Article

Ancient West African foragers in the context of African population history

https://doi.org/10.1038/s41586-020-1929-1

Received: 27 November 2018

Accepted: 29 November 2019

Published online: 22 January 2020

Mark Lipson[1]*, Isabelle Ribot[2], Swapan Mallick[1,3,4], Nadin Rohland[1], Iñigo Olalde[1,26], Nicole Adamski[1,4], Nasreen Broomandkhoshbacht[1,4,27], Ann Marie Lawson[1,4], Saioa López[5], Jonas Oppenheimer[1,4,28], Kristin Stewardson[1,4], Raymond Neba'ane Asombang[6], Hervé Bocherens[7,8], Neil Bradman[5,29], Brendan J. Culleton[9], Els Cornelissen[10], Isabelle Crevecoeur[11], Pierre de Maret[12], Forka Leypey Mathew Fomine[13], Philippe Lavachery[14], Christophe Mbida Mindzie[15], Rosine Orban[16], Elizabeth Sawchuk[17], Patrick Semal[16], Mark G. Thomas[5,18], Wim Van Neer[16,19], Krishna R. Veeramah[20], Douglas J. Kennett[21], Nick Patterson[1,22], Garrett Hellenthal[5,18], Carles Lalueza-Fox[23], Scott MacEachern[24], Mary E. Prendergast[1,25,30] & David Reich[1,3,4,22,30]

Our knowledge of ancient human population structure in sub-Saharan Africa, particularly prior to the advent of food production, remains limited. Here we report genome-wide DNA data from four children—two of whom were buried approximately 8,000 years ago and two 3,000 years ago—from Shum Laka (Cameroon), one of the earliest known archaeological sites within the probable homeland of the Bantu language group[1–11]. One individual carried the deeply divergent Y chromosome haplogroup A00, which today is found almost exclusively in the same region[12,13]. However, the genome-wide ancestry profiles of all four individuals are most similar to those of present-day hunter-gatherers from western Central Africa, which implies that populations in western Cameroon today—as well as speakers of Bantu languages from across the continent—are not descended substantially from the population represented by these four people. We infer an Africa-wide phylogeny that features widespread admixture and three prominent radiations, including one that gave rise to at least four major lineages deep in the history of modern humans.

The deposits at Shum Laka, a rockshelter located in the Grassfields region of western Cameroon, are among the most important archaeological sources for the study of Late Pleistocene and Holocene prehistory in western Central Africa[1–4]. The oldest human-occupied layers at the site date to about 30,000 calendar years before present (BP; taken as AD 1950 in accordance with radiocarbon calibration convention), but of special interest are artefacts and skeletons dating to between the end of the Later Stone Age (about 8,000 BP) and the beginning of the Iron Age (about 2,500 BP) (Extended Data Fig. 1, Supplementary Information section 1). This transitional period—sometimes referred to as the 'Stone to Metal Age'—featured a gradual appearance of new stone tools, as well as pottery[3–5]. Subsistence evidence in the rockshelter during the Stone to Metal Age points primarily to foraging, but with an increasing use of the fruits of *Canarium schweinfurthii* that coincided with developments in material culture and served as a foundation for later agriculture[3] (Supplementary Information section 1, Supplementary Table 1). These cultural changes and their early appearance at Shum Laka are particularly intriguing because during the late Holocene epoch, the area around the present-day border between Cameroon and Nigeria was probably the cradle of the Bantu language group, and of populations whose descendants would spread across much of the southern half of Africa between about 3,000 and 1,500 BP, resulting in the vast range and diversity of Bantu languages today[6–11].

A total of 18 human skeletons have been discovered at Shum Laka, comprising 2 distinct burial phases[1–3] (Supplementary Information section 1). We attempted to retrieve DNA from six petrous-bone samples

[1]Department of Genetics, Harvard Medical School, Boston, MA, USA. [2]Département d'Anthropologie, Université de Montréal, Montreal, Quebec, Canada. [3]Medical and Population Genetics Program, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [4]Howard Hughes Medical Institute, Harvard Medical School, Boston, MA, USA. [5]UCL Genetics Institute, University College London, London, UK. [6]Department of Arts and Archaeology, University of Yaoundé I, Yaoundé, Cameroon. [7]Department of Geosciences, Biogeology, University of Tübingen, Tübingen, Germany. [8]Senckenberg Research Centre for Human Evolution and Palaeoenvironment, University of Tübingen, Tübingen, Germany. [9]Institutes of Energy and the Environment, Pennsylvania State University, University Park, PA, USA. [10]Department of Cultural Anthropology and History, Royal Museum for Central Africa, Tervuren, Belgium. [11]CNRS, UMR 5199-PACEA, Université de Bordeaux, Bordeaux, France. [12]Faculté de Philosophie et Sciences Sociales, Université Libre de Bruxelles, Brussels, Belgium. [13]Department of History and African Civilization, University of Buea, Buea, Cameroon. [14]Agence Wallonne du Patrimoine, Service Public de Wallonie, Namur, Belgium. [15]Department of Arts and Archaeology, University of Yaoundé I, Yaoundé, Cameroon. [16]Royal Belgian Institute of Natural Sciences, Brussels, Belgium. [17]Department of Anthropology, Stony Brook University, Stony Brook, NY, USA. [18]Department of Genetics, Evolution and Environment, University College London, London, UK. [19]Department of Biology, University of Leuven, Leuven, Belgium. [20]Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY, USA. [21]Department of Anthropology, University of California, Santa Barbara, CA, USA. [22]Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, USA. [23]Institute of Evolutionary Biology (CSIC-UPF), Barcelona, Spain. [24]Division of Social Science, Duke Kunshan University, Kunshan, China. [25]Department of Sociology and Anthropology, Saint Louis University, Madrid, Spain. [26]Present address: Institute of Evolutionary Biology (CSIC-UPF), Barcelona, Spain. [27]Present address: Department of Anthropology, University of California, Santa Cruz, CA, USA. [28]Present address: Department of Biomolecular Engineering, University of California, Santa Cruz, CA, USA. [29]Present address: The Henry Stewart Group, London, UK. [30]These authors jointly supervised this work: Mary E. Prendergast, David Reich. *e-mail: mlipson@genetics.med.harvard.edu

Table 1 | Details of the four Shum Laka individuals in the study

Identifier	Age at death	Date	Radiocarbon date	Sex	Mt haplogroup	Y haplogroup	Coverage	SNPs	Mt and X contamination (%)
2/SE I	4 ± 1	7,920–7,690	6,985 ± 30 (PSUAMS-6307)	M	L0a2a1	B	0.70	564,164	1.0 and 1.0
2/SE II	15 ± 3	7,970–7,800	7,090 ± 35 (PSUAMS-6308)	M	L0a2a1	A00	7.71	1,082,018	1.5 and 0.6
4/A	8 ± 2	3,160–2,970	2,940 ± 20 (PSUAMS-6309)	M	L1c2a1b	B2b	3.83	935,777	0.3 and 0.5
5/B	4 ± 1	3,210–3,000	2,970 ± 25 (PSUAMS-6310)	F	L1c2a1b	NA	6.41	1,014,618	0.5 and NA

Age at death is given in years (mean ± s.e.m.), and was determined from skeletal remains[2]. Sex was determined from genetic data. Date is given in calibrated years BP as a 95.4% confidence interval (Methods). Radiocarbon date is given in uncalibrated radiocarbon years before present (mean ± s.e.m.), with the laboratory sample code shown in parentheses. Coverage refers to the average sequencing coverage, and contamination to the estimated contamination from mtDNA (Mt) or the X chromosome (X). NA, not applicable. Y, Y chromosome. Additional information is provided in Supplementary Table 2.

and obtained working data from two individuals of the early Stone to Metal Age and two of the late Stone to Metal Age (about 8,000 and 3,000 BP, respectively) (Table 1, Supplementary Table 2). The two earlier individuals—a boy of 4 ± 1 years old at time of death (given the identifying code 2/SE I) lying on top of the lower limbs of an adolescent male of 15 ± 3 years old (denoted 2/SE II)[2]—were recovered from a primary double burial, and the two later individuals—a boy of 8 ± 2 years (denoted 4/A) and a girl of 4 ± 1 years (denoted 5/B)[2]—were in adjacent primary single burials.

We extracted DNA from bone powder and prepared 2 or 4 libraries per individual for Illumina sequencing, enriching for about 1.2 million target single-nucleotide polymorphisms (SNPs) across the genome (Methods, Supplementary Table 2). Final coverage ranged from 0.7 to 7.7× (from 0.56 to 1.08 million SNPs). The authenticity of the data was supported by the observed rate of apparent C-to-T substitutions in the final base of sequenced fragments (4–10%, within the expected range given our library preparation strategy[14]) and of heterozygosity for mitochondrial DNA (mtDNA) and for the X chromosome in males (estimated contamination 0.3–1.5%). We also generated whole-genome shotgun sequence data for individuals 2/SE II (about 18.5× coverage) and 4/A (about 3.9× coverage), as well as genome-wide data (about 598,000 SNPs) for 63 individuals from 5 present-day Cameroonian populations (Extended Data Table 1, Supplementary Table 3).

Uniparental markers and kinship analysis

All of the mtDNA and Y chromosome haplogroups we observe at Shum Laka are associated today with sub-Saharan Africans. The two earlier individuals carry mtDNA haplogroup L0a (specifically L0a2a1), which is widespread in Africa, and the two later individuals carry L1c (specifically L1c2a1b), which is found among both farmers and hunter-gatherers in Central and West Africa[15,16]. Individuals 2/SE I and 4/A have Y chromosomes from macrohaplogroup B (often found today in hunter-gatherers from Central Africa[17]), and 2/SE II has the rare Y chromosome haplogroup A00, which was discovered in 2013 and is present at appreciable frequencies only in Cameroon—in particular, among the Mbo and Bangwa in the western part of the country[12,13]. A00 is the oldest known branch of the modern human Y chromosome tree, with a split time of about 300,000–200,000 BP from all other known lineages[12,18,19]. At 1,666 positions (from whole-genome sequence data; Supplementary Table 4) that differ between present-day A00[18] and all other Y chromosomes, the sequence of the Shum Laka individual carries the non-reference allele at a total of 1,521, translating to a within-A00 split at about 37,000–25,000 BP (95% confidence interval) (Fig. 1, Methods).

