



Published in final edited form as:

Nature. 2021 February ; 590(7844): 103–110. doi:10.1038/s41586-020-03053-2.

A genetic history of the pre-contact Caribbean

A full list of authors and affiliations appears at the end of the article.

Abstract

Humans settled the Caribbean ~6,000 years ago, with ceramic use and intensified agriculture marking a shift from the Archaic to the Ceramic Age ~2,500 years ago^{1–3}. We report genome-wide data from 174 individuals from The Bahamas, Hispaniola, Puerto Rico, Curaçao, and Venezuela co-analyzed with published data. Archaic Age Caribbean people derive from a deeply divergent population closest to Central and northern South Americans; contrary to previous work⁴, we find no support for ancestry contributed by a population related to North Americans. Archaic lineages were >98% replaced by a genetically homogeneous ceramic-using population related to Arawak-speakers from northeast South America who moved through the Lesser Antilles and into the Greater Antilles at least 1,700 years ago, introducing ancestry that is still present. Ancient Caribbean people avoided close kin unions despite limited mate pools reflecting small effective population sizes which we estimate to be a minimum of $N_e=500–1500$ and a maximum of $N_e=1530–8150$ on the combined islands of Puerto Rico and Hispaniola in the dozens of generations before the analyzed individuals lived. Census sizes are unlikely to be more than ten-fold larger than effective population sizes, so previous estimates of hundreds of thousands of people are too large^{5–6}. Confirming a small, interconnected Ceramic Age population⁷, we detect 19 pairs of cross-island cousins, close relatives ~75 kilometers apart in Hispaniola, and low genetic differentiation across islands. Genetic continuity across transitions in pottery styles reveals that cultural changes during the Ceramic Age were not driven by migration of genetically-

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

Correspondence and requests for materials should be addressed to A. Cop. (alfredo.coppa@uniroma1.it), R.P. (ron.pinhasi@univie.ac.at), or D.R. (reich@genetics.med.harvard.edu).

*These authors contributed equally to this work.

Author Contributions W.F.K., A.Cop., M.L., R.P., and D.R. supervised the study. J.S., O.C., C.A.A., E.V.C., R.C., A.Cuc., F.G., C.K., F.L.P., M.L., M.V.M., C.T.M., C.M., I.P., M.P., T.S., C.G.S., and M.V. provided skeletal materials and/or assembled and interpreted archaeological and anthropological information. C.A.A., E.V.C., C.K., M.V.M., C.T.M., C.M., I.P., M.P., T.S., and C.G.S. contributed local perspectives to the interpretation and contextualization of new genetic data. B. M.-T. provided data from present-day populations. N.R., M.M., S.M., N.A., R.B., G.B., N.B., O.C., K.C., F.C., L.D., K.S.D.C., S.F., A.M.L., K.M., J.O., K.Ö., C.S., R.S., K.St., and F.Z. performed ancient DNA laboratory and/or data-processing work. B.J.C., R.J.G., L.E., F.M., W.C.M., F.T. and D.J.K. performed radiocarbon analysis and stable isotope work; D.J.K. supervised this work. D.F., K.Si., H.R., M.M., S.M., I.O., and M.L. analysed genetic data. D.F., K.Si., W.F.K., and D.R. wrote the manuscript with input from all co-authors.

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/s41586-020-03053-2>.

Code availability The custom code used in this study is available from <https://github.com/DReichLab/ADNA-Tools>.

Data availability The aligned sequences are available through the European Nucleotide Archive under accession number PRJEB38555. Genotype data used in analysis are available at <https://reich.hms.harvard.edu/datasets>. Any other relevant data are available from the corresponding authors upon reasonable request.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

differentiated groups from the mainland but instead reflected interactions within an interconnected Caribbean world^{1,8}.

Prior to European colonization, the Caribbean was a mosaic of archaeologically-distinct communities connected by networks of interaction since the first human occupations in Cuba, Hispaniola, and Puerto Rico around 6,000 years ago^{3,7}. The pre-contact Caribbean is divided into three archaeological Ages that denote shifts in material cultural complexes^{1,9}. The Lithic and Archaic Ages are defined by distinct stone-tool technologies^{10–11}, while the Ceramic Age, beginning ~2,500–2,300 years ago, featured an agricultural economy and intensive pottery production. Technological and stylistic changes in material culture across these Ages reflect local developments by connected Caribbean people and also migration from the American continents, although the geographic origins, trajectories, and numbers of migratory waves remain under debate^{1,3,12} (Table 1; Supplementary Information section 1).

We screened 195 individuals and generated genome-wide data passing authenticity criteria for 174 individuals (Supplementary Data 1, 2) who lived ~3100–400 calibrated years before present (calBP; based on 45 new radiocarbon dates, Extended Data Fig. 1a; Supplementary Data 3; Supplementary Information section 3) in The Bahamas, Hispaniola (Haiti and the Dominican Republic), Puerto Rico, Curaçao, and Venezuela (Fig. 1a; Supplementary Information section 2). These individuals had a median of 700,689 SNPs covered (range: 20,063–977,658 SNPs, median of 2.2× coverage of targeted positions (range: 0.02–9.95×), Supplementary Data 1). We co-analyzed the new data alongside 89 previously-published individuals⁴ (Supplementary Information section 4). In what follows, we denote sites with stone tools or radiocarbon dates predating intensive ceramic use as ‘Archaic’ and sites with a preponderance of ceramics as ‘Ceramic’; we use ‘-related’ to refer to ancestry and ‘-associated’ for archaeological affiliation.

Ethics

We acknowledge the ancient individuals whose skeletal remains we analyzed, present-day people who have an Indigenous legacy, and Caribbean-based scholars who were centrally involved in this work. Permission to perform ancient DNA analysis was documented through authorization letters signed by a custodian who represented the remains from each site. Results were discussed prior to submission with members of Indigenous communities who trace their legacy to the pre-contact Caribbean and their feedback was incorporated. Genetic data are a form of knowledge that contributes to understanding the past; they co-exist with oral traditions and other Indigenous knowledge. Genetic ancestry should not be conflated with perceptions of identity, which cannot be defined by genetics alone. A full ethics statement is in the Supplementary Information.

Genetic structure of the pre-contact Caribbean

We performed principal component analysis (PCA), projecting ancient individuals onto axes computed using present-day Indigenous American groups¹³ (Extended Data Fig. 1b; Supplementary Data 4). Ceramic- and Archaic-associated individuals project in separate clusters, while ancient Venezuelans relate to present-day Chibchan-speakers (like Cabécar)

in PCA and ADMIXTURE analysis (Extended Data Figs. 1b, 1c; Supplementary Information sections 5, 6; population self-denominations in Supplementary Data 5). Individuals from Curaçao and Haiti (who are admixed, discussed below) mostly overlap the Ceramic-associated cluster. An exception to within-site genetic homogeneity is at Andrés (a primarily Ceramic-associated site, Dominican Republic), where individual I10126 is dated to the Archaic Age (~3140–2950 calBP, Supplementary Data 3) and appears genetically similar to other Archaic-associated individuals (Extended Data Figs. 1b, 1c). We exclude from subsequent analyses three Archaic-associated individuals from Cueva Roja (~1900 calBP, Dominican Republic) with low coverage ($< \sim 0.05\times$) who are qualitatively similar to other Archaic-associated individuals, and one individual from three pairs of first-degree relatives (Supplementary Data 1).

To study genetic structure independent of archaeologically-based assignments (Supplementary Information section 2), we grouped individuals with increasing resolution based on allele sharing, starting with major ‘clades’ and then ‘sub-clades’ (Supplementary Information section 8). Our nomenclature combined the geographic location encompassing sites in the cluster plus ‘Archaic’ or ‘Ceramic’ (Fig. 1b).

We identified three significantly differentiated major clades. *GreaterAntilles_Archaic* included 50 individuals from Cuba spanning ~3200–700 calBP⁴ and individual I10126 from Andrés (Dominican Republic). *Caribbean_Ceramic* comprised 194 individuals from Ceramic-associated sites dating ~1700–400 calBP. *Venezuela_Ceramic* comprised eight individuals dated ~2350 calBP. Two *Haiti_Ceramic* and five *Curacao_Ceramic* individuals fit as mixtures of major clades (below).

We next identified sub-clades and substructure within them (Supplementary Data 6; Table S6). Within *Caribbean_Ceramic*, *SECoastDR_Ceramic* comprised four sites along 50 kilometers of the southeast coast of the Dominican Republic (from west to east: La Caleta, Andrés, Juan Dolio, and El Soco) (Table S7). These sites were occupied for ~1,400 years, documenting genetic continuity across changes in ceramic styles. All Ceramic-associated sites from The Bahamas and Cuba (spanning ~700 years) grouped as *BahamasCuba_Ceramic*, and further substructure was present in each of five Bahamian islands and two Cuban sites. The two sites in the Lesser Antilles grouped as *LesserAntilles_Ceramic*, and the remaining sites from *Caribbean_Ceramic* grouped as *EasternGreaterAntilles_Ceramic*, showing no cross-site substructure. Pairwise $F_{ST} < \sim 0.01$ indicates a striking degree of homogeneity among these *Caribbean_Ceramic* sub-clades (compared to $F_{ST} \sim 0.1$ between Ceramic- and Archaic-related clades), reflecting high migration rates among islands (discussed below; Extended Data Fig. 2).

To identify *Caribbean_Ceramic* individuals who had an excess of Archaic-related ancestry relative to others within each sub-clade, we used f_4 -statistics (Supplementary Information section 8; Supplementary Data 8). Individual I16539 from La Caleta (Dominican Republic) and the two individuals comprising *Haiti_Ceramic* showed significant evidence of Ceramic-/Archaic-related admixture ($Z = -5.5$; Table S8). In contrast to a previous claim¹¹, we did not detect significant Archaic-related admixture in individual PDI009 from Paso del Indio (Puerto Rico) ($Z = 0.6$; Supplementary Information section 4; Table S3).

Archaic-associated Caribbean people

The *GreaterAntilles_Archaic* clade shares the most genetic drift with Indigenous groups from Central and northern South America belonging to seven language families: Arawakan, Cariban, Chibchan, Chocoan, Guajiboan, Mataco-Guaicuru, and Tupian^{14,15} (Fig. 2a; Supplementary Data 10; Supplementary Information section 11). There is no evidence of excess allele sharing with people from one language family relative to the others or evidence of genetic drift specifically shared with present-day populations from Mesoamerica or North America (Fig. 2a, 2b; Supplementary Data 11). Archaic-associated individuals from Cuba share more alleles with each other than with Dominican individual I10126 (Table S6), demonstrating Archaic substructure; we separate individual I10126 as *Dominican_Andres_Archaic* for some analyses.

We could not replicate a previous claim that a migration by people with affinity to North Americans also contributed ancestry to some Archaic Age Caribbean individuals⁴ (Supplementary Information section 17). This claim was based on a finding of affinity between Early Period individuals from California's Channel Islands (*USA_CA_Early_SanNicolas*) and individual CIP009 from Cueva del Perico (Cuba) relative to individual GUY002 from Guayabo Blanco (Cuba). First, in the symmetry test $f_4(GUY002, CIP009; USA_CA_Early_SanNicolas, Bahamas_Taino)$, the deviation is non-significant ($Z=-0.9$; Table S25). Second, a key statistic underlying this claim was that a *qpWave*-based symmetry test involving CIP009 and GUY (three individuals from Guayabo Blanco) yielded $p=0.013$; however, this is not significant after correcting for the number of sample pairs tested. Third, we computed $f_4(Outgroup, CIP009; USA_CA_Early_SanNicolas, Bahamas_Taino)$, whose negative value was interpreted as evidence for affinity between CIP009 and *USA_CA_Early_SanNicolas*; while we replicated the non-significant statistic ($Z=-1.3$; Table S23), it became positive when we replaced the *Mbuti* outgroup with diverse Eurasians or *Bahamas_Taino*¹⁶ with ancient Bahamian shotgun data newly generated for this study, which should give qualitatively similar results (Tables S24 and S26). Fourth, the (non-significant) Z -scores for attraction to CIP009 were as strong when South American ancient genomes were placed in the position of *USA_CA_Early_SanNicolas*, showing no evidence of a North American-specific relationship (Table S27). Fifth, CIP009 fits best in a simplified version of our *qpGraph* tree on the same node as other Archaic-associated individuals (Supplementary Information section 17; Fig. S34). Thus, to the limits of the resolution of allele sharing methods, all Archaic-associated Caribbean ancestry is consistent with deriving from a single source.

In *qpGraph*, we fit *GreaterAntilles_Archaic* in an early splitting branch containing most ancient Caribbean, Belizean, Brazilian, and Argentinian populations (Fig. 2c). In a maximum likelihood tree allowing admixture events¹⁷, *GreaterAntilles_Archaic* also fits as a divergent Native American group (Extended Data Fig. 3). We could not obtain further evidence of specific affinities to mainland groups using *qpAdm* (Supplementary Information section 9; Table S16) or f_{ST} -statistics (Table S17).

The arrival of ceramic users displaced Archaic-related ancestry in much of the Caribbean. An exception is western Cuba, where Archaic lineages persisted with minimal mixture for

>2,500 years, resonating with archaeological¹⁸ and historical¹⁹ accounts that this region was home to people with a distinct language and cultural traditions as late as the Contact Period.

The spread of ceramic users

Previous analyses have found that Caribbean Ceramic-associated people have genetic affinities to Arawak-speakers in northeastern South America^{16,20,21} (Supplementary Information section 1). Although we are not able to support this conclusion with symmetry f_4 -statistics which show no significant evidence of closer relatedness to Arawak- than to Cariban- or Tupian-speaking populations (Fig. 2b; Supplementary Data 11; Supplementary Information section 11), ADMIXTURE suggests an Arawak affinity, as individuals from each *Caribbean_Ceramic* sub-clade are almost entirely composed of a component found in the highest proportion in modern Arawak speakers (e.g., Piapoco in Extended Data Fig. 1c). We also find support for an Arawak connection in a maximum likelihood tree allowing admixture events, which places all *Caribbean_Ceramic* sub-clades on the same branch as Arawak-speaking Piapoco and Palikur (Extended Data Fig. 3). Further evidence comes from a successful fit with Piapoco as the single source for *Caribbean_Ceramic* in *qpAdm* (Tables S18, S19), and *qpGraph* (Fig. 2c).

We estimate ~0.5–2.0% Archaic-related ancestry in the Ceramic-associated people of the Greater Antilles and The Bahamas when modeled in *qpAdm* as a mixture of *LesserAntilles_Ceramic* and *Dominican_Andres_Archaic* (Table S21). We reject reverse models of *LesserAntilles_Ceramic* deriving from Greater Antilles or Bahamas/Cuba-based sub-clades which fail when Archaic-associated people are included in the reference set ($p=0.001$ – 0.008 , Table S21). This supports a scenario of south-to-north movement of ceramic using ancestors into the Caribbean, whereby ancestry like that in the 1000–650 BP ancient Lesser Antilles individuals (plausibly descended from the first ceramic users of the Lesser Antilles) spread into the Greater Antilles and The Bahamas, displacing the people that lived there with no more than ~2.0% mixture with resident groups.

