

1 Ancient West African foragers in the context of deep
2 human genetic history
3

4 Mark Lipson,^{1,*} Isabelle Ribot,² Swapan Mallick,^{1,3,4} Nadin Rohland,¹ Iñigo Olalde,¹ Nicole
5 Adamski,^{1,4} Nasreen Broomandkhoshbacht,^{1,4,†} Ann Marie Lawson,^{1,4} Saioa López,⁵ Jonas
6 Oppenheimer,^{1,4,†} Kristin Stewardson,^{1,4} Raymond Neba'ane Asombang,⁶ Hervé Bocherens,^{7,8}
7 Neil Bradman,^{5,†} Brendan J. Culleton,⁹ Els Cornelissen,¹⁰ Isabelle Crevecoeur,¹¹ Pierre de
8 Maret,¹² Forka Leypey Matthew Fomine,^{13,†} Philippe Lavachery,¹⁴ Christophe Mbida Mindzie,¹⁵
9 Rosine Orban,¹⁶ Elizabeth Sawchuk,¹⁷ Patrick Semal,¹⁶ Mark G. Thomas,^{5,18} Wim Van Neer,¹⁶
10 Krishna Veeramah,¹⁹ Douglas J. Kennett,²⁰ Nick Patterson,^{1,21} Garrett Hellenthal,⁵ Carles Lalueza-
11 Fox,²² Scott MacEachern,²³ Mary E. Prendergast,^{1,24,‡} David Reich^{1,3,4,‡}

12 ¹Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

13 ²Department of Anthropology, Université de Montréal, Montréal, QC H3C3J7, Canada

14 ³Medical and Population Genetics Program, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

15 ⁴Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA

16 ⁵UCL Genetics Institute, University College London, London WC1E 6BT, UK

17 ⁶Department of Arts and Archaeology, University of Yaoundé I, PO Box 6544 Yaoundé, Cameroon

18 ⁷Department of Geosciences, Biogeology, University of Tübingen, 72074 Tübingen, Germany

19 ⁸Senckenberg Research Centre for Human Evolution and Paleoenvironment, University of Tübingen, 72076 Tübingen,
20 Germany

21 ⁹Institutes for Energy and the Environment, Pennsylvania State University, University Park, PA 16802, USA

22 ¹⁰Department of Cultural Anthropology and History, Royal Museum for Central Africa, 3080 Tervuren, Belgium

23 ¹¹Université de Bordeaux, CNRS, UMR 5199-PACEA, 33615 Pessac Cedex, France

24 ¹²Faculty of Philosophy and Social Sciences, Université Libre de Bruxelles, Brussels, B-1050, Belgium

25 ¹³Department of History, University of Yaoundé I, PO Box 337 Yaoundé, Cameroon

26 ¹⁴Agence wallonne du Patrimoine, Service Public de Wallonie, 5100 Namur, Belgium

27 ¹⁵Department of Arts and Archaeology, University of Yaoundé I, PO Box 14617 Yaoundé, Cameroon

28 ¹⁶Royal Belgian Institute of Natural Sciences, Brussels, B-1000, Belgium

29 ¹⁷Department of Anthropology, Stony Brook University, Stony Brook, NY 11790, USA

30 ¹⁸Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, UK

31 ¹⁹Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY 11794, USA

32 ²⁰Department of Anthropology, University of California, Santa Barbara, CA 93106, USA

33 ²¹Department of Human Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA

34 ²²Institute of Evolutionary Biology (CSIC-UPF), Barcelona 08003, Spain

35 ²³Division of Social Science, Duke Kunshan University, Kunshan, China 215316

36 ²⁴Division of Humanities, Saint Louis University, Madrid 28003, Spain

37 *Email: mlipson@genetics.med.harvard.edu

38 †Present addresses: Department of Anthropology, University of California, Santa Cruz, CA 95064, USA (N.Bro.);

39 Department of Biomolecular Engineering, University of California, Santa Cruz, CA 95064, USA (J.O.); The Henry

40 Stewart Group, London WC1A 1LT, UK (N.Bra.); Department of History, University of Buea, PO Box 63 Buea,

41 Cameroon (F.L.M.F.)

42 ‡These authors equally supervised this work

43

44 **We generated genome-wide DNA data from four children buried roughly 8000 and 3000**
45 **years ago at Shum Laka (Cameroon), one of the earliest archaeological sites potentially**
46 **associated with the origins of Bantu languages. One individual carried the deeply diver-**
47 **gent Y chromosome haplogroup A00, which is found today almost exclusively in the same**
48 **region. However, all four individuals' genome-wide ancestry profiles are most similar to**
49 **West-Central African hunter-gatherers, implying that present-day populations in western**
50 **Cameroon, as well as Bantu speakers across the continent, are not descended substantially**

51 **from the population represented by these four people. Combining the Shum Laka indi-**
52 **viduals with existing data, we derive a detailed phylogenetic model that features eleven**
53 **ancestral admixture events and three prominent radiations within Africa, including one**
54 **giving rise to at least four major lineages deep in the history of modern humans.**

55 The deposits at Shum Laka, a rockshelter located in the Grassfields region of western
56 Cameroon, are among the most important archaeological sources for the study of Late Pleis-
57 tocene and Holocene prehistory in West-Central Africa [1–4]. The oldest human-occupied lay-
58 ers at the site date to approximately 30,000 calendar years before present (BP), but of greatest
59 interest are a series of artifacts and skeletons from about 8000–3000 BP, between the Later
60 Stone Age (LSA) and the Iron Age (Extended Data Fig. 1; Supplementary Information section
61 1). This transitional period, sometimes referred to as the Stone to Metal Age (SMA), featured
62 a gradual appearance of new stone tools as well as pottery [2, 4–6]. Subsistence evidence in
63 the rockshelter during the SMA points primarily to foraging, but with increasing usage of fruits
64 from *Canarium schweinfurthii* coinciding with developments in material culture, and serving
65 as a foundation for later agriculture [4, 7] (Supplementary Information section 1; Supplemen-
66 tary Table 1). In the context of broader African history, these cultural changes and their early
67 appearance at Shum Laka are particularly intriguing because the Cameroon/Nigeria border area
68 during the late Holocene was likely the cradle of Bantu languages, and of populations whose
69 descendants would spread across much of the southern half of Africa between ~3000–1500 BP,
70 resulting in the vast range and diversity of the Bantu language family today [8–18].

71 To explore population history in this region, and more broadly in Africa, we sampled human
72 bones from Shum Laka with the goal of extracting and sequencing ancient DNA (which to our
73 knowledge has not previously been reported from West or Central Africa). A total of eighteen
74 human skeletons (juveniles and adults) have been discovered at Shum Laka, comprising two
75 distinct burial phases (four individuals in the earlier phase and fourteen in the later phase; see

76 Supplementary Information section 1) [1–4]. We attempted to retrieve DNA from six individu-
77 als and obtained working data from four skeletons (three of which were mostly complete, and
78 one more fragmentary): two from the early SMA (~8000 BP) and two from the late SMA
79 (~3000 BP; Table 1, Supplementary Table 2). The two earlier individuals, 2/SE I and 2/SE II,
80 were both boys (2/SE I a child of 4 ± 1 years and 2/SE II an adolescent of 15 ± 3 years [3])
81 and were recovered from primary double burial #2, with the 2/SE I skeleton lying on top of the
82 lower limbs of 2/SE II. The two later individuals, 4/A and 5/B, were also children (4/A a boy of
83 8 ± 2 years and 5/B a girl of 4 ± 1 years [3]) and were found in adjacent primary single burials
84 #4 and #5.

85 We extracted DNA from petrous bone samples and prepared a total of 12 sequencing li-
86 braries (2–4 per individual, all treated with the enzyme uracil-DNA glycosylase (UDG) [19, 20]
87 to reduce the rate of damage-induced cytosine-to-thymine errors), from which we generated
88 genome-wide data by enriching for ~1.2 million single-nucleotide polymorphism (SNP) targets
89 (Methods; Supplementary Table 2). All four individuals returned data at more than 500,000
90 SNPs ($> 0.7\times$ average coverage), and three at more than 900,000 ($> 3.8\times$). Quality met-
91 rics indicated authentic ancient DNA: 4–10% C-to-T deamination damage in the final base
92 of sequenced fragments (relatively low but within the expected range after partial UDG treat-
93 ment [20]), and minimal apparent heterozygosity rates for mtDNA (0.3–1.5% estimated con-
94 tamination) and for the X chromosome in males (0.5–1.0% estimated contamination). The
95 molecular preservation of the samples is especially impressive given the long-term warm and
96 humid climate at Shum Laka [21] (supporting a mixed forest-savannah environment, at an ele-
97 vation of ~1650 meters above sea level). We also generated whole-genome sequence data for
98 individuals 2/SE II (~18.5 \times average coverage) and 4/A (~3.9 \times average coverage), and we re-
99 port new Human Origins array data (~598,000 SNPs) for 63 individuals from five present-day
100 Cameroonian populations (Extended Data Table 1; Supplementary Table 3).

Table 1. Details for the four ancient Shum Laka individuals in the study

ID	Age at death (yrs)	Date (cal BP)	Radiocarbon date (uncal)	Sex	Mt hap	Y hap	Cov	SNPs	Mt/X contam (%)
2/SE I	4 ± 1	7920–7700	6985 ± 30 BP (PSUAMS-6307)	M	L0a2a1	B	0.70	564164	1.0/1.0
2/SE II	15 ± 3	7970–7800	7090 ± 35 BP (PSUAMS-6308)	M	L0a2a1	A00	7.71	1082018	1.5/0.6
4/A	8 ± 2	3160–2970	2940 ± 20 BP (PSUAMS-6309)	M	L1c2a1b	B2b	3.83	935777	0.3/0.5
5/B	4 ± 1	3210–3000	2970 ± 25 BP (PSUAMS-6310)	F	L1c2a1b	..	6.41	1014618	0.5/..

Calibrated direct radiocarbon dates are given as 95.4% CI (see Methods). Age was determined from skeletal remains, and sex from genetic data [3]. Mt/Y hap, mtDNA/Y-chromosome haplogroup; Cov, average sequencing coverage.; Mt/X contam, estimated contamination from mtDNA/X chromosome. See also Supplementary Table 2.

101 Uniparental markers and kinship analysis

102 All of the mtDNA and Y chromosome haplogroups we observe among the Shum Laka indi-
 103 viduals are associated today with sub-Saharan African populations. The two earlier individuals
 104 carry mtDNA haplogroup L0a (specifically L0a2a1), which is widespread in Africa, while the
 105 two later individuals carry L1c (specifically L1c2a1b), which is found among both farmers and
 106 hunter-gatherers in Central and West Africa [22, 23]. Individuals 2/SE I and 4/A have Y chro-
 107 mosomes from macrohaplogroup B, often found today in Central African hunter-gatherers [24],
 108 while 2/SE II has the rare Y chromosome haplogroup A00, which was discovered in 2013
 109 and has subsequently been identified at low frequencies in present-day Cameroon, in particular
 110 among the Mbo and Bangwa groups in the western part of the country [25, 26]. A00 is the oldest
 111 known extant branch of the human Y chromosome tree, with a split time from all other modern
 112 human Y chromosomes of ~200,000–300,000 BP [25, 27, 28]. Our documentation of A00 at
 113 Shum Laka is its first known instance in ancient DNA.

114 We used our whole-genome sequence data to investigate the relationship of 2/SE II’s Y
 115 chromosome to present-day A00 sequences (Supplementary Table 4). Present-day A00 chro-
 116 mosomes are classified into the subtypes A00a, A00b, and A00c, whose divergence times from

117 each other have not been precisely estimated but are quite recent, perhaps only a few thousand
 118 years [25, 26]. At every subtype-specific site for which we had coverage, the Shum Laka A00
 119 carries the ancestral allele. We called genotypes at 1666 positions that differ between (present-
 120 day) A00 [27] and all other modern human Y chromosomes, and we found that the Shum Laka
 121 A00 carries the alternative allele at 1521 (91%). Using published calibrations [28, 29], we esti-
 122 mate a time of 31,000 BP (95% CI: 25,000–37,000 BP; see Methods) for the split of the 2/SE
 123 II A00 from present-day A00 sequences (Fig. 1), meaning that it cannot be directly ancestral to
 124 the present-day subtypes.

