PALEOGENOMICS IN THE AMERICAS: DEMOGRAPHY, ADAPTATIONS, AND RESILIENCE

BY

JOHN LINDO

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Anthropology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2015

Urbana, Illinois

Doctoral Committee:

Associate Professor Ripan S. Malhi, Chair
Associate Professor Charles Roseman
Professor Alejandro Lugo
Assistant Professor Michael DeGiorgio, Pennsylvania State University

ABSTRACT

This work covers the evolutionary history of an indigenous population from the Americas through the lens of ancient DNA. In recent years, major advances in DNA sequencing have allowed for the utilization of ancient DNA on a grand scale. We now have the ability to examine the whole genomes of ancient individuals and examine the forces of evolution though space and time. Through these advances, this project addresses questions that would have otherwise been difficult to address with modern DNA alone. These questions span a period of 10,000 years and broach both local and continental demographic histories. In doing so, a comprehensive genetic picture of an indigenous population, with a history characteristic of other native groups, emerges so as to explain nuanced migration patterns, the historical experiences of Native Americans with European-borne pathogens, and the long-lasting genomic effects of European contact. This work was made possible through the collaborative efforts of Northwest Coast First Nations.

The first topic presented here focuses on the highly debated processes behind the initial peopling of the Americas and, more specifically, the Northwest Coast. Over the past decades, this region has received a great deal of a attention due to the Northwest Coast's proximity to Beringia and its prime location with respect to the coastal migration model, which holds that the first migration wave into the Americas occurred along the Pacific coast. Recent research, utilizing genome-wide data of both ancient and modern native individuals, has pointed to a primary migration that extends an ancestral lineage to most indigenous groups living today. Previous research, however, utilizing paternal and maternal genetic lineages, have found distributions of genetic markers that are difficult to reconcile into a single migration event.

To address this discrepancy, the genomes of three ancient individuals from the Northwest Coast, ranging in age from 10,000 to 1,500 years Before Present (BP), were sequenced and compared to previously described ancient individuals from the Americas. The results suggest that the Northwest Coast was populated by two sources, which may represent two separate migrations into the region. I present analyses suggesting that the first source was part of a primary migration that led to the peopling of both continents. The second source may have been a result of a migration that occurred during the Holocene, which further diversified the gene pool and directly links to the current indigenous populations of the Northwest Coast. This analysis reconciles

modern distributions of genetic markers, which are missing from North America but are prevalent on the southern continent. These markers, which are evident in the most ancient individuals of the region, are gradually replaced through time, which may be indicative of additional gene flow that entered the area after the primary migration wave made its way southward.

The next period of inquiry moves forward to the era of European colonization and addresses its consequences on the indigenous populations of the Americas. A common notion holds that Native Americans were especially susceptible to European borne pathogens at the time of first contact. This has generated a variety of theories from a variety of fields to explain this epidemiological pattern. The most fundamental of these is immunological naiveté, which posits that Native Americans were more susceptible to these diseases simply because they had never encountered them. Others have tried to find a genetic component by exploring the diversity of immune markers in extant indigenous populations. These hypotheses, however, either fail to articulate the exact genetic mechanism or fail to incorporate the genetic diversity of indigenous peoples before the massive population declines associated with European contact. In order to address these issues, I examined the genomes of a Northwest Coast First Nation from two different time frames: before and after European contact. This population, from present day British Columbia, suffered extensive population declines in the mid-Nineteenth Century, mainly due to smallpox epidemics. By comparing the DNA sequences (covering all coding regions) of the contemporary individuals to that of their ancestors, I was able to examine the pre-existing genetic diversity and scan the ancient and living individuals for signals of natural selection. My initial hypothesis held that the ancient individuals would exhibit signs of adaptations to the ancient environments of the Americas, which occurred over several thousand years before the era of European colonization. Remarkably, the gene that showed the strongest signal for selection was indeed an immune gene (*HLA-DQA1*), which is involved in the detection of pathogens and in launching targeted immune responses. Furthermore, components of the ancient version of the gene are now in very low frequency in the living members of this population. This pattern suggests the occurrence of two selection events: the first in ancient times, respective to the ancient environment, followed by the second event occurring after European contact, which favored the present version of the gene.

These findings are consistent with the hypothesis that the ancient population was adapted to a distinct and local epidemiological prehistory. European-borne pathogens like smallpox may have been able to take advantage of this ancient immune configuration, thus leading to increased virulence. My research suggests that the full explanation for Native American disease susceptibility appears to involve local adaptations to distinct and ancient immunological scenarios. These adaptations may have been advantageous in the ancient environment but proved deleterious once the environment was radically changed upon the onset of European contact.

The final period of indigenous evolutionary history explored here extends into modern times. From a general perspective, the genomic patterns of a population are molded by a variety of evolutionary factors, ranging from demographic to cultural. The indigenous peoples of the Americas have a particularly nuanced evolutionary history—involving founder effects, migrations, severe bottlenecks, and gene flow with non-native populations. While all of these events lend to the shaping of genomic patterns, theoretical expectations suggests that the severe population decline suffered after European contact will have left an equally severe deleterious mark on the indigenous populations of the Americas. This expectation is mainly based on the loss of genetic diversity, which would increase the appearance of genomic features that could prove harmful in terms of both disease and the ability to adapt. Upon comparing the genomic patterns of the ancient and contemporary individuals from a Northwest Coast First Nation, however, I found a resilience that belayed expectations. I show that in less than 7 generations since the bottleneck due to European-borne epidemics in the 19th Century, the contemporary population's genetic pool shows reduced levels of rare variants, lower homozygosity, and lower overall potentially damaging variants with respect to the ancient individuals. This can be explained by gene flow from non-native populations, which has increased genetic diversity and broken-up potentially damaging stretches of homozygous regions. This demographic process has mitigated, in a rapid fashion, the predicted deleterious effects of the population declines suffered after European contact. This feature not only demonstrates the resilience of Native Americans but also, in a broader sense, the beneficial effects of gene flow between populations as cultural and political boundaries begin to fade in the modern age of globalization.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Ripan Malhi, for all of his efforts with this project and Michael DeGiorgio for his tremendous expertise in computational biology. I would also like to thank my committee members Alejanadro Lugo for his helpful feedback and Charles Roseman for his guidance.

I owe the success of this project to the active support and collaboration of the Lax Kw'alaams and Metlakatla First Nations. This project was also made possible by the generous funding from the University of Illinois Graduate College and the National Science Foundation. I also would like to thank Morten Rasmussen and Eske Willerselv for hosting me at the Centre for GeoGenetics in Copenhagen.

Finally, I would also like to thank Robin Kar, who's valuable input contributed greatly to this project, and Jeremy Hobbs, for his long-lasting support throughout this process.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

The evolutionary history of the indigenous peoples of the Americas is both ancient and multi-dimensional. The demographic events and processes within this history have shaped the genomes of indigenous groups on both continents. Over a period of approximately 15,000 years, indigenous peoples established civilizations throughout the Americas and inhabited every type of ecological zone (Thornton, 1998; Ramenofsky, 2003; Mulligan et al., 2004). During this time, indegenious peoples adapted culturally and perhaps genetically to the diverse environments of the Americas. The onset of European colonization, however, radically altered these environments. After thousands of years of relative isolation from peoples of other continents, populations from two hemispheres rapidly merged in a matter of decades. The merging not only introduced different cultures and technologies, but also an overwhelming load of pathogens (Roberts, 1989; Dobyns, 1993; Mancall and Merrell, 2000). European colonization left indigenous civilizations in disarray, with many suffering death or assimilation by non-native settlers (Thornton, 1987; Thornton, 1997). The resilience of indigenous peoples, however, has allowed them to adapt and survive in radically different environments than those encountered by their ancestors.

This dissertation aims to examine the evolutionary history of indigenous populations of the Americas with a specific focus on a First Nation from the Pacific Northwest Coast. The genomic patterns of these individuals tell the story of their unique history, including demographic events characteristic of other indigenous populations. These events include the original peopling of the Americas, the ecological disruption after European contact, and the resilient adaptations into modern times (Thornton, 1987; Thornton, 1998; Ramenofsky, 2003). By examining these events in a systematic manner, with the aid of ancient DNA sequencing, I reveal how the genomic patterns of a human population change through space and time in response to both environmental and demographic changes. These changes are not only important to understand from an indigenous perspective, in terms of their impact, but also have global implications. The events that shaped the genomes of modern worldwide populations can best be understood by examining how these patterns have changed through time and the potential consequences of these changes. Understanding these global genomic patterns can also elucidate the future in an age of

globalization, where the boundaries between human populations are diminishing and their connectivity is ever increasing due to gene flow (Smith et al., 2007; McMichael, 2013).

In recent decades, a distinction has developed between scientific research conducted on Native Americans and mutually beneficial partnerships between scientists and indigenous groups. The latter represents not only the most appropriate method for conducting research, with any human group, but also the most effective because of its dynamic relationship between researchers and participants. In order to comprehend the evolutionary history of any population, the foundation must first come through the fundamental understanding of the community. Not only should collaborations involving indigenous groups benefit the participants but they should also shape it. Asking scientific questions of humans devoid of any ethical or community concerns will yield research that is incomplete and lacking in any basic understanding of the participants involved. Any question that is asked without this understanding, regardless of the objective and method, will yield an outcome overwhelmed by the scientist's own cultural perspectives (McInnes, 2011).

The history of research on Native Americans is not a pleasant one, hindered by ethical and moral misuse, with accompanying negative impacts for the communities involved (Schroeder et al., 2006). A disconcerting trend in previous years involved dissemination of data without the consent of the indigenous participants and without respect or understanding of local traditions (Macaulay, 1994; Hodge et al., 2000). These research outcomes are not only marred by culturally-centric interpretations but also parallel the government's paternalistic and pernicious treatment of indigenous communities in both Canada and the United States.

The historical government-based treatment of the indigenous peoples in North America is riddled with disturbing factors stemming from self-interested regulations and relocation programs. These legal actions have impacted communities in extremely negative ways, including tribal disbandment and the prohibition of ancient cultures and languages (Caldwell, 2005). Critical issues still face modern indigenous communities as an impact of previous and current political agendas that effects their culture, economy, and health. Despite these issues, however, many communities demonstrate resilience through both cultural and spiritual resurgence (Caldwell,

2005). This resurgence is reflected in the actions of many groups to assume active roles in scientific research agendas (Schroeder et al., 2006).

The work presented here is made possible through an ongoing partnership with Northwest Coast First Nations. This partnership was established by my thesis advisor, Ripan Malhi, in 2009, where annual visits to the communities and constant communication with tribal representatives allows for not only active participation but also guidance and shaping of the research itself. The project proposals that have led to this dissertation were presented to the community in 2010 and proceeded only with their expressed consent. All facets of the research described here were evaluated and approved by tribal representatives before publication. Through this active collaboration, I present three chapters reflecting three critical periods in the evolutionary history of these communities.

The first period is examined in Chapter 2 and concerns the initial peopling of the northwest coast of North America, on both the local and broader continental levels. Despite a plethora of research spanning several decades, the specifics of how and when the Americas were first populated remains a debated issue. In recent years, however, major advances in DNA sequencing, and its application to ancient DNA research, has allowed for an unprecedented amount of genome-wide data from both contemporary and ancient individuals from the Americas. This work has revealed many different pieces of demographic puzzles with greater resolution, some of which have unified previous genetic, linguistic, and archaeological hypotheses (Greenberg et al., 1986; Mulligan et al., 2004; Pitblado, 2011; Reich et al., 2012; Gravel et al., 2013; Rasmussen et al., 2014).

The Northwest Coast of North America is an intriguing region for ancient genomic studies because it lies in the path of a migration route from Beringia hypothesized to have occurred along the Pacific coast. This migration route may have been utilized by a settlement wave, which left its genetic signature on indigenous groups living on both continents today (Bodner et al., 2012; Reich et al., 2012).

While the South American continent shows evidence of a single and rapid migration wave, the North American continent exhibits possible gene flow from multiple population movements (Tamm et al., 2007; Reich et al., 2012; Raghavan et al., 2014). In fact, populations residing in the artic regions may be the descendants of a much younger wave from Asia, dating back to approximately 6,000 years, with a subsequent wave less than 1,000 years ago (Raghavan et al., 2014). Other demographic events that may have occurred further south into the interior, however, are still poorly understood from a genomic perspective.

I present the low coverage genome sequences of three ancient individuals from the Northwest Coast, spanning in age from approximately 10,000 to 1,800 calendar years before present (BP). By comparing these individuals to two previously described ancient genomes, the Anzick-1 (~12,000 calendar years BP, from Montana) (Rasmussen et al., 2014) and 939 (~6,500 calendar years BP, from coastal British Columbia (BC)) (Raghavan et al., 2015), I address the issue of the first peopling of the region by genetically assessing ancestral relationships through space and time. In doing so, I demonstrate a nuanced picture of how the living indigenous population relates to the first settlers of the area. The ~10,000 year old sample from Alaska shows a closer affinity to Anzick-1 than to the younger ancient individuals from BC, as well as to the current indigenous population of the region. This is despite the significant distance from Alaska to Montana and the close proximity from Southeast Alaska to the BC coast. This is also notwithstanding the fact that both the Alaska and the Anzick-1 samples show the same mitochondrial lineage as the ~6,500 year old sample from British Columbia. Since the current Northwest Coast population carry a different maternal marker, this suggests that the peoples of the Pacific Northwest coast share a lineage from two source populations. The first source may stem from a migration into the Americas that may have populated both continents and the second from a later expansion into the coastal region.

In Chapter 3, I examine the evolutionary effects of European contact on the indigenous peoples of the Americas. From the 15th to 19th Centuries, the influx of non-native populations into the Americas resulted in a major ecological shift. The indigenous inhabitants suffered dramatic population declines due to warfare, alterations in social structure, and the overwhelming introduction of European-borne pathogens (Cook, 1998; Fenn, 2001; Patterson and Runge, 2002).

The latter component is perhaps best categorized during the early times of colonization, with three major European-borne diseases exhibiting the greatest affect—smallpox, tuberculosis, and measles—and each exhibiting a differential effect on Native Americans when compared to Europeans (Merbs, 1992; Thornton, 1997; Fenn, 2001). These diseases are reported to have severely impacted native population mortality, with many tribes suffering a complete collapse (Thornton et al., 1991). The Northwest Coast First Nation, despite a relatively late contact period, suffered massive epidemics as well. The most severe of which came from smallpox, where epidemics in the 19th Century carried an estimated 30% to 70% mortality rate across various communities (Boyd, 1990).

If the historical accounts relating to the impact of European-borne disease on the indigenous peoples are accurate, mortality rates are significantly higher than those experienced by non-native populations, with additional factors such as nutrition and population density only partially explaining certain cases (Ramenofsky, 2003). The historical epidemiological discrepancies between the indigenous peoples of the Americas and European colonizers has prompted an array of theories from a variety of fields that propose a heightened vulnerability to European-borne pathogens (Bianchine and Russo, 1992; Dobyns, 1993; Patterson and Runge, 2002). Among them are biological hypotheses that employ the rudimentary concept that exposure to a pathogen will cause genetic resistance to it (Black, 1975; Motulsky, 1989). On these theories, the "naïve" status of a population with respect to a particular disease should be sufficient to explain these kinds of differences in susceptibility, although the precise genetic mechanisms for these differences are typically left unarticulated. Black (Black, 1992) took this line of reasoning one step further by suggesting that pre-Columbian Native Americans may have been particularly susceptible to disease due to immunological homogeneity, specifically with respect to the genes involved in the human leukocyte antigen (HLA) system. This assumption of homogeneity, however, was based on a survey of living Native Americans, which represent the surviving members of communities that suffered dramatic population declines, thus failing to assess the level of diversity that may have existed before European contact.

I test the hypothesis of a genetic component that may have contributed to this disease susceptibility with the collaborative efforts of Northwest Coast First Nations. The genetic

diversity of the population is explored in two different time frames, before and after European contact, via genome-wide sequencing of both ancient and living individuals. The resulting sequencing data allowed for non-candidate based scans for genes under the influence of positive natural selection. The scans revealed that the ancient individuals were under positive selection with respect to a variant of a HLA immune gene (*HLA-DQA1*), which drops to low frequency in the living population. This drop is consistent with pathogen-driven selection affecting the frequency of the gene variant between the two time frames. It is also consistent with a scenario for local adaption to ancient pathogens, or another component of the environment, which drove the variant to high frequency. This version of the gene may have proved incompatible with the environment, however, when the environment was altered with the influx of European-borne pathogens. Although HLA is also revealed in this study, it is revealed in an evolutionary framework and not as a simple case of low genetic diversity, as Black (1992) suggests. This work demonstrates the importance in understanding the evolutionary history of the indigenous peoples of the Americas with respect to their historical experiences with European-borne pathogens.

In Chapter 4, I explore the impact of European contact on the overall genomic patterns of the Northwest Coast First Nation and their ancestors. As previously mentioned above, the indigenous populations of the Americas suffered extensive population declines after European contact. This is the second significant demographic event resulting in a population contraction, however, that the indigenous peoples of the Americas may have experienced. The first stems from the initial peopling of the Americas, where the associated founder effect carried a subset of the genetic diversity that existed in the source populations (Mulligan et al., 2004; Schurr and Sherry, 2004; Tamm et al., 2007). These demographic events will have left their mark on the patterns of genetic variation in the living indigenous populations. These patterns are, however, further complicated by early population expansions and later gene flow from non-native groups, post European contact. Although previous studies have explored the genomic diversity of contemporary Native American populations (Hunley and Healy, 2011; Reich et al., 2012; Gravel et al., 2013), as well as the recent admixture events between native and non-native groups (Wall et al., 2011; Moreno-Estrada et al., 2013; Verdu et al., 2014), the effects of colonization have not

been examined with the aid of ancient DNA and the exploration of ancient indigenous genetic diversity.

The genomic patterns of the Northwest Coast First Nation population, once again examined from two different time frames, before and after European contact, reveal several interesting features that arose from the earliest and most recent chapters of Native American history. First, the genomes of the ancient individuals exhibit evidence of a founder effect followed by an expansion, which could be the signature of the initial peopling of the Americas. Second, when the effective population size is compared through time, the Northwest Coast populations seem to have entered a period of decreased diversity, or even a population decline, before European contact, while living individuals show great variation in this regard. Third, the prediction for "deleterious" alleles and the potential damaging factors associated with the presence of long stretches of homozygous genotypes is increased in the ancient individuals when compared to the living. Given the recent population decline due to European contact, this last feature is a surprising one since theoretical expectations would expect the living population to exhibit the potentially "damaging" excess of homozygosity (Nei et al., 1975; Chakraborty and Nei, 1977). The conflict between expectation and observation can be reconciled, however, when taking into account the known gene flow between non-native and indigenous groups in the area. This gene flow has increased the genetic diversity of the living individuals and thus mitigated, in a rapid fashion, the potentially "deleterious" effects predicted with such a dramatic population decline.

CHAPTER 2

A LOW COVERAGE GENOME OF AN EARLY HOLOCENE INDIVIDUAL FROM ALASKA REVEALS A GENETIC DIVERSIFICATION OF THE NORTHWEST COAST INDIGENOUS POPULATIONS

Abstract

Recent genome-wide studies of both ancient and contemporary indigenous peoples of the Americas have shed light on the demographic processes involved during the first peopling. The Northwest Coast is an especially intriguing focus for these studies due to its association with coastal migration models and nuanced genetic ancestral patterns that are difficult to reconcile with modern DNA alone. Here we report the low coverage genome of a male recovered from the On Your Knees Cave (OYKC) in southeastern Alaska. The human remains date to 9,200±50 ^{14}C years before present (BP), which is approximately 10,300 calendar years BP. We also present the genomes of two younger individuals from the coast of British Columbia, dating to 2440±75 ^{14}C and 1820±55 ^{14}C years BP, which is approximately 2,500 and 1,750 calendar years BP, respectively. With the aid of previously described ancient samples, we show that the Northwest Coast exhibits genetic continuity for the past ~6,500 years. The OYKC individual, however, shares a greater affinity with contemporary individuals from South America than to both ancient and contemporary individuals from the Northwest Coast, despite its proximate location. OYKC also shares a closer genetic relationship with the previously described Clovis associated individual, Anzick-1. This pattern suggests that the peopling of this region occurred in a minimum of two phases: first, by an expansion that ultimately led to the peopling of the southern continent, followed by a secondary source of gene flow that resulted in a genetic diversification of the region.

Introduction

The initial peopling of the Northwest Coast has received a great deal of attention due its close proximity to Beringia and the associated implications for migrations into the Americas. The Coastal Migration hypothesis proposes that the migration into the Americas first occurred via a coastal route stemming from Asia, which quickly reached both continents by means of watercraft

(Fladmark, 1979; Erlandson et al., 2007; Dixon et al., 2014). Both genetic and archaeological evidence supports this theory to a certain extent. Genetic evidence based on mitochondrial genomes shows the presence of founding haplogroups predominantly observed on the coast of both continents, in ancient and living individuals (Schurr and Sherry, 2004; Fix, 2005; Perego et al., 2009; Achilli et al., 2013). One such founding haplogroup is D4h3a, which is primarily observed in high frequency in contemporary indigenous populations close in proximity to the Pacific Coast (Kemp et al., 2007; Perego et al., 2009). The distribution of the haplogroup suggests that the lineage was part of a first migration, which rapidly reached South America. Archaeological evidence also supports this theory, as some of the earliest sites on the coasts of both continents exhibit material culture geared towards maritime adaptations (Erlandson et al., 2007; Dillehay et al., 2008; Dixon et al., 2014). If this model is correct, then the Northwest Coast was one of the first areas populated by the initial migrants into the Americas.

The peopling of the Northwest Coast may, however, be complicated by later demographic events in the region. Studies based on mitochondrial and Y-chromosomal markers, have hypothesized that the region likely experienced demic expansions after the initial migrations from Asia (Schurr and Sherry, 2004; Dulik et al., 2012). Reich et al. (2012), utilizing genome-wide data, inferred ancient gene flow into North America that likely stemmed from subsequent population movements long after a primary settlement wave. The study, however, did not consider indigenous populations from the Pacific Northwest, leaving the question open as to whether multiple source populations are reflected in the genomes of these peoples.

The distribution of the mitochondrial haplogroup D4h3a is also of particular significance for the Northwest Coast region. Previous work on ancient individuals from North America revealed an interesting spatial distribution of this lineage through space and time. The oldest sampled individual to carry the haplogroup, thus far, is the Anzick-1, dating back to ~12,600 calendar years BP (Rasmussen et al., 2014). Anzick-1 is associated with Clovis technology and was unearthed from present day Montana (Lahren and Bonnichsen, 1974). The Clovis tradition describes the cultural remnants of settlements throughout North America, dating back to 13,000 calendar years BP and geared towards terrestrial game-hunting (Haynes, 1964; Waters and Stafford, 2007). The age and distribution of these sites has led some to believe that the peopling

of the Americas was spear-headed by Clovis hunters, who traveled southward via the interior of the ice free corridor of North America (Fladmark, 1979; Goebel, 1999; Fiedel, 2000).

The maternal lineage is later found on the Prince of Whales Island, Alaska, with an individual known as OYKC—an acronym for the "On Your Knees Cave" in which he was found (Kemp et al., 2007). OYKC is not associated with Clovis culture but instead with a maritime tradition, consistent with a coastal migration model (Dixon et al., 2014). Approximately 300 km southeast of the OYKC site is Lucy Island, off the coast of British Columbia. This is the location of another individual bearing the lineage, known as 939, which dates back to ~6,500 calendar years BP (Cui et al., 2013). 939 is ancestral to the Coast Tsimshian, who currently live in the same region and are a traditionally maritime adapted peoples (Cybulski, 2001). This lineage, however, disappears from the region (or is found in extremely low frequency) in later ancient and living individuals (Cui et al., 2013). The change in mitogenome frequencies suggests gene flow from a distinct population into the Northwest Coast, occurring in the early Holocene[1], or, to the more extreme, a population replacement. On a broader continental context, the sharing of OYKC's founding mitochondrial genome with a Late Pleistocene individual (Bolnick, 2006; Rasmussen et al., 2014), found further inland of North America, may suggest a link between early coastal and interior populations.

In order to further explore these demographic scenarios, we present genome-wide data from OYKC, which dates to approximately 10,300 calendar years BP (Kemp et al., 2007) (Table 1; Fig. 9). Two additional ancient low coverage genomes from Prince Rupert Harbor (PRH), British Columbia (Table 1; Fig. 9), are also analyzed, dating to 1,750 and 2,500 calendar years BP. In comparing the genome-wide data of the three ancient individuals with previously described ancient genomes from the Americas, we aim to test two hypotheses with regard to the peopling of the Northwest Coast. We first test the hypothesis that the people of the region demonstrate genetic continuity dating back to at least 6,500 calendar years BP. Second, we test the hypothesis that the ancestors of the Northwest Coast peoples experienced additional gene flow in the early Holocene.

[1] The Holocene is a geological epoch commencing at 11,700 years BP and continuing into the present.

Sample	Age ^{14}C BP	Age Calendar Years BP	mtDNA Haplogroup	Sex[1]	Sequencing Strategy	Libraries Sequenced	Average Read Depth	Nucleotides Covered (hg19)
OYKC	9,200±50	10,344±83	D4h3a	XY	Shotgun (0.22% Endogenous DNA)	1	1.19	8.1 x 10^6
OYKC					Whole Genome Enrichment	8	2.85	1.7 x 10^8
PRH 302	2440±75	2498±142	A2d	XY	Whole Genome Enrichment	8	4.01	1.6 x 10^9
PRH 443	1820±55	1750±70	A2p	XX	Whole Genome Enrichment	8	4.74	2.0x10^9
939[2]	5710±40	6487±50	D4h3a7	XX	N/A	N/A	5.15	1.1x10^9

Table 1. Sequencing results. [1]Sex was determined from the sequence reads, using the method described in Skoglund et al. 2013). [2]Previously described in Raghavan et al. (2015).

Archeological Context

OYKC Individual

The skeletal remains of OYKC are dated to approximately 9,200±50 ^{14}C years BP (Dixon, 1999; Kemp et al., 2007). The spatial distribution of the remains within the cave suggests that the individual was not intentionally buried but instead was placed under accidental circumstances (Dixon et al., 2014). The fossil record of the southeast Alaskan coast suggests that the area was a coastal refugia during the last glacial maximum, with continual use starting at about 17,200 years BP (Heaton and Grady, 2003). Humans may have made use of the cave as early as 12,000 years BP.

Isotope analysis of the bone collagen revealed a long-term diet of marine foods with little sustenance derived from land sources (Dixon et al., 2014). The stone tools associated with the individual were manufactured with materials that stemmed from proximate islands and inland sources. This evidence suggest that the population associated with OYKC utilized coastal navigation to reach various deposits of raw materials. On a broader context, the combined evidence implies a seafaring population with established trade networks (Dixon et al., 2014).

PRH Ancient Individuals

The archaeological context of the ancient individuals recovered from the PRH region, ranging in age from approximately 6,500 to 1,750 calendar years BP, demonstrate cultural

continuity between the prehistory of the region and the Coast Tsimshian communities described in historical times. Two of the ancient individuals (302 and 443) derive from the PRH excavation sites, while individual 939 (previously described in Cui et al. (2013) and Raghavan et al. (2015)) stems from the Lucy Islands site, 19 km west of PRH. The region shows seasonal occupation starting as early as 5,000 years BP, with housing structures that show continuity between the sites associated with Lucy and the younger archaeological sites, covering a period from 2,500 to 1,500 years BP, as well as to the historical dwelling structures of the Coast Tsimshian (Fladmark et al., 1990; Archer, 2011).

