Origins and genetic legacies of the Caribbean Taino

Hannes Schroeder[1,2], Martin Sikora[1], Shyam Gopalakrishnan[1], Lara M. Cassidy[3], Pierpaolo Maisano Delser[3,4], Marcela Sandoval Velasco[1], Joshua G. Schraiber[5], Simon Rasmussen[6], Julian R. Homburger[7], María C. Ávila-Arcos[8], Morton E. Allentoft[1], José Victor Moreno Mayar[1], Gabriel Renaud[1], Alberto Gómez Carballa[9], Jason E. Laffoon[2,10], Rachel J.A. Hopkins[11], Thomas F.G. Higham[11], Robert S. Carr[12], William C. Schaffer[13,14], Jane S. Day[15], Menno Hoogland[2], Antonio Salas[9], Carlos D. Bustamante[7], Rasmus Nielsen[1,16], Daniel G. Bradley[3], Corinne L. Hofman[2], Eske Willerslev[1,4,17]

[1]Center for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5–7, 1350 Copenhagen, Denmark, [2]Faculty of Archaeology, Leiden University, Van Steenis Building, Einsteinweg 2, 2333 CC Leiden, The Netherlands, [3]Smurfit Institute of Genetics, Trinity College Dublin, Dublin 2, Ireland, [4]Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK, [5]Department of Biology, BioLife Building (015-00), 1900 North 12th Street, Temple University, Philadelphia, PA 19122-6078, USA, [6]Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Kemitorvet, Building 208, 2800 Kongens Lyngby, Denmark, [7]Department of Genetics, School of Medicine, Stanford University, 300 Pasteur Drive, Lane Building, Stanford, CA 94305, USA, [8]International Laboratory for Human Genome Research, National Autonomous University of Mexico, Juriquilla 76230, Santiago de Querétaro, Mexico, [9]Unidade de Xenética, Departamento de Anatomía Patolóxica e Ciencias Forenses, Instituto de Ciencias Forenses, Facultade de Medicina, Universidade de Santiago de Compostela, and GenPoB Research Group, Instituto de Investigaciones Sanitarias (IDIS), Hospital Clínico Universitario de Santiago, 15872 Galicia, Spain, [10]Faculty of Sciences, Vrije Universiteit, De Boelelaan 1085, 1081 HV, Amsterdam, The Netherlands, [11]Oxford Radiocarbon Accelerator Unit, Research Laboratory for Archaeology and the History of Art, University of Oxford, South Parks Road, Oxford OX1 3QY, UK, [12]Archaeological and Historical Conservancy Inc., Davie, FL 33341, USA, [13]Liberal Arts Department, Phoenix College, Phoenix, AZ 85013, USA, [14]Center for Bioarchaeological Research, School of Human Evolution and Social Change, Arizona State University, Tempe, AZ 85281, USA, [15]Research Atlantica Inc., Boca Raton, FL 33432, USA, [16]Department of Integrative Biology, University of California, Berkeley, 4134 Valley Life Sciences Building, Berkeley, California 94720, USA, [17]Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK

Submitted to Proceedings of the National Academy of Sciences of the United States of America

The Caribbean was one of the last parts of the Americas to be settled by humans, but how and when the islands were first occupied remains a matter of debate. Ancient DNA can help answering these questions, but the work has been hampered by poor DNA preservation. We report the genome sequence of a 1,000-year-old Lucayan Taino individual recovered from the site of Preacher's Cave in the Bahamas. We sequenced her genome to 12.4-fold coverage and show that she is genetically most closely related to present-day Arawakan speakers from northern South America, suggesting that the ancestors of the Lucayans originated there. Further, we find no evidence for recent inbreeding or isolation in the ancient genome, suggesting that the Lucayans had a relatively large effective population size. Finally, we show that the native American components in some present-day Caribbean genomes are closely related to the ancient Taino, demonstrating an element of continuity between pre-contact populations and present-day Latino populations in the Caribbean.

ancient DNA | archaeology | migration | palaeogenomics

When Columbus set foot in the Americas, the so-called "Taino" were the dominant group in the Greater Antilles, the northern Lesser Antilles and the Bahamas, where they were known as the Lucayans (1). The ancestors of the Taino are thought to have been Arawakan speakers who entered the Caribbean from South America, starting as early as 2,500 years cal BP (2). The Bahamas were not settled until a thousand years later, as part of the Ostionoid expansion that started around 1,400 years cal BP (1). Opinions vary as to where these migrations originated, but archaeological and linguistic evidence suggests strong links with South America (2). Some scholars trace their origins to the Amazon basin, where the Arawakan languages developed (3). Others have argued for an origin further west in the Colombian Andes, connected with the Arhuaco and other Chibchan-speaking groups (4). The differences in opinion illustrate the difficulty of tracing population movements based on a patchy archaeological record.

Modern DNA studies (5, 6) also point to South America, but they are complicated by the fact that modern Caribbean genomes are largely composed of African and European ancestry and that only relatively little indigenous Caribbean ancestry remains (5–7). Furthermore, it is unclear whether this native component reflects Taino ancestry or whether it reached the Caribbean as a result of later population movements and migrations. The key to solving these issues lies in ancient DNA, but so far ancient DNA studies in the Caribbean have been hampered by poor preservation (8) and the few studies that exist are limited to mitochondrial DNA and, therefore, lack in resolution (9–11).