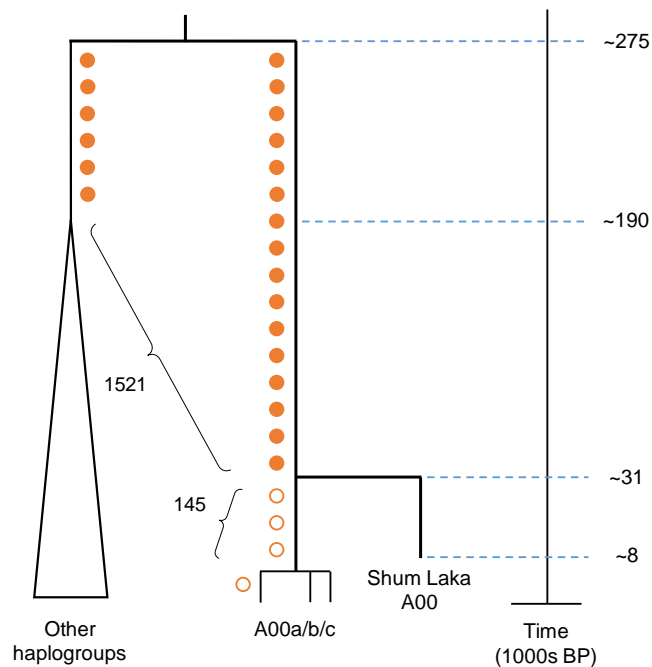


Figure 1. Y chromosome phylogeny. Circles represent mutations along the (unrooted) A00 lineage at sites where we observe the alternative (filled) or reference (empty) allele state in the Shum Laka A00.

125 On a genome-wide level, we computed rates of allelic identity for each pair of individuals
 126 to infer degrees of relatedness, leveraging the effects of chromosomal segments shared identical

127 by descent (IBD). Both contemporaneous pairs display elevated identity, with 2/SE I and 2/SE
128 II at the level of fourth-degree relatives and 4/A and 5/B at the level of second-degree relatives
129 (Extended Data Fig. 2), supporting archaeological interpretations that the rockshelter was used
130 as an extended family cemetery during both burial phases [3]. We would expect more recent
131 average shared ancestry for the contemporaneous pairs even if they were not closely related,
132 but when computing allele matching along the genome, we observe clear signatures of long
133 IBD segments, meaning that the genome-wide levels indeed reflect family relatedness (and
134 confirming that both pairs indeed lived close in time; Supplementary Information section 2).
135 All four individuals also show evidence of relatively recent inbreeding, both from genome-wide
136 identity and window-based analysis (Extended Data Fig. 2; Supplementary Information section
137 2). For 4/A and 5/B, because both died as children, we can eliminate a grandparent-grandchild
138 relationship, and the lack of long segments with both homologous chromosomes shared IBD
139 implies that they are not double cousins (the few ostensible double-IBD stretches are likely a
140 result of inbreeding). Thus, we can conclude that they were either uncle and niece (or aunt and
141 nephew) or half-siblings.

142 **PCA and allele-sharing statistics**

143 We visualized the genome-wide relationships between the Shum Laka individuals and diverse
144 present-day and ancient sub-Saharan African populations (Extended Data Table 1) using prin-
145 cipal component analysis (PCA). Initially, we computed axes using East and West Africans
146 and Southern and East-Central African hunter-gatherers, projecting the Shum Laka individuals
147 together with other populations for comparison (Fig. 2A). Along PC1, the overall trend is for
148 (historically) farming and pastoralist populations to fall toward the left and hunter-gatherers
149 toward the right. The position of the Shum Laka individuals is to the right of Bantu speak-

150 ers and related West African populations (Chewa, Mbo, and Mende), closest to present-day
 151 West-Central hunter-gatherers from Cameroon (Baka, Bakola, and Bedzan [30]) and the Cen-
 152 tral African Republic (Aka, often known as Biaka). To confirm this signal, we carried out a
 153 second PCA using only West and East Africans and Aka to compute the axes, and again the
 154 Shum Laka individuals project in the direction of West-Central hunter-gatherers (Fig. 2B). By
 155 contrast, present-day Niger-Congo-speaking groups from western Cameroon, including Mbo
 156 and Bangwa, cluster tightly with other West Africans when projected onto PCs 1 and 2 (Fig. 2;
 157 Extended Data Fig. 3A). In both plots, the two earlier Shum Laka individuals fall slightly closer
 158 to West and East Africans than do the more recent individuals, but all four appear to be quite
 159 similar in their ancestry relative to other populations, and we therefore grouped them together
 160 in the analyses that follow unless otherwise noted.

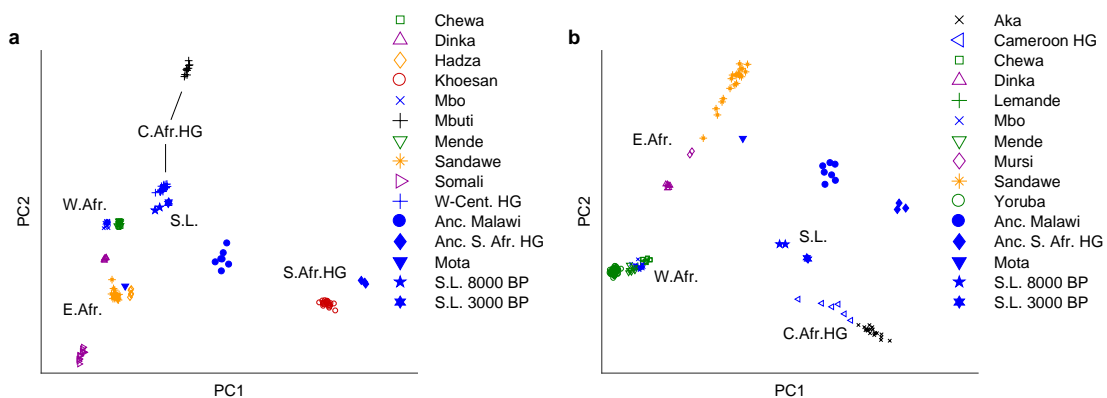


Figure 2. PCA results. (A) Broad-scale analysis. (B) Narrow-scale analysis. Groups shown in blue (including all ancient individuals, with filled symbols) were projected onto axes computed using the other populations. HG, hunter-gatherers; S.L., Shum Laka. The W-Cent. HG grouping in (A) (closest to Shum Laka) consists of Aka and Cameroon hunter-gatherers (Baka, Bakola, and Bedzan). For both analyses, we used SNPs from the Human Origins array (Extended Data Table 1).

161 To refine these observations, we used f -statistics [31] (Fig. 3A) to investigate levels of “deep
 162 ancestry,” which we define as ancestry from sources diverging earlier than the split between

163 non-Africans and the majority of sub-Saharan Africans (i.e., above the point (2) in the tree in
164 Fig. 4A). Initially, we employed the statistic $f_4(X, \text{Mursi}; \text{South Africa HG, Han})$, which is
165 expected to be increasingly positive for increasing proportions (and, to some degree, for earlier
166 divergences) of deep ancestry in population X , with a baseline of zero set by Mursi, Nilotic-
167 speaking pastoralists from western Ethiopia [30] (and assuming no specific allele-sharing be-
168 tween X and either ancient South African hunter-gatherers [32, 33] or non-Africans). We find
169 that Shum Laka has a large positive statistic, comparable to that of West-Central African hunter-
170 gatherers (Fig. 3A, top), suggesting that the genetic affinity between the Shum Laka individ-
171 uals and present-day hunter-gatherers could be due to a shared component of deep ancestry.
172 Other West Africans (e.g., Yoruba and Mende) yield smaller but significantly positive values,
173 as do East African hunter-gatherers (present-day Hadza from Tanzania and the ~ 4500 BP Mota
174 individual from Ethiopia [34]). We also computed related statistics in which we used differ-
175 ent reference groups in place of Mursi, South African hunter-gatherers, and Han, allowing us
176 both to confirm the robustness of our inferences and to extend them to include additional test
177 populations (Extended Data Table 2). From the statistics $f_4(\text{Mursi/Agaw, Han}; \text{South Africa}$
178 $\text{HG, Yoruba})$, we find minimal differences in deep ancestry proportions among Han, Mursi,
179 and Agaw (an Afroasiatic-speaking population from Ethiopia [30]); from $f_4(X, \text{Mursi}; \text{Chimp,}$
180 $\text{Yoruba})$, we obtain a value for South African hunter-gatherers that is roughly twice as large
181 as for Central African hunter-gatherers (using chimpanzee as a deep outgroup symmetric to all
182 human populations).

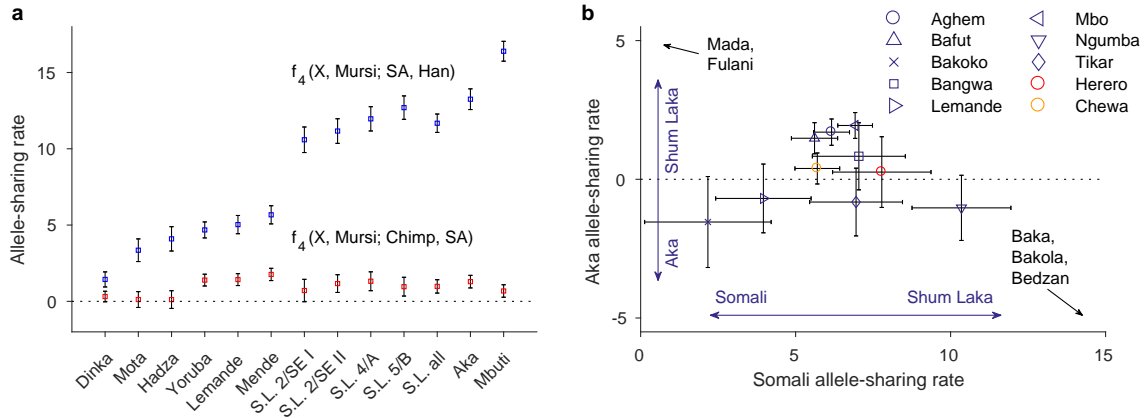


Figure 3. Allele-sharing statistics. (A) Statistics sensitive to ancestry from a deeply-splitting lineage (multiplied by 1000; blue, deeper than non-Africans; red, deeper than South African hunter-gatherers). Bars show two standard errors in each direction. S.L., Shum Laka; SA, ancient South African hunter-gatherers. (B) Relative allele sharing (multiplied by 10,000) with Shum Laka versus East Africans ($f_4(X, \text{Yoruba}; \text{Shum Laka}, \text{Somali})$; x-axis) and versus Aka ($f_4(X, \text{Yoruba}; \text{Shum Laka}, \text{Aka})$; y-axis) for present-day populations from Cameroon (blue points) and southern and eastern Bantu speakers (Herero in red and Chewa in orange). Bars show one standard error in each direction. See Extended Data Fig. 3B for wider plot.

183 We also explored how much, if any, of this deep ancestry is from sources (potentially in-

184 cluding archaic humans) diverging more deeply than Southern African hunter-gatherers (the

185 modern human population with the oldest known average split date [32, 35, 36]). For this pur-

186 pose, we employed the statistic $f_4(X, \text{Mursi}; \text{Chimp}, \text{South Africa HG})$. Previous work has

187 shown that Southern African hunter-gatherers are not a symmetric outgroup relative to other

188 sub-Saharan Africans, with West Africans (especially Mende) having excess affinity toward

189 deeper outgroups [33]. In agreement with this observation, we find that our test statistic is max-

190 imized in Mende and other West Africans (Fig. 3A, bottom). Hadza and Mota have values close

191 to zero, and Shum Laka and Central African hunter-gatherers are intermediate. While some

192 populations yield positive values for both f_4 -statistics (Fig. 3A), the fact that the two sets are

193 poorly correlated implies that they reflect at least partially separate signals.

