



# Population interconnectivity over the past 120,000 years explains distribution and diversity of Central African hunter-gatherers

Cecilia Padilla-Iglesias<sup>a,1</sup>, Lane M. Atmore<sup>b</sup>, Jesús Olivero<sup>c</sup>, Karen Lupo<sup>d</sup>, Andrea Manica<sup>e</sup>, Epifanía Arango Isaza<sup>f</sup>, Lucio Vinicius<sup>g</sup>, and Andrea Bamberg Migliano<sup>a</sup>

Edited by Brenna Henn, University of California, Davis, CA; received July 28, 2021; accepted April 5, 2022 by Editorial Board Member James F. O'Connell

The evolutionary history of African hunter-gatherers holds key insights into modern human diversity. Here, we combine ethnographic and genetic data on Central African hunter-gatherers (CAHG) to show that their current distribution and density are explained by ecology rather than by a displacement to marginal habitats due to recent farming expansions, as commonly assumed. We also estimate the range of hunter-gatherer presence across Central Africa over the past 120,000 years using paleoclimatic reconstructions, which were statistically validated by our newly compiled dataset of dated archaeological sites. Finally, we show that genomic estimates of divergence times between CAHG groups match our ecological estimates of periods favoring population splits, and that recoveries of connectivity would have facilitated subsequent gene flow. Our results reveal that CAHG stem from a deep history of partially connected populations. This form of sociality allowed the coexistence of relatively large effective population sizes and local differentiation, with important implications for the evolution of genetic and cultural diversity in *Homo sapiens*.

hunter-gatherers | Central Africa | ecological niche modelling | environmental change | population dynamics

The evolutionary history of African hunter-gatherers may hold key insights into patterns and processes behind the evolution of modern human diversity. Recent genomic studies have revealed that these populations represent the oldest and most diverse human genetic lineages and have been genetically differentiated from one another since the origin of humans (1–3) (*SI Appendix, Table S1*). Therefore, a first question is whether their current ecological niches were also characteristic of early *Homo sapiens* populations. However, genetic data alone can neither determine the geographic distribution of hunter-gatherers in the past nor demonstrate a deep history of adaptation of hunter-gatherers to their current environments. In fact, various studies have proposed that farming expansions within the past 5,000 years (in particular by the ancestors of Bantu speakers) would have only recently displaced hunter-gatherers to marginalized regions less favorable to agriculture (such as rainforests and deserts) (4–7).

For example, the central part of Africa, between latitudes 5°N and 5°S currently is inhabited by ~20 scattered hunter-gatherer ethnic groups (8). These Central African hunter-gatherers (CAHG) form a genetic clade thought to have diverged from other African populations as far back as 120,000 to 200,000 years ago (2, 9). The lack of any major linguistic specificity between them is often implied to reflect extensive contacts with surrounding farmer populations (8, 10), and seen as evidence of recent displacement into marginal forest environments by expanding farming populations. However, anthropologists have remarked on the huge variability in lifestyle, habitat, techniques, and tools between CAHG (11), suggestive of long-term cultural diversification and adaptation to forest environments. Research on the drivers of demography and adaptation of CAHG populations remains extremely limited, which can be partially attributed to the lack of archaeological and osteological data resulting from a rapid disintegration of fossil remains in the rainforest's acidic soils, in addition to social instability in the region (12). Therefore, we are still left with crucial questions regarding the time depth of occupation of Central Africa by hunter-gatherers, the breadth of the niche exploited by earlier populations in the region, and variations in levels of interconnectivity at different points in time.

To address those questions, we first compiled ethnographic data on the distribution of 749 camps from 11 hunter-gatherer groups extending from West to East Central Africa. We used them as inputs for environmental niche models (ENMs) to determine the relative influence of several bioclimatic and ecological factors, as well as the

## Significance

We combined ethnographic, archaeological, genetic, and paleoclimatic data to model the dynamics of Central African hunter-gatherer populations over the past 120,000 years. We show, against common assumptions, that their distribution and density are explained by changing environments rather than by a displacement following recent farming expansions, and that they have maintained large population sizes and genetic diversity, despite fluctuations in niche availability. Our results provide insights into the evolution of genetic and cultural diversity in *Homo sapiens*.

Author affiliations: <sup>a</sup>Institute of Anthropology, University of Zurich, 8057, Zurich, Switzerland; <sup>b</sup>Centre for Ecological and Evolutionary Synthesis, University of Oslo, 0316, Oslo, Norway; <sup>c</sup>Departamento de Biología Animal, Universidad de Málaga, 29071, Málaga, Spain; <sup>d</sup>Department of Anthropology, Southern Methodist University, Dallas, TX 75275-0233; <sup>e</sup>Department of Zoology, University of Cambridge, CB23EJ, Cambridge, United Kingdom; and <sup>f</sup>Institute of Evolutionary Biology, University of Zurich, 8057, Zurich, Switzerland

Author contributions: C.P.-I., L.V., and A.B.M. designed research; C.P.-I. and A.M. acquired data; C.P.-I. and K.L. curated data; C.P.-I. and K.L. performed research; C.P.-I., J.O., A.M., and E.A.I. contributed new reagents/analytic tools; C.P.-I. and L.M.A. analyzed data; K.L. and L.V. provided guidance; A.M. and L.V. provided supervision; C.P.-I. wrote the paper; and K.L., A.M., and L.V. reviewed and edited the manuscript.

The authors declare no competing interest.

This article is a PNAS Direct Submission. B.H. is a guest editor invited by the Editorial Board.

Copyright © 2022 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution License 4.0 (CC BY).

<sup>1</sup>To whom correspondence may be addressed. Email: cecilia.padillaiglesias@uzh.ch.

This article contains supporting information online at <http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2113936119/-DCSupplemental>.

Published May 17, 2022.

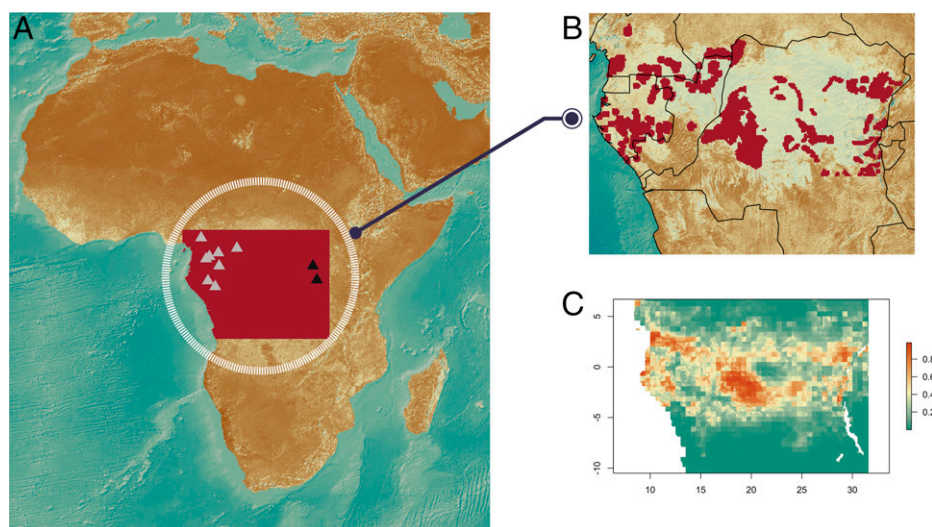
presence of farming populations, on the distribution and abundance of CAHG (13, 14). Then, we used high-resolution paleoclimatic reconstructions and topographic maps to make continuous predictions about where CAHG could have lived over the past 120,000 years and the potential extension of their interaction networks. Next, we compiled all reliably dated archaeological assemblages ascribed to hunter-gatherer groups in the Congo Basin ( $n = 168$ ) and confirmed the model's ability to predict the location and date of the sites. We further contextualized genomic estimates of population divergences with changes in population densities and interpopulation connectivity predicted by our model. Last, we complemented these analyses with a detailed assessment of present and historical gene flow between nine CAHG populations ( $n = 265$  individuals), which we used to assess recent interactions between previously diverged CAHG populations, after farming expansions. Our study therefore provides a causal link between past environmental changes and human population dynamics over evolutionary time, by predicting where and when populations across Central Africa could have exchanged genetic and/or cultural information throughout their evolutionary history.