Leveraging the effects of chromosomal segments that are shared identical by descent (IBD), we computed rates of allele matching for each pair of individuals to infer degrees of relatedness. Both of the contemporaneous pairs display elevated levels of matching: 2/SE I and 2/SE II share alleles at the level of fourth-degree relatives, and 4/A and 5/B at the level of second-degree relatives (either uncle and niece, aunt and nephew or half-siblings) (Extended Data Fig. 2), supporting archaeological interpretations that—during both burial phases—the rockshelter was used as a cemetery for extended families[2]. We would expect more recent shared ancestry for the contemporaneous pairs even if they were not closely related, but we observe clear signatures of long IBD segments across the genome, which confirms their close family relatedness (Supplementary Information section 2). All four individuals also have evidence of intra-individual IBD, and thus of recent inbreeding.

PCA and allele-sharing statistics

We visualized the genome-wide relationships between the Shum Laka individuals and diverse present-day and ancient sub-Saharan Africans (Extended Data Table 1) using principal component analysis (PCA). Initially, we computed axes using East and West Africans and hunter-gatherers from southern Africa and eastern Central Africa (Fig. 2a). The Shum Laka individuals project to the right of present-day West African populations and speakers of Bantu languages (hereafter, Bantu-speakers) and are closest to present-day hunter-gatherers from Cameroon (Baka, Bakola and Bedzan[20]) and the Central African Republic (Aka, often known as Biaka). We then carried out a second PCA using only West and East Africans and Aka to compute the axes, and again the Shum Laka individuals project in the direction of hunter-gatherers from western Central Africa (Fig. 2b). By contrast, present-day groups from western Cameroon, who speak languages from the Niger–Congo family, cluster tightly with other West Africans (Fig. 2, Extended Data Fig. 3a). In both plots, the two earlier Shum Laka individuals fall slightly closer to West and East Africans, but—on the basis of their overall similarity—we grouped all four Shum Laka individuals together for most subsequent analyses.

Using *f*-statistics (Fig. 3a), we investigated components of 'deep ancestry' from sources that diverged earlier than the split between non-Africans and most sub-Saharan Africans (above point (2) in Fig. 4a). We began with the statistic $f_4(X, \text{Mursi}; \text{ancient South African hunter-gatherers}, \text{Han})$, which is expected to be positive if deep ancestry in

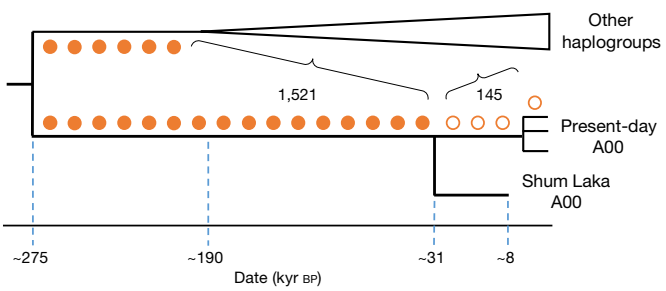


Fig. 1 | Y chromosome phylogeny. Circles represent mutations along the (unrooted) A00 lineage where we observe the alternative (filled) or reference (empty) allele in the A00 sequence carried by Shum Laka individual 2/SE II. Branch lengths are not drawn to scale. kyr, thousand years.

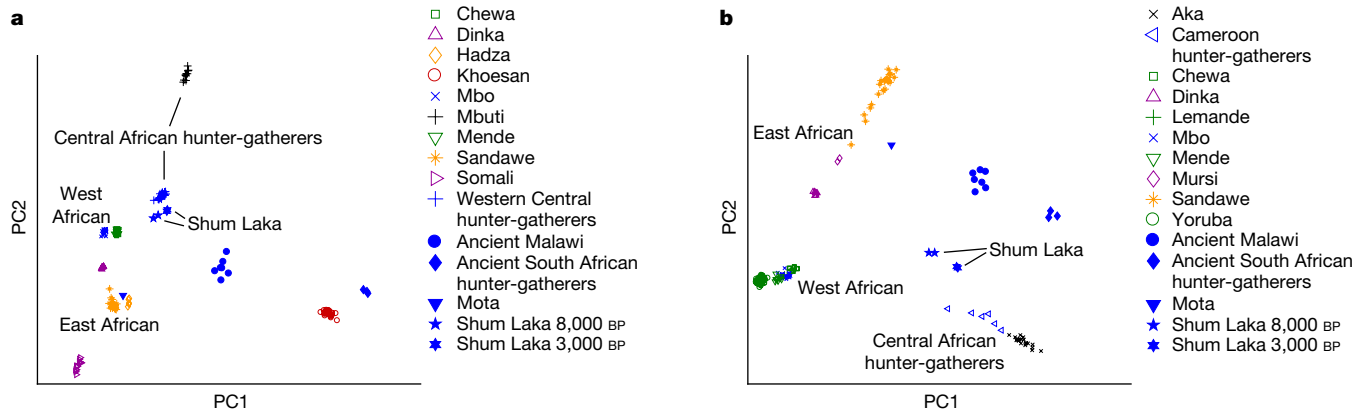


Fig. 2 | PCA results. a, Broad-scale analysis. **b**, Narrow-scale analysis. Groups in blue (including ancient individuals (filled symbols)) were projected onto axes computed using the other populations, using 593,124 SNPs (Methods). The western Central hunter-gatherer group in **a** comprises the Aka plus Cameroon hunter-gatherers (Baka, Bakola and Bedzan).

population *X* induces allele-sharing between *X* and ancient South African hunter-gatherers[21,22] (with a baseline of zero set by the Mursi, a group of pastoralists from western Ethiopia who speak a Nilotic language[20]). The Shum Laka individuals show a large positive statistic that is comparable to that of hunter-gatherers from western Central Africa (Fig. 3a top), whereas other West Africans (for example, Yoruba and Mende) yield smaller—but still significantly positive (Fig. 3a)—values, as do East African hunter-gatherers (the Hadza from Tanzania, as well as the ancient individual from Mota Cave in Ethiopia (hereafter, Mota individual), dating to approximately 4,500 BP[23]). We also obtained consistent results from analogous statistics using different reference groups (Extended Data Table 2).

Next, using chimpanzees as an outgroup that is symmetric to all human populations, we computed f_4(*X*, Mursi; chimpanzee, ancient South African hunter-gatherers) to evaluate whether any of this deep ancestry is from sources that diverged more deeply than southern African hunter-gatherers (the modern human lineage with the oldest known average split date[21,24,25]). Previous work has shown that southern African hunter-gatherers are not a symmetric outgroup relative to other sub-Saharan Africans: West Africans (especially the Mende) have excess affinity towards deeper outgroups[22]. Indeed, our test statistic is maximized in Mende and other West Africans (Fig. 3a, bottom). The Hadza, and the Mota individual, have values close to zero, and the Shum Laka individuals and Central African hunter-gatherers are intermediate. Some populations yield positive values for both f_4-statistics (Fig. 3a), but the two sets are poorly correlated, which implies that they—at least in part—reflect separate signals.

Combining our newly genotyped individuals with published data[20], we searched for differential allele-sharing between the Shum Laka individuals (compared to either East Africans (Somali) or to the Aka) and present-day Cameroonians (Fig. 3b, Extended Data Fig. 3b). We identified three distinct clusters: (1) Mada and Fulani, (2) hunter-gatherers and (3) other populations who speak languages in the Niger–Congo family (shown in close-up in Fig. 3b). Within the third cluster are the only groups—Mbo, Aghem and Bafut (all of whom live close to the site of Shum Laka today)—with significantly Shum-Laka-directed statistics in both dimensions, consistent with small proportions of Shum-Laka-related admixture (a maximum of about 7–8%) (Supplementary Information section 3).

Admixture graph analysis

Finally, we built an admixture graph (Fig. 4a, Extended Data Fig. 4, Methods) comodelling the Shum Laka, Mota and ancient South African hunter-gatherer individuals; present-day Mbuti, Aka, Agaw (speakers of an Afroasiatic language, from Ethiopia[20]), Yoruba, Mende and

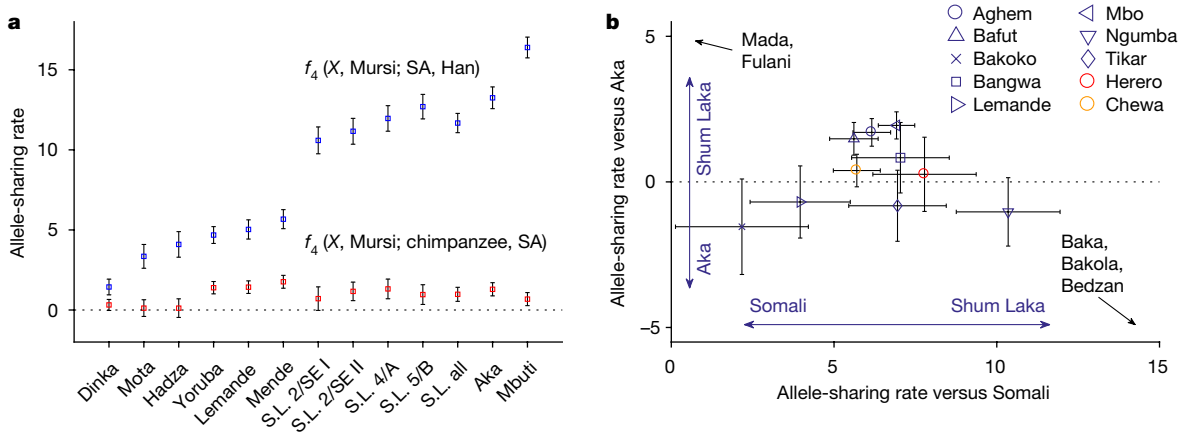
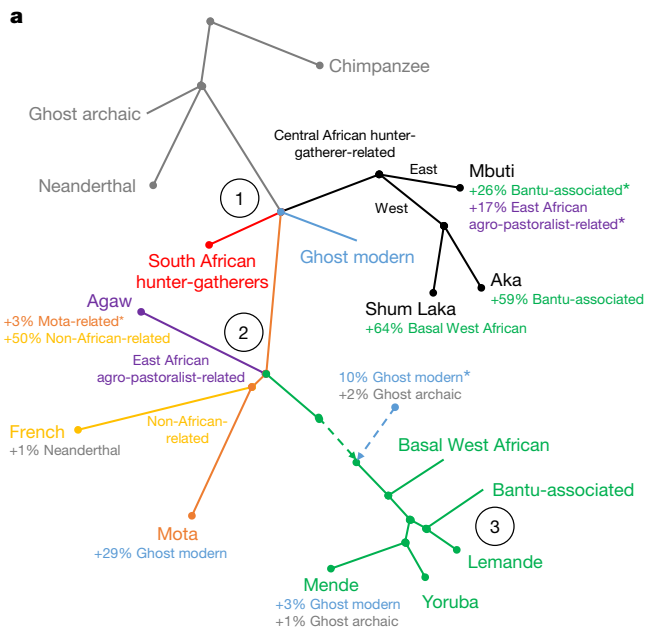


Fig. 3 | Allele-sharing statistics. a, Statistics sensitive to deep ancestry (mean ± 2 s.e.m., multiplied by 1,000). Blue, deeper than non-Africans; red, deeper than southern African hunter-gatherers; computed on 1,121,119 SNPs. SA, ancient South African hunter-gatherers; S.L., Shum Laka. **b**, Relative allele sharing (mean ± s.e.m., multiplied by 10,000; computed on 538,133 SNPs) with the Shum Laka individuals versus East Africans (f_4(*X*, Yoruba; Shum Laka, Somali); *x* axis) and versus Aka (f_4(*X*, Yoruba; Shum Laka, Aka); *y* axis) for present-day populations from Cameroon (blue) and southern (Herero, red) and eastern (Chewa, orange) Bantu-speakers. See also Extended Data Fig. 3b.



