Ancient West African foragers in the context of African population history
Mark Lipson, Isabelle Ribot, Swapan Mallick, Nadin Rohland, Iñigo Olalde, Nicole Adamski, Nasreen Broomandkhoshbacht, Ann Lawson, Saioa López, Jonas Oppenheimer, et al.

To cite this version:
Mark Lipson, Isabelle Ribot, Swapan Mallick, Nadin Rohland, Iñigo Olalde, et al.. Ancient West African foragers in the context of African population history. Science, American Association for the Advancement of Science, 2020, 577, pp.665 - 670. 10.1038/s41586-020-1929-1. hal-02990305

HAL Id: hal-02990305
https://hal.archives-ouvertes.fr/hal-02990305
Submitted on 16 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Ancient West African foragers in the context of African population history

Mark Lipson1, Isabelle Ribot2, Swapan Mallick1,4, Nadin Rohland1, Eli J. Ogbuji1,2,3, Nicole Adamski1,4, Nasreen Broomandkhoshbacht4,6,7, Ann Marie Lawson1,8, Saisa López5, Jonas Oppenheimer1,4, Kristin Stewardson1,4, Raymond Neba’ane Asombang6, Hervé Bocherens9,6, Neil Bradman10,11, Brendan J. Cullleton1,12, Elsa Cornelissen1,12, Isabelle Crevecoeur14, Pierre de Maret15, Forka Leypey Matthew Fomine16,20, Philippe Lavachery14, Christophe Mbida Mindzie15, Rosine Orban16, Elizabeth Sawchuk17, Patrick Semal16, Mark G. Thomas5,18, Wim Van Neer16,19, Krishna R. Veeramah20, Douglas J. Kieffer15, Nick Patterson12,22, Garrett Hellenthal3, Carles Lalueza-Fox13, Scott MacEachern14, Mary E. Prendergast20,29,30 & David Reich13,14,19,21

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 group1–12. One individual carried the deeply divergent Y chromosome haplogroup A00, which today is found almost exclusively in the same region12,13. 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 Africa1–4. 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 (3,000 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 pottery5–7. 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 agriculture1 (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 today4–11.

A total of 18 human skeletons have been discovered at Shum Laka, comprising 2 distinct burial phases1–3 (Supplementary Information section 1). We attempted to retrieve DNA from six petrous-bone samples...
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)—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)—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) 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 5/B (4±1× 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. Individuals 2/SE I and 4/A have Y chromosomes from macrohaplogroup B (often found today in hunter-gatherers from Central Africa), 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. 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. At 1,666 positions (from whole-genome sequence data; Supplementary Table 4) that differ between present-day A00 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 yr (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. 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), 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 fS(X, Mursi; ancient South African hunter-gatherers, Han), which is expected to be positive if deep ancestry in

![Fig. 1: Y chromosome phylogeny.](image)

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. kyr, thousand years.
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 y.a$^{23}$). 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 $f_2(X, Mursi; chimpanzee, ancient South African hunter-gatherers)$ 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 $f_2$-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: (a) Mada and Fulani, (b) hunter-gatherers and (c) 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 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 (a) Central African hunter-gatherers, (b) southern African hunter-gatherers and (c) other modern human populations, along with (d) a ‘ghost’ source that contributes a minority of the ancestry in West Africa.
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 $f_2(\text{Shum Laka 8,000 bp, Shum Laka 3,000 bp; Yoruba, Aka}) (Z = 4.2)$ (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, $f_2$-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$^{22–31}$.

**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—altematively—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$^{21}$. 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$^{30}$), 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$^{13,32}$.
Moreover, gene flow from the north before 8,000 bp is plausible in light of a short period of Saharan and Sahelian aridification33, which could have contributed to population movements. Present-day groups in northern West Africa and the Sahel have substantial admixture connected to later migrations44, 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 I). 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)21,22. The divergence time of A00 from other modern human haplogroups, of about 300,000–200,000 bp34,35, 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 process4,6–10. 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, 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 bp34,35. Such an expansion is also consistent with mtDNA phylogeography—specifically the diversification of haplogroup L3, which probably originated in East Africa about 70,000 bp36,37—and potentially with the origins of clade CT in the Y chromosome tree at a similar time depth36,38.

