Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Elizabeth K. Wagman April 10, 2023

Genome-wide patterns of selection in pre- and post-European contact Caribbean populations

by

Elizabeth K. Wagman

Dr. John Lindo
Adviser

Anthropology

Dr. John Lindo

Adviser

Dr. Kristin Phillips

Committee Member

Dr. Dietrich Stout

Committee Member

Sophie Joseph

Committee Member

2023

Genome-wide patterns of selection in pre- and post-European contact Caribbean populations

By

Elizabeth K. Wagman

Dr. John Lindo

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Anthropology

2023

Abstract

Genome-wide patterns of selection in pre- and post-European contact Caribbean populations
By Elizabeth K. Wagman

In recent years, the study of human population genetics has opened up the possibility to reconstruct ancestral and physiological changes from a biological perspective. Supplementing historical and cultural analyses, the use of ancient and modern genomes can improve our understanding of complex turning points in human history. Beginning in the late 15th century in the Americas, colonization acted as a powerful selective pressure on the genomes of Indigenous populations due to the rapid influx of novel pathogens, shifting labor and agricultural practices, and a population wipeout. In this study, recent methods from the field of computational genomics will be applied to pre- and post-contact Caribbean populations to examine ancestral cluster components, visualize demographic history, and investigate positive selection on genes involved with a multitude of functions. Demographic analyses were conducted using principal components analysis, maximum likelihood trees, and model-based cluster analysis to estimate population substructure. Following demographic analysis, allele frequency-based selection scans were performed to identify gene candidates for positive selection. In the ancient cohort's selection scan, variants associated with genes coding for proteins involved in skin pigmentation lightening and tumor suppressor pathways exhibited signs of positive selection. Moreover, in the contemporary Caribbean cohort, polymorphisms associated with a number of genes linked to DNA repair, maintenance of genomic and cellular stability, immune system regulation, and immune cell development and activation showed positive selection signals. Given the colonial history of the Caribbean islands, as well as the complex pattern of migration and admixture between Indigenous American, European, and African populations, it is expected that genes related to cell repair, genomic stability, and immune regulation exhibit signs of positive selection in the modern Caribbean cohort. Taken together, the results of the demographic analyses and selection scans, alongside a review of the relevant literature, supports a more holistic understanding of post-colonial Caribbean legacies and the lasting impact of European colonization on contemporary populations in the region.

Genome-wide patterns of selection in pre- and post-European contact Caribbean populations

By

Elizabeth K. Wagman

Dr. John Lindo

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Anthropology

2023

Acknowledgements

I would like to first thank my thesis advisor Dr. John Lindo for all his support and guidance in the conception, structuring, and editing of my project. I would also like to thank my committee members, Dr. Kristin Phillips, Dr. Dietrich Stout, and Sophie Joseph for their commitment to my project and growth. I would like to give a special thanks to Sophie Joseph again for her close mentorship and friendship throughout this past year and over the course of my tenure as an undergraduate researcher in the Lindo Ancient DNA lab. Additionally, I would like to thank Dr. Robert Paul, Heather Carpenter, and the Department of Anthropology for all their guidance in navigating the honors thesis process. Finally, I'd like to thank my parents and friends for all their support throughout this journey.

Table of Contents

Figures

Tables

Introduction

The colonization of the Caribbean during the 15th century marked a turning point in the socioeconomic, cultural, and demographic history of the region. Settlement by many European countries not only brought the onslaught of slavery across the Atlantic, but also the introduction of novel pathogens, such as the L4 tuberculosis lineage, bubonic plague, measles, smallpox, mumps, chickenpox, influenza, cholera, diphtheria, typhus, malaria, leprosy, and yellow fever (Brynildsrud et al. 2018; Jones 2003; Martin and Goodman 2002). Many diseases originating in the pre-Columbus Americas were originally theorized to be largely chronic and episodic, with many deaths also attributed to nutritional insufficiencies (Larsen 1994).

Following initial contact by Christopher Columbus in 1492, and subsequent contact periods between Europeans and Indigenous individuals, many Indigenous groups were infected rapidly and simultaneously with novel pathogens, causing what Alfred Crosby coined as "virgin soil epidemics" (Crosby 1976). While not entirely unidirectional in the nature of infection, this sudden change in environmental conditions brought on by European colonization in the Caribbean may have served as a powerful selective pressure on the immune system within Indigenous populations (Larsen 1994).

Due to recent computational, biomolecular, and methodological advancements in the 21st century, the field of genomics can now be applied to the study of both ancient and modern populations, illuminating many demographic and evolutionary events that are not covered in historical texts (Shapiro and Hofreiter 2014). The study of ancient DNA allows researchers to disentangle complex hominin histories, and, in conjunction with rapidly developing sample-processing and analytical methods, can provide a fairly reliable picture of ancient history (Slatkin and Racimo 2016). When examined alongside paleoarcheological, historical, and anthropological

findings, the study of ancient and modern human genomes helps paint a more complete picture of demographic patterns, evolution, and major events in human history.

The field of ancient genomics has only recently been applied to population studies in the Americas and within Indigenous communities. Owing to poor ancient DNA preservation in the Caribbean climate, few studies have been done investigating genomic insights. In the past, studies have been performed using mitochondrial DNA, but this does not sufficiently represent a comprehensive picture of the population history due to its limitation to short fragments of the maternal line (Shapiro and Hofreiter 2014). Therefore, the genomic story of European colonization in the Caribbean remains incompletely understood. It was not until recently, in a work by Fernandes and colleagues in 2021, that pre-contact Indigenous nuclear genomes originating in the Caribbean were published with open access (Fernandes et al. 2021). While the data represented are only portions of the whole genome (using '1240K' capture), this marks a turning point in the study of Caribbean ancestry using ancient genomics due to the availability of nuclear DNA.

Using ancient nuclear DNA from Indigenous individuals, carbon-dated to pre-European contact, and modern nuclear DNA from individuals with Puerto Rican ancestry from Puerto Rico, the present study will construct a population-specific history using multiple computational methods (Fernandes et al. 2021). Primary investigation will include various demographic analyses, including a principle component analysis, maximum likelihood tree modeling, and population structure inference (Chiu et al. 2022; Pickrell and Pritchard 2012; Zheng et al. 2012). These computational methods will allow further exploration of the relationships between ancient Caribbean and modern global populations from the 1000 Genomes Project and the Simons Genome Diversity Project (Auton et al. 2015; Mallick et al. 2016). Selection scans using an

allele frequency-based method will be used to ascertain whether population-specific positive selection has occurred between the pre- and post-contact Caribbean cohorts, namely on genes associated with immunological functions (Cheng, Mailund, and Nielsen 2017). A literature review of the relevant anthropological, archaeological, and historical texts surrounding European contact with the Americas has been included as a comparative tool to supplement and support genomic findings. This investigation into the historical, anthropological, and genomic insights of the colonization of the Caribbean islands will help shed light on the processes that have shaped the current landscape of Indigenous Caribbean ancestry.

Literature Review

a. The Pre-Columbian period and the immunological history of the Caribbean

Archaeological evidence suggests the peopling of the Caribbean occurred roughly 8,000 calibrated years before the present (cal yr B.P.) (Keegan and Hofman 2017). Around 2,800 cal yr B.P., groups spread across the Caribbean islands, starting what is known as the Ceramic Age (Nägele et al. 2020). During this time, a diverse set of new pottery and agricultural techniques were developed, further differentiating emerging Indigenous cultures across the Caribbean. By the time European settlers arrived in the Caribbean in the late 15th century, there were distinct Indigenous communities across the islands.

When Columbus landed in the Bahamas in 1492, he first encountered the Taíno, an Indigenous group living in the Greater Antilles, the northern Lesser Antilles, and the Bahamas (present day Cuba, Jamaica, Haiti, the Dominican Republic, Puerto Rico, and the Virgin Islands). Speaking the Arawakan language, the Taíno people are thought to have originated from groups around the Orinoco River in South America (Schroeder et al. 2018). They developed a rich culture in the region known as present-day Hispaniola, with formally structured villages, intensified farming practices, and distinct material art and ritual, revealed in the archaeological record through evidence of bone and shell craftsmanship ("Taíno Culture History" 2017). The Indigenous Taíno can be further broken down into different cultural subdivisions, with the most complex and developed societies and the Chican-Ostionoid material culture associated with the Central or "Classic" Taínos in Hispaniola. "Western" Taínos, sometimes referred to as Lucayans, are linked to Ostionoid-Meillacan material tradition, and lived in central Cuba, Jamaica, the Bahamas, and parts of Hispaniola.

Finally, “Eastern” Taínos occupied regions of the Virgin Islands and the Leeward Islands of the Lesser Antilles (“Taíno Culture History” 2017).

A number of additional Indigenous groups lived in the Caribbean alongside and following the Taíno during the pre-Columbian period. The first group to settle the Caribbean islands in roughly 300 B.C. was known as the Ciboney. The Kalingo people, also referred to as the Carib in the colonial record, were the last Indigenous group to migrate to the Caribbean as settlers were encountering native groups (Beckles 1992). The Kalingo started their migration into the Caribbean in roughly 1000 A.D., and were in the midst of conquering land from the Taíno in the Greater and Lesser Antilles as Europeans arrived in the northern Caribbean, which resulted in their depiction in the colonial archive as an aggressive people (Ross et al. 2020).

While the arrival of European settlers in the late 15th century certainly brought a wave of infectious disease and novel pathogens to Indigenous communities in the Americas, such as the L4 tuberculosis lineage, bubonic plague, measles, smallpox, mumps, chickenpox, influenza, cholera, diphtheria, typhus, malaria, leprosy, and yellow fever, current research in the developing fields of paleomicrobiology and paleogenomics have illuminated a much more complex picture of the immunological history of the pre-contact Americas (Brynildsrud et al. 2018; Martin and Goodman 2002; Jones 2003; Joseph and Lindo 2022). Much of the previous literature surrounding records of immunology in the pre-Columbian period was informed by subjective narratives told by European colonial missionaries. It is clear now, through evidence-based research, that there was not a unidirectional transfer of all pathogens from Europe to the Americas (Joseph and Lindo 2022). Instead, there existed a diverse microbiological landscape in the Americas predating European contact. While the traditional

theory of one-way infection does hold true in some specific cases, it is important to question and further investigate long-standing narratives surrounding colonization, infectious disease, and selective pressures.

Tuberculosis (TB), transmitted between humans by the bacterium Mycobacterium tuberculosis through respiratory droplets, primarily affects the lungs through mechanisms of tissue damage. However, the origin of TB infection is highly contested. The leading theory in the 20th century was a zoonotic origin of TB from M. bovis in Europe during the transition to domestic cattle in the Neolithic Age (Barbier and Wirth 2016; Buzic and Giuffra 2020). This assumption was also informed by a colonial-centric approach to the historical record; owing to the extreme decimation of Indigenous populations following European contact, it was concluded that TB did not exist in the Americas due to a lack of immune response (Stead et al. 1995). However, according to recent studies in paleomicrobiology using next generation sequencing (NGS) techniques, the narrative of the origin of TB has shifted from European-borne to one that resulted from zoonosis in the Americas. That said, a study by Brynildsrud et al. (2018) showed that the L4 tuberculosis lineage, the most common strain today, originated in Europe, and patterns of infection coincided with waves of colonization in the Americas. Thus, it is plausible that TB existed within Indigenous populations in the Caribbean prior to European contact, but was outcompeted by the L4 European lineage following contact (Bos et al. 2014; Brynildsrud et al. 2018).

With the recent advancements in molecular genomics and NGS, researchers are now able to paint a clearer picture of the history of infectious disease in the Americas among Indigenous populations. Previous analyses, such as Crosby's model of "virgin soil epidemics," have theorized that the majority of modern infectious diseases did not exist in

the Americas prior to European colonization (Crosby 1976). Informed by missionary reports and colonial records, it was previously thought that morbidity and mortality in the pre-Columbus Americas were due to chronic and episodic diseases, nutritional insufficiencies, and sexually-transmitted bacterial infections (Larsen 1994). As evidenced by the example of TB, while the L4 lineage brought to the New World on waves of colonization certainly devastated Indigenous communities, it may not have been the first strain of TB to exist globally, evidenced by older strains of TB in fossilized animal and human remains in the Americas (Bos et al. 2014; Brynildsrud et al. 2018; Lee et al. 2012; Rothschild et al. 2001). Previously established narratives of the origin and transmission of specific diseases associated with European colonization, such as TB, have been complicated by developments in the fields of paleogenomics and paleomicrobiology. While the introduction of European-borne pathogens—as well as other factors such as social reorganization and warfare—did cause a significant Indigenous population decline, Crosby's virgin soils hypothesis has been put into question. Population decrease alone is not sufficient evidence to conclude that Indigenous communities were not exposed to diseases like TB prior to colonization (Joseph and Lindo 2022; Lindo et al. 2016).

b. Post-contact Caribbean: Genetic and social legacies

As shown by recent studies co-analyzing ancient and modern Caribbean populations, there is evidence of genetic continuity of Indigenous ancestry through time in present-day Caribbean genomes. A study by Schroeder and colleagues in 2018 showed that the native American components of some modern Caribbean genomes are closely genetically related to the ancient Taíno, "demonstrating an element of continuity between pre-contact populations and present-day Latino populations in the Caribbean" (Schroeder et al. 2018). While modern

Caribbean genomes are admixed due to gene flow, colonialism, and migrations, there is evidence of a lasting Indigenous American ancestral component in modern individuals, which will be explored in greater depth in later demographic analyses using maximum likelihood trees and model-based cluster analysis.