Results and Discussion

Here we report the genome sequence of a Lucayan Taino individual who lived in the Bahamas ~500 years prior to European contact. We sequenced the genome to an average depth of 12.4×, using whole genome enrichment and High-Throughput Sequencing. The sequence was obtained from a tooth excavated at the site of Preacher's Cave, which is located on the island of Eleuthera in the Bahamas (SI Appendix, Fig. S1) (12). The tooth was directly dated to 1,082 ±29 ^{14}C years BP (cal AD 776-992) (SI Appendix, section 3) and strontium isotope analysis suggests that the individual grew up locally in the Bahamas (SI Appendix,

Significance

Ancient DNA has revolutionized the field of archaeology, but in the Caribbean and other tropical regions of the world the work has been hampered by poor DNA preservation. We present the first ancient human genome from the Caribbean and use it to shed light on the early peopling of the islands. We demonstrate that the ancestors of the so-called "Taino" who inhabited large parts of the Caribbean in pre-Columbian times, originated in northern South America, and we find evidence that they had a comparatively large effective population size. We also show that the native components in some modern Caribbean genomes are closely related to the ancient Taino, suggesting that indigenous ancestry in the region has survived until the present day.

Reserved for Publication Footnotes

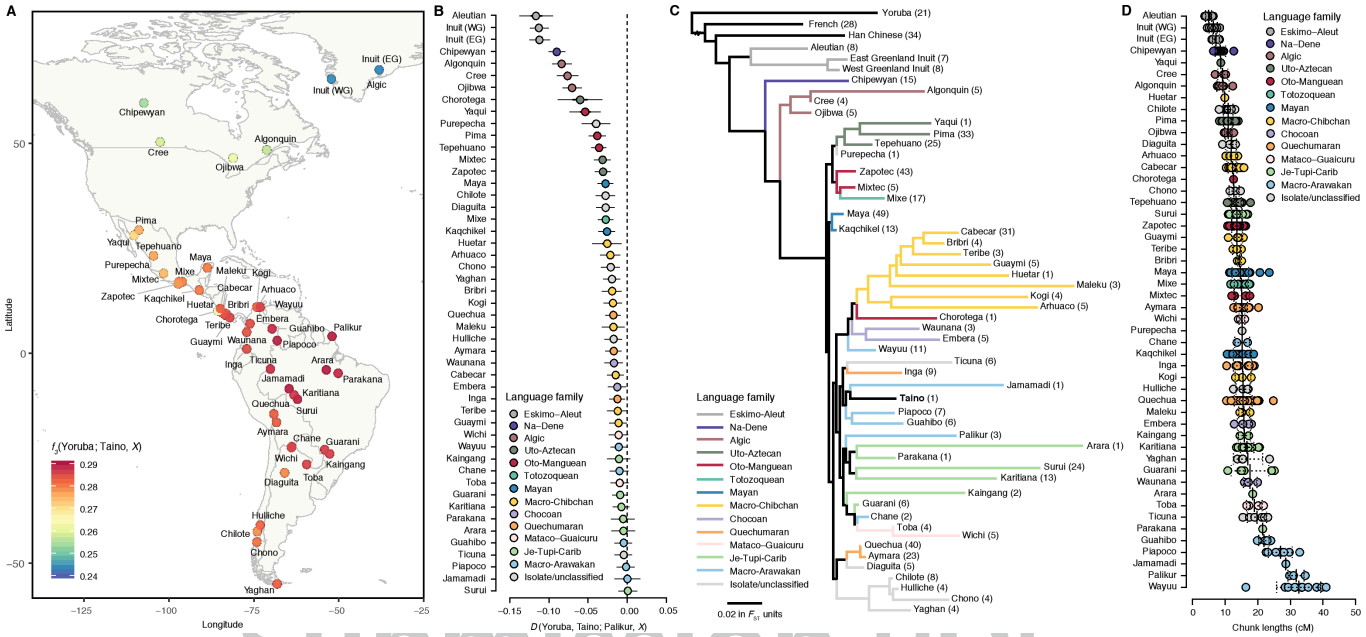

Fig. 1. The genetic origins of the Taino. The individual from Preacher's Cave is most closely related to Arawakan and Cariban speakers from the Amazon and Orinoco basins. **(A)** Heat map of outgroup f_3-statistics testing (Yoruba; Taino, *X*) where *X* is one of 50 Native American populations (24). Warmer colors indicate higher levels of allele sharing. **(B)** We computed *D*-statistics of the form *D*(Yoruba, Taino; Palikur, *X*) to test if any other group is more closely related to the Taino than the Palikur. Thick and thin whiskers represent 1 and 3 standard errors, respectively. **(C)** Neighbor-joining tree based on F_{st} distances. **(D)** Expected total length (cM) of shared haplotypes between the Taino and 50 Native American groups based on ChromoPainter analysis (26). To avoid the confounding effects of missing data, we ran ChromoPainter (26) on the unmasked dataset. Horizontal bars mark mean values plus or minus the standard deviation. For language classification, see SI Text, section 1.