## Results

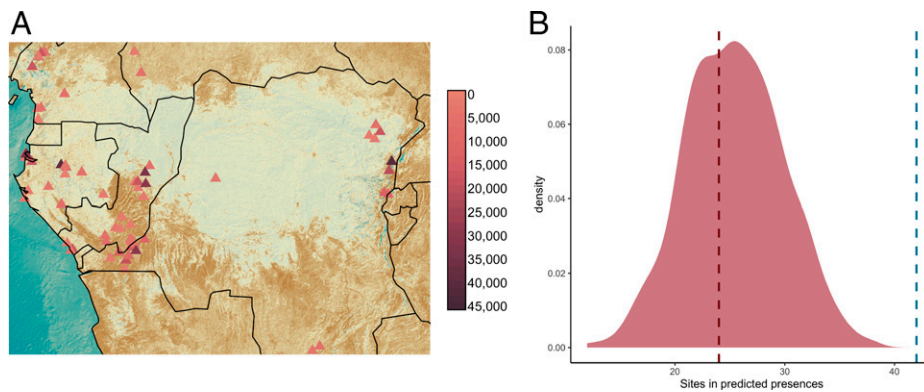
**Current CAHG Distribution Reflects Long-Term Ecological Adaptation Rather Than a Recent Bantu Expansion.** We first asked whether the current distribution and density of CAHG is a product of long-term adaptation to life in the rainforest or instead a recent product of the Bantu Expansion. We compiled ethnographic data on the geographical location of 749 camps from 11 CAHG populations (Dataset S1), and then applied the maximum entropy (MaxEnt) machine-learning algorithm (15) to determine the relative influence of several bioclimatic and ecological factors on the distribution of CAHG (13, 14) (Fig. 1 *A* and *B*) (see *Materials and Methods* for alternative model fitting algorithms). Our ENM performed well on hold-out data (true skill statistic (TSS) = 0.55 and area under the curve (AUC) = 0.87; *SI Appendix, Table S2*), correctly classifying presences and absences 81% and 74% of the time, respectively. The main ecological factor rendering particular areas unsuitable for

hunter-gatherer presence was precipitation seasonality (*BIO15*) followed by annual temperature range (*BIO7*) (*SI Appendix, Table S3* and *Figs. S1* and *S2*), confirming results from other African regions (16, 17). Including predictors relating to Bantu farming populations (rural population density or distance to populated places) neither improved model fit (TSS = 0.51 and AUC = 0.85 for both models) nor significantly altered the relative contribution of the other variables. This suggests that the population dynamics of hunter-gatherers and farmers are independent and subject to environmental factors rather than competitive exclusion (18).

Next, we tested whether the same ecological factors could also predict CAHG population density (18). For the 50 map cells where camp size data were available ( $n = 75$  camps), we found a positive linear relationship between environmental suitability (estimated by MaxEnt) and population density ( $b = 43.379$ ,  $t = 2.238$ ,  $P = 0.030$ ; *SI Appendix, Fig. S3*). We used this relationship to estimate the current CAHG metapopulation size at 191,118 individuals, in close agreement with reported census data between 160,000 (11) and 204,000 (19). In line with previous research, environmental suitability was an even better predictor of the upper limit of population density, as it may be a satisfactory proxy for an absolute constraint on carrying capacity by environmental features (13) (*SI Appendix, Fig. S3*). The slopes of suitability from linear quantile regressions were always significant above the 75th percentile of density, with slopes and  $R^2$  values increasing with the percentile (*SI Appendix, Table S4*). To test whether those ecological drivers were specific to CAHG, we applied general additive models to predict ecological drivers of farming population densities (*SI Appendix, Text, Figs. S4* and *S5*, and *Table S5*). Results showed that bioclimatic variables and biome types affect the two populations differently. For example, high rural population density is associated with temperate forests and large temperature annual ranges, in contrast to hunter-gatherer prevalence which is associated with more stable tropical climates. Thus, our results confirm that the ecological niche of CAHG and farmers are distinct, and that the current range, distribution, and population sizes of CAHG can be predicted by ecological conditions alone (7, 20).



**Fig. 1.** Geographical location of Central African hunter-gatherers. (A). Dark and light triangles designate respectively the sources of genomic samples from Eastern (Batwa and Mbuti) and Western CAHG (Baka from Cameroon and Gabon, Bakoya, Biaka, Bedzan, and Southern and Eastern Babongo). (B) Locations of hunter-gatherer camps used in our bioclimatic environmental niche model (see *Materials and Methods*). (C) Estimated present niche of CAHG using the MaxEnt modeling algorithm.



**Fig. 2.** Ecological modeling of archaeological site distribution in Central Africa since 47,000 BP. (A) 118 archaeological sites included in our analyses and their  $^{14}\text{C}$  dates. (B) MaxEnt predicted the presence of 42 sites in suitable cells (dashed blue line). Distribution shows number of sites in predicted presences across 1,000 randomizations of site dates. Dashed pink line indicates expected cumulative number of sites in predicted presences when randomizing their spatial location at each time period.

### Ecology Predicts Viable CAHG Populations in Tropical Forests throughout the Past 120,000 years.

After establishing that ecological factors are the main determinants of locations of current hunter-gatherers in Central Africa, we then applied our model to predict locations of past CAHG and to estimate how their level of interconnectivity may have varied across regions over time. We employed a bias-corrected time series of global terrestrial climate and vegetation (16) to project our ENM model into 1000- or 2000-year time slices from the present up to 120,000 BP (see *Materials and Methods*), obtaining suitability maps in our area of interest for each time slice. Similarly, we used our estimated linear relationship between environmental suitability and grid cell population density to obtain metapopulation size estimates at each time slice. Our results indicate that repeated episodes of abrupt climatic changes and habitat fragmentation in Central Africa since the Last Interglacial (16, 21, 22) would have caused drastic expansions and reductions of the suitable range for CAHG groups (see Fig. 3B and *SI Appendix, Fig. S6* and *Movie S1*). While this would have resulted in demographic fluctuations, hunter-gatherer populations may have remained viable in Central Africa, with a potential census metapopulation size never falling below 160,000 individuals (*SI Appendix, Fig. S7*). This suggests that CAHG would have been able to maintain their niche and relatively stable population sizes throughout their evolutionary history. Unlike approaches relying exclusively on genetic or linguistic evidence, our model directly situates geographically and ecologically the locations of putative ancestors of CAHG. Hence, contrary to claims that Central African environments would not sustain the presence of human groups independent of agriculture (6, 7, 19), our results add further support to claims from previous linguistic and genetic studies that modern CAHG are descended from ancestral groups whose adaptations to the rainforest may extend back into the Late Pleistocene (1, 8, 9, 23).

### Ecological Predictions Match the Central African Archaeological Record.

Among other factors, high temperatures and abundant precipitation in rainforest environments limit the preservation of organic remains (12, 24). Thus, the archaeological record of the region is limited and biased toward nonforest areas. Therefore, we used the archaeological record to test our model's ability to predict the location and dating of known archaeological sites occupied by ancient hunter-gatherers, rather than to directly delimit the niche of CAHGs in the past (25). We compiled all published  $^{14}\text{C}$  dates ( $n = 168$ ) from hunter-gatherer archaeological sites from the Middle Stone Age onwards