Previous research has demonstrated that the amount of Indigenous ancestry varies dramatically across the islands in modern Caribbean populations (Gravel et al. 2013). Some individuals have a greater Native American ancestral component than others, ranging from 10-15%, alongside a majority of European and African ancestry, due in large part to the colonial history of the region. European settlers displaced the vast majority of Indigenous Caribbean communities by the mid-16th century, simultaneously importing large enslaved African populations by means of the trans-Atlantic slave trade for colonial agriculture production (Bryc et al. 2010). Despite the large amount of admixture observed in modern Caribbean genomes (in this thesis, modern Puerto Ricans from the 1000 Genomes Study), the ancestral cluster that aligns with Indigenous American ancestry exhibits clear similarities to Arawakan speakers (originating in northeast South America) and the ancient Taíno (Schroeder et al. 2018).

Alongside these definitions and delineations of indigeneity established by studies in population genomics, it is important to recognize the effects of colonialism on Indigenous identity in the post-contact Caribbean. The cultural diaspora of the contemporary Caribbean can only be understood in the context of ethnic identity, informed by African, Indian, Chinese, Syro-Lebanese, Portuguese, Spanish, French, Dutch, and English influences from colonial settlers and the import of enslaved and indentured human labor through systems of exploitation (Murdoch 2021). It is also important to recognize the Caribbean region and the

formation of identity in the context of the successive inflows and outflows of populations; "[i]t is this arc of encounter and transformation that allows us to posit the Caribbean as both a diasporized and a creolized society, where the complex process of ethnocultural admixture mediated the commingling of peoples from elsewhere that provided the framework for regional identity" (Murdoch 2021). While the use of genomic data and population demography can help researchers quantify Caribbean ancestry, it is even more vital to consider the sociopolitical underpinnings of contemporary Caribbean identity, as informed by patterns of colonization, migration, and cultural interconnectedness.

c. Human population genetics

Newly emerging technologies and methods in the field of human population genetics and the study of ancient DNA allow researchers to better understand the complex history of human migration, social organization, and natural selection. The underlying goal of population genetics is "to infer the past history of populations and describe the evolutionary forces that have shaped their genetic variation" (Tataru et al. 2017). Moreover, not only does the field provide us with the technical tools needed to research concrete scientific hypotheses regarding genetic data, but when used in conjunction with paleoanthropological and archaeological findings and historical texts, researchers can draw broader conclusions about human culture. While the foundational principles of the field have existed since the mid-20th century, after the first whole human genome was sequenced and published in 2001, the field of population genetics was opened up to new breakthroughs (Lander et al. 2001; Venter et al. 2001). Since 2001, thousands of high-coverage genomes have been published and studied, including those of modern and ancient global human populations, non-human primates, and other organisms.

Specifically, studies using ancient and modern DNA can be used to illuminate trends of natural selection during critical points in history. Computational methods in the field have allowed researchers to search for specific loci or haplotypes that have been conserved over time in specific populations (Slatkin and Racimo 2016). This phenomenon, also known as positive selection, is defined as "the tendency of beneficial traits to increase in prevalence (frequency) in a population" and is the driving force behind adaptation (Schaffner and Sabeti 2008). In this thesis, the example of European colonization in the Americas is used as the historical framework for the selection scan analyses on pre- and post-contact Caribbean cohorts.

This study includes an allele frequency-based method ("Ohana"), which uses a likelihood ratio test to discern alleles in a given population that have strongly deviated from a genome-wide covariance structure (Cheng, Mailund, and Nielsen 2017). Ohana relies solely on the genomic data to infer ancestral population assignments of each sample based on allele frequencies, instead of requiring the input of assigned population labels to each individual genome or the use of phased genomes for haplotype information. It also uses the results of a cluster analysis to determine the number of ancestral populations. Ohana was utilized to highlight the variants that deviated most strongly from the genome-wide covariance structure, which generates likelihood ratios for each locus in the dataset. A higher likelihood ratio signifies a strong deviation from the genome-wide covariance structure among multiple individuals, which indicates a high probability of a positive selection signal.

d. Previous ancient genomic research in Caribbean population history

A number of various studies have been conducted in the field of ancient genomics and population demography with regard to the peopling of the Caribbean, although analyses are

restricted due to poor ancient sample preservation in the Caribbean island environment (Schroeder et al. 2015). Additionally, much of the prior research has been limited due to the use of mitochondrial DNA (mtDNA) (Lalueza-Fox et al. 2001; 2003; Mendisco et al. 2015; Moreno-Estrada et al. 2013). The use of solely mtDNA in genomic research is insufficient to paint the full picture of a population's history due to representation of only fragments of the maternal line (Shapiro and Hofreiter 2014). With the advancement of molecular genomic laboratory techniques and archaeological preservation, more studies are done using nuclear DNA, which contains both maternal and paternal genetic information, thus improving the quality of published data.

With the development of improved sequencing techniques came the recent publication of high-quality genome-wide data from ancient individuals from the Caribbean. Specifically, an article by Fernandes and colleagues reported 174 newly sequenced genomes from pre-contact individuals from the Bahamas, Haiti and the Dominican Republic (together, Hispaniola), Puerto Rico, Curaçao, and Venezuela, which were co-analyzed with 89 previously published genomes. The goal of that study was to assess the interconnectedness of different cultures across the islands. Fernandes et al. also performed studies to investigate the effective population sizes and social structure during the Ceramic Age (Fernandes et al. 2021). This publication provided the basis upon which this thesis was built; the demography studies and selection scans outlined below contain the partial genomes of 19 pre-contact ancient Caribbean individuals initially reported by Fernandes and colleagues (Table S1).

In addition to the ancient Caribbean samples introduced above, this study includes 101 modern whole-genomes from individuals with Caribbean heritage in order to provide a temporal comparison to the pre-European contact cohort (Table S2). These samples were

obtained from the 1000 Genomes Project, and are part of the Puerto Rican in Puerto Rico dataset (Auton et al. 2015). Additional samples from the 1000 Genomes Project and the Simons Genome Diversity Project were used to supplement relevant demographic analyses and selection scans (Mallick et al. 2016) (Table S3).

e. Broader applications, implications, and ethics

In such an emerging field like population genomics, there is still much discussion surrounding the applications and impacts of scholarship and research among the scientific community and the general public alike. While the main goal of population geneticists is to answer the question of how populations vary through time and space, there are also useful applications of findings to the field of public health and precision medicine (Roberts et al. 2021). Moreover, as shown by its use in this study, population genomics can serve as a supplement to previously established historical and archeological records, providing further evidence of the biological impacts of selective pressures like colonization.

With the increased accessibility of at-home ancestry testing and advancing technologies in population genomics, there has been a reversion to older schools of thought that conflate molecular genetic data with notions of identity. Especially when it comes to discussions surrounding indigeneity, these recent scientific developments seem to trump Indigenous peoples' own culturally-bound definitions of identity. As stated by TallBear in her review of the intersections between indigeneity and genomics, "indigenous peoples' 'ancestry' is not simply genetic ancestry evidenced in 'populations' but biological, cultural, and political groupings constituted in dynamic, long-standing relationships with each other and with living landscapes that define their people-specific identities and, more broadly, their indigeneity"

(TallBear 2013). This is a crucial distinction to keep in mind as we explore the impacts of colonization on Indigenous populations in this study.

More directly, studies in population genetics tend to misrepresent Indigenous ancestry by solely focusing on the biological basis of difference from the colonial population. In most research, geneticists differentiate and recognize Indigenous groups as those distinct from invading colonial powers, therefore being less admixed and biologically and culturally tied to a geographic land base compared to other global populations. These outsider formulations of indigeneity "cannot account for resistance to the state and indigenous attempts to survive and flourish that underpin contemporary indigeneity" (TallBear 2013). Further, these definitions and classifications entirely omit the way "Indigenous" is used by native peoples to emphasize their relationships to the original peoples globally, united not on a biological or racial front but by colonial similarities throughout history. Oftentimes, the field of molecular anthropology and population genetics recasts indigeneity as simply a population categorization without deep consideration of the complex historical, social, and cultural underpinnings of identity.

Historically, ideas of race, ethnicity, and science have been intertwined, embodied in the infamous eugenics movement. Conceptualized by Francis Galton at the turn of the 20th century and later weaponized by the Nazi regime during World War II, the eugenics movement brought to light a dangerous conceptualization of racial categories and cultural evolution, postulating that some races are "superior" to others, or, in other words, more evolved. Backed by developing statistical testing, phrenology, and heredity studies, this popular pseudoscientific claim came at the detriment of many different individuals, who suffered from persecution, discrimination, and in some cases, sterilization (Turda 2022).

Today, while the negative relationship between racial identity and science has been largely dismantled, population geneticists still need to reconcile the social and biological construction of race in the presentation of global populations in their research, specifically in the representation of human diversity. As the field of population genetics attempts to make inferences regarding racial identity and indigeneity through demographic studies, we must carefully examine the positionality of the researcher, the categorization of human groups, and the broader sociocultural implications of these claims on the conceptualization of identity as a whole.

Methods

a. Principal components analysis (PCA)

Eigenvectors were calculated for individuals in the ancient and modern Caribbean cohorts with outgroup samples from Europe, East Asia, Africa, and the Americas after removing missing data using Plink v1.9 (--remove-indels, --max-missing 1.0). The SNPRelate Bioconductor R package v3.15 was used for pruning the whole genome dataset for SNPs (single nucleotide polymorphisms) in linkage disequilibrium using the snpgdsLDpruning() option. Linkage disequilibrium describes the degree to which one allele of a genetic variant is correlated to an allele of a different nearby variant; it is important to prune SNPs in linkage disequilibrium in order to prevent false positive signals of statistical association between SNPs (Bush and Moore 2012; Slatkin 2008). The first two principal components (PC1 and PC2) were visualized with the snpgdsPCA() option, using default parameters (Zheng et al. 2012) (Figure 1B). The PCA was used to visualize the relatedness of the individuals in the dataset, and to remove three outliers from the modern Caribbean cohort that did not align with the modern Caribbean cluster.

b. Ancestry clustering

A model-based clustering demographic analysis was performed using the likelihood-free approach introduced in a recent publication by Chiu et al. SCOPE uses latent subspace estimation and alternating least squares to infer admixture fractions from individual allele frequencies (Chiu et al. 2022). To estimate K, the number of ancestral populations in the dataset, the K value was chosen with the lowest cross-validation index from the ADMIXTURE program after testing K=2 through K=8 (Alexander, Novembre, and Lange 2009) (Figure S1). K=4 was discovered to have the lowest cross-validation index, which

meant that a model containing K=4 ancestral clusters best fit the dataset. The SCOPE program was then run using K=4 ancestral clusters (--k 4) with default parameters and visualized with PONG (Behr et al. 2016) (Figure 1C). Two iterations of the dataset were run using the SCOPE program: one including the full dataset, and one without the ancient Caribbean cohort. Due to the partial genome availability of the ancient samples (samples were sequenced using '1240K' capture methods) as opposed to whole genome availability, we were not able to properly present the findings of the entire dataset under the parameters of SCOPE. This represents one limitation of the '1240K' capture method of DNA extraction and sequencing that appears in later demographic analyses.

c. Ancestry clustering for maximum likelihood trees

The TreeMix program was applied to the dataset to create maximum likelihood trees from the allele frequency data. The Mbuti from the Simons Genome Diversity Project dataset was used to root the tree using the command --root Mbuti (Mallick et al. 2016). Linkage disequilibrium was accounted for by grouping M adjacent sites with the –k option. M was chosen such that a dataset with L sites will have approximately $L/M \approx 20{,}000$ independent sites. Following the analysis m=1, we performed a global rearrangement (with the -global option). Admixture scenarios were considered with m=0, m=1, m=2, m=3, and m=4 migration events for the full dataset of ancient and modern Caribbean individuals (Figure S2). Each migration model was run with 10 replicates, and the trial with the highest likelihood was chosen to depict the maximum likelihood tree for the given migration scenario (Pickrell and Pritchard 2012) (Figure 2).

d. Selection scans

The selscan program, contained within a suite of programs called Ohana, is an allele frequency-based selection scan and uses a likelihood ratio test to discern alleles in a given population that have strongly deviated from a genome-wide covariance structure (Cheng, Mailund, and Nielsen 2017). This program does not require the input of assigned population labels to each individual genome, nor does it utilize phased genomes for haplotype information. Alternatively, Ohana relies solely on the genomic data to infer ancestral population assignments of each sample based on allele frequencies. It also uses the results of a cluster analysis to determine the number of ancestral populations.

Two Ohana selection scans were conducted, one for the ancient Caribbean cohort and one for the modern Caribbean cohort, and two outgroups were included (Utah residents with Northern and Western European ancestry (code CEU) and Yoruba in Ibadan, Nigeria (code YRI)) from the 1000 Genomes Project (Auton et al. 2015). To estimate K, the number of ancestral populations in the tree, the K value was chosen with the lowest cross-validation index from the ADMIXTURE program after testing K=2 through K=15 (Alexander, Novembre, and Lange 2009). K=3 was used, which aligns with the number of distinct major population groups used as input in the Ohana program (Caribbean test population, and CEU and YRI outgroup populations). VCFtools was used to prepare the VCF by filtering sites out of Hardy-Weinberg Equilibrium with a p-value below 10e-4 and removing indels and missing data. The dataset was then downsampled at random to 5% of the original variants to estimate the correlation structure of each individual to the population tree. Then, using Ohana's qpas function, we inferred component covariances to produce admixture-corrected allele frequencies. Ohana's selscan function was utilized to highlight the variants that

deviated most strongly from the genome-wide covariance structure, which generates likelihood ratios for each locus in the dataset. A higher likelihood ratio signifies a strong deviation from the genome-wide covariance structure, which equates to a stronger positive selection signal. Due to the nature of this population-specific scan (as opposed to a global estimate of covariance), the method requires a scalar addition of 10 (h=10) to the position of the covariance matrix that corresponds to the focal population of the selection scan (ancient or modern Caribbean in this case).