194 Combining our new genotype array data with published individuals from Cameroon [30],
195 we searched for differential relatedness between the Shum Laka individuals and present-day
196 Cameroonians (Fig. 3B, Extended Data Fig. 3B). We computed allele-sharing statistics using
197 Yoruba as a baseline and either East Africans (Somali) or Aka in the outgroup position and
198 identified three distinct clusters: (a) Mada and Fulani, (b) hunter-gatherers, and (c) a relatively
199 tight grouping of Niger-Congo-speaking populations (shown in closeup in Fig. 3B). Within
200 the third cluster, we find the only subset of populations with significantly positive (i.e., Shum
201 Laka-oriented) values in both dimensions; these groups (Mbo, Aghem, and Bafut), who also live
202 close to the site of Shum Laka, thus have evidence of slight excess relatedness to the Shum Laka
203 individuals despite their low overall genome-wide differentiation from other West Africans (Ex-
204 tended Data Fig. 3A).

205 **Admixture graph analysis**

206 To validate and extend these signals of admixture as part of an integrated phylogenetic model,
207 we built an admixture graph (Fig. 4A, Extended Data Fig. 4) co-modeling the ancient Shum
208 Laka, Mota, and South African hunter-gatherer individuals and present-day Mbuti, Aka, Agaw,
209 Yoruba, Mende, and Lemande, together with non-Africans (French) and two outgroups (Al-
210 tai Neanderthal and chimpanzee). The final model provides a good fit to the data, with all
211 f -statistics relating subsets of the populations predicted to within 2.3 standard errors of their
212 observed values. Initially, we detected a slight but significant signal (max $Z = 2.5$) of allele-
213 sharing between Shum Laka and non-Africans, which we hypothesize is due to a small amount
214 of DNA contamination. To prevent this effect from influencing our results, we included a
215 “dummy” admixture of non-African ancestry into Shum Laka (inferred 1.1%, consistent with
216 mtDNA- and X chromosome-based contamination estimates), although model parameters with-

217 out the dummy admixture are also very similar (Extended Data Table 3, Supplementary Infor-
 218 mation section 3). To check the robustness of our inferences, we also fit versions of the model
 219 using alternative SNP ascertainment schemes and with additional populations (Hadza, Mbo,
 220 Herero, Chewa, Mursi, Baka, Bakola, Bedzan, Mada, Fulani, and ancient individuals from
 221 Taforalt in Morocco [37]) and obtained qualitatively concordant results in all cases (Extended
 222 Data Table 3; Supplementary Information section 3).

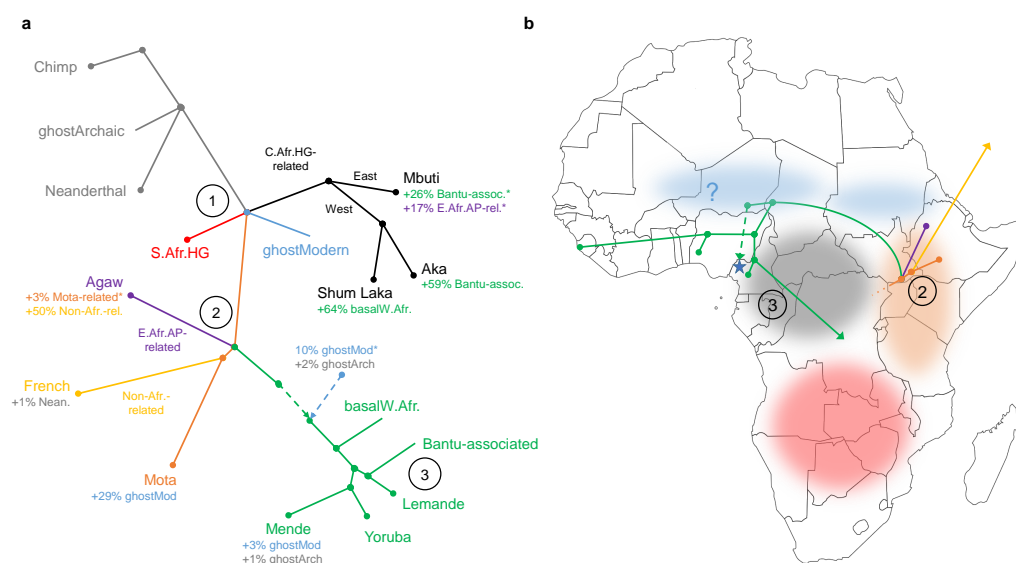


Figure 4. Admixture graph results. In both panels, points at which multiple lineages are shown diverging simultaneously indicate splits occurring in short succession (whose order we cannot confidently assess) but are not meant to represent exact multifurcations. Key points correspond to (1) early modern human split, (2) East African divergences, and (3) Bantu expansion. Branch lengths are not drawn to scale. (A) Full model; see Extended Data Fig. 4 for branch lengths. HG, hunter-gatherer; AP, agro-pastoralist. *Proportion not well constrained (for Mbuti, the sum of the two indicated proportions is well constrained but not the separate values). (B) Geographical structure: shaded areas correspond to rough hypothesized historical locations of lineages descended from split point (1) in panel (A), and branching order is shown for populations descended from split point (2). For ease of visualization, we select one ancestry component per population. Leaf nodes are placed at sampling locations, but the locations of internal nodes (and ancestral populations more generally) are not known. The blue star represents Shum Laka, with a possible direction of gene flow for one component of its ancestry (dashed green line).

223 Along the modern human lineage, the deepest-splitting branch in our model is inferred to
224 be the one leading to Central African hunter-gatherers, although four lineages diverge in a very
225 short span: those contributing the primary ancestry to (a) Central African hunter-gatherers, (b)
226 Southern African hunter-gatherers, and (c) other modern human populations, along with (d) a
227 “ghost” source contributing a minority of the ancestry in West Africans and the Mota individual.
228 Among Central African hunter-gatherers, the first split [38] is between East (Mbuti) and West,
229 with the latter then branching into components represented in Aka and Shum Laka. The next
230 major feature of the topology is a second cluster of divergences involving West Africans, two
231 East African lineages (hunter-gatherer-associated and agro-pastoralist-associated), and non-
232 Africans, the latter tentatively inferred to be a sister group to Mota but with no deep “ghost”
233 ancestry. Within the West African clade, we identify Yoruba and Mende as sister populations,
234 with Lemande as an outgroup, and most basally a separate West African-related lineage con-
235 tributing the majority of the ancestry for Shum Laka (64%). A Bantu-associated source (most
236 closely related to Lemande) contributes 59% of the ancestry in Aka and 26% in Mbuti [39], with
237 the latter also receiving ancestry (17%) from an East African agro-pastoralist-related source.
238 We can also obtain a good fit for the Shum Laka individuals in a less-parsimonious alternative
239 model using three components, replacing the basal West African source with a combination of
240 ancestry from inside the clade defined by the other West African populations and from a source
241 splitting between the East and West Africans (similar to the split point for one component con-
242 tributing to Taforalt; Extended Data Fig. 5, Supplementary Information section 3). However,
243 two-component models for Shum Laka with the majority component splitting along a differ-
244 ent branch create significant deviations from the observed data ($Z = 7.1$ closer to other West
245 Africans; $Z = 3.7$ closer to East Africans).

246 The West African clade (green in Fig. 4) is also distinguished by admixture from a deep
247 source that can be modeled as a combination of modern human and archaic ancestry. The mod-

ern human component is inferred to diverge at almost the same point as Central and Southern African hunter-gatherers and to be related to the deep source that contributes ancestry to the Mota individual, but how closely related is not well determined in our model. The archaic component fits best as being from a lineage that diverged close to the split between Neanderthals and modern humans (either slightly basal to the split or along the Neanderthal lineage in different versions; Supplementary Information section 3). The signals of deep ancestry in West African-related groups (Fig. 3A) can be explained parsimoniously by two admixture events: one along the ancestral West African lineage, and a second, smaller contribution ($\sim 4\%$) to Mende from the same deep source (Fig. 4A). In particular, statistics testing for ancestry basal to Southern African hunter-gatherers (Fig. 3A, bottom) are highly correlated to inferred proportions of ancestry from the West African clade (Extended Data Fig. 6). In our primary model, we estimate the shared admixture to introduce 10% deep modern human and 2% archaic ancestry, although the first proportion is not well constrained and is as high as $\sim 30\%$ in some versions of the graph (Extended Data Table 3). We also note that an alternative model with no archaic component, in which the West African clade receives deep ancestry from a single source splitting before the primary early modern human divergence point [33], also provides a good fit to the data, although modestly worse (Supplementary Information section 3). The two versions are quite similar overall, but combined with previous evidence for archaic ancestry in sub-Saharan African populations [40–48], we prefer the model in Fig. 4A.

Shum Laka in genetic context

Our analyses show that the four sampled children from Shum Laka can be modeled as admixed with $\sim 35\%$ ancestry related to West-Central African hunter-gatherers and $\sim 65\%$ from a basal West African-related source (originating outside of a clade containing diverse present-

271 day Niger-Congo speakers), or alternatively as a mixture of hunter-gatherer-related ancestry
272 plus two additional components, one from inside the clade of present-day West Africans and
273 one splitting between East and West Africans. The first component, given its relatedness to
274 hunter-gatherers still living in West-Central Africa, plausibly represents ancestry present in this
275 area since at least the LSA, whereas the second component (third in the alternative model) may
276 represent a lineage originally from outside the region. Although the scope of our sampling is
277 limited to two individuals at either end of the SMA, the observed genetic similarity across a
278 span of almost 5000 years suggests long-term continuity in the region, at least for one popu-
279 lation who repeatedly used the Shum Laka rockshelter for various activities, including burying
280 their dead (Supplementary Information section 1). The later pair did have slightly but signifi-
281 cantly more Central African hunter-gatherer-related ancestry than the earlier pair (e.g., $f_4(\text{Shum}$
282 $\text{Laka 8000 BP, Shum Laka 3000 BP; Yoruba, Aka}) > 0$, $Z = 4.2$; $\sim 5\%$ more from admixture
283 graph modeling, Supplementary Information section 3), which could reflect a minor resurgence
284 of local hunter-gatherer-related ancestry as in Neolithic Europe [49–51]. The genetic conti-
285 nuity we infer is also consistent with morphometric analyses of the remains (Supplementary
286 Information section 1).

287 Given the phylogenetic position of the basal West African-related ancestry component in
288 the Shum Laka individuals, together with the geography and phylogeny of other sampled West
289 African populations, a possible hypothesis is that this component had an origin farther to the
290 north (Fig. 4B). The chronology of the archaeological record at the site suggests a possible
291 northern influence on cultural developments during the SMA [4, 16]; these include changes
292 in stone tools, which can be interpreted as a fusion of local LSA tool-making traditions with
293 new macrolithic technologies introduced from the north [4], and the appearance of ceramics
294 (four sherds found in the early SMA burial layer, and more abundant and distinct ceramics
295 in later SMA deposits) perhaps derived from earlier pottery-working traditions in the Sahara

296 and Sahel [2, 4, 52]. Gene flow from the north before 8000 BP is also plausible due to a short
297 period of Saharan and Sahelian aridification [4, 53]. Present-day populations in northern West
298 Africa and the Sahel have extensive admixture connected to later migrations [54, 55], however,
299 so pinpointing the source of the Shum Laka ancestry will likely require additional ancient DNA
300 data.