We found only three individuals from two Ceramic-associated sites in Hispaniola with significant Archaic-related admixture, who we estimate using *qpAdm* to have Archaic-related ancestry in proportions ranging between $11.8\pm 1.9\%$ (I16539 from La Caleta, Dominican Republic; Table S9) and $18.5\pm 2.1\%$ (two individuals from Diale 1, Haiti; Tables S12, S13). Using DATES²², we estimate that admixture occurred $\sim 16\pm 3$ generations (~ 350 – 500 years) before these individuals from Haiti lived (Supplementary Information section 14).

Venezuela_Ceramic's affinities with Chibchan speakers in ADMIXTURE and f -statistics (Fig. 2a, 2b; Extended Data Fig. 1c) are confirmed in *qpAdm* where *Venezuela_Ceramic* fits as a clade with Cabécar (Tables S18, S19). Thus, although Las Locas is located in a hypothesized source region for the Ceramic expansion and the individuals date to near the beginning of the Ceramic Age, our analysis increases the weight of evidence that this expansion had more easterly origins. We model ceramic users from Curaçao as $74.5\pm 3.7\%$ *LesserAntilles_Ceramic*-related ancestry and $25.5\pm 3.7\%$ *Venezuela_Ceramic*-related ancestry (Table S15), suggesting that Curaçao's Ceramic Age population was derived from

the admixture of two groups: one related to the population that also spread to the Antillean Caribbean at the onset of the Ceramic Age, and the other associated with the Dabajuroid ceramic styles linking sites like Las Locas to Curaçao.

Although a study of cranial morphology suggested a possible Carib migration from western Venezuela ~1,150 years ago²³, we find no evidence of a new ancestry, as might be expected for such an event. In simulations using *Venezuela_Ceramic*, *LesserAntilles_Ceramic*, or present-day Cariban-speaking Arara as proxies for Caribs, we can detect as little as ~2–8% ancestry from such groups (Supplementary Information section 13). The genetic data shows no evidence for a separate migration, although we cannot rule out migration from an unsampled continental group genetically more similar to Caribbean ceramic people than the proxies we used for simulation, or who contributed less than 2% of their ancestry.

Social structure and population size estimates

We screened 202 individuals from our co-analysis dataset with >400,000 SNPs covered for runs of homozygosity (ROH) >4 centimorgan (cM)²⁴ (Supplementary Data 12; Supplementary Information section 7; Fig. S21). Large sums of long ROH (>20cM) indicate parental relatedness within the last few generations, whereas an abundance of shorter ROH signals background parental relatedness and restricted mating pools²⁵. Only two out of 202 individuals had more than 100cM of their genome in ROH>20cM blocks (~135cM is the average in offspring of first cousins), indicating that close kin unions were rare. In contrast, 48 individuals had at least one ROH>20cM, indicating that many unions took place between individuals as close as second or third cousins, suggesting limited local population sizes.

As further evidence of low population sizes, we detected abundant short and mid-size ROH across the Caribbean. We estimated effective population size (N_e) using the length distribution of all ROH 4–20cM, which arise from co-ancestry mostly within the last ~50 generations (Figs. 3a, 3b). N_e estimates can be used to infer census population size, which in humans is typically three- and up ten-fold greater^{26,27}. N_e for Ceramic-associated Caribbean sites are larger (N_e ~500–1500, similar to previous estimates^{16,20}) than for Archaic-associated sites (N_e ~200–300) (Extended Data Fig. 4a; Extended Data Table 1), pointing to increased population density with the intensification of agriculture. This is also reflected in higher heterozygosity in Ceramic- than Archaic-associated groups (Extended Data Fig. 5).

N_e estimates from the ROH signal represent lower bounds on pan-Caribbean effective population size as they could reflect restricted gene pools for people living just at those sites, rather than interconnected gene pools. We therefore also analyzed long shared segments (IBD blocks) between the X chromosomes of pairs of males (Supplementary Information section 7). Focusing on shared segments of long IBD 12–20cM, which reflect the size of the shared ancestor pool from within the last ~20 generations (Fig. 3a), we find that the rate of such segments decreases with geographic distance (Fig. 3c), as expected if people exchange more genes with people living closer to them. However, we still detect 19 pairs of individuals who share segments of at least 8.7cM across islands (Extended Data Table 2), revealing that people across the Caribbean shared common ancestors in the hundreds of years prior to the time they lived (as expected given a small pan-Caribbean population size).

A comparison between the two major clades in Hispaniola and Puerto Rico gives an estimate of $N_e=3082$ (1530–8150, 95% CI; estimates in Fig. 3 legend). This provides an upper bound for the recent effective size of the joint population living in Hispaniola and Puerto Rico, as limited migration reduces the rate of distant cousins and IBD sharing across sites. Multiplying N_e estimates by three- to ten-fold to obtain census size, we infer that pre-contact population size estimates of hundreds of thousands or even millions for large islands such as Hispaniola⁵ (based on outdated reports or poorly-documented population counts⁶) are too large.

We also identified 57 pairs of closely related individuals (up to third- to fourth-degree relatives; Extended Data Fig. 6; Supplementary Information section 7). Most were within La Caleta (Dominican Republic), where 37 out of 63 individuals studied had one or several close relatives, although the rate was not significantly greater than within other sites (95% CI 1.5%–2.8% for La Caleta versus 1.4%–4.6% for other sites). As further evidence of an interconnected population, we identified male relatives buried ~75 kilometers apart in the southern Dominican Republic: a father/son pair from Atajadizo and their second and third-degree relative from La Caleta.

Pre-contact ancestry persists in the present-day Caribbean

We tested for genetic affinity between the Indigenous ancestry found in present-day²¹ and ancient Caribbean people by computing $f_4(\text{European}, \text{Test}, \text{Cuba_Archaic}, \text{Caribbean_Ceramic})$. We obtained a signal for relatedness between Puerto Ricans and Ceramic-associated individuals ($|Z|=3.4$ and 4.6 for two datasets) (Supplementary Data 14). Our results are consistent with entirely Ceramic-related but not entirely Archaic-related ancestry (Supplementary Information section 14). We carried out the same test separately for 15 provinces of Cuba²⁸ and found two provinces and eight municipalities with weakly significant evidence of Ceramic-related ancestry ($2.0 < |Z| < 3.4$) and only a single municipality (Guines, western Cuba) with marginally significant evidence of Archaic-related ancestry ($Z=2.0$) (Supplementary Data 14). Thus while the available ancient data show the perpetuation of unadmixed Archaic-related ancestry in parts of Cuba into the last millennium, it was substantially replaced by Ceramic-related ancestry prior to the present day.

Previous reports have also found pre-contact Indigenous ancestry in present-day Caribbean people in uniparental haplogroups^{29–32}. We add to this by identifying a previously undocumented deep branch of mitochondrial DNA (mtDNA) haplogroup C1d at a frequency of ~7% across *Caribbean_Ceramic* sub-clades as well as in a modern Puerto Rican individual from the 1000 Genomes Project dataset³³ (Supplementary Data 9; Supplementary Information section 10). This provides direct evidence that Indigenous matrilineal ancestry persisted in the Caribbean since pre-contact times and cannot be explained by colonial-era movements from the American continents.

Discussion

This study addresses multiple debates about the people of the pre-contact Caribbean (Table 1). First, the ancestry present in the Greater Antilles during the Archaic Age was consistent with deriving from a single source, with only subtle differences among Archaic-associated individuals spanning ~2,500 years. We cannot distinguish between a Central or South American origin for the source population of Archaic-associated people, but find a North American origin to be unlikely (though we note that there is a paucity of comparative genetic data from North America).

Second, our data are consistent with a migratory movement accompanying the introduction and spread of intensive ceramic use in the Caribbean³⁴. Ceramic-associated individuals show an affinity to present-day Arawak speakers, consistent with archaeological and linguistic evidence of northeastern South American origin³⁵. In line with hypotheses that Arawak-speaking populations split as they migrated northeast from Amazonian South America, with some groups moving further along the Orinoco and into the Antilles and others toward the western Venezuela coast²⁹, Curaçao individuals have ancestry related to that in *LesserAntilles_Ceramic*. While the earliest ceramic sites in the Caribbean are in Puerto Rico and the northern Lesser Antilles, and there is no archaeological evidence that the Windward Islands of the Lesser Antilles were settled until ~1,800 years ago, the sharing of some ancestry between individuals from Curaçao and those from the Lesser Antilles but not the Greater Antilles supports a south-to-north stepping stone trajectory into the Caribbean⁴.

Third, we find no association between our *Caribbean_Ceramic* sub-clades and the traditional Caribbean ceramic typologies (Saladoid, Ostionoid, Meillacoid, Chicoid), providing no support for a culture-history model that views stylistic transitions as the result of major movements of new people. Instead, the ancestry profile in regions such as the southeastern coast of the Dominican Republic spans more than a millennium across stylistic transitions in material culture. While we cannot rule out that migrations of populations from the Americas genetically similar to Caribbean people drove some of the cultural changes, our findings increase the weight of evidence that connectivity among ceramic using groups within the Caribbean catalyzed stylistic transitions.

Fourth, we provide the first evidence of admixture between Archaic-/Ceramic-related ancestry in three individuals in Hispaniola. This finding also confirms a previous inference⁴ that admixture between people of Archaic- and Ceramic-associated ancestry in the Caribbean was extremely rare (seen here in only three out of 201 ceramic-using Caribbean individuals).

Fifth, we confirm that people living in some parts of the Caribbean (especially Puerto Rico and Cuba) today carry proportions of pre-contact Indigenous ancestry. In Cuba, Archaic-related ancestry persisted nearly until the Contact Period; however, the Indigenous ancestry in Cuba today is mostly not derived from this source. This could reflect post-colonial movement of Indigenous people, although at least some of it likely reflects pre-contact

events as Ceramic-related ancestry was present in individuals from western and central Cuba dated to ~500 calBP.

Sixth, our data provide insights into social structure and demography. Analyzing ROH, we document an avoidance of unions between close relatives during both the Archaic and Ceramic Ages and detect large proportions of cumulative ROH across most of the Caribbean, reflecting a small population size³⁶. We identify male relatives buried ~75 kilometers apart, suggesting networks of connectivity between archaeological sites analyzed today as separate entities. As further evidence of connectivity, we observe shared haplotypes across islands (19 distant cousin pairs) at a rate expected for an effective population size of $N_e=3082$ (95% CI 1530–8150) across the large islands of Hispaniola and Puerto Rico. Although these estimates represent the last ~20 generations since the analyzed individuals lived, they point to a census size across these large islands being substantially less than estimates of hundreds of thousands to millions at contact suggested in some literature^{1,37}. While our population size estimates are lower than those from historical reports and population counts^{5,6}, the devastating impact that European colonization, expropriation, and systematic killing of Indigenous people had on Caribbean populations is indisputable.

The ancestry and legacy of pre-contact Caribbean people persists today, and the study of ancient DNA helps us to better appreciate this. Present-day Caribbean people harbor mixtures of genetic ancestry in different proportions, primarily comprising pre-contact Indigenous populations (~4% on average in Cuba, ~6% in the Dominican Republic, and ~14% in Puerto Rico according to our estimation by *qpAdm*), immigrant Europeans (~70% in Cuba, ~56% in the Dominican Republic, and ~68% in Puerto Rico), and Africans who were brought to this region during the course of the trans-Atlantic slave trade (~26% in Cuba, ~38% in the Dominican Republic, and ~18% in Puerto Rico) (Extended Data Table 3). All three groups contributed in central ways to the present-day people of the Caribbean and continue to shape the legacy of the interconnected Caribbean world.

METHODS

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

Ancient DNA analysis

We generated powder from the skeletal remains of all individuals excavated from sites throughout the Caribbean (see Supplementary Information section 2 for archaeological site information and Figures S1-S11 for maps showing the location of the islands and/or sites studied). Powder was produced from a cochlea^{38,39}, tooth, phalanx, or ossicle⁴⁰ from each individual in a clean room facility at Harvard Medical School (Boston, USA), University College Dublin (Dublin, Ireland), or the University of Vienna (Vienna, Austria); see Supplementary Data 2 for the skeletal element used for each individual and location of powder preparation.

We extracted DNA in dedicated ancient DNA laboratories at Harvard Medical School or the University of Vienna following published protocols^{41–43}. From the extracts, we prepared dual-barcoded double-stranded⁴⁴ or dual-indexed single-stranded libraries⁴⁵, both treated with uracil-DNA glycosylase (UDG) to reduce the rate of characteristic ancient DNA damage⁴⁶. Double-stranded libraries were treated in a modified partial UDG preparation⁴⁴ ('half'), leaving a reduced damage signal at both ends (5' C-to-T, 3' G-to-A). Single-stranded libraries were treated with *E. coli* UDG (USER from NEB) that inefficiently cuts the 5' Uracil and does not cut the 3' Uracil. For a subset of individuals, we increased coverage by preparing multiple libraries; see Supplementary Data 2 for the number of libraries analyzed for each individual.

To generate SNP capture data, we used in-solution target hybridization to enrich for sequences that overlap the mitochondrial genome and ~1.24 million genome-wide SNPs^{47–50} ("1240k"), either in two separate enrichments or simultaneously (Supplementary Data 2). We then added two 7-base-pair indexing barcodes to the adapters of each double-stranded library (single-stranded libraries are already indexed from the library preparation) and sequenced libraries using either an Illumina NextSeq500 instrument with 2×76 cycles or an Illumina HiSeqX10 instrument with 2×101 cycles and reading the indices with 2×7 cycles (double-stranded libraries) or 2×8 cycles (single-stranded libraries).

Prior to alignment, we merged paired-end sequences, retaining reads that exhibited no more than one mismatch between the forward and reverse base if base quality was ≥20, or 3 mismatches if base quality was <20. A custom toolkit was used for merging and trimming adapters and barcodes (available at <https://github.com/DReichLab/ADNA-Tools>). Merged sequences were mapped to the reconstructed human mtDNA consensus sequence (RSRS)⁵¹ and the human reference genome version hg19 using the `samse` command in BWA v.0.7.15-r1140⁵² with the parameters `-n 0.01, -o 2, and -l 16500`. Duplicate molecules (those exhibiting the same mapped start and end position and same stand orientation) were removed after alignment using the Broad Institute's Picard MarkDuplicates tool (available at <http://broadinstitute.github.io/picard/>). We trimmed two terminal bases from UDG-half libraries to reduce damage-induced errors.

We evaluated the authenticity of the isolated DNA by retaining individuals with a minimum of 3% of cytosine-to-thymine substitutions at the end of the sequenced fragments⁴⁴ for double stranded libraries and 10% for single-stranded libraries, point estimates of mitochondrial DNA (mtDNA) contamination below 5% using `contamMix v.1.0–12`⁴⁷, and point estimates of X chromosome contamination (in males) below 3%⁵³; we also used `contamLD`⁵⁴ to confirm low contamination rates (<~6%) (Supplementary Data 2). Eight single-stranded libraries from Ceramic Age individuals did not reach our 10% cytosine-to-thymine substitution threshold but had at least an 8% substitution rate, and therefore assessed as authentic given the relatively recent dates for these individuals; all eight libraries also were within the expected range for the other two authenticity metrics and had <1% contamination as assessed by `contamLD`. Multiple libraries from I10333 and I10334 as well as one library from I12341 showed poor match rates to the mtDNA consensus sequence, but this is likely due to low mtDNA coverage (0.5–2.1×). Two libraries from I7977 and one from I15596 were also slightly below this threshold (6–10% mismatch rate), but also

surpassed thresholds for the other two metrics and had ~1.1% contamination as assessed by contamLD.