After sorting the results of the selscan function in order of decreasing likelihood ratio (the genetic loci with the strongest selection signal listed at the top), the ANNOVAR program was utilized to annotate the Ohana results (Wang, Li, and Hakonarson 2010). This program identified and recorded the gene names associated with the chromosome and positions of the selection candidates, as described in the Results section. The top genetic hits for both the ancient and modern Caribbean selection scans are visualized using Manhattan plots and histograms depicting the distribution of the likelihood ratio scores created using the qqman R package (Turner 2014) (Figure 3).

Results

a. Demographic analyses

This study includes 19 ancient genomes from Indigenous individuals across the Caribbean islands predating European contact (Table S1) and 101 modern individuals from the Puerto Rico cohort in the 1000 Genomes Project (Auton et al. 2015; Fernandes et al. 2021) (Table S2, Figure 1A). To facilitate a selection study with continuity between population groups, we conducted demographic analyses. We created a principal component analysis (PCA), maximum likelihood trees, and model-based cluster analysis to estimate population substructure (Chiu et al. 2022; Pickrell and Pritchard 2012; Zheng et al. 2012). The maximum likelihood trees were constructed including the 19 ancient Caribbean samples. Due to the incomplete nature of the '1240k' capture genomic data available for individuals from this region, they were not able to be included in the cluster analyses. This incompatibility of genomic capture data with certain demographic modeling tools represents one limitation of this study.

The PCA includes a modern Caribbean cohort, ancient Caribbean cohort, and individuals with African, American, East Asian, and European ancestry from the Simons Genome Diversity Project (SGDP) dataset (Mallick et al. 2016) (Figure 1B). The ancient Caribbean group (green) is observed to overlap with some of the modern Caribbean individuals (dark blue), demonstrating shared ancestry. Nearby are the Europe (purple) and Americas (red) clusters, which align with the region's historical narrative including Indigenous American ancestry and post-European colonization admixture. The cluster analysis, including the modern Caribbean cohort and individuals from the SGDP, revealed four distinct ancestral clusters (K=4) (Chiu et al. 2022). The modern Caribbean individuals exhibit predominantly

European ancestry (red), with smaller proportions of African (green) and Indigenous American (aqua) ancestry as well (Figure 1C).

Our maximum likelihood trees yielded similar results; the modern Caribbean cohort formed a branch adjacent to European populations from England and Spain, in close proximity to populations from the Americas (Pickrell and Pritchard 2012). Moreover, there is a contribution from the Quechua to the modern Caribbean population shown by the migration arrow. The Quechua are the descendants of the Indigenous people of Peru; while many are native to Peru, there are contemporary communities living in Ecuador, Bolivia, Chile, Colombia, and Argentina as well. Many modern Caribbean populations have mixed Indigenous, African, and European ancestry due to colonization starting in the late 15th century, demonstrated by the cluster analysis and maximum likelihood trees and supported by the historical record (Moreno-Estrada et al. 2013; Sans 2000). The ancient Caribbean individuals can be seen forming a branch with the Piapoco, which is an Arawakan speaking Indigenous American population from eastern Colombia and western Venezuela (Figure 2). These results are consistent with previous demographic analyses done comparing genomes from Ceramic Age ancient Caribbean individuals to modern Indigenous American populations (Fernandes et al. 2021).

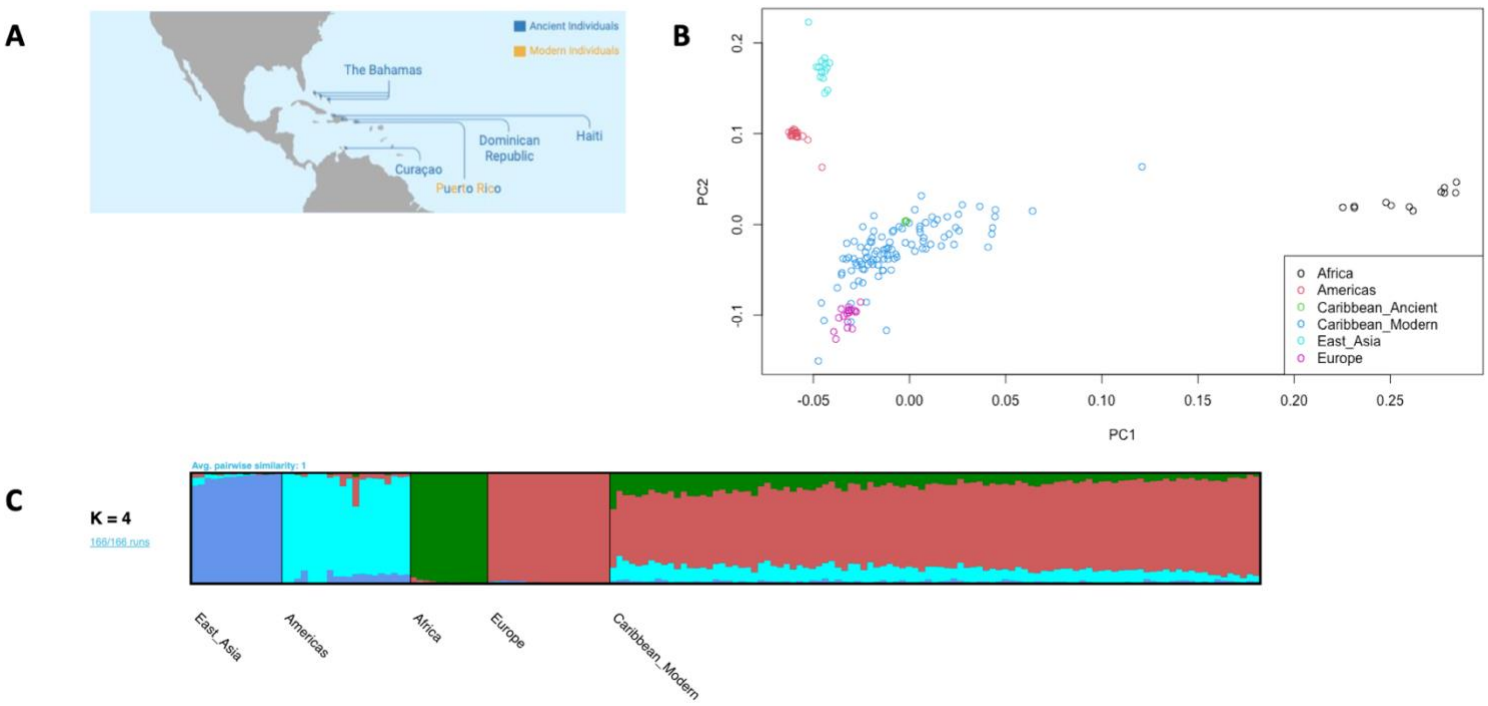


Figure 1. Demographic analysis.

A. Map of population locations for ancient and modern Caribbean individuals.

B. Principal components analysis showing first two principal components, including individuals from this study and individuals from Africa, Europe, East Asia, and the Americas obtained from the SGDP dataset.

C. Visualization of cluster analysis at K=4, which exhibited the lowest Bayesian information criterion (BIC) value.

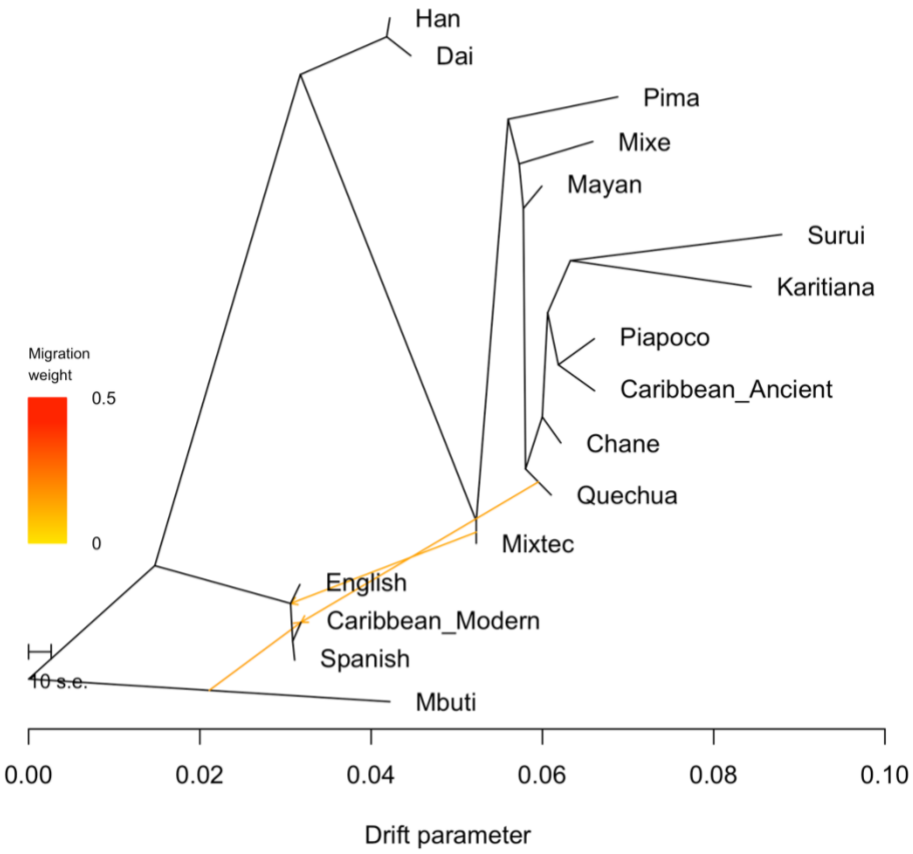


Figure 2. Maximum likelihood tree generated by TreeMix (m=3) showing ancestry relationships between Caribbean cohorts and individuals in the SGDP dataset.

b. Selection scans

After determining an ancestral connection between the ancient and modern Caribbean cohorts using demographic analyses and historical inference, I performed an allele frequency-based selection scan to identify gene candidates for positive selection (Cheng, Mailund, and Nielsen 2017). Due to a significant amount of admixture in the modern Caribbean population due to the effects of colonization, an admixture-aware method was used to account for this extensive gene flow (Cheng, Mailund, and Nielsen 2017). The Ohana program is a more effective way of detecting positive selection signals in admixed populations due to its utilization of cluster analyses to determine the number of underlying ancestral populations. Using this method, we were able to predict the genome-wide global covariance matrix, visualized using a Newick tree. If an allele in the Caribbean populations differed significantly from the predicted genome-wide covariance model, measured by a likelihood ratio test, then it was assumed to be under positive selection.

However, given the three-way admixture (majority European, followed by African and Indigenous American ancestries) in the contemporary Caribbean cohort, it is likely that the selection scan is only detecting ancient selection in Europeans (Figure 1C). This represents one limitation of this study; while the Ohana program does account for admixture in the target population, this method is not fully effective given the specific modern genomic dataset used. Future directions include a repetition of the modern Caribbean natural selection scan using genomes from individuals with a greater percentage of Indigenous ancestry. The results from this scan are included below to serve as a temporal comparison to the ancient Caribbean cohort.

Among the strongest selection signals in the ancient Caribbean cohort were intergenic and intronic SNPs associated with the *SLC24A5* and *SLC45A2* genes, respectively. Both the *SLC24A5* and *SLC45A2* genes play a key role in skin pigmentation pathways, specifically in skin lightening (Basu Mallick et al. 2013; Crawford et al. 2017; Hernandez-Pacheco et al. 2017). Additionally, an intronic polymorphism associated with the *FBXO31* gene had a positive selection signal in the ancient Caribbean selection scan. *FBXO31* encodes for a protein involved in the response to DNA damage and functions as a tumor suppressor by promoting oncogenic MDM2 degradation, resulting in an increase in p53 levels (a tumor suppressor protein) (Malonia et al. 2015).

In the modern Caribbean cohort, many top genes under positive selection are related to cell replication, repair, and immunological functions. Intronic SNPs of the *ZRANB3* and *MCM6* genes, for example, are associated with DNA replication fork formation and maintenance of genomic stability (Ciccia et al. 2012; Weitzman and Fradet-Turcotte 2018; Zeng et al. 2021). Additionally, intronic polymorphisms of the *R3HDM1* and *DARS1* genes, and an intergenic polymorphism of the *BMP4* gene have been correlated to functions of the immune system, such as T cell regulation and cellular sensors in viral infections (Feng et al. 2022; Huang et al. 2021; Liu et al. 2022) (Table 1). In a recent study investigating potential crosstalk genes involved in immunological mechanisms in two inflammatory diseases, researchers found that the *R3HDM1* gene was positively correlated with type 1 T helper (Th1) cells in both diseases (Liu et al. 2022). Th1 cells are immune cells that "activate macrophages and are responsible for cell-mediated immunity and phagocyte-dependent protective responses" (Romagnani 1999). In addition, the *DARS1* gene codes for an aminoacyl-tRNA synthetases (aaRSs), which are essential enzymes in translation by linking

amino acids onto their cognate tRNAs during protein synthesis (Feng et al. 2022). aaRSs play a key role in immune cell maturation, transcription, recruitment, and activation through complex cellular mechanisms and downstream processes. Additionally, aaRSs are crucial in host viral and bacterial responses by upregulating the production of pro-inflammatory cytokines, activating macrophages, and inducing chemokine production and phagocytosis (Nie et al. 2019).