301 Today, the large majority of the ancestry in populations from Cameroon is more closely
302 related to that of other West Africans rather than to the group represented by the ancient in-
303 dividuals from Shum Laka. Present-day hunter-gatherers in Cameroon are also not descended
304 substantially from this specific group, as they do not share the same signal of ancestry from
305 outside the main portion of the West African clade (Supplementary Information section 3). We
306 do observe slightly elevated allele-sharing between the Shum Laka individuals and present-
307 day Grassfields populations, consistent with small proportions of Shum Laka-related admixture
308 (maximum $\sim 7\text{--}8\%$; Supplementary Information section 3). This pattern is reminiscent of pre-
309 vious results for Malawi, where ancient hunter-gatherers from the sites of Hora, Chencherere,
310 and Fingira ($\sim 8000\text{--}2500$ BP) were largely continuous in their ancestry but highly differenti-
311 ated from present-day populations [33] (98% Bantu-associated and 2% southern African hunter-
312 gatherer-related ancestry for Chewa in our extended admixture graph results; $Z = 3.8$ without
313 admixture; Supplementary Information section 3). We also observe an A00 Y chromosome
314 carried by the adolescent boy 2/SE II, suggesting that the concentration of this haplogroup in
315 western Cameroon [25, 26] may have a long history. The phylogenetic position of the Shum
316 Laka A00, well outside of the A00a/b/c clade, additionally implies that A00 may have been
317 more diverse during the LSA and SMA, and it is unlikely to have been introduced to present-
318 day populations by recent archaic introgression. The $\sim 200,000\text{--}300,000$ BP divergence time
319 of A00 from other modern human haplogroups [27, 28] could support its association either with
320 the Central African hunter-gatherer-related ancestry component of the Shum Laka individuals

321 or with the deep modern human portion of their West African-related ancestry.

322 Linguistic and genetic evidence points to western Cameroon as the most likely area for the
323 development of Bantu languages and as the ultimate source of subsequent migrations of Bantu
324 speakers, and while the mid-Holocene archaeological record of the region is sparse, Shum Laka
325 has been speculated to have been an important site in the early phase of this process [8–18].
326 However, the genetic profiles of our four sampled individuals—even by 3000 BP, when the
327 spread of Bantu languages and of ancestry associated with Bantu-speaking populations was al-
328 ready underway—are very different from those of most Niger-Congo speakers today, implying
329 that these individuals are not representative of the primary source population(s) ancestral to
330 present-day Bantu speakers. These results do not contradict a central role for the Grassfields
331 area in the origins of Bantu-speaking peoples, but it may be that multiple, highly differentiated
332 populations formerly lived in the region, with potentially either high or low levels of linguistic
333 diversity. In fact, it would not be surprising if the Shum Laka site itself was used (either suc-
334 cessively or concurrently) by multiple groups with different ancestry, cultural traditions, and/or
335 languages [1], evidence of which may not be visible from the collection of remains as preserved
336 today.

337 **Implications for broader African population history**

338 As in other parts of the world, present-day genetic diversity in Africa has been heavily influ-
339 enced by recent population movements, especially the expansion of Bantu language speakers
340 across the continent [8–15, 56]. For example, among the signals of recent migration and ad-
341 mixture in our results are the Bantu-associated ancestry components in Central African hunter-
342 gatherers (59% in Aka, who speak a Bantu language, and 26% in Mbuti) and likely the East
343 African-related ancestry in Mbuti (who speak both Bantu and Sudanic languages). At the same

344 time, building a phylogenetic model including Shum Laka and other diverse groups, aided by
345 the availability of ancient DNA data from past populations, allows us to gain new insights into
346 more distant relationships [13, 33, 56–60]. In the deepest portion of the model, our findings sup-
347 port previous evidence of archaic ancestry in African populations [40–48], with the particular
348 signal we identify being specific to the West African clade. Among modern humans, mean-
349 while, our results are relevant to open questions about the time depths of different elements of
350 African population structure both today and in the past [61–63].

351 First, we infer a series of closely spaced population splits in our admixture graph model
352 involving West African-related and two East African-related lineages, as well as non-Africans
353 (point (2) in Fig. 4A). Based on the populations involved, the center of this radiation was likely
354 in East Africa (Fig. 4B), with a date of $\sim 60,000$ – $80,000$ BP based on estimated divergences
355 of African and non-African populations [35, 64–67]. The existence of such an expansion is
356 also consistent with human mtDNA phylogeographic patterns—specifically the diversification
357 of haplogroup L3, likely originating in East Africa roughly 70,000 BP [68, 69]—and potentially
358 with the origins of clade CT in the Y chromosome tree at a similar time depth [27, 70].

359 Equally noteworthy is the earliest major phase of diversification in our model, involving at
360 least four lineages early in the history of modern humans (point (1) in Fig. 4A). Recent con-
361 sensus has been that Southern African hunter-gatherers represent the deepest sampled branch
362 of the modern human population tree [32, 35, 36], but we show that the Central African hunter-
363 gatherer clade (which accounts for about a third of the ancestry at Shum Laka) split at close to
364 the same time or perhaps slightly earlier. We also infer the presence of at least one additional
365 deep lineage splitting near the same point and contributing ancestry to West Africans and some
366 East Africans. The signal for East African hunter-gatherers is in line with previous reports
367 of admixture in Hadza and Sandawe from a deeply splitting source [58], but we find that the
368 best fit for the deep ancestry in Hadza and in the ancient Mota individual (as well as in West

369 Africans) is from a source that is not specifically related to either Southern or Central African
370 hunter-gatherers (Supplementary Information section 3). The presence of this “ghost” deep lin-
371 eage [33] contributing to the West African clade (including Shum Laka), separate from Central
372 African hunter-gatherer-related ancestry, is notable in light of the regional Pleistocene archae-
373 ological record, which, although thin [6, 71], includes *Homo sapiens* fossils dated to ~300,000
374 BP in northwestern Africa [72], as well as an individual buried ~12,000 BP in southwestern
375 Nigeria (the oldest known human fossil from West Africa proper) with archaic morphological
376 features [73]. Middle Stone Age artifacts have also been found in parts of West Africa into the
377 terminal Pleistocene [74], despite the development of LSA technologies elsewhere (e.g., Shum
378 Laka). Thus, the available material and fossil evidence is concordant with our genetic results in
379 indicating elements of long-term population structure and admixture [61].

380 Beyond the specific populations involved in the earliest phase, the presence of multiple
381 closely-spaced splits suggests that Southern and Central African hunter-gatherers diverged as
382 part of a large-scale modern human radiation within the continent. Previous estimates, specifi-
383 cally for the split of Southern African hunter-gatherers from other populations, place this period
384 at approximately 200,000–250,000 BP [32, 35, 36]. Further work is necessary to investigate the
385 extent to which this radiation was associated with biological, technological, environmental, or
386 other factors, and whether some of the studied lineages might be further admixed in ways we
387 are not yet able to detect. It is very possible that there was additional broad population structure
388 among early modern humans, including groups only known to us through fossil remains [61, 62],
389 but the persistence of ancestry today (in admixed form) from at least four lineages marks this
390 period as an important one in human evolution.

391 **Methods**

392 **Ancient DNA sample processing**

393 We obtained bone powder from the Shum Laka skeletons (see Supplementary Information sec-
394 tion 1 for more information on the site and burials) by drilling cochlear portions of petrous bone
395 samples in a clean room facility at the Royal Belgian Institute of Natural Sciences. In dedicated
396 clean rooms at Harvard Medical School, we extracted DNA using published protocols [75, 76].
397 From the extracts, we prepared barcoded double-stranded libraries treated with uracil-DNA gly-
398 cosylase (UDG) to reduce the rate of characteristic ancient DNA damage [19, 20] in a modified
399 partial UDG preparation including magnetic bead cleanups [20, 77]. For the SNP capture data,
400 we used two rounds of in-solution target hybridization to enrich for sequences overlapping the
401 mitochondrial genome and approximately 1.2 million genome-wide SNPs [50, 78–81]. We then
402 added 7-base-pair indexing barcodes to the adapters of each library [82] and sequenced on an
403 Illumina NextSeq 500 machine with 76-base-pair paired-end reads. For individuals 2/SE II and
404 4/A, we also generated whole-genome shotgun data from the same libraries but without the
405 target enrichment step. Sequencing was performed at the Broad Institute on an Illumina HiSeq
406 X Ten machine, using 19 lanes for 2/SE II (yielding approximately $18.5\times$ average coverage,
407 including 1,216,658 sites covered from the set of target SNPs used in most analyses) and two
408 lanes for 4/A ($3.9\times$ average coverage, 1,158,884 sites covered).

409 From the raw sequencing results, we retained reads with no more than one mismatch per
410 read pair to the library-specific barcodes. Prior to alignment, we merged paired-end sequences
411 based on forward and reverse mate overlaps and trimmed barcodes and adapters. Preprocessed
412 reads were then mapped to both the mitochondrial reference genome RSRS [69] and the human
413 reference genome (version hg19) using the “samse” command with default parameters in BWA
414 (version 0.6.1) [83]. Duplicate molecules (having the same mapped start and end positions and

415 strand orientation) were removed post-alignment. We filtered the mapped sequences (requiring
416 mapping quality scores of at least 10 for targeted SNP capture and 30 for whole-genome shotgun
417 data) and trimmed two terminal bases to eliminate (almost all) damage-induced errors.

418 For mitochondrial DNA, we called haplogroups using HaploGrep2 [84]. For nuclear DNA
419 obtained from SNP capture and for the whole-genome shotgun data for individual 4/A, we se-
420 lected one allele at random per site to create pseudo-haploid genotypes. For the whole-genome
421 shotgun data for individual 2/SE II, we used a previously described reference-bias-free diploid
422 genotype calling procedure [36], converting resulting genotypes into a fasta-like encoding al-
423 lowing for extraction of data at specified sites via *casertain* and *cTools* [36]. We determined
424 the sex of each individual by examining the fractions of sequences mapping to the X and Y
425 chromosomes [85], and we determined Y-chromosome haplogroups by comparing sequence-
426 level SNP information to the tree established by the International Society of Genetic Genealogy
427 (<http://www.isogg.org>). To ensure authenticity, we computed the proportion of C-to-T errors in
428 terminal positions of sequenced molecules and evaluated possible contamination via heterozy-
429 gosity at variable sites in haploid genome regions, using *contamMix* [78] and *ANGSD* [86] for
430 mtDNA and the X chromosome (in males), respectively.

431 **Radiocarbon dates**

432 At the Pennsylvania State University (PSU) Radiocarbon Laboratory, we generated direct ra-
433 diocarbon dates via accelerator mass spectrometry (AMS) for the four analyzed individuals,
434 using fragments of the same temporal bone portions that were sampled for ancient DNA. The
435 resulting dates are in good agreement with previously reported direct dates for different bones
436 from individuals 2/SE II (8160–7790 cal BP, 7150 ± 70 BP, OxA-5203) and 4/A (3380–3010
437 cal BP, 3045 ± 60 BP, OxA-5205) [1]. We performed calibrations using *OxCal* [87] version
438 4.3.2 with a mixture of the *IntCal13* [88] and *SHCal13* [89] curves, specifying “U(0,100)” to

439 allow for a flexible combination [87, 90], and rounding final results to the nearest 10 years (see
440 also Supplementary Information section 1).

441 **New present-day data**

442 We generated genome-wide SNP genotype data for 63 individuals from five present-day Cameroo-
443 nian populations on the Human Origins array: Aghem (28), Bafut (11), Bakoko (1), Bangwa
444 (2), and Mbo (21) (Extended Data Table 1; Supplementary Table 3). Samples were collected
445 with informed consent, with collection and analysis of samples approved by the UCL/UCLH
446 Committee on the Ethics of Human Research, Committee A and Alpha.

447 **A00 Y chromosome split time estimation**

448 To estimate the split time of the Shum Laka A00 Y chromosome, we called genotypes for in-
449 dividual 2/SE II at a set of positions where sequences from two present-day individuals with
450 haplogroup A00 [27] differ from all non-A00 individuals. To avoid needing to determine the
451 status of mutations as ancestral or derived, we considered the entire unrooted lineage specific
452 to A00 (see Fig. 1). The total time span represented by this lineage is approximately 359,000
453 years, using published values of $\sim 275,000$ BP for the divergence of the A00 lineage from other
454 modern human haplogroups [28] and $\sim 191,000$ BP for the next-oldest split within macrohap-
455 logroup A [29]. With a requirement of at least 90% agreement among the reads at each site,
456 we called 1521 positions as having the alternative allele (i.e., matching present-day A00 and
457 differing from the human reference sequence) and 145 as having the reference allele (taking
458 the average of 143 and 147 for the two present-day individuals). The fraction $145/(145+1521)$
459 then defines the position of the Shum Laka split along the (unrooted) A00 lineage. We note that
460 split times computed either from all sites (relaxing the 90% threshold and using the majority

461 allele), or from additionally requiring at least two reads per site, differ from our primary esti-
462 mate by only a few hundred years. To produce a confidence interval, we used the variance in
463 the published estimates and assumed an independent Poisson sampling error for the number of
464 observed reference alleles.

