid point data and (Greenberg-based) family classifications for every known living African language were obtained from Ethnologue (https://www.ethnologue.com/). We then applied the a-LoCoH algorithm to construct 'utilization distributions' for each of the five major African language families (Niger-Congo, Afro-Asiatic, Nilo-Saharan, Khoisan, and Austronesian), using values equal to the longest geodesic distance between any two languages in a family, to accommodate variable point densities. This produced a set of layered isopleths for each language family representing decile occurrence probabilities. These isopleths were then plotted with overlaid language point data to visualize the extent and density of language distributions by family in relation to historical migration rate estimates as determined by EEMS (Figure S2). Putative linguistic isolates were determined according to Kibebe and Blench.[8,78]

Runs of homozygosity

We determined runs of homozygosity (RoH) in the autosomes of the Ethiopian populations and other African hunter-gatherers, the Hadza, Sandawe, Batwa, Biaka, and Mbuti. The Hadza and Sandawe were assayed on the Illumina 550k array.[25] We removed SNPs with more than 5% missingness or a less than 1% MAF from all datasets and removed SNPs that were not in Hardy-Weinberg equilibrium within each population ($p < 0.001$).[57,65,66] We also thinned each dataset to approximately match the number of SNPs in the smallest dataset, leaving approximately 470,000 SNPs per population for analysis. We then identified RoH in each individual using PLINK, defining a run as having at least 30 SNPs and being at least 500 kb in length, allowing for no more than two missing and one heterozygous SNP per run.[57,65,66]

Despite varying these parameters, we found many instances of two RoH within a single individual closely flanking a low SNP density region. We chose to join such segments post hoc with a custom script (see 'data and code availability') by defining low density regions as 1Mb windows that fell in the lower 5% of SNP count when compared to the entire genome. We also observed genome regions where unusually high numbers of individuals in a population carried a RoH segment. We defined such outlier regions as being more than three standard deviations above the mean depth of RoH in the population. We added these regions to a list of previously identified low density regions and known low complexity regions (i.e. heterochromatin, telomere, centromere, and short arm regions). We then removed all RoH segments that overlapped by 85% or more with one of these regions using a custom script (see 'data and code availability'). However, we found that none of these post hoc adjustments made a qualitative difference to RoH distributions at the population level.

Genomic autozygosity regions likelihood-based inference (GARLIC)

In order to analyze RoH in separate classes corresponding to the relative age of the events that produced them, we also identified RoH using GARLIC.[55] This algorithm implements a population model-based method of inferring RoH in 'short', 'intermediate' and 'long' size classes.[79] We ran GARLIC on the datasets used for PLINK RoH analysis using the following parameters: 'error' of 0.001, 'winsize' of 30, 'auto-winsize', and 'auto-winsize-step' of 5. As with PLINK RoH, we then joined segments that flanked regions of low SNP density, and updated their size class accordingly.

Historical effective population size inference

In order to infer segments of the genome shared IBD across individuals in a population, we first performing phasing using SHAPEIT2. We started with the unthinned RoH datasets described above and phased all individuals assayed on a given platform together using a reference panel of phased individuals from the 1000 Genomes project Phase 3 dataset the –duohmm option, and a window size of 5 Mb.[59,62] We converted the output of SHAPEIT2 to vcf file and then used hap-ibd v1.0 to identify tracts shared identical-by-descent (IBD) across all individuals in a given dataset.[56] We then 'repaired' these IBD segments using the merge-ibd-segments script.[80] We filtered out any IBD segments that were shorter than 4 centiMorgans, and used these to estimate historical N_e with the 2015 version of IBDNe.[23]

Demographic simulations

In order to evaluate the effects of sample size and demographic history on IBDNe inference, we performed a series of simulations using *msprime* with an African-American recombination map and a mutation rate of 1×10^{-8}.[24,81,82] We used a standard coalescent model until 100 generations ago (ga), at which point our simulations switch to using a discrete time Wright-Fisher model. The latter has been shown to produce more realistic and unbiased patterns of recent IBD, which is especially important for our application.[83] We simulated 3 basic demographic scenarios: a population with N_e declining from 10,000 to 1,000 starting 50 ga, a population holding steady at 5,000, and a population increasing from 5,000 to 10,000 starting 10 ga. We also simulated scenarios where a single pulse of gene flow (either 10% or 25%) occurs at 10, 30, and 50 ga from a population that diverged 600 ga. In order to emulate the effect of ascertainment bias, we used a custom script (see 'data and code availability') to filter our simulated genotype data to approximately match the allele frequency distribution of the 1000 Genomes project Phase 3 dataset and the SNP density of the MEGA array dataset in non-overlapping 1 Mb windows.[62] We inferred IBD and ran IBDNe on the resulting dataset as described above.

