



**University of  
Zurich**<sup>UZH</sup>

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2024

---

## **Deep history of cultural and linguistic evolution among Central African hunter-gatherers**

Padilla-Iglesias, Cecilia ; Blanco-Portillo, Javier ; Pricop, Bogdan ; Ioannidis, Alexander G ; Bickel, Balthasar ;  
Manica, Andrea ; Vinicius, Lucio ; Migliano, Andrea Bamberg

DOI: <https://doi.org/10.1038/s41562-024-01891-y>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-260745>

Journal Article

Published Version



The following work is licensed under a Creative Commons: Attribution 4.0 International (CC BY 4.0) License.

Originally published at:

Padilla-Iglesias, Cecilia; Blanco-Portillo, Javier; Pricop, Bogdan; Ioannidis, Alexander G; Bickel, Balthasar; Manica, Andrea; Vinicius, Lucio; Migliano, Andrea Bamberg (2024). Deep history of cultural and linguistic evolution among Central African hunter-gatherers. *Nature Human Behaviour*:online.

DOI: <https://doi.org/10.1038/s41562-024-01891-y>

# Deep history of cultural and linguistic evolution among Central African hunter-gatherers

Received: 7 March 2023

Accepted: 18 April 2024

Published online: 27 May 2024

 Check for updates

Cecilia Padilla-Iglesias<sup>1</sup>✉, Javier Blanco-Portillo<sup>2</sup>, Bogdan Pricop<sup>3</sup>, Alexander G. Ioannidis<sup>4</sup>, Balthasar Bickel<sup>3,5</sup>, Andrea Manica<sup>6</sup>, Lucio Vinicius<sup>1</sup> & Andrea Bamberg Migliano<sup>1,5</sup>✉

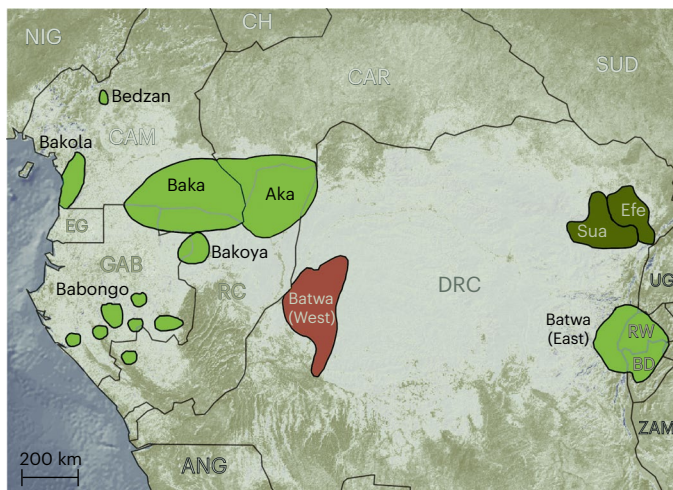
Human evolutionary history in Central Africa reflects a deep history of population connectivity. However, Central African hunter-gatherers (CAHGs) currently speak languages acquired from their neighbouring farmers. Hence it remains unclear which aspects of CAHG cultural diversity results from long-term evolution preceding agriculture and which reflect borrowing from farmers. On the basis of musical instruments, foraging tools, specialized vocabulary and genome-wide data from ten CAHG populations, we reveal evidence of large-scale cultural interconnectivity among CAHGs before and after the Bantu expansion. We also show that the distribution of hunter-gatherer musical instruments correlates with the oldest genomic segments in our sample predating farming. Music-related words are widely shared between western and eastern groups and likely precede the borrowing of Bantu languages. In contrast, subsistence tools are less frequently exchanged and may result from adaptation to local ecologies. We conclude that CAHG material culture and specialized lexicon reflect a long evolutionary history in Central Africa.

Recent fossil and genetic findings have revised the origins of *Homo sapiens* from less than 120,000 years<sup>1</sup> ago to almost half a million years ago<sup>2–5</sup>. Genomic analyses have revealed that some of the oldest human lineage divergences are represented by various extant African hunter-gatherer groups (that is, populations that primarily practise a foraging lifestyle for subsistence). Such groups include Khoisan-speaking hunter-gatherer groups in southern and eastern Africa<sup>6–8</sup>. In contrast, other studies have postulated that an encapsulation and isolation of Central African hunter-gatherers (CAHGs) from each other<sup>9–11</sup> and gradual intermixing with farming neighbours<sup>12</sup> would have resulted in a shallow cultural history<sup>12–15</sup>. Such a view has also been proposed based on their universal adoption of languages from neighbouring farmers, and therefore the absence of a common CAHG

language or language family<sup>16,17</sup> (although see refs. 17,18). For example, the 10 CAHG groups in our study (7 western and 3 eastern groups) speak 13 languages from 3 highly differentiated linguistic families that are the product of farming expansions (Fig. 1 and Supplementary Tables 1–3).

Recent ecological, genetic and archaeological analyses have cast doubt on this interpretation, revealing various signs of long-term adaptation of CAHGs to current environments and independence from Bantu demographic history. For example, ref. 19 showed that ecological and climatic features of the Congo Basin can successfully predict 120,000 years of hunter-gatherer between-group interconnectivity, genetic exchange and continuous forest occupation. These results are in agreement with archaeological evidence<sup>20</sup>, the deep genetic coalescence of CAHGs with other human lineages<sup>3,21</sup> and identification of

<sup>1</sup>Human Evolutionary Ecology Group, Institute of Evolutionary Anthropology, University of Zurich, Zurich, Switzerland. <sup>2</sup>Department for Biology, Stanford University, Stanford, CA, USA. <sup>3</sup>Department of Comparative Language Science, University of Zurich, Zurich, Switzerland. <sup>4</sup>Department of Biomedical Data Science, Stanford Medical School, Stanford, CA, USA. <sup>5</sup>Center for the Interdisciplinary Study of Language Evolution, University of Zurich, Zurich, Switzerland. <sup>6</sup>Department of Zoology, University of Cambridge, Cambridge, UK. ✉e-mail: [cecilia.padillaiglesias@uzh.ch](mailto:cecilia.padillaiglesias@uzh.ch); [andrea.migliano@uzh.ch](mailto:andrea.migliano@uzh.ch)



**Fig. 1 | Distribution of sampled CAHGs.** Genetic data were not available for the Western Batwa (shown in pink). Efe and Sua (dark green) were grouped into a single genetic unit (Mbuti). Cultural groups shown in light green correspond to a single genetic group. Abbreviated country names are indicated in capital letters. ANG, Angola; BD, Burundi; CAM, Cameroon; CAR, Central African Republic; CH, Chad; DRC, Democratic Republic of Congo; EG, Equatorial Guinea; GAB, Gabon; NIG, Nigeria; RC, Republic of Congo; RW, Rwanda; SUD, South Sudan; UG, Uganda; ZAM, Zambia. Map made with Natural Earth.

genes under selective pressure<sup>21–25</sup>. Finally, ethnographic studies have also questioned the isolation of CAHGs from one another, highlighting their widespread trips throughout the Congo Basin<sup>26,27</sup> to participate in rituals or search for spouses<sup>27–29</sup>.

As proposed by the foraging niche hypothesis<sup>30</sup>, a deep history of interpopulation connectivity across Africa would have various implications for our understanding of human origins and cultural diversity<sup>31,32</sup>. For example, rather than being the product of recent borrowing from farming groups, as is undoubtedly the case with language, other dimensions of CAHG culture such as musical instruments or subsistence tools may be the outcome of long-term, independent cultural evolution<sup>12,33–35</sup>. A long history of cultural exchange in eastern and southern Africa is exemplified by a system of bead transfers dated to 50,000 years ago<sup>36</sup> and by the long-distance trade of obsidian over 160 km nearly 200,000 years ago<sup>37</sup>. However, no study has investigated the structure of cultural diversity among hunter-gatherers in the large Central African region where CAHG groups may live as far as 2,000 km apart. Hence, showing the antiquity of cultural exchanges among CAHGs would provide key evidence for the role of long-term, continent-wide social networks in the cultural evolution of *H. sapiens*<sup>4,30,38</sup>.

To investigate the possible effects of genetic history and group interconnectivity on cultural evolution in the Congo Basin, we integrated genetic data and cultural data on musical instruments, foraging tools and specialized vocabulary of CAHGs (Fig. 1). These three cultural domains are thought to be subject to different cultural evolutionary pressures due to the different functions they serve in society<sup>32,39,40</sup>. We compiled a comprehensive dataset of musical instrument repertoires, subsistence tools and their names in ten CAHG groups from ethnographies, photographs and museum collections (Fig. 1, Methods and Supplementary Dataset 1). We also obtained genome-wide single-nucleotide polymorphism (SNP) data from  $N = 382$  individuals from 10 populations grouped into 9 units (Sua and Efe were grouped into a Mbuti group; Supplementary Table 4). We introduced a procedure for establishing the timing of cultural exchange events by performing genetic analyses of CAHG groups at three evolutionary depths: a full dataset incorporating segments resulting from Bantu introgression; a dataset including only genomic segments of CAHG origin, after masking away Bantu segments; and a set including only

genomic segments of deep CAHG ancestry not resulting from recent gene flow among CAHG populations. Our procedure shows that splitting genetic evidence into ancestry layers can reveal deep associations with cultural exchange and long-term interconnectivity among CAHGs independently from farming groups. We conclude that the early evolutionary history of CAHGs involved cultural exchanges across large-scale social networks and that the specialized cultural lexicons in CAHGs can provide insights into hunter-gatherer languages predating the expansion of farming groups in Central Africa.

