

## A genetic history of the pre-contact Caribbean

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

<sup>1</sup> Department of Evolutionary Anthropology, University of Vienna, 1090 Vienna, Austria.

<sup>2</sup> CIAS, Department of Life Sciences, University of Coimbra, 3000-456 Coimbra, Portugal.

<sup>3</sup> Department of Genetics, Harvard Medical School, Boston, MA 02115, USA.

<sup>4</sup> Department of Human Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA.

<sup>5</sup> Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA.

<sup>6</sup> Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA.

<sup>7</sup> Institute of Evolutionary Biology, CSIC-Universitat Pompeu Fabra, 08003 Barcelona, Spain.

<sup>8</sup> Institutes of Energy and the Environment, The Pennsylvania State University, University Park, PA 16802, USA.

<sup>9</sup> Department of Legal Medicine, Toxicology and Physical Anthropology, University of Granada, Granada, Spain.

<sup>10</sup> Superintendency of Archaeology, Fine Arts and Landscape for the city of Cagliari and the provinces of Oristano and South Sardinia, Cagliari, Italy.

<sup>11</sup> Department of Anthropology, The Pennsylvania State University, University Park, PA, 16802, USA.

<sup>12</sup> Department of Mathematics and Physics, Campania University "Luigi Vanvitelli", 81100 Caserta, Italy.

<sup>13</sup> CIBIO - InBIO, University of Porto, Campus de Vairão, 4485-661 Vairão, Portugal.

<sup>14</sup> Museo Antropológico Montané, University of Havana, Havana, Cuba.

<sup>15</sup> Matanzas University of Medical Sciences, Carretera de Quintanilla Km 101, 40100 Matanzas, Cuba.

<sup>16</sup> Peabody Museum of Natural History, Yale University, New Haven, CT 06511, USA.

<sup>17</sup> Facultad de Ciencias Antropológicas, Universidad Autónoma de Yucatán, Mérida, Yucatán, Mexico.

<sup>18</sup> Department of Environmental Biology, Sapienza University of Rome, Square Aldo Moro, 5, Rome 00185, Italy.

<sup>19</sup> National Archaeological-Anthropological Memory Management (NAAM), 13 Johan van Walbeeckplein, Willemstad, Curaçao.

<sup>20</sup> DANTE Laboratory of Diet and Ancient Technology, Sapienza University of Rome, Rome, Italy.

<sup>21</sup> Universidad Autónoma de Santo Domingo, Sanchez, San Francisco de Macorís 31000, Dominican Republic.

<sup>22</sup> National Center of Medical Genetics, Medical University of Havana, Havana, Cuba.

<sup>23</sup> Museo del Hombre Dominicano, Plaza de la Cultura Juan Pablo Duarte, Av. Pedro Henríquez Ureña, Santo Domingo 10204, Dominican Republic.

<sup>24</sup> Instituto de Investigaciones Bioantropológicas y Arqueológicas, Universidad de Los Andes, Mérida, Venezuela.

<sup>25</sup> Turks and Caicos National Museum Foundation, Front St, Cockburn Town TKCA 1ZZ, Turks and Caicos Islands.

<sup>26</sup> Department of Biology, University of The Bahamas, P.O. Box N-4912, Nassau, The Bahamas.

<sup>27</sup> National Geographic Society, Washington, DC 20036, USA.

<sup>28</sup> Department of Anthropology, University of California, Santa Barbara, CA 93106, USA.

<sup>29</sup> Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA.

<sup>30</sup> Present address: Department of Anthropology, University of California, Santa Cruz, CA 95064, USA.

<sup>31</sup> Present address: Department of Biomolecular Engineering, University of California, Santa Cruz, CA 95064, USA.

<sup>32</sup> These authors jointly supervised this work: William F. Keegan, Alfredo Coppa, Mark Lipson, Ron Pinhasi, David Reich.

\*These authors contributed equally to this work.

Humans settled the Caribbean ~6,000 years ago, with intensified agriculture and ceramic use marking a shift from the Archaic Age to the Ceramic Age ~2,500 years ago. To shed new light on the history of Caribbean people, we report genome-wide data from 184 individuals predating European contact from The Bahamas, Cuba, Hispaniola, Puerto Rico, Curaçao, and northwestern Venezuela. A largely homogeneous ceramic-using population most likely originating in northeastern South America and related to present-day Arawak-speaking groups moved throughout the Caribbean at least 1,800 years ago, spreading ancestry that is still detected in parts of the region today. These people eventually almost entirely replaced Archaic-related lineages in Hispaniola but not in northwestern Cuba, where unadmixed Archaic-related ancestry persisted into the last millennium. We document high mobility and inter-island connectivity throughout the Ceramic Age as reflected in relatives buried ~75 kilometers apart in Hispaniola and low genetic differentiation across many Caribbean islands, albeit with subtle population structure distinguishing the Bahamian islands we studied from the rest of the Caribbean and from each other, and long-term population continuity in southeastern coastal Hispaniola differentiating this region from the rest of the island. Ceramic-associated people avoided close kin unions despite limited mate pools reflecting low effective population sizes ( $2N_e=1000-2000$ ) even at sites on the large Caribbean islands. While census population sizes can be an order of magnitude larger than effective population sizes, pan-Caribbean population size estimates of hundreds of thousands are likely too large. Transitions in pottery styles show no evidence of being driven by waves of migration of new people from mainland South America; instead, they more likely reflect the spread of ideas and people within an interconnected Caribbean world.

Prior to European colonization, the Caribbean islands were a mosaic of archaeologically-distinct cultures reflecting intricate networks of interaction and multiple instances of cultural change since the first human occupation ~6,000 years ago. There is debate about the extent to which these changes reflect local developments and intra-Caribbean interactions or migrations from the American continents<sup>1,2</sup>, and the connections between demographic movements and three distinct archaeological periods, classified as the Lithic, Archaic, and Ceramic Ages remain unclear<sup>3</sup> (Supplementary Information section 1).

The Lithic Age represents the earliest archaeological evidence of human occupation in the Caribbean (~6,000 years ago) and is found in Cuba, Hispaniola, and Puerto Rico. The Archaic Age is defined by the appearance of ground-stone artifacts as early as 5,000 years ago, and has been hypothesized to reflect a second spread of technology and possibly people from South America<sup>4,5</sup>. The Ceramic Age, beginning 2,500-2,300 years ago, is characterized by an agricultural economy and intensive pottery production; it is widely accepted as reflecting at least one migration of people from the mouth of the Orinoco River in Venezuela and the Guianas who spoke a language related to present-day Arawak languages. There are debates about the extent to which distinct ceramic styles ('series') correlate to the movements of people during the Ceramic Age<sup>1-3</sup> (Supplementary Information section 1), as well as about pan-Caribbean population size during this time<sup>6-8</sup>.

Ancient DNA data from the Caribbean to date comes from mitochondrial DNA<sup>9-13</sup> as well as genome-wide data from a 1,000 year-old individual from The Bahamas<sup>14</sup> and two ancient individuals from Puerto Rico<sup>13</sup>. We screened 208 individuals for evidence of authentic ancient DNA (Supplementary Data 1) and generated genome-wide data passing standard criteria for authenticity for 184 unique ancient individuals who lived between ~3150-300 calibrated years (cal. yr) before the present (BP, taken as 1950 CE in accordance with radiocarbon calibration convention) in the Bahamas, Cuba, Hispaniola (which we separate into Haiti and the Dominican Republic in our analyses using present-day borders for higher geographic resolution), Puerto Rico, Curaçao, and northwestern Venezuela (Fig. 1; Supplementary Information section 2). We leverage these data and 52 new direct radiocarbon dates (Supplementary Information section 3; Supplementary Data 3) to address key debates about population history in the pre-contact Caribbean. We summarize these debates and how they are advanced by new genetic data in Table 1. In what follows, we use 'Archaic' to denote sites with an abundance of stone tools and/or site dates predating the spread of intensive ceramic use, and 'Ceramic' to denote sites with a preponderance of ceramics.

**Table 1.** Archaeological debates addressed by genetic data generated in the present study.

Key debates from archaeology	New genetic data exploring this debate
Geographic origin, timing, trajectory of Archaic Age migration into the Caribbean; number of migrations during the Archaic Age	Genome-wide data from 16 Archaic-associated individuals from Cuba and the Dominican Republic suggest that there was a homogenous Archaic ancestry profile throughout the Greater Antilles. This ancestry persisted with minimal admixture in some areas for over 2,000 years. The Archaic Age migration originated within a population that is not closely related to any present-day genotyped Indigenous groups, though we find a North American origin for this migration unlikely. We find no evidence of multiple migrations during the Archaic Age.
Geographic origin, timing, trajectory of Ceramic Age migration into the Caribbean; number of migrations during the Ceramic Age	Genome-wide data from 153 individuals across The Bahamas and Greater Antilles spanning 1,400 years point to a single wave of migration into the Caribbean in the Early Ceramic Age at least 1,800 years ago with a likely connection to the geographic region where Arawak-speaking peoples live today (Venezuela and the Guianas). This accords with the presence of Arawak languages in the pre-contact Caribbean. A caveat is that multiple waves of migration from source populations that were highly genetically similar (such as may have been widespread throughout northern South America in the past even though such populations are not widespread there today) could go undetected with ancient DNA.
The relationship between transitions of material culture styles ('series') and the movement of people during the Ceramic Age	We find a signal of genetic homogeneity during the Ceramic Age that is independent of ceramic typology. This increases the weight of evidence that changes in pottery styles throughout the Ceramic Age were not associated with multiple waves of migration of genetically-distinct peoples from the American continents, and instead suggests that cultural diffusion or movement of genetically homogenous people within the Caribbean drove stylistic change. We detect relatives ~75km apart on Hispaniola, providing direct evidence of mobility. Geography explains the subtle population structure that we do observe; for example, individuals from The Bahamas form a group relative to ceramic users from the Greater Antilles, and people from four sites on the southeast coast of Hispaniola form a group relative to people from other parts of the island.
Nature of interactions between Archaic-associated peoples of the Caribbean and ceramic users who moved into this region	We find instances of admixture between Archaic- and Ceramic-associated people in some places, and the parallel existence of Archaic- and Ceramic-associated populations with little if any admixture in others. Admixture occurred at the Ceramic-associated site of Diale 1 (Haiti), where two individuals had ~19% Archaic-related ancestry. An individual from La Caleta (Dominican Republic) had significantly more Archaic-related ancestry (12%) than other Ceramic Age people who lived in this part of Hispaniola. In contrast, unadmixed Archaic-related ancestry persisted for ~2,000 years at Canimar Abajo (Cuba).
Genetic distinction between people living in the early or late part of the Ceramic Age	Our data do not support genetic differentiation across time during the Ceramic Age. Instead, we identify long-term genetic continuity at sites such as Andrés (Dominican Republic), where a homogenous ancestry profile persists for almost 1,000 years.
Presence of pre-contact ancestry in modern Caribbean people	Our data reveal Ceramic-related, but not Archaic-related, ancestry in present-day peoples from Puerto Rico and parts of Cuba in amounts up to ~14%. We also identify a previously undocumented mtDNA haplogroup - a deep branch of C1d - that is unique to Caribbean people and is identified in a single individual from Puerto Rico from the 1000 Genomes Project, providing additional direct evidence of Indigenous ancestry in the Caribbean today.
Population size and social organization in the Caribbean prior to European colonization	We estimate that effective population sizes ranged around $2N_e=1000-2000$ people in the pre-contact Caribbean. Census population size is typically not more than ten times larger, suggesting that previous inferences of population sizes in the hundreds of thousands are overestimated. While Caribbean people avoided unions of first cousins or closer, they had a limited pool of potential mates, indicated by many unions as close as second or third cousins.
Origins of the people who colonized The Bahamas	Genome-wide data from 24 Ceramic Age individuals from five islands of The Bahamas show most genetic affinity to ceramic users from the Greater Antilles, providing evidence against their origin in Archaic populations from Cuba. We show that neither individuals from an Archaic-associated site in the western part of Cuba (Canimar Abajo), nor an Archaic Age individual from the Dominican Republic (Andrés) have genetic affinity to the ancient people of the Bahamas that we analyze.
Possible migration of Carib peoples around 800 CE	Our data point to genetic continuity throughout the Ceramic Age and provide no definitive genetic evidence of a Carib migration during this time, although we cannot rule out admixture proportions up to ~8%. Additional caveats are that <i>Venezuela_Ceramic</i> and <i>Arara</i> may be imprecise proxies for ancient Carib-related ancestry, that the limited number of ancient individuals studied here are not fully representative of the entire Ceramic Age Caribbean (where some places not studied here could have Carib-related ancestry), and that a widely-shared genetic profile throughout northern South America (possibly including the Caribs) at the time of these posited migrations makes them difficult to identify using ancient DNA data.