465 **PCA and allele-sharing statistics**

466 We performed PCA using smartpca (with the “lsqproject” and “autoshrink” options) [91, 92]
467 and computed f_4 -statistics using ADMIXTOOLS (with standard errors estimated via block
468 jackknife) [31]. We projected all ancient individuals in PCA rather than using them to com-
469 pute axes in order to avoid artifacts caused by missing data. In each PCA, we also projected
470 a subset of the present-day populations to allow controlled comparisons with ancient individ-
471 uals. In most cases, reported f_4 -statistics are based on the approximately 1.15M autosomal
472 SNPs from our target capture set. For PCA and for f_4 -statistics testing differential relatedness
473 to Shum Laka, we used autosomal SNPs from the Human Origins array (a subset of the target
474 capture set), with some populations in the analyses only genotyped on this subset (see Extended
475 Data Table 1). For these latter f_4 -statistics, we excluded for all populations a set of roughly 40k
476 SNPs having high missingness in the present-day Cameroon data.

477 **Admixture graphs**

478 We fit admixture graphs with the ADMIXTUREGRAPH (qpGraph) program in ADMIXTOOLS
479 (with the options “outpop: NULL,” “lambdascale: 1,” “inbreed: YES,” and “diag: 0.0001”) [31,
480 51, 93], using the 1.15M autosomal SNPs from our target capture set by default, and other sets
481 of SNPs in alternative model versions as specified. The program requires as input the branch-
482 ing order of the populations in the graph and a list of admixture events, and it then solves

483 for the optimal parameters of the model (branch lengths and mixture proportions) via an ob-
484 jective function measuring the deviation between predicted and observed values of a basis set
485 of f -statistics. From the inferred parameters, poorly fitting topologies (including positions of
486 admixture sources) can be corrected by changing split orders at internal nodes that appear as tri-
487 furcations under the constraints enforced by the input (see Supplementary Information section
488 3).

489 To evaluate the fit quality of output models, we employed two metrics: first, a list of resid-
490 ual Z -scores for all f -statistics relating the populations in the graph, and second, a combined
491 approximate log-likelihood score. The first metric is useful for identifying particularly poorly
492 fitting models and the elements that are most responsible for the poor fits, while the second
493 provides a means for comparing the overall fits of separate models (Supplementary Information
494 section 3). In order to assess the degree of constraint on individual parameter inferences, we
495 were guided primarily by the variability across different model versions (using different pop-
496 ulations and SNP sets; see Extended Data Table 3 and Supplementary Information section 3),
497 which reflects both statistical uncertainty and changes in model-specific assumptions.

498 **Data availability**

499 The aligned sequences are available through the European Nucleotide Archive under accession
500 number PRJEB32086. Genotype data used in analysis are available at <https://reich.hms.harvard.edu/datasets>.

501 **References**

- 502 1. de Maret, P. Shum Laka (Cameroon): human burials and general perspectives. In Pwiti,
503 G. & Soper, R. (eds.) *Aspects of African Archaeology: Papers from the 10th Congress*

- 504 of the Panafrican Association of Prehistory and Related Studies, 274–279 (University of
505 Zimbabwe Publications, Harare, 1996).
- 506 2. Lavachery, P. *De la pierre au métal: archéologie des dépôts holocènes de l’abri de Shum*
507 *Laka (Cameroun)*. Ph.D. thesis, Université libre de Bruxelles (1997).
- 508 3. Ribot, I., Orban, R. & de Maret, P. The Prehistoric Burials of Shum Laka Rockshelter
509 (North-West Cameroon). In *Annales du Musée Royal de l’Afrique Centrale*, vol. 164
510 (Musée Royal de l’Afrique Centrale, Tervuren, Belgium, 2001).
- 511 4. Lavachery, P. The Holocene archaeological sequence of Shum Laka rock shelter (Grass-
512 fields, western Cameroon). *African Arch. Rev.* **18**, 213–247 (2001).
- 513 5. de Maret, P. Archaeologies of the Bantu expansion. In Mitchell, P. & Lane, P. (eds.) *The*
514 *Oxford Handbook of African Archaeology*, 319–328 (Oxford University Press, 2013).
- 515 6. Cornelissen, E. Hunting and gathering in Africa’s tropical forests at the end of the Pleis-
516 tocene and in the early Holocene. In Mitchell, P. & Lane, P. (eds.) *The Oxford Handbook*
517 *of African Archaeology*, 319–328 (Oxford University Press, 2013).
- 518 7. de Maret, P., Clist, B. & Van Neer, W. Résultats des premières fouilles dans les abris
519 de Shum Laka et d’Abeke au Nord-Ouest du Cameroun. *L’Anthropologie* **91**, 559–584
520 (1987).
- 521 8. Vansina, J. New linguistic evidence and ‘the Bantu expansion’. *The Journal of African*
522 *History* **36**, 173–195 (1995).
- 523 9. Ehret, C. Bantu expansions: re-envisioning a central problem of early African history.
524 *Int. J. African Hist. Stud.* **34**, 5–41 (2001).

- 525 10. Holden, C. J. Bantu language trees reflect the spread of farming across sub-Saharan
526 Africa: a maximum-parsimony analysis. *Proc. R. Soc. London B* **269**, 793–799 (2002).
- 527 11. Phillipson, D. W. Language and farming dispersals in sub-Saharan Africa, with par-
528 ticular reference to the Bantu-speaking peoples. In Bellwood, P. & Renfrew, C. (eds.)
529 *Examining the farming/language dispersal hypothesis*, 177–187 (McDonald Institute for
530 Archaeological Research, 2002).
- 531 12. Bellwood, P. *First farmers: The origins of agricultural societies* (Blackwell, Oxford,
532 2005).
- 533 13. Tishkoff, S. A. *et al.* The genetic structure and history of Africans and African Ameri-
534 cans. *Science* **324**, 1035–1044 (2009).
- 535 14. Berniell-Lee, G. *et al.* Genetic and demographic implications of the Bantu expansion:
536 insights from human paternal lineages. *Mol. Biol. Evol.* **26**, 1581–1589 (2009).
- 537 15. Russell, T., Silva, F. & Steele, J. Modelling the spread of farming in the Bantu-speaking
538 regions of Africa: an archaeology-based phylogeography. *PLoS One* **9**, e87854 (2014).
- 539 16. Bostoen, K. *et al.* Middle to late Holocene Paleoclimatic change and the early Bantu
540 expansion in the rain forests of Western Central Africa. *Curr. Anthropol.* **56**, 354–384
541 (2015).
- 542 17. Grollemund, R. *et al.* Bantu expansion shows that habitat alters the route and pace of
543 human dispersals. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 13296–13301 (2015).
- 544 18. Bostoen, K. The Bantu Expansion. In *Oxford Research Encyclopedia of African History*
545 (Oxford University Press, 2018).

- 546 19. Briggs, A. W. *et al.* Removal of deaminated cytosines and detection of in vivo methyla-
547 tion in ancient DNA. *Nucleic Acids Res.* **38**, e87 (2010).
- 548 20. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-
549 DNA-glycosylase treatment for screening of ancient DNA. *Phil. Trans. R. Soc. B* **370**,
550 20130624 (2015).
- 551 21. Giresse, P., Maley, J. & Brenac, P. Late Quaternary palaeoenvironments in the Lake
552 Barombi Mbo (West Cameroon) deduced from pollen and carbon isotopes of organic
553 matter. *Palaeogeography, Palaeoclimatology, Palaeoecology* **107**, 65–78 (1994).
- 554 22. Gonder, M. K., Mortensen, H. M., Reed, F. A., de Sousa, A. & Tishkoff, S. A. Whole-
555 mtDNA genome sequence analysis of ancient African lineages. *Mol. Biol. Evol.* **24**,
556 757–768 (2006).
- 557 23. Batini, C. *et al.* Phylogeography of the human mitochondrial L1c haplogroup: genetic
558 signatures of the prehistory of Central Africa. *Mol. Phyl. Evol.* **43**, 635–644 (2007).
- 559 24. Wood, E. T. *et al.* Contrasting patterns of Y chromosome and mtDNA variation in Africa:
560 evidence for sex-biased demographic processes. *Eur. J. Hum. Genet.* **13**, 867 (2005).
- 561 25. Mendez, F. L. *et al.* An African American paternal lineage adds an extremely ancient
562 root to the human Y chromosome phylogenetic tree. *Am. J. Hum. Genet.* **92**, 454–459
563 (2013).
- 564 26. Krahn, T., Schrack, B., Fomine, F. L. M. & Krahn, A.-M. Searching for our most distant
565 (paternal) cousins in Cameroon. Institute for Genetic Genealogy 2016 Conference, San
566 Diego (2016).

- 567 27. Karmin, M. *et al.* A recent bottleneck of Y chromosome diversity coincides with a global
568 change in culture. *Genome Res.* **25**, 459–466 (2015).
- 569 28. Mendez, F. L., Poznik, G. D., Castellano, S. & Bustamante, C. D. The divergence
570 of Neandertal and modern human Y chromosomes. *Am. J. Hum. Genet.* **98**, 728–734
571 (2016).
- 572 29. Jobling, M. A. & Tyler-Smith, C. Human Y-chromosome variation in the genome-
573 sequencing era. *Nat. Rev. Genet.* **18**, 485 (2017).
- 574 30. Fan, S. *et al.* African evolutionary history inferred from whole genome sequence data of
575 44 indigenous African populations. *Genome Biol.* **20**, 82 (2019).
- 576 31. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065–1093
577 (2012).
- 578 32. Schlebusch, C. M. *et al.* Southern African ancient genomes estimate modern human
579 divergence to 350,000 to 260,000 years ago. *Science* **358**, 652–655 (2017).
- 580 33. Skoglund, P. *et al.* Reconstructing prehistoric African population structure. *Cell* **171**,
581 59–71 (2017).
- 582 34. Gallego Llorente, M. *et al.* Ancient Ethiopian genome reveals extensive Eurasian admix-
583 ture in Eastern Africa. *Science* **350**, 820–822 (2015).
- 584 35. Gronau, I., Hubisz, M., Gulko, B., Danko, C. & Siepel, A. Bayesian inference of ancient
585 human demography from individual genome sequences. *Nat. Genet.* **43**, 1031–1034
586 (2011).
- 587 36. Mallick, S. *et al.* The Simons Genome Diversity Project: 300 genomes from 142 diverse
588 populations. *Nature* **538**, 201–206 (2016).

- 589 37. van de Loosdrecht, M. *et al.* Pleistocene North African genomes link Near Eastern and
590 sub-Saharan African human populations. *Science* **360**, 548–552 (2018).
- 591 38. Patin, E. *et al.* Inferring the demographic history of African farmers and Pygmy hunter-
592 gatherers using a multilocus resequencing data set. *PLoS Genet.* **5**, e1000448 (2009).
- 593 39. Loh, P.-R. *et al.* Inferring admixture histories of human populations using linkage dise-
594 quilibrium. *Genetics* **193**, 1233–1254 (2013).
- 595 40. Plagnol, V. & Wall, J. D. Possible ancestral structure in human populations. *PLoS Genet.*
596 **2**, e105 (2006).
- 597 41. Hammer, M. F., Woerner, A. E., Mendez, F. L., Watkins, J. C. & Wall, J. D. Genetic
598 evidence for archaic admixture in Africa. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 15123–
599 15128 (2011).
- 600 42. Hsieh, P. *et al.* Model-based analyses of whole-genome data reveal a complex evolution-
601 ary history involving archaic introgression in Central African Pygmies. *Genome Res.* **26**,
602 291–300 (2016).
- 603 43. Durvasula, A. & Sankararaman, S. Recovering signals of ghost archaic admixture in the
604 genomes of present-day Africans. *bioRxiv* preprint 285734 (2018).
- 605 44. Hey, J. *et al.* Phylogeny estimation by integration over isolation with migration models.
606 *Mol. Biol. Evol.* **35**, 2805–2818 (2018).
- 607 45. Ragsdale, A. P. & Gravel, S. Models of archaic admixture and recent history from two-
608 locus statistics. *bioRxiv* preprint 489401 (2018).
- 609 46. Speidel, L., Forest, M., Shi, S. & Myers, S. A method for genome-wide genealogy
610 estimation for thousands of samples. *bioRxiv* preprint 550558 (2019).

