Article

# A genetic history of the pre-contact Caribbean

https://doi.org/10.1038/s41586-020-03053-2

Received: 31 May 2020

Accepted: 10 November 2020

Published online: 23 December 2020

Check for updates

A list of authors and their affiliations appears at the end of the paper.

Humans settled the Caribbean about 6,000 years ago, and ceramic use and intensified agriculture mark a shift from the Archaic to the Ceramic Age at around 2,500 years ago[1–3]. Here we report genome-wide data from 174 ancient individuals from The Bahamas, Haiti and the Dominican Republic (collectively, Hispaniola), Puerto Rico, Curaçao and Venezuela, which we co-analysed with 89 previously published ancient individuals. Stone-tool-using Caribbean people, who first entered the Caribbean during the Archaic Age, derive from a deeply divergent population that is closest to Central and northern South American individuals; contrary to previous work[4], we find no support for ancestry contributed by a population related to North American individuals. Archaic-related lineages were >98% replaced by a genetically homogeneous ceramic-using population related to speakers of languages in the Arawak family from northeast South America; these people 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 that reflect small effective population sizes, which we estimate to be a minimum of 500–1,500 and a maximum of 1,530–8,150 individuals on the combined islands of Puerto Rico and Hispaniola in the dozens of generations before the individuals who we analysed lived. Census sizes are unlikely to be more than tenfold larger than effective population sizes, so previous pan-Caribbean estimates of hundreds of thousands of people are too large[5,6]. Confirming a small and interconnected Ceramic Age population[7], we detect 19 pairs of cross-island cousins, close relatives buried around 75 km 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 differentiated groups from the mainland, but instead reflected interactions within an interconnected Caribbean world[1,8].

Prior to European colonization (hereafter, pre-contact), the Caribbean was a mosaic of archaeologically distinct communities that were 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, which denote shifts in material cultural complexes[1,9]. The Lithic and Archaic Ages are defined by distinct stone tool technologies[10,11], and the Ceramic Age—which began about 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 as well as migration from the American continents, although the geographical origins, trajectories and numbers of migratory waves remain under debate[1,3,12] (Table 1, Supplementary Information section 1).

We screened 195 ancient individuals and generated genome-wide data that passed authenticity criteria for 174 individuals (Supplementary Data 1, 2); on the basis of 45 newly generated radiocarbon dates, these individuals lived between about 3,100 and 400 calibrated years before present (AD 1950) (cal. BP) (Extended Data Fig. 1a, Supplementary Data 3, Supplementary Information section 3) in The Bahamas, Hispaniola, Puerto Rico, Curaçao and Venezuela (Fig. 1a, Supplementary Information section 2). These individuals had a median of 700,689 single-nucleotide polymorphisms (SNPs) covered (range of 20,063–977,658 SNPs; median of 2.2× coverage of targeted positions (range of 0.02–9.95×)) (Supplementary Data 1). We co-analysed the newly generated data alongside 89 previously published individuals[4] (Supplementary Information section 4). In what follows, we denote sites with stone tools or radiocarbon dates that predate 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 analysed, 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 before submission with members of Indigenous communities who trace their legacy to the pre-contact Caribbean and their feedback was incorporated. Genetic

**Table 1 | Archaeological debates addressed by our analyses**

| Debates | Genetic inferences |
|---|---|
| Archaic Age migration(s) | Archaic-associated individuals have ancestry more closely related to published Central and South American individuals than to North American individuals. 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 American individuals. |
| 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 | Significant admixture between Archaic- and Ceramic-associated peoples was extremely rare; we identify it in 3 out of 201 ceramic-using Caribbean individuals. Unadmixed Archaic-related ancestry persisted as late as 700 BP in Cuba, but it was replaced by Ceramic-related ancestry in Hispaniola beginning at least a millennium before. |
| Demographic history | $N_e$ values for Ceramic-associated sites were larger (about 500–1,500) than for Archaic-associated sites (about 200–300), and are estimated at around 1,500–8,000 across islands. A small pan-Caribbean gene pool and interconnected population is also supported by our identification of 19 cross-island relative pairs and very low genetic differentiation across the Ceramic Age Caribbean. As census size is unlikely to be >10× larger than $N_e$, population estimates in the hundreds of thousands are probably too large. Ancient Caribbean people avoided unions of first cousins or closer. |
| Persistence of ancestry today | We identify up to around 14% Ceramic-related ancestry in present-day Puerto Rican and Cuban individuals, and identify a previously undocumented mtDNA haplogroup that is unique to the Caribbean and was present in pre-contact times as well as today. |

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 provided 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 groups[13] (Extended Data Fig. 1b, Supplementary Data 4). Ceramic- and Archaic-associated individuals project in separate clusters, whereas ancient Venezuelan individuals relate to present-day Chibchan-speakers (such as Cabécar) in PCA and ADMIXTURE analysis (Extended Data Fig. 1b, c, Supplementary Information sections 5, 6; population self-denominations are provided in Supplementary Data 5). Individuals from Curaçao and Haiti (who are admixed, discussed in 'The spread of ceramic users') mostly overlap the Ceramic-associated cluster. An exception to within-site genetic homogeneity is at Andrés (a primarily Ceramic-associated site in the Dominican Republic), where individual I10126 is dated to the Archaic Age (about 3,140–2,950 cal. BP) (Supplementary Data 3) and appears genetically similar to other Archaic-associated individuals (Extended Data Fig. 1b, c). We exclude from subsequent analyses three Archaic-associated individuals from Cueva Roja (about 1,900 cal. BP, in the Dominican Republic) with low coverage (less than about 0.05×) 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 on the basis of allele-sharing, starting with major 'clades' and then 'subclades' (Supplementary Information section 8); we denote groups defined using genetics by asterisks[14]. Our nomenclature for these genetic groupings combines the geographical location that encompasses sites in the cluster plus 'Archaic' or 'Ceramic' (Fig. 1b).

We identified three significantly differentiated major clades (Figs. 1b, 2). *GreaterAntilles_Archaic included 50 individuals from Cuba, spanning about 3,200–700 cal. BP[4], as well as individual I10126 from Andrés.*Caribbean_Ceramic comprised 194 individuals from Ceramic-associated sites, dating to between about 1,700 and 400 cal. BP. *Venezuela_Ceramic comprised 8 individuals dated to around 2,350 cal. BP. Two *Haiti_Ceramic and five *Curacao_Ceramic individuals fit as mixtures of major clades.

We next identified subclades and substructure within them (Supplementary Data 6, Supplementary Table 6). Within *Caribbean_Ceramic, southeast coast Dominican Republic Ceramic (*SECoastDR_Ceramic) comprised 4 sites along 50 km of coastline (from west to east, La Caleta, Andrés, Juan Dolio and El Soco) (Supplementary Table 7). These sites were occupied for about 1,400 years, documenting genetic continuity across changes in ceramic styles. All Ceramic-associated sites from The Bahamas and Cuba (spanning around 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 $F_{ST}$ values of less than about 0.01 indicate a notable degree of homogeneity among these *Caribbean_Ceramic subclades (compared to $F_{ST}$ values of about 0.1 between Ceramic- and Archaic-related clades), reflecting high migration rates among islands (as discussed in 'Social structure and population size estimates') (Extended Data Fig. 2).

To identify *Caribbean_Ceramic individuals who had an excess of Archaic-related ancestry relative to others within each subclade, 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- and Archaic-related admixture ($Z = -5.5$) (Supplementary Table 8). In contrast to a previous claim[4], we did not detect significant Archaic-related admixture in individual PDI009 from Paso del Indio (Puerto Rico) ($Z = 0.6$) (Supplementary Information section 4, Supplementary Table 3).

## Archaic-associated Caribbean people

The *GreaterAntilles_Archaic clade shares the most genetic drift with Indigenous groups from Central and northern South America who belong to seven language families: Arawakan, Cariban, Chibchan,

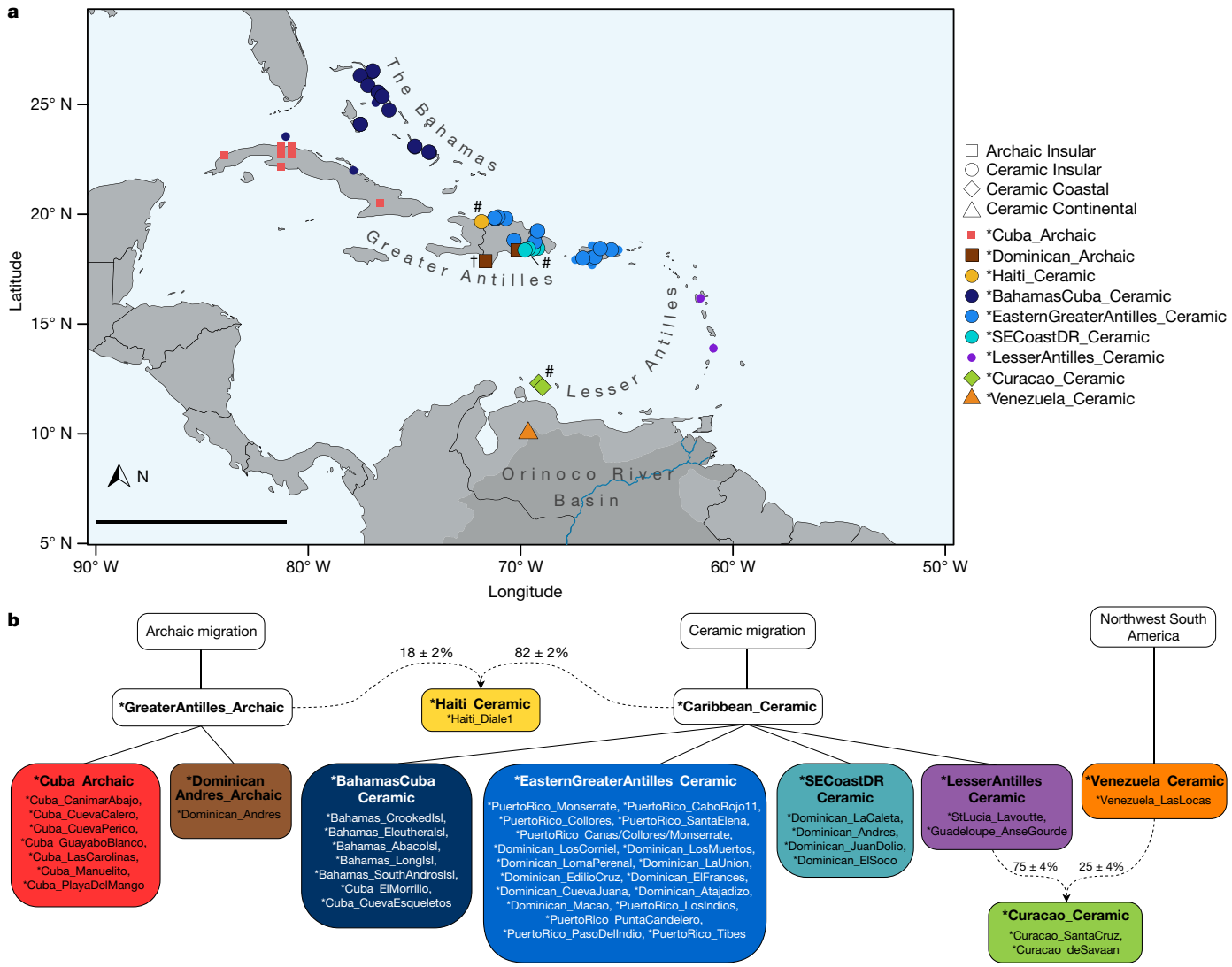



**Fig. 1 | Geography and genetic structure. a**, Newly reported data are shown as large bordered shapes; co-analysed data[4] are shown as small nonbordered shapes. †Archaic-associated site of Cueva Roja (excluded from our main analyses owing to low coverage); # denotes a site with admixed individuals. Andrés is represented as *SECoastDR_Ceramic and *Dominican_Archaic. Numbers of individuals and temporal distribution are shown in Extended Data Fig. 1a. Map generated with the R package maps (https://CRAN.R-project.org/package=maps). Scale bar, 1,000 km. **b**, Relationships reconstructed from allele-sharing (Supplementary Information section 8). Solid lines connect subgroupings comprising a larger group; dashed lines represent admixture. Coloured boxes represent final subclades; colour scheme matches that in **a**. Isl, island.

