

1 **Hunter-gatherer genomes reveal diverse demographic trajectories following the rise of farming in East**

2 **Africa**

3

4 Shyamalika Gopalan^{1†}, Richard E.W. Berl^{2†}, Gillian Belbin³, Christopher R. Gignoux⁴, Marcus W.

5 Feldman⁵, Barry S. Hewlett^{*6}, Brenna M. Henn^{*7,8}.

6

7 1 Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY 11794, USA.

8 2 Department of Human Dimensions of Natural Resources, Colorado State University, Fort Collins, CO

9 80523, USA.

10 3 Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.

11 4 Department of Medicine, University of Colorado, Anschutz Medical Campus, Aurora, CO 80045, USA.

12 5 Department of Biology, Stanford University, Stanford, CA 94305, USA.

13 6 Department of Anthropology, Washington State University, Vancouver, WA 98686, USA.

14 7 Department of Anthropology, University of California, Davis, CA 95616, USA.

15 8 UC Davis Genome Center, University of California, Davis, CA 95616, USA.

16 * Corresponding author. E-mail: hewlett@wsu.edu; bmhenn@ucdavis.edu.

17 † These authors contributed equally to this work.

18

19 **Abstract:**

20 A major outstanding question in human prehistory is the fate of hunting and gathering populations
21 following the rise of agriculture and pastoralism. Genomic analysis of ancient and contemporary
22 Europeans suggests that autochthonous groups were either absorbed into or replaced by expanding
23 farmer populations. Many of the hunter-gatherer populations persisting today live in Africa, perhaps
24 because agropastoral transitions occurred later on the continent. Here, we present the first genomic
25 data from the Chabu, a relatively isolated and marginalized hunting-and-gathering group from the
26 Southwestern Ethiopian highlands. The Chabu are a distinct genetic population that carry the highest
27 levels of Southwestern Ethiopian ancestry of any extant population studied thus far. This ancestry has
28 been in situ for at least 4,500 years. We show that the Chabu are undergoing a severe population
29 bottleneck which began around 40 generations ago. We also study other Eastern African populations
30 and demonstrate divergent patterns of historical population size change over the past 60 generations
31 between even closely related groups. We argue that these patterns demonstrate that, unlike in Europe,
32 Africans hunter-gatherers responded to agropastoralism with diverse strategies.

33

34 Since the beginning of the Holocene 12,000 years ago (ya), the dominant mode of human subsistence
35 has shifted from hunting and gathering to agriculture through a process known as the Neolithic
36 transition. Whether this transition occurred primarily through the mass movement of people from
37 centers of domestication (demic diffusion) or through the cultural transmission of agricultural practices
38 (cultural adoption) is still debated in archaeology, genetics and anthropology (1, 2). As this transition
39 largely concluded by 4,000 ya in Europe and Asia, there remains little direct evidence of on-the-ground
40 interaction between hunter-gatherers and agriculturalist migrants. Theoretically, in the face of
41 displacement and conflict over resources, hunter-gatherer populations might respond in a variety of
42 ways: 1) intermarry with the migrant group and adopt their agropastoral subsistence practices
43 (substantial genetic exchange); 2) adopt the subsistence practices of the migrants without intermarriage
44 (limited genetic exchange); 3) reduce their geographic range or resource acquisition (leading to a decline
45 in population size); 4) enter into an economic-symbolic exchange relationship with the migrant group; or
46 5) move to an ecological region that is marginal for pastoralism or agriculture (3, 4). These are not
47 mutually exclusive; the history of any particular hunter-gatherer group may involve multiple modes of
48 response. In Europe, the current consensus is that early Near Eastern farmers facilitated the spread of
49 agriculture, completely replacing groups in some areas, and in others – particularly in southern Europe –
50 admixing with hunter-gatherers (5–8). However, the extent to which the Neolithization processes of
51 Europe occurred on other continents remains unclear. In Africa, the genetic and cultural landscape has
52 been significantly shaped by recent expansions of agriculture and pastoralism, such as the Bantu
53 migration (9, 10), but a lack of data precludes detailed characterization of these expansion events.

54 We sought to understand the demographic impact of the expanding Neolithic in Southwest
55 Ethiopia, a region of high ethnic and linguistic diversity that is home to several of the world's remaining
56 hunter-gatherer groups, but has a relatively sparse archaeological record (11). We present new genomic
57 data from the Chabu hunter-gatherers and their immediate neighbors, the Majangir and Shekkacho,
58 which we analyze together with other East African populations. The Chabu people are poorly known
59 even among anthropologists. They inhabit the highland forests that straddle the border between the
60 Oromia Regional State, Gambella Regional State, and Southern Nations, Nationalities, and Peoples'
61 Region (SNNPR) (12). Recent estimates of their census size range between 1,700 and 2,500 individuals
62 (12). The Chabu claim to be the original inhabitants of these forests, an assertion that their Majangir and
63 Shekkacho neighbors generally support. The Chabu are thought to be a linguistic isolate, while the
64 Majangir are Nilo-Saharan Surmic-speakers who practice small-scale cultivation and limited hunting and

65 gathering and the Shekkacho are Afro-Asiatic Omotic-speakers who practice intensive agriculture (13–
66 15).

67 We investigate two alternative hypotheses regarding the origins of the Chabu. The first is that
68 the Chabu are the descendants of an earlier population that occupied Southwest Ethiopia prior to
69 encroachment by agriculturalists. Archaeological research from Ajilak, roughly 100 km west of the Chabu
70 forest and 500 m lower in elevation, strongly suggests that hunter-gatherers were present in the area as
71 late as 800-1,000 years ago (16). The presence of iron tools from Ajilak indicates that foraging groups
72 interacted with lowland Sudanese pastoralist populations in the past. Based on archaeological and
73 ethnographic evidence, it has previously been suggested that the Chabu, the Majangir, and Koman-
74 speakers could be the living descendants of the Ajilak hunter-gatherers (16). A competing hypothesis is
75 that the Chabu were previously an agricultural or agropastoral group that transitioned to foraging in
76 order to exploit an unused ecological niche in the forest highlands. This hypothesis implies close
77 relatedness between the Chabu and another Ethiopian or Sudanese population, from which they would
78 have diverged relatively recently. Although a transition to hunting and gathering from other subsistence
79 modes is presumed to be uncommon historically, it has occurred more than once (17, 18). In Eastern
80 Africa specifically, some primarily pastoralist groups are known to fluctuate between subsistence
81 strategies (19, 20).

82 In order to test whether the Chabu are descendants of earlier hunter-gatherer inhabitants, we
83 performed unsupervised clustering of autosomal single nucleotide polymorphism (SNP) data from
84 populations of Ethiopia, Sudan, South Sudan, and Kenya (Methods). Importantly, we also included
85 genomic data from Mota, a 4,500 year old individual found in the nearby Gamo highlands who lived
86 prior to any evidence of agriculture or pastoralism in the region (21). In the clustering algorithm, we
87 varied 'K', the hypothesized number of ancestral source populations for the dataset, from 2 to 12. We
88 focus on the pattern that emerges at K=5, and refer to these five genetic components as 'Nilo-Saharan',
89 'Afro-Asiatic', 'Southwestern Ethiopian', 'Niger-Congo', and 'Near Eastern' ancestries based on their
90 frequencies across geographic space and language families (Fig. 1). The Chabu and their neighbors are
91 primarily characterized by differences in frequency of the first three components. Nilo-Saharan ancestry
92 is most concentrated in southern Sudan and South Sudan, overlapping slightly with western Ethiopia
93 (i.e. at high frequency in the Nilo-Saharan-speaking populations of the region) (Fig. 1B). Afro-Asiatic
94 ancestry is highest in northeast Ethiopia, Eritrea, and northeast Sudan, declining steadily towards the
95 south and west (Fig. 1C). Southwestern Ethiopian ancestry is found in all language groups but is
96 concentrated in the Southwestern Ethiopian highlands (Fig. 1D). Importantly, the ancient Mota genome

97 carries a high proportion of this ancestry, suggesting that it has characterized this region for at least
98 4,500 years. Among the contemporary groups we studied, the Chabu carry the highest levels of
99 Southwestern Ethiopian ancestry, as well as moderate levels of Nilo-Saharan ancestry (Fig. 1A). The
100 Majangir and the Gumuz are characterized by the same components, each present at ~50%. The Chabu
101 are distinct from Afro-Asiatic Ethiopian and Sudanese agropastoral populations, as well as from
102 traditionally pastoral Nilotic speakers, who include the Anuak, Dinka, and Shilluk (Fig. 1A). Similarly, in
103 principal component (PC) space, the Chabu anchor the second PC and are closest to Mota, the Majangir,
104 and Gumuz (Fig. S1). Notably, the Majangir and Gumuz are not readily distinguishable from each other in
105 either K=5 ADMIXTURE plots or a biplot of the first two PCs. We conclude that the Chabu carry the major
106 Southwestern Ethiopian genetic component also identified in Mota, and are distinct from agricultural
107 and/or pastoralist groups. Therefore, they are likely to be direct descendants of ancient Southwest
108 Ethiopian hunter-gatherer groups; secondary adoption of hunting-and-gathering is not supported.