Gene flow can obscure the true patterns of historical N_e; sample size, admixture proportions, and timing all affect correct estimation. Specifically, gene flow caused an inflation of the estimate of N_e, leading to an overall trajectory that looks like decline or fluctuation (Figure S4). This is expected given that gene flow with a divergent group will introduce new, unrelated haplotypes and reduce the extent of IBD sharing in the population. We found that, in general, the severity of these errors declined with increased age of the gene flow event, decreased gene flow proportion, and increased sample size. Finally, we found that SNP ascertainment had essentially no effect on the accuracy of our N_e estimates (see 'data and code availability').

Political boundaries in maps

The boundaries depicted in the maps do not imply the expression of an opinion by any of the authors of this paper regarding the legal status or political boundaries of any country or territory.

Current Biology, Volume *32*

Supplemental Information

Hunter-gatherer genomes reveal diverse demographic trajectories during the rise of farming in Eastern Africa

Shyamalika Gopalan, Richard E.W. Berl, Justin W. Myrick, Zachary H. Garfield, Austin W. Reynolds, Barnabas K. Bafens, Gillian Belbin, Mira Mastoras, Cole Williams, Michelle Daya, Akmel N. Negash, Marcus W. Feldman, Barry S. Hewlett, and Brenna M. Henn

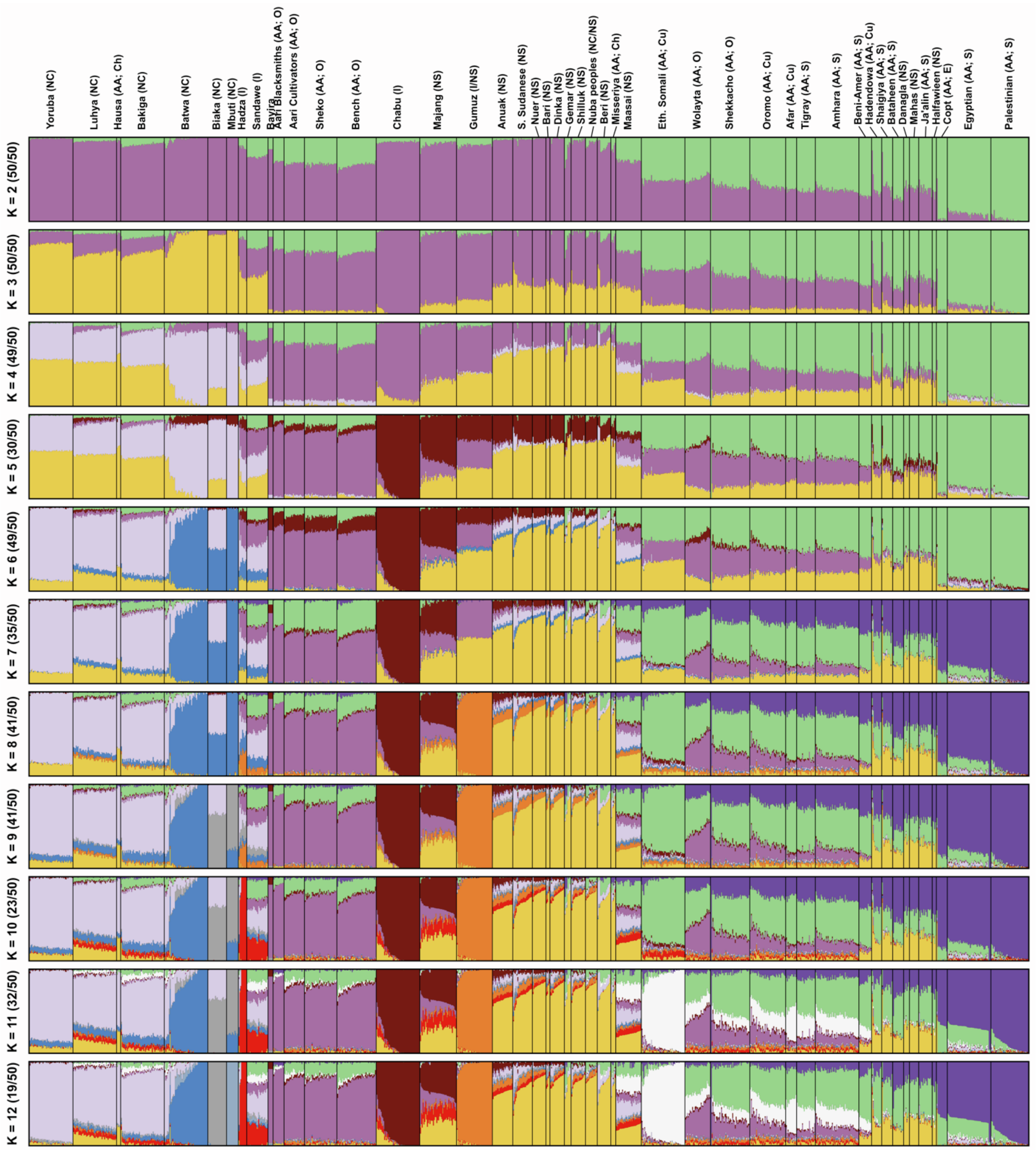


Figure S1. Global Ancestry Proportions of Individuals for K=2 to K=12, Related to Figure 1 and STAR Methods.