The mitochondrial haplogroup of the 939 ancient individual is D4h3a, sharing the same mitochondrial haplogroup with the OYKC individual—who lived some 4,000 years earlier. Individuals 302 and 443 belong to the more common haplogroup A2, which is found in high frequency in indigenous populations on the Northwest Coast (Eshleman et al., 2004; Malhi et al., 2004; Achilli et al., 2013).

Results

To assess the shared ancestry between the ancient individuals and 156 worldwide populations, both ancient and contemporary, outgroup f_3-statistics were employed (Patterson et al., 2012; Raghavan, 2013). The data set included 52 Native American populations, which were masked for European ancestry, as well as additional masked indigenous groups from North America (Reich et al., 2012; Raghavan et al., 2015). All four ancient individuals (OYKC, 939, 443, and 302) show greater affinity with Native American groups than with other global populations (Figs. 1-3, Raghaven et al. 2015). We also ranked the individual outgroup f_3-statistics and found that PRH 443 shares a greater affinity with Northwest Coast groups (Fig. 4b). The OYKC individual ostensibly shows closer affinities to South American groups, such as the Surui and Karitiana, however, due to the low coverage of the sample, the relationship is not statistically significant (Fig. 4c).

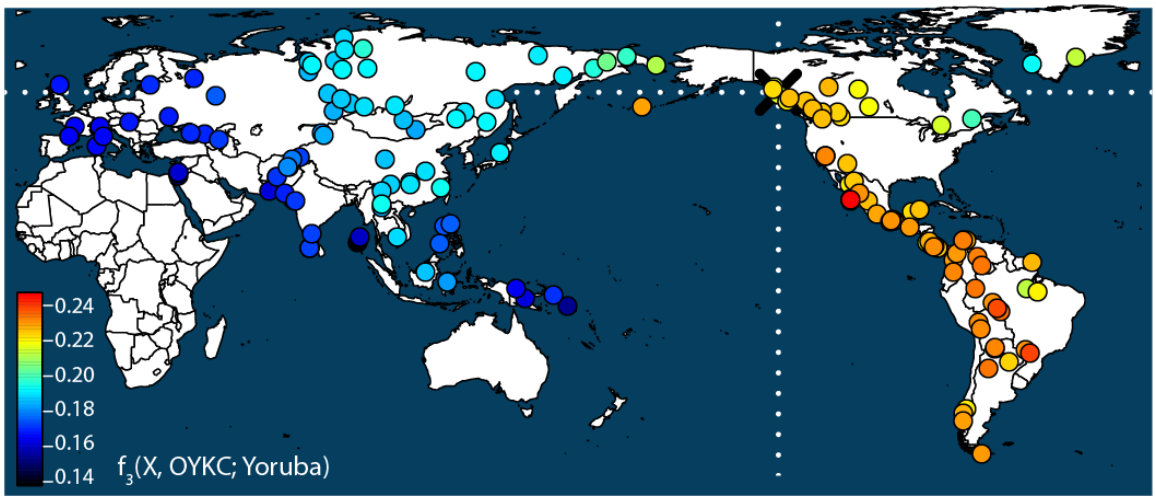

Figure 1. Genetic affinity of the OYKC individual to contemporary human populations. The OYKC individual shows greater genetic affinity with Native American groups than other global populations. Heat map represents the outgroup f_3-statistics estimating the amount of shared genetic drift between OYKC and each of the 156 contemporary populations since their divergence with the African Yoruban population.

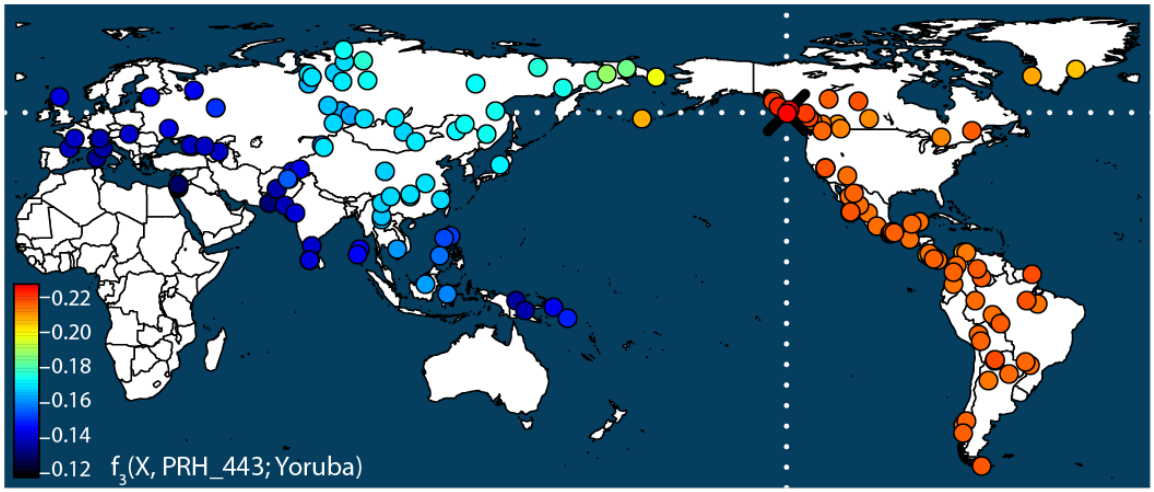

Figure 2. Genetic affinity of the 443 PRH individual to contemporary human populations. The 443 PRH individual shows greater genetic affinity with Native American groups than other global populations. Heat map represents the outgroup f_3-statistics estimating the amount of shared genetic drift between PRH 443 and each of the 156 contemporary populations since their divergence with the African Yoruban population.

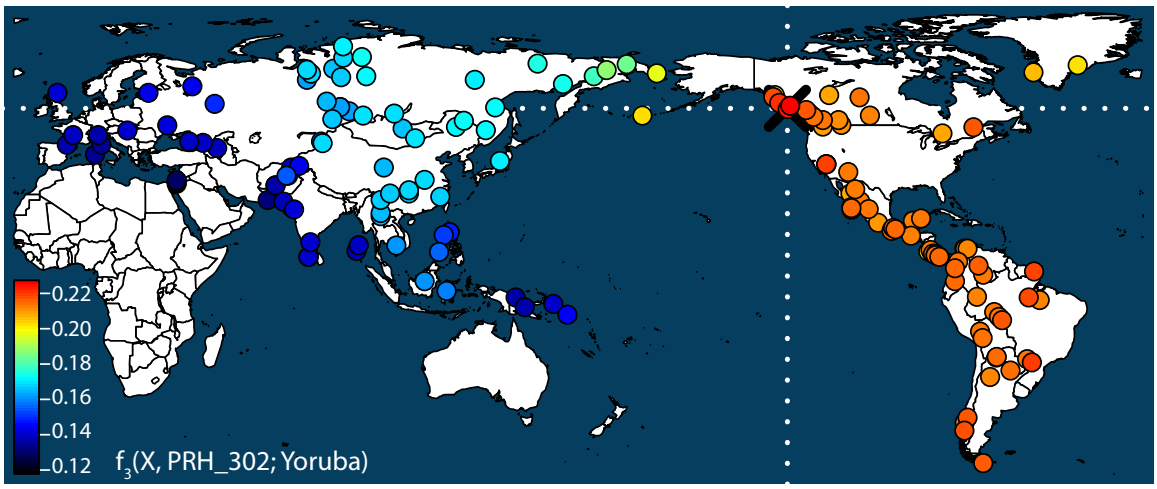

Figure 3. Genetic affinity of the 302 PRH individual to contemporary human populations. The 302 PRH individual shows greater genetic affinity with Native American groups than other global populations. Heat map represents the outgroup f_3-statistics estimating the amount of shared genetic drift between PRH 302 and each of the 156 contemporary populations since their divergence with the African Yoruban population.

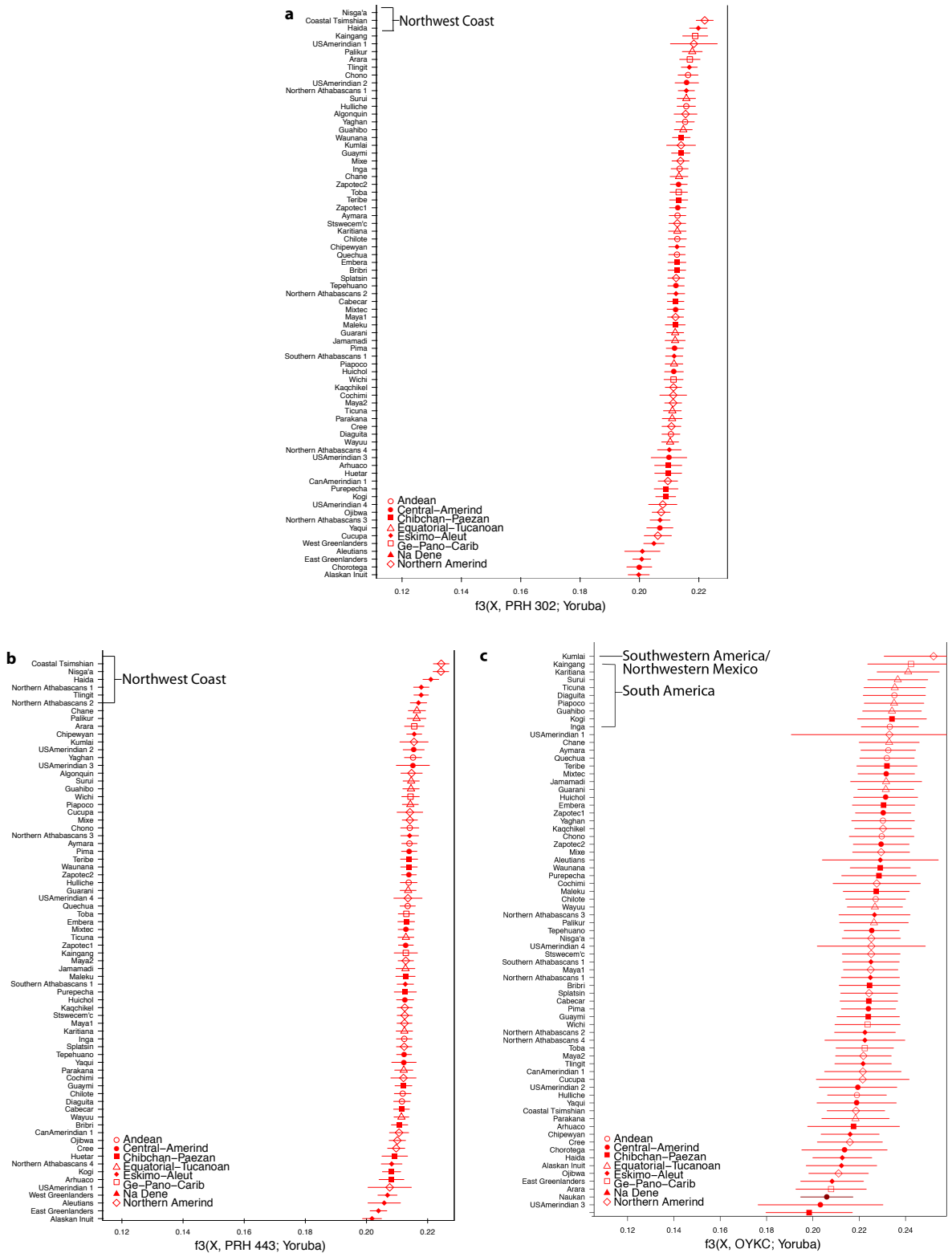

Figure 4: Ranked outgroup f_3-statistics examining shared genetic history. (**a)** PRH 302 and **(b)** PRH 443 both show a close relationship to indigenous populations from the Northwest Coast. **(c)** OYKC shows a closer affinity to Southwestern and South American groups.

To further elucidate the relationship among the ancient individuals, as well as their relationship to contemporary populations, we examined maximum likelihood trees created with *TreeMix* (Pickrell and Pritchard, 2012). Nucleotide transitions were removed from the ancient samples to guard against DNA damage. We see that PRH 443 and 302 form a sister clade to the contemporary Tsimshian (masked for European ancestry) (Fig. 6a, b). 939 is basal to both the Tsimshian and Athabascan (Raghaven et al. 2015). OYKC is placed at the root of all native groups, except for the ancient Saqqaq individual from Greenland (Rasmussen et al., 2010), but appears closer to South American populations (Fig. 6c).

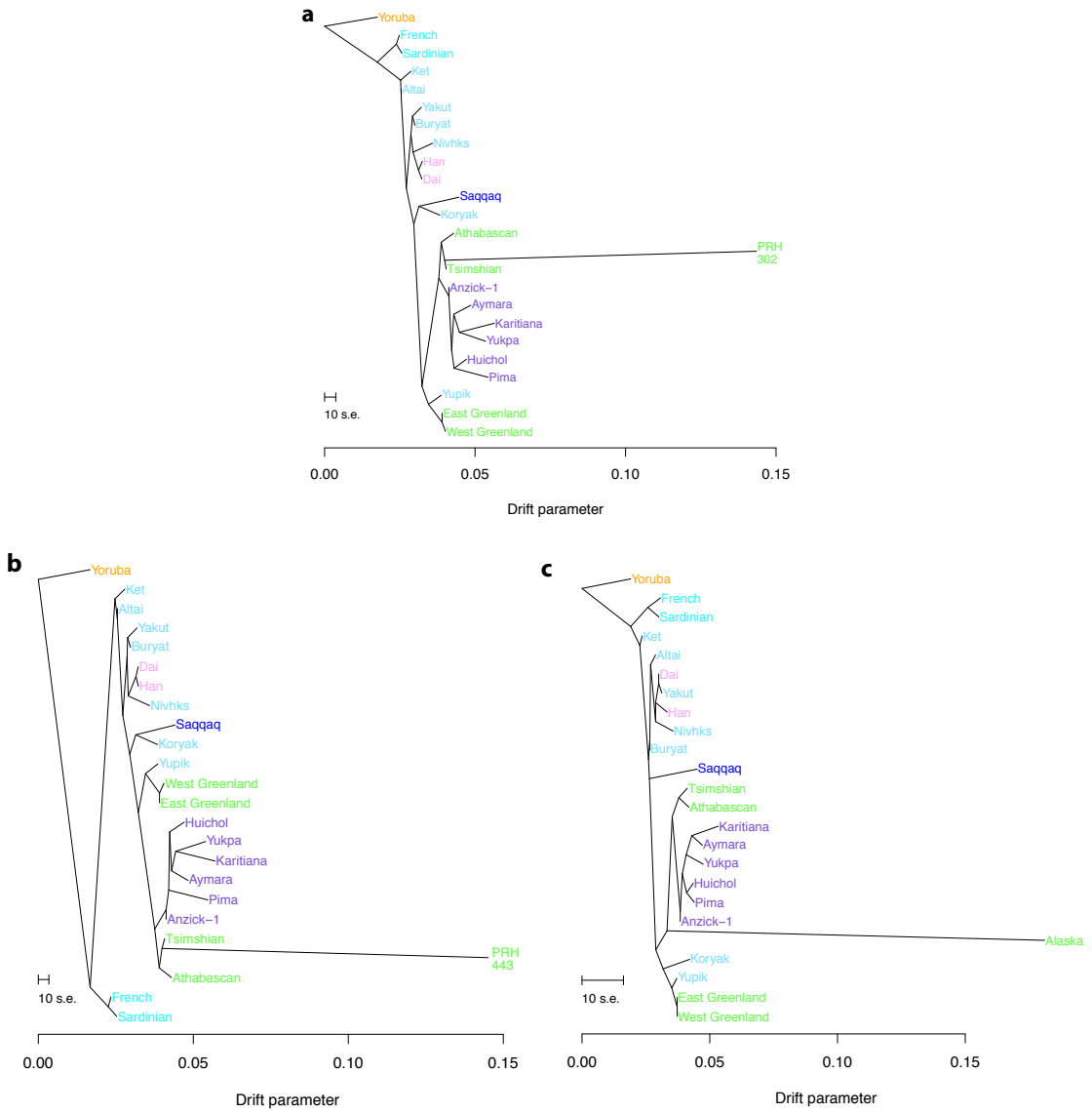

Figure 5: Maximum liklihood trees. Maximum likelihood tree generated by *TreeMix* using whole-genome sequencing data, with the Tsimshian genome masked for European ancestry. HGDP, Human Genome Diversity Project. Trees shown have 0 migrations for **(a)** PRH 302, **(b)** PRH 443, and **(c)** OYKC.

In order to further test the hypothesis that the Tsimshian exhibit a close genetic relationship with ancient individuals from the PRH region, we employed the *D*-statistic (Patterson et al., 2012). This statistical test operates on a four taxon tree, where the null hypothesis, *D*=0, implies that the tree is correct. If *D* significantly deviates from zero, then gene flow is implied between the branches of the tree. We once again guarded against possible biases from DNA damage by removing transitions from the ancient samples and tested different models of genetic relationships among ancient and contemporary individuals. The first model (Fig. 7a) supports the hypothesis where all three PRH ancient individuals have a shared ancestry. The second model (Fig. 7b) extends the shared ancestry to the local Tsimshian population (masked for European ancestry) and is supported by our data. The model that directly links the ancestry of the PRH individuals to OYKC is, however, rejected by our data (Fig. 7d). The final model directly linking the ancient PRH individuals to a South American group (Karitiana) is also not supported (Fig. 7c).

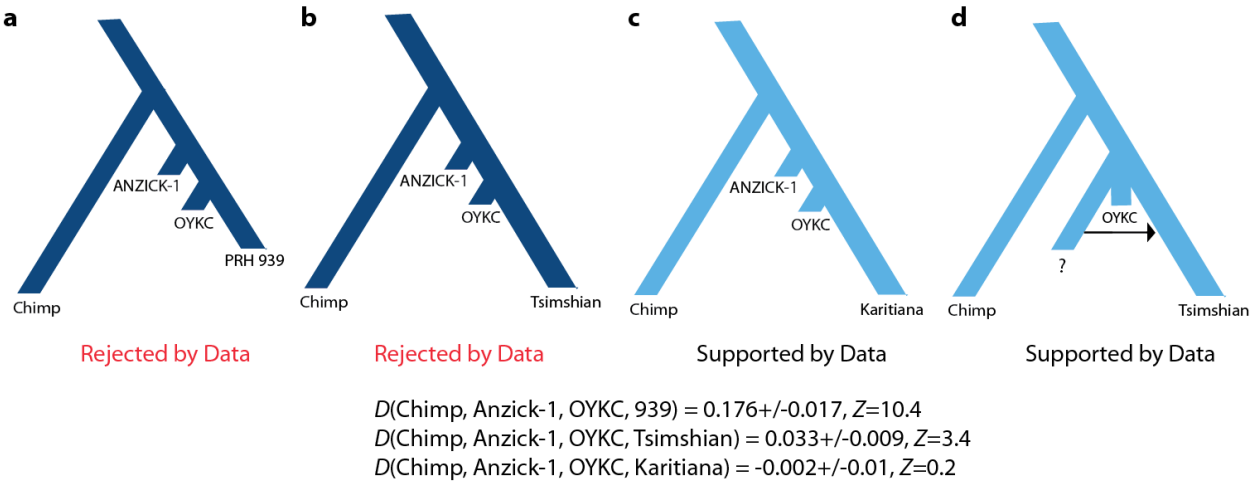

Figure 6. Genetic models of the regional peopling of the Pacific Northwest Coast. We tested alternative models of shared ancestry of PRH individuals with other contemporary and ancient populations. **(a)** The data supports a simple lineage, where PRH 939 and 302 are both basal to 443. **(b)** The data also supports a simple lineage where PRH 302 and 443 are basal to the contemporary Tsimshian. **(c)** The data does not support a lineage where the PRH 302 and 443 are basal to the South American Karitiana. **(d)** The data also did not support a lineage where OYKC is basal to PRH 939 and 443.

The next models examine a less direct relationship between the OYKC and PRH individuals, in addition to the spread of the D4h3a mitochondrial haplogroup across the Americas. Since the Anzick-1 individual from Montana predates the OYKC individual (~12,600 vs. ~10,300 calendar years BP) and shares the same mitochondrial genome, we incorporated the sample into the

analysis to assess their relationship. The first two models (Fig. 7a, b) place the Anzick-1 individual basal to both the OYKC and PRH individuals, both of which are rejected by our data. This result is intriguing, since the test for treeness also fails when OYKC is made basal to the PRH individuals (Fig. 5c). The third (Fig. 7c) involves Anzick-1 and OYKC being basal to a South American group and is supported by our data. The final model (Fig. 7d), especially supported in conjunction with the *TreeMix* analysis, involves a basal diversification of the OYKC lineage, which predates the PRH individuals but contributes to their genetic makeup.

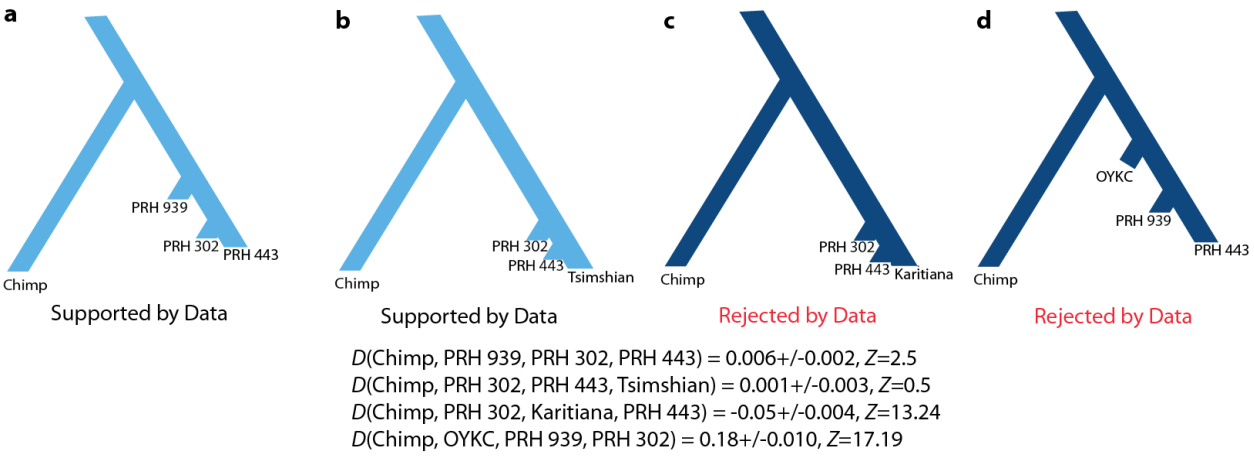

Figure 7. Genetic models relating Anzick-1 to OYKC. We tested alternative models of shared ancestry of the Anazick-1 and OYKC individuals with other contemporary and ancient populations. (**a)** The data did not support a simple lineage, where Anzick-1 and OYKC are both basal to PRH 939. (**b)** The data also failed to support a simple lineage where Anzick-1 and OYKC are basal to the contemporary Tsimshian. (**c**) We tested the hypothesis that both the Anzick-1 and OYKC were basal to the South American Karitiana, which is supported by the data. (**d)** An additional hypothesis is also supported where the branch deriving from the the OYKC meta population contributes to the genetic lineage of the contemporary Tsimshian.

Discussion

Our data supports a shared ancestry for the indigenous peoples of the Northwest Coast, dating back to at least ~6,500 calendar years BP. The individual that establishes this is PRH 939, which carries the D4h3a mitochondrial haplogroup. The haplogroup, however, is not shared with later ancient individuals or the contemporary population of the same region. Furthermore, although PRH 939 shares the mitochondrial haplogroup with both the Anzick-1 and OYKC, their genetic relationship is not closely shared. Instead, PRH 939 exhibits a greater infinity with PRH 302 (~2,400 calendar years BP) and 443 (~1750 calendar years BP), both of which share the A2 mitochondrial haplogroup and reflect the maternal lineage of the contemporary population. The *TreeMix* analysis, however, reveals a more complicated ancestral picture for PRH 939, which

places it basal to both the Tsimshian and the Athabascan—a distinct indigenous population from the Northwest (Raghaven et al. 2015).

The pattern could be explained by the genetic relationship of OYKC to other populations. Although OYKC shows greater genetic affinity to indigenous groups of the Americas than to other worldwide populations, OYKC exhibits a closer relationship to South American groups. This is the same result found with Anzick-1 (Rasmussen et al., 2014). OYKC also exhibits a closer relationship to Anzick-1 than to the PRH individuals and shares the D4h3a haplogroup. This suggests that OYKC was part of a population closely related to the ancestral population. PRH 939, however, seems to exhibit dual genetic affinities. The individual also shares the D4h3a maternal lineage, but its genomic ancestral signatures begin to align towards the Northwest Coast populations, as implied by the individual's basal relationship to two modern indigenous groups from the area. The genetic diversification appears complete by the time of 302 and 443, which form a sister clade with the Tsimshian and reflect a distinct maternal lineage (Fig. 8). The hypothesis is also compatible with the disappearance of D4h3a from the Northwest Coast. The first inhabitants of the region may have been part of the wave that led to South American groups, with gene flow from later expansions contributing to the later genetic diversification. A genetic diversification without external gene flow is also a possibility, but this would not explain the disappearance of the mitochondrial haplogroup.

Figure 8. Site location of the ancient individuals and their associated mitochondrial haplogroups. The Anzick-1, OYKC, and 939 all share the D4h3a mitochondrial haplogroup. Later ancient individuals from the Prince Rupert Harbour region show a shift in mitochondrial linages to A2. Glaciation patterns are as described in Dyke *et al.* (Dyke et al., 2002).

The archaeological record also shows a shift in the region in terms of technology, which coincides with the 4,000 year time span between OYKC and PRH 939. Many of the sites along the coast of Alaska and British Columbia, predating 9,000 calendar years BP, exhibit a bifacial maritime stone industry (Fedje and Christensen, 1999). This technology, however, begins to shift around 8,900 calendar years BP with the addition of microblades, which completely replaces the bifacial tools by 8,000 calendar years B.P. (Fedje and Christensen, 1999). Interestingly, the new technology is thought to derive from a Berigian maritime paleoarctic adaptation, which endures until the mid-Holocene and reaches into the time of PRH 939 (Dixon, 1993). While this technology may have been diffused along trading routes, it is also possible that the tradition may have been accompanied by new peoples, which may have diversified the gene pool of the region.

Lastly, we broach the topic of migration routes taken during the first peopling of the Americas. Previous studies have examined the distribution of the D4h3a mitochondrial haplogroup in order to support competing migration models. Perego et al. (2009) hypothesized that the haplotype distribution exhibited by living Native American populations is a result of a coastal migration that was independent of a movement along the interior of North America. However, the appearance of D4h3a in ancient individuals found away from the coast, such as Anzick-1, belays this theory (Bolnick, 2006; Rasmussen et al., 2011). While we cannot speak of evidence for either model, we do show that the Anzick-1 and OYKC might be from the same ancestral population, which had an early presence in North America. These individuals, on a regional level, however, demonstrate major differences in adaptive technologies: marine vs. terrestrial. If they were indeed part of the same ancestral population, they were perhaps not limited to one movement or survival stratagem.

