Reconstructing Prehistoric African Population Structure

Pontus Skoglund[1*], Jessica C. Thompson[2], Mary E. Prendergast[3], Alissa Mittnik[4,5,+], Kendra Sirak[2,6,+], Mateja Hajdinjak[7,+], Tasneem Salie[8,+], Nadin Rohland[1], Swapan Mallick[1,9], Alexander Peltzer[4,10], Anja Heinze[7], Iñigo Olalde[1], Matthew Ferry[1,11], Eadaoin Harney[1,11,12], Megan Michel[1,11], Kristin Stewardson[1,11], Jessica Cerezo-Roman[13], Chrissy Chiumia[14], Alison Crowther[15], Elizabeth Gomani-Chindebvu[14], Agness O. Gidna[16], Katherine M. Grillo[17], I. Taneli Helenius[18], Garrett Hellenthal[18], Richard Helm[19], Mark Horton[20], Saioa López[18], Audax Z.P. Mabulla[16], John Parkington[21], Ceri Shipton[22,23], Mark G. Thomas[18], Ruth Tibesasa[24], Menno Welling[25,26], Vanessa M. Hayes[27,28,29], Douglas J. Kennett[30], Raj Ramesar[8], Matthias Meyer[7], Svante Pääbo[7], Nick Patterson[2,8], Alan G. Morris[21], Nicole Boivin[4], Ron Pinhasi[6,31,@], Johannes Krause[4,5,@] and David Reich[1,8,11,*,@$]

+ These authors contributed equally

@ Senior authors

[1]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA
[2]Department of Anthropology, Emory University, Atlanta, Georgia 30322, USA
[3]Radcliffe Institute for Advanced Study, Harvard University, Cambridge, Massachusetts 02138, USA
[4]Max Planck Institute for the Science of Human History, Jena 07745, Germany
[5]Institute for Archeological Sciences, Eberhard-Karls-University, Tuebingen 72070 Germany
[6]School of Archaeology and Earth Institute, University College Dublin, Dublin 4, Ireland
[7]Max Planck Institute for Evolutionary Anthropology, Leipzig 04103, Germany
[8]Division of Human Genetics, Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town 7925, South Africa
[9]Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA
[10]Integrative Transcriptomics, Centre for Bioinformatics, University of Tuebingen, Tuebingen 72076, Germany
[11]Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115, USA
[12]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA
[13]Department of Geography and Anthropology, California State Polytechnic University, Pomona, CA 91768, USA
[14]Malawi Department of Museums and Monuments, Lilongwe 3, Malawi
[15]School of Social Science, The University of Queensland, Brisbane, Queensland 4072, Australia
[16]National Museums of Tanzania, Dar es Salaam, Tanzania
[17]Department of Archaeology and Anthropology, University of Wisconsin - La Crosse, La Crosse, Wisconsin 54601, USA
[18]Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, UK
[19]Canterbury Archaeological Trust, Canterbury CT1 2LU, UK
[20]Department Archaeology and Anthropology, University of Bristol, Bristol BS8 1UU, UK
[21]Department of Archaeology, University of Cape Town, 7700 Cape Town, South Africa
[22]McDonald Institute for Archaeological Research, Cambridge CB2 3ER, UK
[23]British Institute in Eastern Africa, Nairobi, Kenya
[24]University of Pretoria, Department of Anthropology and Archaeology, Pretoria, South Africa
[25]African Studies Centre Leiden, Leiden University, Leiden 2300 RB, Netherlands
[26]African Heritage Ltd, Zomba, Malawi
[27]Genomics and Epigenetics Division, Garvan Institute of Medical Research, Darlinghurst, NSW 2010, Australia
[28]Central Clinical School, University of Sydney, Camperdown, NSW 2050, Australia
[29]School of Health Systems and Public Health, University of Pretoria, Gezina 0031, South Africa
[30]Department of Anthropology and Institutes for Energy and the Environment, Pennsylvania State University, University Park, PA 16802, USA
[31]Department of Anthropology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria

$ Lead Contact

* Correspondence to: skoglund@genetics.med.harvard.edu & reich@genetics.med.harvard.edu

Summary

We assembled genome-wide data from 16 prehistoric Africans. We show that the anciently divergent lineage that comprises the primary ancestry of the southern African San had a wider distribution in the past, contributing ~2/3 of the ancestry of Malawi hunter-gatherers ~8100-2500 years ago, and ~1/3 of Tanzanian hunter-gatherers ~1400 years ago. We document how the spread of farmers from western Africa involved complete replacement of local hunter-gatherers in some regions, and we track the spread of herders by showing that the population of a ~3100 year-old pastoralist from Tanzania contributed ancestry to people from northeast to southern Africa, including a ~1200-year-old southern African pastoralist. The deepest diversifications of African lineages were complex, involving long-distance gene flow, or a lineage more deeply diverging than that of the San contributing more to some western Africans than others. We finally leverage ancient genomes to document episodes of natural selection in southern African populations.

Introduction

Africa harbors more genetic diversity than any other part of the world (Cann et al., 1987; Tishkoff et al., 2009). This is reflected both in a higher average number of differences between sub-Saharan African genomes than between non-African genomes (Cann et al., 1987; Ramachandran et al., 2005), and in the fact that the ancestry found outside of Africa is largely a subset of that within it (Tishkoff et al., 2009). Today, some of the earliest branching African lineages are present only in populations with relatively small census sizes, including the southern African Khoe-San (see **STAR Methods** for terminology), central African rainforest hunter-gatherers, and Hadza of Tanzania (Gronau et al., 2011; Schlebusch et al., 2012; Veeramah et al., 2012). However, the population structure of Africa prior to the expansion of food producers (pastoralists and agriculturalists) remains unknown (Busby et al., 2016; Gurdasani et al., 2015; Patin et al., 2017). Bantu-speaking agriculturalists originating in western Africa are thought to have brought farming to eastern Africa by ~2000 years calBP (calibrated radiocarbon years Before Present, defined by convention as years before 1950 CE) and to southern Africa by ~1500 BP, thereby spreading the largest single ancestry component to African genomes today (Russell et al., 2014; Tishkoff et al., 2009). Earlier migration(s), which brought ancestry related to the ancient Near East (Lazaridis et al., 2016; Pagani et al., 2012; Pickrell et al., 2014), brought herding to eastern Africa by ~4000 BP (Marshall et al., 1984), and to southern Africa by ~2000 BP (Sadr, 2015).

Results

To reconstruct African population structure prior to the spread of food production, we generated genome-wide data from 15 ancient sub-Saharan Africans (**Table 1; Table S1; Table S2; STAR**

91   **Methods**). For three individuals from the western Cape of South Africa (~2300-1300 BP) we
92   carried out direct shotgun sequencing to 0.7-2.0-fold coverage. For 12 individuals from eastern
93   and south-central Africa, we used in-solution enrichment of ~1.2 million single nucleotide
94   polymorphisms (SNPs). These included 4 individuals from the coastal region of Kenya and
95   Tanzania (~1400-400 BP), one from interior Tanzania (~3100 BP), and 7 from Malawi (ranging
96   over ~8100-2500 BP) (**Fig. S1**). All individuals had post-mortem degradation characteristic of
97   ancient DNA (**Table 1**), and we confirmed that key results are unlikely to be artifacts of
98   contamination by restricting to sequences with post-mortem degradation (Skoglund et al., 2012;
99   Skoglund et al., 2014a) (**Fig. S2**). We merged the new ancient DNA data with previously
100   reported shotgun sequence data from a ~4500 BP Ethiopian highland individual (Llorente et al.,
101   2015), and with SNP genotypes from 584 present-day African individuals from 59 diverse
102   populations (including new data from 34 Malawi individuals; **STAR Methods**) (Lazaridis et al.,
103   2014; Patterson et al., 2012), as well as 300 high-coverage genomes from 142 worldwide
104   populations (Mallick et al., 2016).
105
106   **An ancient cline of southern and eastern African hunter-gatherers.** We used principal
107   component analysis (PCA) (Patterson et al., 2006), and automated clustering (Alexander et al.,
108   2009) to relate the 16 ancient individuals to present-day sub-Saharan Africans (**Fig. 1**). Whereas
109   the two individuals buried in ~2000 BP hunter-gatherer contexts in South Africa share ancestry
110   with southern African Khoe-San populations in the PCA, 11 of the 12 ancient individuals who
111   lived in eastern and south-central Africa between ~8100-400 BP form a gradient of relatedness to
112   the eastern African Hadza on one hand, and southern African Khoe-San on the other (**Fig. 1A**).
113   The genetic cline correlates to geography, running along a north-south axis with ancient
114   individuals from Ethiopia (~4500 BP), Kenya (~400 BP), Tanzania (both ~1400 BP) and Malawi
115   showing increasing affinity to southern Africans (both ancient individuals and present-day Khoe-
116   San). The seven individuals from Malawi show no significant heterogeneity, indicating a
117   longstanding and distinctive population in ancient Malawi that persisted for at least ~5,000 years
118   (the minimum span of our radiocarbon dates), but no longer exists today.
119
120   We constructed a model where ancient and present-day African populations trace their ancestry
121   to a putative set of nine ancestral populations. As proxies for these populations we used three
122   different ancient Near Eastern populations and six African populations that, according to our
123   analyses, harbor substantial ancestry related to major lineages present in Africa today. The
124   Mende from Sierra Leone are used in this model to represent a component of ancestry that exists
125   in high proportions in western African populations, the ancient southern African genomes
126   (South_Africa_2000BP) are used to represent the ancestry of southern Africa before agriculture,
127   the Ethiopian individual (Ethiopia_4500BP) is used to represent northeastern African ancestry
128   before agriculture, the Mbuti are used to represent central African rainforest hunter-gatherer
129   ancestry,   the   individual   from   an   eastern   African   pastoralist   context
130   (Tanzania_Luxmanda_3100BP) is used to represent an early pastoralist lineage from eastern

131 Africa (see below), and the Dinka (from Sudan) are used to represent distinctive ancestry found
132 in Nilotic speakers today. The ancient Near Eastern populations were representative of Anatolia,
133 the Levant, and Iran, respectively (Lazaridis et al., 2016; Mathieson et al., 2015). We used
134 *qpAdm* (Haak et al., 2015), a generalization of f_4 symmetry statistics, to successively test 1-
135 source, 2-source, or 3-source models and admixture proportions for all other ancient and present-
136 day African populations, with a set of 10 non-African populations as outgroups (**STAR**
137 **Methods**).
138
139 We find that ancestry closely related to the ancient southern Africans was present much farther
140 north and east in the past than is apparent today. This ancient southern African ancestry
141 comprises up to 91% of the ancestry of Khoe-San groups today (**Table S3**), and also 31% ± 3%
142 of the ancestry of Tanzania_Zanzibar_1400BP, 60% ± 6% of the ancestry of
143 Malawi_Fingira_6100BP, and 65% ± 3% of the ancestry of Malawi_Fingira_2500BP (**Fig. 2A**).
144 Notably, the Khoe-San-related ancestry in ancient individuals from Malawi and Tanzania is
145 symmetrically related to the two previously identified lineages present in the San (Z<2; **Fig. S2**),
146 estimated to have diverged at least 20,000 years ago (Mallick et al., 2016; Pickrell et al., 2012;
147 Schlebusch et al., 2012), implying that this was an ancient divergent branch of this group that
148 lived in eastern Africa at least until 1400 BP. However, it was not present in all eastern Africans,
149 as we do not detect it in the ~400-year old individual from coastal Kenya, nor in the present-day
150 Hadza.
151
152 **Displacement of forager populations in eastern Africa.** Both unsupervised clustering (**Fig.**
153 **1B**) and formal ancestry estimation (**Fig. 2B**) suggests that individuals from the Hadza group in
154 Tanzania can be modeled as deriving all their ancestry from a lineage related deeply to ancient
155 eastern Africans such as the Ethiopia_4500BP individual (**Fig. 3A; Table S3**). However, this
156 lineage appears to have contributed little ancestry to present-day Bantu-speakers in eastern
157 Africa, who instead trace their ancestry to a lineage related to present-day western Africans, with
158 additional components related to the Nilotic-speaking Dinka and to the
159 Tanzania_Luxmanda_3100BP pastoralist (see below and **Fig. 2**). The Sandawe, another
160 population that like the Hadza uses click consonants in their spoken language, are modeled as
161 having ancestry similar to the Hadza but also admixture related to that of neighboring
162 populations (**Fig. 3A; Table S3**) consistent with previous findings (Henn et al., 2011; Tishkoff et
163 al., 2009). Population replacement by incoming food-producers appears to have been nearly
164 complete in Malawi, where we detect little if any ancestry from the ancient individuals who lived
165 ~8100-2500 BP. Instead, present-day Malawian individuals are consistent with deriving all their
166 ancestry from the Bantu expansion of ultimate western African origin (**Fig. 3**).
167
168 Among the ancient individuals analyzed here, only a ~600 BP individual from the Zanzibar
169 archipelago has a genetic profile similar to present-day Bantu-speakers (**Fig. 1**). Notably, this
170 individual is consistent with having even more western African-related ancestry than the present-
171 day Bantu-speakers we analyzed from Kenya, who also derive some of their ancestry from

3

172 lineages related to Dinka and Tanzania_Luxmanda_3100BP (**Fig. 1B**). Using linkage
173 disequilibrium, we estimate that this admixture between western and eastern African related
174 lineages occurred an average of 800-400 years ago (**STAR Methods**). This suggests a scenario
175 of genetic isolation between early farmers and previously established foragers during the initial
176 phase of the Bantu expansion into eastern Africa (Crowther et al., 2017; Ribot et al., 2010), a
177 barrier that broke down over time as mixture occurred. This parallels the patterns previously
178 observed in genomic analyses of the Neolithic expansion into Europe (Haak et al., 2015;
179 Skoglund et al., 2012), and the East Asian farming expansion into Remote Oceania (Skoglund et
180 al., 2016). However, this process of delayed admixture did not always apply in Africa, as is
181 evident in the absence of admixture from previously established hunter-gatherers in present-day
182 Malawians.

183

184 **Early Levantine farmer-related admixture in a ~3100-year-old pastoralist from Tanzania.**
185 Western Eurasian-related ancestry is pervasive in eastern Africa today (Pagani et al., 2012;
186 Tishkoff et al., 2009), and the timing of this admixture has been estimated to be ~3000 BP on
187 average (Pickrell et al., 2014). We found that the ~3100 BP individual
188 (Tanzania_Luxmanda_3100BP), associated with a Savanna Pastoral Neolithic archeological
189 tradition, could be modeled as having 38±1% of her ancestry related to the nearly 10,000 year
190 old pre-pottery farmers of the Levant (Lazaridis et al., 2016), and we can exclude source
191 populations related to early farmer populations in Iran and Anatolia. These results could be
192 explained by migration into Africa from descendants of pre-pottery Levantine farmers, or
193 alternatively by a scenario in which both pre-pottery Levantine farmers and
194 Tanzania_Luxmanda_3100BP descend from a common ancestral population that lived thousands
195 of years earlier in Africa or the Near East. We fit the remaining approximately 2/3 of
196 Tanzania_Luxmanda_3100BP as most closely related to the Ethiopia_4500BP (P=0.029) or,
197 allowing for 3-way mixture also from a source closely related to the Dinka (P=0.18; the
198 Levantine-related ancestry in this case was 39±1%) (**Table S3**).

199

200 While these findings show that a Levant Neolithic-related population made a critical contribution
201 to the ancestry of present-day eastern Africans (Lazaridis et al., 2016), present-day Cushitic-
202 speakers such as the Somali cannot be fit simply as being of Tanzania_Luxmanda_3100BP
203 ancestry. The best fitting model for the Somali includes Tanzania_Luxmanda_3100BP ancestry,
204 Dinka-related ancestry, and 16% ± 3% Iranian Neolithic-related ancestry (P=0.015). This
205 suggests that ancestry related to the Iranian Neolithic appeared in eastern Africa after earlier
206 gene flow related to Levant Neolithic populations, a scenario that is made more plausible by the
207 genetic evidence of admixture of Iranian Neolithic-related ancestry throughout the Levant by the
208 time of the Bronze Age (Lazaridis et al., 2016) and in ancient Egypt by the Iron Age
209 (Schuenemann et al., 2017).

210

Direct evidence of migration bringing pastoralism to eastern and southern Africa. In contrast to the Malawi and Zanzibar individuals, all three ancient southern Africans show affinities to the ancestry predominant in present-day Tuu speakers in the southern Kalahari more than to present-day Ju'|hoan speakers in the northern Kalahari (**Fig. S2B; Fig. S2C**). However, the ~1200 BP sample from the western Cape that is found in a pastoralist context has a specific similarity in clustering analyses to present-day Khoe-Khoe-speaking pastoralist populations such as the Nama (**Fig. 1B**), and like them has affinity to three groups: Khoe-San, western Eurasians and eastern Africans. This supports the hypothesis that a non-Bantu-related population carried eastern African and Levantine ancestry to southern Africa by at least around 1200 BP, providing direct evidence for claims previously made based on analysis of present-day populations (Pickrell et al., 2014).

We used our modeling framework to show that the South_Africa_1200BP pastoralist individual from the Western Cape is consistent with being a mixture of just two streams of ancestry relative to non-southern African populations, with 40.3% ± 2.3%, ancestry related to the Tanzania_Luxmanda_3100BP individual (54% ± 7% when restricting to sequences with postmortem damage), and the remainder being related to the South_Africa_2000BP hunter-gatherers (**Table S3**). This supports the hypothesis that the Savanna Pastoral Neolithic archaeological tradition in eastern Africa is a plausible source for the spread of herding to southern Africa. Even the Ju_hoan_North, the San individuals with the least genetic affinity to eastern Africans, have 9% ± 1% of their ancestry most closely related to Tanzania_Luxmanda_3100BP, consistent with previous findings that the ancestries of all present-day San and Khoe were affected by agro-pastoralist migrations in the last two millennia (Pickrell et al., 2014).

The earliest divergences among modern human populations. Previous studies have suggested that the primary ancestry in the San is from a lineage that separated from all other lineages represented in modern humans today, before the latter separated from each other (Gronau et al., 2011; Veeramah et al., 2012). Such a model emerges when we automatically fit a tree without admixture to the data (**Fig. 3A**), but we also find that a tree-like representation is a poor fit (**Fig. S4A**), in the sense that ancient southern Africans who lived ~2000 BP were not strictly an outgroup to extant lineages in other parts of sub-Saharan Africa. In particular, we find that ancient southern Africans, who have none of the eastern African admixture that is ubiquitous today, share significantly more alleles with present-day and ancient eastern Africans (including Dinka, Hadza and Ethiopia_4500BP), than they do with present-day western Africans (**Fig. 3B**). Even within present-day western Africans, the genetic differences between Yoruba from Nigeria and the Mende from Sierra Leone are inconsistent with descent from a homogeneous ancestral population isolated from ancient southern Africans. The asymmetry between Yoruba and Mende is also observed with non-Africans, but no stronger than in eastern Africans (the most closely related Africans to the ancestral out-of-Africa population), and thus these signals are not driven

251   by admixture from outside Africa, and instead likely reflect demographic events entirely within
252   Africa (**Fig. 3C**).
253
254   We carried out admixture graph modeling of the allele frequency correlations and found two
255   parsimonious models that fit the data. The first posited that present-day western Africans harbor
256   ancestry from a basal African lineage that contributed more to the Mende than it did to the
257   Yoruba, with the other source of western African ancestry being related to eastern Africans and
258   non-Africans (**Fig. 3D; Fig. S4; Fig. S5; Table S6**). The second model posited that long-range
259   and long-standing gene-flow has connected southern and eastern Africa to some groups in
260   western Africa (*e.g.* the ancestors of the Yoruba) to a greater extent than to other groups in
261   western Africa (*e.g.* the ancestors of the Mende) (**Fig. 3E**) (Pleurdeau et al., 2012). The possible
262   basal western African population lineage would represent the earliest known divergence of a
263   modern human lineage that contributed a major proportion of ancestry to present-day humans.
264   Such a lineage must have separated before the divergence of San ancestors, which is estimated to
265   have begun on the order of 200 to 300 thousand years ago (Scally and Durbin, 2012). Such a
266   model of basal western African ancestry might support the hypothesis that there has been ancient
267   structure in the ancestry of present-day Africans, using a line of evidence independent from
268   previous findings based on long haplotypes with deep divergences from other human haplotypes
269   (Hammer et al., 2011; Lachance et al., 2012; Plagnol and Wall, 2006). One scenario consistent
270   with this result could involve ancestry related to eastern Africans (and the out-of-Africa
271   population) expanding into western Africa and mixing there with more basal lineages. Our
272   genetic data do not support the theory that this putative basal lineage diverged prior to the
273   ancestors of Neandertals, since the African populations we analyze here are approximately
274   symmetrically related to Neandertals (Mallick et al., 2016; Prufer et al., 2014).
275
276   **A selective sweep targeting a taste receptor locus in southern Africa**. The availability of
277   ancient African genomes provides an opportunity to search for genomic footprints of natural
278   selection manifested as regions of greater allele frequency differentiation between ancient and
279   present-day populations than predicted by the genome-wide background. We compared to the
280   two ancient southern African ~2000 BP shotgun sequence genomes to six present-day high-
281   coverage San genomes with minimal recent mixture. The small number of ancient individuals
282   does not permit inference of changing allele frequencies at single loci, so we performed a scan
283   for high allele frequency differentiation in 500 kb windows with a step size of 10 kb. Using ~500
284   windows spaced at least 5 million base pairs apart as a null distribution, we found that the most
285   differentiated locus was 15 standard deviations from the observed genome-wide mean and
286   overlapped a cluster of eight taste-receptor genes on chromosome 12 (**Fig. 4A; Table 2**). Taste
287   receptor genes have previously been identified as targets of natural selection in humans, as they
288   modulate the ability to detect poisonous compounds in plants (Campbell et al., 2011).
289

6

290 **Polygenic adaptation.** Natural selection on phenotypic traits in humans is expected to only
291 occasionally take the form of sweeps on a single locus, instead acting on multiple genes
292 simultaneously to drive phenotypic adaptation (Coop et al., 2009). While a lack of genome-wide
293 association studies in eastern and southern Africans has left the genetic basis of phenotypic traits
294 far less well documented than it is for other populations, a variety of studies have linked broad
295 functional classes of genes to phenotypic traits. To test for evidence of selection on specific
296 functional categories of genes in present-day San since the divergence of the two ancient
297 genomes from southern Africa (**Fig. 4B**), we estimated allele frequency differentiation for 208
298 gene ontology categories with 50 or more genes in each, and computed weighted block jackknife
299 standard errors. The functional category that displays the most extreme allele frequency
300 differentiation between present-day San and ancient southern Africans is "response to radiation"
301 ($Z = 3.3$ compared to the genome-wide average). To control for the possibility that genes in this
302 category show an inflated allele frequency differentiation in general, we computed the same
303 statistic for the Mbuti central African rainforest hunter-gatherer group, but found no evidence for
304 selection affecting the "response to radiation" category (**Fig. 4C**). Instead, the top category for
305 the Mbuti is "response to growth", suggesting the possibility that the small stature of rainforest
306 hunter-gatherer populations such as the Mbuti may be an acquired adaptation (although we have
307 no ancient central African genome and thus no information about the time frame of selection).
308 We speculate that the signal for selection in the "response to radiation" category in the San could
309 be due to exposure to sunlight associated with the life of ‡Khomani and Ju'|hoan North in the
310 Kalahari Basin, which has harbored a larger proportion of San populations in the last millenia
311 due to encroachment from pastoralist and agriculturalist groups (Morris, 2002).
312
313 **Discussion**
314
315 This study, which multiplies by 16-fold the number of individuals with genome-wide ancient
316 DNA data from sub-Saharan Africa, highlights the power of ancient African genomes to provide
317 insights into prehistoric events that are difficult to discern based solely on analysis of present-day
318 genomes. We reveal the presence of a hitherto unknown cline of geographically structured
319 hunter-gatherer populations stretching from Ethiopia to South Africa, which we show existed
320 prior to the great populations transformations that occurred in the last few thousand years in
321 association with the spread of herders and farmers. We also document deeper structure in
322 western Africa, possibly predating the divergence of the ancestors of southern African hunter-
323 gatherers from other population lineages. We finally provide case examples of how populations
324 in eastern and southern Africa were transformed by the spread of food producers, and show how
325 the process gave rise to interactions with the previously established hunter-gatherers, with the
326 outcomes ranging from no detectable mixture in present-day populations to substantial mixture.
327 Our documentation of a radically different landscape of human populations before and after the
328 spread of food producers highlights the difficulty of reconstructing the African past based solely
329 on analysis of present-day populations and the importance of using ancient DNA to study deep

African population history in an era in which technological improvements have now made it feasible. It is clear that ancient DNA studies with larger sample sizes and covering a broader chronological and geographic range have the potential to make major progress in improving our understanding of African prehistory.