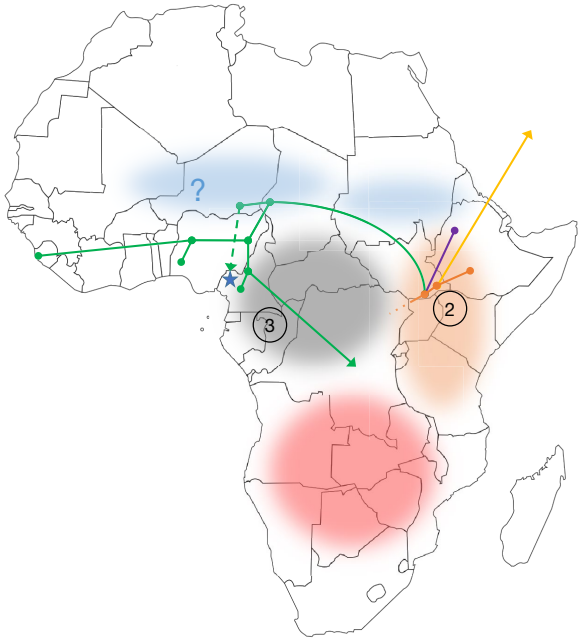


Fig. 4 | Admixture graph results. Points at which multiple lineages are shown diverging simultaneously indicate splits occurring in short succession (the order of which we cannot confidently assess) but do not represent exact multifurcations. Key points are (1) the early modern human split, (2) East African divergences and (3) the expansion of ancestry associated with Bantu-speakers. Branch lengths are not drawn to scale. **a**, Full model (Extended Data Fig. 4). *Proportion not well-constrained. **b**, Geographical structure. Shaded areas denote the hypothesized historical locations of lineages descended from split point (1) in **a**, and branching order is shown for populations descended from split point (2) (one ancestry component per population, with leaf nodes at sampling locations). The blue star represents Shum Laka (dashed line, possible direction of gene flow).

Lemande; non-Africans (French); and two outgroups (Neanderthal and chimpanzees). We also fit versions of the model using alternative SNP ascertainments and additional populations (Hadza, Mbo, Herero, Chewa, Mursi, Baka, Bakola, Bedzan, Mada, Fulani and ancient individuals from Taforalt in Morocco[26]) and obtained similar results (Extended Data Table 3, Supplementary Information section 3).

Among modern humans, the deepest-splitting branch is inferred to be the one that leads to Central African hunter-gatherers, although four lineages diverge in a very short span: those that contribute the primary ancestry to (1) Central African hunter-gatherers, (2) southern African hunter-gatherers and (3) other modern human populations, along with (4) a 'ghost' source that contributes a minority of the ancestry in West Africans and the Mota individual. Central African hunter-gatherers separate into eastern (Mbuti) and western clades; the latter then branches into components represented in Aka and the Shum Laka individuals. Next, a second cluster of divergences involves West Africans, two East African lineages (one associated with hunter-gatherers and another with agro-pastoralists) and non-Africans (who are tentatively inferred to split closest to the Mota individual, but with no deep ghost ancestry). Within the West African clade, we identify Yoruba and Mende as sister groups (with Lemande as an outgroup), and—most basally—a separate lineage that contributed to the Shum Laka individuals (64%). A source associated with Bantu-speakers (most closely related to Lemande) contributes 59% of the ancestry in Aka and 26% in Mbuti (who also have ancestry (17%) from a source related to agro-pastoralists of East Africa). In a model separating the two pairs of Shum Laka individuals, the pair dating to 3,000 BP have about 5% more ancestry related to hunter-gatherers from Central Africa (as confirmed by the significantly positive statistic f_4(Shum Laka 8,000 BP, Shum Laka 3,000 BP; Yoruba, Aka) ($Z = 4.2$)) (Supplementary Information section 3).

We can also obtain a good fit for the Shum Laka individuals in a less-parsimonious alternative model using three components, replacing the basal West African source with a combination of ancestry from inside the clade defined by the other West African populations and from a source entirely outside the West African clade (near one lineage that contributes to the Taforalt individuals) (Extended Data Fig. 5, Supplementary Information section 3). However, two-component models for the Shum Laka individuals that have the majority source splitting closer to other West or East Africans are rejected ($Z = 7.1$ and $Z = 3.7$, respectively).

The West African clade is distinguished by admixture from a deep source that can be modelled as a combination of modern human and archaic ancestry. The modern human component diverges at almost the same point as Central and southern African hunter-gatherers and is tentatively related to the deep source that contributes ancestry to the Mota individual, and the archaic component diverges close to the split between Neanderthals and modern humans (Supplementary Information section 3). The signals of deep ancestry in groups related to the West African clade (Fig. 3a) can be explained by two admixture events: one along the ancestral West African lineage, and a second, smaller contribution (about 4%) to Mende from the same source (Fig. 4a). Accordingly, f_4-statistics testing for ancestry basal to southern African hunter-gatherers (Fig. 3a, bottom) are well-correlated with the inferred proportions of ancestry from the West African clade (Extended Data Fig. 6). We estimate the shared admixture to introduce 10% deep modern human and 2% archaic ancestry, although the first proportion is not well-constrained (Extended Data Table 3). An alternative model with no archaic component—in which the West African clade receives deep ancestry from a single source[22] that splits before point (1) in Fig. 4a—also provides a reasonable fit to the data (Extended Data Fig. 7, Supplementary Information section 3), although it does not account for previous evidence of archaic ancestry in sub-Saharan Africans[27–31].

Genetics and archaeology at Shum Laka

Our analyses show that the 4 sampled children from Shum Laka can be modelled as admixed with about 35% ancestry related to hunter-gatherers from western Central Africa and about 65% from a basal West African source, or—alternatively—as a mixture of ancestry related to

hunter-gatherers plus two additional components, one from inside the clade of present-day West Africans and one that splits between East and West Africans. The first component plausibly represents ancestry that has been present in the area since at least the Later Stone Age (prior to 8,000 BP), whereas the second component (or the third in the alternative model) may have its origins farther to the north, given the geography and phylogeny of the other populations we studied (Fig. 4b). The chronology of the archaeological record at Shum Laka also suggests a possible northern influence on cultural developments during the Stone to Metal Age[3,9]. These developments include (1) changes in stone tools (which can be interpreted as a fusion of local tool-making traditions of the Later Stone Age with new macrolithic technologies that were introduced from the north[3]), and (2) the appearance of ceramics (four sherds have been found in the burial layer associated with the early Stone to Metal Age, and more abundant and distinct ceramics are found in later Stone to Metal Age deposits), which are potentially related to earlier pottery-working traditions in the Sahara and Sahel[3,32]. Moreover, gene flow from the north before 8,000 BP is plausible in light of a short period of Saharan and Sahelian aridification[3,33], which could have contributed to population movements. Present-day groups in northern West Africa and the Sahel have substantial admixture connected to later migrations[34], so identifying the exact source area may require additional ancient DNA studies.

Although the scope of our sampling is limited to two individuals at either end of the Stone to Metal Age, the observed genetic similarity across a span of almost 5,000 years—a similarity that is consistent with skeletal morphometric analyses—suggests a long-term presence of related peoples who used the rockshelter for various activities, including burying their dead (Supplementary Information section 1). However, most populations in Cameroon today are more closely related to other West Africans than to the group represented by these individuals. Present-day hunter-gatherers in Cameroon are also not descended substantially from this specific group, as they lack the signal of basal West African ancestry (Supplementary Information section 3). We do observe elevated levels of allele-sharing between the Shum Laka individuals and present-day populations of the Grassfields region, so the genetic discontinuity is not absolute. Additionally, the adolescent male 2/SE II carried an A00 Y chromosome, which suggests both that the concentration of this haplogroup in western Cameroon may have a long history and that A00 was formerly more diverse (given that the Shum Laka sequence falls outside of known present-day variation)[12,13]. The divergence time of A00 from other modern human haplogroups, of about 300,000–200,000 BP[18,19], could support its association either with the component of the ancestry of the Shum Laka individuals that is related to Central African hunter-gatherers or with the deep modern human portion of their West-African-related ancestry.

Linguistic and genetic evidence points to western Cameroon as the most likely area for the development of Bantu languages and as the ultimate source of subsequent migrations of Bantu-speakers, and—although the regional mid-Holocene archaeological record is sparse—Shum Laka has previously been highlighted as a site that was potentially important in the early phase of this process[1–4,6–11]. However, the genetic profiles of our four sampled individuals—even by about 3,000 BP, when the spreads of Bantu languages and of ancestry associated with Bantu-speakers were already underway—are very different from those of most speakers of Niger–Congo languages today, which implies that these individuals are not representative of the primary source population(s) that were ancestral to present-day Bantu-speakers. These results neither support nor contradict a central role for the Grassfields area in the origins of Bantu-speakers, and it may be that multiple, highly differentiated populations formerly lived in the region—with potentially either high or low levels of linguistic diversity. It would not be surprising if the Shum Laka site itself was used (either successively or concurrently) by multiple groups with different ancestry, cultural traditions or languages[1], evidence of which may not be visible from the collection of remains as preserved today.

Implications for deep population history

By analysing data from Shum Laka and other ancient individuals in conjunction with present-day groups, we gain insights into African population structure on multiple timescales. First, we infer a series of closely spaced population splits that involve one West-African-related lineage and two East-African-related lineages, as well as non-Africans (point (2) in Fig. 4a). The geography of the populations involved suggests the centre of this radiation was plausibly in East Africa (Fig. 4b), and estimated divergences of African and non-African populations place its date at about 80,000–60,000 BP[24,35]. Such an expansion is also consistent with mtDNA phylogeography—specifically the diversification of haplogroup L3, which probably originated in East Africa about 70,000 BP[36,37]—and potentially with the origins of clade CT in the Y chromosome tree at a similar time depth[18,38].