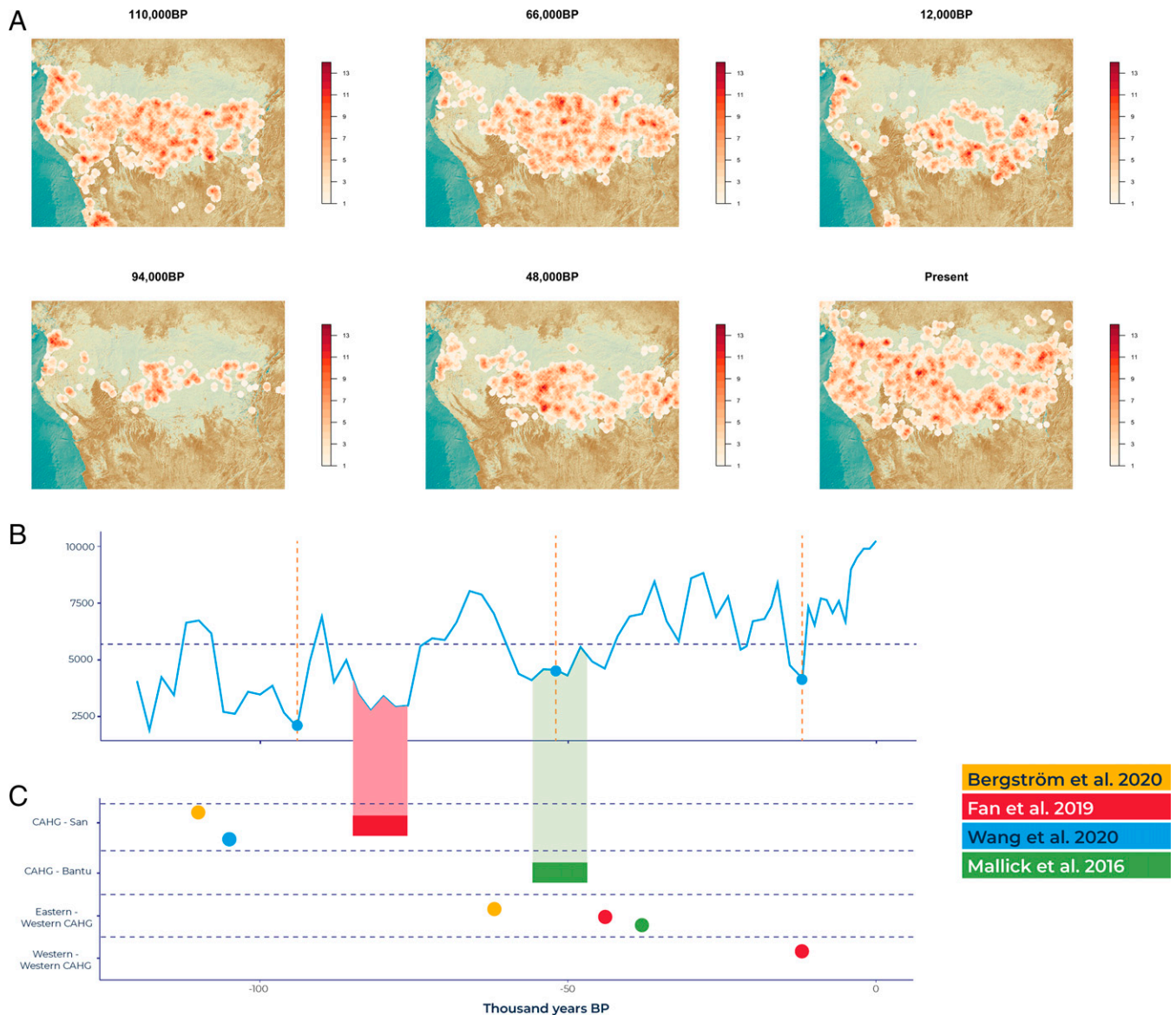
(*Datasets S2–S4, SI Appendix, Figs. S8 and S9, Fig. 2A, and Materials and Methods*). We removed sites with multiple dates within the same 1000-year time slice, resulting in a sample of  $n = 118$  dated sites. Analyses showed that archaeological sites were almost twice as likely to be found in cells predicted to be suitable by our model than in randomly selected cells in their corresponding time slices (Observed = 42, Expected = 23,  $\chi^2 = 12.677$ ,  $df = 1$ ,  $p < 0.0001$ ) (Fig. 2B). We then performed 1,000 random permutations of the  $^{14}\text{C}$  dates of our CAHG sites and confirmed that the number of sites falling in suitable cells when assigned their correct date was higher than predicted by any of the permutations (Fig. 2B; mean across permutations = 25.401,  $SD = 5.082$ ;  $t = 76.074$ ,  $df = 217383$ ,  $p < 0.0001$ ). Furthermore, our model performance was not related to the sites' dates ( $r = 0.003$ ,  $t = 0.393$ ,  $df = 107$ ,  $P = 0.968$ ). We also verified that the high suitability of areas where archaeological sites were located was not mediated by reduced soil acidity, since none of the ecological variables included in our ENM was significantly correlated to soil pH (*SI Appendix, Figs. S8 and S10*). This means that the high suitability of areas with sites was not simply the product of better preservation.

Our results indicate that the location of archaeological sites is not only determined by time-invariant geographic or topographic features but by how suitable such locations would have been for hosting hunter-gatherer populations at different points in time. The combined results of our models and the archaeological site distributions support the idea that CAHG have long been adapted to life in tropical environments, and expansions and contractions of suitable ranges have influenced the demography of CAHGs. An increasing fossil record is also consistent with an ancient occupation of Central African environments by hunter-gatherers (see *SI Appendix, Table S6* for a comprehensive summary of recent evidence) (26).

### Ecology-Driven Changes in the Range Size and Connectivity of CAHG Populations Explain Population Divergence.

To understand how environmental changes in Central Africa affected the deep evolutionary histories of African hunter-gatherers and their levels of genetic and cultural exchange, we modeled metapopulation sizes and connectivity over the past 120,000 years. For each time slice, we plotted all camps predicted by our ENM. To estimate their degree of interconnectivity, we calculated the cost of movement around each camp using Tobler's hiking function (27), which takes into account topographical features, rivers, and water masses that represent challenges to mobility. Then, we computed the total number of neighboring



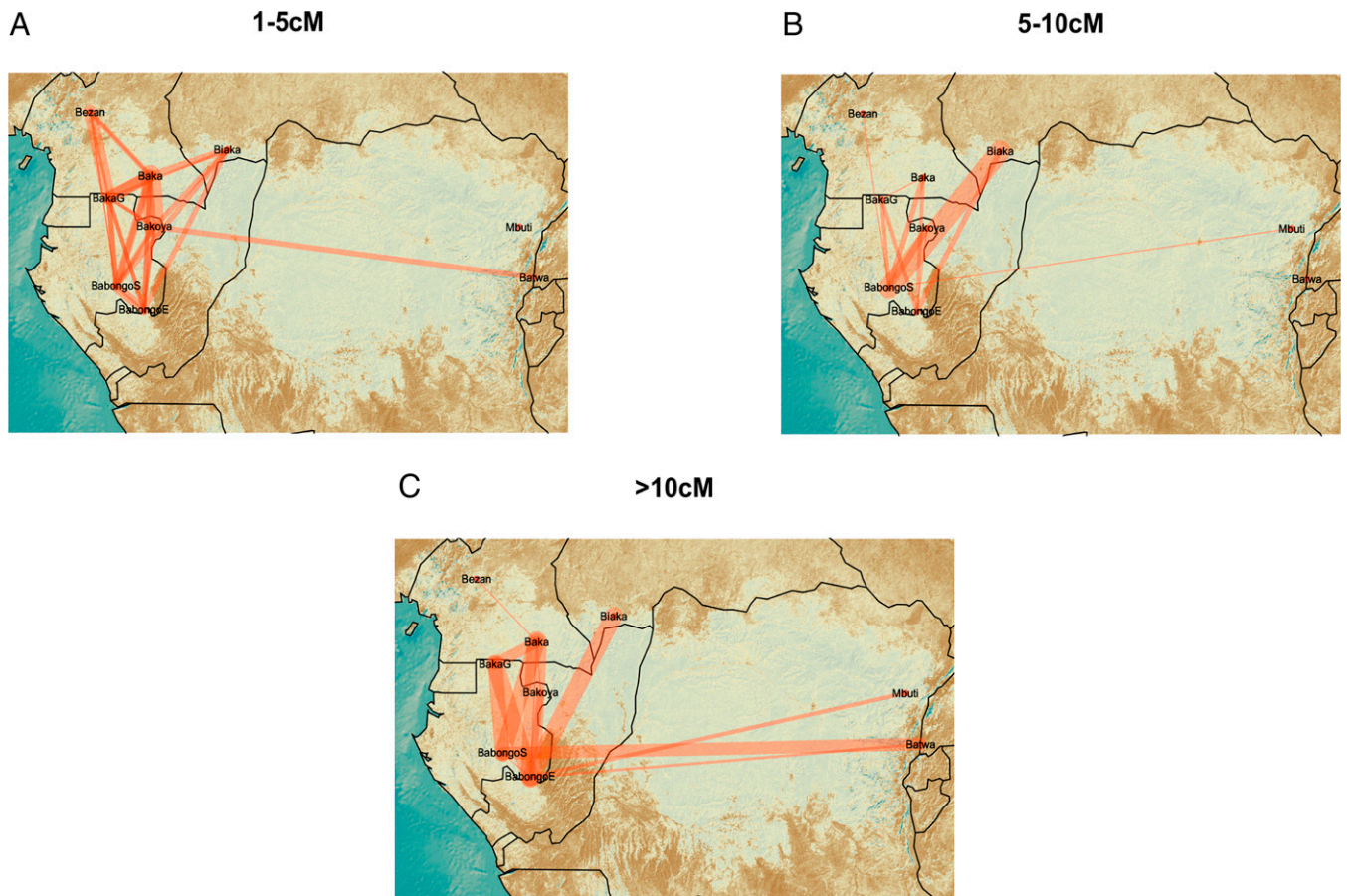


**Fig. 3.** Predicted dates of genetic splits between CAHG populations match decreases in predicted CAHG range sizes and group interconnectivity. (A) Predicted connectivity from our *MaxEnt* model at the time periods corresponding to the genetic estimates of population divergence times as well as before and after them. The predicted number of camps at each time period was randomly distributed across the cells with predicted presences, and the number of other camps within a 7-h walk of each camp calculated. Darker shading indicates a greater number of camps within a 7-h walking distance of one another. (B) Range size of CAHG across time predicted by our *MaxEnt* model. Vertical dashed lines indicate estimates of genetic split times by averaging all available studies for a particular event. Horizontal dashed line indicates average size of suitable range for CAHG. (C) Estimates of divergence times between African hunter-gatherer populations using cross-coalescence methods on whole genomes. Colored segments represent midpoint of cross-coalescence rates.

camps within a predicted 7-h walk from it. This radius corresponds to Cavalli-Sforza and Hewlett’s empirically derived (28) average “half range” of Aka hunter-gatherers and matches posterior estimates of life-time ranges of other CAHG populations (29, 30). To evaluate how changes in connectivity affected genetic differentiation among populations, we compiled recent genomic studies of CAHG demographic histories and their predicted divergences from other human lineages as well as from one another and compared our predicted range sizes and interconnectivity levels to genomic estimates of population splits obtained from whole-genome studies using cross-coalescence methods (Fig. 3C; see *SI Appendix, Table S1* for a compilation of all genomic estimates of divergence times between CAHG lineages).