References

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

Behar, D., Van Oven, M., Rosset, S., Metspalu, M., Loogvali, E.-L., Silva, N., Kivisild, T., Torroni, A., and Villems, R. (2012). A "Copernican" reassessment of the human mitochondrial DNA tree from its root. Am J Hum Genet *90*, 675 - 684.

Briggs, A.W., and Heyn, P. (2012). Preparation of next-generation sequencing libraries from damaged DNA. Ancient DNA: Methods and Protocols, 143-154.

Briggs, A.W., Stenzel, U., Meyer, M., Krause, J., Kircher, M., and Pääbo, S. (2010). Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. Nucleic acids research *38*, e87-e87.

Bronk Ramsey, C. (2009). Bayesian analysis of radiocarbon dates. Radiocarbon *51*, 337-360.

Brown, T.A., Nelson, D.E., Vogel, J.S., and Southon, J.R. (1988). Improved collagen extraction by modified Longin method. Radiocarbon *30*, 171-177.

372 Busby, G.B.J., Band, G., Si Le, Q., Jallow, M., Bougama, E., Mangano, V.D., Amenga-Etego, L.N.,
373   Enimil, A., Apinjoh, T., Ndila, C.M., *et al.* (2016). Admixture into and within sub-Saharan Africa.
374   eLife *5*, e15266.
375 Campbell, M.C., Ranciaro, A., Froment, A., Hirbo, J., Omar, S., Bodo, J.-M., Nyambo, T., Lema, G.,
376   Zinshteyn, D., and Drayna, D. (2011). Evolution of functionally diverse alleles associated with PTC
377   bitter taste sensitivity in Africa. Molecular biology and evolution, msr293.
378 Cann, R.L., Stoneking, M., and Wilson, A.C. (1987). Mitochondrial DNA and human evolution. Nature
379   *325*, 31-36.
380 Chami, F. (2009). Zanzibar and the Swahili coast from c. 30,000 years ago (E&D Vision Pub.).
381 Chami, F., Khator, J., and Hamis Ali, A. (2009). The excavation of Mapangani cave, Pemba island,
382   Zanzibar. Studies in the African Past *9*, 74-101.
383 Clark, J. (1972). Prehistoric origins. The Early History of Malawi Longman, London, 17-27.
384 Clark, J.D. (1956). Prehistory in Nyasaland. The Nyasaland Journal *9*, 92-119.
385 Clark, J.D. (1959). The prehistory of southern Africa (Penguin Books).
386 Clark, J.D., and Clerk, J.D. (1973). ARCHAEOLOGICAL INVESTIGATION OF A PAINTED ROCK
387   SHELTER AT MWANA WA CHENCHERERE, NORTH OF DEDZA, CENTRAL MALAŴI In
388   July to September, 1972. The Society of Malawi Journal *26*, 28-46.
389 Cole-King, P. (1973). Kukumba Mbiri Mu Malaŵi: A Summary of Archaeological Research to March,
390   1973 (Government Press).
391 Coop, G., Pickrell, J.K., Novembre, J., Kudaravalli, S., Li, J., Absher, D., Myers, R.M., Cavalli-Sforza,
392   L.L., Feldman, M.W., and Pritchard, J.K. (2009). The role of geography in human adaptation. PLoS
393   genetics *5*, e1000500.
394 Crader, D.C. (1984). Faunal remains from Chencherere II rock shelter, Malawi. The South African
395   Archaeological Bulletin, 37-52.
396 Crowther, A., Prendergast, M.E., Fuller, D.Q., and Boivin, N. (2017). Subsistence mosaics, forager-
397   farmer interactions and the transition to food production in eastern Africa. Quaternary International.
398 Crowther, A., Veall, M.-A., Boivin, N., Horton, M., Kotarba-Morley, A., Fuller, D.Q., Fenn, T., Haji, O.,
399   and Matheson, C.D. (2015). Use of Zanzibar copal (Hymenaea verrucosa Gaertn.) as incense at
400   Unguja Ukuu, Tanzania in the 7–8th century CE: chemical insights into trade and Indian Ocean
401   interactions. Journal of Archaeological Science *53*, 374-390.
402 Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., García,
403   N., Pääbo, S., and Arsuaga, J.-L. (2013). Complete mitochondrial genome sequence of a Middle
404   Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proceedings of the National
405   Academy of Sciences *110*, 15758-15763.
406 Dabney, J., and Meyer, M. (2012). Length and GC-biases during sequencing library amplification: a
407   comparison of various polymerase-buffer systems with ancient and modern DNA sequencing
408   libraries. Biotechniques *52*, 87-94.
409 DeBusk, G.H. (1997). The distribution of pollen in the surface sediments of Lake Malawi, Africa, and the
410   transport of pollen in large lakes. Review of Palaeobotany and Palynology *97*, 123-153.
411 Fleisher, J., and LaViolette, A. (2013). The early Swahili trade village of Tumbe, Pemba Island, Tanzania,
412   AD 600–950. Antiquity *87*, 1151-1168.
413 Fu, Q., Hajdinjak, M., Moldovan, O., Constantin, S., Mallick, S., Skoglund, P., Patterson, N., Rohland,
414   N., Lazaridis, I., and Nickel, B. (2015). An early modern human from Romania with a recent
415   Neanderthal ancestor. Nature.
416 Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S.M., Bondarev, A.A., Johnson, P.L.F., Aximu-Petri, A.,
417   Prufer, K., de Filippo, C., *et al.* (2014). Genome sequence of a 45,000-year-old modern human from
418   western Siberia. Nature *514*, 445-449.
419 Gansauge, M.-T., and Meyer, M. (2013). Single-stranded DNA library preparation for the sequencing of
420   ancient or damaged DNA. Nature protocols *8*, 737-748.

421 Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai,
422   W.W., Fritz, M.H.Y*., et al.* (2010). A Draft Sequence of the Neandertal Genome. Science *328*, 710-
423   722.
424 Gronau, I., Hubisz, M.J., Gulko, B., Danko, C.G., and Siepel, A. (2011). Bayesian inference of ancient
425   human demography from individual genome sequences. Nature genetics *43*, 1031-1034.
426 Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K.,
427   Karthikeyan, S., Iles, L., Pollard, M.O., Choudhury, A*., et al.* (2015). The African Genome Variation
428   Project shapes medical genetics in Africa. Nature *517*, 327-332.
429 Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S.,
430   Harney, E., and Stewardson, K. (2015). Massive migration from the steppe is a source for Indo-
431   European languages in Europe. arXiv preprint arXiv:150202783.
432 Hammer, M.F., Woerner, A.E., Mendez, F.L., Watkins, J.C., and Wall, J.D. (2011). Genetic evidence for
433   archaic admixture in Africa. Proceedings of the National Academy of Sciences *108*, 15123-15128.
434 Helm, R., Crowther, A., Shipton, C., Tengeza, A., Fuller, D., and Boivin, N. (2012). Exploring
435   agriculture, interaction and trade on the eastern African littoral: preliminary results from Kenya.
436   Azania: Archaeological Research in Africa *47*, 39-63.
437 Helm, R.M. (2000). Conflicting histories: the archaeology of the iron-working, farming communities in
438   the central and southern coast region of Kenya (University of Bristol).
439 Henn, B.M., Gignoux, C.R., Jobin, M., Granka, J.M., Macpherson, J.M., Kidd, J.M., Rodríguez-Botigué,
440   L., Ramachandran, S., Hon, L., Brisbin, A*., et al.* (2011). Hunter-gatherer genomic diversity
441   suggests a southern African origin for modern humans. Proceedings of the National Academy of
442   Sciences *108*, 5154-5162.
443 Hogg, A.G., Hua, Q., Blackwell, P.G., Niu, M., Buck, C.E., Guilderson, T.P., Heaton, T.J., Palmer, J.G.,
444   Reimer, P.J., and Reimer, R.W. (2013). SHCal13 Southern Hemisphere calibration, 0–50,000 years
445   cal BP. Radiocarbon *55*, 1889-1903.
446 Jostins, L., Xu, Y., McCarthy, S., Ayub, Q., Durbin, R., Barrett, J., and Tyler-Smith, C. (2014). YFitter:
447   Maximum likelihood assignment of Y chromosome haplogroups from low-coverage sequence data.
448   arXiv preprint arXiv:14077988.
449 Juma, A. (2004). Unguja Ukuu on Zanzibar: An archaeological study of early urbanism.
450 Kennett, D.J., Plog, S., George, R.J., Culleton, B.J., Watson, A.S., Skoglund, P., Rohland, N., Mallick, S.,
451   Stewardson, K., Kistler, L*., et al.* (2017). Archaeogenomic evidence reveals prehistoric matrilineal
452   dynasty. *8*, 14115.
453 Kircher, M. (2012). Analysis of high-throughput ancient DNA sequencing data. In Ancient DNA
454   (Springer), pp. 197-228.
455 Kircher, M., Sawyer, S., and Meyer, M. (2011). Double indexing overcomes inaccuracies in multiplex
456   sequencing on the Illumina platform. Nucleic acids research, gkr771.
457 Korlević, P., Gerber, T., Gansauge, M.-T., Hajdinjak, M., Nagel, S., Ayinuer-Petri, A., and Meyer, M.
458   (2015). Reducing microbial and human contamination in DNA extractions from ancient bones and
459   teeth. BioTechniques *59*, 87-93.
460 Lachance, J., Vernot, B., Elbers, Clara C., Ferwerda, B., Froment, A., Bodo, J.-M., Lema, G., Fu, W.,
461   Nyambo, Thomas B., Rebbeck, Timothy R*., et al.* (2012). Evolutionary History and Adaptation from
462   High-Coverage Whole-Genome Sequences of Diverse African Hunter-Gatherers. Cell *150*, 457-469.
463 Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D.C., Rohland, N., Mallick, S., Fernandes, D., Novak, M.,
464   Gamarra, B., Sirak, K*., et al.* (2016). Genomic insights into the origin of farming in the ancient Near
465   East. Nature *536*, 419-424.
466 Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P.H., Schraiber,
467   J.G., Castellano, S., Lipson, M*., et al.* (2014). Ancient human genomes suggest three ancestral
468   populations for present-day Europeans. Nature *513*, 409-413.
469 Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform.
470   Bioinformatics *25*, 1754-1760.

471   Llorente, M.G., Jones, E.R., Eriksson, A., Siska, V., Arthur, K.W., Arthur, J.W., Curtis, M.C., Stock, J.T.,
472   Coltorti, M., Pieruccini, P., *et al.* (2015). Ancient Ethiopian genome reveals extensive Eurasian
473   admixture in Eastern Africa. Science *350*, 820-822.
474   Loh, P.-R., Lipson, M., Patterson, N., Moorjani, P., Pickrell, J.K., Reich, D., and Berger, B. (2012).
475   Inference of admixture parameters in human populations using weighted linkage disequilibrium.
476   Lohse, J.C., Madsen, D.B., Culleton, B.J., and Kennett, D.J. (2014). Isotope paleoecology of episodic
477   mid-to-late Holocene bison population expansions in the Southern Plains, USA. Quaternary Science
478   Reviews *102*, 14-26.
479   Longin, R. (1971). New method of collagen extraction for radiocarbon dating. Nature *230*, 241-242.
480   Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N.,
481   Nordenfelt, S., Tandon, A., *et al.* (2016). The Simons Genome Diversity Project: 300 genomes from
482   142 diverse populations. Nature *538*, 201-206.
483   Manhire, A. (1993). A report on the excavations at Faraoskop Rock Shelter in the Graafwater district of
484   the south-western Cape. Southern African Field Archaeology *2*, 3-23.
485   Maricic, T., Whitten, M., and Pääbo, S. (2010). Multiplexed DNA sequence capture of mitochondrial
486   genomes using PCR products. PloS one *5*, e14004.
487   Marshall, F., Stewart, K., and Barthelme, J. (1984). Early domestic stock at Dongodien in northern
488   Kenya. AZANIA: Journal of the British Institute in Eastern Africa *19*, 120-127.
489   Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S.A., Harney, E.,
490   Stewardson, K., Fernandes, D., Novak, M., *et al.* (2015). Genome-wide patterns of selection in 230
491   ancient Eurasians. Nature *528*, 499-503.
492   Meyer, M., Fu, Q., Aximu-Petri, A., Glocke, I., Nickel, B., Arsuaga, J.-L., Martínez, I., Gracia, A., de
493   Castro, J.M.B., and Carbonell, E. (2014). A mitochondrial genome sequence of a hominin from Sima
494   de los Huesos. Nature *505*, 403-406.
495   Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target
496   capture and sequencing. Cold Spring Harbor Protocols *2010*, pdb. prot5448.
497   Meyer, M., Kircher, M., Gansauge, M.-T., Li, H., Racimo, F., Mallick, S., Schraiber, J.G., Jay, F., Prüfer,
498   K., de Filippo, C., *et al.* (2012). A High-Coverage Genome Sequence from an Archaic Denisovan
499   Individual. Science *338*, 222-226.
500   Miller, S.F. (1971). The age of Nachikufan industries in Zambia. The South African Archaeological
501   Bulletin *26*, 143-146.
502   Moorjani, P., Patterson, N., Hirschhorn, J.N., Keinan, A., Hao, L., Atzmon, G., Burns, E., Ostrer, H.,
503   Price, A.L., and Reich, D. (2011). The history of African gene flow into Southern Europeans,
504   Levantines, and Jews. PLoS genetics *7*, e1001373.
505   Morris, A.G. (2002). Isolation and the origin of the Khoisan: Late Pleistocene and Early Holocene human
506   evolution at the southern end of Africa. Human Evolution *17*, 231-240.
507   Morris, A.G., Heinze, A., Chan, E.K., Smith, A.B., and Hayes, V.M. (2014). First ancient mitochondrial
508   human genome from a prepastoralist southern African. Genome biology and evolution *6*, 2647-2653.
509   Morris, A.G., and Ribot, I. (2006). Morphometric cranial identity of prehistoric Malawians in the light of
510   sub-Saharan African diversity. American journal of physical anthropology *130*, 10-25.
511   Odner, K. (1972). Excavations at Narosura, a Stone Bowl Site in the Southern Kenya Highlands. Azania:
512   Archaeological Research in Africa *7*, 25-92.
513   Pagani, L., Kivisild, T., Tarekegn, A., Ekong, R., Plaster, C., Gallego Romero, I., Ayub, Q., Mehdi, S.Q.,
514   Thomas, Mark G., Luiselli, D., *et al.* (2012). Ethiopian Genetic Diversity Reveals Linguistic
515   Stratification and Complex Influences on the Ethiopian Gene Pool. The American Journal of Human
516   Genetics *91*, 83-96.
517   Parkington, J., and Dlamini, N. (2015). First People: Ancestors of the San. Creda Communications.
518   Patin, E., Lopez, M., Grollemund, R., Verdu, P., Harmant, C., Quach, H., Laval, G., Perry, G.H.,
519   Barreiro, L.B., Froment, A., *et al.* (2017). Dispersals and genetic adaptation of Bantu-speaking
520   populations in Africa and North America. Science *356*, 543.

521   Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T.,
522   and Reich, D. (2012). Ancient admixture in human history. Genetics *192*, 1065-1093.
523   Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS genetics *2*,
524   e190.
525   Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., and Nieselt, K. (2016). EAGER:
526   efficient ancient genome reconstruction. Genome biology *17*, 1.
527   Phillipson, D.W. (1976). The Early Iron Age in eastern and southern Africa: a critical re-appraisal.
528   AZANIA: Journal of the British Institute in Eastern Africa *11*, 1-23.
529   Pickrell, J.K., Patterson, N., Barbieri, C., Berthold, F., Gerlach, L., Guldemann, T., Kure, B., Mpoloka,
530   S.W., Nakagawa, H., Naumann, C.*, et al.* (2012). The genetic prehistory of southern Africa. Nat
531   Commun *3*, 1143.
532   Pickrell, J.K., Patterson, N., Loh, P.-R., Lipson, M., Berger, B., Stoneking, M., Pakendorf, B., and Reich,
533   D. (2014). Ancient west Eurasian ancestry in southern and eastern Africa. Proceedings of the
534   National Academy of Sciences *111*, 2632-2637.
535   Pickrell, J.K., and Pritchard, J.K. (2012). Inference of population splits and mixtures from genome-wide
536   allele frequency data. PLoS genetics *8*, e1002967.
537   Plagnol, V., and Wall, J.D. (2006). Possible ancestral structure in human populations. PLoS Genet *2*,
538   e105.
539   Pleurdeau, D., Imalwa, E., Détroit, F., Lesur, J., Veldman, A., Bahain, J.-J., and Marais, E. (2012). "Of
540   Sheep and Men": Earliest Direct Evidence of Caprine Domestication in Southern Africa at Leopard
541   Cave (Erongo, Namibia). PLOS ONE *7*, e40340.
542   Prendergast, M.E., Mabulla, A.Z.P., Grillo, K.M., Broderick, L.G., Seitsonen, O., Gidna, A.O., and
543   Gifford-Gonzalez, D. (2013). Pastoral Neolithic sites on the southern Mbulu Plateau, Tanzania.
544   Azania: Archaeological Research in Africa *48*, 498-520.
545   Prendergast, M.E., Rouby, H., Punnwong, P., Marchant, R., Crowther, A., Kourampas, N., Shipton, C.,
546   Walsh, M., Lambeck, K., and Boivin, N.L. (2016). Continental island formation and the archaeology
547   of defaunation on Zanzibar, eastern Africa. PloS one *11*, e0149565.
548   Prufer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G.,
549   Sudmant, P.H., de Filippo, C.*, et al.* (2014). The complete genome sequence of a Neanderthal from
550   the Altai Mountains. Nature *505*, 43-49.
551   Ramachandran, S., Deshpande, O., Roseman, C.C., Rosenberg, N.A., Feldman, M.W., and Cavalli-
552   Sforza, L.L. (2005). Support from the relationship of genetic and geographic distance in human
553   populations for a serial founder effect originating in Africa. Proceedings of the National Academy of
554   Sciences of the United States of America *102*, 15942-15947.
555   Reich, D., Green, R.E., Kircher, M., Krause, J., Patterson, N., Durand, E.Y., Viola, B., Briggs, A.W.,
556   Stenzel, U., Johnson, P.L.F.*, et al.* (2010). Genetic history of an archaic hominin group from
557   Denisova Cave in Siberia. Nature *468*, 1053-1060.
558   Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., Parra, M., Rojas, W., Duque,
559   C., Mesa, N.*, et al.* (2012). Reconstructing Native American population history. Nature *488*, 370 -
560   374.
561   Reich, D., Thangaraj, K., Patterson, N., Price, A.L., and Singh, L. (2009). Reconstructing Indian
562   population history. Nature *461*, 489-494.
563   Reimer, P.J., Bard, E., Bayliss, A., Beck, J.W., Blackwell, P.G., Bronk Ramsey, C., Buck, C.E., Cheng,
564   H., Edwards, R.L., Friedrich, M.*, et al.* (2013). IntCal13 and Marine13 Radiocarbon Age Calibration
565   Curves 0–50,000 Years cal BP.
566   Renaud, G., Kircher, M., Stenzel, U., and Kelso, J. (2013). freeIbis: an efficient basecaller with calibrated
567   quality scores for Illumina sequencers. Bioinformatics, btt117.
568   Renaud, G., Slon, V., Duggan, A.T., and Kelso, J. (2015). Schmutzi: estimation of contamination and
569   endogenous mitochondrial consensus calling for ancient DNA. Genome biology *16*, 1.
570   Ribot, I., Morris, A.G., Sealy, J., and Maggs, T. (2010). Population history and economic change in the
571   last 2000 years in KwaZulu-Natal, RSA. Southern African Humanities *22*, 89-112.

572   Robinson, K., and Sandelowsky, B. (1968). The Iron Age of northern Malawi: recent work. AZANIA:
573   Journal of the British Institute in Eastern Africa *3*, 107-146.
574   Robinson, K.S.R. (1982). Iron Age of northern Malaŵi: an archaeological reconnaissance (Malaŵi Govt.
575   Ministry of Education and Culture).
576   Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., and Reich, D. (2015). Partial uracil–DNA–
577   glycosylase treatment for screening of ancient DNA. Philosophical Transactions of the Royal
578   Society of London B: Biological Sciences *370*.
579   Rohland, N., and Hofreiter, M. (2007). Comparison and optimization of ancient DNA extraction.
580   Biotechniques *42*, 343-352.
581   Russell, T., Silva, F., and Steele, J. (2014). Modelling the spread of farming in the Bantu-speaking
582   regions of Africa: an archaeology-based phylogeography. PLoS One *9*, e87854.
583   Sadr, K. (2015). Livestock First Reached Southern Africa in Two Separate Events. PloS one *10*,
584   e0134215.
585   Sadr, K., Smith, A., Plug, I., Orton, J., and Mütti, B. (2003). Herders and foragers on Kasteelberg: interim
586   report of excavations 1999-2002. The South African Archaeological Bulletin, 27-32.
587   Sandelowsky, B. (1972). Later Stone Age Assemblages from Malawi and their Technologies (Ph. D.
588   dissertation, University of California, Berkeley, Unpublished).
589   Scally, A., and Durbin, R. (2012). Revising the human mutation rate: implications for understanding
590   human evolution. Nature Reviews Genetics *13*, 745-753.
591   Schlebusch, C. (2010). Issues raised by use of ethnic-group names in genome study. Nature *464*, 487-487.
592   Schlebusch, C.M., Skoglund, P., Sjödin, P., Gattepaille, L.M., Hernandez, D., Jay, F., Li, S., De Jongh,
593   M., Singleton, A., Blum, M.G.B.*, et al.* (2012). Genomic Variation in Seven Khoe-San Groups
594   Reveals Adaptation and Complex African History. Science *338*, 374-379.
595   Schuenemann, V.J., Peltzer, A., Welte, B., van Pelt, W.P., Molak, M., Wang, C.-C., Furtwängler, A.,
596   Urban, C., Reiter, E., Nieselt, K.*, et al.* (2017). Ancient Egyptian mummy genomes suggest an
597   increase of Sub-Saharan African ancestry in post-Roman periods. *8*, 15694.
598   Sealy, J., Patrick, M., Morris, A., and Alder, D. (1992). Diet and dental caries among later stone age
599   inhabitants of the Cape Province, South Africa. American Journal of Physical Anthropology *88*, 123-
600   134.
601   Shipton, C., Crowther, A., Kourampas, N., Prendergast, M.E., Horton, M., Douka, K., Schwenninger, J.-
602   L., Faulkner, P., Quintana Morales, E.M., Langley, M.C.*, et al.* (2016). Reinvestigation of Kuumbi
603   Cave, Zanzibar, reveals Later Stone Age coastal habitation, early Holocene abandonment and Iron
604   Age reoccupation. Azania: Archaeological Research in Africa *51*, 197-233.
605   Shriver, M.D., Kennedy, G.C., Parra, E.J., Lawson, H.A., Sonpar, V., Huang, J., Akey, J.M., and Jones,
606   K.W. (2004). The genomic distribution of population substructure in four populations using 8,525
607   autosomal SNPs. Human Genomics *1*, 274.
608   Sinclair, P., Juma, A., and Chami, F. (2006). Excavations at Kuumbi Cave on Zanzibar 2005.
609   Skoglund, P., Mallick, S., Bortolini, M.C., Chennagiri, N., Hünemeier, T., Petzl-Erler, M.L., Salzano,
610   F.M., Patterson, N., and Reich, D. (2015). Genetic evidence for two founding populations of the
611   Americas. Nature.
612   Skoglund, P., Malmström, H., Raghavan, M., Storå, J., Hall, P., Willerslev, E., Gilbert, M.T.P.,
613   Götherström, A., and Jakobsson, M. (2012). Origins and genetic legacy of Neolithic farmers and
614   hunter-gatherers in Europe. Science *336*, 466-469.
615   Skoglund, P., Northoff, B.H., Shunkov, M.V., Derevianko, A.P., Pääbo, S., Krause, J., and Jakobsson, M.
616   (2014a). Separating endogenous ancient DNA from modern day contamination in a Siberian
617   Neandertal. Proceedings of the National Academy of Sciences.
618   Skoglund, P., Posth, C., Sirak, K., Spriggs, M., Valentin, F., Bedford, S., Clark, G.R., Reepmeyer, C.,
619   Petchey, F., Fernandes, D.*, et al.* (2016). Genomic insights into the peopling of the Southwest
620   Pacific. Nature *538*, 510-513.
621   Skoglund, P., Sjödin, P., Skoglund, T., Lascoux, M., and Jakobsson, M. (2014b). Investigating Population
622   History Using Temporal Genetic Differentiation. Molecular Biology and Evolution *31*, 2516-2527.

