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.

Jack Warn April 10, 2024

An aDNA Approach to Investigating Post-Admixture Selection in Mexico: Assessing the (Immuno)Genetic Consequences of European Colonization

by

Jack Warn

Dr. John Lindo
Advisor

Biology

Dr. John Lindo
Advisor

Dr. Karen Conneely
Committee Member

Dr. Daniel Weissman
Committee Member

2024

An aDNA Approach to Investigating Post-Admixture Selection in Mexico: Assessing the (Immuno)Genetic Consequences of European Colonization

by

Jack Warn

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

Biology

2024

Abstract

An aDNA Approach to Investigating Post-Admixture Selection in Mexico: Assessing the (Immuno)Genetic Consequences of European Colonization

By Jack Warn

The development and advancement of ancient DNA technologies have opened up new possibilities in recent years to gain insight into past demographic histories, population movements, and to understand the various evolutionary processes that have led to the modern day. The onset of European colonization in the 15th century marked a significant turning point in the history of the Americas and the world. The emergence of the trans-Atlantic slave trade was accompanied by the interchange of populations previously geographically isolated and genetically distinct from each other. With this increased contact between populations, various pathogens were introduced to certain populations of Indigenous Americans which, when combined with warfare and slavery, caused devastating population collapse. As such, modern genomic methods have the potential to illustrate how post-admixture selection shaped our evolution, particularly with respect to genes regulating the immune system and response. However, whole-genome data for Latin American and Indigenous American populations are drastically underrepresented in genomics research. This not only stifles biological discovery, but makes our knowledge disproportionately relevant to populations of European descent, thereby corrupting understandings of health and genetics.

Recently, 12 whole-genome ancient samples were discovered in Mexico. including 4 high-coverage (>1x) samples. With these samples, there is a potential to further understanding of evolutionary histories in the region. By sequencing the recently-discovered genomes of ancient Mexican individuals alongside modern Mexican samples, we hope to identify several gene pathways and proteins that have been affected as a result of European colonization in Mexico. Moreover, it is our goal that these findings can be useful in understanding not only how post-admixture selection in this context shaped the modern population, but also in directing population health interventions in the region.

Using computational genomic methods, modern and ancient Mexican genomes will be analyzed for signals of selection, with particular focus on genes affecting the immune system and human health. Utilizing a combination of demographic analyses and selection scans, the study aims to achieve a deeper understanding of the processes of evolution that have shaped the contemporary Mexican population.

An aDNA Approach to Investigating Post-Admixture Selection in Mexico: Assessing the (Immuno)Genetic Consequences of European Colonization

by

Jack Warn

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

Biology

2024

Acknowledgements

I would like to first thank my advisor, Dr. John Lindo, for his guidance throughout this process. Without his lab and his passion for ancient genomics, I never could have pursued this project. And to my committee members, Dr. Karen Conneely and Dr. Daniel Weissman, I am honored and grateful to receive your wisdom and insight. I'd like to give a special thank you to Sophie Joseph for the immense support and mentorship you've provided throughout the year. Additionally, to Dr. Alexander Escobar and the Biology Department, I'm incredibly grateful for the opportunity to participate in this program and the education I've received over the past four years. Lastly, I'd like to thank my family and friends for their constant support.

Table of Contents

Introduction

The trans-atlantic slave trade and European colonization of the Americas brought about an unprecedented level of admixture. Populations prior to contact experienced an upturn in genetic diversity through genetic drift and mutation, whilst remaining somewhat geographically isolated (Norris et al). Modern-day Latin American populations possess novel haplotypes -- combinations of genetic variants which tend to be inherited together -- that previously never existed together. In recent years, there has been a growing interest in studying post-admixture selection -- that is, the selective pressures on populations of mixed genetic ancestries. In a 2018 study, Norris and colleagues analyzed 347 admixed Latin American genomes along with 1102 putative ancestral source genomes. In doing so, they identified several SNPs in which the modern population contained a genetic frequency unexpected based on their admixture ratios, coining these "ancestry-enriched SNPs". Reading this paper sparked an interest in incorporating ancient DNA analysis into studies of selection in admixed Latin American populations. Notably, many scholars suggest particular utility in studying selection by simultaneous analysis of ancient and modern genomes, as it is necessary to incorporate data of pre-contact immunity to more accurately assess the ways in which colonization-related selective pressures shaped Latin American genomes (Collen et al., 2022; Rishishwar et al., 2015; Norris et al., 2020; Vicuña et al., 2020; Mendoza-Revilla et al., 2022)

Of particular interest to the population geneticists and evolutionary biologists is the human health implications of post-admixture selection and these "ancestrally-enriched SNPs". Statistically-significant ancestry-enriched haplotypes or SNPs are hypothesized to arise out of some sort of adaptive advantage due to the existing selective pressures. Understanding the mechanism by which admixed populations undergo natural selection is crucial to understanding

modern human genetics. Recent studies that have used ancestry-enrichment analysis have identified at a variety of health-related loci, including those related to pathways involving adaptive and innate immunity.

In 2023, Villa-Islas and colleagues conducted a demographic study of Mexico with 12 whole-genome ancient samples discovered along with various mitochondrial genomes (Villa-Islas et al., 2023). This provided an opportunity to incorporate whole-genome ancient data into studies of post-admixture selection and immune system evolution. The potential for these samples to translate into meaningful results and insights in the field cannot be understated, especially given the rarity of ancient samples and particularly Latin American ancient samples. Along with previously published modern samples, I began to plan on conducting a study to identify signals of selection on the genome in response to the introduction of novel pathogens brought about by European colonization.

Literature Review

a. Ancient individuals and pre-contact Mexico

Villa-Islas and colleagues conducted a study in 2022 presented whole-genome data for 12 individuals alongside mitochondrial DNA for 27 individuals across eight distinct pre-hispanic archaeological sites in Mexico. This study presents a collection of mitogenomic and whole-genome data that is unprecedented among the pre-hispanic Mexican genetic record (Villa-Islas et al 2022). Because ancient genomic data from Latin America is so limited, this provides a crucial opportunity to uncover the genetic variation and population history of the region.

Prior to colonization, Mexico was geographically divided into Aridoamerica in the North and Mesoamerica in the South. While primarily hunter-gatherers made up the population of

Aridoamerica, many of the larger agriculture-based societies in history thrived in Mesoamerica (Villa-Islas et al 2022). Due to multiple droughts over time, the border between these two regions moved south between 900 and 1300 CE. This climate change, also known as the Medieval Warm Period, led to a variety of population shifts and migrations in Mexican pre-hispanic populations. These demographic and climate shifts would have a particular impact on groups along the Northern Frontier of Mesoamerica (NFM), of which this study investigates several individuals. To start, the extent of endogenous ancient DNA was evaluated via shotgun sequencing for the recovered samples. Subsequently, whole-genome capture was performed with informative amounts of data (0.01 - 4.7 x) obtained for 12 individuals. These individuals came primarily from Mesoamerica: Sierra Gorda (n=4), Cañada de la Virgen (n=4), and Michoacán (n=2). Ancient genomes of two individuals from Aridoamerica, specifically in the Sierra Tarahumara region, were also recovered. The genetic ancestries of the central Mexican individuals were revealed to be more complex, as individuals from those regions did not appear to have a higher shared genetic drift with any present-day Indigenous population according to f3 outgroup values. This result is consistent with the understanding that central mexican populations have undergone extensive gene flow between them and are, as such, not completely genetically diverged from each other (Villa-Islas et al 2022).

b. Demographic History of Mexico

The peopling of Mesoamerica -- the region that ultimately became home to a complex array of Indigenous Central American communities up until European contact -- is fascinating and highly contested (Haines and Steckel 2000). Moreover, the demographic history of Mexico can be divided into four main time periods: Ancient (A.D. - 1519), Colonial (1519-1821), National (1821 - 1910), and Modern (1910 - present). Estimates of the original peopling event of the Americas are highly contested, ranging from 20,000 to 70,000 years ago. Additionally, the number of Asian

migration waves into the Americas is still unknown (Haines and Steckel 2000). There is evidence of continuous human occupation in the Tehuacán Valley specifically for at least the past 12,000 years. Notably, population estimates for the Central Mexico Basin are incredibly high -- up to around 1 - 1.12 million in 1519. The paleodemographic record shows exposure to pathogens, anemia, osteoarthritis, stress, and more among these populations. In fact, Mesoamerican life expectancy estimates suggest that those who lived until 15 would typically die between 28 and 44 (Haines and Steckel 2000). Yet, very little is known about how, given poor nutrition, ill health, and other stressors, this population was able to sustain itself. In studying ancient Mexico, the questions of how these civilizations survived and how these Indigenous civilizations perished both must be answered. Some evidence suggests that many Indigenous Mexican populations -- prior to the arrival of Hernán Cortes in 1519 -- were in a high-pressure demographic system where early mating and reproduction (high fertility) counteracted low life expectancies (high mortality).

The population decrease brought about by European colonization in Mexico is documented as a demographic catastrophe -- "one of the worst in the history of humanity". Cook and Borah suggest that the native population went from 25.2 million in in 1519 to 6.3 million in 1545, 2.5 million in 1570, and ultimately 1.2 million in 1620 -- a population loss of around 90-96% which is, largely, corroborated by various independent estimates conducted by other scholars -- thought the lack of pre-colombian records complicates attempts to estimate population loss (Haines and Steckel 2000). An epidemic of smallpox immediately following contact was responsible for wiping out an estimated 20-30% of the Mexican population (Acuna-Soto et al 2002). Moreover, this was followed by periods of unfavorable environmental conditions and the spread of various other transmittable diseases which ultimately left the population decimated. Historians, by and large, have reached a consensus that much of this monumental population decrease in Mexico and across

the Americas was due to waves of infectious disease brought via the transatlantic slave trade and European colonization. Entire families were often struck by smallpox, leaving no one able to provide. Thus, many Indigenous Mexicans died of hunger and dehydration before dying of smallpox (Haines and Steckel 2000). European colonization left a once booming population of Indigenous Mexicans in shambles.

Regarding the samples used in the Villa-Islas study, the Mesoamerican Indigenous groups occupying Sierra Tarahumara, La Mina, Sierra Gorda, and Cañada de la Virgen show evidence of continuous occupation in the region. Notably, however, the Sierra Tarahumara and Cañada de la Virgen populations show genetic ancestry from two distinct ghost populations.

c. Genetics and Health

The completion of the human genome project inspired and opened the door to the field of genomic study. Because of the sheer scope of the project, it allows the scientific community to better understand genetic diversity across populations, study evolutionary processes, and link disease patterns to genetics. Scientists largely agree that when examining two randomly-selected individuals, their genomes contain about 0.1% of differences or variations, called polymorphism. Genetic variation or polymorphism arises due to mutation. Comparative studies done on identical and fraternal twins suggest that this genetic polymorphism contributes to disease susceptibility, among other things (Shastry 2007).

The simplest form of genetic variation that occurs among individuals is a single nucleotide polymorphism or SNP, that is, the substitution of one single nucleotide for another. SNPs occur at an estimated frequency of 1 in 1,000 bp throughout the genome. These variations can have an array of effects on the genome. Around 50% of SNPs occur in noncoding regions, 25% lead to a

missense mutation (alter encoded amino acids), and 25% are silent mutations (do not alter) encoded amino acids). While nonsynonymous SNPs (ones that change-encoded amino acids) may produce pathology, both nonsynonymous and synonymous SNPs influence promoter activity and pre-mRNA conformation and stability, thus having the potential to affect protein function, disease susceptibility, and genome evolution, and consequently may be under forces of natural selection.

