Prioritization of Evolutionary Events for Immune Response Studies: Introgression, Selection, and Infectious Disease Among Indigenous Americans

by

Obed A.C. Garcia

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Anthropology)
in the University of Michigan
2020

Doctoral Committee:

      Associate Professor Abigail W. Bigham, UCLA, Co-Chair
      Associate Professor Jacinta Beehner, Co-Chair
      Associate Professor John Kingston
      Professor Jun Li

Obed Garcia

oag@umich.edu

ORCID iD: 0000-0001-6150-6118

DEDICATION

This dissertation is dedicated to my parents, Obed and Olga. Without their support and

sacrifices, I would not be where I am today.

ACKNOWLEDGEMENTS

I started graduate school in the Fall of 2012, a journey that has lasted 8 years. This period of my life is filled with unforgettable memories, surrounded me with incredible people, and gave me the opportunity to conduct fieldwork in various parts of Latin America. For bringing me to Michigan, I am eternally grateful to my advisor and role model Abigail Bigham. She imparted her expertise on genetics/genomics, spent countless hours spent pouring over my writing, and mentored me so that I could become a better scholar. I could not have asked for a more supportive advisor, someone who let me chase my dreams of conducting fieldwork in Guatemala and was inspiring in moments of despair and desperation.

I gained invaluable experience conducting fieldwork in the birthplace of my parents, Guatemala. Although I had visited Guatemala before, primarily to visit family while I was growing up, I wanted to establish research there and give back to the local communities. I initially struggled to start a dengue project in Guatemala and faced various setbacks, I became intimately aware of how health disparities and bureaucracy affect Guatemalans and hinders progress. However, my setbacks changed when I met Leticia Castillo-Signor who also shared my passion to help Guatemalans. She taught me how to navigate the Ministry of Public Health and Social Assistance in Guatemala and seek out collaborators. Without her help and support, I would still be attempting to

A thank you to all the support I received from the rest of the University of Michigan and the Department of Anthropology, particularly the office staff. Great appreciation to my cohort, that became my extended family at Michigan: Irisa Arney, Jeffrey Bradshaw, Barry Brillantes, Adrian Deoanca, John Doering-White, Georgia Ennis, Matan Kaminer, Nama Khalil, Adrienne Lagman, Prash Naidu, Sandhya Narayanan, Niku T'archechu T'arhesi, and Warren Thompson. And we cannot forget other close friends and colleagues in the department whom I also consider family: Vincent Battista, Abagail Breidenstein, Ainash Childebayeva, Paloma Contreras, Andrew Foster, Rachna Reddy, Aaron Sandel, and Elizabeth Werren. Thank you to BAGS and BAM, who always provided feedback on various presentations and grants. A special thank you to John Mitani, who was a mentor to me outside of my committee. I'm grateful to Rob Pettigrew, who helped me out to do some last-minute formatting edits on my dissertation. Many thanks to the various internal and external fellowships and awards, particularly the Wenner Gren Dissertation Fieldwork Grant and the NSF DDRIG that made life and research much easier.

My gratitude to the people I intersected with at Northwestern, where I had my first hands on experiences in conducting research: Christine Ackerman, Scott Budinger, Jiwang Chen, Andrea Dunaif, Geoffrey Hayes, Karen Ridge, Iasha Sznajder, Richard Wunderink, Margrit Urbanek, and many, many others. Many thanks to Dan Eisenberg, who somehow convinced me to become an anthropologist.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

A history of exposure to infectious diseases has influenced the biology of Mesoamericans. From the urbanization of Mesoamerican communities to colonial contact in the 15th century, immune response to various pathogens have shaped the available diversity at various genomic loci providing immunological resistance to pathogens. This dissertation explores the population history of Mesoamerican populations as it relates to exposures to and immunity against infectious diseases. I use an evolutionary lens to look at three time scales: 1) archaic introgression in immune response pathways that took place thousands of years prior to the peopling of the Americas 30,000 years ago, 2) signatures of natural selection in immune response genes from the establishment of city-states to colonial contact, shaped 10,000 years ago to the 15th century, 3) epidemiological risk factors for dengue-infection in Guatemala between the years 2018 and 2019.

I use publicly available DNA sequences from Indigenous Americans, SNP array genotype data, and ethnographic surveys to answer the following questions: 1) Are the detectable regions of archaic introgression in modern Mesoamericans related to immunity? 2) Are we able to detect signatures of natural selection in Mesoamericans at immune response loci? 3) What risk factors contribute to dengue-infection in Mesoamerican populations today?

The introgression analysis in an indigenous American population from Mexico finds that multiple genes lying in the TGF-β signaling pathway were introgressed by Neanderthals. Similarly, various genes associated with cytoplasm and exosomes, part of the innate immune system, were introgressed by Denisovans. The presence of those segments in important immune response pathways and responses highlights that variants deep in evolutionary history may continue to be relevant today.

Similarly, our natural selection scan in a Mesoamerican cohort demonstrated that the Major Histocompatibility Complex (MHC) and the peroxisome proliferator-activated receptor gamma (PPAR-γ) signaling pathway showed signatures of positive selection. This demonstrates that in the more recent past, various pathogens including smallpox exerted strong evolutionary pressure on the human genome, thus shaping host responses to infection.

Our epidemiological survey demonstrated various risk factors for dengue infection in Guatemalan populations of Mesoamerican ancestry. We found that factors related to water accumulation including water security, water storage, and nearby trash increased the risk for dengue-infection likely by providing additional breeding grounds for mosquito. This study suggests where the Ministry of Public Health should focus their efforts for dengue prevention. Furthermore, it identifies the epidemiological risk-factors that future studies of dengue infection should account for in Guatemala.

This dissertation highlights the importance of adaptation on immunity among Mesoamerican populations whether deep in our evolutionary past, as recent as colonial contact, or even continuously shaped by more recent infectious diseases like dengue.

Furthermore, it demonstrates how to prioritize candidate genomic regions influencing host susceptibility or resistance to modern infectious disease using an evolutionary approach. This emphasis on compiling outlying genes will be particularly beneficial to studies conducted on smaller sample sizes as it provides greater statistical power given that these outlying genes have greater effects physiologically more so than other genomic regions.

Chapter 1: Introduction and Background

Approaches to medicine based on evolution provide a more holistic view for understanding modern health and disease. The framework entangling both medicine and evolution is known as evolutionary medicine or Darwininan medicine. Evolutionary medicine seeks to incorporate perspectives from medicine, public health, and anthropology (among other fields) to understand how humans evolve, how we continue evolving, and how this affects our overall health in the present (Nesse and Stearns 2008). This framework allows scholars to understand both susceptibility and resistance to modern disease, discover more about human physiology, and better treat disease—not just the symptoms presented. As Dobhzansky famously claimed, "nothing in biology makes sense except in light of evolution" (Dobzhansky 1973).

This dissertation explores the population history of Mesoamerican populations with respect to immunity and exposures to infectious diseases through an evolutionary lens. It draws on the growing literature of human evolutionary history and expands on it. It demonstrates how to leverage knowledge of archaic adaptive introgression and past events of natural selection to incorporate evolution into the study of infectious disease. These same approaches can be expanded to include autoimmune disorders, cardiac health, among others.

In more ways than one, this dissertation is a deeply personal journey into Mesoamerican history, given my family ties to Guatemala—part of the admixed local

population known as Ladinos. Observing the blatant health disparities that are apparent in both Guatemala and Latinx populations in the US, I attempted to understand health and disease from an evolutionary perspective. Representation in genomics has improved over the last decade. Nonetheless, there remains a critical gap in studies focusing on Latinx populations and even more so on populations of Indigenous American ancestry (Popejoy and Fullerton 2016). This dissertation incorporates an evolutionary framework in order to holistically understand the origin and impact of Indigenous American genomic variation on modern immune response.

Mesoamerica most commonly refers to the geographic region encompassing modern-day Northern/Central Mexico to parts of Costa Rica. Today there are roughly 6 major Indigenous civilizations or ethnicities who speak languages belonging to one of 13 separate language families. Early crop domestication including maize and cacao led to the establishment of agriculture (Flannery 2019). This in turn led to population growth. By 3,000 years ago, Mesoamericans established powerful, dense, and populous city-states (Flannery 2019). The urbanization of Mesoamerican city states allowed for the circulation of infectious diseases including: chagas, tuberculosis, syphilis, and hepatitis (Bos et al. 2014; Klaus et al. 2010; Merbs 1992; Steverding 2014). This environment undoubtedly played a role in shaping indigenous Mesoamerican population's genomes through natural selection by shifting the frequencies of advantageous alleles that conferred resistance to pathogenic infection (Mummert et al. 2011; Smith 2016).

Fast forward in history to the year 1492, the year European colonization of the Americas began. Colonial expansion brought a devastating blow to the population. Along with contact came novel infectious diseases. Historical records describe symptoms consistent with smallpox that devastated Mesoamerican communities. In fact, several scholars estimate that up to 90% of the population was decimated (Feldman 1999; Leon-Portilla 2011; Lockhart 1992; Restall et al. 2005). Again, allele frequencies that conferred resistance to infection shifted from a neutral distribution to being positively selected, and those advantageous alleles would be present at higher frequencies in the surviving segments of the population.

Today, globalizing economies and rapid urbanization coupled with substandard infrastructure of Latin American cities has introduced modern day Mesoamerican populations to novel infectious diseases (Alirol et al. 2011; Gubler 2011). Trade, particularly of ships carrying used tires, spread new infectious diseases as the stagnant water in the tires was a prime breeding ground for mosquito vectors (Sprenger 1987). Now, climate change expands and threatens populations with novel, emerging infectious diseases as the ecological niche of the vectors also expands (Reiter 2001). Underlying genetic variation in a population will ultimately influence the immune response to infection. Among Mesoamericans, this underlying genetic variation undoubtedly has been shaped by the cyclical history of exposure to infectious diseases over time.

Immune system

      In order to understand how the immune system is shaped by pathogens, we must understand it within a broad framework of evolutionary pressures. This framework posits hosts and pathogens locked in an arms race for survival, which in turn drives evolutionary change. This concept is known as the Red Queen Hypothesis (Smith 1976; Van Valen 1973). Hosts and pathogens push each other to adapt, either one defeats the other or they both co-exist. The pathogen lifecycle is much shorter; therefore, it can evolve rapidly. Human hosts on the other hand have much longer lifespans. To overcompensate for having a longer lifespan than the pathogen, humans have evolved more complex immune systems. With complex immune systems, they are better equipped to fight the development of the rapidly evolving pathogens.

      The immune system is complex and diverse—necessarily so, to ensure individual survivability. The immune system forms is composed of both innate and adaptive factors that work together to respond to infection with pathogens (Galli et al. 2011). Some of these factors such as ion channels, receptors like the macrophage mannose receptor, among others are inborn. The genome has coded that various forms of these receptors or ion channels on the cells will be physically present. Others like antibodies are acquired throughout an organism's lifetime, some within an acute response to the pathogen, others as part of the immunological memory of infection. Within anthropology, there has been an increased focus on the plasticity and shaping of the immune system, specifically how early life exposures affect downstream health—particularly within the

framework known as life-history theory (Gluckman et al. 2010; Kuzawa and Thayer 2011; Silveira et al. 2007; Thayer and Kuzawa 2014).

Life-history assumes a zero-sum game of energy, where the body has a set amount of energy to allocate between growth, maintenance, and reproduction (McDade 2003). The immune system forms a subsection of maintenance in the body and the energy devoted to it must trade-off with another subsection. An example of this tradeoff is when an individual has too high a pathogen load, then the energy the body devotes to fighting off infections then generally reduces the energy available for growth (McDade 2003). When we consider the overall health development of individuals, we can see how it begins to affect the population as a whole and most cost-effective ways of fighting off infection will be selected for in a population. One of the ways that the body attempts to fight off unknown pathogens is by developing a generalized immune system that is plastic, that is a varied phenotype dependant on the environment.

The amount of plasticity of a specific trait is dependent on genetic variation (Auld et al. 2010). Without genetic variation allowing such a flexibility in a particular trait, then the desired phenotype is not possible. Even so, not every trait can be plastic as plasticity itself may have costs. For example, possessing a smaller range of plasticity for phenotypes is perhaps more cost effective with respect to energy (Auld et al. 2010). According to the phenotype first hypothesis, plasticity is only a short-term response. Should that trait be favorable and pressures for maintaining that phenotype remain for multiple generations, genetic change would follow—selecting for mutations that hard-wire for that trait (Kuzawa and Bragg 2012). With respect to immunity, when a pathogen

has persisted in a population then we see that genetic change or signature of selection for a trait that protects against infection. We are interested in disentangling that underlying genetic variation that influences the immune system and using computational methods to detect signatures of natural selection.

Signatures of Natural Selection

When infectious diseases shape the immune system, they leave distinctive patterns in the population known as signatures of natural selection—thereby altering the genetic diversity found at those loci. Deep down beneath the functions of the immune system, genetics remain a major part of immune-response phenotypes whether contributing to innate or adaptive mechanisms as part of the entanglement of genetics, culture, and the environment (Baker 1997).

Natural selection acts on the variation available to it selecting for adaptive phenotypes. These adaptive phenotypes are shaped by underlying alleles that are functionally different than wild-type alleles. The entire process of selection leaves behind unique signatures of selection that can be teased apart statistically from the rest of the neutrally evolving genome (Fumagalli et al. 2015; Wisser et al. 2008). Consistent with evolutionary theory, these patterns demonstrate that infectious disease, skin color, and metabolic traits show the strongest selection signatures in the genome—traits that manifest themselves physiologically thus improving chances of survival and reproduction (Bamshad and Wooding 2003; Grossman et al. 2013; Jablonski and Chaplin 2000, 2013; Sabeti et al. 2007).

Two examples of selection are positive selection and balancing selection.

Positive selection or directional selection refers to when a particular trait is advantageous, therefore the frequency of that trait and the underlying advantageous alleles increase dramatically. Furthermore, not just the advantageous allele increases, so does the region surrounding that allele as they are physically linked on a particular chromosome. Balancing selection refers to when an advantageous trait is encoded by more than one allele, therefore the patterns keep both in the population at optimum frequencies. Many genes that have been subject to local positive natural selection (e.g. CD40) or balancing natural selection (e.g. CCR5, genes of the MHC complex) are associated with susceptibility to infectious disease (Bamshad et al. 2002; Hughes and Yeager 1998a; Sabeti et al. 2002). Identifying signatures of natural selection in immune response genes is a means to prioritize candidate genes based on evolutionary signatures of selection.

Selection mapping, which is the process of detecting signatures of selection in the genome, can be used to identify risk alleles for the same infectious disease that acted as the selective pressure. Here if a particular disease has persisted throughout various generations and the evolutionary pressures are strong, evolutionary processes would have left a signature around loci. For example, Karlsson et al. (2013) combined selection and association studies to identify risk alleles for cholera infection.

However, selection mapping can also be used to identify risk alleles to modern pathogens that were not the selective agent. Since immune function is also generalized, a protective trait for one infectious disease can also prove to be protective to a newly

introduced infectious disease. In this particular case the previous infectious disease would have been responsible for leaving a signature, even though it is protective against the newer infectious disease.   For example, HIV is a recently evolved human pathogen that has existed for less than 100 years in Europe. Studies on HIV demonstrated that a deletion mutation in the chemokine C-C motif receptor 5 known as the *CCR5-Δ32* allele provides protection from HIV infection (Liu et al. 1996). The region containing the *CCR5-Δ32* deletion demonstrates a pattern of selection in Northern Europeans (Galvani and Novembre 2005). Given that the pathogen has existed for less than 100 years, it would not be responsible for leaving such a strong signature in the population, a previous pathogen would have been responsible for leaving it. HIV just benefited from how generalized immune response is to pathogen response. Through the reconstruction of historical events, scholars hypothesize smallpox to be the leading candidate for the selective agent (Galvani and Slatkin 2003). This example highlights that variation selected in the past affect disease susceptibility to modern infectious diseases.

The work presented in this dissertation takes a complimentary but opposite approach. Instead of retroactively looking for signatures of selection around loci associated with a particular infectious disease, we directly apply signatures of selection to identify susceptibility or resistance loci to a modern pathogen, dengue virus. Our hypothesis that these signatures of selection affect immune response to dengue infection, rests on the previous literature of genetic immune response and how the immune system is generalized and prepared to fight unknown pathogens. In this

approach, we are certain that dengue virus did not shape genetic variation in Mesoamerican populations given that dengue has existed in the Americas at epidemic proportions for only two generations. Through selection mapping, the effects of natural selection can be inferred from patterns of variation and used to: 1) find functional alleles, those alleles that are different from wild-type alleles, 2) determine which alleles might be most useful in association studies, by prioritizing those alleles to the study of current infectious diseases affecting Mesoamerican populations, and 3) identify alleles that contribute to disease susceptibility, by nominating the signatures of selection as candidate loci to study with dengue infection.

From here we can create a list of candidate genes to prioritize based on the map of selection signatures.

Overview of Dissertation

The dissertation is organized into three parts that trace portions of Mesoamerican evolutionary history, in particular with respect to how infectious diseases shape the population. Each chapter seeks to answer one of the following research questions:

1. *Are the detectable regions of archaic introgression in modern Mesoamericans related to immunity?*

To answer this question, we explore a few periods of time in Chapter 2. The first looks at Neanderthal and Denisovan introgression in Indigenous American genomes.

Neanderthal introgression into Eurasian populations occurred 52,000–58,000 years ago, although potentially more pulses of admixture occurred later into Asian populations after their split from European populations (Villanea and Schraiber 2019). Denisovan introgression occurred around 50,000 years ago (Browning et al. 2018b). Ongoing investigations continue to resolve our understanding of archaic introgression. Introgression in Indigenous American genomes would have occurred prior to the peopling of the Americas, occurred as early as 33,000 years ago and certainly prior to 20,000 years ago (Ardelean et al. 2020).

We determined introgression using IBDMix, a novel program for detecting introgression, in a cohort of indigenous Americans from Mexico that are part of the Simons Genome Diversity Project (Chen et al. 2020; Mallick et al. 2016). This statistic was run twice, once using the Atlai Neanderthal genome to represent Neanderthals, and the second using the Denisovan genomes. We pulled out all introgressed segments over 50 kb segment length and had an sLOD score of at least four. We retained segments that overlapped in at least five individuals so that our results were more likely to be representative of being at a higher frequency in the population as described by Chen et al. (2020). These regions were then annotated and run through DAVID 6.8 pathway analysis to determine overrepresented pathways and gene categories (Huang da et al. 2009a, b). Here we determine that both Neanderthals and Denisovan introgression has persisted in genomic regions associated with immune response and are therefore potentially examples of adaptive introgression in Mesoamericans today.

2.   Are we able to detect signatures of natural selection in Mesoamericans given the prevalent history of exposure to infectious diseases left?

For Chapter 3, we used statistical tools aimed at detecting natural selection in a Mesoamerican cohort from various populations in southern Mexico. Our first steps before conducting the scan included addressing the admixture in our Mesoamerican cohort. To do this, we ran global estimates of admixture using fastStructure and refined the regions of admixture by assigning the ancestral population to each haplotype using RFmix, a program for determining localized admixture on each chromosome. Then, we removed all haplotypes assigned to non-indigenous American ancestry and imputed them using a panel of indigenous Americans from throughout the Americas. Our last step was to run the natural selection scan, for which we ran LSBL, XP-EHH, and iHS, statistics to determine natural selection and departure from neutral expectations. We paired each of the haplotype tests (XP-EHH and iHS) with an allele frequency test (LSBL) in order to reduce false positives. These regions were then annotated and run through DAVID 6.8 pathway analysis to determine overrepresented pathways and gene categories (Huang da et al. 2009a, b). Here, we were able to demonstrate various pathways and genes that were under selection, some of which have been previously associated with infectious diseases in other populations, which were known to be prevalent throughout the history of Mesoamericans.

3.   *What risk factors contribute to Dengue-Infection in a modern Mesoamerican*

        population today?

For Chapter 4, we looked at epidemiological data collected in Guatemala during the 2018-2019 rainy season. The participants were of indigenous American ancestry, either self-identifying as *Ladino*, the name used for *Mestizo* in Guatemala, or *Maya*. Here we conducted epidemiological surveys, using the standardized questionnaires from the Ministry of Public Health and Social Assistance of Guatemala for cases. These questionnaires were replicated for a set of controls. From there, we conducted t-tests, chi-squared, calculated odds ratios, and ran a combined logistic regression to determine whether there were certain risk factors between the cases and controls. Here, we found that certain social factors such as waste on property or water security play a significant role in the risk for infection. This study identifies what risk factors need closer attention in order to reduce the risk of dengue transmission.

Significance

This dissertation highlights the importance of immunity and adaptation among Mesoamerican populations whether deep in our evolutionary past, recent colonial contact, or still being shaped by more recent infectious diseases. Anthropological perspectives to applications of the study on modern health and disease place the population within a particular place and time—which provides greater context to modern ailments. Our work incorporating evolutionary perspectives to the study of infectious

diseases will ultimately aid in understanding the pathways associated with DENV-infection, facilitate the identification of different antigens targeting specific biological pathways for treatment and prevention measures such as in-vitro mouse models and ultimately vaccines. Our study will provide the groundwork for participant stratification based on DENV-infection susceptibility in future vaccine trials.

Modern humans have been recently shaped by selection more than previously thought—and immune response is only a small portion of it (Hawks et al. 2007; Sabeti et al. 2006). Furthermore, many diseases could be the result of risk variants being proximate to regions selected for and therefore those regions warrant a closer look, an example is the identification of heart disease loci in previously selected regions (Ko et al. 2014; Zanetti et al. 2015). This dissertation demonstrates how to prioritize outlying regions in studies of modern infectious diseases in Mesoamerican population: 1) regions from archaic populations like Neanderthals or Denisovans, 2) regions demonstrating strong signatures from previous natural selection.

By using approaches that incorporate introgressed regions or loci under recent natural selection as candidate loci for studies, these candidate loci will ultimately contribute to understanding gene function by tying genotype to phenotype. The identification of immune response pathways under selection will aid in understanding human physiology in response to infection. Furthermore, this approach will also aid future endeavors to reconstruct population history and evolution particularly relating to the adaptations of indigenous American communities—specifically Mesoamericans. The context in which this project is formulated will inform researchers investigating the

intersections of shared ancestry and how they are tied to the environmental and political

history in the "local biologies" of Mesoamerica (Lock and Nguyen 2010). For instance,

the dissertation ties in political history and epidemiological data to demonstrate

environmental risk factors which go hand in hand with genetic factors that contribute

overall to the overall response. In more ways than one, historical events (past and

present) that manifest themselves on a cultural or structural level are deeply embedded

in the genomic history of populations contributing to how bodies today respond to

infection with modern pathogens. Our results show the effect of past selection and

introgression on immune response pathways thus bridging the gap between the

medicine and anthropology—and directions to which anthropology can steer the

conversation towards evolutionary framing.

Chapter 2: Neanderthal and Denisovan Introgression Shape Immune Response Genes in the Americas

ABSTRACT

We identified archaic introgression using IBDMix in a cohort of indigenous Americans from Mexico that are part of the Simons Genome Diversity Project (SGDP). We tested for Neanderthal introgression using the Atlai Neanderthal genome and for Denisovan introgression using the Denisovan genomes. We applied a segment length cut off of 50 kb with a sLOD score of at least four. We further limited our findings to regions that were overlapping in at least five individuals from our study population. These regions were annotated and run through DAVID 6.8 pathway analysis to determine overrepresented pathways and gene categories. We found evidence for Neanderthal introgression into the TGF-β signaling pathway with 17 genes identified as potentially adaptively introgressed. We also identified various loci related to cytoplasm and exosomes that were introgressed by Denisovans. These findings demonstrate that events of introgression played a role in shaping Mesoamerican genetic diversity with respect to immune response.