623   Smith, A.B. (1992a). Kasteelberg. In Guide to Archaeological Sites in the South-western Cape, A.B.
624   Smith, and B. Mutti, eds. (Cape Town: South African Association of Archaeologists Conference),
625   pp. 28-30.
626   Smith, A.B. (1992b). Pastoralism in Africa (Johannesburg: Witwatersrand University Press).
627   Smith, B.W. (1995). Rock art in south-central Africa: a study based on the pictographs of Dedza District,
628   Malawi and Kasam District, Zambia (University of Cambridge).
629   Stafford, T.W., Hare, P.E., Currie, L., Jull, A.J.T., and Donahue, D.J. (1991). Accelerator radiocarbon
630   dating at the molecular level. Journal of Archaeological Science *18*, 35-72.
631   Stuiver, M., and Polach, H.A. (1977). Discussion: reporting of 14 C data. Radiocarbon *19*, 355-363.
632   Tishkoff, S.A., Reed, F.A., Friedlaender, F.R., Ehret, C., Ranciaro, A., Froment, A., Hirbo, J.B.,
633   Awomoyi, A.A., Bodo, J.-M., Doumbo, O.*, et al.* (2009). The Genetic Structure and History of
634   Africans and African Americans. Science *324*, 1035-1044.
635   Van Klinken, G.J. (1999). Bone collagen quality indicators for palaeodietary and radiocarbon
636   measurements. Journal of Archaeological Science *26*, 687-695.
637   Van Oven, M., and Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human
638   mitochondrial DNA variation. Hum Mutat *30*, E386 - E394.
639   Veeramah, K.R., Wegmann, D., Woerner, A., Mendez, F.L., Watkins, J.C., Destro-Bisol, G., Soodyall,
640   H., Louie, L., and Hammer, M.F. (2012). An early divergence of KhoeSan ancestors from those of
641   other modern humans is supported by an ABC-based analysis of autosomal resequencing data.
642   Molecular biology and evolution *29*, 617-630.
643   Wang, Y., and Nielsen, R. (2012). Estimating population divergence time and phylogeny from single-
644   nucleotide polymorphisms data with outgroup ascertainment bias. Molecular Ecology *21*, 974-986.
645   Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J., Kronenberg,
646   F., Salas, A., and Schönherr, S. (2016). HaploGrep 2: mitochondrial haplogroup classification in the
647   era of high-throughput sequencing. Nucleic acids research *44*, W58-W63.
648   Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z.X.P., Pool, J.E., Xu, X., Jiang, H., Vinckenbosch,
649   N., Korneliussen, T.S.*, et al.* (2010). Sequencing of 50 Human Exomes Reveals Adaptation to High
650   Altitude. Science *329*, 75-78.
651   Zubieta, L.F. (2016). Learning through practise: Cheŵa women's roles and the use of rock art in passing
652   on cultural knowledge. Journal of Anthropological Archaeology *43*, 13-28.
653

Main Figure titles and legends

655

656 **Figure 1. Overview of ancient genomes and African population structure.** a) Map of
657 sampling locations in Africa and principal component analysis of all individuals. Present-day
658 individuals are indicated with gray circles. b) Automated clustering of key ancient- and present-
659 day populations (for K=7 cluster components). Present-day populations are labeled in gray.

660

661 **Figure 2. Ancestral components in eastern and southern Africa.** We show bar plots with the
662 proportions inferred for the best model for each target population. We used a model that inferred
663 the ancestry of each target population as 1-source, 2-source, or 3-source mixture of a set of
664 potential source populations. In (a) we show an example of the inferred model for
665 South_Africa_1200BP, an early pastoralist. A filled circle symbol in each panel indicates the
666 geographic location of the sample that we use as a representative of the source population. We
667 show five sources: b) South_Africa_2000BP representing forager populations in southern Africa
668 and a component of prehistoric Malawi and Tanzania that is no longer extant; c)
669 Ethiopia_4500BP which is today found in the Hadza but in the past was characteristic of eastern
670 African hunter-gatherers; d) the Mende from Sierra Leone which is related deeply to the western
671 African ancestry that was spread with the Bantu expansion of agriculturalists; e) the Savanna
672 Pastoral Neolithic sample Tanzania_Luxmanda_3100BP which provides a missing link of the
673 pastoralist population that brought ancestry most closely related to the ancient Levant to southern
674 Africa, and which is also closely but not exclusively related to present-day Cushitic speakers;
675 and f) ancestry more closely related to the Iran Neolithic than what is found in
676 Tanzania_Luxmanda_3100BP, and which may have entered the Horn of Africa in later
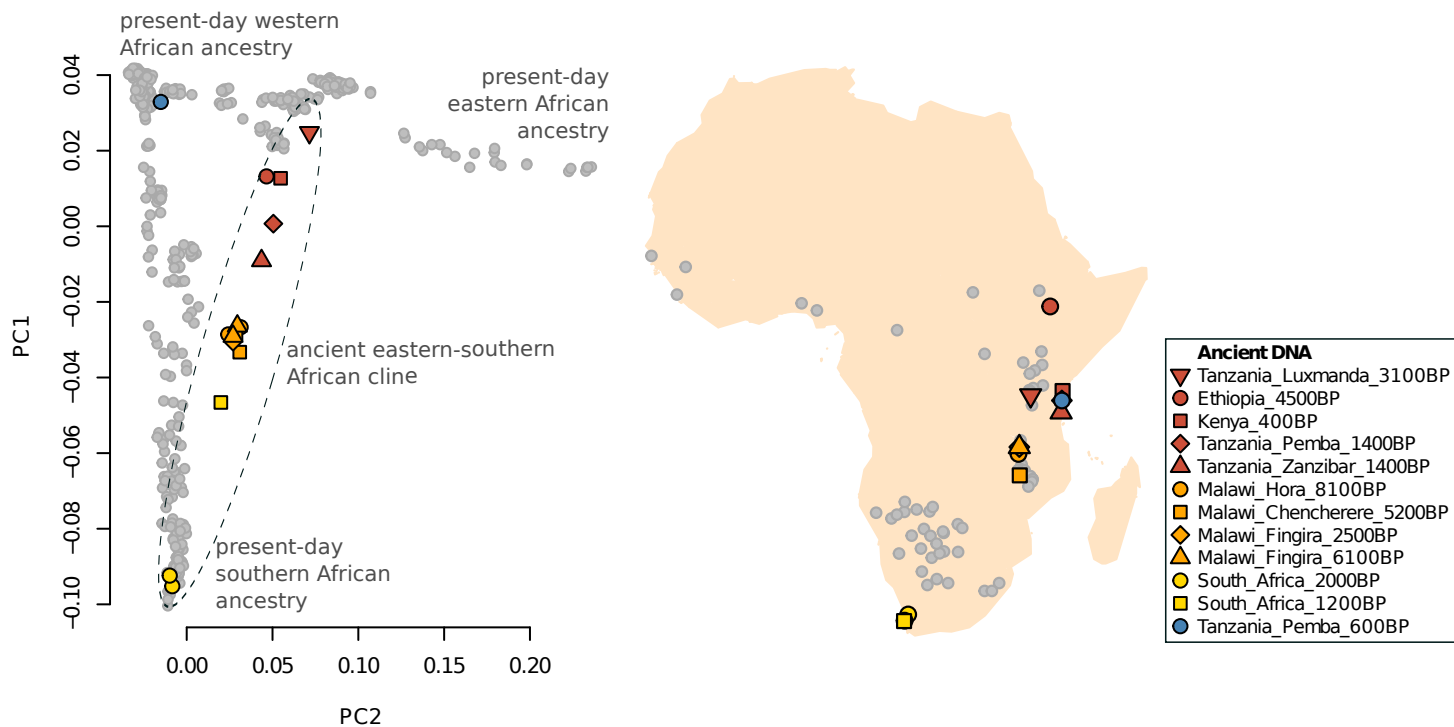677 migrations.

678

679 **Figure 3. Mixture events in the deeper population history of continental African lineages.**
680 A) Maximum likelihood tree of genome sequences from present-day and ancient populations,
681 excluding populations with evidence of asymmetrical allele sharing with non-Africans indicative
682 of recent gene flow (**Table S5**). Nodes with bootstrap support >95% are indicated with a circle.
683 B) A symmetry test of the hypothesis that ancient southern Africans are an outgroup lineage to
684 other African populations, which can be rejected for most pairs. C) Asymmetry between western
685 African Mende and Yoruba in the 1000 Genomes Project data is maximized in the Yoruba's
686 excess affinity to eastern Africans and non-Africans, but highly significant also for groups as
687 distant as southern Africans. D) Admixture Graph solution where Mende from Sierra Leone and
688 Yoruba from Nigeria have ancestry from a basal western African lineage. The other source of
689 western African ancestry is most closely related to eastern Africans and non-Africans (**Fig.
690 S5D**), which could be consistent with an expansion from eastern Africa. Note that the exact
691 proportion 'West Africa A' ancestry is not well constrained by the model, but the difference
692 between Yoruba and Mende is highly significant (panel C). E) Admixture graph solution where

693 the Yoruba have gene flow from a population related to both southern and eastern Africa, which
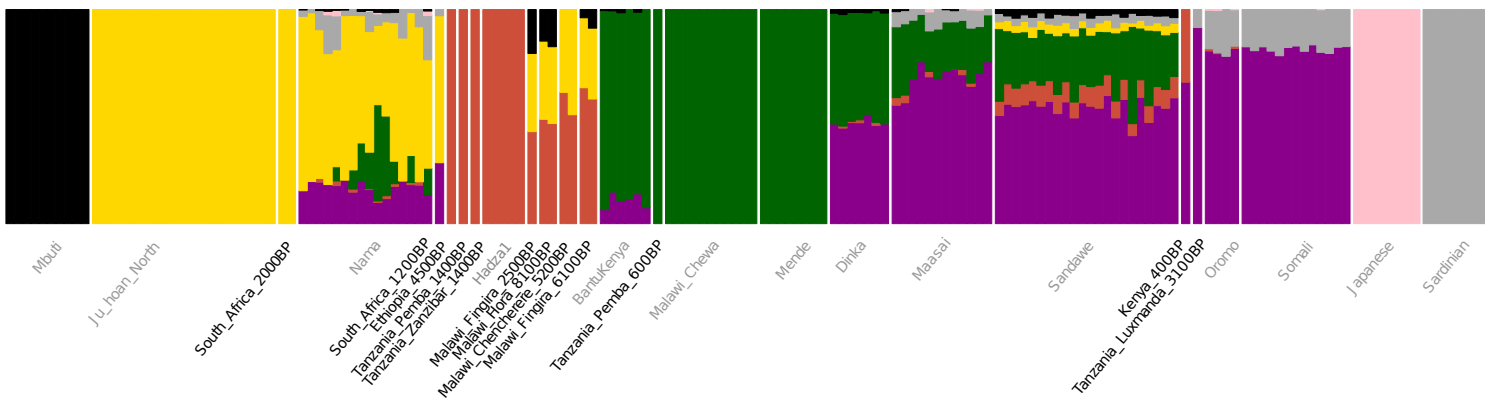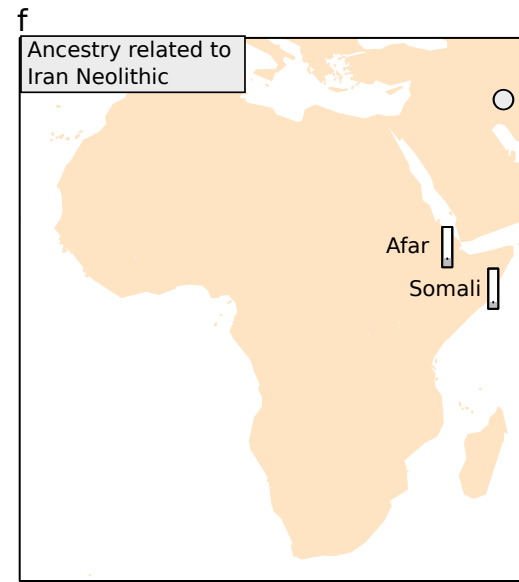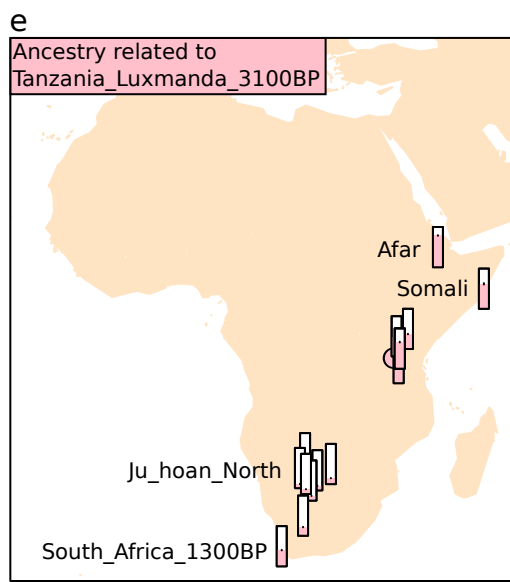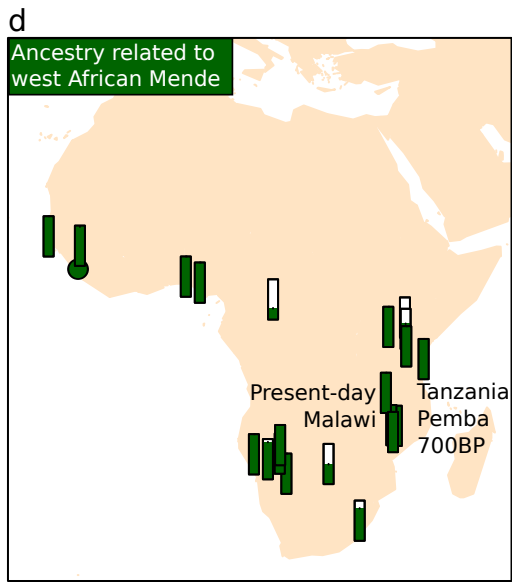694 could be consistent with a more complex pattern of isolation-by-distance in the continent.
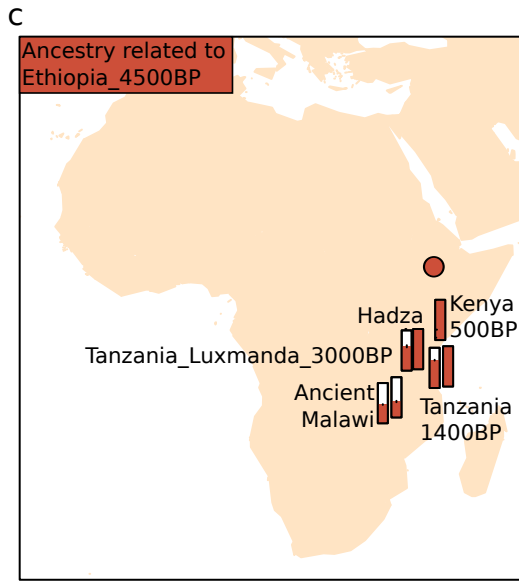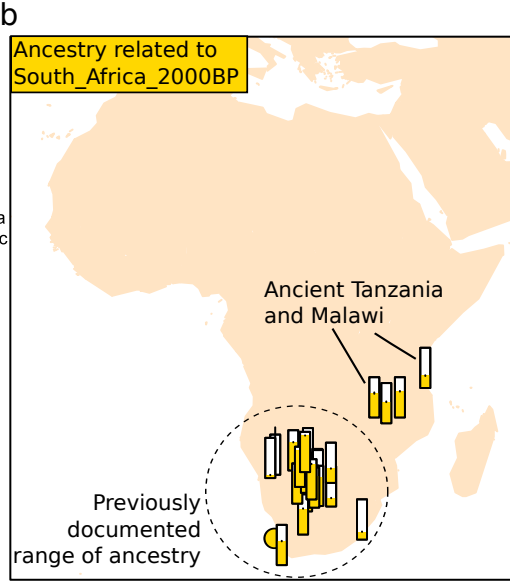695
696 **Figure 4. Ancient genomes provide evidence of natural selection in present-day southern**
697 **African San populations.** We computed branch-specific allele frequency differentiation in 6
698 present-day high-coverage San genomes compared to a pool of two ~2000 BP South African
699 genomes as an outgroup, using two approaches. A) We computed the statistic in windows of
700 500kb separated by 10kb. We also estimated genome-wide average and standard deviation of the
701 statistic using windows separated by at least 5 Mb, and transformed the genome wide
702 distribution of the sliding windows to be approximately normal (right panel). We observe
703 outliers 15 standard deviations from the mean in a taste receptor gene cluster on chromosome 12,
704 and a secondary peak in the Keratin Associated Protein 4 gene cluster. The outgroup used was 4
705 Central African Mbuti genomes. See **Table 2** for details on all major outlying regions. B)
706 Illustration of the branch-specific allele frequency differentiation approach. C) We computed the
707 statistic and block jackknife standard errors for 208 gene ontology categories with at least 50
708 genes each (y-axis). The outgroup used was western Africans. As a control to confirm that
709 outlier categories do not show larger magnitudes of allele frequency differentiation across
710 populations, we replaced the present-day San with the central African Mbuti (x-axis).

a

present-day western African ancestry

present-day eastern African ancestry

ancient eastern-southern African cline

present-day southern African ancestry

PC1

PC2

Ancient DNA

- Tanzania_Luxmanda_3100BP
- Ethiopia_4500BP
- Kenya_400BP
- Tanzania_Pemba_1400BP
- Tanzania_Zanzibar_1400BP
- Malawi_Hora_8100BP
- Malawi_Chencherere_5200BP
- Malawi_Fingira_2500BP
- Malawi_Fingira_6100BP
- South_Africa_2000BP
- South_Africa_1200BP
- Tanzania_Pemba_600BP

b

Mbuti

Ju_hoan_North

South_Africa_2000BP

Nama

South_Africa_1200BP
Ethiopia_4500BP
Tanzania_Pemba_1400BP
Tanzania_Zanzibar_1400BP

Hadza1

Malawi_Fingira_2500BP
Malawi_Hora_8100BP
Malawi_Chencherere_5200BP
Malawi_Fingira_6100BP

BantuKenya

Tanzania_Pemba_600BP

Malawi_Chewa

Mende

Dinka

Maasai

Sandawe

Kenya_400BP
Tanzania_Luxmanda_3100BP

Oromo

Somali

Japanese

Sardinian

a

Ust Ishim
Loschbour
Denisova
Orcadian
Mixe
Punjabi
Ami
Georgian
Anatolia Neolithic
Iranian
Iran Neolithic
Greek
Levant Neolithic
Mende
Mbuti
South Africa 2000BP
Dinka
Tanzania Luxmanda 3100BP
Ethiopia 4500BP

46±7%   54±7%

South Africa 1200BP

b Ancestry related to South_Africa_2000BP

Ancient Tanzania and Malawi

Previously documented range of ancestry

c Ancestry related to Ethiopia_4500BP

Hadza
Kenya 500BP
Tanzania_Luxmanda_3000BP
Ancient Malawi
Tanzania 1400BP

d Ancestry related to west African Mende

Present-day Malawi
Tanzania Pemba 700BP

e Ancestry related to Tanzania_Luxmanda_3100BP

Afar
Somali
Ju_hoan_North
South_Africa_1300BP

f Ancestry related to Iran Neolithic

Afar
Somali

A — Phylogenetic tree:
Sardinian, English, Han, Japanese, Ethiopia_4500BP, Dinka, Yoruba, Esan, Mende, Gambian, Mandenka, Luo, Luhya, BantuKenya, Biaka, Mbuti, Ju'/hoan North, Khomani San, South Africa 2000BP, Denisova, Altai, Vindija, Mezmaiskaya

B — $f_4(\text{Archaic, South Africa; Mende, Test}) \pm 1\ \text{SE}$

Malawi_Holocene, Ethiopia_4500BP, Dinka, Mbuti, Biaka, Yoruba, Mandenka, Esan, Gambian

Legend: ● South Africa 2000BP (yellow), ● San (Ju_hoan_North) (tan)

C — $f_4(\text{Archaic, Test; Mende, Yoruba}) \pm 1\ \text{SE}$

Dinka, Mixe, Japanese, Sardinian, Papuan, Ust_Ishim, Han, Ethiopia_4500BP, Biaka, Mbuti, Malawi_Holocene, South_Africa_2000BP, Khomani_San, Ju_hoan_North

D — Archaic genomes; West Africa A, West Africa B, Ethiopia 4500BP, South Africa 2000BP; Mende (Sierra Leone) 13%, 9%, Yoruba (Nigeria)