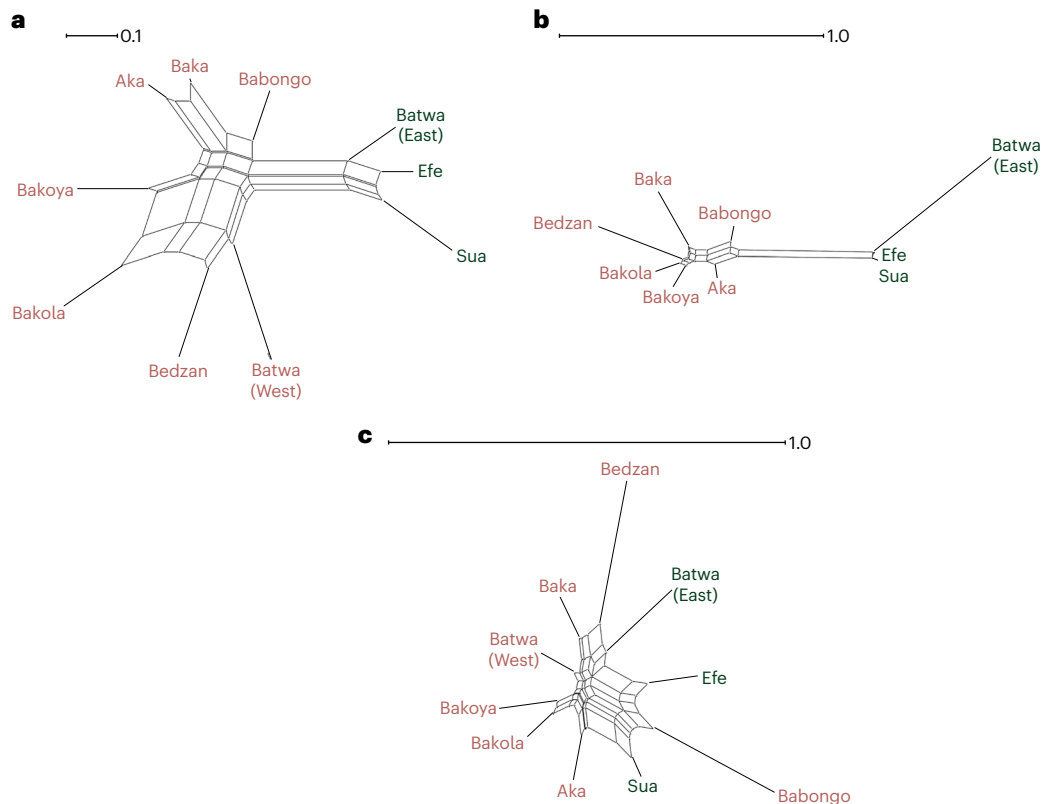
## Results

### Musical instruments of CAHGs result from cultural exchange

We first asked whether cultural features of current CAHGs result from a deep history of evolution or were recently adopted from farmers in the same manner as their languages. We examined whether the distribution of CAHG musical instruments showed geographical patterning indicative of a shared genetic ancestry, or instead revealed stronger similarities between each CAHG group and its farming neighbours as a sign of recent cultural borrowing from farmers<sup>41–44</sup>. We calculated Jaccard distances between populations<sup>45,46</sup> based on the presence or absence of the 44 instrument types from our compiled dataset across 10 CAHG groups (mean of 14 instrument types per group; range between 10 and 25; Supplementary Fig. 1). The visualized results of a principal coordinate analysis (PCoA) using the NeighborNet algorithm revealed a clear geographical patterning, with defined western (Bedzan, Western Batwa, Bakola, Bakoya, Aka, Baka and Babongo) and eastern groups (Efe, Sua and Eastern Batwa) (Fig. 2a and Supplementary Figs. 2 and 3). Mantel tests and congruence among distance matrices (CADM) analyses provided statistical confirmation of the pattern, revealing a significant correlation between geography and music (Mantel statistic  $\rho = 0.408$ , adjusted  $P = 0.024$ ; see Supplementary Table 5 for multiple matrix regressions and Supplementary Table 6 for CADM).

Next we assessed a possible link between the distribution of musical instruments and genetic affinities between CAHG populations. We began by examining whether CAHG genetic data also showed similar signs of geographical substructuring and differentiation independent of farmers. For this, we masked away Bantu-associated ancestry from all CAHG samples using a local ancestry inference method (Gnomix<sup>47,48</sup>). Following imputation for masked SNPs, multidimensional scaling (MDS; Methods) revealed affinities among CAHGs that largely mirrored geographical distances<sup>49–51</sup>, with the first dimension segregating western from eastern CAHG groups and a second dimension separating northwestern and southwestern groups (Supplementary Fig. 4). Such spatial patterning was less clear when using either the unmasked dataset or the dataset including only genomic segments associated with Bantu ancestry (Supplementary Figs. 5–9). We found a significant correlation between genetic distances (pairwise fixation index ( $F_{ST}$ ) measures exclusively based on CAHG ancestry genomic segments) and geographical proximity (Mantel statistic  $\rho = 0.769$ ,  $P = 0.001$ ; Supplementary Tables 6 and 7 for CADM results) even after controlling for ecological distance (measured as similarity in biome composition between CAHG territories) using multiple matrix regressions (Supplementary Figs. 10 and 11 and Supplementary Table 8). This was also revealed by our NeighborNet plot of between-group genetic distances (pairwise  $F_{ST}$ )<sup>52</sup> (Fig. 2b).

Having found evidence for the spatial structuring of both CAHG cultural and genetic diversity, we evaluated their potential associations. We implemented an isolation-by-distance model using SpaceMix<sup>53</sup> (Methods) whose null hypothesis is that the spatial distribution of CAHG genetic or cultural diversity reflects neutral diffusion among neighbouring populations, rather than substantial demic processes such as population replacement. We obtained a map in which mean group distances based on CAHG genetic pseudo-coordinates are smaller than their actual geographical distances ( $t$ -test with 80 degrees of freedom)  $t(80) = 3.8849$ ,  $P < 0.001$ , indicating widespread gene flow across Central Africa and hence deviation from the null hypothesis (Supplementary



**Fig. 2 | Genetic and cultural distances between CAHGs. a–c,** NeighborNet graphs based on dimensionality reduced distance matrices for musical instruments (a), genes using exclusively CAHG ancestry segments (b) and

subsistence tools (c) (see Methods for details). In pink, Western CAHG groups; in green Eastern CAHG groups. Scale bars indicate the distance from the Neighbournet joining algorithm in the original units of the data.

Fig. 12). SpaceMix analysis detected signs of recent gene flow even in the most geographically isolated populations, such as the Bedzan (12.8% of their genome originating in other CAHGs) or the Eastern Batwa (6.3%) (Supplementary Fig. 13 and Supplementary Table 9). Similarly, musical instrument pseudo-coordinates were closer to the actual CAHG geographical coordinates than to genetic pseudo-coordinates and also overall closer to one another than their actual geographical locations (Supplementary Fig. 12 and Supplementary Table 10). Overall our results suggest that demographic exchanges have led to the transfer of both genes and musical instruments among CAHG groups.

### Identical-by-descent segments reveal history of musical instrument exchange

We then investigated the time depth of cultural exchanges among CAHGs. Although the origin of CAHG musical instruments cannot be directly dated, they must be either a product of long-term evolution in Central Africa or only recently exchanged among CAHGs after farming populations expanded into the region and admixed with local hunter-gatherers. The length of shared identical-by-descent (IBD) fragments has been widely used for dating genetic exchanges of recent but nonetheless variable antiquity between populations<sup>50,54–56</sup>, including CAHGs<sup>19</sup>, with longer shared IBD blocks being the product of more recent admixture events (as fewer recombination events have taken place in admixed individuals)<sup>57</sup> (Supplementary Fig. 14). Therefore, we propose a method that searches for associations between cultural exchange and genetic events dated by IBD segments (Supplementary Text 1). This method is based on estimating correlations between musical instrument similarity and genetic similarity after splitting our CAHG genomic dataset into three levels: first, all genomic fragments, including those recently introgressed from Bantu populations; second, all fragments of CAHG ancestry (by masking away Bantu genomic segments); and, third, the remaining CAHG fragments after the exclusion of IBD fragments recently

exchanged among CAHGs. It should be noted that the last set contains both fragments of more ancient ancestry and others whose antiquity cannot be established (Fig. 3 and Supplementary Tables 11 and 12).

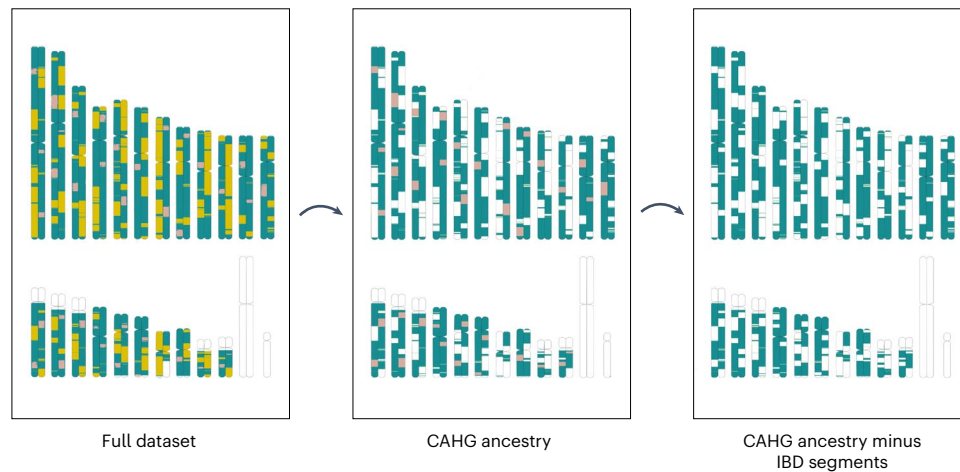
This approach implies various predictions: (1) if we only find a significant correlation between musical instrument and genetic similarity in the full genomic dataset, musical instruments have most likely been recently introgressed from neighbouring farmers, together with genes; (2) if we instead find correlations between similarity in musical instrument repertoires and genomic components exchanged among CAHGs, musical instruments are more likely to have been originated before the arrival of Bantu farmers, but continued to be exchanged among hunter-gatherers in recent times; and (3) if musical instruments only correlate with the remaining genomic fragments after the exclusion of IBD fragments exchanged among CAHGs, the distributions of both genetic fragments and musical instruments are likely to be the result of the pre-farming shared history of CAHGs.

Mantel tests showed that CAHG genetic similarity based on the full dataset including Bantu fragments did not correlate with musical instrument similarity, a result also obtained when we used only Bantu fragments (Table 1; Supplementary Table 6 for CADM results). In contrast, genetic similarity based either on the set of all fragments of CAHG ancestry or on the set excluding IBD fragments recently shared between CAHGs shows significant and virtually identical correlations with musical instrument similarity. This suggests that genetic and cultural exchanges between hunter-gatherer groups are ancient and have remained mostly unaffected after recent farming expansions, pointing to a deep history of hunter-gatherer networks and uninterrupted cultural interconnectivity across the Congo Basin.