We conclude that the Northwest Coast was populated as a result of two demographic expansions into the region. The OYKC individual, who lived some 10,300 years ago, was part of an ancestral population that may have first populated the region but ultimately led to the peopling of the southern continent. Although the ancestral population left its genetic mark on the later occupants of the Northwest Coast, it appears that a separate demographic event, occurring during the later Holocene, further shaped the genetic architecture of the region. This is supported by the dual genetic nature of PRH 939, which shares attributes akin to both the OYKC and later ancient individuals of Prince Rupert Habour. The results presented here begin to elucidate the processes involved behind the peopling of the Northwest Coast.

Methods

DNA Extraction

DNA extractions were completed in an ancient DNA laboratory facility at the University of Illinois Urbana-Champaign. Ancient DNA extractions and PCR amplification setups were completed in the ancient DNA laboratory facility at the University of Illinois. The ancient DNA lab is a positively pressured clean room with hepa-filtered air. The clean room contains an anteroom and air flows from the ancient DNA lab to the anteroom to the hallway. Personnel working in the ancient DNA lab wear disposable hairnets, facemasks, laboratory coveralls and

booties. All equipment, reagents and consumables are dedicated for use in the ancient DNA laboratory. The ancient DNA lab is routinely cleaned with bleach and all containers are wiped with DNA Away before placed in the ancient DNA lab. Personnel are restricted from entering the ancient DNA after being in a contemporary DNA laboratory. A database containing mitochondrial control region sequence is maintained of all personnel working in the ancient laboratory and of any personnel who may have come into contact with the human remains prior to DNA analysis. Contamination controls were used with every DNA extraction and PCR setup in order to detect any contamination from reagents. Also, series of negative controls are routinely performed in the ancient DNA lab.

Teeth were utilized from each ancient individual for the extraction. Each tooth was soaked in 6% sodium hypochlorite for 3 minutes, rinsed three times with UV-irradiated molecular grade water, and dried in a UV Crosslinker for 10 minutes, so as to remove surface contamination. Approximately 0.20 grams of tooth powder was incubated in 4 ml of demineralization/lysis buffer (0.5 M EDTA, 33.3 mg/ml Proteinase K, 10% N-lauryl sarcosine) for 24 hours at 37ºC. The digested sample was then concentrated to approximately 250 µl using Amicon centrifugal filter units. Following concentration, the digest was run through silica columns using the MinElute Qiagen PCR Purification Kit (Qiagen, Hilden, Germany) and eluted in 60 µl of DNA extract.

DNA Screening for mtDNA

In order to test for viable DNA before proceeding with library building and sequencing, each ancient individual was amplified for the hypervariable region I of mitochondrial DNA from 2 µl of extract, utilizing the reagents and conditions described in Malhi *et al.* (2003). All 3 ancient samples exhibited Native American mitochondrial haplogroups.

Library Build

Libraries were created using the New England Biolabs Ultra Kit (E7370S, Ipswich, MA) following the manufacturer's protocol with the following modifications: DNA fragmentation was not performed. DNA purifications were done using the MinElute Reaction Cleanup Kit (Qiagen, Valencia, CA). Library amplification was done in two steps. The first round of amplification

utilized the kit's reagents and protocol with 12 cycles of (10s at 98°C, 30s at 65°C and 30s at 72°C). For the second round, we achieved a sufficient DNA concentration for the whole-genome enrichment (~500ng), without excessive amplification, by creating 4 PCR reactions from the initial amplified product and then pooling them before applying the Qiagen MinElute PCR Clean-up kit. For the 2nd PCR, we created a 50 µl reaction, utilizing 0.2µM of primers P5 (5'-AATGATACGGCGACCACCGA) and P7 (5' CAAGCAGAAGACGGCATACGA) (Meyer and Kircher, 2010), 5µl from the initial PCR, 25µl of Phusion® High-Fidelity PCR Master Mix with HF Buffer (New England Biolabs, Ipswich, MA), 3% DMSO (New England Biolabs, Ipswich, MA), 0.2mg/ml BSA (New England Biolabs, Ipswich, MA). PCR conditions were as follows: 4m at 98°C, 10 cycles of (10s at 98°C, 30s at 62°C and 30s at 72°C), with a final extension at 72°C for 10m. Library fragment sizes were confirmed via a BioAnalyzer High Sensitivity assay to be above 130bp.

Genome Enrichment and Illumina Sequencing

For each of the PRH ancient samples (302 and 443) and OYKC, 8 libraries were genome enriched using the MyBaits Whole Genome Capture kit (Mycroarray, Ann Arbor, MI). Each of the 3 samples were sequenced on 4 lanes, pooling 2 libraries in each lane, on the Illumina HiSeq 2000 at the High-Throughput Sequencing Division of the W.M. Keck Biotechnology Center at the University of Illinois Urbana-Champaign. For the capture, the manufacture's protocol was used with the following modifications: the hybridization temperature was decreased to 50°C for 24 hours, Qiagen MinElute PCR Clean-up kit was used instead of beads, a final heat elution from the Streptavidin beads was performed with molecular biology grade water at 90°C for 2 minutes, and the post-capture amplification involved 12 cycles.

Mapping and Genotyping

Raw data from the Illumina HiSeq 2000 platform was base called with CASAVA 1.8.2. Sequences were de-multiplexed with a requirement for a full match of the 6 nucleotide indexes that was used for library preparation. Illumina adapter sequences were trimmed using Trimmomatic-0.32 (Bolger et al., 2014) with a minmum length of 25 and removing leading and trailing quality or N bases below a quality score of 3. Five bases were additionally trimmed at each end of the reads to minimize the effects of DNA damage. Trimmed reads were aligned to

the human reference genome hg19 using Bowtie2-2.2.5 (Ben Langmead et al., 2009) with a local realignment option and a seed length set to 1000. SAMtools-1.2 (Li et al., 2009) was used to sort and remove duplicate reads based on mapping positions. Due to the low average read depth of the PRH and OYKC individuals, genotypes were not called. The resulting alignment files were further filtered for downstream statistical analysis using a minimum mapping quality of 30, a minimum base quality of 30.

DNA damage patterns

DNA damage (type I and type II) was assessed by comparing T–C/G–A and C–T/A –G transitions, respectively using MapDamage 2.0 (Jonsson et al., 2013). A specific pattern of DNA damage has been identified in other ancient DNA studies (Rasmussen et al., 2010; 2011). These studies show a pattern of increased type II DNA damage at the beginning and end of degraded DNA fragments. The MapDamage program was run on the shotgun sequencing runs, without trimming. The results show signatures of DNA, which suggests that the PRH ancient sequences originate from ancient DNA templates and not modern contaminants (Fig. 9).

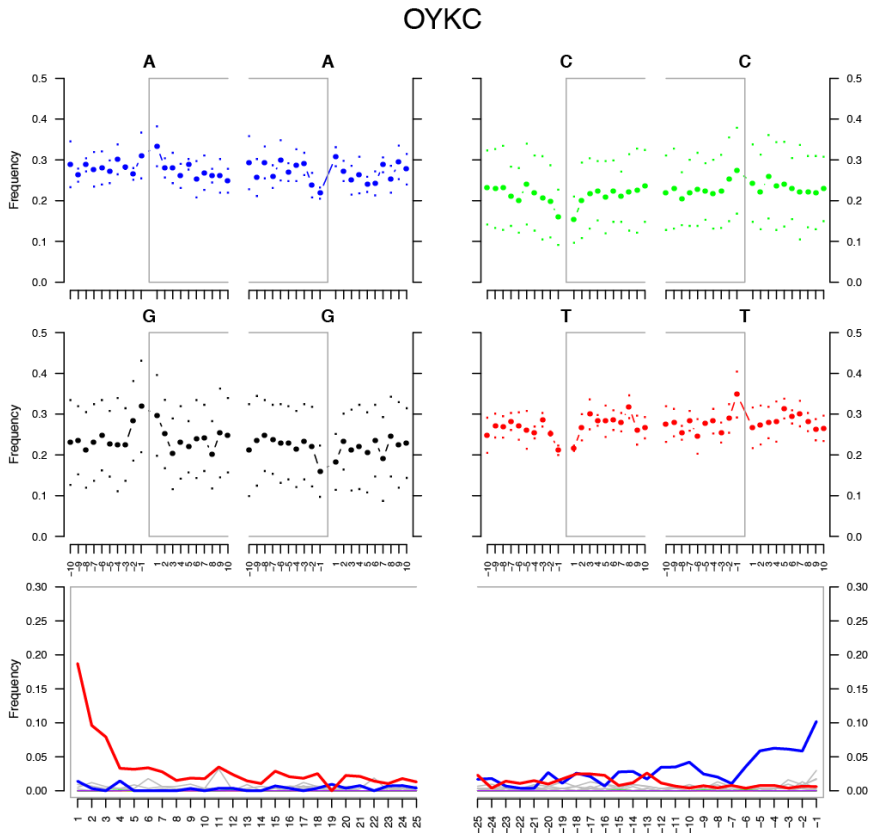

Figure 9 (cont.)

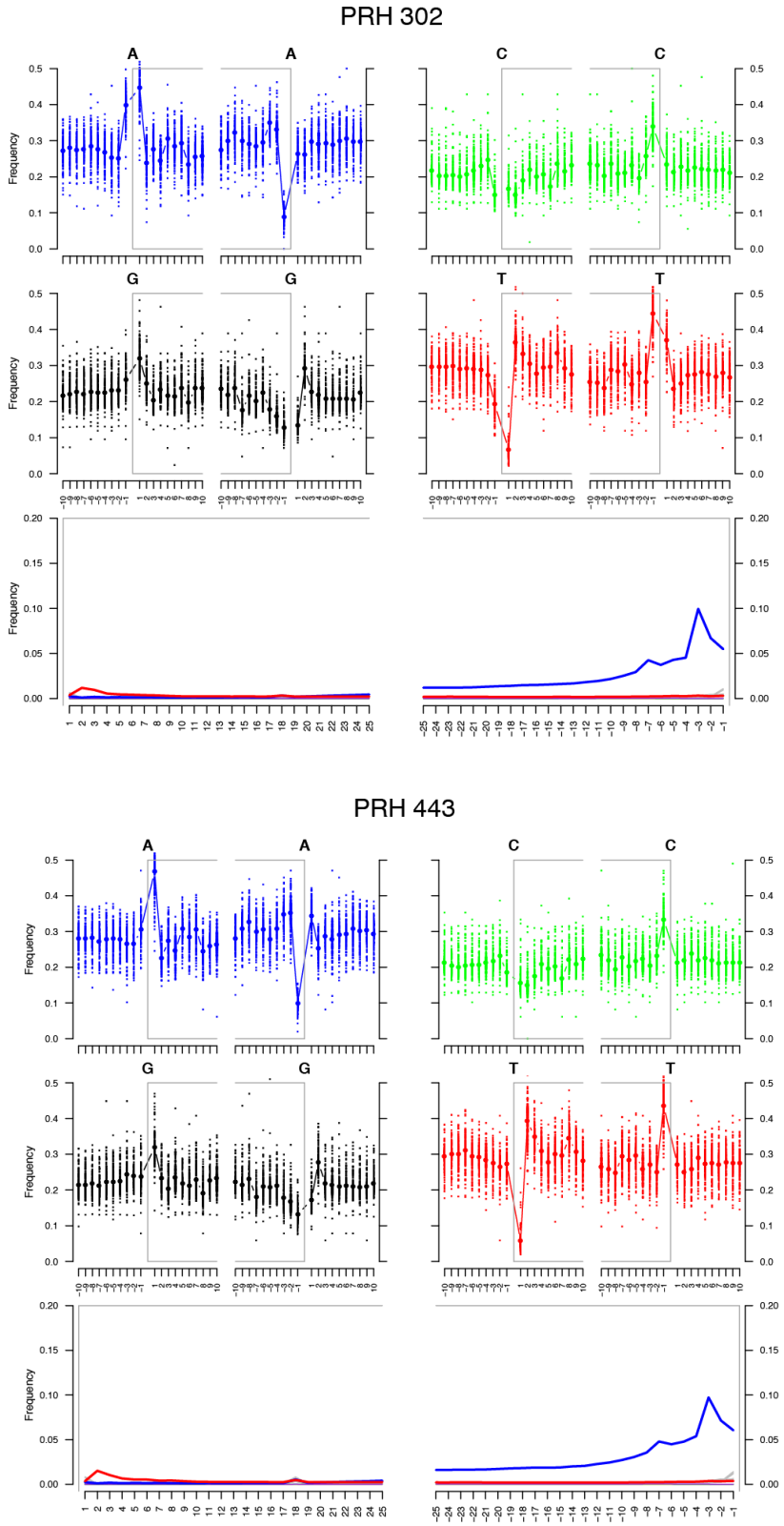

Figure 9. DNA damage patterns for the ancient individuals. Random subset of all mapped reads for each PRH Ancient individual. The mismatch frequency is relative to the reference as function of read position, C to T in red and G to A in blue.

Contamination Estimates

Contamination estimates utilizing the mitochondrial genome where run on all three samples, using an Approximate Bayesian computation method described in Korneliussen *et al.* (2014) (Korneliussen et al., 2014) through the ANGSD software suite (http://popgen.dk/angsd). Since OYKC and 443 were typed as male, using the method described in Skoglund *et al.* (Skoglund et al., 2013), contamination based on the X chromosomes was also performed for these samples. The OYKC sample did not have sufficient coverage along the X chromosome to perform the estimate.

Sample	Method	Contamination Estimate	*P* value
OYKC	Mitochondria	2.16%	0.005
OYKC	X Chromosome	Insufficient Coverage	-
443	X Chromosome	0.80%	1.23×10^{-9}
443	Mitochondria	0.84%	0.005
302	Mitochondria	0.77%	0.005

Table 2. Contamination estimates.

f_3 and D statistics

To test the genetic affinity of the ancient individuals with global populations, we preformed outgroup f_3 out-group statistics, using the method outlined by Patterson et al. (2012). Mapping and nucleotide qualities below 30 were not considered. The genetic affinities of the 4 ancient individuals with different combinations of global populations are shown on Figures 1-4 . We also examined the genetic affinities of each individual and various populations using ranked f_3 statistics, shown in Figure 5.

The data set used for global comparisons included 2,081 individuals derived from the published study by Raghavan et al. (2015). This data set, comprised of Native American and Siberian genome-wide SNPs, was masked for European or African descent. For the genetic affinity analysis to global populations, the reads were aligned to hg19.

To examine the relationship between the ancient individuals (OYKC, PRH 939, PRH 443, PRH 302, and Anzick-1), we performed an ABBA-BABA test or *D*-Statistic (Green et al., 2010) using the definition employed by ANGSD (Korneliussen et al., 2014). The statistic tests the tree ((P1,

P2), P3, O), where O is an outgroup to populations P1, P2, and P3. If the null hypothesis, where D=0, cannot be rejected, then there is no evidence for gene flow between either P1 and P3 or between P2 and P3. If there is a significant deviation from 0, then we can reject the hypothesis of the tree ((P1, P2), P3, O), and the data are consistent with gene flow between P1 and P3 or between P2 and P3. A Z score was employed to determine the significance of the test, whereby an absolute value greater than 3 ($|Z|>3$) was used as a critical value. The chimpanzee genome was used for the outgroup. The tests also included the masked whole genome of a contemporary Tsimshian (T60) ((Raghavan et al., 2015), which was masked for European ancestry, and the whole genome of a Karitiana individual (HGDP00998) (Rasmussen, 2014). In order to guard against DNA damage from the ancient individuals, transitions were not considered during the tests.

CHAPTER 3

ANCIENT EXOMES FROM INDIGENOUS PEOPLES OF THE NORTHWEST COAST REVEAL IMMUNE-BASED ADAPTATION TO THE AMERICAS

Abstract

The susceptibility of Native Americans to specific infectious disease has been postulated as a major factor for the population decline after European colonization. In order to investigate if a genetic component contributed to this phenomenon, the genetic diversity of an indigenous population before and after European contact was captured by sequencing 50 exomes of both living and ancient individuals from the Pacific Northwest Coast. Genes showing allele frequency changes between the two sampling time frames were identified, which represent strong candidates for immune-based adaptation. The strongest immune-based selection signal in the ancient individuals came from the human leukocyte antigen *(HLA)-DQA1,* which is involved in antigen presentation and the activation of the adaptive immune response. The living individuals show a strong frequency decrease of these ancient alleles (the most pronounced variant showing a 64% difference), which may correlate to the smallpox epidemics of the 19th century. The frequency changes suggest ancient local adaptation to the environment of the Americas and pathogen-driven selection after European colonization.

Introduction

The decline of native populations of the Americas after European contact has been linked to several factors including warfare, alterations in social structure, and the overwhelming introduction of European-borne pathogens (Cook, 1998; Fenn, 2001; Patterson and Runge, 2002). Although, the extent of the population decline is still contentious, the decline due to epidemics may have disproportionately contributed to this phenomenon (Dobyns, 1993; Boyd, 1999). This has prompted biologists to explore the possibility of disease susceptibility, where low genetic variation and immunological naïve populations are held responsible for exacerbating pathogen associated mortality rates (Black, 1975; Motulsky, 1989; Black, 1992). Assumptions of low diversity amongst immunological genes, such as HLA, however, are based on surveys of living Native Americans, which represent the surviving members of communities affected by

colonization, thus failing to assess the level of diversity that may have existed before European contact and leaving the possibility of a preexisting genetic component unarticulated.

Given the demographic history of the indigenous peoples of the Americas, the immunological story is likely much more complex. As these populations expanded into the Americas, approximately 15,000 to 20,000 years BP (Tamm et al., 2007; Gravel et al., 2013; Raghavan et al., 2015), they settled into environments with distinct ecologies and with presumably little gene flow from other continental populations until the era of colonization. As a consequence, these populations may have adapted locally to local pathogens, resulting in long-lasting changes on immune related loci. These type of ancient immune adaptations are suspected to have occurred throughout human prehistory as populations spread into varying environments across the globe (Fumagalli, 2011). For the indigenous peoples of the Americas, these adaptations may have proven robust in the ancient ecologies but deleterious after European contact altered the environment through the large influx of pathogens, some of which may have been novel. Therefore, pre-existing genetic adaptations may have contributed to the disease related population decline of the indigenous peoples of the Americas, after European contact.

To investigate possible immune-related genes under selection before and after European contact, I sequenced the exomes (containing all of the coding regions of the genome) of ancient and living individuals of a Northwest Coast First Nation from Prince Rupert Harbor (PRH), British Columbia (Fig. 10). The ancient individuals sampled for this study are the presumed ancestors of the living members of the First Nation, as evidenced by their oral histories, archaeology, and genetics (Cybulski, 2001; Martindale and Marsden, 2003; Cui et al., 2013). The archaeological context of the ancient individuals, unearthed from PRH and Lucy Island, ranging in age from approximately 6,500 to 500 calendar years before present (BP) (Fig. 10, Table 3), demonstrate cultural continuity between the ancient history of the region and the First Nation communities described in historical times[2] (Fladmark et al., 1990).

[2] For a more detailed archeological context of the ancient samples *see* Chapter 4.

This indigenous population did not begin regular contact with Europeans until approximately the 1780s, with the development of the maritime fur trade. In 1834, the Hudson's Bay Company was established at Fort Simpson and was later accompanied by an influx of European settlers (Sequin and Halpin, 1990). Shortly thereafter, the indigenous peoples suffered two major smallpox epidemics in 1836 and 1862, resulting in an estimated 70% population decline (Boyd, 1999). Today, the decadents of these peoples have recovered to near pre-contact levels and still reside in the same region as their ancestors(Sequin and Halpin, 1990; Census of Canada, 2006).

In this chapter, I test the hypothesis that the indigenous peoples of the Americas were adapted to the ancient environments, which may have effected their interaction with European borne pathogens. The test entails genomic scans for natural selection, on both the ancient and living individuals, which represent two time frames: before and after European contact. In doing so, I hope to reveal evidence for ancient adaptations and the molecular impact of European contact.

Figure 10. **Archaeological sites in the Prince Rupert Harbor region of British Columbia, Canada.**

Catalog ID	Archaeological Site	Osteology-Based Sex Estimation	Sequence-Based Sex Estimation	Specimen	Conv ^{14}C Years BP	Cal Years BP	Dated via Associated Cultural Remains	mtDNA Haplo-group
XVII-B-125	GBTo-23	M	Not Assigned	mand right 3rd molar	2260± 40	2288± 58		A*
XVII-B-158	GBTo-23	F	XX	mand left 3rd molar	2290± 50	3179± 83		A2
XVII-B-163	zGBTo-18	M	XY	mand right 2nd molar	n/a	n/a	"Pre-historic"	A2
XVII-B-167	GBTo-23	M	Not Assigned	mand right 2nd molar	n/a	n/a	cal 800 to 350 BC	A*
XVII-B-168	zGBTo-18	F	XX	molar	2650± 75	2760± 60		D4
XVII-B-181	GBTo-23	F	XX	mand left 3rd molar	2620± 40	2743± 20		A2
XVII-B-300	GBTo-30	F	XX	mand left 3rd molar	1650± 75	1544± 88		D
XVII-B-302	GBTo-30	F	XX	mand left 3rd molar	2440± 75	2498± 142		A2
XVII-B-311	GBTo-31	M	Not Assigned	mand. right 3rd molar	2090± 60	2058± 87		A*
XVII-B-318	GBTo-31	M	Not Assigned	mand right 2nd molar	1550± 50	1432± 68		D*
XVII-B-322	GBTo-31	M	Not Assigned	mand right 3rd molar	2050± 50	2003± 65		A*
XVII-B-357	GBTo-31	M	XX	mand right 3rd molar	n/a	n/a	cal 50 BC to 690 AD	A2
XVII-B-365	GBTo-31	F	XX	molar	2270± 65	2290± 75		A*
XVII-B-386	GBTo-31	M	Not Assigned	mand left 2nd molar	1060± 40	962±33		A*
XVII-B-406	GBTo-31	F	XX	mand right 3rd molar	n/a	n/a	cal 50 BC to 690 AD	A*
XVII-B-412	GBTo-31	F	Not Assigned	mand right 3rd molar	1940± 40	1883± 40		A2
XVII-B-413	GBTo-23	F	XX	molar	1970± 42	1913± 42		A2
XVII-B-443	GBTo-31	M	Not Assigned	max left 2nd molar	1820± 55	1750± 70		A2
XVII-B-468	GBTo-33	M	XY	mand right 3rd molar	1940± 45	1883± 46		A2
XVII-B-470	GBTo-33	M	XY	mand right 3rd molar	1600± 40	1488± 52		A2
XVII-B-507	GBTo-36	M	Not Assigned	mand right 3rd molar	2320± 65	2335± 61		A*
XVII-B-516	GBTo-36	M	Not Assigned	mand right 2nd molar	n/a	n/a	cal 300 BC to 440 AD	A2
XVII-B-525	GBTo-31	M	Not Assigned	mand right 2nd molar	1860± 40	1800± 51		A2
XVII-B-532	GBTo-36	M	XY	molar	n/a	n/a	3200 to 1500 BP	A2
XVII-B-939	GBTp-1	M	XX	mand left 2nd molar	5710± 40	6487± 50		D4h3a

Table 3. Ancient sample characterization. Sequence-based sex determination was completed as described in Skoglund *et al.* (Skoglund et al., 2013). Mitochondrial haplogroups designated with * were assigned via RFLP.

Results

To investigate possible immune-related genes under selection before and after European contact, the exomes of ancient and living indigenous individuals from the Northwest Coast First Nation were sequenced. This was followed by non-candidate population genomic scans for natural selection and functional characterization of genes with the strongest signals. Exomes of 25 living individuals from the community were sequenced to an average read depth of 9.66x. The exomes of 25 ancient individuals from the PRH sites were sequenced to a mean depth of 7.97x. Only individuals dated older than 1,000 years BP (Table 3) were selected for sequencing to assure pre-European contact status. Contamination estimates using the exome-wide data revealed an average contamination of 0.684% at a 95% credibility interval ((Cibulskis et al., 2011), Table 9). All 25 individuals exhibited deamination patterns consistent with post-mortem DNA damage ((Briggs et al., 2007; Seguin-Orlando et al., 2013), Fig. 15). Mitochondrial haplogroups were determined for each ancient individual, all showing Native American lineages (Table 3; (Cui et al., 2013)).

Before proceeding with the selection scans, the genetic relationship between the ancient and living individuals was first explored, in order to ensure that a single population was being examined through time. For these analyses, transitions were not consider in order to prevent potential bias resulting from DNA damage. Genotypes were not called due to the low coverage of the ancient samples, which may result in statistical uncertainty. I therefore relied on a probabilistic model that can take these uncertainties into account through genotype likelihoods and an estimation of allele frequencies (Nielsen, 2012). A multi-dimensional scaling (MDS) plot representing genetic distances between individuals, including populations from the 1,000 genomes project (http://1000genomes.org, Fig. 11a), reveals an association between the two indigenous groups, with the living population drifting towards Europeans, as expected from known admixture (Verdu et al., 2014). The genetic affinity of the ancient individuals to both contemporary and ancient genomes was also assessed via f_3-statistics, which revealed that the closest relationship is to Native American populations, in particular to the living First Nation (Fig. 17). Next, the individuals were compared via a maximum likelihood tree using the program *TreeMix* (Pickrell and Pritchard, 2012). At 1 migration event, the ancients exhibit minimal drift

and appear ancestral to the living, with an admixture event occurring between the living and European populations (Fig. 11b). A cluster analysis completed with the program *ADMIXTURE* (Alexander et al., 2009) also supports this relationship, where *(at K=5)* the living show a large ancient component, as well as a component from the European population (Fig. 11c).