E — Archaic genomes; Mende (Sierra Leone), Ethiopia 4500BP, South Africa 2000BP; 31%, Yoruba (Nigeria) 2%

a

Z-score

TAS2R
cluster

KRTAP5
cluster

500kb windows separated by 10kb

b

Outgroup

South Africa
2000-2300BP

Allele frequency
differentiation

Present-day San

c

Mbuti genome-wide
estimate

Response to radiation

Response to abiotic stimulus

*Regulation of immune
system process*

San genome-wide
estimate

Regulation of growth

Allele frequency differentiation in present-day San
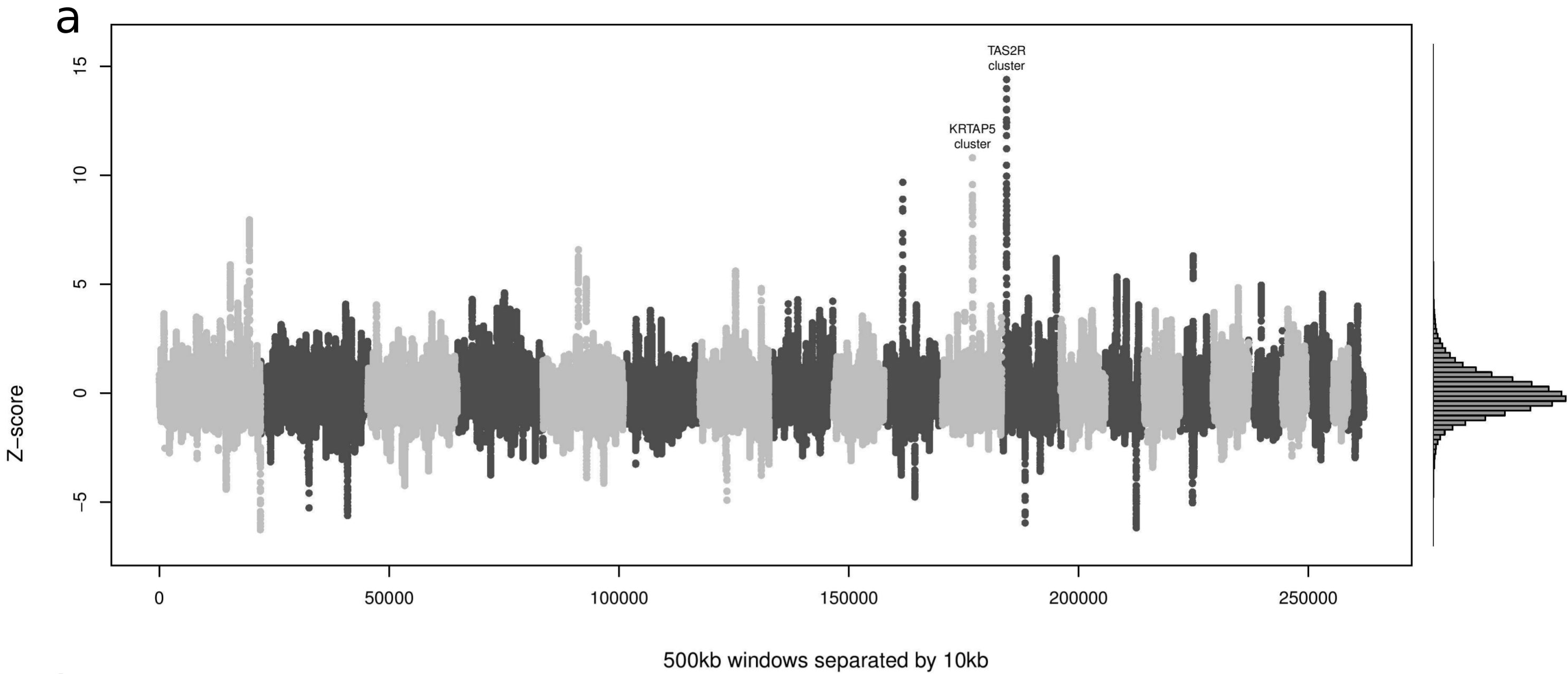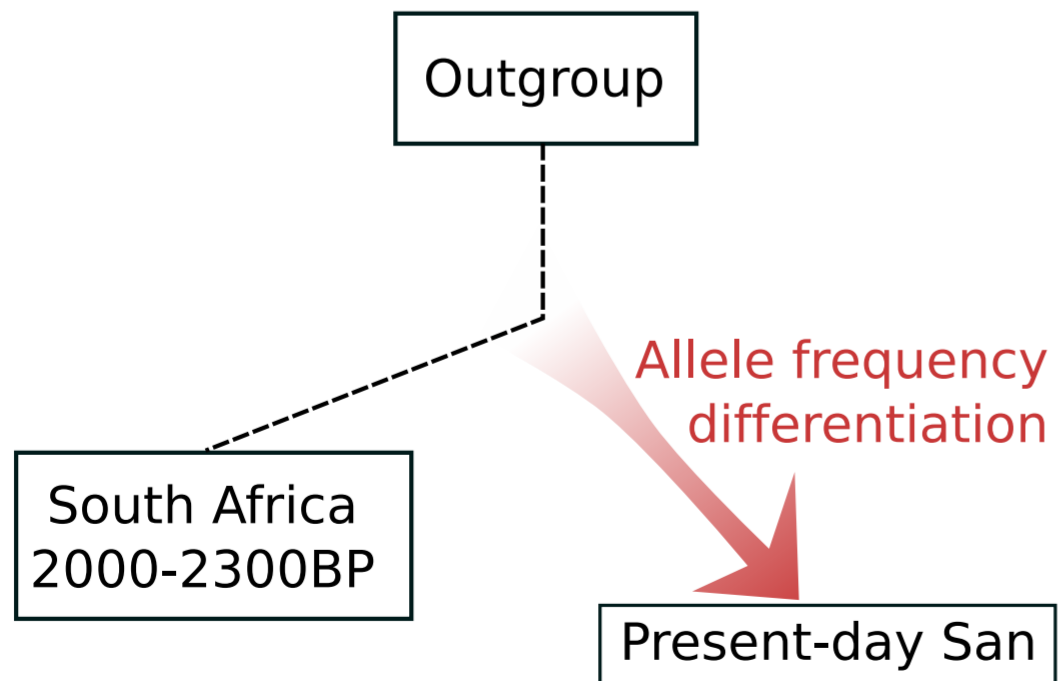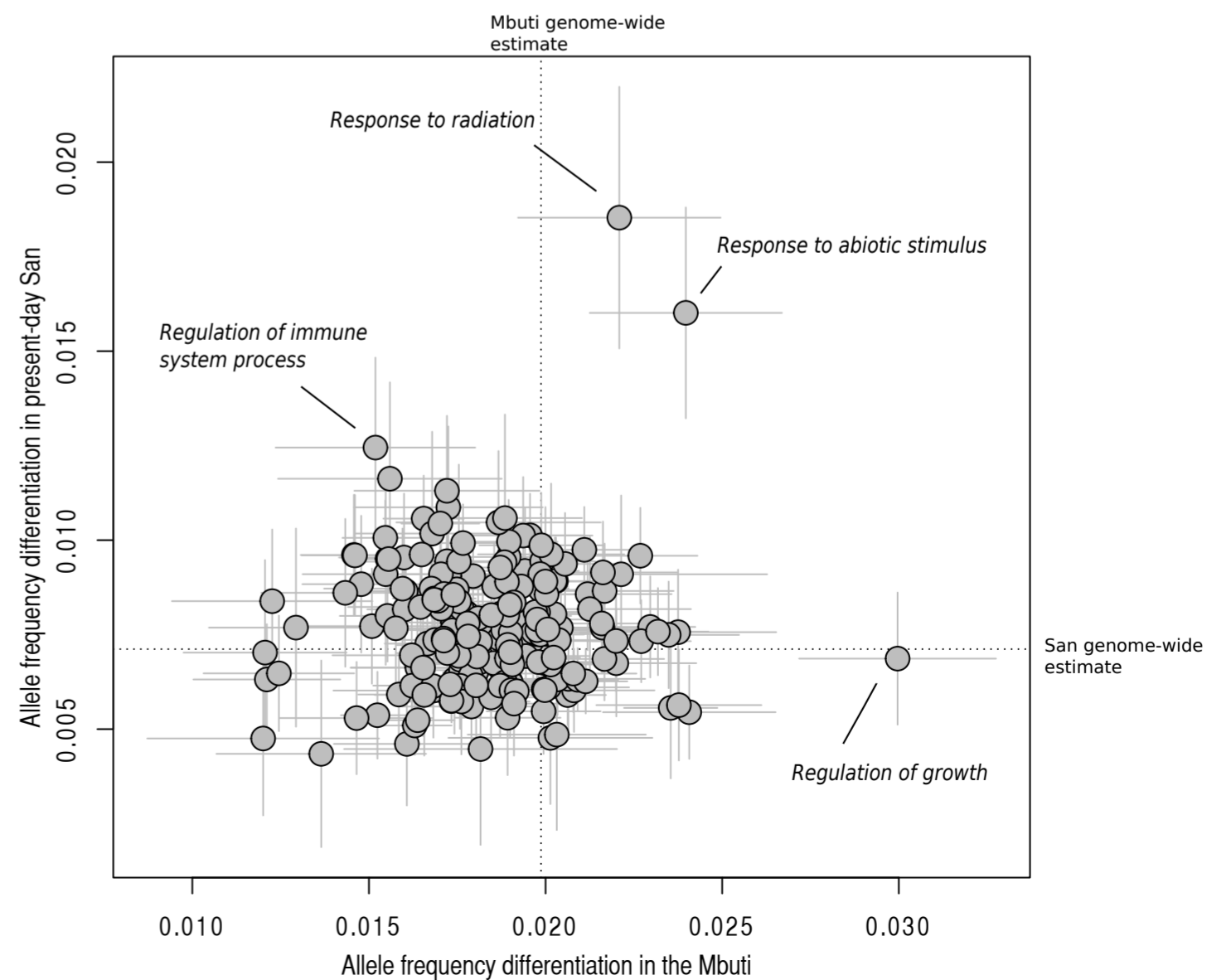
Allele frequency differentiation in the Mbuti

Main Tables and legends

Table 1. Summary of ancient DNA from 15 prehistoric individuals newly reported in this study.

ID	Population label	Date* In calibrated years before present defined as 1950 (calBP)	Location	Lat.	Long.	Y chromosome haplogroup	mtDNA hg	Damage rate at 5' CpG dinucleotides	SNPs hit on 1.2M auto-somal targets
I9028	South_Africa_2000BP	2241-1965 calBP (2330±25 BP, UGAMS-7255)	St. Helena, South Africa	-32.8	18.0	A1b1b2a	L0d2c1[$]	28%	731,098 (1.1X shotgun)
I9133	South_Africa_2000BP	2017-1748 calBP (2000±50 BP, Pta-5283)	Faraoskop Rock Shelter, South Africa	-32.0	18.5	A1b1b2a	L0d1b2b1b	33%	1,028,904 (2.3X shotgun)
I9134	South_Africa_1200BP	1282-1069 calBP (1310±50 BP, Pta-4373)	Kasteelberg, South Africa	-32.8	17.9	Female	L0d1a1a	21%	641,971 (0.8X shotgun)
I4427	Malawi_Fingira_6100BP	6175-5913 calBP (5270±25 BP, UCIAMS-186346)	Fingira, Malawi	-10.8	33.8	BT	L0d1b2b	37%	99,341
I4468	Malawi_Fingira_6100BP	6177-5923 calBP (5290±25 BP, UCIAMS-186347)	Fingira, Malawi	-10.8	33.8	BT	L0d1c	49%	30,257
I4421	Malawi_Chencherere_5200BP	5400-4800 calBP (radiocarbon dating was unsuccessful; the date is based on context of other materials from the same site)	Chencherere, Malawi	-14.4	33.8	Female	L0k2	28%	59,470
I4422	Malawi_Chencherere_5200BP	5293-4979 calBP (4525±25 BP, UCIAMS-186348)	Chencherere, Malawi	-14.4	33.8	Female	L0k1	42%	9,355
I2966	Malawi_Hora_8100BP	10000-5000 calBP (radiocarbon dating was unsuccessful; the date is based on context of other materials from the same site)	Hora, Malawi	-11.7	33.6	BT	L0k2 (PMDS>3)	54%	610,605
I2967	Malawi_Hora_8100BP	8173-7957 calBP (7230±60 BP, PSUAMS-2535)	Hora, Malawi	-11.7	33.6	Female	L0a2	48%	65,686
I4426	Malawi_Fingira_2500BP	2676-2330 calBP [2676-2343 calBP (2425±20 BP, PSUAMS-1734), 2483-2330 calBP (2400±20 BP, PSUAMS-1881)]	Fingira, Malawi	-10.8	33.8	Female	L0f	39%	635,427
I3726	Tanzania_Luxmanda_3100BP	3141-2890 calBP (2925±20 BP, ISGS-A3806)	Luxmanda, Tanzania	-4.3	35.3	Female	L2a1	65%	845,016
I0589	Tanzania_Zanzibar_1400BP	1370-1303 calBP (1479±23 BP, OxA-31427)	Kuumbi Cave, Zanzibar Island, Tanzania	-6.4	39.5	Female	L4b2a2c	22%	752,917
I1048	Tanzania_Pemba_1400BP	1421-1307 calBP (1520±30 BP, Beta-434912)	Makangale Cave, Pemba Island, Tanzania	-4.9	39.6	Female	L0a	42%	168,117
I2298	Tanzania_Pemba_600BP	639-544 calBP (623±20 BP, Wk-43308)	Makangale Cave, Pemba Island, Tanzania	-4.9	39.6	Female	L2a1a2	29%	695,242
I0595	Kenya_400BP	496-322 calBP (388±27 BP, OxA-30803)	Panga ya Saidi, Kenya	-3.7	39.7	E1b1b1b2	L4b2a2	30%	150,383

* Table S2 provides detailed information on the direct radiocarbon dating measurements.
$ Consistent with previously published mtDNA sequence by (Morris et al., 2014).

717

Table 2. Top five candidate regions identified in genome-wide scan for selective sweeps in present-day San populations in southern Africa compared to ancient genomes.

Rank	Chrom.	start-end	f_3-statistic	Z-score	Genes in top 500kb window in peak region
1	12	11,123,548-11,623,548	0.163	14.3	TAS2R43, PRH1-PRR4, TAS2R20, TAS2R50, TAS2R42, TAS2R46, TAS2R30, TAS2R31, PRB1, PRB2, PRB3, PRB4, LOC100129361, TAS2R19
2	11	71,208,258-71,708,258	0.125	10.8	LOC100129216, KRTAP5-7, DEFB108B, KRTAP5-8, NADSYN1, KRTAP5-9, FAM86C1, RNF121, ALG1L9P, LOC100133315, KRTAP5-10, KRTAP5-11
3	10	46,069,893-46,569,893	0.113	9.7	DQ577099, PTPN20B, PTPN20A, ZFAND4, AGAP4, DQ588224, FAM21C
4	1	224,960,062-225,460,062	0.095	8.0	DNAH14
5	5	82,375,629-82,875,629	0.08	6.6	VCAN, XRCC4

720

Supplemental Figure titles and legends

722

723 **Figure S1. Archaeological context and dating of newly reported human remains from**
724 **Malawi.** Related to Figure 1. A) Locations of the three sites in Malawi. B) Plan of the Hora rock
725 shelter showing locations of the original Hora 1 (male) and Hora 2 (female) burials relative to
726 new excavations from 2016. C) Plan of Fingira Cave showing locations of human remains
727 recovered in 1966 (green dots) and 2016 (green stars). D) Plan of Chencherere II rock shelter
728 redrawn from Crader (1984), showing approximate locations where human remains were
729 recovered in 1972 (red and green stars), and the approximate locations of the remains sampled
730 from Nguludi for this study (blue stars). E) Ultrafiltered collagen FTIR spectra for three
731 individuals from Malawi (MW2 = UCIAMS-186346, MW3 = UCIAMS-18647, MW4 =
732 UCIAMS-186348) compared to well-preserved standards and samples (Modern Carabou, BU4,
733 RP01, POR1). These samples show the typical features for well-preserved collagen. The top of
734 the figure shows the FTIR spectra for a single XAD purified sample (MW1 = PSUAMS-1734)
735 compared against a well preserved standard (Beauford Whale). These data, along with the stable
736 isotopes and C:N ratios, indicate that these samples are well preserved and that the AMS ^{14}C
737 dates are reliable.

738

739 **Figure S2. Ancient individuals and African population structure.** Related to Figure 1. A) full
740 legend and PC1-PC2 scatter plot (same analysis as Figure 1) with all present-day populations
741 indicated with separate symbols. B) PC1-PC2 computed only with southern African populations
742 (and west African Yoruba). C) Symmetry statistics confirming that ancient individuals from the
743 western Cape region in South Africa are more closely related to southern San such as the
744 Khomani rather than more northern San populations such as the Ju_hoan_North.

745

746 **Figure S3. ADMIXTURE clustering analysis of K=2 to K=7 clusters.** Related to Figure 1.
747 Each bar shows inferred cluster membership in each individual for a given number of clusters.
748 This analysis used 431,134 SNPs that were not in a CpG dinucleotide context. We also show an
749 inset with cluster membership inferred for damage-restricted ancient DNA sequences (PMD-
750 score >3) in a parallel run with only 111,208 transversion SNPs.

751

752 **Figure S4. Admixture models of African population history.** Related to Figure 3. A) A tree-
753 like model. Tests of all possible triplet topologies predicted to be symmetric by the tree obtained
754 for major African population lineages in the maximum likelihood model in **Fig. 3A**. B-H) Under
755 each graph we show the outlier f_4 statistic with the greatest Z-score between model and empirical
756 values. When the same group appears twice in an f_4 statistic (e.g. f_4(Archaic, South_Africa_HGs;
757 South_Africa_HGs,YRI), this corresponds to an f_3 statistic. Graphs F and H are the most
758 parsimonious fits with only two admixture events.

759   **Figure S5. Automatically adding additional populations to a skeleton graph.** Related to
760   Figure 3. A) Skeleton graph invoking widespread admixture from a hypothesized ancestral
761   northeast African population. B-D) Best fitting insertion points for Malawi_Hora_8100BP
762   shotgun sequence data (B), Mbuti (C), and Japanese non-Africans (D), in the skeleton graph
763   displayed in (A). For details see **Table S6**.

764

STAR Methods text

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, David Reich (reich@genetics.med.harvard.edu)

EXPERIMENTAL MODEL AND SUBJECT DETAILS

We generated new genome-wide data from skeletal remains of 15 prehistoric individuals: 5 from eastern Africa, 7 from south-central Africa, and 3 from southern Africa (**Table 1; Table S1; Table S2**). One of these individuals, from St. Helena Bay and directly dated to ~2100 BP, previously yielded a complete mitochondrial genome (Morris et al., 2014). We directly dated a second South African individual buried in a hunter-gatherer context from Faraoskop to ~2000 BP, and a third individual buried in a pastoralist context from Kasteelberg to ~1,200 BP. We also directly dated and used in-solution enrichment to obtain genome-wide DNA from four individuals from coastal eastern Africa: From the cave site of Panga ya Saidi in the coastal region of southeastern Kenya (~400BP), Kuumbi Cave in the southeast of Zanzibar Island (Tanzania; ~1,400 BP), and Makangale Cave in the northwest of Pemba Island (Tanzania; ~1,400 BP and ~600 BP). We also obtained genome-wide data from a ~3100 BP individual from a pastoralist context in north-central Tanzania, and ~8100-2500 year old individuals from Malawi.

Terminology. There is no widely accepted term with neutral connotations for indigenous communities in southern Africa (Schlebusch, 2010). In this manuscript, we follow San council recommendations in using population-specific terms whenever possible, and alternatively the terms San for Tuu and K'xaa speaking hunter-gatherer groups and Khoe for Khoe-khoe speakers. When necessary we collectively refer to groups with southern Africa-specific ancestry as Khoe-San, or as having San-related ancestry.

Panga ya Saidi Cave, Kilifi County, Kenya (n=1). Panga ya Saidi is a large limestone cave complex formed within an escarpment c. 15 km from the Indian Ocean coast in southern Kenya. Excavated in multiple Sealinks Project campaigns, the cave's long and complex depositional sequence spans, discontinuously, more than 76,000 years, and contains mainly Later Stone Age (LSA) deposits, overlain by Middle Iron Age (MIA) and Later Iron Age (LIA) deposits dating to the last two thousand years (Helm et al., 2012). The sampled specimen (I0595, Kenya_400BP) is a phalanx recovered from an *in situ* burial (context 403) and directly AMS radiocarbon dated to 496-322 calBP (388±27 BP, OxA-30803). The individual was a tall, robust young adult male. He was buried in a shallow grave in a crouched position with two hands and one foot in the small of the back and the skull disarticulated and placed by the knees. The individual was buried by

803   sediment containing marine shell beads, small knapped stone tools, and Tana Tradition
804   potsherds. The associated faunal remains are exclusively wild, with the exception of a single
805   possible caprine bone. Large numbers of remains of birds, rodents, and other microfauna suggest
806   that the cave may have only been sporadically occupied when the human remains were
807   deposited. We infer from the material culture and fauna that the cave was occupied by foragers
808   during the time the individual was buried, although food producers were present at nearby
809   settlements such as Mtsengo and Mbuyuni (Helm, 2000). Crop remains of African sorghum,
810   pearl millet and finger millet found at the site suggest these foragers had access to agricultural
811   resources.

812

813   **Makangale Cave, Pemba, Tanzania (n=2).** This limestone cave at the northern end of Pemba
814   Island in the Zanzibar archipelago has been excavated in multiple campaigns, the most recent
815   two seasons conducted by the ERC-funded Sealinks Project at Oxford University in 2012 and
816   then the Max Planck Institute for the Science of Human History in 2016 (unpublished; see also
817   (Chami et al., 2009)). The sequence shows clear evidence of human occupation beginning
818   around 1400 BP with an escargotière layer of giant African land snail shells, pottery, and
819   disarticulated human remains. Above this layer, the sequence shows regular human use of the
820   cave into the last thousand years. The first sampled specimen (I0589, Tanzania_Pemba_1400BP)
821   is a sacral vertebra from context 204 (Sealinks Project faunal catalog no. 15336), directly dated
822   to 1421-1307 calBP (1520±30 BP, Beta-434912). The second specimen (I2298,
823   Tanzania_Pemba_600BP) is a lower molar from context 301 (Sealinks Project faunal catalog no.
824   15624), which lies just below the surface and was dated to 639-544 calBP (623±20 BP, Wk-
825   43308). Both specimens are associated with a highly unusual faunal assemblage, dominated by
826   fragmented crocodile (*Crocodylus* cf. *niloticus*) remains and diverse microfauna, including
827   *Rattus rattus* (Asian black rat), a nonnative rodent that must have arrived to the area via maritime
828   exchange routes. There are no taphonomic indicators in the faunal assemblage of hunting by
829   humans, nor of crocodile predation on humans. During both of the occupational phases targeted
830   in this study, there were nearby settlements occupied by farmers whose ancestors likely came
831   from the mainland, for example at the sites of Tumbe (c. 1400-1000 BP) and Chwaka (1000-400
832   BP) (Fleisher and LaViolette, 2013).

833

834   **Kuumbi Cave, Zanzibar, Tanzania (n=1).** Kuumbi Cave is a limestone solutional cave
835   excavated in multiple campaigns (Chami, 2009; Sinclair et al., 2006). Sealinks Project
836   excavations in 2012 documented a complex depositional sequence stretching discontinuously
837   over 20,000 years, containing evidence of LSA and MIA occupations in five discernible phases
838   (Shipton et al., 2016). The analyzed specimen (I0589, Tanzania_Zanzibar_1400BP) is a
839   complete second phalanx of an adult (Sealinks Project faunal catalog no. 4353). It was recovered
840   from context 1011, in association with local Tana Tradition ceramics typical of the MIA,
841   moderately-sized limestone lithic artifacts, and diverse wild game animals, but no additional
842   human remains (Prendergast et al., 2016). The specimen is directly dated to 1370-1303 calBP

843 (1479±23 BP, OxA-31427), thus placing it at the beginning of the MIA phase. While Kuumbi
844 Cave is interpreted as a forager site, elsewhere on the island at this time, large settlements such
845 as Unguja Ukuu emerge, occupied by farmers whose origins are likely on the mainland
846 (Crowther et al., 2015; Juma, 2004).

847

848 **Luxmanda, Babati District, Tanzania (n=1)**. Luxmanda is an open-air settlement sitting atop
849 the Rift Valley escarpment (1878 m above sea level) at the southern edge of the Mbulu Plateau,
850 just north of Lake Balangida and Mount Hanang. Excavations in 2012, 2013, and 2015 by the
851 RAPT project (Research on the Archaeology of Pastoralism in Tanzania) have shown Luxmanda
852 to be the largest and southernmost known settlement site of the Pastoral Neolithic (PN), the era
853 corresponding to the spread of mobile livestock herding in eastern Africa (Prendergast et al.,
854 2013) (Grillo, Prendergast et al. forthcoming). Despite its isolated location, Luxmanda shows
855 strong material culture affinities to sites of southern Kenya classified as Savanna Pastoral
856 Neolithic (SPN), in particular the Narosura type-site (Odner, 1972); Luxmanda's ties to SPN
857 sites are further supported by sourcing of obsidian stone tools to the Naivasha Basin in the
858 Central Rift Valley of Kenya. Faunal remains from Luxmanda indicate a diet almost exclusively
859 focused on sheep, goat, and cattle; botanical remains are currently under study. A suite of eleven
860 radiocarbon dates provides a tightly constrained window of occupation c. 3000-2900 calBP,
861 which is at the early end of the range for SPN sites. The analyzed specimen
862 (Tanzania_Luxmanda_3100BP) is a petrous bone from a perinatal infant. The infant was found
863 complete and buried just to the west of, and c. 35 cm below, a burnt earth feature, interpreted as
864 a hearth. The burnt earth feature was then overlain by domestic refuse. Collagen from the same
865 petrous bone was AMS radiocarbon dated to 3141-2890 calBP (2925±20 BP, ISGS-A3806), a
866 date nearly identical to those of charcoal samples taken from the overlying burned earth feature
867 and domestic refuse.

868

869 **Hora, Malawi (n=2).** The Hora 1 and Hora 2 skeletons from Malawi were excavated from the
870 Hora 1 site in the Mzimba District of Malawi, located on the northeastern side of Mount Hora.
871 Mount Hora lies south of the Nyika Plateau and west of the Viphya Mountains, where miombo
872 vegetation grades southwest into edaphic grasslands (DeBusk, 1997). The region divides the
873 Luangwa River Basin in Zambia from Lake Malawi ~130km to the east, which receives all water
874 from the district via the South Rukuru River (**Fig. S1**). The South Rukuru flows south-to-north
875 along the Zambian border before turning east to intersect the Kasitu, a major interior waterway
876 that flows past Mount Hora and divides the mountains to the east from plains to the west. It is
877 notable that in the Hora region, rock art, stone tools, and burial practices all have substantial
878 differences from those even in the nearby Luangwa Valley of Zambia, suggesting cultural
879 subdivisions across relatively small areas (Clark, 1959; Phillipson, 1976; Sandelowsky, 1972).

880

881 Hora is a distinctive granite-gneiss inselberg that rises 110m from superficial floodplain deposits.
882 Hora 1 is a large overhang at the base (1,420m AMSL) that covers ~80 m^2. Although the shelter

has no surviving rock art, at least four other shelters on the inselberg contain paintings that include white stars and abstract red "gridirons" or "nets", both of which are motifs replicated at other sites on Mount Hora and nearby localities (Clark, 1956; Cole-King). Hora 1 was excavated by Clark in 1950, and produced a 2.2m cultural sequence containing pottery and iron slag at the top, faunal remains and LSA lithic assemblages with mollusk shell beads below this, and two human burials approximately 70cm below the surface (Clark, 1956). The first burial to be revealed was Hora 1 (UCT-242), a short-statured male in his thirties or forties at death, who may have been buried with a flaked stone axe (Morris and Ribot, 2006). The skeleton was left partially *in situ* during the original excavation and later exhumed by Rangeley (Clark, 1956). About 3 m to the south, the skeleton of a female was recovered (Hora 2, UCT-243); she was of similar stature to the male and died in her early twenties (Morris and Ribot, 2006). The female was found in a flexed position on her left side, and most of the bones of the hands and feet were missing or had been displaced – suggesting a degree of exposure prior to burial (Clark, 1956).