INTRODUCTION

Recent advancements in ancient genomics have unpinned our understanding of population history—particularly with the sequencing of Neanderthal and Denisovan genomes. These developments have demonstrated not only that we are able to

sequence the genomes of extinct Hominini revealing the extent of diversity within and

between ancient hominins, but also that admixture occurred between modern humans

and others from the genus *Homo* (Green et al. 2010; Green et al. 2006; Meyer et al.

2012; Prufer et al. 2014). These events of admixture are known as introgression, and

they have been detected across global human populations. Estimates of Neanderthal

introgression in non-African populations fall between 1-2% (Green et al., 2010; Prufer et

al., 2014), whereas estimates of Denisovan ancestry range from 0.2% in East Asians to

3% in Oceanians (Reich et al., 2010; Meyer et al., 2012; Prufer et al., 2014).

Studies of archaic introgression have highlighted the role of adaptive

introgression in shaping the human genome (Racimo et al. 2017; Racimo et al. 2015).

Adaptive introgression is characterized by genomic regions introduced by archaic

populations such as Neanderthals or Denisovans into the human gene pool at some

point in the past that led to increased fitness for the carrier in a specific environment.

One particularly noteworthy study demonstrated that the Tibetan version of *EPAS1,* a

gene involved in oxygen sensing and metabolism that contributes to high-altitude

adaptation, most likely introgressed into Tibetan populations from Denisovans or

populations closely related to them. The introgressed version of *EPAS1* that is found at

high frequency in Tibetans is one of the first examples of selection acting to increase

the frequency of an introgressed variant that increases the fitness of the carrier (Huerta-

Sanchez et al. 2014). Adaptive introgression also has been identified in immune

response genes including the MHC complex and toll like receptors (TLRs) across

populations from Europe and Asia (Abi-Rached et al. 2011; Dannemann et al. 2016;

Mendez et al. 2012b). These findings highlight the importance of adaptive introgression in shaping modern resistance to infectious diseases. However, it is important to stress that not every introgressed segment has a selective advantage. Rather, several studies have demonstrated that potentially introgressed regions into European and Turkish populations from Neanderthals or Denisovans may be involved in metabolic or autoimmune disorders (Lin et al. 2015; Taskent et al. 2017). Together, these findings highlight the impact of both advantageous and deleterious introgressed variation in shaping modern human phenotypic variation and disease susceptibility.

Despite recent advances in our understanding of the functional impact of introgressed variation, considerable gaps in our knowledge remain concerning the amount and phenotypic impact of adaptive introgression in most human populations. One particularly understudied region with respect to archaic introgression is the Americas. Introgression from both Neanderthals and Denisovans has been documented among populations from the Americas (Chen et al. 2020; Racimo et al. 2017; Sankararaman et al. 2014). However, our understanding of introgression among Indigenous American populations is limited to grouping Indigenous Americans together as a single continental group. Doing so ignores population-specific evolutionary events that have shaped introgression levels on a local scale. Rather, introgression should be determined on a population-by-population basis. Characterizing population-specific introgression is an important step in understanding the impact of introgression in particular populations. Furthermore, results from population-level analyses will highlight the unique histories of populations that followed the peopling of the Americas.

To date, we do not have a clear understanding of the specific introgressed segments that have been maintained in particular populations. As demonstrated in populations from outside of the Americas, introgression has biological consequences on health, whether it is adaptive or predisposes to different conditions. Determining the regions that have been introgressed into Mesoamerican populations will shed light on their unique evolutionary history and contribute to our understanding of gene function in a modern environment.

Given that other studies have pointed to adaptive introgression in immunity, we were particularly interested in understanding the overall genomic profile of immunity for Mesoamericans given their unique history of exposure to infectious diseases. This history ranges from early city-states, to the diseases brought by colonial contact, to modern infectious diseases. Here we ask whether we are able to detect an representation of introgression at immune response genes in Mesoamerican populations. To answer our research question, we joined together and formed a cohort of 11 Indigenous American individuals from Mexico that were sequenced as part of the Simons Genome Diversity Project (SGDP) (Mallick et al. 2016) to identify regions of introgression in Mesoamericans,

METHODS

Archaic Populations: The Altai Neanderthal was obtained from the Max Planck Institute through the Neanderthal Genome Project http://cdna.eva.mpg.de/neandertal/altai/AltaiNeandertal/VCF/ (Prufer et al. 2014). The

Denisova genome also was obtained from the Max Planck Institute through their online

repository http://cdna.eva.mpg.de/denisova/VCF/hg19_1000g/ (Meyer et al. 2012). The

Neanderthal and Denisova map35_50 minimal filters were downloaded and applied to

the dataset https://bioinf.eva.mpg.de/altai_minimal_filters/. The filter used required that

50% of the segments used to align the archaic genome, which had a minimum base

pair length of 35, could not have a matching position to any other portion of the genome,

but could have up to one mismatch (Prufer et al. 2014). These filters are to reduce the

probability of mapping errors influencing the data observed.

Modern Populations: We obtained 11 fully public Indigenous Americans from Mexico

(gVCFs) from the Simons Diversity Genome Project (SGDP) available through the

Seven Bridges Cancer Genomics Cloud (Lau et al. 2017; Mallick et al. 2016). These

include: 2 Maya, 3 Mixe, 2 Mixtec, 2 Pima, and 2 Zapotec. We limited our analysis to

Indigenous Americans from Mexico to be representative of Mesoamerica and thus

reduced the potential effects of population structure that could have been present had

we included other distantly related Indigenous Americans populations in the SGDP. This

sample size is sufficient to determine population based signals of introgression, given

that Chen and colleagues determined the recommended sample size for the IBDmix

method we applied here to be ten individuals (Chen et al. 2020). All markers flagged as

"Low Quality" by the assembly process for the SGDP samples by Mallick et al. (2016)

were dropped. For analysis, we kept only those markers which were successfully

genotyped in a minimum of ten of the 11 individuals. The gVCFs were checked against

the HG19/GRCh37 with decoy sequences obtained at: ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/technical/reference/phase2_reference_assembly_sequence/hs37d5.fa.gz.

Indels and Segmental Duplications: To prepare the files for introgression analysis, all indels and multi-allelic SNVs were removed and trimmed (+/- 5 bp) from both the archaic and modern humans and the following steps were taken as recommended by Chen et al. (2020). First, the Neanderthal and Denisovan indels were queried directly from the gVCF. Second, the indels for the SGDP samples were queried from the BGT files using the FermiKit small variant calls available at: https://github.com/lh3/sgdp-fermi. Third, segmental duplications were obtained from the UCSC genome browser and removed from all humans http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/genomicSuperDups.txt.gz.

Genome Preparation: Files were manipulated and prepared for introgression analysis using VCFtools, BCFtools, and bedtools (Danecek et al. 2011; Li et al. 2009; Quinlan 2014; Quinlan and Hall 2010).

Introgression analysis: We used IBDMix, which is based on the principles of Identity by Descent (IBD), to calculate the Logarithm of the Odds (LOD) score for regions, better represented as a sLOD score, or the average of LOD for a segment (s) (Chen et al. 2020). We ran one archaic individual, whether it was the Denisovan or Neanderthal, in

comparison with the 11 indigenous American individuals from Mexico. CpG regions

were masked in our analysis by preparing a bed file containing all CpG islands from the

UCSC genome browser: http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/.

We considered all segments longer than 50 kb with a sLOD score higher than four as

candidates for introgression. Chen and colleagues previously determined that 50 kb

length paired with a sLOD score of four was powered to detect introgressed segments

in the population while reducing the amount of false positives (Chen et al. 2020). In

pedigree studies, a LOD score of three is considered to be significant as it means that

the odds are one thousand to one (10^3) that the segment was inherited through identity

by descent. In this case, an sLOD score of four means that the odds are ten thousand

to one (10^4) that the segment was inherited by identity by descent.

Annotation of Regions: Windows were annotated for genes using the bedmaps option

from BEDOPS tools (Neph et al. 2012).

RESULTS

We determined introgression using IBDMix (Chen et al. 2020) in a cohort of 11

Indigenous Americans from Mexico obtained from the SGDP. The IBDMix statistical

program was run twice, once using the Atlai Neanderthal genome to represent

Neanderthals and again using the Denisovan genome. To reduce false positives, we

pulled out all introgressed segments that were over 50 kb in length and that had an

sLOD score of at least four as optimized by Chen et al. (2020). We retained segments

that overlapped in at least five individuals so that our results were more likely to

represent regions of adaptive introgression in the population as described by Chen et al.

(2020). Neanderthal and Denisovan introgression were detected on all of the

autosomes (Figure 2.1). On average the frequency of Denisovan introgression in our 11

individuals was higher (1.68% of the genome) than our estimate of Neanderthal

introgression (0.68% of the genome) (Table 2.1).
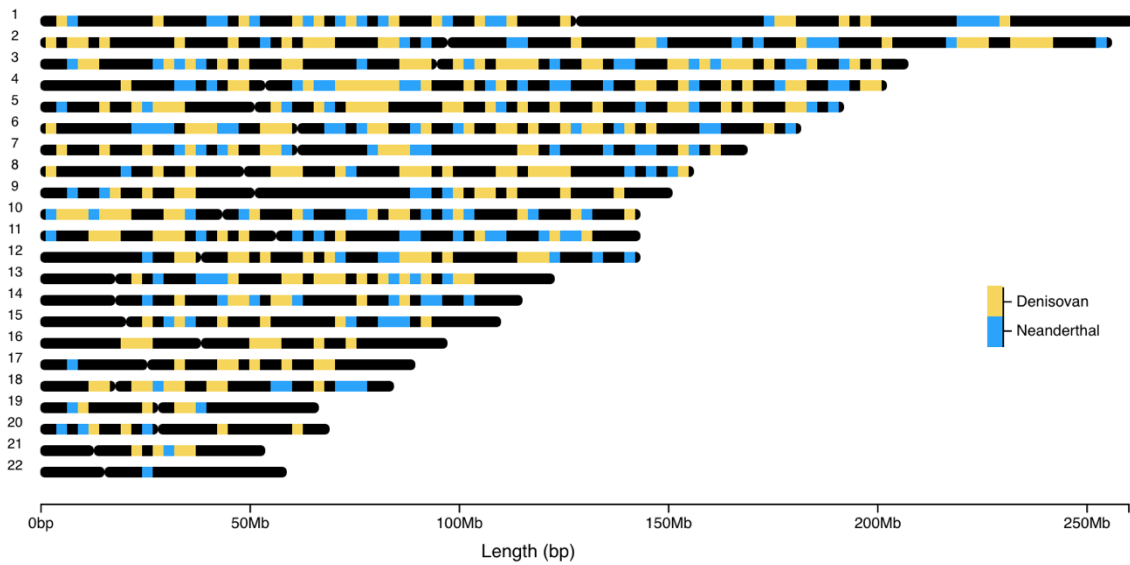
Figure 2.1. Composite map of the proposed regions of introgression across all 11 Indigenous American individuals from Mexico. Neanderthal and Denisovan introgression were detected on all chromosomes. Fragment size differs per individual and the start and end coordinates are staggered and overlapping at different ends. There are no overlapping Neanderthal and Denisovan segments as those were previously removed to control for incomplete lineage sorting.

Table 2.1. Total Amount of Introgression Characteristics for Indigenous Americans from Mexico

	% of Autosomes	St. Dev	Average Mbp	St. Dev
Neanderthal	0.86	0.08	24.85	2.43
Denisovan	1.68	0.46	48.43	13.15

Mbp, Mega base pair
St.Dev, Standard deviation

Segment Lengths

Introgressed segment lengths differed by archaic genome. With the exception of a few outliers, Neanderthal segment lengths were longer than Denisovan segment lengths (Figure 2.2). Neanderthal segment lengths ranged from 50,001 to 1,300,492 bp with a median length of 12,1130, whereas Denisovan segment lengths ranged from 50,188 to 24,982,278 bp with a median length of 107,450 bp (Figure 2.2). We determined that the segment length inherited across individuals was similar for both Neanderthal and Denisovan introgression (Figure 2.3a, b). Individuals inherited similar total amounts of Neanderthal sequences, but the range for inherited Denisovan sequences was much larger (Figure 2.3c). Neanderthal introgression ranged from 20.87 to 29.49 mega base pairs (Mbp). Denisovan ranged from 28.39 to 63.49 Mbp. Given that there were several outlying segment lengths for Denisovan introgression, these were plotted to observe the distribution across segment lengths by sLOD scores (Figure 2.4). We identified 7 outlying segments for Denisovan introgression that ranged from 24.98 Mbp to 23.85 Mbp in length. They ranged in sLOD score from 5.02 to 18.15. These may be artifacts of incomplete lineage sorting as the lengths are incredibly large and outside the normal distribution of values.
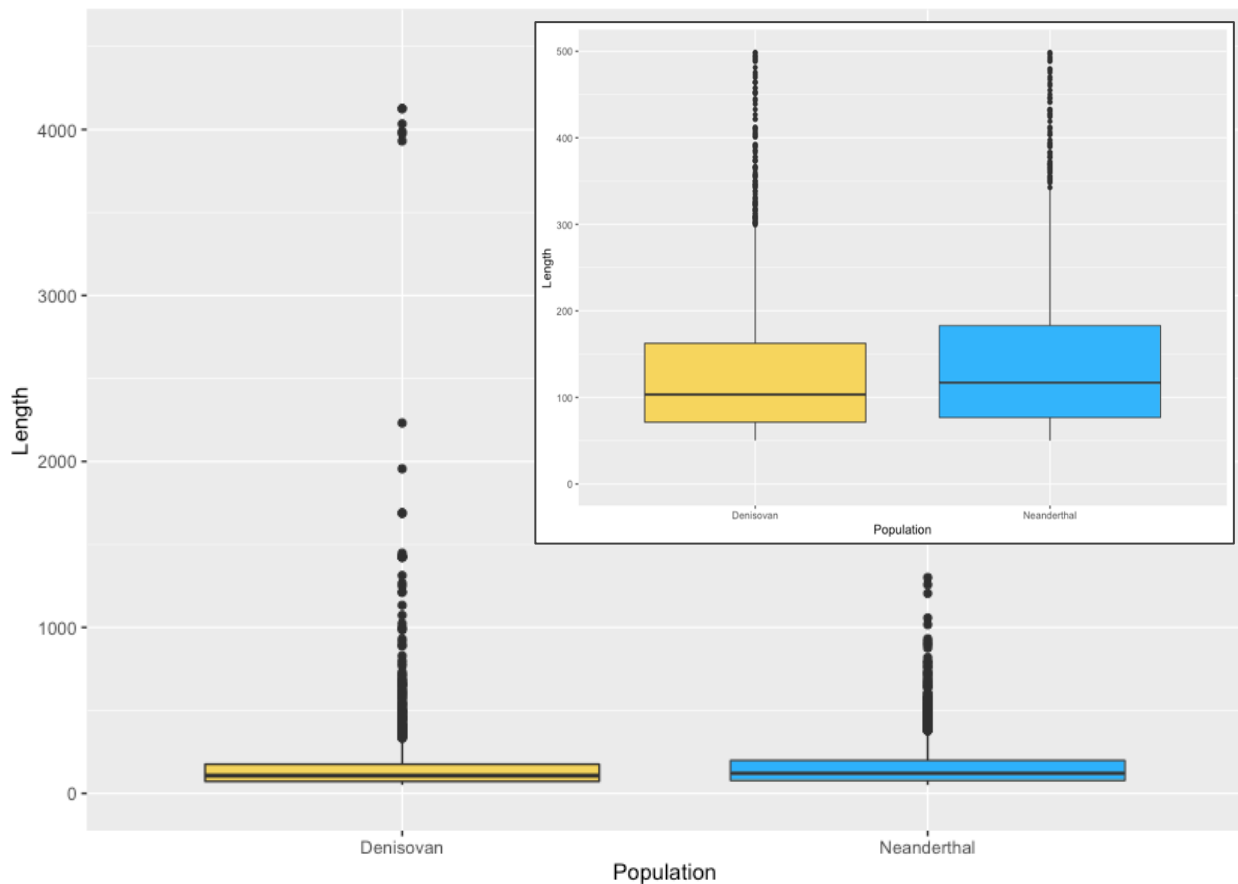
Figure 2.2. Neanderthal and Denisovan fragment lengths in Indigenous Americans from Mesoamerica. The overall distribution of introgressed fragment lengths detected in Indigenous Americans from the SGDP from Denisovan (yellow) and Neanderthal (blue). Fragment length is depicted on the x-axis and is scaled in kb. The inset is a zoomed in view of the overall distribution of introgressed fragment lengths, looking at the y-axis from 0-500. There are 152 total data points missing in the inset from the original figure. The inset highlights that the Neanderthal fragment lengths are on average longer than the Denisovan fragment lengths

A.

B.

26
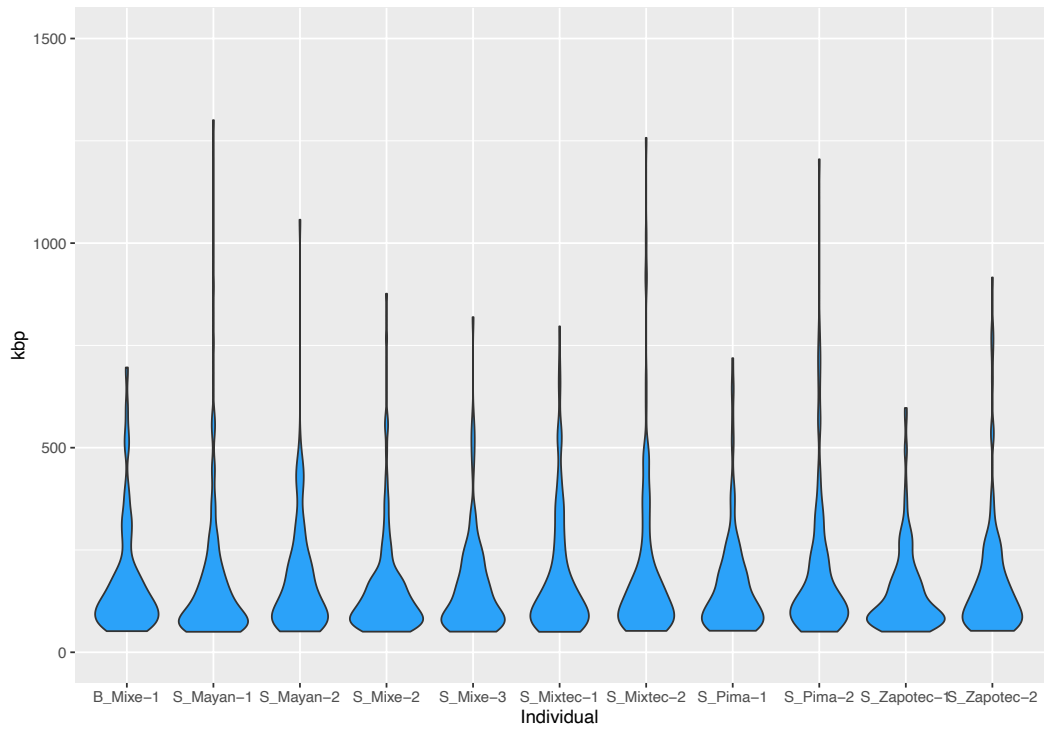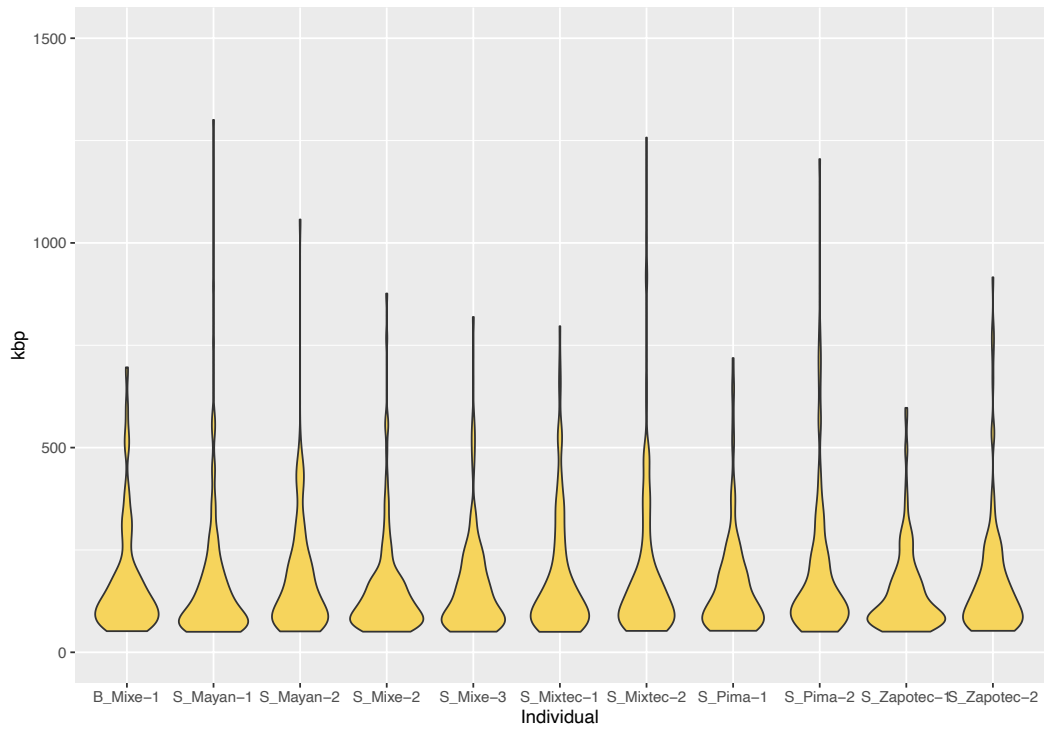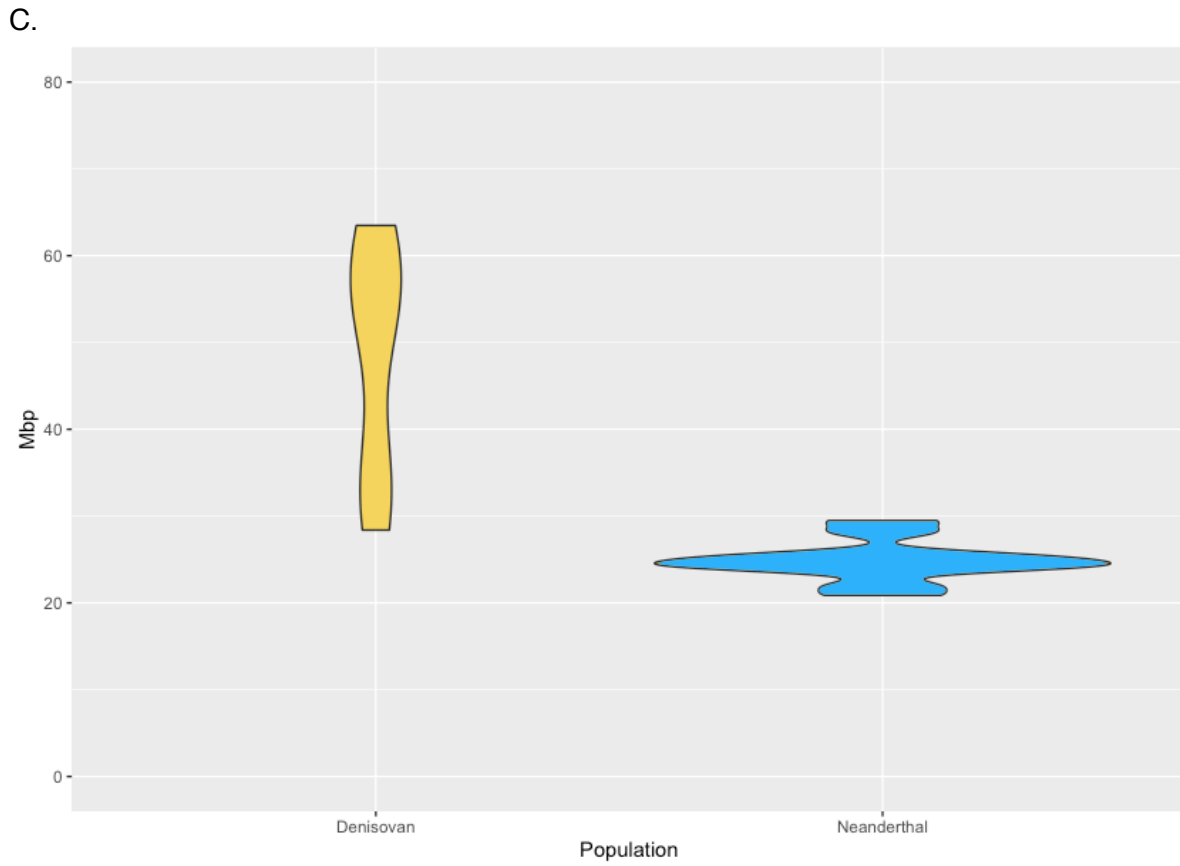
Figure 2.3. Total amount of Neanderthal and Denisovan introgressed sequences across individuals. Violin plot showing the amounts of introgression per individual. The thicker parts of the plot demonstrate a higher density of introgressed segments for that length. A. Neanderthal introgressed segment lengths in each Indigenous American individual in kilo base-pair length. Individual sample IDs are provided on the x-axis. Overall, the density is higher for shorter fragments. B. Denisovan introgressed segment lengths in each Indigenous American individual in kilo base-pair length. Overall, the density is higher for shorter fragments. C. The total amount of introgressed sequences per individual represented in Mbp.