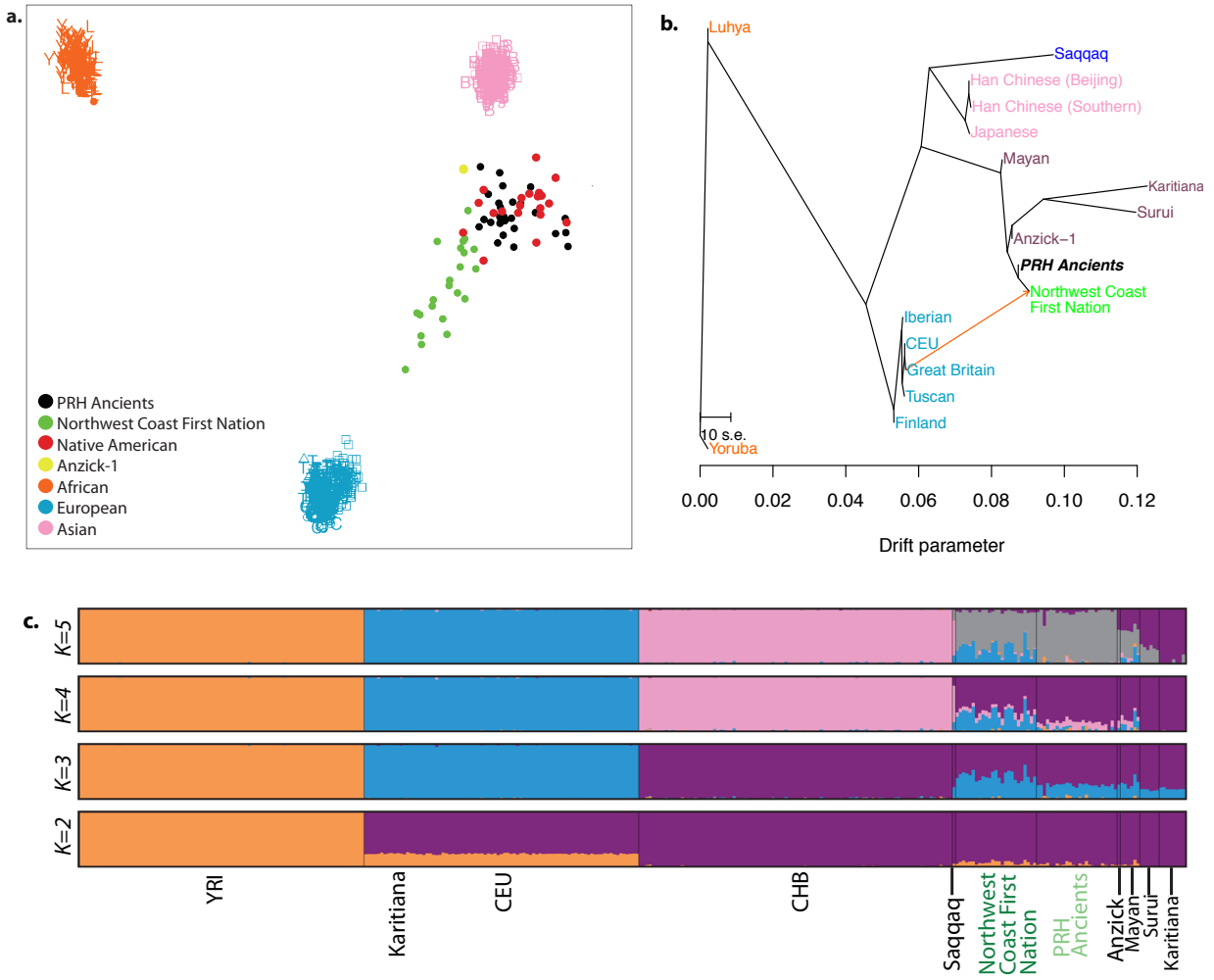

Figure 11. Population structure of the ancient individuals from Prince Rupert Harbour. a) Multidimensional Scaling Analysis (MDS) Plot includes 1,000 genomes reference populations and Native American Populations sampled from the Karitiana, Surui, and Maya. b) Maximum likelihood tree created with TreeMix, showing one migration event. c) ADMIXTURE Cluster Analysis depicting ancestry proportions assuming the number of genetic components, K, is equal to 1 through 5, for a set of 329 living Eurasian, African and Native American populations, along with 2 ancient Native Americans. The model at K=5 was found to have the best predictive power, as determined by the lowest cross validation index values.

The population history of the indigenous population was also modeled by taking into account the severe bottleneck that occurred after European contact (Larsen, 1994; Thornton, 1997; O'Fallon and Fehren-Schmitz, 2011) (*see* Methods). Utilizing both ancient and modern genome-wide data, the population decline was modeled with the mutli-dimensional frequency spectra of potentially synonymous sites with respect to hg19. The best-fitting model suggests that a bottleneck occurred approximately 150 years ago in the First Nation, with the split between the East Asian and ancestral native population occurring approximately 14,800 years ago. The timing of the bottleneck coincides with the smallpox epidemics of the 19th Century and historical reports of large population declines (Boyd, 1990; Sequin and Halpin, 1990).

In order to prevent false-positive signals of selection due to the apparent admixture of the living individuals with Europeans, an admixture correction was performed (*see* Methods). The genetic differentiation between the ancients and the admixed-corrected living individuals was relatively low ($F_{ST=}$~0.02) and represent an appropriate comparison. The populations were scanned for selection signals utilizing the Populations Branch Statistic (PBS; (Yi et al., 2010)). PBS has proven effective in detecting selected loci amongst high-altitude populations (Yi et al., 2010; Huerta-Sanchez et al., 2013) and takes into account a 3rd more distantly related population. 25 Han individuals from Beijing, part of the 1,000 genomes project (http://www.1000genomes.org), served as the third comparative population. This allowed us to compare the pairwise F_{ST} values of all three populations in order to determine the frequency change that occurred in the ancient since the divergence from the living individuals (Shriver et al., 2004). The resulting PBS scores (Table 4) correspond to the magnitude of the allele frequency change at a given locus in the ancient individuals.

Gene	PBS	*P* value	SNPs	Function
HLA-DQA1	0.278776809	0.00032	16	Major histocompatibility complex, class II, DQ alpha 1.
BLZF1	0.227968256	0.0016	56	Basic leucine zipper nuclear factor 1. Required for normal Golgi structure and for protein transport from the endoplasmic reticulum through the Golgi apparatus to the cell surface.
CELA3A	0.207270149	0.00311	27	Chymotrypsin-like elastase family, member 3A. Serine protease that hydrolyze many proteins in addition to elastin.
C1orf216	0.190562841	0.00555	65	Chromosome 1 open reading frame 216. Function not known.
OR6Q1	0.175425197	0.00985	12	Olfactory receptor, family 6, subfamily Q, member 1.
TOMM7	0.171970388	0.01137	19	Translocase Of Outer Mitochondrial Membrane 7 Homolog. Regulates the assembly and stability of the translocase complex.
CCDC181	0.166360505	0.01391	25	Coiled-Coil Domain Containing 181.
PTF1A	0.146513385	0.02793	9	Pancreas specific transcription factor, 1a. May be involved in the maintenance of exocrine pancreas-specific gene expression including ELA1 and amylase.
OR3A1	0.13890171	0.0364	22	Olfactory receptor, family 3, subfamily A, member 1.
BTNL2	0.131518242	0.04735	24	Butyrophilin-like 2 (MHC class II associated. Negative regulator of T-cell proliferation.

Table 4. Genes with strongest frequency changes in the ancient individuals.

Ancient genes showing the most extreme PBS values represent strong candidates for selection, of which 5 are directly involved with immune function (Table 4*).* Enriched gene ontologies were also identified from the ranked list of genes generated from the PBS scan, which highlight immune function related to antigen presentation (Table 5). In order to confirm the selection signals, with respect to alternative demographic explanations, the extreme PBS scores were compared against neutral simulations using our demographic history model, which allowed for the calculation of *P* values for each score (*see* Methods). Variants from the top candidate with the most pronounced frequency changes were confirmed via Sanger sequencing in all ancient samples reporting data (*see* Methods).

Ontology Enrichment Term	P-value	FDR q-value
Olfactory receptor activity	3.17E-09	1.30E-05
MHC class II receptor activity	5.34E-07	1.09E-03
Peptide antigen binding	3.09E-06	4.21E-03
Receptor activity	1.26E-05	1.28E-02
G-protein coupled receptor activity	1.38E-05	1.13E-02
Peptidase inhibitor activity	4.36E-05	2.97E-02
Molecular transducer activity	5.13E-05	3.00E-02
Peptidase regulator activity	1.16E-04	5.95E-02
Transmembrane signaling receptor activity	1.75E-04	7.95E-02
Endopeptidase inhibitor activity	1.86E-04	7.62E-02
Signaling receptor activity	2.15E-04	7.99E-02
Endopeptidase regulator activity	2.92E-04	9.95E-02
Antigen binding	3.01E-04	9.47E-02
Serine-type endopeptidase activity	4.75E-04	1.39E-01
Cytokine activity	5.15E-04	1.40E-01
Serine-type peptidase activity	5.25E-04	1.34E-01
Beta-3 adrenergic receptor binding	5.74E-04	1.38E-01
Serine hydrolase activity	6.22E-04	1.41E-01
Nucleoside triphosphate adenylate kinase activity	7.59E-04	1.64E-01

Table 5. Enriched gene ontologies. Derived from the PBS selection scan on the ancient individuals; calculated via GOrilla (http://cbl-gorilla.cs.technion.ac.il).

Discussion

The most extreme PBS score belonged to the human leukocyte antigen *(HLA)-DQA1,* which encodes for the alpha chain of the major histocompatibility complex (MHC), class II, DQ1 isoform. The *HLA-DQA1* SNP with the most pronounced frequency difference between the ancient (100%) and living (36%) falls in an UTR5 region of the gene, which may be indicative of selection acting on the regulation of the gene (Table 10). The chromosomal region where the gene is located also shows strong differentiation in the ancient individuals when compared to the living (Fig 12). An analysis for the enrichment of gene ontology terms also revealed a significance for MHC Class II receptor activity and peptide antigen binding in the ancient individuals (Fig. 13).

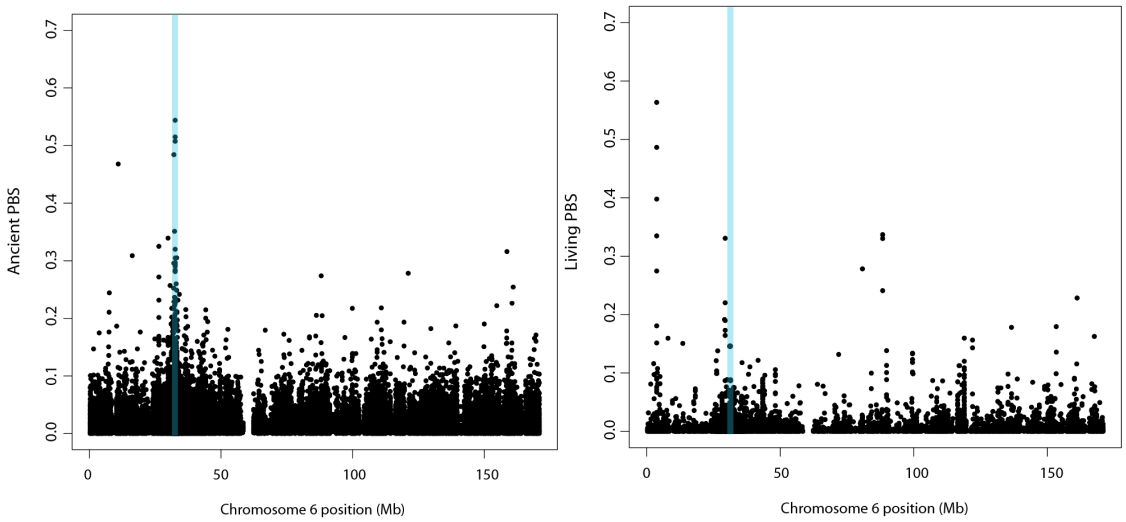

Figure 12. PBS score vs. gene position on chromosome 6. The highlighted region shows PBS scores for SNPs within 10kb of the *HLA-DQA1* gene. The living were corrected for European admixture.

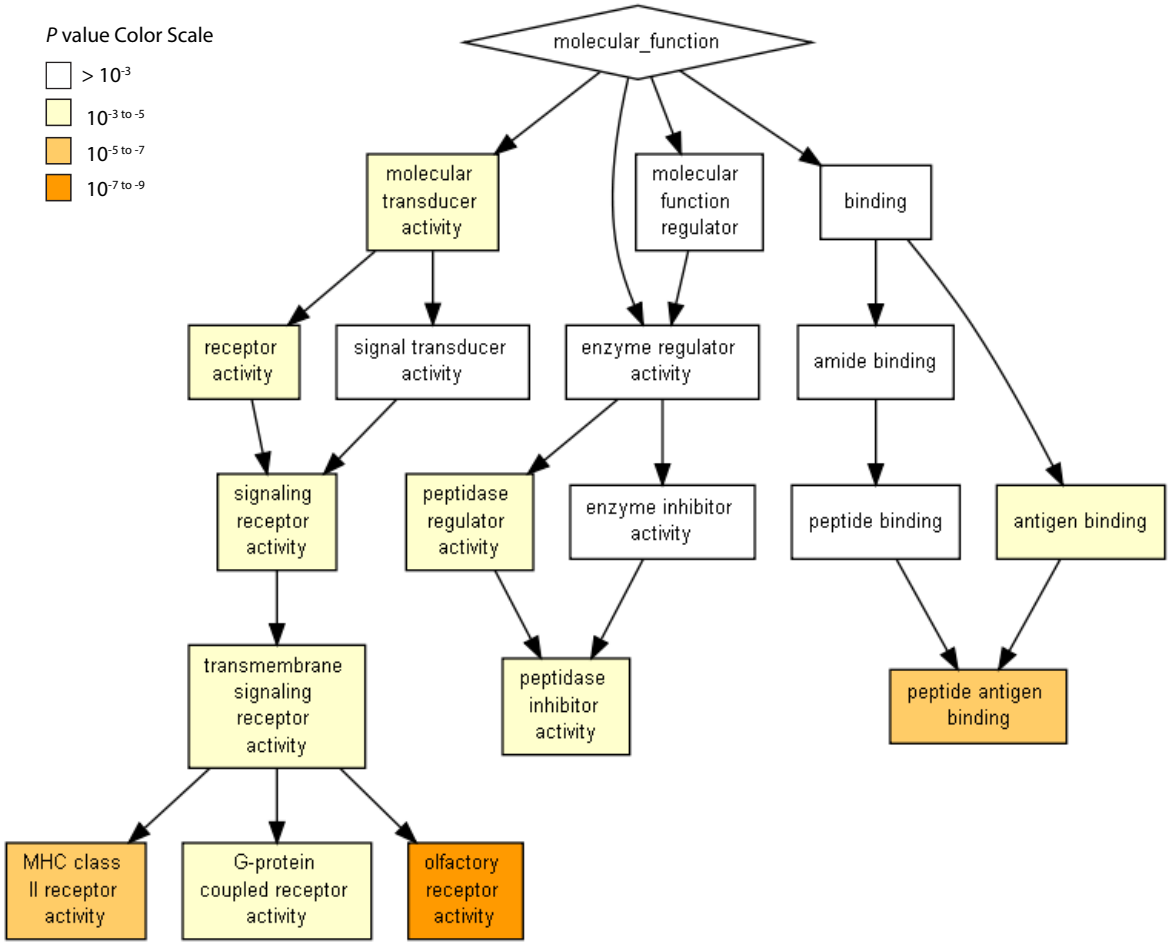

Figure 13. Gene Ontology Enrichment for the ancient individuals amongst PBS genes. Data produced with the gene list derived from the PBS selection scan and GOrilla (http://cbl-gorilla.cs.technion.ac.il).

HLA-DQ is one of the three main types of MHC class II molecules, along with DR and DP, and are expressed mainly expressed on antigen presenting cells (Jones et al., 2006). MHC class II molecules are responsible for binding to extracellular pathogen peptides and presenting them to CD4+T helper cells, which activate a targeted adaptive immune response towards the associated microbe (Roche and Furuta, 2015). These molecules are known to be highly polymorphic, mainly due to sequence differences corresponding to the binding domain of the molecule, which impact the type of peptides that can be bound(Wang et al., 2008). Because of this variety in binding domains, differing MHC class II isoforms can have differing disease outcomes due to the restriction imposed on T-cell activation (Roche and Furuta, 2015). The polymorphic nature of these molecules across different populations, however, would not explain the heightened differentiation in the ancient individuals with respect to their living descendants.

Alleles of *HLA-DQ* have been associated in a variety of colonization era infectious diseases, including measles (Moss and Griffin, 2006; Ovsyannikova et al., 2006), tuberculosis (Kim et al., 2005; Delgado et al., 2006), and with regard to the adaptive immune response to the vaccinia virus, which is an attenuated form of smallpox (Ovsyannikova et al., 2011; 2014). Although it is not possible to know if the ancient alleles putatively under selection may pose a differential disease outcome with respect to the smallpox virus directly (due to its eradication), the use of ancient DNA affords a unique opportunity to examine the ancient frequencies of these alleles. If we consider the low genetic differentiation between the living and ancient individuals, we may examine the frequency of the *HLA-DQA1* alleles as they fluctuate through time in a single population. The significant frequency changes between the two time periods, before and after European contact, suggests a case for positive selection followed by negative frequency-dependent selection of the *HLA* gene (Takahata and Nei, 1990). Selection of the gene in the ancestral population may have driven it to high frequency and produced a relatively common set of alleles, but may have proven disadvantageous when the environment drastically changed upon European contact. The common form of the *HLA-DQA1* gene, at the time smallpox was introduced to the Pacific Northwest Coast, may have increased the virulence of the virus in the indigenous population, thereby decreasing the frequency of the allele in modern times. Further microbiological studies will be needed to understand the alleles' potential affects on the

virulence of the attenuated form of smallpox (vaccinia virus), as well as their affect on downstream target genes.

Another possible correlation exists between olfactory and HLA receptors, which both show significant selection signals and ontology term enrichment (Fig. 13, Table 5). The correlation may provide an alternative hypothesis, involving sexual selection, as to why the associated immune genes may have been driven to high frequency in the ancestral population. HLA has been theorized to influence mating preference by affecting an individual's unique odor (Jacob et al., 2002; Ziegler et al., 2005; Havlicek and Roberts, 2009). The olfactory receptors that detect these molecules are also theorized to have come under selection, given their role in processing the socially complex signals (Linnen and Hoekstra, 2010; Pause, 2011; Milinski et al., 2013). Although this correlation would require further exploration on a molecular level, the result would be the same in terms of creating a negative frequency dependent scenario upon the introduction of a potentially novel pathogen, which may have been able to take advantage of the ancient immunological state of the population.

Additional genes identified in this study may relate to other ancient adaptive events (64). These include the aforementioned olfactory genes (*OR6Q1 and OR3A1),* which code for receptors that bind to odorant molecules. Olfactory receptors have been a known target of evolution in humans, where putatively selected variants within these genes can alter odor perception (Gilad et al., 2003; Keller et al., 2007). Another pair of candidate gene correlate with digestion. *PTF1A* codes for a pancreas specific transcription factor, which may be involved in the expression of amylase and elastase 1 (Kawaguchi, 2002; Dong et al., 2007). A related gene, *CELA3A*, codes for a pancreatic enzyme linked to digestive function in the intestine (Whitcomb, 2007). It may also be involved in the transport of and metabolization of cholesterol (Tani et al., 1988). These selection signals and their potential link to adaptive traits, in terms of digestion and olfactory sense, may have proven advantageous in the ancient environment of the Americas.

Selection scans were also run on the living population, with and without the European admixture correction. Neither list show significant selection on any immune-related genes, which may suggest that pathogen driven selection has acted negatively on this functional class (Table 6, 7).

The gene ontology enrichment analyses, however, reveal two interesting patterns. The enrichment list derived from the admixed corrected selection scan shows a correlation with the negative regulation of the intestinal absorption of cholesterol (Fig. 14). Two genes are associated with this enrichment, *ABCG5* and *ABCG8*. These genes have been shown to not only regulate the cholesterol absorption but also have been correlated to the specific regulation of sterol absorption from shellfish, which can be toxic if absorbed at high levels (Lee et al., 2001; Yu et al., 2002; Nguyen et al., 2012). This is an intriguing potential adaptation since the First Nation have a long standing history of a marine based diet (Cybulski, 2001). Furthermore, this may also represent a distinct adaptation in lieu of the more common *ABCA1* allele—found in other Native American populations that regulates cholesterol absorption in communities with agricultural-based diets (Plat and Mensink, 2002; Acuña-Alonzo et al., 2010).

Examining the gene ontology enrichment list for the selection scan on the living individuals, *without* European admixture correction, reveals a different pattern. There seems to be an enrichment related to the metabolic processing of alcohol dehydrogenase (Fig. 15). The four genes associated with this enrichment are *ADH5, ADH6, ALDH2*, and *ADH1B*. These genes may be involved in the rate at which alcohol and acetaldehyde are metabolized and have been associated with both positive and negative effects on alcohol induced organ damage and dependence (Mulligan et al., 2003; Ehlers et al., 2012; Ramchandani, 2013). Given that this enrichment is not revealed in the admixed corrected analysis, it may be possible that these alleles have entered the population as a result of gene flow from Europeans.

In conclusion, the results of the selection scans presented here reveal a nuanced evolutionary history of Native Americans, which is greatly complicated by European contact. Given the pronounced frequency changes of immune-related genes showing the strongest signals of selection, there appears to be an evolutionary shift in the indigenous communities of the Northwest Coast after the severe population declines associated with the smallpox epidemics of the 19th Century. The evolutionary trend points towards rapid adaptive events focusing on negative selection of previously selected immune related alleles, that may function in a frequency dependent manner. The removal of ancient alleles adapted to the ancient environment may have mitigated the virulence of specific pathogens introduced after European colonization,

allowing for increased survivability during the epidemics of the 19th Century. Although detailed microbiological studies are needed to understand the possible effects of these genes with a given pathogen, such as smallpox, an interesting picture has emerged depicting the genetic resilience of Native Americans in the face of rapid environmental change. This may help to better understand the historical experiences of Native Americans with European borne disease by integrating their complex evolutionary history, which spans thousands of years.

Gene	PBS Score	*P* value	SNPs	Function
PDCL3	0.2998444	5.00×10^{-05}	32	Phosducin-like 3. Acts as a chaperone for the angiogenic VEGF receptor KDR/VEGFR2, controlling its abundance and inhibiting its ubiquitination and degradation. Modulates the activation of caspases during apoptosis.
CNOT11	0.173709579	0.00545	112	CCR4-NOT transcription complex, subunit 11. Component of the CCR4-NOT complex which is one of the major cellular mRNA deadenylases and is linked to various cellular processes including bulk mRNA degradation, miRNA-mediated repression, translational repression during translational initiation and general transcription regulation.
PSG5	0.168520517	0.00656	16	Pregnancy specific beta-1-glycoprotein 5. The human pregnancy-specific glycoproteins (PSGs) are a group of molecules that are mainly produced by the placental syncytiotrophoblasts during pregnancy.
RAB6C	0.165826759	0.00732	12	Member RAS oncogene family.
SUMO2	0.159406912	0.00915	14	Small ubiquitin-like modifier 2.
GPR21	0.148679547	0.01341	22	Ubiquitin-like protein that can be covalently attached to proteins as a monomer or as a lysine-linked polymer.
PDYN	0.140294814	0.0185	47	Prodynorphin. Involved in physiologic functions, including pain perception and responses to stress.
PXDC1	0.136720516	0.02123	95	PX domain containing 1.
POLE4	0.135276994	0.0223	37	Polymerase (DNA- directed), epsilon 4, accessory subunit. May play a role in allowing polymerase epsilon to carry out its replication and/or repair function.
OR11A1	0.132702324	0.02443	36	Olfactory receptor, family 11, subfamily A, member 1.
NEUROG3	0.131154492	0.02577	47	Neurogenin 3. Acts as a transcriptional regulator.
OR10C1	0.128797097	0.02805	28	Olfactory receptor, family 10, subfamily C, member 1 (gene/pseudogene).
TCP10L2	0.1287457	0.02811	6	T-Complex 10-Like 2.
CDH10	0.127837578	0.02898	111	Cadherin 10, type 2 (T2-cadherin).
HIST1H4H	0.126585295	0.03045	12	Histone cluster 1, H4h. Core component of nucleosome.
CR1L	0.121785583	0.03647	53	Complement component (3b/4b) receptor 1-like.
ADH6	0.121539163	0.03683	56	Alcohol dehydrogenase 6 (class V).
CCDC74B	0.114047862	0.04937	23	Coiled-coil domain containing 74B.

Table 6. Functions of Genes with *P* values below 0.05 for the living individuals with admixture correction.
Gene functions were derived from GeneCards (www.gencards.org).

Gene	PBS Score	*P* value	SNPs	Function
CYP4Z1	0.329829344	9.0x10-05	17	Mitochondrial translational release factor 1-like.
PDCL3	0.321767386	8.0x10-05	39	Phosducin-like 3. Acts as a chaperone for the angiogenic VEGF receptor KDR/VEGFR2, controlling its abundance and inhibiting its ubiquitination and degradation. Modulates the activation of caspases during apoptosis.
PSG5	0.169863548	0.00772	19	Pregnancy specific beta-1-glycoprotein 5. The human pregnancy-specific glycoproteins (PSGs) are a group of molecules that are mainly produced by the placental syncytiotrophoblasts during pregnancy.
CNOT11	0.169410523	0.00782	160	CCR4-NOT transcription complex, subunit 11. Component of the CCR4-NOT complex which is one of the major cellular mRNA deadenylases and is linked to various cellular processes including bulk mRNA degradation, miRNA-mediated repression, translational repression during translational initiation and general transcription regulation.
SUMO2	0.156278054	0.02801	18	Small ubiquitin-like modifier 2.
PDYN	0.154671374	0.02965	84	Ubiquitin-like protein that can be covalently attached to proteins as a monomer or as a lysine-linked polymer.
RAB6C	0.143426055	0.01298	13	Member RAS oncogene family.
GPR21	0.141307594	0.04202	32	Ubiquitin-like protein that can be covalently attached to proteins as a monomer or as a lysine-linked polymer.

Table 7. **Functions of Genes with *P* values below 0.05 for the living individuals *without* correcting for admixture.** Gene functions were derived from GeneCards (www.gencards.org).

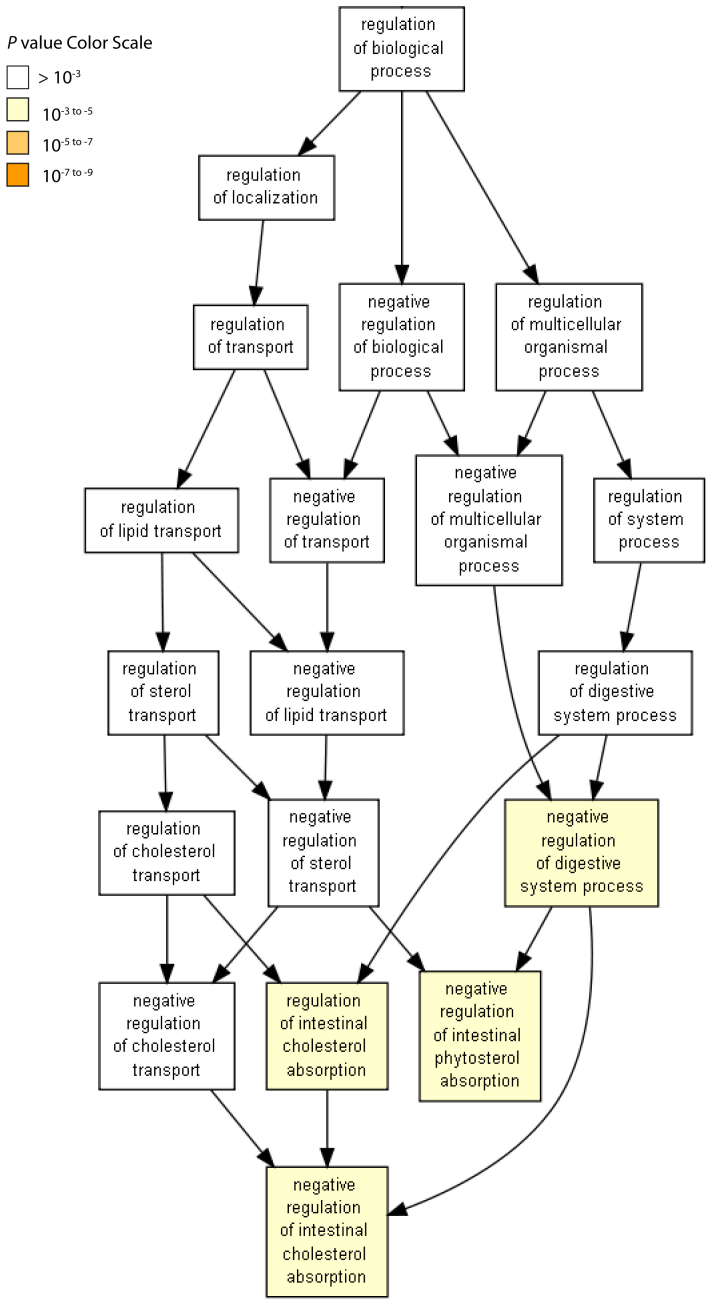

Figure 14. Gene Ontology Enrichment for the living individuals amongst PBS genes, with correction for admixture with Europeans. Data produced with the gene list derived from the PBS selection scan and GOrilla (http://cbl-gorilla.cs.technion.ac.il).