Second, we infer a phase of divergences that involved at least four lineages early in the history of modern humans (point (1) in Fig. 4a). Recent consensus has been that southern African hunter-gatherers, who split from other populations about 250,000–200,000 BP, represent the deepest sampled branch of modern human variation[21,24,25]. Our results suggest that Central African hunter-gatherers split at close to the same time (or perhaps slightly earlier), and thus that both clades—as well as the lineage that would later diversify at point (2) (Fig. 4a)—originated as part of a large-scale radiation.

In addition to the well-characterized deep lineages, we also detect at least one deep ghost source that contributed to West Africans and East African hunter-gatherers. This signal corroborates previous evidence for the Hadza and Sandawe[39] and for West Africans[22], although we find that the best fit is a source that splits near the same point as southern and Central African hunter-gatherers. Our results are also consistent with previous reports of archaic ancestry in African populations[27–31], specifically in West Africans. The presence of deep ancestry in the West African clade is notable in light of the Pleistocene archaeological record[5,40], which is limited but includes *Homo sapiens* fossils dated to about 300,000 BP in northwestern Africa[41], as well as an individual with archaic features buried about 12,000 BP in southwestern Nigeria (the oldest known human fossil from West Africa proper)[42]. Middle Stone Age artefacts have also been found in parts of West Africa into the terminal Pleistocene[43], despite the development of Later Stone Age technologies elsewhere (as, for example, at Shum Laka). Thus, the available material and fossil evidence is concordant with our genetic results in indicating long-term African population structure and admixture[44,45].

Further genetic studies may reveal additional complexities in deep human population history, although some early human groups will probably remain known only through fossils[44,45]. On the basis of our current understanding, the presence of at least 4 modern human lineages that diversified about 250,000–200,000 BP and are represented in people living today provides further support for archaeological evidence that suggests this era was a pivotal period for human evolution in Africa.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-1929-1.

1. de Maret, P. in *Aspects of African Archaeology: Papers from the 10th Congress of the Pan-African Association of Prehistory and Related Studies* (eds Pwiti, G. & Soper, R.) 274–279 (Univ. of Zimbabwe Publications, Harare, 1996).
2. Ribot, I., Orban, R. & de Maret, P. *The Prehistoric Burials of Shum Laka Rockshelter (North-West Cameroon) (Annales du Musée Royal de l'Afrique Centrale vol. 164)* (Musée Royal de l'Afrique Centrale, Tervuren, 2001).
3. Lavachery, P. The Holocene archaeological sequence of Shum Laka rock shelter (Grassfields, western Cameroon). *Afr. Archaeol. Rev.* **18**, 213–247 (2001).

4. de Maret, P. in *The Oxford Handbook of African Archaeology* (eds Mitchell, P. & Lane, P.) 627–643 (Oxford Univ. Press, 2013).
5. Cornelissen, E. in *The Oxford Handbook of African Archaeology* (eds Mitchell, P. & Lane, P.) 403–417 (Oxford Univ. Press, 2013).
6. Vansina, J. New linguistic evidence and 'the Bantu expansion'. *J. Afr. Hist.* **36**, 173–195 (1995).
7. Tishkoff, S. A. et al. The genetic structure and history of Africans and African Americans. *Science* **324**, 1035–1044 (2009).
8. Berniell-Lee, G. et al. Genetic and demographic implications of the Bantu expansion: insights from human paternal lineages. *Mol. Biol. Evol.* **26**, 1581–1589 (2009).
9. Bostoen, K. et al. Middle to late Holocene Paleoclimatic change and the early Bantu expansion in the rain forests of Western Central Africa. *Curr. Anthropol.* **56**, 354–384 (2015).
10. Patin, E. et al. Dispersals and genetic adaptation of Bantu-speaking populations in Africa and North America. *Science* **356**, 543–546 (2017).
11. Bostoen, K. in *Oxford Research Encyclopedia of African History* (ed. Spear, T.) https://oxfordre.com/africanhistory/view/10.1093/acrefore/9780190277734.001.0001/acrefore-9780190277734-e-191 (Oxford Univ. Press, 2018).
12. Mendez, F. L. et al. An African American paternal lineage adds an extremely ancient root to the human Y chromosome phylogenetic tree. *Am. J. Hum. Genet.* **92**, 454–459 (2013).
13. Krahn, T., Schrack, B., Fomine, F. L. M. & Krahn, A.-M. Searching for our most distant (paternal) cousins in Cameroon. *Institute for Genetic Genealogy 2016 Conference, San Diego* (2016).
14. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Phil. Trans. R. Soc. Lond. B* **370**, 20130624 (2015).
15. Gonder, M. K., Mortensen, H. M., Reed, F. A., de Sousa, A. & Tishkoff, S. A. Whole-mtDNA genome sequence analysis of ancient African lineages. *Mol. Biol. Evol.* **24**, 757–768 (2007).
16. Batini, C. et al. Phylogeography of the human mitochondrial L1c haplogroup: genetic signatures of the prehistory of Central Africa. *Mol. Phylogenet. Evol.* **43**, 635–644 (2007).
17. Wood, E. T. et al. Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes. *Eur. J. Hum. Genet.* **13**, 867–876 (2005).
18. Karmin, M. et al. A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Res.* **25**, 459–466 (2015).
19. Mendez, F. L., Poznik, G. D., Castellano, S. & Bustamante, C. D. The divergence of Neandertal and modern human Y chromosomes. *Am. J. Hum. Genet.* **98**, 728–734 (2016).
20. Fan, S. et al. African evolutionary history inferred from whole genome sequence data of 44 indigenous African populations. *Genome Biol.* **20**, 82 (2019).
21. Schlebusch, C. M. et al. Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. *Science* **358**, 652–655 (2017).
22. Skoglund, P. et al. Reconstructing prehistoric African population structure. *Cell* **171**, 59–71 (2017).
23. Gallego Llorente, M. et al. Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa. *Science* **350**, 820–822 (2015).
24. Gronau, I., Hubisz, M. J., Gulko, B., Danko, C. G. & Siepel, A. Bayesian inference of ancient human demography from individual genome sequences. *Nat. Genet.* **43**, 1031–1034 (2011).
25. Mallick, S. et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016).
26. van de Loosdrecht, M. et al. Pleistocene North African genomes link Near Eastern and sub-Saharan African human populations. *Science* **360**, 548–552 (2018).
27. Plagnol, V. & Wall, J. D. Possible ancestral structure in human populations. *PLoS Genet.* **2**, e105 (2006).
28. Hammer, M. F., Woerner, A. E., Mendez, F. L., Watkins, J. C. & Wall, J. D. Genetic evidence for archaic admixture in Africa. *Proc. Natl Acad. Sci. USA* **108**, 15123–15128 (2011).
29. Durvasula, A. & Sankararaman, S. Recovering signals of ghost archaic admixture in the genomes of present-day Africans. Preprint at bioRxiv https://doi.org/10.1101/285734 (2018).
30. Hey, J. et al. Phylogeny estimation by integration over isolation with migration models. *Mol. Biol. Evol.* **35**, 2805–2818 (2018).
31. Ragsdale, A. P. & Gravel, S. Models of archaic admixture and recent history from two-locus statistics. *PLoS Genet.* **15**, e1008204 (2019).
32. Huysecom, E. et al. The emergence of pottery in Africa during the 10th millennium calBC: new evidence from Ounjougou (Mali). *Antiquity* **83**, 905–917 (2009).
33. Gasse, F. Hydrological changes in the African tropics since the Last Glacial Maximum. *Quat. Sci. Rev.* **19**, 189–211 (2000).
34. Triska, P. et al. Extensive admixture and selective pressure across the Sahel belt. *Genome Biol. Evol.* **7**, 3484–3495 (2015).
35. Laval, G., Patin, E., Barreiro, L. B. & Quintana-Murci, L. Formulating a historical and demographic model of recent human evolution based on resequencing data from noncoding regions. *PLoS ONE* **5**, e10284 (2010).
36. Soares, P. et al. The expansion of mtDNA haplogroup L3 within and out of Africa. *Mol. Biol. Evol.* **29**, 915–927 (2012).
37. Behar, D. M. et al. A "Copernican" reassessment of the human mitochondrial DNA tree from its root. *Am. J. Hum. Genet.* **90**, 675–684 (2012).
38. Poznik, G. D. et al. Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. *Nat. Genet.* **48**, 593–599 (2016).
39. Pickrell, J. et al. The genetic prehistory of southern Africa. *Nat. Commun.* **3**, 1143 (2012).
40. Scerri, E. in *Oxford Research Encyclopedia of African History* (ed. Spear, T.) https://oxfordre.com/africanhistory/view/10.1093/acrefore/9780190277734.001.0001/acrefore-9780190277734-e-137 (Oxford University Press, 2017).
41. Hublin, J.-J. et al. New fossils from Jebel Irhoud, Morocco and the pan-African origin of *Homo sapiens*. *Nature* **546**, 289–292 (2017).
42. Harvati, K. et al. The later Stone Age calvaria from Iwo Eleru, Nigeria: morphology and chronology. *PLoS ONE* **6**, e24024 (2011).
43. Scerri, E. M., Blinkhorn, J., Niang, K., Bateman, M. D. & Groucutt, H. S. Persistence of Middle Stone Age technology to the Pleistocene/Holocene transition supports a complex hominin evolutionary scenario in West Africa. *J Archaeol. Sci. Rep.* **11**, 639–646 (2017).
44. Scerri, E. M. L. et al. Did our species evolve in subdivided populations across Africa, and why does it matter? *Trends Ecol. Evol.* **33**, 582–594 (2018).
45. Henn, B. M., Steele, T. E. & Weaver, T. D. Clarifying distinct models of modern human origins in Africa. *Curr. Opin. Genet. Dev.* **53**, 148–156 (2018).