Due to the relationships between diseases, heredity, and genetic polymorphism, a principal goal of the human genome project was to facilitate the application of genomic study to the treatment and understanding of disease. According to the Journal of Human Genetics, single nucleotide polymorphism (SNP) technologies can serve as important tools in identifying an array of disease-causing genes in humans (Shastry 2007). Possible applications of SNP-technology is using findings to implement individualized genome-based diets and medicines that are safer and more effective. Interestingly, analyzing single nucleotide polymorphisms in the genome can elucidate evolutionary histories and pathways. Throughout a given gene, the rate, type, and site of nucleotide substitutions and selection pressure on codons is not uniform. Oftentimes, residues that evolve under strong selective pressures are found to be strongly associated with human disease. When deleterious mutations arise affecting the biological function of a protein, these mutations over-time can be effectively rejected from the gene pool. When these substituted nucleotides are fixed during evolution they may provide selection advantages, be neutral, or cause pathology. As such, disease-associated SNPs and evolution are often highly related. Genetic variants have been proposed to not only be responsible for modulating disease risk and individual variance, but also molecular evolution. Genetic evolution is reliant upon a balance between natural selection and environmentally-driven mutation.

d. Immunogenetic Histories

Latin American and Eurasian populations have distinct immunogenetic histories. Paleomicrobiological evidence suggests that several pathogens emerged in conjunction with shifts of agriculturalization and urbanization. These high mortality rates indicate that immune adaptation is a complex process (Collen et al 2022). Zoonotic pathogens tend to be increasingly likely to be the cause for emerging diseases among humans. Many of these zoonotic pathogens ultimately proved the most detrimental to Indigenous populations: Measles, mumps, smallpox, diphtheria, and influenza. In Europe, there is believed to have been time for adaptive immunity to develop against a variety of these deadly pathogens due to the history of the Silk Road and population movement across Eurasia. In the Americas, however, the domestication of animals evolved distinctly from Eurasia and involved different animals. Camelids and guinea pigs, the animal species domesticated first, are not associated with any known disease-causing pathogens. Moreover, indigenous American urban areas were characterized by more advanced and effective systems of water storage and distribution which likely improved sanitation and limited microbial spread. Despite having a longer time to adapt to a variety of pathogens, some diseases -- particularly smallpox and measles -- continued to infect the Eurasian population with high rates of fatalities up until the development of vaccines (Collen et al 2022). Because the development of immunity is not always associated with pathogenic exposure and due to significant under-sampling of Latin American subjects, it is very difficult to fully assess the extent to which different evolutionary histories and different pathogenic histories contributed to the vast depopulation of Indigenous Americans at the hands of infectious diseases.

Understanding pathogen-related events of depopulation is crucial given that infectious diseases are one of the most significant driving forces of selection on the human genome. Although

some anthropologists have proposed an innate susceptibility of Indigenous populations to pathogenic diseases, many argue that disease, along with crucial sociological factors are what drove depopulation. European contact involved the degradation of Indigenous ways of life, ultimately bringing about poor sanitation, birth rate decline, loss of infrastructure, wars, famine, and more. Incorporating these aspects into the explanation for the loss of Indigenous life and livelihood is crucial.

e. Similar Studies

The study of genetic predisposition to disease in Latin American populations has generally involved the analysis of old world populations, with only recent publications studying admixed populations. Vidal and colleagues used ANNOVAR to identify SNPs of potential functional impact among eleven Mapuche-Huilli individuals from Chile. Using ANNOVAR and filtering genetic variants through the CADD and FATHMM-MLK databases allowed for certain variants to be labeled as "Variants with Potential Functional Impact" or VPFIs. With the use of WebGestalt tools, variants predicted to be deleterious were assessed for diseases already identified as overrepresented. Notably, various IPFIs corresponded to and were able to explain various health outcomes, including predispositions to gallbladder cancer and neural tube defects in the Mapuche population. Additionally, seven VPFIs were found amongst five hypercholesterolemia disease-associated genes. The authors indicate that future studies looking more broadly into variants contributing to hypercholesterolemia across Native American and admixed latinx populations would be interesting (Vidal et al 2019).

Kerner and colleagues highlight how admixture is able to foster adaptation to various pathogens; in fact, interbreeding between previously genetically isolated human populations in history have greatly shaped modern-day immune responses (Kerner et al 2021). Kerner and

colleagues begin their paper by acknowledging that immunity is a site of incredible biological relevance due to its frequent subjection to natural selection, an understanding which has driven various population genetic studies (Kerner et al 2021). Our evolutionary histories continue to be of significant relevance to modern immunological issues. Even with the COVID-19 pandemic, a risk haplotype that increases odds of hospitalization by 60% is of Neanderthal origin. Another haplotype of Neanderthal origin, however, has shown to reduce risk of developing severe COVID by 22%. They also point to admixture as a significant explanation for various immune outcomes, specifically identifying two main periods of human genetic admixture. Firstly, the late pleistocene, when populations of hunter-gatherers dispersed around the globe. During this spread, they encountered archaic hominins and diversified. Secondly, the holocene, a more recent period, characterized by large-scale population movements facilitated by various technological and demographic changes. According to ancient DNA studies, admixture has proven to have occurred from around 7500 years onward in Western Europe, for example. The modern-day genetic diversity of humans is, in large part, due to admixture events. Interestingly, adaptive admixture has been proposed. That is, the acquisition of adaptive traits through intraspecial admixture. Notably, intraspecies admixture has been associated with pathogen resistance. Ultimately, ancient DNA techniques have opened up new avenues for immune understanding and uncovering histories of pathogen exposure (Kerner et al 2021).

Gene set enrichment analysis (GSEA) has previously been conducted in order to identify specific pathways and phenotypes particularly affected by post-admixture selection occurring after contact. From there, in order to assess the driving factor behind the selection for these health-related ancestry-enriched SNPs, investigation into the functional annotation of the SNPs is required. Norris and Colleagues specifically compared findings across four Latin American

populations -- Columbia (n=94), Mexico (n=64), Puerto Rico (n=104), and Peru (n=85) -- in order to shed light on the ways in which post-admixture selection differs amidst distinct social and physical environments. Norris and colleagues found that across all four Latin American populations studied, there were ancestry-enriched SNPs associated with cytokine receptor interaction, T cell receptor signaling, and antigen processing and presentation pathways (Norris et al).

In 2018, Lindo and colleagues examined the genetic prehistory of the Peruvian Andean Highlands 7000 BP. Due to the 2500m altitude in this environment, inhabitants underwent a series of complex cultural, biological, and genetic adaptations (Lindo et al 2018). Notably, the Andean Highlands have long been considered an ideal location for studying human genetic adaptations despite the relative lack of understanding of the genetic background of this population. The genetic history of a group already facing an array of selective pressures becomes only more interesting upon European contact, which brought on an additional set of economic, social, and pathogenic changes. The study tested first for high allele frequency divergence due to the high altitude environment by contrasting ancient highlands populations with ancient lowland populations. Then, the study targeted selection hypotheses related to adapting to European-introduced pathogens. The Population Branch Statistic (PBS) was used to do this; PBS highlights alleles that appeared at high rates in one population relative to others. When comparing the ancient Rio Uncallane highlands population (pre-contact) with the Aymara, a modern Andes population (post-contact), the strongest selection signal originated from SNVs in the vicinity of *CD83*, a gene involved in various immune pathways, including T cell receptor signaling (Lindo et al 2018). Specifically, *CD83* is involved in encoding an immunoglobulin receptor and has been shown to be upregulated in response to vaccinia virus infection (virus used for the smallpox vaccine). Moreover, the

differentiated SNPs associated with the CD83 signal show a variety of chromatin alterations in monocytes, T cells, and natural killer cells -- types of immune cells (Lindo et al 2018). This is suggestive that the genetic polymorphisms are functional. Additionally, another gene in the top 0.01% of the scan, IL-36R, codes for an interleukin receptor that has been associated with the skin inflammatory pathway and vaccinia infection. The second most highly differentiated SNPs were in the vicinity of RPS29, which codes for a ribosomal protein and is involved in viral mRNA translation and metabolism, including those of influenza. The strength of these signals and roles of associated genes may suggest that selective pressures brought about by contact favored alleles that directly affect the pathogenicity of the diseases encountered by the ancestors of the epidemic survivors (Lindo et al 2018).

f. Genomics as a Clinical Intervention and Biological Implications

With the rise of whole-genome sequencing technology availability and accessibility, WGS has the potential to be used as a valuable clinical tool when assessing patient susceptibility to diseases, family history, and ultimately even aid in the diagnosis of rare genomic disorders (Hayeems et al 2020). In terms of therapeutic efficacy, WGS was determined to be highly effective. Specifically, WGS and singleton whole-exome sequencing (ES) were found to be significantly more effective than microarray testing, with a p-value less than 0.0001 when looking at changes in the clinical status follow genomic intervention (Hayeems et al 2020). Microarray testing is typically used for the genetic testing of developmental and learning disabilities, autism spectrum disorder, and congenital abnormalities and malformations (Miller et al 2010).

Moreover, population genomics has the ability to revolutionize public health understanding and clinical methods in regards to both communicable and non-communicable diseases.

Integrating genomic methods into public health research will allow for a broader understanding of how evolutionary and biological factors interact with social, environmental, and behavioral determinants of health. Whole-genome sequencing is becoming the standard method of testing for characterizing infectious disease as well as guiding public health interventions. Moreover, WGS can shed light on vaccine antigens and antimicrobial resistance. Functional genomic characterization of host response can direct the identification of treatments more suitable for specific population groups or individuals. Although this comes with some ethical concerns, WGS also has the potential to drive more effective clinical treatment and implementation of targeted epidemiological preventative measures (Khoury and Holt 2021).

g. Limitations

Working with ancient genomes, there are inherent limitations due to the quality of the samples. Over time, DNA samples not kept in optimal storing conditions will shorten as well as develop residues of deaminated cytosines (Dabney et al 2013). While the DNA of living cells is maintained relatively stable through various enzymatic and homeostatic processes, these processes cease once an individual dies. Thus, when sampling and analyzing ancient individuals, researchers often are faced with reduced DNA fragment size, lesions blocking replication (needed for various analysis techniques), and lesions causing incorrect nucleotides to be incorporated during replication (Dabney et al 2013). Genomic data with missing nucleotides hinders the identification of SNVs of significance as well as the assessment of individual and population-level ancestral components and traits.

Moreover, gaps in genomic methods and technologies persist -- in spite of the significant developments made in recent years. Mathieson investigated selection at the FADS locus in Native

American populations. Looking at present-day Inuit and Siberian populations, they are observed to have strong selection at the ancestral haplotype, hypothesized to be an adaptation to PUFA-rich arctic diet (Mathieson 2020). However, similar signals have been detected in Native American populations, thereby suggesting there was early selection on the ancestral haplotype in the Indigenous American lineage. The population branch statistic (PBS) was used to compare Native American (NA), European (EUR), and East Asian (EAS) populations. Through assessing genetic differentiation between the populations, the test is able to identify if any of the branches have "excess" differentiation, thereby indicating selection. However, this test assumes that the three branches are independent from each other. If, for example populations B and C have undergone parallel selection, PBS will misattribute selection to A due to the high divergence of the other two branches from it. In this case, there is a much lower frequency of the derived allele in Upper Paleolithic Eurasia -- the group ancestral to Indigenous North and South American populations -- compared to European and East Asian groups, we would expect PBS(NA,(EUR,EAS)) to find signals of selection in the NA population. Ultimately, these findings show the difficulty of assessing selection at locusts which have been the site of complex and confounding processes of selection (Mathieson 2020). Ultimately these limitations demonstrate the importance of carefully using the PBS as a genomic tool and the value of incorporating ancient DNA evidence to contextualize and corroborate modern-day findings.