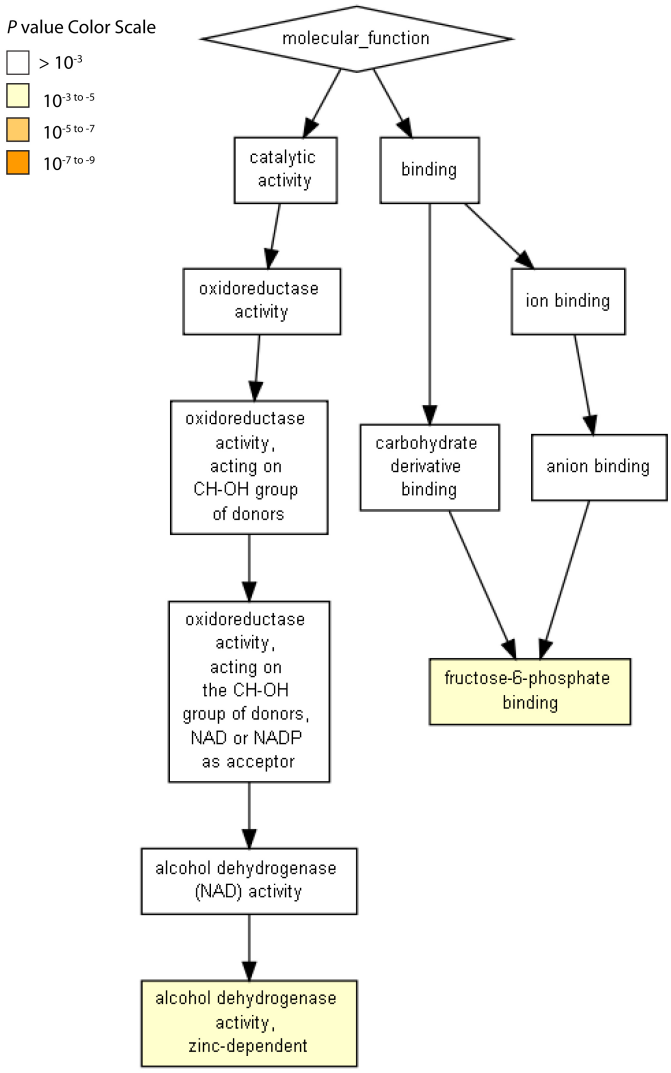

Figure 15. Gene ontology enrichment for the living individuals amongst PBS genes, with *no correction* for admixture with Europeans. Data produced with the gene list derived from the PBS selection scan and GOrilla (http://cbl-gorilla.cs.technion.ac.il).

Methods

DNA Extraction

DNA extractions and PCR amplification setups were completed in an ancient DNA laboratory facility at the University of Illinois Urbana-Champaign. The ancient DNA lab is a positively pressured clean room with hepa-filtered air. The clean room contains an anteroom and air flows from the ancient DNA lab to the anteroom to the hallway. Personnel working in the ancient DNA lab wear disposable hairnets, facemasks, laboratory coveralls and booties. All equipment, reagents and consumables are dedicated for use in the ancient DNA laboratory. The ancient

DNA lab is routinely cleaned with bleach and all containers are wiped with Takara DNA Off (Mountain View, CA) before placed in the ancient DNA lab. Personnel are restricted from entering the ancient DNA after being in a contemporary DNA laboratory. A database containing mitochondrial control region sequence is maintained of all personnel working in the ancient laboratory and of any personnel who may have come into contact with the human remains prior to DNA analysis. Contamination controls were used with every DNA extraction and PCR setup in order to detect any contamination from reagents. A series of negative controls are routinely performed in the ancient DNA lab.

Each tooth was soaked in 6% sodium hypochlorite for 3 minutes, rinsed three times with UV-irradiated molecular grade water, and dried in a UV Crosslinker for 10 minutes, so as to remove surface contamination. Approximately 0.20 grams of tooth powder was incubated in 4 ml of demineralization/lysis buffer (0.5 M EDTA, 33.3 mg/ml Proteinase K, 10% N-lauryl sarcosine) for 24 hours at 37ºC. The digested sample was then concentrated to approximately 250 µl using Amicon centrifugal filter units. Following concentration, the digest was run through silica columns using the MinElute Qiagen PCR Purification Kit (Qiagen, Hilden, Germany) and eluted in 60 µl of DNA extract.

DNA Screening for mtDNA

In order to test for viable DNA before proceeding with library building and exome sequencing, each ancient individual was amplified for the hypervariable region I of mitochondrial DNA from 2 µl of extract, utilizing the reagents and conditions described in Malhi *et al.* (Malhi et al., 2003). Native American maternal lineage was also confirmed for each sample via restriction fragment length polymorphism analysis for Native American haplogroups (A, B, C, D, and X) (Kaestle and Smith, 2001), or via off-target reads from the exome capture and previously performed captures of the mitochondrial genome, as described in Cui *et al.* (2013,Table 3). All 25 ancient samples demonstrate Native American mitochondrial haplogroups.

Library Build and Exome Sequencing

Libraries were created using the New England Biolabs Ultra Kit (E7370S, Ipswich, MA) following the manufacturer's protocol with the following modifications: DNA fragmentation was

not performed. DNA purifications were done using the MinElute Reaction Cleanup Kit (Qiagen, Valencia, CA). Library amplification was done in two steps. The first round of amplification utilized the kit's reagents and protocol with 12 cycles of (10s at 98°C, 30s at 65°C and 30s at 72°C). For the second round, I achieved a sufficient DNA concentration for the exome enrichment (~500ng), without excessive amplification, by creating 4 PCR reactions from the initial amplified product and then pooling them before using a Qiagen MinElute PCR Clean-up kit. For the 2nd PCR, I created a 50 µl reaction, utilizing 0.2µM of primers P5 (5'-AATGATACGGCGACCACCGA) and P7 (5' CAAGCAGAAGACGGCATACGA) ((Meyer and Kircher, 2010)), 5µl from the initial PCR, 25µl of Phusion® High-Fidelity PCR Master Mix with HF Buffer (New England Biolabs, Ipswich, MA), 3% DMSO (New England Biolabs, Ipswich, MA), 0.2mg/ml BSA (New England Biolabs, Ipswich, MA). PCR conditions were as follows: 4m at 98°C, 10 cycles of (10s at 98°C, 30s at 62°C and 30s at 72°C), with a final extension at 72°C for 10m. Library fragment sizes were confirmed via a BioAnalyzer High Sensitivity assay to be above 130bp.

Exome Enrichment and Illumina Sequencing

For the living samples, a combination of the Illumina TruSeq Exome Enrichment Kit and the Nextera Rapid Capture Exome Kit (Illumina, San Diego, CA) were used (Table 8), following the manufacturer's protocol. One library per individual was sequenced (single-end reads) and pooled for a total of 4 libraries per lane on the Illumina HiSeq 2000 at the High-Throughput Sequencing Division of the W.M. Keck Biotechnology Center at the University of Illinois Urbana-Champaign.

For the PRH ancient samples, only the Illumina TruSeq Exome Enrichment Kit was used. For each ancient individual, 4 libraries were captured and then pooled for sequencing on one lane. For the capture, the manufacture's protocol was used with the following modifications: the Qiagen MinElute PCR Clean-up kit was used instead of beads, post-capture amplification involved 12 cycles instead of 10, and the hybridization temperature was decreased to 50°C.

Living Samples	Exome Capture Method	Capture Libraries Sequenced	Illumina 62MB On-target Hits	Average Read Depth	**Ancient Samples**	Exome Capture Method	Capture Libraries Sequenced	Illumina 62MB On-target Hits	Average Read Depth
S001	Nextera	1	0.904283758	6.3422	**125**	Tru-Seq	4	0.328200145	2.03254
S002	Tru-Seq	1	0.930950435	8.62802	**158**	Tru-Seq	4	0.536221871	5.25787
T004	Tru-Seq	1	0.705098484	2.80743	**163**	Tru-Seq	4	0.065611919	1.75727
T008	Nextera	1	0.801474597	10.6555	**167**	Tru-Seq	4	0.341898258	1.85701
T012	Tru-Seq	1	0.663371532	2.57248	**168**	Tru-Seq	4	0.84163529	12.8529
T015	Tru-Seq	1	0.90549279	9.87665	**181**	Tru-Seq	4	0.887003	12.7429
T018	Nextera	1	0.829727742	12.59	**300**	Tru-Seq	4	0.580635306	3.09703
T023	Tru-Seq	1	0.767662613	3.049	**302**	Tru-Seq	4	0.912515032	33.115
T024	Nextera	1	0.860356081	14.41	**311**	Tru-Seq	4	0.665996613	3.8207
T026	Nextera	1	0.917903548	7.61586	**318**	Tru-Seq	4	0.263561226	2.37845
T028	Nextera	1	0.862068403	4.83134	**322**	Tru-Seq	4	0.528905694	2.78005
T036	Tru-Seq	1	0.846779968	4.96615	**357**	Tru-Seq	4	0.772120629	10.0139
T052	Nextera	1	0.869650113	4.83776	**365**	Tru-Seq	4	0.851621435	13.2557
T054	Nextera	1	0.921754565	22.0136	**386**	Tru-Seq	4	0.326494726	1.66748
T055	Nextera	1	0.840876871	4.13476	**406**	Tru-Seq	4	0.477336452	2.71423
T057	Nextera	1	0.842303194	5.00326	**412**	Tru-Seq	4	0.326017484	1.65456
T058	Nextera	1	0.939344113	18.4778	**413**	Tru-Seq	4	0.716144032	6.33255
T059	Nextera	1	0.885633032	11.1151	**443**	Tru-Seq	4	0.864574258	27.2342
T060	Nextera	1	0.875366774	16.1518	**468**	Tru-Seq	4	0.914177242	17.9042
T061	Nextera	1	0.946905661	17.8558	**470**	Tru-Seq	4	0.929462935	15.2033
T066	Nextera	1	0.929130468	12.3193	**507**	Tru-Seq	4	0.602631839	2.78834
T067	Nextera	1	0.922596113	9.82765	**516**	Tru-Seq	4	0.710534081	6.07661
T200	Tru-Seq	1	0.818794145	4.52119	**525**	Tru-Seq	4	0.298353194	3.39217
T201	Tru-Seq	1	0.928923984	19.96	**532**	Tru-Seq	4	0.701980726	4.35001
T202	Tru-Seq	1	0.862945242	6.95669	**939**	Tru-Seq	4	0.342927726	4.93794

Table 8. Exome enrichment results.

Mapping and Genotyping

Raw data from the Illumina HiSeq 2000 platform was base called with CASAVA 1.8.2. Sequences were de-multiplexed with a requirement for a full match of the 6 nucleotide indexes that was used for library preparation. Illumina adapter sequences were trimmed using Trimmomatic-0.32 (Bolger et al., 2014) with a minmum length of 25 and removing leading and trailing quality or N bases below a quality score of 3. Trimmed reads were aligned to the human reference genome hg19 using Bowtie2-2.1.0 (Langmead, 2012) with a local realignment option and a seed length set to 1000. SAMtools-1.1 (Li et al., 2009) was used to sort and remove duplicate reads based on mapping positions. Due to the low average read depth of the ancients, genotypes were not called directly. Instead, the program suite ANGSD (Nielsen, 2012) was used to compute genotype likelihoods using the SAMtools model and called SNPs via its estimation of allele frequencies. Each alignment used in the estimation was filtered for a minimum mapping quality of 30, a minimum base quality of 20, trimmed at each end for 5 bp to minimize biases

from DNA deamination, and a minimum p-value threshold of 10^{-6}. Sites were also filtered for violation of a one-tailed test for Hardy-Weinberg Equilibrium at a p-value < 10^{-4} (Shan, 2005).

DNA damage patterns

DNA damage (type I and type II) was assessed by comparing T–C/G–A and C–T/A –G transitions, respectively using MapDamage 2.0 (Jonsson et al., 2013). A specific pattern of DNA damage has been identified in other ancient DNA studies (Rasmussen et al., 2010; 2011). These studies show a pattern of increased type II DNA damage at the beginning and end of degraded DNA fragments. The MapDamage results show signatures of DNA damage, consistent with use of both the AT overhang library technique and blunt end (Seguin-Orlando et al., 2013), which suggests that the ancient sequences originate from ancient DNA templates and not modern contaminants (Fig. 16).

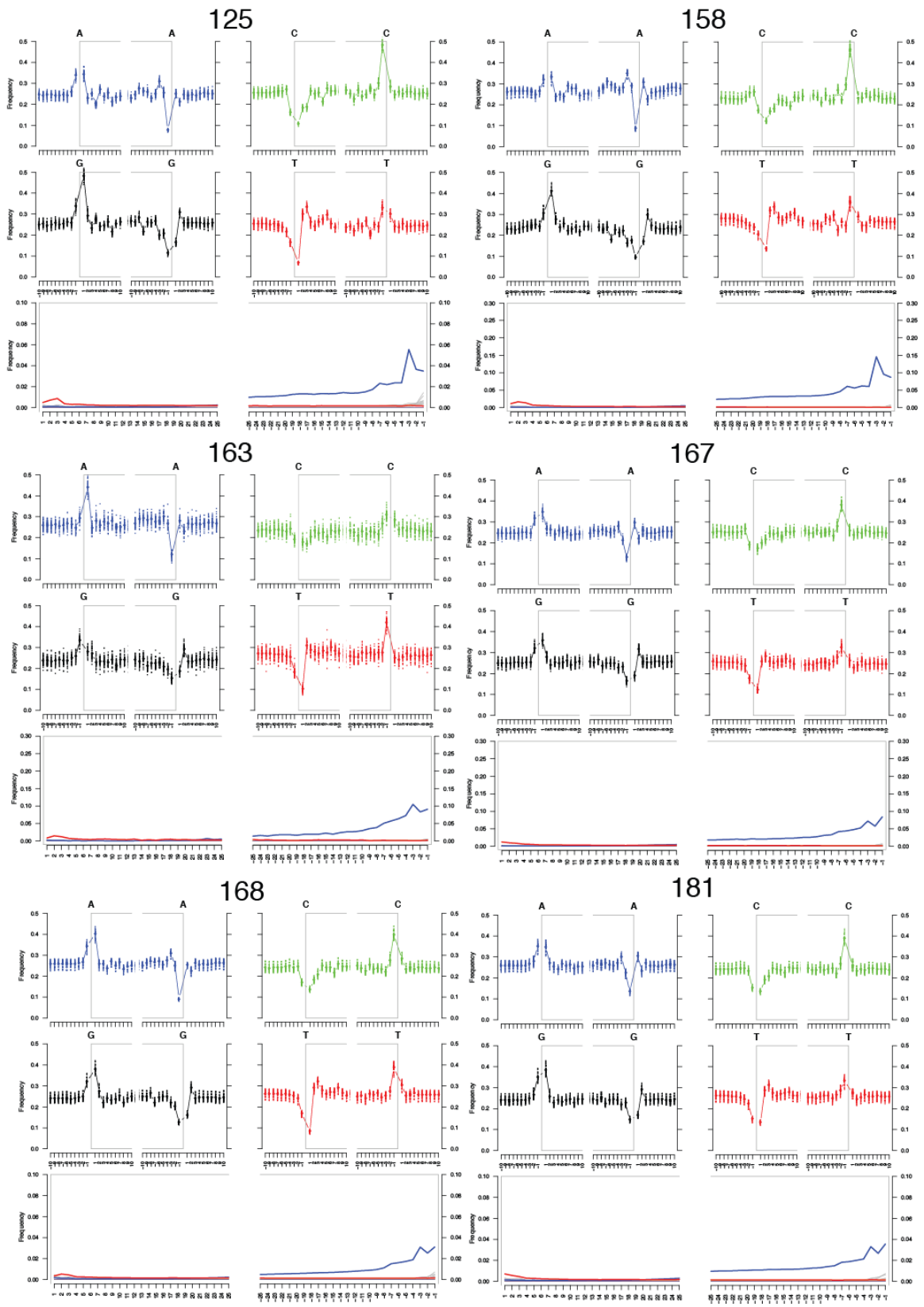

Figure 16 (Cont.)

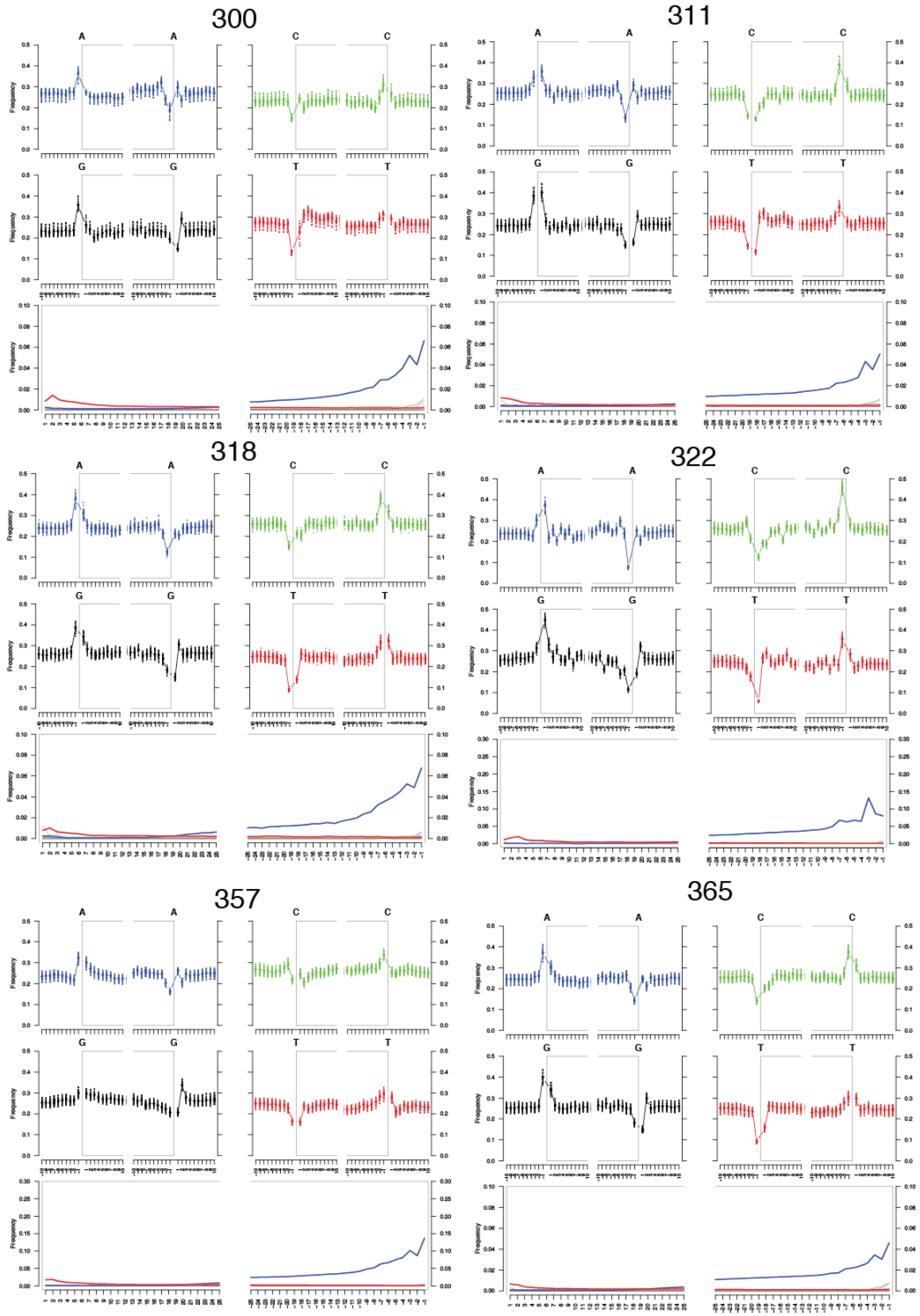

Figure 16 (Cont.)

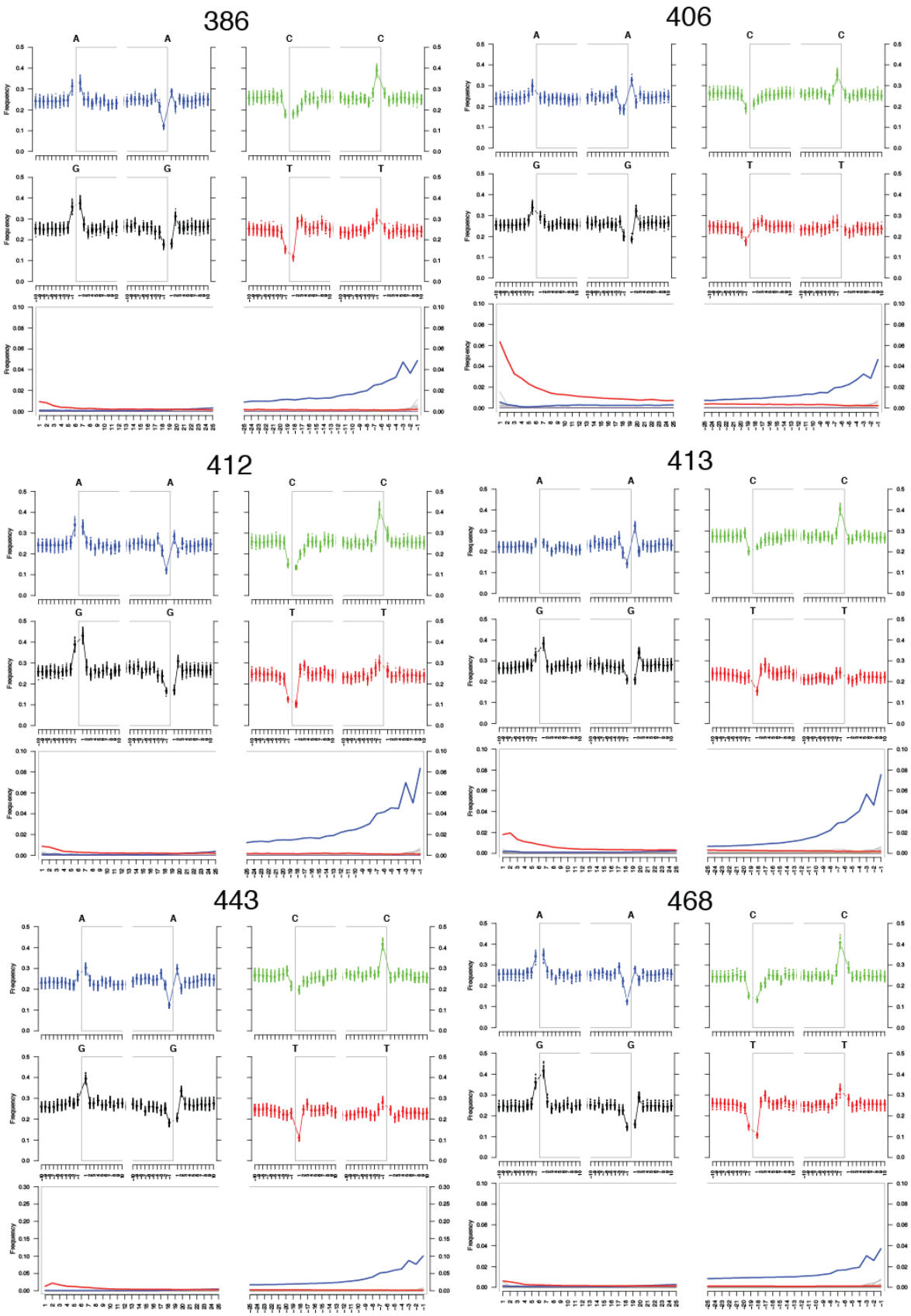

Figure 16 (Cont.)

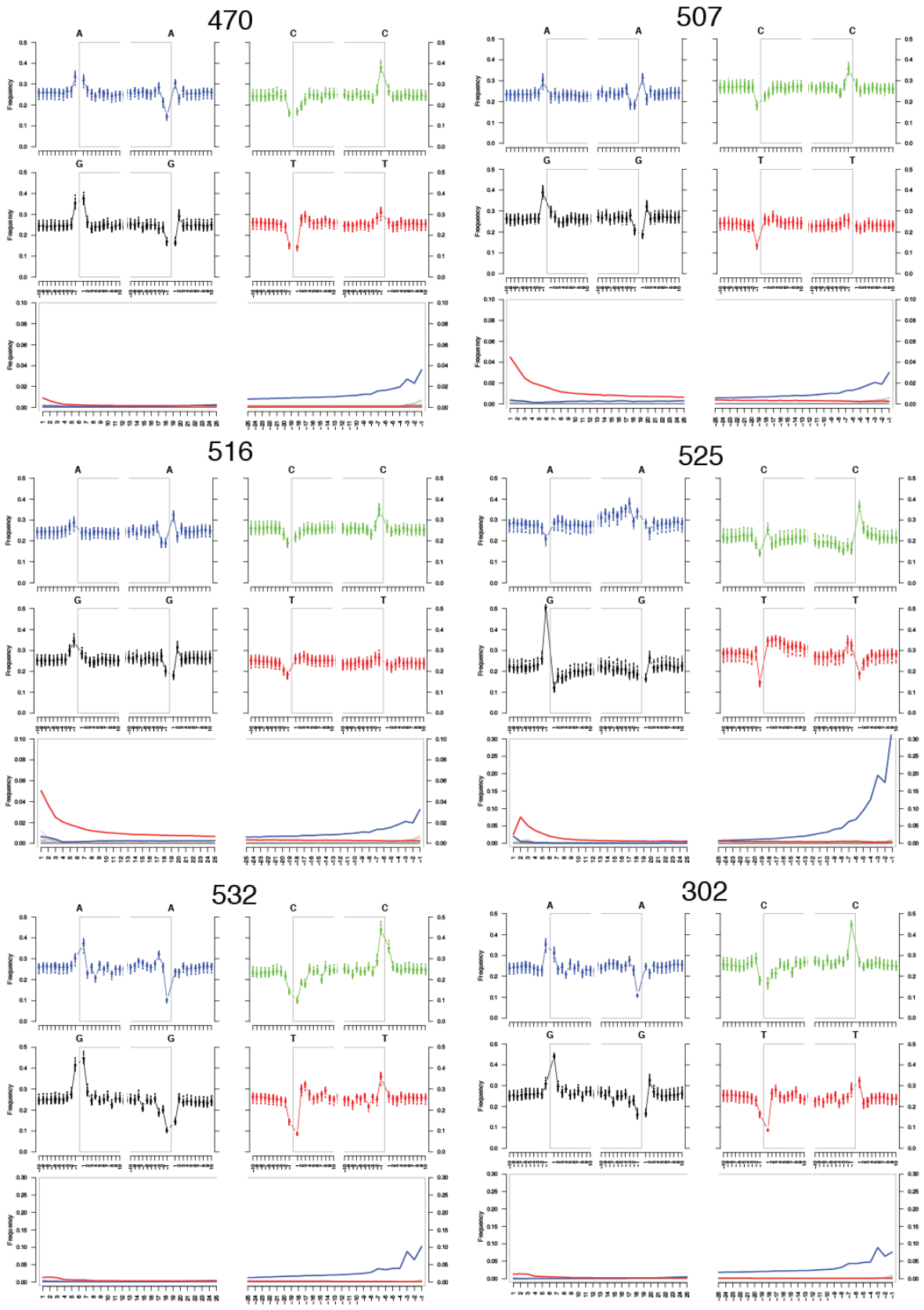

Figure 16 (Cont.)