### Music lexicon predates the arrival of Bantu languages

Although CAHGs speak widely differentiated and mutually unintelligible languages recently acquired from farming groups (Supplementary





**Fig. 3 | Fractional analysis of IBD fragments.** Schematic representation of the three genetic sets obtained through a two-step masking process on a human karyotype. The first set includes all genomic segments including segments of Bantu origin (yellow) and shared IBD segments between individuals from different CAHG populations (green and pink). The second set includes only

genomic segments of CAHG ancestry (green and pink) after the removal of Bantu segments. The third set includes the remaining genomic segments of CAHG ancestry (green) after removing all IBD segments exceeding 1 cM shared between individuals from different CAHG populations.

**Table 1 | Mantel tests of correlations between cultural distances (musical instruments and subsistence tools) and geographic, ecological and genetic distances**

	Musical instruments		Subsistence tools	
	Mantel rho statistic	Adjusted P	Mantel rho statistic	Adjusted P
Geography	<b>0.408</b>	<b>0.024</b>	0.207	0.152
Ecology	0.116	0.441	<b>0.480</b>	<b>0.034</b>
Genes (full dataset)	0.312	0.135	0.087	0.411
Genes (CAHG ancestry)	<b>0.416</b>	<b>0.032</b>	0.312	0.113
Genes (CAHG ancestry without IBD segments)	<b>0.422</b>	<b>0.032</b>	0.215	0.162
Genes (Bantu ancestry)	-0.01	0.508	-0.242	0.729

Genetic data were split into distinct ancestry levels: all genomic segments, segments of CAHG ancestry, and segments of CAHG ancestry excluding shared IBD segments between CAHG and segments of Bantu ancestry. *P* values were adjusted using the Benjamini-Hochberg procedure. Bold numbers represent significant statistics at an adjusted significance threshold of 0.05.

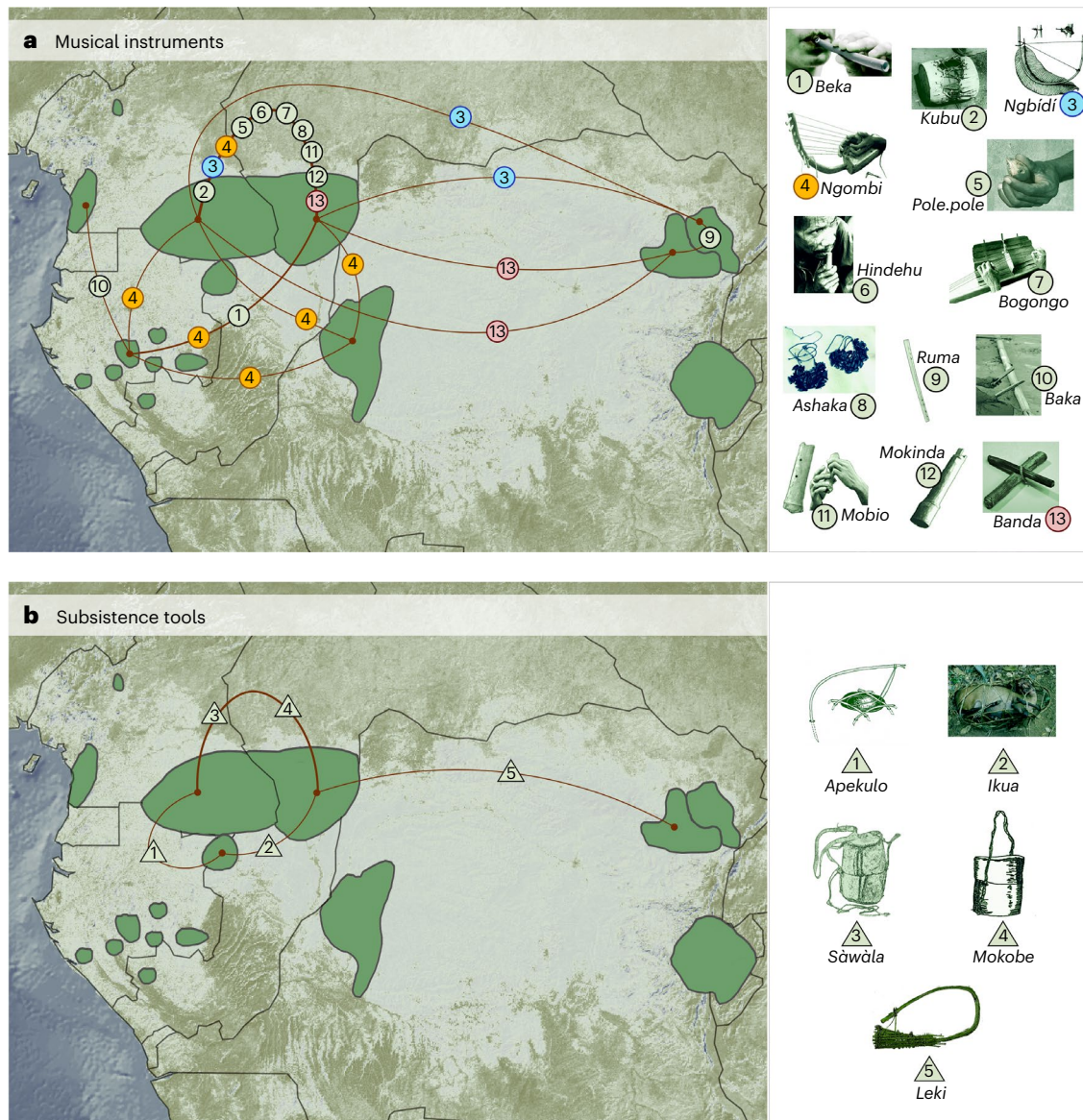
Figs. 15 and 16 and Supplementary Tables 2 and 3), our results suggest that their musical instruments have a more ancient origin. Hence, either CAHGs adapted words from farmer languages to name their instruments or they preserved the original names from their lost languages. We conducted an analysis of the specialized lexicon for musical instruments across CAHG groups to identify music-related words possibly predating the spread of farming groups and their languages. We compiled from primary sources a total of 183 words designating musical instruments from 10 groups (mean words per group = 18, range = 3–46). Despite having lost their original languages, CAHGs share 23 music-related words: 17 words shared between 2 groups, 3 among 3 groups, 3 among 4 groups and 1 among 5 groups (Supplementary Figs. 15 and 17, Supplementary Table 13, Supplementary Dataset 2 and Methods). Of those 23 words, 15 are exclusive to hunter-gatherer populations and not found among any of the farming neighbours of CAHGs (Fig. 4).

For example, the word *ngbidi* (see Supplementary Table 13 for variants) denotes a musical bow with two strings that is played exclusively by female CAHGs in the exact same manner by both eastern (Efe) and western (Aka and Baka) groups living over 2,000 km apart and not found in any other population<sup>58</sup>. Another shared word exclusive to

CAHGs is *ngombi*, denoting the harp present in the Baka, Babongo, Aka and Batwa in the west (the latter being at least 400 km away from the other 3 groups). It is also relevant that Aka (Bantu C10 speakers) and Baka (Ubangian speakers) neighbours speak very distinct languages and yet share various unique words denoting exclusively shared musical instruments, such as *haka* (ankle rattles), *bogongo* (zither harp), *pole.pole* (seed whistle), *mobio* (flute) and *mokinda* (single-skinned drum). The presence of these instruments and words shared between different CAHG groups but not with their respective Bantu-speaking and Ubangian farming neighbours nor with other farming groups speaking related languages has previously been interpreted as evidence for their descent from a common ancient population<sup>43,59</sup>.

To confirm that the sharing of words for musical instruments between populations was not the product of inheritance from the same or closely related farmer languages, we calculated linguistic distances between all pairs of CAHG languages using three methods. First, we used the Automated Similarity Judgment Program (ASJP) database to calculate pointwise mutual information (PMI) distances between all pairs of CAHG languages<sup>60,61</sup>. PMI has been shown to be informative for performing distance-based phylogenetic inference and it provided a measure of general vocabulary similarity between CAHG groups (Supplementary Fig. 16)<sup>61,62</sup>. We then calculated phylogenetic distances by counting the number of internal nodes between each pair of CAHG languages using the Glottolog trees<sup>63</sup>. For the subset of Bantu languages in our sample, we also obtained phylogenetic distances based on the most recent dated phylogeny of Bantu languages<sup>64</sup> (Methods). The correlation between PMI distances and Glottolog distances, as well as between PMI distances, and phylogenetic (patristic) distances were high (Spearman  $\rho = 0.60$ ,  $S = 6353.1$ ,  $P < 0.001$ ; and Spearman  $\rho = 0.86$ , (Spearman's rank order correlation)  $S = 2223.7$ ,  $P < 0.001$ , respectively). Our rationale was that if the sharing of words for musical instruments between CAHG groups results from their current languages being closely related, groups speaking more closely related languages should share more music-related words. If instead shared musical instrument words are (as with the musical instruments themselves) the product of a long cultural evolutionary history predating recent language replacement, genetically closer CAHG populations should share more words for musical instruments than linguistically related populations.

Zero-inflated Poisson models showed that genetic and geographical proximity were significant predictors of the number of shared musical instrument words between populations, whereas phylogenetic linguistic proximity and overall language similarity (that is, smaller



**Fig. 4 | Networks of specialized vocabulary for musical instruments and subsistence tools exclusive to CAHGs. a,b**, Shared vocabulary representing musical instruments (a) and subsistence tools (b) in CAHGs. Edge widths reflect the number of unique shared words between two populations. Each symbol (circles for musical instruments; triangles for subsistence tools) indicates one shared word between populations. Numbers inside symbols denote the corresponding musical instrument or subsistence tool. Symbol colours were

added to facilitate visualization of instruments for which the corresponding word was shared between more than two groups. Map made with Natural Earth. Image credits for panel a: 1, 6, 10, 11, Africa Museum Musical Instrument collection, Tervuren; 2, 4, 5, 7–9, 12, 13, Susanne Fűrinniss; 3, ref. 43. Image credits for panel b: 1, 5, ref. 110; 2, Nikhil Chaudhury; 3, Africa Museum Musical Instrument collection, Tervuren; 4, ref. 18. All images redrawn by Rodolph Schlaepfer.

patristic, PMI or Glottolog distances) were not (Supplementary Tables 14–19). In fact, when considering patristic distances, we found the opposite to be true, with groups most distantly related linguistically sharing more musical instrument words (Supplementary Tables 16 and 19). This association between shared musical instrument lexicon with deep population history, but not with overall linguistic relationships or similarities, held both when considering all shared terms between CAHGs and when only considering those unique to CAHGs. For example, the two groups with the most similar general lexicons were the Aka and the Eastern Batwa (PMI = 0.71), which share no musical instrument words. In contrast, the Aka and the Baka are the genetically closest groups in our sample and share the greatest number of musical instrument words ( $N = 15$ , 12 of which are unique to CAHGs) but their general lexicons are substantially more distant (PMI = 0.81) than our sample average (PMI = 0.74).