109 Despite carrying moderate to high levels of Southwestern Ethiopian ancestry, there are clear
110 differences between the Chabu-Majangir-Gumuz and the Aari-Mota genetic profiles. At K=5, Mota and
111 the Aari carry Afro-Asiatic ancestry but no Nilo-Saharan ancestry. In a biplot of the first two PCs, the Aari
112 populations pull away from the Chabu-Majangir-Gumuz towards the Wolayta, the Afro-Asiatic group
113 who are their nearest neighbors today (Fig S1). These ancestry differences may relate to differential
114 gene flow with neighbors or may reflect ancient patterns of divergence.

115 Next, we modelled spatial population structure by using genetic data to estimate the effective
116 migration surface (EEMS) (22). This analysis reveals corridors of and barriers to migration that closely
117 correspond to the spatial distribution of ancestral components (Fig. 1G). Some migration barriers also
118 correspond closely with major geographic features such as deserts, certain high elevation areas, and
119 bodies of water. However, other major features, such as the Nubian Desert and Northeastern Ethiopian
120 Highlands, do not appear to have been historical barriers to migration (Fig. S2). Furthermore, areas with
121 low rates of migration tend to lie along the boundaries between the Nilo-Saharan, Afro-Asiatic, and
122 Niger-Congo language families, while corridors of high gene flow lie within them (Fig. S2C). Together,
123 these results emphasize the close association between geography and language in determining gene
124 flow between groups (23). Furthermore, the Chabu lie directly in the center of a language contact area
125 with negative effective migration rates, perhaps indicating relative isolation from neighboring groups.

126 During the last decade, the Chabu have faced mounting challenges to their survival. Their land
127 claims are not recognized, and as a result their traditional forests are often violently co-opted for coffee
128 plantations and other development (12). We sought to understand the effects of recent and historical

129 isolation and demographic pressures faced by the Chabu by analyzing runs of homozygosity (RoH)
130 (Methods). Compared to their neighbors, the Chabu carry more of their genome in RoH (Fig. S3). As the
131 Chabu have no cultural tradition of close relative marriage and practice clan exogamy, this demonstrates
132 an increased level of isolation relative to their neighbors (12). We compared the Chabu to other African
133 hunter-gatherers, including the Biaka, Mbuti, and Hadza, as well as recent descendants of hunter-
134 gatherer groups; these include the Majangir, Gumuz, Sandawe, and Aari. The former three groups are
135 primarily small-scale farmers today, but ethnographic studies indicate that they regularly hunted and
136 gathered in the recent past, while the genetic similarity of the Aari to Mota indicate that they are also
137 recent descendants of hunter-gatherers (14, 21, 24–27). Compared to other Ethiopian populations that
138 carry a large proportion of Southwest Ethiopian ancestry, the Chabu have the highest levels of total RoH
139 (Fig. 2). However, the Hadza of Tanzania, who were previously shown to carry the highest proportion of
140 their genomes in RoH (total RoH) among Africans, exceed the Chabu in this measure (Fig S4) (26).
141 Overall, African hunter-gatherer and hunter-gatherer descendant groups vary greatly with respect to
142 total RoH, which suggests that they have experienced different degrees of isolation and/or population
143 size decline in the past. In fact, several hunter-gatherer populations are indistinguishable from farming
144 populations with respect to total RoH (Fig. S5). Interestingly, when using a model-based method of
145 inferring RoH created by ancient, intermediate, and recent events, the Chabu showed significantly
146 elevated total RoH in only the recent (longest) class (Fig. 2). Similar patterns were observed in the Aari
147 Blacksmiths and the Hadza (Fig. 2, S4).

148 Given this signature of recent demographic pressure on the Chabu, Aari Blacksmiths, and Hadza,
149 we sought to estimate precisely when these population declines occurred. We used a non-parametric
150 method of estimating the effective size (N_e) of a population through time using the distribution of
151 segments that are shared identical by descent (IBD) across pairs of individuals (28). In each population,
152 we optimized the parameters used to infer IBD by comparing RoH, and total the total amount of IBD
153 shared between kin, with ‘truth sets’ generated under independent inference methods (Methods, Figure
154 S6). Through this approach, we estimated historical N_e from 4-60 generations ago (ga) and found that
155 the Chabu, Majangir, and Aari Blacksmiths have all faced recent declines in N_e , while the N_e of Aari
156 Cultivators, Gumuz, and Shekkacho have all increased (Fig. 3). The decline in the Chabu and Aari
157 Blacksmiths starting approximately 40 and 50 ga, respectively, is consistent with the RoH results, but the
158 recent Majangir decline starting approximately 50 ga was not suggested by their patterns of RoH. The
159 Biaka, Mbuti, Hadza and Sandawe all have modest estimated N_e 60 ga but end up lower by 4 ga, with

160 this decline varying in severity between all four groups. Furthermore, the N_e of the Hadza and Mbuti
161 both appear to increase initially (Fig. 3).

162 Previous research has shown that European hunter-gatherers that adopted agriculture or were
163 absorbed into agricultural groups experienced population growth (8, 29). Within Africa, the expansion of
164 Bantu-speaking agriculturalists resulted in both the assimilation and extinction of local hunter-gatherers
165 (9). By examining a large number of Eastern African groups, our results demonstrate a diversity of
166 hunter-gatherer responses to the intensification and spread of agropastoralism, of which we discuss the
167 five outlined at the beginning of this paper. The Ethiopian Majangir and Gumuz are primarily farmers
168 today, but regularly hunted and gathered in the recent past; both also exhibit many characteristic
169 features of hunter-gatherer societies such as high degrees of egalitarianism and reciprocity (14, 25).
170 Notably, we show that they have nearly identical ancestry profiles (Fig. 1A, Fig S1). Despite this, their
171 population size trajectories suggest striking historical differences. The Majangir have declined by over
172 85% from 60 to 4 generations ago, while the Gumuz have increased by more than 50% over the same
173 period. Neither group has seen recent demonstrable gene flow from agropastoralist groups. Similarly,
174 the Sandawe from Tanzania, a ‘click-speaking’ isolate that transitioned to agropastoralism within the
175 past 500 years (27), have experienced a pattern of decline that is similar to the Majangir in both
176 trajectory and magnitude. These results may suggest that the ‘late-adopters’, the Majangir and
177 Sandawe, responded to ecological pressure and declining population size by adopting farming with
178 limited gene flow (Response 2).

179 We also observe opposite demographic trends in the Aari Blacksmiths and Aari Cultivators.
180 Previous studies have shown that these two groups diverged within the last 4,500 years, and are both
181 probable descendants of a ‘Mota-like’ hunter-gatherer population (21, 24). Evidence for a recent
182 bottleneck in the Aari Blacksmiths was also reported (24). Our results support these previous findings,
183 and estimate that the decline in the Aari Blacksmiths began approximately 50 ga. We also find that the
184 N_e of the Aari Cultivators was relatively steady while that of the Aari Blacksmiths began to decline, and
185 then rapidly increased by half by 4 generations ago. Today, the Aari Blacksmiths are a marginalized
186 group of craftspeople, who are geographically proximate to the Aari Cultivators and Wolayta cultivators,
187 with whom they engage in mutual economic exchange (Response 4) (24). Archeological evidence for
188 blacksmithing, today considered a marginal activity across southern Ethiopia (as is foraging and eating
189 wild foods), appears in nearby regions between 1,000 and 3,000 years ago (30). It is possible that the
190 genetic divergence of the two Aari populations was associated with the adoption of different cultural

191 practices (i.e. blacksmithing versus farming). These transitions may have influenced the later divergent
192 patterns of historical N_e that we observe (Responses 1, 3).

193 Finally, we show that the Chabu, descendants of earlier Southwestern Ethiopian hunter-
194 gatherers, have experienced a precipitous decline from an ancestral N_e of $\sim 5,700$ to ~ 140 over 56
195 generations. Recent ethnographic work describes the loss of Chabu land to agriculturalists in just the
196 past two decades (12); we hypothesize that this is a continuation of a trend that began 40 ga (Response
197 3). We found that the Chabu and the Aari are descended from Southwestern Ethiopian ancestors that
198 must have once comprised a wide-ranging population (Fig. 1B). Within just the last decade, the Chabu
199 have shown a marked shift in marriage preferences, with an increasing proportion of Chabu men
200 preferring to take a Majangir, Shekkacho, or Amhara spouse (12). This may indicate that the Chabu are
201 moving towards assimilating into neighboring agricultural groups, which may lead to further changes in
202 subsistence and culture (Response 1). We found that, similarly to the Chabu, the Hadza of Tanzania
203 underwent a population bottleneck that accelerated over the past 25 generations, reaching a minimum
204 N_e of ~ 160 . The Hadza today live around Lake Eyasi, an area unsuitable for cultivation or pastoralism,
205 which may explain their continued persistence as hunter-gatherers (Responses 3, 5) (31).

206 We characterize nuanced hunter-gatherer responses to recent cultural and demographic
207 changes associated with the spread of agriculture and pastoralism in Eastern Africa. While a shift to
208 agricultural subsistence has been linked to increases in effective population size (29), we show a
209 corresponding decline in populations that appear to resist this cultural change. Continued ethnographic
210 and genetic work in collaboration with the Chabu and other marginalized groups will provide valuable
211 insights into the interactions between farmers and hunter-gatherers and the drivers of major cultural
212 transitions.