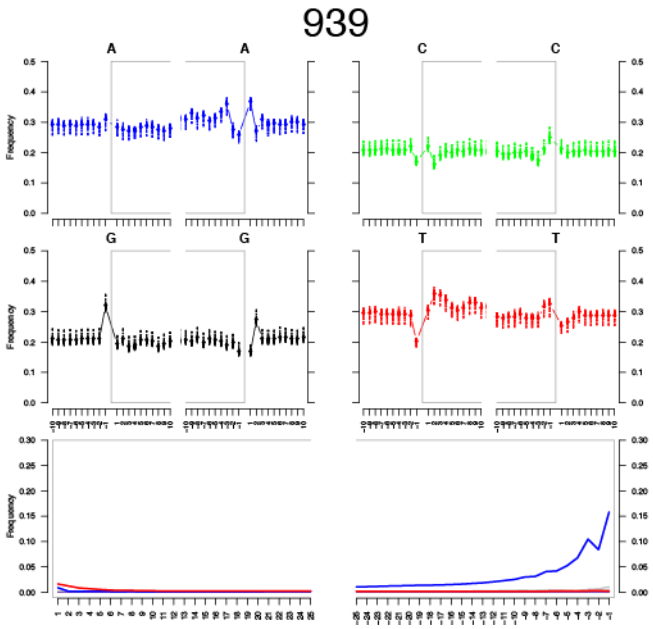

Figure 16. DNA damage patterns for the ancient individuals. Random subset of all mapped reads for each ancient individual. The mismatch frequency is relative to the reference as function of read position, C to T in red and G to A in blue.

Contamination Estimates

In order to estimate contamination across the genome-wide data, the ContEst tool was used (Cibulskis et al., 2011). This uses a Bayesian approach to calculate both the posterior and the maximum a posteriori probability of contamination level within a BAM file of an individual. This method has been shown effective in detecting contamination in exomes with low coverage. All ancient samples demonstrated contamination at or below 1%, except for sample 163 (Table 9). Due to the error issues associated with such a contamination level in detecting mutations, the sample was not used in the statistical analyses. HapMap_3.3 global populations frequencies for each SNP, mapped to b37, were used for the estimates.

Ancient Sample	Contamination Estimate Level at 95% Confidence	Confidence Interval 95 Low	Confidence Interval 95 High	Sites
125	0.6%	0.4%	0.7%	3221
158	0.4%	0.3%	0.5%	3957
163	5.8%	5.4%	6.3%	2497
167	0.8%	0.7%	0.9%	3097
168	0.8%	0.8%	0.9%	2991
181	0.6%	0.6%	0.7%	2416
300	0.7%	0.7%	0.8%	3094
302	0.3%	0.2%	0.4%	3693
311	0.1%	0.1%	0.2%	2605
318	0.4%	0.4%	0.5%	3501
322	0.7%	0.7%	0.9%	2627
357	0.5%	0.5%	0.6%	3402
365	0.8%	0.8%	0.9%	3139
386	0.3%	0.2%	0.4%	3441
406	0.8%	0.8%	0.9%	3309
412	0.4%	0.4%	0.5%	2835
413	0.7%	0.7%	0.8%	3740
443	0.4%	0.3%	0.5%	3471
468	0.8%	0.8%	0.9%	2563
470	0.5%	0.4%	0.6%	3819
507	0.1%	0.1%	0.2%	2899
516	0.1%	0.1%	0.2%	2827
525	0.2%	0.2%	0.3%	3726
532	0.2%	0.1%	0.3%	2645
939	0.1%	0.1%	0.2%	2074

Table 9. Ancient sample contamination estimates.

Multidimensional scaling (MDS) applied to exome sequencing data

For this analysis all transitions were removed from each population to mitigate possible biases due to DNA damage. For each sample, a single read was sampled uniformly at random at each site and the allele sampled was representative of the individual at that site. An allele sharing distance matrix was created as in Raghavan et al. (2014a), and classical multidimensional scaling was applied to the matrix.

Clustering Analysis of the ancient and living individuals with the 1,000 Genomes and contemporary Native American populations

To assess the genetic relationship of the ancient and living to each other and global populations, genome-wide data was compared from phase 3 of the 1,000 Genomes projects (YRI=112, CEU=90, CHB=109), three living Native American populations from South America (Maya, Surui, and Karitiana), and two ancient Native American individuals (Anzick (Rasmussen et al., 2014) and Saqqaq (Rasmussen et al., 2010)). For the following analyses, transitions were removed from each population to mitigate possible biases due to DNA damage. A total of 36,745 SNPs were used from the ancients and 32,827 from the living.

The program ADMIXTURE (Alexander et al., 2009) was used to assess model-based estimation of global ancestry. The runs for K=2 through K=5 are shown in figure 11c. At K=5 the ancients show a major genetic component depicted in gray (Fig. 11c). This component is also in the ancient and living Native American populations, suggesting a linked ancestry. It should be noted that the gray genetic component decreases as the populations move south, while the ancient Saqqaq individual from Greenland shares very little, suggesting a different migration wave.

Next, the genetic relatedness was assessed of the ancients to each of the living populations covered in phase 3 of the 1,000 genomes project, as well as the South American populations and ancient Saqqaq mentioned above, by computing outgroup f_3-statistics (Raghavan, 2013). This statistic shows the genetic sharing between two populations as they diverge from 0, with respect to their unrooted history with the designated outgroup. Pairwise affinities were compared to either the ancients, living, Saqqaq, Great Britain, or Han to the affinities of either the ancients, Karitiana, living, or Saqqaq. When compared to the Saqqaq, the ancients show greater affinity to the living and South American populations (Figure 17). When compared to the Great Brittan and Han populations, the ancients also cluster closer with the Native American populations.

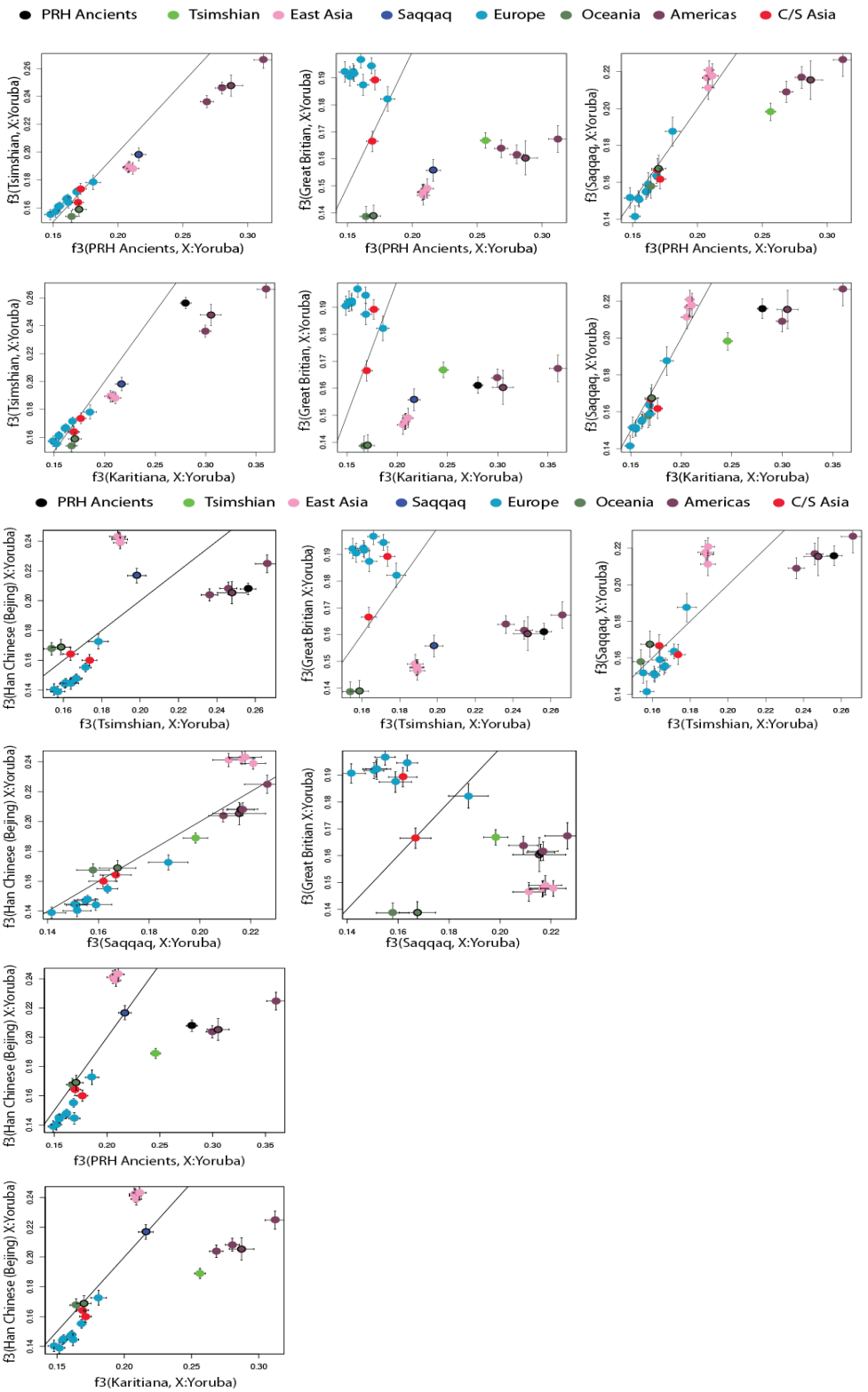

Figure 17. Top: Pairwise outgroup f_3 statistics comparisons using PRH Ancients and Karitiana on the *x*-axis and the Tismshian, Great Britain, and Saqqaq on the *y*-axis. Bottom: Pairwise outgroup f_3 statistics comparisons using the Tsimshian, Saqqaq, PRH Ancients and Karitiana on the *x*-axis and the Han Chinese, Great Britain, and Saqqaq on the *y*-axis.

In order to relate the ancients to global populations, using the same whole genome data described above, maximum likelihood trees were generated with *TreeMix* (Pickrell and Pritchard, 2012). Figure 11b depicts a maximum likelihood tree with a single migration event. The Native American populations split off from the Asian populations, with the exception of the Saqqaq, which, as noted above, may be due to a distinct migration. The Mayan are branched off from the other Native American groups, likely due to heavy admixture. The ancient Anzick individual seems splits off from the ancients and living. Again, this may be due to a different migration, where Anzick is ancestral to the Surui and Karitiana populations and the ancients are ancestral to the living. Figure 18 also generated by *TreeMix*, shows a residual fit plot from the maximum likelihood tree in Fig. 11b. Residual scores above zero represent populations that are more closely related to each other than in the tree. This excess relatedness can be indicative of admixture events. The living have very strong residuals across the European populations, indicating a possible admixture event between the two populations. The living and the ancients also exhibit an extreme residual, which demonstrates their ancestral link.

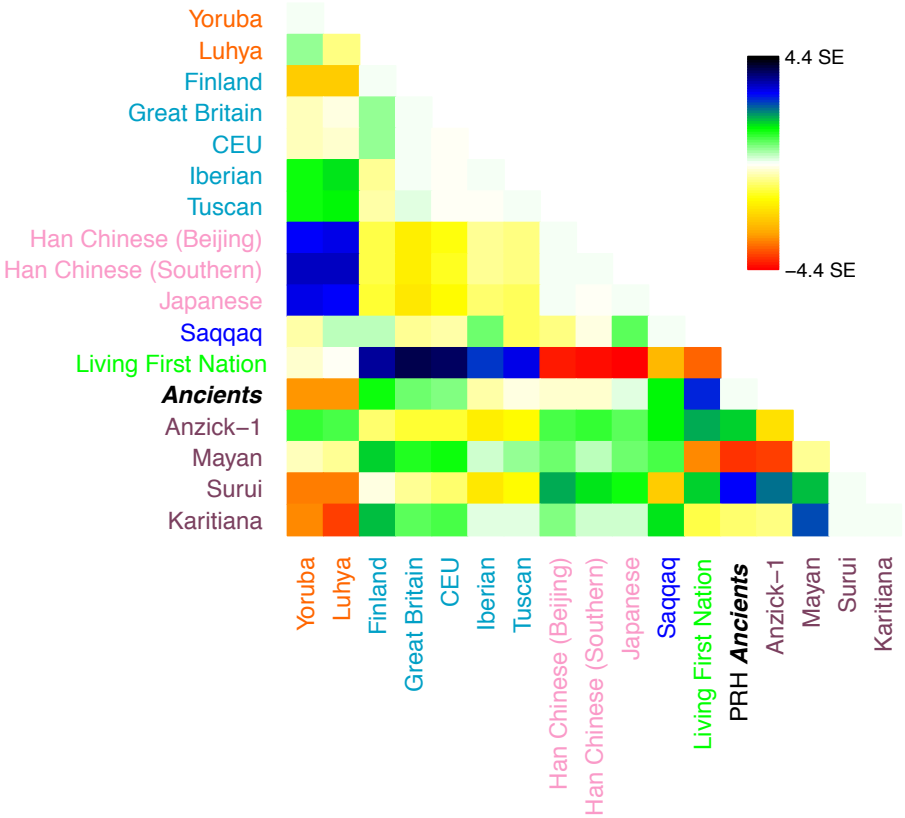

Figure 18. ***TREEMIX*** **residual with a single migration event.**

Selection Scans

To detect regions under selection in both the ancient and the living, the Population Branch Statistic was utilized (PBS; (Yi et al., 2010)). This statistic has proven powerful in detecting hypoxia adaptation in high altitude populations(Yi et al., 2010; Huerta-Sanchez et al., 2013). This statistic uses three populations with an evolutionary relationship depicted in Figure 19. The 1,000 Genomes Beijing Han population (CHB) is assumed to be the outgroup to the living and the ancient.

Under a scenario where the evolutionary forces are a factor of genetic drift, increased genetic similarity between the ancient and the living is expected, as compared to the ancient and the Han or the living and the Han. However, if local adaptation occurred to the differing ecological landscapes (*e.g.* pathogens) of pre and post colonial environments of the Americas, then it is expected that loci will be highly diverged between the ancient and the living.

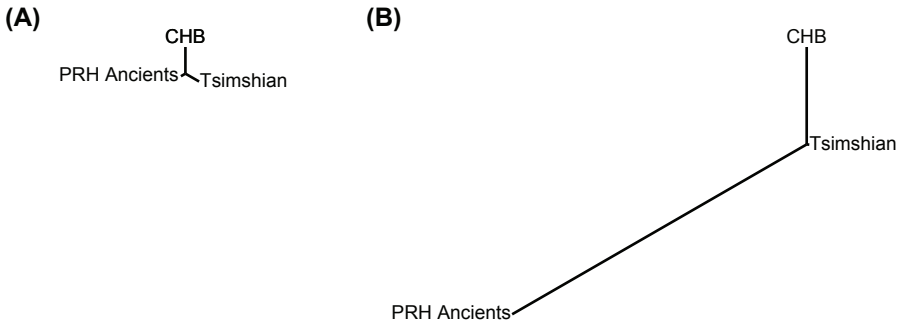

Figure 19. Selection signal from the *HLA-DQA1* gene. (A) The branch lengths of the F_{ST}-based genomic average between the three comparative populations. (B) The branch lengths for *HLA-DQA1*, with the branch for the PRH Ancients exhibiting strong differentiation.

Gene annotation were derived from RefSeq (http:// genome.ucsc.edu/), utilizing the longest RefSeq identifier. F_{ST} was calculated as in Reynolds *et al.*(Reynolds et al., 1983).

For the Ancient PBS was calculated as follows:

$$PBS_{Ancients} = \frac{T^{A,L} + T^{A,CHB} - T^{L,CHB}}{2}$$

where $T^{PRH,F} = -\log(1 - F_{ST}^{PRH,F})$ is an estimate of the divergence time between the ancient and the living individuals, as are the other *Times*, respective of their comparative populations. For detecting selection in the living , the following formula was employed:

$$PBS_{Living} = \frac{T^{L,A} + T^{L,CHB} - T^{A,CHB}}{2}$$

PBS was calculated for each gene with the requirement that a focus SNP was accompanied by a F_{ST} calculation based on at least 10 alleles from each population. Transitions were not removed due to their potentially informative value in this analysis. However, DNA damage across the ancient dataset is evident at very low frequency and does not affect more than 3 individuals at any one particular site. Therefore, the requirement for 10 alleles will prevent DNA damage from entering the selection calculus. Both variants and monomorphic sites were used.

Since the *Treemix* simulations (Figure 11b) indicate a likely admixture event between the living and Europeans, the Living were corrected for admixture with the method described in Huerta-Sa´nchez *et al.* (Huerta-Sanchez et al., 2013). The correction functions to limits the rate of false positive selection signals from admixture by creating a pseudo-unadmixed allele frequency for the Living.

HLA-DQA1 SNP Confirmation

The *HLA-DQA1* SNP showing the highest frequency changes (located at positions chr6: 32605189 and chr6: 32605197) between the ancient and living were confirmed via Sanger sequencing from 18 of the ancient individuals (Table 10, Fig. 20). The remaining 7 ancient samples either did not amplify in the specified region or the sequence was not readable. The extraction method was the same as described above. Forward and reverse PCR primers were constructed as follows: CCTCACAATTACTCTACAGCTCAG and CTCATGCACTCACCCACAA.

PCR reactions were performed with 25ul of Q5 High-Fidelity 2X master mix (New England Biolabs, Ipswich, MA), 0.2μM of each PCR primer, 3% DMSO (New England Biolabs, Ipswich, MA), and 0.2mg/ml BSA (New England Biolabs, Ipswich, MA). PCR conditions were as follows: 30s at 98°C, 50 cycles of 10s at 98°C, 30s at 60°C and 30s at 72°C, with a final extension at 72°C for 2m.

Gene Segment	Chr	Position	Ancestral Allele	Derived Allele	Function	PRH Ancient Frequency	Living Frequency	CHB Frequency
UTR5	6	32605216	G	A	N/A	1	0.43	0.11
exonic	6	32605257	C	A	nonsynonymous	1	0.55	0.51
UTR5	6	32605197	G	A	N/A	1	0.41	0.16
UTR5	6	32605189	A	G	N/A	1	0.37	0.46
exonic	6	32605271	C	T	synonymous	1	0.42	0.12
exonic	6	32609173	G	C	synonymous	0.89	0.67	0.79
exonic	6	32609181	T	C	synonymous	0.84	1	1
exonic	6	32609192	A	G	nonsynonymous	0.77	0.27	0.14
exonic	6	32609172	G	A	synonymous	0.11	0	0
exonic	6	32605284	G	A	nonsynonymous	0.10	0.05	0.12
exonic	6	32609161	G	A	nonsynonymous	0.06	0	0
exonic	6	32605264	G	A	nonsynonymous	0.04	0.05	0
exonic	6	32605266	G	A	nonsynonymous	0	0.04	0.06
exonic	6	32609199	G	A	synonymous	0	0	0.02
UTR5	6	32605207	G	C	N/A	0	0.55	0.49
exonic	6	32609195	G	A	nonsynonymous	0	0	0.28

Table 10. Population Frequencies for the HLA-DQa1 SNPs.

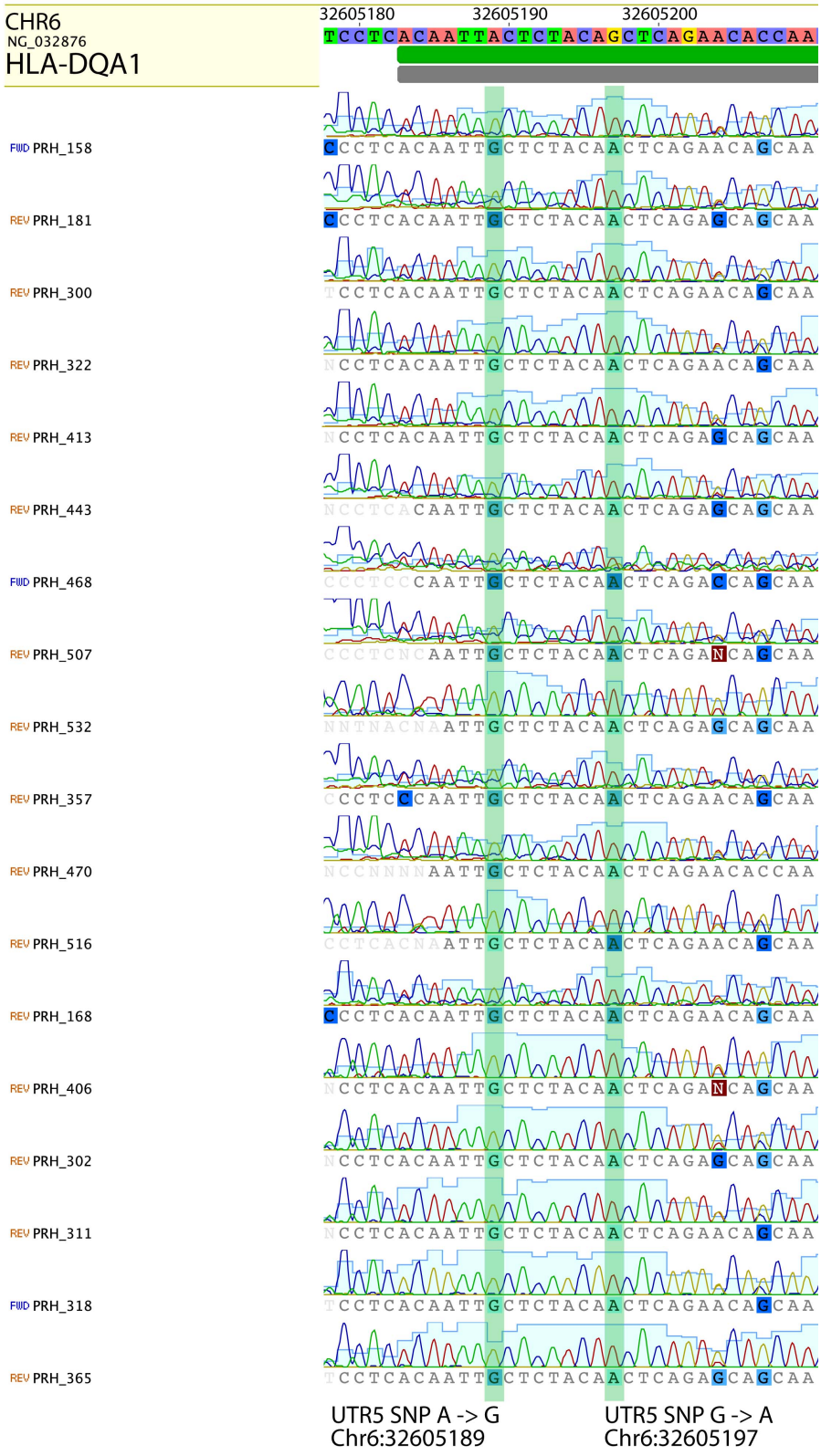

Figure 20. Confirmation of HLA-DQA1 high frequency SNPs via Sanger sequencing in 18 of the ancient individuals.

Demographic Model

Parameters for the demographic model (Fig. 21) were inferred with FastSimCoal2 (Excoffier et al., 2013). 100 optimizations were run for the inferred values. The optimizations utilized a multi-dimensional SFS containing the CHB, ancient, and living samples. The GBR served as a ghost population in the model. This SFS contained 7.4 Mb monomorphic and polymorphic sites based on hg19 potential synonymous sites, where data was reported for each individual. The mutation rate was set to $2x10^{-8}$. A parametric bootstrapping approach was used to construct the 95% confidence intervals. The approach called for 40 sets of simulated data, using the parameter file with the best likelihood from the 100 optimizations. The data was simulated with an effective sequence length of 7.4 Mb and a mutation rate of $2x10^{-8}$. Each of the simulated data sets were put through 100 optimizations, taking the best likelihood parameters from each of the 40 sets of runs. The inferred parameters and confidence intervals are listed in Table 11.

Parameter	Inferred Value	95% CI
$N_{A\_Split\_CHB\_Ancients}$	242 (Diploid Individuals)	156-767
$N_{Ancients}$	11,320 (Diploid Individuals)	6,452-14,679
N_{Living}	2,195 (Diploid Individuals)	1,372-7,035
$T_{Split\_CHB\_Ancients}$	14,800 years	13,925-16,750
$T_{Bottleneck\_European\_Contact}$	150 years	125-225

Table 11. Parameter estimated for the model displayed in Figure 21.

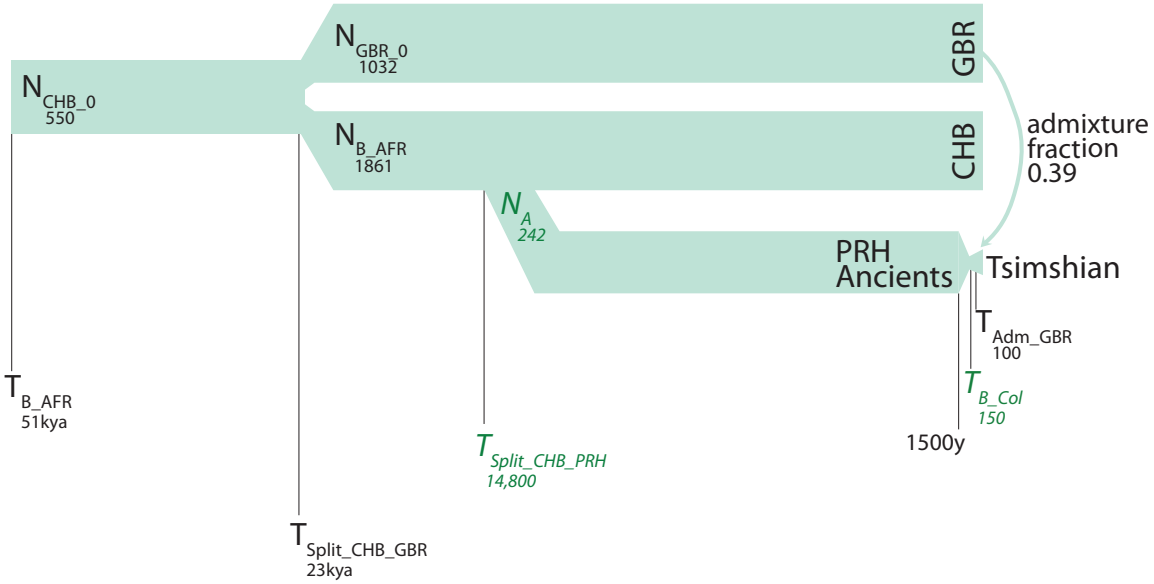

Figure 21. Demographic Model. Best fitting model depicting the Living bottleneck after European colonization and the subsequent admixture with Europeans. Fixed demographic parameters are from (Gravel et al., 2011)and admixture parameters from (Verdu et al., 2014). The inferred parameters are in green.