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

Ancient DNA sample processing

We obtained bone powder from the Shum Laka skeletons (see Supplementary Information section 1 for more information on the site and burials) by drilling cochlear portions of petrous bone samples in a clean room facility at the Royal Belgian Institute of Natural Sciences. In dedicated clean rooms at Harvard Medical School, we extracted DNA using published protocols[46,47]. From the extracts, we prepared barcoded double-stranded libraries treated with uracil-DNA glycosylase (UDG) to reduce the rate of characteristic ancient DNA damage[14,48] in a modified partial UDG preparation that included magnetic bead clean-ups[14,49]. For the SNP capture data, we used two rounds of in-solution target hybridization to enrich for sequences that overlap the mitochondrial genome and approximately 1.2 million genome-wide SNPs[50–54]. We then added 7-base-pair indexing barcodes to the adapters of each library[55] and sequenced on an Illumina NextSeq 500 system with 76-base-pair paired-end reads. For individuals 2/SE II and 4/A, we also generated whole-genome shotgun data from the same libraries but without the target enrichment step. Sequencing was performed at the Broad Institute on Illumina HiSeq X Ten systems, using 19 lanes for 2/SE II (yielding approximately 18.5× average coverage, including 1,216,658 sites covered from the set of target SNPs used in most analyses) and two lanes for 4/A (3.9× average coverage, 1,158,884 sites covered).

From the raw sequencing results, we retained reads with no more than one mismatch per read pair to the library-specific barcodes. Before alignment, we merged paired-end sequences on the basis of forward and reverse mate overlaps and trimmed barcodes and adapters. Pre-processed reads were then mapped to both the mitochondrial reference genome RSRS[37] and the human reference genome (version hg19) using the samse command with default parameters in BWA (version 0.6.1)[56]. Duplicate molecules (with the same mapped start and end positions and strand orientation) were removed after alignment. We filtered the mapped sequences (requiring mapping quality scores of at least 10 for targeted SNP capture and 30 for whole-genome shotgun data) and trimmed 2 terminal bases to eliminate almost all damage-induced errors.

For mtDNA, we called haplogroups using HaploGrep2[57]. For nuclear DNA obtained from SNP capture and for the whole-genome shotgun data for individual 4/A, we selected one allele at random per site to create pseudohaploid genotypes. For the whole-genome shotgun data for individual 2/SE II, we used a previously described reference-bias-free diploid-genotype calling procedure[25], converting the resulting genotypes into a fasta-like encoding that allows for extraction of data at specified sites using cascertain and cTools[25]. We determined the sex of each individual by examining the fractions of sequences mapping to the X and Y chromosomes[58], and we determined Y chromosome haplogroups by comparing sequence-level SNP information to the tree established by the International Society of Genetic Genealogy (http://www.isogg.org).

To ensure authenticity, we computed the proportion of C-to-T deamination errors in terminal positions of sequenced molecules and evaluated possible contamination via heterozygosity at variable sites in haploid genome regions, using contamMix[50] and ANGSD[59] for mtDNA and the X chromosome (in males), respectively. Observed damage rates (4–10%) were relatively low but within the expected range after partial UDG treatment[14], and apparent heterozygosity rates for mtDNA (0.3–1.5% estimated contamination) and the X chromosome (0.5–1.0% estimated contamination) were minimal. The molecular preservation of the samples is impressive given the long-term warm and humid climate at Shum Laka[60] (which supports a mixed forest–savannah environment, at an elevation of about 1,650 m above sea level).

Radiocarbon dates

At the Pennsylvania State University (PSU) Radiocarbon Laboratory, we generated direct radiocarbon dates via accelerator mass spectrometry (AMS) for the four analysed individuals, using fragments of the same temporal bone portions that were sampled for ancient DNA. We extracted and purified amino acids using a modified XAD process[61] and assessed sample quality using stable isotope analysis. C:N ratios for all 4 samples fell between 3.3 and 3.4, well within the nominal range of 2.9–3.6 that indicates good collagen preservation[62]. The PSU AMS dates were in good agreement with previously reported direct dates for different bones from individuals 2/SE II (7,150 ± 70 uncalibrated radiocarbon years before present, calibrated to 8,160–7,790 BP (Oxford Radiocarbon Accelerator Unit sample code OxA-5203)) and 4/A (3,045 ± 60 uncalibrated radiocarbon years before present, calibrated to 3,380–3,010 BP, OxA-5205)[1,2,63,64], but on the basis of a (modestly) aberrant date[65] from a rib of individual 2/SE I (Supplementary Table 5), we restricted our final reported results to the temporal bones. We performed calibrations using OxCal[66] version 4.3.2 with a mixture of the IntCal13[67] and SHCal13[68] curves, specifying 'U(0,100)' to allow for a flexible combination[66,69], and rounding final results to the nearest 10 years (Supplementary Information section 1).

Present-day data

We generated genome-wide SNP genotype data for 63 individuals from 5 present-day Cameroonian populations on the Human Origins array: Aghem (28), Bafut (11), Bakoko (1), Bangwa (2) and Mbo (21) (Extended Data Table 1, Supplementary Table 3). Samples were collected with informed consent, with collection and analysis approved by the UCL/UCLH Committee on the Ethics of Human Research, Committee A and Alpha.

A00 Y chromosome split time estimation

Present-day A00 Y chromosomes are classified into the subtypes A00a, A00b and A00c, the divergence times of which from each other have not been precisely estimated but are quite recent—perhaps only a few thousand years ago[12,13]. To estimate the split time of the Shum Laka A00 Y chromosome from present-day A00, we called genotypes for individual 2/SE II (from our whole-genome sequence data) at a set of positions at which sequences from two present-day individuals with haplogroup A00[18] differ from all non-A00 individuals. At every subtype-specific site for which we had coverage, the Shum Laka A00 carries the ancestral allele. To avoid needing to determine the status of mutations as ancestral or derived, we considered the entire unrooted lineage specific to A00 (Fig. 1). The total time span represented by this lineage is approximately 359,000 years, using published values of about 275,000 BP for the divergence of the A00 lineage from other modern human haplogroups[19] and about 191,000 BP for the next-oldest split within macrohaplogroup A[70]. With a requirement of at least 90% agreement among the reads at each site, we called 1,521 positions as having the alternative allele (that is, matching the present-day A00 and differing from the human reference sequence) and 145 as having the reference allele (taking the average of 143 and 147 for the 2 present-day individuals). The fraction 145/(145 + 1,521) then defines the position of the split of the Shum Laka individual along the (unrooted) A00 lineage. Split times computed either from all sites (relaxing the 90% threshold and using the majority allele), or from additionally requiring at least two reads per site, differ from our primary estimate by only a few hundred years. To produce a confidence interval, we used the variance in the published estimates and assumed an independent Poisson sampling error for the number of observed reference alleles. The final point estimate was about 31,000 BP (95% confidence interval, 37,000–25,000 BP), which means that the A00 of the Shum Laka individual (with a sample date of about 8,000 BP) cannot be directly ancestral to the present-day subtypes.

PCA and allele-sharing statistics

We performed PCA using smartpca (with the lsqproject and autoshrink options)[71,72] and computed f_4-statistics using ADMIXTOOLS (with standard errors estimated via block jackknife over 5 cM chromosomal segments)[73]. We projected all ancient individuals in PCA rather than using them to compute axes to avoid artefacts caused by missing data. In each PCA, we also projected a subset of the present-day populations to enable controlled comparisons with ancient individuals. In most cases, reported f_4-statistics are based on the approximately 1.15 million autosomal SNPs from our target capture set. For PCA and for f_4-statistics testing differential relatedness to the Shum Laka individuals, we used autosomal SNPs from the Human Origins array (a subset of the target capture set), with some populations in the analyses only genotyped on this subset (Extended Data Table 1). For these latter f_4-statistics, we excluded for all populations a set of roughly 40,000 SNPs having high missingness in the present-day Cameroon data.

Admixture graphs

We fit admixture graphs with the ADMIXTUREGRAPH (qpGraph) program in ADMIXTOOLS (with the options 'outpop: NULL', 'lambdascale: 1', 'inbreed: YES' and 'diag: 0.0001')[73–75], using the 1.15 million autosomal SNPs from our target capture set by default, and other sets of SNPs in alternative model versions as specified. The program requires as input the branching order of the populations in the graph and a list of admixture events, and it then solves for the optimal parameters of the model (branch lengths and mixture proportions) via an objective function measuring the deviation between predicted and observed values of a basis set of *f*-statistics. From the inferred parameters, poorly fitting topologies (including positions of admixture sources) can be corrected by changing split orders at internal nodes that appear as trifurcations under the constraints enforced by the input (Supplementary Information section 3).

To evaluate the fit quality of output models, we used two metrics: first, a list of residual *Z*-scores for all *f*-statistics relating the populations in the graph, and second, a combined approximate log-likelihood score. The first metric is useful for identifying particularly poorly fitting models and the elements that are most responsible for the poor fits, and the second provides a means for comparing the overall fits of separate models (Supplementary Information section 3). To assess the degree of constraint on individual parameter inferences, we were guided primarily by the variability across different model versions (using different populations and SNP sets) (Extended Data Table 3, Supplementary Information section 3), which reflects both statistical uncertainty and changes in model-specific assumptions. In our primary model, all *f*-statistics relating subsets of the populations are predicted to within 2.3 standard errors of their observed values.

Initially, we detected a slight—but significant—signal (maximum $Z = 2.5$) of allele-sharing between the Shum Laka individuals and non-Africans, which we hypothesize is due to a small amount of DNA contamination. To prevent this effect from influencing our results, we included a 'dummy' admixture of non-African ancestry into the Shum Laka individuals (inferred 1.1%, consistent with mtDNA- and X chromosome-based contamination estimates), although model parameters without the dummy admixture are also very similar (Extended Data Table 3, Supplementary Information section 3).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The aligned sequences are available through the European Nucleotide Archive under accession number PRJEB32086. Genotype data used in analysis are available at https://reich.hms.harvard.edu/datasets. Any other relevant data are available from the corresponding author upon reasonable request.