Global ancestry proportions of individuals inferred from unsupervised clustering of genotype data of Eastern and Central African individuals, the Yoruba from Nigeria, and Palestinians, assuming values of K from 2 to 12. For each K, the most prevalent pattern (major mode) that appeared over 50 replicates is plotted. Each color corresponds to a genetic component and each vertical bar represents one individual (Bayira is plotted 5 times wider than other individuals for visualization purposes). Population labels include linguistic codes in brackets; Afro-Asiatic (AA), Niger-Congo (NC), Nilo-Saharan (NS), linguistic isolates (I). Within Afro-Asiatic speakers, we further differentiate between Chadic (Ch), Cushitic (Cu), Egyptian (E), Omotic (O), and Semitic (S) speakers.

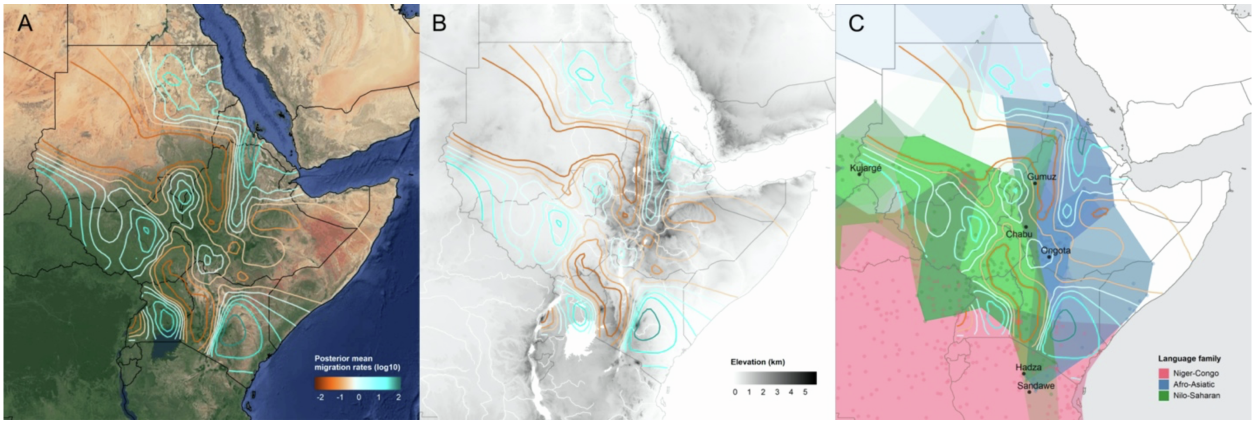


Figure S2. **Visualization of Effective Migration Rate Estimates Over Geographic and Linguistic Features, Related to Figure 1 and STAR Methods.**
Effective migration surfaces are depicted as contour lines over A) satellite imagery, B) elevation and water features, and C) the geographic distribution of major Eastern African language families. For contour lines, cool colors correspond to effective migration corridors, while warm colors correspond to effective migration barriers. A) Some geographic features, such as the Libyan Desert (Sahara) in northwestern Sudan, the northwestern Ethiopian Highlands, the East African Rift, and Nalubaale/Lolwe/Nyanza (Lake Victoria), correspond with historical migration barriers. Map data from Google; imagery © 2018 TerraMetrics. B) Some regions of high elevation or roughness, such as the northwestern Ethiopian Highlands and the volcanic range along the East African Rift, correspond with historical migration barriers. However, the northeastern Ethiopian Highlands and valleys feature high estimated rates of historical migration, potentially as a preferred route around the inhospitable Danakil Depression and Afar Triangle to the east. Elevation data at 30 arc-second (~1 km) resolution are from the U.S. Geological Survey GOTOPO30 digital elevation model, accessed through EarthExplorer (https://earthexplorer.usgs.gov/). Physical vectors for rivers and lakes at 1:10 million scale are from Natural Earth (https://www.naturalearthdata.com/). C) Darker isopleths represent higher occurrence probabilities for languages within each family. Putative linguistic isolates, including the Chabu, are depicted as black points. Language data are from Ethnologue (www.ethnologue.com/).