Chocoan, Guajiboan, Mataco–Guaicuru and Tupian[15,16] (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, b, Supplementary Data 11). Archaic-associated individuals from Cuba share more alleles with each other than with Dominican individual I10126 (Supplementary Table 6), which demonstrates Archaic-related 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 American individuals also contributed ancestry to some Archaic Age Caribbean individuals[4] (Supplementary Information section 17). This claim was based on a finding of affinity between Early Period individuals from the Channel Islands of California (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 not significant ($Z = -0.9$) (Supplementary Table 25). 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). The negative value of this statistic was previously interpreted as evidence for affinity between CIP009 and USA_CA_Early_SanNicolas; although we replicated the non-significant statistic ($Z = -1.3$) (Supplementary Table 23), it became positive when we replaced the Mbuti outgroup with diverse Eurasian individuals or Bahamas_Taino[17] with ancient Bahamian shotgun data newly generated for this study, which should give qualitatively similar results (Supplementary Tables 24, 26). 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 that there is no evidence for a North-American-specific relationship (Supplementary Table 27). 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, Supplementary

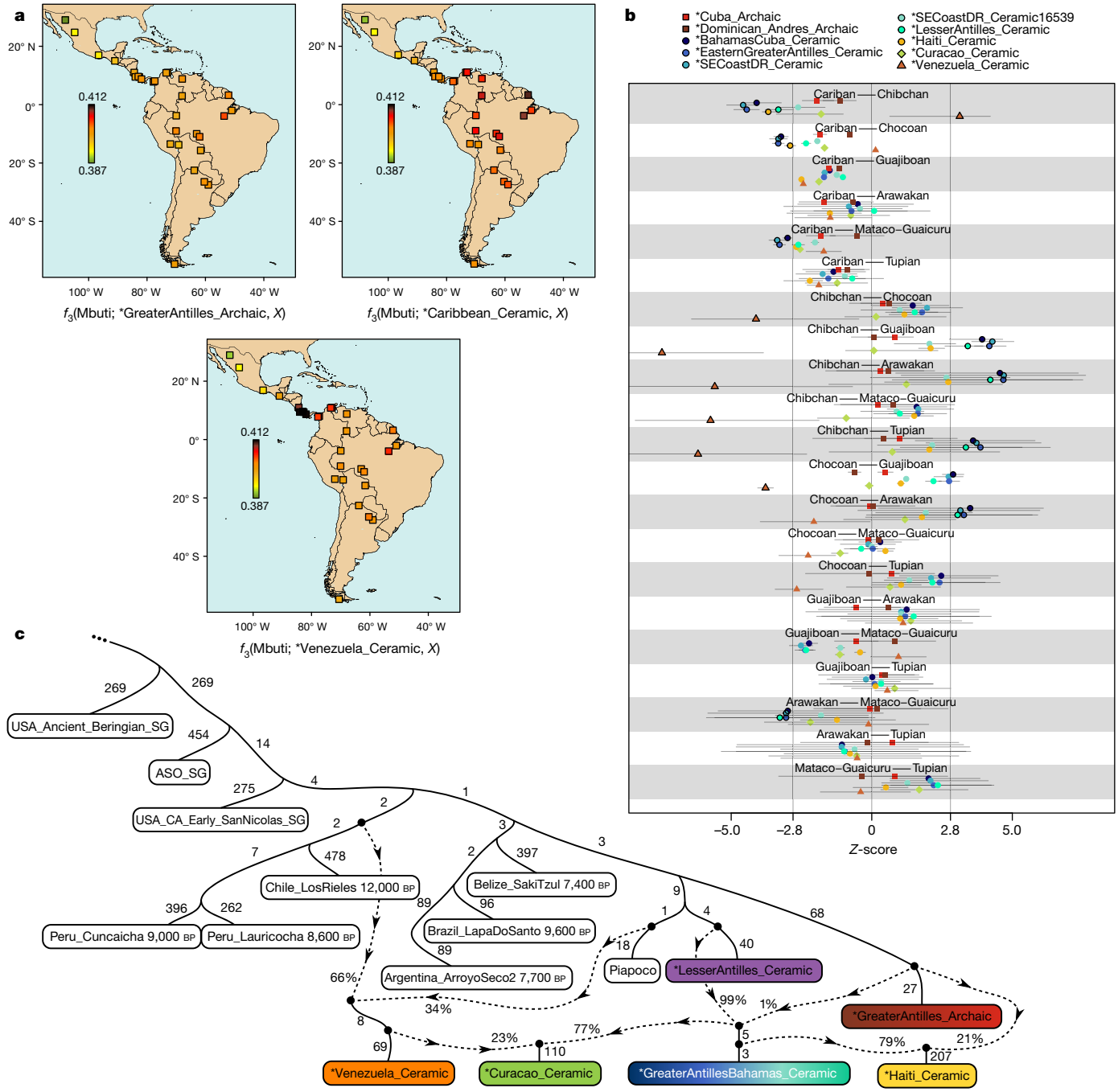



**Fig. 2 | Genetic affinities of ancient Caribbean people. a**, Outgroup $f_3$-statistics measuring the relatedness of the clades *GreaterAntilles_Archaic, *Caribbean_Ceramic and *Venezuela_Ceramic to present-day populations (squares). Map generated with the R package maps (https://CRAN.R-project.org/package=maps). **b**, We computed $f_4$(Mbuti, test; language group 1 population, language group 2 population) evaluating whether each test subclade 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. Outlined symbols represent averages above a significance threshold corresponding to a 99.5% confidence interval (denoted by vertical lines). **c**, 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 subclades and the clade *Venezuela_Ceramic fit as mixtures. The worst *Z*-score comparing observed-to-expected *f*-statistics is |3.6|. ASO, ancient southwestern Ontario; SG, sample with whole-genome shotgun data.

Fig. 34). Thus, to the limits of the resolution of allele-sharing methods, all Archaic-related Caribbean ancestry is consistent with deriving from a single source.

In qpGraph, we fit *GreaterAntilles_Archaic in an early splitting branch that contains most ancient Caribbean, Belizean, Brazilian and Argentinian populations (Fig. 2c). In a maximum-likelihood tree allowing admixture events[18], *GreaterAntilles_Archaic also fits as a divergent Indigenous American group (Extended Data Fig. 3). We could not obtain further evidence of specific affinities to mainland groups using qpAdm (Supplementary Information section 9; Supplementary Table 16) or $f_4$-statistics (Supplementary Table 17).

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

2,500 years, resonating with archaeological[19] and historical[20] accounts that this region was home to people with a distinct language and cultural traditions as late as the time of European contact.

## The spread of ceramic users

Previous analyses have found that *Caribbean_Ceramic-associated people have genetic affinities to Arawak-speakers in northeastern South America[17,21,22] (Supplementary Information section 1). Although we are not able to support this conclusion with our 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 analysis suggests an Arawak affinity, as individuals from each *Caribbean_Ceramic subclade are almost entirely composed of a component found in the highest proportion in modern Arawak-speakers (for example, 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 subclades 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 (Supplementary Tables 18, 19), and qpGraph (Fig. 2c).

We estimate about 0.5–2.0% Archaic-related ancestry in the Ceramic-associated people of the Greater Antilles and The Bahamas when modelled in qpAdm as a mixture of *LesserAntilles_Ceramic and *Dominican_Andres_Archaic (Supplementary Table 21). We reject reverse models of *LesserAntilles_Ceramic deriving from any of the *Caribbean_Ceramic subclades, which fail when Archaic-associated people are included in the reference set ($P = 0.001–0.008$) (Supplementary Table 21). The simplest explanation for these observations is a scenario of south-to-north movement of ceramic-using ancestors into the Caribbean, in which ancestry similar to that in the 1,000–650 cal. BP 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 around 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 ± 1.9% (I16539 from La Caleta) (Supplementary Table 9) and 18.5 ± 2.1% (two individuals from Diale 1, Haiti) (Supplementary Tables 12, 13). Using the software DATES[23], we estimate that admixture occurred around 16 ± 3 generations (about 350–500 years) before these individuals from Haiti lived (Supplementary Information section 14).

The affinities of *Venezuela_Ceramic with Chibchan-speakers in ADMIXTURE and *f*-statistics (Fig. 2a, b, Extended Data Fig. 1c) are confirmed in qpAdm, in which *Venezuela_Ceramic fits as a clade with Cabécar (Supplementary Tables 18, 19). Thus, although Las Locas is located in a hypothesized source region for the expansion of ceramic-associated cultures 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 having 74.5 ± 3.7% *LesserAntilles_Ceramic-related ancestry and 25.5 ± 3.7% *Venezuela_Ceramic-related ancestry (Supplementary Table 15), which suggests that the Ceramic Age population of Curaçao 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 that link sites such as Las Locas to Curaçao.

Although a study of cranial morphology suggested a possible Carib migration from western Venezuela about 1,150 years ago[24], we find no evidence of a new ancestry at this time, as might be expected for such an event. In simulations using *Venezuela_Ceramic, *LesserAntilles_Ceramic or present-day Cariban-speaking Arara as proxies for Carib peoples, we can detect as little as around 2–8% ancestry from such groups (Supplementary Information section 13). The genetic data show no evidence for a separate migration, although we cannot rule out migration from an unsampled continental group that is genetically more similar to *Caribbean_Ceramic-associated 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 more than 400,000 SNPs covered for runs of homozygosity (ROH) of over 4 centimorgan (cM)[25] (Supplementary Data 12, Supplementary Information section 7, Supplementary Fig. 21). Large sums of long ROH (more than 20 cM) indicate parental relatedness within the past few generations, whereas an abundance of shorter ROH signals indicate background parental relatedness and restricted mating pools[26]. Only 2 out of 202 individuals had more than 100 cM of their genome in blocks of ROH of more than 20 cM (about 135 cM is the average in offspring of first cousins), which indicates that close kin unions were rare. By contrast, 48 individuals had at least one ROH of over 20 cM, which indicates that many unions took place between individuals as close as second or third cousins and suggests 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 of 4–20 cM, which arise from co-ancestry mostly within the past 50 generations (Fig. 3a, b). Estimates of $N_e$ can be used to infer census population size, which in humans is typically threefold, and up to tenfold, greater[27,28]. $N_e$ values for Ceramic-associated Caribbean sites are larger ($N_e$ of around 500–1,500, similar to previous estimates[17,21]) than for Archaic-associated sites ($N_e$ of around 200–300) (Extended Data Fig. 4a, Extended Data Table 1), which points 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).