Included in Table 1 is an expressive quantitative trait loci (eQTL) analysis, which identifies genetic variants that directly affect the expression of another gene (Lonsdale et al. 2013). Additionally, the genes of interest with the highest likelihood ratios are visualized using a Manhattan plot, with the highlighted peaks representing the genes under strong positive selection and the chromosome number depicted on the x-axis (Turner 2014) (Figure 3). The top 0.012% of SNP selection candidates for the ancient Caribbean cohort and the top 0.000037% of SNP selection candidates for the modern Caribbean cohort are depicted in supplementary tables S5 and S6. Included alongside the Manhattan plots are histograms depicting the distribution of the likelihood ratio scores for the selection scans.

Table 1. SNPs of interest under high probability of selection described in the Results section. Related to Figure 3.

Chr*	Pos*	Annotated gene	eQTL**	rsID	Derived allele frequency in ancient cohort	Derived allele frequency in modern cohort
15	48400199	*SLC24A5*	*SLC12A1*	rs2675345	0.789474	0.197115
5	33952106	*SLC45A2*	*TTC23L-AS1*	rs185146	0.0263158	0.677885
16	87394386	*FBXO31*	*RP11-482M8.3*	rs1834022	0	0.100962
2	136138627	*ZRANB3*	*MCM6, TMEM163, DARS-AS1, UBXN4, MAP3K19, DARS, CCNT2*	rs3940549	1	0.798077
2	136608646	*MCM6*	*DARS-AS1, UBXN4, DARS, CCNT2, MAP3K19*	rs4988235	N/A	0.206731
2	136429366	*R3HDM1*	*MCM6, TMEM163, DARS-AS1, UBXN4, MAP3K19, DARS, CCNT2, ZRANB3*	rs62168795	N/A	0.793269
2	136707982	*DARS1*	*MCM6, DARS-AS1, UBXN4, MAP3K19, ZRANB3*	rs6754311	1	0.793269
14	54711654	*BMP4*	N/A	rs72709956	N/A	0.125

*Hg19 coordinates

**(Lonsdale et al. 2013)

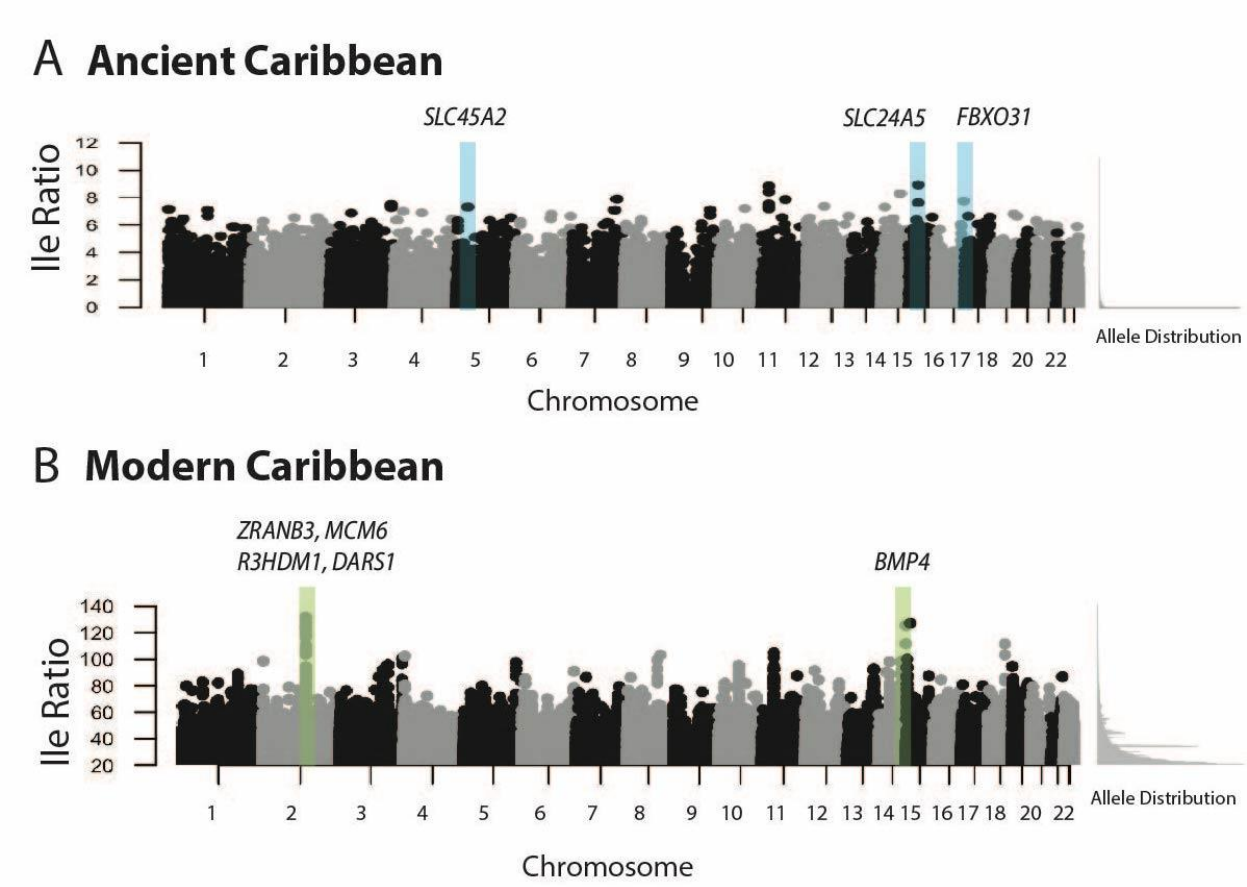


Figure 3. Selection scans highlighting the strongest signals of positive selection.

Ohana selscan Manhattan plot and histogram of the allele distribution of the corresponding likelihood ratio (lle ratio) for the ancient Caribbean cohort (A) and modern Caribbean cohort (B).

Discussion

This study aimed to contribute to previous archaeological, historical, and anthropological investigations into the population history of the Caribbean, with particular attention to post-colonial shifts in demography and genetic processes. The PCA, maximum likelihood trees, and model-based cluster analysis described in the above Results corroborate previous studies illustrating the relationship between pre- and post-European contact Caribbean populations, as well as exhibit the complexities of admixture and global population relatedness in the modern Caribbean cohort (Gravel et al. 2013; Schroeder et al. 2018).

Specifically, the results of the PCA align with previously established insights into ancient and modern Caribbean ancestry (Zheng et al. 2012) (Figure 1B). The ancient cohort (Caribbean_Ancient), represented by the light green color, is shown to cluster directly over the modern cohort (Caribbean_Modern) in dark blue. Also clustering near a portion of the modern Caribbean cohort are individuals from Europe shown by a purple color. This can be contextualized using the broader colonial history of the Caribbean basin, as well as with the maximum likelihood trees and cluster analyses from the present study (Bryc et al. 2010; Murdoch 2021). Moreover, the Indigenous American cohort, shown in red, clusters nearby the modern and ancient Caribbean individuals, illuminating a more distant shared ancestral relationship between them. Despite extensive admixture from African populations which began during the trans-Atlantic slave trade, the African individuals (black) clusters away from the modern Caribbean cluster. This is likely due in part to African populations having generally greater levels of genetic diversity, fewer SNPs in linkage disequilibrium, and extensive population substructure compared to other populations around the world (Campbell and Tishkoff

2008). Additionally, there is no African source population that represents the direct ancestry of those enslaved and transported to the Caribbean basin in this analysis.

Postcolonial admixture is also exhibited genetically in maximum likelihood trees and model-based cluster analysis in order to estimate population substructure (Pickrell and Pritchard 2012). In the maximum likelihood tree constructed using the full dataset of individuals with m=3 migrations, the ancient Caribbean cohort is seen forming a branch with the Piapoco population, an Arawakan speaking group originating in eastern Colombia and western Venezuela. The ancient Caribbean and Piapoco diverge from another branch containing Chané (Arawakan-speakers from Paraguay, Brazil, Bolivia, and Argentina), Karitiana (Indigenous Brazilian), and Surui (Indigenous Brazilian) populations.

These results are consistent with the results from Fernandes and colleagues, who found that the Piapoco contribute to the greatest proportion of ancestry to ancient Caribbean individuals from the Ceramic Age using 1- and 2-way qpAdm modeling, TreeMix, and qpGraph tests. These results help support hypotheses regarding the original peopling of the Caribbean from northeast South America roughly 6,000 years ago (Fernandes et al. 2021). However, the maximum likelihood tree containing these findings must be carefully considered and reevaluated using whole-genome data for ancient Caribbean individuals. Further research directions using high quality ancient whole-genomes from the Caribbean could be explored given technological advancements in preservation and genetic sequencing of highly-degraded samples.

In the same tree, the modern Caribbean individuals are shown forming a branch alongside individuals from England and Spain, which is expected given the colonial history of the region. Forming an adjacent cluster of branches are Indigenous populations from the Americas (Figure 2). The modern Caribbean cohort is also shown to have a migratory contribution from Quechuan

individuals, who are descendants of Indigenous Peruvians, and have communities in Peru, Ecuador, Bolivia, Chile, Colombia, and Argentina today. On the maximum likelihood tree, Quechuan individuals branch in close proximity to the ancient Caribbean individuals, the Chané (Indigenous Argentinians), and the Mayans (from Central America). One explanation for this relatedness could be due to a population split in northeast South America during the early peopling of the continent and migration to the Caribbean islands (Lindo et al. 2018).

Due to the incomplete genomic sequencing of the ancient Caribbean individuals and the data parameters required for ancestry cluster modeling, the ancient Caribbean cohort was not included in the final demographic model-based cluster analysis (Cheng, Mailund, and Nielsen 2017). In the visualization of the cluster analysis, individuals from East Asia (blue) are shown on the left side of the figure, followed by populations from the Americas (aqua), Africa (green), Europe (red), and the modern Caribbean samples (Caribbean_Modern), revealing four distinct ancestry clusters (K=4) (Figure 1C). In the modern Caribbean cohort, the predominating ancestral cluster aligns with the European population, followed by proportions of African and Indigenous American ancestry. Few modern Caribbean genomes had East Asian ancestral components. These observations, alongside the maximum likelihood trees, align with previous historical and cultural constructions of contemporary Caribbean ancestry following colonization (Schroeder et al. 2018).

Moreover, the findings from the allele frequency-based natural selection scans in both the ancient and modern Caribbean cohorts highlight gene candidates for positive selection (Figure 3; Tables 1, S5, and S6). This can inform us of the impact of a selective pressure, such as European colonization, on the individual genome and genetic adaptation. Individuals living in the Caribbean following European contact undoubtedly underwent many cultural and physiological

changes in the face of genocide, large waves of settlement, and agricultural shifts to the landscape through slave labor and plantation development. The present study aimed to highlight the genetic foundations of these phenotypic adaptations. However, due to the limited effectiveness of this natural selection scan method with highly admixed populations, the scan described below for the modern Caribbean cohort may be detecting positive selection signals on SNPs from the ancestral European component of the modern Caribbean genome. These tests should be replicated using unadmixed modern Caribbean genomes in a future study.

In the selection scan performed for the ancient Caribbean individuals, among the strongest selection signals was an intergenic SNP associated with the *SLC24A5* gene, and an intronic SNP associated with the *SLC45A2* gene. Both are implicated in pathways affecting skin pigmentation, specifically skin lightening. Skin pigmentation is essential to protect the body against harmful UV radiation. The production of melanin by melanocytes, the skin's main pigmentation molecule, is regulated by complex interactions and is involved in many other bioregulatory and metabolic pathways, both in the skin and in other organ systems. Skin color varies widely across global populations, and is strongly correlated with latitude due its protective effects against UV radiation. Additionally, many other selective pressures have impacted the variation of skin pigmentation in different regions, such as protection against folate photolysis, appropriate vitamin D level maintenance, and improving the skin barrier's ability to retain water and provide cutaneous antimicrobial defense (Hernandez-Pacheco et al. 2017).

At first glance, positive selection on skin lightening genes does not align with the environment of the Caribbean basin, which is located near the equator and has a high amount of UV radiation. However, there is genetic evidence demonstrating that the light pigmentation variant at *SLC24A5* was introduced into East Africa by gene flow from non-Africans prior to the

peopling of the Caribbean (Crawford et al. 2017). The *SLC24A5* and *SLC45A2* genes contribute to the genetic basis of lighter skin pigmentation in European populations, suggesting a genetic drift event, such as the founder effect, between ancestral European and Caribbean populations (Basu Mallick et al. 2013). Another explanation for this observation is convergent evolution, or the independent evolution of specific genetic adaptations in distinct populations over time. Further directions include the application of a program called RFmix, which is a method of investigating ancestral components for specific SNPs of interest using haplotype phased genetic data (Maples et al. 2013). However, this program requires the use of a phased dataset, which requires high-coverage genomes. With improved sequencing methods, higher-coverage ancient Caribbean genomes could be produced and analyzed using RFmix in a future study.