- 611 47. Wang, K., Mathieson, I., O'Connell, J. & Schiffels, S. Tracking human population
612 structure through time from whole genome sequences. *bioRxiv* preprint 585265 (2019).
- 613 48. Lorente-Galdos, B. *et al.* Whole-genome sequence analysis of a Pan African set of
614 samples reveals archaic gene flow from an extinct basal population of modern humans
615 into sub-Saharan populations. *Genome Biol.* **20**, 77 (2019).
- 616 49. Brandt, G. *et al.* Ancient DNA reveals key stages in the formation of Central European
617 mitochondrial genetic diversity. *Science* **342**, 257–261 (2013).
- 618 50. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European
619 languages in Europe. *Nature* **522**, 207–211 (2015).
- 620 51. Lipson, M. *et al.* Parallel palaeogenomic transects reveal complex genetic history of
621 early European farmers. *Nature* **551**, 368–372 (2017).
- 622 52. Huysecom, E. *et al.* The emergence of pottery in Africa during the 10th millennium
623 calBC: new evidence from Ounjougou (Mali). *Antiquity* **83**, 905–917 (2009).
- 624 53. Gasse, F. Hydrological changes in the African tropics since the Last Glacial Maximum.
625 *Quat. Sci. Rev.* **19**, 189–211 (2000).
- 626 54. Henn, B. M. *et al.* Genomic ancestry of North Africans supports back-to-Africa migra-
627 tions. *PLoS Genet.* **8**, e1002397 (2012).
- 628 55. Triska, P. *et al.* Extensive admixture and selective pressure across the Sahel belt. *Genome*
629 *Biol. Evol.* **7**, 3484–3495 (2015).
- 630 56. Patin, E. *et al.* Dispersals and genetic adaptation of Bantu-speaking populations in Africa
631 and North America. *Science* **356**, 543–546 (2017).

- 632 57. Schlebusch, C. M. *et al.* Genomic variation in seven Khoe-San groups reveals adaptation
633 and complex African history. *Science* **338**, 374–379 (2012).
- 634 58. Pickrell, J. *et al.* The genetic prehistory of southern Africa. *Nat. Comm.* **3** (2012).
- 635 59. Busby, G. B. *et al.* Admixture into and within sub-Saharan Africa. *eLife* **5**, e15266
636 (2016).
- 637 60. Choudhury, A., Aron, S., Sengupta, D., Hazelhurst, S. & Ramsay, M. African genetic
638 diversity provides novel insights into evolutionary history and local adaptations. *Hum.*
639 *Mol. Genet.* **27**, R209–R218 (2018).
- 640 61. Scerri, E. M. *et al.* Did our species evolve in subdivided populations across Africa, and
641 why does it matter? *Trends Ecol. Evol.* **33**, 582–584 (2018).
- 642 62. Henn, B. M., Steele, T. E. & Weaver, T. D. Clarifying distinct models of modern human
643 origins in Africa. *Curr. Opinion Genet. Devel.* **53**, 148–156 (2018).
- 644 63. Klein, R. G. Population structure and the evolution of Homo sapiens in Africa. *Evol.*
645 *Anthropol.* DOI: 10.1002/evan.21788 (2019).
- 646 64. Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H. & Bustamante, C. D. Infer-
647 ring the joint demographic history of multiple populations from multidimensional SNP
648 frequency data. *PLoS Genet.* **5**, e1000695 (2009).
- 649 65. Laval, G., Patin, E., Barreiro, L. & Quintana-Murci, L. Formulating a historical and de-
650 mographic model of recent human evolution based on resequencing data from noncoding
651 regions. *PLoS ONE* **5**, e10284 (2010).
- 652 66. Gravel, S. *et al.* Demographic history and rare allele sharing among human populations.
653 *Proc. Natl. Acad. Sci. U. S. A.* **108**, 11983–11988 (2011).

- 654 67. Schiffels, S. & Durbin, R. Inferring human population size and separation history from
655 multiple genome sequences. *Nat. Genet.* **46**, 919–925 (2014).
- 656 68. Soares, P. *et al.* The expansion of mtDNA haplogroup L3 within and out of Africa. *Mol.*
657 *Biol. Evol.* **29**, 915–927 (2012).
- 658 69. Behar, D. M. *et al.* A “Copernican” reassessment of the human mitochondrial DNA tree
659 from its root. *Am. J. Hum. Genet.* **90**, 675–684 (2012).
- 660 70. Poznik, G. D. *et al.* Punctuated bursts in human male demography inferred from 1,244
661 worldwide Y-chromosome sequences. *Nat. Genet.* **48**, 593 (2016).
- 662 71. Scerri, E. The Stone Age archaeology of West Africa. In *Oxford Research Encyclopedia*
663 *of African History* (Oxford University Press, 2017).
- 664 72. Hublin, J.-J. *et al.* New fossils from Jebel Irhoud, Morocco and the pan-African origin
665 of *Homo sapiens*. *Nature* **546**, 289–292 (2017).
- 666 73. Harvati, K. *et al.* The later Stone Age calvaria from Iwo Eleru, Nigeria: morphology and
667 chronology. *PLoS One* **6**, e24024 (2011).
- 668 74. Scerri, E. M., Blinkhorn, J., Niang, K., Bateman, M. D. & Groucutt, H. S. Persistence of
669 middle stone age technology to the pleistocene/holocene transition supports a complex
670 hominin evolutionary scenario in west africa. *J Archaeol. Sci. Rep.* **11**, 639–646 (2017).
- 671 75. Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave
672 bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U. S. A.* **110**,
673 15758–15763 (2013).
- 674 76. Korlević, P. *et al.* Reducing microbial and human contamination in DNA extractions
675 from ancient bones and teeth. *BioTechniques* **59**, 87–93 (2015).

- 676 77. Lipson, M. *et al.* Ancient genomes document multiple waves of migration in Southeast
677 Asian prehistory. *Science* **361**, 92–95 (2018).
- 678 78. Fu, Q. *et al.* DNA analysis of an early modern human from Tianyuan Cave, China. *Proc.*
679 *Natl. Acad. Sci. U. S. A.* **110**, 2223–2227 (2013).
- 680 79. Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal ancestor.
681 *Nature* **524**, 216–219 (2015).
- 682 80. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature*
683 **528**, 499–503 (2015).
- 684 81. Lazaridis, I. *et al.* Genomic insights into the origin of farming in the ancient Near East.
685 *Nature* **536**, 419–424 (2016).
- 686 82. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in mul-
687 tiplex sequencing on the Illumina platform. *Nucl. Acids Res.* **40**, e3 (2011).
- 688 83. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler trans-
689 form. *Bioinformatics* **26**, 589–595 (2010).
- 690 84. Weissensteiner, H. *et al.* HaploGrep 2: Mitochondrial haplogroup classification in the
691 era of high-throughput sequencing. *Nucleic Acids Res.* **44**, W58–W63 (2016).
- 692 85. Skoglund, P., Storå, J., Götherström, A. & Jakobsson, M. Accurate sex identification of
693 ancient human remains using DNA shotgun sequencing. *J Archaeol. Sci.* **40**, 4477–4482
694 (2013).
- 695 86. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of next generation
696 sequencing data. *BMC Bioinformatics* **15**, 356 (2014).

- 697 87. Ramsey, C. B. & Lee, S. Recent and planned developments of the program OxCal.
698 *Radiocarbon* **55**, 720–730 (2013).
- 699 88. Reimer, P. J. *et al.* IntCal13 and Marine13 radiocarbon age calibration curves 0-50,000
700 years cal BP. *Radiocarbon* **55**, 1869–1887 (2013).
- 701 89. Hogg, A. G. *et al.* SHCal13 Southern Hemisphere calibration, 0–50,000 years cal BP.
702 *Radiocarbon* **55**, 1889–1903 (2013).
- 703 90. Marsh, E. J. *et al.* IntCal, SHCal, or a Mixed Curve? Choosing a 14C Calibration Curve
704 for Archaeological and Paleoenvironmental Records from Tropical South America. *Ra-*
705 *diocarbon* **60**, 925–940 (2018).
- 706 91. Patterson, N., Price, A. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.*
707 **2**, e190 (2006).
- 708 92. Liu, L. T., Dobriban, E. & Singer, A. ePCA: High dimensional exponential family PCA.
709 <https://arxiv.org/abs/1611.05550> (2016).
- 710 93. Lipson, M. & Reich, D. A working model of the deep relationships of diverse modern
711 human genetic lineages outside of Africa. *Mol. Biol. Evol.* **34**, 889–902 (2017).
- 712 94. Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Moun-
713 tains. *Nature* **505**, 43–49 (2014).
- 714 95. Moeyersons, J., Cornelissen, E., Lavachery, P. & Doutrelepont, H. L’abri sous-roche de
715 Shum Laka (Cameroun Occidental): Données climatologiques et occupation humaine
716 depuis 30.000 ans. *Géo-Eco-Trop* **20**, 39–60 (1996).

717 **Acknowledgments**

718 We thank Iosif Lazaridis, Vagheesh Narasimhan, and Kendra Sirak for discussions and com-
719 ments; Monika Karmin for help with Y chromosome data; Laurie Eccles for help with radio-
720 carbon dating; Brad Erkkila for help with isotopic analysis; Rebecca Bernardos, Matthew Mah,
721 and Zhao Zhang for other technical assistance; and Jean-Pierre Warnier for his role in locating
722 the site of Shum Laka. The Shum Laka excavations were supported by the Belgian Fund for
723 Scientific Research (FNRS), the Université Libre de Bruxelles, the Royal Museum for Central
724 Africa, and the Leakey Foundation. The collection of samples from present-day individuals in
725 Cameroon was supported by Neil Bradman and the Melford Charitable Trust. I.R. was sup-
726 ported by a Université de Montréal exploration grant (2018-2020). M.G.T. was supported by
727 Wellcome Trust Senior Investigator Award Grant 100719/Z/12/Z. G.H. was supported by a Sir
728 Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (grant
729 number 098386/Z/12/Z). M.E.P. was supported by a fellowship from the Radcliffe Institute for
730 Advanced Study at Harvard University during the development of this project. D.R. was sup-
731 ported by the National Institutes of Health (NIGMS GM100233) and by an Allen Discovery
732 Center grant, and is an Investigator of the Howard Hughes Medical Institute.

733 **Author contributions**

734 N.R., G.H., M.E.P., and D.R. supervised the study. I.R., R.N.A., H.B., E.C., I.C., P.d.M., P.L.,
735 C.M.M., R.O., E.S., P.S., W.V.N., C.L.-F., S.Mac., and M.E.P. provided samples and assem-
736 bled archaeological and anthropological materials and information. S.L., N.Bra., F.L.M.F.,
737 M.G.T., K.V., and G.H. provided data from present-day populations. S.Mal., N.R., N.A.,
738 N.Bro., A.M.L., J.O., K.S., and D.R. performed ancient DNA laboratory and data-processing
739 work. B.J.C. and D.J.K. performed radiocarbon analysis. M.L., S.Mal., I.O., N.P., and D.R.

740 analyzed genetic data. M.L., I.R., H.B., E.S., C.L.-F., S.Mac., M.E.P., and D.R. wrote the
741 manuscript.

742 **Author information**

743 Reprints and permissions information is available at www.nature.com/reprints. The authors
744 declare no competing financial interests. Correspondence and requests for materials should be
745 addressed to M.L. (mlipson@genetics.med.harvard.edu).