Estimates of $N_e$ 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 analysed long shared segments (identical-by-descent (IBD) blocks) between the X chromosomes of pairs of males (Supplementary Information section 7). Focusing on shared segments of long IBD of 12–20 cM (which reflect the size of the shared ancestor pool from within the past approximately 20 generations) (Fig. 3a), we find that the rate of such segments decreases with geographical distance (Fig. 3c), as expected if people exchange more genes with people who live closer to them. However, we still detect 19 pairs of individuals who share segments of at least 8.7 cM across islands (Extended Data Table 2), which reveals that people across the Caribbean shared common ancestors in the hundreds of years before 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 $N_e = 3,082$ (95% confidence interval of 1,530–8,150) (estimates are given in the legend of Fig. 3). 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 tenfold to obtain census size, we infer that estimates of pre-contact population size of hundreds of thousands or even millions for large islands such as Hispaniola[5] (on the basis of outdated reports or poorly documented population counts[6]) 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, where 37 out of 63 individuals studied had one or several close relatives, although

the rate per tested pair was not significantly greater than within other sites (95% confidence interval of 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 around 75 km 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 in present-day populations

We tested for genetic affinity between the Indigenous ancestry found in present-day[22] and ancient Caribbean people by computing $f_4$(European, test; *Cuba_Archaic, *Caribbean_Ceramic). We obtained a signal for relatedness between Puerto Rican individuals 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 Cuba[29] and found 2 provinces and 8 municipalities with weakly significant evidence of Ceramic-related ancestry ($2.0 < |Z| < 3.4$) and only a single municipality (Guines, in western Cuba) with marginally significant evidence of Archaic-related ancestry ($Z = 2.0$) (Supplementary Data 14). Thus, although the available ancient data show the perpetuation of unadmixed Archaic-related ancestry in parts of Cuba into the past millennium, it became heavily admixed with Ceramic-related ancestry before the present day.

Previous reports have found pre-contact Indigenous ancestry in present-day Caribbean people in uniparental haplogroups[30–33]. We build on these findings by identifying—to our knowledge—a previously undocumented deep branch of mitochondrial (mt)DNA haplogroup C1d at a frequency of about 7% across Caribbean Ceramic subclades as well as in a modern Puerto Rican individual from the 1000 Genomes Project dataset[34] (Supplementary Data 9, Supplementary Information section 10). This provides direct evidence that Indigenous matrilineal ancestry has 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 is consistent with deriving from a single source, with only subtle differences among Archaic-associated individuals who span about 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 (although 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 Caribbean[35]. Ceramic-associated individuals show a genetic affinity to present-day Arawak-speakers, consistent with archaeological and linguistic evidence of northeastern South American origin[36]. Consistent 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 towards the western Venezuela coast)[30], individuals from Curaçao have ancestry related to that in *LesserAntilles_Ceramic. Although the earliest Ceramic Age 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 about 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 Caribbean[4].

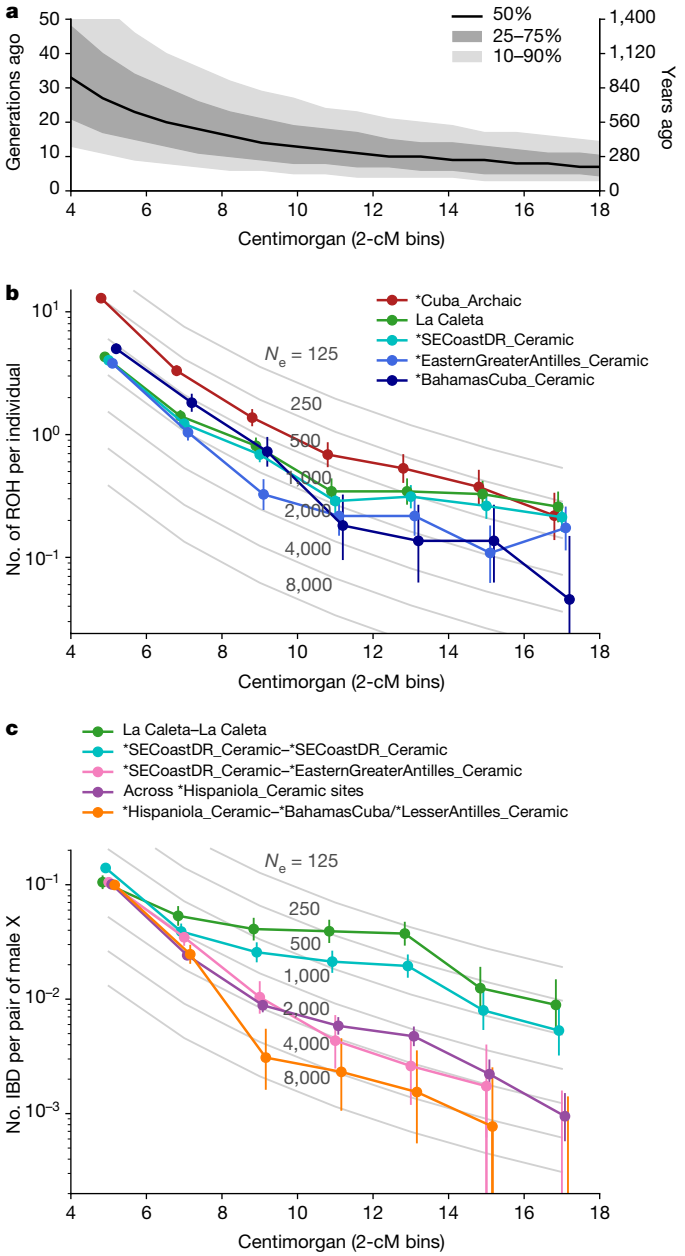



**Fig. 3 | Estimates of $N_e$ from shared haplotypes. a**, Number of generations since two chromosomes with a shared segment of a specific size shared a common ancestor, assuming a constant population size $N_e = 1{,}000$. **b**, Average rate of ROH segments in different length bins after excluding highly consanguineous individuals (defined as having more than 50 cM of their genome in blocks of ROH >20 cM in length). **c**, Rates of IBD segments shared on the X chromosome between pairs of males within length bins after excluding closely related individuals (defined as sharing more than 25 cM of their X chromosome in IBD blocks >20 cM in length). For the $N_e$ estimates, we use the pool of 12–20 cM segments; for comparisons between the two major clades *SECoastDR_Ceramic and *EasternGreaterAntilles_Ceramic, this gives $N_e = 3{,}082$ (95% confidence interval 1,530–8,150). In **b**, **c**, confidence intervals correspond to one s.d. (68% coverage) assuming a Poisson distribution in each bin (vertical bars). Point estimates (circles) are placed at the centre of each 2-cM bin, with jitter added for visual separation. Grey lines depict expectations for panmictic populations of various sizes. Further details are provided in Supplementary Information section 7.

Third, we find no association between our *Caribbean_Ceramic subclades and the traditional Caribbean ceramic typologies (Saladoid, Ostionoid, Meillacoid and Chicoid), providing no support for a

culture-history model that views these 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. Although 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 catalysed stylistic transitions.

Fourth, to our knowledge, we provide the first evidence of admixture between Archaic- and Ceramic-related ancestry in three individuals in Hispaniola. This finding also confirms a previous inference[4] that admixture between people of Archaic- and Ceramic-related ancestry in the Caribbean was extremely rare (seen here in only 3 out of 201 ceramic-using Caribbean individuals).

Fifth, we confirm that people living in some parts of the Caribbean (especially Puerto Rico and Cuba) at present 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 probably reflects pre-contact events (as Ceramic-related ancestry was present in individuals from western and central Cuba dated to around 500 cal. BP).

Sixth, our data provide insights into social structure and demography. By analysing 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, which reflects a small population size[37]. We identify male relatives buried about 75 km apart, which suggests networks of connectivity between archaeological sites that have otherwise been analysed 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 3,082 (95% confidence interval of 1,530–8,150) across the large islands of Hispaniola and Puerto Rico. Although these estimates represent the past approximately 20 generations since the analysed individuals lived, they point to a census size across these large islands that was substantially less than the estimates of hundreds of thousands to millions at contact that have been suggested in some of the literature[1,38]. Although our estimates of population size are lower than those from historical reports and population counts[5,6], the devastating effects 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 into the present, and the study of ancient DNA helps us to better appreciate this. Present-day Caribbean people contain mixtures of genetic ancestry in different proportions, primarily comprising pre-contact Indigenous populations (about 4% on average in Cuba, about 6% in the Dominican Republic and about 14% in Puerto Rico according to our estimation by qpAdm), immigrant European individuals (about 70% in Cuba, about 56% in the Dominican Republic and about 68% in Puerto Rico) and African individuals who were brought to this region during the course of the trans-Atlantic slave trade (about 26% in Cuba, about 38% in the Dominican Republic and about 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.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-03053-2.


1. Rouse, I. *The Tainos: Rise & Decline of the People who Greeted Columbus* (Yale Univ. Press, 1992).
2. Maggiolo, M. V. *La Isla de Santo Domingo antes de Colón* (Banco Central de la Republica Dominicana, 1993).
3. Keegan, W. F. & Hofman, C. L. *The Caribbean before Columbus* (Oxford Univ. Press, 2017).
4. Nägele, K. et al. Genomic insights into the early peopling of the Caribbean. *Science* **369**, 456–460 (2020).
5. Cook, S. F. & Borah, W. *Essays in Population History* Vol. 1, 376–410 (Univ. California Press, 1971).
6. Henige, D. On the contact population of Hispaniola: history as higher mathematics. *Hisp. Am. Hist. Rev.* **58**, 217–237 (1978).
7. Wilson, S. M. *The Archaeology of the Caribbean* (Cambridge Univ. Press, 2007).
8. Rodríguez Ramos, R. in *Oxford Handbook of Caribbean Archaeology* (eds Keegan, W. F. et al.) 155–170 (Oxford Univ. 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, R. T. in *Oxford Handbook of Caribbean Archaeology* (eds Keegan, W. F. et al.) 285–295 (Oxford Univ. Press, 2013).
11. Siegel, P. E. et al. Paleoenvironmental evidence for first human colonization of the eastern Caribbean. *Quat. Sci. Rev.* **129**, 275–295 (2015).
12. Oliver, J. R. *The Archaeological, Linguistic and Ethnohistorical Evidence for the Expansion of Arawakan into Northwestern Venezuela and Northeastern Colombia*. PhD thesis, Univ. Illinois at Urbana-Champaign (1989).
13. Reich, D. et al. Reconstructing Native American population history. *Nature* **488**, 370–374 (2012).
14. Eisenmann, S. et al. Reconciling material cultures in archaeology with genetic data: The nomenclature of clusters emerging from archaeogenomic analysis. *Sci. Rep.* **8**, 13003 (2018).
15. Greenberg, J. H. *Language in the Americas* (Stanford Univ. Press, 1987).
16. Salzano, F. M., Hutz, M. H., Salamoni, S. P., Rohr, P. & Callegari-Jacques, S. M. Genetic support for proposed patterns of relationship among lowland South American languages. *Curr. Anthropol.* **46**, S121–S128 (2005).
17. Schroeder, H. et al. Origins and genetic legacies of the Caribbean Taino. *Proc. Natl Acad. Sci. USA* **115**, 2341–2346 (2018).
18. Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).
19. Chinique de Armas, Y., Roksandic, M., Suárez, R. R., Smith, D. G. & Buhay, W. M. in *Cuban Archaeology in the Circum-Caribbean Context* (ed. Roksandic, I.) 125–146 (Univ. Press Florida, 2016).
20. Lovén, S. E. *Origins of the Tainan Culture, West Indies* (Elanders, 1935).
21. Nieves-Colón, M. A. et al. Ancient DNA reconstructs the genetic legacies of precontact Puerto Rico communities. *Mol. Biol. Evol.* **37**, 611–626 (2020).
22. Moreno-Estrada, A. et al. Reconstructing the population genetic history of the Caribbean. *PLoS Genet.* **9**, e1003925 (2013).
23. Narasimhan, V. M. et al. The formation of human populations in South and Central Asia. *Science* **365**, eaat7487 (2019).
24. Ross, A. H., Keegan, W. F., Pateman, M. P. & Young, C. B. Faces divulge the origins of Caribbean prehistoric inhabitants. *Sci. Rep.* **10**, 147 (2020).
25. Ringbauer, H., Novembre, J. & Steinrucken, M. Human parental relatedness through time - detecting runs of homozygosity in ancient DNA. Preprint at https://doi.org/10.1101/2020.05.31.126912 (2020).
26. Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M. & Wilson, J. F. Runs of homozygosity: windows into population history and trait architecture. *Nat. Rev. Genet.* **19**, 220–234 (2018).
27. Frankham, R. Effective population size/adult population size ratios in wildlife: a review. *Genet. Res.* **89**, 491–503 (2007).
28. Browning, S. R. & Browning, B. L. Accurate non-parametric estimation of recent effective population size from segments of identity by descent. *Am. J. Hum. Genet.* **97**, 404–418 (2015).
29. Fortes-Lima, C. et al. Exploring Cuba's population structure and demographic history using genome-wide data. *Sci. Rep.* **8**, 11422 (2018).
30. Toro-Labrador, G., Wever, O. R. & Martínez-Cruzado, J. C. Mitochondrial DNA analysis in Aruba: strong maternal ancestry of closely related Amerindians and implications for the peopling of northwestern Venezuela. *Caribb. J. Sci.* **39**, 11–22 (2003).
31. Mendizabal, I. et al. Genetic origin, admixture, and asymmetry in maternal and paternal human lineages in Cuba. *BMC Evol. Biol.* **8**, 213 (2008).
32. Vilar, M. G. 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).
33. 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).
34. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
35. Hofman, C. L. & Reid, B. A. in *Encyclopedia of Caribbean Archaeology* (eds Reid, B. & Gilmore, G.) 300–303 (Univ. Press Florida, 2014).
36. Roksandic, I. & Roksandic, M. in *New Perspectives on the Peopling of the Americas* (eds Harvati, K. et al.) 199–223 (Kerns, 2018).
37. Keegan, W. *The People Who Discovered Columbus* (Univ. Press Florida, 1992).
38. Anderson-Córdova, K. F. *Surviving Spanish Conquest: Indian Fight, Flight, and Transformation in Hispaniola and Puerto Rico* (Univ. Alabama Press, 2017).