LOD and Segment Length Distribution
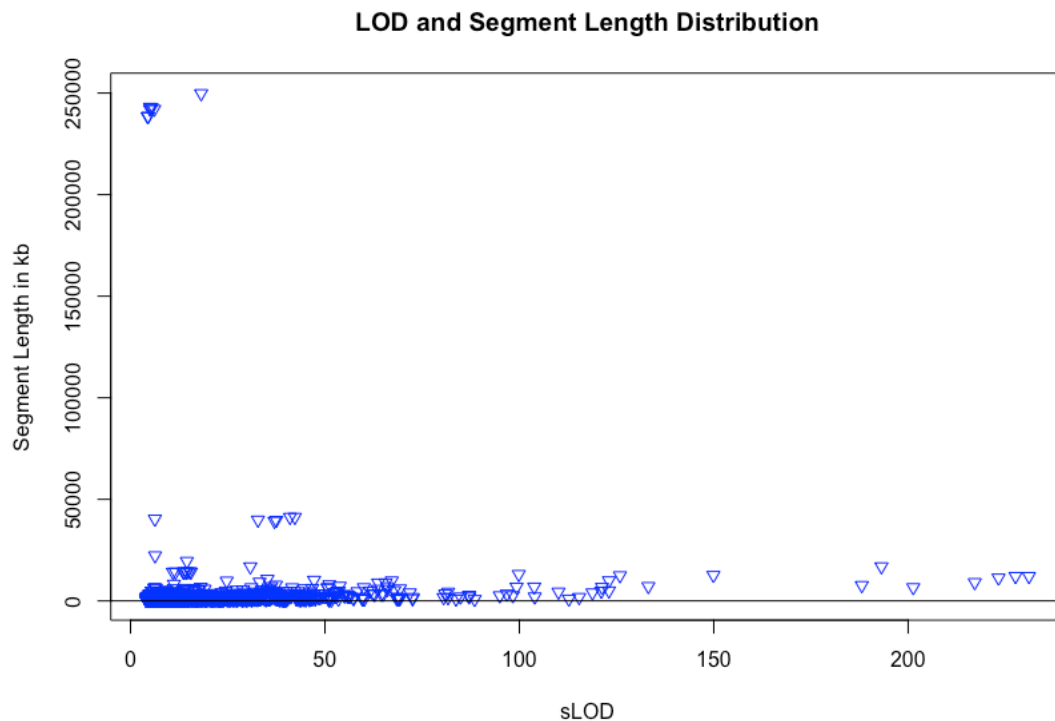
Figure 2.4. Distribution of LOD score and segment length for Denisovan introgression into indigenous Americans from Mexico. The distribution of the sLOD values (x-axis) with segment length (y-axis) for Denisovans. The black line is the line of best fit for the data. Of note are the few outlying segments in the upper right corner. These may be artifacts of incomplete lineage sorting. Higher LOD scores determine that there is a higher probability that the segment was introgressed into Mesoamericans.

Pathway Analysis

Introgressed regions were annotated and run through DAVID 6.8 to determine overrepresented pathways and gene categories (Huang da et al. 2009a, b). Our annotations were limited to genes within the windows of introgression in order increase chances that the introgressed segment will be functional. For Neanderthal introgression, we identified 36 GO term categories (Table 2.2). We identified the following categories of interest among Indigenous Americans with respect to immune function: "response to virus" consisting of 7 genes, "lipid catabolic process" consisting of 6 genes, "insulin

28

secretion" consisting of 4 genes, and the "negative regulation of TGF-β receptor

pathway", consisting of 6 genes. We chose to highlight these categories given the

importance of the TGF-β pathway for both protective immunity as well as its role in

regulating insulin and metabolic disorders (Batlle and Massague 2019; Worthington et

al. 2012; Yadav et al. 2011).

Table 2.2. Gene Ontology grouping and annotations of the Neanderthal introgressed regions.

Go Term	Gene Count	Genes	Fold Enrich.	p-val	Benjamini
receptor localization to synapse	4	NLGN1, NRXN1, DLG2, NETO1	16.22	0.002	0.93
collagen catabolic process	8	MMP10, MMP20, COL19A1, COL13A1, MMP8, MMP27, MMP3, MMP1	4.56	0.002	0.80
negative regulation of cardiac muscle cell proliferation	4	TGFBR2, GJA1, KCNK2, TP73	14.60	0.002	0.72
coenzyme biosynthetic process	3	PANK1, SLC25A16, PPCDC	15.64	0.014	1.00
translesion synthesis	5	RFC3, POLI, DTL, POLD2, UBE2L6	5.07	0.016	1.00
positive regulation of apoptotic process	16	ING5, AIFM2, BCL2A1, ARHGEF16, ARHGEF12, ECT2, ST20, PDCD1, TP73, PDCD2, ITGA6, IRF5, C1QBP, BOK, PDCD5, GADD45A	1.95	0.018	1.00
serotonin metabolic process	3	RNF180, HTR1A, BTBD9	13.69	0.019	0.99
activation of GTPase activity	7	EVI5, GPR65, ARHGEF16, TBC1D1, TBC1D30, ECT2, USP6NL	3.23	0.021	0.99
establishment of protein localization	5	ITGA8, NLGN1, NRXN1, PHLDB2, USH2A	4.56	0.023	0.99
release of cytochrome c from mitochondria	4	GGCT, BOK, BCL2A1, ST20	6.35	0.024	0.99
negative regulation of transforming growth factor beta receptor signaling pathway	6	TGFBR2, ONECUT2, SMAD3, TGFBR3, PRDM16, PBLD	3.42	0.030	0.99

protein targeting to membrane	4	RTP5, PARD3, ATG4B, ATG3	5.62	0.033	0.99
oxidative demethylation	3	ALKBH1, CYP1A2, TET1	9.13	0.041	1.00
locomotory exploration behavior	3	SLC4A10, LSAMP, DPP4	9.13	0.041	1.00
extracellular matrix organization	11	NPHP3, COL9A1, SMOC2, LAMB3, COL19A1, ITGA6, COL13A1, ITGA8, MYF5, CCDC80, KDR	2.05	0.043	1.00
heterophilic cell-cell adhesion via plasma membrane cell adhesion molecules	5	REG3A, NECTIN1, NLGN1, NRXN1, CD200	3.65	0.047	1.00
protein localization to synapse	3	NLGN1, NRXN1, PCLO	8.42	0.048	1.00
biosynthetic process	4	AADAT, SPTLC2, APMAP, PBLD	4.87	0.048	0.99
insulin secretion	4	CPLX3, RAPGEF4, EIF2AK3, PCLO	4.71	0.052	1.00
formate metabolic process	2	MTHFS, MTHFD1L	36.50	0.054	0.99
regulation of cellular localization	2	PARD3, CAMK2D	36.50	0.054	0.99
positive regulation of protein oligomerization	3	TRABD2B, MMP3, MMP1	7.82	0.055	0.99
telomere maintenance via recombination	4	DNA2, RFC3, POLE, POLD2	4.56	0.056	0.99
extracellular matrix disassembly	6	MMP10, MMP20, LAMB3, MMP8, MMP3, MMP1	2.88	0.057	0.99
skeletal muscle tissue regeneration	3	MYF6, EYS, GJA1	7.30	0.062	0.99
DNA strand elongation involved in DNA replication	3	GINS1, RFC3, POLD2	7.30	0.062	0.99
activation of cysteine-type endopeptidase activity involved in apoptotic process	6	NOD1, BOK, SMAD3, EIF2AK3, NLRP1, PDCD2	2.64	0.077	1.00
vesicle fusion	5	C2CD5, SYT14, SYTL2, YKT6, TSNARE1	3.09	0.078	1.00
regulation of cilium assembly	4	EVI5, TBC1D1, ATG3, USP6NL	3.95	0.080	1.00
synaptic vesicle clustering	2	NLGN1, NRXN1	24.34	0.080	1.00
response to virus	7	IFIT3, IFIT1, IFIH1, CREBZF, PRKRA, CXCL12, DCLK1	2.32	0.081	1.00
lipid catabolic process	6	NCEH1, PLCB4, LRCOL1, LIPH, PLCXD2, OC90	2.58	0.083	1.00

		CPLX2, VIPAS39, COL13A1, CCDC169-SOHLH2, ONECUT2, TGFBR2, SPATA31E1, SPATA31C2, SPATA31C1, SOHLH2, NKAPL, PTPRQ, CYLC2, COL19A1, SPATA16, SLC9C1, BRDT,			
cell differentiation	19	CSK, ATOH7	1.50	0.084	1.00
memory	5	ITGA8, LMX1A, KCNK2, NETO1, KCNK10	2.94	0.089	1.00
chloride transport	4	SLC4A10, GABRG3, GABRA5, ANO7	3.65	0.095	1.00

For Denisovan introgression, we identified 24 GO term categories (Table 3). Of these categories, five play a crucial role in immunity. They include "cytoplasm" with 219 genes associated, "cytosol" with 133 genes, "cilium" with 12 genes, "primary cilium" with 8 genes, and "extracellular exosome" with 123 genes associated (Table 3). For example, exosome signaling, and the cytoplasm are part of the innate immune system (Baquir and Hancock 2017; Kagan 2012). Our combined results demonstrate the importance of introgression to immune response loci and point to their adaptive importance for the environments of the evolutionary past through to the present.

Table 2.3. Gene Ontology grouping and annotations of the Denisovan introgressed regions.

GO Term	Gene Count	Genes	Fold Enrich.	p-val	Benjimini
intermediate filament	25	KRTAP15-1, KRT12, KRTAP13-4, KRTAP13-3, KRT10, KRTAP27-1, KRT20, KRT23, KRTAP13-2, KRTAP13-1, KRTAP26-1, KRT24, KRT26, KRT39, KRT25, KRT28, KRT27, KRTAP25-1, KRT40, KRTAP23-1, KRT222, KRTAP19-3, KRTAP19-4, KRTAP19-1, KRTAP19-2	6.20	0.00	0.00

synapse	16	*PPFIA2, GABRA2, CPEB3, CADM2, GABRB2, WASF1, SDK1, PCDH15, GRIN3A, NRCAM, TULP1, PACSIN1, RAPSN, SNTB2, CYFIP1, UNC5C*	2.48	0.00	0.38
cytoplasm	219	*CTHRC1, RP9, CMTR1, PTPN20, TECRL, RBM8A, VPS4A, EIF2B3, LRRC7, SYK, NAA25, KRT10, OPTN, CTNNA2, GRB10, DCAF6, RFC1, ROR1, PDE4DIP, HAS3, DNAH12, ASB14, KRT20, TIPRL, KRT24, DTD1, KRT25, KRT28, TMEM50A, KRT27, RCHY1, PABPC1, OLFM2, DHX8, RANBP17, SMAD5, DUSP22, PCDH15, RPS8, PTPN11, PXDNL, SRSF3, RNF8, CBLB, CCT4, BTBD10, FAM161A, MAP7, RGSL1, RIT2, GRK2, BTBD11, DPP3, CREBRF, MTRNR2L5, LXN, CPEB3, NAAA, NFKB1, CFAP36, RSPH4A, NSMCE3, EEF2K, FBXL12, DPP8, HHIP, ADCY10, TUBB1, LOC653513, ALS2CL, NIP7, PRSS50, KCTD20, CDR2, MAST2, BNIP1, ADM, FANCD2, PLEKHH2, NUP205, MAP3K19, SIAH1, ARL4D, CRYBA1, PPFIA2, NBPF20, CEP128, NUFIP2, MLF1, WDR19, ANKS6, WDR11, BCAS3, AGL, LRSAM1, NBPF10, NASP, SRGAP2C, APPL1, SVIL, RBM14, NBPF19, MPHOSPH8, RBM14-RBM4, UBXN4, PGK2, ALDH1L1, STK38, NIT2, RBM4, IDE, LRRC4C, MCM10, WNT1, PACSIN1, KLHL5, PACSIN3, DNAJC9, CPOX, LIX1L, TWF1, MADD, NUDT3, ARNTL, PNPLA1, TNKS2, MRE11, PSMA2, ANKRD13D, FAM72B, FAM72C, FAM72D, SNTG1, HSPB1, LRRK2, FILIP1L, ARFGAP2, MRPS16, HACD3, PKHD1, SSH3, PNPT1, SNX4, WT1, COMMD10, IRAK4, ZNF322, EIF3E, NPM1, HECTD4, RPL10A, RHEBL1, ZMYM2, ZMYM5, BRCA1, PLK3, PSMC3, SPTBN2, MYLK, PC, HECW1, TACR3, FIGNL1, GABBR2, LARP1, AKR1C3, AKR1C2, DCAF13, FRMD6, AKR1C4, DDX60, HEY2, ARHGAP1, SNTB2, NBPF4, CNTLN, NBPF6, NBPF9, NBPF8, PPIAL4C, IRAK2, PPIAL4A, PPIAL4G, STXBP4, DDN, CTH, AKR1C8P, ERN1, KPNA5, COMMD1, EMC2,*	1.18	0.00	0.36

		UBAP2, SLC7A6OS, BCLAF1, CDH1, CDH3, DNAH5, UBE2R2, TUBGCP5, ROPN1B, ECD, PPP1R12A, PNPO, CCS, BDH2, TNPO1, TERF2, GEMIN5, PPP4R3B, TXNIP, IPP, CENPE, UBL5, RGS20, ACTRT2, RSRC1, PSPC1, IRF4, URI1			
collagen trimer	10	*MARCO, C1QTNF9B, CTHRC1, COL21A1, P4HA1, SFTPA2, SFTPD, COL12A1, SFTPA1, ADIPOQ*	3.05	0.01	0.46
extracellular exosome	123	*RARRES1, ALDH1L1, CHMP4C, NIT2, GABRB2, RAB3GAP1, CUL3, WNT1, PACSIN3, TGFBI, ST3GAL6, RPL26L1, DYNC2H1, COL12A1, VPS4A, RPL12, ITFG1, CRISP3, NUDT3, CHTF8, ERP29, KRT12, ABCC11, KRT10, ACTN3, CFAP70, PSMA2, NPC1, RFC1, NBR1, TMEM106A, F2, HSPB1, RPS10, LRRK2, MDH2, FAM171A1, GC, PKHD1, ACP2, RIMS2, EPHB1, KRT24, ALCAM, KRT26, KRT25, KRT28, UBE2D3, KRT27, EIF3E, PABPC1, SCARB2, RPL10A, ACMSD, PCDH15, RPS8, SERPINI2, CCT4, CYFIP1, MYLK, UQCRC2, DPP3, LXN, NAAA, FIGNL1, PRKAG1, AKR1C3, ART3, AKR1C4, ARHGAP1, SNTB2, TUBB1, TUBA1A, CNTLN, TUBA1B, AKR1C1, PPIAL4C, KNG1, DDC, PPIAL4A, CTSZ, SDK1, ATP6V1H, STXBP4, ALK, VAT1, C1QTNF9B, CTH, CNTN1, COMMD1, CTSH, PROS1, CTSF, PPFIA2, GPR180, FKBP5, CDH1, SPHKAP, ANXA7, AKR1A1, PNPO, BRK1, SUCNR1,*	1.23	0.01	0.60

		BDH2, COL8A1, TNPO1, DNM3, GSTA3, PLA2G15, TMC5, FBN1, RPL27, SCRN2, APPL1, MAN1C1, ADIPOQ, PEX11B, PLSCR1, YWHAG, EYS, MUC19, MPHOSPH8, PGK2			
cilium	12	*CFAP70, TULP1, WDR19, DNAH12, CCP110, PKHD1, GPR161, BBS9, ANKS6, DYNC2LI1, ADCY10, ARL13B*	2.23	0.02	0.77
primary cilium	8	*WDR19, PKHD1, GPR161, DYNC2LI1, DYNC2H1, GRK2, TNPO1, ARL13B*	2.88	0.02	0.74
focal adhesion	23	*TWF1, HACD3, WASF1, RPL27, CDH1, ACTN3, HMGA1, RPS8, ALCAM, YWHAG, DCAF6, RPL6, SVIL, NPM1, PPP1R12A, SNTB2, CYFIP1, HSPB1, RPS10, RPL10A, PABPC1, RPL12, SCARB2*	1.65	0.02	0.73
lamellar body	3	*SFTPA2, SFTPD, SFTPA1*	10.51	0.03	0.79
SCAR complex	3	*WASF1, CYFIP1, BRK1*	10.51	0.03	0.79
lysosomal lumen	8	*GC, GPC5, NAAA, GPC6, ACP2, SCARB2, MANBA, CTSF*	2.64	0.03	0.76
cytoplasmic microtubule	6	*KIF2C, RGS20, DYNC2LI1, BCAS3, TUBA1A, TUBA1B*	3.30	0.03	0.76
ubiquitin ligase complex	9	*MED6, RNF8, FBXL12, ASB14, RCHY1, RNF20, BRCA1, MPHOSPH8, DCUN1D5*	2.34	0.04	0.77
postsynapse	4	*GABRA2, RAB3GAP1, NPAS4, LRRK2*	5.10	0.04	0.77

motile cilium	7	*WDR19, RSPH4A, ROPN1B, DYNC2LI1, MKKS, ADCY10, ARL13B*	2.58	0.05	0.83
intracellular ribonucleoprotein complex	10	*MRPL10, NPM1, SNRPD1, RPL27, PABPC1, RBM14, LRRK2, BRCA1, RBM14-RBM4, RPS8*	2.06	0.05	0.81
cell junction	24	*SLC8A3, SVOP, GABRA2, GABRB2, CPEB3, DLGAP2, SYT12, SDK1, CDH1, GABBR2, LRRC4C, GRIN3A, RIMS2, TULP1, DCAF13, PACSIN1, RAPSN, DDB2, SPTBN2, KCTD16, UNC5C, LRRK2, OLFM2, LRRC7*	1.47	0.06	0.84
postsynaptic density	12	*NETO2, DNM3, GRM3, RGS20, DAB1, DLGAP2, CPEB3, EEF2K, GRIN3A, LRP4, LRRC7, CTNNA2*	1.83	0.07	0.83
axoneme	7	*BBS1, RSPH4A, DYNC2LI1, DYNC2H1, TCTEX1D4, DNAH5, ARL13B*	2.45	0.07	0.81
late endosome membrane	8	*NPC1, ANXA8, CHMP4C, SNX16, VPS4A, ADAM30, OSBPL11, SCARB2*	2.22	0.07	0.81
cytosol	133	*RNASEL, CNOT8, ALDH1L1, CHMP4C, GABRB2, DYNC2LI1, IDE, SNRPD1, CNOT4, CUL3, MAP3K5, PACSIN1, DAB1, RBM8A, RPL26L1, DYNC2H1, VPS4A, PELI3, RPL12, EIF2B3, SYK, ROCK1, MADD, NUDT3, ACTN3, OPTN, MRE11, TNKS2, CTNNA2, PSMA2, GRB10, SULT1B1, RIOK3, RASGRF1, NBR1, HSPB1, RPS10, LRRK2, GC, ARFGAP2, ERBB4, BBS9, CEP164, EPHB1, IFI35, PIN1, IRAK4, UBE2D3, RPL6, EIF3E, NPM1, PABPC1, RPL10A, ZMYM2, AGBL1, CKAP5, OSBPL7, SMAD5, ACMSD, RPS8, PTPN11, TULP1, CBLB, CDKN1A, CCT4, PSMC3, SPTBN2, CYFIP1, GRK2, MYLK, PC, HECW1, PHKB, PRKAG1, MYBPC3, NFKB1, AKR1C3, KIF2C, AKR1C4, PRMT7, ARHGAP1, EEF2K, MKKS, SULT1E1, RHOD, ADCY10, TUBA1A, CABLES1, AKR1C1, IRAK2, DDC, ATP6V1H, PDE10A, HMGA1, LDLRAP1, BBS1, CTH, EIF4A2, KPNA5, SIAH1, CTSH, PPFIA2, SNX16, ATG13, PRDM16, DCT, ANXA7, ANXA8, TUBGCP5, AKR1A1, PPP3CB, PPP1R12A, PNPO, CCS, BRK1, BDH2, TNPO1, AGL, GEMIN5, TXNIP, GSTA3,*	1.12	0.08	0.86

		CCP110, VHL, RPL27, POLR3GL, CENPE, APPL1, POLR3E, PLSCR1, YWHAG, SP6, ALOX5, IRF4			
ciliary membrane	5	*BBS1, GPR161, BBS9, HHIP, ARL13B*	2.98	0.09	0.85
large ribosomal subunit	3	*RPL26L1, RPL10A, MRPL32*	6.01	0.09	0.84
clathrin-coated endocytic vesicle	3	*SFTPA2, SFTPD, SFTPA1*	5.61	0.10	0.87

DISCUSSION

Determining archaic introgression, or the contribution of genetic material from other *Homo* subspecies to our own, can prove useful for identifying population specific genomic regions that effect human health and disease. In fact, as has been shown in several studies of immunity and autoimmune disorders, identifying adaptively introgressed segments can be used as a tool to prioritize loci with functional phenotypic effects (Dannemann et al. 2016; Lin et al. 2015; Mendez et al. 2012a; Pajic et al. 2016;

36

Taskent et al. 2017). These loci could be adaptive at the present time or were adaptive

at some point in the past. If the latter, changing environments would have rendered their

former adaptive significance neutral. These regions can be better understood and

classified as adaptive when gene function is determined and tied to introgressed alleles.

Similar to approaches identifying outlying genes subject to past natural selection, these

outlying introgressed loci could be nominated as candidate loci for the study of

infectious diseases susceptibility or autoimmune disorders given that they are high

frequency haplotypes that are potentially different from haplotypes present in other

populations.