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 variation21,24,25. 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 contributes to West Africans and East African hunter-gatherers. This signal corroborates previous evidence for the Hadza and Sandawe46 and for West Africans47, 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 populations48–50, specifically in West Africans. The presence of deep ancestry in the West African clade is notable in light of the Pleistocene archaeological record49, which includes Homo sapiens fossils dated to about 300,000 bp in northwestern Africa50, 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)51. Middle Stone Age artefacts have also been found in parts of West Africa into the terminal Pleistocene52, 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 admixture53–55.

Further genetic studies may reveal additional complexities in deep human population history, although some early human groups will probably remain known only through fossils48–50. 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 Pwtz, G. & Soep, R.) 274–279 (Univ. of Zimbabwe Publications, Harare, 1996).
2. Ribo, 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. Corneliussen, 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. Bernelot-Leel, E. 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 Paleoclimate 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. Speare, 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 Y chromosome phylogenetic tree. Am. J. Hum. Genet. 92, 454–459 (2013).
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 the Harvard Medical School, we extracted DNA using published protocols45-47. From the extracts, we prepared barcoded double-stranded libraries treated with uracil-DNA glycosylase (UDG) to reduce the rate of characteristic ancient DNA damage48,49 in a modified partial UDG preparation that included magnetic bead cleanups48-49. 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 SNPs50-54. We then added 7-base-pair indexing barcodes to the adapters of each library55 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. Preprocessed reads were then mapped to both the mitochondrial reference genome RSR55 and the human reference genome (version hg19) using the same command with default parameters in BWA (version 0.6.1)43. 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 mitochondrial DNA, we called haplogroups using HaploGrep257. 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 procedure60, converting the resulting genotypes into a fasta-like encoding that allows for extraction of 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 contamMix61 and ANGSD62 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 treatment61, 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 Laka49 (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 process63 and assessed sample quality using stable isotope analysis. C:N ratios for all four samples fell between 3.3 and 3.4, well within the nominal range of 2.9–3.6 that indicates good collagen preservation55. 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 Ox-A-5203)) and 4/A (3,045 ± 60 uncalibrated radiocarbon years before present, calibrated to 3,380–3,010 BP, Ox-A-5205)54,64, but on the basis of a (modestly) aberrant date65 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 version 4.3.2 with a mixture of the IntCal1367 and SHCal1368 curves, specifying ‘U(0,100)’ to allow for a flexible combination68,69, 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 ago21,21. 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 (ref. 89) 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 haplogroups20 and about 191,000 BP for the next-oldest split within macrohaplogroup A20. 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 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 haplogroups20 and about 191,000 BP for the next-oldest split within macrohaplogroup A20. 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 Isolproject and autoshrink options) and computed $f_2$-statistics using ADMIXTOOLS (with standard errors estimated via block jackknife over 5 cM chromosomal segments). 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 $f_2$-statistics are based on the approximately 1.15 million autosomal SNPs from our target capture set. For PCA and for $f_2$-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 $f_2$-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,’ ‘lambdafieldset: 1,’ ‘inbreed: YES’ and ‘diag: 0.0001’) 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 measurement defining the deviation between predicted and observed values of 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 truncations 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 PRJEB2086. Genotype data used in analysis are available at https://reich.lims.harvard.edu/datasets. Any other relevant data are available from the corresponding author upon reasonable request.