**Daniel M. Fernandes[1,2,33], Kendra A. Sirak[3,4,33], Harald Ringbauer[3,4], Jakob Sedig[3,4], Nadin Rohland[3,5], Olivia Cheronet[1], Matthew Mah[3,4,5,6], Swapan Mallick[3,4,5,6], Iñigo Olalde[3,7], Brendan J. Culleton[8], Nicole Adamski[3,6], Rebecca Bernardos[3,6], Guillermo Bravo[1,9], Nasreen Broomandkhoshbacht[3,6,31], Kimberly Callan[3,6], Francesca Candilio[10], Lea Demetz[1], Kellie Sara Duffett Carlson[1], Laurie Eccles[11], Suzanne Freilich[1], Richard J. George[12], Ann Marie Lawson[3,6], Kirsten Mandl[1], Fabio Marzaioli[13], Weston C. McCool[12], Jonas Oppenheimer[3,6,32], Kadir T. Özdogan[1], Constanze Schattke[1], Ryan Schmidt[14], Kristin Stewardson[3,6], Filippo Terrasi[13], Fatma Zalzala[3,6], Carlos Arredondo Antúnez[15], Ercilio Vento Canosa[16], Roger Colten[17], Andrea Cucina[18], Francesco Genchi[19], Claudia Kraan[20], Francesco La Pastina[19], Michaela Lucci[21], Marcio Veloz Maggiolo[22], Beatriz Marcheco-Teruel[23], Clenis Tavarez Maria[24], Christian Martínez[24], Ingeborg París[25], Michael Pateman[26,27], Tanya M. Simms[28], Carlos Garcia Sivoli[25], Miguel Vilar[29], Douglas J. Kennett[12], William F. Keegan[30,34], Alfredo Coppa[1,3,19,34] ✉, Mark Lipson[3,4,34], Ron Pinhasi[1,34] ✉ & David Reich[3,4,5,6,34] ✉**

[1]Department of Evolutionary Anthropology, University of Vienna, Vienna, Austria. [2]CIAS, Department of Life Sciences, University of Coimbra, Coimbra, Portugal. [3]Department of Genetics, Harvard Medical School, Boston, MA, USA. [4]Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, USA. [5]Broad Institute of Harvard and MIT, Cambridge, MA, USA. [6]Howard Hughes Medical Institute, Harvard Medical School, Boston, MA, USA. [7]Institute of Evolutionary Biology, CSIC–Universitat Pompeu Fabra, Barcelona, Spain. [8]Institutes of Energy and the Environment, The Pennsylvania State University, University Park, PA, USA. [9]Department of Legal Medicine, Toxicology and Physical Anthropology, University of Granada, Granada, Spain. [10]Superintendency of Archaeology, Fine Arts and Landscape for the city of Cagliari and the provinces of Oristano and South Sardinia, Cagliari, Italy. [11]Department of Anthropology, The Pennsylvania State University, University Park, PA, USA. [12]Department of Anthropology, University of California, Santa Barbara, CA, USA. [13]Department of Mathematics and Physics, Campania University 'Luigi Vanvitelli', Caserta, Italy. [14]CIBIO–InBIO, University of Porto, Vairão, Portugal. [15]Museo Antropológico Montané, University of Havana, Havana, Cuba. [16]Matanzas University of Medical Sciences, Matanzas, Cuba. [17]Peabody Museum of Natural History, Yale University, New Haven, CT, USA. [18]Facultad de Ciencias Antropológicas, Universidad Autónoma de Yucatán, Mérida, Mexico. [19]Department of Environmental Biology, Sapienza University of Rome, Rome, Italy. [20]National Archaeological–Anthropological Memory Management (NAAM), Willemstad, Curaçao. [21]DANTE Laboratory of Diet and Ancient Technology, Sapienza University of Rome, Rome, Italy. [22]Universidad Autónoma de Santo Domingo, San Francisco de Macorís, Dominican Republic. [23]National Center of Medical Genetics, Medical University of Havana, Havana, Cuba. [24]Museo del Hombre Dominicano, Santo Domingo, Dominican Republic. [25]Instituto de Investigaciones Bioantropológicas y Arqueológicas, Universidad de Los Andes, Mérida, Venezuela. [26]Turks and Caicos National Museum Foundation, Cockburn Town, Turks and Caicos Islands. [27]AEX Bahamas Maritime Museum, Freeport, Bahamas. [28]Department of Biology, University of The Bahamas, Nassau, Bahamas. [29]National Geographic Society, Washington, DC, USA. [30]Florida Museum of Natural History, University of Florida, Gainesville, FL, USA. [31]Present address: Department of Anthropology, University of California, Santa Cruz, CA, USA. [32]Present address: Department of Biomolecular Engineering, University of California, Santa Cruz, CA, USA. [33]These authors contributed equally: Daniel M. Fernandes, Kendra A. Sirak. [34]These authors jointly supervised this work: William F. Keegan, Alfredo Coppa, Mark Lipson, Ron Pinhasi, David Reich. ✉e-mail: alfredo.coppa@uniroma1.it; ron.pinhasi@univie.ac.at; reich@genetics.med.harvard.edu

# 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; Supplementary Information section 2 provides archaeological site information, and Supplementary Figs. 1–11 show the location of the islands and/or sites studied. Powder was produced from a cochlea[39,40], tooth, phalanx or ossicle[41] from each individual in a clean room facility at Harvard Medical School, University College Dublin or the University of Vienna; Supplementary Data 2 provides 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[42–44]. From the extracts, we prepared dual-barcoded double-stranded[45] or dual-indexed single-stranded libraries[46], both treated with uracil-DNA glycosylase (UDG) to reduce the rate of characteristic ancient DNA damage[47]. Double-stranded libraries were treated in a modified partial UDG preparation[45] ('half'), leaving a reduced damage signal at both ends (5′ C-to-T, 3′ G-to-A). Single-stranded libraries were treated with *Escherichia 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; Supplementary Data 2 gives the number of libraries analysed for each individual.

To generate SNP capture data, we used in-solution target hybridization to enrich for sequences that overlap the mitochondrial genome and about 1.24 million genome-wide SNPs[48–51] ('1240K'), either in two separate enrichments or simultaneously (Supplementary Data 2). We then added two seven-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).

Before 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 (available at https://github.com/DReichLab/ADNA-Tools) was used for merging and trimming adapters and barcodes. Merged sequences were mapped to the reconstructed human mtDNA consensus sequence (RSRS)[52] and the human reference genome version hg19 using the samse command in BWA v.0.7.15-r1140[53] 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 Picard MarkDuplicates tool of the Broad Institute (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 fragments[45] 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[48], and point estimates of X chromosome contamination (in males) below 3%[54]; we also used contamLD[55] to confirm low contamination rates (less than about 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 were assessed as authentic given the relatively recent dates for these individuals; all 8 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 probably 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 2 metrics and had around 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 (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, Università degli Studi della Campania Luigi Vanvitelli). The sample preparation methodology at PSU was carried out as previously reported[23]. Bone collagen was extracted and purified using a modified Longin method with ultrafiltration[56] (>30 kDa gelatin); if collagen yields were low, a modified XAD process[57] (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 preservation[56]. We also evaluated diet in these individuals (for example, marine versus terrestrial) and compared the results to reference data from 242 ancient Caribbean and Maya individuals (Supplementary Figs. 12–14). Attenuated total reflectance Fourier transform infrared (ATR-FTIR) spectra were generated to assess post mortem changes in the apatite crystal structure of the bone samples; ATR-FTIR spectra of all samples are displayed in Supplementary Fig. 15 and quality-control parameters are reported in Supplementary Table 1. Ultimately, all calibrated $^{14}$C ages were computed using OxCal v.4.4[58] using the IntCal20[59] after our stable isotope analysis detected minimal consumption of marine resources. Sample preparation at CIRCE was carried out following the laboratory-adapted Longin method[60]; 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 individuals[4] with newly generated data from 174 ancient individuals for co-analysis, retaining 89 of them for a final co-analysis dataset comprising 263 individuals; the details of the merging are in Supplementary Information section 4. We leverage these previously published data to revisit statistics and analyses reported in that work[4] (Supplementary Tables 2, 23, 29) and carry out additional analyses using these data (Supplementary Tables 3, 24–28, Supplementary Figs. 33, 34).

We merged these 263 ancient individuals that passed screening into a base dataset that included 61 previously published ancient American

individuals[17,21,61–64], and 36 modern Indigenous American groups sourced from SNP array genotyping datasets or whole-genome sequencing datasets (Supplementary Data 5): (1) '1240K SNPs', whole-genome sequencing data restricted to a canonical set of 1,233,013 SNPs[48–51,65,66]; (2) 'Human Origins dataset', 597,573 SNPs[67–69]; and (3) 'Illumina dataset' (unmasked and unadmixed individuals only), 352,432 SNPs[13].