Across different studies, the lineages of CAHG and of the San hunter-gatherers of Southern Africa (the earliest human genetic split) coalesce during the periods 105,000 to 110,000

BP and 68,000 to 87,000 BP (1, 31). These estimates match our model predictions of two periods of drastic reduction in the suitable range for hunter-gatherers, as well as decreased connectivity peaking at 94,000 BP, the midpoint of the divergence estimates between CAHG and the San across studies (Fig. 3A and B). The few hunter-gatherer populations remaining in Central Africa at this time would have been small and relatively isolated, compatible with separation between San and CAHGs (Fig. 3B and *SI Appendix, Fig. S11*). The next period where our model predicts a drastic contraction of the hunter-gatherer niche and decreased connectivity between Eastern and Western Central Africa lasts from 66,000 to 42,000 BP and encompasses the coalescence between CAHG lineages and those of farmers whose ancestry can be traced to the Bantu expansion (Fig. 3A–C) as well as between all Eastern and Western CAHG lineages (1, 31, 32).



**Fig. 4.** Recent genetic connectivity between CAHG populations. Network visualizations of the average number of IBD segments shared per cross-population individual pairs with identified IBD blocks in the range of: (A) 1 to 5 cM (2,500 to 1,500 y ago), (B) 5 to 10 cM (1,500 to 500 y ago), (C) and over 10 cM (500 to 0 y ago). Thicker lines indicate greater gene flow as identified by a higher probability of sharing IBD blocks of the specified length.

Finally, our model situates a last period of niche contraction and decreased connectedness starting at 17,000 BP and peaking at 12,000 BP, predicting the fewest number of suitable cells during the past 50,000 y. This is precisely the time range at which all Western CAHG lineages coalesce (*SI Appendix, Table S1* and Fig. 3C) (1). In the West, suitable ranges were restricted to the coast of what is now Cameroon. In summary, our data show that genetic estimates of splits among CAHG mostly take place during periods when the suitable range falls below average and connectivity between regions is reduced. A recent study (*SI Appendix, Table S1*) testing different Central African demographic models using exome data situated the Eastern-Western CAHG divergence during the 12,000 to 18,000 BP period, but we wish to note that these estimates are at odds with those relying on whole genome data.

**Evidence of Gene Flow between Eastern and Western Hunter-Gatherers in the Past 2,500 years.** Besides periods of ecological niche contraction favoring population splits, our model also predicts periods of noticeable niche expansion and gene flow between previously separated hunter-gatherer groups. We thus used patterns of sharing of identical-by-descent genetic segments (IBD) between populations to provide insights into recent contact among populations with deep common ancestries. We analyzed IBD blocks in three categories: 1 to 5, 5 to 10, and >10 centimorgans (cM) (Fig. 4), roughly corresponding to time intervals of 2,500 to 1,500 years ago, 1,500 to 500 years ago, and 500 to 0 years ago, respectively (33, 34). During the three time intervals, analyses identified widespread gene

flow among Western groups, as well as between Eastern and Western CAHG (Fig. 4 and *SI Appendix, Tables S7 and S8*). High levels of connectivity in the past 2,500 years between geographically and genetically differentiated hunter-gatherer groups argue against population decrease and fragmentation after the Bantu expansion (35, 36).

Although our data do not allow us to date intergroup connectivity at more ancient times, we also find evidence of distinct sharing of private alleles between all pairs of CAHG populations (*SI Appendix, Fig. S12*), again indicating widespread gene-flow between hunter-gatherer populations across Central Africa (37). Together with our simulations, this strongly suggests that the histories of fragmentation and connectivity among CAHG are equally deep. This pattern is compatible with recent studies indicating that rather than clean splits, genetic separations within Africa were gradual and shaped by ongoing gene flow (23), and with evidence of gene flow between Mbuti (East), Biaka (West), and San (South) until at least 50,000 y ago, and between Biaka and Mbuti (East) until the present day (31, 37). If this scenario is correct, the evolutionary history of *Homo sapiens* must be seen in the light of such long-term demographic dynamics involving both population fragmentation and interconnectivity.

Our ecological model can therefore recover the timing of major genetic events of population fragmentation and divergence both within central Africa, between Central African and other hunter-gatherers, and may account for the relatively high levels of genetic differentiation ( $F_{ST}$ ) among CAHG groups when compared to non-hunter-gatherer populations inhabiting



the same areas (ANOVA:  $F$ -value = 631.6,  $df$  = 2,  $P$  < 0.0001) (38) (*SI Appendix*, Fig. S13). Thus, our prediction of multiple cycles of isolation and reconnections over the past 120,000 years would account both for the maintenance of high effective population sizes throughout the Late Pleistocene (~30,000 individuals) despite the small band sizes documented in ethnographic records (1, 23, 31, 39–41) and of relatively high levels of genetic diversity compared to the more recent descendants of Bantu-speaking farmers. It would also explain the deep population structure of human occupation of Central African environments and remarkable morphological and archaeological diversity of early *Homo sapiens*, evidenced by the increasingly available fossil record for the region. (*SI Appendix*, Table S6) (42, 43).

## Discussion

We have modeled the role of environmental change since the last interglacial on the distribution, density, and dynamics of hunter-gatherers in Central Africa, a key geographical bridge between Northern, Western, Eastern, and Southern Africa but nonetheless traditionally left out from human evolutionary studies (12). Contrary to common assumptions, the current range and abundance of CAHG is mainly determined by long-term adaptation to ecological conditions rather than recent displacement by expanding Bantu-speaking farmers (44–46). Despite vast fluctuations in environment and available niche during the Late Pleistocene and Holocene eras, CAHGs would have always maintained relatively large, viable, and distinct populations (16, 21, 22). Future work could adopt a similar approach to determine the routes and niches followed by Bantu expansions, and in turn assess whether and how the latter may have affected CAHG presence and density over more recent time scales and in finer-grained resolution. Nonetheless, the fact that our modeled relationship between ecology and CAHG presence is consistent with the location of past archaeological assemblages and human remains suggests that as well as having remained genetically distinct for a very long time, contemporary African hunter-gatherers occupy similar habitats to their ancestors. It also supports recent morphological and genetic evidence that early hunter-gatherer populations in Africa were highly structured into semi- rather than fully-isolated groups (1, 9, 23, 42). This form of sociality characterized by a combination of local differentiation and partial connectivity, facilitated by the fluid structure of hunter-gatherer bands, may also explain the ability of early hunter-gatherer groups to maintain relatively large and stable effective population sizes (47–49) despite regular episodes of environmental fragmentation.

Although the increasing availability of paleoenvironmental reconstructions has resulted in a consensus over the extreme climatic variability characterizing African environments throughout human evolution (21, 22), few studies have attempted to directly test the effect of changing ecologies on human population dynamics (20, 50, 51). We hope that future research adopts similar spatio-temporally, ethnographically explicit approaches to further shed light on both the origins of modern human ancestry and the flexibility and adaptive potential of the human foraging niche. By determining the role of changing ecosystems in shaping variation within and between groups of humans across Africa, our study helps to characterize the breadth of the ecological niche exploited by early *Homo sapiens* populations known to extend well beyond coastal and savannah environments, and may provide valuable insights into past between-group relationships, as well as their relative contribution to present genetic, behavioral and cultural diversity.

## Materials and Methods

**Hunter-Gatherer Presence Data.** Our study area extended 6.7°N, 10.5°S, 31.6°E, 8.4°W, comprising five Central African countries (Cameroon, Central African Republic, Gabon, Republic of Congo, and the Democratic Republic of Congo). Within it, we obtained georeferenced location data for a total of 749 documented CAHG camps from a combination of primary literature and a previous study by Olivero et al. (13). We excluded camps that are the result of forced relocation after government resettlement programs (52) (*Dataset S1*).