Extended Data Table 1. Populations used in the study

Population	Country	Language family	Date	Sample size	Data type	Reference
Shum Laka	Cameroon		~8000–3000 BP	4/1/1	1240k/DG/SG	This paper
Ancient Malawi HG	Malawi		~8100–2500 BP	7*	1240k	[33]
Mota	Ethiopia		~4500 BP	1	SG	[34]
Ancient South African HG	South Africa		~2000 BP	3 [†]	SG	[32, 33]
Taforalt	Morocco		~15,000–14,000 BP	6	1240k	[37]
Altai Neanderthal	Russia		~120,000 BP	1	DG	[94]
Aghem	Cameroon	NC	Present	28	HO	This paper
Bafut	Cameroon	NC	Present	11	HO	This paper
Baka	Cameroon	NC	Present	2	DG	[30]
Bakoko	Cameroon	NC	Present	1	HO	This paper
Bakola	Cameroon	NC	Present	2	DG	[30]
Bangwa	Cameroon	NC	Present	2	HO	This paper
Bedzan	Cameroon	NC	Present	2	DG	[30]
Fulani	Cameroon	NC	Present	2	DG	[30]
Lemande	Cameroon	NC	Present	2	DG	[36]
Mada	Cameroon	AA	Present	2	DG	[30]
Mbo	Cameroon	NC	Present	21	HO	This paper
Ngumba	Cameroon	NC	Present	2	DG	[30]
Tikar	Cameroon	NC	Present	2	DG	[30]
Agaw	Ethiopia	AA	Present	2	DG	[30]
Aka (Biaka)	Central African Republic	NC	Present	20/2	HO/DG	[33, 36]
Chewa	Malawi	NC	Present	11	HO	[33]
Dinka	Sudan	NS	Present	7/4	HO/DG	[33, 36]
French	France	IE	Present	3	DG	[36]
Hadza	Tanzania	KS	Present	5(2)/1	HO/DG	[33, 36]
Han	China	ST	Present	4	DG	[36]
Herero	Namibia	NC	Present	2	DG	[36]
Khoesan	Namibia	KS	Present	22	HO	[33]
Mbuti	DR Congo	NC, NS	Present	10/4	HO/DG	[33, 36]
Mende	Sierra Leone	NC	Present	8/2	HO/DG	[33, 36]
Mursi	Ethiopia	NS	Present	2	DG	[30]
Sandawe	Tanzania	KS	Present	22	HO	[33]
Somali	Kenya	AA	Present	13	HO	[33]
Yoruba	Nigeria	NC	Present	70/3	HO/DG	[33, 36]

List of populations used in analyses in the study. Data types are in-solution targeted SNP capture (1240k), whole-genome sequence with pseudo-haploid genotype calls (SG), high-coverage whole-genome sequence with diploid genotype calls (DG), and Human Origins SNP array (HO). For some populations, we used different sample sets for different analyses, indicated by slashes; Human Origins array genotyped individuals were used for PCA and for f -statistics testing differential relatedness to Shum Laka (Fig. 3B, Extended Data Fig. 3B). For Hadza, we used five individuals with Human Origins data for PCA and two of those five individuals for admixture graph modeling. HG, hunter-gatherers; AA, Afroasiatic; IE, Indo-European; KS, Khoesan; NC, Niger-Congo; NS, Nilo-Saharan; ST, Sino-Tibetan.

*Individuals from Hora, Chencherere, and Fingira.

[†]Individuals from Ballito Bay (A and B) and St. Helena Bay.

Extended Data Table 2. Allele-sharing statistics for deep ancestry

Test pop	$f_4(X, \text{Mursi}; \text{SA}, \text{Han})$		$f_4(X, \text{Mota}; \text{SA}, \text{Han})$		$f_4(X, \text{Han}; \text{SA}, \text{Mursi})$		$f_4(X, \text{Mota}; \text{SA}, \text{Mursi})$	
	Value	Z-score	Value	Z-score	Value	Z-score	Value	Z-score
Dinka	1.4	5.8	-2.0	-5.5	0.1	0.2	-6.3	-20.2
Mota	3.4	9.0	0	0	6.3	18.1	0	0
Hadza	4.1	10.3	0.8	1.7	7.3	21.2	1.0	2.7
Yoruba	4.7	17.8	1.3	3.8	5.2	18.2	-1.1	-3.5
Lemande	5.0	16.8	1.7	4.5	5.7	18.2	-0.6	-2.1
Mende	5.7	19.1	2.3	6.3	6.3	20.0	0	0
Shum Laka	11.7	38.7	8.3	22.6	12.7	40.8	6.4	20.5
Aka	13.3	39.1	9.9	25.2	13.6	40.4	7.3	22.0
Mbuti	16.4	50.4	13.0	34.9	16.4	49.9	10.0	31.8
Mursi	0	0	-3.4	-9.0
Agaw	0.1	0.3	-6.2	-18.9
SA
Test pop	$f_4(X, \text{Mursi}; \text{SA}, \text{Mota})$		$f_4(X, \text{Han}; \text{SA}, \text{Mota})$		$f_4(X, \text{Han}; \text{SA}, \text{Yor})$		$f_4(X, \text{Mursi}; \text{Chimp}, \text{Yor})$	
	Value	Z-score	Value	Z-score	Value	Z-score	Value	Z-score
Dinka	0.8	3.3	3.7	11.9	-0.7	-2.8	-0.9	-4.7
Mota	5.7	18.1	5.2	17.7
Hadza	4.1	11.5	7.0	17.7	4.8	15.2	3.4	11.4
Yoruba	4.1	15.7	7.1	21.6
Lemande	4.1	14.5	7.1	21.0
Mende	4.8	17.3	7.8	22.5
Shum Laka	9.1	29.8	12.0	33.7	8.0	28.7	8.3	31.9
Aka	10.3	33.4	13.2	35.5	7.8	24.8	8.5	30.1
Mbuti	12.5	41.8	15.5	44.1	11.6	40.8	11.8	46.3
Mursi	0	0	3.0	8.8	0.6	2.2	0	0
Agaw	-2.4	-7.7	0.6	1.8	0	0.2	-0.2	-0.9
SA	20.3	66.0

Variations of allele-sharing statistics (multiplied by 1000) sensitive to ancestry in the test population X from a deeply-splitting lineage, along with Z -scores for difference from zero. We note that the zero level has a different meaning depending on which population is in the second position in the statistic. Blank entries are statistics that are confounded by specific relationships between the test population and one of the reference populations (in the third or fourth position; either duplication of the same group, Agaw with Han due to non-African-related ancestry, or Yoruba with other West Africans). SA, ancient South African hunter-gatherers; Yor, Yoruba.

Extended Data Table 3. Admixture graph parameter estimates

Model version:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Mixture proportions (%)																							
Shum Laka basal WA	64	66	62	71	64	58	63	61	63	61	64	64	64	64	63	63	63	..	64	61	69	63	67/62*
Aka Bantu-associated	59	59	57	63	59	56	58	57	59	58	59	59	59	59	59	59	58	58	59	58	62	61	59
Mbuti Bantu-associated	26	24	33	19	28	27	26	12	28	30	32	25	24	26	29	28	35	35	25	35	23	36	27
Mbuti East African-related	17	19	10	27	14	9	16	23	15	13	11	19	20	18	13	14	6	6	18	9	23	8	16
West African clade archaic	2	2	4	4	3	3	3	2	2	2	3	2	2	2	3	3	3	2	2
West African clade deep modern human	10	9	17	8	12	29	15	24	11	18	19	9	8	9	14	13	29	29	11
Mende deep ancestry	4	4	4	3	4	3	4	6	5	5	5	4	4	4	5	5	5	5	4	4	4	3	4
Mota deep ancestry	29	29	30	29	30	31	31	30	29	31	29	29	29	28	30	31	29	29	29	30	27	26	29
Branch lengths																							
Basal WA split [†]	2	3	3	3	3	1	3	2	2	2	2	2	3	3	2	2	3	..	2	3	3	1	3
South African HG split [‡]	1	1	0	4	1	-1	1	2	1	1	1	1	1	1	1	1	1	1	1	0	4	0	1
Ghost modern human split [#]	1	1	1	-3	1	1	0	-2	1	0	-1	1	1	1	0	1	1	1	2

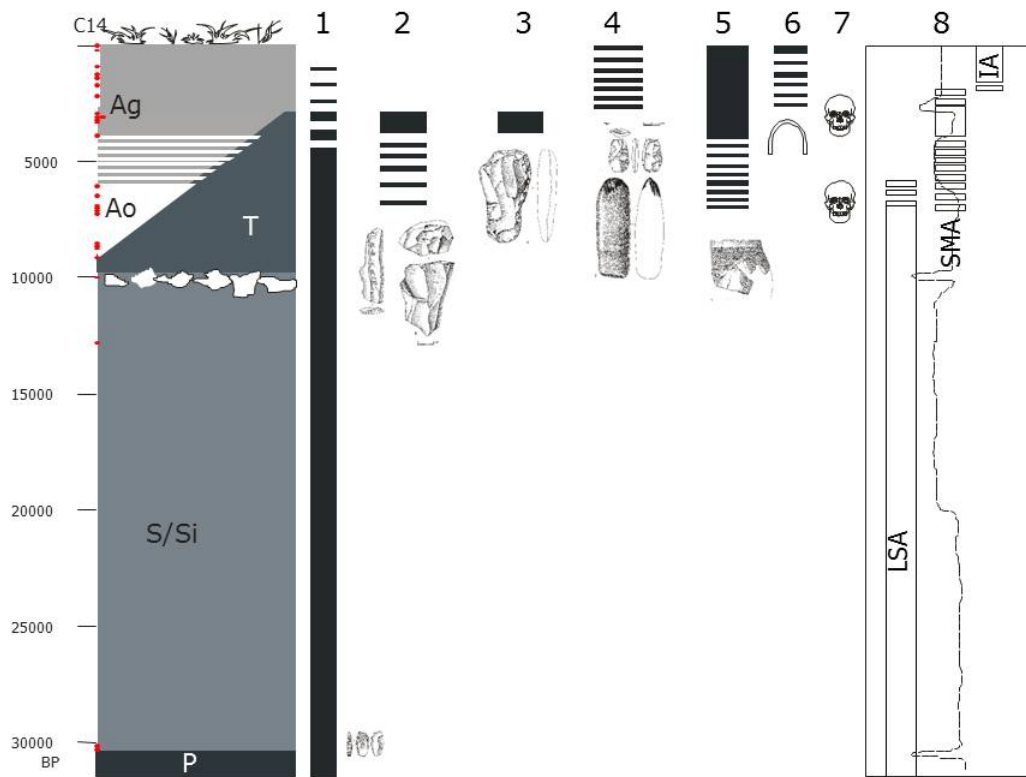
Key admixture graph parameter estimates across different model versions (see Supplementary Information section 3 for full details): 1, primary model; 2, no “dummy” admixture; 3, African-ascertained SNPs; 4, transversion SNPs; 5, Shum Laka whole-genome sequence data; 6, outgroup-ascertained transversions; 7, Hadza added; 8, Mbo in place of Lemande; 9, Herero added; 10, Chewa added; 11, Mursi in place of Agaw; 12, Baka added; 13, Bakola added; 14, Bedzan added; 15, Mada added; 16, Fulani added; 17, Taforalt added; 18, alternative admixture for Shum Laka; 19, alternative deep source; 20, alternative deep source with African-ascertained SNPs; 21, alternative deep source with transversion SNPs; 22, alternative deep source with outgroup-ascertained transversions; 23, Shum Laka pairs fit separately. HG, hunter-gatherers.