All comparative analyses involving present-day Indigenous American populations were performed on the Illumina dataset, whereas for the set of outgroup populations of qpAdm and qpWave ('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; Supplementary Fig. 29 shows the distribution of mtDNA haplogroups, Supplementary Fig. 30 gives details of three mtDNA mutations diagnostic of a previously unobserved mtDNA haplogroup (which is a variant of C1d), and Supplementary Fig. 31 shows the distribution of Y chromosome haplogroups.

### Kinship

We assessed kinship for every pair of individuals newly reported here as well as those that we co-analyse[4] (including individuals from different sites and islands) using a previously described method[70], and we present results for first-, second-, and third- or fourth-degree ('close') relatives in Supplementary Table 5 (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 first-degree relatives, 21 pairs were second-degree relatives and 25 pairs were third-degree relatives or higher. For the data that we co-analysed[4], we identified 13 individuals who were part of 8 relationships (four second-degree and four third-degree relatives or higher). No close relatives were identified between the datasets. Distant cousins detected using IBD analysis are presented in Extended Data Table 2, Supplementary Data 13.

### Analysis of shared genomic segments

We identified ROH within our ancient dataset using the Python package hapROH version 0.1a8 (https://pypi.org/project/hapROH/). Following a previously described method[25], we used 5,008 global haplotypes from the 1000 Genomes Project haplotype panel[34] 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 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 4 length categories: 4–8 cM, 8–12 cM, 12–20 cM and >20 cM and report the total sum in these bins (Supplementary Data 12, Supplementary Fig. 21).

To estimate $N_e$ from ROH, we applied a maximum-likelihood inference framework; Supplementary Information section 7 describes the derivation of the likelihood. We fit the lengths of all genome-wide ROH in the size range of 4–20 cM, and infer the $N_e$ that maximizes the likelihood for ROH lengths observed in a set of individuals. Estimation uncertainties are obtained from the likelihood profile (95% confidence intervals 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 (Supplementary Figs. 22, 23).

We also analysed 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; Supplementary Fig. 24 provides an example of an 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 (Supplementary Information section 7).

### Conditional heterozygosity

We used popstats[69] to compute conditional heterozygosity for all clades and subclades, which we compared with contemporaneous groups from continental South America, such as from the Peruvian Middle and Late Horizon periods[71]. As previously described[72,73], we restricted the analysis to transversion SNPs ascertained in a Yoruba individual (Extended Data Fig. 5).

### PCA

We performed PCA with smartpca v.181623[74], 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 Extended Data Fig. 1. 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 nonprojected individuals to be appropriately covisualized. We projected 92 previously published ancient individuals[4,17,21] and 174 newly described ancient individuals onto the first two principal components computed using 61 individuals from 23 present-day populations (Extended Data Fig. 1b). Supplementary Data 4 provides all individuals included in PCA and the values of principal components 1 and 2 for the PCA shown in Extended Data Fig. 1. For the PCA presented as Supplementary Fig. 19 (Supplementary Information section 5), we used nonrelated, nonoutlier ancient individuals from *Cuba_Archaic, *Venezuela_Ceramic, *EasternGreaterAntilles_Ceramic, *BahamasCuba_Ceramic and *SECoastDR_Ceramic with >500,000 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. Supplementary Figures 16, 17, 18 and 19 show PCA by archaeological site, nonzoomed PCA, PCA excluding CpG sites and PCA with axes computed using ancient individuals, respectively.

### Unsupervised analysis of population structure

We used the software ADMIXTURE v.1.3.0[75,76] to perform unsupervised structure analysis on a dataset comprising autosomal SNPs that overlap between the 1240k and Illumina datasets, and pruned in PLINK1.9[77] 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 lowest cross-validation errors (Supplementary Table 4). 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 6 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 Supplementary Fig. 20 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 s.e. via block-jackknife using smartpca v.181623 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 (as discussed in 'Genetic structure of the pre-contact Caribbean'); 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 analysed are second- or third-degree relatives of at least one other individual from the site (Extended Data Fig. 2).

### Clade grouping framework with qpWave, TreeMix and $f_4$-statistics

We used a multistep 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 (as detailed in Supplementary Information section 8). We started by using qpWave to identify major clades on the basis of shared ancestry, and then used TreeMix and $f_4$-statistics to investigate the existence of subclades. Once all subclades were identified, we used $f_4$-statistics to investigate further substructure between sites within each clade. Geographical 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 example[78].

The software qpWave[13] from ADMIXTOOLS v.6.0[69] 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 whether 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 dataset[68] plus Yukpa[65], 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 and 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 analysis[79]. The 'allsnps: YES' option was developed to increase the number of SNPs analysed in cases in which very little SNP overlap exists between all populations included in a qpWave model[80]. Although it is commonly used when low-coverage data results in the loss of the majority of sites in the initial datasets[79], there is a risk that this option introduces unreliability in the analysis, particularly in cases in which 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) (Supplementary Fig. 25) or clades, and then reassessed the relationships between members within those clades by running the same tests in a 'model competition' approach in which individuals from other sites from within the same clade were added to the right set (step 2) (Supplementary Fig. 26). 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 populations[13] (Supplementary Data 5) to identify within-clade site structure (step 3) (Supplementary Figs. 27, 28) by generating a maximum-likelihood tree. We excluded four Chibchan, Chocoan and Arawak-speaking populations that are possibly admixed with each other from this analysis. We ran TreeMix, grouping the SNPs in windows of 500 (-k 500) to account for linkage disequilibrium, setting Chipewyan as root (-root), allowing random migration (admixture) events (-m) and disabling sample size correction (-noss) to include sites or populations represented by a single individual. We note that single-individual populations present artefactually long branches that do not truly represent population-specific drift. By running TreeMix and allowing consecutive random admixture events, we identified nodes and branches that maintained the same ancient Caribbean sites among the different runs.

We used $f_4$-statistics to evaluate whether sets of sites formed a subclade to the exclusion of the other sites by following the structure of the tree. For each identified intact node among all TreeMix runs, we used each downstream pair of site(s) as test 1 and test 2, and investigated their relationship to upstream sites or pools of sites (step 4). If an upstream node was statistically consistent with all tests, 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: $f_4$(Mbuti, pool; test 1; test 2); $f_4$(Mbuti, test 1; pool, test 2); and $f_4$(Mbuti, test 2; test 1, pool). With test 1 and test 2 expected to be closer to each other than to the pool, the tested relationship finds support if the first test is statistically nonsignificant and at least one of the other two are significant. We used a $Z$-score threshold of 2.8 (associated with a 99.5% confidence interval) to assess significance. These sites were then merged into a subclade inside the major Ceramic clade for further analysis. We did not include the sites of Cueva del Perico I, Los Indios, Punta Candelero and Tibes in the TreeMix and $f_4$-statistics owing to reduced coverage, but evaluated these sites separately to see whether they shared closer affinities to any subclades relative to the others (Supplementary Data 7, Supplementary Information section 8).

After this clade analysis, we used $f_4$-statistics to further investigate potential substructure between sites within each subclade (step 5). For each pairwise site comparison, we randomly divided each site into two groups of individuals, and used a statistic of the form $f_4$(site 1 subset 1, site 2 subset 1; site 1 subset 2, site 2 subset 2) to identify positive statistics suggesting substructure within the same clade. This randomization step was repeated ten times, and the average $Z$-score was calculated. If a site was composed of a single individual, we instead computed statistics of the form $f_4$(Mbuti, site 1 subset 1; site 2 single individual, site 1 subset 2), intended to evaluate whether individuals within site 1 were closer to each other than to the single individual from site 2. No statistics were computed if both sites being tested contained only one individual.

We also used $f_4$-statistics to test whether any specific subclade within the *Caribbean_Ceramic clade had more Archaic-related ancestry than another. Specifically, we used the statistic $f_4$(Mbuti, Greater Antilles Archaic, subclade 1, subclade 2) and interpreted results as significant on the basis of a $|Z| > 2.8$; results are presented in Supplementary Table 20.

### qpAdm

We used qpAdm[50] from ADMIXTOOLS v.6.0[67] with 'allsnps: NO' to identify the most likely sources of ancestry and admixture for our populations or clades. First, we investigated whether the possible outliers *SECoastDR_Ceramic16539, *SECoastDR_Ceramic16520 and *EasternGreaterAntilles_Ceramic7969, as well as the individuals comprising the subclades *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 from Cuba and I10126), *Caribbean_Ceramic (composed of *BahamasCuba_Ceramic, *EasternGreaterAntilles_Ceramic and *SECoastDR_Ceramic, as well as *LesserAntilles_Ceramic, where relevant) and *Venezuela_Ceramic (Supplementary Tables 9, 10, 12–15). We used this information to complete Fig. 1b. We also used qpAdm to evaluate the presence of Archaic-related ancestry in *Caribbean_Ceramic. Then, on the basis of this admixture information, we attempted to obtain more detailed admixture models using the subclades from within *Caribbean_Ceramic and *GreaterAntilles_Archaic as possible sources. Finally, we attempted to identify more distal sources of ancestry by using previously published ancient individuals from the Americas[61–64], in this case for three major clades or groups of qpWave. The base right set used was the same as used for qpWave. We also tested all one-, two- and three-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 and unadmixed populations from the Illumina dataset.

### qpGraph

We used qpGraph and an edited skeleton tree of previously published ancient American populations[64] to construct an admixture graph 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 subclades 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 speakers of Carib languages), *Venezuela_Ceramic (inhabitants of a possible region of origin for this ancient Carib migration), and also *LesserAntilles_Ceramic (possibly representing Island Carib populations), and then assessed at what admixture threshold we were able to reliably detect the latter ancestry type (Supplementary Information section 13, Supplementary Fig. 32). 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 *BahamasCuba_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 subclade, 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 distribution of ancestry tracts of evolutionary signals (DATES)[23] v.3520 (M. Chintalapati et al., manuscript in preparation) method 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 Supplementary Table 22.

### 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 (R. Maier et al., manuscript in preparation) through the statistic $f_4$(European, test; *Cuba_Archaic, *Caribbean_Ceramic). To evaluate statistical power, we compared results for present-day Cuban individuals 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

Analysing 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[81] 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 two 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, and 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[65,69] ('population Y' signal), using the statistic $f_4$(Mbuti, Onge/Papuan; Mixe, ancient clade or subclade) and testing all final subclades in the position of ancient clade or subclade. Here, Mixe is representative of a population that harbours 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-associated individual I10126 from the site of Andrés (Dominican Republic) was significant at $Z = 3.4$. Although 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 nonsignificant 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 $f_4$(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 previous studies that also have not found this signal in ancient individuals from this region[17] and other areas of South America[64].

### Reporting summary

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

## Data availability

The aligned sequences are available through the European Nucleotide Archive under accession number 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.

## Code availability

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


39. Pinhasi, R. et al. Optimal ancient DNA yields from the inner ear part of the human petrous Bone. *PLoS ONE* **10**, e0129102 (2015).
40. Pinhasi, R., Fernandes, D. M., Sirak, K. & Cheronet, O. Isolating the human cochlea to generate bone powder for ancient DNA analysis. *Nat. Protocols* **14**, 1194–1205 (2019).
41. Sirak, K. et al. Human auditory ossicles as an alternative optimal source of ancient DNA. *Genome Res.* **30**, 427–436 (2020).