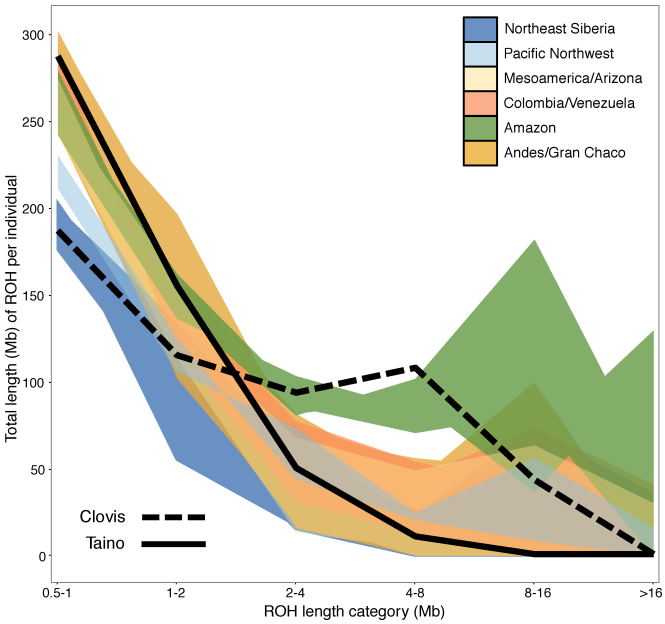

Fig. 2. Taino demography. Total estimated length of genomic ROH for the Taino and the Clovis genome (13) and selected Native American and Siberian genomes (15, 31, 32) in a series of length categories. ROH distributions for modern individuals have been condensed into population-level silhouettes (SI Text, section 14).

section 4). All DNA libraries displayed features typical for ancient DNA, including short average fragment lengths, characteristic fragmentation patterns, and an increased frequency of apparent C-to-T substitutions at the 5′ end of DNA molecules (SI Appendix, section 8). Contamination was estimated to be around 0.1-1.2% (SI Appendix, section 9), which is within the normal range observed for other ancient genomes (13–15) and unlikely to affect downstream analyses (16).

Chromosomal sex and mitochondrial DNA. We determined the sex of the individual to be female, based on the number of reads mapping to the X and Y chromosomes, respectively (SI Appendix, section 10). The mitochondrial genome was sequenced to an average depth of ~167× and was placed at the root of Native American haplogroup B2 (SI Appendix, section 11). As one of the founding lineages of the Americas, B2 has a pan-American distribution among present-day Native Americans (17), although our analysis suggests that it occurs at higher frequency among South Americans (SI Appendix, Fig. S9). A close search of the literature on modern published mtDNAs from the Caribbean (7, 18–23) revealed no matches or closely related sequences (SI Appendix, section 11). Generally speaking, the B2 lineage appears to be quite rare in Caribbean populations today and, interestingly, it has not been previously detected in ancient populations from the region (9–11). It is possible, therefore, that haplotype B2 was relatively rare in the Caribbean in the past. Alternatively, it may have been lost during the dramatic population declines experienced by Caribbean populations after 1492 (5).

Genome-wide affinities. To assess the genome-wide affinities of the ancient Taino, we computed outgroup f_3-statistics of the form f_3(Yoruba; Taino, *X*), where *X* is one of 50 Native American groups from a previously published dataset (24) that we used as reference. Due to high levels of recent European and African admixture in many Native Americans, those genomic segments were excluded prior to analysis (SI Appendix, section 12). We find that the ancient Taino is most closely related to the Palikur and other Arawakan speakers from the Amazon and Orinoco basins (Fig. 1A). We observe similar affinities using *D*-statistics (Fig. 1B), principal component analysis (SI Appendix, Fig. S10), and a neighbor-joining tree based on pairwise F_{ST} distances, which places the Taino on the same branch as other Arawakan speakers (Fig. 1C). These results are further supported by ADMIXTURE

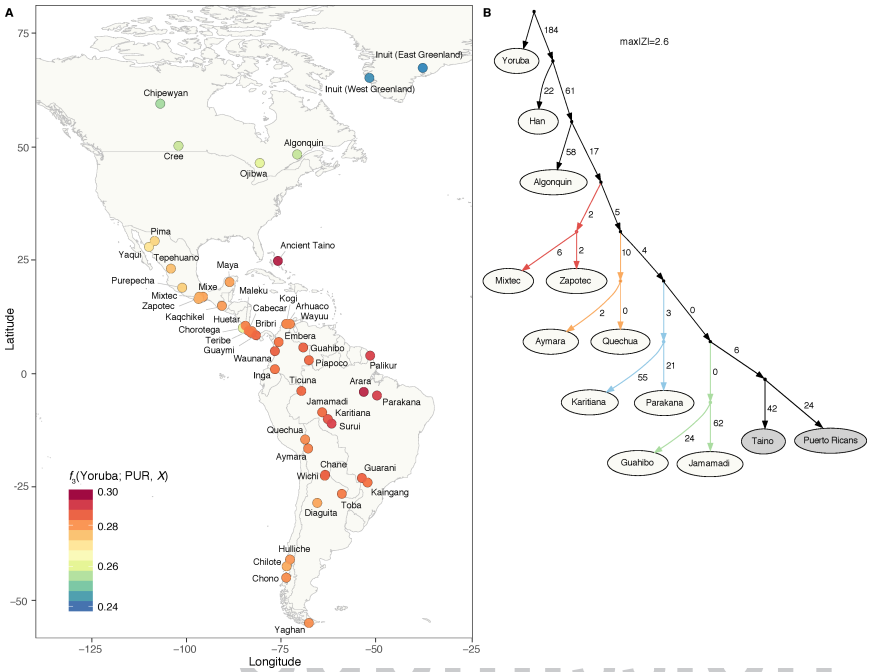

Fig. 3. The genetic legacy of the Taino. (A) Heat map showing the amount of allele sharing between the Native American component in present-day Puerto Ricans, Native Americans, and the Taino. Warmer colors indicate higher levels of allele sharing. **(B)** Model of Native American population history that fits the patterns of observed allele frequencies in our dataset (max|Z|=2.6). The Taino and masked Puerto Ricans form a clade that branches off the South American lineage. Branches are colored by language family (SI Text, section 1). Drift values are shown in units proportional to F_{ST} × 1,000.