We determined SNPs by randomly sampling an overlapping read with minimum mapping quality of 10 and base quality of 20. Individuals with <20,000 covered SNPs were excluded from quantitative analyses. One individual from each of three pairs of first-degree relatives in the dataset was excluded from population genetics analysis; in all cases, we retained the higher coverage individual; see Supplementary Data 1.

We also generated shotgun sequencing data for two Ceramic-associated individuals from The Bahamas, I14922 (Abaco Island) and I14879 (South Andros) using the same system of data generation and processing, although the capture step was not included (Supplementary Data 2). For shotgun data, we report thresholds of mapping quality 30 and base quality 20.

Radiocarbon dates

We report 45 new radiocarbon (^{14}C) dates on bone fragments generated using accelerator mass spectrometry (AMS) (Supplementary Data 3). Most dates ($n=41$) were generated at the Pennsylvania State University (PSU) Radiocarbon Laboratory, and the remainder ($n=4$) were generated at the Center for Isotopic Research on Cultural and Environmental heritage (CIRCE). The sample preparation methodology at PSU was carried out as previously reported²², where bone collagen was extracted and purified using a modified Longin method with ultrafiltration⁵⁵ (>30 kDa gelatin); if collagen yields were low, a modified XAD process⁵⁶ (XAD amino acids) was used. Carbon and nitrogen isotope ratios were then measured (Supplementary Information section 3) as a quality control measure; all C:N ratios fell between 3.15 and 3.44, indicating good collagen or amino acid preservation⁵⁵. We also evaluated diet in these individuals (e.g., marine vs. terrestrial) and compared the results to reference data from 242 ancient Caribbean and Maya individuals (Figures S12-S14). Attenuated Total Reflectance Fourier Transform Infrared (ATR-FTIR) spectra were generated to assess postmortem changes in the apatite crystal structure of the bone samples; ATR-FTIR spectra of all samples are displayed in Figure S15 and quality control parameters are reported in Table S1. Ultimately, all calibrated ^{14}C ages were computed using OxCal v4.4⁵⁷ using the IntCal20⁵⁸ after our stable isotope analysis detected minimal consumption of marine resources. Sample preparation at CIRCE was carried out following the lab-adapted Longin method⁵⁹; isotopic information was not generated for these individuals. Supplementary Data 3 lists the preparation method used for each individual and Supplementary Information section 3 describes the generation of isotopic data in more detail and its use in calibrating the ^{14}C dates generated for the Caribbean individuals.

Dataset assembly

We merged genome-wide data for 93 previously-reported individuals⁴ with newly-generated data from 174 ancient individuals for co-analysis, retaining 89 of them for a final co-analysis dataset comprising 263 individuals (details of merging in Supplementary Information section 4). We leverage these previously published data to revisit statistics and analyses

reported in that work⁴ (Tables S2, S23, S29) and carry out additional analyses using these data (Tables S3, S24, S25, S26, S27, S28, Figures S33, S34).

We merged these 263 ancient individuals that passed screening into a base dataset that included 61 previously published ancient American individuals^{16,20,60–63}, and 36 modern Indigenous American groups sourced from single nucleotide polymorphism (SNP) array genotyping datasets or whole genome sequencing datasets (Supplementary Data 5):

- ‘1240K SNPs’, whole genome sequencing data restricted to a canonical set of 1,233,013 SNPs^{47–50,64,65}
- ‘Human Origins dataset’, 597,573 SNPs^{66–68}
- ‘Illumina dataset’ (unmasked/unadmixed individuals only), 352,432 SNPs¹³

All comparative analyses involving present-day Indigenous American populations were performed on the Illumina dataset, whereas for *qpAdm* and *qpWave*’s set of outgroup populations (“Right”) we used the Human Origins dataset for increased coverage. All genome-wide analyses were performed on autosomal data.

Uniparental haplogroups

We determined mtDNA haplogroups for all individuals using bam files, restricting to reads with MAPQ ≥ 30 and base quality ≥ 20 . We constructed a consensus sequence with samtools and bcftools version 1.3.1 using a majority rule and then determined the haplogroup with HaploGrep2, using Phylotree version 17. We determined Y chromosome haplogroups using sequences mapping to 1240K Y-chromosome targets, restricting to sequences with MAPQ ≥ 30 and base quality ≥ 30 . We called haplogroups by determining the most derived mutation for each individual, using the nomenclature of the International Society of Genetic Genealogy (ISOGG; <http://www.isogg.org>) version 14.76 (April 2019). Mutational differences and corresponding mtDNA haplogroups, and Y chromosome haplogroups and their supporting derived mutations are found in Supplementary Data 9. A discussion of mtDNA and Y chromosome haplogroup distribution in the Caribbean is found in Supplementary Information section 10; see Figures S29 for distribution of mtDNA haplogroups, Figure S30 for details of three mtDNA mutations diagnostic of previously unobserved mtDNA haplogroup which is a variant of C1d, and Figure S31 for distribution of Y chromosome haplogroups.

Kinship

We assessed kinship for every pair of individuals newly-reported here as those that we co-analyze⁴ (including individuals from different sites and islands) using a previously described method⁶⁹, and we present results for 1st-, 2nd-, and 3rd-/4th-degree (‘close’) relatives in Table S5 (Supplementary Information section 7). In our newly-reported dataset of 174 ancient individuals, we identified 49 individuals sharing 49 unique pairwise kin relationships. Three pairs of individuals were identified as 1st-degree relatives, while 21 pairs were 2nd-degree relatives, and 25 pairs were 3rd-degree or higher. For the data that we co-analyze⁴, we identified 13 individuals who were part of eight relationships (four 2nd-degree and four 3rd-degree or higher). No close relatives were identified between the

datasets. Distant cousins detected using IBD analysis are presented elsewhere (Extended Data Table 2; Supplementary Data 13).

Analysis of shared genomic segments

We identified Runs of Homozygosity (ROH) within our ancient dataset using the Python package *hapROH* (<https://test.pypi.org/project/hapsburg/>). Following a previously described method²⁴, we used 5008 global haplotypes from the 1000 Genomes Project haplotype panel³³ as the reference panel. As recommended for datasets with genotypes for 1240K SNPs, we applied our method to ancient individuals with at least 400,000 SNPs covered and ran the method on the pseudo-haploid data to identify ROH longer than 4 centimorgan (cM). We used the default parameters of *hapROH*, which are optimized for ancient data genotyped at 1240K SNPs. For each individual, we group the inferred ROH into four length categories: 4–8cM, 8–12cM, 12–20cM and >20cM and report the total sum in these bins (Supplementary Data 12; Fig. S21).

To estimate effective population size N_e from ROH, we applied a maximum likelihood inference framework (for derivation of the likelihood see Supplementary Information section 7). We fit the lengths of all genome-wide ROH lengths 4–20cM, and infer the effective population size that maximizes the likelihood for ROH lengths observed in a set of individuals. Estimation uncertainties are obtained from the likelihood profile (95% CIs correspond to values within 1.92 units down from the maximum of the log-likelihood function). Tests on simulated data confirmed the ability of our estimator to recover N_e estimates from genome-wide ROH of few individuals (Figs. S22, S23).

We also analyzed shared genomic segments on the X chromosome between pairs of male individuals (“IBD_X”). To call such IBD blocks, we paired pseudo-haploid data of two X chromosomes and ran *hapROH* on read counts of the resulting artificial diploid individual; see Figure S24 for example of IBD segment shared between two individuals. We inferred population sizes from IBD with the same likelihood approach as described for ROH, applying it to all pairs of individuals between two groups of individuals. See Supplementary Information section 7 for details.

Conditional Heterozygosity

We used popstats⁶⁸ to compute conditional heterozygosity for all clades and sub-clades, which we compared with contemporaneous groups from continental South America, such as from the Peruvian Middle and Late Horizon periods⁷⁰. As previously described^{71,72}, we restricted the analysis to transversion SNPs ascertained in a Yoruba individual; see Extended Data Fig. 5.

PCA

We performed principal component analysis (PCA) with smartpca v181623⁷³, using the 1240K + Illumina merged dataset and using the option ‘lsqproject: YES’ to project ancient individuals onto the eigenvectors computed from modern individuals in the version shown in the main manuscript. The approach of projecting each ancient individual onto patterns of variation learned from modern individuals enables us to use data from a large fraction of

SNPs covered in each individual and thereby maximize the information about ancestry that would be lost in approaches that require restriction to a potentially smaller number of SNPs for which there is intersecting data across lower coverage ancient individuals. We used the option ‘newshrink: YES’ to remap the points for the individuals used to generate the PCA onto the positions where they would be expected to fall if they had been projected, thereby allowing the projected and non-projected individuals to be appropriately co-visualized. We projected 92 previously published ancient individuals^{4,16,20} and 174 new ancient individuals onto the first two principal components computed using 61 individuals from 23 present-day populations (Extended Data Fig. 1b). See Supplementary Data 4 for all individuals included in PCA and values of PCs 1 and 2 for the main manuscript PCA. For the PCA presented as Fig. S19 (Supplementary Information section 5), we used non-related, non-outlier ancient individuals from *Cuba_Archaic*, *Venezuela_Ceramic*, *EasternGreaterAntilles_Ceramic*, *BahamasCuba_Ceramic*, and *SECoastDR_Ceramic* with >500K SNPs to compute the eigenvectors and projected all other ancient individuals. We again used the ‘lsqproject: YES’ and ‘newshrink: YES’ options. Individuals used to compute eigenvectors are listed in Supplementary Data 4. For PCA by archaeological site, non-zoomed PCA, PCA excluding CpG sites, and PCA with axes computed using ancient individuals, see Figs. S16-S19.

Unsupervised analysis of population structure

We used the software ADMIXTURE v1.3.0^{74,75} to perform unsupervised structure analysis on a dataset comprised of autosomal SNPs that overlap between the 1240k and Illumina dataset and pruned in PLINK1.9⁷⁶ using --indep-pairwise 200 25 0.4. This left 273,245 SNPs for the analysis. We ran five random-seeded replicates for each K in the interval between 2 and 10 with cross-validation enabled (--cv flag) to identify the runs with the low cross-validation errors (Table S4). For each value of K, we plotted the replicate with the lowest cross-validation error and compared the results. We choose to present K=6 as Extended Data Fig. 1c, as we found that the model with six components had a low cross-validation error and differentiated the components in a useful way for visualization. Results for the other values of K are presented as Fig. S20 in Supplementary Information section 6.

Estimation of F_{ST} coefficients

To measure pairwise genetic differentiation between two groups of individuals, we estimated average pairwise F_{ST} and its standard error via block-jackknife using smartpca v.1.81623 and the options ‘fstonly: YES’ and ‘inbreed: YES.’ We removed the individual with lower coverage of each pair of first degree relatives, as well as ancestry outliers (see main text); we excluded *Haiti_Ceramic*, which comprises only two individuals who share a second-degree relationship as well as Macao, a site in the Dominican Republic from which all four individuals analyzed are 2nd-3rd-degree relatives of at least one other individual from the site. See results in Extended Data Fig. 2.

Clade grouping framework with *qpWave*, *TreeMix* and f_4 -statistics

We used a multi-step framework involving *qpWave*, *TreeMix*, and f_4 -statistics to group sites and individuals, and considered this information together with admixture profiles and proportions from *qpAdm* to produce Fig. 1b (detailed methodology in Supplementary Information section 8). We started by using *qpWave* to identify major clades based on shared

ancestry and then used *TreeMix* and f_{ST} -statistics to investigate the existence of sub-clades. Once all sub-clades were identified, we used f_{ST} -statistics to investigate further substructure between sites within each clade. Geographic and chronological information such as island or cultural affiliation was not considered for these analyses, ensuring all clades and subclades were based solely on genetic information. We examined the association between genetic data and archaeological cultural complexes only after considering the genetic and archaeological information separately, following a previously published example⁷⁷.

The software *qpWave*¹³ from ADMIXTOOLS v6.0⁶⁸ estimates the minimum number of ancestry sources needed to form a group of test populations (“Left”), relative to a set of differentially related reference populations (“Right”). If the “Left” group contains two populations, *qpWave* will evaluate if they can be modelled as descending from the same sources, and hence will determine whether they form a clade. We used 12 present-day Indigenous American populations from the Human Origins dataset⁶⁷ plus Yukpa⁶⁴ representing different language families and ancestries from the American continent as our “Right” reference population set:

Chipewyan, Zapotec, Mixe, Mixtec, Suruí, Cabécar, Piapoco, Karitiana, Yukpa, Quechua, Wayuu, Apalai, Arara

The argument ‘allsnps: NO’ was used, which restricts the analysis SNP set to intersection of all SNPs among all populations and maximizes the reliability of the analysis⁷⁸. The ‘allsnps: YES’ option was developed to increase the number of SNPs analyzed in cases where very little SNP overlap exists between all populations included in a *qpWave* model⁷⁹. While it is commonly used when low coverage data results in the loss of the majority of sites in the initial datasets⁷⁸, there is a risk that this option introduces unreliability in the analysis, particularly in cases where the base population is highly diverged. In this dataset, a high depth of coverage and relatively large sample sizes made it unnecessary for us to use the ‘allsnps: YES’ option. We ran two consecutive steps of *qpWave* analyses, starting with the identification of major groupings (step 1; Figure S25), or clades, and then reassessed the relationships between members within those clades by running the same tests in a “model competition” approach where individuals from other sites from within the same clade were added to the “Right” set (step 2; Figure S26). A significance threshold of $p > 0.01$ was set for accepting a clade between two sites or individuals. The range of covered SNPs was 170,927–827,039, with a median of 672,888.

After identifying the major clades and/or pairs of sites that uniquely formed a clade with one another, we ran *TreeMix* with these clades and 27 previously published present-day Indigenous populations¹³ (Supplementary Data 5) to identify within-clade site structure (step 3; Figures S27, S28) by generating a maximum likelihood tree. We excluded four Chibchan, Chocoan and Arawak-speaking populations possibly admixed with each other from this analysis. We ran *TreeMix*, grouping the SNPs in windows of 500 (flag -k 500) to account for linkage disequilibrium, setting Chipewyan as root (-root), allowing random migration events (-m), and disabling sample size correction (-noss) in order to include sites or populations represented by a single-individual. We note that single-individual populations still present artifactually long branches that do not truly represent population-specific drift. By running *TreeMix* and allowing consecutive random migration/admixture events, we

identified nodes and branches that maintained the same ancient Caribbean sites among the different runs. We then used f_4 -statistics to evaluate if they formed a sub-clade to the exclusion of the other sites by following the tree's structure. For each identified intact node among all *TreeMix* runs we used each downstream pair of site(s) as Test1 and Test2 and investigated their relationship to upstream sites or pools of sites (step 4). If an upstream node was unchanged in all runs, the sites composing it were pooled. However, once the first inconsistency was identified in an upstream node, all sites beyond that node were pooled together. A combination of three statistics per relationship allowed us to evaluate the *TreeMix* structure of the sites being tested:

$$f_4(\text{Mbuti}, \text{Pool}; \text{Test1}; \text{Test2})$$

$$f_4(\text{Mbuti}, \text{Test1}; \text{Pool}, \text{Test2})$$

$$f_4(\text{Mbuti}, \text{Test2}; \text{Test1}, \text{Pool})$$

With Test1 and Test2 expected to be closer to each other than to Pool, the tested relationship finds support if the first test is statistically non-significant and at least one of the other two are significant. We used a Z-score threshold of 2.8 (associated with a 99.5% CI) to assess significance. These sites were then merged into a sub-clade inside the major Ceramic clade for further analysis. We did not include the sites of Cueva del Perico I, Los Indios, Punta Candelerero, and Tibes in the *TreeMix* and f_4 due to reduced coverage, but evaluated these sites separately to see if they shared closer affinities to any sub-clades relative to the others (Supplementary Data 7; Supplementary Information section 8).