     Our study demonstrated the importance of the persistence of genes from

Neanderthal and Denisovan populations that are related to immune response. Here we

propose the TGF-β signaling pathway introgressed from Neanderthal populations as an

important mediator of immunity and metabolic pathways. Furthermore, we propose that

innate immunity via cytoplasm related mechanisms and exosome signaling has been

shaped by introgressed segments of Denisovan origins.

     For the TGF-β signaling pathway, we found various categories associated with

this particular pathway. For instance, in our pathway analysis we identified a group of

genes associated with the negative regulation of this pathway. Additionally, we identified

genes falling into categories of metabolism and insulin sensitivity as well as viral

response. To date, the TGF-β signaling pathway has been studied broadly in Mexican

and Mexican American populations with heart disease, metabolic disorders, and

cancers (Fragoso et al. 2012; Martínez-Campos et al. 2019; Zavala et al. 2013). Further

studies on immune response and the TGF-β signaling pathway have focused primarily

on Hepatitis C infection in Mexico (Fierro et al. 2017). No research has been devoted to

understanding the role of this pathway in T-cell regulated immune response in Mexican

or Mexican American populations despite the fact that this pathway has been largely

documented to regulate T-cells with respect to infectious diseases like Chagas and HIV

in other global populations (Reed 1999; Worthington et al. 2012). Clarifying the

importance of this pathway in related indigenous populations is important given that

Indigenous Americans from Mexico faced pressures of infections from Chagas in their

city states and still continue to have exposure to Chagas. Understanding outlying genes

due to adaptive introgression may potentially improve treatments and infection

outcomes.

Our DAVID pathway analysis of Denisovan introgression revealed the GO

Terms "cytoplasm" and "exosome and cytosol" as the categories with the highest

number of genes. Cytoplasm had over 200 genes and exosome and cytosol had over

100 genes showing evidence of Denisovan introgression. Both of these GO Terms are

important regulators of the immune response. The cytoplasm is involved in innate

immunity. It contains proteins that scrutinize incoming pathogens, identify them as either

viral or bacterial, and classify them as a threat (Kagan 2012). Exosomes function as

part of the immune system by transporting, merging, or binding to receptors, and

signaling between cells (Baquir and Hancock 2017). They are primarily associated with

transporting MHC Class I or II peptides (Robbins and Morelli 2014). Given the

evolutionary importance of the MHC complex in response to infectious diseases and the

prolific exposure of Indigenous Americans to infectious diseases during the Colonial

era, understanding the genetic variation that was introgressed and how it influences the

innate immune system, would aid in better understanding and treating infections in

certain populations.

There are various methods to detect events of introgression. They include the

four population statistic test, the D statistic (ABBA-BABA), the S* Statistic, and most

recently IBDMix (Chen et al. 2020; Green et al. 2010; Plagnol and Wall 2006; Reich et

al. 2009; Vernot and Akey 2014). The D-statistic is the most commonly used measure of

introgression. However, when this statistic is applied to humans, it makes the

assumption that African populations contain no Neanderthal or Denisovan introgression.

Furthermore, it is performed on individual genomes rather than population-level data.

We employed the newest statistic, IBDMix, to determine regions of introgression in a

population of indigenous Americans. This method does not require an outgroup such as

statistics like the D statistic (Chen et al. 2020), thus allowing for the potential to discover

novel regions of introgression previously masked by an outgroup.

In our indigenous American cohort from Mexico, we found a smaller amount of

Neanderthal introgression, 24.85 Mbp, than previously reported by other studies for

Indigenous Americans (Figure 2.2, Table 2.1). For instance, Chen et al. (2020) reported

that there was on average ~50 Mbp per individual among 1000 Genomes Americas

population. This discrepancy could be explained in several ways. First, it may be the

result of an artifact of modern-day admixture in the 1000 Genomes American

populations. Second, it could be the result of variable levels of introgression across

different populations as demonstrated by Racimo et al. (2017). In that study, Peruvians from Lima were identified as having the highest amount of introgressed sequences in the Americas. Therefore, since estimates are provided for the continent, the averages for all of the Americas may be inflated by differing levels of detectable introgression across various indigenous populations. Third, another potential explanation for our smaller than average Neanderthal introgression is population substructure within the Indigenous Mexican population studied here. In fact, Mexican indigenous groups exhibit genetic differentiation and are characterized by small effective population sizes (Moreno-Estrada et al. 2014). For a region to be considered introgressed, it had to fulfill the threshold of being present in five individuals from the study population. This criterion may have affected our introgression estimate. This could be addressed in future studies that estimate introgression in a single, more homogenous population.

Our Denisovan estimate for introgression (1.68%) was higher than the previous estimates for the Americas that placed it at 0.05% of autosomes (Sankararaman et al. 2016). We did expect to detect novel regions with more refined techniques of introgression such as IBDmix used here. However, we suspect that our higher percentage is an artifact of the parameters set forth in IBDmix, which have not been optimized for Denisovan populations. Given that introgression estimates for Denisovans are hypothesized to have occurred before Neanderthal admixture, we would expect shorter segments on average than Neanderthals. The pattern holds mostly true (Figure 2.2). However, we identified several outlying Denisovan segments that were much longer than the average distribution (Figure 2.4). Therefore, these segments most likely

would be false positives and could be artifacts of incomplete lineage sorting (ILS). To further optimize the results, we suggest the following: 1) Apply a stricter threshold for sLOD scores to be greater than four. Just a small shift towards using a stricter sLOD score such as five or six would remove the outlying points. 2) Remove any African-Denisovan introgression results to account for ILS, at least initially until results have been verified by D-Statistics. 3) Optimize the results using an already verified method, such as D-statistics in order to optimize the fragment length and LOD scores to get a consensus of introgressed regions. Implementation of #3 would obviate the need for #2, as one may be masking actual introgressed regions introduced into the African population as a result of back-migrations from other populations (Chen et al. 2020). These suggestions are outside of the scope of this project and will require it to be optimized not only in one population but across multiple global populations. I plan to explore these further in the future with an expanded dataset.

Our study adds to the growing number of studies on archaic introgression and demonstrates that there are varying degrees of introgression in genes related to immunity and metabolism in an Indigenous American population from Mexico. Furthermore, we have created a catalogue of candidate genes with evidence of adaptive introgression. This catalog will be useful in future studies whose aim is to characterize the functional impact of adaptive introgression, and it may be useful in efforts to characterize gene function. Finally, our results shed light on the potential impact of adaptive introgression in the evolutionary history of Indigenous Americans.

We have identified candidate regions of adaptive introgression among Indigenous Americans from Mexico, thus laying the groundwork for determining the function of shared alleles between archaic hominins and modern Mesoamerican populations. Characterizing the function of specific haplotypes will allow us to create better targeted approaches to understanding infectious diseases response, immunity, cancer, and metabolic disorders among others. The goal is not to stop at identifying these alleles, but at tying genotype to phenotype in the near future. Refining the regions of introgression and identifying their functional effects will allow us to reconstruct population history to understand the present and will lead to clearer understandings of the molecular mechanisms underlying phenotypic diversity. Furthermore, functional analysis of introgressed variants will determine how they impact health today.

Chapter 3: Immune Response Genes Under Selection in a Mesoamerican Cohort

ABSTRACT

Infectious diseases are among the strongest selective pressures on the human

genome; and past selective events influence host susceptibility and resistance to

modern diseases. Therefore, searching for signatures of natural selection in genes

related to immune function is a particularly attractive strategy to identify host factors for

infectious disease. To characterize host risk factors in a Mesoamerican population, we

interrogated 525,467 SNPs assayed using the Affymetrix 6.0 genotyping array for

signatures of natural selection in immune response genes. The Mesoamerican

populations included: 25 Maya and 14 Mesoamericans from Mexico. Additionally, we

used available data from 60 Europeans of northern European ancestry and 90 East

Asians from China and Japan. We applied three statistical tests to identify signatures of

natural selection: locus specific branch length (LSBL), the Cross-Population Extended

Haplotype Homozygosity (XP-EHH), and the Integrated Haplotype Score (iHS). Each of

the Haplotype tests (XP-EHH and iHS) were paired with an allele frequency test and

significance was determined at the 1% level. For the paired analyses, we identified 55

statistically significant windows for XP-EHH/LSBL and 70 statistically significant

windows for iHS/LSBL. Among our top immune response loci, we find significant

associations with the Major Histocompatibility Complex (MHC) and the peroxisome

proliferator-activated receptor gamma (PPAR-γ) signaling pathway. These findings

illustrate that Mesoamerican populations immunity has been shaped by exposures to

infectious disease.

INTRODUCTION

In the current genomic era, we have begun to disentangle population

evolutionary history—learning how the environment shapes genetic histories. Despite

advances in understanding evolutionary histories of worldwide populations, such as

migration routes, admixture, and natural selection, there remains a dearth in knowledge

concerning Indigenous Americans. Based on what is known to date of population history

and historical records, we hold the assumption that two major events contributed to the

genomic architecture of Indigenous Americans: 1) migration and subsequent settling of

populations to the Americas 20,000 years ago, and 2) colonial contact starting in the

15th century.

After crossing over into the American continents, Indigenous Americans

encountered new environments ranging from tropics to tundra and from low to high

altitude. These ecologies shaped their genomes through adaptations to local selective

pressures (Reich et al. 2012; Wang et al. 2007). Several genomic studies have

demonstrated local adaptation among various populations from the Americas including

high-altitude adaptation among Andean highlanders, dietary and cold adaptations

among the Inuit and descendant populations, adaptation to physical endurance in

Mexico, and immune adaptation among various Indigenous American groups (Amorim et al. 2017; Avila-Arcos et al. 2020; Bigham et al. 2010; Reynolds et al. 2019).

Given the urbanization and high population density of Mesoamericans during pre-colonial times, infectious disease was likely a strong evolutionary pressure in this region (Mummert et al. 2011; Smith 2016). However, the diversity of infectious diseases present in the Americas prior to colonial contact differed from the infectious diseases of Afro-Eurasia and Oceania. This was an outcome of both geographic isolation and differences in zoonotic biota (e.g. insects and fauna) that served as vectors. Therefore, selection likely did not act on the genomes of Indigenous Americans for variants that protected them from the "Old World" infectious diseases. Rather, pre-colonial populations in the Americas adapted to diseases that were locally prevalent. Examples of diseases confirmed to have been present in the Americas prior to European contact, such as Chagas, tuberculosis, syphilis, and hepatitis, would therefore be responsible for leaving signatures of selection in the population prior to colonial contact (Bos et al. 2014; Klaus et al. 2010; Merbs 1992; Steverding 2014).

Indigenous Americans' isolation from Old World diseases ended with devastating effects from European colonial contact in the late 15th century. Historical records indicate that infectious disease killed upwards of 90% of Mesoamerican communities (Feldman 1999; Leon-Portilla 2011; Lockhart 1992; Restall et al. 2005). Mitochondrial DNA data corroborate these historical accounts by demonstrating a population bottleneck 500 years ago coincident with European contact (O'Fallon and Fehren-Schmitz 2011). Accordingly, the evolutionary pressures for survival were strong.

However, our knowledge of Indigenous American genetic variation in general and at loci

related to infectious disease and immune response in particular is limited. To date, only

a few studies have identified immune response genes under selection in Indigenous

American populations (Avila-Arcos et al. 2020; Lindo et al. 2016; Mychaleckyj et al.

2017; Reynolds et al. 2019). Together, these studies demonstrate that the combination

of precolonial and postcolonial history shaped Indigenous Americans genomes,

particularly in response to infectious diseases.

Here, we interrogated the genomes of Indigenous Mesoamericans for evidence

of natural selection at immune response loci. We expected to find evidence of selection

as a result of the historical burden of infectious diseases in Mesoamerican populations.

Furthermore, we hypothesized that colonial contact would have left strong signatures of

natural selection in their genomes given the high mortality rate from novel infectious

diseases (e.g. variola virus that causes smallpox) during the colonial era. In particular,

we expected to identify a high proportion of immune response genes and pathways

under natural selection given the history of infectious disease exposure among

Mesoamericans across time. These genes under selection can provide valuable insight

into population history and also serve as candidate loci for studying localized, biological

responses to modern infectious and autoimmune disease.

METHODS

Populations: Our Mesoamerican cohort included a total of 39 individuals representing

the following populations: Twenty-five Maya from the Yucatan Peninsula of Mexico, two

Nahua, seven Mixtec, and five Tlapanec speakers from Guerrero, Mexico previously

described in (Bigham et al. 2010). We obtained publicly available data from the 1000

Genomes Project for the following control populations: Sixty Europeans of Northern and

Western European ancestry (CEU), ninety East Asians from Beijing, China (CHB) and

Tokyo, Japan (JPT), and ninety Yorubans from Ibadan, Nigeria (YRI) (International

HapMap 2003).

Genome-wide SNP data: All samples were previously genotyped using the Affymetrix

Genome-Wide Human SNP Array 6.0 containing 906,600 SNPs (Bigham et al., 2010).

We analyzed autosomal SNPs with call rates >95% and that had ancestral allele

information available from the 1000 Genomes Project. The X and Y-chromosome as

well as mitochondrial DNA (mtDNA) SNPs were excluded from our analyses as we

chose to focus on the autosomes. No SNPs were removed based on departure from

Hardy-Weinberg equilibrium (HWE) as this could potentially remove SNPs under

selection that would mimic HWE departures. After QC, we carried out statistical analysis

using 525,467 autosomal SNPs.

Phasing and File Manipulation: All files were haplotype-phased using Beagle 5.1,

processed using PLINK 1.9/2.0, and manipulated using VCFtools, BCFtools, and

bedtools (Browning et al. 2018a; Browning and Browning 2007; Chang et al. 2015;

Danecek et al. 2011; Li et al. 2009; Purcell et al. 2007; Quinlan 2014; Quinlan and Hall

2010). 1000 Genomes Project phase 3 populations were used for phasing the SNP data (Genomes Project et al. 2015).

Relatedness: We calculated relatedness using kinship coefficients as estimated with the Kinship-based INference for Genome-wide association studies (KING) (Manichaikul et al. 2010). We removed six individuals from our dataset that were first, second, and third-degree relatives, leaving us with a sample size of n = 33 individuals.

Admixture analysis: We obtained twenty-two Indigenous American genomes from the Simons Diversity Genome Project (SGDP) to use as a reference panel for our admixture and imputation analysis in our Mesoamerican cohort. This panel was comprised of one Chane, three Karitiana, two Maya, three Mixe, two Mixtec, two Piapoco, two Pima, three Quechua, two Surui, and two Zapotec (Mallick et al. 2016). We used fastSTRUCTURE (K = 4) to obtain a global estimate of admixture for each Mesoamerican individual (n = 39) (Pritchard et al. 2000; Raj et al. 2014). We had previously removed six related individuals, leaving us with n = 33. We then proceeded to remove two individuals whose levels of non-indigenous admixture were estimated in the 30s as their admixture levels were outliers to the rest of our dataset. This also ensured minimal imputation of missing genotypes in our dataset. This left us with a sample size of n = 31 individuals, with only four samples with admixture levels in the 20s. To further determine chromosomal segment ancestry, we utilized RFMix 2 (Maples et al. 2013). For RFMix 2, we assumed approximately 20 generations since initial admixture, which corresponds to admixture

on Spanish encomiendas in the Yucatan (Machuca Gallegos 2016). Haplotypes

demonstrating admixture were extracted using bedtools, set as missing, and

imputed/phased in Beagle 5.1 using the Indigenous American population as a reference

panel from the SGDP (Browning et al. 2018a; Browning and Browning 2007; Mallick et

al. 2016; Quinlan 2014; Quinlan and Hall 2010).

Ancestral alleles: Ancestral alleles were queried from the 1000 Genomes Project phase

3 VCF files using BCFtools (Li et al. 2009). VCF files were recoded using PLINK 2.0 in

order to preserve phasing information (Chang et al. 2015; Purcell et al. 2007).

Identifying genomic signatures of selection among Indigenous Americans: We

employed three statistics to identify regions in the genome showing statistical evidence

of natural selection. They included: 1) Locus specific branch length (LSBL) (Shriver et

al. 2004). 2) Cross-population extended haplotype homozygosity (XP-EHH) (Pickrell et

al. 2009; Sabeti et al. 2007), and 3) Integrated haplotype score (iHS) (Voight et al.

2006).

      LSBL compared Mesoamericans against European Americans and East Asians.

F_{st} values were computed for each SNP using Weir-Cockerham's equation (Akey 2009;

Bigham et al. 2010; Shriver et al. 2004; Weir and Cockerham 1984). Statistical

significance was determined using an empirical distribution. $P_E(x)$ = (number of

loci>x)/(total number loci) using a significance threshold of $\alpha = 0.01$ (Akey et al. 2002).

LSBL results were then aggregated into 200 kilobase pair windows, that matched with the XP-EHH and iHS coordinates.

XP-EHH was calculated in Selscan (Szpiech and Hernandez 2014). We standardized this statistic based on allele frequency bins using the norm function of Selscan. We then used non-overlapping windows of 200 kilobase pairs (kb). We identified regions with the longest haplotypes reaching significance thresholds of α = 0.01. We compared Mesoamericans to East Asians to look specifically for haplotypes present in Mesoamerican populations that arose after their split from Asian populations.

iHS was calculated and normalized using allele frequency bins in Selscan (Szpiech and Hernandez 2014). iHS values were grouped into 200 kilobase pair (kb) non-overlapping windows. These windows were then binned by number of SNPs for quantile estimation of percentile using selscan's norm function. iHS scores were not computed for a MAF < 0.05. Windows with fewer than 10 SNPs were dropped from analysis. Only windows in the 1% tail of bin distribution were considered.

Annotation of Regions: Windows were annotated for genes using the bedmaps option from BEDOPS tools (Neph et al. 2012).

RESULTS

Estimating Individual Admixture

To conduct our selection scan in our Mesoamerican cohort, we first needed to address potential admixture in the population. We estimated global ancestry using fastStructure

(Pritchard et al. 2000). We tested for four-way admixture in the population including

ancestry from the Americas, Europe, Africa, and East Asia. Ancestry estimates are

reported in Table 3.1. Individual admixture estimates ranged from a maximum of 100%

Indigenous American ancestry to a minimum of 74% Indigenous American ancestry.

European admixture was detected in 15 individuals, ranging from 1% to 24%. Ten

individuals had detectable East Asian ancestry ranging from 1% to 5%. African Ancestry

was detected in two individuals at 5% and 2%, respectively. Part A of Figure 3.1 shows

the global ancestry estimates of all Mesoamericans (39), alongside the East Asians,

European Americans, and Yorubans. These results indicate that Mesoamericans

primarily have European admixture. Part B of Figure 3.1 shows the amount of admixture

present in each Mesoamerican individual, after the removal of any potential relatives

and the removal of two participants with only 74% Indigenous American ancestry.

Table 3.1. Mesoamerican Global Ancestry Estimates

Sample	Population	Indigenous American	CEU	YRI	EAS
MX004	Mexico	0.97	0.03	0.00	0.00
MX019	Mexico	1.00	0.00	0.00	0.00
MX024	Mexico	1.00	0.00	0.00	0.00
MX034	Mexico	1.00	0.00	0.00	0.00
MX035	Mexico	1.00	0.00	0.00	0.00
MX037	Mexico	1.00	0.00	0.00	0.00
MX040	Mexico	1.00	0.00	0.00	0.00
MX042	Mexico	1.00	0.00	0.00	0.00
MX043	Mexico	1.00	0.00	0.00	0.00
MX045	Mexico	0.96	0.00	0.00	0.04
MX050	Mexico	1.00	0.00	0.00	0.00
MX056	Mexico	1.00	0.00	0.00	0.00
MX059	Mexico	0.91	0.06	0.00	0.03

MX062	Mexico	1.00	0.00	0.00	0.00
MYN001	Mexico	0.74	0.26	0.00	0.00
MYN002	Mexico	0.89	0.11	0.00	0.00
MYN003	Mexico	0.83	0.14	0.00	0.03
MYN004	Mexico	0.97	0.00	0.00	0.03
MYN006	Mexico	0.88	0.09	0.00	0.03
MYN007	Mexico	0.74	0.24	0.00	0.01
MYN008	Mexico	1.00	0.00	0.00	0.00
MYN009	Mexico	0.98	0.02	0.00	0.00
MYN010	Mexico	1.00	0.00	0.00	0.00
MYN011	Mexico	1.00	0.00	0.00	0.00
MYN012	Mexico	1.00	0.00	0.00	0.00
MYN013	Mexico	0.86	0.14	0.00	0.00
MYN014	Mexico	1.00	0.00	0.00	0.00
MYN015	Mexico	0.99	0.01	0.00	0.00
MYN018	Mexico	0.95	0.05	0.00	0.00
MYN019	Mexico	0.97	0.03	0.00	0.00
MYN020	Mexico	0.95	0.00	0.05	0.00
MYN021	Mexico	1.00	0.00	0.00	0.00
MYN022	Mexico	0.98	0.00	0.00	0.02
MYN023	Mexico	1.00	0.00	0.00	0.00
MYN024	Mexico	0.99	0.00	0.00	0.01
MYN025	Mexico	0.99	0.01	0.00	0.00
MYN027	Mexico	0.92	0.08	0.00	0.00
MYN028	Mexico	0.89	0.07	0.02	0.02
MYN030	Mexico	0.95	0.00	0.00	0.05
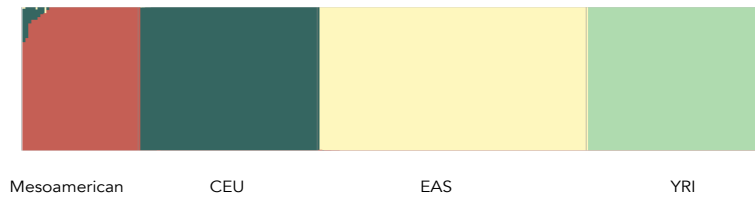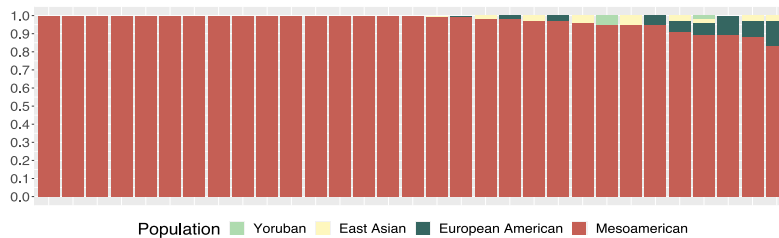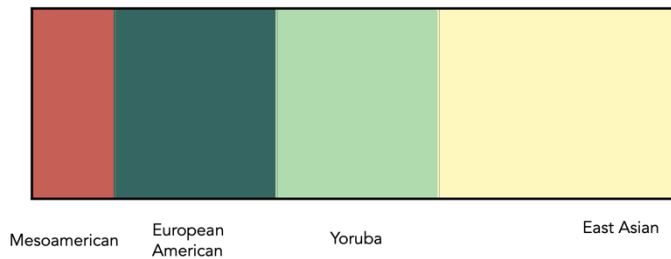
Figure 3.1. Individual Ancestry Estimates. Individual ancestry was estimated for Mesoamerican study participants using fastStructure. A. fastStructure global estimates of ancestry at K=4 for Mesoamericans, CEPH Europeans (CEU), East Asians (EAS), and Yoruba (YRI). B. Detailed image of the global ancestry estimates for Mesoamerican participants that were selected to proceed to the RFMix analysis and then in the downstream selection scan (N=31). Criteria for exclusion were high levels of admixture or relatedness to our other participants. C. fastStructure results, K = 4, after removing admixed segments and imputing missing genotypes from Indigenous American SGDP genomes.