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. Archeol. 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 dateist 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 calibration curves 0–50,000 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. Marshall, E. J. et al. IntCal, SHCal, or a mixed curve? Choosing a 14C calibration curve for archaeological and palaeoecological 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., Dobrinn, E. & Singer, A. ePCA: high dimensional extended 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 paleogenomic transects reveal complex genetic history of early European farmers. Nature 551, 368–372 (2017).
76. Moeyersons, J., Cornelissen, E., Lavachery, P. & Doutrelent, H. Labi 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. Simk for discussions and comments; M. Karmin for help with Y chromosome data; L. Eccles for help with radiocarbon dating; B. Eriklika for help with isotopic analysis; R. Bernardos, M. Mah and Z. Zhang for other technical assistance; J.-P. Wariner 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 (FNSR), 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. 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 (NIH 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., Pd.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.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, Stephen Schiffels, Lluis Quintana-Murci and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at http://www.nature.com/reprints.
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 right60,76), 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. Cornelissen77.
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**, Broadscale 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., multiplied by 10,000, computed on 538,133 SNPs, as in Fig. 3b) with the Shum Laka individuals versus East Africans ($f_1(X, Yoruba; Shum Laka, Somali); x axis$) and versus Aka ($f_2(X, Yoruba; Shum Laka, Aka); y axis$) 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. in each direction.
Extended Data Fig. 4 | Primary inferred admixture graph with full parameters. Of the approximately 1,200,000 targeted SNPs, 932k 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. AP, agro-pastoralist. *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 ($f_2(X, Mursi; chimpanzee, ancient South African hunter-gatherers)$, mean ± 2 s.e. 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 range     | 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           | DG             | [23]      |
| Ancient South | South Africa  |                 | ~2000 BP       | 3*          | DG             | [21, 22] |
| African HG  |                  |                 |                |             |                 |           |
| Taforalt    | Morocco          |                 | ~15,000–14,000 BP | 6          | 1240k          | [26]      |
| Altai Neanderthal | Russia       |                 | ~120,000 BP    | 1           | DG             | [76]      |
| Aghem       | Cameroon         | NC              | Present        | 28          | HO             | This paper |
| Balut       | Cameroon         | NC              | Present        | 11          | HO             | This paper |
| Baka        | Cameroon         | NC              | Present        | 2           | DG             | [20]      |
| Bakofo      | Cameroon         | NC              | Present        | 1           | HO             | This paper |
| Bakola      | Cameroon         | NC              | Present        | 2           | DG             | [20]      |
| Bangwa      | Cameroon         | NC              | Present        | 2           | HO             | This paper |
| Bedzanz     | 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]      |
| Moo         | 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   | NC              | Present        | 20/2        | HO/DG          | [22, 25] |
|             | Republic         |                 |                |             |                 |           |
| 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]      |
| Mouti       | 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. 76.

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

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 $f_4(X, Mursi; Han)$, 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.

| Test pop | $f_4(X, Mursi; SA, Han)$ | $Z$-score | $f_4(X, Mota; SA, Han)$ | $Z$-score | $f_4(X, Han; SA, Mursi)$ | $Z$-score | $f_4(X, Mota; SA, Mursi)$ | $Z$-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      |
| Lemanze  | 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      |
| Mbouti   | 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 | $f_4(X, Mursi; SA, Mota)$ | $Z$-score | $f_4(X, Han; SA, Mota)$ | $Z$-score | $f_4(X, Han; SA, Yor)$ | $Z$-score | $f_4(X, Mursi; Chimpanzee, Yoruba)$ | $Z$-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      | ..                     | ..        | ..                            | ..        |
| Lemanze  | 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      |
| Mbouti   | 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      | ..                            | ..        |
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 | 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 | 58 | 58 | 58 | 58 | 59 | 58 | 62 | 61 | 59 |
| Mbuti Bantu-associated | 26 | 24 | 33 | 19 | 28 | 27 | 26 | 12 | 28 | 30 | 25 | 24 | 26 | 29 | 28 | 35 | 25 | 35 | 25 | 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 | 2 | 2 | 2 | 2 | 2 | 3 | 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 | 4 | 4 | 4 | 5 | 5 | 5 | 4 | 4 | 4 | 3 | 4 |
| Mbuti deep ancestry | 29 | 29 | 30 | 29 | 30 | 31 | 30 | 31 | 29 | 29 | 29 | 29 | 29 | 29 | 29 | 28 | 30 | 31 | 29 | 29 | 27 | 26 | 29 |