After this clading analysis, we used f_4 -statistics to further investigate potential substructure between sites within each sub-clade (step 5). For each pairwise site comparison, we randomly divided each site into two groups of individuals, and used a statistic of the form $f_4(\text{Site1_subset1}, \text{Site2_subset1}; \text{Site1_subset2}, \text{Site2_subset2})$ to identify positive statistics suggesting substructure within the same clade. This randomization step was repeated 10 times, and the average Z-score was calculated. If a site was composed of a single individual we instead computed statistics of the form $f_4(\text{Mbuti}, \text{Site1_subset1}; \text{Site2_singleIndividual}, \text{Site1_subset2})$, intended to evaluate if individuals within Site1 were closer to each other than to the single individual from Site2. No statistics were computed if both sites being tested contained only one individual.

We also used f_4 -statistics to test if any specific sub-clade within the *Caribbean_Ceramic* clade had more Archaic-related ancestry than another. Specifically we used the statistic $f_4(\text{Mbuti}, \text{GreaterAntilles_Archaic}, \text{Sub_Clade1}, \text{Sub_Clade2})$ and interpreted results as significant based on a $|Z| > 2.8$; results are presented in Table S20.

qpAdm—We used *qpAdm*⁴⁹ from ADMIXTOOLS v6.0⁶⁶ with 'allsnps: NO' to identify the most likely sources of ancestry and admixture for our populations/clades. First, we investigated if the possible outliers *SECoastDR_Ceramic16539*, *SECoastDR_Ceramic16520* and *EasternGreaterAntilles_Ceramic7969*, as well as the individuals comprising the sub-clades *LesserAntilles_Ceramic*, *Haiti_Ceramic* and *Curacao_Ceramic*, could be modelled as admixed between the major ancestries represented by *GreaterAntilles_Archaic* (composed of all Archaic-associated individuals Cuba and I10126), *Caribbean_Ceramic* (composed of

BahamasCuba_Ceramic, *EasternGreaterAntilles_Ceramic* and *SECoastDR_Ceramic*, as well as *LesserAntilles_Ceramic* where relevant), and *Venezuela_Ceramic* (see Tables S9, S10, S12-S15). We used this information to complete Fig. 1b. We also used *qpAdm* to evaluate the presence of Archaic-related ancestry in *Caribbean_Ceramic*. Then, based on this admixture information, we attempted to obtain more detailed admixture models using the sub-clades from within *Caribbean_Ceramic* and *GreaterAntilles_Archaic* as possible sources. Lastly, we attempted to identify more distal sources of ancestry by using previously published ancient individuals from the Americas^{60–63}, in this case for *qpWave*'s three major clades/groups. The base “Right” set used was the same used for *qpWave*. We also tested all 1-, 2-, and 3-way models using these “Right” present-day populations as sources by moving them to the “Left” as necessary, and confirmed the results with the same unmasked/unadmixed populations from the Illumina dataset.

qpGraph—We used *qpGraph* and an edited skeleton tree of previously published ancient American populations⁶³ to construct an admixture tree representing the relationships of the new populations analysed in this study along with ref.⁴ and present-day Piapoco, which our other analyses showed to be closely related to *Caribbean_Ceramic* (Fig. 2c). Detailed methodology is provided in Supplementary Information section 12.

Admixture simulations

We investigated the sensitivity of *qpWave* in detecting Carib-related ancestry in the *Caribbean_Ceramic* sub-clades by generating artificially admixed individuals with *Caribbean_Ceramic* ancestry mixed with increasing amounts (1, 2, 5, 8, 10, 20, 30, 40, and 50%) of a plausibly Carib-associated ancestry. For the Carib-associated ancestry we tested Arara (present-day Indigenous Carib speakers), *Venezuela_Ceramic* (inhabitants of a possible region of origin for this ancient Carib migration), and also *LesserAntilles_Ceramic* (possibly representing Island Caribs), and then assessed at what admixture threshold we were able to reliably detect the latter ancestry type (Supplementary Information section 13; Fig. S32). To generate these admixed individuals, we identified common SNPs between the two sources, randomly selected genotypes from the Arara individuals from the Human Origins and Illumina SNP array datasets corresponding to each of the nine percentages to be tested, and added the remaining SNPs from a random individual from *Bahamas_Ceramic*, *EasternGreaterAntilles_Ceramic*, *SECoastDR_Ceramic*, and *LesserAntilles_Ceramic* with over 800,000 SNPs. We then ran *qpWave* with each of the simulated admixed individuals on the “Left” plus their correspondent sub-clade, while using the default 12 “Right” populations (excluding Arara), as described in Supplementary Information section 8, plus the Carib proxy population used to generate those individuals.

Dating admixture

We used the method *DATES* (Distribution of Ancestry Tracts of Evolutionary Signals²² v3520 (Chintalapati, M., Neel, A., Patterson, N. & Moorjani, P. Reconstructing the spatio-temporal patterns of admixture in human history. *In Preparation.*) to estimate the dates of admixture in admixed individuals from Haiti. This method measures the decay of ancestry covariance to infer the time since mixture and estimates jackknife standard errors. Details of

DATES analysis are found in Supplementary Information section 14; results for *Haiti_Ceramic* are found in Table S22.

Relatedness of ancient individuals to present-day admixed Caribbean populations

We computed relative allele-sharing between present-day admixed Caribbean populations (via their Indigenous ancestry) and ancient Archaic-associated versus Ceramic-associated individuals with ADMIXTOOLS 2 (Maier R., Reich D., Patterson N. Rapid inference of demographic history using ADMIXTOOLS 2. *In Preparation.*) through the statistic $f_4(\text{European, Test, Cuba_Archaic, Caribbean_Ceramic})$. In order to evaluate statistical power, we compared results for present-day Cubans alone to results obtained by adding one ancient individual from either the *GreaterAntilles_Archaic* or *Caribbean_Ceramic* clade to the Cuban test population. Full details are found in Supplementary Information section 15.

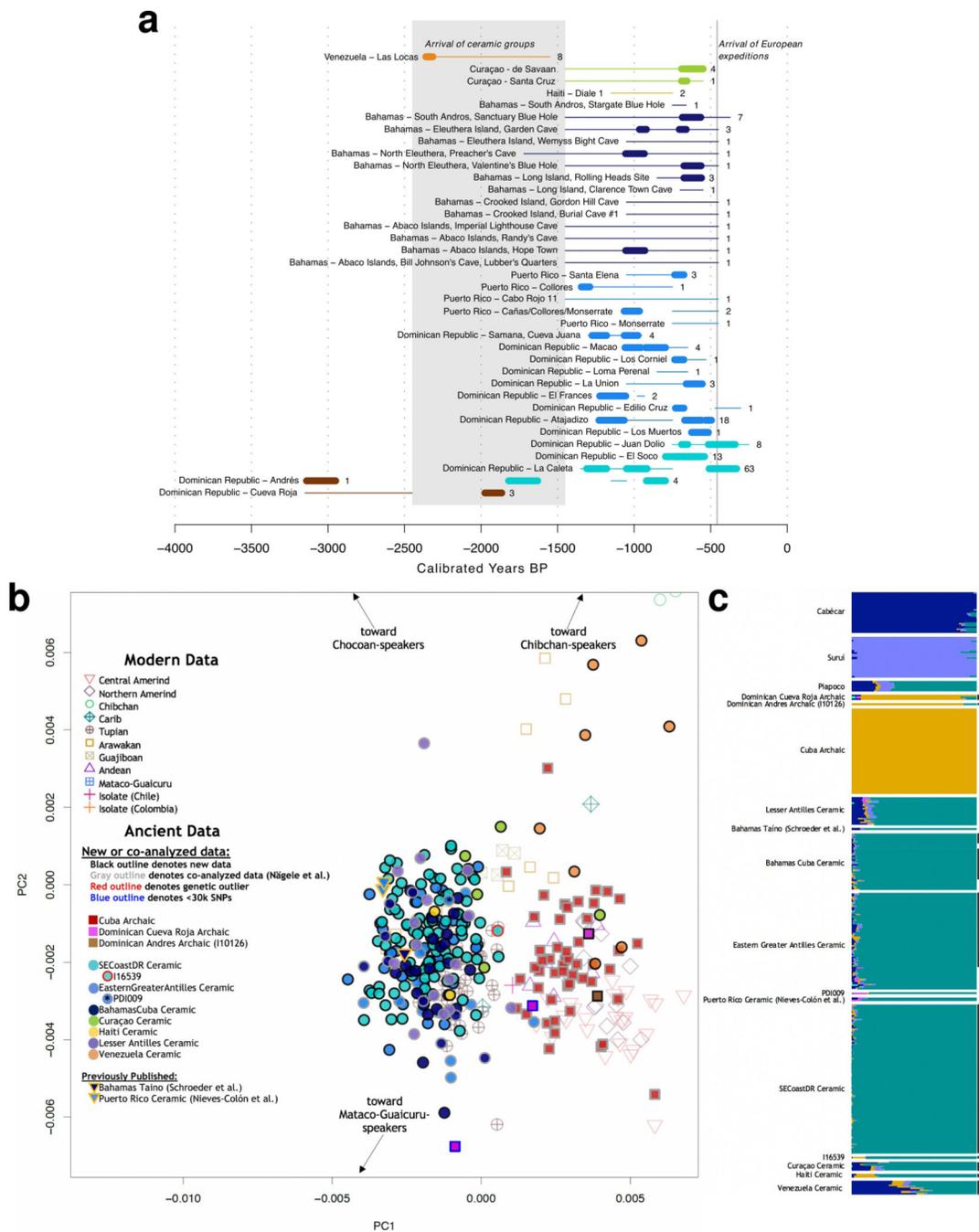
Analysis of phenotypically-relevant SNPs

Analyzing SNPs previously known to be relevant to phenotypic traits allows us to explore their frequencies in the pre-contact Caribbean and Venezuela. We used mpileup in *samtools*⁸⁰ version 1.3.1 with the settings -B -q30 -Q30 to obtain information about each SNP covered by reads from the bam files of our individuals (after trimming 2 base pairs from the molecule ends) and used the fasta file from human genome GRCh37 (hg19) as a reference file for the pileup. We counted the number of reference and alternate alleles, combining counts on the forward and reverse strands. Data are provided in Supplementary Data 15, with a discussion of results in Supplementary Information section 16.

Testing for an Australasian link

We tested for a signal of relatedness to present-day Australasian populations^{64,68} (“Population Y” signal), using the statistic $f_4(\text{Mbuti, Onge/Papuan, Mixe, Archaic/Ceramic})$ and testing all final sub-clades as *Archaic/Ceramic*. Here, Mixe is representative of a population that harbors no Population Y signal. When Onge was used as the Australasian proxy, several of the ancient groups showed weakly positive statistics (Z between 2 and 3), but only the Archaic individual I10126 from the site of Andrés (Dominican Republic) was significant at Z = 3.4. While this signal is significant at p=0.0030 even after performing a Bonferroni correction for the nine hypotheses tested in Extended Data Table 4, the signal is non-significant when Papuan is used as the Australasian proxy (Z=2.2). We also caution that all Population Y statistics are likely to be overinflated in their significance because the original discovery of the Population Y signal carried out extensive hypothesis testing to identify a population in the third position of the statistic $f_4(\text{Mbuti, Onge/Papuan, Mixe, Archaic/Ceramic})$ (Mixe) that maximized the value of the statistic when any other Native American group in was used in the fourth position; thus, there is a further multiple hypothesis testing issue for which our analysis does not correct. The lack of a clear population Y signal is consistent with prior studies that also have not found this signal in ancient individuals from this region¹⁶ and other areas of South America⁶³.

Extended Data



Extended Data Fig. 1: Temporal distribution of newly-reported individuals and overview of population structure.

(a) Numbers represent individuals from each site; thick lines denote direct ¹⁴C dates (95.4% calibrated confidence intervals); thin lines denote archaeological context dating; grey area identifies the first arrivals of ceramic-users in the Caribbean. Colors and labels are consistent with Fig. 1. (b) PCA plot with ancient individuals shown as solid squares or circles (Archaic- or Ceramic-associated individuals, respectively). Newly-reported individuals are