(25) results, which show that the Taino has ancestry proportions similar to those of the Palikur and other Arawakan speakers (SI Appendix, Fig. S11).

To further explore the ancestry of the ancient Taino we used the haplotype-based approach implemented in ChromoPainter (26). By leveraging linkage information, haplotype- based approaches are more powerful in detecting fine-scale structure than those using unlinked loci. To avoid the confounding effects of missing data, we ran ChromoPainter (26) on the unmasked dataset. As expected, we observe the highest levels of shared haplotypes between the Taino and Arawakan speakers, which strikingly provide all the top hits in the analysis, as shown in Figure 1D. Interestingly, this includes admixed groups, such as the Wayuu, who were not picked up in the SNP-based analyses, probably as a result of additional gene flow from the Isthmo-Colombian area, which can be seen as the light blue component in the ADMIXTURE result (SI Appendix, Fig. S11).

We also specifically looked for traces of Australasian ancestry in the Taino genome, since previous studies (27) have found surprising affinities between some Amazonian populations (e.g., Surui) and populations from Melanesia, Australia, and the Andaman Islands. Using *D*-statistics of the form *D*(Yoruba, *X*; Mixe, Taino) computed with the Affymetrix Human Origins SNP array data (28), we do not detect the same excess affinity in the ancient Taino (SI Appendix, Fig. S12), suggesting either that the signal was somehow lost in the Taino or that it entered Amazonian populations after the divergence from the Taino within the last 3,000 years.

Runs of homozygosity. Next, we analysed the ancient genome for runs of homozygosity (ROH) to investigate the demographic history of the Taino (SI Appendix, section 14). ROH can inform about past demography; short ROH being indicative of ancient restrictions in effective population size, while longer ROH reflect recent episodes of isolation and/or inbreeding (29, 30). Figure 2 plots the ROH distributions for the ancient Taino genome and the Clovis genome (13), against a backdrop of 53 modern Native American and Siberian genomes (15, 31, 32). As previously observed (29, 30), all Native American genomes, including the Taino, show clear evidence for having undergone one or more ancestral population bottlenecks, as indicated by the excess of shorter (< 2 Mb) ROH. This is consistent with the proposed occurrence of an extreme founder event on entry to the continent, followed by successive bottlenecks (33, 34). Interestingly, the Clovis genome (13) (~12,600 BP) appears to provide a snapshot of one such early bottleneck. The individual does not share the same excess of shorter runs seen in modern Native Americans, but instead exhibits inflated ROH coverage between 2-8 Mb. The relatively low level of shorter ROH could argue against an extremely long or intense Beringian Incubation Model, which states that the people who eventually colonized the Americas descended from a small population that spent up to 15,000 years isolated on the Bering Land Bridge before entering the Americas (35).

At the other end of the spectrum, the Taino genome displays some of the lowest levels of longer (>8 Mb) ROH of any Native American genomes (Fig. 2). This argues against a history of recent isolation or inbreeding in the Lucayan population and suggests that the Lucayans had a relatively large effective population size. Based on the distribution of longer ROH (≥ 1.6 Mb), we estimate an effective size of around 1,600 individuals, which is considerably higher than our estimates for some present-day South American populations, such as the Karitiana and Surui (Table S13). However, the island of Eleuthera measures only around 518 square km and it is difficult to imagine how this community was able to sustain such a relatively large effective size without outside contact. Current thinking suggests that Caribbean communities were highly mobile and maintained pan-regional networks that extended far beyond the local scale (36, 37). Our results are consistent with this view. Evidently, these networks did not only involve the exchange of goods and ideas, as evidence by archaeology, but also of genes. With the arrival of Europeans, however, these networks were disrupted, which may have contributed to the catastrophic population declines suffered by Caribbean communities soon after contact (38).

Genetic legacies. Previous studies (5–7) have shown that the amount of Native American ancestry in modern Caribbean populations varies widely across the region. While some retain substantial amounts of Native American ancestry, others are largely

composed of African and/or European ancestry (5–7). Puerto Ricans, for example, harbour between 10-15% Native American ancestry; however, it is unclear to what extent this component reflects Taino ancestry. To address this issue, we added 104 modern Puerto Rican genomes from the 1000 Genomes Project (39) to our dataset and used the clustering algorithm ADMIXTURE (25) to estimate the composition of genetic ancestry in each individual (SI Appendix, Fig. S13). Due to the high levels of African and European ancestry in modern Puerto Ricans the native components are difficult to discern; however, when we compare only the estimated ancestry clusters that reflect non-African/European ancestries, there are clear similarities between Puerto Ricans, Arawakan speakers, and the ancient Taino (SI Appendix, Figs. S14 and S15).