Branch lengths

| | 2 | 3 | 3 | 3 | 3 | 1 | 3 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 2 | 2 | 3 | .. | 2 | 3 | 3 | 1 | 3 |
| Basal WA split | 1 | 1 | 0 | 4 | 1 | -1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 4 | 0 | 1 |
| South African Hg split | 1 | 1 | 1 | -3 | 1 | 1 | 0 | -2 | 1 | 0 | -1 | 1 | 1 | 0 | 1 | 1 | 1 | .. | .. | .. | .. | 2 |
| Ghost modern human split | 1 | 1 | 1 | -3 | 1 | 1 | 0 | -2 | 1 | 0 | -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, Tafaralt 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).
Author Queries

Journal: *Nature*
Paper: s41586-020-1929-1
Title: Ancient West African foragers in the context of African population history

**AUTHOR:**
The following queries have arisen during the editing of your manuscript. Please answer by making the requisite corrections directly in the e.proofing tool rather than marking them up on the PDF. This will ensure that your corrections are incorporated accurately and that your paper is published as quickly as possible.

| Query Reference | Reference |
|-----------------|-----------|
| **Q1**          | per preproof comments, an extra affiliation has been added to author Wim Van Neer, and a present address added for author Iñigo Olalde. Please can you check that the extra affiliations are correct? This required a bit of renumbering; please can you double-check to ensure all authors remain linked to the correct affiliations? |
| **Q2**          | This proof has been produced on the basis of your corrections to the preproof. For this later stage of production we use an online ‘eproof’ tool, where you can make corrections directly to the text within the tool and also mark up corrections to the copyedited figures. Please check that the display items are as follows: Figs 0 (black & white); 4 (colour); Tables: one; Boxes: None; Extended Data display items: 7 figures, 3 tables; SI: yes. The eproof contains the main-text figures edited by us and (if present) the Extended Data items (unedited except for the legends) and the Supplementary Information (unedited). Please note that the eproof should be amended in only one browser window at any one time, otherwise changes will be overwritten. Please check the edits to all main-text figures (and tables, if any) very carefully, and ensure that any error bars in the figures are defined in the figure legends. Extended Data items may be revised only if there are errors in the original submissions. If you need to revise any Extended Data items please upload these files when you submit your corrections to this preproof. |
| **Q3**          | Please check your article carefully, coordinate with any co-authors and enter all final edits clearly in the eproof, remembering to save frequently. Once corrections are submitted, we cannot routinely make further changes to the article. |
| **Q4**          | Note that the eproof should be amended in only one browser window at any one time; otherwise changes will be overwritten. |
| **Q5**          | Author surnames have been highlighted. Please check these carefully and adjust if the first name or surname is marked up incorrectly. Note that changes here will affect indexing of your article in public repositories such as PubMed. Also, carefully check the spelling and numbering of all author names and affiliations, and the corresponding email address(es). |
Author Queries

Journal: Nature
Paper: s41586-020-1929-1
Title: Ancient West African foragers in the context of African population history

AUTHOR:
The following queries have arisen during the editing of your manuscript. Please answer by making the requisite corrections directly in the e.proofing tool rather than marking them up on the PDF. This will ensure that your corrections are incorporated accurately and that your paper is published as quickly as possible.

| Query | Reference | Reference |
|-------|-----------|-----------|
| Q6    | You cannot alter accepted Supplementary Information files except for critical changes to scientific content. If you do resupply any files, please also provide a brief (but complete) list of changes. If these are not considered scientific changes, any altered Supplementary files will not be used, only the originally accepted version will be published. |
| Q7    | I have slightly amended one of your preproof crossheadings (changing ‘Shum Laka: genetics and archaeology’ to ‘Genetics and archaeology at Shum Laka’), as we don’t use colons, em dash or similar punctuation in headings. Is the revision OK? Please amend if necessary. |
| Q8    | references 2, 11 and 40 have been edited per preproof comments; please can you check these revisions are correct? |
| Q9    | the html version of the article needs all references to be cited in the text to work correctly (i.e. a reference cannot be only embedded within a display item image). I added the sentence ‘Data for the Altai Neanderthal individual are from ref. 78’ to the legend of Extended Data Table 1 so that this reference is cited; please amend if desired. |
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