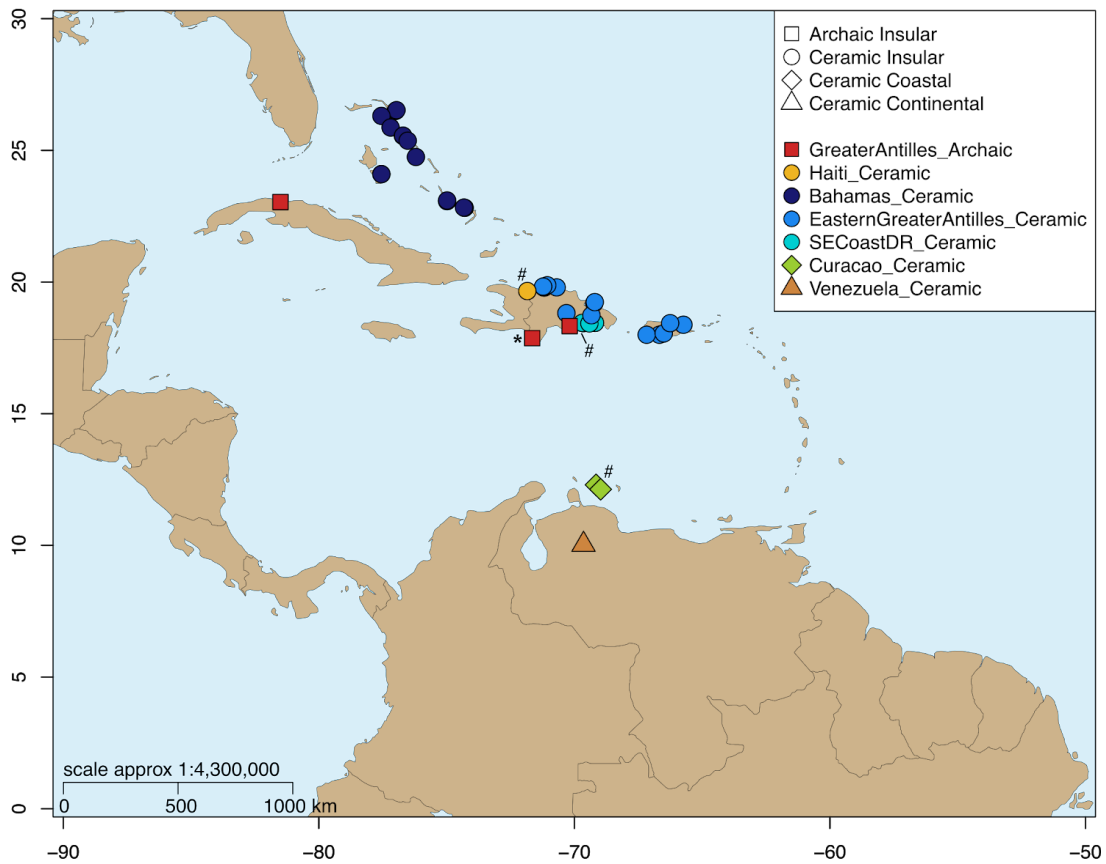
## Ethics

We acknowledge the ancient individuals whose skeletal remains we analyzed, present-day people who identify as having an Indigenous legacy, the community representatives across the Caribbean who provided critical feedback and offered local perspectives, and the Caribbean-based scholars who were centrally involved as collaborators and co-authors. Permissions to perform ancient DNA analysis of the human skeletal remains in this study were documented through authorization letters signed by a custodian who assumed responsibility for the skeletal remains collected from a specific geographic region or site. The conclusions drawn by this project are intended to provide new information about the genetic ancestry of the people of the Caribbean prior to European colonization in the late 15<sup>th</sup> century and the subsequent African slave trade, and were discussed prior to publication with members of Indigenous communities who trace their legacy to the pre-contact Caribbean; feedback from this discussion was then incorporated into our manuscript. While genetic data are one form of knowledge that contributes to understanding the past, oral traditions and other types of Indigenous knowledge can coexist with scientific data. Genetic ancestry should not be conflated with perceptions of identity, which cannot be defined by genetics alone. A full ethics statement is provided in Supplementary Information section 15.

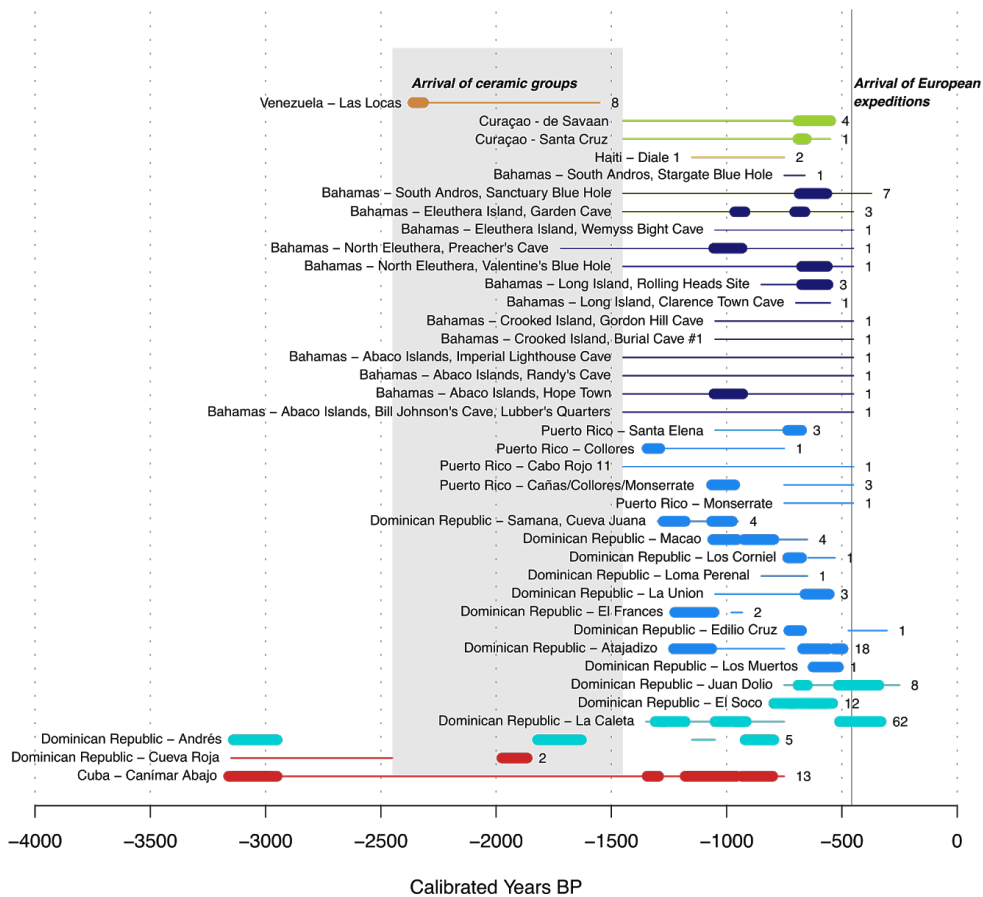
## Genetic-based clustering

We performed principal component analysis (PCA), computing axes using present-day North, Central, and South American groups with no evidence of European or African ancestry<sup>15</sup> (Fig. 2a; Supplementary Data 4; Supplementary Information section 4). Ceramic-associated individuals project in a cluster separate from Archaic-associated individuals, a distinction confirmed by clustering analysis using ADMIXTURE (Fig. 2b; Supplementary Information section 5). Archaic-associated individuals and ancient Venezuelans somewhat overlap on the PCA and are characterized by two primary components in the ADMIXTURE analysis, one maximized in modern-day Chibchan-speaking Cabécar and the other in Ceramic-associated Caribbean individuals. One exception to the genetic homogeneity observed within the majority of sites is at Andrés (Dominican Republic), where most individuals fall within the ceramic cluster (consistent with the primarily ceramic association of this site), but a single individual (I10126) is dated to the Archaic Age (3140-2950 cal. yr BP) and clusters with other Archaic-associated individuals in PCA. Individuals from Curaçao and the Diale 1 site (Haiti) are distinct from either of the two main clusters. We exclude one individual from each of the four pairs of first-degree relatives (Supplementary Information section 6) and the two Archaic-associated individuals from Cueva Roja (Dominican Republic) with low coverage ( $<0.02X$ , or  $\sim 20,000$  SNPs) from subsequent statistical analyses (Supplementary Data 1).

**a**



**b**



**Fig. 1:** Analysed sites. **(a)** Geographical distribution. Shapes represent genetic clusters to which we assigned each sample; asterisk (\*) denotes individuals from the Archaic site of Cueva Roja (Dominican Republic) excluded from main analyses due to low-coverage; hash (#) denotes presence of individuals who harbored mixtures of the main ancestry clusters. We separate sites in Hispaniola using the present-day border between Haiti and the Dominican Republic. **(b)** Temporal distribution. Numbers represent individuals from each site; thick lines denote direct C14 dates (95.4% calibrated confidence intervals); thin lines denote archaeological context dating; grey area identifies the first arrivals of ceramic-users in the Caribbean. Colors and labels are consistent with Fig. 2c.

To study genetic makeup independently from archaeologically-based material culture assignments (Supplementary Information section 2), we used a multi-step *qpWave*, Treemix, and  $f_4$ -statistics-based workflow (Supplementary Information section 7). We sequentially grouped all individuals with increasing resolution based on their degree of allele sharing, starting with the identification of major groupings ('clades'), and refining our understanding of the relationships between groups by assigning sub-groupings ('sub-clades') when they were supported. Clades and sub-clades were named following the formation of groups by combining the most specific geographic location encompassing all sites within that grouping along with 'Archaic' or 'Ceramic', based on chronology and/or the predominant artifacts at the site (Fig. 2c).

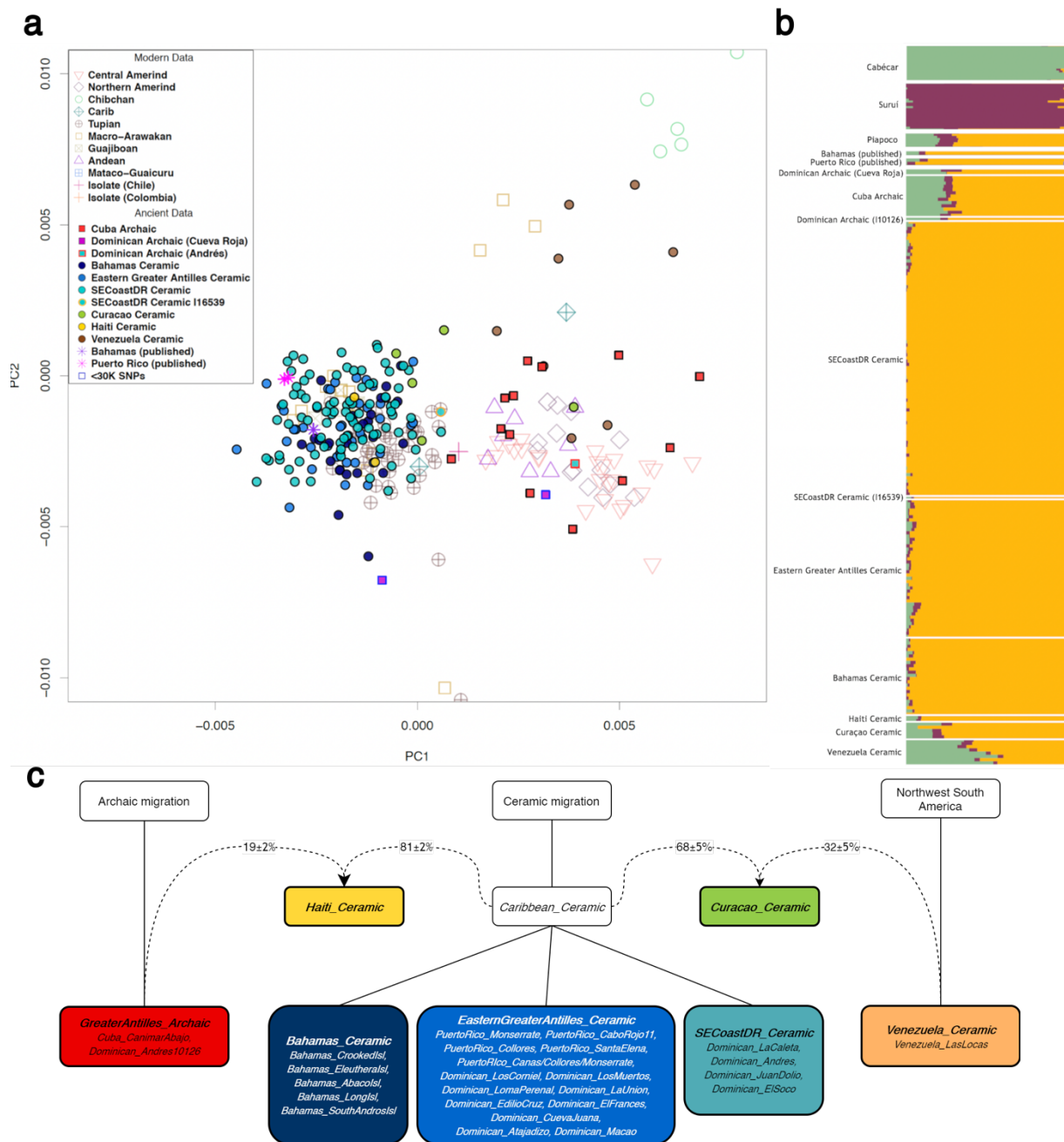
We identified three major clades that were significantly differentiated from each other by *qpWave*. *GreaterAntilles\_Archaic* included all individuals from Canimar Abajo in northwestern Cuba spanning ~3150-800 cal. yr BP and individual I10126 from Andrés; two low coverage Archaic-associated individuals from Cueva Roja (~1900 cal. yr BP) also qualitatively fit into this clade (Fig. 2a, b). *Caribbean\_Ceramic* comprised 153 individuals from The Bahamas, Dominican Republic, and Puerto Rico dating between ~1800-300 cal. yr BP. *Venezuela\_Ceramic* comprised eight individuals from the site of Las Locas, dated to ~2350 cal. yr BP. Two remaining groups, *Haiti\_Ceramic* and *Curacao\_Ceramic*, were best modeled as having mixtures of ancestry related to the major clades (described below).

We next tested for finer population structure, identifying multi-site sub-clades within which individuals were significantly more closely related to each other on average than to individuals from other sites within the same clade (Supplementary Data 5). The sub-clade *SECoastDR\_Ceramic* comprised four sites located along 50 kilometers of the southeast coast of the Dominican Republic (from west to east, La Caleta, Andrés, Juan Dolio, and El Soco) (Table S4). With radiocarbon dates spanning 1,500 years, this grouping documents a high degree of local and stable population structure over time and across changes in archaeological culture, indicating that shifts in ceramic style were not always associated with new spreads of people. All Bahamian sites (spanning ~600 years) separated into the sub-clade *Bahamas\_Ceramic*, within which further substructure was detected specific to each of the five islands included in this study (Abaco Islands, Crooked Island, Eleuthera, Long Island,



Andros Island), pointing to restricted gene flow among the islands; a possible signal of closer interactions is seen between the neighbouring Long and Crooked Islands (Table S3). The remaining sites from the *Caribbean\_Ceramic* clade did not show any specific affinities and were merged to form *EasternGreaterAntilles\_Ceramic*. We detected no between-site substructure within this sub-clade, except at Macao, where a signal of substructure is likely driven by the high proportion of related individuals analyzed. While we observe very limited differentiation in *Caribbean\_Ceramic* spanning over a millennium, we use sub-clades to investigate the local history of ceramic users. We also tested all Ceramic-associated individuals individually for an excess of Archaic-related ancestry relative to others within their sub-clade using  $f_4$ -statistics (Supplementary Information section 8; Supplementary Data 6) and identified individual I16539 from La Caleta (Dominican Republic) as showing significantly closer affinity to *GreaterAntilles\_Archaic* than the rest of the *SECoastDR\_Ceramic* sub-clade (Table S5); this individual was separated as *SECoastDR\_Ceramic16539* for further analyses.