A previous study (Clark, 1956) reports three major stratigraphic units at Hora 1, and places the burials in the upper part of the second unit in association with the "Nachikufan II" – an industry with a type site ~300 km to the east in Zambia. Clark does not consider the burials intrusive, noting that they are overlain exclusively by hunter-gatherer material culture. The earliest occupation layer at the site is "reddish brown earth" that begins at about 1.7m depth, and contains a lithic assemblage assigned by Clark to the Nachikufu I. On the basis of typology, the earliest deposits at the site may date to between 16000-11000 BP, the burials to between 10000-5000 BP, and materials in the uppermost unit are likely as recent as the last few hundred years (Miller, 1971).

As this age range is imprecise, new pilot excavations were conducted in 2016. These confirmed that there is no pottery, slag, or other indication of non-hunter-gatherer material culture within the 0.5m overlying the depth of the burials. Unfortunately, attempts to directly date the Hora 1 skeletons through ^{14}C AMS failed at two labs because of lack of preserved collagen. However, the other Malawi specimens reported here were all directly dated (**Table 1; Table S2**). The Hora 1 and Hora 2 genetic data show that these specimens align genetically with the other ancient individuals from Malawi, which cover a time period from approximately 6200-2300 calBP. Their cultural associations, state of preservation, and genetic affinities therefore together place the Hora specimens in the late Holocene of Malawi.

Fingira, Malawi (n=3). The three Fingira samples derived from *ex situ* human remains (2 adult femora and one subadult femur) recovered in 2016 from Fingira Rock, a large shelter located within the boundaries of Nyika National Park in northern Malawi. The climate is cool and moist compared to the surrounding regions (~1,600 mm annual rainfall, and average daily highs 10 – 20° C). Fingira Rock is an isolated inselberg at ~2100 meters above sea level, near the upper limit of the miombo woodland. Within this inselberg is set a single deep rock shelter with 160m[2]

923 of deposit (**Fig. S1**). First excavated in 1966, Fingira yielded fragmentary remains of at least 15
924 human individuals (subadults and adults), plus one more complete burial near the front, in
925 association with rich lithic, archaeofaunal, and archaeobotanical assemblages. All human
926 remains previously recovered from Fingira were studied by Brothwell, and reported in in a
927 previous study (Sandelowsky, 1972). They are curated at the Natural History Museum in
928 London.

930 In addition to human remains, the Fingira deposits contained bone tools, pigments, and
931 ornaments (Robinson and Sandelowsky, 1968; Sandelowsky, 1972). Two conventional ^{14}C ages
932 obtained near the base of the ca. 50cm-deep excavation were returned in stratigraphic sequence
933 as 3,260 ± 80 BP and 3,430 ± 80 BP (Sandelowsky 1972). Geometric rock paintings at the site
934 exhibit white overpainting on red, which has been interpreted elsewhere in Malawi as re-use by
935 food producers (Smith, 1995; Zubieta, 2016). Ceramics are rare but present at the site, where the
936 large deposit exhibits predominately Later Stone Age material culture. This was confirmed
937 during 2016 pilot excavations, permitted by the Departments of Antiquities and National Parks
938 and Wildlife.

940 The deposits at Fingira had not been backfilled after the 1966 excavation, and had undergone
941 extensive erosion and slumping. As the site is accessible to the public, piles of materials had
942 been pulled from the section and placed on fallen rocks. Two of those specimens comprised the
943 two adult femoral specimens reported here. Using original site plans, we identified the extent of
944 the 1966 excavations prior to slumping, including the relative positions of originally-recovered
945 human remains. A large part of this area had been covered with a central termite mound, and it
946 was within this that the partial skeleton of a neonate was recovered. This comprised the subadult
947 sample from Fingira. Direct ^{14}C AMS ages on these three specimens are reported in **Table S2**,
948 and show that LSA people were using the site as a cemetery for at least 3,700 years, from 6177-
949 5923 calBP (5290±25 BP, UCIAMS-186347) to 2676-2330 calBP [2676-2343 calBP (2425±20
950 BP, PSUAMS-1734), 2483-2330 calBP (2400±20 BP, PSUAMS-1881)]. The two adult
951 specimens therefore significantly pre-date the earliest known age in Malawi for the Bantu
952 expansion, which derives from the Kasitu Valley (containing Hora Mountain) at 1,750 ± 60 BP
953 (Robinson, 1982). If the Bantu expansion into Malawi began more than ~700 years before what
954 is currently known, then the neonate from Fingira could potentially overlap in time with it.
955 However, the overall antiquity of these specimens makes admixture attributable to the Bantu
956 expansion highly unlikely in light of current knowledge.

958 **Chencherere II (n=2).** Mwana wa Chencherere II is a painted rock shelter set in a granitic
959 inselberg at ca. 1,700 meters above sea level in the Chongoni Rock Art Region (Smith, 1995). Its
960 relatively high altitude results in cool, moist year-round conditions between 10 – 20° C. It was
961 excavated by Clark in 1972 (Clark, 1972; Clark and Clerk, 1973), and the faunal assemblage
962 published in detail by Crader (Crader, 1984), who also reports one adult male burial and the

fragmentary remains of seven other individuals (adult and subadult). The site contained large lithic and faunal assemblages, bone and shell tools and ornaments, and increasing abundances of pottery and other evidence of interaction with food-producers over time. The youngest reported date from the site is a conventional ^{14}C age on charcoal from the top of Level 3, at 800 ± 50 BP. The oldest date is from near the base of Level 4, and is $2,480 \pm 200$ BP.

All human remains are reported from Levels 2, 3, and 4 – with most in Level 3. The original description of human remains was by Brothwell, and Crader (1984b:Appendix 2) later listed several more individuals that had been discovered within the faunal material. All human remains recovered from Chencherere II were thought to be held at the Natural History Museum in London, but during a 2016 visit to the Malawi National Repository in Nguludi (near Blantyre), six additional specimens were identified. Five of these derived from a cluster in square A3 (**Fig. S1**): right ilium, left femur (sampled), and 3 partial ribs. These were inferred to belong to the same individual, a subadult aged 3-5. An upper right incisor (sampled) was labeled as deriving from square E2, and therefore although of similar ontogenetic age it was deemed likely to be from a different individual.

Although the genetic analysis confirms that these are two different individuals, insufficient material remained from the incisor root of the second individual for a direct age. The first individual returned a direct ^{14}C AMS age of 5293-4979 calBP (4525±25 BP, UCIAMS-186348). As with Fingira, the direct AMS ages on the human remains are substantially older than the conventional charcoal ages suggested, indicating either intrusive charcoal or problems with the original dates.

St. Helena Bay, South Africa (n=1).   In June 2010, an intact skeleton was excavated by Andrew B Smith along the southwest coastal region of South Africa at St. Helena Bay. The skeleton is stored in the Department of Human Biology at the University of Cape Town under the accession number UCT-606. The body had been placed on an impermeable consolidated dune surface, on its right side in a fully flexed position. The bones originate from a single male who stood no more than 1.5m in height. Dental wear and significant areas of osteoarthritis suggest that he was at least 50 years of age at time of death. Lack of any evidence of tooth decay and excessive occlusal wear suggests a diet typical of hunter-gatherer subsistence. The presence of abnormal bone growths in the right auditory meatus (ear canal opening) caused a condition known as "surfer's ear" (auditory exostosis) and provides evidence that this individual most likely spent considerable time in the cold coastal waters sourcing food No obvious cause of death was evident. The results of carbon-14 to stable carbon-13 isotope ratio analysis of a rib provided a date of 2241-1965 calBP (2330±25 BP, UGAMS-7255) years before present (UGAMS-7255) with a d13C value of -14.6%. Although the minimum date falls right on the edge of the arrival of pastoralism in the Western Cape, anatomical and archaeological analysis of this skeleton and the associated burial site clearly defines this individual as an indigenous Southern African, predating

1001 pastoral arrival into the region. This individual has previously been sampled for mtDNA analysis
1002 (Morris et al., 2014).

1003

1004 **Faraoskop, South Africa (n=1).** The site of Faraoskop is a rock shelter about 30 kilometres
1005 inland from Elands Bay on the west coast of the Western Cape Province of South Africa. The
1006 shelter is on the highest ridge of a small hill at an altitude of 300 metres about the surrounding
1007 plain. Seven skeletons were collected by a local farmer in 1984, but the site was subsequently
1008 excavated under controlled conditions in 1987 and 1988 (Manhire, 1993) and five more
1009 skeletons were collected including the one referred to here as UCT-386. The shelter has no rock
1010 paintings but there is a rich assemblage of Later Stone Age artefacts including stone tools,
1011 ostrich eggshell beads, worked marine shell, leather and twine (Manhire, 1993). No pottery is
1012 associated with the site. Six skeletons have been C14 dated with resulting dates ranging from
1013 2300 years BP to 1900 years BP (Manhire, 1993; Sealy et al., 1992) but all the dates overlap at
1014 the 2nd standard deviation.   UCT-386 is the skeleton of a female about 40-50 years at death. A
1015 bone sample provided a date of 2017-1748 calBP (2000±50 BP, Pta-5283) with a d13C value of
1016 -16.8% (Manhire, 1993). A recent re-evaluation of the site indicates the possibility that all of the
1017 individuals died in one event. Not only do all of the dates overlap, but the excavation data
1018 suggest no separate grave shafts and at least two of individuals show signs of perimortem injury
1019 and violent death (Parkington and Dlamini, 2015). The Faraoskop human skeletons are stored in
1020 the Department of Human Biology at the University of Cape Town.

1021

1022 **Kasteelberg, South Africa (n=1).**   The site of Kasteelberg is on a granite hill on the
1023 Vredenberg Peninsula about 4 kilometres from the coastal village of Paternoster, approximately
1024 150 kilometres north of Cape Town (Smith, 1992a).   There are several sites on the hill including
1025 a small rockshelter, but the human skeleton was excavated from square 22 extension at KBB on
1026 the eastern side of the base of the hill. UCT-437 is the nearly complete skeleton of a child of
1027 about 4 years of age excavated by Lita Webley in 1986. The body was in a shallow pit about 1.5
1028 metres deep and with no stone cover. The specimen has a date of 1282-1069 calBP (1310±50
1029 BP, Pta-4373). There were no grave goods in association with the skeleton. The earliest sites on
1030 top of the hill provide evidence of domestic sheep at around 2100 years ago. The KBB site at the
1031 base of the hill is dated to the latter part of the occupation but present the first appearance of
1032 cattle in the region (Smith, 1992b). Overall, the Later Stone Age occupation of the Kasteelberg
1033 indicates the presence of herder-foragers who practiced seasonal economic systems, sometimes
1034 relying on domestic stock while at other times hunting seals (Sadr et al., 2003). The human
1035 skeleton from Kasteelberg is stored in the Department of Human Biology at the University of
1036 Cape Town.

1037

1038 # METHOD DETAILS

1039

1040

Direct AMS ^{14}C Bone Dates

1041
1042 We report new direct AMS ^{14}C bone dates in this study from multiple laboratories. In general,
1043 bone samples were manually cleaned and demineralized in weak HCl and, in most cases (PSU,
1044 UCIAMS, OxA, ISGS), soaked in an alkali bath (NaOH) at room temperature to remove
1045 contaminating soil humates. Samples were then rinsed to neutrality in Nanopure H_2O and
1046 gelatinized in HCL (Longin, 1971). The resulting gelatin was lyophilized and weighed to
1047 determine percent yield as a measure of collagen preservation (% crude gelatin yield).

1048
1049 Collagen was then directly AMS ^{14}C dated (ISGS, Pta) or further purified using ultrafiltration
1050 (PSUAMS, UCIAMS, OxA, Wk, Beta) (Brown et al., 1988; Kennett et al., 2017) or a modified
1051 XAD method (Lohse et al., 2014; Stafford et al., 1991). It is standard in some laboratories
1052 (PSUAMS, UCIAMS, Wk, OxA) to use stable carbon and nitrogen isotopes as an additional
1053 quality control measure. For these samples, the %C, %N and C:N ratios were evaluated before
1054 AMS ^{14}C dating. C/N ratios for well-preserved samples fall between 2.9 and 3.6, indicating good
1055 collagen preservation (Van Klinken, 1999). Additional quality control work was carried out on
1056 the samples from Malawa using FTIR spectra (**Fig. 1E**).

1057
1058 All ^{14}C ages were δ^{13}C-corrected for mass dependent fractionation with measured $^{13}C/^{12}C$ values
1059 (Stuiver and Polach, 1977) and calibrated with OxCal version 4.3 using the southern hemisphere
1060 calibration curve (SHCal13). Given their proximity to the equator, AMS ^{14}C dates for sites in
1061 coastal Kenya and Tanzania were calibrated using OxCal v. 4.3 (Bronk Ramsey, 2009) at
1062 95.4% probability employing a mixed curve that combines the SHCal13 (Hogg et al., 2013)
1063 and IntCal13 (Reimer et al., 2013) curves at ratios of 70:30 to account for the differential effects
1064 of the intertropical convergence zone.

1065
Ancient DNA sample processing in Leipzig: St. Helena Bay sample

1066
1067
1068 *DNA extraction and library preparation*
1069 30.4 mg of bone powder was removed from the internal root canal of the tooth (SP2809) using a
1070 sterile dentistry drill in the clean room facilities of the Max Planck Institute for Evolutionary
1071 Anthropology in Leipzig, Germany. A DNA extract (E649) was prepared with a silica-based
1072 method, described in detail previously (Rohland and Hofreiter, 2007). 15 µL (15% of the total
1073 volume) of the extract was converted into a single-stranded DNA library (A5354) using a
1074 modified version of the single-stranded DNA library preparation protocol (Gansauge and Meyer,
1075 2013; Korlević et al., 2015). Library positive and negative controls were carried throughout the
1076 library preparation process. Library A5354 was pre-treated with the USER enzyme, a mixture of
1077 uracil-DNA glycosylase (UDG) and endonuclease VIII, in order to remove uracils from the
1078 internal parts of ancient DNA molecules (Briggs et al., 2010; Meyer et al., 2012). The number of
1079 DNA molecules in the library (**Table S1**) was determined by digital droplet PCR (Bio-Rad
1080 QX200), using 1 µL of a 5,000-fold dilution of the library in EBT buffer (10 mM Tris-HCl pH

1081   8.0, 0.05% Tween 20) as template in an Eva Green assay (Bio-Rad) with primers IS7 and IS8
1082   (Meyer and Kircher, 2010). The library was amplified into the PCR plateau in a 100 µL reaction
1083   with AccuPrime Pfx DNA polymerase (Dabney and Meyer, 2012) using a pair of primers with
1084   two unique index sequences according to a double indexing scheme described in detail elsewhere
1085   (Kircher et al., 2011). 50 µL of amplification products were purified using the MinElute PCR
1086   Purification Kit (Qiagen) and eluted in 30 µL TE buffer (10 mM Tris-HCl pH 8.0, 1 mM
1087   EDTA). The DNA concentration of the indexed library (A5369) was determined using a
1088   NanoDrop 1000 Spectrophotometer.
1089
1090   *Size fractionation for shotgun sequencing*
1091   From the amplified library A5369, 1 µl was taken as template for a second round of
1092   amplification in a 100 µl PCR reaction using primers IS5 and IS6 (Meyer and Kircher, 2010)
1093   with Herculase II Fusion DNA polymerase (Agilent Technologies) under the conditions
1094   described in detail previously (Dabney and Meyer, 2012). The concentration of the final library
1095   was determined on a Bioanalyzer 2100 instrument (Agilent Technologies) using a DNA-1000
1096   chip. Library A5369 was pooled and sequenced with libraries from another experiment,
1097   occupying 33% of one lane of a flow cell on the Illumina HiSeq 2500 platform in rapid mode,
1098   using a double index configuration (2 x 76 + 2 x 76) (Kircher et al., 2011).
1099
1100   For a more effective use of sequencing capacity, 10 µL of the library A5369 was additionally
1101   separated on a Criterion Precast polyacrylamide gel (10% TBE, BioRad), and the fraction of
1102   library molecules with insert sizes larger than 40 bp was gel-excised as described in detail
1103   previously (Meyer et al., 2012). Gel-excised library molecules were subjected to a second round
1104   of amplification and the concentration of the final library (A5386) was determined using a DNA
1105   1000 chip on the Bioanalyzer 2100.
1106
1107   **Ancient DNA sample processing in Tübingen: Faraoskop and Kasteelberg samples**
1108
1109   *Sampling and extraction*
1110   Sampling took place in the clean room facilities of the Institute for Archaeological Sciences at
1111   the University of Tübingen. Both samples were irradiated with UV light for 10 minutes from all
1112   sides to remove surface contamination. The tooth from the South African forager from
1113   Faraoskop (UCT386) was sawed apart transversally at the border of crown and root, and dentine
1114   from inside the crown was sampled using a sterile dentistry drill, resulting in 56 mg dentine
1115   powder. For the femur fragment from the South African pastoralist from Kasteelberg (UCT437),
1116   the surface layer from the sampling area was removed with a dentistry drill prior to obtaining
1117   four aliquots between 51 and 80 mg of bone powder from the inside of the bone by drilling.
1118
1119   Extraction was performed following a protocol optimized for the recovery of small ancient DNA
1120   molecules (Dabney et al., 2013), resulting in 100µl of DNA extract per sample. Three of the

1121   bone powder aliquots from UCT437 underwent a 10 minute pre-digestion step after which the
1122   extraction buffer was removed (pre-digest) and replaced by fresh extraction buffer followed by
1123   over-night digestion (ON-digest), the powder from UCT386 and one aliquot of UCT437 were
1124   extracted without the pre-digestion step (full-digest). All eight resulting extracts were taken
1125   along for further library preparation. Negative controls were included in the extraction and taken
1126   along for all further processing steps.
1127
1128   *Screening*
1129   Two double-indexed libraries were produced from an aliquot of 20μl of the full-digest
1130   extractions of UCT386 and UCT437 (Kircher et al., 2011; Meyer and Kircher, 2010). Positive
1131   and negative controls were included in library preparation and taken along into sequencing.
1132   Libraries were enriched for human mitochondrial DNA (Maricic et al., 2010) and both enriched
1133   and shotgun libraries were sequenced on a HiSeq2500 with 2x101+8 cycles. Processing by the
1134   EAGER pipeline (Peltzer et al., 2016) and *schmutzi* (Renaud et al., 2015) resulted in an
1135   endogenous DNA content of 39% and 8% and an estimated mitochondrial contamination of 0-
1136   2% and 1-3% for UCT386 and UCT437, respectively.
1137
1138   *Library preparation for shotgun sequencing*
1139   For UCT386 four more libraries were produced from an aliquot of 20ul of full-digest extract
1140   each, including a DNA repair step with UDG and endonuclease VIII to remove deaminated bases
1141   (Briggs and Heyn, 2012). For UCT437, six additional UDG-treated libraries were produced from
1142   20ul each of extract from the three pre-digest and the three ON-digest extracts. After indexing
1143   PCR (Kircher et al., 2011), aliquots of the UDG-treated libraries were size selected on a PAGE
1144   gel to remove fragments of sizes below 35 and above 80 bp (Meyer et al., 2012).
1145
1146   **Ancient DNA sample processing in Dublin: Malawi samples**
1147   Sampling took place in ancient DNA-dedicated clean room facilities at University College
1148   Dublin. The petrous part of the temporal bone was selected for analysis from each individual
1149   (n=2). Each complete petrous was UV irradiated for 10 minutes on each side prior to processing
1150   to reduce surface contamination. Any remaining sediment was removed using a Renfert Basic
1151   Classic Sandblaster (Renfert GmbH) at low power. Bone powder was retrieved from the petrous
1152   of UCT242 (Hora 1) by drilling a small hole on the superior surface of the petrous with a 4.8mm
1153   High Speed Cutter (Dremel) until the cochlea was accessible. Bone powder was then collected
1154   directly from the cochlea using a 3.2mm Tungsten Carbide Cutter (Dremel). The petrous from
1155   UCT243 (Hora 2) was cut from anterior to posterior using a Dremel drill at a location that
1156   transected the cochlea. The powder was collected directly from the cochlea using a 3.2mm
1157   Tungsten Carbide Cutter (Dremel). Powder aliquots from both samples were then UV irradiated
1158   for 5 minutes and placed in 2.0mL Eppendorf tubes.
1159

Ancient DNA sample processing in Boston: Tanzania samples, Kenya samples, and Malawi sample powder

Sampling and DNA extraction

In a dedicated ancient DNA facility at Harvard Medical School, samples were UV-irradiated for 10 minutes in a UVP crosslinker. At the chosen part of each sample (root for the tooth and compact part for bones) the surface was removed with a sanding disk. About 75 mg (±10 mg) of fine powder was obtained by drilling into the physically cleaned part with a sterile dentist drill bit and collected for DNA extraction (**Table S1**). In the case of KC-10-1011(4353) (I0589), additional bone powder was collected for a second DNA extraction attempt. The seven Malawi_Holocene samples arrived in the Boston laboratory as powders prepared in Dublin, Ireland. Starting from the sample powder, we followed the Dabney et al. 2013 extraction protocol for all samples, but replaced the funnel/MinElute assemblage with the pre-assembled Roche columns (Korlević et al., 2015) and eluted two times in 45μl for a total of 90μl DNA extract.

Initial library preparation

One initial barcoded library (L1) was prepared from 30μl DNA extracts for all but two samples (Hora1 and Hora2), which were discolored and we reduced the volume to 3μl (reducing the volume seems to mitigate library preparation inhibition, which we often find to be associated with discolored DNA extracts) following protocols published previously (Rohland et al., 2015) (**Table S1**). For three samples (I0589, I0595, I1048) the initial library was UDG-treated (Briggs et al., 2010) following a modification from Rohland et al. 2015 (partial UDG treatment) that is tailored to inefficiently remove terminal Uracils therefore leaving the aDNA authenticity signal in the terminal bases while efficiently removing miscoding damage within the molecules. The initial libraries for the other samples (I2298, I2966, I2967) were not UDG-treated. The last step of the library preparation, the amplification with universal primers, was set up in the cleanroom, but the thermal cycling happened in another laboratory physically separated from the cleanroom. The final products of our barcoded libraries cannot be sequenced right away; an additional PCR step is needed to finalize the adapter sites. This is advantageous in that we can incorporate a second set of barcodes through dual indexing to differentiate two or more experiments done with the same barcoded library within the same sequencing run (see below).

Screening

Each initial library underwent screening that consisted, first, of shallow shotgun sequencing after an indexing PCR (that adds dual indices to each library, (Kircher et al., 2011), and second, target capture enrichment for mitochondrial DNA and a varying number of nuclear loci to assess mitochondrial haplogroup, mitochondrial contamination, aDNA authenticity and nuclear complexity (Meyer et al., 2014; Rohland et al., 2015). This experiment is finished by adding unique index combinations to each captured library, which is then subsequently pooled with the

14

1200 shotgun indexing PCR product for sequencing. Sequencing was done on an Illumina NextSeq500
1201 with 2x76cycles + 2x7cycles.

1202

1203 We demultiplexed reads to be sample-specific requiring that that the 7bp P5 and P7 indices
1204 matched (allowing one mismatch). Sample identification was further ensured by requiring that
1205 additional 7bp internal barcodes matched, again allowing one mismatch. We merged with a
1206 modified form of *SeqPrep* (github.com/jstjohn/SeqPrep) (the modification ensures that the
1207 highest quality base is retained in the overlap region), requiring at least 15bp overlap between
1208 forward and reverse reads, allowing one mismatch, retaining only reads of length greater than or
1209 equal to 30 base pairs, generating single ended reads.

1210

1211 Reads were then aligned using the *samse* algorithm of BWA (version 0.6.1) (Li and Durbin,
1212 2009) using ancient parameters to allow an increased mismatch rate, and to disable seeding ("-n
1213 0.01 -0 2 -l 16500"). Multiple sequencing runs were run to increase coverage, and merged
1214 together. Duplicates were then removed by identifying clusters of reads which have the same
1215 start and stop position, and the same mapped orientation. The highest base quality representative
1216 of each set is used to represent the cluster. The mix of mitochondrial and nuclear loci
1217 necessitates two different references for the alignment process: for mitochondrial analysis, we
1218 use the RSRS (Behar et al., 2012) mitochondrial genome; for nuclear analysis we use the
1219 hg19/GRCh37, 1000 Genomes release reference genome.