42. Dabney, J. et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl Acad. Sci. USA* **110**, 15758–15763 (2013).
43. Korlević, P. et al. Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* **59**, 87–93 (2015).
44. 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. Protocols* **13**, 2447–2461 (2018).
45. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Phil. Trans. R. Soc. Lond. B* **370**, 20130624 (2015).
46. Gansauge, M.-T., Aximu-Petri, A., Nagel, S. & 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. *Nat. Protocols* **15**, 2279–2300 (2020).
47. Briggs, A. W. et al. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res.* **38**, e87 (2010).
48. Fu, Q. et al. A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23**, 553–559 (2013).
49. Fu, Q. et al. An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216–219 (2015).
50. Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
51. Mathieson, I. et al. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).
52. Behar, D. M. et al. A "Copernican" reassessment of the human mitochondrial DNA tree from its root. *Am. J. Hum. Genet.* **90**, 675–684 (2012).
53. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
54. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* **15**, 356 (2014).
55. Nakatsuka, N. et al. ContamLD: estimation of ancient nuclear DNA contamination using breakdown of linkage disequilibrium. *Genome Biol.* **21**, 199 (2020).
56. Kennett, D. J. et al. Archaeogenomic evidence reveals prehistoric matrilineal dynasty. *Nat. Commun.* **8**, 14115 (2017).
57. Lohse, J. C., Culleton, B. J., Black, S. L. & Kennett, D. J. A precise chronology of Middle to Late Holocene bison exploitation in the far southern Great Plains. *J. Texas Arch. Hist.* **1**, 94–126 (2014).
58. Bronk Ramsey, C. Bayesian analysis of radiocarbon dates. *Radiocarbon* **51**, 337–360 (2009).
59. Reimer, P. J. et al. The IntCal20 northern hemisphere radiocarbon age calibration curve (0–55 cal kBP). *Radiocarbon* **62**, 725–757 (2020).
60. Passariello, I. et al. Characterization of different chemical procedures for $^{14}C$ dating of buried, cremated, and modern bone samples at Circe. *Radiocarbon* **54**, 867–877 (2012).
61. Lindo, J. et al. The genetic prehistory of the Andean highlands 7000 years BP though European contact. *Sci. Adv.* **4**, eaau4921 (2018).
62. Moreno-Mayar, J. V. et al. Early human dispersals within the Americas. *Science* **362**, eaav2621 (2018).
63. Scheib, C. L. et al. Ancient human parallel lineages within North America contributed to a coastal expansion. *Science* **360**, 1024–1027 (2018).
64. Posth, C. et al. Reconstructing the deep population history of Central and South America. *Cell* **175**, 1185–1197.e22 (2018).
65. Raghavan, M. et al. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* **349**, aab3884 (2015).
66. Mallick, S. et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016).
67. Patterson, N. et al. Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
68. Lazaridis, I. et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014).
69. Skoglund, P. et al. Genetic evidence for two founding populations of the Americas. *Nature* **525**, 104–108 (2015).
70. Olalde, I. et al. The genomic history of the Iberian peninsula over the past 8000 years. *Science* **363**, 1230–1234 (2019).
71. Nakatsuka, N. et al. A paleogenomic reconstruction of the deep population history of the Andes. *Cell* **181**, 1131–1145.e21 (2020).
72. Skoglund, P. et al. Genomic insights into the peopling of the southwest Pacific. *Nature* **538**, 510–513 (2016).
73. Harney, É. et al. Ancient DNA from Chalcolithic Israel reveals the role of population mixture in cultural transformation. *Nat. Commun.* **9**, 3336 (2018).
74. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
75. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
76. Alexander, D. H. & Lange, K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* **12**, 246 (2011).
77. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
78. Fu, Q. et al. The genetic history of Ice Age Europe. *Nature* **534**, 200–205 (2016).
79. Lipson, M. Applying $f_4$-statistics and admixture graphs: theory and examples. *Mol. Ecol. Resour.* **00**, 1–10 (2020).
80. Harney, É., Patterson, N., Reich, D. & Wakeley, J. Assessing the performance of qpAdm: a statistical tool for studying population admixture. Preprint at https://doi.org/10.1101/2020.04.09.032664 (2020).
81. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).



**Acknowledgements** We acknowledge the ancient people who were the source of the skeletal material analysed 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 analysed. This work was supported by a grant from the National Geographic Society to M. Pateman to facilitate analysis of skeletal material from The Bahamas and by a grant from the Italian 'Ministry of Foreign Affairs and International Cooperation' (Italian archaeological, anthropological and ethnological missions abroad, DGPSP Ufficio VI). 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 J. Avilés, J. Acayaguana Delvalle, J. Estevez, D. T. Golding Frankson, J. Gregory, L. A. Guitar, L. Kelly, G. A. Lopez Castellano, K. R. Nibonri and O. Patterson for comments on early versions of this manuscript and discussions that improved the presentation of this work; V. A. Forbes-Pateman and N. Albury for their assistance compiling descriptions for archaeological sites in The Bahamas; E. Harney, R. Maier and N. Nakatsuka for help with data processing; and M. Chintalapati, P. Moorjani and N. Patterson for advice on analysis. We dedicate this article to the memory of F. Luna Calderon, who would have been a co-author had he not passed away in the course of the work for this study.


**Author contributions** W.F.K., A. Coppa, M. Lipson, R.P. and D.R. supervised the study. J.S., O.C., C.A.A., E.V.C., R.C., A. Cucina, F.G., C.K., F.L.P., M. Lucci, M.V.M., C.T.M., C.M., I.P., M.P., T.M.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.M.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.T.Ö., C.S., R.S., K.S. 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.M.F., K.A.S., H.R., M.M., S.M., I.O. and M. Lipson analysed genetic data. D.M.F., K.A.S., 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**

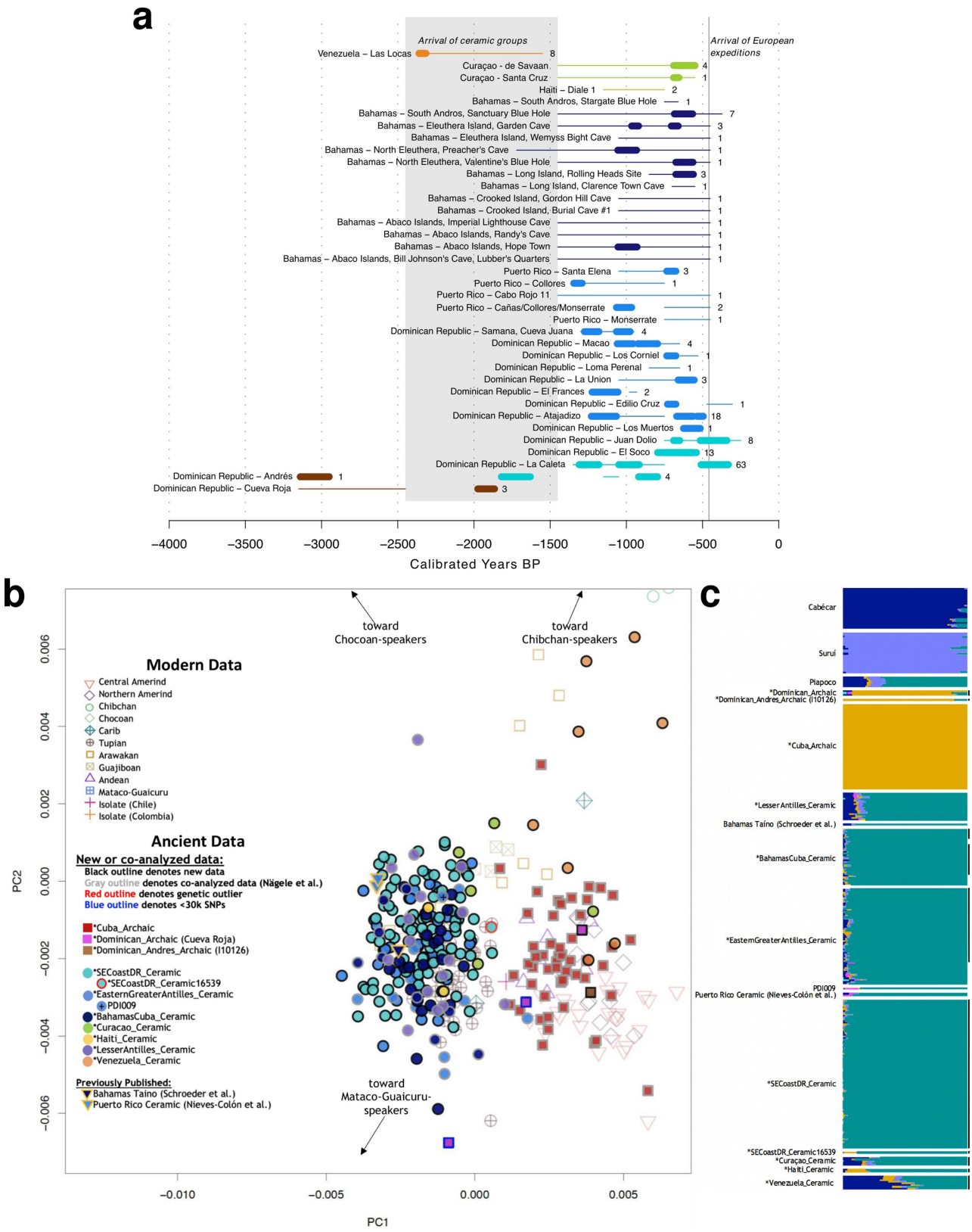

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41586-020-03053-2.



**Correspondence and requests for materials** should be addressed to A.C., R.P. or D.R.



**Peer review information** *Nature* thanks John Lindo, Alice Samson and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