We called Y chromosome and mitochondrial (mtDNA) haplogroups (Supplementary Data 7; Supplementary Information section 9) and found that the majority of Y chromosome haplogroups belong to the Q-M3 lineage that is characteristic of Indigenous peoples of the Americas<sup>16,17</sup>, while all four major pan-American mtDNA haplogroups (A2, B2, C1, and D1<sup>18,19</sup>) are represented in our dataset. Consistent with previous work<sup>13</sup>, mtDNA haplogroup C1b2 was the most common haplogroup in *Caribbean\_Ceramic*, and we find evidence of a previously undocumented deep branch of C1d present at a frequency of ~7% and in all sub-clades of *Caribbean\_Ceramic* (Supplementary Information section 9). The discovery of this branch—which we do not detect in any ancient or modern individuals outside the Caribbean—indicates that ancestry of pre-contact people persists in the Caribbean today. It provides new evidence that Indigenous ancestry in the present-day Caribbean cannot simply be explained as imported from mainland groups after contact, as we find a single instance of this mtDNA haplogroup in a modern Puerto Rican individual in a screen of the 1000 Genomes Project dataset<sup>20</sup> (Supplementary Information section 9).



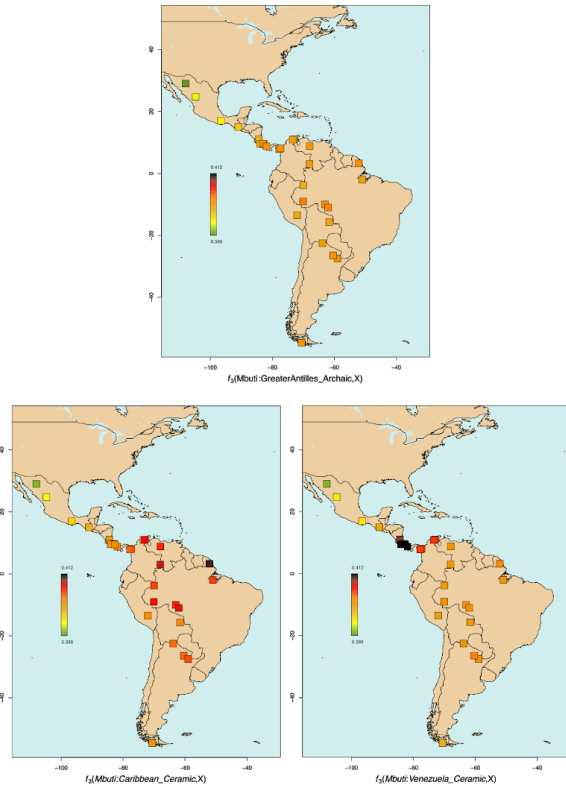
**Fig. 2: Overview of population structure.** (a) PCA of ancient individuals projected onto present-day variation. Newly reported individuals are shown as solid symbols outlined in black, red, or blue (<30K SNPs). The plot here is zoomed for focus on ancient individuals and excludes some present-day populations; full plot provided in Supplementary Information section 4. (b) Visualization of model-based ancestry analysis using ADMIXTURE at five-fold cross-validation supported K=5 ancestral elements. Three modern-day populations are included for reference (Suruí, Cabécar, Piapoco); full results are provided in Supplementary Information section 5. (c) Relationships reconstructed with *qpWave*, Treemix, and *f<sub>4</sub>*-statistics (Supplementary Information section 7). Solid lines connect sub-groupings derived from a larger group; dashed lines represent ancestry contribution to admixed groups. Colored boxes represent the final sub-groupings.

## The early peoples of Cuba and the Dominican Republic

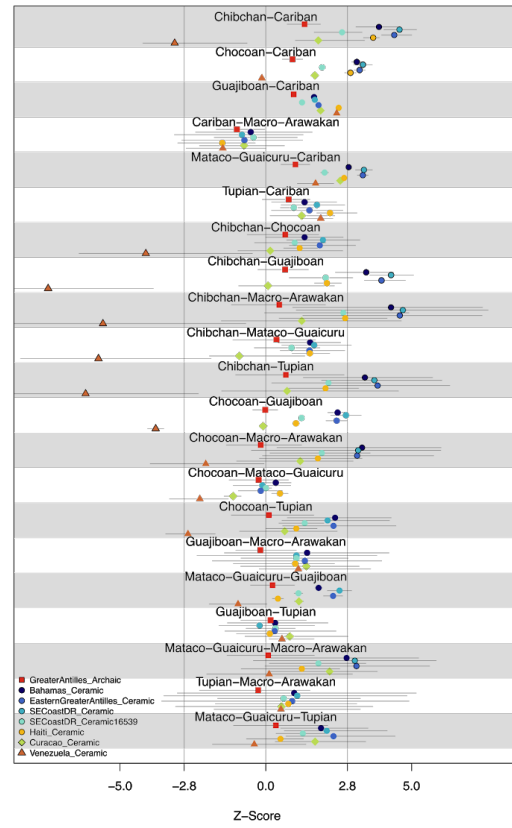
All Archaic-associated individuals from Canimar Abajo (Cuba) and I10126 form a single clade in *qpWave* despite spanning over 2,000 years. This points to a well-defined Archaic ancestry profile consistent with deriving from a single source spread across the Greater Antilles that persisted with minimal mixture in some regions well into the archaeologically-defined Ceramic Age. To investigate the affinities of *GreaterAntilles\_Archaic* to populations from the American continents, we applied outgroup- $f_3$  analysis (Supplementary Information section 10; Supplementary Data 8). *GreaterAntilles\_Archaic* shares the most genetic drift with Indigenous groups from Central and northern South America (Fig. 3a) belonging to seven language families: Tupian, Macro-Arawakan, Cariban, Chibchan, Chocoan, Guajiboan, and Mataco-Guaicuru<sup>21,22</sup>. We find no evidence of excess allele sharing with people from any one family relative to the others (all  $|Z| < 2.8$ , Fig. 3b; Supplementary Data 9).

We used *qpAdm* to explore the most likely source(s) of ancestry for *GreaterAntilles\_Archaic* using published ancient DNA data older than 2000 years. We found a single model using individuals from southern Brazil as a source that passed a threshold of  $p > 0.05$  (*Brazil\_Laranjal\_6700BP*,  $p = 0.343$ ) (Supplementary Information section 8; Table S11). However, statistics of the form  $f_4(\text{Mbuti}, \text{GreaterAntilles_Archaic}, \text{Brazil_Laranjal_6700BP}, \text{Test})$  with *Test* as six other possible sources did not confirm significant affinities between *GreaterAntilles\_Archaic* and *Brazil\_Laranjal\_6700BP* ( $|Z| < 1.451$ ), suggesting that the fit of *Brazil\_Laranjal\_6700BP* as a source may reflect limited power to reject the model (Table S12). Using present-day populations, a similar result of no specific affinities was found (Table S13). In *qpGraph*, we fit *GreaterAntilles\_Archaic* in several positions along a skeleton admixture graph from Posth et al.<sup>23</sup>, again reflecting poor power to differentiate between a Central or South American origin (Supplementary Information section 11). Likewise, in a maximum likelihood phylogenetic tree based on allele frequency covariances<sup>24</sup>, *GreaterAntilles\_Archaic* is inferred to split before all South and some Central American populations (Fig. 3c). Together, these results point to the origin of *GreaterAntilles\_Archaic* in a deeply divergent Native American population that is not particularly closely related to any sampled present-day groups.

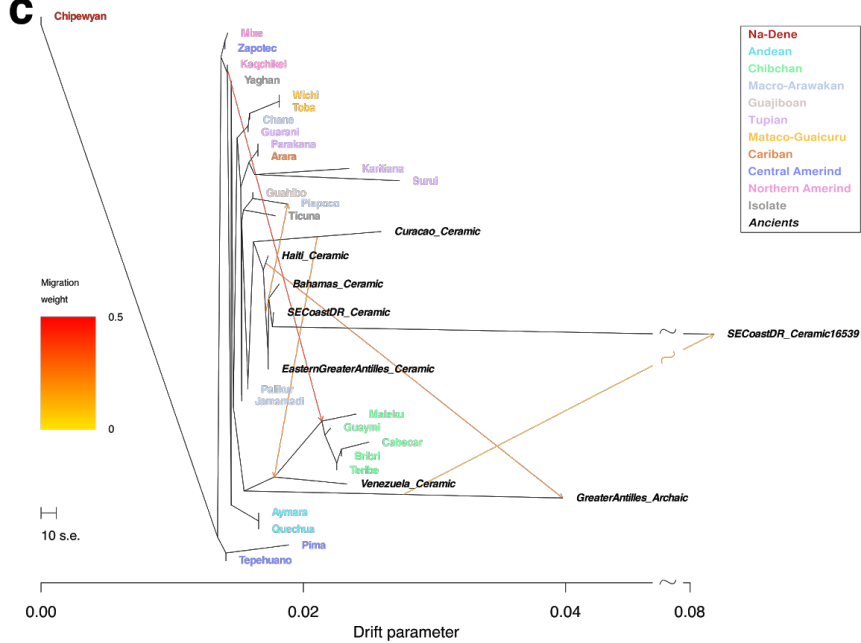
**a**



**b**



**c**



**Fig. 3: Affinities of the ancient populations from each region to modern Indigenous populations.** (a) Outgroup  $f_3$ -statistics performed on *GreaterAntilles\_Archaic*, *Caribbean\_Ceramic*, and *Venezuela\_Ceramic*. Each square represents a modern population. (b) Symmetry tests using ancient sub-clades, analysing the average relatedness to non-pooled modern populations from seven language families found in South America. We show the average among symmetry tests of the form  $f_4(\text{Mbuti, Test; LanguageGroup1Pop, LanguageGroup2Pop})$ , where *LanguageGroupXPop* is a population from the corresponding language family; horizontal lines represent the range. Vertical lines represent statistical significance threshold of  $Z=\pm 2.8$  (corresponding to a 99.5% CI). (c) Maximum likelihood population tree from allele frequencies using Treemix showing the *Caribbean\_Ceramic* sub-clades on the same branch as modern Arawak-speaking groups (Palikur, Jamamadi). Orange arrows represent admixture events, although the indicated direction of admixture might be inaccurate based on observations from analyses such as *qpAdm* admixture modeling - e.g., we believe it likely that there is *GreaterAntilles\_Archaic* admixture into *Haiti\_Ceramic* rather than the reverse direction of flow (Supplementary Information section 8).

### The spread of ceramic users in the Caribbean

When projected onto the PCA (Fig. 2a), all *Caribbean\_Ceramic* individuals (dating ~1800-300 cal. yr BP) fall in a single cluster, as expected if a genetically homogeneous population moved into and throughout the Caribbean from a single source. We confirm this with pairwise  $F_{ST} < 0.01$  between the *Caribbean\_Ceramic* sub-clades (very subtle absolute genetic differentiation). Values of  $F_{ST} > 0.1$  (comparable to present-day inter-continental genetic distances) support the substantial genetic differentiation between this clade and *GreaterAntilles\_Archaic* (Extended Data Fig. 1a).

Previous DNA studies have pointed to Arawak-speaking South American groups as being most closely related to Caribbean Ceramic-associated people<sup>10,11,14,25</sup> (Supplementary Information section 1), consistent with the presence of Arawak languages in the Caribbean at the time of European contact<sup>2</sup>. ADMIXTURE analysis suggests that individuals from each *Caribbean\_Ceramic* sub-clade are almost entirely composed of a component found in the highest proportion in modern Arawak-speaking Piapoco and Palikur (Fig. 2b), an affinity that is also supported by the position of all *Caribbean\_Ceramic* sub-clades on the same branch as Piapoco and Palikur in a maximum likelihood tree allowing for admixture events (Fig. 3c). Similar to previous work<sup>13,14</sup>, we are unable to identify a statistically significant signal of closer relatedness to Arawak- than to Cariban- or Tupian-speaking populations using  $f_4$ -statistics (Fig. 3b; Supplementary Information section 10; Supplementary Data 9), although we obtained a successful 1-way model with the Arawak-speaking Piapoco in *qpAdm* ( $p=0.519$  or  $p=0.249$ , depending on the dataset used for Piapoco; Tables S13 and S14). Thus, our analyses also suggest an Arawak connection and highlight a genetic affinity between the Ceramic Age Caribbean and northeastern South Americans.

Aside from subtle population structure (described above) likely shaped by the geographic separation among the islands and potentially also influenced by different (but sub-significant) levels of admixture with Archaic groups (though a near-replacement of Archaic-related ancestry eventually

did occur in many regions), we find no evidence that substantial proportions of different ancestry profiles were introduced to the Greater Antilles and The Bahamas throughout the Ceramic Age, in contrast to analyses of skeletal morphology that suggest a migration of Carib peoples from Venezuela ~1,150 years ago<sup>26</sup>. Our simulations show that ~2-8% Carib- or *Venezuela\_Ceramic*-related ancestry would be required for detection, so we cannot rule out contributions less than this proportion; this analysis also requires that our selected proxies accurately represent Carib-related ancestry (Supplementary Information section 12). Thus, if a major Carib expansion occurred into the Antilles around 800 CE, it cannot have derived large proportions of ancestry from groups related to our present-day Cariban-speaking proxy or to ancient *Venezuela\_Ceramic*-related people, and instead must have been derived to a mainland population from the northern coast of South America with an ancestry profile similar to that of present-day Caribbean Islanders. We are not aware of any modern groups from northern South America with such an ancestry profile, but the present-day distribution of Indigenous groups could plausibly be quite different from that of a millennium ago.

We find that only a minimal amount of Archaic-related ancestry may have persisted in ceramic-using populations, identifying isolated signals of Archaic-related admixture in three individuals from two Ceramic-associated sites in Hispaniola. *SECoastDR\_Ceramic16539* (Dominican Republic) has 12.1±2.0% Archaic-related ancestry (Table S6), resulting from admixture that we estimate using the DATES software occurred an average of 39±14 generations before the individual lived ( $Z=2.77$ ; Supplementary Information section 13), while two individuals from Diale 1 (Haiti) have 18.6±2.1% Archaic-related ancestry, which we estimate to have resulted from admixture ~11±5 generations before the individuals lived ( $Z=2.24$ ; mixture proportions in Tables S7 and S8). In contrast, all individuals from the site of Canimar Abajo (Cuba) retained unadmixed Archaic-related ancestry spanning over 2,000 years to the limits of our resolution, consistent with archaeological<sup>27</sup> and historical<sup>28</sup> accounts that this region was home to people with a different language and cultural traditions than more easterly parts of Cuba as late as the Contact Period.