213

214 **Acknowledgments:**

215 The authors would like to thank the Stanford Center for Computational, Evolutionary, and Human
216 Genomics and Dr. Carlos Bustamante for providing funding for data generation, and Alexandra Sockell
217 for her assistance with sample preparation. This work was supported by a L.S.B. Leakey Foundation
218 grant to BSH. REWB was supported by an Exploration Fund grant from The Explorers Club, the NSF IGERT
219 Program in Evolutionary Modeling at Washington State University, and the Max Planck Institute for the
220 Science of Human History.

221

222 **Ethics statement:**

223 REWB collected the Chabu samples after months of ethnographic research with the communities by
224 Samuel Dira (SD) and BSH. The research with the Chabu and their neighbors is part of a larger formal
225 collaborative research and capacity building relationship between the Departments of Anthropology at
226 Hawassa University, Ethiopia (HU) and Washington State University (WSU). The collaboration involves
227 training several HU faculty, such as SD, in the WSU PhD program and cooperative participation in
228 research projects in Southwestern Ethiopia. Authorization for the research was obtained from HU and
229 WSU.

230

231 **Materials and Methods:**

232 ***Collection, Consent, Ethics approval:***

233 Samples from the Chabu, Majangir, and Shekkacho were collected by REWB in May 2013, using
234 Oragene•DISCOVER (OGR-500) kits for the Chabu and generic 5 ml tubes with Norgen preservation
235 solution for the other two groups. Prior approvals for the project were obtained from the leadership of
236 each group, from the School of Behavioral Sciences at Hawassa University (#BS/502/05), and from the
237 Majangir Zone Council of the Gambella Regional State (#901/ጥዳዘ1/ፈ4). Ethical approval for human
238 subjects research was obtained from the Institutional Review Board of Washington State University
239 under proposals #12972 and #13134. Informed consent was obtained from each participant after
240 reading or hearing the approved text translated into their local language and providing their signature,
241 or a fingerprint in lieu of a signature for non-literate participants.

242

243 ***Data generation and processing:***

244 50 individuals each of the Chabu, Majangir, and Shekkacho were genotyped using the Illumina Infinium
245 MultiEthnic Global Array, which assays over 1.7 million genetic markers. Genotypes were initially called
246 using Illumina GenomeStudio software. We removed 5 samples (1 Majangir, 4 Shekkacho) that had a call
247 rate below 90%. Calls for rare variants, defined as those having a minor allele frequency (MAF) < 5%,
248 were then replaced by using zCall following their published procedure (32). Variants with more than 15%
249 missing data, an observed heterozygosity greater than or equal to 80%, or with cluster separation less
250 than or equal to 2% were removed from the dataset. In preparation for merging with additional
251 datasets, we converted all variants to the Illumina top strand and oriented them to match the 1000
252 Genomes reference. We renamed SNPs to match dbSNP version 144, and removed all indels and A/T or
253 C/G transversion variants, leaving over 1.3 million SNPs in the final dataset.

254

255 ***Unsupervised clustering and principal components analyses of genotype data:***

256 We merged our Ethiopian SNP data with previously published or publically available genotype data from
257 other Ethiopian, Somali, Sudanese, South Sudanese, as well as the 1000 Genomes Yoruba, Luhya,
258 Maasai, and Iberians (33–36). We removed SNPs that did not overlap across datasets, were out of
259 Hardy-Weinberg equilibrium ($p < 0.001$) in any population, or had a missingness rate of over 5% in the
260 merged dataset. We also removed SNPs with a MAF of 5% or less and SNPs in linkage disequilibrium ($r^2 >$
261 0.1 within a 50 kbp window, stepping 10 kbp at a time), leaving 59,700 SNPs and 1233 individuals for
262 analysis across 37 populations, plus Mota. Using this dataset, we identified and removed related
263 individuals within populations using kinship statistics calculated by PLINK ($PI_HAT > 0.1$). Of the
264 remaining individuals, we randomly discarded a set of individuals such that no population, as defined by
265 our population labels, had more than 50 individuals. We then applied a less strict linkage disequilibrium
266 filter ($r^2 > 0.3$ within a 50 kbp window, stepping 10 kbp at a time) to the MAF filtered data, leaving a
267 dataset of 106,997 SNPs from 828 individuals across 37 populations plus Mota. We ran the ADMIXTURE
268 algorithm for K s of 2 to 12 with 10 replicates each (37). We used *pong* to visualize the concordance
269 between different runs and, for each K , identify the most frequent mode among the 10 replicates (38).
270 For extant populations with known sampling or ethnographic coordinates (see ‘Estimated effective
271 migration surfaces’), we also plotted the population averages of each ancestry component
272 geographically, interpolating between datapoints across the landscape as in Uren et al. 2016 (39). We
273 also used the merged and LD filtered genotype dataset to perform a principal components analysis
274 (PCA) using *smartpca* (40).

275

276 ***Estimated effective migration surfaces:***

277 We used a method of estimating effective migration surfaces (EEMS) to visualize variation in migration
278 rates across East Africa (22). The algorithm takes geo-referenced genetic SNP data as input and
279 simulates migration across a grid under a stepping-stone model, returning a spatial depiction of
280 estimated historical rates of gene flow. We prepared the SNP dataset by following the same procedure
281 as for ADMIXTURE and PCA, but excluded some samples before relatedness and LD filtering. The
282 samples we removed were Mota, the ancient sample, the Somali and Sudanese populations from Pagani
283 et al. 2012, due to lack of specific geographical information, and the Iberians and Yoruba because they
284 live outside of the region of interest. The Baria were excluded in this filtering process for having
285 excessive missing data. This left a dataset of 131,892 SNPs and 571 individuals across 32 populations.
286 We primarily used the coordinates in the original publication with some adjustments. As this is a spatial

287 analysis based on historical population locations, the coordinates for Tigray individuals were changed
288 from their sampling location near Addis Ababa to their traditional homeland in Eritrea, using the
289 Glottolog coordinates for Tigrinya (41). For similar reasons, we excluded the 3 Hausa individuals given
290 their recent migration from outside the region of interest within the last 100 years (42). We found that
291 neither of these changes led to major qualitative differences in results. We performed a number of runs
292 under a range of starting parameters (number of demes specified as 200, 300, 400, and 500, each under
293 three different starting seed values) and averaged the results to mitigate the possible bias of any single
294 run. Each run was allowed to proceed for 30 million MCMC iterations to ensure convergence, with the
295 first 15 million discarded as burn-in and the remaining 15 million thinned to retain 1 out of every 15,000
296 data points. Proposal variances were tuned so that proposals were accepted between 20% and 30% of
297 the time for all runs.

298

299 ***Estimating the distributions of major language families:***

300 In order to determine the correspondence between EEMS-inferred migration barriers and corridors and
301 linguistic boundaries, we calculated kernel estimates of language family distributions using the adaptive
302 radius local convex hull (α -LoCoH) method (43). Language centroid point data and (Greenberg-based)
303 family classifications for every known living African language were obtained from Ethnologue
304 (<https://www.ethnologue.com/>). We then applied the α -LoCoH algorithm to construct ‘utilization
305 distributions’ for each of the five major African language families (Niger-Congo, Afro-Asiatic, Nilo-
306 Saharan, Khoisan, and Austronesian), using values of α equal to the longest geodesic distance between
307 any two languages in a family, to accommodate variable point densities. This produced a set of layered
308 isopleths for each language family representing decile occurrence probabilities. These isopleths were
309 then plotted with overlaid language point data to visualize the extent and density of language
310 distributions by family in relation to historical migration rates estimates as determined by EEMS (Fig.
311 S2C). Putative linguistic isolates were determined according to Blench, 2017 (44).

312

313 ***Runs of homozygosity:***

314 We determined runs of homozygosity (RoH) in the autosomes of the Ethiopian populations and other
315 African hunter-gatherers, the Hadza, Sandawe, Biaka, and Mbuti. We used the GenomeStudio and zCall
316 processed dataset for the Chabu, Majangir, and Shekkacho, a merged dataset of Ethiopians assayed on
317 the Illumina Infinium Omni 1M and Omni 2.5M arrays (33, 34), a merged dataset of Southern African
318 hunter-gatherers and Western African hunter-gatherers assayed on the Illumina 550k and 660k arrays,

319 respectively (26, 36). We removed SNPs with more than 5% missingness or a less than 1% MAF from all
320 datasets. We also randomly thinned the Ethiopian datasets to approximately match the SNP density of
321 the merged Southern and Western African hunter-gatherer dataset. Lastly, we removed SNPs that were
322 not in Hardy-Weinberg equilibrium within each population ($p < 0.001$), which left approximately 470,000
323 SNPs per population for analysis. We then identified RoH in each individual using PLINK, defining a run
324 as having at least 50 SNPs and being at least 1 Mb in length, allowing for no more than two missing and
325 one heterozygous SNP per run. Despite varying these parameters, we found many instances of two RoH
326 within a single individual closely flanking a low SNP density region. We chose to join such segments post
327 hoc with a custom script by defining low density regions as 1Mb windows that fell in the lower 5% of
328 SNP count when compared to the entire genome. We also observed genome regions where unusually
329 high numbers of individuals in a population carried a RoH segment. We defined such outlier SNPs as
330 being more than three standard deviations above the mean depth of RoH in the population. We joined
331 nearby outlier SNPs into larger regions and added them to a list of previously identified low density
332 regions, known low complexity regions (i.e. heterochromatin, telomere, centromere, and short arm
333 regions). We removed all RoH segments that overlapped by 85% or more with one of these regions.