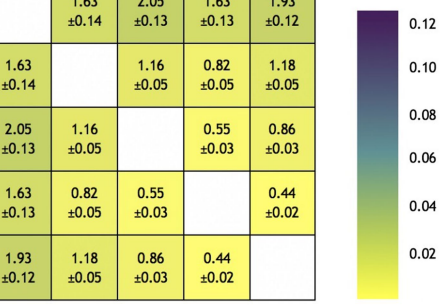



**Extended Data Fig. 1** | See next page for caption.

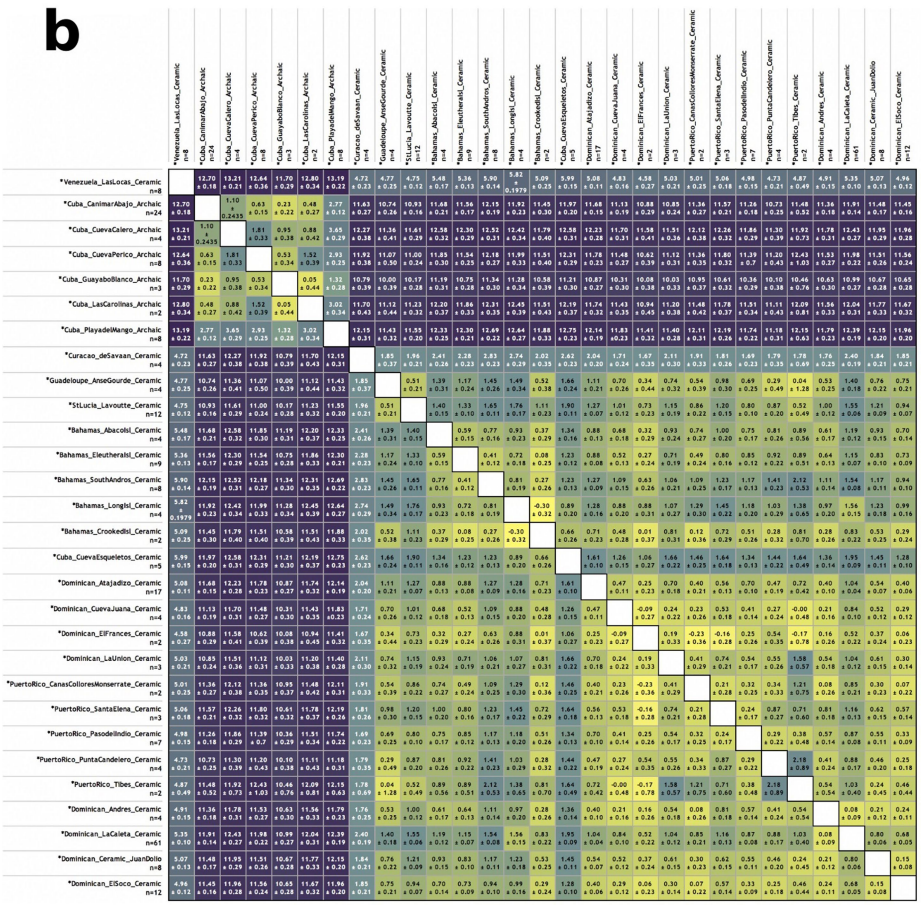
**Extended Data Fig. 1 | Temporal distribution of newly reported individuals and overview of population structure. a**, Numbers represent individuals from each site; thick lines denote direct $^{14}$C dates (95.4% calibrated confidence intervals); thin lines denote archaeological-context dating; grey area identifies the first arrivals of ceramic users in the Caribbean. Colours and labels are consistent with Fig. 1. **b**, PCA plot with ancient individuals shown as solid squares or circles for 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 subclades, and three individuals from the site of Cueva Roja (Dominican Republic) who were excluded from clading analysis analysis are labelled *Dominican_Archaic (Cueva Roja) and coloured magenta. Individual PDI009, previously assessed elsewhere as an outlier[11], is denoted with a cross. Three previously published ancient Caribbean individuals[9,10] are shown as inverted triangles outlined in grey and coloured for the subclade that encompasses the geographical 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 Supplementary Fig. 17. **c**, ADMIXTURE analysis best supports $K = 6$ ancestral elements. Newly reported and co-analysed individuals are clustered by subclade; 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 **b** are included, and three present-day populations (Suruí, Cabécar and Piapoco) are shown for reference.

**a**

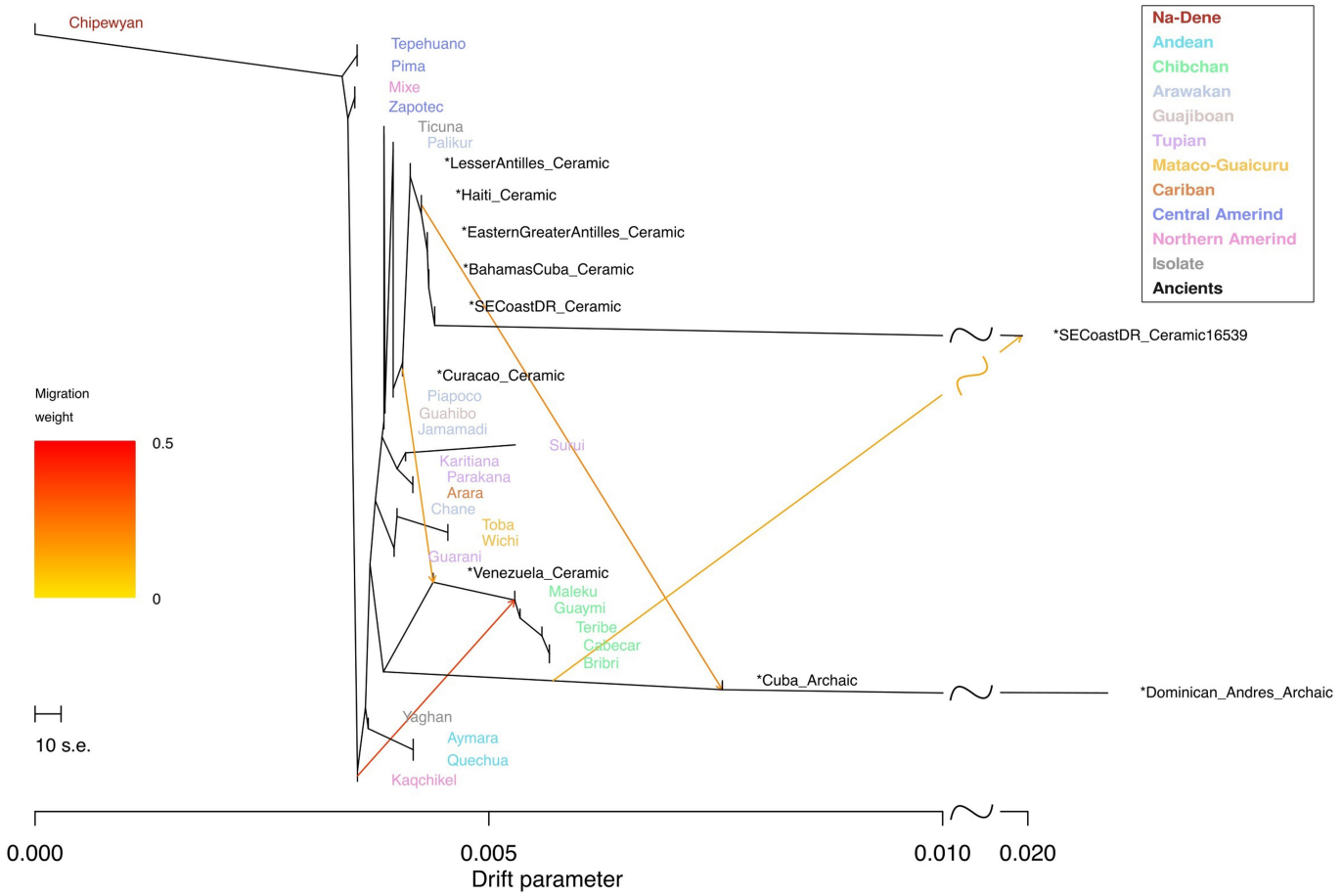
| | *Venezuela_Ceramic n=8 | *Cuba_Archaic n=50 | *Curacao_Ceramic n=5 | *LesserAntilles_Ceramic n=16 | *BahamasCuba_Ceramic n=33 | *EasternGreaterAntilles_Ceramic n=52 | *SECoastDR_Ceramic n=85 |
|---|---|---|---|---|---|---|---|
| *Venezuela_Ceramic n=8 | | 12.45 ±0.17 | 4.58 ±0.17 | 4.70 ±0.12 | 5.31 ±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 | |