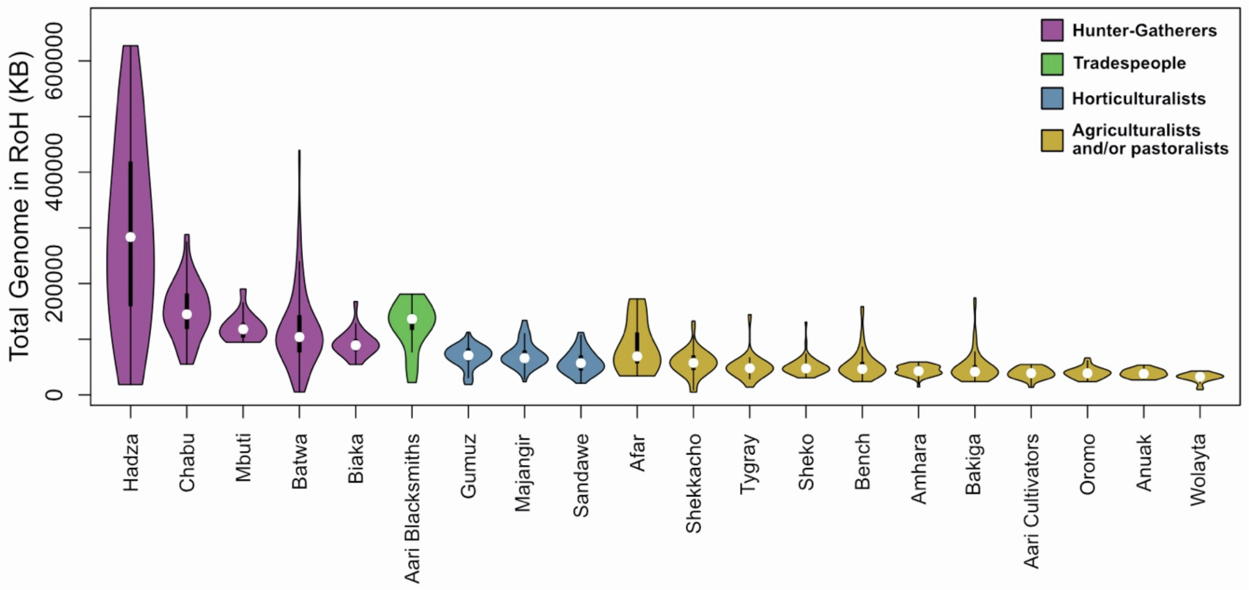


Figure S3. Runs of Homozygosity (RoH) Across Economic Strategies, Related to Figure 3. Distributions of the total amount of the genome in runs of homozygosity in Ethiopian populations and eastern/central African hunter-gatherers. The white point represents the median value of the distribution, and the black rectangle represents values between the lower and upper quartiles. The thin black "whiskers" extend to data points that lie within 1.5 times the interquantile range below or above the lower and upper quantiles, respectively. Populations are grouped according to their primary subsistence or economic strategy. African hunter-gatherer descendant groups vary greatly with respect to total RoH, suggesting significant heterogeneity in the degree of population isolation and decline experienced by these groups in the past.