334

335 **GARLIC:**

336 In order to analyze RoH in separate classes corresponding to the age of the events that produced them,
337 we also used GARLIC software to identify RoH in each population (45). This algorithm implements a
338 population model-based method of inferring RoH in 'short', 'intermediate' and 'long' size classes (46).
339 We ran GARLIC on the missingness- and MAF-filtered datasets used for PLINK RoH analysis, after
340 removing SNPs that were out of Hardy-Weinberg equilibrium within the population ($p < 0.001$), using
341 the following parameters: 'error' of 0.001, 'winsize' of 30, 'auto-winsize', and 'auto-winsize-step' of 5. As
342 with PLINK RoH, we then joined segments that flanked regions of low SNP density, and updated their
343 size class accordingly.

344

345 **IBDNe:**

346 Working from the missingness- and MAF-filtered datasets used for PLINK RoH analysis, we phased all
347 individuals in a given dataset together using SHAPEIT2 with a window size of 5 Mb and the duoHMM
348 option and haplotypes from the 1000 Genomes phase 3 dataset as a reference. We converted the
349 output of SHAPEIT2 to a 'phased' PLINK file format using a custom script. We then used GERMLINE2 to
350 identify tracts shared identical-by-descent (IBD) between chromosomes across all individuals in a given

351 dataset using the --w_extend and --haploid flags (47). We then joined IBD segments that, perhaps due to
352 errors in phasing or genotyping, were separated by a gap of less than 0.6 centiMorgans that contained
353 no more than one discordant SNP (28). As described in the 'Runs of homozygosity' section, we also
354 filtered out segments that had high overlap with regions of excess depth, low complexity, or low SNP
355 density. In inferring IBD tracts, we varied the 'bits' (from 5 to 200, increasing by 5) and 'errhom'
356 (between 0, 1 and 2) parameters and decided the optimal combination for each population using three
357 metrics. These parameters control the minimum number of exactly matching SNPs required to call an
358 IBD segment and the number of mismatches allowed, respectively (47). First, we compared the
359 distribution of RoH inferred by GERMLINE2 to that inferred by PLINK, and calculated how much of the
360 RoH inferred by GERMLINE2 was not inferred by PLINK, normalizing by the total amount of RoH inferred
361 by PLINK. Second, we calculated how much of the RoH inferred by PLINK was not inferred by
362 GERMLINE2, normalizing by the total amount of RoH inferred by PLINK. Finally, we compared the total
363 amount of the genome inferred to be IBD between pairs independently by PLINK and GERMLINE2, and
364 calculated the residuals. We selected the GERMLINE2 parameter combination that produced the IBD
365 segment distribution that most closely represented the PLINK results based on the average of these
366 three metrics (Table S1). We then used IBDNe to estimate the historical effective population size from
367 all IBD segments shared within a population that were 4 centiMorgans or longer (28).

368

369 ***Political boundaries in maps:***

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

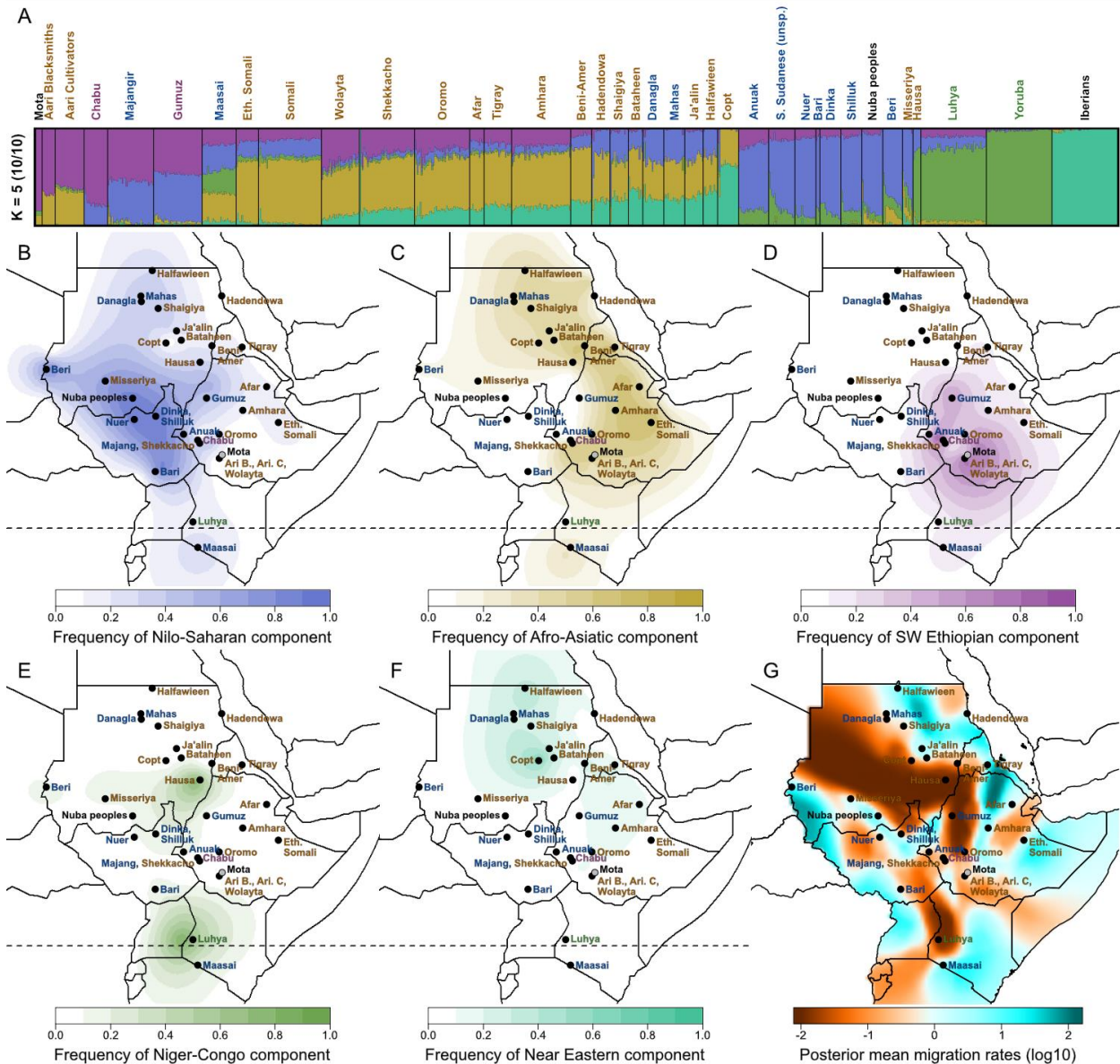
372

373 **References**

- 374 1. R. Dennell, *European Economic Prehistory: A New Approach* (Academic Press, 1983).
- 375 2. A. J. Ammerman, L. L. Cavalli-Sforza, *The Neolithic Transition and the Genetics of Populations in*
376 *Europe* (Princeton University Press, Princeton, 1984).
- 377 3. M. Zvelebil, The agricultural transition and the origins of Neolithic society in Europe. *Doc.*
378 *Praehist.* 28, 1-26 (2001).
- 379 4. A. Crowther, M. E. Prendergast, D. Q. Fuller, N. Boivin, Subsistence mosaics, forager-farmer
380 interactions, and the transition to food production in eastern Africa. *Quat. Int.* 489, 101–120 (2018).
- 381 5. J. Fort, Demic and cultural diffusion propagated the Neolithic transition across different regions
382 of Europe. *J. R. Soc. Interface.* 12 (2015).
- 383 6. M. Sikora et al., Population genomic analysis of ancient and modern genomes yields new
384 insights into the genetic ancestry of the Tyrolean Iceman and the genetic structure of Europe. *PLoS*
385 *Genet.* 10, e1004353 (2014).
- 386 7. G. González-Fortes et al., Paleogenomic evidence for multi-generational mixing between
387 Neolithic farmers and Mesolithic hunter-gatherers in the Lower Danube Basin. *Curr. Biol.* 27, 1801–1810
388 (2017).
- 389 8. M. Lipson et al., Parallel palaeogenomic transects reveal complex genetic history of early
390 European farmers. *Nature.* 551, 368–372 (2017).
- 391 9. M. Sikora, H. Laayouni, F. Calafell, D. Comas, J. Bertranpetit, A genomic analysis identifies a
392 novel component in the genetic structure of sub-Saharan African populations. *Eur. J. Hum. Genet.* 19,
393 84–88 (2011).
- 394 10. E. Patin et al., Dispersals and genetic adaptation of Bantu-speaking populations in Africa and
395 North America. *Science.* 356, 543–546 (2017).
- 396 11. E. A. Hildebrand, S. A. Brandt, J. Lesur-Gebremariam, The Holocene archaeology of Southwest
397 Ethiopia: new insights from the Kafa Archaeological Project. *African Archaeol. Rev.* 27, 255–289 (2010).
- 398 12. S. J. Dira, B. S. Hewlett, The Chabu hunter-gatherers of the highland forests of Southwestern
399 Ethiopia. *Hunt. Gatherer Res.* 3, 323-352 (2017).
- 400 13. A. Fentaw, “A history of the Shekacho (1898-1974)”, thesis, Addis Ababa University (2007).
- 401 14. J. Stauder, *The Majangir: Ecology and Society of a Southwest Ethiopian People* (Cambridge
402 University Press, 1971).
- 403 15. T. Schnoebelen, paper presented at the Language Documentation & Linguistic Theory 2, London,
404 UK, 13-14 November 2009.
- 405 16. A. González-Ruibal, C. Marín Suárez, M. Sánchez-Elípe, J. Lesur, C. Martínez Barrio, Late hunters
406 of Western Ethiopia: the sites of Ajilak (Gambela), ca. 1000-1200 AD. *Azania Archaeol. Res. Africa.* 49,
407 64–101 (2014).