We model five ancient individuals from Curaçao as having 68.1±5.1% *Caribbean\_Ceramic*-related ancestry and 31.9±5.1% *Venezuela\_Ceramic*-related ancestry ( $p=0.317$ ; Tables S9 and S10), indicating that people related to those who moved into the Greater Antilles also reached Curaçao and left a genetic legacy in the individuals we analyzed (we were unable to estimate dates for this admixture event due to limited statistical power). The *Venezuela\_Ceramic* clade is distinct from any coastal or island population studied here, with ADMIXTURE suggesting a major component maximized in present-day Chibchan-speaking Cabécar (Fig. 2b); significant affinities with Chibchan speakers are also seen in statistics of the form  $f_4(\text{Mbuti}, \text{Test}; \text{LanguageGroup1Pop}, \text{LanguageGroup2Pop})$  (Fig. 3a, b). We are also able to model *Venezuela\_Ceramic* using *qpAdm* with a single Chibchan-related component of ancestry from Cabécar (Tables S13 and S14). Thus, despite the location of *Venezuela\_Ceramic* in a hypothesized source region for the peopling of the Caribbean and the dates of the analyzed individuals near the beginning of the Caribbean Ceramic Age, our analysis increases the weight of

evidence for the Ceramic expansion having more easterly origins in northern South America. These results point to a scenario in which Curaçao's Ceramic Age population was derived from admixture of two groups: one that was related to the population that spread to the Antilles at the beginning of the Caribbean Ceramic Age, and the other associated with the Dabajuroid ceramic style that linked Venezuelan sites like Las Locas to Curaçao.

### Insights into social structure and demographic history

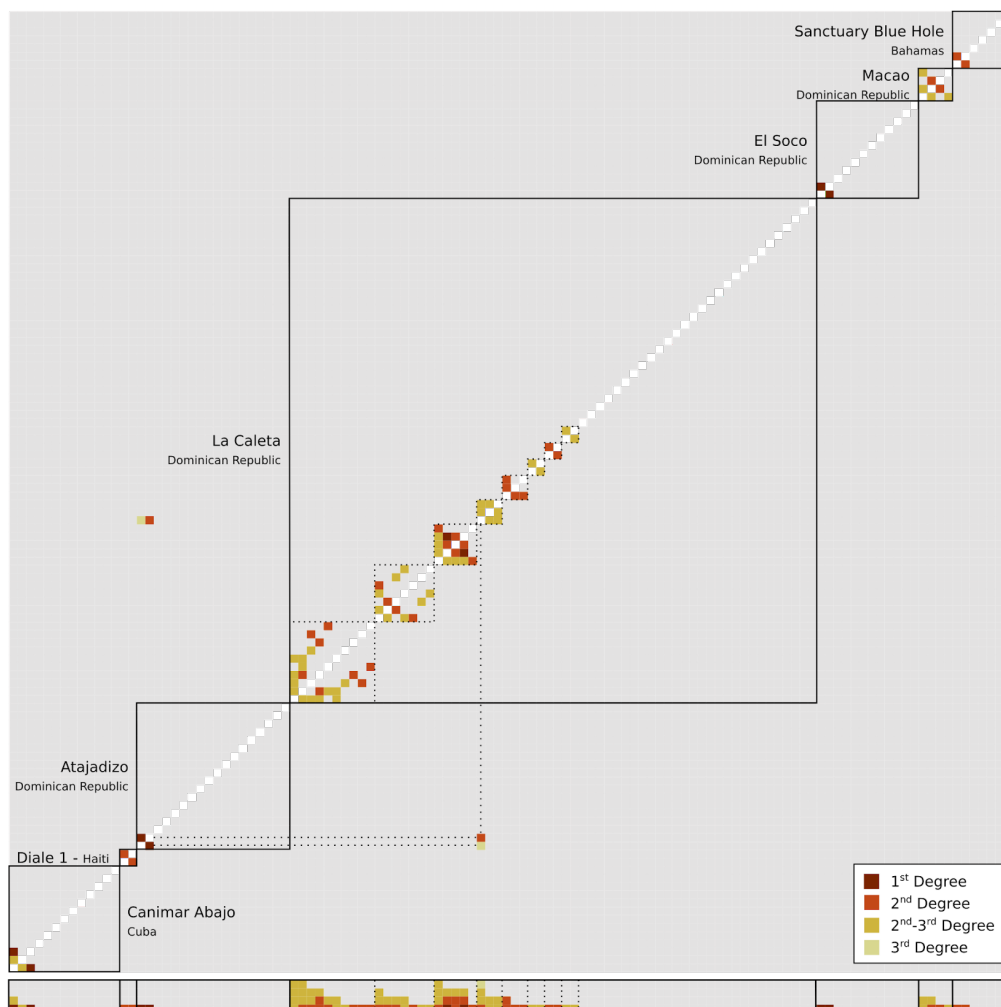
We assessed kinship between the ancient individuals and identified a total of 43 first-, second-, or third-degree relationships. The majority were within the site of La Caleta, where 34 out of 62 individuals studied had relatives in our dataset, although the proportion of related pairs was not significantly larger than that from other sites (32 related out of 1891 pairs tested, 1.69%, 95% CI 1.16%-2.38%, versus six related out of 302 within-site pairs tested in other Caribbean Ceramic sites, 1.99%, 95% CI 0.73%-4.25%; Fig. 4a; Supplementary Information section 6). We also identified relatives (all genetic males) buried ~75 kilometers apart in the southern Dominican Republic: father I17906 and son I17903 from Atajadizo were second and third-degree relatives of I15601 from La Caleta. This inter-regional kinship provides direct evidence of mobility and connectivity during the Ceramic Age.

We screened 154 individuals with at least 400,000 SNPs covered for runs of homozygosity (ROH)  $>4cM$ <sup>29</sup> (Supplementary Information section 6; Supplementary Data 10; Figure S17). Large sums of long ROH blocks ( $>20cM$ ) indicate recent parental relatedness (Fig. 4b), whereas an abundance of shorter ROH is a signature of background parental relatedness, a consequence of restricted mating pools<sup>30</sup>. We identified only 2 of 154 individuals with at least 100cM of their genome in ROH $>20cM$  blocks (~135cM is the average value for blocks of this size in offspring in first cousins, Fig. 4b, Figure S17). This shows that close kin unions were rare in the pre-contact Caribbean; instead, many unions took place between people as close as second and third cousins. However, we detected a general abundance of short and mid size ROH across our ancient samples (Fig. 4b). To quantify this signal, we applied a maximum likelihood method to estimate effective population size  $N_e$  using the length distribution of all ROH within 4-20cM, which arise from co-ancestry within the last few dozen generations (see Supplementary Information section 6 for details). We infer that population sizes for Ceramic Caribbean sites are larger ( $2N_e$  values ranging 1000-2000) than population sizes for Archaic individuals older than 2000 BP ( $2N_e$ ~500 for two sites) (Fig. 4c; Extended Data Table 1), pointing toward an increase of population density with the arrival of agricultural subsistence strategies, an observation in agreement with conditional heterozygosity levels (Extended Data Fig. 2). However, effective population size estimates across the Ceramic Caribbean remain low throughout the Ceramic Age, within the range of previous estimates made using a single ancient genome from The Bahamas<sup>14</sup> ( $2N_e=3200$ ) and using coalescence rates of Caribbean mtDNA clades in Puerto Rico<sup>13</sup> (pre-contact female  $2N_e<10,000$ ). While short ROH measures population sizes of the past few dozen generations, the consistent signal across all Caribbean sites (including large and small islands) indicates that these estimates share some signal of larger meta-populations, as expected due to a high rate of mobility

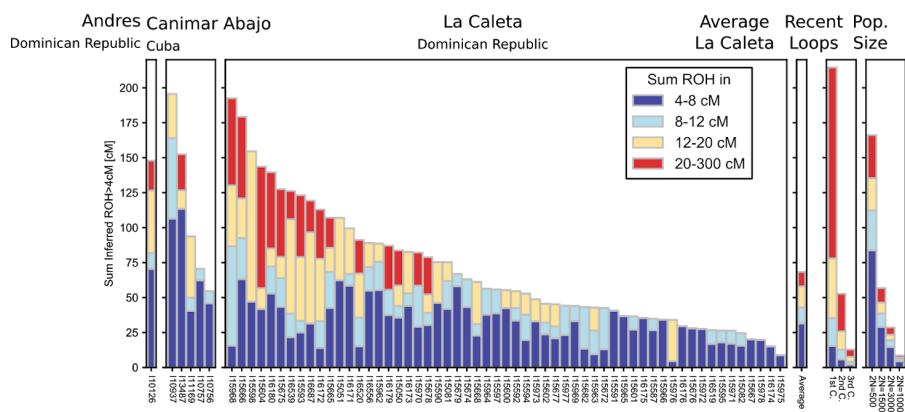
(and as confirmed by our detection of relative buried ~75 kilometers apart). For humans, census population sizes are typically up to an order of magnitude larger than effective population sizes<sup>31</sup>, with a factor of three estimated for inferences based on shared sequence blocks<sup>32</sup>. Even using the larger ratios, our results indicate that the population size estimates of hundreds of thousands on major Caribbean islands such as Hispaniola prior to European colonization that are discussed in the literature<sup>6</sup> are substantial overestimates. At the same time, our results confirm the magnitude of population collapse following European colonization and the killing of Indigenous people. The severity of this collapse is evidenced, for example, in the frequencies of mitochondrial haplogroups such as B2 and D1 that were common in our data but are rare in present-day Caribbean populations<sup>33</sup>.



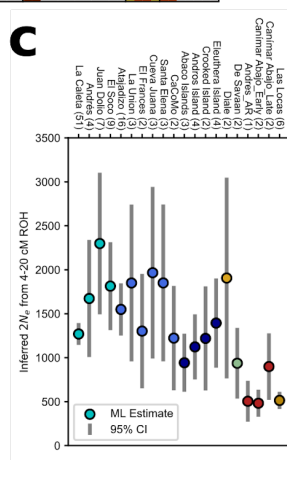
**a**



**b**



**c**



**Fig. 4:** (a) Pairwise kinship estimates for all individuals from sites where relatives were identified. Dotted lines identify family clusters and inter-site relationships. Bottom rows correspond to relationships per individual. (b) Sums of inferred ROH larger than 4cM, per individual, for the Archaic-associated site of Canimar Abajo (Cuba) and the Ceramic-associated site of La Caleta (Dominican Republic). Remaining sites shown in Figure S17. “Recent loops” and “Pop. Size” represent expected distributions from simulated data for parental relationships as shown (C., cousin), and varying population sizes, respectively. (c) Inferred effective population sizes for individuals grouped per site (note that we here differentiate between an earlier and later occupation at Canimar Abajo). These estimates are derived from ROH blocks 4-20cM and a likelihood model (Supplementary Information section 6). Colors as per sub-clades, numbers in brackets denote sample size with sufficient coverage for ROH estimation).

### Pre-Contact Ancestry Persists in Modern Caribbean Populations

We used genotype data from modern Caribbean populations<sup>25</sup> to test for affinities between the Indigenous ancestry found in present-day Caribbean groups and the ancient Caribbean individuals. For three different populations (Cuba, Dominican Republic, and Puerto Rico) as *Test*, we measured affinity to the Archaic-associated individuals from Cuba versus Ceramic-associated individuals by computing the statistic  $f_4(\text{European}, \text{Test}; \text{Cuba\_Archaic}, \text{Caribbean\_Ceramic})$ . Despite reduced statistical power due to relatively low proportions of Indigenous ancestry in present-day populations, we obtained a significant signal ( $Z=3.0$ ) for Puerto Rico of greater relatedness to Ceramic-associated individuals, which we also replicated using Puerto Ricans from the 1000 Genomes Project ( $Z=4.1$ ) (Supplementary Data 11). We performed an empirical power analysis (Supplementary Information section 14) for the present-day Cubans and determined that our results are consistent with entirely Ceramic-related ancestry but not with entirely Archaic-related ancestry. We also carried out the same test separately for present-day individuals from all 15 provinces of Cuba<sup>34</sup> and found three provinces and five municipalities with weakly significantly Ceramic-related ancestry ( $2 < |Z| < 3.1$ ) but none with significant evidence of Archaic-related ancestry ( $|Z| < 2$ ), including the region of northwestern Cuba that was the source of our ancient individuals (Supplementary Data 11). Thus, while our ancient DNA data from northwestern Cuba do not identify any Ceramic-related ancestry by ~900 cal. yr BP, such ancestry did eventually reach the island (Extended Data Tables 2 and 3). We also find support for the persistence of Indigenous ancestry in uniparental haplogroups, where major Indigenous-specific haplogroups identified in our ancient individuals are still found in the Caribbean today<sup>33,35-37</sup>, despite the introduction of European and African haplogroups over the last ~500 years. Detection of our newly discovered deep branch of haplogroup C1d (the only haplogroup in our dataset unique to the Caribbean) in a modern individual from Puerto Rico from the 1000 Genomes Project adds to the weight of this evidence.

## Discussion

Our analysis provides insight into debates surrounding the timing, trajectory, and geographic origin of demographic movements into the pre-contact Caribbean and the role of such movements in changes in technology and material culture that took place in this region over the past 6,000 years.

First, we show that the ancestry in the Greater Antilles during the Archaic Age was consistent with deriving from a single continental source, though a limitation of our data is that the individuals we were able to analyze at high resolution were recovered from just two sites, albeit importantly coming from two different islands (Cuba and Hispaniola). While we cannot currently distinguish between models of a Central or South American origin, our analyses suggest that a North American origin for these Archaic-associated Caribbean peoples was unlikely.

Second, in line with evidence suggesting that the people who introduced intensive ceramic usage and agriculture into the Greater Antilles and The Bahamas moved from a source along the Orinoco River basin in Venezuela and the Guianas and arrived in the Caribbean Islands ~2,500 years ago<sup>38</sup>, our data are consistent with a movement from a single source accompanying the transition in material culture from predominant lithic-usage to a preponderance of ceramics. Ceramic-associated individuals from the parts of the Caribbean that we analyzed show an affinity to present-day Arawak speakers, consistent with an origin in northeastern South America as evidenced by archaeological and linguistic data<sup>39</sup>. Consistent with hypotheses that proto-Arawak-speakers experienced population splits at river junctions as they migrated from Amazonian South America toward the northeast part of South America, with some groups moving further along the Orinoco and into the Antilles and others moving toward the western Venezuela coast<sup>35</sup>, Ceramic-associated individuals from Curaçao have some ancestry related to that found in the *Caribbean\_Ceramic* clade and additional ancestry such as is seen in the ceramic-users from Las Locas, suggesting that these populations came back together through admixture in the ancestry of the studied individuals in Curaçao.