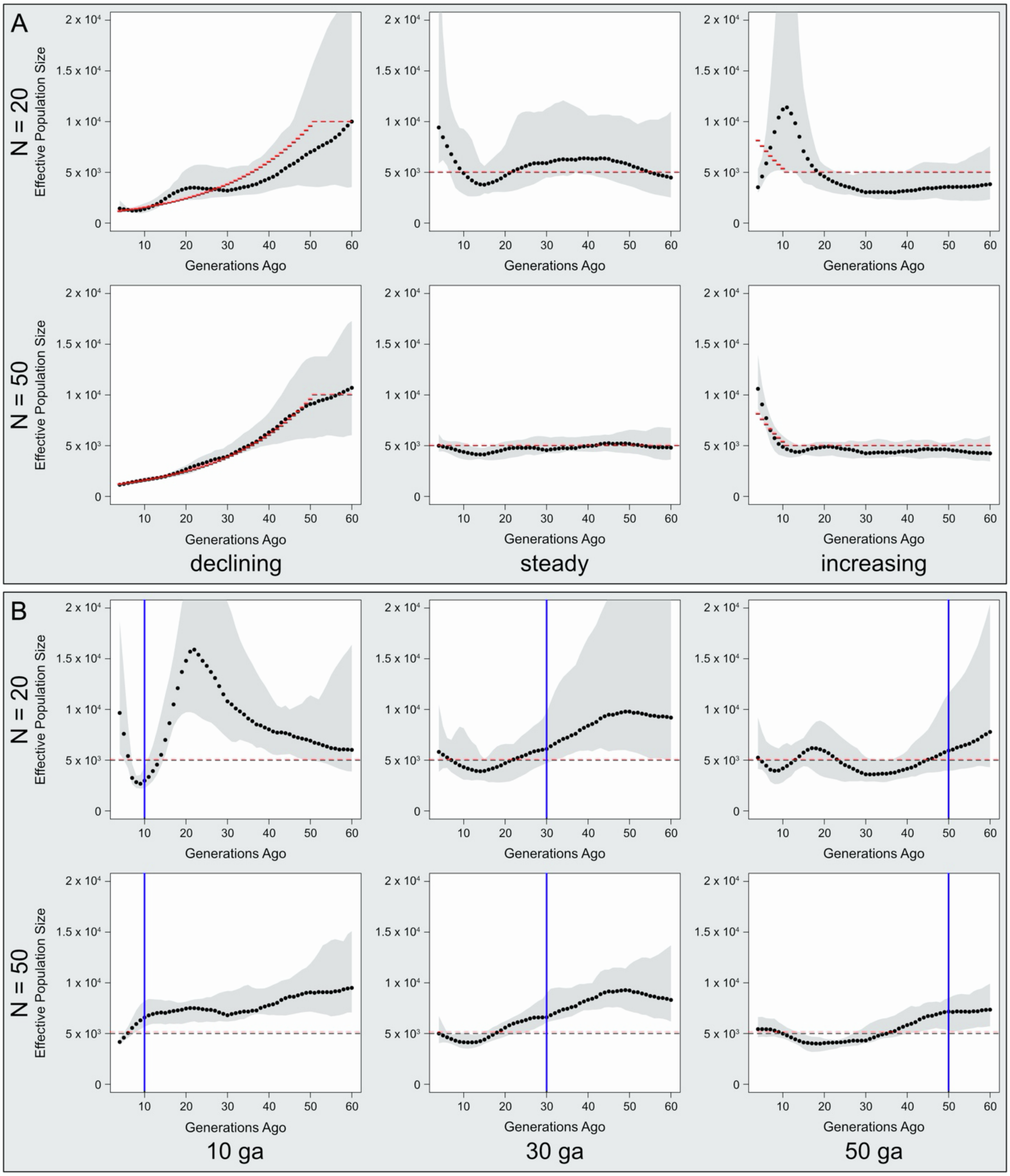


Figure S4. Population Size Histories Estimated from Simulated Genetic Data, Related to STAR Methods.

Effective population sizes (N_e) estimated using IBDNe from 4 to 60 generations ago (ga) for various demographic histories both A) without and B) with gene flow from a diverged population (STAR methods). The true Ne, as simulated *msprime* is depicted by a red dashed line. The estimated Ne at a given generation is represented by a filled circle. Colored ribbons indicate bootstrapped 95% confidence intervals around these estimates. Some of these were cropped for visualization purposes. B) The time of gene flow (admixture proportion of 25%) is indicated by a blue vertical line.

Population	**Sample Size**	**Language Family**	**Latitude**	**Longitude**	**Source**
Aari Blacksmith	17	Afro-Asiatic	6	37	Pagani et al. 2012
Aari Cultivator	24	Afro-Asiatic	6	37	Pagani et al. 2012
Afar	12	Afro-Asiatic	12	41	Pagani et al. 2012
Amhara	51	Afro-Asiatic	10	39	Pagani et al. 2012, Pagani et al. 2015
Anuak	23	Nilo-Saharan	8	34	Pagani et al. 2012
Bakiga	61	Niger-Congo	-1	29.7	Perry et al. 2014
Bari	5	Nilo-Saharan	4.8	31.6	Hollfelder et al. 2017
Bataheen	16	Afro-Asiatic	15.9	33.8	Hollfelder et al. 2017
Batwa	169	Niger-Congo	-1	29.7	Perry et al. 2014
Bench	48	Afro-Asiatic	7	35.6	**Present study**
Beni-Amer	16	Afro-Asiatic	15.4	36.4	Hollfelder et al. 2017
Beri	16	Nilo-Saharan	13.4	22.4	Hollfelder et al. 2017
Biaka	22	Niger-Congo	NA	NA	Bergström et al. 2020
Chabu	83	Isolate	7.5	35.2	**Present study**
Copts	16	Afro-Asiatic	15.6	32.5	Hollfelder et al. 2017
Danagla	16	Nilo-Saharan	19.1	30.5	Hollfelder et al. 2017
Dinka	16	Nilo-Saharan	9.5	31.6	Hollfelder et al. 2017
Egyptian	100	Afro-Asiatic	NA	NA	Pagani et al. 2015
Ethiopian Somali	65	Afro-Asiatic	NA	NA	Pagani et al. 2012, Pagani et al. 2015
Gemar	7	Nilo-Saharan	12	24.9	Hollfelder et al. 2017
Gumuz	44	Isolate/Nilo-Saharan	11	35.9	Pagani et al. 2012, Pagani et al. 2015
Hadendowa	14	Afro-Asiatic	19.6	37.2	Hollfelder et al. 2017
Hadza	21	Isolate	-3.6	34.7	Henn et al. 2011
Halfawieen	11	Nilo-Saharan	21.8	31.4	Hollfelder et al. 2017
Hausa	7	Afro-Asiatic	14	35.4	Hollfelder et al. 2017
Ja'alin	15	Afro-Asiatic	16.7	33.4	Hollfelder et al. 2017
Luhya	116	Niger-Congo	0.6	34.8	Auton et al. 2015
Maasai	31	Nilo-Saharan	-1.5	35.2	Auton et al. 2015
Mahas	16	Nilo-Saharan	19.6	30.4	Hollfelder et al. 2017
Majang	49	Nilo-Saharan	7.2	35.3	**Present study**
Mbuti	16	Niger-Congo/Nilo-Saharan	1	29	Mallick et al. 2016, Bergström et al. 2020
Misseriya	8	Afro-Asiatic	12.4	27.4	Hollfelder et al. 2017
Bayira (Mota)	1	Unknown	NA	NA	Gallego Llorente et al. 2015