*Earlier pair/later pair

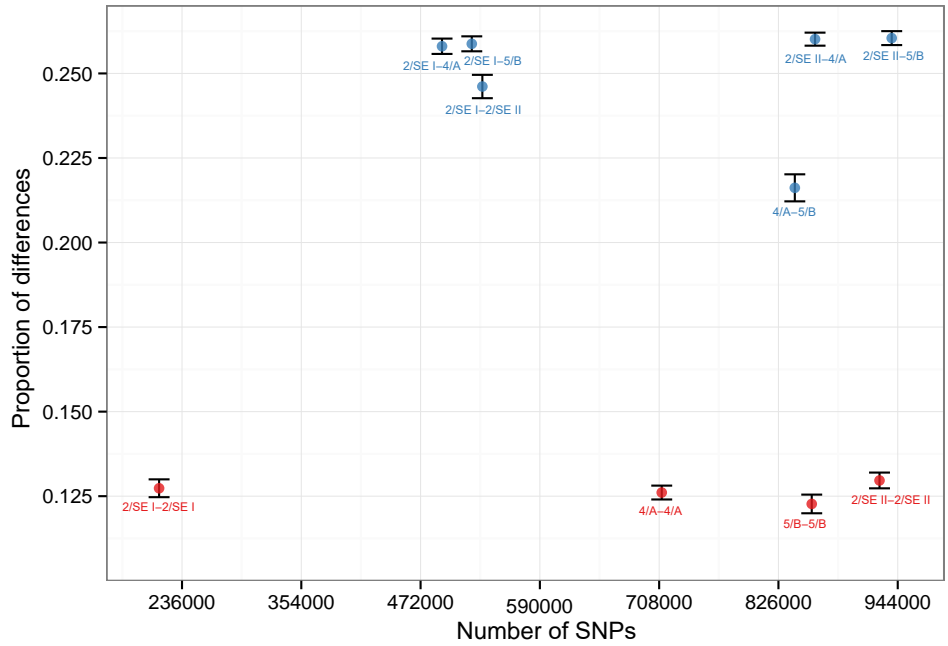
[†]Units above the main West African clade

[‡]Units below the split of the Central African hunter-gather lineage (negative value indicates distance above)

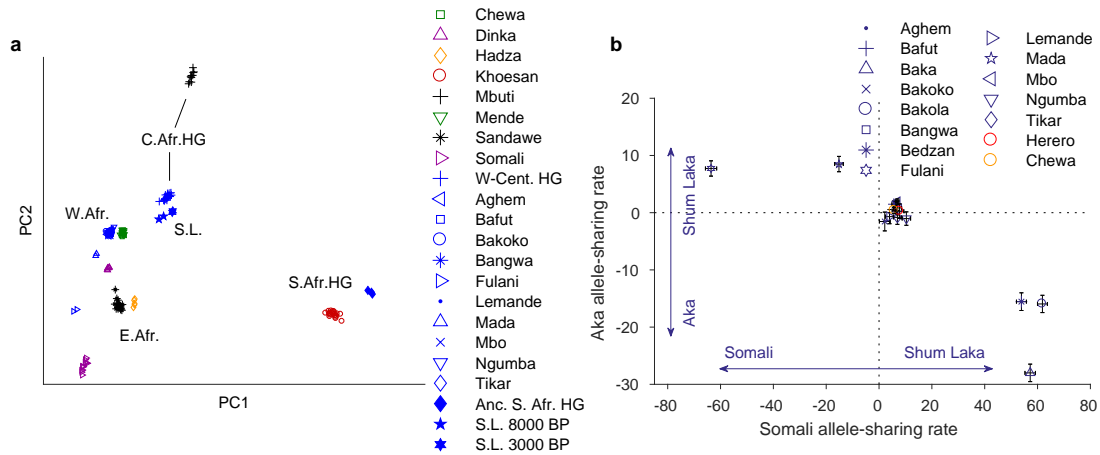
[#]Units along the Central African hunter-gather lineage (negative values indicate distances along an adjacent edge)



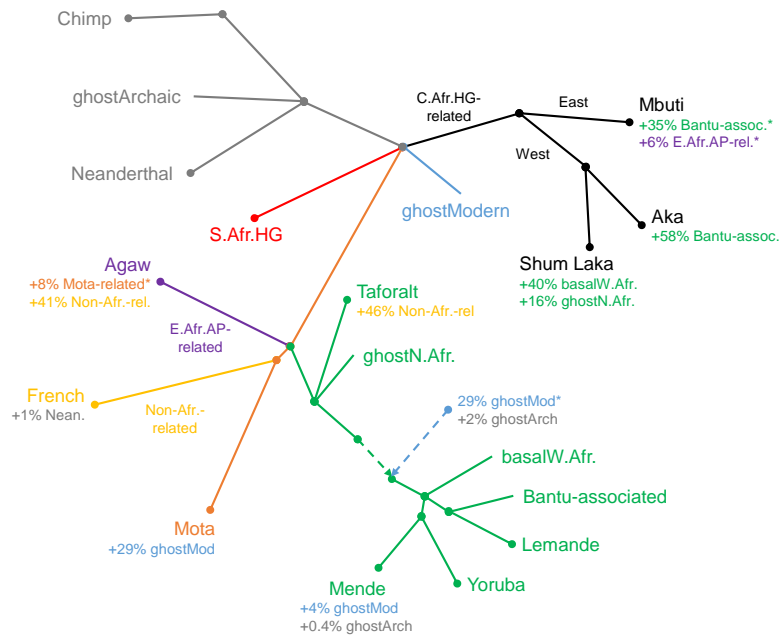
Extended Data Figure 1. Overview of the site of Shum Laka. The left column represents generalized stratigraphy, with radiocarbon dates (uncalibrated) shown as red dots on the y-axis, and deposits indicated by their archaeological nomenclature (P, S/Si = Pleistocene; T, A = Holocene; Ao = Holocene ochre ashy layer; Ag = Holocene gray ashy layer; after ref. [95]). Columns 1–6 display chronological extents of technological traditions: 1, microlithic quartz industry; 2, macrolithic flake and blade industry on basalt; 3, bifaces of the axe-hoe type; 4, pecked grounded adze and arrow heads; 5, pottery; and 6, iron objects. Column 7 indicates the two Shum Laka burial phases. Column 8 shows climatic reconstructions based on carbon stable isotopes and pollen from organic matter extracted from sediment cores at Lake Barombi Mbo in western Cameroon (more arid conditions to the left and more humid conditions to the right [21, 95]), along with archaeological eras (LSA, Later Stone Age; SMA, Stone to Metal Age; IA, Iron Age). Drawings: Y. Paquay, composition © Royal Museum of Central Africa.



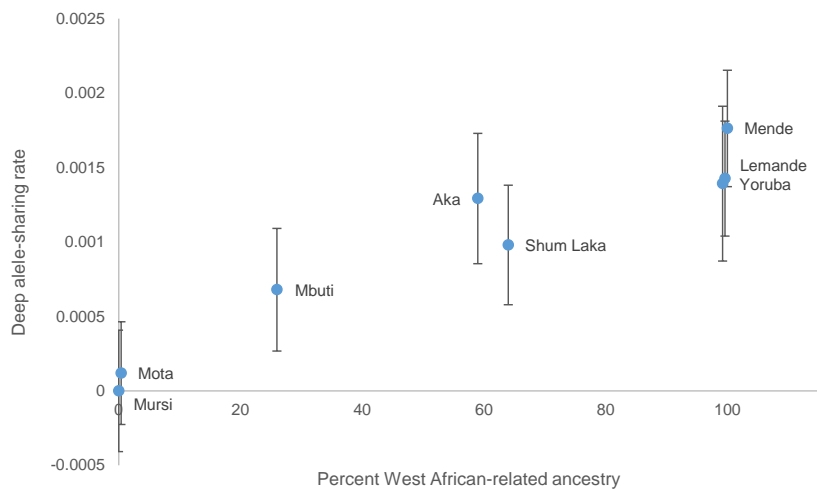
Extended Data Figure 2. Average genome-wide allelic mismatch rates for each pair of individuals, as well as intra-individual comparisons. We selected one read per individual at random at each targeted SNP. Monozygotic twins (or intra-individual comparisons) are expected to have a value one-half as large as unrelated individuals; first-degree relatives, halfway between monozygotic twins and unrelated individuals; second-degree relatives, halfway between first-degree relatives and unrelated individuals; and so on. The presence of inbreeding also serves to reduce the rates of mismatches. Bars show 99% confidence intervals (computed by block jackknife).



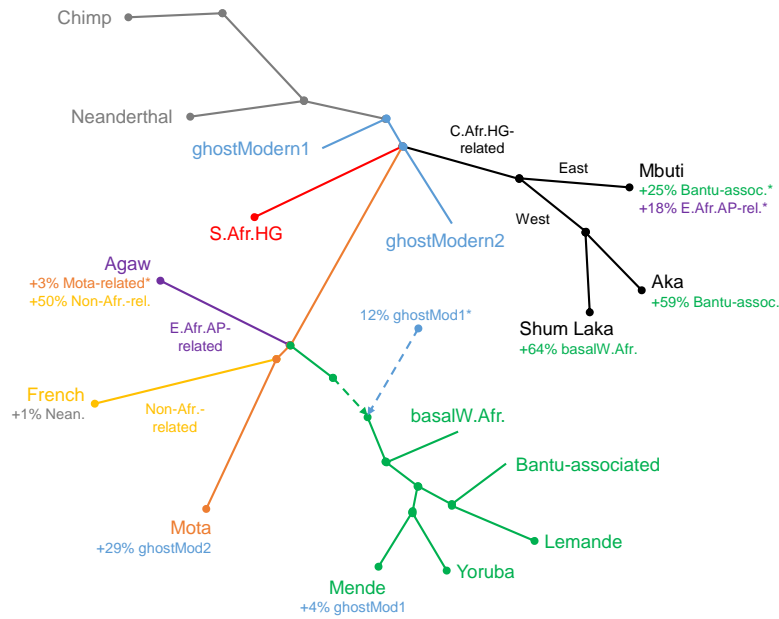
Extended Data Figure 3. (A) Broad-scale PCA (differing from Fig. 2A by projecting all present-day Cameroon populations). Groups shown in blue were projected onto axes computed using the other populations. HG, hunter-gatherers; S. L., Shum Laka. The W-Cent. HG grouping consists of Aka and Cameroon hunter-gatherers (Baka, Bakola, and Bedzan). The majority of the present-day Cameroon individuals fall in a tight cluster near other West Africans and Bantu speakers. (B) Relative allele sharing (multiplied by 10,000, as in Fig. 3B) with Shum Laka versus East Africans ($f_4(X, \text{Yoruba}; \text{Shum Laka}, \text{Somali})$; x-axis) and versus Aka ($f_4(X, \text{Yoruba}; \text{Shum Laka}, \text{Aka})$; y-axis) for present-day populations from Cameroon (blue points) and southern and eastern Bantu speakers (Herero in red and Chewa in orange). Mada and Fulani share more alleles with Shum Laka than with Aka, but this is likely a secondary consequence of admixture from East or North African sources (as reflected in greater allele sharing with Somali; see also Supplementary Information section 3). Bars show one standard error in each direction.



Extended Data Figure 5. Schematic of alternative admixture graph results including ancient individuals from Taforalt in Morocco associated with the Iberomaursian culture, with the Shum Laka individuals modeled as having a mixture of hunter-gatherer-related ancestry plus two additional components: one from within the main portion of the West African clade, and one splitting at nearly the same point as one of the sources contributing ancestry to Taforalt. Branch lengths are not drawn to scale. Points at which multiple lineages are shown diverging simultaneously indicate splits occurring in short succession (whose order we cannot confidently assess) but are not meant to represent exact multifurcations. HG, hunter-gatherer; AP, agro-pastoralist. *Proportion not well constrained (for Mbuti, the sum of the two indicated proportions is well constrained but not the separate values). See Supplementary Information section 3 for full inferred model parameters.



Extended Data Figure 6. Allele-sharing statistic sensitive to ancestry splitting more deeply than South African hunter-gatherers ($f_4(X, \text{Mursi}; \text{Chimp}, \text{South Africa HG})$, as in Fig. 3A), as a function of West African-related ancestry (from admixture graph results; Mota, Yoruba, and Lemande shifted slightly away from the boundaries for legibility). Bars show two standard errors in each direction. The (relative) allele-sharing rate for Mursi is identically zero according to the definition of the statistic.



Extended Data Figure 7. Schematic of admixture graph results with alternative deep source for West Africans. Branch lengths are not drawn to scale. Points at which multiple lineages are shown diverging simultaneously indicate splits occurring in short succession (whose order we cannot confidently assess) but are not meant to represent exact multifurcations. HG, hunter-gatherer; AP, agro-pastoralist. *Proportion not well constrained (for Mbuti, the sum of the two indicated proportions is well constrained but not the separate values). See Supplementary Information section 3 for full inferred model parameters.