Given the identification of non-Indigenous American ancestry within a subset of

our study participants, we assigned locus-specific ancestry to each chromosomal

segment/haplotype using RFMix (Maples et al. 2013). We set non-Indigenous American

53

ancestry segments to missing and imputed the missing genotypes using the SGDP

Indigenous Americans in Beagle 5.1 (Browning et al. 2018a; Browning and Browning

2007). We chose to use the SGDP Indigenous Americans given the diversity of the

populations representing indigenous populations across the Americas. Furthermore, the

SGDP populations were preferable to the 1000 Genomes Project populations given

their low level of admixture. We then re-ran fastStructure on the masked and imputed

Mesoamerican dataset. The results indicated that this analysis effectively eliminated

European, East Asian, and African ancestry from the Mesoamerican genomes (Figure

1C).

Selection Scan: Major findings

We identified genomic signals of natural selection using three statistics including: 1)

LSBL (Shriver et al. 2004), 2) XP-EHH (Pickrell et al. 2009; Sabeti et al. 2007), and 3)

iHS (Voight et al. 2006). LSBL identified 4,047 SNPs falling in the top 1% of the

empirical distribution. XP-EHH and iHS identified 141 and 101 statistically significant

windows at the 1% level, respectively. Figure 3.2 depicts the results for each of the

three tests for individual SNP values after data normalization.
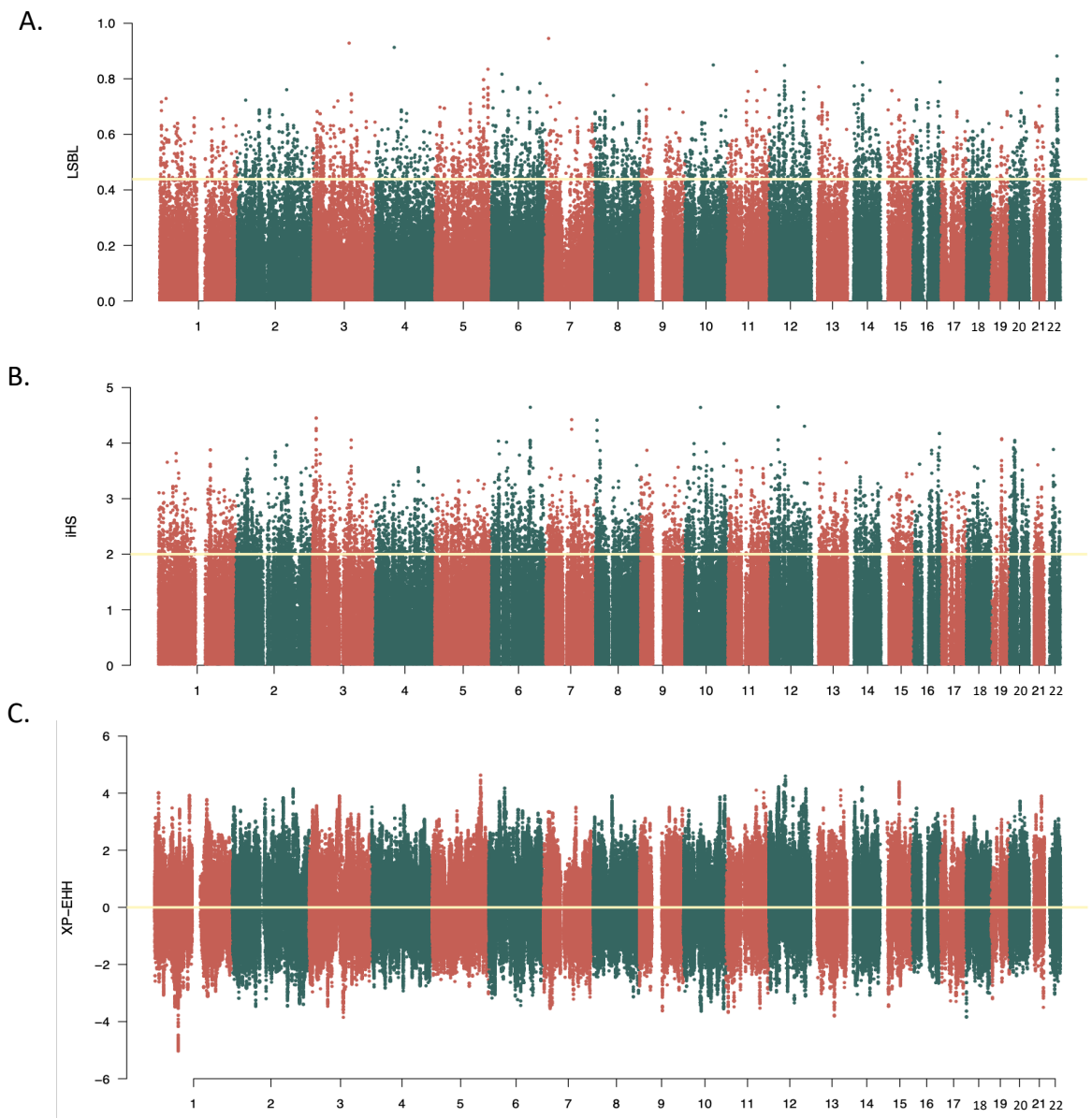
Figure 3.2. Manhattan Plots of Selection-Scan Statistics. For each plot, the value of the statistic is represented on the y-axis. Chromosome location is depicted along the x-axis. Yellow lines indicate the 1% significance level for each test statistic, with the exception of XP-EHH, which shows the axis. A. Distribution of LSBL values across the genome for Mesoamerican branch length calculated using East Asians and CEPH European Americans as outgroups. B. Plot of the absolute value of iHS scores for Mesoamericans. C. Plot of XP-EHH comparing Mesoamericans to East Asians. Values above 0 indicate directional selection in the Mesoamerican population.

To reduce false positives, we identified regions of the genome showing statistical significance for LSBL and at least one of the two haplotype tests, XP-EHH and iHS. The following results are the windows identified that passed the 1% thresholds for both the haplotype test and the allele frequency test. Chromosome 6 contained the most significant regions of any chromosome for both the XP-EHH/LSBL and iHS/LSBL analysis (Figure 3.3). For XP-EHH/LSBL, chromosome 6 had 14 windows that passed the 1% threshold for both tests. Similarly, for iHS/LSBL, chromosome 6 had 12 outlying windows on chromosome 6. Given the presence of the Major Histocompatibility Complex (MHC) on chromosome 6, this result was anticipated. In fact, the majority of the significant results for this region were identified in and around the MHC region. In addition to the MHC region, we identified clusters of statistically significant results on the q arm of chromosome 6.
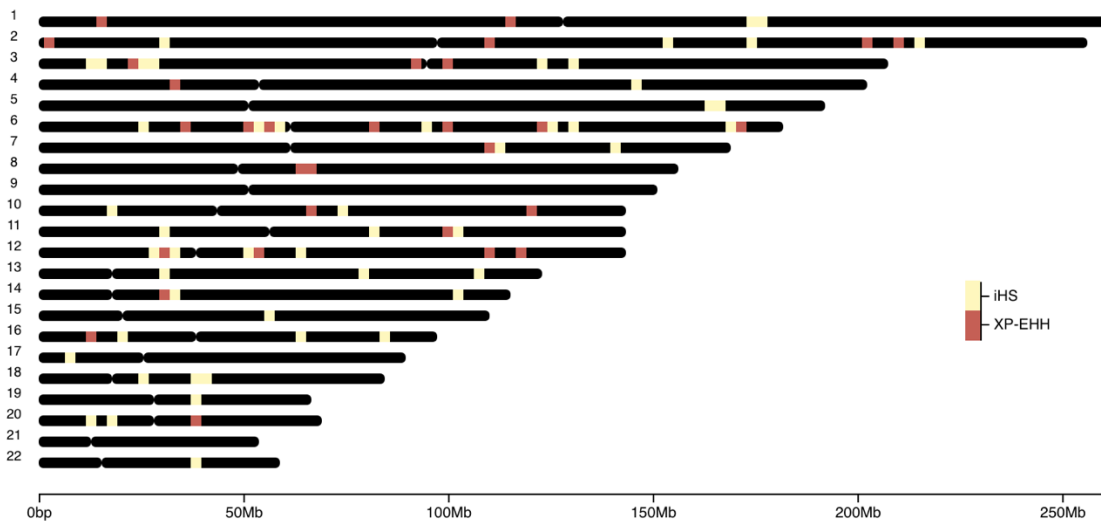
Figure 3.3. Distribution of 1% windows when paired with LSBL. iHS/LSBL regions are shown in yellow and XP-EHH/LSBL regions are shown in red. For both of the combined statistical tests, the majority of windows were found on chromosome 6, followed by chromosome 3.

In addition to our findings on chromosome six, we identified several other chromosomal regions containing windows with significant results. Our most compelling result was found on chromosome 3 (chr3:12,200,001 – 12,800,001). This region consisted of three, back-to-back significant 200 kb-windows contained a significant fraction of both LSBLs and iHS values in the top 1% of results. Here, our highest LSBL value was 0.68 and the highest iHS value was 3.68. This region contains the following genes: *PPARG*, *SYN2*, *TIMP4, MKRN2*, *MKRN2OS, TSEN2, RAF1*, and *TMEM40*.

A second compelling result was a 200 kb-window on chromosome five (chr5: 15,3800,001-15,4000,001) with both a significant fraction of LSBLs and XP-EHH values in the 1% of results. Here our highest LSBL value for the window was 0.57 and the highest XP-EHH value was 4.63. This window contained the following genes: *GALNT10*, *HAND1*, *MIR314*, *SAP30L*, *SAP30L-AS1*.

In addition to these results, we identified several other genes involved in immune response pathways that stood out given what is known about Mesoamerican population history. These were identified by looking at the top windows with extreme values of our selection scan and then doing a literature review of those genes in the top windows. These potentially may be related to infections with smallpox, tuberculosis, or other important immune signaling pathways. This is summarized in Table 3.2.

Table 3.2. Immune response genes significant for LSBL/XP-EHH or LSBL/ iHS.

Gene Name	Window (hg19/)	max Test	max LSBL	Test	Info
IL18R1, IL1RL1, IL1RL2	Chr2:102800001-103000001	3.79	0.68	XP-EHH	IL-1 receptor activity and IL-33 activity; IL18R1 has been associated with adverse effects of the Smallpox vaccine.
SFI1	Chr6:159400001-159600001	2.3	0.43	iHS	NF-κB pathway; regulates inflammation
ILRUN	Chr6:34400001-34600001	3.64	0.46	XP-EHH	ILRUN interacts with IRF3 as part of the IFN- γ production and the NF-KB pathways by inhibiting IRF3 and regulating cytokine production; associated with antiviral response
CYP7A1	Chr8:59400001-59600001	3.73	0.74	XP-EHH	Part of TLR signaling; previously associated with Tuberculosis
AGAP2	Chr12:58000001-58200001	2.91	0.35	iHS	IL-6-type cytokine receptor ligand genes; part of the TGF- β signaling
CIITA	Chr16:10800001-11200001	3.06	0.51	XP-EHH	In the HLA region; regulates MHC class I gene transcription

Pathway Analysis

We used gene ontology pathway enrichment analysis in DAVID to determine

overrepresented associations of genes and gene groups. We focused on regions that

were significant for the LSBL/XP-EHH or the LSBL/iHS analysis. For LSBL/XP-EHH,

DAVID analysis identified eight functional-related gene groups (Table 3.3). Fold

Enrichment in our tables represented how many times the category was

overrepresented in the database over random chance.

Table 3.3. LSBL and XP-EHH 1% genes nominally in overrepresented pathways.

GO Term	Genes	Count	Fold Enrich.	P-val	Benjamini
chromosome condensation	*PRM3*, *PRM1*, *PRM2*	3	71.97	0.001	0.277
DNA packaging	*PRM1*, *PRM2*	2	79.96	0.024	0.996
chitin catabolic process	*OVGP1*, *CHIA*	2	68.54	0.028	0.986
glucocorticoid receptor signaling pathway	*YWHAH*, *NCOA6*	2	53.31	0.036	0.984
positive regulation of chemokine secretion	*CHIA*, *IL1RL1*	2	53.31	0.036	0.984
cardiac septum morphogenesis	CHD7, HAND1	2	47.98	0.040	0.974
negative regulation of proteasomal ubiquitin-dependent protein catabolic process	*CLEC16A, SDCBP*	2	21.81	0.087	0.999
immune response	*CIITA, IL18R1, CHIA, IL1RL1, TINAG*	5	2.85	0.095	0.998

GO, Gene Ontology

For the LSBL/iHS analysis, DAVID analysis identified enriched 13 GO terms (Table 3.4).

For each DAVID analysis, the p-value did not remain significant after correcting for

multiple tests.

Table 3.4. LSBL and iHS 1% genes nominally in overrepresented pathways.

Term	Genes	Count	Fold Enrich.	P-val	Benjamini
detection of light stimulus involved in visual perception	ATP8A2, CNGB1, RGS9BP	3	26.70	0.005	0.971
amino acid transmembrane transport	SLC38A4, SLC7A9, SLC7A10	3	18.15	0.012	0.977
amino acid transport	SLC38A4, SLC7A9, SLC7A10	3	12.97	0.022	0.992
cellular response to hyperoxia	ATG7, PPARG	2	43.22	0.045	0.999
negative regulation of endopeptidase activity	SERPINA5, SERPINA4, SERPINA3, TIMP4	4	5.00	0.045	0.998
negative regulation of cAMP-dependent protein kinase activity	PRKAR2B, PKIB	2	37.82	0.051	0.997
L-serine transport	SERINC1, SLC7A10	2	33.62	0.057	0.996
cilium morphogenesis	CEP89, PKHD1, CFAP20, AVIL	4	4.45	0.060	0.993
vitamin D metabolic process	CUBN, CYP27B1	2	30.26	0.064	0.991
positive regulation of GTPase activity	DIS3, ANKRD27, RSU1, TAGAP, ARHGEF25, ARHGAP15, AGAP2, DEPDC5	8	2.14	0.078	0.995
neutral amino acid transport	SLC7A9, SLC7A10	2	23.27	0.082	0.994
white fat cell differentiation	CEBPA, PPARG	2	23.27	0.082	0.994
lipoprotein transport	CUBN, PPARG	2	20.17	0.094	0.995

DISCUSSION

Our results confirmed our hypothesis that natural selection at immune response genes is central to the evolutionary history of Mesoamerican populations and was shaped by their history of exposures to infectious diseases. We were able to demonstrate that patterns of genomic diversity at the population level show convincing

evidence of past natural selection. Our findings thus add additional support to previous

studies of selection that identified in Indigenous American ancestry.

We identified signatures of natural selection at several chromosomal locations

that contained clusters of immune response genes. Our most compelling result was

found on chromosome 3 (chr3:12,200,001- 12,800,001). Within this region, we found

the following genes: *PPARG, SYN2, TIMP4, MKRN2, MKRN2OS, TSEN2, RAF1*, and

TMEM40. Of this group, *PPARG, RAF1*, and *MKRN* are part of various immune

response pathways. As part of the PPAR-γ signaling pathway *PPARG,* or *Peroxisome*

Proliferator Activated Receptor Gamma, would be the most likely candidate for selection

in response to pathogenic infection. This gene is a ligand-activated transcription factor

that contributes to gene regulation. The PPAR-γ signaling pathway regulates lipid and

glucose metabolism through the expression of cytokines and chemokines (Le Menn

and Neels 2018). Importantly, the PPAR-γ signaling pathway has known immune

functions by activating both pro- and anti- inflammatory macrophages (Chawla 2010).

However, it should be cautioned, given the continuous legacy of infectious disease

exposure in Mesoamerica, any of the genes in this region could have been the target of

selection.

A second significant region located on chromosome 5 contained five genes:

GALNT10, HAND1, MIR314, *SAP30L*, and *SAP30L-AS1*. Of these genes, *GALNT10* is

known to be responsible for regulating CD4+T cells infiltration and decreasing

granzyme B expression in CD8+T cells (Zhang et al. 2020). T cell regulation is crucial

for an adequate immune response. For instance, CD8 cells are essential for protection

against viruses, intracellular bacterial infection, and tumor cells, while CD4+T cells are crucial to immune memory (Worthington et al. 2012). *GALNT10* also is known to interact with the MHC complex genes including various interleukin cytokines (Kakoola et al. 2014).

Chromosome 6 contained the highest number of significant LSBL/XP-EHH and LSBL/iHS windows. This is likely the result of the MHC, a region of known high genomic diversity (de Bakker et al. 2006; Trowsdale 1993). MHC is a complex that contains 224 genes, largely related to immunity (MHC sequencing consortium 1999). The MHC complex has been largely detected in natural selection scans, not only in populations outside the Americas, but also across many other mammalian and aquatic species (Hughes and Yeager 1998b).

We identified several other genes under selection that belong to various cell-signaling pathways that have been associated with immune response to infections such as tuberculosis, intestinal parasites, and variola in other populations. These genes include *IL18R1*, *CYP7A1*, and *CIITA* among others. This highlights the importance of immunity and adaptation among Mesoamerican populations whether deep in our evolutionary past, as recent as colonial contact, or even continuously shaped by more recent infectious diseases.

Our selection scan was designed to identify outlying genes in Mesoamericans when compared to the other distantly related populations, in this particular case East Asians and Europeans. Given the lack of publicly available data for larger cohorts of Indigenous American populations, we did not compare Mesoamerican genomes to other

Indigenous American genomes to identify region-specific selective events. Therefore, this particular design was not be able to distinguish if a selective event was specific to Mesoamericans or affected Indigenous American populations more broadly. An example of this is our identification of *CIITA* as an outlying immune response gene in Mesoamericans. This gene also has been demonstrated to be under selection in Brazil (Mychaleckyj et al. 2017). This could very well be due to shared demographic history or the result of selection at this locus occurring prior to the population split of these two groups.

Similarly, we identified evidence of selection for a cluster of three genes, *IL18R1, IL1RL1,* and *IL1RL2* located on chromosome three, that are part of the IL-1 and IL-33 cytokine activity. Reynolds et al. (2019) identified *IL1R1*, a related gene to *IL1RL1* and belonging to the same pathway, as part of the selection scan performed for an indigenous population of the southeastern United States. This is further evidence that not all of the outlying genes identified by our selection scan are due to a shared past, but rather it highlights the possibility that similar pathogens may have affected local populations. In this particular case, similar pathogenic exposure could have influenced genes along the IL-1 and IL-33 pathway without necessarily affecting the same genes. Furthermore, it highlights the importance of cytokines and/or the cytokine pathway in the body's defense system against pathogenic infection and as targets of past natural selection across human populations.

There are several limitations to our study. First, our results could be influenced by SNP design and ascertainment bias of the Affymetrix 6.0 chip. For instance, the

design of the Affymetrix chip was designed to capture the diversity and haplotypes from the HapMap Project, which source populations were the European Americans, East Asians, and Yorubans. Therefore, linkage disequilibrium blocks and distribution of SNPs would be different in Mesoamerican populations. Second, given that our haplotype tests require ancestral allele information, many SNPs without this information were dropped from analysis. Therefore, many regions were excluded from the XP-EHH and iHS analysis due to insufficient SNP density in a given window. Either a better designed SNP chip or the interrogation of sequencing data would remedy these caveats in future studies.

Studying variation in outlying regions shaped by natural selection would be beneficial to studies of current pathogens. Within the bounds of human evolution, evolutionary theory predicts that genes in pathways that control the body's response to infectious disease, climate, altitude, and metabolic traits show the strongest selection signatures in the genome—as they are traits that manifest themselves physiologically, thus improving chances of survival and reproduction (Grossman et al. 2013; Sabeti et al. 2007). Therefore, these genes under selection are more likely than other parts of the neutrally evolving genome to have some sort of phenotypic effect (Fumagalli et al. 2010; Wisser et al. 2008). Therefore, this makes proposing regions under selection to the study of modern infectious disease an attractive strategy, given that they will manifest themselves physiologically and are known to be differentiated from other global populations. This strategy would work as our immune system is highly redundant and compensates for factors such as novel genetic variation that may be detrimental to

specific pathways. Therefore, any small changes manifest themselves physiologically

are worthy of study (Nish and Medzhitov 2011). Focusing on genes under selection has

proven to be beneficial in smaller sample sizes as demonstrated by several studies

taking this approach (Karlsson et al. 2013; Park et al. 2012; Perry et al. 2014;

Schwarzenbacher et al. 2012). Hence, we recommend that studies with a limited-sized

study population whose goal is to identify susceptibility or resistance factors for

infectious diseases such as dengue, zika, chikungunya, or coronavirus should start off

by targeting genes under natural selection. Similarly, targeting immune response genes

subject to past natural selection can aid in the study of population specific variants

related to auto-immune diseases or cancer given that many of the pathways are

overlapping.

Our results are not only useful in studies related to immunity. Many of these

genes identified here also are associated with various metabolic traits such as

cholesterol and insulin. For example, *PPARG* and *CYP7A1* regulate cholesterol

homeostasis and metabolism (Baker et al. 2010; Chinetti et al. 2001; Li et al. 2013).

High cholesterol in Latinx populations in the US have been documented as risk factors

for heart disease (Daviglus et al. 2012; Li et al. 2017). However, an evolutionary past

could have selected for those traits, given that cholesterol also plays a role in activating

the TLR signaling pathways as part of the innate immune system (Björkbacka et al.

2004; Ghosh 2011). Using signatures of natural selection to study modern health

problems was implemented in a heart disease study among a Mexican population. Ko

and colleagues pulled out SNPs whose allele frequencies differed from European

populations for the focus of their analysis (Ko et al. 2014). In so doing, they were able to

successfully identify genes and risk alleles involved in dyslipidemia that were under

selection. Follow-up studies to our selection scan using highly differentiated alleles in

populations of Mesoamerican ancestry would aid in the identification of genotype-

phenotype associations by increasing the power to find associations with such

metabolic traits.

Chapter 4: Epidemiological Risk Factors for Dengue-Infection in Guatemala

NOTA BENE

Originally, this chapter was structured to be a continuation of the other chapters, by nominating the signatures of selection as candidate loci and test our hypothesis that these outlying alleles do affect modern infectious diseases immune response by looking at dengue infection in a modern Mesoamerican cohort. Furthermore, we were hoping to not only find allelic associations with dengue, but also tie genotype to phenotype by looking at cytokine response to infection and how our statistically significant alleles may affect cytokine expression. Unfortunately, lab closures due to a novel infectious disease, SARS-CoV-2 hindered our ability to finish this project. Here we present the data that we were able to analyze in the midst of a pandemic that paralyzed the world.

ABSTRACT

Dengue Virus (DENV) is a mosquito borne infectious disease that has expanded exponentially in Latin America in the last 50 years. In Guatemala, a country located in Central America, DENV-infection has exponentially grown in the last 50 years as a result of rapid urbanization, substandard infrastructure, the of widening socio-economic gaps, and the cessation of vector controls by the Pan American Health Organization. Furthermore, nearly half of the Guatemalan population lives below the established poverty line, and dengue infection is known to disproportionately affect those living in

poverty given that several worldwide risk factors for DENV are tied to water security, water storage, and nearby trash. This study identifies the epidemiological risk factors plaguing Guatemalan populations that increase their risk for DENV infection, by looking at epidemiological surveys in a case-control cohort. Our cohort consisted of 82 participants with dengue symptomology and 159 uninfected controls recruited in 2018-2019. Each question asked on the epidemiological surveys were analyzed for differences in the case-control cohort using Welch Two Sample t-test and Pearson's Chi-squared test with Yates' continuity correction. Furthermore, we also combined the variables into a single model using logistic regression. We find that waste on property is the greatest risk factor for dengue-infection across all models. These findings will aid the Ministry of Public Health and Social Assistance of Guatemala in determining where to direct their efforts in dengue prevention.