Nuba peoples	16	Niger-Congo/Nilo-Saharan	11	29.7	Hollfelder et al. 2017
Nuer	16	Nilo-Saharan	9.2	29.8	Hollfelder et al. 2017
Oromo	46	Afro-Asiatic	8	37	Pagani et al. 2012, Pagani et al. 2015
Palestinian	46	Afro-Asiatic	NA	NA	Bergström et al. 2020
Sandawe	35	Isolate	-4.5	35.5	Henn et al. 2011
Shaigiya	15	Afro-Asiatic	18.7	31.9	Hollfelder et al. 2017
Shekkacho	46	Afro-Asiatic	7.2	35.4	**Present study**
Sheko	50	Afro-Asiatic	7	35.5	**Present study**
Shilluk	16	Nilo-Saharan	9.5	31.6	Hollfelder et al. 2017
Tigray	21	Afro-Asiatic	15.3	38.9	Pagani et al. 2012
Wolayta	32	Afro-Asiatic	6	37	Pagani et al. 2012, Pagani et al. 2015
Yoruba	189	Niger-Congo	NA	NA	Auton et al. 2015

Table S1. Linguistic and Geographic Data for All Analyzed Populations, Related to Figure 1 and STAR Methods.
Latitude and longitude coordinates indicate the average reported sampling location, except for Tigray, where coordinates indicate the Glottolog coordinates for the Tigrinya language. Populations with NA in the latitude and longitude columns were not used in any spatial analysis (i.e. geographic plotting of ADMIXTURE components or EEMS).