46. Dabney, J. et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl Acad. Sci. USA* **110**, 15758–15763 (2013).
47. Korlević, P. et al. Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* **59**, 87–93 (2015).
48. Briggs, A. W. et al. Removal of deaminated cytosines and detection of *in vivo* methylation in ancient DNA. *Nucleic Acids Res.* **38**, e87 (2010).
49. Lipson, M. et al. Ancient genomes document multiple waves of migration in Southeast Asian prehistory. *Science* **361**, 92–95 (2018).
50. Fu, Q. et al. DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl Acad. Sci. USA* **110**, 2223–2227 (2013).
51. Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
52. Fu, Q. et al. An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216–219 (2015).
53. Mathieson, I. et al. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).
54. Lazaridis, I. et al. Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419–424 (2016).
55. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3 (2012).
56. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
57. Weissensteiner, H. et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **44**, W58–W63 (2016).
58. Skoglund, P., Storå, J., Götherström, A. & Jakobsson, M. Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J. Archaeol. Sci.* **40**, 4477–4482 (2013).
59. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* **15**, 356 (2014).
60. Giresse, P., Maley, J. & Brenac, P. Late Quaternary palaeoenvironments in the Lake Barombi Mbo (West Cameroon) deduced from pollen and carbon isotopes of organic matter. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **107**, 65–78 (1994).
61. Lohse, J. C., Culleton, B. J., Black, S. L. & Kennett, D. J. A precise chronology of middle to late Holocene bison exploitation in the far southern Great Plains. *J. Texas Archeol. Hist.* **1**, 94–126 (2014).
62. van Klinken, G. J. Bone collagen quality indicators for palaeodietary and radiocarbon measurements. *J. Archaeol. Sci.* **26**, 687–695 (1999).
63. Lavachery, P. *De la Pierre au Métal: Archéologie des Dépôts Holocènes de l'Abri de Shum Laka (Cameroun).* PhD thesis, Université Libre de Bruxelles (1997).
64. Bronk Ramsey, C., Higham, T. F., Owen, D., Pike, A. & Hedges, R. E. Radiocarbon dates from the Oxford AMS system: archaeometry datelist 31. *Archaeometry* **44**, 1–150 (2002).
65. Ward, G. K. & Wilson, S. R. Procedures for comparing and combining radiocarbon age determinations: a critique. *Archaeometry* **20**, 19–31 (1978).
66. Ramsey, C. B. & Lee, S. Recent and planned developments of the program OxCal. *Radiocarbon* **55**, 720–730 (2013).
67. Reimer, P. J. et al. IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal bp. *Radiocarbon* **55**, 1869–1887 (2013).
68. Hogg, A. G. et al. SHCal13 Southern Hemisphere calibration, 0–50,000 years cal BP. *Radiocarbon* **55**, 1889–1903 (2013).
69. Marsh, E. J. et al. IntCal, SHCal, or a mixed curve? Choosing a ^{14}C calibration curve for archaeological and paleoenvironmental records from tropical South America. *Radiocarbon* **60**, 925–940 (2018).
70. Jobling, M. A. & Tyler-Smith, C. Human Y-chromosome variation in the genome-sequencing era. *Nat. Rev. Genet.* **18**, 485–497 (2017).
71. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
72. Liu, L. T., Dobriban, E. & Singer, A. ePCA: high dimensional exponential family PCA. Preprint at https://arxiv.org/abs/1611.05550 (2016).
73. Patterson, N. et al. Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
74. Lipson, M. & Reich, D. A working model of the deep relationships of diverse modern human genetic lineages outside of Africa. *Mol. Biol. Evol.* **34**, 889–902 (2017).
75. Lipson, M. et al. Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature* **551**, 368–372 (2017).
76. Moeyersons, J., Cornelissen, E., Lavachery, P. & Doutrelepont, H. L'abri sous-roche de Shum Laka (Cameroun Occidental): données climatologiques et occupation humaine depuis 30.000 ans. *Geo. Eco. Trop.* **20**, 39–60 (1996).
77. Cornelissen, E. in *Field Manual for African Archaeology* (eds Smith, A. L. et al.) 168–173 (Royal Museum for Central Africa, 2017).
78. Prüfer, K. et al. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).

Acknowledgements We thank I. Lazaridis, V. Narasimhan and K. Sirak for discussions and comments; M. Karmin for help with Y chromosome data; L. Eccles for help with radiocarbon dating; B. Erkkila for help with isotopic analysis; R. Bernardos, M. Mah and Z. Zhang for other technical assistance; J.-P. Warnier for his role in locating the site of Shum Laka; and O. Graf for proofreading, photograph editing and other figure assistance for the Supplementary Information. The Shum Laka excavations were supported by the Belgian Fund for Scientific Research (FNRS), the Université Libre de Bruxelles, the Royal Museum for Central Africa and the Leakey Foundation. The collection of samples from present-day individuals in Cameroon was supported by N. Bradman and the Melford Charitable Trust. The genotyping of the

present-day individuals sampled from Cameroon was supported by the Biotechnology and Biological Sciences Research Council (grant number BB/L009382/1). I.R. was supported by a Université de Montréal exploration grant (2018-2020). M.G.T. was supported by Wellcome Trust Senior Investigator Award Grant 100719/Z/12/Z. G.H. was supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (grant number 098386/Z/12/Z). C.L-F. was supported by Obra Social La Caixa 328, Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement de la Generalitat de Catalunya (GRC 2017 SGR 880), and a FEDER-MINECO grant (PGC2018-095931-B-100). Radiocarbon work was supported by the NSF Archaeometry program (grant BCS-1460369) to D.J.K. and B.J.C. M.E.P. was supported by a fellowship from the Radcliffe Institute for Advanced Study at Harvard University during the development of this project. D.R. was supported by the National Institutes of Health (NIGMS GM100233), by an Allen Discovery Center grant and by grant 61220 from the John Templeton Foundation, and is an Investigator of the Howard Hughes Medical Institute.

Author contributions 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., C.M.M., R.O., E.S., P.S., W.V.N., C.L.-F., S. MacEachern and M.E.P. provided samples and assembled archaeological and anthropological materials and information. S.L., N. Bradman, F.L.M.F., M.G.T., K.R.V. and G.H. provided data from present-day populations. S. Mallick, N.R., N.A., N. Broomandkhoshbacht, A.M.L., J.O., K.S. and D.R. performed ancient DNA laboratory and data-processing work. B.J.C. and D.J.K. performed radiocarbon analysis. M.L., S. Mallick, I.O., N.P. and D.R. analysed genetic data. M.L., I.R., H.B., E.S., C.L.-F., S. MacEachern, M.E.P. and D.R. wrote the manuscript.

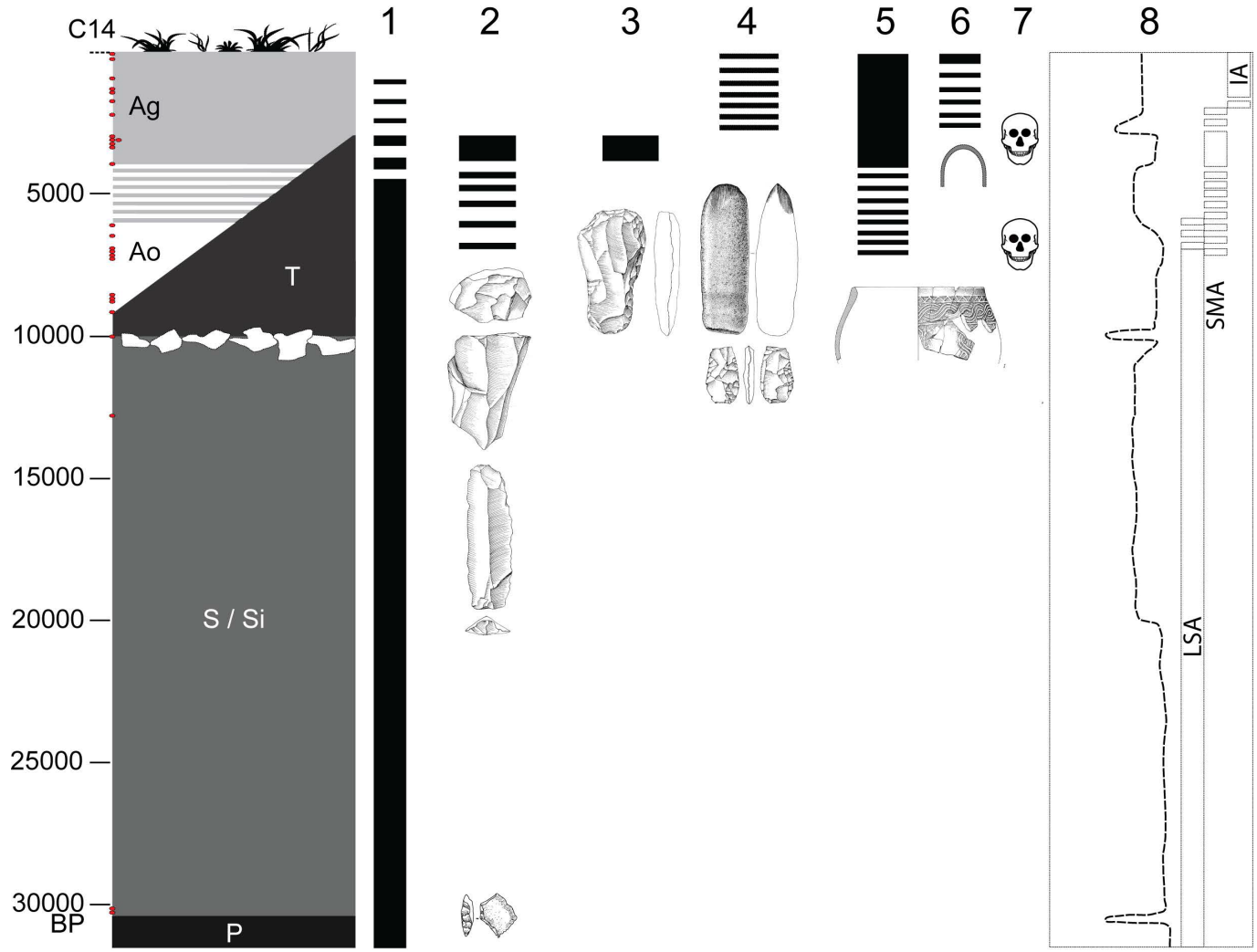
Competing interests The authors declare no competing interests.

Additional information

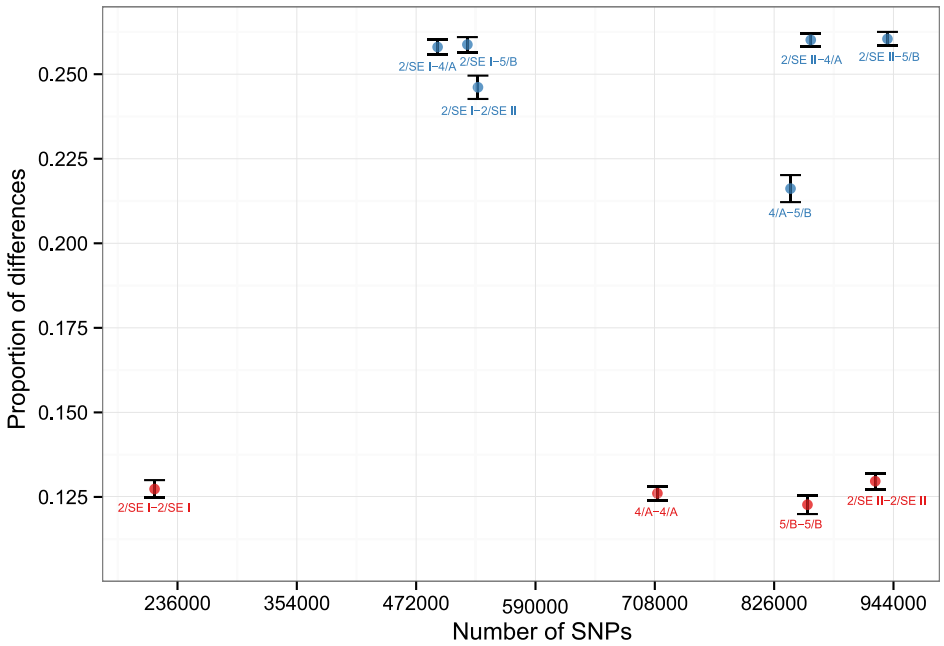
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-1929-1.