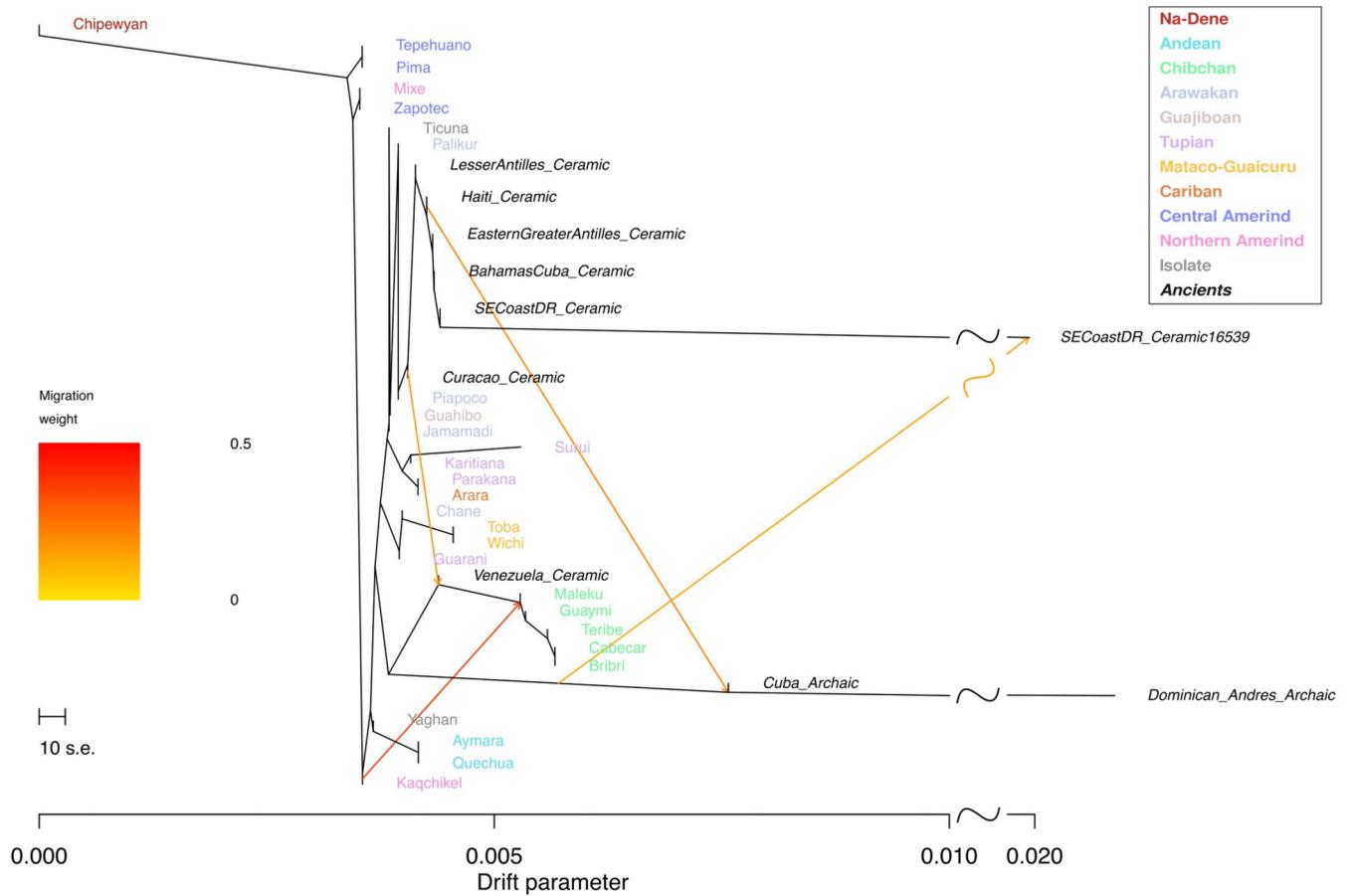
outlined in black, genetic outliers are outlined in red, and individuals with <30,000 SNPs are outlined in blue. Individuals are separated by sub-clades, and three individuals from the site of Cueva Roja (Dominican Republic) who were excluded from clading analysis are labeled “Dominican Cueva Roja Archaic” and colored magenta. Individual PDI009, assessed elsewhere as an outlier¹¹, is denoted with an asterisk. Three previously-published ancient Caribbean individuals^{9,10} are shown as inverted triangles outlined in gray and colored for the sub-clade that encompasses the geographic region with which they are associated. This plot focuses on ancient individuals and does not show some present-day populations; a full plot is provided as Fig. S17. (c) ADMIXTURE analysis best supports K=6 ancestral elements. Newly-reported and co-analyzed individuals are clustered by sub-clade; all newly-reported individuals are identified by a black bar to the side of the plot. The same three previously-published individuals^{9,10} shown in Extended Data Fig. 1b are included, and three modern-day populations are shown for reference (Suruí, Cabécar, Piapoco).

a

	Venezuela_Ceramic n=8	Cuba_Archaic n=50	Curacao_Ceramic n=5	LesserAntilles_Ceramic n=16	BahamasCuba_Ceramic n=33	EasternGreaterAntilles_Ceramic n=52	SECoastDR_Ceramic n=85
Venezuela_Ceramic n=8		12.45 ±0.17	4.58 ±0.17	4.70 ±0.12	5.31 ±0.10	4.81 ±0.10	5.11 ±0.10
Cuba_Archaic n=50	12.45 ±0.17		11.22 ±0.22	10.64 ±0.15	11.34 ±0.15	10.98 ±0.14	11.40 ±0.14
Curacao_Ceramic n=5	4.58 ±0.17	11.22 ±0.22		1.63 ±0.14	2.05 ±0.13	1.63 ±0.13	1.93 ±0.12
LesserAntilles_Ceramic n=16	4.70 ±0.12	10.64 ±0.15	1.63 ±0.14		1.16 ±0.05	0.82 ±0.05	1.18 ±0.05
BahamasCuba_Ceramic n=33	5.31 ±0.10	11.34 ±0.15	2.05 ±0.13	1.16 ±0.05		0.55 ±0.03	0.86 ±0.03
EasternGreaterAntilles_Ceramic n=52	4.81 ±0.10	10.98 ±0.14	1.63 ±0.13	0.82 ±0.05	0.55 ±0.03		0.44 ±0.02
SECoastDR_Ceramic n=85	5.11 ±0.10	11.40 ±0.14	1.93 ±0.12	1.18 ±0.05	0.86 ±0.03	0.44 ±0.02	

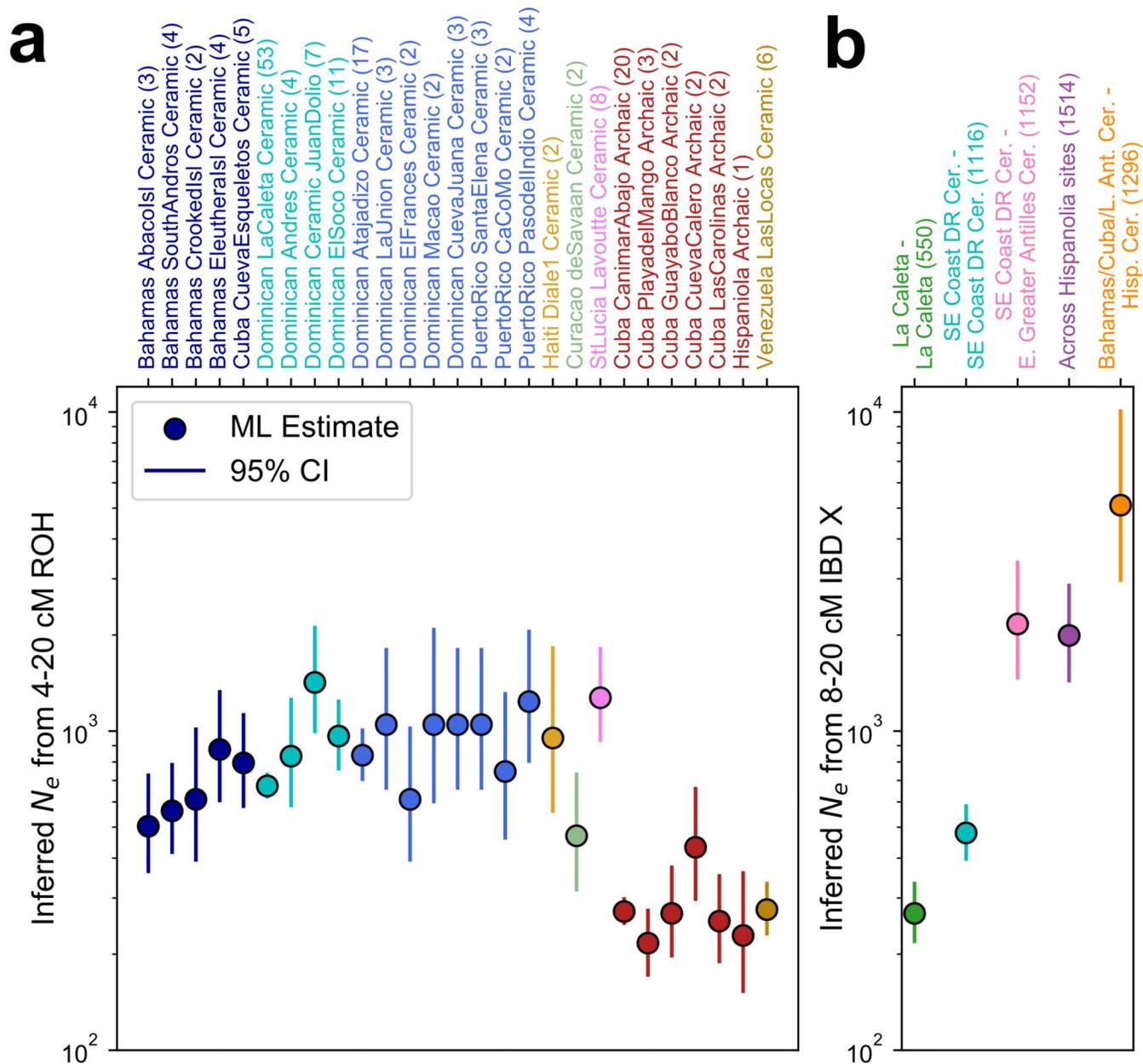
b

Extended Data Fig. 2|. F_{ST} distances.
Average pairwise F_{ST} distances and standard errors (x100) between (a) clades and (b) sites with more than two unrelated individuals, demonstrating both overall high levels of genetic similarity between the *Caribbean_Ceramic* sub-clades and the sites composing them, as well as the magnitude of genetic differentiation between those and the groups with Archaic- and Venezuela-related ancestries.



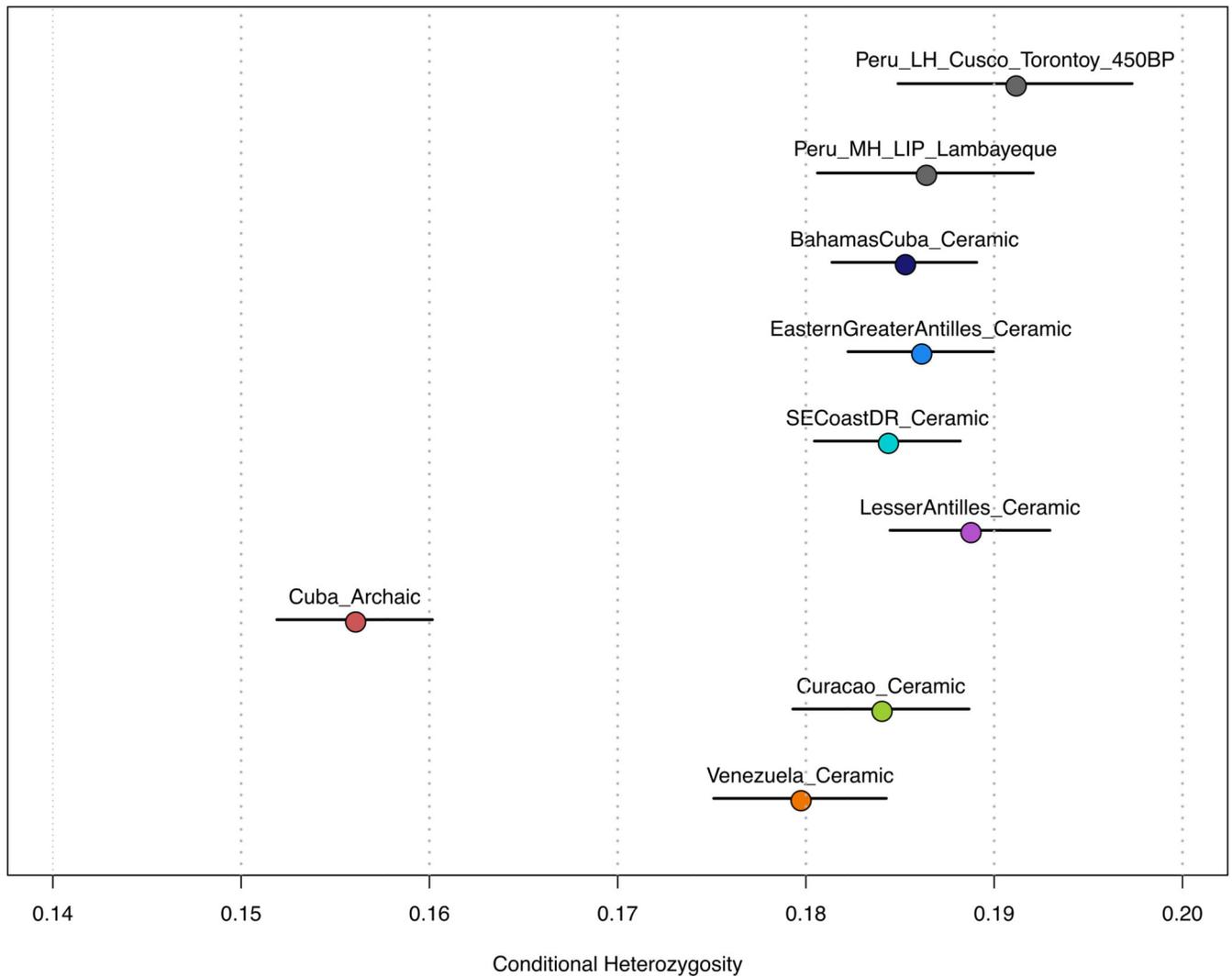
Extended Data Fig. 3: Maximum likelihood population tree from allele frequencies using *Treemix*.

The *Caribbean_Ceramic* sub-clades are shown on the same branch as modern Arawak-speaking groups (Palikur, Jamamadi). Orange arrows represent admixture events, although observations from other analyses (e.g., *qpAdm* admixture modeling) suggest that the indicated direction of admixture may be inaccurate (e.g., we believe it is more likely that there is *GreaterAntilles_Archaic* admixture into *Haiti_Ceramic* than the reverse scenario; Supplementary Information section 9).



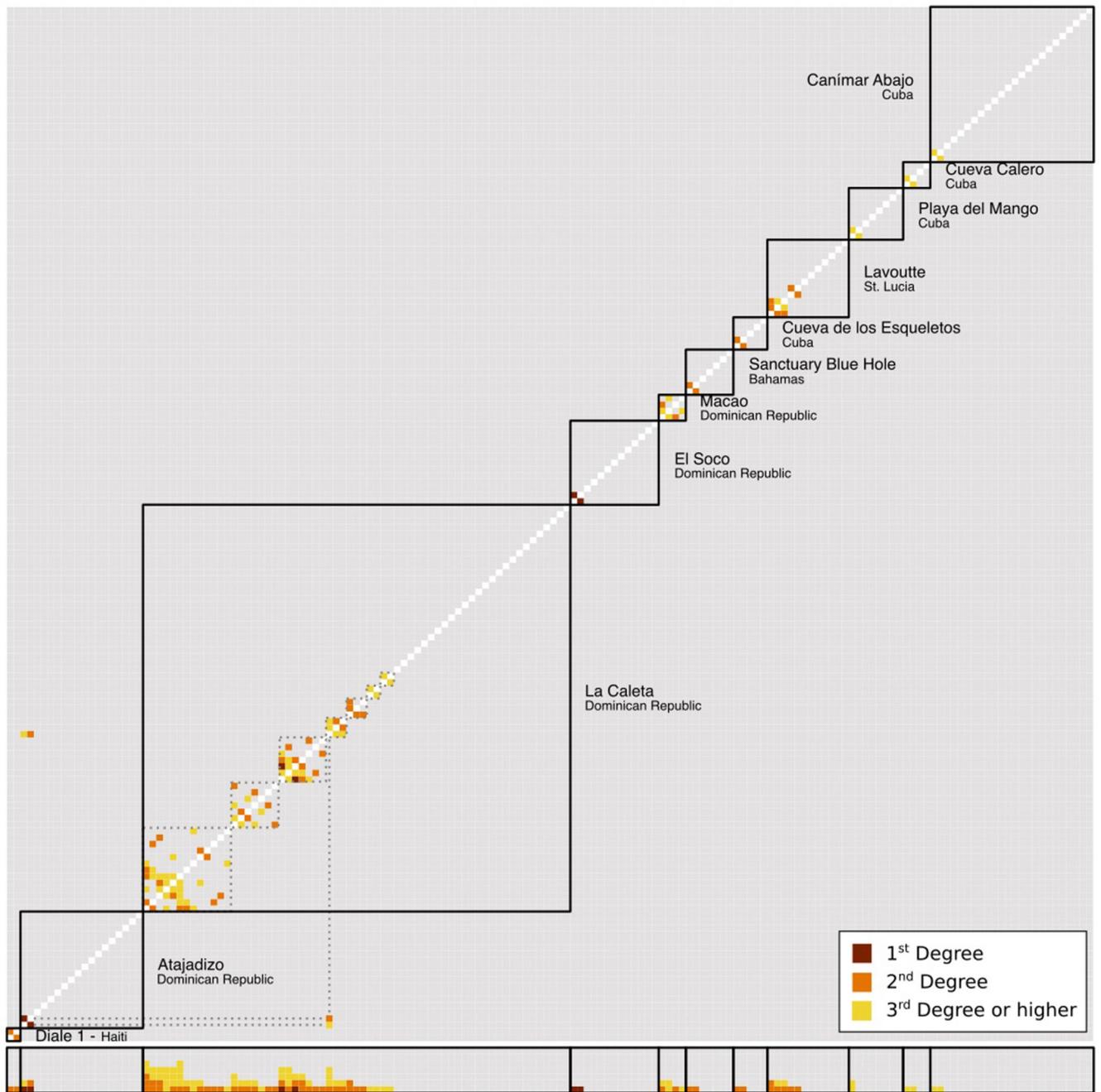
Extended Data Fig. 4: Estimated effective population sizes.