Following Olivero et al. (13), we generated a buffer zone of 20 km of land around each camp representing a theoretical unit area of land liable to be exploited for natural resources. To determine this area, Olivero et al. (13) first determined the mean radius ( $18.5 \pm 1.0$  km) encircling a camp by using average subsistence area. They also obtained the average travel distance covered for subsistence activities from 36 studies ( $21.0 \pm 3.65$  km) (29, 53–55). From these two measures, the buffer zone around each camp was chosen to be of 20 km.

We then applied this buffer to all camps in our database to plot onto a  $0.1^\circ \times 0.1^\circ$  map of the study area. This resulted in 4,577 grid cells of CAHG presence, out of the total of 36,277 cells that covered the entire study area (Fig. 1B). We then considered absences or background (depending on the modeling algorithm used) to be those grid cells not included in the presence-grid-cells (see below).

**Environmental and Paleoenvironmental Data.** We made use of the same newly published high-resolution ( $0.5^\circ$ ) bias-corrected time series of global terrestrial climate and vegetation data covering the past 120,000 years (16) for both building our models and extrapolating them into the past. Gridded reconstructions are available at 2,000-year time steps between 120,000 and 22,000 BP, and 1,000-year time steps between 21,000 BP and the modern era. The data include 17 bioclimatic variables, which have been used extensively in environmental niche models as well as reconstructions of global biomes, leaf area index and net primary productivity. For interpretability purposes, we aggregated the 27 biomes simulated in Beyer et al. (16) into the same megabiomes as described in the original paper. In addition to bioclimatic variables, we obtained topographical and hydrographical variables from the US Geological Survey (56) as well as from HydroSHEDS (*SI Appendix*, Table S1).

After obtaining a selection of climatic, bioclimatic, and vegetation variables that are thought to affect the distribution of forager populations from previous literature, to avoid multicollinearity issues, we checked for pairwise correlations among them (*SI Appendix*, Fig. S14). Following Dormann et al. (57), we used an adapted version of the *select07* method to identify all pairs of variables with  $|r| < 0.7$  and remove the less important variable in terms of explained deviance. The final models included nine predictors (*SI Appendix*, Fig. S15 and Tables S2 and S3).

To fit the set of models (see below) designed to test the relative influence of the presence of farming populations whose ancestry can be traced to the Bantu expansion on the current distribution of CAHG, we also obtained data on the rural population density at each cell in our map as well as its distance to populated places. The former was taken from the Relational World Database II and the later from the LandScan 2008 High Resolution Global Population Data Set (58), excluding any areas less than 2-km far from urban areas. Although indirect, these two measures in conjunction are good proxies for assessing whether the range and density of these farming populations affect CAHG presence and density.

**Archaeological Data.** We compiled all published  $^{14}\text{C}$  dates from archaeological sites in the area of Central Africa covered by our study dating from the Middle Stone Age onwards. Their dates range from the present to 46,000, corresponding to the limit of radiocarbon dating (35, 36, 59). First, we removed  $^{14}\text{C}$  dates based either on carbonates—possibly affected by carbon reservoir effects—or with associated errors greater than 1,000 year (the resolution of our paleoclimatic) as well as dates associated with dubious stratigraphy or missing laboratory codes. (*Dataset S3*). The final sample after this stage comprised  $n = 930$  sites (*Dataset S3*). However, for the present study, only sites occupied by hunter-gatherers are relevant, hence, we needed to remove any sites that could have been occupied by Bantu populations.

From the primary literature, we obtained the chrono-cultural affiliation of each of the sites as well as checked the description of the assemblages to make

sure the criteria coincided. Presence of farming populations in archaeological sites was inferred from the occurrence of *Pennisetum glaucum* (Pearl millet), *Elaeis guineensis* (Oil palm), iron metallurgy, or pit features (35). Those sites that did not have such features but that had been assigned to either the “Neolithic,” “Early Iron Age,” or “Recent Iron Age” by their original authors were also considered to represent farming populations and hence excluded from our analyses. In addition, we checked the agreement of our chronocultural affiliation and assessment of dating reliability with that reported in recent summaries of the archaeological findings in Central Africa by Garcin et al. (36), Oslisly et al. (60), Morin-Rivat et al. (61), Cornelissen (24), and Seidensticker et al. (35). The remaining dates, that is, those with a chronocultural affiliation within the Stone Age (Middle or Later), no evidence of farming population occupation and reliable  $^{14}\text{C}$  dates (according to the criteria outlined above) were assumed to represent hunter-gatherer sites in the past. We classified these dates into two categories, the first one including the whole sample ( $n = 168$  dates) and the second including the subset of sites with associated lithics from hunter-gatherer populations ( $n = 151$  dates) (Dataset S2). We did not recalibrate the dates for the present study.

This sample is likely to be overly conservative, as our selection criteria implies that any sites that would have been co-occupied by hunter-gatherers and farming populations, where trade between foraging and farming groups would have traded tools or where hunter-gatherers would have utilized or cultivated any of the above-mentioned plants (something that is very common among present populations) (62, 63), would have automatically been removed from the sample. However, given that our aim was to test the strength of the evidence for an ancient occupation of Central Africa by hunter-gatherers, a conservative approach is desired. To further verify the robusticity of our results to different site-selection criteria, we ran the analyses on two additional datasets: one excluding several disputed dates and another one excluding sites attributed to hunter-gatherers but with evidence of pottery use (64) (Dataset S2 and SI Appendix, Figs. S16 and S17).

**Genetic Data.** Illumina array data were accessed from three published datasets (4, 65, 66). These data were filtered for relatedness  $>0.0886$  with KING (67). A random individual was chosen from each related pair to give a total of 1,325 individuals (SI Appendix, Table S7). The data were then pruned for MAF  $<0.05$  with PLINK (68). A pruned version of the dataset was created with PLINK—*indep-pairwise* 50 5 0.2 to account for linkage disequilibrium. The nonpruned dataset comprised 1,325 individuals and 555,630 single nucleotide polymorphisms (SNPs), and the pruned dataset comprised 1,325 individuals and 213,927 SNPs. The pruned dataset was used to assess the structure of genetic diversity within our area of study by calculating Wright’s  $F_{ST}$  statistics between all pairs of populations (38) using PLINK and for decomposing genetic variants into principal components using smartPCA with 10 outlier iterations (69).

**Environmental Niche Modeling of Distribution of CAHG.** Environmental Niche Models (ENM) or Species distribution models (SDM) are an extremely powerful tool from ecology (70). They mathematically relate occurrences of a particular species (in this case “species” corresponds to georeferenced hunter-gatherer camps) and the bioclimatic or ecological features of the areas it inhabits to produce a model that by identifying the “realized niche” of the species (71) can predict its potential geographic distribution based on suitable environmental conditions. Increasingly, ENMs have started being used to also understand the nonrandom distribution of human populations or cultural traditions in the present and in the past (13, 14, 25).

To model the potential distribution of CAHG throughout Central Africa, we used maximum entropy (MaxEnt) (70), one of the most widespread and best-performing techniques for ENM. Nonetheless, to minimize the potential of our model predictions to be reliant on the particular model algorithm used, we also performed the models using the Favorability function obtained from generalized linear models (GLMs), and compared their output with that obtained from MaxEnt (SI Appendix, Text). MaxEnt is a type of nonparametric machine-learning (ML) method. It takes a list of species presence locations as input, as well as a set of environmental predictors across a landscape divided into grid cells. From this landscape, MaxEnt extracts a sample of background locations that it contrasts against the presence locations. The model then estimates a relative occurrence rate (ROR) at each grid cell, as a function of the environmental predictors at that location (71). The ROR is the relative probability that a cell is contained in

a collection of presence samples. MaxEnt allows fitting very complex, highly non-linear response shapes (70). However, it also limits model complexity—and, hence, protects against overfitting by regularization: a penalty for each term included in the model and for higher weights given to a term (72). This is important given our primary aim to use the model to make predictions in environments that differ to those used to fit (in our case, paleoenvironments). We used all feature classes (linear, hinge, quadratic, product, threshold, and discrete), as our predictors are a mixture of categorical and continuous variables.

A key objective of our study was to use the fitted species distribution models to project hunter-gatherer distributions under past climate scenarios. MaxEnt is also the preferred model for extrapolating species’ distributions to new environments because it is “clamped,” that is, it extrapolates in a horizontal line from the most extreme environmental values in the training data set (70). Previous studies have also used this algorithm to model paleolithic niches of human populations (25).

Predicted suitabilities obtained from MaxEnt were converted to presence/absence predictions using the optimum threshold value maximizing the sum of sensitivity and specificity (73).