To explore these relationships further, we then masked segments of African and European ancestry in the Puerto Rican genomes (SI Appendix, section 12) and computed a set of outgroup f_3-statistics to assess the amount of shared drift between Puerto Ricans, other present-day Native Americans, and the ancient Taino. The results are shown in Figure 3A and demonstrate that Puerto Ricans share more drift with the Taino than any other native American group in our dataset. To formally test this relationship, we then computed two sets of D-statistics of the form D(YRI, Taino; PUR, X) and D(YRI, X; Taino, PUR), where X is the test population. Results are consistent with Puerto Ricans and the ancient Taino forming a clade without any significant gene-flow post-divergence (SI Appendix, Fig. S16). To test whether other present-day Latino populations in the Caribbean share the same affinities with the ancient Taino, we repeated the analyses with SNP array data for a more diverse set of Caribbean populations from Haiti, Cuba, and the Dominican Republic (5); however, due to the low amounts of Native American ancestry in these populations we were unable to replicate the results.

Finally, we tried to fit both the ancient Taino and masked Puerto Ricans on a previously defined admixture graph (24). Figure 3B shows a model that is a good fit to the data in the sense that none of the predicted f-statistics are more than three standard errors from what is observed ($\max|Z|=2.6$). In this model, the ancient Taino and masked Puerto Ricans form a clade that branches off the main South American lineage. By contrast, a model where Puerto Ricans are added as direct descendants of the Taino does not fit the data (SI Appendix, Fig. S18). To determine if patterns of allele frequencies in modern Puerto Ricans and the ancient Taino individual are compatible with direct ancestry we then used a recently developed likelihood ratio test (40). While the test rejects the hypothesis of direct ancestry, it also shows that the ancient Taino only recently diverged from the ancestors of modern Puerto Ricans (SI Appendix, Table S15). This result is mirrored in the ChromoPainter analysis (26), which shows that Puerto Ricans share large parts of their genomes with the ancient Taino, despite significant European and African admixture (SI Appendix, Fig. S19).

Conclusion

Our study provides a first glimpse of the initial peopling of the Caribbean from an ancient genome perspective. Specifically, we were able to show that the Lucayan Taino were genetically most closely related to present-day Arawakan speakers from northern South America, suggesting that their ancestors originated there. However, we note that this does not preclude the possibility of other/earlier migrations to the Caribbean that originated elsewhere and more data will need to be collected to address this issue. Further, we find no evidence for recent isolation or inbreeding in the ancient genome, suggesting that the Lucayans had a comparatively large effective population size despite their island location. This is consistent with archaeological evidence, which suggests that indigenous Caribbean communities were highly mobile and maintained complex regional networks of interaction and exchange that extended far beyond the local scale. Lastly, we find that the native component in present-day Puerto Rican genomes is closely related to the ancient Taino, demonstrating an element of continuity between pre-contact populations and present-day Latino populations in the Caribbean despite the disruptive effects of European colonization.

Material and Methods

Samples. The samples for this study were excavated at the site of Preacher's Cave, which is located on the northern part of the island of Eleuthera in the Bahamas (SI Appendix, Fig. S1). During excavations in 2007 (SI Appendix, section 2), a total of six Lucayan primary burials were discovered within the cave, three of which were well preserved (12). The three burials belonged to two adult males and one female, aged 20-35 years at the time of death (12). For the present study, we sampled five of the burials for isotopic and ancient DNA analysis and in the absence of petrous bones, we opted for teeth.

Radiocarbon dating. Radiocarbon dating was performed at the Oxford Radiocarbon Accelerator Unit (SI Appendix, section 3). The standard method for radiocarbon dating is measuring the amount of ^{14}C in collagen from bone or dentine. However, in tropical environments, collagen is often only poorly preserved or not at all (41) and since part of the dentine was used for DNA analysis, the remaining sample was very small. Consequently, we turned to the enamel fraction. Chemical pre-treatment was done as described in (42) in order to remove labile carbonates on crystal surfaces and grain boundaries. While the procedure is still far from being standardised, it is thought to provide a reliable *terminus ante quem* (SI Appendix, section 3).

Isotope analyses. We conducted multiple isotope (Sr, C, O) analyses in order to determine whether the individuals buried in the cave were of local or non-local origin. The logic behind this approach is that it cannot be reasonably argued that an individual's ancestors were local and thus 'representative' of a particular population if the individual was in fact not local, but a first-generation migrant, especially if the results indicate long-distance migration. This is an especially important consideration for the ancient Antilles where high rates of migration have been documented for various time periods (43). The analytical procedure is described in the supplement (SI Appendix, section 4).

DNA extraction and library preparation. DNA was extracted from ~100 mg of starting material (SI Appendix, section 5). Thirty µl of each DNA extract was then built into DNA libraries using Illumina specific adapters (44). Ten µl of the DNA libraries were then amplified and indexed in 50 µl PCR reactions using sample-specific barcodes, as described in (45). The optimal number of PCR cycles was determined by qPCR. The amplified libraries were purified using AMPure XP beads (Beckman Coulter), quantified on a 2200 TapeStation (Agilent Technologies), pooled in equimolar amounts, and sequenced on an Illumina HiSeq 2500 run in SR mode. The results of the screening run are shown in Table S4. As expected, all the samples yielded extremely low endogenous DNA contents, except one (PC537), which turned out to be exceptionally well preserved (SI Appendix, Table S4).