INTRODUCTION

Dengue Virus (DENV) is a mosquito-borne RNA virus belonging to the family Flaviviridae. DENV is widespread in subtropic and tropic environments, with estimates as high as 390 million infections each year, of which 25,000 result in death (Horstick et al. 2015). It results in wide variation of clinical presentations and outcomes. As outlined in the classic model of DENV-infection, phenotypes range from asymptomatic infection to dengue fever (DF) to dengue hemorraghic fever/dengue shock syndrome (DHF/DSS). Recently, the WHO proposed revisions to this classification system by focusing on the fever inducing stages to improve clinical treatment outcomes and curb

disease progression. The new system has the following categories: dengue infection

without warning signs, dengue with warning signs, and severe dengue (Diseases 2009).

Four confirmed DENV serotypes (DENV–1, 2, 3, 4) exist, with a possible fifth

discovered within the last four years (Normile 2013). Infection with one serotype

typically results in asymptomatic infection and provides long-lasting host immunity to

that strain. Heterotypic secondary infection demonstrates the same clinical course as

infection with one serotype, but with increased risk of developing DHF/DSS or severe

dengue phenotypes (Holmes and Twiddy 2003; Whitehorn and Simmons 2011).

Dengue, formerly characterized as a neglected infectious disease, is now

characterized as an emerging infectious disease given its wide distribution and

increasing prevalence in "developing" countries of Southeast Asia and Latin America.

According to the World Health Organization (WHO), the incidence of DENV-infections

has increased 30-fold over the past 50 years, reaching epidemic proportions in Asia and

Latin America (Navarro-Sanchez et al. 2005). Today, rates of infection in Latin America

are growing at the fastest rate in the world, with complete coverage over all the tropics

and subtropics in a little over 30 years (Kyle and Harris 2008).

In Guatemala, a country in Central America, DENV was first reported in 1978,

after an epidemic in Honduras that spread to nearby countries (Brathwaite Dick et al.

2012; Pinheiro and Corber 1997). Today, DENV continues to grow exponentially.

Several factors have contributed to its growth. First, a major earthquake in Guatemala in

1976 left 1.2 million people homeless (Riding 1977). As people struggled to remake

their lives, they fled to cities, leading to the rapid expansion of urban environments

without proper infrastructure such as running water (Oliver-Smith 1991). This created a prime environment for the *Aedes* mosquito to flourish in Guatemala. Concurrently the global market and trade in Latin America blossomed, and trade through ships expanded. Unfortunately, simultaneous with trade was the spread of the vectors and novel infectious diseases, as ships contained stagnant water through which mosquitos bred (Sprenger 1987). Second, government cuts to funding exacerbated infection rates. For instance, the rapid expansion of DENV-infection in Guatemala in particular and Latin American more generally is attributable to the 1970's cessation of a Pan American Health Organization campaign to eradicate the *Aedes* mosquito (Gubler and Clark 1995; Kendall et al. 1991).

Most recently, cases tripled from the year 2018 to the year 2019 based on statistics from the Ministry of Public Health. This is in large part the result of changes to climatic conditions that influence risk. In Guatemala, risk is no longer cyclical, in sync with the rainy season, but year-round. This highlights how climate change increases the incidence of disease in areas previously affected seasonally. More broadly, climate change has contributed to range expansions of the *Aedes aegypti* mosquito, DENV's primary vector, causing DENV and other related diseases, such as Yellow Fever and Chikungunya, to spread to novel populations (Muehlenbein 2010).

We established this study in Guatemala to determine the risk factors contributing to dengue infection. Determining these social factors will ultimately assist public health officials in issuing prevention strategies and will bring to light the much-needed infrastructure both in quality of life and in healthcare. Reducing risk is necessary as it

not only affects the odds of infection, but repeat and chronic exposures also affect overall health outcomes at the point of infection.

Despite the Ministry of Public Health and Social Assistance's work to prevent dengue infection, we predict that water-based sources remain a major factor in risk for dengue infection. Here we partnered with the Ministry of Public Health and Social Assistance of Guatemala to document and survey risk factors for water security, water storage, household density, waste on property, and waste near property, to determine the risk factors contributing to dengue infection in Guatemala.

PARTICIPANTS AND METHODS

This study and all its materials were approved by the University of Michigan's Institutional Review Board and the National Committee of Ethics from the Ministry of Public Health and Social Assistance from the Republic of Guatemala.

Study participants

In a collaboration with the Ministry of Public Health and Social Assistance of Guatemala, we recruited participants primarily in two urban centers, Guatemala City and Chiquimula, using a national hospital and a few public health posts. In order to be included in our study, participants needed to be over the age of 18 and either Maya or Ladino. Maya participants are those indigenous to Guatemala, who have maintained their linguistic and cultural traditions. Ladino represents the category known in other parts of Latin America as "mestizo". Ladino's are an admixed population with both

European and Indigenous American ancestry. However, the category of Ladino is a social category that is not always indicative of ancestry. For example, Maya individuals may assimilate into the Ladino population within their lifetime if they shed their cultural and linguistic ties (Berghe 1968). Similarly, Maya individuals may be admixed, but still be identified as Maya as long as they were primarily raised in an indigenous community and keep their linguistic and cultural practices.

Interviews for Epidemiological risk factors

Our participants were of similar socioeconomic status; however, we saw it necessary to tease apart specific risk factors associated with dengue-infection. All cases and controls were interviewed to collect a demographic information about specific risk factors for dengue infection. Here, we tested several factors for increased risk to dengue infection including, water storage, water security, improper storage of waste, and proximity to cemetery or scrap shops. The exact translations of questions asked were as follows:

1) Do you store water in uncovered water basins, barrels, or other recipients?

2) Number of people in your household?

3) Are there scraps, useless containers, and/or tires on your patio or surrounding it?

4) Do you live less than 100 meters from a cemetery, junk yard, or tire yard?

5) Do you have running water in your household?

Statistics

For continuous variables, we used the Welch Two Sample t-test to test for differences between cases and controls. These tests were conducted in R using the package "dplyr."

For categorical variables, we used the Pearson's Chi-squared test with Yates' continuity correction to test for deviations from expectations in cases and controls. For sex, we set the case/control status as the outcome and male/female as the exposure variable. For the epidemiological risk factors (i.e. household density, water availability, water storage, waste on property, and waste near property), we set the risk factor as the exposure variable in order to determine the odds for people contracting dengue as the outcome variable. When the relationship was significant, we calculated the odds ratio (OR) for the statistic. These tests were conducted in R using the "MASS" package for Chi-squared and the "oddsratio" package for Odds Ratio. Lastly, we tested all the variables in a single model using logistic regression. The logistic regression was modeled in R using the package "aod."

RESULTS

Participant Characteristics

Our study participants came from six Guatemalan departamentos or states, with five participants from Honduras who were recruited in Guatemala. The states included Guatemala, Chiquimula, Jalapa, Zacapa, Alta Verapaz, and Jutiapa (Figure 4.1). We recruited 82 participants with dengue symptomology and 159 uninfected controls (Table 4.1). The majority of participants (n = 149) were recruited in Chiquimula, followed by 69

participants recruited in Guatemala. Six, six, four and two participants were recruited in Jalapa, Zacapa, Alta Verapaz, and Jutiapa, respectively. (Table 2). Five cases were residents of a border town in Honduras, Copan Ruinas, just a 1.5-hour drive to the Chiquimula hospital. These participants were recruited in Chiquimula. Through the C-4 agreement, Guatemala, El Salvador, Honduras, and Nicaragua, share open borders. Therefore, people go to the closest hospital even if it is across an international border.

Figure 4.1 Study sites across Guatemala. The six departmentos where study participants were from are indicated in green. The stars represent the two primary city locations where recruitment was performed. The dark green dot underneath the name Chiquimula, shows the approximate location of the five cases who came from the Honduran town, Copan Ruinas.

Table 4.1 Participants' departamento of origin

State	Cases	Controls	Total
Guatemala	12	57	69
Chiquimula	59	90	149
Jalapa	4	2	6
Zacapa	0	6	6
Alta Verapaz	1	3	4
Jutiapa	1	1	2
Honduras	5	0	5
Total	82	159	241

Participants from each recruitment location came from various geographic backgrounds, as all the states are a mixture of urban, rural, and urban-rural. Urban-Rural communities were small dense communities that were surrounded by farmlands. We conducted a Pearson's Chi-squared test with Yates' continuity correction to determine if there was a difference between different types of living communities in our cohort (i.e. urban, rural, and urban-rural). We set the case/control status as the outcome variable and the living location as the exposure variable. Although the living locations are different, the participants would be from roughly similar socioeconomic status. We found a significant difference between cases and controls for the three groups ($\chi^2 = 43.06$, $p < 0.001$) (Table 4.2).

Table 4.2 Participant Characteristics

Variable	Cases (n = 82)	Controls (n = 159)
Urban %	39 ± 10.98	77 ± 6.5
Rural %	20 ± 8.95	15 ± 5.56
Urban-Rural %	41 ± 11.04	8 ± 4.11
Age, yr	39.8 ± 19	36.8 ± 17.7
Female %	71 ± 9.85	73 ± 6.90
Household Size, persons	5.7 ± 2.29	5.32 ± 2.68

All characteristics are reported as the mean ± the standard deviation or the 95% confidence interval for proportions.

Cases and controls had similar age ranges and mean ages (Figure 4.2). Our case population was slightly older on average with a mean age of 39.8 years old. The age range for cases was from 18 to 81. Our control population age ranged from 18 to 92 with a mean age of 36.8 (Table 4.2). There was no difference in the means for age in our case-control cohort (T-test = -1.1857 (95%CI: -7.98-1.99), p = 0.23) (Table 4.2). We recruited 174 female participants and 67 male participants. However, cases and controls were not different by sex (χ^2 = 0.05, p = 0.83)
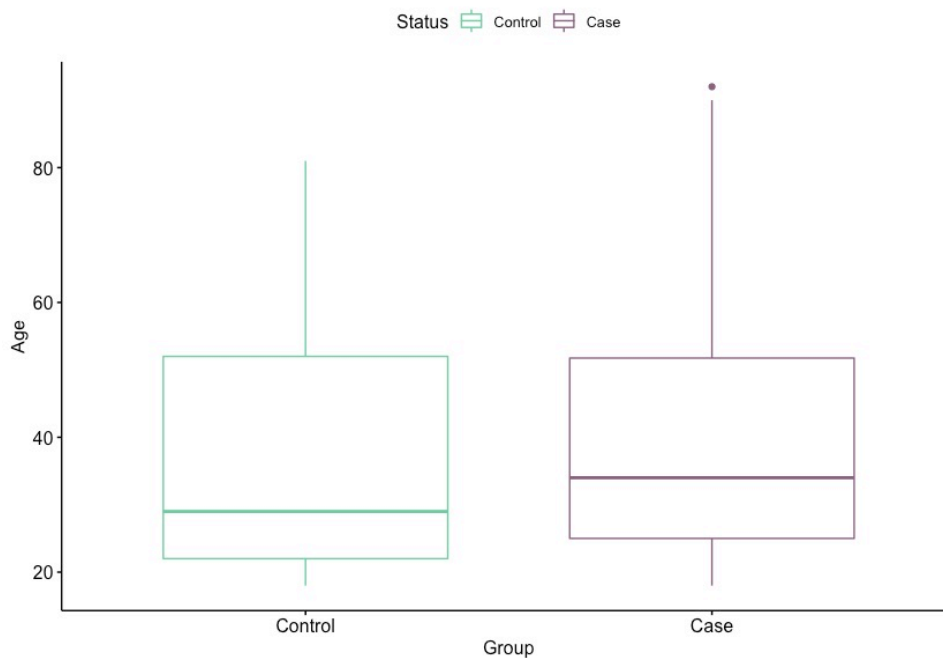
Figure 4.2. Box Plot of age in cases and controls. This boxplot shows the distribution of age in our cases and controls. In the left, we have a mean age of 36.8 for controls. In the right, we have a mean age of 39.8 for cases. The differences between the two groups were determined not to be significant.

Epidemiological Risk Factors

Several practices common in Guatemala can increase the risk for the mosquito vector to reproduce, thus increasing the chance of dengue transmission. Through participant interviews, we gathered data on five epidemiological risk factors including: people in household, running water to household, water storage, waste on property, and waste near property.

People in Household

The amount of people in a household can be a potential risk factor as it increases the population density in an area, allowing for easier transmission via the *Aedes* mosquito. Cases and controls reported similar numbers of people living in their household (Figure 4.3). Cases had a mean of 5.70 individuals in their household, whereas controls reported a mean of 5.32 individuals in their household (Table 4.2). Cases and controls were not different for this risk factor (t-test = -1.08 (95% CI -1.07 - 0.32), $p = 0.283$).
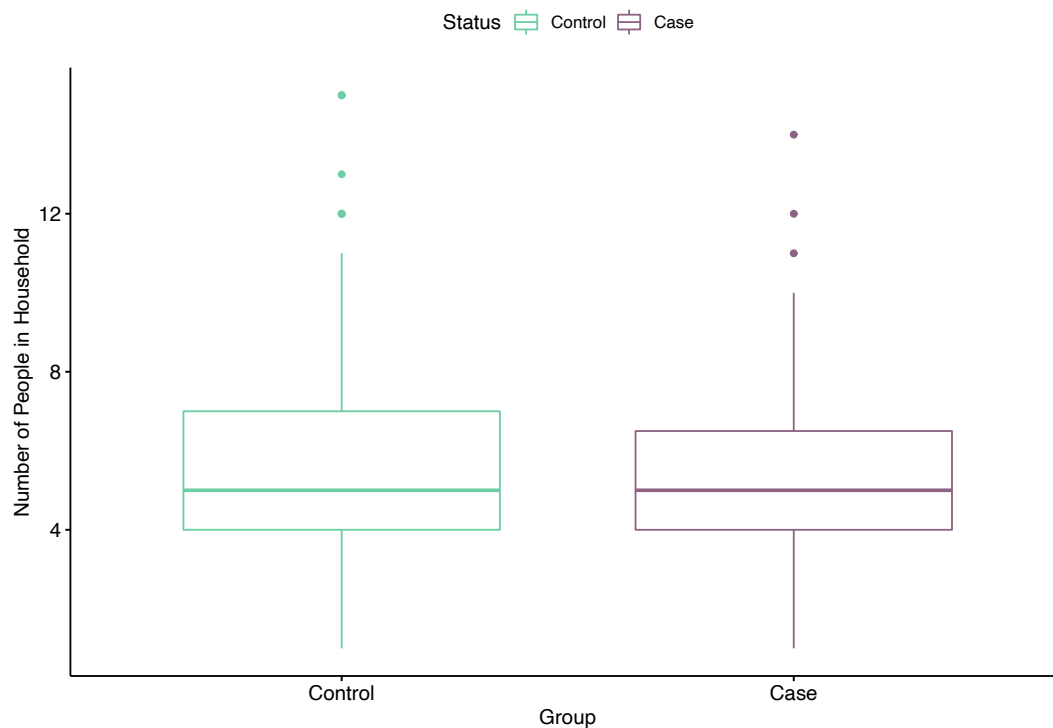
Figure 4.3. Box Plot of the number of people per household in cases and controls. This boxplot shows the distribution of Number of People in Household in our cases and controls. In the left, we have a mean number of 5.32 people living in a household for controls. In the right, we have a mean number of 5.70 people living in a household for cases. The differences between the two groups were determined not to be significant.

Water Security

Water security can be an important risk factor contributing to infection rates for

mosquito-borne diseases. Without access to running water, individuals must obtain

water outside of the home and store it in containers. This stored water remains stagnant

and is a known breeding ground for mosquitos. Furthermore, a lack of water security

can be indicative of and a proxy for high poverty levels. Here, we assessed water

security by determining if study participants had running water in their household. Cases

and controls were significantly different in their access to running water in their

households (χ = 7.76, p $\leq$ 0.005) (Table 4.3). The odds of dengue infection for people

with water security were smaller than the odds of those without water security (OR = 0.35, p-value < 0.004) (Table 4.3).

Table 4.3 Epidemiological risk factors for dengue infection

Variable	Cases (n = 82)	Controls (n = 159)	OR	Odds of Dengue
Water Security*	52	140	0.35	0.37
Water Storage**	52	135	0.30	0.39
Waste on Property**	28	16	5.32	1.75
Waste Near Property	20	40	NR	NR

Values for cases and controls represent the number of study participants who answered "yes" to the presence of the epidemiological variable.
Odds of dengue (ORs) are reported for the presence of the epidemiological factor
* $p \leq 0.005$
** $p < 0.001$
OR, odds ratio; NR, not reported

Water Storage

Water storage in uncovered containers is an important risk factor as it can be a breeding ground for the mosquito vector. Since people do not receive water every day, they will collect water to account for shortages the following days. We asked if participants stored water in water basins, barrels, or other uncovered receptacles in their household. We found a significant difference between the groups ($\chi^2 = 12.79$, p < 0.001) (Table 4.3). Next, we found that the odds of dengue infection for people with uncovered water storage were smaller than the odds of dengue infection for people without uncovered water storage (OR = 0.30, p-value < 0.003) (Table 4.3).

Waste on Property

Waste on property is considered a risk factor because they can accumulate water when it rains, and mosquitos can breed in those locations. We determined if study participants had scrap metal, useless containers, and/or used tires in their yard or surroundings. Here we found a significant association between the groups ($\chi^2 = 23.07$, $p < 0.001$) (Table 4.3). We found that the odds of dengue for people that store waste on their property were greater than the odds of those without waste on their property (OR = 5.32, $p < 0.001$) (Table 4.3).

Waste Near Property

Similar to waste on property, being proximate to a tire yard, cemetery, or scrap yard, places where water can accumulate through rainfall, can be a risk factor for dengue infection. We determined if study participants lived with waste near their property by specifically asking if they lived less than 100 meters from a cemetery, scrap metal yard, or tire yard. We failed to find any difference between the two groups ($\chi^2 = 4.678\text{e-}30$).

Logistic Regression for all the variables combined

Table 4.4 Epidemiological risk factors for dengue infection based on logistic regression

Variable	Estimate	Std. Error	z-value	p-value
(Intercept)**	-2.92	0.92	-3.19	1.4e-3
Sex	-0.26	0.41	-0.63	0.52
Age*	0.02	0.01	2.21	0.03
Community***	0.91	0.23	3.98	6.89e-5
Household	0.09	0.08	1.11	0.27
Water Storage**	-1.44	0.47	-3.10	1.91e-3
Water Security	-0.20	0.51	-0.39	0.69
Waste on Property***	2.21	0.46	4.76	1.91e-6

Waste near Property	-0.07	0.44	-0.163	0.87

Values for cases and controls represent the number of study participants who answered "yes" to the presence of the epidemiological variable. Community was coded by decreasing population density, with Urban as the base, followed by Urban-Rural, and finally with Rural.

* $p < 0.05$
** $p < 0.01$
*** $p < 0.001$

Our combined model results demonstrate that Community and Waste on Property are the biggest contributors to risk of dengue infection, followed by Water Storage and Age. This is mostly supported by our previous analysis looking at each variable at a time in cases and controls. The biggest contributor to risk was waste on property with an odds of 2.21 for dengue infection ($p < 0.001$). The living type community plays a risk in the risk of infection, with 0.91 increased odds for each unit step up, from Urban to Urban-Rural to Rural ($p < 0.001$). For those that answered that they store water in uncovered containers had 1.44 decreased odds for dengue infection ($p < 0.01$). We find that there is a decreased risk for dengue infection. For age, we find a very small 0.02 increase in odds for risk of dengue infection for every increment in age values (p-val < 0.05).

DISCUSSION

In conjunction with the Ministry of Public Health and Social Assistance, we recruited 82 dengue-infected participants and 159 uninfected controls. We determined the social factors that contributed to the risk of dengue infection in our cohort. Understanding these social factors will aid public health officials in identifying risk factors for dengue infection and help to develop their efforts for mitigating that risk.

The largest contributing factors to dengue infection are social living conditions and practices that grew out of rapid urbanization with substandard infrastructure. In Guatemala, socioeconomic class dictates whether you will have running water or consistent water supply. These notions were further confirmed by the significant associations primarily with type of community they live in, scrap metal on property, and storage of uncovered water. We determined that a high number of densities in the household and living near property with large amounts of trash such as junk yards were not contributors to dengue infection in our study cohort.

For water security, we found that the odds of those with no water security was 2.8 times greater than those with water security. However, this association was not supported by our logistic regression model. Some participants described having to purchase water from trucks carrying water once a week to their communities, while others described having to collect tap water from a community tap in order to obtain water for their household. It is important to note, that not everyone who had water security in their home, had water running every day. Instead, based on conversations with participants, they reported a consistent water supply that they received once to twice a week. They stored water until the next time they could collect water.

For water storage, we found that people that do not store water in uncovered water containers were at greater risk of contracting dengue than those that do. This data was supported by both analyses. This is inconsistent with other studies of dengue risk factors in Latin America that have found that uncovered water containers were breeding grounds for mosquitos placing people at increased risk for dengue infection

(San Martín and Brathwaite-Dick 2007; Villegas-Trejo et al. 2011). However, given the tight living conditions of people in urbanized locations, this question does not get to the practices of neighbors who may be improperly storing water as well. Alternatively, based on interviews with participants, it may be that the Ministry of Public Health's strategy of visiting the urban neighborhoods every few months and putting an insecticide in the water basins and barrels to reduce risk is paying off. Furthermore, these participants with water containers were more likely to receive instructional materials from Ministry of Public Health officials for reducing dengue risk. These efforts not only equalize this risk factor, but greatly reduce the odds of infection. This result can be juxtaposed with the storage of waste and metal scraps on the property, which have not received the same amount of attention or require more effort for developing preventative strategies from the Ministry of Public Health.

The odds for dengue-infection for those that store scraps, useless containers, and/or tires was significantly larger than those that did not. In fact, the logistic regression model demonstrated that the event of infection was 2.2x more likely to occur. Metal scraps and tires generally store small amounts of water that provide enough space for breeding of the Aedes mosquito (Nogueira et al. 1999). Unlike the water storage, the pools of water might be smaller in these breeding grounds and therefore are less likely to receive attention from the Ministry of Public Health and Social Assistance. Therefore, the removal of or aiding in properly storing them should be a priority for the Guatemalan government.

The limitations of our study were the nature of the questions and study sample size. Our questions were limited to those that were already approved by the Ministry of Public Health and Social Assistance for their epidemiological survey of risk factors, which we then extended to asking the control participants. Better ethnographic methods would increase our understanding about their responses to these questions. Another limitation was that we primarily were focused on urban centers and did not capture as many people living in rural communities. Therefore, our statistics of difference between the groups was most likely a reflection of recruitment strategy. Finally, even though we are able to identify significant associations between risk factors, a larger sample size would allow us to further break down the categories and fine map the risk categories within different segments of the population.

This study found that certain social risk factors contribute to the overall risk of dengue infection. Poverty levels were indicative of immune function and immunity, as the stress on the bodies ultimately might shift resources from developing a robust immune system as a trade off with other forms of growth, development, and reproduction (Baker 1997; Chen et al. 2003; McDade 2003; Owen et al. 2003). This study can be expanded to look at serum for immune response with those that reported certain risk factors as a proxy for socioeconomic status and determine if there is a difference in this population for dengue infection outcome. This study not only provides associations which will be helpful to public health officials as to what to target to reduce the risk of infection in these sectors of Guatemala, but it also identifies several potential covariates for future genomic and serological studies on dengue-infection.

Chapter 5: Concluding Remarks

This dissertation explored the population history of Mesoamerican populations with respect to immunity and exposures to infectious diseases. In order to capture snapshots of the Mesoamerican immune response over time, we focused on three events that span deep into the evolutionary past through to the present.