CHAPTER 4

PATTERNS OF CODING VARIATION BETWEEN ANCIENT AND LIVING NORTH AMERICAN INDIGENOUS PEOPLES REVEAL UNEXPECTED EFFECTS OF EUROPEAN COLONIZATION

Abstract

Standard theoretical models suggest that the effects of European colonization on the genomes of Native Americans should have produced excesses of potentially deleterious features, mainly due to the severe reductions in population size and the associated loss of genetic diversity. These models do not, however, take into account actual genomic patterns that existed before colonization; nor do they adequately capture the effects of non-native admixture with indigenous populations. In this study, we analyze the whole-exome sequences of a Northwest Coast Living with a demographic history characteristic of other indigenous populations from the Americas. We compare this with the exomes of their presumed ancient ancestors. Contrary to standard theoretical expectations, we show that in approximately ten generations from initial European contact, the living indigenous individuals show reduced levels of private and rare variants, reduced levels of potentially deleterious alleles, and increased levels of heterozygosity, when compared to their ancestors. This pattern, which is unexpected after such a dramatic population decline, can be explained by certain mitigating factors, including the increased genetic diversity stemming from admixture with non-native groups. This study examines the genomic consequences of colonialism on an indigenous population and describes the role of gene flow in rapidly mitigating deleterious features.

Introduction

The indigenous populations of the Americas suffered extensive population declines caused by modern European contact. Although the precise extent of this decline is contested (Thornton et al., 1991; O'Fallon and Fehren-Schmitz, 2011), and likely varied with local circumstance (Patterson and Runge, 2002), these events should have affected the genetic variation within surviving indigenous populations. These effects are, however, further complicated by patterns of

gene flow from non-native immigrant groups. Although previous studies have explored the genetic diversity of contemporary Native American populations (Wang et al., 2007; Gravel et al., 2011; Hunley and Healy, 2011; Reich et al., 2012; Verdu et al., 2014), the effects of colonization have not been examined with the aid of studies concerning *ancient* Native American genetic diversity. Here we examine the effects of colonization by comparing the genomic patterns of a single indigenous population from two different time frames: before and after European contact.

Broadly speaking, genetic variation among human populations can result from many different sources, including random processes (e.g., genetic drift), natural selection, and differences in demographic histories (Cavalli-Sforza and Feldman, 2003). With the advent of cost-effective, whole-exome sequencing, it has become easier to study genome-wide patterns in many individuals simultaneously. Statistical analysis of large data sets can be used to reconstruct key events in human evolutionary history. These include events such as the initial and widely-accepted Out of Africa dispersal, as well as numerous founder effects and population expansions that subsequently occurred as early humans spread throughout the globe (Li et al., 2008; Lohmueller et al., 2008; Coventry et al., 2010). This has resulted in the identification of demographic events discernable from the allele frequency spectra in various global populations (Keinan and Clark, 2012; Tennessen et al., 2012; Casals et al., 2013). For example, the accumulation of rare variants within a single modern population has been interpreted as evidence of a past demographic bottleneck coupled with subsequent population growth (Gravel et al., 2011; Nelson et al., 2012; Casals et al., 2013). With this type of demographic process, the variants in question tend to reflect increased putatively deleterious alleles in the coding regions of the genome (Tennessen et al., 2012). The accumulation of potentially deleterious alleles can be greatly amplified if the evolutionary time scale of the demographic event that caused them is relatively short and if purifying selection has had insufficient time to act on them (Casals et al., 2013).

Some facts about population history can also be inferred from individual genomic data. For example, it is now common to study runs of homozygosity (ROH), which are defined as long stretches of homozygous genotypes in an individual genome. The presence of these homozygous regions can provide information on aspects of population history. This is because the *length* of

ROH can be used to infer the relative age of these segments and the processes that likely generated them (Pemberton et al., 2012; Szpiech et al., 2013). When combined with other genomic patterns, such as site frequency spectra, ROH studies provide additional, powerful tools for understanding the evolutionary histories of populations.

Inferring the demographic history of indigenous populations in the Americas has proven difficult due in part to its multi-faceted nature—which, we now know, involved at least a combination of bottlenecks, founder effects, population growth, and recent non-indigenous admixture. The short evolutionary timescale of the effects caused by European colonization further complicate the picture since most of the population level genomic patterns previously identified involve much longer periods of time, spanning thousands of years instead of hundreds (Consortium et al., 2010; Gravel et al., 2011; Creanza et al., 2015). Many of the statistical methods used to identify these patterns are, accordingly, best suited to identify demographic patterns that emerge over longer periods of evolutionary time. Hence, while early Native American migrations have been thoroughly explored (Reich et al., 2012; Gravel et al., 2013), as have the admixture effects of colonization (Wall et al., 2011; Moreno-Estrada et al., 2013; Verdu et al., 2014), these recent admixture events have not yet been studied with the aid of comprehensive data concerning patterns of *ancient* genomic diversity in these populations. As shown below, this type of analysis reveals novel features of the genomic patterning that arose from this most recent chapter of Native American history.

The genomic patterns of the Northwest Coast Living population before and after European contact

The data in this study offer a unique opportunity to compare the genomic patterns of an ancient indigenous population from the Americas (i.e., before any effects of European colonization) with the genomic patterns of their living descendants. Through this comparison, we can study the genomic evolution of a single population through time. We can identify the changes in genomic patterns that have resulted from multiple demographic processes—including some that took place over a relatively short period of time in more recent history.

To that end, we analyze the whole-exome sequence data from the Northwest Coast Living individuals and compare various population level statistics and genomic patterns to those of their

presumed ancient ancestors—dating to before European contact. This analysis offers an enriched understanding of the genomic impact of the special demographic histories experienced by the indigenous populations of the Americas. Speaking in broadest terms, and as shown in Figure 1, this demographic history includes a founder effect during the initial peopling of the Americas, a long period of relative isolation of the ancestral population from other global populations, a population expansion during this period of relative isolation, a dramatic population bottleneck during European colonization, and some recent admixture with non-native groups (Mulligan et al., 2004).

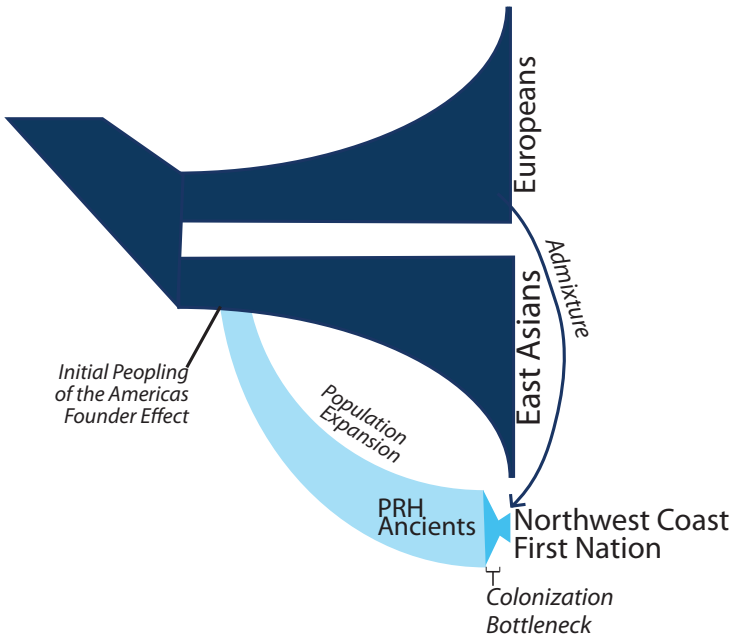

Figure 22. Illustration of the demographic history of an indigenous population from the Americas.

The predominant prediction, with respect to a population that has undergone a recent and severe bottleneck, would be to find a reduction in overall fitness between modern and ancestral populations, with some evidence of this at both the individual and population level (Charlesworth and Charlesworth, 2003; Grueber et al., 2008; Charlesworth, 2009). This is because population bottlenecks typically lead to a decrease in heterozygosity due to reductions in effective population size (Fox et al., 2008; Bouzat, 2010). Another potential outcome of a population bottleneck is an increase in stretches of homozygous genotypes, known as runs of homozygosity (Gibson et al., 2006; Pemberton et al., 2012). The frequency of homozygous

segments on an individual level is typically positively correlated with consanguinity (and hence with reductions in population size of a single, related population) but can also arise from other features of demographic history. For example, population bottlenecks, small effective population size, geographical isolation, the degree of endogamy (within population marriage), have all been theorized to effect observed genome-wide patterns of homozygous segments (Woods et al., 2006; Kirin et al., 2010; Humphreys et al., 2011).

Ancient Sample	Site	Burial	Conv ^{14}C Years BP	Standard Error	Calendar Years BP	Dated via Associated Cultural Remains
125	Garden	XVII-B-125	2260	40	2288±58	
158	Garden	XVII-B-158	2990	50	3179±83	
163	Dodge	XVII-B-163	n/a	n/a	n/a	"Pre-contact"
167	Garden	XVII-B-167	n/a	n/a	n/a	cal 800 to 350 BC
168	Dodge	XVII-B-168	2650	75	2760±60	
181	Garden	XVII-B-181	2620	40	2743±20	
300	Parizeau	XVII-B-300	1650	75	1544±88	
302	Parizeau	XVII-B-302	2440	75	2498±142	
311	Boardwalk	XVII-B-311	2090	60	2058±87	
318	Boardwalk	XVII-B-318	1550	50	1432±68	
322	Boardwalk	XVII-B-322	2050	50	2003±65	
357	Boardwalk	XVII-B-357	n/a	n/a	n/a	cal 50 BC to 690 AD
365	Boardwalk	XVII-B-365	2270	65	2290±75	
386	Boardwalk	XVII-B-386	1060	40	962±33	
406	Boardwalk	XVII-B-406	n/a	n/a	n/a	cal 50 BC to 690 AD
412	Boardwalk	XVII-B-412	1940	40	1883±40	
413	Garden	XVII-B-413	1970	40	1913±42	
443	Boardwalk	XVII-B-443	1820	55	1750±70	
468	Lachane	XVII-B-468	1940	45	1883±46	
470	Lachane	XVII-B-470	1600	40	1488±52	
507	Baldwin	XVII-B-507	2320	60	2335±61	
516	Baldwin	XVII-B-516	n/a	n/a	n/a	cal 300 BC to 440 AD
525	Boardwalk	XVII-B-525	1860	40	1800±51	
532	Baldwin	XVII-B-532	n/a	n/a	n/a	3200 to 1500 BP
939	Lucy	XVII-B-939	5710	40	6487±50	

Table 12: Archaeological context of the ancient individuals from PRH and Lucy Island.

Results

Through the use of exome sequencing, we set out to investigate how the patterns of genetic variation in a Native American population have been altered before and after European contact. Using previously described data (Lindo et al. 2015), we detected 59,858 high-confidence SNPs from 25 living individuals from the Northwest Coast First Nation ("living") and 60,973 from 25 of the PRH and Lucy Island ancient individuals ("ancient"), derived from 62 Mb of targeted regions. Due to the uncertainty of low frequency transitions in the ancient individuals that may be a result of DNA damage, transition frequencies below 0.10 were removed in both data sets (see *Methods*).

Compared to the ancient individuals, the living have higher levels of heterozygosity within coding regions (living 30% vs. 27% for the ancient, on average) (Table 13). This observation is *inconsistent* with expectations for a population that has experienced a recent and dramatic intervening bottleneck. We also extended the heterozygosity measure to calculate changes in effective population size through time, which included the 10,300 year old OYKC sample described in chapter 2 (Fig. 23). We see a significant correlation between increasing effective population size and increasing time BP (Kendall's τ, $P< 0.0001$). The population exhibits a general trend of population decline before European contact, which has been observed in other studies of indigenous populations of the Americas (Mulligan et al., 2008). There is also variation in the ancients within particular archaeological sites, which may suggest local demographic factors at play. The living individuals similarly show variation in effective population size, which may correlate to varying levels of gene flow from non-native groups.

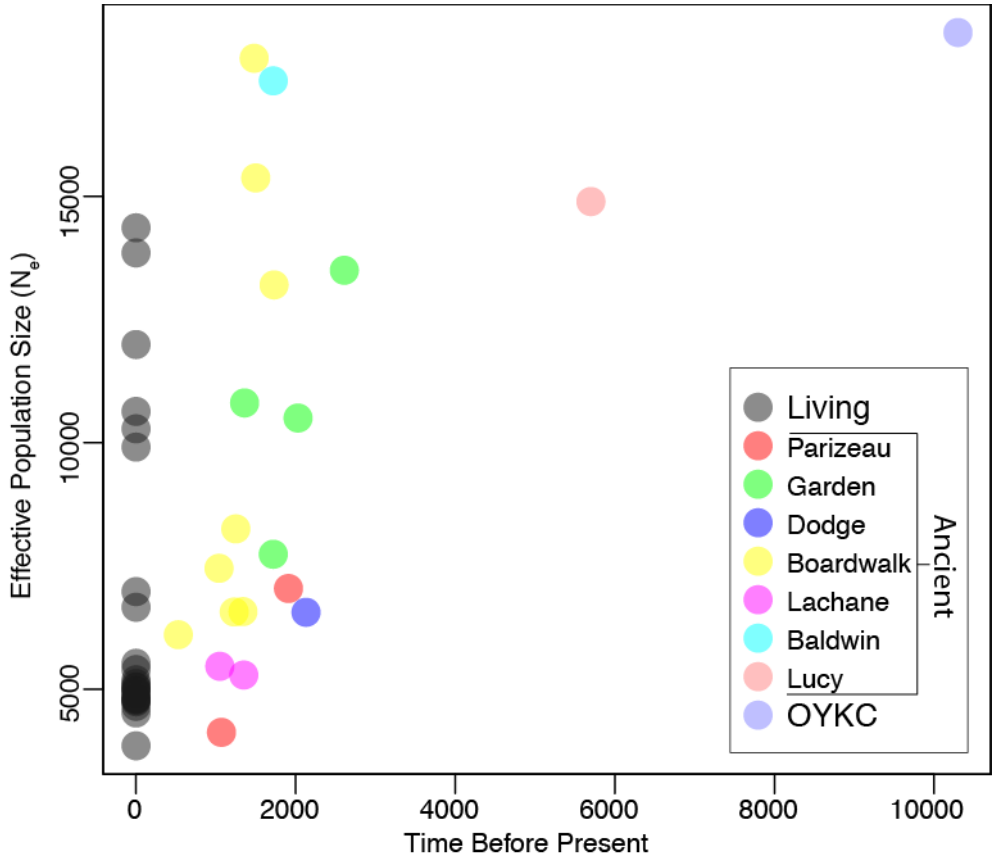

Figure 23. Effective population size through time. The ancient individuals are color-coded according to their burial sites. The OYKC is an ancient individual from present day Alaska, which dates back to approximately 10,300 calendar years BP (*see* Chapter 2).

Next, we examined the derived frequency spectrum (Fig. 24). The ancient individuals show a significant increase over the living in variants below a frequency of 0.2. This result is likely due to a population expansion after the founder effect from the initial peopling of the Americas. Further evidence for this early expansion, which is consistent with the archaeological evidence, arises from the fact that the ancient individuals exhibit lower Tajima *D* values than the living Living (z-test P = 1.73e-6; Table 13). In comparison, the living allelic distribution suggests the potential *loss* of low-frequency variants.

	Individuals	Total SNPs	Coding SNPs	Functional SNPs	Average Depth	Het	Tajima's D	Ts/Tv
Living	25	59,858	28,727	11,430	8.79	0.3	1.10	3.259
Ancient	25	60,973*	31,205*	16,071	6.95	0.27	0.77	3.833
Shared		*40,069*	*19,088*	*8,061*				

Table 13. Genetic measures for the 62 Mb targeted regions for the living and ancient individuals. Heterozygosity was measured per individual at variants sites. Tajima's *D* was measured as an average across coding sites, with a minimum of 5 variants. Depth was calculated per individual using coding sites. Ts/Tv designates the transition/transversion ratio within coding regions; a ratio above 3 is expected (Fu et al., 2012). *These totals are with the removal of transitions below a frequency of 0.1, which are potentially due to DNA damage.

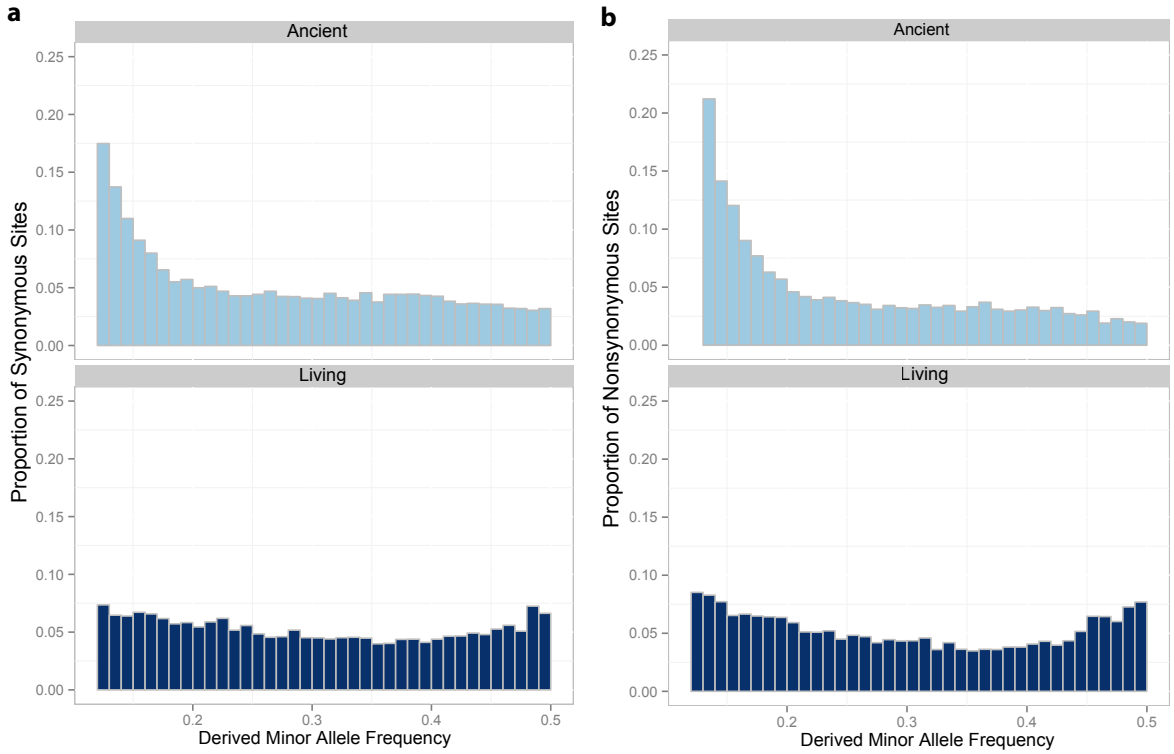

Figure 24. Derived allele site frequency spectra. Spectra depict the synonymous (a) and nonsynonymous (b) derived allele frequencies of the ancient and the living individuals.

Of the combined SNPs, 49% are shared between the ancient and living. This percentage drops to 41%, however, when comparing functional SNPs (viz., nonsense, missense, and splice sites) (Table 14), which are enriched for alleles with frequencies of less than 20%. This genetic feature is especially pronounced in the ancient individuals with respect to the living. Despite disparate features between the two sampling time frames (*i.e.*, before and after European contact), a look at the F_{ST} scores corresponding to the 40,069 shared variants between the ancient and contemporary individuals remains low (F_{ST}=0.027), which is indicative of low genetic differentiation.

	Total SNPs	Functional SNPs	Private SNPs	Private Possibly and Probably Damaging
Living	59,858	11,430	1,032	175
Ancient	60,973	16,071	13,539	3,514
Shared	40,069	8,061	242	40

Table 14. Shared Variants between the ancient and living individuals for the 62 Mb targeted sites. Private SNPs were calculated by filtering SNPs from dbSNP 142 and the 1,000 Genomes variants from the October 2014 release. Damaging sites were calculated with PolyPhen2.

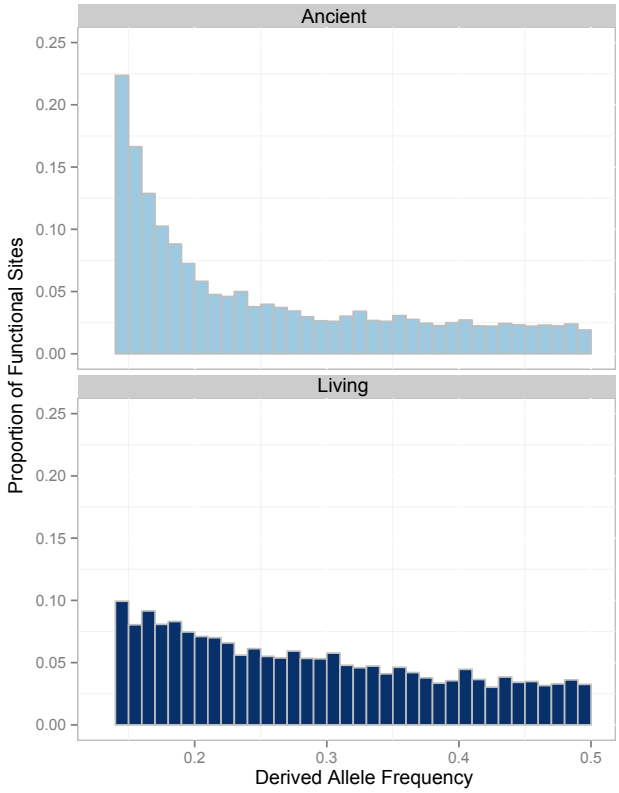

Figure 25. Derived allele frequencies at functional sites. Comparison of the derived allele frequency of functional sites (falling within the exon) between the ancient and living individuals. Both sets of individuals had transitions below 0.1 removed to safeguard against bias from damaged sites in the ancient.

When examining the derived frequency of functional SNPs, the ancient PRH individuals exhibit an excess of variants below a frequency of 0.2 (Fig. 25), as compared to the living. Because these differences are considerable, we discuss the effects of these variants. We do this, first, by testing for differences in the ratio of missense to synonymous changes within the derived site frequency spectrum (SFS) (Fig. 24). The missense to synonymous ratio of 1.92 in the ancient individuals, for SNPs with a frequency below 0.1, points to a major fraction of potentially deleterious SNPs. The same ratio in the living is 1.17 (z-test P <.0002), thus suggesting a significantly larger proportion of rare mutations at missense sites in the ancient individuals. This difference between the two sampling time frames is pronounced, especially when compared to previous findings from other global populations (~1.37 - 1.49), as well as theoretical expectations (~2.2) (Kryukov et al., 2007; Casals et al., 2013). For the more common variants, with frequencies greater than 0.2, the ancient and living individuals exhibit, however, similar missense to synonymous ratios to each other.

Second, because previous studies predict that rare nonsynonymous variants tend to be deleterious (Bodmer and Bonilla, 2008; Coventry et al., 2010; Tennessen et al., 2012), we consider whether this difference between the missense/synonymous ratios correlate to a shift in the burden of potentially damaging alleles (Casals et al., 2013). To assess this possibility, we examine the predicted effects of nonsynonymous variants using Genomic Evolutionary Rate Profiling (GERP) (Cooper et al., 2005). GERP is a conservation measure, calculated across 34 mammalian species, and can be used to annotate potentially damaging effects of genetic variants inferred through the level of constraint. Neutral sites tend to score near zero, whereas constrained sites generally score positively. Using this measure, the ancient individuals demonstrate evidence for an excess of potentially damaging mutations below a frequency of 20%, when compared to the living (Fig. 26).

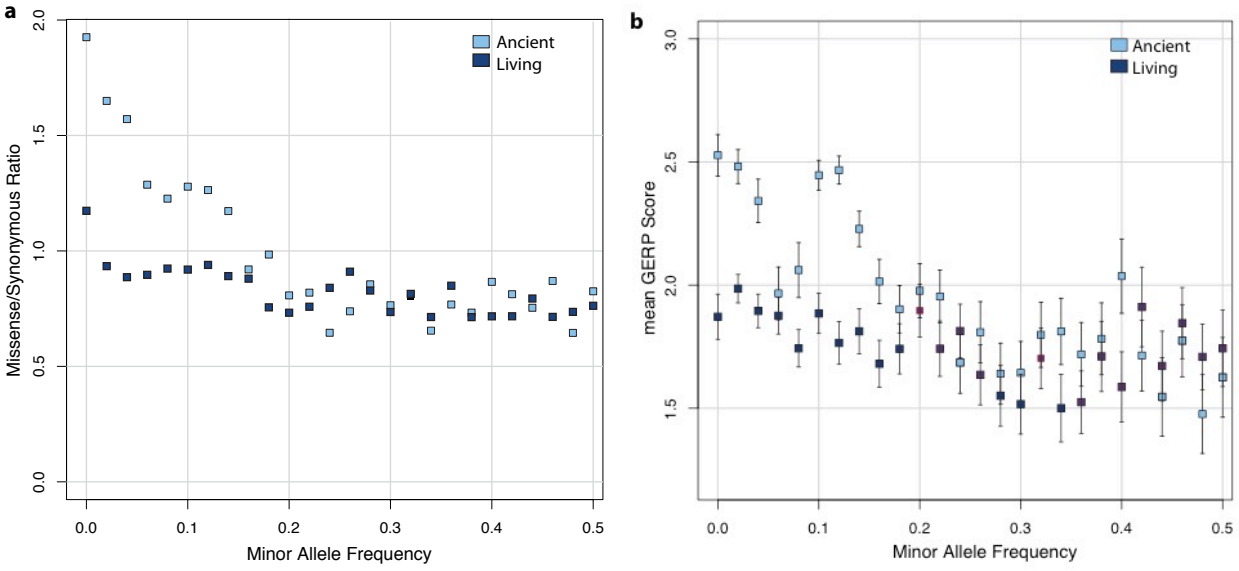

Figure 26. Excess of functional variants in the ancient individuals. a) Ratio of nonsynonymous to synonymous changes in the ancients and living individuals for variants grouped by minor allele frequency. b) Average GERP score of the functional changes for each frequency class in the ancient and living individuals.

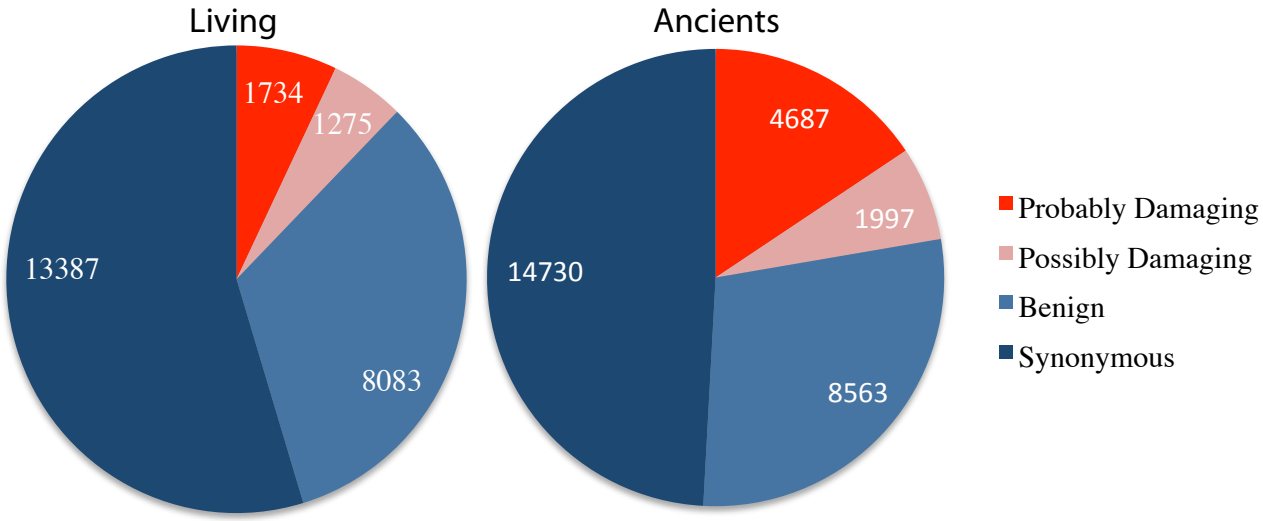

Figure 27. PolyPhen2 Classification of high-confidence SNPs in each population. PolyPhen2 predicts possible impact of an amino acid substitution on the structure and function of a human protein. The living individuals show a reduced proportion of probable and possible damaging sites, when compared to the ancient.