Methods

a. Data Processing

During the process of variant calling, sample genomes, once aligned, are compared to reference genomes as means of highlighting locations of genetic differentiation across the whole genome. Generally, variant calling uncovers millions of single nucleotide variants or SNVs per

individual human sample (Zverinova and Guryev 2022). For variant calling, the hg19 reference genome was used alongside bcftools commands *mpileup* and *call* in order to call genotypes.

Working with ancient DNA (aDNA) samples, DNA degradation becomes a primary concern for genomic researchers. After death, microbial colonization and DNA degradation via deamination and fragmentation cause the vast majority of aDNA samples to have low depth of coverage (Sousa da Mota et al 2023). This hinders genotype calling, as bases are read with reduced confidence. As such, DNA imputation can be used to fill in the gaps of low coverage ancient genomes. Imputation is the process of inferring missing sites using reference haplotype data. From there genotypes can be genotype uncertainty of ancient samples are captured by likelihoods. The growing availability of reference data from the 1000 Genomes Project, Haplotype Reference Consortium (HRC), and TOPMed have facilitated the improvement of imputation methods (Martiniano et al; Haber et al; Saupe et al; Clemente et al; Cox et al; Allentoft et al). Although there are skeptics regarding the efficacy of imputation, Sousa da Mota and colleagues conducted a study using 43 ancient samples to determine if imputation methods are effective. They found that imputation accuracy is similar between ancient samples and present-day imputed (high-coverage) genomes. Notably, accuracy is higher at more common variants with rarer variants being harder to impute and for higher coverage samples. (Sousa da Mota et al 2023) Although imputing Indigenous samples with varying admixture patterns could be a concern for this project, Sousa da Mota and colleagues find the imputation process to be accurate for indigenous Americans populations across low (e.g. Puerto Rican) and high (e.g. Peruvian) admixture populations (Sousa da Mota et al 2023).

b. Demographic analyses

For this project, demographic continuity was assessed using SNPRelate (Zheng et al 2012), SCOPE (Chiu et al., 2022), and TreeMix (Pickrell and Pritchard 2012). These analyses were done in order to confirm population continuity among the ancient and modern samples. In order to make accurate inferences regarding selection, the modern population must be descended from the ancient groups.

SNPRelate is a program designed to analyze relatedness based on principal component analysis (PCA) by calculating sample and SNP eigenvectors. PCA is a test done to analyze SNP data and identify outliers, ultimately aiding in the determination of population structure (Abraham and Inouye 2014). Next, ancestral components are identified. SCOPE (scalable population structure inference) is a program which further complements methods of determining population structure. In approaching population structure, typically admixture proportions of individuals and ancestral allele frequencies are determined given genetic variance data (Chiu et al 2022). Relying upon a pre-existing proposed likelihood-free framework (estimating the individual allele frequency matrix via latent subspace estimation) and incorporating randomized eigendecomposition and avoiding matrix formation, SCOPE offers an accurate and efficient (1800x faster than comparable methods) means of inferring ancestry and admixture. TreeMix is a tool which describes the often complex demographic history of a population which can involve gene flow, population size changes, and population splits using a maximum likelihood framework(Pickrell and Pritchard). Although this model is accurate, it does make several assumptions which can cause some limitations. It does depict migration events to occur at instantaneous time points which is not the case and therefore is better suited for when gene flow occurs over shorter periods of time as opposed to populations who underwent continuous migration (Pickrell and Pritchard). Ultimately, a maximum-likelihood tree graph is produced using genome-

wide allele frequency data to show genetic differentiation of various populations with sequential migrations shown.

c. Genome-wide selection scan

The population branch statistic, PBS, is a method of assessing selection based on the level of divergence between three different branches. More divergent branches are assumed to have undergone a greater level of selection in order to become more differentiated from the other populations (Jiang and Assis 2020). PBS relies upon the use of the F_{ST} statistic as a measure of genetic differentiation. F_{ST} quantitatively describes the differences in allele frequencies of populations at specific sites (Yi et al 2010). By including a distantly related group, three Fst values are able to be compared to each other. The comparison of three Fst values allows for a better estimate of the strength of divergence that has occurred between the two closely-related populations since they diverged from the outgroup, which, in the case of this study, is the Yoruba. As such, the population branch statistic is better able to account for more recent events of selection while also providing more insight into when and in what population selection for a specific variant occurred. PBS was conducted between pre- and post- admixture Mexican populations utilizing ancient individuals previously published by Villa-Islas and colleagues along with and modern-day Simon's Genome Diversity Project (SGDP) individuals to assess the level of selective pressure brought on by colonization in Latin America, with the Yoruba used as an outgroup (*Supplementary Table S1*.). Investigating patterns of selection in admixed populations is made difficult due to how independent signals from source populations may be obscured by each other (Yelmen et al; Huerta-Sánchez et al. 2014; Galaverni et al. 2017; Pierron et al. 2018). Moreover, it can be difficult to determine if the signal originated pre or post admixture (Yelmen et al). For pre-admixture selection

signals, it is then a task to determine which source population was subject to the selective pressure. The population branch statistic is a method of assessing selection based on the level of divergence between three different branches. More divergent branches are assumed to have undergone a greater level of selection in order to become more differentiated from the other populations (Jiang and Assis 2020).

Results

a. Demographic analyses

In order to compare the genomes of pre and post colonization Mexican individuals, demographic analyses needed to be conducted in order to confirm population continuity. Especially given the complex demographic history of the region, it is key to ensure that the modern Mexicans do contain Indigenous ancestry with continuity to the genetic makeup of the ancient Mexicans.

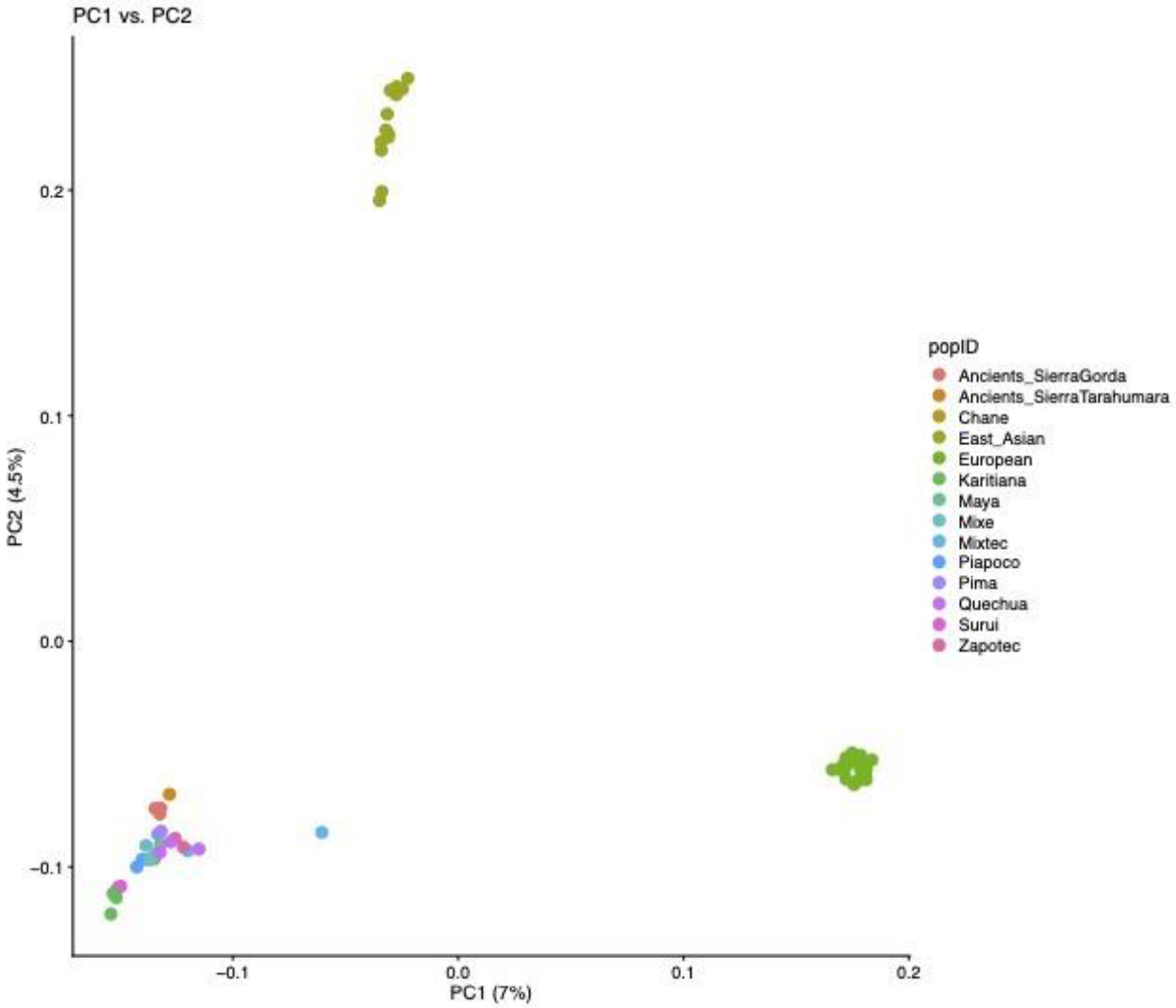
For the PCA, this included the ancient samples from Sierra Tarahumara (n=1) and Sierra Gorda (n=3), modern day Indigenous Mexican (Mixe, Mixtec, Zapotec, Pima, and Maya) and Latin American samples (Karitiana, Piapoco, Quechua, Surui, and Chane) from SGDP, as well as East Asian and European samples from SGDP. Previous demographic analyses showed population-specific clusterings on the PCA when run for different indigenous groups across Mexico -- across Mesoamerica and Aridoamerica -- thereby suggesting that geographic origin shaped the genetic composition of these individuals (García-Ortiz et al 2021). Running the PCA with Latin American indigenous groups, Europeans, and East Asians was expected to result in the formation of three general clusters where individuals from those respective groups would reside; as such, some of the genetic differentiation between Mexican subpopulations and other groups

would likely be concealed by this PCA. Yet, it was still expected that some population-specific clustering would occur even within the larger clusters of Latin Americans. From the results of the PCA, we see ancient and modern American individuals largely clustering in the bottom left corner, with East Asian individuals clustered in the top central and European individuals clustered in the bottom left. Ancient individuals appear to cluster together at the top of the Latin America cluster and are generally closer to modern Mexican populations than the South American groups, which are largely located towards the bottom of the cluster. These results are largely in-line with our expectations and ultimately suggest relative genetic continuity and similarity across Indigenous American groups -- both ancient and modern -- as well as differentiation from European and East Asian groups.