- 408 17. H. Oota et al., Recent origin and cultural reversion of a hunter–gatherer group. *PLoS Biol.* 3, 537-
409 542 (2005).
- 410 18. P. Bellwood, A hypothesis for Austronesian origins. *Asian Perspect.* 26, 107–117 (1984).
- 411 19. S. A. Tishkoff et al., The genetic structure and history of Africans and African Americans. *Science.*
412 324, 1035–1044 (2009).
- 413 20. H. K. Schneider, *Livestock and Equality in East Africa: The Economic Basis for Social Structure*
414 (Indiana University Press, 1980).
- 415 21. M. G. Llorente et al., Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern
416 Africa. *Science.* 350, 820–822 (2015).
- 417 22. D. Petkova, J. Novembre, M. Stephens, Visualizing spatial population structure with estimated
418 effective migration surfaces. *Nat. Genet.* 48, 94–100 (2016).
- 419 23. N. Creanza et al., A comparison of worldwide phonemic and genetic variation in human
420 populations. *Proc. Natl. Acad. Sci.* 112, 1265–1272 (2015).
- 421 24. L. van Dorp et al., Evidence for a common origin of Blacksmiths and Cultivators in the Ethiopian
422 Ari within the last 4500 years: lessons for clustering-based inference. *PLoS Genet.* 11, e1005397 (2015).
- 423 25. E. E. Alemayehu, Mapping the socio-cultural landscape of the Gumuz Community of Metekel,
424 Northwestern Ethiopia. *African J. Hist. Cult.* 7, 209–218 (2015).
- 425 26. B. M. Henn et al., Hunter-gatherer genomic diversity suggests a southern African origin for
426 modern humans. *Proc. Natl. Acad. Sci.* 108, 5154–5162 (2011).
- 427 27. J. L. Newman, *The Ecological Basis for Subsistence Change among the Sandawe of Tanzania*
428 (National Academy of Sciences, 1970).
- 429 28. S. R. Browning, B. L. Browning, Accurate non-parametric estimation of recent effective
430 population size from segments of identity by descent. *Am. J. Hum. Genet.* 97, 404–418 (2015).
- 431 29. C. R. Gignoux, B. M. Henn, J. L. Mountain, Rapid, global demographic expansions after the
432 origins of agriculture. *Proc. Natl. Acad. Sci.* 108, 6044–6049 (2011).
- 433 30. D. W. Phillipson, *African Archaeology* (Cambridge University Press, 2005).
- 434 31. N. Blurton Jones, *Demography and Evolutionary Ecology of Hadza Hunter-Gatherers* (Cambridge
435 University Press, 2016).
- 436 32. J. I. Goldstein et al., zCall: a rare variant caller for array-based genotyping. *Bioinformatics.* 28,
437 2543–2545 (2012).
- 438 33. L. Pagani et al., Ethiopian genetic diversity reveals linguistic stratification and complex influences
439 on the Ethiopian gene pool. *Am. J. Hum. Genet.* 91, 83–96 (2012).
- 440 34. L. Pagani et al., Tracing the route of modern humans out of Africa by using 225 human genome
441 sequences from Ethiopians and Egyptians. *Am. J. Hum. Genet.* 96, 1–6 (2015).

- 442 35. N. Hollfelder et al., Northeast African genomic variation shaped by the continuity of indigenous
443 groups and Eurasian migrations. *PLoS Genet.* 13, e1006976 (2017).
- 444 36. A. Auton et al., A global reference for human genetic variation. *Nature.* 526, 68–74 (2015).
- 445 37. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated
446 individuals. *Genome Res.* 19, 1655–1664 (2009).
- 447 38. A. A. Behr, K. Z. Liu, G. Liu-Fang, P. Nakka, S. Ramachandran, pong: fast analysis and visualization
448 of latent clusters in population genetic data. *Bioinformatics.* 32, 2817–2823 (2016).
- 449 39. C. Uren et al., Fine-scale human population structure in southern Africa reflects ecological
450 boundaries. *Genetics.* 204, 303–314 (2016).
- 451 40. N. Patterson, A. L. Price, D. Reich, Population structure and eigenanalysis. *PLoS Genet.* 2, 2074–
452 2093 (2006).
- 453 41. H. Hammarström, R. Forkel, M. Haspelmath, *Glottolog 3.3* (2018; <https://glottolog.org/>).
- 454 42. S. Benson, M. Duffield, Women’s work and economic change: the Hausa in Sudan and in Nigeria.
455 *IDS Bull.* 10, 13–19 (1979).
- 456 43. W. M. Getz et al., LoCoH : Nonparametric kernel methods for constructing home ranges and
457 utilization distributions. *PLoS One.* 2, e207 (2007).
- 458 44. R. Blench, “African language isolates” in *Language Isolates*, L. Campbell, Ed. (2017), pp. 176–
459 206.
- 460 45. Z. A. Szpiech, A. Blant, T. J. Pemberton, GARLIC: genomic autozygosity regions likelihood-based
461 inference and classification. *Bioinformatics.* 33, 2059–2062 (2017).
- 462 46. T. J. Pemberton et al., Genomic patterns of homozygosity in worldwide human populations. *Am.*
463 *J. Hum. Genet.* 91, 275–292 (2012).
- 464 47. A. Gusev et al., Whole population, genome-wide mapping of hidden relatedness. *Genome Res.*
465 19, 318–326 (2009).