Our findings also match popular legends in Central Africa describing a deep shared ancestry of musical instruments predating CAHG divergence and the common origins of the single-stringed musical bow *gongo* (shared across different CAHG groups) and of the CAHGs themselves<sup>65,66</sup>. Overall music-related words used by current CAHGs seem to represent words that originated in their extinct languages and the distribution of these words offers a glimpse at a shared linguistic ancestry across CAHGs<sup>43,59</sup>.

### Subsistence tools represent adaptations to local environments

We also compiled complete subsistence tool repertoires from 10 CAHG groups (Methods), obtaining a total of 40 and average of 17 types of subsistence tool per group (range = 9–25; Supplementary Fig. 1). The distribution of tools among CAHG groups followed a distinct pattern



from that of musical instruments. First, the number of subsistence tools and musical instruments in groups were not correlated (Spearman  $\rho = -0.073$ ,  $S = 116.95$ ,  $P = 0.839$ ). Mantel tests, multiple matrix regressions and CADM analyses identified no correlations between tool sharing and geographical distance (Table 1 and Supplementary Tables 5 and 7). SpaceMix analyses showed that the distance between the geographical coordinates of groups and their pseudo-coordinates estimated from subsistence tools was generally much greater (for 7 of 9 CAHG groups) than those between geographical coordinates and pseudo-coordinates estimated from musical instruments (Wilcoxon signed-rank test statistic  $W = 20$ ,  $P = 0.038$ ; Supplementary Fig. 18). As a result, the structure of spatial variation in subsistence tools showed a much worse match with the geographical distribution of groups (Supplementary Fig. 12). A NeighborNet visualization of PCoAs based on subsistence toolkit repertoires failed to identify the separation between western and eastern groups (Fig. 2c), between adjacent populations (Aka and Baka or Sua and Efe; Table 1 and Supplementary Fig. 2) or any clear geographical pattern. Finally, we found no correlation between tool similarity and genetic similarity when considering any of the three genomic sets (Table 1 and Supplementary Table 7). This indicates that diffusion of subsistence tools between groups has not regularly occurred during events of genetic exchange.

The only significant predictor of foraging tool distribution was similarity in ecology (biome composition; Table 1 and Supplementary Table 7 for CADM results). We noticed a nearly universal presence across populations of certain items, such as spears, hunting bows or arrows, intrinsically linked to the hunter-gatherer niche and therefore not expected to respond to ecological variation. However, the dataset also included more sparsely distributed items related to resource exploitation in specific ecological settings, such as hooks present only in the Baka and Bedzan, which are two geographically distant groups that are equally dependent on aquatic resources. The association with ecology but independence from demographic exchanges and cultural diffusion suggests that the distribution of such items may be more parsimoniously explained by convergent adaptation to similar environments.

Next we asked whether an ecology-driven pattern would also underlie the distribution of words for subsistence tools. From primary sources we identified 89 words designating subsistence tools in 9 populations (mean number of words per group = 9, range = 0–28; Supplementary Fig. 17 and Supplementary Table 13). In contrast to the more extensive sharing of musical terms, only ten tool names were shared between two groups and one name between three groups, whereas only three names denoted objects unique to CAHGs. We found that words designating musical instruments were more likely to be shared between CAHG groups than words designating subsistence tools ( $W = 204$ ,  $P < 0.001$ ) but the limited number of shared words related to tools precluded further statistical assessment. In summary, tool names are much less commonly shared than musical terms, even among groups sharing similar ecologies.

## Discussion

Our results provide evidence that long-term, large-scale social networks underlie a deep history of material cultural and lexical evolution among CAHGs. Such a deep history is one of the many social and evolutionary consequences of the high rates of between-group mobility evolved in hominin ancestors and is still observed in extant hunter-gatherers<sup>30,67</sup>. These consequences include the maintenance of high levels of cooperation in the absence of explicit punishment mechanisms<sup>68</sup> or the unique pattern of coresidence of unrelated individuals observed in hunter-gatherer camps<sup>69</sup>. At a regional scale, between-group mobility generates multilevel networks in which household clusters and camps are interconnected into larger, multicamp structures<sup>70</sup>. Such structures represent properties that facilitate cultural evolution and knowledge specialization through dedicated channels of social transmission<sup>31,32,71</sup>.

Recently, we provided evidence for the long-term existence of such networks at a much larger scale, spanning the entire Central Africa for at least the last 120,000 years based on paleoclimatic reconstruction and archaeological data<sup>19</sup>. By combining cultural, linguistic and genetic data in the current study, we have shown that although some aspects of CAHG culture such as language are evolutionarily recent and borrowed from Bantu- and Ubangian-speaking groups, other domains reveal signs of long-term evolution<sup>12,17</sup>. Although extensive demographic and cultural exchanges with incoming farming populations undeniably took place over the last millennia, interconnectivity between CAHGs was not interrupted, thus preventing their isolation from each other<sup>5,19,72,73</sup>. For this reason, CAHG material culture is both the product of long-term contact with other hunter-gatherers and of recent exchanges with farming groups. This ability of maintaining large-scale social networks throughout the evolutionary history of our hunter-gatherer ancestors might represent a fundamental step in the divergence between hominins and our closest relatives<sup>30</sup>.

The intricacy of CAHG cultural evolution is expressed in the distinct patterns of geographical distribution observed in musical instruments, subsistence tools and specialized lexicon. Musical instruments present a distribution suggesting exchanges over thousands of kilometres between western and eastern regions of the Congo Basin. We have introduced an analytical approach that decomposes genomic fragments into sets representing distinct ancestry depths and in this way we were able to show that the distribution of musical instruments across CAHGs is linked to genomic fragments not derived from recent exchanges with Bantu farmers. This suggests that splitting the genomic evidence into fractions of distinct antiquity can at least reject certain evolutionary and demographic scenarios, as we have exemplified above by ruling out a recent Bantu origin for musical instruments found in CAHG groups. Importantly, this approach can provide a clear link between the distribution of a cultural trait and an underlying pattern of demographic contact among populations.

As shown above, we identified an association between demographic exchanges among CAHGs with musical instruments not interrupted by contact with farming groups. This was not the case for subsistence tools, which suggests a possible long-term role of music in establishing cultural group identities. Music is an integral part of CAHG rituals that often differs among cultural groups and brings together individuals living far apart<sup>28,29</sup>. A common function of CAHG rituals is to allow members from distinct groups to meet unrelated cooperative or marriage partners, which may provide a social and mechanistic explanation for the correlation between diffusion of genes and musical instruments<sup>28,29</sup>. In fact, it has been reported that rituals in which the Aka and Baka use similar musical instruments are also named the same in both populations, such as the ritual to invoke the success of a quest (*zòbòkò*, *ndàmbo*, *è.sà* and *mbèlà*), to prepare for the death of an animal (*mò.nzòli* and *kóbá*) or to celebrate the first capture of an important animal by a young man (*mò.póndí*)<sup>42</sup>.

In contrast, the significant correlation between the distribution of subsistence tools and local ecological conditions provides strong evidence for long-term adaptation of CAHG groups to forest environments in the Congo Basin. The deep adaptive history of foraging tools may explain why ecological variables can successfully predict the distribution of both current and past hunter-gatherers in Central Africa<sup>19</sup>. A similar pattern was found among San hunter-gatherer groups across the Kalahari Desert, where shared environments could explain 90% of the variance in their projectile points<sup>74</sup>.

We have also shown that although CAHGs have adopted languages from farmers, music-related words were most likely inherited from ancestral populations together with the corresponding musical instruments. In contrast, the fact that very few tool-related words are shared between CAHGs may reflect either a convergent origin of the similar tools in groups with divergent languages (and hence distinct words created for similar tools) or an ancient common ancestor to both

subsistence tools and tool-related words, with tool (but not word) differentiation over time being limited by ecological constraints. Overall the distribution of music-related words across Central Africa points to an ancient linguistic network interconnecting western and eastern populations<sup>43</sup>. We believe that extending our procedure may lead to the identification of other ancient shared linguistic terms in cultural domains known to precede the arrival of farming groups, such as vocabulary related to forest, forest life, rituals regarding the forest and kin relationships<sup>18,66</sup>.

In summary, our results provide cultural, linguistic and genetic evidence that extant CAHGs are the outcome of a long history of between-group interconnectivity and occupation of forest environments in the Congo Basin. They also point to a key role played by extended social networks in circulating material culture, words and genes. Although current CAHGs have also maintained extensive contact with neighbouring farmers, we have shown that genetic and cultural evolution has taken place in Central Africa as a result of long-term and large-scale exchanges among CAHGs groups. Future studies should involve more comprehensive genetic, linguistic and cultural analyses including CAHGs alongside other African hunter-gatherers from southern and eastern Africa. Moreover, they should aim at including multivariate analyses of cultural and linguistic variation (beyond presence-absence) to pinpoint the drivers of biological and cultural evolution at a continental scale.

## Methods

### Material cultural corpus

We obtained data on musical instruments and subsistence-related tools (hunting, gathering and fishing) from  $N = 115$  primary sources, including ethnographic accounts and photographic and video material specifically designed for documenting the cultural repertoires of a specific group in those different domains, and the ethnographic collection of the Royal Museum for Central Africa in Tervuren (Belgium) (Supplementary Dataset 1). Note that the latter comprised multiple primary sources. All the records were from the twentieth and twenty-first century. In all cases, we contrasted the data provided by every source with at least one other source and, in most instances, personally contacted the relevant ethnographers to confirm the presence or absence of each of the recorded objects. As different names are sometimes used to designate the same CAHG population, we only included data for which specific collection locations were indicated. When available, we also recorded (from the same source) the name or names used in the local language to designate each of the objects. We were able to obtain  $N = 1,366$  material cultural records comprising full repertoires of musical instruments and subsistence items from 10 CAHG groups (Supplementary Table 1 and Supplementary Dataset 1). Overall we documented 44 different types of musical instrument and 40 different types of subsistence tool. At the same time, we compiled  $N = 183$  different words to designate musical instruments from 10 different populations and  $N = 89$  words to designate subsistence tools from 9 populations (see Supplementary Fig. 17, Supplementary Table 13, and Supplementary Datasets 1 and 2 for variation in terminology).