Whole genome capture and deep-sequencing. Following the initial screening run, we built three more libraries for PC537 and enriched them using the MYbaits Human Whole Genome Capture Kit (MYcroarray), following the manufacturer's instructions (46). The method makes use of biotinylated RNA probes transcribed from genomic DNA libraries to capture the human DNA in the library. The captured libraries were amplified for 10-12 cycles using primers IS5 and IS6 (44), purified using AMPure XP beads (Beckman Coulter), quantified, and sequenced as above. After capture, the endogenous fraction increased from around 13% to 35%, albeit with some loss in complexity. We then

sequenced the ancient genome to an average depth of 12.4× using a combination of shotgun and captured libraries.

Mapping. Basecalling was done with CASAVA-1.8.2. Only reads with correct indexes were kept. FASTQ files were filtered using AdapterRemoval (47) to remove adapter sequences, low quality stretches, and ambiguous bases at the ends of reads. The minimum length allowed after trimming was 25 nucleotides. Trimmed and filtered reads were mapped to GRCh37/hg19 (NCBI build 37.1) using bwa-0.7.5 (48), with the seed disabled to allow for better sensitivity (49) and filtering for reads with a minimum mapping quality of 30. The mitochondrial sequence in the reference was replaced by the revised Cambridge Reference Sequence (rCRS) (50). Clonal reads were removed using samtools-1.2.1 (51) rmdup function and bam files from different sequencing runs were merged using samtools-1.2.1 (51) merge.

Genotype calling. Diploid genotypes were called using samtools-1.2.1 (51) mpileup function (-C50 option) and bcftools-1.2.4 call with the consensus caller enabled. Genotype calls were filtered for a minimum depth of one third and a maximum depth of two times the average depth of coverage (12.4×). Subsequently, clustered variants were filtered out by removing SNPs/indels that were called within 5 base pairs of each other. Variants were also filtered for a Phred posterior probability of less than 30, strand bias or end distance bias p-value < 10^-4. Overall and type-specific error rates were estimated using ANGSD (52) (SI Appendix, section 7).

Genetic affinities. To explore the genetic affinities of the Taino individual, we merged the called ancient genome with a previously published SNP-chip dataset (24), which includes 493 individuals from 50 Native American populations genotyped at 346,465 SNPs (SI Appendix, Table S7) for which segments of European and African ancestry had been masked. In addition, we included 21 Yoruba, 34 Han Chinese and 28 French individuals from HGDP (53). Merging was done using PLINK 1.9 (54). While transition SNPs are sensitive to post-mortem damage, we opted to include all SNPs, since only ∼1% of the sites included in the reference panel involve transversions, corresponding to 4,856 sites. Genotypes where the ancient genome had a different allele to the ones observed in the reference panel were set to missing, resulting in a final dataset of 325,139 SNPs.

Outgroup f_3- and D-statistics. Outgroup f_3- and D-statistics were computed using *AdmixTools* (55). To estimate the amount of shared drift between the ancient Taino and present-day Native Americans we computed outgroup f_3-statistics of the form f_3(Yoruba; Taino, X), where X is one of 50 Native American populations in our dataset (24) (Fig. 1A). We then computed a set of allele-frequency based D-statistics of the form D(Yoruba, Taino; Palikur, X), where X is the test population, to test whether any other other Native American group in our dataset is more closely related to the Taino than the Palikur (Fig. 1B).

F_{ST} tree. We used EIGENSOFT (102) to calculate pairwise F_{ST} distances based on allele frequencies in our dataset. The distance matrix was then used to build a neighbor-joining tree using custom R scripts (available upon request). The phylogenetic tree in Figure 1C was rendered using FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

ADMIXTURE analysis. We ran ADMIXTURE (25) both on the masked and unmasked datasets using default parameters for K=2 to K=14 and diploid genotype calls for both the ancient genome and the modern reference populations. For the masked dataset, we removed individuals with more than 60% missing genotypes and any variants with call rates of less than 40%, resulting in a final dataset of 466 individuals typed at 346,418 SNPs. The unmasked dataset includes 1,112 individuals typed at 346,465 SNPs. For both datasets, we ran 100 replicates for each K and picked the one with the highest log-likelihood as result for that K (SI Appendix, Fig. S11 and S13). Figure S15 shows the ancestry proportions estimated for the unmasked dataset after removing the first three ancestry components (corresponding to African, Western European and East Asian ancestries), and normalizing the remaining ancestry clusters such that they sum to 1. Figure S15 displays the estimated ancestry proportions averaged by population/language group.

ChromoPainter analysis. To avoid the confounding effects of missing data, ChromoPainter (26) was run on the unmasked dataset. The dataset was split by chromosomes and phased using SHAPEIT 2.r837 (56). For the estimation of haplotypes, the 1000 Genome Project Phase3 dataset was used as reference panel including 2,504 individuals from 26 populations. Hap files were converted into ChromoPainter (26) format using the "impute2-chromopainter.pl" script, while recombination maps were produced with "convertrecfile.pl" (both scripts are available for download on the ChromoPainter website).

Runs of homozygosity. For the ROH analysis we merged the ancient Taino genome with 109 other modern Native American and Siberian genomes (96–98). The ancient Clovis genome (13) was also included. The dataset was then filtered for missingness and minor allele frequency, retaining only transversions, resulting in a final dataset of 583,623 SNPs. ROH were estimated using PLINK 1.9 (110), as described in the supplement (SI Appendix, section 14).