466

467

468

469

470

471

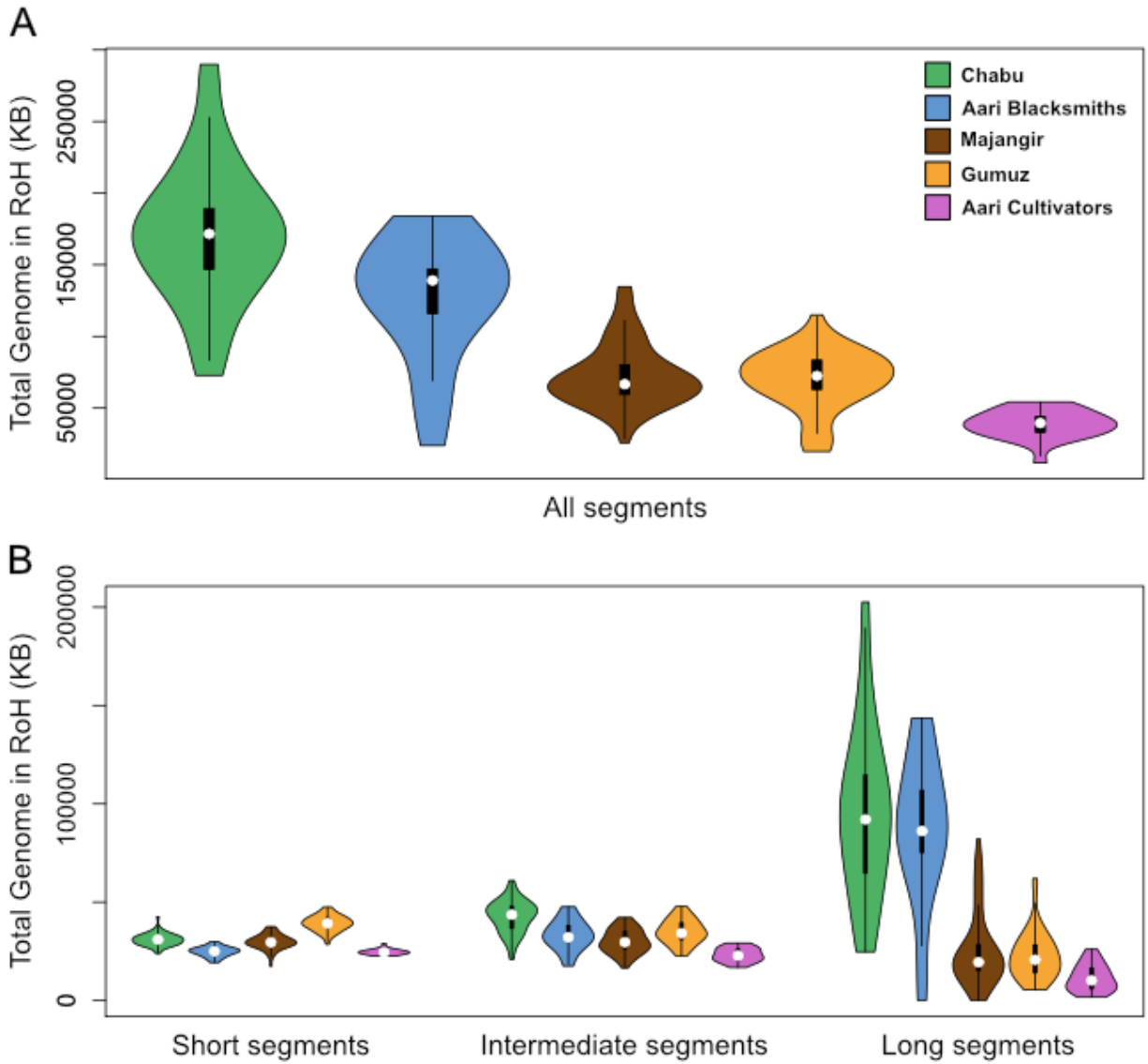
472

473

474

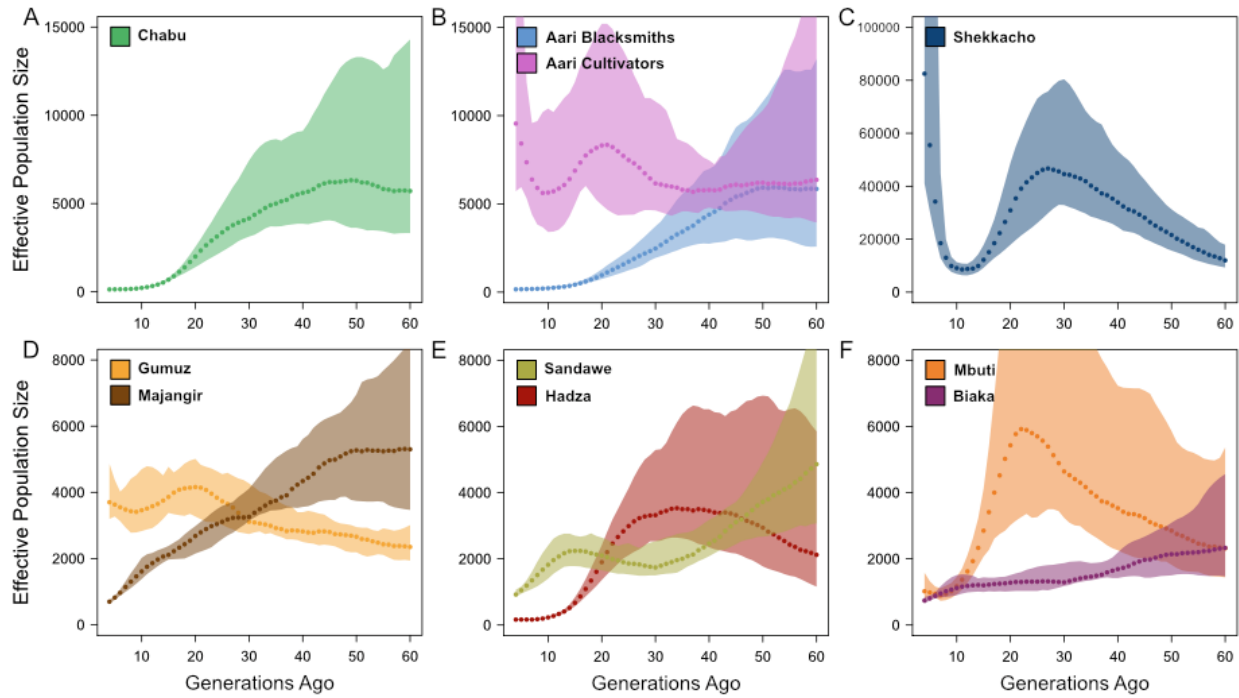
Figure 1. Global ancestry proportions of northeast African individuals, the Yoruba from Nigeria and Iberians from Spain, inferred using unsupervised clustering. A) Each color corresponds to a genetic component and each vertical bar represents one individual (Mota is plotted 5 times wider for visualization purposes). The population labels are colored according to linguistic affiliation, with green denoting Niger-Congo speakers, blue Nilo-Saharan speakers, yellow Afro-Asiatic speakers, and purple speakers of linguistic isolates. All ten replicates at K=5 ancestral components converged on the same overall pattern of partitioning the genetic variation. B-F) The geographic distributions of these components are depicted, with the intensity of the color corresponding to the mean population proportion of the respective ancestry.

475 G) Effective migration surfaces, inferred using the rate of decay of genetic similarity across space, are
476 depicted. Cool colors correspond to effective migration corridors, while warm colors correspond to
477 effective migration barriers.
478



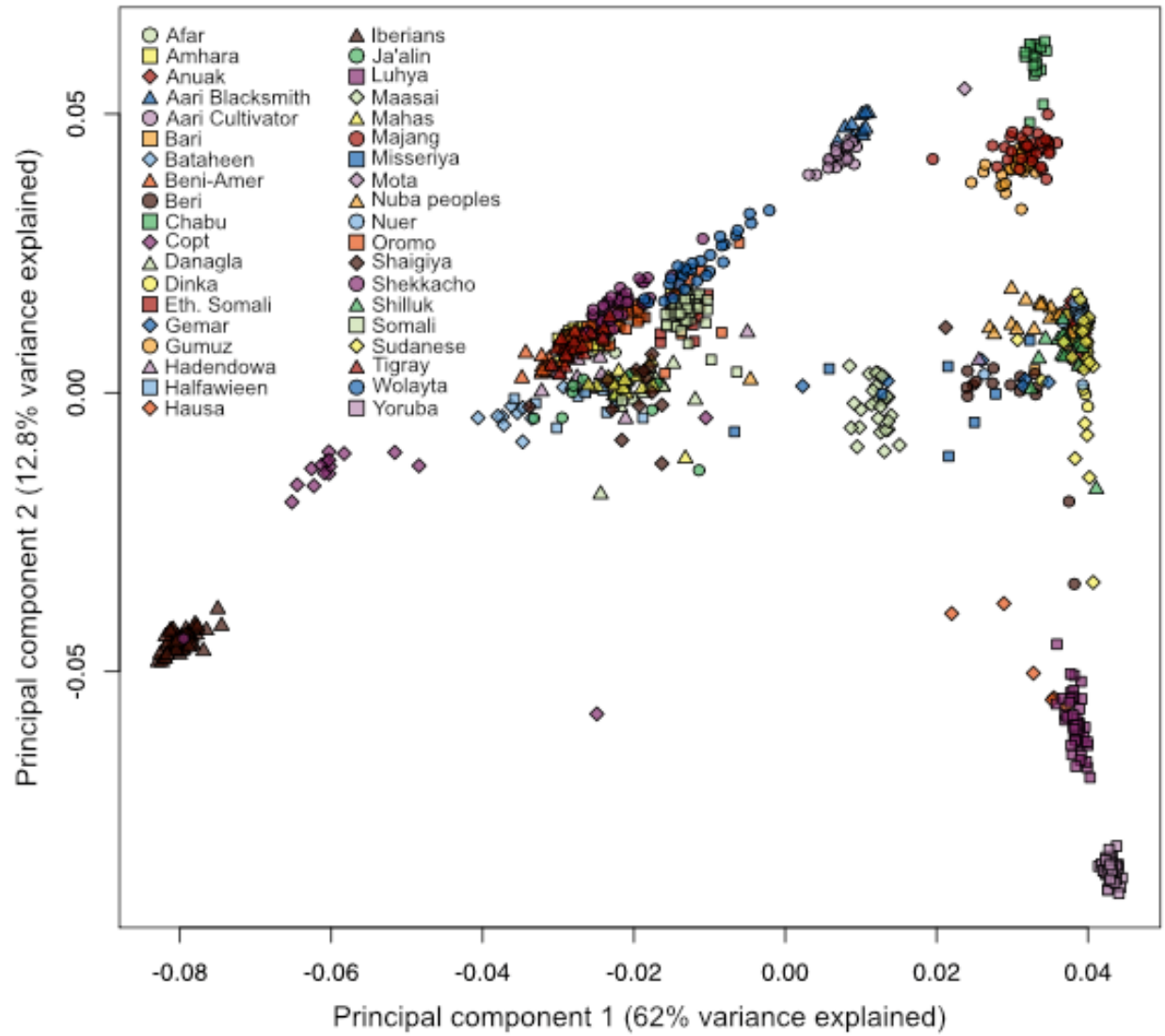
479
480
481
482
483
484
485

Figure 2. The distributions of the total amount of the genome in runs of homozygosity (RoH) in the Chabu, Aari Blacksmiths, Majangir, Gumuz, and Aari Cultivators for A) all RoH segments and B) in each size class.