For the model-based clustering analysis (Chiu et al 2022), it appears to show relative genetic continuity across the ancient and modern Mexican individuals, as well as with other Latin American indigenous groups as evidenced by similar coloring patterns. In the SCOPE cluster analysis, we see European ancestry represented in blue and East Asian in green. Indigenous Latin American ancestry is represented by red. This model-based clustering analysis was run for K=2 to K=9, with K referring to the number of ancestral population groups. Given the demographic history of the region and samples included in this demographic analysis, it was hypothesized that K=3 or K=4 would be the best-fit model. However, the K=2 and K=3 models are found to be the best fit, with K=2 having a slightly lower k value, thereby indicating a better fit. This discrepancy in expected outcome and result could be explained by the low sample size for the ancient Mexican populations and limitations of the method. Still, the results of the model-based clustering analysis are sufficient to proceed with selection scans, as there is evidence of shared ancestry and genetic

continuity among the ancient and modern Mexican groups. As such, reasonable inferences can be made regarding the selective pressures these groups have undergone.

Lastly, a maximum likelihood tree was generated to model the population histories of the populations. The maximum likelihood was calculated for a single migration event (m=1), which shows Europeans migrating to the Mixtec population. However, likely due in part to the aforementioned issues with some of the ancient and modern samples, the results are somewhat compromised. For example, we see the Mexican Ancestry in LA population branching off from the European population, which obviously is not accurately representing the population history. Instead, we should see a European migration to the Mexican branch between the ancients and Moderns, with the Mexican Ancestry in LA individuals stemming off from a subsequent branch along with groups like the Zapotec, Mixtec, Pima, and Mixe. Moreover, the ancients from the Sierra Gorda site are located quite far from the other ancient groups on the tree. We would expect them to show a closer relationship to the other ancient populations compared to modern groups. However, the maximum likelihood tree does appear to assess genetic relatedness among some of the modern indigenous populations better, with the South American groups (Surui, Karitiana, Piapoco, and Quechua) branching off together and also showing close relationships between the Maya and Mixe and Zapotec and Mixtec. We attempted to create another maximum likelihood tree with the Mbuti people as an outgroup, but ran into computational issues with the samples and were unable to complete it in time.


PC1 vs. PC2
PC2 (4.5%)
PC1 (7%)
0.2
0.1
0.0
−0.1
−0.1
0.0
0.1
0.2
popID
Ancients_SierraGorda
Ancients_SierraTarahumara
Chane
East_Asian
European
Karitiana
Maya
Mixe
Mixtec
Piapoco
Pima
Quechua
Surui
Zapotec

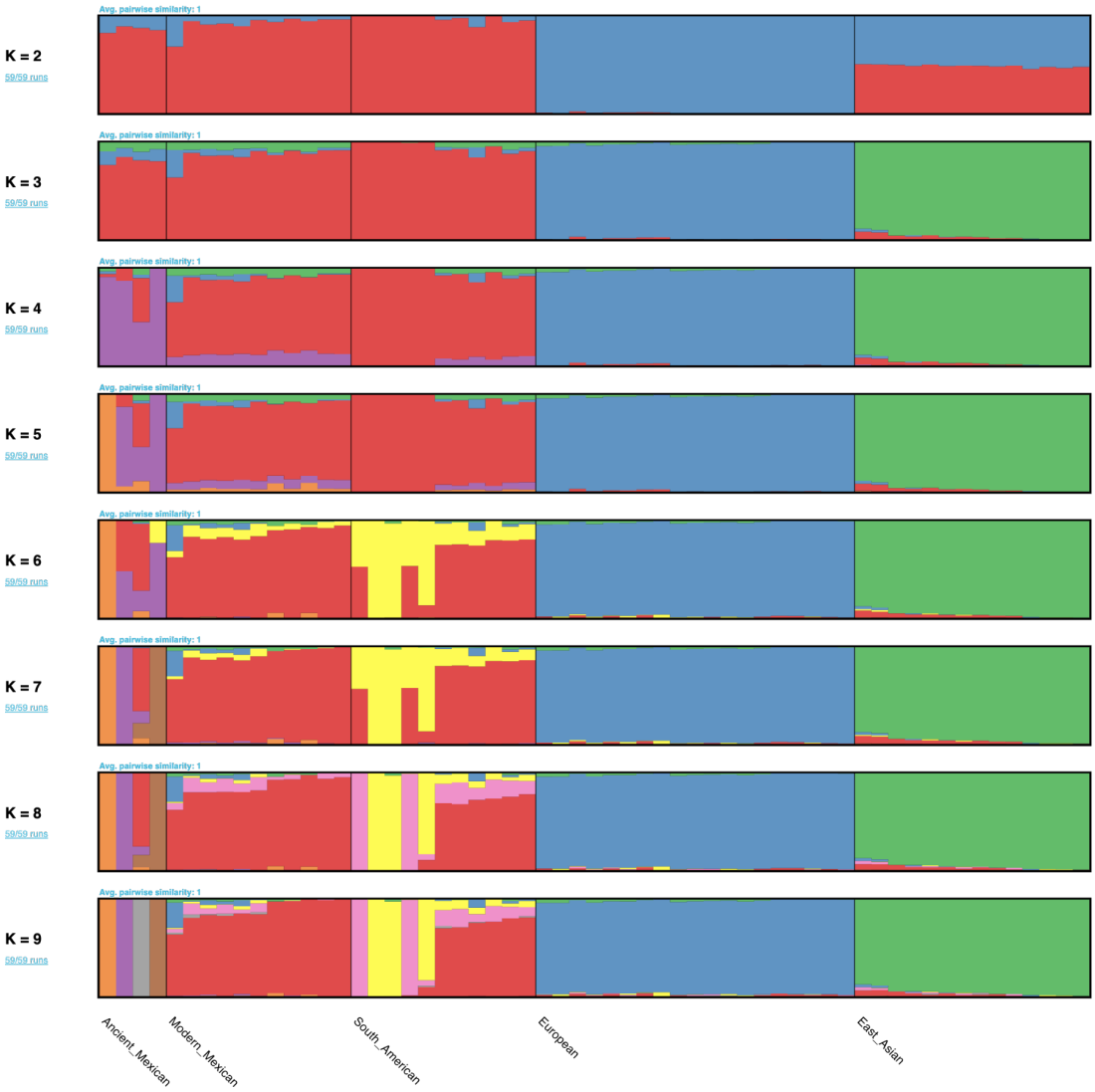


Figure 4:

a) Principal components analysis of samples

b) Model-based clustering analysis (SCOPE) visually demonstrative the ancestral makeup of samples

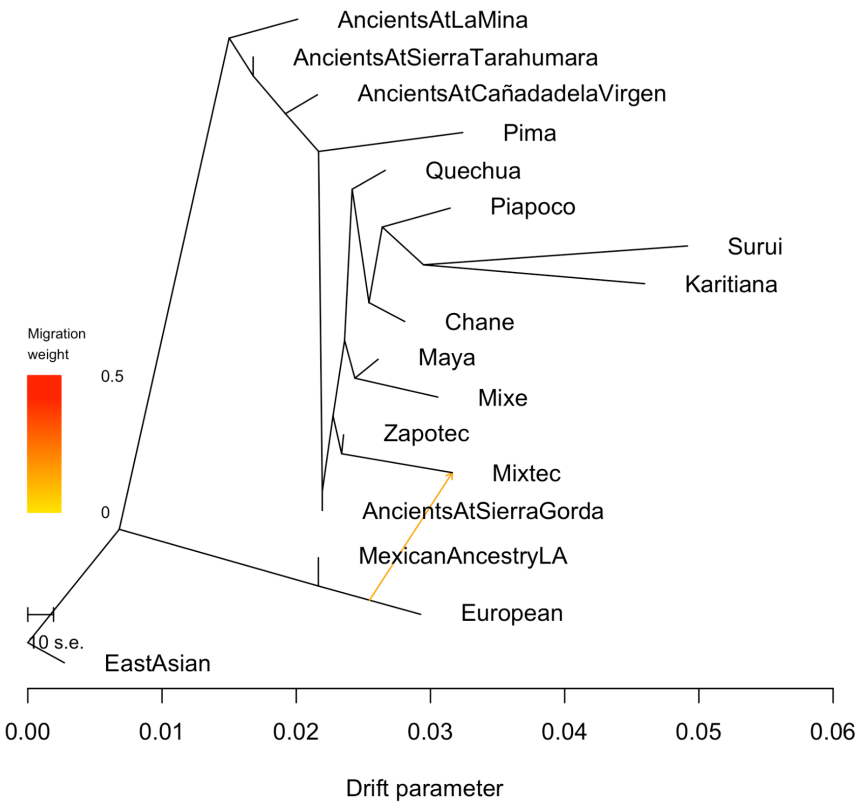


Figure 5: Maximum likelihood tree generated by TreeMix (m=1) showing ancestry relationships between sampled populations.

b. Selection scan

Immune System Selection

The population branch statistic branch conducted on the Ancient Mexican (n=4), Modern Mexican (n=11), and Yoruba populations (n=20) demonstrated evidence of selection on an array of single nucleotide variants.I investigated the top 0.1% of variants putatively under selection with the highest PBS scores. Of these, 17 variants were associated with genes involved in some aspect of the acute immune response, which would be implicated by the pathogenic effects of European colonization. Notably, many chronic diseases are associated with prolonged periods of excessive activation of the adaptive immune response and inflammation. Many SNVs were in gene regions related to chronic immunity, inflammation, and oncogenic pathways, which, although immune-related, are more likely evidence of selection for increased immunity to cancer and other chronic diseases overtime. Instead, particular attention was paid to SNVs involved in the innate immune system. The innate immune system is responsible for the body's initial defense against new pathogens as well as subsequent initiation of the adaptive immune response (Alberts et al 2002). In the case of this study, we would expect relevant selection to primarily occur on the acute (innate and adaptive) immune responses to favor individuals more adapted to a world increasingly defined by gene flow and the spread of pathogens. Any selection related to chronic inflammation and immune responses with regards to chronic diseases are unlikely to be related to European colonization of the Americas.

Several SNV's were also hits in the Genotype-Tissue Expression (GTEx) portal, indicating they control the expression of other genes. The Genotype-Tissue Expression (GTEx) portal is an initiative to further study the functional impact of genetic variants, as many variants uncovered by genome-wide association studies have not been studied for their impacts on health and disease (Lonsdale et al 2013). For example, rs12459654 on chromosome 19 was a hit on the gTEX portal

for the genes *ADGRE2*, *ORC71*, and *ZNF333.* As such, we can conclusively determine that this variant has a functional impact on the gene functions and phenotypes related to these genes.

The variant under the strongest selective pressure according to its PBS score of 2.95 was rs367721985 on chromosome 20. This is an intergenic SNV between the *NFATC* and *ATP9A* genes. Notably, the *NFATC* gene produces a transcription factor involved in immune pathways. *NFATC2* is an activator of curdlan-mediated genes. In addition to *NFATC2* targeting genes involving several dendritic cell-derived cytokines and chemokines, *NFATC2* is a key component of the epigenetic modification involved in recruiting the H3K4me3 mark (Yu et al). This epigenetic marker is associated with heightened activation of curdlan-mediated dendritic cells, thereby implicating the role of *NFATC*.

Of immune related variants, there were several under strong selective pressure. rs202199280, an intergenic variant between the *TRIML1* and *LINC01060* genes, is involved in the innate immune system. Although the *TRIM* family of proteins are largely understudied, their involvement in the immune system is widely accepted. Specifically, researchers have found that *TRIM* proteins are involved in host defense and retrovirus restriction and hypothesize that they facilitate T cell proliferation, as *TRIML1* is associated with increased *IL-2* expression (Ozato et al). More studies will have to be done to isolate *TRIML1's* role in immune processes, but it is known that the *TRIM* family of proteins are all activated in response to interferons. Interferons are released by infected host cells of viral and bacterial pathogens and trigger the innate immune response (Mertowska et al), thus implicating the *TRIM* proteins in immune-related pathways. Also located on chromosome 4, variant rs75096357 is between the *ANKRD50* and *FAT4* genes. *FAT4*, FAT atypical cadherin 4, has been found to have a significant and positive relationship with a variety of innate immune response cells, including CD4+ T cells, neutrophils, macrophages, and

dendritic cells (Li et al 2023). Ultimately, *FAT4* plays a significant role in host defense against pathogens.