1220

1221 *Additional libraries and processing*
1222 To collect more nuclear data for a subset of the samples, we prepared additional barcoded
1223 libraries (I0589, I0595, I1048, I3726) without UDG-treatment from existing DNA extracts (L2-
1224 L5). For sample I0589 we collected more bone powder than necessary for one extraction, and
1225 therefore prepared three additional libraries from a newly prepared DNA extract (E2). Four
1226 additional libraries for one sample (I3726) were prepared on an Agilent Bravo Workstation using
1227 an automated protocol based on the partial UDG protocol that replaced the MinElute cleanups
1228 with magnetic bead cleanups.

1229

1230 All libraries underwent the same procedure as outlined, screening (see above) and nuclear target
1231 enrichment (see below), with the exception that up to 4 libraries from the same sample were
1232 pooled in equimolar concentrations before screening and nuclear target capture (**Table S1**).
1233 Preprocessing and alignment for nuclear data used the same procedure as performed in
1234 screening, without requiring the mitochondrial alignments.

1235

1236 **Shotgun genome sequencing**
1237 Shotgun sequencing of the ancient South African from St. Helena Bay was performed at the Max
1238 Planck Institute in Leipzig, Germany, on four lanes of the Illumina HiSeq 2500 platform in rapid
1239 mode, using a double index configuration (2 x 76 + 2 x 76) (Kircher et al., 2011). An indexed

1240 ΦX174 control library was spiked in prior to sequencing. Base calling was done with the
1241 machine-learning algorithm freeIBIS (Renaud et al., 2013). Overlapping pair-end reads were
1242 merged (Kircher, 2012) and mapped to the human reference genome (hg19/GRCh37, 1000
1243 Genomes release) using the Burrows-Wheeler Aligner (BWA) (Li and Durbin, 2009). BWA
1244 parameters were adjusted for ancient DNA sequences ("-n 0.01 –o 2 –l 16500"), to allow for
1245 more mismatches and indels and to turn off the seeding (Meyer et al., 2012). A total of
1246 64,128,220 raw sequences were obtained from the first shotgun sequencing of the library A5369.
1247 Another 800,205,849 raw sequences were generated from the size-selected library using four
1248 lanes of the Illumina HiSeq 2500. Only mapped sequences longer than 35 bp were retained and
1249 duplicates removed (bam-rmdup; https://github.com/udo-stenzel/biohazard), leaving 9,880,908
1250 sequences from the first shotgun run and 52,551,348 sequences from sequencing the size-
1251 fractionated library. Duplication rates were 1.02 and 1.06, respectively, indicating that both
1252 libraries were not sequenced to exhaustion. The proportion of sequences ≥35bp that mapped to
1253 the human reference genome was ~28%.

1255 Shotgun sequencing of the ancient South African from St. Helena Bay (UCT386) and the ancient
1256 South African from Kasteelberg (UCT437) was performed at the University of Tuebingen using
1257 two lanes of an Illumina HiSeq2500 instrument for 2x101+8 cycles (UCT386 non-UDG-treated
1258 library), on 50% of two lanes of an Illumina NextSeq500 instrument for 2x151+8 cycles
1259 (UCT437 non-UDG-treated library), on a complete run of an Illumina HiSeq2500 instrument for
1260 2x125+8 cycles (four UCT386 UDG-treated and size-selected libraries and six UCT437 UDG-
1261 treated and size-selected libraries), and on 5.5 lanes of an Illumina HiSeq2500 instrument for
1262 2x125+8 cycles (four UCT386 UDG-treated libraries without size-selection and three UCT437
1263 UDG-treated full-digest libraries without size-selection). The samples were processed using the
1264 EAGER pipeline (Peltzer et al., 2016), clipping adapters and merging reads subsequently with an
1265 overlap of 10bp. Resulting reads were then mapped within the pipeline against the human
1266 reference genome GrCh37 and using BWA 0.7.5 (Li and Durbin, 2009) for further downstream
1267 analysis.

1269 Shotgun sequencing of the Malawi_Hora_8100BP samples was performed at Harvard Medical
1270 School using an Illumina NextSeq500 instrument. Preprocessing and alignment used the same
1271 procedure as performed in screening.

1273 **In-solution nuclear target enrichment**
1274 After libraries passed screening QC (that is, there was evidence of authentic ancient DNA), we
1275 performed nuclear target enrichment of the short (but barcoded) libraries following (Fu et al.,
1276 2015) aiming to enrich for about 1.24M SNPs in total (Fu et al., 2015; Haak et al., 2015;
1277 Mathieson et al., 2015) using a semi-automated approach on a Perkin Elmer Evolution P3 liquid
1278 handler. For two libraries (S0589.E1.L1 and S2595.E1.L1) the desired 1.24M targeted SNPs
1279 were captured in two independent reactions by enriching, first, for about 0.39M SNPs, and

1280 second, for 0.84M SNPs. The other first libraries (L1) were enriched in one single reaction
1281 (1240k). After indexing PCRs with dual indices and equimolar pooling sequencing was
1282 performed on an Illumina NextSeq500 with 2x76cycles + 2x7cycles. Preprocessing and
1283 alignment used the same procedure as performed in screening, without requiring the
1284 mitochondrial alignments.
1285

1286 **Genotyping and initial processing of 34 present-day individuals from Malawi**
1287 We newly report data from 34 present-day individuals from Malawi, genotyped on the
1288 Affymetrix Human Origins SNP array (Patterson et al., 2012). Quality control of the data prior to
1289 merging involved screening for outlier individuals, excess missingness, as well as deviations
1290 from Hardy-Weinberg equilibrium, and was performed in a manner similar to what has
1291 previously been described (Lazaridis et al., 2014).
1292

1293 **Data processing and preparation**
1294 We extracted genotypes from the ancient genomes by drawing a random sequence read at each
1295 position, ignoring the first and last 3 bp of every read and any read containing insertions or
1296 deletions in their alignment to the human reference genome. If the randomly drawn haploid
1297 genotype of an ancient individual did not match either of the alleles of the biallelic SNP in the
1298 reference panel, we set the genotype of the ancient individual as missing.
1299

1300 We added these pseudohaploid genotypes to 17 million dinucleotide transversion SNPs
1301 identified between present-day genomes from the Simons Genome Diversity Panel (which
1302 includes human-fixed differences to chimpanzee). We also added the ancient genotypes to 550
1303 individuals from 56 African populations genotyped on the Affymetrix Human Origins array
1304 (Lazaridis et al., 2014; Patterson et al., 2012; Pickrell et al., 2012; Pickrell et al., 2014; Skoglund
1305 et al., 2015). To all these datasets we added diploid genotypes from two archaic human genomes
1306 – a Neanderthal and a Denisovan (Meyer et al., 2012; Prufer et al., 2014). The populations
1307 shown in **Fig. S2** are individuals from the Affymetrix Human Origins array, when we in the text
1308 refer to Khomani_San, they are the individuals from the Simons Genome Diversity panel and so
1309 are not shown in the legend in **Fig. S2**.
1310

1311 # QUANTIFICATION AND STATISTICAL ANALYSIS
1312

1313 Population genetic approaches that quantify shared genetic drift, such as *f*-statistics and
1314 admixture graph fitting, are maximally robust when ascertainment of SNPs are performed in an
1315 outgroup (Patterson et al., 2012; Wang and Nielsen, 2012), such that there is no bias in allele
1316 frequencies between the analyzed populations and the polymorphism that appeared by mutation
1317 in the ancestral population of all analyzed populations. Whereas the Human Origins Array
1318 comprises 13 different panels ascertained in modern humans (Patterson et al., 2012), none of
1319 these can be regarded as outgroup-ascertained for the purpose of African populations. To obtain

1320 an outgroup-ascertained set of SNPs for African populations, we identified 814,242 transversion
1321 SNPs polymorphic between the archaic Denisovan (Meyer et al., 2012) and Neanderthal (Prufer
1322 et al., 2014) genomes (together labeled as 'Archaic' here). Since the ancestors of Denisovans and
1323 Neanderthals are consistent with having diverged from sub-Saharan lineages before those
1324 lineages separated from each other (Green et al., 2010; Mallick et al., 2016; Meyer et al., 2012;
1325 Prufer et al., 2014; Reich et al., 2010), the ascertained SNPs that were also present as
1326 polymorphisms in sub-Saharan Africa were highly likely to have been polymorphic before the
1327 African populations diversified. We extracted these positions from the 1000 genomes project
1328 MSL (Mende from Sierra Leone; 81 unrelated individuals), and YRI (Yoruba from Ibadan,
1329 Nigeria; 107 unrelated individuals), to increase power. These 1000 genomes project sequences
1330 were processed by sampling a random sequence at each position as for the ancient data, setting
1331 the genotype as missing if it did not match either of the two alleles in the ascertained SNP set.
1332
1333 **Principal component analysis and ADMIXTURE clustering analyses**
1334 We used *smartpca* (Patterson et al., 2006) to compute principal components using all
1335 transversion and transitions SNPs, and the present-day populations shown **in Fig. 1** and **Fig. S2**.
1336 We projected the ancient individuals the option *lsqproject: YES*, on eigenvectors computed using
1337 the present-day populations. To deal with the confounder factor of recent admixture with western
1338 Eurasian-related populations on the PCA, we removed all northern Africans, eastern African
1339 Cushitic speakers, Nama, and 8 Khomani individuals that had 5% or more cluster membership in
1340 the shared with Europeans in an ADMIXTURE analysis.
1341
1342 For our main ADMIXTURE clustering analysis (Alexander et al., 2009) (**Fig. 1B; Fig. S3**) we
1343 excluded 166,439 SNPs that were in a CpG context and thus retain postmortem damage, and
1344 used 431,134 SNPs and 208 selected ancient- and present-day individuals genotyped on the
1345 Human Origins Array. In **Fig. 1B**, we show only eight representative individuals for the non-
1346 African Japanese (originally n=29), and Sardinian (n=27) populations. For the authentication
1347 analysis investigating evidence of contamination, we used *PMDtools* (Skoglund et al., 2014a) to
1348 isolate sequences from each sample that had clear evidence of contamination according the
1349 postmortem damage score (PMD score > 3, using only based with phred-scaled quality of at least
1350 30 to compute the score), and performed clustering analysis only on 111,208 transversion SNPs
1351 (**Fig. S3**). The exclusion of transition SNPs is due to the PMD score approach enriching for C>T
1352 and G>A substitutions indicative of ancient DNA.
1353
1354 **Symmetry statistics and admixture tests**
1355 *D*-statistics*, f₄*-statistics, and *f₃*-statistics (Patterson et al., 2012; Reich et al., 2009) were
1356 computed with *POPSTATS* (Skoglund et al., 2015). f_4-statistics test whether two pairs of
1357 populations are symmetric with respect to one another, and quantify any asymmetry arising from
1358 admixture. More specifically, if p_1, p_2, p_3, and p_4 are the derived allele frequencies at a biallelic
1359 SNP locus in population 1, population 2, population 3, and population 4, we can estimate f_4 as a

sum over all SNP loci $f_4 = \sum(p_1 - p_2)(p_3 - p_4)$ (Reich et al., 2009). D-statistics (Green et al., 2010), which are also used in this study, are a version of f_4-statistics with a denominator to normalize for heterozygosity, but in practice both statistics have similar power to detect deviations from the null model, and f_4-statistics have the additional advantage of being directly informative about admixture proportions and shared genetic drift (Patterson et al., 2012).

For the statistics in **Fig. 3B and Fig. 3C**, we used 814,242 transversion SNPs polymorphic between the archaic Denisovan (Meyer et al., 2012) and Neanderthal (Prufer et al., 2014) genomes (together labeled as 'Archaic' here). We extracted these loci from the 1000 genomes project MSL (Mende from Sierra Leone; 81 unrelated individuals), and YRI (Yoruba from Ibadan, Nigeria; 107 unrelated individuals). The statistics in **Fig. S2C** used either complete genome sequences from the Simons Genome Diversity Panel (Mallick et al., 2016), or panel 5 of the Human Origins Array, which comprises 119,413 SNPs that were originally ascertained as polymorphic positions in a single Yoruba individual. We used this set to test whether the ancient individuals were closer to one of two San groups because some of the SNPs on the Human Origins array were ascertained in one of the San groups, potentially affecting the statistic. In **Table S5** we report multiple D-statistics for different configurations of populations using transversion SNPs in complete genomes from the Simons Genome Diversity Project and ancient shotgun sequences.

Y-chromosomal and mitochondrial haplogroups

For Y-chromosome haplogroup calling, we filtered reads with mapping quality <30 and bases with base quality <30, and determined the most derived mutation for each sample using the tree of the International Society of Genetic Genealogy (http://www.isogg.org) version 11.110 (21 April 2016). We also used *Yfitter* (Jostins et al., 2014) to confirm the haplogroups of the male Faraoskop and St. Helena Bay individuals using the entire shotgun sequence data, with identical haplogroup calls as the other approach.

For mitochondrial DNA haplogroups, we used *Haplogrep2* (Weissensteiner et al., 2016) with Phylotree 17 (Van Oven and Kayser, 2009), restricting to sites with base quality 10, and depth 1. These relatively permissive thresholds were used to maximize coverage on the mitogenome. For sample I2966, which was found to have mitochondrial contamination, we first restricted to damaged reads using a PMD score threshold of 3 (Skoglund et al., 2014a).

Ancestry model and estimates with *qpAdm*

Clustering analyses and PCA are sensitive to genetic drift, such as the genetic drift that occurs in a population after the time ancient individuals lived (Skoglund et al., 2014b), and may thus not provide an accurate view of shared ancestry between ancient and present-day individuals. We employed a framework for estimating ancestry proportions that is based on f_4-symmetry statistics, taking advantage of the fact that f_4-statistics are proportional to admixture proportions

and genetic drift. In the well-documented case of Neanderthal admixture into non-African populations, for example, the statistic f_4(chimpanzee, Neanderthal; African, non-African) is proportional to αx, where α is the proportion of Neanderthal-related ancestry (approximately 2%) and x is proportional to the amount of genetic drift that occurred from the divergence of Neanderthal ancestors and African ancestors, to the divergence of the sampled Neanderthal genome and the Neanderthal population that admixed with non-Africans. By analyzing many such f_4-statistics, (Lazaridis et al., 2014) and (Haak et al., 2015) showed that it is possible to estimate admixture proportions for a target population without detailed assumptions about population phylogeny, and also to perform hypothesis tests for whether a particular mixture model fits the data (Reich et al., 2012) and to estimate standard errors for admixture proportions with a weighted block jackknife procedure over large segments over the genome (in this study 5 cM). This has been implemented as the *qpAdm* algorithm in the *ADMIXTOOLS* package and requires the proposal of a set of source populations as well as a set of outgroups that are proposed to not share drift with the target population more recently than the source populations. In other words, appropriate source populations do not need to be the true source populations but instead, need only be more closely related to the true source populations than they are to any of the outgroups. Violations of these assumptions can be detected as an increase in rank in the matrix of f_4-statistics computed (Reich et al., 2012). We also analyze a statistic using fitted allele frequencies predicted using the estimated mixture proportions, f_4(Target population, Fitted Target population; Mbuti, Test). Deviations of this statistic from 0 are informative about whether some outgroups have an excess, or deficiency, of shared drift with the Target population under the fitted mixture proportions.

Here, we used a model with 19 populations (Mbuti, Dinka, Mende, South_Africa_2000BP, Tanzania_Luxmanda_3100BP, Ethiopia_4500BP, Levant_Neolithic (PPNB), Anatolia_Neolithic, Iran_Neolithic, Denisova, Loschbour, Ust_Ishim, Georgian, Iranian, Greek, Punjabi, Orcadian, Ami, and Mixe), using previously published complete genomes (Fu et al., 2014; Lazaridis et al., 2014; Mallick et al., 2016; Meyer et al., 2012) and ancient DNA data enriched using the 1240k SNP set (Lazaridis et al., 2016; Mathieson et al., 2015) to maximize the power to infer admixture proportions for the ancient African populations. These populations, and in particular the ones from Africa, were chosen to capture major strands of ancestry and extremes in population differentiation found in sub-Saharan Africa (**Fig. 1**)

We then successively moved a set of candidate source populations (Mende, Dinka, Mbuti, South_Africa_2000BP, Tanzania_Luxmanda_3100BP, Ethiopia_4500BP, PPNB Anatolia_Neolithic, Iran_Neolithic) from the outgroup set to test if they fit as sources in admixture models. Using these 9 candidate sources populations, for each target population we thus tested 9 one-source ancestry models, $\binom{9}{2} = 36$ two-source admixture models, and $\binom{9}{3} = 84$ three source admixture models, for a total of 9+36+84=129 models. In **Fig. S2** and **Table S3**, we show admixture proportions for the model with the lowest chi-square score (or highest p-value),

1440 if that model had a p-value >0.01. If a one-source model did not fulfill this criterion, we
1441 considered two-source models, and then subsequently three-source models if no two-source
1442 model fulfilled the criteria.
1443
1444 We successfully obtained mixture models for 55 Target populations, comprising both ancient
1445 populations (we excluded Malawi_Chencherere_5200BP due to low SNP coverage) and
1446 populations genotyped on the Affymetrix Human Origins array, all shown in **Table S3**. In one
1447 analysis, Tanzania_Luxmanda_3100BP was also used as a target population, and in these
1448 analyses it was dropped from the outgroup set. We highlight some notable mixture models
1449 inferred here:

1450 • Kenya_400BP, Tanzania_Pemba_1400BP and Hadza1 are all fitted as having ~100%
1451 Ethiopia_4500BP-related ancestry. The other group of Hadza samples are fitted as having
1452 19% ± 8% Dinka-related ancestry (the remainder being Ethiopia_4500BP-related).

1453 • Tanzania_Pemba_600BP, Malawi_Chewa, Malawi_Ngoni, Malawi_Tumbuka,
1454 Malawi_Yao, Yoruba, Esan, Gambian, Luo, BantuKenya, BantuSA_Ovambo, Himba,
1455 Wambo, BantuSA_Herero are all fitted as consistent with having ~100% Mende-related
1456 western African-related ancestry. The Mandenka, from the western African coast, are
1457 fitted as having 2.8% ± 0.6% Levant Neolithic-related ancestry (PPNB).

1458 • The Luhya, an eastern Bantu-speaking group, are fitted as having 40% ± 6% Dinka-
1459 related ancestry, with the remainder being western African Mende-related ancestry.

1460 • The Biaka, a western rainforest hunter-gatherer Pygmy group in Cameroon, is fitted as
1461 having 72% ± 2% Mbuti-related ancestry (the Mbuti are an eastern rainforest hunter-
1462 gatherer Pygmy group), with the remainder being western African Mende-related
1463 ancestry.

1464 • The minimum indigenous southern African ancestry observed in Khoe-groups and Bantu-
1465 speakers in southern Africa is ~8% ± 2% in the Damara, and the remainder is western
1466 African-related.

1467 • The maximum indigenous southern African ancestry observed in the present-day
1468 populations is the 91% ± 1% inferred for the Ju_hoan_North, with the remainder being
1469 related to Tanzania_Luxmanda_3100BP.

1470 • Some populations in northern and eastern Africa are fitted as having large proportions of
1471 Tanzania_Luxmanda_3100BP related ancestry. This includes the Maasai (49% ± 2%)
1472 and Datog (66% ± 3%) who have ancestry also related to the Dinka; the Kikuyu (63% ±
1473 2%) who also have ancestry related to the Mende; and finally, the Afar (79% ± 3%) and
1474 Somali (62% ± 6%) who have large amounts of inferred Tanzania_Luxmanda_3100BP-
1475 related ancestry in addition to ancestry related to the Iran Neolithic.
1476

1477 **Maximum likelihood tree model**
1478 We used the four ancient African shotgun genomes together with complete genomes from
1479 African populations in the Simons Genome Diversity project (Mallick et al., 2016), excluding

1480 populations with evidence of asymmetrical allele sharing with non-Africans indicative of gene
1481 flow (**Table S5**), to reconstruct a maximum likelihood tree using Treemix v1.12 (Pickrell and
1482 Pritchard, 2012). We performed 100 bootstrap replicates to assess the uncertainty of the fitted
1483 model. While this tree is not an adequate representation of human population history in Africa,
1484 we found 100% bootstrap support for the Ethiopian_4500BP Mota being most closely related to
1485 the ancestral population of all non-Africans (**Fig. 3A**).
1486
1487 **Testing a tree-like model of African population history**
1488 The maximum-likelihood tree based on allele frequency covariance between the ancient African
1489 genomes and complete genomes from the SGDP panel (**Fig. 3A**) recapitulates many aspects of
1490 previous analyses of African populations (Pickrell et al., 2012; Schlebusch et al., 2012). When
1491 African populations are forced into a tree (not allowing for mixture), southern Africans diverge
1492 first, followed by Pygmies (e.g. Mbuti), West Africans, ancient and present-day eastern Africans
1493 (Dinka, Hadza), then non-Africans.
1494
1495 To scrutinize this tree-like model in more detail, we computed all 35 *D*-statistics that include the
1496 outgroup and have an expected value of 0 if a tree-like model (chimpanzee,
1497 (South_Africa_2000BP, (Mbuti, (Mende, (Dinka, Ethiopia_4500BP))))) is true. We find that
1498 most statistics computed are inconsistent with the null model (**Fig. S4A**). Notably, we reject the
1499 hypothesis that the ancient South Africans are an outgroup to other African populations for
1500 several pairs of present-day populations (**Table S5; Fig. S4A**). For instance, with genome
1501 sequence data we find that the Ethiopian_4500BP Mota genome shares more derived alleles with
1502 the ancient South Africans than with present-day western Africans (*D*[chimpanzee,
1503 South_Africa_2000BP; Yoruba, Ethiopia_4500BP] > 0). This is also true when West Africans
1504 are compared to Mbuti (*D*[chimpanzee, South_Africa_2000BP; Yoruba, Mbuti] > 0). Similar
1505 excess of shared derived alleles is observed for the eastern Pygmies (Mbuti) compared to western
1506 Pygmies, and even when contrasting western African populations such as the Yoruba and Mende
1507 (**Fig. 3B; Fig. 3C**; **Table S5; Fig. S4**). This could be explained in two ways: 1) there has been
1508 gene flow between ancient southern Africans and a broad set of other populations that has
1509 resulted in a gradient of southern Africans relatedness, or 2) there is a gradient of ancestry in
1510 western Africans that is basal to southern Africans, causing an attraction to the outgroup
1511 (chimpanzee in this case) (see below).
1512
1513 **Testing admixture graph models of African population history**
1514 To reconstruct admixture graph models relating the histories of African populations, we used
1515 South_Africa_2000BP (South Africa), Mende (MSL; West Africa), and Ethiopia_4500BP (East
1516 Africa) to represent major lineages contributing to present-day Africans. In addition, we sought
1517 to explain one of the most surprising observations in our data, that the Mende and Yoruba West
1518 African populations are not symmetrically related to South_Africa_2000BP, and so we also
1519 included the Yoruba (YRI) in these analyses. For all admixture graph analyses, we used 814,242

1520 transversion SNPs polymorphic between the archaic Denisovan (Meyer et al., 2012) and
1521 Neanderthal (Prufer et al., 2014) genomes (together labeled as 'Archaic' here).
1522
1523 *A tree-like model does not fit the data*
1524 We first tested a strict tree-like model with no admixture edges, hypothesizing that the topology
1525 obtained from basic tree reconstructions where southern Africans are the earliest diverging
1526 lineage, and Yoruba and Mende are a clade to the exclusion of the Ethiopia_4500BP, is true
1527 (**Fig. S4B**). We found that this model is strongly rejected by the data with 19 predicted f_4-
1528 statistics deviating from the empirically observed data by |Z| >3. The most deviating f_4-statistic
1529 was f_4 (Archaic, Ethiopia_4500BP; MSL, YRI), which is predicted to be 0 in the tree-like model
1530 but is empirically observed to be f_4=0.000427, Z=9.157. Insight into the imperfect fit can be
1531 obtained by inspecting all four significant f_4-statistics that are predicted to be zero in the model:
1532 • f_4(Archaic, Ethiopia_4500BP; MSL, YRI)=0.000427, Z = 9.157
1533 • f_4(Archaic, South_Africa_2000BP; MSL, YRI)=0.000193, Z = 4.742
1534 • f_4(South_Africa_2000BP, Ethiopia_4500BP; MSL, YRI)=0.000235, Z = 4.583
1535 • f_4(Archaic, South_Africa_2000BP; Ethiopia_4500BP, MSL)=-0.000728, Z = -3.068
1536
1537 The three most significant deviations all test the (MSL,YRI) clade. The deviations thus indicate
1538 shared history either between MSL and the outgroup Archaics, or between YRI and
1539 South_Africa_2000BP/Ethiopia_4500BP. These observations could be parsimoniously explained
1540 by any of the following gene flow events
1541 • Gene flow from a basal human lineage (separating from the ancestors of all sub-Saharan
1542   Africans before their separation from each other) into the ancestors of MSL to a greater
1543   extent than YRI (there is no evidence of specifically Neanderthal/Denisovan gene flow
1544   since f_4(chimpanzee, Archaics; MSL,YRI) = -1.6 in this data)
1545 • Gene flow related to YRI into the ancestors of Ethiopia_4500BP and
1546   South_Africa_2000BP
1547 • Gene flow related to Ethiopia_4500BP into the ancestors of YRI more than MSL
1548 • Gene flow most closely related to MSL into the common ancestors of the archaic
1549   Neanderthals and Denisovans (we exclude this as implausible)
1550
1551 The fourth deviating statistic f_4(Archaic, South_Africa_2000BP; Ethiopia_4500BP, MSL) (Z =
1552 3.068) suggests that Ethiopia_4500BP and MSL are not a clade with respect to
1553 South_Africa_2000BP. This weakens the case for gene flow *into* Yoruba alone as a sufficient
1554 explanation, since the Yoruba do not enter into this statistic. Instead, either excess basal ancestry
1555 in the Mende and gene flow between South_Africa_2000BP and Ethiopia_4500BP could explain
1556 this particular statistic.
1557
1558 *Admixture models with gene flow events*

1559 We proceeded by testing models with one gene flow event positing that YRI have mixture
1560 related to either South_Africa_2000BP or Ethiopia_4500BP, or that MSL have ancestry from a
1561 basal lineage **(Fig. S4C; Fig. S4D; Fig. S4E)**. We found that neither of these fit the data, with
1562 between 4 and 13 f_4-statistic outliers with $|Z|>3$.