486
487
488
489
490
491
492

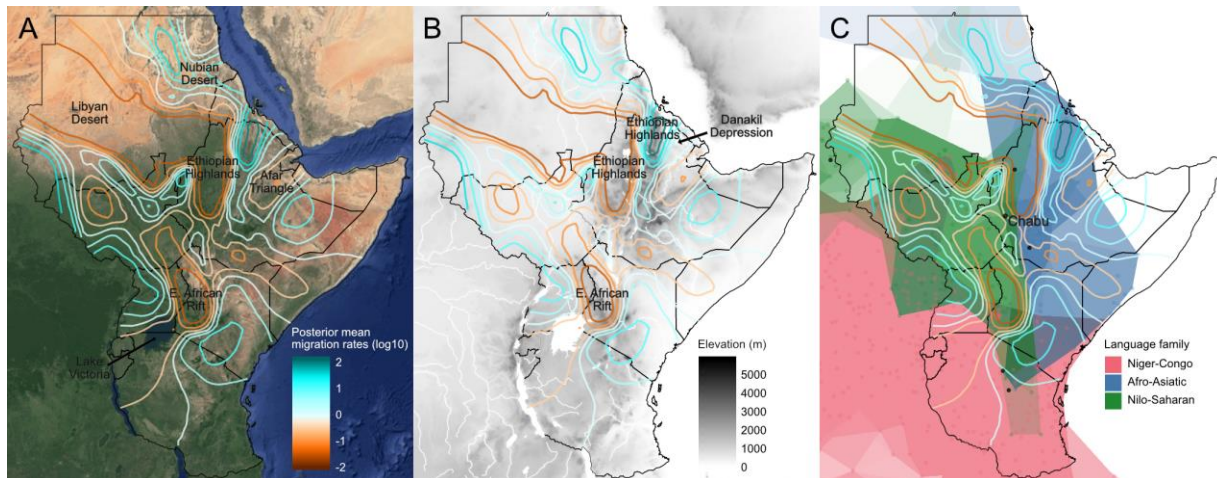
Figure 3. Historical effective population sizes (N_e), from 4 to 60 generations ago, as inferred from distributions of segments that are identical-by-descent across pairs of individuals within A) the Chabu, B), the Aari populations, C) the Shekkacho, D) the Majangir and Gumuz, E) the Sandawe and Hadza, and F) the Mbuti and Biaka. Colored ribbons indicate bootstrapped confidence intervals.



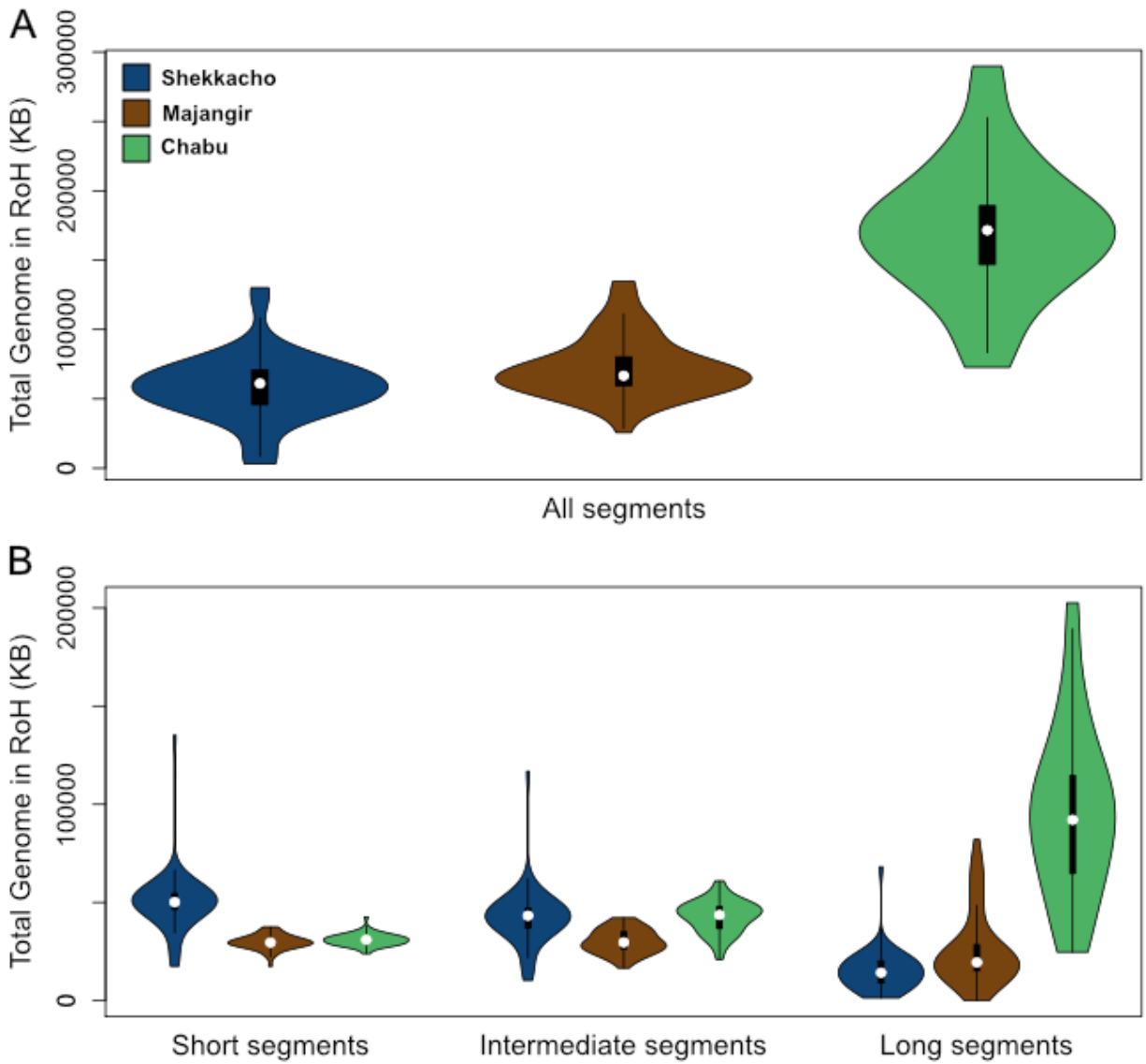
493
494

495 **Figure S1.** Principal components analysis on genotype data from populations of northeast Africa,
496 including Mota, an ancient individual, as well as the Yoruba from Nigeria and Iberians from Spain.

497



498
499 **Figure S2.** Effective migration surfaces depicted as contour lines over A) satellite imagery, B) elevation
500 and water features, and C) the geographic distribution of major language families in Eastern Africa (see
501 Methods). For contour lines, cool colors correspond to effective migration corridors, while warm colors
502 correspond to effective migration barriers. A) Some geographic features, such as the Libyan Desert
503 (Sahara) in northwestern Sudan, the northwestern Ethiopian Highlands, the Afar Triangle in the Ethiopian
504 Rift down through the East African Rift, and Nalubaale/Lolwe/Nyanza (Lake Victoria), correspond with
505 historical migration barriers. Map data from Google; imagery © 2018 TerraMetrics. B) Some regions of
506 high elevation or roughness, such as the northwestern Ethiopian Highlands and the volcanic range along
507 the East African Rift, correspond with historical migration barriers. However, the northeastern Ethiopian
508 Highlands feature high estimated rates of historical migration, potentially as a preferred route around the
509 inhospitable Danakil Depression to the east. Elevation data at 30 arc-second (~1 km) resolution are from
510 the U.S. Geological Survey GTOPO30 digital elevation model, accessed through EarthExplorer
511 (<https://earthexplorer.usgs.gov/>). Physical vectors for rivers and lakes at 1:10 million scale are from
512 Natural Earth (<https://www.naturalearthdata.com/>). C) Darker isopleths represent higher occurrence
513 probabilities for languages within each family. Putative linguistic isolates, including the Chabu, are
514 depicted as black points. Language data are from Ethnologue (<https://www.ethnologue.com/>).
515