### Geographical location and territories

A geographical area was assigned to each of the cultural groups by georeferencing a previously published map<sup>16,17</sup> and spatial polygon objects were created from the territory occupied by each cultural group (Fig. 1). Given that groups inhabited relatively small areas, we took the coordinates for the centroid of each territory as the location for each cultural group<sup>16,17</sup>.

### Genetic samples

Genetic data from CAHGs were obtained by merging 4 different publicly available datasets<sup>24,75–77</sup> comprising individuals from 13 African populations (Supplementary Table 4). Ethical approval for using data in

each of the datasets was granted by the relevant data access committees and signing officials for each of the publicly available data repositories (EGA and dbGaP) and by the Ethics Commission (Ethikkommission) of the University of Zurich (permit number 20.2.8). Each dataset was previously filtered for relatedness and Hardy–Weinberg equilibrium before we accessed the data. The data were merged using PLINK<sup>78</sup> and R<sup>79</sup> so that only SNPs common to all studies were retained. We removed multi-allelic loci, filtered data for relatedness  $>0.0886$  with KING<sup>80</sup> and a random individual was chosen from each related pair. Using PLINK, the data were then filtered to remove SNPs with a minor allele frequency  $<0.05$ . A pruned version of the dataset was created with PLINK –indep-pairwise 50 50.2 to account for linkage disequilibrium.

The non-pruned dataset comprised 511 individuals and 431,867 SNPs from 13 different African populations and the pruned dataset comprised 511 individuals and 193,718 SNPs. For the purposes of this study, we merged BakaG and Baka as a single Baka population and BabongoE and BabongoS as a single Babongo population (Supplementary Table 4). The unpruned dataset was used for Gnomix, SpaceMix, all ancestry-specific analyses and for calculating runs of homozygosity. The pruned dataset was used for ADMIXTURE analyses<sup>81</sup>. All samples were analysed on the GRCh37 (hg19) genome build.

We recorded the geographical location from which each sample was obtained along with the cultural group to which it belonged (Supplementary Table 1). Note that the Mbuti genetic samples attributed to one population encompassed more than one cultural group (Efe and Sua). In this case, the two cultural groups were assumed to not be genetically differentiated from one another.

### Unsupervised ancestry clustering analyses

We performed unsupervised ancestry clustering using ADMIXTURE<sup>81</sup> on both our full pruned dataset and on a reduced subsample of our pruned dataset that included exclusively CAHG populations and a few unadmixed farming populations commonly used as references in population genomic analyses (Supplementary Table 4). ADMIXTURE estimates for every individual the proportions of the genome originating from  $K$  ancestral populations, with  $K$  specified a priori. The program was run at  $K$  values from 1–8 with cross-validation error estimation, default values for fold iteration ( $\nu = 5$ ) and a random seed for each  $K$  value (Supplementary Figs. 19–21). Therefore, for each value of  $K$ , we ran as many iterations as required for the log-likelihood to increase by less than  $\epsilon = 10^{-4}$  between iterations ( $\epsilon$  is the difference in log likelihood; see Supplementary Table 20 for number of iterations required for each  $K$  value). See Supplementary Table 21 for average ancestry proportions.

### Local ancestry inference and masking

We used Gnomix<sup>47</sup> on our dataset to perform semisupervised local ancestry inference with references from  $K = 2$  ancestry clusters, 1 corresponding to CAHG ancestry components and the other to Bantu-associated agriculturalist ancestry. Gnomix has been shown to achieve higher accuracy than any other available local ancestry inference method<sup>47</sup>, even when data at much higher resolutions are available<sup>47</sup>. Following refs. 24,82, reference panels for each ancestry cluster were created by selecting unadmixed agriculturalist and CAHG individuals (with  $>98\%$  agriculturalist- or CAHG-associated ancestry) indicated by the unsupervised global ADMIXTURE  $K = 2$  (ref. 81) clustering run described above, selecting a balanced number of individuals for each panel. The choice of reference samples was also corroborated by previous analyses of the same genomes. This resulted in a reference panel of  $N = 31$  CAHG and  $N = 38$  agriculturalist individuals. As recommended by the developers of the software, we used 15% of the data for training set 1, 80% for training set 2 and 5% for validation, a window size of 0.2 cM and an  $r_{\text{admixed}}$  parameter of 1 in Gnomix<sup>47</sup>. Samples were first phased by Beagle v.5.3 (ref. 83) with default settings. Across chromosomes, the classification accuracy for haplotypes of CAHG ancestry was 88% (range = 83–95%) and for



haplotypes of agriculturalist ancestry the classification accuracy was 84% (range = 76–88%) (see Supplementary Table 22 for the full confusion matrix). This classification accuracy is much higher than what is normally achieved by other local ancestry inference methods, even when data at much higher resolutions are available<sup>47</sup>. We removed Bantu-associated chromosomal segments in a procedure known as masking<sup>84,85</sup>: SNPs located in certain ancestry segments, as classified by Gnomix, were removed, or masked, from downstream analyses. We refer to the remaining unmasked chromosomal segments as CAHG chromosomal segments and we refer to analyses that use only these segments as CAHG ancestry-specific analyses.

To eliminate between-population shared long IBD segments, which are the product of recent genetic exchanges between CAHG groups, we took the re-phased CAHG ancestry components output from Gnomix and used refinedIBD<sup>86</sup> to identify shared IBD blocks between each pair of individuals >1 cM. This threshold allows the identification of individuals that shared common ancestors within the last 2,500 years while ensuring a high-enough power to detect IBD blocks and an extremely low false positive rate<sup>50,54–57</sup>. A key date for this study is 2500 BP because this period is thought to represent the initial north–south migration of Bantu-speaking communities across the Equator following a fragmentation of rainforest that would have created corridors for human migration<sup>87–89</sup> (but see ref. 64 for putative evidence of an earlier crossing of the rainforest by Bantu speakers). We then merged IBD blocks within a 0.6 cM gap and allowed only one inconsistent genotype between the gap and block regions using the program merge-ibd-segments from Beagle utilities<sup>90</sup>. Finally, we selected only those IBD blocks shared by individuals belonging to two different populations and masked them in their corresponding haplotypes on the already masked file (see Supplementary Text 1 for further details into IBD analyses).

### Ancestry-specific genotype frequency matrix

Following ref. 82 for each haplotype in a given diploid individual, masking was done for SNPs located in non-CAHG chromosomal segments. The individual's two haplotypes were averaged, creating a genotype frequency vector with (0, 0.5 or 1) average alternate allele count. Sites at which an individual had no CAHG ancestry on either haplotype are marked as missing. Matrix completion was used on the  $N$  individuals  $\times$   $p$ -genotyped markers SNP matrix following ref. 82.

### Ancestry-specific MDS

As described in ref. 82, we performed MDS on the ancestry-specific genotype matrix described above. For a distance metric, we used the average number of pairwise differences, which varies directly with genetic drift<sup>91</sup>.

### Ancestry-specific population statistics

Following ref. 82, the two population statistics described below were computed on population variant frequency ( $f$ ) vectors created by computing, for each site,

$$f_i = \frac{a_i}{n_i}$$

where  $a_i$  is the minor allele count in population  $i$  considering only CAHG chromosomal segments, and  $n_i$  is the total count of CAHG minor and major alleles in  $i$ . Any sites located in one or fewer CAHG segment were removed from the dataset for all populations, so  $n_i > 1 \forall i$ . This filtering resulted in the loss of 9,738 SNPs from the total 431,867 SNPs, leaving 422,129 SNPs across all populations for computation of the following population allele frequency statistics.

We used  $F_{ST}$  statistics to compute pairwise genetic distances between the cultural groups (Supplementary Figs. 7–9). Pairwise  $F_{ST}$  is the proportion of the total genetic variance due to between-population differences and is a convenient measure because it does not depend on

the actual magnitude of the genetic variance. In other words, genetic markers that evolve slowly are expected to have the same  $F_{ST}$  value as markers that evolve more rapidly because the total variance is decomposed into within-population and between-population components<sup>62</sup>. We used this measure because it can account for homozygosity within populations and population histories were comparable<sup>21,76</sup>.

We use Hudson's estimator for  $F_{ST}$ :

$$F_{ST}^{\text{Hudson}} = \frac{(f_A - f_B)^2 - \frac{f_A(1-f_A)}{n_A-1} - \frac{f_B(1-f_B)}{n_B-1}}{n_A(1-n_B) + n_B(1-n_A)},$$

for a given SNP, where A and B are different populations. For more than one site, the numerator and denominator were averaged across all SNPs separately and then the division was performed<sup>92</sup>.

### Environmental similarities

The similarity in the environments inhabited by each of the cultural groups was calculated by using differences in biome composition between the cultures' territories. The biome composition at each 0.5° grid cell in the map was obtained from a recently published global vegetation model<sup>93</sup>. Two distance matrices were built: one based on Gower dissimilarities in the percentage of the culture's territory occupied by each of the biomes included in the dataset from ref. 93 and the other based on presence or absence of each biome type.

### Measuring cultural distances

To compute between-culture similarities in cultural repertoires we used Jaccard distances, equivalent to the ratio between the sum of the number of traits present in one culture but not in another and the sum of the number of traits that are present in one or both the cultures. We did this for musical instruments and subsistence tools separately. The resulting numerical index is bounded between 0 (identical presences in the two sites) and 1 (complete absence of shared traits) and is extremely useful for this type of data as it does not consider negative matches (that is, shared absences) that are common in sparse matrices with many absences<sup>45,46,94</sup>.

### Assessing the drivers of cultural distances

We estimated the relative effects of geographic, biome and ancestry-specific genetic distances on similarities in musical instrument and subsistence toolkit repertoires using Mantel tests<sup>95</sup>, which assess correlations between dissimilarity matrices using 1,000 permutations to obtain  $P$  values that were then adjusted using the Benjamini–Hochberg procedure<sup>45</sup>. We then used multiple matrix regressions to test for the effect of each variable while controlling for the impact of a third matrix, which allows investigating causal relationships between distance matrices rather than the paired vectors themselves<sup>45,96</sup>, also using 1,000 permutations to obtain  $P$  values. Mantel tests and multiple matrix regressions are routinely used in population studies, ecology and increasingly in archaeological and anthropological studies aiming precisely at assessing the drivers of the structure of cultural diversity<sup>45,46,94,97,98</sup>.