Moreover, an intronic polymorphism associated with the *FBXO31* gene exhibited signs of positive selection for the ancient Caribbean cohort. *FBXO31* codes for an F-box protein, which is responsible for regulating levels of the murine double minute 2 protein (MDM2) in the body. *MDM2* functions as an oncogene, coding for a protein that downregulates p53 function, which plays a critical role in tumor suppression. When cells are under stress, such as during DNA damage events, p53 levels increase and induce many protective biological responses. MDM2 is important in the degradation of p53 under normal cellular conditions; the function of the FBXO31 F-box protein in regulating the levels of MDM2 is crucial to the functioning of the downstream tumor suppressor p53 (Malonia et al. 2015). Positive selection on such a gene in the ancient Caribbean cohort could signify an upregulation in protective cellular defense mechanisms and the promotion of anti-tumor development.

In the modern Caribbean cohort, many top genes under positive selection are related to cell replication and DNA repair, which aligns with the region's cultural and physiological

adaptive history to colonization. The upregulation of cellular repair mechanisms would likely prove to be advantageous for individuals living in the aftermath of colonization due to the increase in novel infectious diseases brought to the Caribbean by Europeans. Among the top genetic candidates for positive selection are intronic SNPs of the *ZRANB3* and *MCM6* genes. The *ZRANB3* gene is recruited by poly-ubiquitinated proliferating cell nuclear antigen (PCNA) to promote DNA replication fork restart following replication arrest. *ZRANB3* helps maintain genomic stability at stressed or collapsed replication forks by mimicking stalled replication forks and disassembling unnecessary recombination intermediates (Ciccia et al. 2012). The ability to repair stalled or collapsed replication forks and protect genomic integrity is a crucial protective function of the cell against viral and bacterial infections, such as Listeria monocytogenes (Benedetti et al. 2021; Weitzman and Fradet-Turcotte 2018). Additionally, the *MCM6* gene plays a significant regulatory role in DNA replication, thus helping to sustain the cell cycle (Zeng et al. 2021). It codes for one of six proteins in the minichromosome maintenance (MCM) family, which is a heterohexameric DNA helicase involved in the initiation of DNA replication and elongation (Schrader et al. 2005; Zeng et al. 2021). Studies have also implicated MCM family proteins in DNA damage response, transcription, and chromatin structure (Forsburg 2004). The functions of *MCM6* all contribute to the maintenance of genomic stability in the cell, which, similar to *ZRANB3*, helps protect an individual from viral or bacterial infection.

Moreover, some of the top genetic hits for the natural selection scan in the modern Caribbean cohort have direct immunological functions. Intronic polymorphisms of the *R3HDM1* and *DARS1* genes, and an intergenic polymorphism of the *BMP4* gene indicated strong positive selection signals. The protein-coding *R3HDM1* gene has been associated with Th1 immune cells; these cells help activate the body's immune response against intracellular pathogens, such as

viruses and bacteria (Liu et al. 2022; Romagnani 1999). The *DARS1* gene also has an immunological function, coding for a protein that mediates the attachment of amino acids to their corresponding tRNAs (in this case, L-aspartate to tRNA). Studies have also demonstrated that tRNA acylation networks are involved in mediating cholera and tuberculosis progression (Duffy et al. 2019; Ellis et al. 2015).

SNPs associated with the *BMP4* gene also exhibited positive selection signals in the modern Caribbean cohort. *BMP4* encodes for a protein that is understood to promote bone regeneration and regulate T cell development in the thymus. *BMP4* also moderates glycolysis of T cells after activation, which is a crucial function for high-energy activation and cytokine production (Huang et al. 2021). As described in earlier sections, European colonization brought a vast number of novel infectious diseases into the New World, which decimated Indigenous populations and placed a large selective pressure on surviving native and imported populations. Given the colonial history of the Caribbean islands, as well as the complex pattern of migration and gene flow, it is expected that genes related to cell repair, genomic stability, and immune regulation and response exhibit signs of positive selection in the modern Caribbean population.

Conclusion

The goal of this study was to explore the genomic data that underpins Caribbean cultural identity. Through various demographic analyses and genome-wide selection scans, I hoped to shed light on the social and biological complexities of colonization, slavery, and indigeneity in the Caribbean basin. As the PCA, maximum likelihood trees, and model-based cluster analysis illustrated, modern Caribbean genetic ancestry is mostly comprised of European ancestry, followed by African and Indigenous American ancestral components. After establishing genetic continuity between the ancient and modern Caribbean cohorts, natural selection scans were performed to further illuminate genetic variants under positive selection. Among the SNPs under positive selection in the ancient Caribbean cohort were two related to skin pigmentation lightening pathways, which raises further questions regarding gene flow, admixture, and the peopling of the Caribbean (Crawford et al. 2017). In the modern Caribbean selection scan, a number of genetic variants exhibited signs of positive selection, including those associated with genes related to DNA repair, replication, and immune regulation.

This result is expected given the broader context of Caribbean colonial history. Beginning in the 16th century, European colonization placed a large selective pressure on the surviving Indigenous and enslaved African populations in the Caribbean, which undoubtedly led to social and physiological adaptations. The purpose of the natural selection scan was to provide genetic evidence for the aforementioned adaptations. Positive selection on protein-coding genes related to cellular repair and immune function in the modern Caribbean cohort illustrates an adaptive genetic history following the introduction of novel pathogens post-colonization.

To further this line of inquiry, additional research should be done in the collection and sequencing of pre-contact Caribbean genomes. As described in earlier sections, this study

includes only portions of the ancient Caribbean genomes, as opposed to the whole-genome sequences used for all modern individuals. As genetic sequencing techniques improve for highly degraded samples, we will be able to provide a more complete picture of the demographic landscape of Caribbean ancestry. Moreover, given the vast amount of admixture in the modern Caribbean cohort, investigation into the ancestral components for specific SNPs of interest using haplotype phased genetic data using the RFmix program would help further explain the origins of these genetic adaptations (Maples et al. 2013). Further, replication of the natural selection scan with a modern Caribbean cohort with a higher proportion of Indigenous ancestry should be performed to confirm the positive selection candidates outlined in this study due to the limited compatibility of the Ohana program with highly admixed populations. Finally, given the evidence for the existence of TB in the Americas prior to European contact, further research into human-pathogen coevolution with regard to TB and other infectious diseases could contribute to our understanding of the mechanisms by which colonization decimated Indigenous populations (Joseph et al. 2023).

References

Alexander, David H., John Novembre, and Kenneth Lange. 2009. "Fast Model-Based Estimation of Ancestry in Unrelated Individuals." *Genome Research* 19 (9): 1655–64. https://doi.org/10.1101/gr.094052.109.

Auton, Adam, Gonçalo R. Abecasis, David M. Altshuler, Richard M. Durbin, Gonçalo R. Abecasis, David R. Bentley, Aravinda Chakravarti, et al. 2015. "A Global Reference for Human Genetic Variation." *Nature* 526 (7571): 68–74. https://doi.org/10.1038/nature15393.

Barbier, Maxime, and Thierry Wirth. 2016. "The Evolutionary History, Demography, and Spread of the Mycobacterium Tuberculosis Complex." *Microbiology Spectrum* 4 (4): 4.4.53. https://doi.org/10.1128/microbiolspec.TBTB2-0008-2016.

Basu Mallick, Chandana, Florin Mircea Iliescu, Märt Möls, Sarah Hill, Rakesh Tamang, Gyaneshwer Chaubey, Rie Goto, et al. 2013. "The Light Skin Allele of SLC24A5 in South Asians and Europeans Shares Identity by Descent." *PLoS Genetics* 9 (11): e1003912. https://doi.org/10.1371/journal.pgen.1003912.

Beckles, HILARY McD. 1992. "Kalinago (Carib) Resistance to European Colonisation of the Caribbean." *Caribbean Quarterly* 38 (2/3): 1–124.

Behr, Aaron A., Katherine Z. Liu, Gracie Liu-Fang, Priyanka Nakka, and Sohini Ramachandran. 2016. "Pong: Fast Analysis and Visualization of Latent Clusters in Population Genetic Data." *Bioinformatics (Oxford, England)* 32 (18): 2817–23. https://doi.org/10.1093/bioinformatics/btw327.

Benedetti, Francesca, Sabrina Curreli, Robert C. Gallo, and Davide Zella. 2021. "Tampering of Viruses and Bacteria with Host DNA Repair: Implications for Cellular Transformation." *Cancers* 13 (2): 241. https://doi.org/10.3390/cancers13020241.

Bolnick, Deborah A., Jennifer A. Raff, Lauren C. Springs, Austin W. Reynolds, and Aida T. Miró-Herrans. 2016. "Native American Genomics and Population Histories." *Annual Review of Anthropology* 45 (1): 319–40. https://doi.org/10.1146/annurev-anthro-102215-100036.

Bos, Kirsten I., Kelly M. Harkins, Alexander Herbig, Mireia Coscolla, Nico Weber, Iñaki Comas, Stephen A. Forrest, et al. 2014. "Pre-Columbian Mycobacterial Genomes Reveal Seals as a Source of New World Human Tuberculosis." *Nature* 514 (7523): 494–97. https://doi.org/10.1038/nature13591.

Bryc, Katarzyna, Christopher Velez, Tatiana Karafet, Andres Moreno-Estrada, Andy Reynolds, Adam Auton, Michael Hammer, Carlos D. Bustamante, and Harry Ostrer. 2010. "Genome-Wide Patterns of Population Structure and Admixture among Hispanic/Latino Populations." *Proceedings of the National Academy of Sciences* 107 (supplement_2): 8954–61. https://doi.org/10.1073/pnas.0914618107.

Brynildsrud, Ola B., Caitlin S. Pepperell, Philip Suffys, Louis Grandjean, Johana Monteserin, Nadia Debech, Jon Bohlin, et al. 2018. "Global Expansion of Mycobacterium Tuberculosis Lineage 4 Shaped by Colonial Migration and Local Adaptation." *Science Advances* 4 (10): eaat5869. https://doi.org/10.1126/sciadv.aat5869.

Bush, William S., and Jason H. Moore. 2012. "Chapter 11: Genome-Wide Association Studies." *PLOS Computational Biology* 8 (12): e1002822. https://doi.org/10.1371/journal.pcbi.1002822.

Buzic, I., and V. Guiffra. 2020. "The Paleopathological Evidence on the Origins of Human Tuberculosis: A Review." *Journal of Preventive Medicine and Hygiene* 61 (1 Suppl 1): E3–8. https://doi.org/10.15167/2421-4248/jpmh2020.61.1s1.1379.

Campbell, Michael C., and Sarah A. Tishkoff. 2008. “AFRICAN GENETIC DIVERSITY: Implications for Human Demographic History, Modern Human Origins, and Complex Disease Mapping.” *Annual Review of Genomics and Human Genetics* 9: 403–33. https://doi.org/10.1146/annurev.genom.9.081307.164258.

Cheng, Jade Yu, Thomas Mailund, and Rasmus Nielsen. 2017. “Fast Admixture Analysis and Population Tree Estimation for SNP and NGS Data.” *Bioinformatics* 33 (14): 2148–55. https://doi.org/10.1093/bioinformatics/btx098.

Chiu, Alec M., Erin K. Molloy, Zilong Tan, Ameet Talwalkar, and Sriram Sankararaman. 2022. “Inferring Population Structure in Biobank-Scale Genomic Data.” *The American Journal of Human Genetics* 109 (4): 727–37. https://doi.org/10.1016/j.ajhg.2022.02.015.

Ciccia, Alberto, Amitabh V. Nimonkar, Yiduo Hu, Ildiko Hajdu, Yathish Jagadheesh Achar, Lior Izhar, Sarah A. Petit, et al. 2012. “The ZRANB3 Translocase Associates with Poly-Ubiquitinated PCNA to Promote Fork Restart and Limit Recombination after Replication Stress.” *Molecular Cell* 47 (3): 396–409. https://doi.org/10.1016/j.molcel.2012.05.024.

Crawford, Nicholas G., Derek E. Kelly, Matthew E. B. Hansen, Marcia H. Beltrame, Shaohua Fan, Shanna L. Bowman, Ethan Jewett, et al. 2017. “Loci Associated with Skin Pigmentation Identified in African Populations.” *Science (New York, N.Y.)* 358 (6365): eaan8433. https://doi.org/10.1126/science.aan8433.

Crosby, Alfred W. 1976. “Virgin Soil Epidemics as a Factor in the Aboriginal Depopulation in America.” *The William and Mary Quarterly* 33 (2): 289–99. https://doi.org/10.2307/1922166.

Deagan, Kathleen. 1988. “The Archaeology of the Spanish Contact Period in the Caribbean.” *Journal of World Prehistory* 2 (2): 187–233.

Duffy, Fergal J., January Weiner, Scott Hansen, David L. Tabb, Sara Suliman, Ethan Thompson, Jeroen Maertzdorf, et al. 2019. “Immunometabolic Signatures Predict Risk of Progression to Active Tuberculosis and Disease Outcome.” *Frontiers in Immunology* 10: 527. https://doi.org/10.3389/fimmu.2019.00527.

Ellis, Crystal N., Regina C. LaRocque, Taher Uddin, Bryan Krastins, Leslie M. Mayo-Smith, David Sarracino, Elinor K. Karlsson, et al. 2015. “Comparative Proteomic Analysis Reveals Activation of Mucosal Innate Immune Signaling Pathways during Cholera.” *Infection and Immunity* 83 (3): 1089–1103. https://doi.org/10.1128/IAI.02765-14.