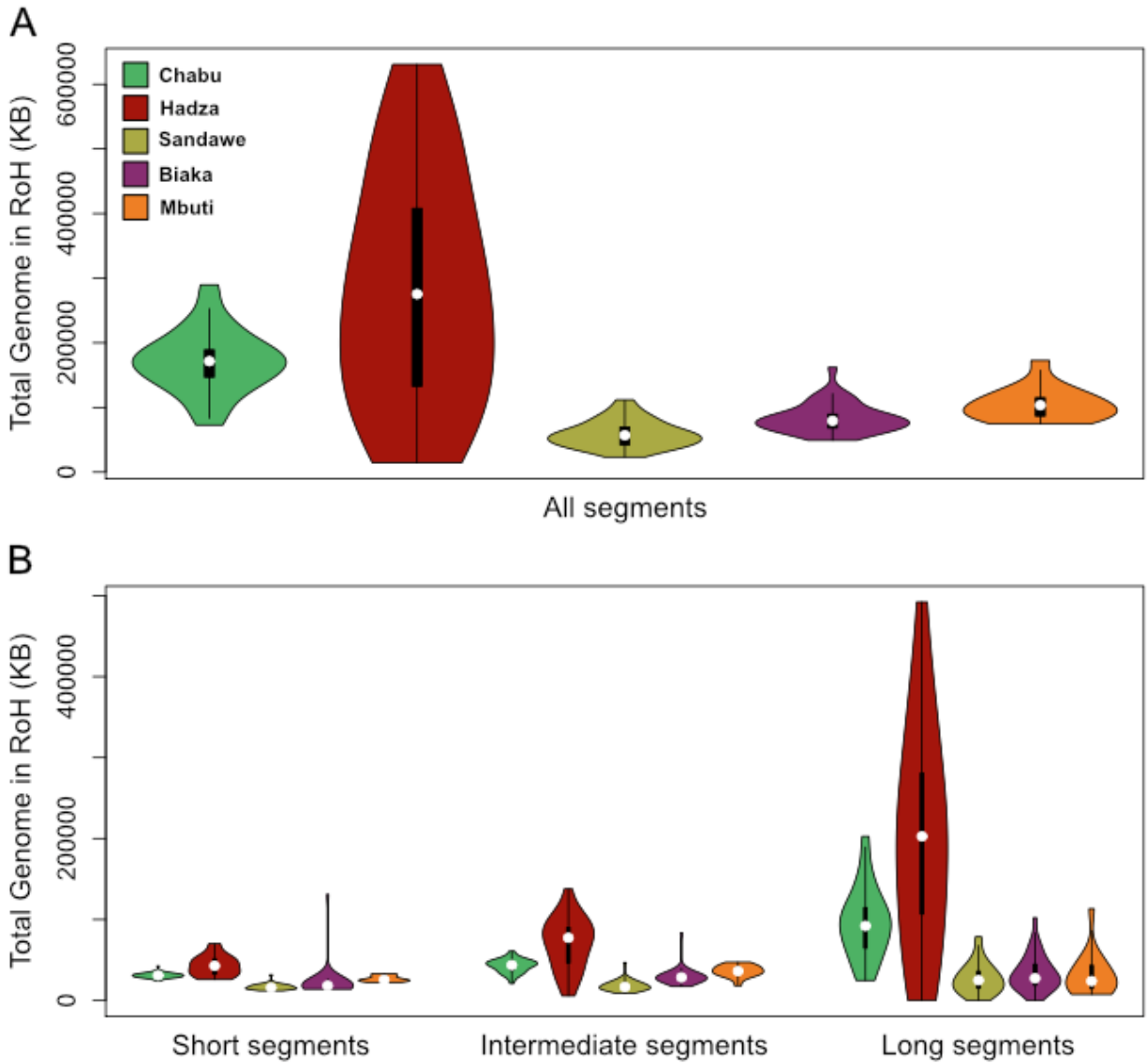
516

517

518 **Figure S3.** The distributions of the total amount of the genome in runs of homozygosity (RoH) in the

519 Shekkacho, Majangir, and Chabu for A) all RoH segments and B) in each size class.

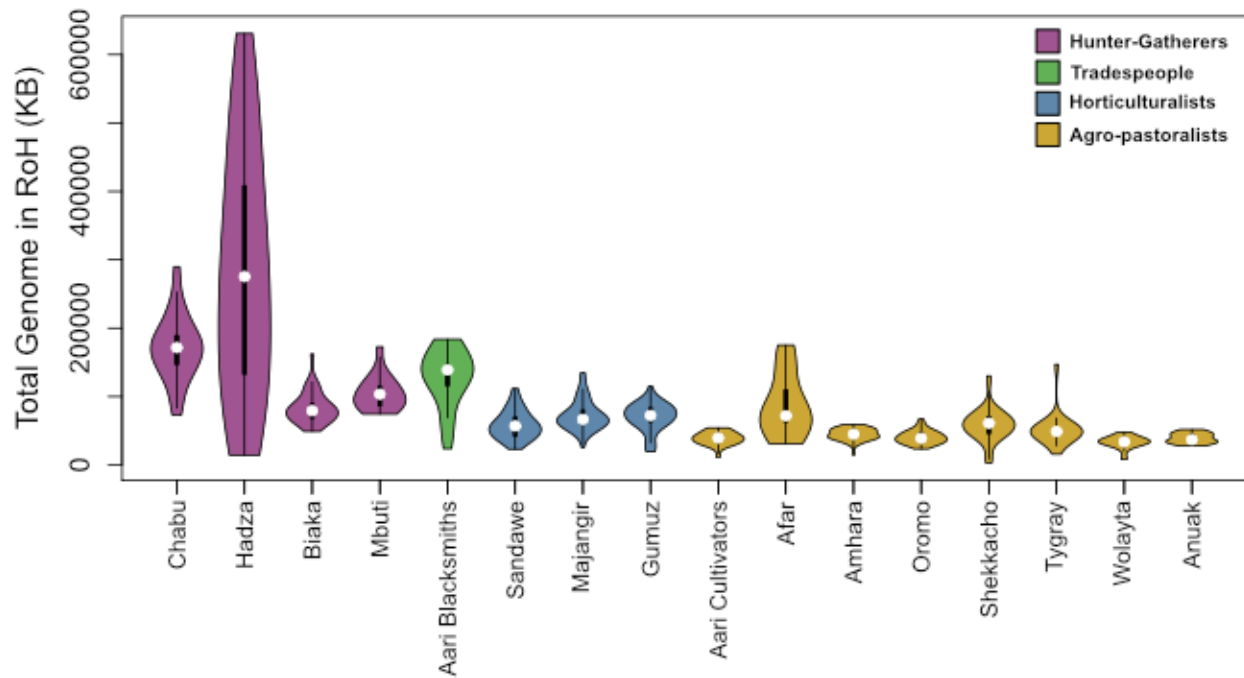
520



521
522

523 **Figure S4.** The distributions of the total amount of the genome in runs of homozygosity (RoH) in the
524 Chabu, Hadza, Sandawe, Biaka, and Mbuti for A) all RoH segments and B) in each size class.

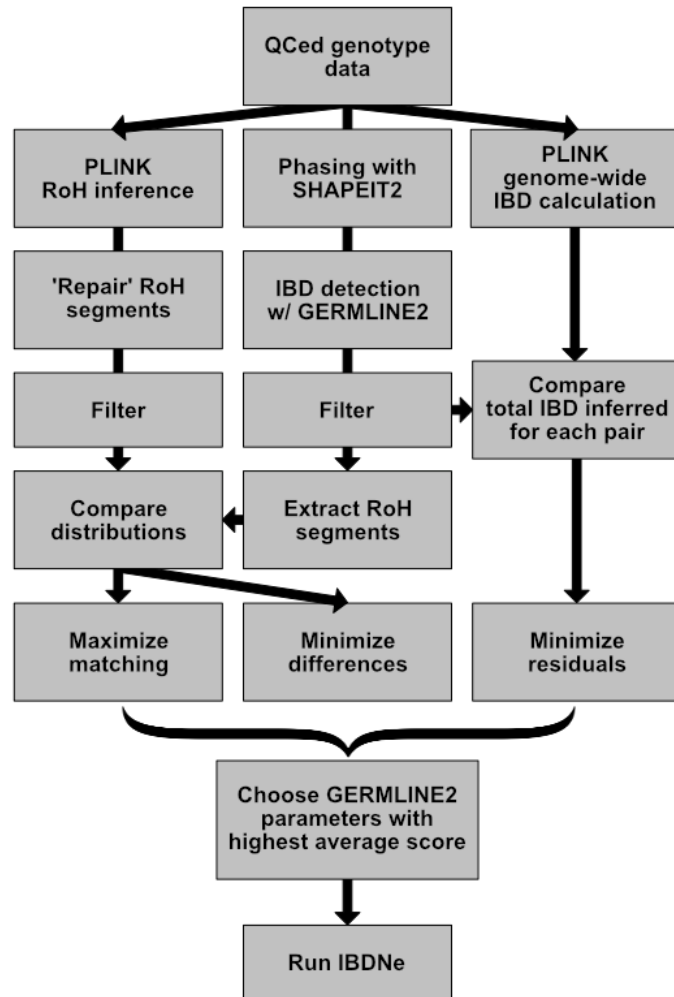
525



526
527

528 **Figure S5.** The distributions of the total amount of the genome in runs of homozygosity (RoH) in all
529 Ethiopian populations and African hunter-gatherers. Populations are grouped and colored according to
530 their primary subsistence or economic strategy.

531



532

533

534 **Figure S6.** A schematic of the strategy used to select the optimal parameters for GERMLINE2 to infer

535 the IBD segment distribution for each population.

536

537 **Table S1.**

538 The GERMLINE2 parameters chosen for each population by our optimization strategy for the inference of
539 historical effective population size (N_e) (see Materials and Methods).

Population	'Bits' parameter	'Errhom' parameter
Chabu	65	2
Aari Blacksmiths	30	1
Aari Cultivators	40	2
Shekkacho	100	2
Gumuz	100	2
Majangir	55	1
Sandawe	155	1
Hadza	185	1
Mbuti	20	2
Biaka	30	2

540