To further assess the robustness of the relationship between geographic, genetic and cultural similarities we analysed the CADM<sup>99</sup> using the ape R package<sup>100</sup>. It has been shown that CADM has the correct rate of type I error and good power when applied to independently generated distance matrices. For each genetic dataset (full genetic dataset, CAHG ancestry segments, CAHG ancestry segments minus IBD segments and Bantu ancestry segments only), we first used the global test of significance, CADM.global, to assess the congruence of the genetic distance matrix based on such a genetic dataset, with the matrices of geographical distance, ecological distance (that is, distance in biome composition) and cultural distance based on either musical instrument or subsistence tool repertoires. For the global test of significance, all distance matrices are permuted at random,

independently of one another. The null hypothesis for this test is the incongruence of all matrices. The alternative hypothesis is that at least two matrices are congruent, with similar rankings of the distances (Supplementary Tables 6, 7). We then performed a posteriori tests with the function `CADM.post` to identify groups of congruent matrices. In these tests, the null hypothesis is the incongruence of the matrix subjected to the test, with respect to all other matrices included. The alternative hypothesis is that this matrix is congruent with at least one other matrix in the set, having similar rankings of the distances. To preserve a correct or error rate, the probabilities are adjusted for multiple testing using the procedure as it is generally recommended for non-independent tests<sup>101</sup>.

Finally, to visualize the structure of cultural and genetic diversity between groups while controlling for multicollinearity, we performed a PCoA on the distance matrices of pairwise Jaccard distances for musical instruments and subsistence tools and of pairwise  $F_{ST}$  (Supplementary Figs. 2 and 3). Similar to a principal component analysis, a PCoA produces a set of orthogonal axes whose importance is measured by eigenvalues. However, in contrast to the principal component analysis, non-Euclidean distance matrices can be used<sup>62</sup>. We then visualized the PCoAs using the NeighborNet algorithm in `phangorn`<sup>102</sup>.

### Assessing the sharing of terms to design musical instruments and subsistence tools

Words were counted as shared between two populations speaking different languages when they used words designating the same object that are sufficiently similar to suggest borrowing and common origin, as opposed to accidental lookalikes. To judge this, we used the Glottocodes of each of the CAHG languages to collect a set of grammars, word lists and other language descriptions to ascertain whether observed differences between similar words can be explained by language-specific patterns in morphology (for example, an added affix might make forms look more different than they are) or by differences in superficial conventions in the orthography used for transcription (for example, whether low tone is transcribed or whether prenasalized fricatives are written *mf*, *mv*, *nfor* or *nv*). See Supplementary Dataset 2, with notes on the rationale behind our judgements and bibliographical references, and Supplementary Text 2 for a more detailed explanation of the methodology.

After compiling the final list of shared musical instrument and subsistence tool terminology between CAHG groups, we also noted whether the objects those terms depicted were unique to CAHGs (as opposed to present in other African non-hunter-gatherer groups) and whether the terms themselves were unique to CAHGs (see Supplementary Dataset 2 for notes). We also noted when the relationships between terms recorded in different languages were likely but there was insufficient evidence to conclusively back up the claim of relatedness. We then created two datasets of shared specialized vocabulary: one (broader dataset) including all shared terms between CAHG groups, and a restricted dataset including only those terms for which we had conclusive evidence of their relatedness and that were unique to CAHGs. We believe both datasets are important as CAHGs are known in Africa for their extensive music and musical instrument repertoires. Hence, many Bantu (and other farmer) populations rely on these instruments to sing and dance during their own ceremonies and rituals. This often results in ethnographers reporting that farmers use the hunter-gatherer word for particular instruments because the instrument comes from the hunter-gatherers and is often only used in contexts in which hunter-gatherers are involved<sup>42,43</sup>.

### Calculating linguistic distances between populations

Although all CAHGs are thought to speak languages borrowed from neighbouring agriculturalist populations following large farming expansions, in some instances they are considered to have diverged

enough to constitute distinct languages<sup>16</sup> (Supplementary Tables 2 and 3). Nonetheless, several CAHGs are thought to speak dialects of languages spoken by neighbouring farming groups. To calculate an estimate of linguistic phylogenetic distance between pairs of CAHG languages, we obtained the ISO 639-3 code of either the CAHG language itself or the farmer language from which the CAHG dialect was most closely related (Supplementary Table 3). In the latter case, we confirmed the match between the classification of farmer language and CAHG dialect using ref. 17.

Next we estimated phylogenetic distances between each pair of languages using three methods. First, we used machine-learning techniques on the word lists of the ASJP database<sup>60,61</sup>. The ASJP contains lists of the 40 most stable items from each language from the 100-item Swadesh list<sup>103</sup> and is thought to best reflect language relatedness<sup>61,104</sup>. We measured linguistic distances based on alignments weighted by sound correspondence probabilities estimated by PMI. PMI-weighted distances correspond more closely to distances in recognized phylogenies than other available estimates of lexical distances<sup>61,62,105</sup>. For the Babongo, the only linguistically heterogeneous population in our sample, we averaged pairwise distances based on each of its languages.

Second, we calculated phylogenetic distances using the global language trees from Glottolog v.4.8 (ref. 63). To do this, we merged the Central Sudanic and Atlantic–Congo subtrees and counted the number of internal nodes between each pair of CAHG languages. Third, we calculated patristic distances between each pair of languages using the augmented phylogenetic tree from the latest published phylogeny of Bantu languages by ref. 64, matching languages by their Glottocodes. For the three languages that were not part of the Bantu family, we took the maximum tree depth as the distance between each of those languages and every other language in the sample.

### Assessing the drivers of shared vocabulary items across CAHG groups

To determine whether shared words between groups were the product of a deep shared ancestry as CAHGs or instead of the borrowing of similar terms from closely related farming languages, we fitted zero-inflated models with a Poisson link function for the count model and a binomial link function for the binary part. Our choice of model formulation is due to the nature of the dataset containing a large number of zeros (pairs not sharing words) for structural and sampling reasons<sup>106</sup>. In the models, we used either pairwise PMI distances, pairwise Glottolog distances or patristic phylogenetic distances, and either pairwise genetic distance based exclusively on CAHG genomic segments or spatial distance between groups to predict their number of shared musical instrument words (Supplementary Tables 14–19). All predictor variables were standardized for comparability of coefficients. Given that genetic and spatial distances had a correlation coefficient of 0.760 ( $P < 0.001$ ), we could not include both predictors in the same models as that could lead to a spurious reversal of estimates from one of the correlated predictors<sup>107</sup>. We could not fit similar models of subsistence tool terminology between groups because cases of shared terminologies were extremely limited.

Lastly, we determined the significance of cultural domain in word sharing by performing a paired Wilcoxon signed-rank test comparing the mean number of shared words between cultural dyads across cultural domains (that is, musical instruments versus subsistence tools).

### Inferring processes of admixture in genetic and cultural data

To assess the relationship between geographic, cultural and genetic distances between populations, we performed three independent `SpaceMix`<sup>53</sup> analyses aimed at retrieving the ten genetic and cultural pseudo-spatial positions of the CAHG groups in our sample with available genetic data and musical instrument and subsistence tool repertoires. For the genetic run, we used our masked data and the allele.

frequency option of SpaceMix, with the sample\_means option specifying for each of the 422,128 loci remaining in our dataset after masking, the average number of individuals in each population carrying a CAHG allele for that locus. For the musical instrument and subsistence tool runs, we used the count/total option of SpaceMix by generating a pseudo-genetic file in which each population was represented as a single diploid individual and the presence of a given instrument or tool was registered as a homozygous trait<sup>108</sup>. The geno-geographical and cultural-geographical coordinates generated were compared with the actual geographical coordinates visually and numerically.

In addition, we used SpaceMix runs on genetic data to assess evidence for recent gene flow between CAHG groups following agriculturalist expansions by computing and visualizing sources and direction of recent admixture between populations<sup>109</sup>. For each population, admixture proportions were obtained by calculating the proportion of alleles estimated to come from that population's inferred admixture source location (Supplementary Table 9 and ref. 109) as opposed to the population geo-genetic location. We then selected those populations with inferred admixture proportions >1% from the Markov chain Monte Carlo run with the highest posterior probability, which we plotted in Supplementary Fig. 13.

### Reporting summary

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

### Data availability

Processed genetic data and raw cultural and linguistic data required to reproduce all the results reported in the article and Supplementary Information are available in the following GitHub repository: <https://github.com/ceciliapad/GenCultEvo>. Raw genome-wide SNP data were obtained from the datasets reported in refs. 24,75–77 and are available in the public servers EGA (accession numbers EGAS00001002078 and EGAC00001000139) and dbGaP (accession numbers phs000449.v2.p1 and phs001780.v1.p1). Language phylogenies were obtained from ASJP<sup>60,61</sup>, Glottolog v.4.8 (ref. 63) and ref. 64.

### Code availability

The code required to reproduce all the results reported in the article and Supplementary Information are available in the following GitHub repository: <https://github.com/ceciliapad/GenCultEvo>.