A SNP, rs12459654, was discovered on chromosome 19 between the *ADGRE2* and *OR7C1* genes. gTEx eQTL confirms that this particular variant is associated with altered gene functions for both neighboring genes, along with the *ZNF333* gene. The *ADGRE2* gene is responsible for coding the adhesion G protein-coupled receptor E2 protein, which is implicated in the production of various immune cells, particularly in CD16+ blood monocytes and BDCA-3+ myeloid dendritic cells. CD16+ are one of the major types of blood monocytes and are associated with pro-inflammatory pathways (Ziegler-Heitbrock 2007). Randolph and colleagues performed a study which showed that CD16+ monocytes preferentially migrate and develop into dendritic cells with superior allostimulatory capacity, which refers to the potential for accurate antigen reception. Ultimately, it is suggested that CD16+ monocytes are able to produce dendritic cells with improved antigen-presenting capacity. Antigen-presenting cells such as dendritic cells are responsible for informing the adaptive immune response of the presence of an infection. Interestingly, there is an increase in CD16+ concentration particularly in the acute infection phase, but also for weeks following infection. This phenomenon of CD16+ increase is observed for a variety of inflammatory and infectious diseases such as tuberculosis and HIV.

At chromosome 12, rs2012290 was found to be associated with the *A2MP1* gene via qTEX eQTL. *A2MP1* is an lncRNA produced by the pseudogene alpha-2-macroglobulin pseudogene 1. Ren and colleagues found differential expression of 103 differentially expressed (up- or down-regulated) lncRNAs and mRNAs in bicep tendon inflammation. They isolated the four most differentiated RNAs, which included *A2MP1*, which was upregulated. Based on GO functional analysis, it was concluded that *A2MP1* and other related lncRNAs and mRNAs were involved in

the regulation of the innate immune response, specifically determining that *A2MP1* targets *NFKB2* (Ren et al). The *NFKB* family of transcription factors is involved in a variety of immune signaling pathways, but most notably for our study is its role in both the adaptive and innate immune response. *NFKB* signaling pathways are activated in response to pathogen exposure and produce pro-inflammatory cytokines. Moreover, *NFKB* dysfunction is associated with defects in B and T cell proliferation and activation (Li and Verma 2002). SNV rs138889705 on chromosome 16 is shown to have an effect on *LPCAT2* function according to gTEx eQTl findings. *LPCAT2*, lysophosphatidylcholine acetyltransferase 2, is part of the *LPCAT* family of enzymes and is specifically associated with the inflammatory response in response to bacterial ligands such as LPS. Abate and colleagues conducted a study to investigate the exact role *LPCAT2* and related enzymes play in the immune response. si-RNA knockdown of *LPCAT2* is necessary for macrophage cytokine gene expression as well as their response to binding *TLR4* and *TLR2*, which are bacterial, fungal, parasital, and viral ligands (Abate et al 2020). Cells overexpressing *LPCAT2* are found to have a heightened immune and inflammatory response. The researchers conclude that *LPCAT2* plays a major role in the regulation of the inflammatory response in response to bacterial TLR ligands. Notably, part of the *LPCAT2* mechanism involves activation of *NFKB* and *MAPK* pathways, which release inflammatory cytokines. Unregulated, excessive release of cytokines is associated with several deleterious conditions, including sepsis. Understanding the regulation of cytokine release and inflammatory pathways activated by TLRs is crucial (Abate et al 2020). Notably, the binding of TLR4 to LPS activates the NFKB pathway (Shih et al 2018). Evidently, both *A2MP1* and *LPCAT2* are relevant to the immune response via implication in *NFKB* pathways; is a lnRNA implicated in key immune responses which would be relevant to the pathogenic exposure of Mexicans with the onset of Spanish colonialism.

An intergenic variant, rs6110964, was discovered to be under strong selective pressure . This variant is located between the *MACROD2* and *KIF16B* genes. Gelemanović and colleagues conducted a genome-wide meta-analysis of variants which can predict susceptibility and outcomes of individuals to human infectious diseases. They identify *MACROD2* as a gene involved in the *NFKB* pathway and one of relevance when assessing susceptibility to infectious disease (Gelemanović et al 2023; Khong et al 2016). Specifically, increasing *MACROD2* expression is associated with increased *NFKB* stimulation, which facilitates an increase in cell survival and an improved immune response.

Two intronic variants were discovered on the *EVT6* gene on chromosome 12, rs10676577 and rs200633262. *EVT6* is a transcription factor which has been found to be involved in optimizing dendritic cell gene expression. Dendritic cells, through the use of pattern-recognition receptors such as TLRs, are able to recognize pathogens. In this way, DCs unite the innate and adaptive immune responses and facilitate antigen-specific T-cell responses (Lau et al 2018). *ETV6* is found to be involved in DC development and differentiation in vitro, as well as be responsible for T cell cross-priming and antigen specificity. Another intronic variant, rs9747624, on chromosome 17 is located in the *Abr* gene. *Abr*, ABR Activator Of RhoGEF And GTPase, is known to be involved in the regulation of the innate immune system. Specifically, *Abr* is involved in the regulation of Rac and macrophage activity (Cho et al 2007). *Abr* is often implicated in septic pathways as a negative inhibitor of excess inflammation. Future study into the potential role of *Abr* in infectious diseases and regulating the immune response to pathogens could prove quite interesting.

Lastly, two intronic variants of immune relevance were discovered to be under strong selective pressures, with PBS scores of 2.44. Variant rs1352846 is located in the GC gene on chromosome 4. GC, the GC vitamin D binding protein, is involved in the immune-system due to

vitamin D receptors being found on the surface of a variety of immune cells, including B cells, T cells, and antigen-presenting cells (Aranow 2011). Interestingly, vitamin D is part of a proposed regulatory system where it is able to act in autocrine fashion on the immune response. Moreover, vitamin D deficiency is associated with increased susceptibility to disease. In the past, vitamin D was even used to treat tuberculosis prior to the development of effective antibiotics. The other variant, rs114343309, is located in the *CARMIL1* gene on chromosome 6. Stark and colleagues find that in certain models, *CARMIL1* (and *CARMIL2*) are involved in stimulating the *NFKB* pathway (Stark et al 2017). However, they acknowledge contrary findings in studies on mice. Either way, *CARMIL1* is implicated in innate immune processes, but there should be future research to clearly elucidate the role it plays in shaping immune responses.

Health-Related Pathways

Aside from the acute immune-system related selective pressures, I noticed several categories of variants whose associated genes were tied to various health outcomes. I divided those into three general categories: Physiological diseases (diabetes, blood pressure, obesity, cardiovascular disease), cancer, and chronic inflammatory (neurodegenerative) disease.

A variant on chromosome 15 at position 84238192, rs1896802, had a gTEX eQTL hit with a variety of genes: *RP11-182J1.14*, *GOLGA6L4*, *FAM103A1*, *RP11-90B9.2*, and *GOLGA6L4*. Many of these genes are implicated in cancers. The *RP11* family of long non-coding RNAs are associated with colorectal cancer (Chen et al 2021). Some models argue for the utility of *FAM103A1* as a prognostic marker for breast cancer, as related upstream miRNAs are associated with immune infiltration (Zhang et al 2022). Moreover, variants rs12833182 on chromosome 12 and rs13128039 on chromosome 4 also had a gTEX eQTL hit for *RP11* non-coding RNAs. Two

variants on chromosome 8, rs72720809 and rs60845745, both had a functional impact on *RNU1-106P*. Notably, mutations in U1 spliceosomal small nuclear RNAs are present in <.001% of all cancer cases but around 97% of adult sonic hedgehog medulloblastoma (Suzuki et al 2019). As such, these variants have the potential to be strong predictors of Shh-MB as well as targets for effective therapeutics. *SCRG1*, a gene implicated by variant rs13128039, was shown to be the only gene out of 9 studied to be an independent prognostic indicator of resectable stomach adenocarcinoma and breast cancer survival, as it was found to be associated with metastatic propensity and, consequently, significantly contribute to poorer patient outcomes (Liu et al 2022) Variant rs12833182 was found to have a functional consequence on *AMDHD1*, a gene with known variants contributing to breast cancer outcomes and risk (Wang et al 2020).

For variants of genes associated with chronic inflammatory diseases, there was only one that corresponded to a gTEX eQTL hit: rs12496388 on chromosome 3. This variant is associated with a functional consequence on *CGGBP1*. Interestingly, *CGGBP1* has been associated with differential expression in ALS (amyotrophic lateral sclerosis), a common chronic inflammatory neurodegenerative disease; specifically *CGGBP1* was seen in elevated levels in both motor neurons and skeletal muscle of ALS patients (Mamoor 2022).

Variant rs2125801 on chromosome 4 has a known functional consequence on *GUCY1B1*, a gene which codes for soluble guanylyl cyclase. Guanylyl cyclase facilitates the relaxation of smooth muscle cells through the conversion of GTP to cGMP; studies have shown that mutations leading to impaired function of soluble guanylyl cyclase are associated with increased risk of hypertension (Curtis 2024). The *PDE1C* gene is affected by variant rs11773452 on chromosome 7. Notably, *PDE1C* deficiency is strongly associated with incidence of abdominal aortic aneurysm (AAA) along with aortic dilation and elastin degradation, ultimately suggesting that targeting

PDE1C has the potential for therapeutic benefit (Zhang et al 2021). Variant rs375303324 on chromosome 12 is associated with functional consequences on two genes, *RFC5* and *WSB2*. *RFC5* has been associated with obesity and diabetes, and it is believed that epigenetic modifiers of this gene region are associated with shaping health outcomes (Remely et al).

This analysis of genetic variants under selection contributing to health outcomes is limited only to variants under the highest selective pressure with known functional genetic consequences. Much more remains to be studied regarding population health and how genetic variants, particularly those common in underrepresented genomic groups, can be utilized to inform health interventions and future therapeutics.

Discussion

The aim of this study was to contribute to understandings of Mexican demographic history and processes of genetic selection as a result of the transition from pre-colonial to post-colonial and show the utility of ancient DNA for enhancing our understanding of these issues. Notably, several variants associated in the *NFKB* pathway, including *ADGRE2*, *A2MP1*, *LPCAT2*, *MACROD2*, and *CARMIL1* along with several variants that are hypothesized to influence the production of immune cells or involved in the regulation of inflammatory processes. This is largely consistent with expected results. Due to the sheer magnitude of the loss of life inflicted on Indigenous Latin American populations as a result of novel pathogens being brought over by European colonizers, selective pressure would be expected on the innate and adaptive immune systems (Collen et al 2022). That being said, the results were not entirely straightforward. The genetic impact of several variants were largely unknown. Moreover, issues in genome imputations compromised the results of the demographic analyses. Throughout this study, we were reminded

that studies of Latin American genetics are faced by several obstacles. These include but are not limited to the recency of genomic sequencing and analysis technologies designed for admixed populations, knowledge gap of many variants of significance (particularly those affecting Latin American populations), and a lack of sample availability. This was expected, as, according to Collen, only a handful of studies have attempted to identify post admixture immunogenetic signals incorporating aDNA evidence (Collen et al 2022).