(a) Estimates per site are based on ROH blocks 4–20 cM long using a likelihood model (Supplementary Information section 7). Colors as per sub-clades, numbers denote the count of analyzed individuals. Highly consanguineous individuals with a sum of ROH > 20 above 50 cM were excluded. (b) Same as (a) but for IBD segments 8–20 cM long shared on the X chromosome between all pairs of males. Closely related pairs of individuals with a sum of IBD X > 20 above 25 cM were excluded. Numbers denote counts of all remaining pairs. In (a) and (b) points represent maximum likelihood estimate and vertical bars represent 95% CI.



Extended Data Fig. 5: Conditional heterozygosity by clade.

Conditional heterozygosity in the ancient Caribbean was similar to that of contemporaneous groups from Peru⁷⁰, except for the Archaic-associated groups and *Venezuela_Ceramic*. First- and second-degree relatives were excluded from the analysis, including the pair of related individuals representing *Haiti_Ceramic*. Colored circles represent point estimates (color scheme matching Fig. 1); bars represent three standard errors.



Extended Data Fig. 6: Pairwise kinship estimates for all individuals from sites where close relatives were identified using autosomal data.
 Dotted lines identify family clusters and inter-site relationships; bottom rows correspond to relationships per individual.

Extended Data Table 1:
 N_e estimates for each site.

Table includes all individuals where ROH analysis is possible and excludes individuals with more than 50cM sum of 20cM long ROH.

N_e Estimate	N_e STD	CI(low)	CI(high)	n	Locality	Country	Clade
503	93	321	684	3	Abaco Island	Bahamas	BahamasCuba_Ceramic
562	94	377	747	4	South Andros Island	Bahamas	BahamasCuba_Ceramic
610	151	314	906	2	Crooked Island	Bahamas	BahamasCuba_Ceramic
873	181	519	1228	4	Eleuthera Island	Bahamas	BahamasCuba_Ceramic
793	140	518	1068	5	Cueva de los Esqueletos	Cuba	BahamasCuba_Ceramic
675	34	608	742	53	La Caleta	Dominican Republic	SECoastDR_Ceramic
837	170	504	1170	4	Andres	Dominican Republic	SECoastDR_Ceramic
1416	280	867	1966	7	Juan Dolio	Dominican Republic	SECoastDR_Ceramic
962	126	715	1208	11	El Soco	Dominican Republic	SECoastDR_Ceramic
839	83	677	1002	17	Atajadizo	Dominican Republic	EasternGreaterAntilles_Ceramic
1050	274	512	1588	3	La Union	Dominican Republic	EasternGreaterAntilles_Ceramic
612	151	315	909	2	El Frances	Dominican Republic	EasternGreaterAntilles_Ceramic
1051	336	391	1710	2	Macao	Dominican Republic	EasternGreaterAntilles_Ceramic
1049	274	512	1587	3	Cueva Juana	Dominican Republic	EasternGreaterAntilles_Ceramic
1049	274	512	1587	3	Santa Elena	Puerto Rico	EasternGreaterAntilles_Ceramic
744	202	348	1141	2	Canas/ Collores/ Monserrate	Puerto Rico	EasternGreaterAntilles_Ceramic
1238	303	643	1832	4	Paso del Indo	Puerto Rico	EasternGreaterAntilles_Ceramic
953	291	382	1524	2	Diale 1	Haiti	Haiti_Ceramic
469	103	267	670	2	de Savaan	Curacao	Curacao_Ceramic
1275	224	836	1715	8	Lavoutte	St. Lucia	LesserAntilles_Ceramic
273	15	244	302	20	Canimar Abajo	Cuba	Cuba_Archaic
216	27	162	270	3	Playa del Mango	Cuba	Cuba_Archaic
268	46	178	357	2	Guayabo Blanco	Cuba	Cuba_Archaic
432	91	254	610	2	Cueva Calero	Cuba	Cuba_Archaic

**Extended Data Table 2:
Subset of cross-site relatives from different islands,
identified through IBD analysis.**

We measured the X chromosome length and IBD map lengths as $\frac{2}{3}$ of the map length of female X. Complete table including cross-site distant relatives within islands in Supplementary Data 13.

ID1	ID2	Evidence	Site 1	Site 2
113320	115973	X chromosome IBD segment of 10.0 cM	Bahamas, Abaco Island	Dominican Republic, La Caleta
113318	PDI010	X chromosome IBD segment of 14.0 cM	Bahamas, Crooked Island	Puerto Rico, Vega Baja, Paso delIndio
113321	112344	X chromosome IBD segment of 12.7 cM	Bahamas, Eleuthera Island	Dominican Republic, El Soco
113321	113196	X chromosome IBD segment of 10.7 cM	Bahamas, Eleuthera Island	Dominican Republic, Juan Dolio
113321	113326	X chromosome IBD segment of 12.0 cM	Bahamas, Eleuthera Island	Puerto Rico, Monserrate
113737	CDE001	X chromosome IBD segment of 10.7 cM	Bahamas, Long Island, Clarence Town, Rolling Heads Site	Cuba, Camaguey, Sierra de Cubitas, Cueva de los Esqueletos 1
114880	112344	X chromosome IBD segment of 8.7 cM	Bahamas, South Andros, SanctuaryBlue Hole	Dominican Republic, El Soco
114879	115963	X chromosome IBD segment of 10.0 cM	Bahamas, South Andros, SanctuaryBlue Hole	Dominican Republic, La Caleta
18549	114879	X chromosome IBD segment of 10.0 cM	Dominican Republic, Andres	Bahamas, South Andros, SanctuaryBlue Hole
117903	114875	X chromosome IBD segment of 14.7 cM	Dominican Republic, Atajadizo	Bahamas, Abaco, Bill Johnson's Cave, Lubber's Quarters
113441	114880	X chromosome IBD segment of 10.7 cM	Puerto Rico, Cabo Rojo 11	Bahamas, South Andros, SanctuaryBlue Hole
113441	113189	X chromosome IBD segment of 10.0 cM	Puerto Rico, Cabo Rojo 11	Dominican Republic, El Soco
113441	115676	X chromosome IBD segment of 10.0 cM	Puerto Rico, Cabo Rojo 11	Dominican Republic, La Caleta
113441	114992	X chromosome IBD segment of 9.3 cM	Puerto Rico, Cabo Rojo 11	Dominican Republic, Los Muertos
113326	112344	X chromosome IBD segment of 11.3 cM	Puerto Rico, Monserrate	Dominican Republic, El Soco
PDI012013	115963	X chromosome IBD segment of 9.3 cM	Puerto Rico, Vega Baja, Paso delIndio	Dominican Republic, La Caleta
113318	114880	X chromosome IBD segment of 22.7 cM	Bahamas, Crooked Island	Bahamas, South Andros, SanctuaryBlue Hole
113318	114879	X chromosome IBD segment of 10.0 cM	Bahamas, Crooked Island	Bahamas, South Andros, SanctuaryBlue Hole
113321	113320	X chromosome IBD segment of 12.0 cM	Bahamas, Eleuthera Island	Bahamas, Abaco

Extended Data Table 3:
Ancestry proportion estimates with *qpAdm* in present-day Caribbean individuals from Cuba (and its provinces), Dominican Republic, and Puerto Rico^{21,28}.

Top half, proportions across countries.

Country	Caribbean_Ceramic		1000 Genomes CEU		1000 Genomes YRI	
	Proportion	SE	Proportion	SE	Proportion	SE
Cuba(SGDP)	0.029	0.002	0.722	0.004	0.249	0.002
Cuba(1000G1)	0.042	0.002	0.703	0.002	0.255	0.001
Dominican Republic (SGDP)	0.058	0.003	0.558	0.006	0.384	0.004
Dominican Republic (1000G1)	0.062	0.002	0.558	0.004	0.379	0.003
Puerto Rico (SGDP)	0.132	0.004	0.686	0.006	0.182	0.003
Puerto Rico (1000G1)	0.140	0.003	0.676	0.003	0.184	0.002

Cuban Province	Caribbean_Ceramic		1000 Genomes CEU		1000 Genomes YRI		1000 Genomes CHB	
	Proportion	SE	Proportion	SE	Proportion	SE	Proportion	SE
Artemisa(1000G2)	0.038	0.004	0.834	0.005	0.100	0.003	0.028	0.004
Camaguey(1000G2)	0.074	0.003	0.616	0.004	0.297	0.002	0.013	0.003
Ciego_de_Avila (1000G2)	0.057	0.003	0.788	0.004	0.145	0.002	0.010	0.003
Cienfuegos(1000G2)	0.028	0.003	0.740	0.004	0.220	0.003	0.012	0.003
Granma(1000G2)	0.145	0.003	0.567	0.003	0.271	0.002	0.018	0.002
Guantanamo(1000G2)	0.083	0.002	0.549	0.003	0.363	0.003	0.004	0.002
Holguin(1000G2)	0.095	0.002	0.655	0.003	0.237	0.002	0.013	0.002
La_Habana (1000G2)	0.033	0.002	0.694	0.003	0.257	0.002	0.015	0.002
Las_Tunas (1000G2)	0.113	0.005	0.725	0.007	0.161	0.004	0.001	0.005
Matanzas(1000G2)	0.016	0.003	0.818	0.003	0.140	0.002	0.026	0.003
Mayabeque(1000G2)	0.012	0.004	0.889	0.005	0.094	0.003	0.005	0.004
Pinar_del_Rio (1000G2)	0.036	0.002	0.727	0.003	0.227	0.002	0.010	0.002
Sancti_Spiritus (1000G2)	0.065	0.003	0.809	0.003	0.108	0.002	0.018	0.003
Santiago_de_Cuba (1000G2)	0.076	0.002	0.501	0.003	0.417	0.002	0.006	0.002
Villa_Clara (1000G2)	0.066	0.002	0.812	0.003	0.106	0.002	0.016	0.002

CEU = European source; YRI = African source; CHB = East Asian source; SGDP = Simons Genome Diversity Project outgroup populations Karitiana, Mixe, Yakut, Ulchi, Papuan, Mursi, and Mbuti; 1000G1 = 1000 Genomes outgroup populations PEL, PJL, JPT, and MSL. Bottom half, proportions across different Cuban provinces. 1000G2 = 1000 Genomes outgroup populations PEL, PJL, JPT, MSL and GIH.

Extended Data Table 4:

Statistics testing for an Australasian link.

Test	$f_d(\text{Mbuti, Onge; Mixe, Test})$	Z-score	SNPs used
Cuba_Archaic	0.000606	2.330	1115829
Dominican_Andres_Archaic	0.001291	3.380	741742
BahamasCuba_Ceramic	0.000590	2.497	1104937
EasternGreaterAntilles_Ceramic	0.000528	2.358	1110135
SECoastDR_Ceramic	0.000548	2.420	1112602
Haiti_Ceramic	0.000720	2.102	1015357
Curacao_Ceramic	0.000595	2.180	984268
LesserAntilles_Ceramic	0.000490	2.098	1096317
Venezuela_Ceramic	0.000633	2.447	957964

Test	$f_d(\text{Mbuti, Papuan; Mixe, Test})$	Z-score	SNPs used
Cuba_Archaic	0.000325	1.315	1116502
Dominican_Andres_Archaic	0.000696	1.853	742248
BahamasCuba_Ceramic	0.000383	1.806	1105601
EasternGreaterAntilles_Ceramic	0.000445	2.192	1110808
SECoastDR_Ceramic	0.000401	1.950	1113277
Haiti_Ceramic	0.000377	1.243	1015971
Curacao_Ceramic	0.000399	1.573	984884
Lesser_Antilles_Ceramic	0.000338	1.599	1096963
Venezuela_Ceramic	0.000225	0.923	958591

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Daniel M. Fernandes^{*,1,2}, Kendra A. Sirak^{*,3,4}, Harald Ringbauer^{3,4}, Jakob Sedig^{3,4}, Nadin Rohland^{3,5}, Olivia Cheronet¹, Matthew Mah^{3,4,5,6}, Swapan Mallick^{3,4,5,6}, Iñigo Olalde^{3,7}, Brendan J. Culleton⁸, Nicole Adamski^{3,6}, Rebecca Bernardos^{3,6}, Guillermo Bravo^{1,9}, Nasreen Broomandkhoshbacht^{3,6,31}, Kimberly Callan^{3,6}, Francesca Candilio¹⁰, Lea Demetz¹, Kellie Sara Duffett Carlson¹, Laurie Eccles¹¹, Suzanne Freilich¹, Richard J. George¹², Ann Marie Lawson^{3,6}, Kirsten Mandl¹, Fabio Marzaioli¹³, Weston C. McCool¹², Jonas Oppenheimer^{3,6,32}, Kadir T. Özdoğan¹, Constanze Schattke¹, Ryan Schmidt¹⁴, Kristin Stewardson^{3,6}, Filippo Terrasi¹³, Fatma Zalzal^{3,6}, Carlos Arredondo Antúnez¹⁵, Ercilio Vento Canosa¹⁶, Roger Colten¹⁷, Andrea Cucina¹⁸, Francesco Genchi¹⁹, Claudia Kraan²⁰, Francesco La Pastina¹⁹, Michaela Lucci²¹, Marcio Veloz Maggiolo²², Beatriz Marcheco-Teruel²³, Clenis Tavarez Maria²⁴, Christian Martínez²⁴, Ingeborg París²⁵, Michael Pateman^{26,27}, Tanya M. Simms²⁸, Carlos Garcia Sivoli²⁵, Miguel Vilar²⁹,

Douglas J. Kennett¹², William F. Keegan^{30,33}, Alfredo Coppa^{1,3,19,33}, Mark Lipson^{3,4,33}, Ron Pinhasi^{1,33}, David Reich^{3,4,5,6,33}