### References

- Stringer, C. B. & Andrews, P. Genetic and fossil evidence for the origin of modern humans. *Science* **239**, 1263–1268 (1988).
- Stringer, C. The origin and evolution of *Homo sapiens*. *Philos. Trans. R. Soc. Lond. B* **371**, 20150237 (2016).
- Bergström, A., Stringer, C., Hajdinjak, M., Scerri, E. M. L. & Skoglund, P. Origins of modern human ancestry. *Nature* **590**, 229–237 (2021).
- 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).
- Lipson, M. et al. Ancient DNA and deep population structure in sub-Saharan African foragers. *Nature* **603**, 290–296 (2022).
- Skoglund, P. et al. Reconstructing prehistoric African population structure. *Cell* **171**, 59–71 (2017).
- Schlebusch, C. M. et al. Khoe-San genomes reveal unique variation and confirm the deepest population divergence in *Homo sapiens*. *Mol. Biol. Evol.* **37**, 2944–2954 (2020).
- Ragsdale, A. P. et al. A weakly structured stem for human origins in Africa. *Nature* **617**, 755–763 (2023).
- Patin, E. et al. Inferring the demographic history of African farmers and Pygmy hunter-gatherers using a multilocus resequencing data set. *PLoS Genet.* **5**, e1000448 (2009).
- Hsieh, P. et al. Model-based analyses of whole-genome data reveal a complex evolutionary history involving archaic introgression in Central African Pygmies. *Genome Res.* **26**, 291–300 (2016).
- Lopez, M. et al. The demographic history and mutational load of African hunter-gatherers and farmers. *Nat. Ecol. Evol.* **2**, 721–730 (2018).
- Le Bomin, S., Lecomte, G. & Heyer, E. The evolution of musical diversity: the key role of vertical transmission. *PLoS ONE* **11**, e0151570 (2016).
- Bailey, R. C. et al. Hunting and gathering in tropical rain forest: is it possible? *Am. Anthropol.* **91**, 59–82 (1989).
- Hart, T. B. & Hart, J. A. The ecological basis of hunter-gatherer subsistence in African rain forests: the Mbuti of Eastern Zaire. *Hum. Ecol.* **14**, 29–55 (1986).
- Blench, R. in *Challenging Elusiveness: Central African Hunter-Gatherers in a Multi-Disciplinary Perspective* (eds Biesbrouck, K. et al.) 41–60 (Research School of Asian, African and Amerindian Studies (CNWS), Univ. Leiden, 1999).
- Bahuchet, S. Changing language, remaining Pygmy. *Hum. Biol.* **84**, 11–43 (2012).
- Bahuchet, S. Languages of African rainforest 'Pygmy' hunter-gatherers: language shifts without cultural admixture. In *Historical Linguistics and Hunter-Gatherer Populations in Global Perspective* (Max Planck Institute, 2006).
- Bahuchet, S. Languages of African rainforest "Pygmy" hunter-gatherers: language shifts without cultural admixture. In *Hunter-Gatherers and Linguistic History: A Global Perspective* (eds Güldemann, T. et al.) 1–32 (Cambridge Univ. Press, 2006).
- Padilla-Iglesias, C. et al. Population interconnectivity over the past 120,000 years explains distribution and diversity of Central African hunter-gatherers. *Proc. Natl Acad. Sci. USA* **119**, e2113936119 (2022).
- Clist, B. et al. Using the radiocarbon dates of Central Africa for studying long-term demographic trends of the last 50,000 years: potential and pitfalls. *Azania* **58**, 235–293 (2023).
- Fan, S. et al. African evolutionary history inferred from whole genome sequence data of 44 indigenous African populations. *Genome Biol.* **20**, 82 (2019).
- Lopez, M. et al. Genomic evidence for local adaptation of hunter-gatherers to the African rainforest. *Curr. Biol.* **29**, 2926–2935 (2019).
- Bergey, C. M. et al. Polygenic adaptation and convergent evolution on growth and cardiac genetic pathways in African and Asian rainforest hunter-gatherers. *Proc. Natl Acad. Sci. USA* **115**, E11256–E11263 (2018).
- Perry, G. H. et al. Adaptive, convergent origins of the pygmy phenotype in African rainforest hunter-gatherers. *Proc. Natl Acad. Sci. USA* **111**, E3596–E3603 (2014).
- Fan, S. et al. Whole-genome sequencing reveals a complex African population demographic history and signatures of local adaptation. *Cell* **186**, 923–939 (2023).
- MacDonald, D. H. & Hewlett, B. S. Reproductive interests and forager mobility. *Curr. Anthropol.* **40**, 501–524 (1999).
- Boyette, A. H., Lew-Levy, S., Jang, H. & Kandza, V. Social ties in the Congo Basin: insights into tropical forest adaptation from BaYaka and their neighbours. *Philos. Trans. R. Soc. B* **377**, 20200490 (2022).
- Lewis, J. Where goods are free but knowledge costs: hunter-gatherer ritual economics in western Central Africa. *Hunt. Gatherer Res.* **1**, 1–27 (2015).
- Cavalli-Sforza, L. L. & Hewlett, B. Exploration and mating range in African Pygmies. *Ann. Hum. Genet.* **46**, 257–270 (1982).
- Migliano, A. B. & Vinicius, L. The origins of human cumulative culture: from the foraging niche to collective intelligence. *Philos. Trans. R. Soc. B* **377**, 20200317 (2022).



31. Migliano, A. B. et al. Hunter-gatherer multilevel sociality accelerates cumulative cultural evolution. *Sci. Adv.* **6**, eaax5913 (2020).
32. Salali, G. D. et al. Knowledge-sharing networks in hunter-gatherers and the evolution of cumulative culture. *Curr. Biol.* **26**, 2516–2521 (2016).
33. Pamjav, H., Juhász, Z., Zalán, A., Németh, E. & Damdin, B. A comparative phylogenetic study of genetics and folk music. *Mol. Genet. Genomics* **287**, 337–349 (2012).
34. Savage, P. E., Brown, S., Sakai, E. & Currie, T. E. Statistical universals reveal the structures and functions of human music. *Proc. Natl Acad. Sci. USA* **112**, 8987–8992 (2015).
35. Aguirre-Fernández, G., Blasi, D. E. & Sánchez-Villagra, M. R. Panpipes as units of cultural analysis and dispersal. *Evol. Hum. Sci.* **2**, e17 (2020).
36. Miller, J. M. & Wang, Y. V. Ostrich eggshell beads reveal 50,000-year-old social network in Africa. *Nature* **601**, 234–239 (2022).
37. Blegen, N. The earliest long-distance obsidian transport: evidence from the ~200 ka Middle Stone Age Sibilo School Road Site, Baringo, Kenya. *J. Hum. Evol.* **103**, 1–19 (2017).
38. Mounier, A. & Lahr, M. M. Deciphering African late middle Pleistocene hominin diversity and the origin of our species. *Nat. Commun.* **10**, 3406 (2019).
39. Hodder, I. The distribution of material culture items in the Baringo district, western Kenya. *Man* **12**, 239–269 (1977).
40. Derex, M. & Mesoudi, A. Cumulative cultural evolution within evolving population structures. *Trends Cogn. Sci.* **24**, 654–667 (2020).
41. Turnbull, C. in *Anthropological Papers of the American Museum of Natural History* Vol. 50 (eds Shapiro, H. L. et al.) 139–282 (American Museum of Natural History, 1965).
42. Fūrniß, S. Musiques aka et baka: une parenté de référence. *J. Afr.* **82**, 107–136 (2012).
43. Fūrniß, S. & Bahuchet, S. Existe-t-il des instruments de musique pygmées? (1995).
44. Klieman, K. A. *The Pygmies Were Our Compass: Bantu and Batwa in the History of West Central Africa, Early Times to C. 1900 CE.* (Bloomsbury, 2003).
45. Blinkhorn, J. & Grove, M. Explanations of variability in Middle Stone Age stone tool assemblage composition and raw material use in Eastern Africa. *Archaeol. Anthropol. Sci.* **13**, 14 (2021).
46. Shennan, S. J., Crema, E. R. & Kerig, T. Isolation-by-distance, homophily, and ‘core’ vs. ‘package’ cultural evolution models in Neolithic Europe. *Evol. Hum. Behav.* **36**, 103–109 (2015).
47. Hilmarsson, H. et al. High resolution ancestry deconvolution for next generation genomic data. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.09.19.460980> (2021).
48. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267 (2006).
49. Wang, C. et al. Comparing spatial maps of human population-genetic variation using Procrustes analysis. *Stat. Appl. Genet. Mol. Biol.* **9**, 13 (2010).
50. Ioannidis, A. G. et al. Native American gene flow into Polynesia predating Easter Island settlement. *Nature* **583**, 572–577 (2020).
51. Novembre, J. et al. Genes mirror geography within Europe. *Nature* **456**, 98–101 (2008).
52. Hudson, R. R., Slatkin, M. & Maddison, W. P. Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**, 583–589 (1992).
53. Bradburd, G. S., Ralph, P. L. & Coop, G. M. Disentangling the effects of geographic and ecological isolation on genetic differentiation. *Evolution* **67**, 3258–3273 (2013).
54. Liu, D. et al. Extensive ethnolinguistic diversity in Vietnam reflects multiple sources of genetic diversity. *Mol. Biol. Evol.* **37**, 2503–2519 (2020).
55. Botigue, L. R. et al. Gene flow from North Africa contributes to differential human genetic diversity in southern Europe. *Proc. Natl Acad. Sci. USA* **110**, 11791–11796 (2013).
56. Palamara, P. F., Lencz, T., Darvasi, A. & Pe’er, I. Length distributions of identity by descent reveal fine-scale demographic history. *Am. J. Hum. Genet.* **91**, 809–822 (2012).
57. Ralph, P. & Coop, G. The geography of recent genetic ancestry across Europe. *PLoS Biol.* **11**, e1001555 (2013).
58. Fūrniß, S. in *Quel avenir pour les Pygmées à l’orée du XXI<sup>e</sup> siècle?: Qui sont-ils, que subissent-ils, comment font-ils face?* (eds Kulesza, P. & Robillard, M.) Chap. 1 (L’Harmattan, 2019).
59. Güldemann, T. & Winkhart, B. The \*Baakaa and other puzzles. *Anthropol. Linguist.* **62**, 259–306 (2020).
60. Wichmann, S., Rama, T. & Holman, E. W. Phonological diversity, word length, and population sizes across languages: the ASJP evidence. *Linguist. Typology* **15**, 177–197 (2011).
61. Jäger, G. Global-scale phylogenetic linguistic inference from lexical resources. *Sci. Data* **5**, 180189 (2018).
62. Matsumae, H. et al. Exploring correlations in genetic and cultural variation across language families in northeast Asia. *Sci. Adv.* **7**, eabd9223 (2021).
63. Hammarström, H., Forkel, R., Haspelmath, M. & Bank, S. glottolog/ glottolog: Glottolog database 4.8. *Zenodo* <https://doi.org/10.5281/ZENODO.8131084> (2023).
64. Koile, E., Greenhill, S. J., Blasi, D. E., Bouckaert, R. & Gray, R. D. Phylogeographic analysis of the Bantu language expansion supports a rainforest route. *Proc. Natl Acad. Sci. USA* **119**, e2112853119 (2022).
65. Arom, S. & Thomas, J. M. *Les Mimbo, génies du piégeage, et le monde surnaturel des Ngbaka-Ma’bo, République Centrafricaine* Vol. 44 (Peeters Publishers, 1974).
66. Bahuchet, S. *Dans la forêt d’Afrique Centrale: les pygmées Aka et Baka* Vol. 8 (Peeters Publishers, 1992).
67. Kelly, R. L. Hunter-gatherer mobility strategies. *J. Anthropol. Res.* **39**, 277–306 (1983).
68. Lewis, H. M., Vinicius, L., Strods, J., Mace, R. & Migliano, A. B. High mobility explains demand sharing and enforced cooperation in egalitarian hunter-gatherers. *Nat. Commun.* **5**, 5789 (2014).
69. Dyble, M. et al. Sex equality can explain the unique social structure of hunter-gatherer bands. *Science* **348**, 796–798 (2015).
70. Migliano, A. B. et al. Characterization of hunter-gatherer networks and implications for cumulative culture. *Nat. Hum. Behav.* **1**, 0043 (2017).
71. Garg, K., Padilla-Iglesias, C., Restrepo Ochoa, N. & Knight, V. B. Hunter-gatherer foraging networks promote information transmission. *R. Soc. Open Sci.* **8**, 211324 (2021).
72. Bergström, A. et al. Insights into human genetic variation and population history from 929 diverse genomes. *Science* **367**, eaay5012 (2020).
73. Lipson, M. et al. Ancient West African foragers in the context of African population history. *Nature* **577**, 665–670 (2020).
74. Wiessner, P. Style and social information in Kalahari San projectile points. *Am. Antiq.* **48**, 253–276 (1983).
75. Patin, E. et al. The impact of agricultural emergence on the genetic history of African rainforest hunter-gatherers and agriculturalists. *Nat. Commun.* **5**, 3163 (2014).
76. Patin, E. et al. Dispersals and genetic adaptation of Bantu-speaking populations in Africa and North America. *Science* **356**, 543–546 (2017).
77. Jarvis, J. P. et al. Patterns of ancestry, signatures of natural selection, and genetic association with stature in Western African Pygmies. *PLoS Genet.* **8**, e1002641 (2012).
78. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

79. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2018).
80. Manichaikul, A. et al. Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
81. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
82. Ioannidis, A. G. et al. Paths and timings of the peopling of Polynesia inferred from genomic networks. *Nature* **597**, 522–526 (2021).
83. Browning, B. L., Tian, X., Zhou, Y. & Browning, S. R. Fast two-stage phasing of large-scale sequence data. *Am. J. Hum. Genet.* **108**, 1880–1890 (2021).
84. Schroeder, H. et al. Origins and genetic legacies of the Caribbean Taino. *Proc. Natl Acad. Sci. USA* **115**, 2341–2346 (2018).
85. Moreno-Estrada, A. et al. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. *Science* **344**, 1280–1285 (2014).
86. Browning, B. L. & Browning, S. R. Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics* **194**, 459–471 (2013).
87. Garcin, Y. et al. Early anthropogenic impact on Western Central African rainforests 2,600 y ago. *Proc. Natl Acad. Sci. USA* **115**, 3261–3266 (2018).
88. González-Santos, M. et al. Exploring the relationships between genetic, linguistic and geographic distances in Bantu-speaking populations. *Am. J. Biol. Anthropol.* **179**, 104–117 (2022).
89. Grollemund, R. et al. Bantu expansion shows that habitat alters the route and pace of human dispersals. *Proc. Natl Acad. Sci. USA* **112**, 13296–13301 (2015).
90. Browning, S. R. et al. Ancestry-specific recent effective population size in the Americas. *PLoS Genet.* **14**, e1007385 (2018).
91. Peter, B. M. Admixture, population structure, and F-statistics. *Genetics* **202**, 1485–1501 (2016).
92. Bhatia, G., Patterson, N., Sankararaman, S. & Price, A. L. Estimating and interpreting FST: the impact of rare variants. *Genome Res.* **23**, 1514–1521 (2013).
93. Beyer, R. M., Krapp, M. & Manica, A. High-resolution terrestrial climate, bioclimate and vegetation for the last 120,000 years. *Sci. Data* **7**, 236 (2020).
94. Buchanan, B. & Hamilton, M. J. A formal test of the origin of variation in North American early Paleoindian projectile points. *Am. Antiq.* **74**, 279–298 (2009).
95. Smouse, P. E., Long, J. C. & Sokal, R. R. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Syst. Zool.* **35**, 627–632 (1986).
96. Legendre, P., Fortin, M.-J. & Borcard, D. Should the Mantel test be used in spatial analysis? *Methods Ecol. Evol.* **6**, 1239–1247 (2015).
97. Ross, R. M., Greenhill, S. J. & Atkinson, Q. D. Population structure and cultural geography of a folktale in Europe. *Proc. R. Soc. B* **280**, 20123065 (2013).
98. Hutchison, D. W. & Templeton, A. R. Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution* **53**, 1898–1914 (1999).
99. Legendre, P. & Lapointe, F.-J. Assessing congruence among distance matrices: single-malt scotch whiskies revisited. *Aust. NZ J. Stat.* **46**, 615–629 (2004).
100. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
101. Wright, S. P. Adjusted *P*-values for simultaneous inference. *Biometrics* **48**, 1005 (1992).
102. Schliep, K. P. phangorn: phylogenetic analysis in R. *Bioinformatics* **27**, 592–593 (2011).
103. Brown, C. H., Holman, E. W., Wichmann, S. & Velupillai, V. Automated classification of the world's languages: a description of the method and preliminary results. *STUF – Language Typology and Universals* **61**, 285–308 (2008).
104. Wichmann, S. et al. Glottochronology as a heuristic for genealogical language relationships. *J. Quant. Linguist.* **17**, 303–316 (2010).
105. Greenhill, S. J. Levenshtein distances fail to identify language relationships accurately. *Comput. Linguist.* **37**, 689–698 (2011).
106. Feng, C. X. A comparison of zero-inflated and hurdle models for modeling zero-inflated count data. *J. Stat. Distrib. Appl.* **8**, 8 (2021).
107. McElreath, R. *Statistical Rethinking: Texts in Statistical Science* (CRC Press, 2015).
108. Bortolini, E. et al. Inferring patterns of folktale diffusion using genomic data. *Proc. Natl Acad. Sci. USA* **114**, 9140–9145 (2017).
109. Bradburd, G. S., Ralph, P. L. & Coop, G. M. A spatial framework for understanding population structure and admixture. *PLoS Genet.* **12**, e1005703 (2016).
110. Bahuchet, S. Les Pygmées Aka et la forêt centrafricaine, ethnologie écologique; [https://www.researchgate.net/publication/278779706\\_Les\\_Pygmees\\_Aka\\_et\\_la\\_foret\\_centrafricaine\\_ethnologie\\_ecologique](https://www.researchgate.net/publication/278779706_Les_Pygmees_Aka_et_la_foret_centrafricaine_ethnologie_ecologique) (SELAF, 1985).

## Acknowledgements

We thank F. Marandola, S. Fűrniß, R. Jadinon, N. Chaudhury and E. Cornelissen for their help curating the cultural corpus, and E. Patin for help with assembling the genetic dataset. We also thank M. Sánchez Villagra and G. Aguirre for helpful methodological discussions. We thank R. Schlaepfer for help with the article figures. We thank the University of Zurich for financially supporting A.B.M., L.V. and C.P.-I. We are grateful to Fundación La Caixa and Forschungskredit Candoc for the funding of this study (grants LCF/BQ/EU19/11710043 and FK-19-083 awarded to C.P.-I.). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

## Author contributions

C.P.-I., L.V. and A.B.M. conceptualized the study. C.P.-I. compiled the data. C.P.-I., J.B.-P., A.G.I., B.P., B.B., A.M., L.V. and A.B.M. developed the methodology. C.P.-I. analysed the data. C.P.-I., L.V. and A.B.M. wrote the article. All authors reviewed the article, provided comments and approved the final draft.

## Funding

Open access funding provided by University of Zurich.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41562-024-01891-y>.

**Correspondence and requests for materials** should be addressed to Cecilia Padilla-Iglesias or Andrea Bamberg Migliano.

**Peer review information** *Nature Human Behaviour* thanks Cesar Fortes-Lima, Ezequiel Koile, Rowan McLaughlin and David Wright for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the

source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024



## Reporting Summary

Nature Portfolio 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 Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- 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
- A full description of the statistical parameters 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)
- 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.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

*Provide a description of all commercial, open source and custom code used to collect the data in this study, specifying the version used OR state that no software was used.*

Data analysis

All data and code required to reproduce the results in the manuscript and supplementary information is available in the following GitHub repository: <https://github.com/ceciliapad/GenCultEvo>. We used the following software:

- Gnomix (<https://github.com/AI-sandbox/gnomix>)
- PLINK (PLINK v1.90b6.12)
- R (R version 4.3.1)
- KING (<https://www.kingrelatedness.com>)
- Spacemix (<https://github.com/gbradburd/SpaceMix>)
- ADMIXTURE (ADMIXTURE/1.3.0)
- Beagle 5.3
- RefinedIBD(<https://faculty.washington.edu/browning/refined-ibd.html>)
- ape R package (version 5.7)
- phangorn R package (version 2.10)

The remaining packages are specified not in the methods but in the scripts included in the Github repository.

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Processed genetic data, and raw cultural and linguistic data required to reproduce all the results reported in the manuscript and supplementary are available in the following Github repository: <https://github.com/ceciliapad/GenCultEvo>. Raw genome-wide SNP data is available in public servers EGA (Accession number EGAS00001002078 and EGAC00001000139) and dbGaP (Accession numbers phs000449.v2.p1 and phs001780.v1.p1). Language phylogenies were obtained from the ASJP database (<https://asjp.cild.org>), Koile et al. (2022) and Glottolog v.4.8 (<https://glottolog.org>)

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	No subjects were involved.
Reporting on race, ethnicity, or other socially relevant groupings	No subjects were involved but cultural information on Central African hunter-gatherer populations were obtained from museum collections, ethnographers and primary literature. Population named reflect their recognised ethnicities and primary literature.
Population characteristics	We compiled published cultural, genetic and location data from each population and museum data on cultural artifacts.
Recruitment	No subjects
Ethics oversight	Research did not require human subjects or animals and no ethics oversight was required, but a data access committee gave ethics approval for our use of the above-mentioned genetic datasets for the purpose of this project.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences  Behavioural & social sciences  Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Ecological, evolutionary & environmental sciences study design

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

Study description	This study presents an analysis of cultural, genetic and linguistic data from Central African hunter-gatherer populations.
Research sample	Published data on words, genetic sequences and museum data on cultural artifacts from Central African hunter-gatherer populations.
Sampling strategy	We collected all data available on cultural artifacts and vocabulary items.
Data collection	Primary literature and museum visits
Timing and spatial scale	A few months of visits to libraries and museums. Spatial scale is the entire Congo Basin.
Data exclusions	No exclusion
Reproducibility	All data and code are available
Randomization	Not applicable
Blinding	Not applicable

Did the study involve field work?  Yes  No

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

## Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Plants

### Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

### Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

### Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.