1563

1564 Testing admixture graphs with two admixture events, we found that a model where both YRI and
1565 MSL have ancestry from a basal African lineage had its single outlier in an f_4 statistic f_4(Archaic,
1566 Ethiopia_4500BP; Ethiopia_4500BP, YRI) that is more negative in the data than the model **(Fig.**
1567 **S4F)**. This f_4-statistic has Ethiopia_4500BP appearing twice, and can thus be rearranged to be an
1568 f_3 statistic f_3(Archaic, YRI; Ethiopia_4500BP) that is not positive enough (Z=3.157), and can
1569 thus be interpreted as the model underrepresenting the external drift in the Ethiopia_4500BP
1570 genome. We do not consider this outlier to compromise the model, as the processing of the
1571 Ethiopia_4500BP (Mota) genome that we use is pseudo-haploid (single random sequence read),
1572 and thus there is no real information about the external drift of the Mota lineage. This outlier
1573 may thus reflect a difficulty in modeling external drift for pseudo-haploid samples.

1574

1575 In addition, we found that an admixture graph where both Ethiopia_4500BP and YRI are mixed
1576 between MSL- and South_Africa_2000BP lineages does not fit the data (**Fig. S4G**), with 3
1577 outliers, and the worst being Z=-4.835. However, a model where the YRI has ~2% ancestry from
1578 a population that is mixed between South_Africa_2000BP and Ethiopia_4500BP fits the data
1579 with 2 outliers that are not too surprising after correcting for multiple hypothesis testing **(Fig.**
1580 **S4H)**. These outliers are f_4(Archaic, South_Africa_2000BP; Ethiopia_4500BP, MSL) (Z =
1581 3.068) which was also significant for the tree model without admixture, and f_4(Archaic,
1582 South_Africa_2000BP; Ethiopia_4500BP, YRI) (Z = 3.018). Both these outliers could be
1583 consistent with the presence of basal African ancestry in YRI and MSL, or unmodeled gene flow
1584 between the ancestors of South_Africa_2000BP and Ethiopia_4500BP.

1585

1586 We thereby conclude that the most parsimonious admixture graph models identified here posit
1587 either the presence of basal African ancestry in Mende and Yoruba (**Fig. 3D; Fig. S4F**), or
1588 alternatively admixture from a source related to both South_Africa_2000BP and
1589 Ethiopia_4500BP in the Yoruba but not the Mende (with evidence also for gene flow between
1590 South_Africa_2000BP and Ethiopia) (**Fig. 3F; Fig. S4G**).

1591

1592 **Automated grafting of populations onto a skeleton admixture graph**
1593 Using the admixture graph model in which the Yoruba and Mende both carry ancestry from a
1594 basal western African population, we automatically added additional populations to each
1595 possible node in the graph. We evaluated the fit in terms of the number and deviation of outlying
1596 f_4-statistics, as well as whether the added branches had zero drift length. We show the results of
1597 this procedure in **Table S6**, with the key to the node labels used shown in **Fig. S5A**. We
1598 highlight topologies that we consider optimal fits in **Fig. S5B-5D**.

1599

1600 We find that the Malawi_Hora_8100BP can be best fitted as mixed between the lineage leading
1601 to South_Africa_2000BP, and the lineage related to Ethiopia_4500BP that also is fitted as
1602 forming part of the ancestry of YRI and MSL (**Fig. S5B**). Similarly, the Mbuti can be best fitted
1603 as mixed between the lineage related to Ethiopia_4500BP that also is fitted as forming part of the
1604 ancestry of YRI and MSL, and secondly a lineage that diverged prior to
1605 South_Africa_2000BPbut after the basal West African lineage (**Fig. S5C**). Non-Africans
1606 (Japanese) are fitted as having part Archaic ancestry (Green et al., 2010), with the remainder of
1607 their ancestry again being derived from the lineage related to Ethiopia_4500BP that also is fitted
1608 as forming part of the ancestry of YRI and MSL (**Fig. S5D**). This analysis is consistent with the
1609 possibility that the same human lineage contributed ancestry both to the source of non-Africans
1610 and many African populations today.

1611

1612 **Support for a single out-of-Africa founding population**
1613 Simple tree models suggest that non-African variation represented by Sardinian, English, Han
1614 Chinese and Japanese falls within the variation of African populations. To test whether non-
1615 Africans are indeed consistent with being descended from a homogeneous population that
1616 separated earlier from the ancestors of a subset of African populations – beyond the known
1617 effects of archaic admixture in non-Africans – we used African populations with little or no
1618 known West Eurasian mixture (South_Africa_2000BP, Mbuti, Biaka, Mende, Ethiopia_4500BP,
1619 Dinka) and tested whether they are consistent with being an unrooted clade with respect to a
1620 diverse set of non-Africans (Orcadian, Onge, Mixe, Motala_Mesolithic, Japanese,
1621 Anatolia_Neolithic) using *qpWave* (Patterson et al., 2012; Reich et al., 2012). We found that this
1622 model was consistent with the data (P=0.53) (transition SNPs excluded to a final set of 110,507
1623 transversion SNPs). Even when we add New Guinean highlanders to the set of non-Africans, the
1624 single-source model for the out-of-Africa founders is not rejected (P=0.11).

1625

1626 **Date of admixture between expanding agriculturalists and previously established foragers**
1627 We estimated the date of admixture between expanding agriculturalists related to western
1628 Africans (probably Bantu-speakers) and the previously established foragers using the ALDER
1629 software (Loh et al., 2012), which uses the rate of decay per generation of the linkage
1630 disequilibrium that is created in admixed populations and that can be detected using signed
1631 linkage disequilibrium weighted by allele frequency differences between populations taken as
1632 proxies for the ancestral populations (Moorjani et al., 2011). To maximize statistical power, we
1633 used the full 1240k data merged with 1000 genomes phase 3 genotype data. We then estimated
1634 weighted linkage disequilibrium in 99 unrelated LWK individuals (Luhya from Webuye, Kenya)
1635 a Bantu-speaking group, using 85 unrelated West African MSL individuals (Mende from Sierra
1636 Leone) and 4 ancient eastern African hunter-gatherer individuals (Ethiopia_4500BP,
1637 Tanzania_Pemba_1400BP, Tanzania_Zanzibar_1400BP, Kenya_400BP). The analysis used a
1638 total of 1,070,197 SNPs. We obtained significant evidence for one-reference weighted LD decay

1639   for both putative source populations (Z ~ 6), as well as for the two-reference weighted LD decay
1640   (Z=4.89, P=10^{-5}). The estimated date of mixture was 16.8 generations ago, with a standard error
1641   of 3.4. Assuming a generation time of 30 years, this suggests that admixture occurred on average
1642   380 to 760 years ago (95% confidence interval). We note that a previous study (Pickrell et al.,
1643   2014) did not obtain high-confidence support for West African related admixture in the Luhya,
1644   and we hypothesize that our clear demonstration of this is due to both the availability of a more
1645   accurate ancestral population in the form of the ancient eastern Africans, as well as to the
1646   increased leverage from the large samples size 1000 Genomes data.
1647

1648 **Evidence for selective sweeps in the ancestry of present-day San**
1649   We performed a genome-wide scan for large genomic regions with excessive allele frequency
1650   between present-day San (Khomani and Ju_hoan_North) and the two ancient
1651   South_Africa_2000BP, with the Mbuti as a second outgroup. Previous statistics such as the
1652   Locus-Specific Branch length (LSBL) (Shriver et al., 2004) and the derivative Population Branch
1653   Statistic (PBS) (Yi et al., 2010) also estimate a branch length in a three-population phylogeny,
1654   but use F_{ST} as their base, which can be undefined when used for loci with fixed differences.
1655   Thus, we computed the statistic f_3=(p_{San}-$p_{South\_Africa\_2000BP}$)(p_{San}-p_{Mbuti}) for windows of 500kb,
1656   separated by 10kb. We only retained windows with at least 50 SNPs, resulting in a total of
1657   l=262,047 autosomal loci. We approximated the neutral genome-wide average μ_f and its standard
1658   deviation σ_f by subsampling 546 of the windows, requiring that these were separated by at least 5
1659   million base pairs (Mb) and thus approximately independent. We then standardized the
1660   distribution of the test statistic f_w in each window as $Z(f_w)=(f_w - \mu_f)/\sigma_f$. We show the most
1661   deviating windows in **Table 2**.
1662

1663 **Evidence for polygenic selection**
1664   Investigating evidence for polygenic selection in African populations is complicated by the fact
1665   that most information on the genomic basis of human phenotypic variation has been based on
1666   analysis of highly differentiated populations such as Europeans. In the absence of rich
1667   phenotypic information about sub-Saharan African populations, we used gene ontology (GO)
1668   information, which draws on information from a wide array of studies on humans and non-
1669   human model organisms. We focused on 208 GO categories that contained at least 50 genes
1670   each, which allows us to compute genome-wide block jackknife standard errors using the entire
1671   genic regions. We focus on f_3-statistics that measure the length of one of the branches in a
1672   hypothetical three-way tree-like population history (assuming no admixture). The three
1673   populations we focused on were the ancient South_Africa_2000BP (n=2 pseudohaploid draft
1674   genomes), the present-day San (n=6 complete genomes), and the Mbuti (n=4 complete
1675   genomes). The availability of the two ancient South_Africa_2000BP genomes can in principle
1676   inform us about selection in the last ~2000 years since these individuals lived, or starting further
1677   back in time in case they are not from a direct ancestral population of the present-day San (our
1678   data suggest that they are more closely related to the Khomani San than to the Ju'/hoan). We thus

1679 computed the statistic f_3(Mbuti, South_Africa_2000BP; San) which is proportional to allele
1680 frequency differentiation in the present-day San compared to the other two populations.
1681
1682 We find that the "RESPONSE_TO_RADIATION" GO category is an outlier that shows the
1683 greatest degree of differentiation in this analysis. However, this could also be due to genes in this
1684 category constantly being under rapid evolution or having other differences compared to other
1685 categories. To test this, we computed the statistic f_3(San, South_Africa_2000BP; Mbuti), and
1686 found no strong signal in the response to radiation category. Instead, the category with the
1687 strongest evidence of differentiation in the Mbuti lineage since the divergence from the San
1688 groups is "REGULATION_OF_GROWTH", suggesting the possibility of relatively recent
1689 evolution of the shorter stature of present-day rainforest hunter-gatherer populations.
1690
1691 **DATA AVAILABILITY**
1692 Raw sequence data (bam files) from the 15 newly reported ancient individuals is available from
1693 the European Nucleotide Archive under accession no. PRJEB21878. The newly reported SNP
1694 genotyping data is available to researchers who send a signed letter to D.R. containing the
1695 following text: "(a) I will not distribute the data outside my collaboration; (b) I will not post the
1696 data publicly; (c) I will make no attempt to connect the genetic data to personal identifiers for the
1697 samples; (d) I will use the data only for studies of population history; (e) I will not use the data
1698 for any selection studies; (f) I will not use the data for medical or disease-related analyses; (g) I
1699 will not use the data for commercial purposes."

Supplemental Table titles and legends

Table S1. Related to Table 1. Details of sequence libraries prepared from ancient samples.

Table S2. Related to Table 1. Details of direct AMS radiocarbon bone dates.

Table S3. Related to Figure 2. Details of ancestry proportions inferred using qpAdm.

Table S4. Related to Table 1. Y-chromosome haplogroup informative polymorphisms inferred for male individuals with sufficient data.

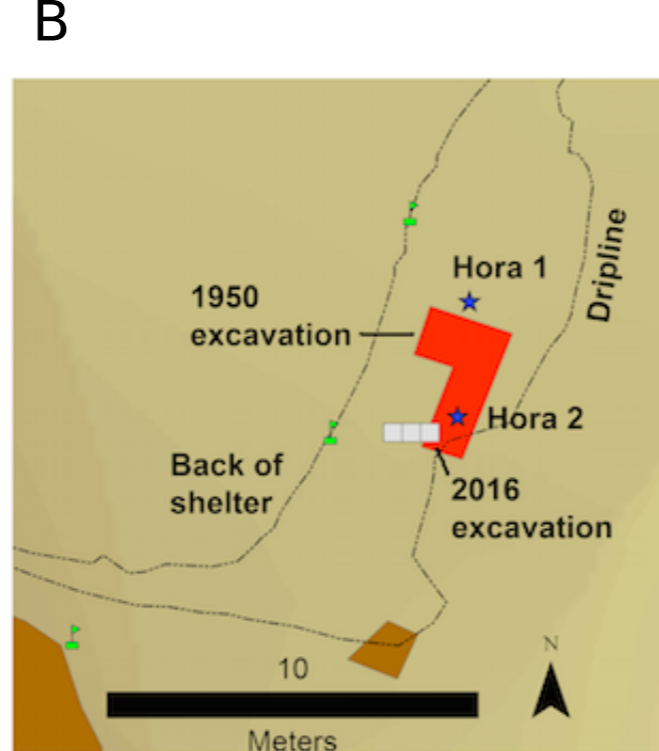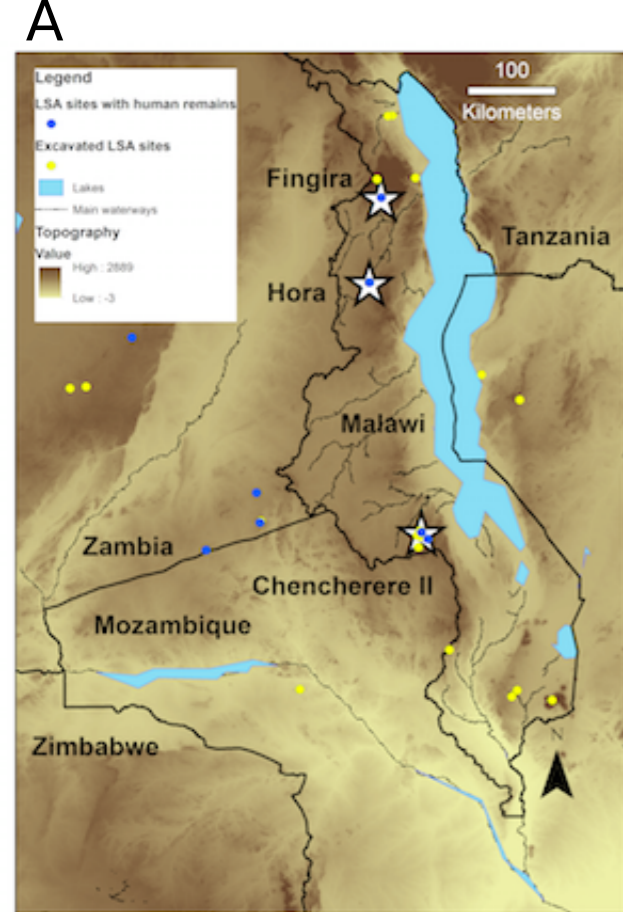**Table S5.** Related to Figure 1. Z-scores (computed for D-statistics) in complete genomes from present-day and ancient Africans.

Table S6. Related to Figure 3. Results of systematically inserting additional populations into all possible edges of the skeleton graph in Fig. S6A.