Affiliations

¹Department of Evolutionary Anthropology, University of Vienna, 1090 Vienna, Austria. ²CIAS, Department of Life Sciences, University of Coimbra, 3000-456 Coimbra, Portugal. ³Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. ⁴Department of Human Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA. ⁵Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA. ⁶Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA. ⁷Institute of Evolutionary Biology, CSIC-Universitat Pompeu Fabra, 08003 Barcelona, Spain. ⁸Institutes of Energy and the Environment, The Pennsylvania State University, University Park, PA 16802, USA. ⁹Department of Legal Medicine, Toxicology and Physical Anthropology, University of Granada, Granada, Spain. ¹⁰Superintendency of Archaeology, Fine Arts and Landscape for the city of Cagliari and the provinces of Oristano and South Sardinia, Cagliari, Italy. ¹¹Department of Anthropology, The Pennsylvania State University, University Park, PA, 16802, USA. ¹²Department of Anthropology, University of California, Santa Barbara, CA 93106, USA. ¹³Department of Mathematics and Physics, Campania University “Luigi Vanvitelli”, 81100 Caserta, Italy. ¹⁴CIBIO - InBIO, University of Porto, Campus de Vairão, 4485-661 Vairão, Portugal. ¹⁵Museo Antropológico Montané, University of Havana, Havana, Cuba. ¹⁶Matanzas University of Medical Sciences, Carretera de Quintanilla Km 101, 40100 Matanzas, Cuba. ¹⁷Peabody Museum of Natural History, Yale University, New Haven, CT 06511, USA. ¹⁸Facultad de Ciencias Antropológicas, Universidad Autónoma de Yucatán, Mérida, Yucatán, Mexico. ¹⁹Department of Environmental Biology, Sapienza University of Rome, Square Aldo Moro, 5, Rome 00185, Italy. ²⁰National Archaeological-Anthropological Memory Management (NAAM), 13 Johan van Walbeeckplein, Willemstad, Curaçao. ²¹DANTE Laboratory of Diet and Ancient Technology, Sapienza University of Rome, Rome, Italy. ²²Universidad Autónoma de Santo Domingo, Sanchez, San Francisco de Macorís 31000, Dominican Republic. ²³National Center of Medical Genetics, Medical University of Havana, Havana, Cuba. ²⁴Museo del Hombre Dominicano, Plaza de la Cultura Juan Pablo Duarte, Av. Pedro Henríquez Ureña, Santo Domingo 10204, Dominican Republic. ²⁵Instituto de Investigaciones Bioantropológicas y Arqueológicas, Universidad de Los Andes, Mérida, Venezuela. ²⁶Turks and Caicos National Museum Foundation, Front St, Cockburn Town TKCA 1ZZ, Turks and Caicos Islands. ²⁷AEX Bahamas Maritime Museum, Port Lucaya Marketplace, Seahorse Road, Bell Channel Way, Freeport, Grand Bahama, The Bahamas. ²⁸Department of Biology, University of The Bahamas, P.O. Box N-4912, Nassau, The Bahamas. ²⁹National Geographic Society, Washington, DC 20036, USA. ³⁰Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA. ³¹Present address: Department of Anthropology, University of California, Santa Cruz, CA 95064, USA. ³²Present address: Department of Biomolecular Engineering, University of California, Santa Cruz, CA 95064, USA. ³³These authors

jointly supervised this work: William F. Keegan, Alfredo Coppa, Mark Lipson, Ron Pinhasi, David Reich.

Acknowledgements

We acknowledge the ancient people who were the source of the skeletal material analyzed in this study as well as modern people from the Caribbean who have a genetic or cultural legacy from some of the ancient populations we analyzed. This work was supported by a grant from the National Geographic Society to Michael Pateman to facilitate analysis of skeletal material from The Bahamas. D.R. was funded by NSF HOMINID grant BCS-1032255, NIH (NIGMS) grant GM100233, the Paul Allen Foundation, the John Templeton Foundation grant 61220, and the Howard Hughes Medical Institute. We thank Juan Avilés, Juan Acayaguana Delvalle, Jorge Estevez, Dianne T. Golding Frankson, Jenna Gregory, Lynne A. Guitar, Lisa Kelly, Gerald Alexander Lopez Castellano, Kalaan Robert Nibonri, and Orlando Patterson for comments on early versions of this manuscript and discussions that improved the presentation of this work. We thank Vanessa A. Forbes-Pateman and Nancy Albury for their assistance compiling descriptions for archaeological sites in The Bahamas; Eadaoin Harney, Robert Maier, and Nathan Nakatsuka for help with data processing; and Manjusha Chintalapati, Priya Moorjani, and Nick Patterson for advice on analysis. We dedicate this article to the memory of Fernando Luna Calderon, who would have been a co-author had he not passed away in the course of the work for this study.

References

1. Rouse I. *The Tainos: Rise & Decline of the People who Greeted Columbus*. (Yale University Press, 1992).
2. Maggiolo MV *La isla de Santo Domingo antes de Colón* (Banco Central de la Republica Dominicana, 1993).
3. Keegan WF & Hofman CL *The Caribbean before Columbus*. (Oxford University Press, 2017).
4. Nägele K. et al. Genomic insights into the early peopling of the Caribbean. *Science* 369, 456–460 (2020). [PubMed: 32499399]
5. Cook SF & Borah W. *The Aboriginal Population of Hispaniola*. vol. 1 376–410 (University of California Press, 1971).
6. Henige D. On the Contact Population of Hispaniola: History as Higher Mathematics. *Hispanic American Historical Review* 58, 217–237 (1978).
7. Wilson SM *The Archaeology of the Caribbean*. (Cambridge University Press, 2007).
8. Rodríguez Ramos R. Isthmo–Antillean Engagements. in *Oxford Handbook of Caribbean Archaeology* (eds. Keegan WF, Hofman CL & Rodríguez Ramos R.) 155–170 (Oxford University Press, 2013).
9. Bérard B. About boxes and labels: A periodization of the Amerindian occupation of the West Indies. *Journal of Caribbean Archaeology* 19, 51–67 (2019).
10. Callaghan RT *Archaeological Views of Caribbean Seafaring in Oxford handbook of Caribbean archaeology* (eds. Keegan WF, Hofman C. & Rodriguez RR) 285–295 (Oxford University Press, 2013).
11. Siegel PE et al. Paleoenvironmental evidence for first human colonization of the eastern Caribbean. *Quaternary Science Reviews* 129, 275–295 (2015).
12. Oliver JR *The archaeological, linguistic and ethnohistorical evidence for the expansion of Arawakan into northwestern Venezuela and northeastern Colombia*. (University of Illinois at Urbana-Champaign, 1989).
13. Reich D. et al. Reconstructing Native American population history. *Nature* 488, 370–374 (2012). [PubMed: 22801491]
14. Greenberg JH *Language in the Americas*. (Stanford University Press, 1987).
15. Salzano FM, Hutz MH, Salomoni SP, Rohr P. & Callegari-Jacques SM Genetic Support for Proposed Patterns of Relationship among Lowland South American Languages. *Current Anthropology* 46, S121–S128 (2005).

16. Schroeder H. et al. Origins and genetic legacies of the Caribbean Taino. *Proc. Natl. Acad. Sci. U. S. A* 115, 2341–2346 (2018). [PubMed: 29463742]
17. Pickrell JK & Pritchard JK Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 8, e1002967 (2012).
18. Chinique de Armas Y., Roksandic M., Suárez RR, Smith DG & Buhay WM Isotopic Evidence of Variations in Subsistence Strategies and Food Consumption Patterns among ‘Fisher-Gatherer’ Populations of Western Cuba in Cuban Archaeology in the Circum-Caribbean Context (ed. Roksandic I.) (University Press of Florida, 2016).
19. Lovén SE Origins of the Tainan Culture, West Indies. (Elanders Bokfryckeri Akfiebolog, 1935).
20. Nieves-Colón MA et al. Ancient DNA reconstructs the genetic legacies of pre-contact Puerto Rico communities. *Molecular Biology and Evolution* 37, 611–626 (2020). [PubMed: 31710665]
21. Moreno-Estrada A. et al. Reconstructing the population genetic history of the Caribbean. *PLoS Genet.* 9, e1003925 (2013).
22. Narasimhan VM et al. The formation of human populations in South and Central Asia. *Science* 365, (2019).
23. Ross AH, Keegan WF, Pateman MP & Young CB Faces Divulge the Origins of Caribbean Prehistoric Inhabitants. *Sci. Rep* 10, 147 (2020). [PubMed: 31924849]
24. Ringbauer H., Novembre J. & Steinrücken M. Detecting runs of homozygosity from low-coverage ancient DNA. [bioRxiv.org](https://doi.org/10.1101/2020.05.14.200987) (2020).
25. Ceballos FC, Joshi PK, Clark DW, Ramsay M. & Wilson JF Runs of homozygosity: windows into population history and trait architecture. *Nat. Rev. Genet* 19, 220–234 (2018). [PubMed: 29335644]
26. Frankham R. Effective population size/adult population size ratios in wildlife: a review. *Genet. Res* 89, 491–503 (2007). [PubMed: 18976539]
27. Browning SR & Browning BL Accurate Non-parametric Estimation of Recent Effective Population Size from Segments of Identity by Descent. *Am. J. Hum. Genet* 97, 404–418 (2015). [PubMed: 26299365]
28. Fortes-Lima C. et al. Exploring Cuba’s population structure and demographic history using genome-wide data. *Sci. Rep* 8, 11422 (2018).
29. Toro-Labrador G., Wever OR & Martínez-Cruzado JC Mitochondrial DNA Analysis in Aruba: Strong Maternal Ancestry of Closely Related Amerindians and Implications for the Peopling of Northwestern Venezuela. *Caribbean Journal of Science* 39, (2003).
30. Mendizabal I. et al. Genetic origin, admixture, and asymmetry in maternal and paternal human lineages in Cuba. *BMC Evol. Biol* 8, 213 (2008). [PubMed: 18644108]
31. Vilar MG et al. Genetic diversity in Puerto Rico and its implications for the peopling of the Island and the West Indies. *Am. J. Phys. Anthropol* 155, 352–368 (2014). [PubMed: 25043798]
32. Benn Torres J. et al. Genetic Diversity in the Lesser Antilles and Its Implications for the Settlement of the Caribbean Basin. *PLoS One* 10, e0139192 (2015).
33. Consortium T. 1000 G. P. & The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* 526, 68–74 (2015). [PubMed: 26432245]
34. Hofman CL & Reid BA The Saladoid. in *Encyclopedia of Caribbean archaeology* (eds. Reid B. & Gilmore G.) 300–303 (University of Florida Press, 2014).
35. Roksandic I. & Roksandic M. *Peopling of the Caribbean. 199–223* (Kerns Verlag, 2018).
36. Keegan W. *The People Who Discovered Columbus.* (University Press of Florida, 1992).
37. Anderson-Córdova KF Hispaniola and Puerto Rico: Indian Acculturation and Heterogeneity, 1492–1550. (University Microfilms International, 1990).
38. Pinhasi R. et al. Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous Bone. *PLoS One* 10, e0129102 (2015).
39. Pinhasi R., Fernandes DM, Sirak K. & Cheronet O. Isolating the human cochlea to generate bone powder for ancient DNA analysis. *Nat. Protoc* 14, 1194–1205 (2019). [PubMed: 30842617]
40. Sirak K. et al. Human auditory ossicles as an alternative optimal source of ancient DNA. *Genome Res.* 30, 427–436 (2020). [PubMed: 32098773]

41. Dabney J. et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U. S. A* 110, 15758–15763 (2013).
42. Korlevi P. et al. Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* 59, 87–93 (2015). [PubMed: 26260087]
43. Rohland N., Glocke I., Aximu-Petri A. & Meyer M. Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. *Nat. Protoc* 13, 2447–2461 (2018). [PubMed: 30323185]
44. Rohland N., Harney E., Mallick S., Nordenfelt S. & Reich D. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. Lond. B Biol. Sci* 370, 20130624 (2015).
45. Gansauge M-T, Aximu-Petri M., Nagel K. & Meyer M. Manual and automated preparation of single-stranded DNA libraries for the sequencing of DNA from ancient biological remains and other sources of highly degraded DNA. *Nature Protocols* 15, 2279–3000 (2020). [PubMed: 32612278]
46. Briggs AW et al. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res.* 38, e87 (2010). [PubMed: 20028723]
47. Fu Q. et al. A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol* 23, 553–559 (2013). [PubMed: 23523248]
48. Fu Q. et al. An early modern human from Romania with a recent Neanderthal ancestor. *Nature* 524, 216–219 (2015). [PubMed: 26098372]
49. Haak W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522, 207–211 (2015). [PubMed: 25731166]
50. Mathieson I. et al. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* 528, 499–503 (2015). [PubMed: 26595274]
51. Behar DM et al. A ‘Copernican’ reassessment of the human mitochondrial DNA tree from its root. *Am. J. Hum. Genet* 90, 675–684 (2012). [PubMed: 22482806]
52. Li H. & Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589–595 (2010). [PubMed: 20080505]
53. Korneliusen TS, Albrechtsen A. & Nielsen R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* 15, 356 (2014). [PubMed: 25420514]
54. Nakatsuka N. et al. ContamLD: estimation of ancient nuclear DNA contamination using breakdown of linkage disequilibrium. *Genome Biol.* 21, 199 (2020). [PubMed: 32778142]
55. Kennett DJ et al. Archaeogenomic evidence reveals prehistoric matrilineal dynasty. *Nat. Commun* 8, 14115 (2017).
56. Lohse JC, Culleton BJ, Black SL & Kennett DJ A Precise Chronology of Middle to Late Holocene Bison Exploitation in the Far Southern Great Plains. *Journal of Texas Archeology and History* 1, 94–126 (2014).
57. Ramsey CB Bayesian Analysis of Radiocarbon Dates. *Radiocarbon* 51, 337–360 (2009).
58. Reimer PJ et al. The IntCal20 Northern Hemisphere Radiocarbon Age Calibration Curve (0–55 cal kBP). *Radiocarbon* 62, 725–757 (2020).
59. Passariello I. et al. Characterization of Different Chemical Procedures for ¹⁴C Dating of Buried, Cremated, and Modern Bone Samples at Circe. *Radiocarbon* 54, 867–877 (2012).
60. Lindo J. et al. The genetic prehistory of the Andean highlands 7000 years BP though European contact. *Sci Adv* 4, eaau4921 (2018).
61. Moreno-Mayar JV et al. Early human dispersals within the Americas. *Science* 362, (2018).
62. Scheib CL et al. Ancient human parallel lineages within North America contributed to a coastal expansion. *Science* 360, 1024–1027 (2018). [PubMed: 29853687]
63. Posth C. et al. Reconstructing the Deep Population History of Central and South America. *Cell* 175, 1185–1197.e22 (2018). [PubMed: 30415837]
64. Raghavan M. et al. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* 349, aab3884 (2015).