Feng, Yajuan, Kang Tang, Qi Lai, Jingxian Liang, Min Feng, Zhong-Wei Zhou, Haissi Cui, Xiangjun Du, Han Zhang, and Litao Sun. 2022. “The Landscape of Aminoacyl-TRNA Synthetases Involved in Severe Acute Respiratory Syndrome Coronavirus 2 Infection.” *Frontiers in Physiology* 12 (January): 818297. https://doi.org/10.3389/fphys.2021.818297.

Fernandes, Daniel M., Kendra A. Sirak, Harald Ringbauer, Jakob Sedig, Nadin Rohland, Olivia Cheronet, Matthew Mah, et al. 2021. “A Genetic History of the Pre-Contact Caribbean.” *Nature* 590 (7844): 103–10. https://doi.org/10.1038/s41586-020-03053-2.

Forsburg, Susan L. 2004. “Eukaryotic MCM Proteins: Beyond Replication Initiation.” *Microbiology and Molecular Biology Reviews* 68 (1): 109–31. https://doi.org/10.1128/MMBR.68.1.109-131.2004.

Gautier, Mathieu, Alexander Klassmann, and Renaud Vitalis. 2017. “Rehh 2.0: A Reimplementation of the R Package Rehh to Detect Positive Selection from Haplotype Structure.” *Molecular Ecology Resources* 17 (1): 78–90. https://doi.org/10.1111/1755-0998.12634.

Gravel, Simon, Fouad Zakharia, Andres Moreno-Estrada, Jake K. Byrnes, Marina Muzzio, Juan L. Rodriguez-Flores, Eimear E. Kenny, et al. 2013. "Reconstructing Native American Migrations from Whole-Genome and Whole-Exome Data." *PLoS Genetics* 9 (12): e1004023. https://doi.org/10.1371/journal.pgen.1004023.

Hernandez-Pacheco, Natalia, Carlos Flores, Santos Alonso, Celeste Eng, Angel C. Y. Mak, Scott Hunstman, Donglei Hu, et al. 2017. "Identification of a Novel Locus Associated with Skin Colour in African-Admixed Populations." *Scientific Reports* 7 (1): 44548. https://doi.org/10.1038/srep44548.

Huang, Feng, Lei Hu, Yuanmin Zhang, Xingmin Qu, and Junji Xu. 2021. "BMP4 Moderates Glycolysis and Regulates Activation and Interferon-Gamma Production in CD4+ T Cells." *Frontiers in Immunology* 12. https://www.frontiersin.org/articles/10.3389/fimmu.2021.702211.

Jones, David S. 2003. "Virgin Soils Revisited." *The William and Mary Quarterly* 60 (4): 703–42. https://doi.org/10.2307/3491697.

Joseph, Sophie K., and John Lindo. 2022. "The Evolutionary History of Infectious Disease in the Ancient Americas and the Pathogenic Consequences of European Contact." *American Journal of Biological Anthropology* n/a (n/a). https://doi.org/10.1002/ajpa.24595.

Joseph, Sophie K., Nicola Rambaldi Migliore, Anna Olivieri, Antonio Torroni, Amanda C. Owings, Michael DeGiorgio, Wladimir Galarza Ordóñez, et al. 2023. "Genomic Evidence for Adaptation to Tuberculosis in the Andes before European Contact." *IScience* 26 (2): 106034. https://doi.org/10.1016/j.isci.2023.106034.

Keegan, William F., and Corinne Lisette Hofman. 2017. *The Caribbean before Columbus*. New York, NY: Oxford University Press.

Lalueza-Fox, C., F. Luna Calderón, F. Calafell, B Morera, and J. Bertranpetit. 2001. "MtDNA from Extinct Tainos and the Peopling of the Caribbean." *Annals of Human Genetics* 65 (2): 137–51. https://doi.org/10.1046/j.1469-1809.2001.6520137.x.

Lalueza-Fox, C., M.t.p. Gilbert, A.j. Martínez-Fuentes, F. Calafell, and J. Bertranpetit. 2003. "Mitochondrial DNA from Pre-Columbian Ciboneys from Cuba and the Prehistoric Colonization of the Caribbean." *American Journal of Physical Anthropology* 121 (2): 97–108. https://doi.org/10.1002/ajpa.10236.

Lander, Eric S., Lauren M. Linton, Bruce Birren, Chad Nusbaum, Michael C. Zody, Jennifer Baldwin, Keri Devon, et al. 2001. "Initial Sequencing and Analysis of the Human Genome." *Nature* 409 (6822): 860–921. https://doi.org/10.1038/35057062.

Larsen, Clark Spencer. 1994. "In the Wake of Columbus: Native Population Biology in the Postcontact Americas." *American Journal of Physical Anthropology* 37 (S19): 109–54. https://doi.org/10.1002/ajpa.1330370606.

Lee, Oona Y.-C., Houdini H. T. Wu, Helen D. Donoghue, Mark Spigelman, Charles L. Greenblatt, Ian D. Bull, Bruce M. Rothschild, Larry D. Martin, David E. Minnikin, and Gurdyal S. Besra. 2012. "Mycobacterium Tuberculosis Complex Lipid Virulence Factors Preserved in the 17,000-Year-Old Skeleton of an Extinct Bison, Bison Antiquus." *PLOS ONE* 7 (7): e41923. https://doi.org/10.1371/journal.pone.0041923.

Lindo, John, Randall Haas, Courtney Hofman, Mario Apata, Mauricio Moraga, Ricardo A. Verdugo, James T. Watson, et al. 2018. "The Genetic Prehistory of the Andean Highlands 7000 Years BP Though European Contact." *Science Advances* 4 (11): eaau4921. https://doi.org/10.1126/sciadv.aau4921.

Lindo, John, Emilia Huerta-Sánchez, Shigeki Nakagome, Morten Rasmussen, Barbara Petzelt, Joycelynn Mitchell, Jerome S. Cybulski, Eske Willerslev, Michael DeGiorgio, and Ripan S. Malhi. 2016. "A Time Transect of Exomes from a Native American Population before and after European Contact." *Nature Communications* 7 (1): 13175. https://doi.org/10.1038/ncomms13175.

Liu, Shuqin, Yun Fu, Dirk Ziebolz, Simin Li, Gerhard Schmalz, and Fan Li. 2022. "Transcriptomic Analysis Reveals Pathophysiological Relationship between Chronic Obstructive Pulmonary Disease (COPD) and Periodontitis." *BMC Medical Genomics* 15 (1): 130. https://doi.org/10.1186/s12920-022-01278-w.

Lonsdale, John, Jeffrey Thomas, Mike Salvatore, Rebecca Phillips, Edmund Lo, Saboor Shad, Richard Hasz, et al. 2013. "The Genotype-Tissue Expression (GTEx) Project." *Nature Genetics* 45 (6): 580–85. https://doi.org/10.1038/ng.2653.

Mallick, Swapan, Heng Li, Mark Lipson, Iain Mathieson, Melissa Gymrek, Fernando Racimo, Mengyao Zhao, et al. 2016. "The Simons Genome Diversity Project: 300 Genomes from 142 Diverse Populations." *Nature* 538 (7624): 201–6. https://doi.org/10.1038/nature18964.

Malonia, Sunil K., Parul Dutta, Manas Kumar Santra, and Michael R. Green. 2015. "F-Box Protein FBXO31 Directs Degradation of MDM2 to Facilitate P53-Mediated Growth Arrest Following Genotoxic Stress." *Proceedings of the National Academy of Sciences* 112 (28): 8632–37. https://doi.org/10.1073/pnas.1510929112.

Maples, Brian K., Simon Gravel, Eimear E. Kenny, and Carlos D. Bustamante. 2013. "RFMix: A Discriminative Modeling Approach for Rapid and Robust Local-Ancestry Inference." *American Journal of Human Genetics* 93 (2): 278–88. https://doi.org/10.1016/j.ajhg.2013.06.020.

Martin, Debra L, and Alan H Goodman. 2002. "Health Conditions before Columbus: Paleopathology of Native North Americans." *Western Journal of Medicine* 176 (1): 65–68.

Mendisco, F., M. H. Pemonge, E. Leblay, T. Romon, G. Richard, P. Courtaud, and M. F. Deguilloux. 2015. "Where Are the Caribs? Ancient DNA from Ceramic Period Human Remains in the Lesser Antilles." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 370 (1660): 20130388. https://doi.org/10.1098/rstb.2013.0388.

Moreno-Estrada, Andrés, Simon Gravel, Fouad Zakharia, Jacob L. McCauley, Jake K. Byrnes, Christopher R. Gignoux, Patricia A. Ortiz-Tello, et al. 2013. "Reconstructing the Population Genetic History of the Caribbean." *PLOS Genetics* 9 (11): e1003925. https://doi.org/10.1371/journal.pgen.1003925.

Murdoch, H. Adlai. 2021. "Diaspora, Creolization, and the Boundaries of Caribbean Identity." In *Diasporas, Cultures of Mobilities, 'Race' 1 : Diasporas and Cultures of Migrations*, edited by Judith Misrahi-Barak and Claudine Raynaud, 271–93. Horizons Anglophones. Montpellier: Presses universitaires de la Méditerranée. http://books.openedition.org/pulm/9055.

Nägele, Kathrin, Cosimo Posth, Miren Iraeta Orbegozo, Yadira Chinique de Armas, Silvia Teresita Hernández Godoy, Ulises M. González Herrera, Maria A. Nieves-Colón, et al. 2020. "Genomic Insights into the Early Peopling of the Caribbean." *Science* 369 (6502): 456–60. https://doi.org/10.1126/science.aba8697.

Nie, Anzheng, Bao Sun, Zhihui Fu, and Dongsheng Yu. 2019. "Roles of Aminoacyl-TRNA Synthetases in Immune Regulation and Immune Diseases." *Cell Death & Disease* 10 (12): 1–14. https://doi.org/10.1038/s41419-019-2145-5.

Pickrell, Joseph K., and Jonathan K. Pritchard. 2012. “Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data.” *PLOS Genetics* 8 (11): e1002967. https://doi.org/10.1371/journal.pgen.1002967.

Reich, David, Nick Patterson, Desmond Campbell, Arti Tandon, Stéphane Mazieres, Nicolas Ray, Maria V. Parra, et al. 2012. “Reconstructing Native American Population History.” *Nature* 488 (7411): 370–74. https://doi.org/10.1038/nature11258.

Roberts, Megan C., Alison E. Fohner, Latrice Landry, Dana Lee Olstad, Amelia K. Smit, Erin Turbitt, and Caitlin G. Allen. 2021. “Advancing Precision Public Health Using Human Genomics: Examples from the Field and Future Research Opportunities.” *Genome Medicine* 13 (1): 97. https://doi.org/10.1186/s13073-021-00911-0.

Romagnani, S. 1999. “Th1/Th2 Cells.” *Inflammatory Bowel Diseases* 5 (4): 285–94. https://doi.org/10.1097/00054725-199911000-00009.

Ross, Ann H., William F. Keegan, Michael P. Pateman, and Colleen B. Young. 2020. “Faces Divulge the Origins of Caribbean Prehistoric Inhabitants.” *Scientific Reports* 10 (1): 147. https://doi.org/10.1038/s41598-019-56929-3.

Rothschild, Bruce M., Larry D. Martin, Galit Lev, Helen Bercovier, Gila Kahila Bar-Gal, Charles Greenblatt, Helen Donoghue, Mark Spigelman, and David Brittain. 2001. “Mycobacterium Tuberculosis Complex DNA from an Extinct Bison Dated 17,000 Years before the Present.” *Clinical Infectious Diseases* 33 (3): 305–11. https://doi.org/10.1086/321886.

Sans, Monica. 2000. “Admixture Studies in Latin America: From the 20th to the 21st Century.” *Human Biology* 72 (1): 155–77.

Schrader, C., D. Janssen, W. Klapper, J.-U. Siebmann, P. Meusers, G. Brittinger, M. Kneba, M. Tiemann, and R. Parwaresch. 2005. “Minichromosome Maintenance Protein 6, a Proliferation Marker Superior to Ki-67 and Independent Predictor of Survival in Patients with Mantle Cell Lymphoma.” *British Journal of Cancer* 93 (8): 939–45. https://doi.org/10.1038/sj.bjc.6602795.

Schroeder, Hannes, María C. Ávila-Arcos, Anna-Sapfo Malaspinas, G. David Poznik, Marcela Sandoval-Velasco, Meredith L. Carpenter, José Víctor Moreno-Mayar, et al. 2015. “Genome-Wide Ancestry of 17th-Century Enslaved Africans from the Caribbean.” *Proceedings of the National Academy of Sciences* 112 (12): 3669–73. https://doi.org/10.1073/pnas.1421784112.

Schroeder, Hannes, Martin Sikora, Shyam Gopalakrishnan, Lara M. Cassidy, Pierpaolo Maisano Delser, Marcela Sandoval Velasco, Joshua G. Schraiber, et al. 2018. “Origins and Genetic Legacies of the Caribbean Taino.” *Proceedings of the National Academy of Sciences* 115 (10): 2341–46. https://doi.org/10.1073/pnas.1716839115.

Shapiro, B., and M. Hofreiter. 2014. “A Paleogenomic Perspective on Evolution and Gene Function: New Insights from Ancient DNA.” *Science (New York, N.Y.)* 343 (6169): 1236573. https://doi.org/10.1126/science.1236573.

Slatkin, Montgomery. 2008. “Linkage Disequilibrium — Understanding the Evolutionary Past and Mapping the Medical Future.” *Nature Reviews. Genetics* 9 (6): 477–85. https://doi.org/10.1038/nrg2361.