Australasian ancestry. To test whether the Taino genome harboured any traces of Australasian ancestry, we merged the ancient genome with the Human Origins dataset (28), which contains several other ancient genomes, as well as 2,345 contemporary humans typed at ∼600,000 SNPs on the Affymetrix Human Origins array. We then computed three sets of D-statistics of the form D(Yoruba, X; Mixe, Surui/Taino/Clovis), where X is one a subset of 59 populations in the Human Origins dataset (28), including Australians, Onge, and Papuans. We find that the Taino and the Clovis genome do not share the same excess affinity with Australasians as the Surui (SI Appendix, Fig. S12).

Genetic legacies. To explore the relationship between the ancient Taino and modern Caribbean populations, we added 104 modern Puerto Rican genomes from the 1000 Genomes Project (39) to our dataset and performed ADMIXTURE analysis (25) as described above (SI Appendix, Fig. S13). Outgroup f_3- and D-statistics were computed using *AdmixTools* (55), but due to the high levels of European and African ancestry in the Puerto Rican genomes those segments were masked prior to analysis (SI Text, section 12). The direct ancestry test was also performed on the masked data (SI Appendix, section 16) and the admixture graphs (Fig. 3B and SI Appendix, Fig. S18) were fitted using *qpGraph* from the *AdmixTools* package (55) (SI Appendix, section 17). The ChromoPainter (26) analysis was run on the unmasked dataset, as described above (SI Appendix, Fig. S19).

ACKNOWLEDGMENTS. We thank the staff at the Danish National High-Throughput Sequencing Centre for technical support and Juan Carlos Martínez-Cruzado and Tom Gilbert for their input and helpful discussions. We also thank Jorge Estevez from the Union Higuayagua for providing guidance and sharing his insights into the story of his people. We acknowledge the DJEI/DES/SFI/HEA Irish Centre for High-End Computing (ICHEC) for access to computational facilities and support and we are grateful to the genotoul bioinformatics platform Toulouse Midi-Pyrenees (Bioinfo Genotoul) for providing computing and storage resources. The research leading to these results was funded by the European Research Council (FP7/2007-2013, grant no. 319209, Synergy project NEXUS1492) and the Danish National Research Foundation (DNRF94) with additional support from the HERA Joint Research Programme "Uses of the Past" (CitiGen) and the European Commission through Horizon 2020 under grant agreement no. 649307. L.M.C. was funded by the Irish Research Council Government of Ireland Scholarship Scheme (GOIPG/2013/1219). J.S. received funding by the United States National Institutes of Health (NIH grant no. R35-GM124745). Author contributions: H.S. and C.L.H. conceived the study. H.S. and E.W. designed the experiments. H.S. and M.S.V. performed laboratory work. H.S., M.S., S.G., L.M.C., P.M.D., J.S., J.V.M.M., G.R., M.E.A., A.G.C., A.S., M.C.A.A, S.R., and J.R.H. performed genetic analyses. J.E.L. performed isotopic analyses. R.J.A.H and T.F.G.H. performed radiocarbon dating. R.S.C., W.C.S, M.H. and J.S.D.

provided samples and archaeological background information. H.S. wrote the manuscript with input from L.M.C., M.S., C.L.H., E.W. and the remaining authors. The mapped data are available for download through the European Nucleotide Archive under accession no. PRJEB22578.