Third, we find no evidence of differential ancestry during the Ceramic Age across changes in ceramic styles that have been hypothesized to reflect new migrations (though we present several caveats in Table 1). A recent study of facial morphology<sup>26</sup> proposed a second expansion of Carib peoples to the Greater Antilles ~1,150 years ago, but we do not detect significant shifts in genetic ancestry throughout the entirety of the Ceramic Age, despite the fact that genome-wide analysis of hundreds of thousands of independent SNPs should in principle have more power to detect population substructure than morphological analysis, which at most can analyze only dozens of independent characters. More generally, we find no clear association between our genetic clusters within the Ceramic Age and the traditional ceramic stylistic typologies (Saladoid, Ostionoid, Meillacoid, Chicoid; Supplementary Information section 1), thereby providing no evidence for the culture-history model that considered stylistic transitions as the result of major movements of new people replacing established groups. Instead, we show that the ancestry profile of the southeastern coastal region of

the Dominican Republic spans more than a millennium across major transitions in material cultural styles. While we cannot rule out scenarios of migrations of genetically very similar populations among the islands helping to drive some of the observed cultural changes, our findings increase the weight of evidence that exchange among established groups within the Caribbean was a key driver of stylistic changes.

Fourth, by assembling the first dataset from the Caribbean from both the Archaic and Ceramic Ages, we shed light on the interaction between representatives of both ancestry types. Our data from Canímar Abajo confirms that the direct descendants of people of the Archaic Age were contemporaries with ceramic-users, users consistent with the archaeological evidence of continuity of Archaic Age cultures in western Cuba. Furthermore, the presence of admixed individuals in Hispaniola (one from La Caleta and two from Diale 1), documents that a complete replacement of local Archaic-related ancestry did not immediately take place in all regions.

Fifth, we find that peoples living in some parts of the Caribbean (including Puerto Rico and certain areas of Cuba) today carry a signal of Ceramic-related ancestry despite centuries of admixture from European and African individuals. In Cuba, our results (combined with linguistic and physical anthropological evidence) suggest the possibility of the persistence of Archaic groups up until the Contact Period; however, the Indigenous admixed ancestry found in Cuba today is most likely not derived from this source. This could reflect post-colonial movement of Indigenous people between Caribbean islands or a more complex landscape of ancestry in Cuba prior to European colonization. Denser sampling of ancient individuals in the centuries surrounding colonization and from under-sampled locales such as Jamaica will provide insight on the events that contributed to the current distribution of Indigenous ancestry in Caribbean peoples.

Finally, our data provide new insights into social structure and demography. We identify instances of mobility, with related male individuals buried ~75 kilometers apart. We observe an active avoidance of unions between close relatives during both the Archaic and Ceramic Ages and large amounts of cumulative ROH across most of the Caribbean sites analysed compared to almost all ancient cultures studied to date with ancient DNA<sup>29</sup>. This reflects limited pools of mates due to small effective population sizes (1,000-2,000) averaged over the millennium prior to the time the individuals lived that are difficult to reconcile with theories that pre-contact population sizes in the Caribbean were extremely large. Even allowing for the fact that census sizes can be an order of magnitude larger than effective population sizes, our results point to a census size for the population of which Hispaniola was a part being no more than 10,000-20,000, which is far less than estimates of over 100,000 or even over 1,000,000 that have been discussed in the literature<sup>3,40</sup>. While our estimates are lower than some previous ones, the devastating impact that European colonization, expropriation, and systematic killing of Indigenous people had on Caribbean populations is clear. Today, modern Caribbean people are mixtures in different proportions of three groups, all of which

have contributed importantly to present-day populations: Indigenous populations that experienced extreme post-colonial bottlenecks (~4% on average in Cuba, ~6% in Dominican Republic, and ~14% in Puerto Rico according to our estimation by *qpAdm*), immigrant Europeans (~70% in Cuba, 56% in the Dominican Republic, and 68% in Puerto Rico), and immigrant Africans who arrived in the course of the trans-Atlantic slave trade (~26% in Cuba, ~38% in the Dominican Republic, and ~18% in Puerto Rico) (Extended Data Tables 2 and 3). Altogether, these results reveal dynamic webs of biological and cultural connectivity in the ancient Caribbean, transformed over the last 500 years by demographic upheaval, but nevertheless persisting in many of the present-day people of this region.

1. Oliver, J. R. The archaeological, linguistic and ethnohistorical evidence for the expansion of Arawakan into northwestern Venezuela and northeastern Colombia. (University of Illinois at Urbana-Champaign, 1989).
2. Keegan, W. F. & Hofman, C. L. The Caribbean before Columbus. (Oxford University Press, 2017).
3. Rouse, I. The Tainos: Rise & Decline of the People who Greeted Columbus. (Yale University Press, 1992).
4. Callaghan, R. T. Archaeological Views of Caribbean Seafaring. in Oxford handbook of Caribbean archaeology (eds. Keegan, W. F., Hofman, C. & Rodriguez, R. R.) 285-295 (Oxford University Press, 2013).
5. Siegel, P. E. et al. Paleoenvironmental evidence for first human colonization of the eastern Caribbean. *Quaternary Science Reviews* 129, 275-295 (2015).
6. Henige, D. On the Contact Population of Hispaniola: History as Higher Mathematics. *Hispanic American Historical Review* 58, 217-237 (1978).
7. Dobyns, H. F. Their Number Become Thinned: Native American Population Dynamics in Eastern North America. (University of Tennessee Press, 1983).
8. Keegan, W. The People Who Discovered Columbus. (University Press of Florida, 1992).
9. Lleonart, R., Riego, E., Suárez, R. R., Ruiz, R. T. & de la Fuente, J. Analyses of DNA from ancient bones of a pre-Columbian Cuban woman and a child. *Genetics and Molecular Biology* 22, 285-289 (1999).
10. Lalueza-Fox, C., Calderón, F. L., Calafell, F., Morera, B. & Bertranpetit, J. MtDNA from extinct Tainos and the peopling of the Caribbean. *Ann. Hum. Genet.* 65, 137-151 (2001).
11. Lalueza-Fox, C., Gilbert, M. T. P., Martínez-Fuentes, A. J., Calafell, F. & Bertranpetit, J. Mitochondrial DNA from pre-Columbian Ciboneys from Cuba and the prehistoric colonization of the Caribbean. *Am. J. Phys. Anthropol.* 121, 97-108 (2003).
12. Mendisco, F. et al. Where are the Caribs? Ancient DNA from ceramic period human remains in the Lesser Antilles. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370, 20130388 (2015).
13. Nieves-Colón, M. A. et al. Ancient DNA reconstructs the genetic legacies of pre-contact Puerto Rico communities. *Molecular Biology and Evolution* 37, 611-626 (2020).
14. Schroeder, H. et al. Origins and genetic legacies of the Caribbean Taino. *Proc. Natl. Acad. Sci. U. S. A.* 115, 2341-2346 (2018).
15. Reich, D. et al. Reconstructing Native American population history. *Nature* 488, 370-374 (2012).
16. Underhill, P. A., Jin, L., Zeman, R., Oefner, P. J. & Cavalli-Sforza, L. L. A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. *Proceedings of the National Academy of Sciences* 93, 196-200 (1996).
17. Kivisild, T. The study of human Y chromosome variation through ancient DNA. *Human Genetics* 136, 529-546 (2017).

18. Schurr, T. G. & Sherry, S. T. Mitochondrial DNA and Y chromosome diversity and the peopling of the Americas: evolutionary and demographic evidence. *Am. J. Hum. Biol.* 16, 420-439 (2004).
19. Achilli, A. et al. The phylogeny of the four pan-American MtDNA haplogroups: implications for evolutionary and disease studies. *PLoS One* 3, e1764 (2008).
20. Consortium, T. 1000 G. P. & The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* 526, 68-74 (2015).
21. Greenberg, J. H. *Language in the Americas*. (Stanford University Press, 1987).
22. 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. *Current Anthropology* 46, S121-S128 (2005).
23. Posth, C. et al. Reconstructing the Deep Population History of Central and South America. *Cell* 175, 1185-1197.e22 (2018).
24. Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 8, e1002967 (2012).
25. Moreno-Estrada, A. et al. Reconstructing the population genetic history of the Caribbean. *PLoS Genet.* 9, e1003925 (2013).
26. 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).
27. Chinique de Armas, Y., Roksandic, M., Suárez, R. R., Smith, D. G. & Buhay, W. M. Isotopic Evidence of Variations in Subsistence Strategies and Food Consumption Patterns among 'Fisher-Gatherer' Populations of Western Cuba. in *Cuban Archaeology in the Circum-Caribbean Context* (ed. Roksandic, I.) (University Press of Florida, 2016).
28. Lovén, S. E. *Origins of the Tainan Culture, West Indies*. (Elanders Bokfryckeri Akfiebölag, 1935).
29. Ringbauer, H., Novembre, J. & Steinrücken, M. Detecting runs of homozygosity from low-coverage ancient DNA. *bioRxiv.org* (2020).
30. 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).
31. Frankham, R. Effective population size/adult population size ratios in wildlife: a review. *Genet. Res.* 89, 491-503 (2007).
32. 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).
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. Fortes-Lima, C. et al. Exploring Cuba's population structure and demographic history using genome-wide data. *Sci. Rep.* 8, 11422 (2018).
35. 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. *Caribbean Journal of Science* 39, (2003).
36. Mendizabal, I. et al. Genetic origin, admixture, and asymmetry in maternal and paternal human lineages in Cuba. *BMC Evol. Biol.* 8, 213 (2008).
37. 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).
38. Hofman, C. L. & Reid, B. A. The Saladoid. in *Encyclopedia of Caribbean archaeology* (eds. Reid, B. & Gilmore, G.) 300-303 (University of Florida Press, 2014).
39. Roksandic, I. & Roksandic, M. *Peopling of the Caribbean*. 199-223 (Kerns Verlag, 2018).
40. Anderson-Córdova, K. F. *Hispaniola and Puerto Rico: Indian Acculturation and Heterogeneity, 1492-1550*. (University Microfilms International., 1990).

## 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. Powder was produced from a cochlea<sup>41,42</sup>, tooth, phalanx, or ossicle<sup>43</sup> from each individual in a clean room facility at Harvard Medical School (Boston, USA), University College Dublin (Dublin, Ireland), or the University of Vienna (Vienna, Austria). See Supplementary Data 2 for the skeletal element used for each individual and location of powder preparation.

We extracted DNA in dedicated ancient DNA laboratories at Harvard Medical School or the University of Vienna following published protocols<sup>44-46</sup>. From the extracts, we prepared dual-barcoded double-stranded<sup>47</sup> or dual-indexed single-stranded libraries<sup>48,49</sup>, both treated with uracil-DNA glycosylase (UDG) to reduce the rate of characteristic ancient DNA damage<sup>50</sup>. Double-stranded libraries were treated in a modified partial UDG preparation<sup>47</sup> ('half'), leaving a reduced damage signal at both ends (5' C-to-T, 3' G-to-A). Single-stranded libraries were treated with *E. coli* UDG (USER from NEB) that inefficiently cuts the 5' Uracil and does not cut the 3' Uracil. For a subset of samples, we increased coverage by preparing multiple libraries; see Supplementary Data 2 for the number of libraries analyzed for each individual.

To generate SNP capture data, we used in-solution target hybridization to enrich for sequences that overlap the mitochondrial genome and ~1.24 million genome-wide SNPs<sup>51-54</sup> ("1240k"), either in two separate enrichments or simultaneously (Supplementary Data 2). We then added two 7-base-pair indexing barcodes to the adapters of each double-stranded library and sequenced libraries using either an Illumina NextSeq500 instrument with 2x76 cycles or an Illumina HiSeqX10 instrument with 2x101 cycles and reading the indices with 2x7 cycles (double-stranded libraries) or 2x8 cycles (single-stranded libraries).

Prior to alignment, we merged paired-end sequences, retaining reads that exhibited no more than one mismatch between the forward and reverse base if base quality was  $\geq 20$ , or 3 mismatches if base quality was  $< 20$ . A custom toolkit was used for merging and trimming adapters and barcodes (available at <https://github.com/DReichLab/ADNA-Tools>). Merged sequences were mapped to the reconstructed human mtDNA consensus sequence (RSRS)<sup>55</sup> and the human reference genome version hg19 using the `samse` command in BWA v.0.7.15-r1140<sup>56</sup> with the parameters `-n 0.01, -o 2 and -l 16500`. Duplicate molecules (those exhibiting the same mapped start and end position and same stand

orientation) were removed after alignment using the Broad Institute's Picard MarkDuplicates tool (available at <http://broadinstitute.github.io/picard/>). We trimmed two terminal bases from UDG-half libraries to reduce damage-induced errors.

We evaluated the authenticity of the isolated DNA by retaining individuals with a minimum of 3% of cytosine-to-thymine substitutions at the end of the sequenced fragments<sup>47</sup>, point estimates of mitochondrial DNA (mtDNA) contamination below 5% using contamMix v.1.0-12<sup>51</sup>, and point estimates of X chromosome contamination (in males) below 3%<sup>57</sup> (Supplementary Data 2); three individuals (I7977, I7594, and I16540) were slightly below the set threshold for one authenticity metric, most likely due to low coverage (0.02-0.10X nuclear coverage and <23X mtDNA coverage), but passed the threshold for the other two metrics. SNPs were determined by randomly sampling an overlapping read with minimum mapping quality of  $\geq 10$  and base quality of  $\geq 20$ . Individuals with less than 20,000 covered SNPs were excluded from quantitative analyses. One individual from each pair of first-degree relatives in the dataset was excluded from population genetics analysis; in all cases, we retained the higher coverage individual, listed in Supplementary Data 1.

### Radiocarbon dates

We report 52 new radiocarbon (<sup>14</sup>C) dates on bone fragments generated using accelerator mass spectrometry (AMS) (Supplementary Data 3). Most dates (n=48) were generated at the Pennsylvania State University (PSU) Radiocarbon Laboratory, and the remainder (n=4) were generated at the Center for Isotopic Research on Cultural and Environmental heritage (CIRCE). The sample preparation methodology at PSU was carried out as previously reported<sup>58</sup>, where bone collagen was extracted and purified using a modified Longin method with ultrafiltration<sup>59</sup> (>30 kDa gelatin); if collagen yields were low, a modified XAD process<sup>60</sup> (XAD amino acids) was used. Sample preparation at CIRCE was carried out following the lab-adapted Longin method<sup>61</sup>. Supplementary Data 3 lists the preparation method used for each sample.