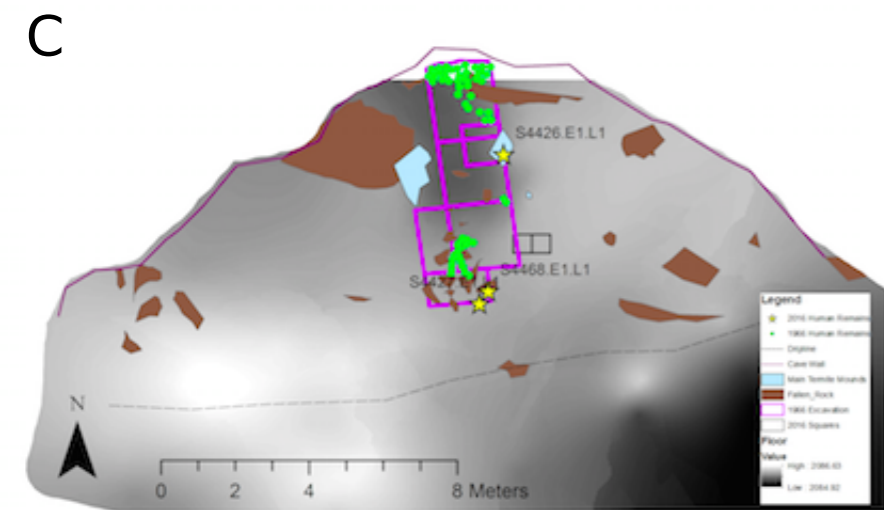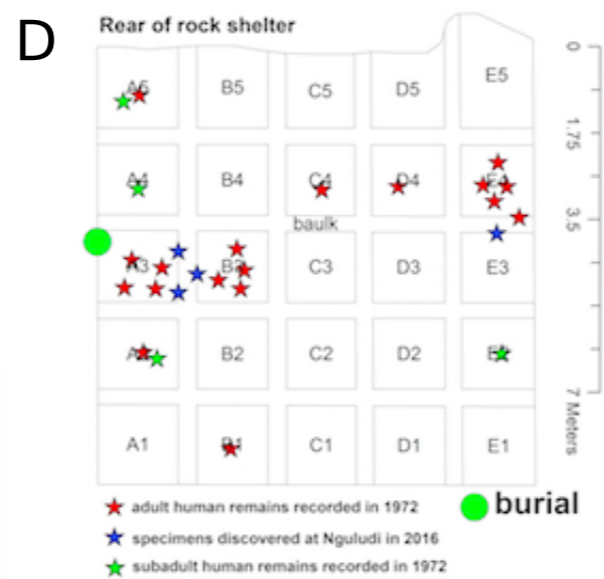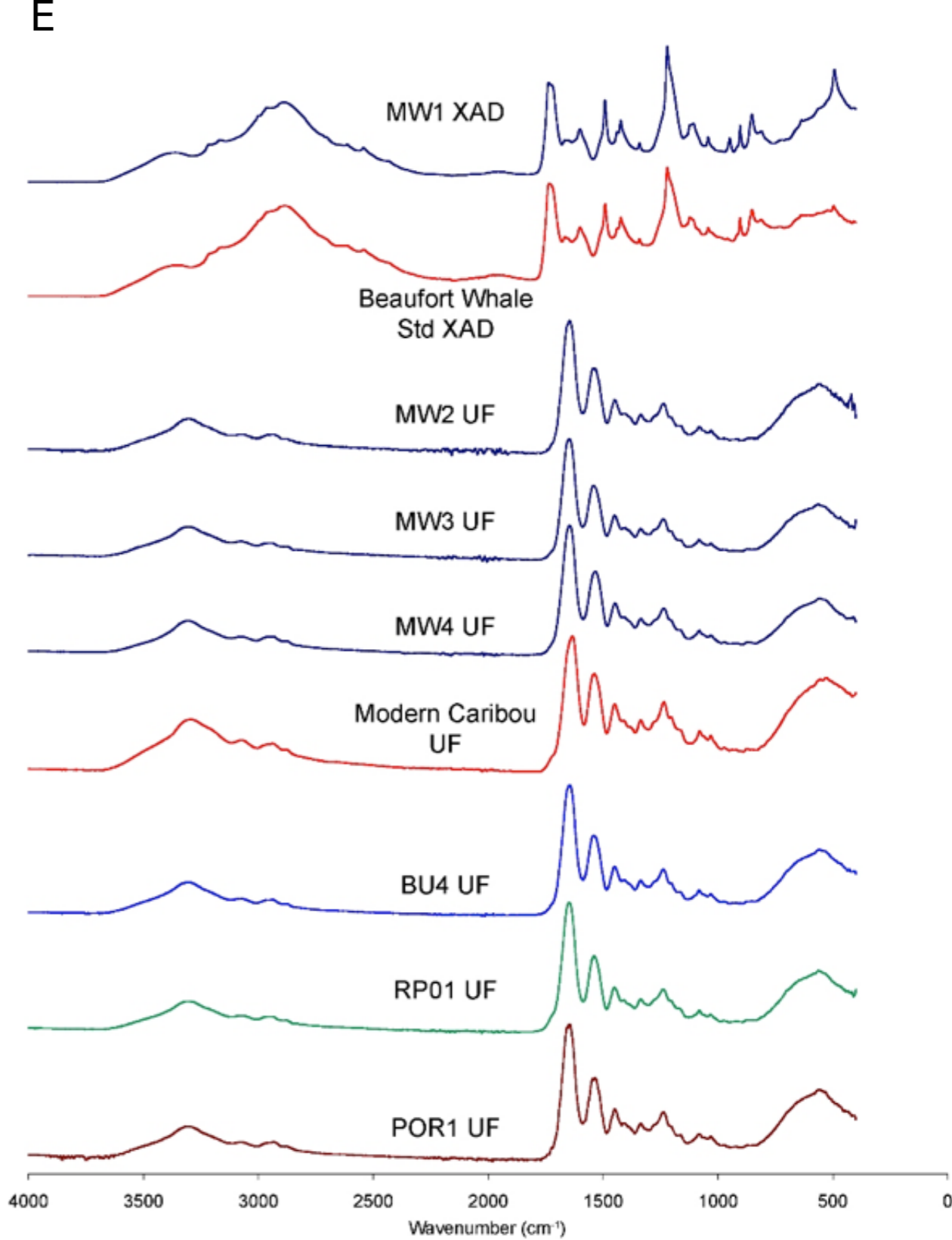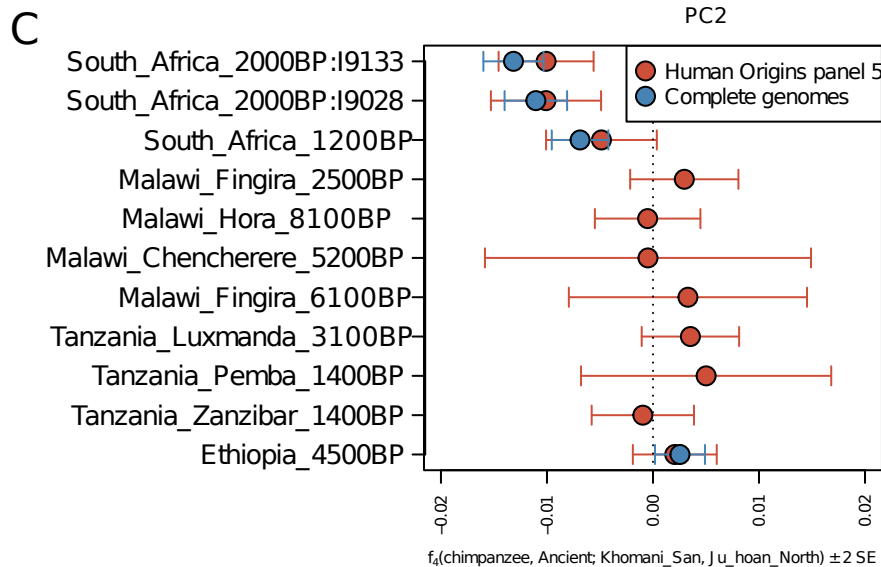**Table S7.** Related to Table 1. MtDNA haplogroup informative polymorphisms.
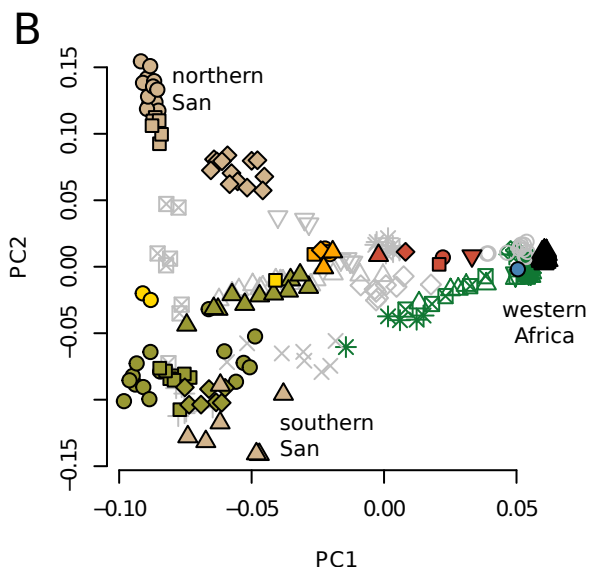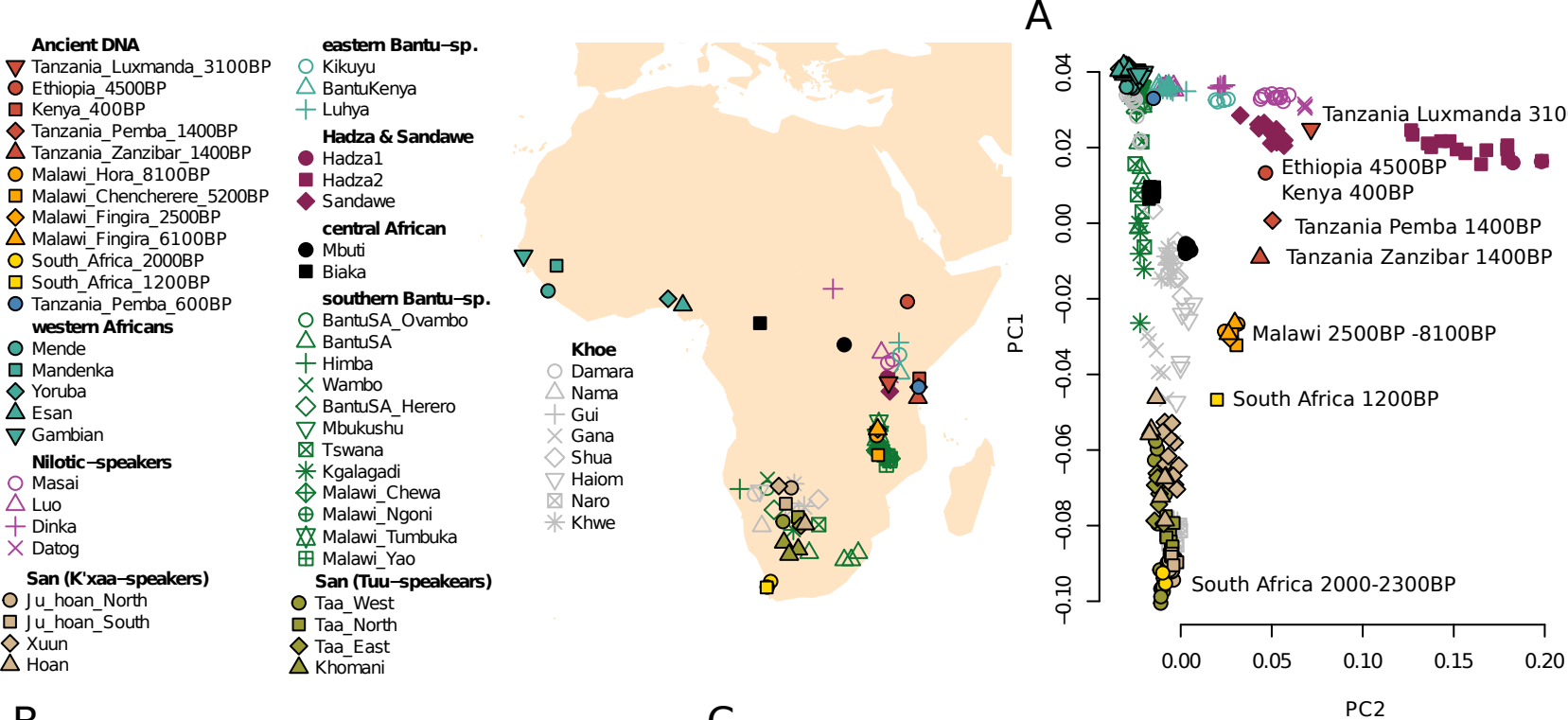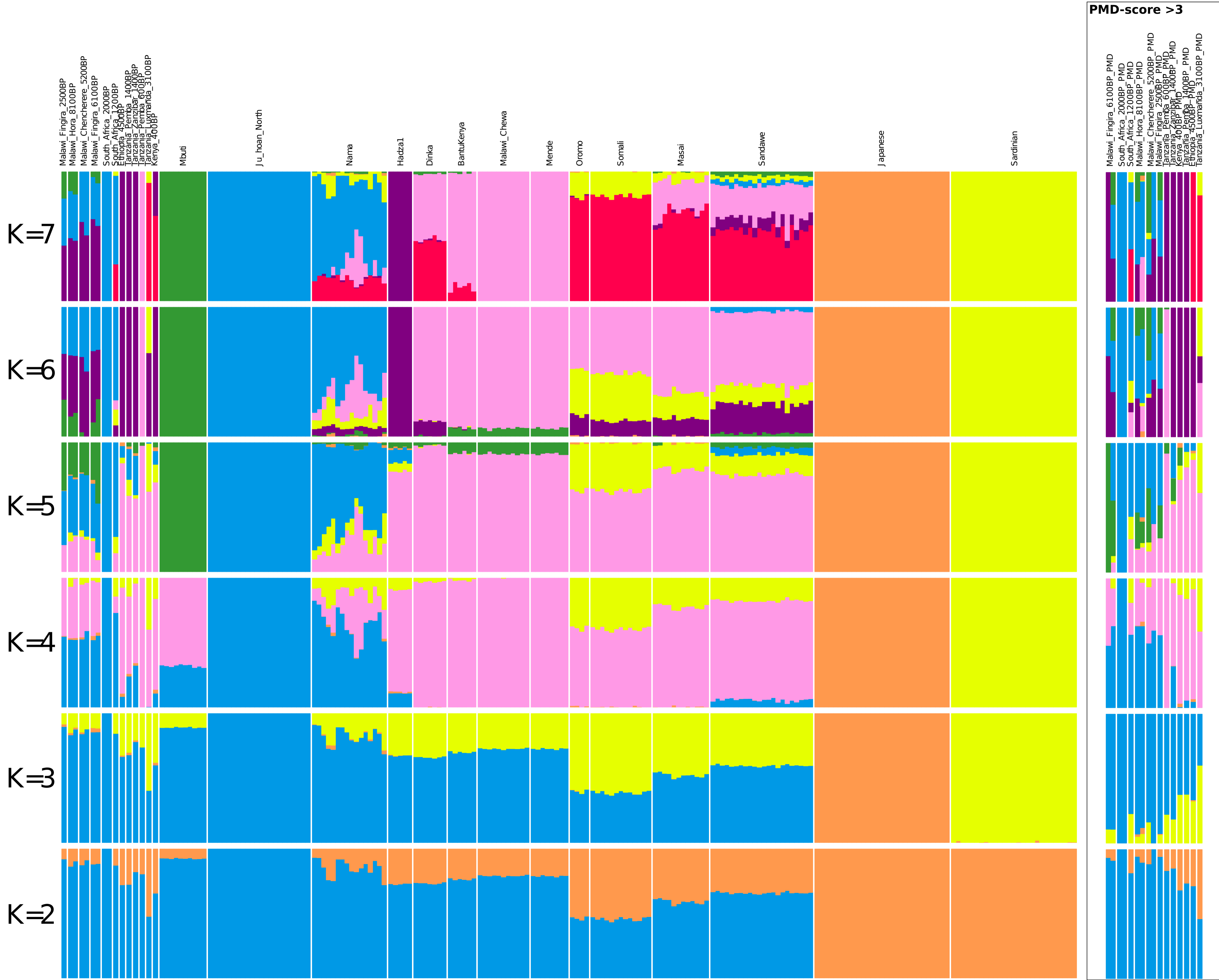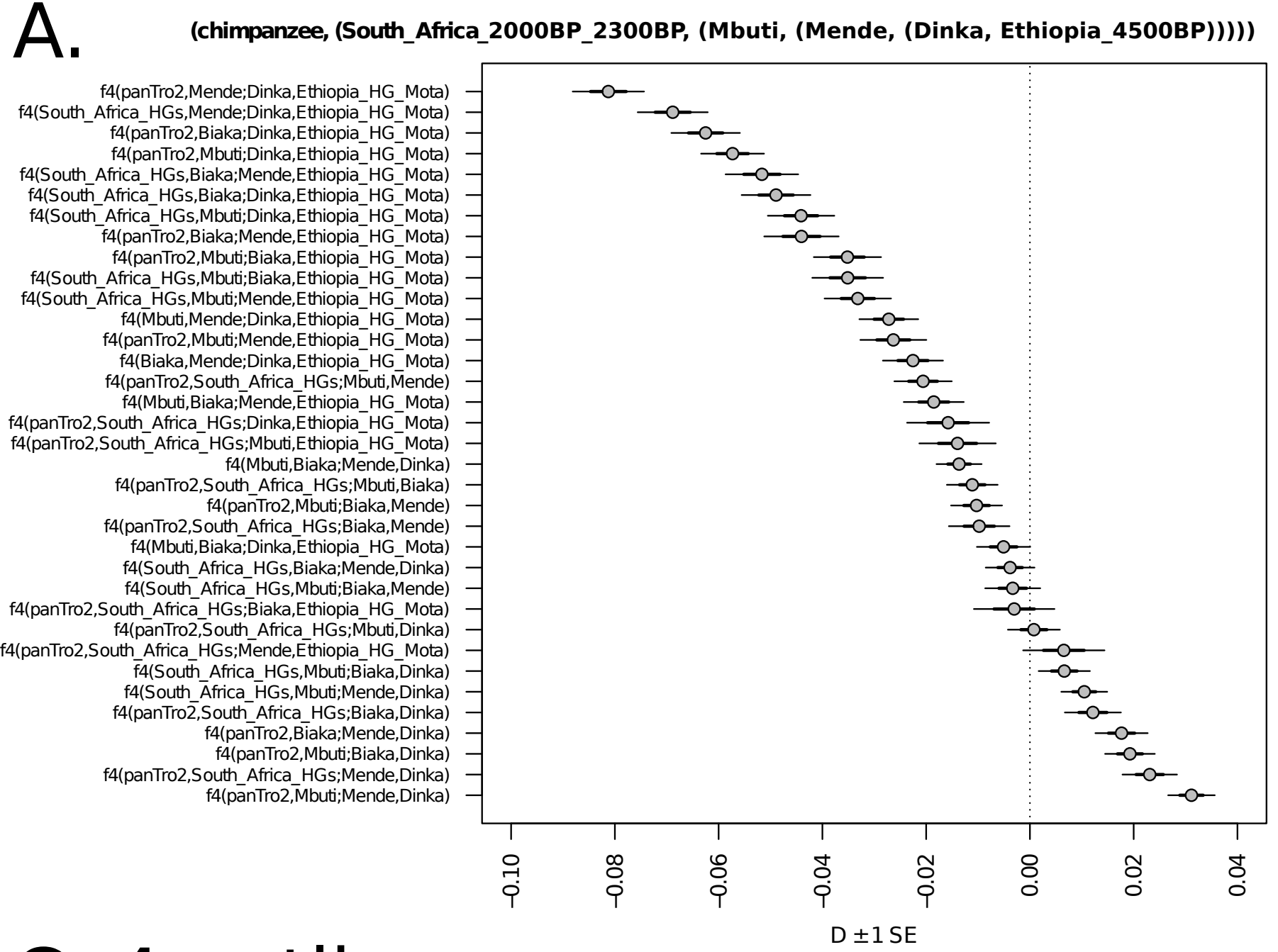
A

B Plan of Hora

C Plan of Fingira
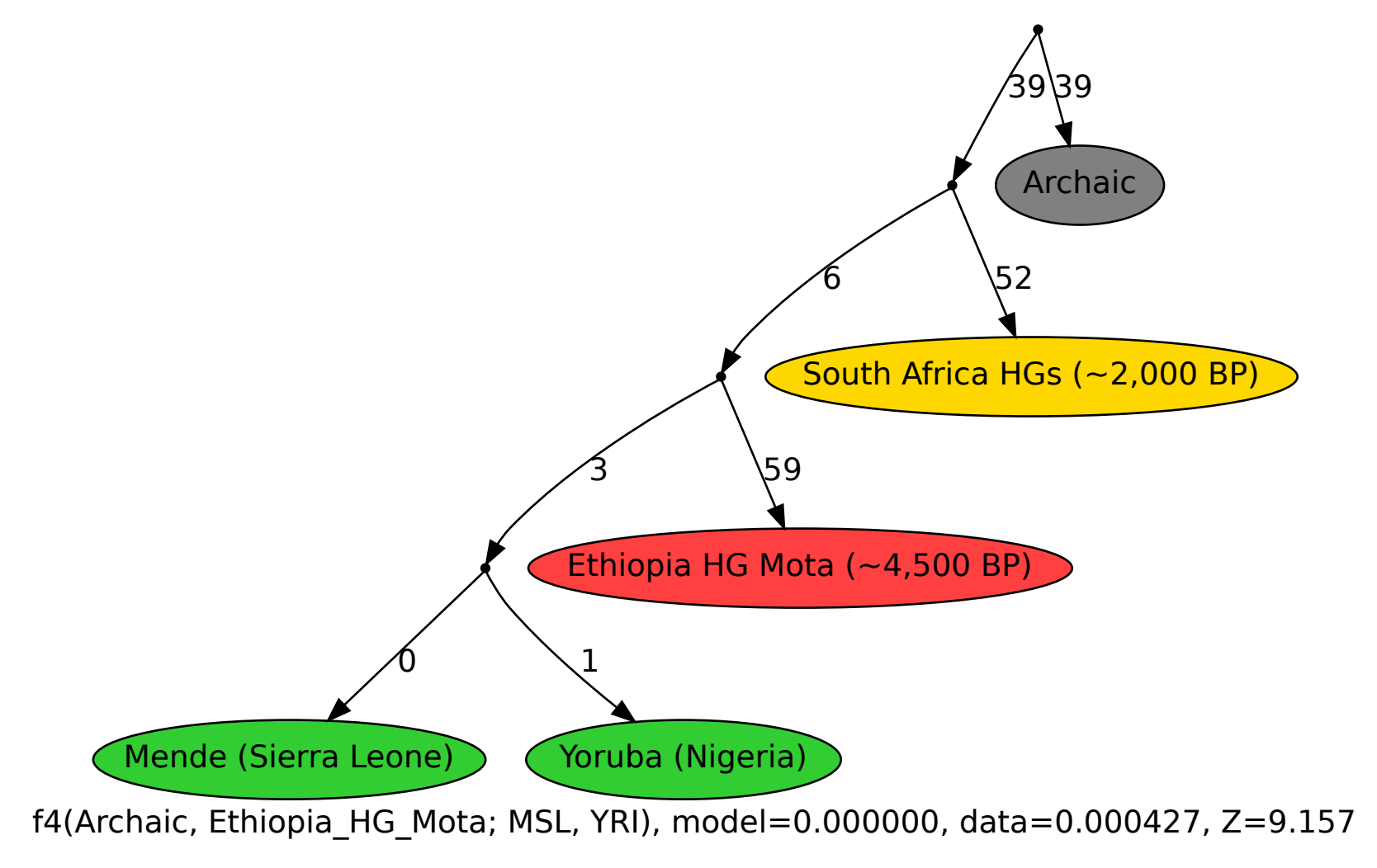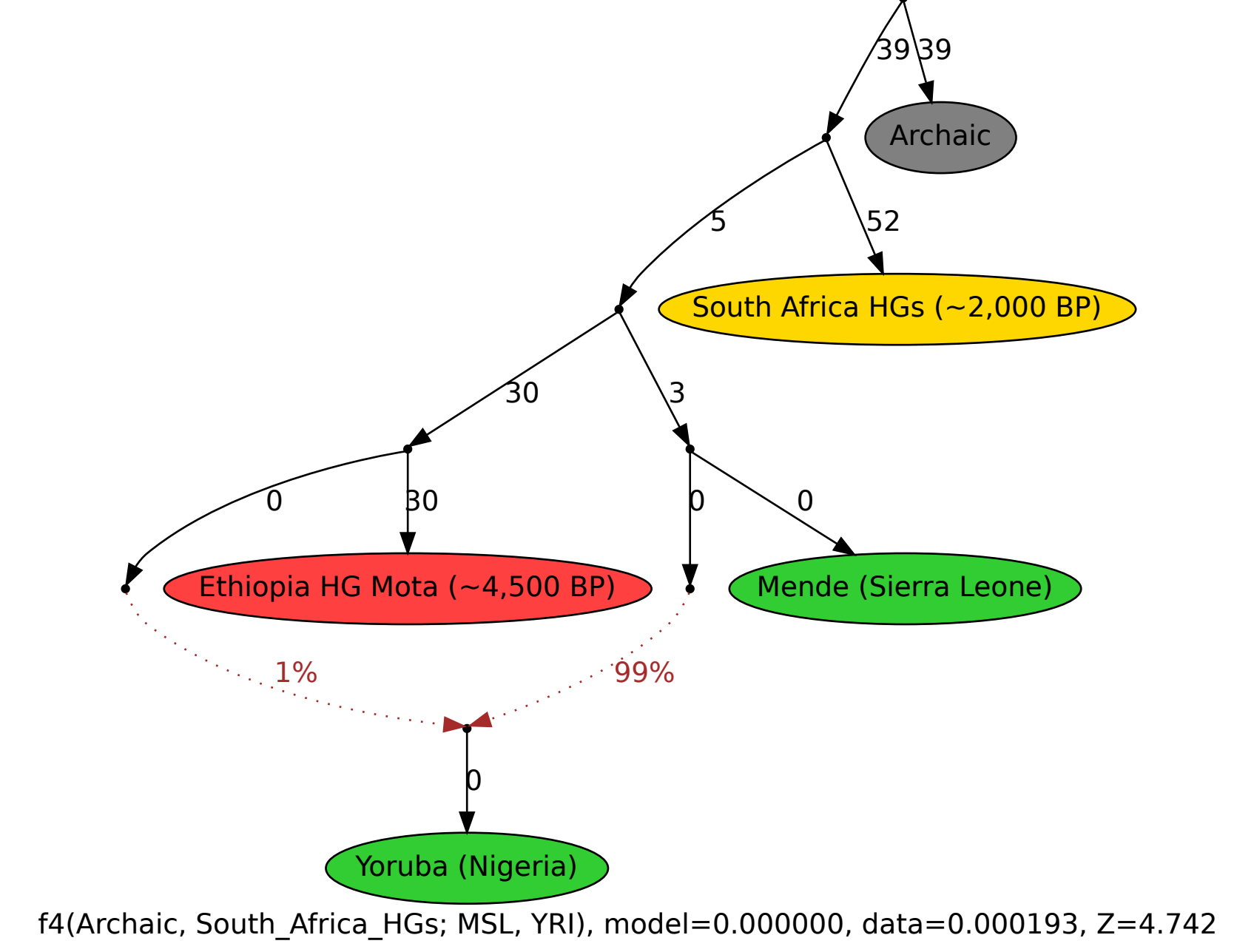
D Plan of Chencherere II
(redrawn from Crader [1984])

E

Legend (left panel):

Ancient DNA
- ▽ Tanzania_Luxmanda_3100BP
- ● Ethiopia_4500BP
- ● Kenya_400BP
- ◆ Tanzania_Pemba_1400BP
- ▲ Tanzania_Zanzibar_1400BP
- ● Malawi_Hora_8100BP
- ■ Malawi_Chencherere_5200BP
- ◆ Malawi_Fingira_2500BP
- ▲ Malawi_Fingira_6100BP
- ● South_Africa_2000BP
- ■ South_Africa_1200BP
- ● Tanzania_Pemba_600BP

western Africans
- ■ Mende
- ■ Mandenka
- ◆ Yoruba
- ▲ Esan
- ▽ Gambian

Nilotic-speakers
- ○ Masai
- △ Luo
- + Dinka
- × Datog

San (K'xaa-speakers)
- ● Ju_hoan_North
- ■ Ju_hoan_South
- ◆ Xuun
- ▲ Hoan

eastern Bantu-sp.
- ○ Kikuyu
- △ BantuKenya
- + Luhya

Hadza & Sandawe
- ● Hadza1
- ■ Hadza2
- ◆ Sandawe

central African
- ● Mbuti
- ■ Biaka

southern Bantu-sp.
- ○ BantuSA_Ovambo
- △ BantuSA
- + Himba
- × Wambo
- ◇ BantuSA_Herero
- ▽ Mbukushu
- ⊠ Tswana
- ✳ Kgalagadi
- ⊕ Malawi_Chewa
- ⊕ Malawi_Ngoni
- ⊠ Malawi_Tumbuka
- ⊞ Malawi_Yao

San (Tuu-speakears)
- ● Taa_West
- ■ Taa_North
- ◆ Taa_East
- ▲ Khomani

Khoe
- ○ Damara
- △ Nama
- + Gui
- × Gana
- ◇ Shua
- ▽ Haiom
- ⊠ Naro
- ✳ Khwe

A

PC1 vs PC2 scatter plot with labels: Tanzania Luxmanda 3100BP; Ethiopia 4500BP; Kenya 400BP; Tanzania Pemba 1400BP; Tanzania Zanzibar 1400BP; Malawi 2500BP -8100BP; South Africa 1200BP; South Africa 2000-2300BP

B

PC1 vs PC2 scatter plot with labels: northern San; western Africa; southern San

C

f_4(chimpanzee, Ancient; Khomani_San, Ju_hoan_North) ± 2 SE

- South_Africa_2000BP:I9133
- South_Africa_2000BP:I9028
- South_Africa_1200BP
- Malawi_Fingira_2500BP
- Malawi_Hora_8100BP
- Malawi_Chencherere_5200BP
- Malawi_Fingira_6100BP
- Tanzania_Luxmanda_3100BP
- Tanzania_Pemba_1400BP
- Tanzania_Zanzibar_1400BP
- Ethiopia_4500BP

Legend: ● Human Origins panel 5; ● Complete genomes

A. **(chimpanzee,(South_Africa_2000BP_2300BP,(Mbuti,(Mende,(Dinka,Ethiopia_4500BP)))))**

B. 19 outliers

f4(Archaic, Ethiopia_HG_Mota; MSL, YRI), model=0.000000, data=0.000427, Z=9.157

C. 4 outliers

f4(Archaic, South_Africa_HGs; MSL, YRI), model=0.000000, data=0.000193, Z=4.742

D. 13 outliers

f4(Archaic, South_Africa_HGs; South_Africa_HGs, YRI), model=-0.051655, data=-0.053170, Z=-3.986

E. 8 outliers

f4(Archaic, Ethiopia_HG_Mota; MSL, YRI), model=-0.000002, data=0.000427, Z=9.195

F. 1 outlier

f4(Archaic, Ethiopia_HG_Mota; Ethiopia_HG_Mota, YRI), model=-0.059468, data=-0.060880, Z=-3.157

G. 2 outliers

f4(Archaic, South_Africa_HGs; South_Africa_HGs, Ethiopia_HG_Mota), model=-0.050538, data=-0.052634, Z=-4.835

H. 2 outliers

f4(Archaic, South_Africa_HGs; Ethiopia_HG_Mota, MSL), model=0.000000, data=-0.000728, Z=-3.068