Slatkin, Montgomery, and Fernando Racimo. 2016. “Ancient DNA and Human History.” *Proceedings of the National Academy of Sciences* 113 (23): 6380–87. https://doi.org/10.1073/pnas.1524306113.

Stead, W W, K D Eisenach, M D Cave, M L Beggs, G L Templeton, C O Thoen, and J H Bates. 1995. "When Did Mycobacterium Tuberculosis Infection First Occur in the New World? An Important Question with Public Health Implications.," 2.

Stephen Schaffner & Pardis Sabeti. 2008. "Evolutionary Adaptation in the Human Lineage" 1 (1): 14.

"Taíno Culture History." 2017. Florida Museum. September 20, 2017. https://www.floridamuseum.ufl.edu/histarch/research/haiti/en-bas-saline/taino-culture/.

TallBear, Kim. 2013. "Genomic Articulations of Indigeneity." *Social Studies of Science* 43 (4): 509–33. https://doi.org/10.1177/0306312713483893.

Tataru, Paula, Maria Simonsen, Thomas Bataillon, and Asger Hobolth. 2017. "Statistical Inference in the Wright–Fisher Model Using Allele Frequency Data." *Systematic Biology* 66 (1): e30–46. https://doi.org/10.1093/sysbio/syw056.

Turda, Marius. 2022. "Legacies of Eugenics: Confronting the Past, Forging a Future." *Ethnic and Racial Studies* 45 (13): 2470–77. https://doi.org/10.1080/01419870.2022.2095222.

Turner, Stephen D. 2014. "Qqman: An R Package for Visualizing GWAS Results Using Q-Q and Manhattan Plots." bioRxiv. https://doi.org/10.1101/005165.

Venter, J. Craig, Mark D. Adams, Eugene W. Myers, Peter W. Li, Richard J. Mural, Granger G. Sutton, Hamilton O. Smith, et al. 2001. "The Sequence of the Human Genome." *Science* 291 (5507): 1304–51. https://doi.org/10.1126/science.1058040.

Wang, Kai, Mingyao Li, and Hakon Hakonarson. 2010. "ANNOVAR: Functional Annotation of Genetic Variants from High-Throughput Sequencing Data." *Nucleic Acids Research* 38 (16): e164. https://doi.org/10.1093/nar/gkq603.

Weitzman, Matthew D., and Amélie Fradet-Turcotte. 2018. "Virus DNA Replication and the Host DNA Damage Response." *Annual Review of Virology* 5 (1): 141–64. https://doi.org/10.1146/annurev-virology-092917-043534.

Zeng, Tian, Yang Guan, Yu-kun Li, Qing Wu, Xiao-jun Tang, Xin Zeng, Hui Ling, and Juan Zou. 2021. "The DNA Replication Regulator MCM6: An Emerging Cancer Biomarker and Target." *Clinica Chimica Acta* 517 (June): 92–98. https://doi.org/10.1016/j.cca.2021.02.005.

Zheng, Xiuwen, David Levine, Jess Shen, Stephanie M. Gogarten, Cathy Laurie, and Bruce S. Weir. 2012. "A High-Performance Computing Toolset for Relatedness and Principal Component Analysis of SNP Data." *Bioinformatics (Oxford, England)* 28 (24): 3326–28. https://doi.org/10.1093/bioinformatics/bts606.

Supplemental Information

Supplementary Table S1. Information about sequencing, geographic locations, carbon dating, and mean read depth for the ancient Caribbean population cohort. Related to Figures 1, 2, and 3.

Sample ID	Population	14C BP	Mean read depth	Reference
I12344	Dominican Republic	1200-1400 [based on three direct dates from the site]*	8.05	Fernandes et al. 2021
I12347	Dominican Republic	1200-1400 [based on three direct dates from the site]*	9.95	Fernandes et al. 2021
I12350	Dominican Republic	1200-1400 [based on three direct dates from the site]*	7.49	Fernandes et al. 2021
I13189	Dominican Republic	855 ± 20	4.91	Fernandes et al. 2021
I14992	Dominican Republic	536 ± 18	6.84	Fernandes et al. 2021
I15973	Dominican Republic	600-1600*	5.15	Fernandes et al. 2021
I8118 (SM)	Dominican Republic	690 ± 20	4.9	Fernandes et al. 2021
I13207	Dominican Republic	1200-1700 [based on two direct dates from the site]*	4.53	Fernandes et al. 2021
I13208	Dominican Republic	710 ± 20	4.49	Fernandes et al. 2021
I13318	The Bahamas	900-1500*	4.38	Fernandes et al. 2021
I13319	The Bahamas	900-1500*	4.38	Fernandes et al. 2021
I13320	The Bahamas	900-1500*	4.38	Fernandes et al. 2021
I13321	The Bahamas	900-1500*	4.37	Fernandes et al. 2021
I13322	The Bahamas	900-1500*	4.26	Fernandes et al. 2021
I13323	Puerto Rico	785 ± 20	4.22	Fernandes et al. 2021
I13324	Puerto Rico	900-1300*	4.21	Fernandes et al. 2021
I13326	Puerto Rico	800-1100*	4.1	Fernandes et al. 2021
I10758	Curaçao	720 ± 20	4.96	Fernandes et al. 2021
I12575	Haiti	800-1200*	5.31	Fernandes et al. 2021

*Indicates cultural dating

Supplementary Table S2. Information about modern Caribbean population cohort (Auton et al. 2015). Related to Figures 1, 2, and 3.

Sample ID	Reference		Sample ID	Reference
HG00740	1000 Genomes Project (PUR)		HG00551	1000 Genomes Project (PUR)
HG01161	1000 Genomes Project (PUR)		HG00638	1000 Genomes Project (PUR)
HG01173	1000 Genomes Project (PUR)		HG00731	1000 Genomes Project (PUR)
HG01197	1000 Genomes Project (PUR)		HG00736	1000 Genomes Project (PUR)
HG01200	1000 Genomes Project (PUR)		HG00743	1000 Genomes Project (PUR)
HG01205	1000 Genomes Project (PUR)		HG00640	1000 Genomes Project (PUR)
HG01248	1000 Genomes Project (PUR)		HG01094	1000 Genomes Project (PUR)
HG00554	1000 Genomes Project (PUR)		HG01102	1000 Genomes Project (PUR)
HG00732	1000 Genomes Project (PUR)		HG01107	1000 Genomes Project (PUR)
HG00737	1000 Genomes Project (PUR)		HG01162	1000 Genomes Project (PUR)
HG00637	1000 Genomes Project (PUR)		HG00553	1000 Genomes Project (PUR)
HG01089	1000 Genomes Project (PUR)		HG01167	1000 Genomes Project (PUR)
HG01104	1000 Genomes Project (PUR)		HG01174	1000 Genomes Project (PUR)
HG01111	1000 Genomes Project (PUR)		HG01198	1000 Genomes Project (PUR)
HG01177	1000 Genomes Project (PUR)		HG01326	1000 Genomes Project (PUR)
HG01191	1000 Genomes Project (PUR)		HG01092	1000 Genomes Project (PUR)
HG01204	1000 Genomes Project (PUR)		HG01097	1000 Genomes Project (PUR)
HG01393	1000 Genomes Project (PUR)		HG01105	1000 Genomes Project (PUR)
HG01398	1000 Genomes Project (PUR)		HG01047	1000 Genomes Project (PUR)
HG01058	1000 Genomes Project (PUR)		HG01054	1000 Genomes Project (PUR)
HG01060	1000 Genomes Project (PUR)		HG01049	1000 Genomes Project (PUR)
HG01247	1000 Genomes Project (PUR)		HG01061	1000 Genomes Project (PUR)
HG01072	1000 Genomes Project (PUR)		HG01051	1000 Genomes Project (PUR)
HG01088	1000 Genomes Project (PUR)		HG01066	1000 Genomes Project (PUR)
HG01077	1000 Genomes Project (PUR)		HG01073	1000 Genomes Project (PUR)
HG01095	1000 Genomes Project (PUR)		HG01063	1000 Genomes Project (PUR)
HG01325	1000 Genomes Project (PUR)		HG01080	1000 Genomes Project (PUR)
HG01110	1000 Genomes Project (PUR)		HG01085	1000 Genomes Project (PUR)
HG01286	1000 Genomes Project (PUR)		HG01070	1000 Genomes Project (PUR)
HG01414	1000 Genomes Project (PUR)		HG01075	1000 Genomes Project (PUR)
HG01312	1000 Genomes Project (PUR)		HG01082	1000 Genomes Project (PUR)
HG01305	1000 Genomes Project (PUR)		HG01395	1000 Genomes Project (PUR)
HG01052	1000 Genomes Project (PUR)		HG01164	1000 Genomes Project (PUR)
HG01064	1000 Genomes Project (PUR)		HG01403	1000 Genomes Project (PUR)

Sample ID	Reference		Sample ID	Reference
HG01069	1000 Genomes Project (PUR)		HG01171	1000 Genomes Project (PUR)
HG01083	1000 Genomes Project (PUR)		HG01176	1000 Genomes Project (PUR)
HG01413	1000 Genomes Project (PUR)		HG01183	1000 Genomes Project (PUR)
HG01402	1000 Genomes Project (PUR)		HG01188	1000 Genomes Project (PUR)
HG01190	1000 Genomes Project (PUR)		HG01079	1000 Genomes Project (PUR)
HG00734	1000 Genomes Project (PUR)		HG01086	1000 Genomes Project (PUR)
HG00739	1000 Genomes Project (PUR)		HG00742	1000 Genomes Project (PUR)
HG01302	1000 Genomes Project (PUR)		HG01168	1000 Genomes Project (PUR)
HG01311	1000 Genomes Project (PUR)		HG01170	1000 Genomes Project (PUR)
HG01323	1000 Genomes Project (PUR)		HG01182	1000 Genomes Project (PUR)
HG01392	1000 Genomes Project (PUR)		HG01187	1000 Genomes Project (PUR)
HG01405	1000 Genomes Project (PUR)		HG01098	1000 Genomes Project (PUR)
HG01412	1000 Genomes Project (PUR)		HG01101	1000 Genomes Project (PUR)
HG00641	1000 Genomes Project (PUR)		HG01303	1000 Genomes Project (PUR)
HG01048	1000 Genomes Project (PUR)		HG01308	1000 Genomes Project (PUR)
HG01055	1000 Genomes Project (PUR)		HG01396	1000 Genomes Project (PUR)
HG01067	1000 Genomes Project (PUR)			

Supplementary Table S3. Outgroup samples utilized from the SGDP dataset (Mallick et al. 2016). Related to Figures 1 and 2.

Sample	Population
S_Mbuti-1	Africa
S_Mbuti-2	Africa
S_Mbuti-3	Africa
B_Mbuti-4	Africa
S_Gambian-1	Africa
S_Gambian-2	Africa
S_Yoruba-2	Africa
S_Esan-1	Africa
S_Yoruba-1	Africa
S_Esan-2	Africa
S_Mandenka-1	Africa
S_Mandenka-2	Africa
S_Mende-1	Africa
B_Mandenka-3	Africa
B_Yoruba-3	Africa
S_Mende-2	Africa
S_Chane-1	Americas
S_Surui-1	Americas
S_Surui-2	Americas
S_Karitiana-1	Americas
S_Karitiana-2	Americas
S_Piapoco-1	Americas
S_Piapoco-2	Americas
S_Pima-1	Americas
S_Pima-2	Americas
S_Mayan-1	Americas
S_Mayan-2	Americas
S_Mixe-2	Americas
S_Mixe-3	Americas
S_Mixtec-1	Americas
S_Mixtec-2	Americas
S_Quechua-3	Americas
S_Quechua-2	Americas
B_Karitiana-3	Americas