65. Mallick S. et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* 538, 201–206 (2016). [PubMed: 27654912]
66. Patterson N. et al. Ancient admixture in human history. *Genetics* 192, 1065–1093 (2012). [PubMed: 22960212]
67. Lazaridis I. et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 513, 409–413 (2014). [PubMed: 25230663]
68. Skoglund P. et al. Genetic evidence for two founding populations of the Americas. *Nature* 525, 104–108 (2015). [PubMed: 26196601]
69. Olalde I. et al. The genomic history of the Iberian Peninsula over the past 8000 years. *Science* 363, 1230–1234 (2019). [PubMed: 30872528]
70. Nakatsuka N. et al. A Paleogenomic Reconstruction of the Deep Population History of the Andes. *Cell* (2020) doi:10.1016/j.cell.2020.04.015.
71. Skoglund P. et al. Genomic insights into the peopling of the Southwest Pacific. *Nature* 538, 510–513 (2016). [PubMed: 27698418]
72. Harney É et al. Ancient DNA from Chalcolithic Israel reveals the role of population mixture in cultural transformation. *Nat. Commun* 9, 3336 (2018). [PubMed: 30127404]
73. Patterson N., Price AL & Reich D. Population structure and eigenanalysis. *PLoS Genet.* 2, e190 (2006).
74. Alexander DH, Novembre J. & Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664 (2009). [PubMed: 19648217]
75. Alexander DH & Lange K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* 12, (2011).
76. Chang CC et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7 (2015). [PubMed: 25722852]
77. Fu Q. et al. The genetic history of Ice Age Europe. *Nature* 534, 200–205 (2016). [PubMed: 27135931]
78. Lipson M. Applying f4-Statistics and Admixture Graphs: Theory and Examples. *Mol Ecol Resour* 00, 1–10 (2020).
79. Harney É, Patterson N., Reich D. & Wakeley J. Assessing the Performance of qpAdm: A Statistical Tool for Studying Population Admixture. *bioRxiv* (2020) doi:10.1101/2020.04.09.032664.
80. Li H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079 (2009). [PubMed: 19505943]

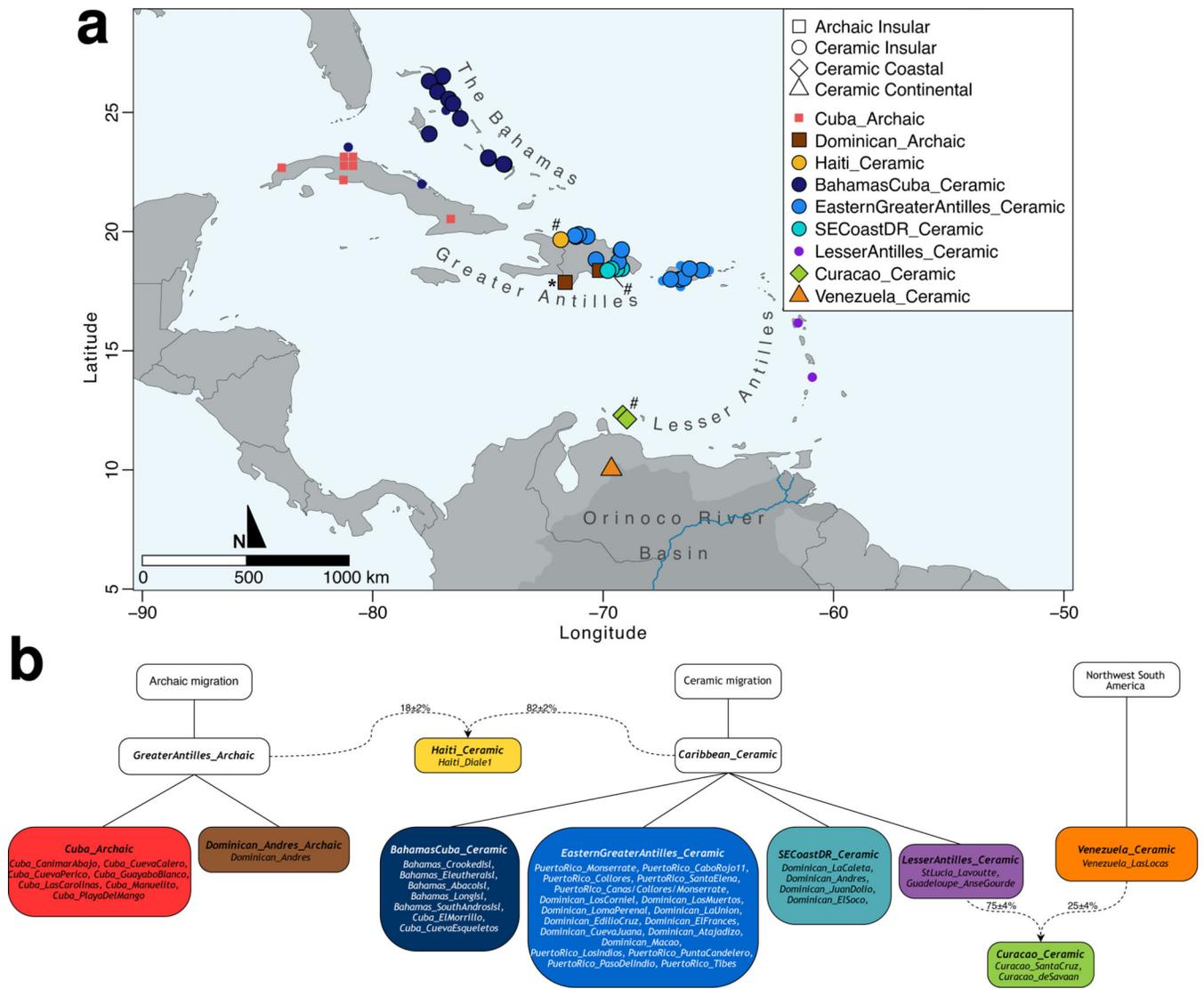


Fig. 1: Geography and significant genetic structure.

(a) Newly-reported data shown as large bordered shapes; co-analyzed data⁴ shown as small non-bordered shapes. Asterisk (*) denotes Archaic-associated site of Cueva Roja (excluded due to low-coverage); hash (#) denotes sites with admixed individuals. Andrés is represented as *SECoastDR_Ceramic* and *Dominican_Archaic*. Numbers of individuals and temporal distribution in Extended Data Fig. 1a. Map generated with the R package “maps” (R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>). (b) Relationships reconstructed from allele sharing (Supplementary Information section 8). Solid lines connect sub-groupings comprising a larger group; dashed lines represent admixture. Colored boxes represent final sub-clades with the color scheme matching Fig. 1a.

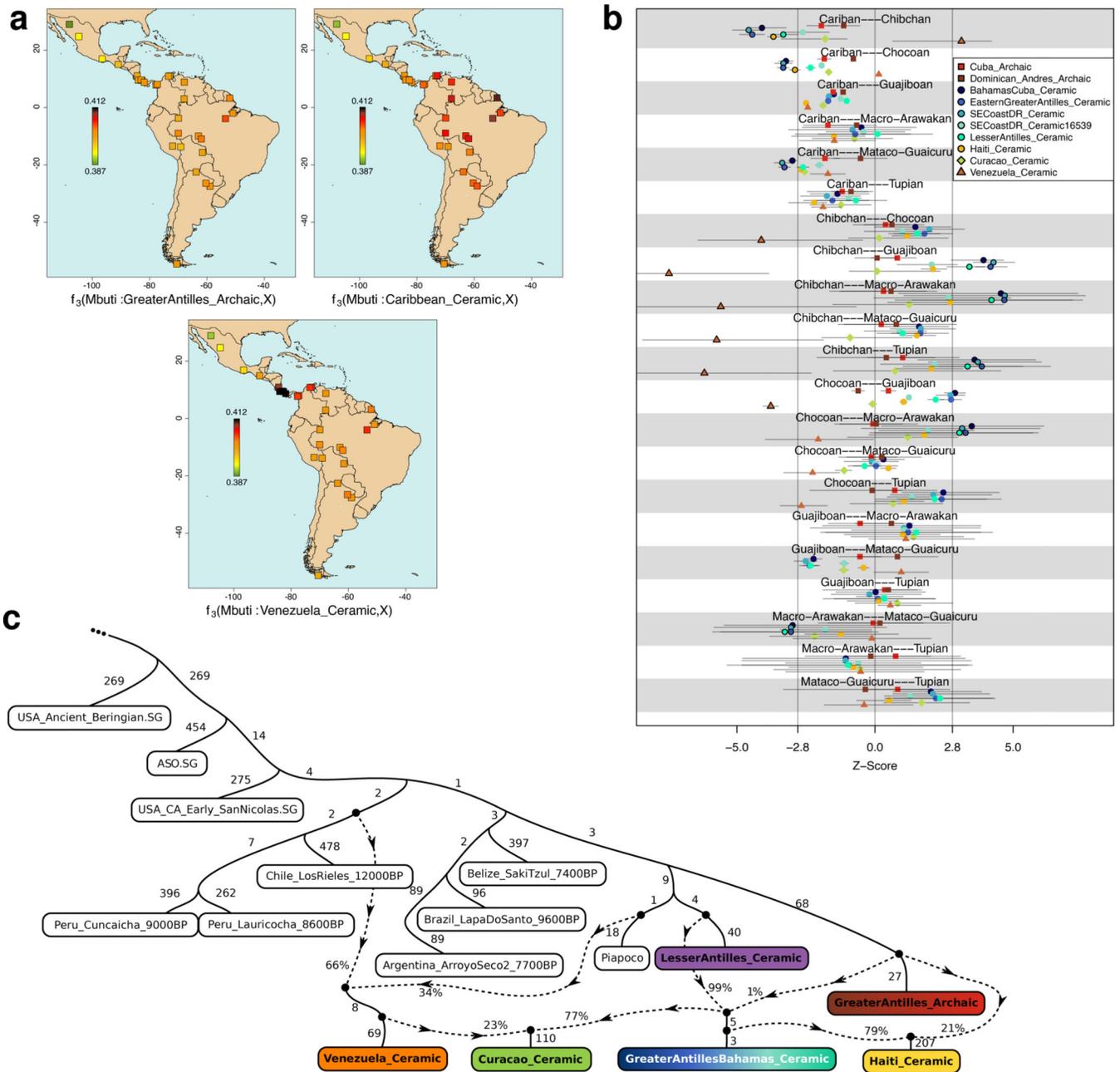


Fig. 2: Genetic affinities of ancient Caribbean people.

(a) Outgroup f_3 -statistics measuring the relatedness of the clades *GreaterAntilles_Archaic*, *Caribbean_Ceramic*, and *Venezuela_Ceramic* to present-day populations (squares). Map generated with the R package “maps” (R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>). (b) We computed $f_4(\text{Mbuti}, \text{Test}; \text{LanguageGroup1Pop}, \text{LanguageGroup2Pop})$ evaluating if each *Test* sub-clade is more closely related to populations belonging to one language family or another. Points represent the average Z-scores among all populations from each pair of language groups tested; horizontal lines show the range across such comparisons. Vertical lines represent a significance threshold

corresponding to a 99.5% CI. (e) Admixture graph modelling of representative ancient Caribbean groupings and select non-Caribbean populations. We fit 12 groups, including the clades *LesserAntilles_Ceramic* and *GreaterAntilles_Archaic*, without mixture; the other three *Caribbean_Ceramic* sub-clades and the clade *Venezuela_Ceramic* fit as mixtures. The worst Z-score comparing observed to expected *f*-statistics is |3.6|, which is not significant after correcting for multiple hypothesis testing.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

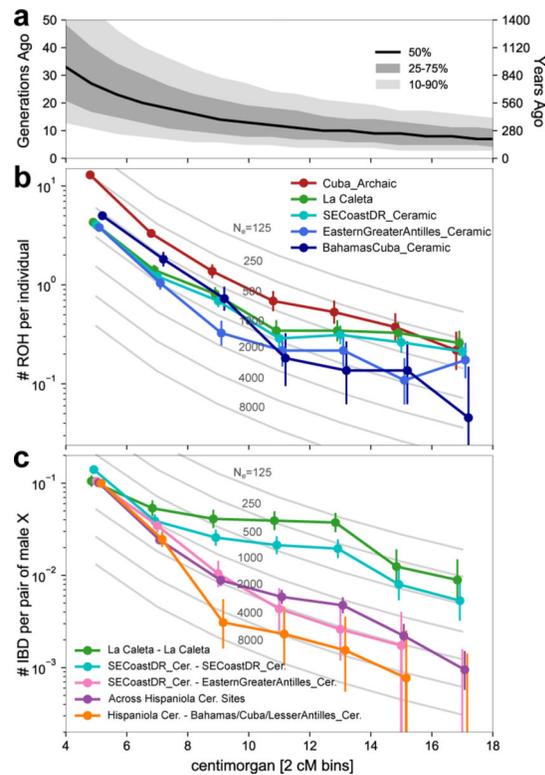


Fig. 3: Estimates of effective population size from shared haplotypes.

Details in Supplementary Information section 7. **(a)** Number of generations since two chromosomes with a shared segment of a specific size shared a common ancestor, assuming a constant population size $N=1000$. **(b)** Average rate of ROH segments in different length bins after excluding highly consanguineous individuals (defined as having a sum of ROH > 20 > 50 cM). **(c)** Rates of IBD segments shared on the X chromosome between pairs of males within length bins after excluding closely related individuals (defined as sum of IBD X > 20 > 25 cM). For the N_e estimates quoted in the paper we use the pool of 12–20 cM segments; for comparisons between the two major clades *SECoastDR_Ceramic* and *EasternGreaterAntilles_Ceramic* this gives $N_e=3082$ (95% CI 1530–8150). In **(b)** and **(c)** confidence intervals correspond to one standard deviation (68% coverage) assuming a Poisson distribution in each bin (vertical bars). Point estimates (circles) placed at the center of each 2 cM bin, with jitter added for visual separation. Gray lines depict expectations for panmictic populations of various sizes.

Table 1.
Archaeological debates addressed by our analyses.

Genetic data provide new insight into open debates inspired by archaeological research.

Debates	Genetic inferences
Archaic Age migration(s)	Archaic-associated individuals have ancestry more closely related to published Central and South Americans than to North Americans. Archaic-related ancestry was >98% replaced by Ceramic-related ancestry in most of the Greater Antilles but persisted with minimal admixture in Cuba for over 2,500 years. All Archaic-associated individuals are consistent with deriving from a single source, contrary to a claim of additional migration with affinity to North Americans.
Ceramic Age migration(s)	The great majority of Ceramic-associated individuals are genetically homogeneous with a connection to northeastern South America, now the homeland of Arawak-speakers. A south-to-north migratory movement of genetically-homogenous people is most parsimonious, although we cannot rule out multiple migrations by genetically similar groups.
Stylistic transitions and migrations	Genetic homogeneity across changes in ceramic styles provides evidence against a scenario of multiple waves of migration of genetically differentiated people from South America. We document over a millennium of genetic continuity in a small region of the southeast coast of Hispaniola.
Archaic/Ceramic interactions	Archaic- and Ceramic-associated admixture was extremely rare; we identify it in 3 of 201 ceramic-using Caribbean individuals. Unadmixed Archaic-related ancestry persisted as late as 700 BP in Cuba, but was replaced by Ceramic-related ancestry in Hispaniola beginning at least a millennium before.
Demographic history	Effective population sizes (N_e) for Ceramic-associated sites were larger (~500–1500) than for Archaic-associated sites (~200–300) and are estimated at ~1500–8000 across islands. A small pan-Caribbean gene pool and interconnected population is also evidenced by 19 cross-island relative pairs and very low genetic differentiation across the Ceramic Age Caribbean. As census size is unlikely to be >10x larger than N_e , population estimates in the hundreds of thousands are likely too large. Ancient Caribbean people avoided unions of first cousins or closer.
Persistence of ancestry today	We identify up to ~14% Ceramic-related ancestry in present-day Puerto Ricans and Cubans and identify a new mtDNA haplogroup unique to the Caribbean present in pre-contact times as well as today.