Correspondence and requests for materials should be addressed to M.L.

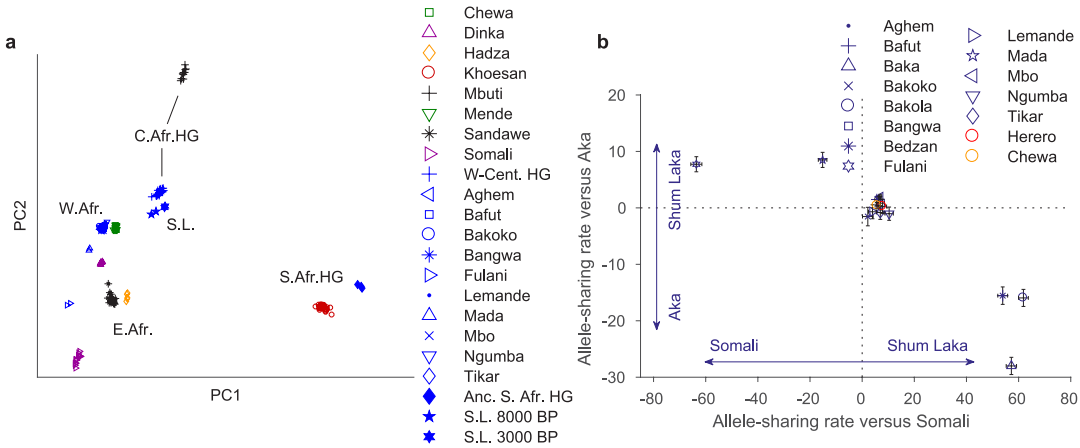
Peer review information *Nature* thanks George Busby, Katerina Harvati, Stephan Schiffels, Lluis Quintana-Murci and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



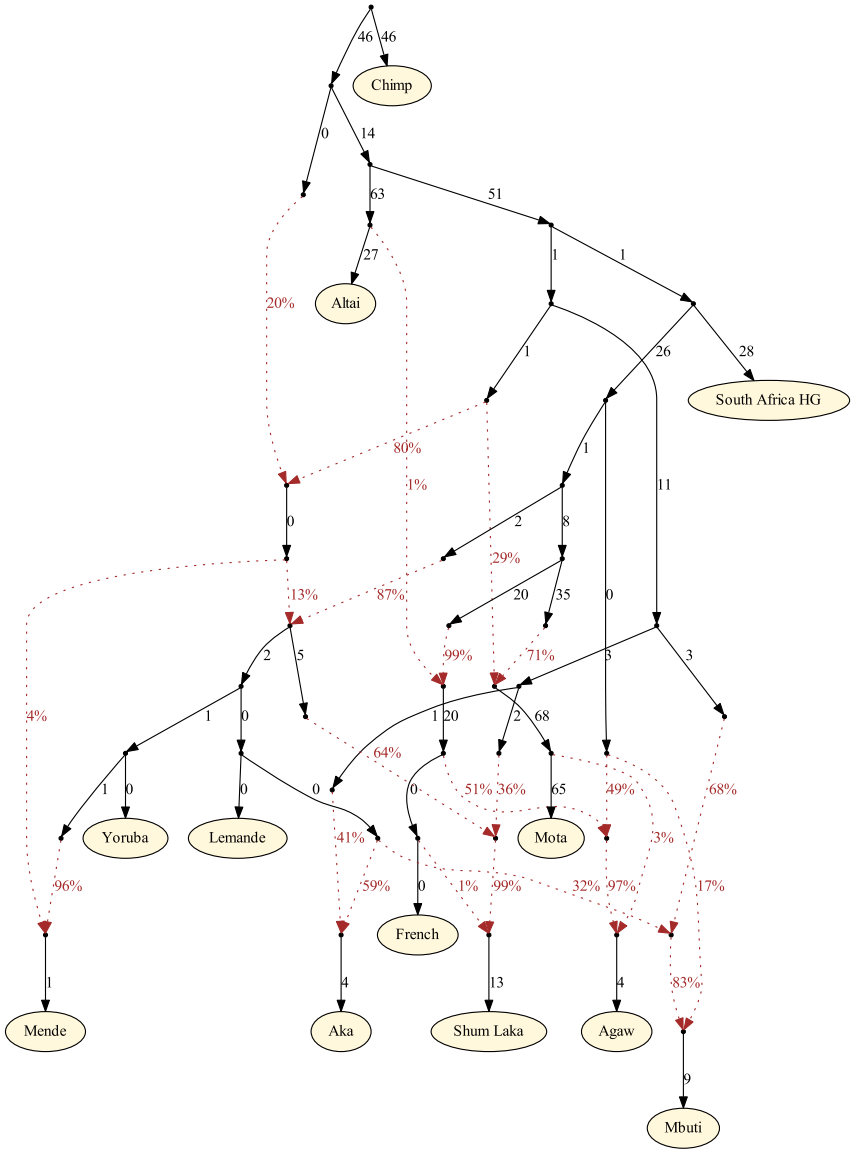
Extended Data Fig. 1 | Overview of the site of Shum Laka. The left column represents generalized stratigraphy, with radiocarbon dates (uncalibrated years before present) shown as red dots on the *y* axis, and deposits indicated by their archaeological nomenclature. P, S/Si, Pleistocene; T, A, Holocene; Ao, Holocene ochre ashy layer; Ag, Holocene grey ashy layer (after ref. [76]). Columns 1–6 display the 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 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[60,76]), along with archaeological eras. IA, Iron Age; LSA, Later Stone Age; SMA, Stone to Metal Age. RMCA Collection; drawings by Y. Paquay, composition © RMCA, Tervuren; modified by E. Cornelissen[77].



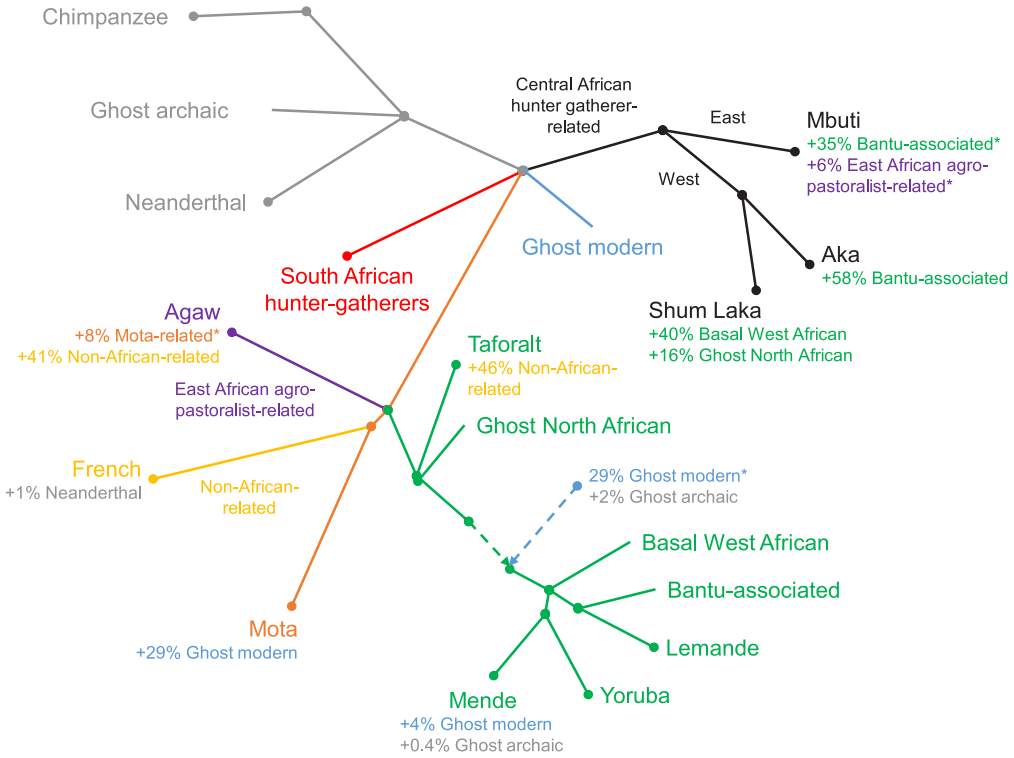
Extended Data Fig. 2 | Kinship analysis. Mean genome-wide allelic mismatch rates for each pair of individuals (blue), as well as intra-individual comparisons (red), are shown. We selected one read per individual at random at each targeted SNP (using all 1,233,013 targeted sites). Monozygotic twins (or 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 rate of mismatches. For 4/A and 5/B, we can eliminate a grandparent–grandchild relationship because both died as children, and the lack of long segments with IBD sharing on both homologous chromosomes implies that they are not double cousins (the few ostensible double-IBD stretches are probably a result of inbreeding (Supplementary Information section 2)). Thus, we can conclude that they were either uncle and niece (or aunt and nephew) or half-siblings. Bars show 99% confidence intervals (computed by block jackknife).