Ultimately, this study served not only as a reminder of the difficulties of working with ancient DNA, but the importance of it. Studies only utilizing modern samples have been unable to replicate the same findings of those that incorporate ancient genomic data, as ancient data is able to aid in the construction of evolutionary histories across time periods and population changes (Collen et al 2022; Kerner et al 2021). The results of this study call upon further study of the region, and post-admixture selection across the board. A variety of questions are left unanswered or unable to be answered. Many of the variants that were found to be under significant selective pressure couldn't be explained by the event of colonization, as expected. Obviously, there have existed numerous other factors that have led to the emergence of selective pressures, thereby shaping the evolution of Mexican populations over centuries. As we sifted through the results, we began to notice several distinct "categories" of variants emerging.

With many variants under selection having functional consequences on genes affecting cancer, diabetes, obesity, neurodegenerative diseases, and more, the value of public health genomics becomes clear. In recent years, many scholars have advocated for the incorporation of genomic findings into population-specific health interventions (Molster et al). While social and environmental factors significantly shape health, the genetic and evolutionary underpinnings of our health cannot be ignored. With the biases that exist in genomic data, it is of incredible

importance to diversify the genomic records of the world to better understand population-specific physiologies and predispositions to various health outcomes. Moreover, given the utility in aDNA analysis to studies of natural selection and evolution, we hope this study can serve as a continued reminder to discover and analyze ancient genomes to progress the fields of genetics and evolutionary biology.

Moving forward with the results of this study, there are many avenues for future research. For example, it would be interesting to determine the ancestral original of the variants of immune significance. If many of these variants are of European origin, it would suggest that the introduction of these variants via admixture into the Mexican population allowed for improved immunity against novel pathogens. Variants of Mexican origin would suggest that variation in immune response existed among the population prior to contact, and the introduction of novel pathogens only selected for those who already had a fitness advantage. Either way, investigating the origin of variants of relevance to the immune system can shed light on the exact method of selection that has shaped the immune response among modern-day Mexicans.Alternatively, future studies into the impact of these variants on gene function could be done. Although the gTEX eQTL is a comprehensive resource for assessing the impact of SNVs on gene function, more studies should be done to bolster the data available. Particularly given how Latin American populations are underrepresented in GWAS studies, the findings of this study can be used to improve understanding of variants common among Mexican populations which can affect immune gene function.

References

Abate, W., et al. “Lysophosphatidylcholine Acyltransferase 2 (LPCAT2) Co-Localises with TLR4 and Regulates Macrophage Inflammatory Gene Expression in Response to LPS.” *Scientific Reports*, vol. 10, June 2020, p. 10355. *PubMed Central*, https://doi.org/10.1038/s41598-020-67000-x.

Abraham, Gad, and Michael Inouye. “Fast Principal Component Analysis of Large-Scale Genome-Wide Data.” *PLOS ONE*, vol. 9, no. 4, Apr. 2014, p. e93766. *PLoS Journals*, https://doi.org/10.1371/journal.pone.0093766.

Acuna-Soto, Rodolfo, et al. “Megadrought and Megadeath in 16th Century Mexico.” *Emerging Infectious Diseases*, vol. 8, no. 4, Apr. 2002, pp. 360–62. *PubMed Central*, https://doi.org/10.3201/eid0804.010175.

Alberts, Bruce, et al. “Innate Immunity.” *Molecular Biology of the Cell. 4th Edition*, Garland Science, 2002. *www.ncbi.nlm.nih.gov*, https://www.ncbi.nlm.nih.gov/books/NBK26846/.

Aranow, Cynthia. “Vitamin D and the Immune System.” *Journal of Investigative Medicine : The Official Publication of the American Federation for Clinical Research*, vol. 59, no. 6, Aug. 2011, pp. 881–86. *PubMed Central*, https://doi.org/10.231/JIM.0b013e31821b8755.

Chen, Huanhuan, et al. “Super-enhancer–Associated Long Noncoding RNA RP11-569A11.1 Inhibited Cell Progression and Metastasis by Regulating IFIT2 in Colorectal Cancer.” *Journal of Clinical Laboratory Analysis*, vol. 35, no. 6, May 2021, p. e23780. *PubMed Central*, https://doi.org/10.1002/jcla.23780.

Chiu, Alec M., et al. “Inferring Population Structure in Biobank-Scale Genomic Data.” *The American Journal of Human Genetics*, vol. 109, no. 4, Apr. 2022, pp. 727–37. *DOI.org (Crossref)*, https://doi.org/10.1016/j.ajhg.2022.02.015.

Cho, Young Jin, et al. “Abr and Bcr, Two Homologous Rac GTPase-Activating Proteins, Control Multiple Cellular Functions of Murine Macrophages.” *Molecular and Cellular Biology*, vol. 27, no. 3, Feb. 2007, pp. 899–911. *PubMed Central*, https://doi.org/10.1128/MCB.00756-06.

Collen, Evelyn Jane, et al. “The Immunogenetic Impact of European Colonization in the Americas.” *Frontiers in Genetics*, vol. 13, 2022. *Frontiers*, https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2022.918227.

Curtis, David. “Genetic Variants Associated with Hypertension Risk: Progress and Implications.” *Pulse*, vol. 12, no. 1, Jan. 2024, pp. 19–26. *Silverchair*, https://doi.org/10.1159/000536505.

Dabney, Jesse, et al. “Ancient DNA Damage.” *Cold Spring Harbor Perspectives in Biology*, vol. 5, no. 7, July 2013, p. a012567. *PubMed Central*, https://doi.org/10.1101/cshperspect.a012567.

Dubey, Preeti. *Stability Analysis and Simulation of Deterministic Models in Epidemiology and Immunology*. 2016.

García-Ortiz, Humberto, et al. "The Genomic Landscape of Mexican Indigenous Populations Brings Insights into the Peopling of the Americas." *Nature Communications*, vol. 12, Oct. 2021, p. 5942. *PubMed Central*, https://doi.org/10.1038/s41467-021-26188-w.

Gelemanović, Andrea, Tatjana Ćatipović Ardalić, et al. "Genome-Wide Meta-Analysis Identifies Multiple Novel Rare Variants to Predict Common Human Infectious Diseases Risk." *International Journal of Molecular Sciences*, vol. 24, no. 8, Apr. 2023. *www.ncbi.nlm.nih.gov*, https://doi.org/10.3390/ijms24087006.

Gelemanović, Andrea, Tatjana Ćatipović Ardalić, et al. "Genome-Wide Meta-Analysis Identifies Multiple Novel Rare Variants to Predict Common Human Infectious Diseases Risk." *International Journal of Molecular Sciences*, vol. 24, no. 8, Apr. 2023, p. 7006. *PubMed Central*, https://doi.org/10.3390/ijms24087006.

Guzmán-Solís, Axel A., et al. "Ancient Viral Genomes Reveal Introduction of Human Pathogenic Viruses into Mexico during the Transatlantic Slave Trade." *eLife*, vol. 10, p. e68612. *PubMed Central*, https://doi.org/10.7554/eLife.68612. Accessed 10 Apr. 2024.

Haines, Michael R., and Richard H. Steckel. *A Population History of North America*. Cambridge University Press, 2000.

Hayeems, Robin Z., et al. "Clinical Utility of Genomic Sequencing: A Measurement Toolkit." *Npj Genomic Medicine*, vol. 5, no. 1, Dec. 2020, pp. 1–11. *www.nature.com*, https://doi.org/10.1038/s41525-020-00164-7.

Jiang, Xueyuan, and Raquel Assis. "Population-Specific Genetic and Expression Differentiation in Europeans." *Genome Biology and Evolution*, vol. 12, no. 4, Apr. 2020, pp. 358–69. *Silverchair*, https://doi.org/10.1093/gbe/evaa021.

Kerner, Gaspard, et al. "New Insights into Human Immunity from Ancient Genomics." *Current Opinion in Immunology*, vol. 72, Oct. 2021, pp. 116–25. *ScienceDirect*, https://doi.org/10.1016/j.coi.2021.04.006.

Khong, Jwu Jin, et al. "Association of Polymorphisms in MACRO Domain Containing 2 With Thyroid-Associated Orbitopathy." *Investigative Ophthalmology & Visual Science*, vol. 57, no. 7, June 2016, pp. 3129–37. *Silverchair*, https://doi.org/10.1167/iovs.15-18797.

Khoury, Muin J., and Kathryn E. Holt. "The Impact of Genomics on Precision Public Health: Beyond the Pandemic." *Genome Medicine*, vol. 13, Apr. 2021, p. 67. *PubMed Central*, https://doi.org/10.1186/s13073-021-00886-y.

Lau, Colleen M., et al. "Transcription Factor Etv6 Regulates Functional Differentiation of Cross-Presenting Classical Dendritic Cells." *The Journal of Experimental Medicine*, vol. 215, no. 9, Sept. 2018, pp. 2265–78. *PubMed Central*, https://doi.org/10.1084/jem.20172323.

Lee, Shin-Ae, et al. "The Role of High Mobility Group Box 1 in Innate Immunity." *Yonsei Medical Journal*, vol. 55, no. 5, Sept. 2014, pp. 1165–76. *PubMed Central*, https://doi.org/10.3349/ymj.2014.55.5.1165.

Li, Jing, et al. “FAT4 Expression in Peripheral Blood Mononuclear Cells Is Associated with Prognosis and Immune Cell Infiltration in Hepatocellular Carcinoma.” *Scientific Reports*, vol. 13, no. 1, Sept. 2023, p. 15735. *www.nature.com*, https://doi.org/10.1038/s41598-023-42560-w.

Li, Qiutang, and Inder M. Verma. “NF-κB Regulation in the Immune System.” *Nature Reviews Immunology*, vol. 2, no. 10, Oct. 2002, pp. 725–34. *www.nature.com*, https://doi.org/10.1038/nri910.

Lindo, John, et al. “The Genetic Prehistory of the Andean Highlands 7000 Years BP Though European Contact.” *Science Advances*, vol. 4, no. 11, Nov. 2018, p. eaau4921. *science.org (Atypon)*, https://doi.org/10.1126/sciadv.aau4921.

Liu, Zining, et al. “A 9-gene Expression Signature to Predict Stage Development in Resectable Stomach Adenocarcinoma.” *BMC Gastroenterology*, vol. 22, Oct. 2022, p. 435. *PubMed Central*, https://doi.org/10.1186/s12876-022-02510-8.

Lonsdale, John, et al. “The Genotype-Tissue Expression (GTEx) Project.” *Nature Genetics*, vol. 45, no. 6, June 2013, pp. 580–85. *www.nature.com*, https://doi.org/10.1038/ng.2653.

Mamoor, Shahan. *Differential Expression of CGGBP1 in Amyotrophic Lateral Sclerosis.* OSF, 20 Sept. 2022. *OSF Preprints*, https://doi.org/10.31219/osf.io/mqytb.

Maples, Brian K., et al. “RFMix: A Discriminative Modeling Approach for Rapid and Robust Local-Ancestry Inference.” *American Journal of Human Genetics*, vol. 93, no. 2, Aug. 2013, pp. 278–88. *PubMed Central*, https://doi.org/10.1016/j.ajhg.2013.06.020.

Mathieson, Iain. “Limited Evidence for Selection at the FADS Locus in Native American Populations.” *Molecular Biology and Evolution*, vol. 37, no. 7, July 2020, pp. 2029–33. *Silverchair*, https://doi.org/10.1093/molbev/msaa064.