We provide conventional radiocarbon dates in years BP and three calibrated dates for each conventional age in Supplementary Data 3. The first date is calibrated using the IntCal13 curve<sup>62</sup>; the second date is calibrated using a mix of the IntCal13 and Marine13<sup>62</sup> curves in a 75%:25% ratio, accounting for a 25% marine protein contribution to the diet; and the third date is calibrated using a mix of the IntCal13 and Marine13 curves in a 50%:50% ratio, accounting for a 50% marine protein contribution to the diet. Supplementary Information section 3 discusses the impact of the marine reservoir effect on these dates. The differences between the median dates obtained for the three different calibration scenarios range from 190-360 years (median 265 years), and so we view this as a measure of the degree of systematic uncertainty in the <sup>14</sup>C dates we obtained. We report all dates



in the main manuscript as calibrated years before present, using the IntCal13 curve data for the calibrated dates.

### Dataset assembly

We merged the 184 ancient individuals that passed screening into a base dataset that included 61 previously published ancient American individuals<sup>13,14,23,63-65</sup>, and 36 modern Indigenous American groups sourced from single nucleotide polymorphism (SNP) array genotyping datasets or whole genome sequencing datasets (Extended Data Table 4):

- ‘1240K SNPs’, whole genome sequencing data restricted to a canonical set of 1,233,013 SNPs<sup>51-54,66,67</sup>
- ‘Human Origins dataset’, 597,573 SNPs<sup>68-70</sup>
- ‘Illumina dataset’ (unmasked/unadmixed individuals only), 352,432 SNPs<sup>15</sup>

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

### Uniparental haplogroups

We determined mtDNA haplogroups using bam files, restricting to reads with MAPQ  $\geq 30$  and base quality  $\geq 20$ . 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  $\geq 30$  and base quality  $\geq 30$ . We called haplogroups by determining the most derived mutation for each sample, 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 7. Discussion of mtDNA and Y chromosome haplogroup distribution in the Caribbean is found in Supplementary Information section 9.

## Kinship, Consanguinity and Conditional Heterozygosity

We assessed kinship for every pair of individuals (including individuals from different sites and islands) using a previously described method<sup>71</sup>, and we present results for first-, second-, and third-degree relatives in Table S2. A number of the identified kin relationships are labeled as “2<sup>nd</sup>-3<sup>rd</sup> degree”, due to uncertainty in the assessment of mismatch rates caused, for example, by increased homozygosity in the population overall.

In our dataset of 184 ancient individuals, we identified 49 individuals sharing 43 unique pairwise kin relationships up to the third degree. Four pairs of individuals were identified as first-degree relatives, while 16 pairs were definitively second-degree relatives, and 1 pair was probable third-degree relatives. The remaining 22 relationships were considered “2<sup>nd</sup>-3<sup>rd</sup>-degree.”

We identified Runs of Homozygosity (ROH) within our ancient dataset using the Python package *hapROH* (<https://test.pypi.org/project/hapROH/>). Following a previously described method<sup>29</sup>, we used 5008 global haplotypes from the 1000G haplotype panel<sup>20</sup> as the reference panel. Following the recommendations for datasets with genotypes for ~1.24 million SNPs, we applied our method to ancient individuals with at least 400,000 SNPs covered (n=154) and ran the method on the pseudo-haploid data to identify ROH longer than 4 centiMorgans (cM). We used the default parameters of *hapROH*, which are optimized for ancient data genotyped at a similar number of sites. For each individual, we group the inferred ROH into four length categories: 4-8 cM, 8-12 cM, 12-20 cM and >20 cM and report the total sum in these bins (Supplementary Data 10; Figure S17).

To estimate effective population size from ROH, we applied a maximum likelihood inference framework (for derivation of the likelihood see Supplementary Information section 6). We fit all genome-wide ROH lengths between 4 and 20 cM long, and infer the effective population size that maximizes the likelihood for ROH lengths observed in a set of individuals. Estimation uncertainties are obtained from the curvature of the likelihood (Fisher Information matrix). Tests on simulated data confirmed the ability of the estimator to recover  $N_e$  estimates from genome-wide ROH of few individuals (Figures S18 and S19).

We used *popstats*<sup>70</sup> to compute conditional heterozygosity for all clades and sub-clades, which we compared with contemporaneous groups from continental South America, such as from the Peruvian Middle and Late Horizon periods<sup>72</sup>. As previously described<sup>73,74</sup>, we restricted the analysis to transversion SNPs ascertained in a Yoruba individual; see Extended Data Fig. 2.

## PCA

We performed principal component analysis (PCA) with *smartpca*<sup>75</sup>, using the option ‘*lsqproject: YES*’ to project ancient individuals onto the eigenvectors computed from modern individuals. The

approach of projecting each ancient sample onto patterns of variation learned from modern samples enables us to use data from a large fraction of SNPs covered in each individual and therefore 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 samples used to generate the PCA onto the positions where they would be expected to fall if they had been projected, thereby allowing the projected and non-projected samples to be appropriately co-visualized. We projected three previously published ancient individuals<sup>13,14</sup> and 184 new ancient individuals onto the first two principal components computed using 61 individuals from 23 present-day populations (Fig. 2a). See Supplementary Data 4 for all individuals included in PCA and values of PCs 1 and 2. For PCA by archaeological site, non-zoomed PCA, and PCA excluding CpG sites, see Figures S13-S15.

### Unsupervised analysis of population structure

We used the software ADMIXTURE<sup>76,77</sup> to perform unsupervised structure analysis on a dataset comprised of SNPs that overlap between the 1240k and Illumina dataset and were pruned in PLINK1.9<sup>78</sup> using `--indep-pairwise 200 25 0.4`. This left 269,091 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 error (Table S1). For each value of K, the replicate with the lowest cross-validation error was plotted and the results were compared. We choose to present K=5 as Fig. 2b as we found that the model with five components had the lowest cross-validation error in four out of five replicates and differentiated the components in a useful way for visualization. Results for the other values of K are presented as Figure S16 in Supplementary Information section 5.

### Estimation of $F_{ST}$ coefficients

To measure pairwise genetic differentiation between two groups of individuals, we estimated average pairwise  $F_{ST}$  and its standard error via block-jackknife over 1000 markers using the function "average\_patterson\_fst" from the package "scikit-allel" (version 1.2.1, DOI 10.5281/zenodo.3238280). We removed the individual with lower coverage of each pair of first degree relatives, as well as ancestry outliers (see main text). For each ancient individual we analyzed one allele per pseudo-haploid call; see Extended Data Fig. 1.

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

We used a multi-step framework involving *qpWave*, Treemix, and  $f_4$ -statistics to group sites and individuals, and considered this information together with admixture profiles and proportions from *qpAdm* to produce Fig. 2c (detailed methodology in Supplementary Information section 7). We started by using *qpWave* to identify major clades based on shared ancestry and then used Treemix and  $f_4$ -statistics to investigate the existence of sub-clades. Once all sub-clades were identified, we used  $f_4$ -statistics to investigate further substructure between sites within each clade. Geographic and chronological information such as island or cultural affiliation was not considered for these analyses, ensuring all clades and subclades were based solely on genetic information. We examined the association between genetic data and archaeological cultural complexes only after considering the genetic and archaeological information separately, following a previously published example<sup>79</sup>.

The software *qpWave*<sup>15</sup> from ADMIXTOOLS<sup>68</sup> estimates the minimum number of ancestry sources needed to form a group of test populations (“Left”), relative to a set of differentially related reference populations (“Right”). If the “Left” group contains two populations, *qpWave* will evaluate if they can be modelled as descending from the same sources, and hence will determine whether they form a clade. We used 12 present-day Indigenous American populations from the Human Origins dataset<sup>69</sup> plus Yukpa from<sup>66</sup> representing different language groups and ancestries from the American continent as our “Right” reference population set:

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

The argument ‘allsnps: NO’ was used. We ran two consecutive steps of *qpWave* analyses, starting with the identification of major groupings (step 1), or clades, and then reassessed the relationships between members within those clades by running the same tests in a “model competition” approach (step 2), such as is implemented in the related software *qpAdm*. A significance threshold of  $p > 0.01$  was set for accepting a clade between two sites or individuals.

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<sup>15</sup> (Extended Data Table 4) to identify within-clade site structure (step 3) by generating a maximum likelihood tree. We excluded four Chibchan, Chocoan and Arawak-speaking populations possibly admixed with each other from this analysis. We ran Treemix, grouping the SNPs in windows of 500 (flag -k 500) to account for linkage disequilibrium, setting Chipewyan as root (-root), allowing random migration events (-m), and disabling sample size correction (-noss) in order to include sites or populations represented by a single-individual. By running Treemix and allowing consecutive random migration/admixture events, we identified the ancient Caribbean sites that consistently shared the same relationships. We then used  $f_4$ -statistics to evaluate if they formed a sub-clade to

the exclusion of the other sites by following the tree's structure. For each identified intact node among all Treemix runs we used each downstream pair of site(s) as Test1 and Test2 and investigated their relationship to upstream sites or pools of sites (step 4). If an upstream node was unchanged in all runs, the sites composing it were pooled. However, once the first inconsistency was identified in an upstream node, all sites beyond that node were pooled together. A combination of three statistics per relationship allowed us to evaluate the Treemix structure of the sites being tested:

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

With Test1 and Test2 expected to be closer to each other than to Pool, the tested relationship finds support if the first test is statistically non-significant and at least one of the other two are significant. We used a Z-score threshold of 2.8 (associated with a 99.5% CI) to assess significance. These sites were then merged into a sub-clade inside the major Ceramic clade for further analysis.

After this clading analysis, we used  $f_4$ -statistics to further investigate potential substructure between sites within each sub-clade (step 5). For each pairwise site comparison, we randomly divided each site into two groups of individuals, and used a statistic of the form  $f_4(\text{Site1\_subset1}, \text{Site2\_subset1}; \text{Site1\_subset2}, \text{Site2\_subset2})$  to identify positive statistics suggesting substructure within the same clade. This randomization step was repeated 10 times, and the average Z-score was calculated. If a site was composed of a single individual we instead computed statistics of the form  $f_4(\text{Mbuti}, \text{Site1\_subset1}; \text{Site2\_singleIndividual}, \text{Site1\_subset2})$ , intended to evaluate if individuals within Site1 were closer to each other than to the single individual from Site2. No statistics were computed if both sites being tested contained only one individual.

### *qpAdm*

We used *qpAdm*<sup>53</sup> from ADMIXTOOLS<sup>68</sup> with 'allsnp: NO' to identify the most likely sources of ancestry and admixture for our populations/clades. First, we investigated if the outliers *SECoastDR\_Ceramic16539* and *EasternGreaterAntilles\_Ceramic7969*, as well as the individuals comprising the sub-clades *Haiti\_Ceramic* and *Curacao\_Ceramic*, could be modelled as admixed between the major ancestries represented by *GreaterAntilles\_Archaic* (composed of all individuals from the Cuban site of Canimar Abajo and I10126), *Caribbean\_Ceramic* (composed of *Bahamas\_Ceramic*, *EasternGreaterAntilles\_Ceramic* and *SECoastDR\_Ceramic*), and *Venezuela\_Ceramic*. We used this information to complete Fig. 2c. Then, based on this admixture information, we attempted to obtain more detailed admixture models using the sub-clades from within *Caribbean\_Ceramic* and the sites from within *GreaterAntilles\_Archaic* as possible sources. Lastly, we attempted to identify more distal sources of ancestry by using previously published ancient individuals from the Americas<sup>23,63-65</sup>, in this case for *qpWave*'s three major clades/groups. The base

“Right” set used was the same used for *qpWave*. We also tested all 1-, 2-, and 3-way models using these “Right” populations as sources by moving them to the “Left” as necessary, and confirmed the results with the same unmasked/unadmixed populations from the Illumina dataset.

### *qpGraph*

Due to the lack of significant *f*-statistics and attraction of *GreaterAntilles\_Archaic* to any other modern or ancient sequenced population, we investigated where *GreaterAntilles\_Archaic* would fit in the skeleton tree of previously published ancient American populations<sup>23</sup> using *qpGraph*. Detailed methodology is provided in Supplementary Information section 11.

### Admixture simulations

We investigated the sensitivity of *qpWave* in detecting Carib-related ancestry in the *Caribbean\_Ceramic* sub-clades by generating artificially admixed individuals with *Caribbean\_Ceramic* ancestry mixed with increasing amounts of proxy Carib-related ancestry (1, 2, 5, 8, 10, 20, 30, 40, and 50%), and then assessing at what admixture threshold we were able to reliably detect the latter ancestry type (Supplementary Information section 12; Figure S28). To generate these admixed individuals, we identified common SNPs between the two sources, randomly selected genotypes from the Arara individuals from the Human Origins and Illumina SNP array datasets, corresponding to each of the nine percentages to be tested, and added the remaining SNPs from a random individual from *Bahamas\_Ceramic*, *EasternGreaterAntilles\_Ceramic* and *SECoastDR\_Ceramic* with over 800,000 SNPs. We then ran *qpWave* with *Bahamas\_Ceramic*, *EasternGreaterAntilles\_Ceramic*, *SECoastDR\_Ceramic* and each of the simulated admixed individuals on the “Left”, while using the default 13 “Right” populations, as described in Supplementary Information section 7, plus the Carib population (Arara) used to generate those individuals.

### Dating admixture

We used the method *DATES* (Distribution of Ancestry Tracts of Evolutionary Signals<sup>58,80</sup>) version 3520 to estimate the time of admixture in admixed individuals from Haiti and Curaçao as well as admixed individual I16539. This method measures the decay of ancestry covariance to infer the time since mixture and estimates jackknife standard errors. Details of *DATES* analysis is found in Supplementary Information section 13.

## 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 through the statistic  $f_4(\text{European}, \text{Test}; \text{Cuba\_Archaic}, \text{Caribbean\_Ceramic})$ . In order to evaluate statistical power, we compared results for present-day Cubans alone to results obtained by adding one ancient individual from either the *GreaterAntilles\_Archaic* or *Caribbean\_Ceramic* clade to the Cuban test population. Full details found in Supplementary Information section 14.