1. Keegan WF (1992) *The people who discovered Columbus: the prehistory of the Bahamas* (University Press of Florida).
2. Keegan WF, Hofman CL (2017) *The Caribbean before Columbus* (Oxford University Press).
3. Rouse I (1993) *The Tainos: Rise and decline of the people who greeted Columbus* (Yale University Press).
4. Meggers BJ, Evans C (1978) Lowland South America and the Antilles. *Ancient Native Americans*, ed Jennings JD (W.H. Freeman, San Francisco), pp 543–591.
5. Moreno-Estrada A, et al. (2013) Reconstructing the Population Genetic History of the Caribbean. *PLoS Genet* 9(11):e1003925.
6. Gravel S, et al. (2013) Reconstructing Native American migrations from whole-genome and whole-exome data. *PLoS Genet* 9(12):e1004023.
7. Bryc K, et al. (2010) Genome-wide patterns of population structure and admixture among Hispanic/Latino populations. *Proc Natl Acad Sci USA* 107(2):8954–8961.
8. Schroeder H, et al. (2015) Genome-wide ancestry of 17th-century enslaved Africans from the Caribbean. *Proc Natl Acad Sci USA* 112(12):3669–3673.
9. Lalueza-Fox C, Calderón F, Calafell F, Morera B, Bertranpetit J (2001) MtDNA from extinct Tainos and the peopling of the Caribbean. *Ann Hum Genet* 65(2):137–151.
10. Lalueza-Fox C, et al. (2003) Mitochondrial DNA from pre-Columbian Ciboneys from Cuba and the prehistoric colonization of the Caribbean. *Am J Phys Anthropol* 121(2):97–108.
11. Mendisco F, et al. (2015) Where are the Caribs? Ancient DNA from ceramic period human remains in the Lesser Antilles. *Philos Trans R Soc Lond B Biol Sci* 370(1660):20130388.
12. Schaffer WC, Carr RS, Day JS, Pateman MP (2012) Lucayan-Taíno burials from Preacher's cave, Eleuthera, Bahamas. *Int J Osteoarchaeol* 22(1):45–69.
13. Rasmussen M, et al. (2014) The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* 506(7487):225–229.
14. Rasmussen M, et al. (2015) The ancestry and affiliations of Kennewick Man. *Nature* 523(7561):455–458.
15. Raghavan M, et al. (2015) Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* 349(6250):aab3884.
16. Racimo F, Renaud G, Slatkin M (2016) Joint Estimation of Contamination, Error and Demography for Nuclear DNA from Ancient Humans. *PLoS Genet* 12(4):e1005972.
17. Perego UA, et al. (2010) The initial peopling of the Americas: a growing number of founding mitochondrial genomes from Beringia. *Genome Res* 20(9):1174–1179.
18. Tajima A, et al. (2004) Genetic background of people in the Dominican Republic with or without obese type 2 diabetes revealed by mitochondrial DNA polymorphism. *J Hum Genet* 49(9):495–499.
19. Mendizabal I, et al. (2008) Genetic origin, admixture, and asymmetry in maternal and paternal human lineages in Cuba. *BMC Evol Biol* 8:213.
20. Wilson JL, et al. (2012) Forensic analysis of mtDNA haplotypes from two rural communities in Haiti reflects their population history. *J Forensic Sci* 57(6):1457–1466.
21. Vilar MG, et al. (2014) Genetic diversity in Puerto Rico and its implications for the peopling of the Island and the West Indies. *Am J Phys Anthropol* 155(3):352–368.
22. Torres JB, et al. (2015) Genetic Diversity in the Lesser Antilles and Its Implications for the Settlement of the Caribbean Basin. *PLoS One* 10(10):e0139192.
23. Madrilejo N, Lombard H, Torres JB (2015) Origins of marronage: Mitochondrial lineages of Jamaica's Accompong Town Maroons. *Am J Hum Biol* 27(3):432–437.
24. Reich D, et al. (2012) Reconstructing Native American population history. *Nature* 488(7411):370–374.
25. Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19(9):1655–1664.
26. Lawson DJ, Hellenthal G, Myers S, Falush D (2012) Inference of population structure using dense haplotype data. *PLoS Genet* 8(1):e1002453.
27. Skoglund P, et al. (2015) Genetic evidence for two founding populations of the Americas. *Nature* 525(7567):104–108.
28. Lazaridis I, et al. (2014) Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 513(7518):409–413.
29. Kirin M, et al. (2010) Genomic runs of homozygosity record population history and consanguinity. *PLoS One* 5(11):e13996.
30. Pemberton TJ, et al. (2012) Genomic patterns of homozygosity in worldwide human populations. *Am J Hum Genet* 91(2):275–292.
31. Mallick S, et al. (2016) The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* 538(7624):201–206.
32. Pagani L, et al. (2016) Genomic analyses inform on migration events during the peopling of Eurasia. *Nature* 538(7624):238–242.
33. Wallace DC, Garrison K, Knowler WC (1985) Dramatic founder effects in Amerindian mitochondrial DNAs. *Am J Phys Anthropol* 68(2):149–155.
34. Ramachandran S, et al. (2005) Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc Natl Acad Sci USA* 102(44):15942–15947.
35. Tamm E, et al. (2007) Beringian standstill and spread of Native American founders. *PLoS One* 2(9):e829.
36. Hofman CL, Bright AJ, Ramos RR (2010) Crossing the Caribbean Sea: towards a holistic view of pre-colonial mobility and exchange. *Journal of Caribbean Archaeology* 10:1–18.
37. Hofman C, Mol A, Hoogland M, Rojas RV (2014) Stage of encounters: migration, mobility and interaction in the pre-colonial and early colonial Caribbean. *World Archaeol* 46(4):590–609.
38. Cameron CM, Kelton P, Swedlund AC (2015) *Beyond Germs: Native Depopulation in North America* (University of Arizona Press).
39. 1000 Genomes Project Consortium, et al. (2015) A global reference for human genetic variation. *Nature* 526(7571):68–74.
40. Schraiber J (2017) Assessing the relationship of ancient and modern populations. *bioRxiv*. Available at: http://biorxiv.org/content/early/2017/03/04/113779.abstract.
41. Pestle WJ, Colvard M (2012) Bone collagen preservation in the tropics: a case study from ancient Puerto Rico. *J Archaeol Sci* 39:2079–2090.
42. Hopkins RJA, Snoeck C, Higham TFG (2016) When Dental Enamel is Put to the Acid Test: Pretreatment Effects and Radiocarbon Dating. *Radiocarbon* 58(4):893–904.
43. Laffoon JE, Rojas RV, Hofman CL (2013) Oxygen and carbon isotope analysis of human dental enamel from the Caribbean: Implications for investigating individual origins. *Archaeometry* 55(4):742–765.
44. Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc* 2010(6):db.prot5448.
45. Kircher M, Sawyer S, Meyer M (2012) Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res* 40(1):e3.
46. MYbaits manual (2016) Available at: http://www.mycroarray.com/pdf/MYbaits-manual-v3-.pdf.
47. Schubert M, Lindgreen S, Orlando L (2016) AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res Notes* 9:88.
48. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760.
49. Schubert M, et al. (2012) Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics* 13:178.
50. Andrews RM, et al. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23(2):147.
51. Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27(21):2987–2993.
52. Korneliussen TS, Albrechtsen A, Nielsen R (2014) ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* 15:356.
53. Li JZ, et al. (2008) Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319(5866):1100–1104.
54. Chang CC, et al. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4:7.
55. Patterson N, et al. (2012) Ancient admixture in human history. *Genetics* 192(3):1065–1093.
56. Delaneau O, Marchini J, Zagury J-F (2012) A linear complexity phasing method for thousands of genomes. *Nat Methods* 9(2):179–181.