Mertowska, Paulina, et al. “Immunomodulatory Role of Interferons in Viral and Bacterial Infections.” *International Journal of Molecular Sciences*, vol. 24, no. 12, June 2023, p. 10115. *PubMed Central*, https://doi.org/10.3390/ijms241210115.

Norris, Emily T., et al. “Genetic Ancestry, Admixture and Health Determinants in Latin America.” *BMC Genomics*, vol. 19, no. 8, Dec. 2018, p. 861. *BioMed Central*, https://doi.org/10.1186/s12864-018-5195-7.

Ozato, Keiko, et al. “TRIM Family Proteins and Their Emerging Roles in Innate Immunity.” *Nature Reviews. Immunology*, vol. 8, no. 11, Nov. 2008, pp. 849–60. *PubMed Central*, https://doi.org/10.1038/nri2413.

Pickrell, Joseph K., and Jonathan K. Pritchard. “Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data.” *PLOS Genetics*, vol. 8, no. 11, Nov. 2012, p. e1002967. *PLoS Journals*, https://doi.org/10.1371/journal.pgen.1002967.

Remely, Marlene, et al. “Obesity: Epigenetic Regulation – Recent Observations.” *Biomolecular Concepts*, vol. 6, no. 3, June 2015, pp. 163–75. *www.degruyter.com*, https://doi.org/10.1515/bmc-2015-0009.

Ren, Yi-Ming, et al. “mRNA and Long Non-Coding RNA Expression Profiles of Rotator Cuff Tear Patients Reveal Inflammatory Features in Long Head of Biceps Tendon.” *BMC*

Medical Genomics, vol. 15, no. 1, June 2022, p. 140. *Springer Link*, https://doi.org/10.1186/s12920-022-01292-y.

Sewell, Jared A., and Juan I. Fuxman Bass. "Cellular Network Perturbations by Disease-Associated Variants." *Current Opinion in Systems Biology*, vol. 3, June 2017, pp. 60–66. *ScienceDirect*, https://doi.org/10.1016/j.coisb.2017.04.009.

Shastry, Barkur S. "SNPs in Disease Gene Mapping, Medicinal Drug Development and Evolution." *Journal of Human Genetics*, vol. 52, no. 11, 11, Nov. 2007, pp. 871–80. *www.nature.com*, https://doi.org/10.1007/s10038-007-0200-z.

Shih, Tzenge-Lien, et al. "Halo-Substituted Chalcones and Azachalcones-Inhibited, Lipopolysaccharited-Stimulated, Pro-Inflammatory Responses through the TLR4-Mediated Pathway." *Molecules*, vol. 23, Mar. 2018, p. 597. *ResearchGate*, https://doi.org/10.3390/molecules23030597.

Silva, Nuno Miguel, et al. "Investigating Population Continuity with Ancient DNA under a Spatially Explicit Simulation Framework." *BMC Genetics*, vol. 18, no. 1, Dec. 2017, p. 114. *BioMed Central*, https://doi.org/10.1186/s12863-017-0575-6.

Sousa da Mota, Bárbara, et al. "Imputation of Ancient Human Genomes." *Nature Communications*, vol. 14, no. 1, 1, June 2023, p. 3660. *www.nature.com*, https://doi.org/10.1038/s41467-023-39202-0.

Stark, Benjamin C., et al. "CARMIL Family Proteins as Multidomain Regulators of Actin-Based Motility." *Molecular Biology of the Cell*, vol. 28, no. 13, July 2017, pp. 1713–23. *molbiolcell.org (Atypon)*, https://doi.org/10.1091/mbc.e17-01-0019.

Suzuki, Hiromichi, et al. "Recurrent Non-Coding U1-snRNA Mutations Drive Cryptic Splicing in Shh Medulloblastoma." *Nature*, vol. 574, no. 7780, Oct. 2019, pp. 707–11. *PubMed Central*, https://doi.org/10.1038/s41586-019-1650-0.

Vidal, Elena A., et al. "Whole Genome Sequence, Variant Discovery and Annotation in Mapuche-Huilliche Native South Americans." *Scientific Reports*, vol. 9, no. 1, 1, Feb. 2019, p. 2132. *www.nature.com*, https://doi.org/10.1038/s41598-019-39391-z.

Villa-Islas, Viridiana. *Demographic History and Genetic Structure in Pre-Hispanic Central Mexico | Science*. https://www.science.org/doi/10.1126/science.add6142. Accessed 8 Feb. 2024.

Wang, Haijiao, et al. "Variants in SNAI1, AMDHD1 and CUBN in Vitamin D Pathway Genes Are Associated with Breast Cancer Risk: A Large-Scale Analysis of 14 GWASs in the DRIVE Study." *American Journal of Cancer Research*, vol. 10, no. 7, 2020, pp. 2160–73.

Wang, Sijia, et al. "Genetic Variation and Population Structure in Native Americans." *PLOS Genetics*, vol. 3, no. 11, Nov. 2007, p. e185. *PLoS Journals*, https://doi.org/10.1371/journal.pgen.0030185.

Yelmen, Burak, et al. "Improving Selection Detection with Population Branch Statistic on Admixed Populations." *Genome Biology and Evolution*, vol. 13, no. 4, Feb. 2021, p. evab039. *PubMed Central*, https://doi.org/10.1093/gbe/evab039.

Yi, Xin, et al. “Sequencing of Fifty Human Exomes Reveals Adaptation to High Altitude.” *Science (New York, N.Y.)*, vol. 329, no. 5987, July 2010, pp. 75–78. *PubMed Central*, https://doi.org/10.1126/science.1190371.

Yu, Hong-Bing, et al. “NFATc2 Mediates Epigenetic Modification of Dendritic Cell Cytokine and Chemokine Responses to Dectin-1 Stimulation.” *Nucleic Acids Research*, vol. 43, no. 2, Jan. 2015, pp. 836–47. *PubMed Central*, https://doi.org/10.1093/nar/gku1369.

Zhang, Chongyang, et al. “Cyclic Nucleotide Phosphodiesterase 1C Contributes to Abdominal Aortic Aneurysm.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 118, no. 31, Aug. 2021, p. e2107898118. *PubMed Central*, https://doi.org/10.1073/pnas.2107898118.

Zhang, Wenchuan, et al. “Prognostic Value of 12 m7G Methylation-Related miRNA Markers and Their Correlation with Immune Infiltration in Breast Cancer.” *Frontiers in Oncology*, vol. 12, Aug. 2022, p. 929363. *PubMed Central*, https://doi.org/10.3389/fonc.2022.929363.

Zheng, Xiuwen, et al. “A High-Performance Computing Toolset for Relatedness and Principal Component Analysis of SNP Data.” *Bioinformatics*, vol. 28, no. 24, Dec. 2012, pp. 3326–28. *PubMed Central*, https://doi.org/10.1093/bioinformatics/bts606.

Ziegler-Heitbrock, Loems. “The CD14+ CD16+ Blood Monocytes: Their Role in Infection and Inflammation.” *Journal of Leukocyte Biology*, vol. 81, no. 3, Mar. 2007, pp. 584–92. *Silverchair*, https://doi.org/10.1189/jlb.0806510.

Zverinova, Stepanka, and Victor Guryev. “Variant Calling: Considerations, Practices, and Developments.” *Human Mutation*, vol. 43, no. 8, Aug. 2022, pp. 976–85. *PubMed Central*, https://doi.org/10.1002/humu.24311.

Supplemental Information

Chr	Pos	Annotated Gene(s)	Type	rsID	gTEx eQTL hit?	PBS Score
chr20	50205014	NFATC2, ATP9A	intergenic	rs367721985	N/A	2.953205394
chr4	189126611	TRIML1, LINC01060	intergenic	rs202199280	N/a	2.73813038
chr4	126080905	ANKRD50, FAT4	intergenic	rs75096357	N/A	2.724327794
chr7	9905097	PER4, MGC4859	intergenic	rs115886780	N/A	2.724327794

chr19	14895594	ADGRE2, OR7C1	intergenic	rs12459654	ADREG2, ZNF333, OR7C1	2.582377763
chr12	30418638	TMTC1, IPO8	intergenic	rs1652143	N/A	2.536942843
chr20	16193876	MACROD2, KIF16B	intergenic	rs6110964	N/A	2.525286266
chr4	129957509	SCLT1	intronic	rs75265815	N/A	2.525286266
chr7	9961509	PER4, MGC4859	intergenic	rs10479888	N/A	2.506248442
chr17	1022796	ABR	intronic	rs9747624	N/A	2.447124267
chr12	11834424	ETV6	intronic	rs10676577, rs200633262	N/A	2.446353212
chr12	9997789	KLRF1	downstream	rs2012290	A2MP1	2.444850861
chr16	55587004	LPCAT2	intronic	rs138889705	LPCAT2	2.444850861
chr4	72617775	GC	intronic	rs1352846	N/A	2.444850861
chr6	25470268	CARMIL1	intronic	rs114343309	N/A	2.444850861

Supplementary Table S1. List of SNPs occurring in gene regions related to immune function.

rsID	**Chr**	**Pos**	**Annotated Gene(s)**	**Type**	**gTEx eQTL hit?**	**Associated Health Pathway**	**PBS Score**
rs2125801	chr4	156698001	ADGRL3-AS1, TECRL	intronic	GUCY1B1	Cardiovascular	2.724327794
rs11773452	chr7	32087464	PDE1C gene	intronic	PDE1C	Cardiovascular	2.724327794
rs6001981	chr22	41017425	MRTFA	intronic	MKL1, XPNPEP3, RP5-1042K10.10, RP5-1042K10.13, RP4-591N18.2	Cancer	2.462610555
rs375303324	chr12	118446459	KSR2, RFC5	intergenic	RFC5, WSB2	Obesity Diabetes	2.428146633

rs12496388	chr3	88363701	C3orf38, CSNKA2IP	intergenic	CGGBP1	Cancer	2.724327794
rs1896802	chr15	84238192	SH3GL3	intronic	RP11-182J1.14, GOLGA6L4, FAM103A1, RP11-90B9.2, GOLGA6L4	Cancer	2.462610555
rs72720809	chr8	128984413	PVT1	ncRNA_intronic	RNU1-106P	Cancer	2.447124267
rs60845745	chr8	128985757	PVT1	ncRNA_intronic	RNU1-106P	Cancer	2.447124267
rs12833182	chr12	96153441	NTN4	intronic	RP11-536G4.2, AMDHD1	Cancer	2.444850861
rs13128039	chr4	174190222	GALNT7	intronic	RP11-10K16.1, SCRG1	Cancer	2.444850861
rs12496388	chr3	88363701	C3orf38, CSNKA2IP	intergenic	CGGBP1	Chronic Inflammation	2.724327794

Supplementary Table S2: List of variants with gTEX eQTL hits affecting health outcomes along with associated pathways, PBS score, and relevant characterization of variant.