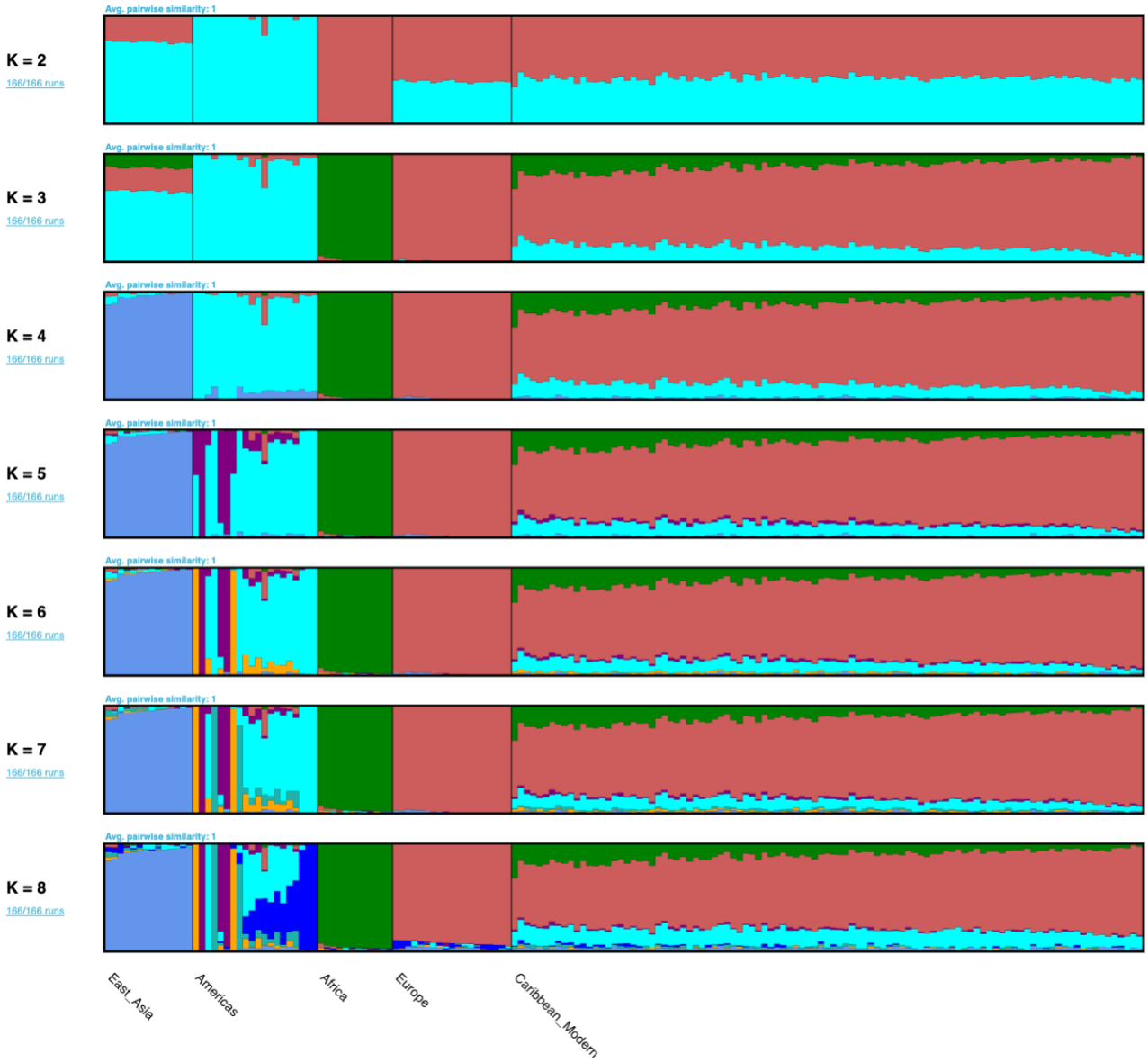
Sample	Population
B_Mixe-1	Americas
S_Quechua-1	Americas
S_Han-2	East_Asia
S_Dai-3	East_Asia
S_Han-1	East_Asia
S_Hezhen-1	East_Asia
S_Hezhen-2	East_Asia
S_Dai-2	East_Asia
S_Dai-1	East_Asia
S_Japanese-1	East_Asia
S_Japanese-2	East_Asia
S_Japanese-3	East_Asia
S_Korean-2	East_Asia
S_Korean-1	East_Asia
B_Dai-4	East_Asia
B_Han-3	East_Asia
S_Bulgarian-2	Europe
S_English-1	Europe
S_English-2	Europe
S_French-1	Europe
S_French-2	Europe
S_Basque-1	Europe
S_Hungarian-2	Europe
S_Hungarian-1	Europe
S_Bergamo-1	Europe
S_Bergamo-2	Europe
S_Tuscan-2	Europe
S_Tuscan-1	Europe
S_Norwegian-1	Europe
S_Polish-1	Europe
S_Spanish-1	Europe
S_Spanish-2	Europe
B_Crete-1	Europe
B_French-3	Europe

Supplementary Table S4. Outgroup samples utilized for the natural selection scans (Auton et al. 2015). Related to Figure 3.

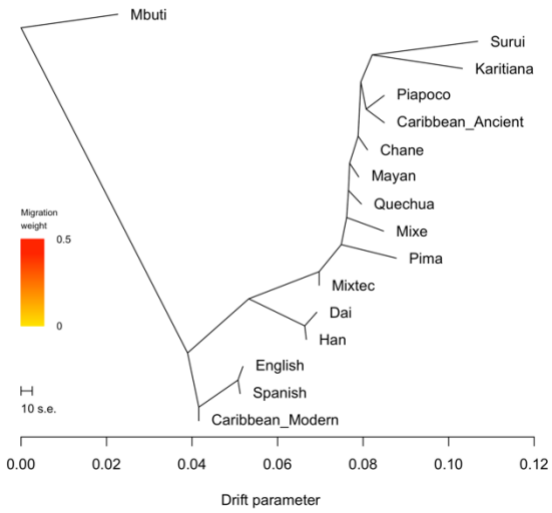
Sample	Population code	Reference
NA12828	CEU	1000 Genomes Project
NA12830	CEU	1000 Genomes Project
NA12842	CEU	1000 Genomes Project
NA12873	CEU	1000 Genomes Project
NA12878	CEU	1000 Genomes Project
NA12546	CEU	1000 Genomes Project
NA12414	CEU	1000 Genomes Project
NA12399	CEU	1000 Genomes Project
NA11832	CEU	1000 Genomes Project
NA12717	CEU	1000 Genomes Project
NA11894	CEU	1000 Genomes Project
NA11919	CEU	1000 Genomes Project
NA11933	CEU	1000 Genomes Project
NA11995	CEU	1000 Genomes Project
NA12006	CEU	1000 Genomes Project
NA12044	CEU	1000 Genomes Project
NA12748	CEU	1000 Genomes Project
NA12750	CEU	1000 Genomes Project
NA12234	CEU	1000 Genomes Project
NA12762	CEU	1000 Genomes Project

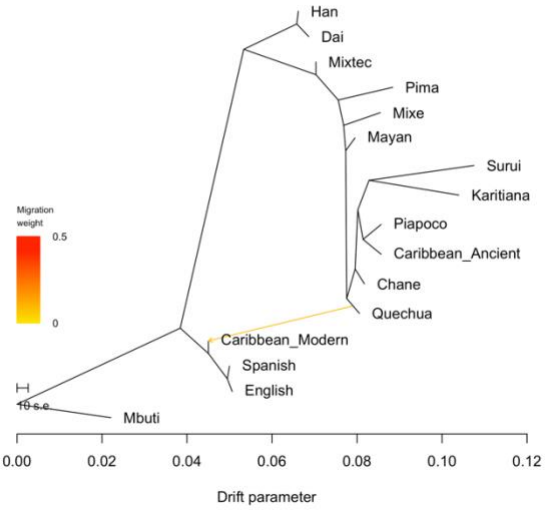
Sample	Population code	Reference
NA19200	YRI	1000 Genomes Project
NA18489	YRI	1000 Genomes Project
NA18504	YRI	1000 Genomes Project
NA19130	YRI	1000 Genomes Project
NA18511	YRI	1000 Genomes Project
NA18516	YRI	1000 Genomes Project
NA18523	YRI	1000 Genomes Project
NA19213	YRI	1000 Genomes Project
NA18508	YRI	1000 Genomes Project
NA18510	YRI	1000 Genomes Project
NA19225	YRI	1000 Genomes Project
NA18522	YRI	1000 Genomes Project
NA18876	YRI	1000 Genomes Project
NA18908	YRI	1000 Genomes Project
NA18910	YRI	1000 Genomes Project
NA18915	YRI	1000 Genomes Project
NA18934	YRI	1000 Genomes Project
NA18933	YRI	1000 Genomes Project
NA19092	YRI	1000 Genomes Project
NA19201	YRI	1000 Genomes Project

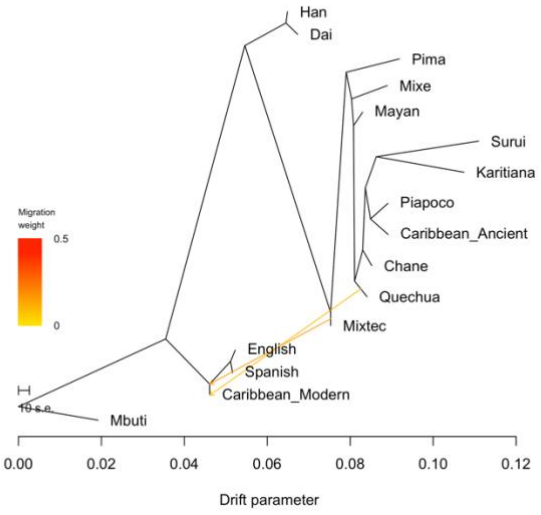
Supplementary Figure S1. Scope cluster analysis visualizations with K=2–8 (Chiu et al. 2022).

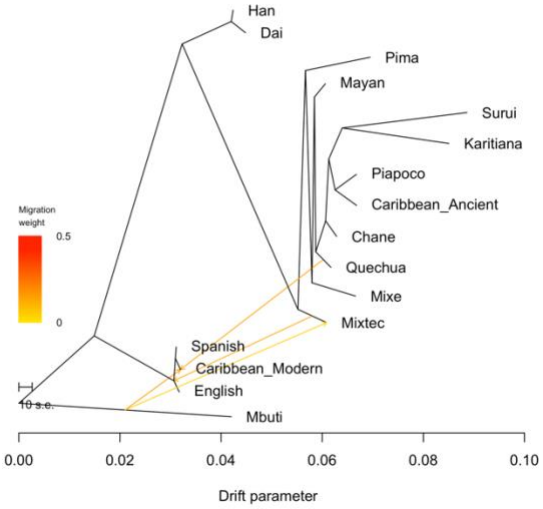


Supplementary Figure S2. Highest likelihood TreeMix scenarios for *m*=0, 1, 2, and 4 migration events (Pickrell and Pritchard 2012).









Supplementary Table S5. Top 30 annotated SNPs under high probability of selection from Ohana selection scan on ancient Caribbean cohort (Cheng, Mailund, and Nielsen 2017; Wang, Li, and Hakonarson 2010).

Chr*	Position*	Likeliood	Annotated gene
15	48400199	8.941E+00	LINC01491(dist=261766),SLC24A5(dist=12988)
11	19620227	8.85E+00	NAV2
11	19696415	8.42E+00	NAV2
14	79244045	8.30E+00	NRXN3
7	137529578	7.90E+00	DGKI
11	70734035	7.87E+00	SHANK2
16	87394386	7.76E+00	FBXO31
15	48218411	7.63E+00	LINC01491(dist=79978),SLC24A5(dist=194776)
11	19629129	7.54E+00	NAV2
3	188694749	7.50E+00	TPRG1-AS1(dist=29321),TPRG1(dist=195011)
12	56565392	7.37E+00	SMARCC2
5	33952106	7.32E+00	SLC45A2
5	33954880	7.32E+00	SLC45A2
11	19622823	7.32E+00	NAV2
14	32183849	7.32E+00	NUBPL(NM_001201574:c.-731730>0)
3	188838591	7.29E+00	TPRG1-AS1(dist=173163),TPRG1(dist=51169)
10	83670102	7.22E+00	NRG3
11	19700177	7.19E+00	NAV2
1	1249187	7.16E+00	INTS11
9	117108665	7.10E+00	AKNA
7	130742066	7.10E+00	LINC-PINT
1	120463230	7.09E+00	NOTCH2,NOTCH2NLC
1	120505996	7.09E+00	NOTCH2,NOTCH2NLC
4	26585311	6.98E+00	TBC1D19(dist=235)
4	83991529	6.89E+00	COPS4
3	66539628	6.86E+00	LRIG1
18	67691520	6.80E+00	RTTN
6	107990428	6.79E+00	SOBP(dist=7918),SCML4(dist=32930)
9	117112884	6.72E+00	AKNA
9	117112886	6.72E+00	AKNA

*Hg19 coordinates

Supplementary Table S6. Top 30 annotated SNPs under high probability of selection from Ohana selection scan on modern Caribbean cohort (Cheng, Mailund, and Nielsen 2017; Wang, Li, and Hakonarson 2010).

Chr*	Position*	Likelihood	Annotated gene
2	135954797	1.32E+02	ZRANB3
2	136138627	1.32E+02	ZRANB3
2	136176540	1.28E+02	ZRANB3
15	32777439	1.27E+02	GOLGA8O(dist=29604),WHAMMP1(dist=34610)
2	136429366	1.26E+02	R3HDM1
14	106331658	1.26E+02	MIR4539(dist=5074),FAM30A(dist=52180)
2	135837906	1.24E+02	RAB3GAP1
2	135907088	1.24E+02	RAB3GAP1
2	136381348	1.24E+02	R3HDM1
2	136098560	1.23E+02	ZRANB3
2	136608646	1.19E+02	MCM6
2	136616754	1.19E+02	MCM6
2	136707982	1.19E+02	DARS1
2	136328890	1.17E+02	R3HDM1
2	136352327	1.17E+02	R3HDM1
14	106330474	1.12E+02	MIR4539(dist=3890),FAM30A(dist=53364)
18	55441724	1.12E+02	ATP8B1
2	136825272	1.10E+02	DARS-AS1(dist=60160),CXCR4(dist=46647)
2	136617805	1.07E+02	MCM6
11	38526006	1.05E+02	LINC02760(dist=549793),LINC02759(dist=113808)
2	136823866	1.04E+02	DARS-AS1(dist=58754),CXCR4(dist=48053)
18	55441626	1.03E+02	ATP8B1
8	105083842	1.03E+02	RIMS2
4	6239015	1.02E+02	LINC02495(dist=3352),WFS1(dist=32562)
11	38556519	1.01E+02	LINC02760(dist=580306),LINC02759(dist=83295)
3	195365807	1.00E+02	APOD(dist=54996),LOC105374297(dist=7699)
3	195365815	1.00E+02	APOD(dist=55004),LOC105374297(dist=7691)
15	23227285	1.00E+02	WHAMMP3(dist=18928),GOLGA8IP(dist=27957)
8	96454433	9.94E+01	C8orf37-AS1
2	4053881	9.86E+01	LOC105373394(dist=31011),LINC01249(dist=621927)

*Hg19 coordinates