				F3 test		ALDER test		
Target Population	**Source Pop. 1**	**Source Pop. 2**	**SNPs**	**Z**	**F3 Statistic**	**Test Status**	**Estimated Time (generations ago)**	**p-value**
Aari Blacksmiths	Chabu	Afar	309532	33.71	0.0166	succeeded*	111.30 +/- 13.79	0.00043†
		Amhara	312787	37.49	0.0174	succeeded*	93.70 +/- 10.31	8.1 x $10^{-11\dagger}$
		Oromo	312355	41.11	0.0188	succeeded*	104.60 +/- 9.52	4.4 x $10^{-28\dagger}$
		Shekkacho	313248	40.84	0.0189	succeeded*	93.36 +/- 13.68	2.2 x $10^{-7\dagger}$
		Wolayta	311241	43.86	0.0193	succeeded*	99.24 +/- 13.25	3.0 x $10^{-11\dagger}$
	Bayira	Afar	836235	6.62	0.0045	failed		
		Amhara	867126	11.92	0.0071	failed		
		Oromo	864011	16.89	0.0101	failed		
		Shekkacho	311969	17.12	0.0106	failed		
		Wolayta	851477	20.81	0.0121	failed		
Aari Cultivators	Chabu	Afar	311024	**-7.44**	**-0.0019**	failed		
		Amhara	313308	**-4.19**	**-0.0009**	failed		
		Oromo	312934	2.31	0.0005	failed		
		Shekkacho	313754	3.89	0.0009	failed		
		Wolayta	312150	7.11	0.0013	failed		
	Bayira	Afar	854453	**-28.77**	**-0.0121**	failed		
		Amhara	873087	**-26.91**	**-0.0093**	failed		
		Oromo	870604	**-18.84**	**-0.0063**	failed		
		Shekkacho	312759	**-18.01**	**-0.0065**	failed		
		Wolayta	863182	**-14.08**	**-0.0042**	failed		
Bench	Chabu	Afar	303126	**-14.49**	**-0.0033**	failed		
		Amhara	438572	**-22.95**	**-0.0039**	succeeded*	52.69 +/- 4.08	4.4 x $10^{-13\dagger}$
		Oromo	437542	**-17.04**	**-0.0027**	succeeded*	50.60 +/- 5.20	4.0 x $10^{-8\dagger}$
		Shekkacho	817001	**-19.97**	**-0.0031**	succeeded*	57.95 +/- 4.48	3.2 x $10^{-38\dagger}$
		Wolayta	436616	**-9.39**	**-0.0013**	succeeded*	77.62 +/- 15.46	5.2 x $10^{-7\dagger}$
	Bayira	Afar	301605	**-28.02**	**-0.0125**	failed		
		Amhara	435164	**-34.85**	**-0.0112**	succeeded*	59.57 +/- 4.49	2.2 x $10^{-20\dagger}$
		Oromo	433795	**-30.37**	**-0.0091**	succeeded*	61.42 +/- 5.68	3.3 x $10^{-27\dagger}$
		Shekkacho	553683	**-33.11**	**-0.0096**	succeeded*	58.76 +/- 7.01	5.1 x $10^{-9\dagger}$
		Wolayta	432602	**-21.25**	**-0.0056**	succeeded*	89.64 +/- 18.88	2.1 x $10^{-6\dagger}$
Chabu	Bayira	Anuak	303372	28.96	0.0143	succeeded*	30.34 +/- 5.37	1.6 x $10^{-8\dagger}$
		Dinka	462369	30.8	0.0143	failed		
		Majang	529068	40.49	0.0143	failed		
		Nuer	467195	29.91	0.0144	succeeded*	43.72 +/- 19.61	0.026
Gumuz	Chabu	Anuak	307894	65.2	0.0138	failed		
		Dinka	407810	58.55	0.0134	failed		
		Nuer	408762	63.35	0.0135	succeeded*	65.30 +/- 21.65	0.0026
	Bayira	Anuak	855628	26.23	0.0093	failed		
		Dinka	1587324	23.09	0.0073	failed		
		Nuer	1591360	23.4	0.0074	failed		
Majang	Chabu	Anuak	307676	**-3.15**	**-0.0004**	failed		
		Dinka	486772	**-6.26**	**-0.0008**	failed		
		Nuer	489934	**-5.47**	**-0.0007**	failed		
	Bayira	Anuak	306269	**-0.85**	**-0.0003**	succeeded*	5.51 +/- 2.38	0.021
		Dinka	480350	**-2.37**	**-0.0008**	failed		
		Nuer	483661	**-1.66**	**-0.0006**	failed		
Shekkacho	Aari B.	Afar	312737	8.22	0.0019	succeeded*	8.35 +/- 1.02	2.3 x $10^{-16\dagger}$
		Amhara	313749	4.86	0.0007	failed		
		Oromo	313561	20.91	0.0033	failed		
		Wolayta	313279	48.05	0.0096	succeeded*	7.80 +/- 3.52	0.027
	Aari C.	Afar	313282	12.09	0.0024	succeeded*	9.20 +/- 0.95	2.7 x $10^{-22\dagger}$
		Amhara	314145	7.66	0.001	failed		
		Oromo	313956	26.66	0.0036	failed		
		Wolayta	313721	57.15	0.0095	succeeded*	8.28 +/- 2.38	0.00051
	Bench	Afar	304566	**-2.58**	**-0.0004**	succeeded*	7.82 +/- 1.09	5.8 x $10^{-13\dagger}$
		Amhara	441707	**-1.23**	**-0.0001**	failed		
		Oromo	441194	20.18	0.0022	failed		
		Wolayta	441064	54.6	0.0077	succeeded*	11.64 +/- 1.79	8.0 x $10^{-11\dagger}$
	Chabu	Afar	313742	**-0.94**	**-0.0002**	succeeded*	7.78 +/- 3.64	0.033
		Amhara	457097	**-3.15**	**-0.0005**	failed		
		Oromo	456500	17.13	0.003	failed		
		Wolayta	456267	46.46	0.0098	succeeded*	20.27 +/- 9.58	0.034