## Testing for an Australasian link

We tested for a signal of relatedness to present-day Australasian populations<sup>66,70</sup> (“Population Y” signal), using the statistic  $f_4(\text{Mbuti}, \text{Onge/Papuan}; \text{Mixe}, \text{Archaic/Ceramic})$ . Here, Mixe is representative of a population that harbors no Population Y signal. When Onge was used as the Australasian proxy, several of our ancient groups showed weak positive statistics ( $Z > 2$ ), but only the Archaic individual I10126 from the site of Andrés in the Dominican Republic ( $Z = 3.4$ ) surpassed our threshold of  $Z > 2.8$  (Extended Data Table 5). The signal was also weaker when Papuan was used as the Australasian proxy ( $Z < 2.6$ ). The lack of a clear population Y signal is consistent with prior studies that also have not found this signal in ancient individuals from this region<sup>14</sup> and other areas of South America<sup>23</sup>.

## Code availability

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

## Data availability

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

41. Pinhasi, R. et al. Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous Bone. *PLoS One* 10, e0129102 (2015).
42. Pinhasi, R., Fernandes, D. M., Sirak, K. & Cheronet, O. Isolating the human cochlea to generate bone powder for ancient DNA analysis. *Nat. Protoc.* 14, 1194-1205 (2019).

43. Sirak, K. et al. Human auditory ossicles as an alternative optimal source of ancient DNA. *Genome Res.* 30, 427-436 (2020).
44. Dabney, J. et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U. S. A.* 110, 15758-15763 (2013).
45. Korlević, P. et al. Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* 59, 87-93 (2015).
46. Rohland, N., Glocke, I., Aximu-Petri, A. & Meyer, M. Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. *Nat. Protoc.* 13, 2447-2461 (2018).
47. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370, 20130624 (2015).
48. Gansauge, M.-T. et al. Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Res.* 45, e79 (2017).
49. Gansauge, M.-T., Aximu-Petri, M., Nagel, K. & Meyer, M. Manual and automated preparation of single-stranded DNA libraries for the sequencing of DNA from ancient biological remains and other sources of highly degraded DNA (In press). *Nature Protocols*.
50. Briggs, A. W. et al. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res.* 38, e87 (2010).
51. Fu, Q. et al. A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* 23, 553-559 (2013).
52. Fu, Q. et al. An early modern human from Romania with a recent Neanderthal ancestor. *Nature* 524, 216-219 (2015).
53. Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522, 207-211 (2015).
54. Mathieson, I. et al. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* 528, 499-503 (2015).
55. 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).
56. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589-595 (2010).
57. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* 15, 356 (2014).
58. Narasimhan, V. M. et al. The formation of human populations in South and Central Asia. *Science* 365, (2019).
59. McClure, S. B., Puchol, O. G. & Culleton, B. J. Ams Dating of Human Bone from Cova De La Pastora: New Evidence of Ritual Continuity in the Prehistory of Eastern Spain. *Radiocarbon* 52, 25-32 (2010).
60. 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. *Journal of Texas Archeology and History* 1, 94-126 (2014).
61. Passariello, I. et al. Characterization of Different Chemical Procedures for <sup>14</sup>C Dating of Buried, Cremated, and Modern Bone Samples at Circe. *Radiocarbon* 54, 867-877 (2012).
62. Reimer, P. J. et al. IntCal13 and Marine13 Radiocarbon Age Calibration Curves 0-50,000 Years cal BP. *Radiocarbon* 55, 1869-1887 (2013).
63. Lindo, J. et al. The genetic prehistory of the Andean highlands 7000 years BP through European contact. *Sci Adv* 4, eaau4921 (2018).
64. Moreno-Mayar, J. V. et al. Early human dispersals within the Americas. *Science* 362, (2018).



65. Scheib, C. L. et al. Ancient human parallel lineages within North America contributed to a coastal expansion. *Science* 360, 1024-1027 (2018).
66. Raghavan, M. et al. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* 349, aab3884 (2015).
67. Mallick, S. et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* 538, 201-206 (2016).
68. Patterson, N. et al. Ancient admixture in human history. *Genetics* 192, 1065-1093 (2012).
69. Lazaridis, I. et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 513, 409-413 (2014).
70. Skoglund, P. et al. Genetic evidence for two founding populations of the Americas. *Nature* 525, 104-108 (2015).
71. Olalde, I. et al. The genomic history of the Iberian Peninsula over the past 8000 years. *Science* 363, 1230-1234 (2019).
72. Nakatsuka, N. et al. A Paleogenomic Reconstruction of the Deep Population History of the Andes. *Cell* (2020) doi:10.1016/j.cell.2020.04.015.
73. Skoglund, P. et al. Genomic insights into the peopling of the Southwest Pacific. *Nature* 538, 510-513 (2016).
74. Harney, É. et al. Ancient DNA from Chalcolithic Israel reveals the role of population mixture in cultural transformation. *Nat. Commun.* 9, 3336 (2018).
75. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* 2, e190 (2006).
76. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655-1664 (2009).
77. Alexander, D. H. & Lange, K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* 12, (2011).
78. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7 (2015).
79. Fu, Q. et al. The genetic history of Ice Age Europe. *Nature* 534, 200-205 (2016).
80. Chintalapati, M., Neel, A., Patterson, N. & Moorjani, P. Reconstructing the spatio-temporal patterns of admixture in human history. (In Preparation).

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

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

**Competing interests** The authors declare no competing interests.

**Additional information**

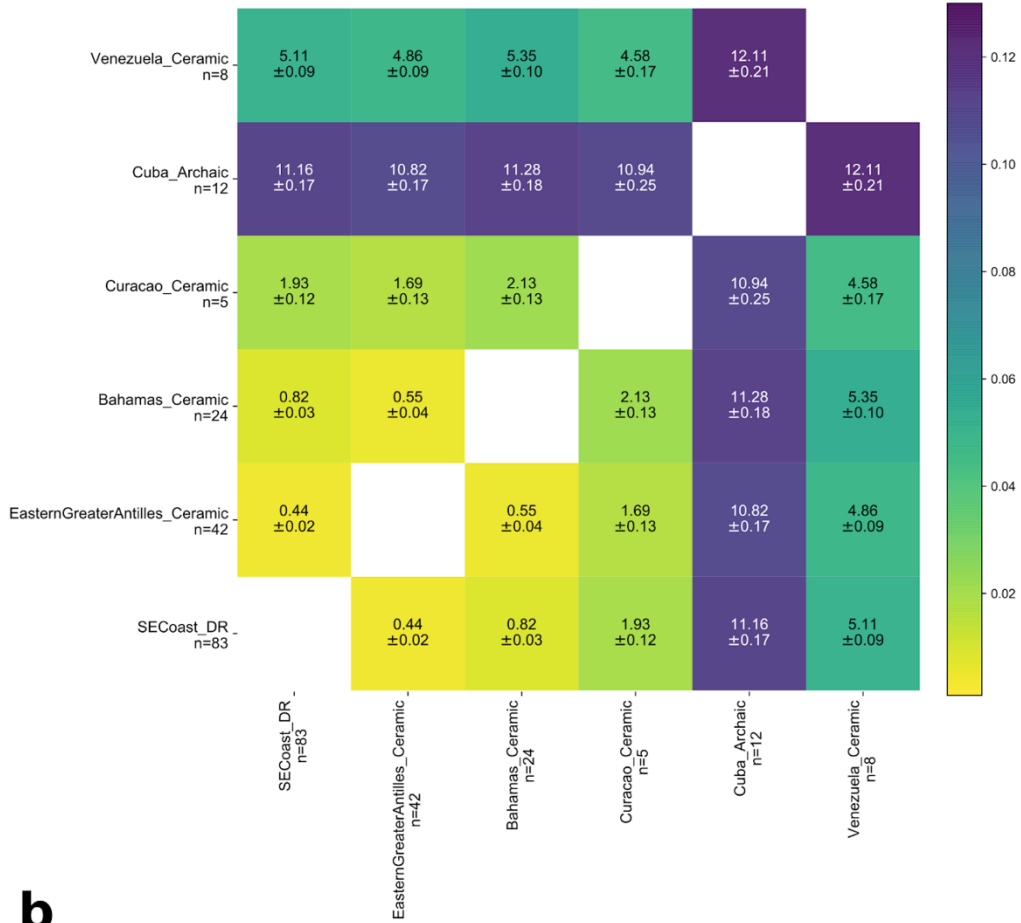
**Extended data** is available for this paper at xxxxxxxx.

**Supplementary information** is available for this paper at xxxxxxxx.

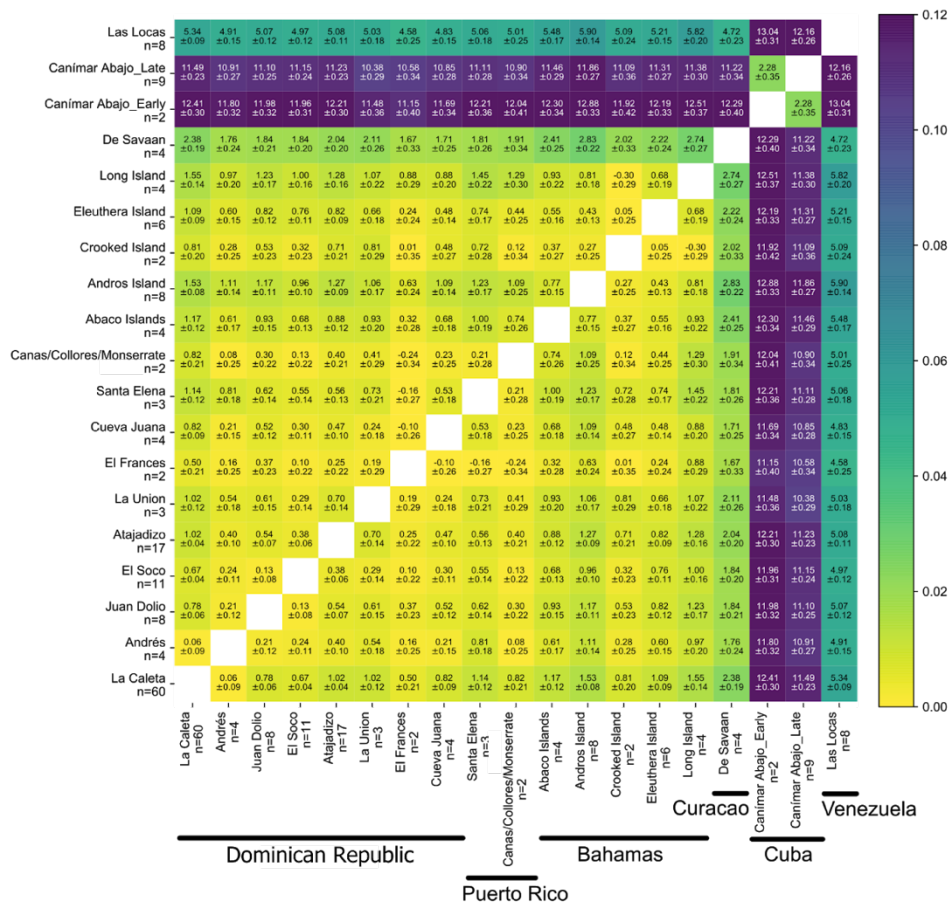
**Reprints and permissions information** is available at xxxxxxxx.

**Correspondence and requests for materials** should be addressed to D.F., K.Si., A.Cop., R.P., or D.R.

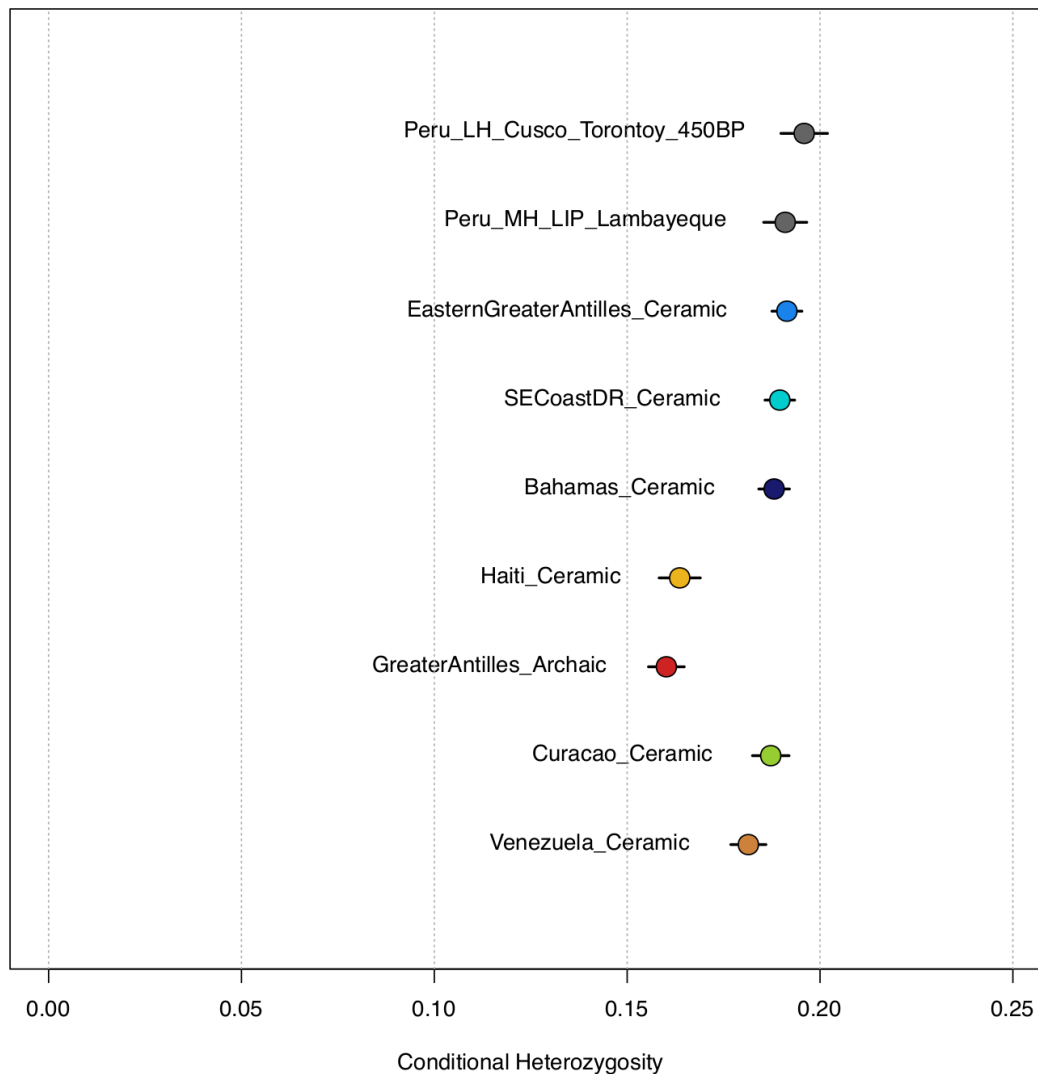
**a**



**b**



**Extended Data Fig. 1 |  $F_{ST}$  distances by site and clades.** Average pairwise  $F_{ST}$  distances (x100) between (a) clades and (b) sites with more than two unrelated individuals, demonstrating both overall high levels of genetic similarity between the *Caribbean\_Ceramic* sub-clades and the sites composing them, as well as the magnitude of genetic differentiation between those and the groups with Archaic- and Venezuela-related ancestries.