The ability of the model to predict hunter-gatherer camp occurrence was assessed using standard cross-validation procedure (80% random sample for calibration and 20% for validation; 500 repeats) (74). The predictive power of the binary models was determined by testing the accuracy of predictions made for the validation dataset (not used to build the model) by calculating the area under the curve of a receiver operating characteristic plot (AUC), Cohen’s Kappa statistic (75) and the TSS that unlike the Kappa statistic is independent of prevalence and therefore in recent years has been considered a more appropriate way of measuring the performance of ENM (76).

Last, to assess the robusticity of our model predictions to potential slight changes in the flow or position of minor rivers, we fitted a second set of models excluding “distance from water masses” as predictor.

**Estimation of Past Distributions of CAHGs.** The logistic output of MaxEnt consists of a grid map with each cell having an index of suitability between 0 and 1. Low values indicate that conditions are environmentally unsuitable for the presence of hunter-gatherer camps, whereas high values indicate that conditions are suitable.

After calibrating our model using the current range of hunter-gatherers (i.e., the present locations of hunter-gatherer camps), we projected it into the each 1,000- or 2,000-year time slice from the present to 120,000 BP to obtain suitability maps in our area of interest for each time period.

**Estimation of Camp and CAHG Population Density in the Present and Past.** The relationship between abundance and distribution range has been extensively studied in biogeography (77, 78). Abundance may be determined by limiting physiological variables or the ecological characteristics of species (18), which do not always exhibit regular spatial patterns (77). If these limiting factors are the same as those that also condition species presence, then models accounting for species occurrence could be useful in providing information on species abundance. Hence, from our distribution models, we estimated past and present population densities.

Since previous studies have found a positive association between population density and environmental suitability, the relationship between hunter-gatherer population density and environmental suitability was examined in 50 grid cells ( $n = 75$  camps) for which camp-size data were available. We calculated CAHG population densities from the sum of all CAHG population figures reported for the same  $0.1^\circ \times 0.1^\circ$  grid cell (123 km<sup>2</sup> at the Equator). We examined the shape of the population–suitability values point cloud after population-size outliers were eliminated following Tukey (79) (i.e., if population size  $>Q3 + 1.5 \times (Q3 - Q1)$ , where Q1 and Q3 are the first and the third quartiles, respectively) and found a typical wedge-shaped relationship.

We then used ordinary linear regression to test the significance of a positive relation between the population size at particular grid cells and their suitability values. In addition, since previous work suggests data on species distribution is particularly useful to predict the upper limit of species abundance [see Olivero et al. (13) for evidence of the same relationship when considering CAHG camps] we also fitted linear quantile regressions to the 50th, 55th, 60th, 65th, 70th, 75th, 80th, 90th, 95th, and 99th percentiles, and the R1 measure (weighted

sum of absolute residuals) was calculated in each percentile as a local measure of goodness-of-fit (80).

We use the term metapopulation here to encompass all spatially separated populations of CAHG groups (note that Olivero et al. (13) define it as all spatially separated CAHG groups that may interact to some extent). To estimate metapopulation size, first, we calculated the average CAHG population size empirically observed in grid cells for which population sizes were available (after removing outliers). Using the coefficient of the slope from the linear regression between suitability and population density, we then calculated the potential population size (PPS) for every grid cell in the study area, according to their suitability values. Finally, we summed all PPS values for the entire study area, but applied the following correction to take territoriality into account:

$$\text{Metapopulation size} = \text{GPPS} \times \text{GCS}/\text{ASA},$$

where the metapopulation size is the net potential population size; GPPS is the gross potential population size resulting from the sum of the PPS values; GCS is the size of a grid cell (i.e., 123 km<sup>2</sup>); and ASA is the average subsistence area estimated for CAHG (i.e., 1,079 km<sup>2</sup>).

**Estimating the Degree of Connectivity of Hunter-Gatherer Populations.** One of the largest challenges for estimating the potential for contact, and therefore the effective population size exchanging genetic and cultural knowledge between past and present hunter-gatherer populations is the very limited availability of data on hunter-gatherer mobility, movement ecology and resource use.

As a first estimate of the degree of connectivity of the predicted hunter-gatherer camps at every time step, we will use Cavalli-Sforza and Hewlett's (28) estimation of the average "half range" of Aka hunter-gatherers, defined as the median of the distance from their current place of residence to the places they had visited at least once during their lifetime. This measure was used to estimate a maximum distance between camps whose members could have interacted with one another. Their empirically derived measure was of 34.5 km or around 7 h of traveling time around a camp, is congruent with posterior estimates of life-time ranges of other CAHG populations (29, 30). Nonetheless, it is important to note that among Aka CAHG, rare movements of 60 to 80 km have also been recorded (28), mostly to visit family members living far away or to obtain work in villages (54).

Since the number of cells with predicted hunter-gatherer presences from our model (at any time period) would be larger than the actual number of hunter-gatherer camps (since all cells within a 20-km radius of a camp are considered to contain "presences"—as this represents the average subsistence area used by camps), we derived the ratio of camps to cells with presences from the original map, which was found to be of 0.16. Hence, for every time period, the predicted number of cells with presences was multiplied by 0.16 to obtain the predicted number of camps at that time-slice. Then, to estimate the distribution of camps from which to calculate connectivity, for each time-slice, the predicted number of camps by each of the algorithms was randomly distributed across the cells with predicted presences.

However, our study area is a mosaic topographical features, rivers, and water masses that would have presented logistical challenges to hunter-gatherers. To take such logistical challenges into account when considering camps' potential connectivity, we used the GTOPO30 Digital Elevation Map (56) to calculate the anisotropic accumulated cost of movement around each camp. We will do this with Tobler's (27) hiking function that estimates speed, taking into account the slope and its direction, and is the most popular cost function in archaeological least-cost path calculations (81, 82). Then, to estimate the degree of connectivity of each camp, we calculated the total number of other camps within 7 h of walk from that camp.

**Validation of the Model Using Archaeological Assemblages.** A great limitation for assessing our the adaptability of the ancestors of CAHG to rainforest environments is that high annual temperatures and abundant precipitation generate acidic conditions in the soil, which limit the preservation of organic remains such as pollen, soft plant tissues, bone, and wood (12, 24) (*SI Appendix, Fig. S8*). This means that the archaeological record of our area of interest is not only limited but biased toward nonforest areas where preservation is better. Consequently, following approaches that rely on the location of the archaeological sites themselves to delimit the niche of hunter-gatherers would not be justified (25).

Instead, we can use the archaeological record to test whether our modeled relationship between ecological variables and CAHG presence in the present (using an exhaustive dataset) also predicts the instances of CAHG presence in the past. In other words, we used the existing archaeological record to assess the continuity of the niche occupied by CAHG through time.

To test whether our estimated suitability landscapes in the past matched areas that were actually occupied by hunter-gatherer populations, we obtained the suitability values of the grid cells where archaeological sites are located at the time corresponding to their <sup>14</sup>C date (rounded to the nearest 1,000-years or 2,000-years according to the resolution of our paleoclimatic reconstructions for each time snap). We used the optimum threshold obtained when building the model for MaxEnt-derived predictions (0.31) to binarize those values, and in so, obtain the number of archaeological sites located in cells with predicted presences. To avoid sampling biases from sites with multiple dates close to one another, only one data point per 0.5 × 0.5° cell per 1,000- or 2,000-year interval was included (*Dataset S4*).

We then compared these values with the number of sites we would expect to observe in cells with predicted presences if sites were distributed randomly across the suitability landscape at each time period (that is, if suitability and archaeological site presence were independent of one another). In this way, we were able to calculate whether the likelihood of finding archaeological sites in cells with predicted presences was above chance.

To calculate the expected number of sites in predicted presences, we took, for each time period:

$$E_T = N_T \times PP_T/TC,$$

where  $E_T$  is the expected number of sites in predicted presences at time period  $T$ ,  $N_T$  is the number of sites with <sup>14</sup>C dates within time period  $T$ ,  $PP_T$  is the number of predicted cells with presences at time period  $T$ , and  $TC$  is the total number of grid cells in our map (36,277).

We added the expected number of cells with predicted presences across all time periods and divided it by the total number of archaeological sites to obtain the proportion of archaeological sites that would lie on cells with predicted presences if suitability and site location were independent of one another. We compared this with the observed proportion of sites located in cells with predicted presences.

We further validated the projections of our model by testing whether changes in the suitability landscapes over time indeed determined the spatial locations occupied by CAHG. To do this, we performed 1,000 random permutations of the <sup>14</sup>C dates of our sites. At each permutation, each site was randomly assigned one of <sup>14</sup>C dates from the list of dates from our sites, and the suitability value at the location of the site, at the time-period corresponding to the assigned <sup>14</sup>C date extracted. Then, for each permutation the total number of sites falling on predicted presences was counted, in so obtaining the number of sites in predicted presences for each of the 1,000 permutations where their <sup>14</sup>C dates were randomized. We then compared this number with the observed number of sites in predicted presences when assigned their real <sup>14</sup>C dates. To verify that any relationship between suitability and archaeological site presence was not mediated by soil acidity, we obtained high resolution data on soil acidity from the ISRIC world soil information dataset (83) and tested whether soil pH had a correlation coefficient above 0.7 with any of the ecological variables included in our model.