	Bayira	Afar	312497	**-6.37**	**-0.0024**	succeeded*	7.50 +/- 2.39	0.0017
		Amhara	454163	**-6.67**	**-0.0015**	failed		
		Oromo	453376	12.76	0.0029	failed		
		Wolayta	453107	40.13	0.0117	succeeded*	16.69 +/- 3.96	2.5 x $10^{-5\dagger}$
	Sheko	Afar	304469	**-1.96**	**-0.0003**	succeeded*	7.43 +/- 1.49	6.6 x $10^{-7\dagger}$
		Amhara	441484	**-1.13**	**-0.0001**	failed		
		Oromo	440926	20.06	0.0023	failed		
		Wolayta	440735	54.32	0.0079	succeeded*	10.71 +/- 2.06	2.1 x $10^{-7\dagger}$
Sheko	Chabu	Afar	302726	**-10.81**	**-0.0025**	failed		
		Amhara	438083	**-16.99**	**-0.003**	succeeded*	63.15 +/- 6.47	4.7 x $10^{-12\dagger}$
		Oromo	436856	**-11.4**	**-0.0019**	succeeded	60.46 +/- 7.15	4.5 x $10^{-9\dagger}$
		Shekkacho	815983	**-13.72**	**-0.0022**	succeeded*	61.35 +/- 4.85	1.1 x $10^{-36\dagger}$
		Wolayta	435764	**-3.44**	**-0.0005**	succeeded*	63.19 +/- 5.90	8.5 x $10^{-27\dagger}$
	Bayira	Afar	300894	**-25.59**	**-0.0115**	failed		
		Amhara	434267	**-31.88**	**-0.0101**	failed		
		Oromo	432637	**-26.43**	**-0.0081**	failed		
		Shekkacho	552868	**-29.78**	**-0.0086**	failed		
		Wolayta	431171	**-17.63**	**-0.0047**	failed		
Wolayta	Aari B.	Afar	865487	**-11.62**	**-0.0023**	failed		
		Amhara	876539	**-26.32**	**-0.0038**	succeeded	37.31 +/- 2.74	2.7 x $10^{-42\dagger}$
		Oromo	875039	**-9.28**	**-0.0013**	succeeded*	41.62 +/- 4.25	4.8 x $10^{-9\dagger}$
		Shekkacho	313279	**-22.03**	**-0.0033**	succeeded*	26.37 +/- 2.79	3.1 x $10^{-21\dagger}$
	Aari C.	Afar	872289	**-9.95**	**-0.0016**	failed		
		Amhara	880504	**-26.13**	**-0.0034**	succeeded*	28.58 +/- 6.03	2.1 x $10^{-6\dagger}$
		Oromo	879167	**-6.13**	**-0.0008**	succeeded*	27.99 +/- 4.93	1.3 x $10^{-8\dagger}$
		Shekkacho	313721	**-27.25**	**-0.0031**	succeeded*	14.40 +/- 6.27	0.022
	Bench	Afar	303520	**-30.42**	**-0.0044**	failed		
		Amhara	437939	**-45.8**	**-0.0041**	failed		
		Oromo	437233	**-21.81**	**-0.0021**	failed		
		Shekkacho	441064	**-7.7**	**-0.0009**	failed		
	Chabu	Afar	312203	**-26.32**	**-0.0063**	failed		
		Amhara	452664	**-44.99**	**-0.0066**	succeeded*	36.33 +/- 6.58	7.0 x $10^{-5\dagger}$
		Oromo	451727	**-21.54**	**-0.0034**	succeeded	41.99 +/- 4.58	1.7 x $10^{-8\dagger}$
		Shekkacho	456267	**-15.1**	**-0.0027**	succeeded*	36.85 +/- 8.22	0.00063
	Bayira	Afar	857831	**-24.4**	**-0.0091**	failed		
		Amhara	1835189	**-35.79**	**-0.0082**	succeeded	41.12 +/- 7.67	4.8 x $10^{-5\dagger}$
		Oromo	1825232	**-13.91**	**-0.0038**	succeeded*	42.01 +/- 8.32	$0.00023^{\dagger}$
		Shekkacho	453107	**-17.69**	**-0.0046**	failed		
	Sheko	Afar	303227	**-30.34**	**-0.0044**	failed		
		Amhara	437439	**-44.71**	**-0.0043**	failed		
		Oromo	436578	**-22.58**	**-0.0022**	failed		
		Shekkacho	440735	**-8.85**	**-0.0011**	failed		

Table S2. F3 Admixture Test Results for Southwest Ethiopian Populations and the Gumuz, Related to STAR Methods. For each test, we hypothesized two sources of ancestry for the target population and computed the F3 statistic using a genome-wide set of SNPs. Negative F3 and Z values (bolded) indicate evidence of gene flow. An asterisk in the test status label (i.e. ‘succeeded*’) indicates that the ALDER test succeeded with a warning that the decay rates were inconsistent. A dagger symbol (†) next to a p-value indicates that it is significant after Bonferroni correction for multiple tests.