This dissertation was presented in a chronological order. Chapter 2 focused on archaic introgression from Neanderthals and Denisovans that were still detectable today in Mesoamerican populations. Chapter 3 identified signatures of natural selection in Mesoamerican populations as they adapted to their local environment as well as those that were potentially left behind by colonial contact. Chapter 4 focused on epidemiological risk factors for dengue infection. This dissertation contributes to the idea that the past experiences of human populations are embodied, continuing to have biological effects today (Lock and Nguyen 2010). These chapters demonstrate that we have addressed our overarching research questions as follows:

1. *Are the detectable regions of archaic introgression in modern Mesoamericans related to immunity?*

In Chapter 2, we identified two archaic introgression events within genic regions of immune response genes. We identified the transforming growth factor-β (TGF-β)

signaling pathway as introgressed from Neanderthal or Neanderthal-like populations into

the ancestral populations to Mesoamericans. This evolutionary ancient pathway

regulates gene expression, and it is an important mediator of immunity and metabolic

pathways. The TGF-β signaling pathway involves at least 87 genes, of which our

introgression analysis identified 17 as potentially adaptively introgressed. Our results

contribute to an understanding of the evolutionarily important loci that today shape the

distribution of heart disease, metabolic disorders, cancer, and infectious diseases in

populations of Mesoamerican ancestry (Fierro et al. 2017; Fragoso et al. 2012; Martínez-

Campos et al. 2019; Zavala et al. 2013). We also identified various loci related to

cytoplasm and exosomes that were introgressed by Denisovan or Denisovan-like

populations. In addition to their roles in cellular homeostasis, these loci are involved in

the body's innate immune response. Their involvement in innate immunity warrants

further investigation identifying their functional role in Mesoamerican populations

(Robbins and Morelli 2014). Understanding the functional role of adaptively introgressed

genesion may potentially provide the basis for novel treatments options for disease and

therefore improve infection outcomes.

2.   *Are we able to detect signatures of natural selection in Mesoamericans at*

      immune response loci?

      Chapter 3's findings indicated that Mesoamerican population history has been

shaped by the pressures of infectious disease during the recent evolutionary past. Our

analysis winnowed down hundreds of immune response genes to identify only a handful

that exhibit evidence of positive directional selection. Mirroring findings from other global populations, our results indicate that as in other populations that the MHC complex has been shaped by natural selection in Mesoamerican populations as it remains the region with the most significant results. We also found that the peroxisome proliferator-activated receptor gamma (PPAR-γ) signaling pathway, which regulates the expression of cytokines and chemokines, shows evidence of recent directional selection (Le Menn and Neels 2018). These genes are an excellent starting point for future research studying immune response in modern day Mesoamerican populations.

3.   *What risk factors contribute to dengue-infection in Mesoamerican populations today?*

In our final research chapter, we identified that a lack of water security, storage of uncovered water, and storage of scrap metals and tires all contributed to individual risk for dengue infection. This finding will contribute to ongoing public health initiatives in the region whose goal is to reduce dengue infection rates by highlighting which intervention strategies are working and which ones need to be improved. The risk factors that we identified are generally correlated with lower socioeconomic status and poverty. Future studies devoted to identifying the biological risk factors (e.g. genetic, serological) for dengue infection and disease severity should include these potential covariates.

Future directions

As a follow-up to the epidemiological analysis of dengue, we will be conducting a case-control analysis of exome sequencing data and measuring pro-inflammatory cytokines in serum. We will incorporate the epidemiological risk factors for dengue infection that we identified in chapter 4 into our genetic and cytokine analyses. This will allow us to more accurately model biological factors contributing to dengue susceptibility and disease severity. In our analysis, we will nominate the regions and pathways identified by both the selection scan and the introgression analysis as candidate loci for study. By focusing our efforts on these candidate loci, we will potentially identify functional variants that will give us insights into modern not only in the present, but a glimpse at our evolutionary past. This will result in a new understanding of how infectious disease shaped the genomes of Mesoamericans over time and how evolutionary events affect their immune response to modern pathogenic infection. We hope that our future efforts will provide a framework for incorporating an evolutionary perspective into the study of modern infectious diseases.

REFERENCES

Abi-Rached L, Jobin MJ, Kulkarni S, McWhinnie A, Dalva K, Gragert L, Babrzadeh F, Gharizadeh B, Luo M, Plummer FA, Kimani J, Carrington M, Middleton D, Rajalingam R, Beksac M, Marsh SG, Maiers M, Guethlein LA, Tavoularis S, Little AM, Green RE, Norman PJ, Parham P (2011) The shaping of modern human immune systems by multiregional admixture with archaic humans. Science 334: 89-94. doi: 10.1126/science.1209202

Akey JM (2009) Constructing genomic maps of positive selection in humans: where do we go from here? Genome Res 19: 711-22. doi: 10.1101/gr.086652.108

Akey JM, Zhang G, Zhang K, Jin L, Shriver MD (2002) Interrogating a high-density SNP map for signatures of natural selection. Genome Res 12: 1805-14. doi: 10.1101/gr.631202

Alirol E, Getaz L, Stoll B, Chappuis F, Loutan L (2011) Urbanisation and infectious diseases in a globalised world. The Lancet Infectious Diseases 11: 131-141. doi: https://doi.org/10.1016/S1473-3099(10)70223-1

Amorim CE, Nunes K, Meyer D, Comas D, Bortolini MC, Salzano FM, Hunemeier T (2017) Genetic signature of natural selection in first Americans. Proc Natl Acad Sci U S A 114: 2195-2199. doi: 10.1073/pnas.1620541114

Ardelean CF, Becerra-Valdivia L, Pedersen MW, Schwenninger J-L, Oviatt CG, Macías-Quintero JI, Arroyo-Cabrales J, Sikora M, Ocampo-Díaz YZE, Rubio-Cisneros II, Watling JG, de Medeiros VB, De Oliveira PE, Barba-Pingarón L, Ortiz-Butrón A, Blancas-Vázquez J, Rivera-González I, Solís-Rosales C, Rodríguez-Ceja M, Gandy DA, Navarro-Gutierrez Z, De La Rosa-Díaz JJ, Huerta-Arellano V, Marroquín-Fernández MB, Martínez-Riojas LM, López-Jiménez A, Higham T, Willerslev E (2020) Evidence of human occupation in Mexico around the Last Glacial Maximum. Nature. doi: 10.1038/s41586-020-2509-0

Auld JR, Agrawal AA, Relyea RA (2010) Re-evaluating the costs and limits of adaptive phenotypic plasticity. Proc Biol Sci 277: 503-11. doi: 10.1098/rspb.2009.1355

Avila-Arcos MC, McManus KF, Sandoval K, Rodriguez-Rodriguez JE, Villa-Islas V, Martin AR, Luisi P, Penaloza-Espinosa RI, Eng C, Huntsman S, Burchard EG, Gignoux CR, Bustamante CD, Moreno-Estrada A (2020) Population History and Gene Divergence in Native Mexicans Inferred from 76 Human Exomes. Mol Biol Evol 37: 994-1006. doi: 10.1093/molbev/msz282

Baker AD, Malur A, Barna BP, Kavuru MS, Malur AG, Thomassen MJ (2010) PPARgamma regulates the expression of cholesterol metabolism genes in alveolar macrophages. Biochem Biophys Res Commun 393: 682-7. doi: 10.1016/j.bbrc.2010.02.056

Baker PT (1997) The Raymond Pearl memorial lecture, 1996: The eternal triangle-genes, phenotype, and environment. Am J Hum Biol 9: 93-101. doi: 10.1002/(SICI)1520-6300(1997)9:1<93::AID-AJHB12>3.0.CO;2-U

Bamshad M, Wooding SP (2003) Signatures of natural selection in the human genome. Nat Rev Genet 4: 99-111. doi: 10.1038/nrg999

Bamshad MJ, Mummidi S, Gonzalez E, Ahuja SS, Dunn DM, Watkins WS, Wooding S, Stone AC, Jorde LB, Weiss RB, Ahuja SK (2002) A strong signature of balancing selection in the 5' cis-regulatory region of CCR5. Proc Natl Acad Sci U S A 99: 10539-44. doi: 10.1073/pnas.162046399

Baquir B, Hancock RE (2017) Exosomes, your body's answer to immune health. Ann Transl Med 5: 81. doi: 10.21037/atm.2017.01.50

Batlle E, Massague J (2019) Transforming Growth Factor-beta Signaling in Immunity and Cancer. Immunity 50: 924-940. doi: 10.1016/j.immuni.2019.03.024

Berghe PLvd (1968) Ethnic membership and cultural change in Guatemala. Social Forces 46: 514-522.

Bigham A, Bauchet M, Pinto D, Mao X, Akey JM, Mei R, Scherer SW, Julian CG, Wilson MJ, Lopez Herraez D, Brutsaert T, Parra EJ, Moore LG, Shriver MD (2010) Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. PLoS Genet 6: e1001116. doi: 10.1371/journal.pgen.1001116

Björkbacka H, Kunjathoor VV, Moore KJ, Koehn S, Ordija CM, Lee MA, Means T, Halmen K, Luster AD, Golenbock DT, Freeman MW (2004) Reduced atherosclerosis in MyD88-null mice links elevated serum cholesterol levels to activation of innate immunity signaling pathways. Nature Medicine 10: 416-421. doi: 10.1038/nm1008

Bos KI, Harkins KM, Herbig A, Coscolla M, Weber N, Comas I, Forrest SA, Bryant JM, Harris SR, Schuenemann VJ, Campbell TJ, Majander K, Wilbur AK, Guichon RA, Wolfe Steadman DL, Cook DC, Niemann S, Behr MA, Zumarraga M, Bastida R, Huson D, Nieselt K, Young D, Parkhill J, Buikstra JE, Gagneux S, Stone AC, Krause J (2014) Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. Nature 514: 494-497. doi: 10.1038/nature13591

http://www.nature.com/nature/journal/v514/n7523/abs/nature13591.html#supplementary
     -information

Brathwaite Dick O, San Martín JL, Montoya RH, del Diego J, Zambrano B, Dayan GH
     (2012) The History of Dengue Outbreaks in the Americas. The American Journal
     of Tropical Medicine and Hygiene 87: 584-593. doi: 10.4269/ajtmh.2012.11-0770

Browning BL, Zhou Y, Browning SR (2018a) A One-Penny Imputed Genome from Next-
     Generation Reference Panels. Am J Hum Genet 103: 338-348. doi:
     10.1016/j.ajhg.2018.07.015

Browning SR, Browning BL (2007) Rapid and accurate haplotype phasing and missing-
     data inference for whole-genome association studies by use of localized
     haplotype clustering. Am J Hum Genet 81: 1084-97. doi: 10.1086/521987

Browning SR, Browning BL, Zhou Y, Tucci S, Akey JM (2018b) Analysis of Human
     Sequence Data Reveals Two Pulses of Archaic Denisovan Admixture. Cell 173:
     53-61.e9. doi: https://doi.org/10.1016/j.cell.2018.02.031

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ (2015) Second-
     generation PLINK: rising to the challenge of larger and richer datasets.
     Gigascience 4: 7. doi: 10.1186/s13742-015-0047-8

Chawla A (2010) Control of macrophage activation and function by PPARs. Circulation
     research 106: 1559-1569. doi: 10.1161/CIRCRESAHA.110.216523

Chen E, Fisher EB, Bacharier LB, Strunk RC (2003) Socioeconomic status, stress, and
     immune markers in adolescents with asthma. Psychosomatic medicine 65: 984-
     992.

Chen L, Wolf AB, Fu W, Li L, Akey JM (2020) Identifying and Interpreting Apparent
     Neanderthal Ancestry in African Individuals. Cell 180: 677-687 e16. doi:
     10.1016/j.cell.2020.01.012

Chinetti G, Lestavel S, Bocher V, Remaley AT, Neve B, Torra IP, Teissier E, Minnich A,
     Jaye M, Duverger N, Brewer HB, Fruchart JC, Clavey V, Staels B (2001) PPAR-
     alpha and PPAR-gamma activators induce cholesterol removal from human
     macrophage foam cells through stimulation of the ABCA1 pathway. Nat Med 7:
     53-8. doi: 10.1038/83348

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE,
     Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, Genomes Project Analysis
     G (2011) The variant call format and VCFtools. Bioinformatics 27: 2156-8. doi:
     10.1093/bioinformatics/btr330

Dannemann M, Andres AM, Kelso J (2016) Introgression of Neandertal- and Denisovan-like Haplotypes Contributes to Adaptive Variation in Human Toll-like Receptors. Am J Hum Genet 98: 22-33. doi: 10.1016/j.ajhg.2015.11.015

Daviglus ML, Talavera GA, Avilés-Santa ML, Allison M, Cai J, Criqui MH, Gellman M, Giachello AL, Gouskova N, Kaplan RC, LaVange L, Penedo F, Perreira K, Pirzada A, Schneiderman N, Wassertheil-Smoller S, Sorlie PD, Stamler J (2012) Prevalence of Major Cardiovascular Risk Factors and Cardiovascular Diseases Among Hispanic/Latino Individuals of Diverse Backgrounds in the United States. JAMA 308: 1775-1784. doi: 10.1001/jama.2012.14517

de Bakker PIW, McVean G, Sabeti PC, Miretti MM, Green T, Marchini J, Ke X, Monsuur AJ, Whittaker P, Delgado M, Morrison J, Richardson A, Walsh EC, Gao X, Galver L, Hart J, Hafler DA, Pericak-Vance M, Todd JA, Daly MJ, Trowsdale J, Wijmenga C, Vyse TJ, Beck S, Murray SS, Carrington M, Gregory S, Deloukas P, Rioux JD (2006) A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. Nature Genetics 38: 1166-1172. doi: 10.1038/ng1885

Diseases tWHOatSPfRaTiT (2009) Dengue: guidelines for diagnosis, treatment, prevention and control -- New edition. WHO Press, Geneva

Dobzhansky T (1973) Nothing in Biology Makes Sense Except in the Light of Evolution. The American Biology Teacher 75: 87-91. doi: 10.2307/4444260

Feldman LH (1999) The war against epidemics in colonial Guatemala,1519-1821.

Fierro NA, Gonzalez-Aldaco K, Roman S, Panduro A (2017) Chapter 9 - The Immune System and Viral Hepatitis. In: Muriel P (ed) Liver Pathophysiology. Academic Press, Boston, pp 129-139

Flannery KV (2019) The Early Mesoamerican Village: Updated Edition. Routledge

Fragoso JM, Martínez-Ríos MA, Alvarez-León E, Vallejo M, Peña-Duque MA, Posadas-Sánchez R, Posadas-Romero C, Furuzawa-Carballeda J, Lima G, Llorente L, Vargas-Alarcón G (2012) The T29C polymorphism of the transforming growth factor-β1 (TGF-β1) gene is associated with genetic susceptibility to acute coronary syndrome in Mexican patients. Cytokine 58: 380-383. doi: https://doi.org/10.1016/j.cyto.2012.03.004

Fumagalli M, Moltke I, Grarup N, Racimo F, Bjerregaard P, Jorgensen ME, Korneliussen TS, Gerbault P, Skotte L, Linneberg A, Christensen C, Brandslund I, Jorgensen T, Huerta-Sanchez E, Schmidt EB, Pedersen O, Hansen T, Albrechtsen A, Nielsen R (2015) Greenlandic Inuit show genetic signatures of diet and climate adaptation. Science 349: 1343-7. doi: 10.1126/science.aab2319

Fumagalli M, Pozzoli U, Cagliani R, Comi GP, Bresolin N, Clerici M, Sironi M (2010) Genome-Wide Identification of Susceptibility Alleles for Viral Infections through a Population Genetics Approach. Plos Genetics 6. doi: ARTN e1000849
10.1371/journal.pgen.1000849

Galli SJ, Borregaard N, Wynn TA (2011) Phenotypic and functional plasticity of cells of innate immunity: macrophages, mast cells and neutrophils. Nat Immunol 12: 1035-44. doi: 10.1038/ni.2109

Galvani AP, Novembre J (2005) The evolutionary history of the CCR5-Delta32 HIV-resistance mutation. Microbes Infect 7: 302-9. doi: 10.1016/j.micinf.2004.12.006

Galvani AP, Slatkin M (2003) Evaluating plague and smallpox as historical selective pressures for the CCR5-Delta 32 HIV-resistance allele. Proc Natl Acad Sci U S A 100: 15276-9. doi: 10.1073/pnas.2435085100

Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR (2015) A global reference for human genetic variation. Nature 526: 68-74. doi: 10.1038/nature15393

Ghosh S (2011) Macrophage cholesterol homeostasis and metabolic diseases: critical role of cholesteryl ester mobilization. Expert review of cardiovascular therapy 9: 329-340. doi: 10.1586/erc.11.16

Gluckman PD, Hanson MA, Mitchell MD (2010) Developmental origins of health and disease: reducing the burden of chronic disease in the next generation. Genome Med 2: 14. doi: 10.1186/gm135

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH, Hansen NF, Durand EY, Malaspinas AS, Jensen JD, Marques-Bonet T, Alkan C, Prufer K, Meyer M, Burbano HA, Good JM, Schultz R, Aximu-Petri A, Butthof A, Hober B, Hoffner B, Siegemund M, Weihmann A, Nusbaum C, Lander ES, Russ C, Novod N, Affourtit J, Egholm M, Verna C, Rudan P, Brajkovic D, Kucan Z, Gusic I, Doronichev VB, Golovanova LV, Lalueza-Fox C, de la Rasilla M, Fortea J, Rosas A, Schmitz RW, Johnson PLF, Eichler EE, Falush D, Birney E, Mullikin JC, Slatkin M, Nielsen R, Kelso J, Lachmann M, Reich D, Paabo S (2010) A draft sequence of the Neandertal genome. Science 328: 710-722. doi: 10.1126/science.1188021

Green RE, Krause J, Ptak SE, Briggs AW, Ronan MT, Simons JF, Du L, Egholm M, Rothberg JM, Paunovic M, Paabo S (2006) Analysis of one million base pairs of Neanderthal DNA. Nature 444: 330-6. doi: 10.1038/nature05336

Grossman SR, Andersen KG, Shlyakhter I, Tabrizi S, Winnicki S, Yen A, Park DJ, Griesemer D, Karlsson EK, Wong SH, Cabili M, Adegbola RA, Bamezai RN, Hill

AV, Vannberg FO, Rinn JL, Genomes P, Lander ES, Schaffner SF, Sabeti PC (2013) Identifying recent adaptations in large-scale genomic data. Cell 152: 703-13. doi: 10.1016/j.cell.2013.01.035

Gubler DJ (2011) Dengue, Urbanization and Globalization: The Unholy Trinity of the 21st Century. Tropical Medicine and Health 39: S3-S11. doi: 10.2149/tmh.2011-S05

Gubler DJ, Clark GG (1995) Dengue Dengue Hemorrhagic-Fever - the Emergence of a Global Health Problem. Emerging Infectious Diseases 1: 55-57.

Hawks J, Wang ET, Cochran GM, Harpending HC, Moyzis RK (2007) Recent acceleration of human adaptive evolution. Proc Natl Acad Sci U S A 104: 20753-8. doi: 10.1073/pnas.0707650104

Holmes EC, Twiddy SS (2003) The origin, emergence and evolutionary genetics of dengue virus. Infect Genet Evol 3: 19-28. doi: 10.1016/s1567-1348(03)00004-2

Horstick O, Tozan Y, Wilder-Smith A (2015) Reviewing dengue: still a neglected tropical disease? PLoS Negl Trop Dis 9: e0003632.

Huang da W, Sherman BT, Lempicki RA (2009a) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37: 1-13. doi: 10.1093/nar/gkn923

Huang da W, Sherman BT, Lempicki RA (2009b) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4: 44-57. doi: 10.1038/nprot.2008.211

Huerta-Sanchez E, Jin X, Asan, Bianba Z, Peter BM, Vinckenbosch N, Liang Y, Yi X, He M, Somel M, Ni P, Wang B, Ou X, Huasang, Luosang J, Cuo ZX, Li K, Gao G, Yin Y, Wang W, Zhang X, Xu X, Yang H, Li Y, Wang J, Wang J, Nielsen R (2014) Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature 512: 194-7. doi: 10.1038/nature13408

Hughes AL, Yeager M (1998a) Natural selection and the evolutionary history of major histocompatibility complex loci. Front Biosci 3: d509-16. doi: 10.2741/a298

Hughes AL, Yeager M (1998b) Natural selection at major histocompatibility complex loci of vertebrates. Annual review of genetics 32: 415-435.

International HapMap C (2003) The International HapMap Project. Nature 426: 789-96. doi: 10.1038/nature02168

Jablonski NG, Chaplin G (2000) The evolution of human skin coloration. J Hum Evol 39: 57-106. doi: 10.1006/jhev.2000.0403

Jablonski NG, Chaplin G (2013) Epidermal pigmentation in the human lineage is an adaptation to ultraviolet radiation. J Hum Evol 65: 671-5. doi: 10.1016/j.jhevol.2013.06.004

Kagan JC (2012) Signaling organelles of the innate immune system. Cell 151: 1168-78. doi: 10.1016/j.cell.2012.11.011

Kakoola DN, Curcio-Brint A, Lenchik NI, Gerling IC (2014) Molecular pathway alterations in CD4 T-cells of nonobese diabetic (NOD) mice in the preinsulitis phase of autoimmune diabetes. Results in immunology 4: 30-45. doi: 10.1016/j.rinim.2014.05.001

Karlsson EK, Harris JB, Tabrizi S, Rahman A, Shlyakhter I, Patterson N, O'Dushlaine C, Schaffner SF, Gupta S, Chowdhury F, Sheikh A, Shin OS, Ellis C, Becker CE, Stuart LM, Calderwood SB, Ryan ET, Qadri F, Sabeti PC, Larocque RC (2013) Natural selection in a bangladeshi population from the cholera-endemic ganges river delta. Sci Transl Med 5: 192ra86. doi: 10.1126/scitranslmed.3006338

Kendall C, Hudelson P, Leontsini E, Winch P, Lloyd L (1991) Urbanization, Dengue, and the Health Transition - Anthropological Contributions to International Health. Medical Anthropology Quarterly 5: 257-268. doi: DOI 10.1525/maq.1991.5.3.02a00050

Klaus HD, Wilbur AK, Temple DH, Buikstra JE, Stone AC, Fernandez M, Wester C, Tam ME (2010) Tuberculosis on the north coast of Peru: skeletal and molecular paleopathology of late pre-Hispanic and postcontact mycobacterial disease. Journal of Archaeological Science 37: 2587-2597. doi: 10.1016/j.jas.2010.05.019

Ko A, Cantor RM, Weissglas-Volkov D, Nikkola E, Reddy PM, Sinsheimer JS, Pasaniuc B, Brown R, Alvarez M, Rodriguez A, Rodriguez-Guillen R, Bautista IC, Arellano-Campos O, Munoz-Hernandez LL, Salomaa V, Kaprio J, Jula A, Jauhiainen M, Heliovaara M, Raitakari O, Lehtimaki T, Eriksson JG, Perola M, Lohmueller KE, Matikainen N, Taskinen MR, Rodriguez-Torres M, Riba L, Tusie-Luna T, Aguilar-Salinas CA, Pajukanta P (2014) Amerindian-specific regions under positive selection harbour new lipid variants in Latinos. Nat Commun 5: 3983. doi: 10.1038/ncomms4983

Kuzawa CW, Thayer ZM (2011) Timescales of human adaptation: the role of epigenetic processes. Epigenomics 3: 221-34. doi: 10.2217/epi.11.11

Kyle JL, Harris E (2008) Global Spread and Persistence of Dengue. Annual Review of Microbiology 62: 71-92. doi: 10.1146/annurev.micro.62.081307.163005

Lau JW, Lehnert E, Sethi A, Malhotra R, Kaushik G, Onder Z, Groves-Kirkby N, Mihajlovic A, DiGiovanna J, Srdic M, Bajcic D, Radenkovic J, Mladenovic V, Krstanovic D, Arsenijevic V, Klisic D, Mitrovic M, Bogicevic I, Kural D, Davis-

Dusenbery B, Seven Bridges CGCT (2017) The Cancer Genomics Cloud: Collaborative, Reproducible, and Democratized-A New Paradigm in Large-Scale Computational Research. Cancer Res 77: e3-e6. doi: 10.1158/0008-5472.CAN-17-0387

Le Menn G, Neels JG (2018) Regulation of Immune Cell Function by PPARs and the Connection with Metabolic and Neurodegenerative Diseases. Int J Mol Sci 19: 1575. doi: 10.3390/ijms19061575

Leon-Portilla M (2011) The Broken Spears 2007 Revised Edition: The Aztec Account of the Conquest of Mexico. Beacon Press

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078-9. doi: 10.1093/bioinformatics/btp352

Li K, Wen M, Henry KA (2017) Ethnic density, immigrant enclaves, and Latino health risks: A propensity score matching approach. Social Science & Medicine 189: 44-52. doi: https://doi.org/10.1016/j.socscimed.2017.07.019

Li T, Francl JM, Boehme S, Chiang JYL (2013) Regulation of cholesterol and bile acid homeostasis by the cholesterol 7α-hydroxylase/steroid response element-binding protein 2/microRNA-33a axis in mice. Hepatology (Baltimore, Md.) 58: 1111-1121. doi: 10.1002/hep.26427

Lin YL, Pavlidis P, Karakoc E, Ajay J, Gokcumen O (2015) The evolution and functional impact of human deletion variants shared with archaic hominin genomes. Mol Biol Evol 32: 1008-19. doi: 10.1093/molbev/msu405

Lindo J, Huerta-Sanchez E, Nakagome S, Rasmussen M, Petzelt B, Mitchell J, Cybulski JS, Willerslev E, DeGiorgio M, Malhi RS (2016) A time transect of exomes from a Native American population before and after European contact. Nat Commun 7: 13175. doi: 10.1038/ncomms13175

Liu R, Paxton WA, Choe S, Ceradini D, Martin SR, Horuk R, MacDonald ME, Stuhlmann H, Koup RA, Landau NR (1996) Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. Cell 86: 367-377.