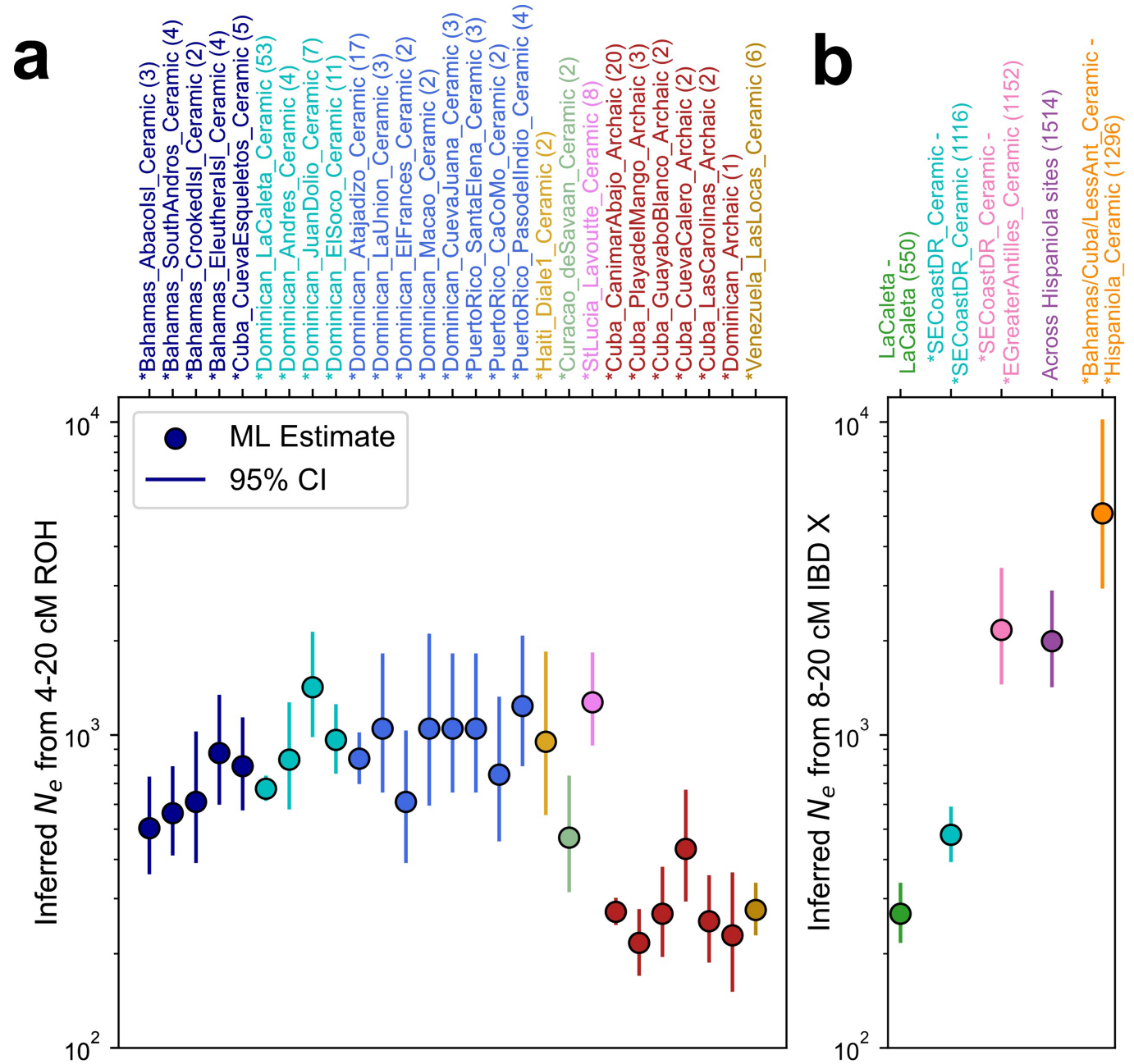

**b**



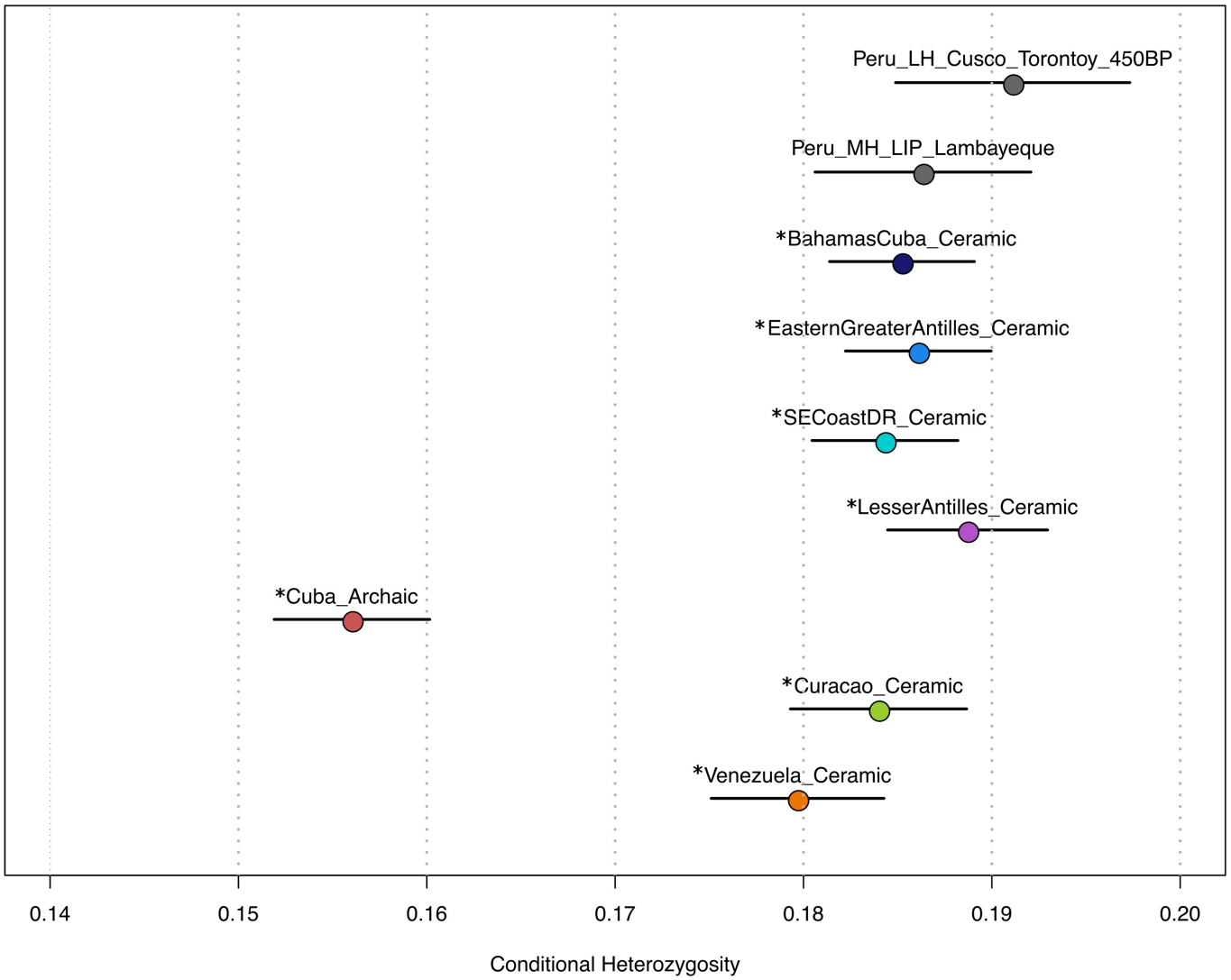

**Extended Data Fig. 2 | $F_{ST}$ distances. a**, **b**, Average pairwise $F_{ST}$ distances and standard errors (×100) between clades (**a**) and sites with more than two unrelated individuals (**b**), demonstrating both overall high levels of genetic similarity between the *Caribbean_Ceramic subclades and the sites composing them, as well as the magnitude of genetic differentiation between those and the groups with Archaic- and Venezuela-related ancestries.



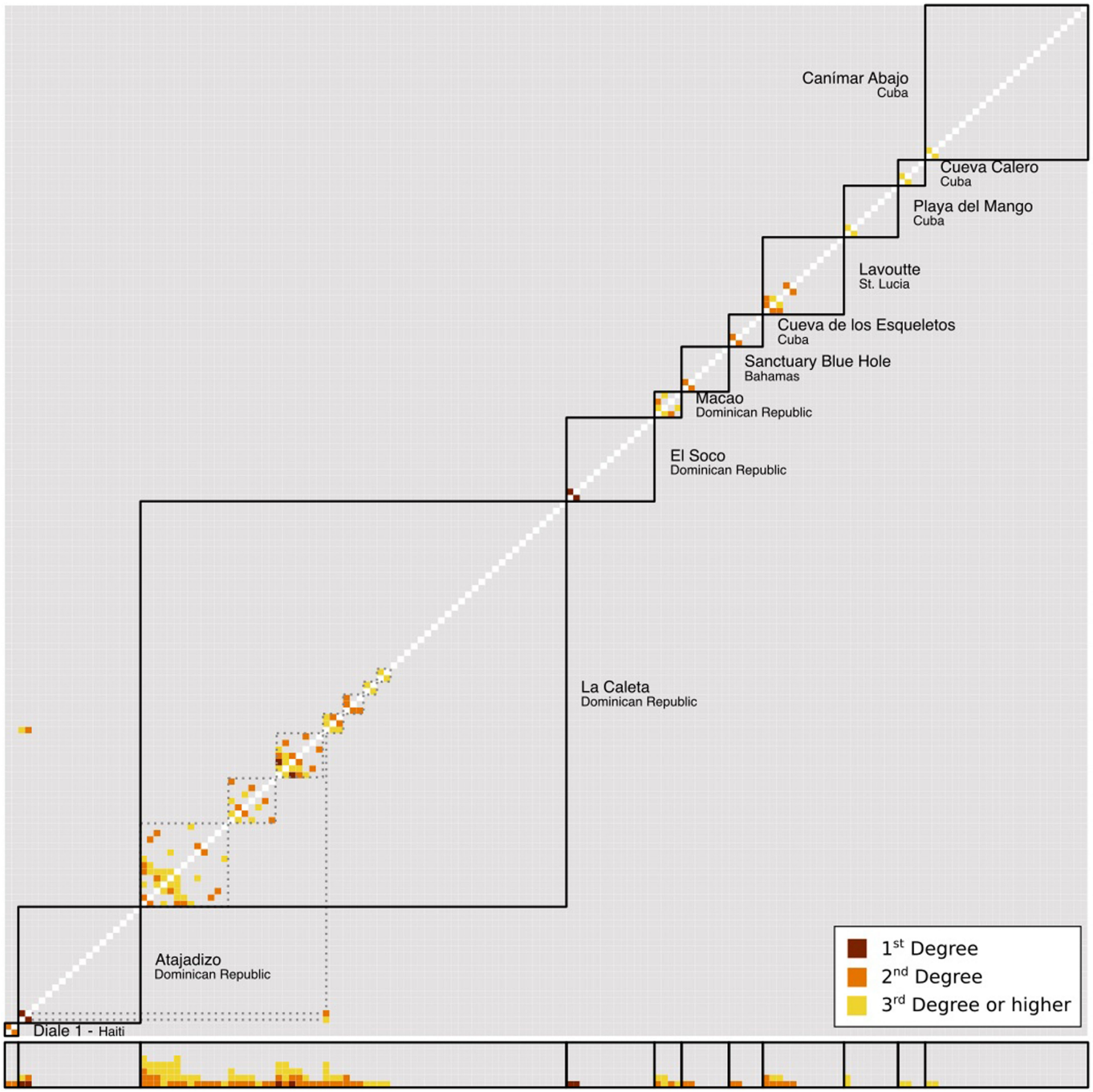

**Extended Data Fig. 3 | Maximum-likelihood population tree from allele frequencies using Treemix.** The *Caribbean_Ceramic subclades are inferred to be on the same branch as modern Arawak-speaking groups (Palikur and Jamamadi). Orange arrows represent admixture events, although observations from other analyses (for example, qpAdm admixture modelling) suggest that the indicated direction of admixture may be inaccurate (for example, 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). Colours as per subclades; numbers denote the count of analysed individuals. Highly consanguineous individuals with a sum of ROH > 20 above 50 cM were excluded. **b**, As in **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**, **b**, points represent maximum-likelihood estimate and vertical bars represent 95% confidence interval.




**Extended Data Fig. 5 | Conditional heterozygosity by clade.** Conditional heterozygosity in the ancient Caribbean was similar to that of contemporaneous groups from Peru[71], 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 who represent *Haiti_Ceramic. Coloured circles represent point estimates (colour scheme matching Fig. 1); bars represent three s.e.




**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 intersite relationships; bottom rows correspond to relationships per individual.

**Extended Data Table 1 | $N_e$ values for each site**

| $N_e$ 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 |

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

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

| ID1 | ID2 | Evidence | Site 1 | Site 2 |
|---|---|---|---|---|
| I13320 | I15973 | X chromosome IBD segment of 10.0 cM | Bahamas, Abaco Island | Dominican Republic, La Caleta |
| I13318 | PDI010 | X chromosome IBD segment of 14.0 cM | Bahamas, Crooked Island | Puerto Rico, Vega Baja, Paso del Indio |
| I13321 | I12344 | X chromosome IBD segment of 12.7 cM | Bahamas, Eleuthera Island | Dominican Republic, El Soco |
| I13321 | I13196 | X chromosome IBD segment of 10.7 cM | Bahamas, Eleuthera Island | Dominican Republic, Juan Dolio |
| I13321 | I13326 | X chromosome IBD segment of 12.0 cM | Bahamas, Eleuthera Island | Puerto Rico, Monserrate |
| I13737 | CDE001 | X chromosome IBD segment of 10.7 cM | Bahamas, Long Island, Clarence Town, Rolling Heads Site | Cuba, Camagüey, Sierra de Cubitas, Cueva de los Esqueletos 1 |
| I14880 | I12344 | X chromosome IBD segment of 8.7 cM | Bahamas, South Andros, Sanctuary Blue Hole | Dominican Republic, El Soco |
| I14879 | I15963 | X chromosome IBD segment of 10.0 cM | Bahamas, South Andros, Sanctuary Blue Hole | Dominican Republic, La Caleta |
| I8549 | I14879 | X chromosome IBD segment of 10.0 cM | Dominican Republic, Andres | Bahamas, South Andros, Sanctuary Blue Hole |
| I17903 | I14875 | X chromosome IBD segment of 14.7 cM | Dominican Republic, Atajadizo | Bahamas, Abaco, Bill Johnson's Cave, Lubber's Quarters |
| I13441 | I14880 | X chromosome IBD segment of 10.7 cM | Puerto Rico, Cabo Rojo 11 | Bahamas, South Andros, Sanctuary Blue Hole |
| I13441 | I13189 | X chromosome IBD segment of 10.0 cM | Puerto Rico, Cabo Rojo 11 | Dominican Republic, El Soco |
| I13441 | I15676 | X chromosome IBD segment of 10.0 cM | Puerto Rico, Cabo Rojo 11 | Dominican Republic, La Caleta |
| I13441 | I14992 | X chromosome IBD segment of 9.3 cM | Puerto Rico, Cabo Rojo 11 | Dominican Republic, Los Muertos |
| I13326 | I12344 | X chromosome IBD segment of 11.3 cM | Puerto Rico, Monserrate | Dominican Republic, El Soco |
| PDI012013 | I15963 | X chromosome IBD segment of 9.3 cM | Puerto Rico, Vega Baja, Paso del Indio | Dominican Republic, La Caleta |
| I13318 | I14880 | X chromosome IBD segment of 22.7 cM | Bahamas, Crooked Island | Bahamas, South Andros, Sanctuary Blue Hole |
| I13318 | I14879 | X chromosome IBD segment of 10.0 cM | Bahamas, Crooked Island | Bahamas, South Andros, Sanctuary Blue Hole |
| I13321 | I13320 | X chromosome IBD segment of 12.0 cM | Bahamas, Eleuthera Island | Bahamas, Abaco |

We measured the X chromosome length and IBD map lengths as two-thirds of the map length of female X. A complete table, including cross-site distant relatives within islands, can be found in Supplementary Data 13.

**Extended Data Table 3 | Ancestry proportion estimates using qpAdm for present-day Caribbean individuals from Cuba (and its provinces), Dominican Republic and Puerto Rico**

| 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 |

Data are from refs. [22,29]. Top half, proportions across countries. 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 | $f_4$(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 | $f_4$(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 |