**Inferring Recent Genetic Connectivity between CAHG.** IBD sharing between populations provides insights into recent contact or recent common ancestry. Since CAHG have a deep divergence (1), shared IBD blocks can be used to infer recent contact (33, 34, 84). We analyzed IBD blocks in three categories: 1 to 5, 5 to 10, and >10 cM.

We used refinedIBD (85) to identify shared IBD blocks between each pair of individuals in our dataset without centromeres and telomeres, and homozygous-by-descent blocks within each individual (34) (*SI Appendix, Fig. S18*). Then, we 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 (86). These results were used to create three data sets based on the length of identified IBD blocks: 1 to 5 cM, 5 to 10 cM, and over 10 cM. For each data set, we summarized IBD sharing between populations, by considering the probability that an individual selected at random from population A shared IBD segments of the specified length with an individual selected at random from population B (84). This was done, for each pair of populations, by dividing the total



number of individual pairs connected by IBD segments of the length of the corresponding dataset by the total number of possible dyads for that pair of populations (87). We only kept the pairs with at least two shared blocks (4 for the range of 1 to 5 cM) to reduce noise and false positives. As a complementary measure of postdivergence gene flow between CAHG populations, we also calculated private allelic richness (per variable site) of alleles shared by pairwise combinations of the nine CAHG populations using ADZE (88) (*SI Appendix, Text*).

**Data Availability.** Data on the size and approximate location of camps are available as supplementary files (*Dataset S1*). Exact geographical coordinates remain confidential for protection of the privacy of the indigenous people involved. Data on the location and dating of hunter-gatherer archaeological sites are available as supplementary files (*Dataset S2*). Data on the location and

dating of initial  $^{14}\text{C}$  date compilation are available as supplementary files (*Dataset S3*). Data on the location and dating of hunter-gatherer archaeological sites after aggregating multiple dates within the same grid cell and 1,000- or 2,000-time interval are available as supplementary files (*Dataset S4*). Access to human genome-wide SNP data in the European Genome-Phenome Archive was granted via accession codes EGAS00001002078 and EGAS00001000908.

**ACKNOWLEDGMENTS.** The authors acknowledge all the anthropologists who have facilitated ethnographic data on hunter-gatherer camps, and Etienne Patin for providing the genetic data. We also thank Matt Peeples for help in calculating the connectivity between hunter-gatherer camps, Luisa Espinós Iglesias for assistance with visualization, and Bastiaan Star for support and guidance. Last, we are grateful to Hanna Kokko for helpful discussions of the manuscript.

- S. Fan *et al.*, African evolutionary history inferred from whole genome sequence data of 44 indigenous African populations. *Genome Biol.* **20**, 82 (2019).
- M. Lipson *et al.*, Ancient West African foragers in the context of African population history. *Nature* **577**, 665–670 (2020).
- C. M. Schlebusch *et al.*, Genomic variation in seven Khoe-San groups reveals adaptation and complex African history. *Science* **338**, 374–379 (2012).
- G. H. Perry *et al.*, Adaptive, convergent origins of the pygmy phenotype in African rainforest hunter-gatherers. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E3596–E3603 (2014).
- R. C. Bailey, T. N. Headland, The tropical rain forest: Is it a productive environment for human foragers? *Hum. Ecol.* **19**, 261–285 (1991).
- T. B. Hart, J. A. Hart, The ecological basis of hunter-gatherer subsistence in African rain forests: The Mbuti of Eastern Zaire. *Hum. Ecol.* **14**, 29–55 (1986).
- R. Blench, "Are the African pygmies an ethnographic fiction" in *Central African Hunter-Gatherers in a Multi-Disciplinary Perspective: Challenging Elusiveness*, K. Biesbrouck, S. Elders, G. Rossel, Eds. (Centre for Non-Western Studies, Leiden, 1999), pp. 41–60.
- S. Bahuchet, Changing language, remaining pygmy. *Hum. Biol.* **84**, 11–43 (2012).
- A. Bergström, C. Stringer, M. Hajdinjak, E. M. L. Scerri, P. Skoglund, Origins of modern human ancestry. *Nature* **590**, 229–237 (2021).
- J. Vansina, *Paths in the Rainforests: Toward a History of Political Tradition in Equatorial Africa* (Currey, 1990).
- S. Bahuchet, *Cultural Diversity of African Pygmies. Hunter-Gatherers of the Congo Basin: Cultures, Histories, and Biology of African Pygmies* (Transaction Publishers, New Brunswick, NJ, 2014), pp. 1–30.
- R. A. Foley, Evolutionary geography and the afro-tropical model of hominin evolution. *BMSAP* **30**, 17–31 (2018).
- J. Olivero *et al.*, Distribution and numbers of pygmies in Central African forests. *PLoS One* **11**, e0144499 (2016).
- N. Antunes, W. Schiefelhövel, F. d'Errico, W. E. Banks, M. Vanhaeren, Quantitative methods demonstrate that environment alone is an insufficient predictor of present-day language distributions in New Guinea. *PLoS One* **15**, e0239359 (2020).
- E. Patin *et al.*, Dispersals and genetic adaptation of Bantu-speaking populations in Africa and North America. *Science* **356**, 543–546 (2017).
- R. M. Beyer, M. Krapp, A. Manica, High-resolution terrestrial climate, bioclimate and vegetation for the last 120,000 years. *Sci. Data* **7**, 236 (2020).
- C. Finlayson, *The Improbable Primate: How Water Shaped Human Evolution* (Oxford University Press, 2014).
- R. D. Sagarin, S. D. Gaines, Geographical abundance distributions of coastal invertebrates: Using one-dimensional ranges to test biogeographic hypotheses. *J. Biogeogr.* **29**, 985–997 (2002).
- R. K. Hitchcock, Foragers and food production in Africa: A cross-cultural and analytical perspective. *World J. Agri. Soil Sci.* **1** (2019).
- R. M. Beyer, M. Krapp, A. Eriksson, A. Manica, Climatic windows for human migration out of Africa in the past 300,000 years. *Nat. Commun.* **12**, 4889 (2021).
- J. Maley *et al.*, Late Holocene forest contraction and fragmentation in central Africa. *Quat. Res.* **89**, 43–59 (2018).
- S. Blockley *et al.*, The resilience of postglacial hunter-gatherers to abrupt climate change. *Nat. Ecol. Evol.* **2**, 810–818 (2018).
- K. Wang *et al.*, Ancient genomes reveal complex patterns of population movement, interaction, and replacement in sub-Saharan Africa. *Sci. Adv.* **6**, eaaz0183 (2020).
- E. Cornelissen, "The later Pleistocene in the northeastern Central African rainforest" in *Africa from MIS 6-2*, S. Jones and B. Stewart, Eds. (Springer, 2016), pp. 301–319.
- F. d'Errico *et al.*, Identifying early modern human ecological niche expansions and associated cultural dynamics in the South African Middle Stone Age. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 7869–7876 (2017).
- R. Lanfranchi, B. Clist, Y. de La Croix, *Aux origines de l'Afrique centrale* (Centre international des civilisations bantu [Sépie], 1991).
- W. Tobler, "Three presentations on geographical analysis and modeling" (Tech. Rep. 93-1, National Center for Geographic Information and Analysis, 1993).
- L. L. Cavalli-Sforza, B. Hewlett, Exploration and mating range in African pygmies. *Ann. Hum. Genet.* **46**, 257–270 (1982).
- H. Yasuoka, Long-term foraging expeditions (Molongo) among the Baka Hunter-Gatherers in the Northwestern Congo Basin, with special reference to the "Wild Yam Question". *Hum. Ecol.* **34**, 275–296 (2006).
- H. Jang, C. Boesch, R. Mundry, S. D. Ban, K. R. L. Janmaat, Travel linearity and speed of human foragers and chimpanzees during their daily search for food in tropical rainforests. *Sci. Rep.* **9**, 11066 (2019).
- A. Bergström *et al.*, Insights into human genetic variation and population history from 929 diverse genomes. *Science* **367**, eaay5012 (2020).
- S. Mallick *et al.*, The simons genome diversity project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016).
- P. Ralph, G. Coop, The geography of recent genetic ancestry across Europe. *PLoS Biol.* **11**, e1001555 (2013).
- D. Liu *et al.*, Extensive ethnolinguistic diversity in Vietnam reflects multiple sources of genetic diversity. *Mol. Biol. Evol.* **37**, 2503–2519 (2020).
- D. Seidensticker *et al.*, Population collapse in Congo rainforest from 400 CE urges reassessment of the Bantu expansion. *Sci. Adv.* **7**, eabd8352 (2021).
- Y. Garcin *et al.*, Early anthropogenic impact on Western Central African rainforests 2,600 y ago. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 3261–3266 (2018).
- C. M. Schlebusch *et al.*, Khoe-San genomes reveal unique variation and confirm the deepest population divergence in Homo sapiens. *Mol. Biol. Evol.* **37**, 2944–2954 (2020).
- B. S. Weir, C. C. Cockerham, Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370 (1984).
- S. Bahuchet, "Spatial mobility and access to resources among the African pygmies" in *Mobility and Territoriality*, M. J. Casimir, A. Rao, Eds. (Routledge, 1992).
- N. Matsuura, Sedentary lifestyle and social relationships among Babongo in Southern Gabon. *Afr. Stud. Monogr.* **33**, 71–93 (2006).
- C. Turnbull, The Mbuti pygmies: An ethnographic survey. *Anthropol. Pap. Am. Mus. Nat. Hist.* **50**, 139–282 (1965).
- E. M. L. Scerri *et al.*, Did our species evolve in subdivided populations across Africa, and why does it matter? *Trends Ecol. Evol.* **33**, 582–594 (2018).
- I. Crevecoeur, A. Brooks, I. Ribot, E. Cornelissen, P. Semal, Late Stone Age human remains from Ishango (Democratic Republic of Congo): New insights on Late Pleistocene modern human diversity in Africa. *J. Hum. Evol.* **96**, 35–57 (2016).
- R. Grollemund *et al.*, Bantu expansion shows that habitat alters the route and pace of human dispersals. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 13296–13301 (2015).
- K. Bostoen *et al.*, Middle to late Holocene Paleoclimatic change and the early Bantu expansion in the rain forests of Western Central Africa. *Curr. Anthropol.* **56**, 367–368 (2015).
- K. A. Klieman, "The Pygmies Were Our Compass": *Bantu and Batwa in the History of West Central Africa, Early Times to c. 1900 C.E.* (Heinemann, 2003).
- M. Sikora *et al.*, Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers. *Science* **358**, 659–662 (2017).
- A. B. Migliano *et al.*, Characterization of hunter-gatherer networks and implications for cumulative culture. *Nature Human Behaviour* **1**, 0043 (2017).
- K. R. Hill, B. M. Wood, J. Baggio, A. M. Hurtado, R. T. Boyd, Hunter-gatherer inter-band interaction rates: Implications for cumulative culture. *PLoS One* **9**, e102806 (2014).
- J. R. Stewart, C. B. Stringer, Human evolution out of Africa: The role of refugia and climate change. *Science* **335**, 1317–1321 (2012).
- A. Eriksson *et al.*, Late Pleistocene climate change and the global expansion of anatomically modern humans. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16089–16094 (2012).
- J. Lewis, *The Batwa Pygmies of the Great Lakes Region* (Minority Rights Group International London, 2000).
- A. Hoare, "Resource rights and timber concessions: Integrating local peoples' land-use practices in forest management in the Congo Basin" (The Rainforest Foundation, London, UK, 2007).
- B. Hewlett, J. M. H. van de Koppel, L. L. Cavalli-Sforza, Exploration ranges of aka pygmies of the Central African Republic. *Man (Lond.)* **17**, 418 (1982).
- J. A. Hart, From subsistence to market: A case study of the Mbuti net hunters. *Hum. Ecol.* **6**, 325–353 (1978).
- US Geological Survey, *USGS EROS Archive - Digital Elevation - Global 30 Arc-Second Elevation (GTOPO30)* (US Geological Survey, Reston, VA, 1996).
- C. F. Dormann *et al.*, Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **36**, 27–46 (2013).
- E. Bright, A. Rose, M. Urban, J. McKee, LandScan 2008 high-resolution global population data set (Oak Ridge National Laboratory, Oak Ridge, TN, 2008).
- B. Clist, *Gabon, 100 000 ans d'histoire* (Centre culturel français Saint-Exupéry, Sépie, 1995).
- R. Oslisly *et al.*, Climatic and cultural changes in the west Congo Basin forests over the past 5000 years. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**, 20120304 (2013).
- J. Morin-Rivat *et al.*, New evidence of human activities during the Holocene in the lowland forests of the northern Congo Basin. *Radiocarbon* **56**, 209–220 (2014).
- K. Kitanishi, Cultivation by the Baka hunter-gatherers in the tropical rain forest of central Africa. *Afr. Study Monogr. Suppl. Issue* **28**, 143–157 (2003).
- C. O. Biloa, The Bagyéli of South-west Cameroon. *Hunt. Gather. Res.* **1**, 317–332 (2015).
- J. R. Denbow, *The Archaeology and Ethnography of Central Africa* (Cambridge University Press, 2014).
- E. Patin *et al.*, The impact of agricultural emergence on the genetic history of African rainforest hunter-gatherers and agriculturalists. *Nat. Commun.* **5**, 3163 (2014).
- M. Fagny *et al.*, The epigenomic landscape of African rainforest hunter-gatherers and farmers. *Nat. Commun.* **6**, 10047 (2015).
- A. Manichaikul *et al.*, Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
- S. Purcell *et al.*, PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