Lock MM, Nguyen V-K (2010) An anthropology of biomedicine. Wiley-Blackwell, Chichester, West Sussex ; Malden, MA

Lockhart J (1992) The Nahuas after the conquest : a social and cultural history of the Indians of central Mexico, sixteenth through eighteenth centuries. Stanford University Press, Stanford, Calif.

Machuca Gallegos L (2016) El ocaso de la encomienda en Yucatán, 1770-1821. Estudios de Historia Novohispana 54: 31-49. doi: 10.1016/j.ehn.2015.02.001

Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, Zhao M, Chennagiri N, Nordenfelt S, Tandon A, Skoglund P, Lazaridis I, Sankararaman S, Fu Q, Rohland N, Renaud G, Erlich Y, Willems T, Gallo C, Spence JP, Song YS, Poletti G, Balloux F, van Driem G, de Knijff P, Romero IG, Jha AR, Behar DM, Bravi CM, Capelli C, Hervig T, Moreno-Estrada A, Posukh OL, Balanovska E, Balanovsky O, Karachanak-Yankova S, Sahakyan H, Toncheva D, Yepiskoposyan L, Tyler-Smith C, Xue Y, Abdullah MS, Ruiz-Linares A, Beall CM, Di Rienzo A, Jeong C, Starikovskaya EB, Metspalu E, Parik J, Villems R, Henn BM, Hodoglugil U, Mahley R, Sajantila A, Stamatoyannopoulos G, Wee JT, Khusainova R, Khusnutdinova E, Litvinov S, Ayodo G, Comas D, Hammer MF, Kivisild T, Klitz W, Winkler CA, Labuda D, Bamshad M, Jorde LB, Tishkoff SA, Watkins WS, Metspalu M, Dryomov S, Sukernik R, Singh L, Thangaraj K, Paabo S, Kelso J, Patterson N, Reich D (2016) The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature 538: 201-206. doi: 10.1038/nature18964

Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM (2010) Robust relationship inference in genome-wide association studies. Bioinformatics 26: 2867-73. doi: 10.1093/bioinformatics/btq559

Maples BK, Gravel S, Kenny EE, Bustamante CD (2013) RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. Am J Hum Genet 93: 278-88. doi: 10.1016/j.ajhg.2013.06.020

Martínez-Campos C, Torres-Poveda K, Camorlinga-Ponce M, Flores-Luna L, Maldonado-Bernal C, Madrid-Marina V, Torres J (2019) Polymorphisms in IL-10 and TGF-β gene promoter are associated with lower risk to gastric cancer in a Mexican population. BMC Cancer 19: 453. doi: 10.1186/s12885-019-5627-z

McDade TW (2003) Life history theory and the immune system: steps toward a human ecological immunology. Am J Phys Anthropol Suppl 37: 100-25. doi: 10.1002/ajpa.10398

Mendez FL, Watkins JC, Hammer MF (2012a) Global genetic variation at OAS1 provides evidence of archaic admixture in Melanesian populations. Mol Biol Evol 29: 1513-20. doi: 10.1093/molbev/msr301

Mendez FL, Watkins JC, Hammer MF (2012b) A haplotype at STAT2 Introgressed from neanderthals and serves as a candidate of positive selection in Papua New Guinea. The American Journal of Human Genetics 91: 265-274.

Merbs CF (1992) A new world of infectious disease. American Journal of Physical Anthropology 35: 3-42. doi: 10.1002/ajpa.1330350603

Meyer M, Kircher M, Gansauge MT, Li H, Racimo F, Mallick S, Schraiber JG, Jay F, Prufer K, de Filippo C, Sudmant PH, Alkan C, Fu Q, Do R, Rohland N, Tandon A, Siebauer M, Green RE, Bryc K, Briggs AW, Stenzel U, Dabney J, Shendure J, Kitzman J, Hammer MF, Shunkov MV, Derevianko AP, Patterson N, Andres AM, Eichler EE, Slatkin M, Reich D, Kelso J, Paabo S (2012) A high-coverage genome sequence from an archaic Denisovan individual. Science 338: 222-6. doi: 10.1126/science.1224344

MHC sequencing consortium T (1999) Complete sequence and gene map of a human major histocompatibility complex. The MHC sequencing consortium. Nature 401: 921-3. doi: 10.1038/44853

Moreno-Estrada A, Gignoux CR, Fernandez-Lopez JC, Zakharia F, Sikora M, Contreras AV, Acuna-Alonzo V, Sandoval K, Eng C, Romero-Hidalgo S, Ortiz-Tello P, Robles V, Kenny EE, Nuno-Arana I, Barquera-Lozano R, Macin-Perez G, Granados-Arriola J, Huntsman S, Galanter JM, Via M, Ford JG, Chapela R, Rodriguez-Cintron W, Rodriguez-Santana JR, Romieu I, Sienra-Monge JJ, del Rio Navarro B, London SJ, Ruiz-Linares A, Garcia-Herrera R, Estrada K, Hidalgo-Miranda A, Jimenez-Sanchez G, Carnevale A, Soberon X, Canizales-Quinteros S, Rangel-Villalobos H, Silva-Zolezzi I, Burchard EG, Bustamante CD (2014) Human genetics. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. Science 344: 1280-5. doi: 10.1126/science.1251688

Muehlenbein MP (2010) Evolutionary Medicine, Immunity, and Infectious Disease. In: Muehlenbein MP (ed) Human evolutionary biology. Cambridge University Press, Cambridge ; New York, pp 459-490

Mummert A, Esche E, Robinson J, Armelagos GJ (2011) Stature and robusticity during the agricultural transition: evidence from the bioarchaeological record. Econ Hum Biol 9: 284-301. doi: 10.1016/j.ehb.2011.03.004

Mychaleckyj JC, Havt A, Nayak U, Pinkerton R, Farber E, Concannon P, Lima AA, Guerrant RL (2017) Genome-Wide Analysis in Brazilians Reveals Highly Differentiated Native American Genome Regions. Mol Biol Evol 34: 559-574. doi: 10.1093/molbev/msw249

Navarro-Sanchez E, Despres P, Cedillo-Barron L (2005) Innate immune responses to dengue virus. Arch Med Res 36: 425-35. doi: 10.1016/j.arcmed.2005.04.007

Neph S, Kuehn MS, Reynolds AP, Haugen E, Thurman RE, Johnson AK, Rynes E, Maurano MT, Vierstra J, Thomas S, Sandstrom R, Humbert R, Stamatoyannopoulos JA (2012) BEDOPS: high-performance genomic feature operations. Bioinformatics 28: 1919-20. doi: 10.1093/bioinformatics/bts277

Nesse RM, Stearns SC (2008) The great opportunity: Evolutionary applications to medicine and public health. Evol Appl 1: 28-48. doi: 10.1111/j.1752-4571.2007.00006.x

Nish S, Medzhitov R (2011) Host defense pathways: role of redundancy and compensation in infectious disease phenotypes. Immunity 34: 629-36. doi: 10.1016/j.immuni.2011.05.009

Nogueira RMR, Miagostovich MP, Schatzmayr HG, dos Santos FB, de Araújo ES, de Filippis AMB, de Souza RV, Zagne SMO, Nicolai C, Baran M (1999) Dengue in the state of Rio de Janeiro, Brazil, 1986-1998. Memórias do Instituto Oswaldo Cruz 94: 297-304.

Normile D (2013) First New Dengue Virus Type in 50 Years. ScienceIncider. AAAS

O'Fallon BD, Fehren-Schmitz L (2011) Native Americans experienced a strong population bottleneck coincident with European contact. Proc Natl Acad Sci U S A 108: 20444-8. doi: 10.1073/pnas.1112563108

Oliver-Smith A (1991) Successes and failures in post-disaster resettlement. Disasters 15: 12-23.

Owen N, Poulton T, Hay FC, Mohamed-Ali V, Steptoe A (2003) Socioeconomic status, C-reactive protein, immune factors, and responses to acute mental stress. Brain, behavior, and immunity 17: 286-295.

Pajic P, Lin YL, Xu D, Gokcumen O (2016) The psoriasis-associated deletion of late cornified envelope genes LCE3B and LCE3C has been maintained under balancing selection since Human Denisovan divergence. BMC Evol Biol 16: 265. doi: 10.1186/s12862-016-0842-6

Park DJ, Lukens AK, Neafsey DE, Schaffner SF, Chang HH, Valim C, Ribacke U, Van Tyne D, Galinsky K, Galligan M, Becker JS, Ndiaye D, Mboup S, Wiegand RC, Hartl DL, Sabeti PC, Wirth DF, Volkman SK (2012) Sequence-based association and selection scans identify drug resistance loci in the Plasmodium falciparum malaria parasite. Proc Natl Acad Sci U S A 109: 13052-7. doi: 10.1073/pnas.1210585109

Perry GH, Foll M, Grenier JC, Patin E, Nedelec Y, Pacis A, Barakatt M, Gravel S, Zhou X, Nsobya SL, Excoffier L, Quintana-Murci L, Dominy NJ, Barreiro LB (2014) Adaptive, convergent origins of the pygmy phenotype in African rainforest hunter-gatherers. Proc Natl Acad Sci U S A 111: E3596-603. doi: 10.1073/pnas.1402875111

Pickrell JK, Coop G, Novembre J, Kudaravalli S, Li JZ, Absher D, Srinivasan BS, Barsh GS, Myers RM, Feldman MW, Pritchard JK (2009) Signals of recent positive

selection in a worldwide sample of human populations. Genome Res 19: 826-37. doi: 10.1101/gr.087577.108

Pinheiro FP, Corber SJ (1997) Global situation of dengue and dengue haemorrhagic fever, and its emergence in the Americas. World health statistics quarterly 50: 161-169.

Plagnol V, Wall JD (2006) Possible ancestral structure in human populations. PLoS Genet 2: e105. doi: 10.1371/journal.pgen.0020105

Popejoy AB, Fullerton SM (2016) Genomics is failing on diversity. Nature 538: 161-164. doi: 10.1038/538161a

Pritchard JK, Stephens M, Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. Genetics 155: 945.

Prufer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G, Sudmant PH, de Filippo C, Li H, Mallick S, Dannemann M, Fu Q, Kircher M, Kuhlwilm M, Lachmann M, Meyer M, Ongyerth M, Siebauer M, Theunert C, Tandon A, Moorjani P, Pickrell J, Mullikin JC, Vohr SH, Green RE, Hellmann I, Johnson PL, Blanche H, Cann H, Kitzman JO, Shendure J, Eichler EE, Lein ES, Bakken TE, Golovanova LV, Doronichev VB, Shunkov MV, Derevianko AP, Viola B, Slatkin M, Reich D, Kelso J, Paabo S (2014) The complete genome sequence of a Neanderthal from the Altai Mountains. Nature 505: 43-9. doi: 10.1038/nature12886

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559-75. doi: 10.1086/519795

Quinlan AR (2014) BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr Protoc Bioinformatics 47: 11 12 1-34. doi: 10.1002/0471250953.bi1112s47

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841-2. doi: 10.1093/bioinformatics/btq033

Racimo F, Marnetto D, Huerta-Sanchez E (2017) Signatures of Archaic Adaptive Introgression in Present-Day Human Populations. Mol Biol Evol 34: 296-317. doi: 10.1093/molbev/msw216

Racimo F, Sankararaman S, Nielsen R, Huerta-Sanchez E (2015) Evidence for archaic adaptive introgression in humans. Nat Rev Genet 16: 359-71. doi: 10.1038/nrg3936

Raj A, Stephens M, Pritchard JK (2014) fastSTRUCTURE: variational inference of population structure in large SNP data sets. Genetics 197: 573-89. doi: 10.1534/genetics.114.164350

Reed SG (1999) TGF-beta in infections and infectious diseases. Microbes Infect 1: 1313-25. doi: 10.1016/s1286-4579(99)00252-x

Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, Parra MV, Rojas W, Duque C, Mesa N, Garcia LF, Triana O, Blair S, Maestre A, Dib JC, Bravi CM, Bailliet G, Corach D, Hunemeier T, Bortolini MC, Salzano FM, Petzl-Erler ML, Acuna-Alonzo V, Aguilar-Salinas C, Canizales-Quinteros S, Tusie-Luna T, Riba L, Rodriguez-Cruz M, Lopez-Alarcon M, Coral-Vazquez R, Canto-Cetina T, Silva-Zolezzi I, Fernandez-Lopez JC, Contreras AV, Jimenez-Sanchez G, Gomez-Vazquez MJ, Molina J, Carracedo A, Salas A, Gallo C, Poletti G, Witonsky DB, Alkorta-Aranburu G, Sukernik RI, Osipova L, Fedorova SA, Vasquez R, Villena M, Moreau C, Barrantes R, Pauls D, Excoffier L, Bedoya G, Rothhammer F, Dugoujon JM, Larrouy G, Klitz W, Labuda D, Kidd J, Kidd K, Di Rienzo A, Freimer NB, Price AL, Ruiz-Linares A (2012) Reconstructing Native American population history. Nature 488: 370-4. doi: 10.1038/nature11258

Reich D, Thangaraj K, Patterson N, Price AL, Singh L (2009) Reconstructing Indian population history. Nature 461: 489-94. doi: 10.1038/nature08365

Reiter P (2001) Climate change and mosquito-borne disease. Environmental health perspectives 109: 141-161.

Restall M, Sousa L, Terraciano K (2005) Mesoamerican voices : native-language writings from Colonial Mexico, Oaxaca, Yucatan, and Guatemala. Cambridge University Press, Cambridge ; New York

Reynolds AW, Mata-Miguez J, Miro-Herrans A, Briggs-Cloud M, Sylestine A, Barajas-Olmos F, Garcia-Ortiz H, Rzhetskaya M, Orozco L, Raff JA, Hayes MG, Bolnick DA (2019) Comparing signals of natural selection between three Indigenous North American populations. Proc Natl Acad Sci U S A 116: 9312-9317. doi: 10.1073/pnas.1819467116

Riding A (1977) US Food Aid Seen Hurting Guatemala. New York Times: 53.

Robbins PD, Morelli AE (2014) Regulation of immune responses by extracellular vesicles. Nature reviews. Immunology 14: 195-208. doi: 10.1038/nri3622

Sabeti P, Usen S, Farhadian S, Jallow M, Doherty T, Newport M, Pinder M, Ward R, Kwiatkowski D (2002) CD40L association with protection from severe malaria. Genes Immun 3: 286-91. doi: 10.1038/sj.gene.6363877

Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, Palma A, Mikkelsen TS, Altshuler D, Lander ES (2006) Positive natural selection in the human lineage. science 312: 1614-1620.

Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R, Schaffner SF, Lander ES, International HapMap C, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Shen Y, Sun W, Wang H, Wang Y, Wang Y, Xiong X, Xu L, Waye MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallee C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, et al. (2007) Genome-wide detection and characterization of positive selection in human populations. Nature 449: 913-8. doi: 10.1038/nature06250

San Martín JL, Brathwaite-Dick O (2007) La estrategia de gestión integrada para la prevención y el control del dengue en la región de las Américas. Revista Panamericana de Salud Pública 21: 55-63.

Sankararaman S, Mallick S, Dannemann M, Prufer K, Kelso J, Paabo S, Patterson N, Reich D (2014) The genomic landscape of Neanderthal ancestry in present-day humans. Nature 507: 354-7. doi: 10.1038/nature12961

Sankararaman S, Mallick S, Patterson N, Reich D (2016) The Combined Landscape of Denisovan and Neanderthal Ancestry in Present-Day Humans. Curr Biol 26: 1241-7. doi: 10.1016/j.cub.2016.03.037

Schwarzenbacher H, Dolezal M, Flisikowski K, Seefried F, Wurmser C, Schlotterer C, Fries R (2012) Combining evidence of selection with association analysis increases power to detect regions influencing complex traits in dairy cattle. BMC Genomics 13: 48. doi: 10.1186/1471-2164-13-48

Shriver MD, Kennedy GC, Parra EJ, Lawson HA, Sonpar V, Huang J, Akey JM, Jones KW (2004) The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. Hum Genomics 1: 274-86. doi: 10.1186/1479-7364-1-4-274

Silveira PP, Portella AK, Goldani MZ, Barbieri MA (2007) Developmental origins of health and disease (DOHaD). J Pediatr (Rio J) 83: 494-504. doi: 10.2223/JPED.1728

Smith JM (1976) A Comment on the Red Queen. The American Naturalist 110: 325-330.

Smith ME (2016) City Size in Late Postclassic Mesoamerica. Journal of Urban History 31: 403-434. doi: 10.1177/0096144204274396

Sprenger PRD (1987) The used tire trade: a mechanism for the worldwide dispersal of container breeding mosquitoes. J Am Mosq Control Assoc 3: 494.

Steverding D (2014) The history of Chagas disease. Parasit Vectors 7: 317. doi: 10.1186/1756-3305-7-317

Szpiech ZA, Hernandez RD (2014) selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. Mol Biol Evol 31: 2824-7. doi: 10.1093/molbev/msu211

Taskent RO, Alioglu ND, Fer E, Melike Donertas H, Somel M, Gokcumen O (2017) Variation and Functional Impact of Neanderthal Ancestry in Western Asia. Genome Biol Evol 9: 3516-3524. doi: 10.1093/gbe/evx216

Thayer ZM, Kuzawa CW (2014) Early origins of health disparities: material deprivation predicts maternal evening cortisol in pregnancy and offspring cortisol reactivity in the first few weeks of life. Am J Hum Biol 26: 723-30. doi: 10.1002/ajhb.22532

Trowsdale J (1993) Genomic structure and function in the MHC. Trends in Genetics 9: 117-122. doi: 10.1016/0168-9525(93)90205-V

Van Valen L (1973) A New Evolutionary Law. Evolutionary Theory 1: 1-30.

Vernot B, Akey JM (2014) Resurrecting surviving Neandertal lineages from modern human genomes. Science 343: 1017-21. doi: 10.1126/science.1245938

Villanea FA, Schraiber JG (2019) Multiple episodes of interbreeding between Neanderthal and modern humans. Nature Ecology & Evolution 3: 39-44. doi: 10.1038/s41559-018-0735-8

Villegas-Trejo A, Che-Mendoza A, González-Fernández M, Guillermo-May G, González-Bejarano H, Dzul-Manzanilla F, Ulloa-García A, Danis-Lozano R, Manrique-Saide P (2011) Control enfocado de Aedes aegypti en localidades de alto riesgo de transmisión de dengue en Morelos, México. salud pública de méxico 53: 141-151.

Voight BF, Kudaravalli S, Wen X, Pritchard JK (2006) A map of recent positive selection in the human genome. PLoS Biol 4: e72. doi: 10.1371/journal.pbio.0040072

Wang S, Lewis CM, Jakobsson M, Ramachandran S, Ray N, Bedoya G, Rojas W, Parra MV, Molina JA, Gallo C, Mazzotti G, Poletti G, Hill K, Hurtado AM, Labuda D,

Klitz W, Barrantes R, Bortolini MC, Salzano FM, Petzl-Erler ML, Tsuneto LT, Llop E, Rothhammer F, Excoffier L, Feldman MW, Rosenberg NA, Ruiz-Linares A (2007) Genetic variation and population structure in native Americans. PLoS Genet 3: e185. doi: 10.1371/journal.pgen.0030185

Weir BS, Cockerham CC (1984) Estimating F-Statistics for the Analysis of Population Structure. Evolution 38: 1358. doi: 10.2307/2408641

Whitehorn J, Simmons CP (2011) The pathogenesis of dengue. Vaccine 29: 7221-7228. doi: http://dx.doi.org/10.1016/j.vaccine.2011.07.022

Wisser RJ, Murray SC, Kolkman JM, Ceballos H, Nelson RJ (2008) Selection mapping of loci for quantitative disease resistance in a diverse maize population. Genetics 180: 583-99. doi: 10.1534/genetics.108.090118

Worthington JJ, Fenton TM, Czajkowska BI, Klementowicz JE, Travis MA (2012) Regulation of TGFbeta in the immune system: an emerging role for integrins and dendritic cells. Immunobiology 217: 1259-65. doi: 10.1016/j.imbio.2012.06.009

Yadav H, Quijano C, Kamaraju AK, Gavrilova O, Malek R, Chen W, Zerfas P, Zhigang D, Wright EC, Stuelten C, Sun P, Lonning S, Skarulis M, Sumner AE, Finkel T, Rane SG (2011) Protection from obesity and diabetes by blockade of TGF-beta/Smad3 signaling. Cell Metab 14: 67-79. doi: 10.1016/j.cmet.2011.04.013

Zanetti D, Carreras-Torres R, Esteban E, Via M, Moral P (2015) Potential Signals of Natural Selection in the Top Risk Loci for Coronary Artery Disease: 9p21 and 10q11. PLoS One 10: e0134840. doi: 10.1371/journal.pone.0134840

Zavala G, Long KZ, García OP, Caamaño MdC, Aguilar T, Salgado LM, Rosado JL (2013) Specific micronutrient concentrations are associated with inflammatory cytokines in a rural population of Mexican women with a high prevalence of obesity. British Journal of Nutrition 109: 686-694. doi: 10.1017/S0007114512001912

Zhang G, Lu J, Yang M, Wang Y, Liu H, Xu C (2020) Elevated GALNT10 expression identifies immunosuppressive microenvironment and dismal prognosis of patients with high grade serous ovarian cancer. Cancer Immunol Immunother 69: 175-187. doi: 10.1007/s00262-019-02454-1