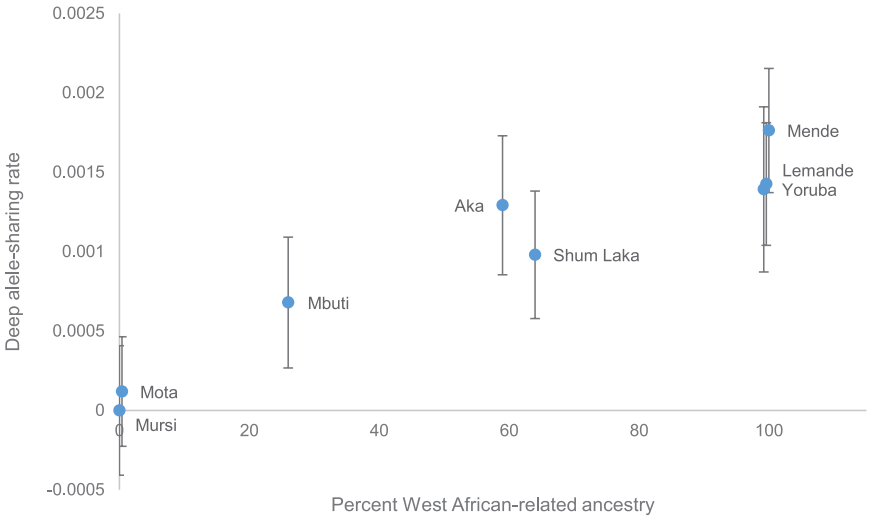
Extended Data Fig. 3 | Alternative PCA and allele-sharing analyses. a, Broad-scale PCA (differing from Fig. 2a by projecting all present-day Cameroon populations; again using 593,124 Human Origins SNPs). Groups shown in blue were projected onto axes computed using the other populations. HG, hunter-gatherers. The grouping marked W-Cent. HG 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 (mean ± s.e.m., multiplied by 10,000, computed on 538,133 SNPs, as in Fig. 3b) with the Shum Laka individuals versus East Africans (f_4(*X*, Yoruba; Shum Laka, Somali); *x* axis) and versus Aka (f_4(*X*, Yoruba; Shum Laka, Aka); *y* axis) for present-day populations from Cameroon (blue points) and southern and eastern Bantu-speakers (Herero in red and Chewa in orange). The Mada and Fulani share more alleles with the Shum Laka individuals than they do with the Aka, but this is probably a secondary consequence of admixture from East or North African sources (as reflected in greater allele sharing with Somali individuals) (Supplementary Information section 3). Bars show one s.e.m. in each direction.



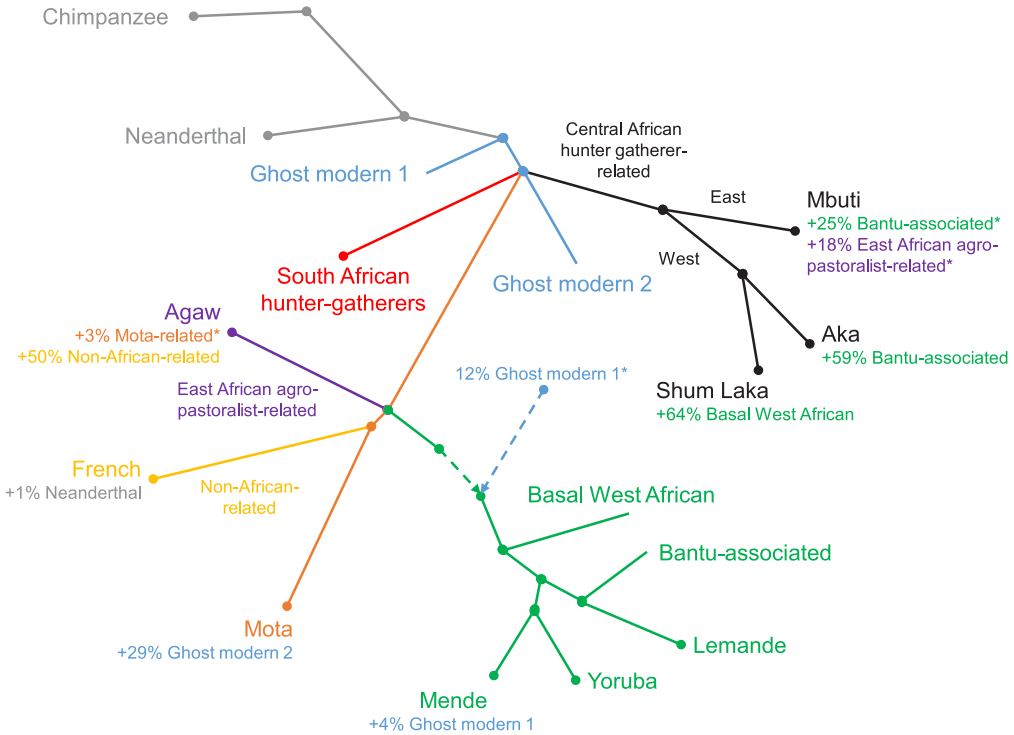
Extended Data Fig. 4 | Primary inferred admixture graph with full parameters. Of the approximately 1,200,000 targeted SNPs, 932,000 are used for fitting (that is, are covered by all populations in the model). Branch lengths (in units of squared allele frequency divergence) are rounded to the nearest integer. All *f*-statistics relating the populations are predicted to within 2.3 standard errors of their observed values.



Extended Data Fig. 5 | Schematic of first alternative admixture graph. Results are shown including ancient individuals from Taforalt in Morocco associated with the Iberomaurusian culture, with the Shum Laka individuals modelled as having a mixture of ancestry related to western Central African hunter-gatherers plus two additional components: one from within the main portion of the West African clade, and one that splits at nearly the same point as one of the sources that contributes ancestry to the Taforalt individuals. Branch lengths are not drawn to scale. Points at which multiple lineages are shown diverging simultaneously indicate splits that occur in short succession (the order of which we cannot confidently assess) but are not meant to represent exact multifurcations. *Proportion not well-constrained (for Mbuti, the sum of the two indicated proportions is well-constrained but not the separate values). Supplementary Information section 3 provides the full parameters of the inferred model.



Extended Data Fig. 6 | Deep ancestry correlation from the West African clade. An allele-sharing statistic sensitive to ancestry that splits more deeply than southern African hunter-gatherers ($f_4(X$, Mursi; chimpanzee, ancient South African hunter-gatherers), mean ± 2 s.e.m. from block jackknife, computed on 1,121,119 SNPs, as in Fig. 3a) is shown as a function of ancestry related to the West African clade (from admixture graph results; the Mota individual, Yoruba and Lemande are shifted slightly away from the boundaries for legibility). The (relative) allele-sharing rate for Mursi is zero according to the definition of the statistic.



Extended Data Fig. 7 | Schematic of the second alternative admixture graph. Results are shown with a single-component deep source for West Africans. Branch lengths are not drawn to scale. Points at which multiple lineages are shown diverging simultaneously indicate splits that occur in short succession (the order of which we cannot confidently assess) but are not meant to represent exact multifurcations. *Proportion is not well-constrained (for Mbuti, the sum of the two indicated proportions is well-constrained but not the separate values). Supplementary Information section 3 provides the full parameters of the inferred model.

Extended Data Table 1 | Populations 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	[*22*]
Mota	Ethiopia		~4500 BP	1	SG	[*23*]
Ancient South African HG	South Africa		~2000 BP	3†	SG	[*21, 22*]
Taforalt	Morocco		~15,000–14,000 BP	6	1240k	[*26*]
Altai Neanderthal	Russia		~120,000 BP	1	DG	[*78*]
Aghem	Cameroon	NC	Present	28	HO	This paper
Bafut	Cameroon	NC	Present	11	HO	This paper
Baka	Cameroon	NC	Present	2	DG	[*20*]
Bakoko	Cameroon	NC	Present	1	HO	This paper
Bakola	Cameroon	NC	Present	2	DG	[*20*]
Bangwa	Cameroon	NC	Present	2	HO	This paper
Bedzan	Cameroon	NC	Present	2	DG	[*20*]
Fulani	Cameroon	NC	Present	2	DG	[*20*]
Lemande	Cameroon	NC	Present	2	DG	[*25*]
Mada	Cameroon	AA	Present	2	DG	[*20*]
Mbo	Cameroon	NC	Present	21	HO	This paper
Ngumba	Cameroon	NC	Present	2	DG	[*20*]
Tikar	Cameroon	NC	Present	2	DG	[*20*]
Agaw	Ethiopia	AA	Present	2	DG	[*20*]
Aka (Biaka)	Central African Republic	NC	Present	20/2	HO/DG	[*22, 25*]
Chewa	Malawi	NC	Present	11	HO	[*22*]
Dinka	Sudan	NS	Present	7/4	HO/DG	[*22, 25*]
French	France	IE	Present	3	DG	[*25*]
Hadza	Tanzania	KS	Present	5(2)/1	HO/DG	[*22, 25*]
Han	China	ST	Present	4	DG	[*25*]
Herero	Namibia	NC	Present	2	DG	[*25*]
Khoesan	Namibia	KS	Present	22	HO	[*22*]
Mbuti	DR Congo	NC, NS	Present	10/4	HO/DG	[*22, 25*]
Mende	Sierra Leone	NC	Present	8/2	HO/DG	[*22, 25*]
Mursi	Ethiopia	NS	Present	2	DG	[*20*]
Sandawe	Tanzania	KS	Present	22	HO	[*22*]
Somali	Kenya	AA	Present	13	HO	[*22*]
Yoruba	Nigeria	NC	Present	70/3	HO/DG	[*22, 25*]

List of populations in the analyses in this study. Data types are in-solution targeted SNP capture (1240k); whole-genome sequences with pseudohaploid genotype calls (SG); high-coverage whole-genome sequences with diploid genotype calls (DG); and Human Origins SNP array (HO). For some populations, we used different sets of samples for different analyses (indicated by forward slashes); individuals genotyped on the Human Origins array were used for PCA and for *f*-statistics testing differential relatedness to Shum Laka individuals (Fig. 3b, Extended Data Fig. 3b). For the Hadza, we used five individuals with Human Origins data for PCA and two of these five individuals for admixture graph modelling. Data for the Altai Neanderthal individual are from ref. [78]. 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

	f_4(X, Mursi; SA, Han)		f_4(X, Mota; SA, Han)		f_4(X, Han; SA, Mursi)		f_4(X, Mota; SA, Mursi)	
Test pop	**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	..	..	..	..	..	..	..	..
	f_4(X, Mursi; SA, Mota)		f_4(X, Han; SA, Mota)		f_4(X, Han; SA, Yor)		f_4(X, Mursi; Chimp, Yor)	
Test pop	**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 1,000; computed on 1,121,119 SNPs) sensitive to ancestry in the test population *X* from a deeply-splitting lineage, along with *Z*-scores for difference from zero. 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 owing to ancestry related to non-Africans, or Yoruba with other West Africans). From the statistics f_4(Mursi or Agaw, Han; ancient South African hunter-gatherers, Yoruba), we find minimal differences in deep ancestry proportions among Han, Mursi, and Agaw; from f_4(*X*, Mursi; chimpanzee, Yoruba), we obtain a value for ancient South African hunter-gatherers that is roughly twice as large as that for Central African hunter-gatherers. 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. Supplementary Information section 3 provides the full details. 1,Primary model; 2, no 'dummy' admixture; 3, African-ascertained SNPs; 4, transversion SNPs; 5, whole-genome sequence data for the Shum Laka individuals; 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 the Shum Laka individuals; 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, pairs of Shum Laka individuals fit separately. WA, West African.

*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).