69. A. L. Price *et al.*, Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
70. J. Elith *et al.*, A statistical explanation of MaxEnt for ecologists: Statistical explanation of MaxEnt. *Divers. Distrib.* **17**, 43–57 (2011).
71. G. E. Hutchinson, The multivariate niche. *Cold Spring Harb. Symp. Quant. Biol.* **22**, 415–421 (1957).
72. S. J. Phillips, M. Dudík, Modeling of species distributions with Maxent: New extensions and a comprehensive evaluation. *Ecography* **31**, 161–175 (2008).
73. C. Liu, M. White, G. Newell, Selecting thresholds for the prediction of species occurrence with presence-only data. *J. Biogeogr.* **40**, 778–789 (2013).
74. D. Zurell *et al.*, A standard protocol for reporting species distribution models. *Ecography* **43**, 1261–1277 (2020).
75. O. Allouche, A. Tsoar, R. Kadmon, Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *J. Appl. Ecol.* **43**, 1223–1232 (2006).
76. Y. Fourcade, A. G. Besnard, J. Secondi, Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Glob. Ecol. Biogeogr.* **27**, 245–256 (2018).
77. R. D. Sagarin, S. D. Gaines, B. Gaylord, Moving beyond assumptions to understand abundance distributions across the ranges of species. *Trends Ecol. Evol.* **21**, 524–530 (2006).
78. J. VanDerWal, L. P. Shoo, C. N. Johnson, S. E. Williams, Abundance and the environmental niche: Environmental suitability estimated from niche models predicts the upper limit of local abundance. *Am. Nat.* **174**, 282–291 (2009).
79. J. W. Tukey, *Exploratory Data Analysis* (Addison-Wesley, Reading, MA, 1977).
80. R. Koenker, J. A. Machado, Goodness of fit and related inference processes for quantile regression. *J. Am. Stat. Assoc.* **94**, 1296–1310 (1999).
81. I. Herzog, A. Yépez, "Least-Cost Kernel density estimation and interpolation-based density analysis applied to survey data" in *Fusion of Cultures. CAA 2010-Proceedings of the 38th Conference on Computer Applications and Quantitative Methods in Archaeology*, F. Contreras, M. Farjas, F. J. Melero, Eds. (Grenada, Spain, 2013), pp. 367–374.
82. D. A. White, S. B. Barber, Geospatial modeling of pedestrian transportation networks: A case study from precolumbian Oaxaca, Mexico. *J. Archaeol. Sci.* **39**, 2684–2696 (2012).
83. L. M. de Sousa *et al.*, "SoilGrids 2.0: Producing quality-assessed soil information for the globe" *Soils and the Natural Environment* (2020). <https://doi.org/10.5194/soil-2020-65>. Accessed 25 November 2021.
84. A. G. Ioannidis *et al.*, Native American gene flow into Polynesia predating Easter Island settlement. *Nature* **583**, 572–577 (2020).
85. B. L. Browning, S. R. Browning, Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics* **194**, 459–471 (2013).
86. S. R. Browning *et al.*, Ancestry-specific recent effective population size in the Americas. *PLoS Genet.* **14**, e1007385 (2018).
87. L. R. Botigué *et al.*, Gene flow from North Africa contributes to differential human genetic diversity in southern Europe. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 11791–11796 (2013).
88. Z. A. Szpiech, M. Jakobsson, N. A. Rosenberg, ADZE: A rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics* **24**, 2498–2504 (2008).