**Extended Data Fig. 2: Conditional heterozygosity by clade.** Conditional heterozygosity in the ancient Caribbean was similar to that of contemporaneous groups from Peru, except for the Archaic-associated groups, *Venezuela\_Ceramic*, and admixed Ceramic-associated populations (Nakatsuka et al. 2020). Bars represent 3 standard errors.

**Extended Data Table 1:  $2N_e$  estimates for each site.** Table includes all individuals where ROH analysis is possible and excludes individuals with more than 50cM sum of 20cM long ROH.

$2N_e$ estimate	STD	95% CI (low)	95% CI (high)	n	Site (Country)	Clade
12670	63	1146	1394	51	La Caleta (Dominican Republic)	Caribbean_Ceramic
1673	340	1007	2339	4	Andrés (Dominican Republic)	Caribbean_Ceramic
2298	411	1493	3104	7	Juan Dolio (Dominican Republic)	Caribbean_Ceramic
1814	255	1314	2314	9	El Soco (Dominican Republic)	Caribbean_Ceramic
1549	152	1252	1846	16	Atajadizo (Dominican Republic)	Caribbean_Ceramic
1850	455	958	2741	3	La Union (Dominican Republic)	Caribbean_Ceramic
1302	332	652	1953	2	El Frances (Dominican Republic)	Caribbean_Ceramic
1966	498	990	2942	3	Cueva Juana (Dominican Republic)	Caribbean_Ceramic
1850	455	959	2742	3	Santa Elena (Puerto Rico)	Caribbean_Ceramic
1223	303	629	1816	2	Canas/Collores/Monserrate (Puerto Rico)	Caribbean_Ceramic
943	169	612	1273	3	Abaco Islands (Bahamas)	Caribbean_Ceramic
1123	189	753	1493	4	Andros Island (Bahamas)	Caribbean_Ceramic
1220	302	628	1812	2	Crooked Island (Bahamas)	Caribbean_Ceramic
1394	259	886	1902	4	Eleuthera (Bahamas)	Caribbean_Ceramic
1906	582	765	3048	2	Diale 1(Haiti)	Haiti_Ceramic
937	205	535	1339	2	De Savaan (Curaçao)	Curacao_Ceramic
504	119	272	737	1	Andrés (Dominican Republic)	GreaterAntilles_Archaic_10126
482	78	328	636	2	Canimar Abajo: Early (Cuba)	GreaterAntilles_Archaic
899	193	521	1277	2	Canimar Abajo: Late (Cuba)	GreaterAntilles_Archaic
514	50	417	612	6	Las Locas (Venezuela)	Venezuela_Ceramic

**Extended Data Table 2: Ancestry proportion estimates with *qpAdm* of Indigenous, European, and African sources in present-day Caribbean individuals from Cuba, Dominican Republic, and Puerto Rico from the GOAL study (Moreno-Estrada et al. 2013). SGDP = Simons Genome Diversity Project outgroup populations Karitiana, Mixe, Yakut, Ulchi, Papuan, Mursi, and Mbuti; 1000G = 1000 Genomes outgroup populations PEL, PJL, JPT, and MSL.**

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 (1000G)	0.042	0.002	0.703	0.002	0.255	0.001
Dominican Republic (SGDP)	0.058	0.003	0.557	0.006	0.384	0.004
Dominican Republic (1000G)	0.062	0.002	0.559	0.004	0.380	0.003
Puerto Rico (SGDP)	0.132	0.004	0.686	0.006	0.182	0.003
Puerto Rico (1000G)	0.140	0.003	0.676	0.003	0.184	0.002

**Extended Data Table 3: Ancestry proportion estimates with *qpAdm* of Indigenous, European, African, and East Asian sources in present-day individuals across different Cuban provinces. Outgroup populations used: PEL, PJJ, JPT, MSL and GIH.**

Province	1000 Genomes CEU		1000 Genomes YRI		Caribbean_Ceramic		1000 Genomes CHB	
	Proportion	SE	Proportion	SE	Proportion	SE	Proportion	SE
Artemisa	0.834	0.005	0.100	0.003	0.038	0.004	0.028	0.004
Camaguey	0.617	0.004	0.297	0.002	0.074	0.003	0.013	0.003
Ciego_de_Avila	0.788	0.004	0.145	0.002	0.057	0.003	0.010	0.003
Cienfuegos	0.740	0.004	0.220	0.003	0.028	0.003	0.012	0.003
Granma	0.567	0.003	0.271	0.002	0.144	0.003	0.018	0.002
Guantanamo	0.549	0.003	0.363	0.003	0.083	0.002	0.004	0.002
Holguin	0.655	0.003	0.237	0.002	0.095	0.002	0.013	0.002
La_Habana	0.694	0.003	0.257	0.002	0.033	0.002	0.015	0.002
Las_Tunas	0.725	0.007	0.161	0.004	0.113	0.005	0.001	0.005
Matanzas	0.818	0.003	0.140	0.002	0.016	0.003	0.026	0.003
Mayabeque	0.889	0.005	0.940	0.003	0.012	0.004	0.005	0.004
Pinar_deL_Rio	0.727	0.003	0.227	0.002	0.036	0.002	0.010	0.002
Sancti_Spiritus	0.809	0.003	0.108	0.002	0.065	0.003	0.018	0.003
Santiago_de_Cuba	0.501	0.003	0.417	0.002	0.076	0.002	0.006	0.002
Villa_Clara	0.813	0.003	0.106	0.002	0.066	0.002	0.016	0.002

**Extended Data Table 4: Present day Indigenous American populations used per analysis, their broad language family attribution, and dataset of origin. IL= Illumina, HO = Human Origins.**

Population	Language Family	Included in PCA	Included in ADMIXTURE	Included in qpWave/qpAdm	Included in Treemix	Included in outgroup-f3	Included in language group f4-statistics
Chipewyan	Na-Dene	-	Yes (IL, n=2)	Yes (HO, n=28), Yes (IL, n=2)	Yes (IL, n=2)	-	-
Aymara	Andean	Yes (IL, n=4)	Yes (IL, n=4)	-	Yes (IL, n=4)	Yes (IL, n=4)	-
Quechua	Andean	Yes (IL, n=1)	Yes (IL, n=1)	Yes (HO, n=7), Yes (IL, n=1)	Yes (IL, n=1)	Yes (IL, n=1)	-
Arara	Carib	Yes (IL, n=1)	Yes (IL, n=1)	Yes (HO, n=4), Yes (IL, n=1)	Yes (IL, n=1)	Yes (IL, n=1)	Yes (IL, n=1)
Apalai	Carib	-	-	Yes (HO, n=4)	-	-	-
Yukpa	Carib	Yes (1240K, n=1)	Yes (1240K, n=1)	Yes (HO, n=1)	-	-	-
Mixtec	Central Amerind	Yes (1240K, n=1)	Yes (1240K, n=1)	Yes (HO, n=10)	-	-	-
Zapotec	Central Amerind	Yes (IL, n=5)	Yes (IL, n=5)	Yes (HO, n=10), Yes (IL, n=5)	Yes (IL, n=5)	Yes (IL, n=5)	-
Nahua	Central Amerind	Yes (1240K, n=2)	Yes (1240K, n=2)	-	-	-	-
Pima	Central Amerind	Yes (IL, n=14)	Yes (IL, n=14)	-	Yes (IL, n=14)	Yes (IL, n=14)	-
Tepehuano	Central Amerind	Yes (IL, n=2)	Yes (IL, n=2)	-	Yes (IL, n=2)	Yes (IL, n=2)	-
Bribri	Chibchan	Yes (IL, n=2)	Yes (IL, n=2)	-	Yes (IL, n=2)	Yes (IL, n=2)	Yes (IL, n=2)
Cabecar	Chibchan	Yes (IL, n=24)	Yes (IL, n=24)	Yes (HO, n=6), Yes (IL, n=24)	Yes (IL, n=24)	Yes (IL, n=24)	Yes (IL, n=24)
Guaymi	Chibchan	Yes (IL, n=5)	Yes (IL, n=5)	-	Yes (IL, n=5)	Yes (IL, n=5)	Yes (IL, n=5)
Kogi	Chibchan	Yes (IL, n=4)	Yes (IL, n=4)	-	-	Yes (IL, n=4)	Yes (IL, n=4)
Maleku	Chibchan	Yes (IL, n=2)	Yes (IL, n=2)	-	Yes (IL, n=2)	Yes (IL, n=2)	Yes (IL, n=2)
Teribe	Chibchan	Yes (IL, n=2)	Yes (IL, n=2)	-	Yes (IL, n=2)	Yes (IL, n=2)	Yes (IL, n=2)
Waunana	Chocoan	Yes (IL, n=3)	Yes (IL, n=3)	-	-	Yes (IL, n=3)	Yes (IL, n=3)
Embera	Chocoan	Yes (IL, n=5)	Yes (IL, n=5)	-	-	Yes (IL, n=5)	Yes (IL, n=5)
Guahibo	Guajiboan	Yes (IL, n=6)	Yes (IL, n=6)	-	Yes (IL, n=6)	Yes (IL, n=6)	Yes (IL, n=6)
Jamamadi	Macro-Arawakan	Yes (IL, n=1)	Yes (IL, n=1)	-	Yes (IL, n=1)	Yes (IL, n=1)	Yes (IL, n=1)
Palikur	Macro-Arawakan	Yes (IL, n=2)	Yes (IL, n=2)	-	Yes (IL, n=2)	Yes (IL, n=2)	Yes (IL, n=2)
Piapoco	Macro-Arawakan	Yes (IL, n=6)	Yes (IL, n=6)	Yes (HO, n=4), Yes (IL, n=6)	Yes (IL, n=6)	Yes (IL, n=6)	Yes (IL, n=6)
Wayuu	Macro-Arawakan	Yes (IL, n=3)	Yes (IL, n=3)	Yes (HO, n=1)	-	Yes (IL, n=3)	Yes (IL, n=3)
Chane	Macro-Arawakan	Yes (IL, n=2)	Yes (IL, n=2)	-	Yes (IL, n=2)	Yes (IL, n=2)	Yes (IL, n=2)
Toba	Mataco-Guaicuru	Yes (IL, n=2)	Yes (IL, n=2)	-	Yes (IL, n=2)	Yes (IL, n=2)	Yes (IL, n=2)
Wichi	Mataco-Guaicuru	Yes (IL, n=4)	Yes (IL, n=4)	-	Yes (IL, n=4)	Yes (IL, n=4)	Yes (IL, n=4)



<b>Kaqchikel</b>	Northern Amerind	Yes (IL, n=1)	Yes (IL, n=1)	-	Yes (IL, n=1)	Yes (IL, n=1)	-
<b>Maya</b>	Northern Amerind	Yes (IL, n=1)	Yes (IL, n=1)	-	-	-	-
<b>Mixe</b>	Northern Amerind	Yes (IL, n=9)	Yes (IL, n=9)	Yes (HO, n=10), Yes (IL, n=9)	Yes (IL, n=9)	Yes (IL, n=9)	-
<b>Guarani</b>	Tupian	Yes (IL, n=3)	Yes (IL, n=3)	-	Yes (IL, n=3)	Yes (IL, n=3)	Yes (IL, n=3)
<b>Karitiana</b>	Tupian	Yes (IL, n=13)	Yes (IL, n=13)	Yes (HO, n=16), Yes (IL, n=13)	Yes (IL, n=13)	Yes (IL, n=13)	Yes (IL, n=13)
<b>Parakana</b>	Tupian	Yes (IL, n=1)	Yes (IL, n=1)	-	Yes (IL, n=1)	Yes (IL, n=1)	Yes (IL, n=1)
<b>Surui</b>	Tupian	Yes (IL, n=24)	Yes (IL, n=24)	Yes (HO, n=12), Yes (IL, n=24)	Yes (IL, n=24)	Yes (IL, n=24)	Yes (IL, n=24)
<b>Ticuna</b>	Ticuna (isolate)	Yes (IL, n=5)	Yes (IL, n=5)	-	Yes (IL, n=5)	Yes (IL, n=5)	Yes (IL, n=5)
<b>Yaghan</b>	Yaghan (isolate)	Yes (IL, n=1)	Yes (IL, n=1)	-	Yes (IL, n=1)	Yes (IL, n=1)	-

**Extended Data Table 5: Statistics testing for an Australasian link.**

Test	$f_A(\text{Mbuti, Onge; Mixe, Test})$	Z-score	SNPs used
<i>Cuba_CanimarAbajo_Archaic</i>	0.000678	2.377	991294
<i>I10126_Archaic</i>	0.001291	3.380	741742
<i>Bahamas_Ceramic</i>	0.000556	2.318	1085703
<i>EasternGreaterAntilles_Ceramic</i>	0.000603	2.660	1105095
<i>SECoastDR_Ceramic</i>	0.000563	2.486	1112395
<i>Haiti_Ceramic</i>	0.000720	2.102	1015357
<i>Curacao_Ceramic</i>	0.000595	2.180	984268
<i>Venezuela_Ceramic</i>	0.000633	2.447	957964

Test	$f_A(\text{Mbuti, Papuan; Mixe, Test})$	Z-score	SNPs used
<i>Cuba_CanimarAbajo_Archaic</i>	0.000229	0.837	991925
<i>I10126_Archaic</i>	0.000696	1.853	742248
<i>Bahamas_Ceramic</i>	0.000384	1.751	1086365
<i>EasternGreaterAntilles_Ceramic</i>	0.000521	2.544	1105767
<i>SECoastDR_Ceramic</i>	0.000205	1.978	1113070
<i>Haiti_Ceramic</i>	0.000377	1.243	1015971
<i>Curacao_Ceramic</i>	0.000399	1.573	984884
<i>Venezuela_Ceramic</i>	0.000225	0.923	958591