Sample ID	**Population**	**Reference**
S_Chane-1	Chane	Simon's Genome Diversity Project
S_Surui-1	Surui	Simon's Genome Diversity Project
S_Surui-2	Surui	Simon's Genome Diversity Project
S_Karitiana-1	Karitiana	Simon's Genome Diversity Project
S_Karitiana-2	Karitiana	Simon's Genome Diversity Project
S_Piapoco-1	Piapoco	Simon's Genome Diversity Project
S_Piapoco-2	Piapoco	Simon's Genome Diversity Project
S_Pima-1	Pima	Simon's Genome Diversity Project

S_Pima-2	Pima	Simon’s Genome Diversity Project
S_Mayan-1	Maya	Simon’s Genome Diversity Project
S_Mayan-2	Maya	Simon’s Genome Diversity Project
S_Zapotec-1	Zapotec	Simon’s Genome Diversity Project
S_Mixe-2	Mixe	Simon’s Genome Diversity Project
S_Mixe-3	Mixe	Simon’s Genome Diversity Project
S_Mixtec-1	Mixtec	Simon’s Genome Diversity Project
S_Mixtec-2	Mixtec	Simon’s Genome Diversity Project
S_Zapotec-2	Zapotec	Simon’s Genome Diversity Project
S_Quechua-3	Quechua	Simon’s Genome Diversity Project
S_Quechua-2	Quechua	Simon’s Genome Diversity Project
S_Quechua-1	Quechua	Simon’s Genome Diversity Project
S_Han-2	East Asian	Simon’s Genome Diversity Project
S_Dai-3	East Asian	Simon’s Genome Diversity Project
S_Han-1	East Asian	Simon’s Genome Diversity Project
S_Hezhen-1	East Asian	Simon’s Genome Diversity Project
S_Hezhen-2	East Asian	Simon’s Genome Diversity Project
S_Dai-2	East Asian	Simon’s Genome Diversity Project
S_Dai-1	East Asian	Simon’s Genome Diversity Project
S_Japanese-1	East Asian	Simon’s Genome Diversity Project
S_Japanese-2	East Asian	Simon’s Genome Diversity Project
S_Japanese-3	East Asian	Simon’s Genome Diversity Project
S_Korean-2	East Asian	Simon’s Genome Diversity Project

S_Korean-1	East Asian	Simon’s Genome Diversity Project
S_Bulgarian-1	European	Simon’s Genome Diversity Project
S_Bulgarian-2	European	Simon’s Genome Diversity Project
S_English-1	European	Simon’s Genome Diversity Project
S_English-2	European	Simon’s Genome Diversity Project
S_French-1	European	Simon’s Genome Diversity Project
S_French-2	European	Simon’s Genome Diversity Project
S_Basque-1	European	Simon’s Genome Diversity Project
S_Hungarian-2	European	Simon’s Genome Diversity Project
S_Hungarian-1	European	Simon’s Genome Diversity Project
S_Bergamo-1	European	Simon’s Genome Diversity Project
S_Bergamo-2	European	Simon’s Genome Diversity Project
S_Tuscan-2	European	Simon’s Genome Diversity Project
S_Tuscan-1	European	Simon’s Genome Diversity Project
S_Norwegian-1	European	Simon’s Genome Diversity Project
S_Polish-1	European	Simon’s Genome Diversity Project
S_Spanish-1	European	Simon’s Genome Diversity Project
S_Spanish-2	European	Simon’s Genome Diversity Project
B_Crete-1	European	Simon’s Genome Diversity Project
B_Dai-4	East Asian	Simon’s Genome Diversity Project
B_French-3	European	Simon’s Genome Diversity Project
B_Han-3	East Asian	Simon’s Genome Diversity Project
B_Karitiana-3	Karitiana	Simon’s Genome Diversity Project

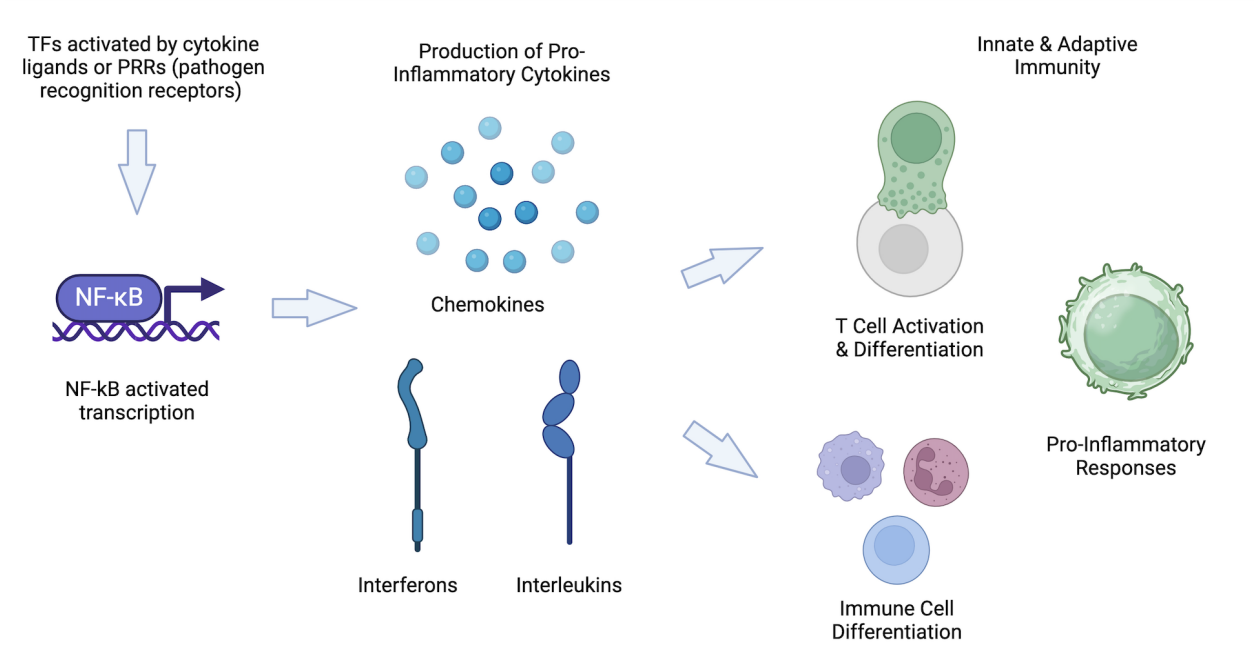
B_Mixe-1	Mixe	Simon's Genome Diversity Project
333B	Ancients at Sierra Gorda	Villa-Islas et al. 2023
2417J	Ancients at Sierra Gorda	Villa-Islas et al. 2023
2417Q	Ancients at Sierra Gorda	Villa-Islas et al. 2023
F9	Ancients at Sierra Tarahumara	Villa-Islas et al. 2023

Supplementary Table S3. A list of samples used in study along with location of origin and reference.

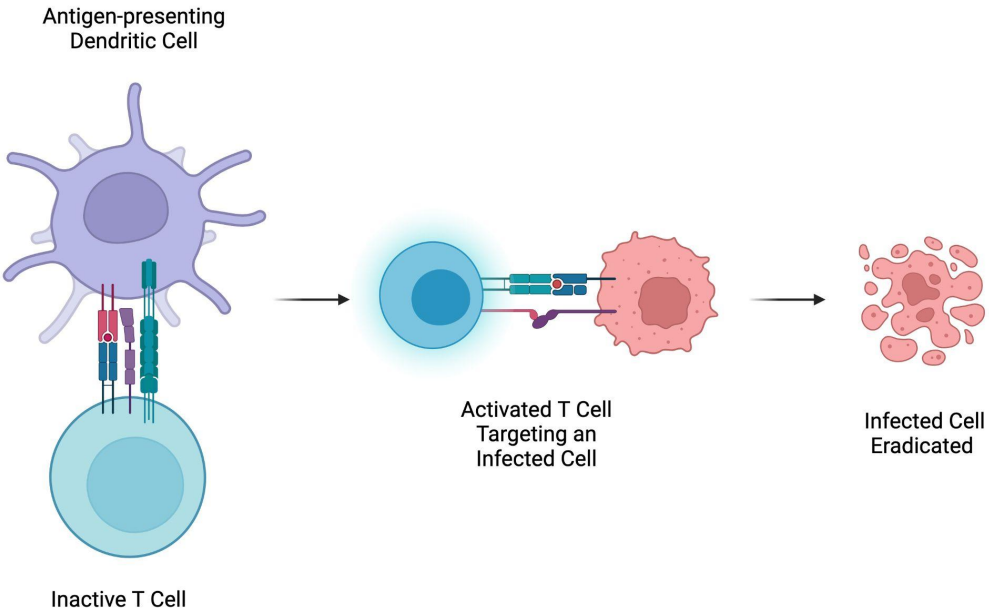
Sample ID	**Population**	**Reference**
F9	Ancients at Sierra Tarahumara	Villa-Islas et al. 2023
2417Q	Ancients at Sierra Gorda	Villa-Islas et al. 2023
2417J	Ancients at Sierra Gorda	Villa-Islas et al. 2023
333B	Ancients at Sierra Gorda	Villa-Islas et al. 2023

Supplementary Table S4. A list of high-coverage (>1x) ancient samples used to conduct population branch statistic (PBS).

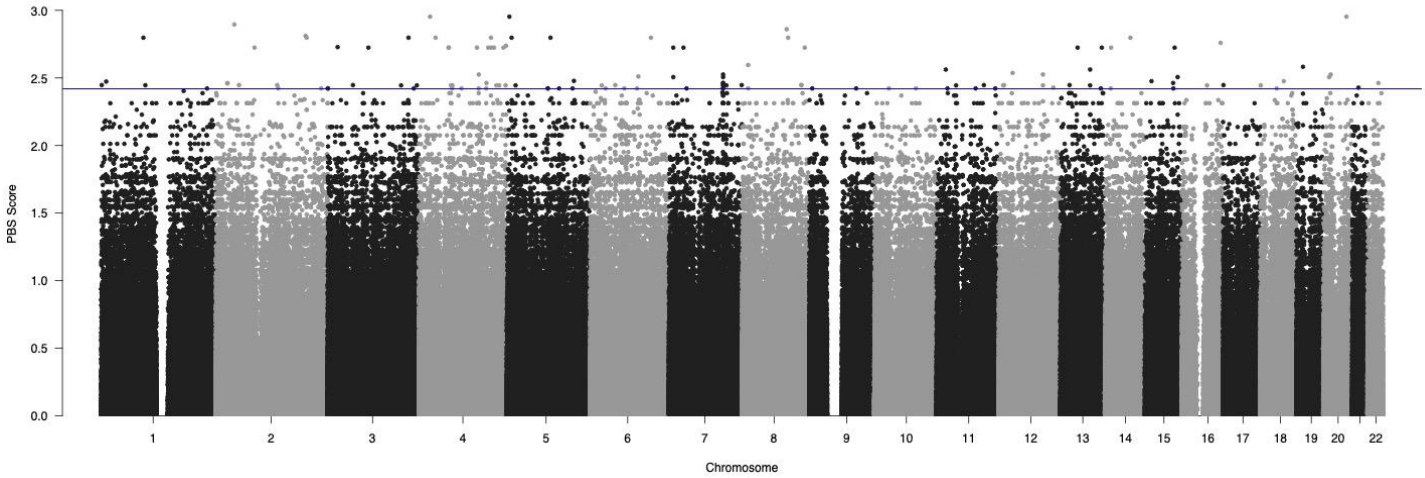
Supplementary Figure F1. Map of high-coverage ancient samples included in the natural selection scan. Created with Biorender.com.



Supplementary Figure F2. Illustration of the NFKB pathway in the innate and adaptive immune system. Created with Biorender.com.



Supplementary Figure F3. Illustration of the function of dendritic cells in the immune system. Created with Biorender.com.



Supplementary Figure F4. Manhattan plot of PBS results. Blue line indicates the threshold above which the top 100 variants under the strongest selective pressure lie. These 100 variants were chosen to be examined in the study.