- [ ] The exact sample size ($n$) for each experimental group/condition, given as a discrete number and unit of measurement
- [X] An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- [X] The statistical test(s) used AND whether they are one- or two-sided
- [ ] Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- [ ] A description of all covariates tested
- [ ] A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- [X] A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- [X] For null hypothesis testing, the test statistic (e.g. $F$, $t$, $r$) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted
  
  Give $P$ values as exact values whenever suitable.
- [X] For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- [X] For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- [X] Estimates of effect sizes (e.g. Cohen's $d$, Pearson's $r$), indicating how they were calculated
- [X] Clearly defined error bars
  
  State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection

- BWA v0.6.1, HaploGrep2 v2.1.19, cTools v11/2016, contamMix v1.0-10, ANGSD v0.923, OxCal v4.3.2

Data analysis

- ADMIXTOOLS v5.1, EIGENSOFT v7.2.1, IBM SPSS Statistics v.25, SYSTAT v.11.5

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on 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.
Ecological, evolutionary & environmental sciences study design

All studies disclose on these points even when the disclosure is negative.

| Study description | Genetic analyses were performed on DNA data generated from ancient human skeletons. Population genetic statistics, primarily testing historical relationships by measuring allele-sharing patterns across populations, were computed using genome-wide SNP genotypes. |
|-------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Research sample   | Four ancient human individuals buried at Shum Laka in Cameroon; 63 newly reported present-day individuals from five populations in Cameroon (for comparative analyses); ancient African individuals from Gallego Llorente et al. 2015, Schlebusch et al. 2017, and Skoglund et al. 2017; present-day African individuals from Mallick et al. 2016 and Fan et al. 2019. |
| Sampling strategy | We sampled available bones from six Shum Laka individuals and obtained working data from four. We targeted approximately 1.2 million genome-wide SNPs, which effectively cover almost all independent loci (due to linkage disequilibrium) and provide good power in population history analyses (all remains in this study yielding good coverage). |
| Data collection   | DNA from the ancient remains was extracted, sequenced, and processed into SNP genotype calls (final data overseen by D.R. and S.Mal.). |
| Timing and spatial scale | Ancient individuals were sampled from the Shum Laka rockshelter in western Cameroon. Newly reported present-day individuals were from western and southwestern Cameroon. |
| Data exclusions   | Two of the sampled skeletons did not yield working data as assessed by pre-established ancient DNA quality criteria. |
| Reproducibility  | All attempts to reproduce were successful. |
| Randomization     | The only subgroup assignments were by date for the ancient individuals and by ethno-linguistic group for the present-day individuals. |
| Blinding          | Analyses were performed either for all four individuals separately, all four grouped together, or for the two temporally separated pairs; other sample-specific features were not relevant to results. |

Did the study involve field work?  No

Reporting for specific materials, systems and methods

| Materials & experimental systems | Methods |
|---------------------------------|---------|
| Involved in the study | Involved in the study |
| n/a | n/a |
| Unique biological materials | ChIP-seq |
| Antibodies | Flow cytometry |
| Eukaryotic cell lines | MRI-based neuroimaging |
| Palaeontology | |
| Animals and other organisms | |
| Human research participants | |

Human research participants

Policy information about studies involving human research participants

Population characteristics | Samples were collected at market places or other locations of a similar nature where many individuals of the ethnic group of interest were expected to reside or assemble. Samples were only collected from males who declared themselves to be 18 years of age or older and not the sibling of another donor. The self-declared ethnic identity of the donor was recorded after the sample was taken, but samples were collected without discrimination on the basis of ethnic identity. There was no genotypic...
data on donors available to collectors; subsequently, for analysis, we confirmed that none of the individuals were substantial genetic outliers relative to their ethno-linguistic groups.

**Recruitment**

Participants were recruited by public announcement at the sampling location with samples taken in the order donors presented themselves. No monetary reward was offered.