Fourth, we examine the number of private variants in both the ancient and living individuals. We define private variants as those that are not found in the phase 3 release of the 1,000 Genomes Project (www.1000genomes.org) or dbSNP 142 (http://www.ncbi.nlm.nih.gov/SNP/). The ancient individuals show an excess of private alleles, when compared to the living (Fig. 28a). This is likely due in part to the overall loss of genetic variation caused by the population bottleneck associated with European colonization. The living individuals show, however, a

significant decrease in the portion of possibly or probably damaging sites within these private alleles, when compared to the ancient (Fig. 28b; living 0.17, ancient 0.26, z-test $P < .0001$). Previous studies have found that rare and private alleles tend to be of a deleterious nature and may have arisen due to population expansions following bottlenecks or founder effects (Nelson et al., 2012; Tennessen et al., 2012; Torkamani et al., 2012). It is thus possible that the relative stability in population size, since the Living's initial recovery from the colonization bottleneck, slowed the introduction of rare and potentially damaging alleles into the population.

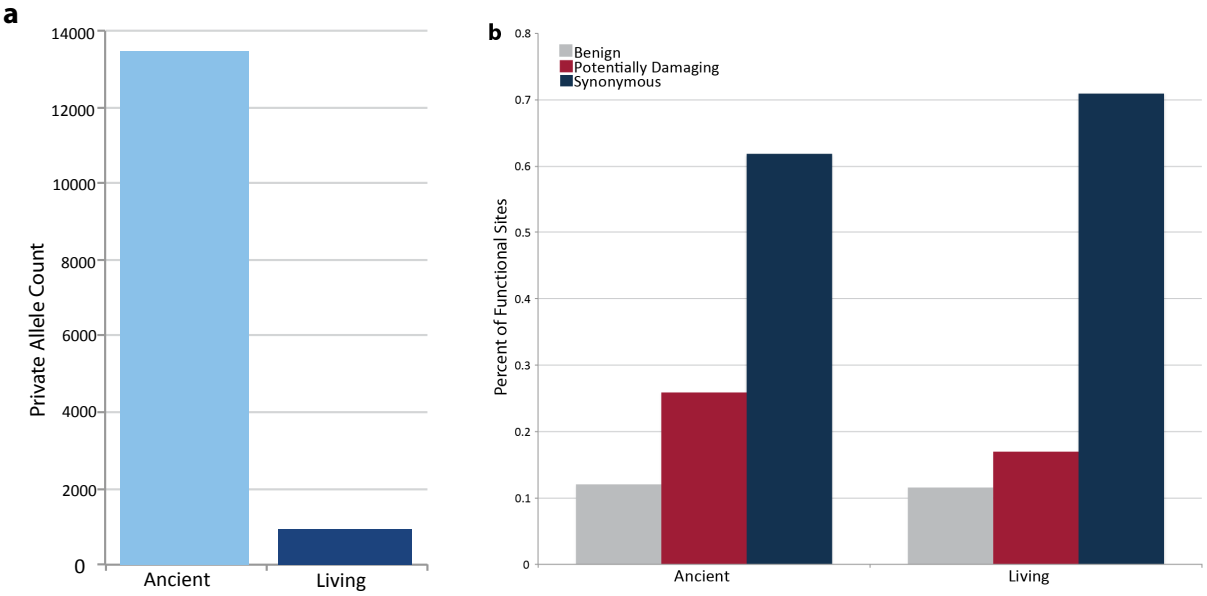

Figure 28. Fraction of private variants that are predicted to be damaging by PolyPhen2.
All transitions below a frequency of 0.1 were removed to guard against bias from low frequency DNA damage. a. Private variants are defined as variants that appears only in a single population, with respect to the reference. The reduced levels of private alleles in the living vs. the ancient may represent the loss of allelic diversity after the population bottleneck associated with European colonization. b. The private missense variants are split into PolyPhen2 predictions of benign and damaging (either possibly or probable). The ancient individuals show a large fraction of potentially damaging alleles when compared to the living.

Finally, we examine comparative data concerning runs of homozygosity (i.e., ROH). Previous work has categorized ROH in many ways, ranging from lengths in base pairs to the minimal number of markers needed to describe a homozygous region (Gibson et al., 2006; Auton et al., 2009; Howrigan et al., 2011; Pemberton et al., 2012). For this study, we followed the model set by Pemberton et al. 2012 (Pemberton et al., 2012) and used a mixture of three normal distributions representing three classes of ROH: class A (< 500 kb), class B (500-1500kb), and class C(>1500kb). Each of these classes has their own distinct demographic and cultural

possibilities as to how they arose within an individual's genome (Table 15). The boundary differences between the ancient individuals are shown in Figure 29a. The method for calculating ROH from each exome took into account the special features of ancient DNA so as to prevent the loss of ROH detection from spurious heterozygous sites caused by DNA damage (*see* Methods).

ROH Class	Mean Size (Kb)		Percentage of Total ROH		Demographic and Cultural Associations
A	Living	164	Living	61%	Ancient. May be the remnants of linkage disequilibrium (LD).
	Ancient	174	Ancient	73%	
B	Living	776	Living	13%	Intermediate formation. May also be the result of LD but on a younger evolutionary scale.
	Ancient	759	Ancient	13%	
C	Living	2,126	Living	26%	Formed recently. May represent endogamous marriages and/or major shifts in demography.
	Ancient	1,985	Ancient	14%	

Table 15. Class sizes and evolutionary significance.

On assessing the ROH genomic patterns between the ancient and living individuals, the first item of significance is an increase in the percentage, with respect to the total ROH, of class C in the living (26%), when contrasted with the ancient (14%) (z-test $P < .0001$). This may be a consequence of increased identity-by-descent inheritance of identical segments and subsequently decreased variation within these segments. This would be evident in much longer segments (i.e., class C ROH), which are younger and have had insufficient time to be broken up by recombination. The distribution of class A segments is also significantly different between the ancient (73%) and the living (61%) (z-test $P < .001$). This difference may be due to the recent bottleneck, which may have resulted in the loss of small homozygous segments. This is also reflected in the density distribution of smaller homozygous fragments, which shows an increase in the ancient individuals (Figure 29b).

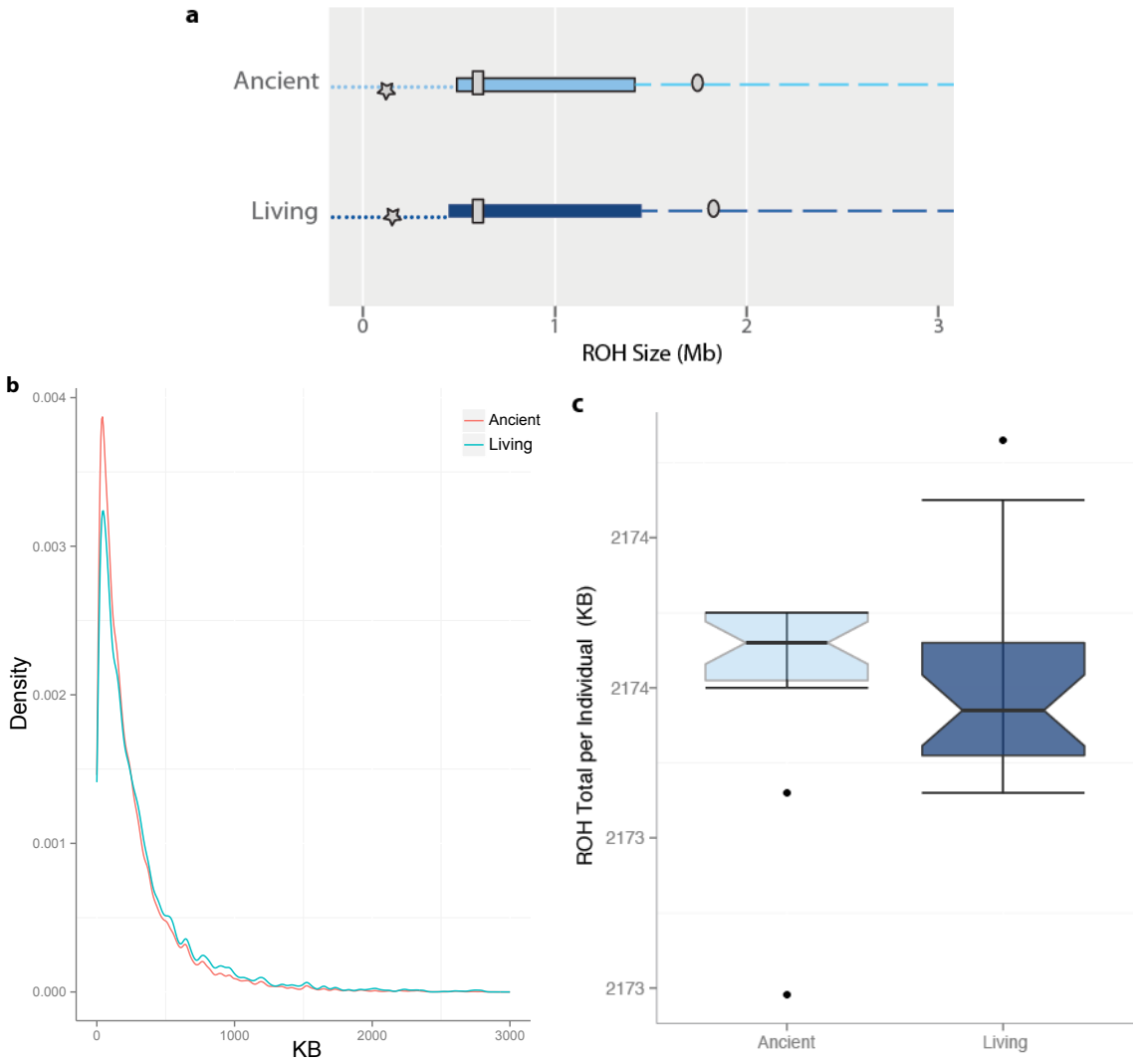

Figure 29. Genomic patterns of homozygosis regions between the ancient and living individuals. a. Mean size of ROH classes and boundaries between classes. Dotted lines and stars indicate the limit and mean size of class A, respectively. Solid lines and squares indicate the limit and mean of class B, respectively. Dashed lines and ellipses describe the limit and mean of class C, respectively. b. Gaussian kernel density estimates of the ROH size distribution in the ancient and living individuals. c. Distribution of total ROH lengths over all individuals in each population, for all three classes combined.

Turning to individual level data, the genome percentage covered by ROH is on average higher in the ancient than in the living (*t*-test, $p < 0.004$) (Fig. 30). This difference in distributions is also reflected in the total ROH mean per individual, where the living have significantly lower mean totals, with respect to the ancient (Fig. 29c). These genomic patterns may be the result of two recent demographic factors. First, the post-colonization bottleneck may have caused some loss of ancient homozygous fragments within the population as a whole, as is reflected in the lower percentage of class A ROH in the living individuals. The second factor may involve the shared

ancestry between the living and European immigrant communities, occurring after contact (Verdu et al., 2014). The resulting gene flow may have decreased the distribution of ROH in the living by increasing the heterozygosity between genetic segments as a consequence of the increased genetic variation.

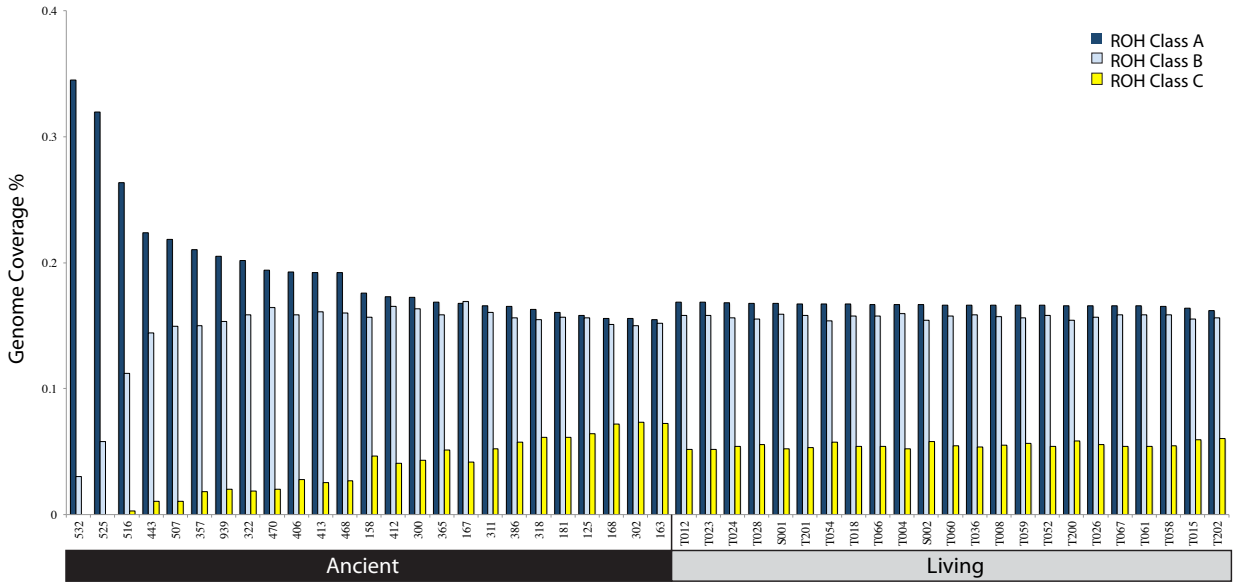

Figure 30. Individual runs of homozygosity. Percentage of individual genomes covered by different ROH classes.

When comparing the number of potentially deleterious alleles within runs of homozygosity that are either probably or possibly damaging according to PolyPhen2 classification, we see a comparative surplus in the ancient individuals over the living individuals within each class (Fig. 32b). For class A ROH, we find that for the smaller and most ancient segments, the proportion of damaging sites is significantly higher in the ancient individuals (ancient, 0.38; living 0.12; z-test P <.0002). Despite having similar run totals for class B (Fig. 32a), the living individuals show a significant decrease in deleterious sites (ancient, 0.21; living 0.12; z-test P <.0002). Turning to the largest and most recent runs of homozygosity, class C, we find a similar pattern. Here the living individuals exhibit a lower proportion of damaging sites than the ancient (ancient, 0.22; living 0.13; z-test P <.0001), despite having little difference in run totals.

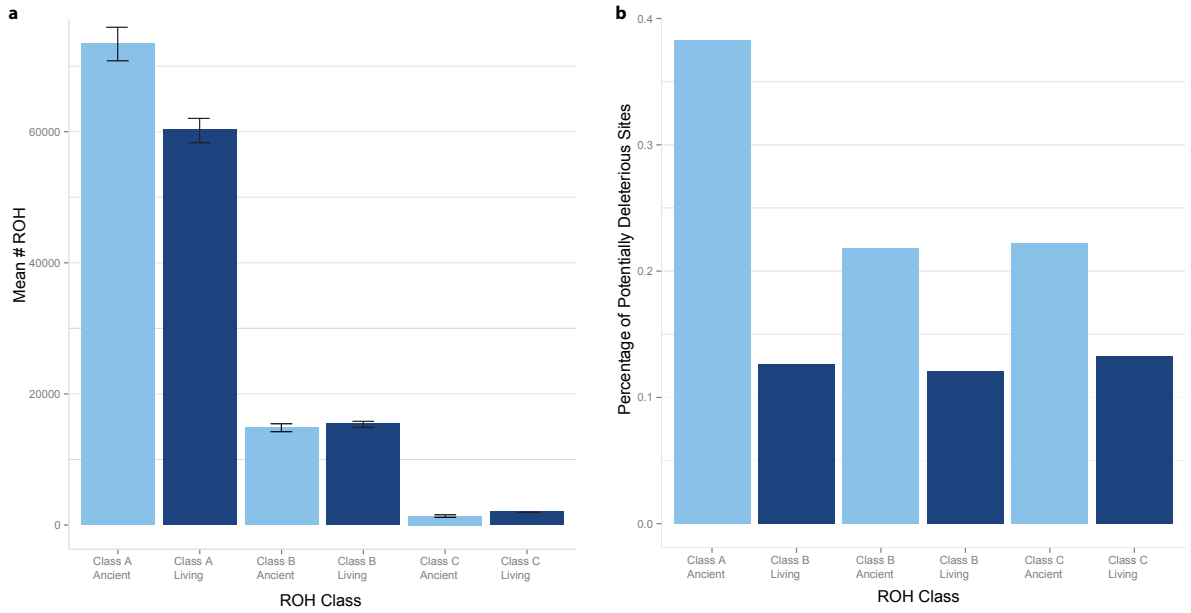

Figure 31. Potentially damaging sites per ROH class. a) Mean number of ROH in the ancient and living individuals. b) Percent of all functional sites that may be potentially damaging within each class of ROH in the ancient and living individuals, as predicted by PolyPhen2, combining both probable and potential. The living show reduced levels of potentially damaging sites within each ROH class, when compared to the ancient.

Discussion

The effects of European colonization has altered the genomes of Native Americans in multiple and dynamics ways. The data discussed in this study suggests that within approximately 9 generations since the time of European contact, the living have significantly less rare and potentially damaging alleles than their ancient ancestors. The differences between the sampling time frames can be partially explained by the population expansion that increased the number of rare alleles in the ancient individuals, following the initial peopling of the Americas. Although the genomic signatures of a human population expansion following a founder effect have been explored in other populations (Cooper et al., 2005; Tennessen et al., 2012; Casals et al., 2013), the genomic consequences of some more recent and severe demographic events seems to carry additional impacts that have thus far been poorly understood.

In particular, while the distribution of allele frequencies in the ancient individuals studied here fit the standard theoretical expectations of a recently expanded population, after an initial founder effect, the distributions relating to living populations do not. Both theory and observation would suggest that after a bottleneck potentially deleterious features would accumulate within the population due to a loss of genetic variability (Nei et al., 1975; Chakraborty and Nei, 1977;

Salmela et al., 2008; Bouzat, 2010). The living have experienced a relatively recent and severe population decline after European contact. This may explain the loss of rare and private alleles when contrasted with the ancient individuals and, in turn, the decrease of potentially deleterious alleles. The effects of such a recent bottleneck, coupled with slow recovery, is, however, typically expected to amplify certain alleles to higher frequency within a population, due to the distortion of allele frequencies by genetic drift (Nei et al., 1975). But these expectations are not born out in our data. We suggest that these differences between theoretical prediction and observation may result from two primary factors: first, the relatively short evolutionary timescale within which these events occurred; and, second, because of recent admixture, which may have increased genetic diversity and countered the stochastic effects of reduced population size (Wright, 1938). Even at very low levels, gene flow from an admixture event has been observed to increase genetic diversity by creating hierarchal higher levels of variation stemming from the connectivity between two populations (Willi et al., 2006; PALSTRA and RUZZANTE, 2008). The Northwest Coast First Nation may have established this increased variation through population connectivity, as evidenced from European admixture (Verdu et al., 2014), and we believe recent and rapid genomic changes like these need further study in a broader range of cases.

Population bottlenecks can also have a different type of effect in terms of the average heterozygosity of a population, which is typically expected to decrease due to the associated loss of alleles (Nei et al., 1975). This decrease is significantly influenced by the rate of population growth after the bottleneck. If the rate of recovery is slow, as in the living, average loss of heterozygosity is more pronounced. The living individuals, however, show a larger than average heterozygosity than their ancestors (Table 13), as well as great variation in their associated effective population size (Fig. 23). This conflicts with the expected heterozygosity recovery time, which should be more pronounced given the size of the bottleneck and its recentness (Chakraborty and Nei, 1977). Once again, this may be a strong indicator that other demographic forces have impacted the living population after the bottleneck, which relates to the population connectivity concept just mentioned. Given the severity of the conflict between standard theory and observation in cases like this, we believe that phenomena like these deserve much greater

attention and study in a broader range of cases relating to the study of human evolutionary history.

A clearer picture arises by examining the effect of a bottleneck on runs of homozygosity on the individual level within a population. ROH can be structured in large chromosomal segments and can occur when an individual's parents share a relatively recent common ancestor, resulting in identity-by-descent (Pemberton et al., 2012). These identical segments, however, can also be the result of a much more distant ancestral relationships or possibly random mating, where the identical segments are inherited from non-common ancestors, or identical-by-state (Lander and Botstein, 1987). Hence, different demographic processes can result in varying sizes of ROH, which can, in turn, be indicative of differing demographic processes and/or cultural practices governing marriage. If a ROH is due to recent endogamous arrangements, the run will, for example, tend to be relatively long since insufficient time has passed for recombination to breakup the identical-by-decent structure. Conversely, shorter ROH are likely due to deeper, if not ancient, ancestral relations where recombination has had more time to break up these runs. It is also possible that some longer stretches of homozygosity persist due to low recombination rates in the chromosomal region where these identical segments exist (Howrigan et al., 2011).

As mentioned above, cultural practices can have a significant effect on how these identical segments are transmitted through arrangements that either link closely related individuals together (*e.g.*, on the cousin level), or restrain the community gene pool to designated members. Demographic processes can also affect the runs of homozygosity within a community through a more broad geographic isolation, which may increase these identical segments within the population by limiting genetic diversity. More dramatic events, such as founder effects and bottlenecks can also increase runs of homozygosity by reducing diversity and increasing identity-by-descent segments that form over time.

The comparison of ROH patterns between the ancient and living is intriguing not only due to varying demographic factors, such as isolation and the colonial-era bottleneck, but also due to the effects of gene flow derived from non-indigenous groups (Verdu et al., 2014). The ROH

patterns taken before and after European contact seem to reflect the overall evolutionary histories of Native Americans through the subtle changes in the associated class distributions.

Apart from having demographic associations, ROH can also have genetic consequences by harboring potentially deleterious alleles, which can become amplified under certain demographic scenarios (Szpiech et al., 2013). Exome sequencing is especially well suited to examine these types of alleles because exome sequencing covers the entirety of the genomic coding regions. Hence, all changes within these coding regions can be examined simultaneously to detect not only changes to the amino acid but also potential harms (viz., by taking into account conservation across species and specific biochemical changes to the protein, as in the above GERP and Phylophen2 analyses) (Fig. 26, 27). These methods can be extended to identify a broader range of potentially damaging alleles within homozygous segments.

The demographic impact of colonization has differentiated the ancient and contemporary genomic patterns, in both the site frequency spectrum and ROH. With respect to deleterious alleles found in ROH, this may be a more complex scenario since deleterious alleles in a homozygous state may unmask its damaging characteristic as opposed to a heterozygous state. For example, a damaging allele in a ROH may correlate to a Mendelian recessive disease or increase the dosage of a damaging gene. Because of these scenarios, it is unlikely that these deleterious alleles will accumulate to any great extent, given the negative viability outcome for the individual. Furthermore, if a particular ROH is stable in a population but deleterious, it is likely to be purged over time by selection (Szpiech et al., 2013). This study suggests that ROH patterns within a population can also be rapidly altered in the presence of admixture and the accompanying increases in genetic diversity.

Therefore, the ROH temporal patterns observed here correlate with the same factors that may have produced the genomic patterns seen in the allele frequency spectra: the recent bottleneck and European admixture. From the general patterns of ROH, we see a strong differentiation of class A segments between the two populations, potentially owing to the loss of these small segments during the living's recent population decline. These sites may have contained potentially damaging alleles that were not passed on to the living. Furthermore, admixture with

European immigrant communities after the bottleneck may have disrupted the transmission of these identical patterns via the introduction of new genetic variation. On the other hand, the living show similar class C segments with respect to the ancient individuals but lack the expected increase in potentially deleterious alleles. This genetic harboring may have been ameliorated by the introgression of alleles from the gene flow with Europeans, thereby reducing transmission of potentially damaging segments via identity-by-descent.

It should also be noted that the term "deleterious" to describe the potential consequence of an allele is problematic when discussing a population as a whole. Here we assess the potential damaging effects of an allele in various ways, including the strength of conservation and the probability that the function protein itself will be altered. These predictions are, however, mainly a tool used by biomedical research to identify a putative disease-causing variant, which are then explored further via other approaches. Given the uncertainty that any allele marked as "deleterious" actually has a disease outcome, we utilized the predictions to examine changes in genomic patterns between two time frames. We do not conclude that ancient Native Americans harbored large reservoirs of deleterious alleles but instead demonstrate a fluctuation of a particular class of alleles through time. The role of novel variants in a population is largely environment-dependent and a negative categorization, especially related to the ancient past, is misleading.

In conclusion, Native American evolutionary history is complex and involves a maelstrom of demographic processes. Some of these complexities are reflected in the genomic patterns observed here in a single Native American population, before and after European contact. Our study suggests that while the ancient individuals exhibit the trademark genomic patterns of a rapid population expansion following a founder effect, their living descendants have genetic patterns that are inconsistent with standard theoretical predictions. Although the effects of the recent bottleneck, associated with European colonization, seems to have removed a large portion of rare and private alleles from the population, this reduction is not accompanied by the same expected increase of potentially deleterious genomic features, such as long stretches of homozygous regions. For reasons discussed, this pattern may represent the ameliorating effects of European allele introgression caused by admixture.

With the use of ancient DNA, we were able to not only uncover the genomic patterns of a population representative of the indigenous peoples of the Americas, but also show how the effects of European colonization have reshaped these patterns in some unexpected ways in more modern times. As human populations continue to intermingle through the effects of globalization, this study highlights the effects of increased gene flow on the genomic patterns between populations with divergent evolutionary histories.

Methods

The exome data presented here, both ancient and contemporary, was generated as described in Lindo et al. 2015. The final variants and genotypes, consisting of 62Mb of captured nucleotides, were called using the ANGSD software suite (http://popgen.dk/angsd). ANGSD is a useful method for calling low to medium coverage sequence data since it does not call variants or genotypes directly, which can be unreliable with lower read depths (Fumagalli et al., 2013). Instead, ANGSD takes into account the genotype uncertainties and sample allele frequencies to determine the most likely call (Nielsen et al., 2011). By doing so, ANGSD allows for the uncertainty associated with the data's coverage to be incorporated into the analysis.

Both genotypes and variants were called by pooling individuals from each population, the Tsimshian and the Ancient s. For all ANGSD procedures, the ancient sequences were trimmed (-trim) by 5 bp at each end, to guard against potential DNA damage. To call genotypes, the following ANGSD parameters were used: -GL 1 -doMaf 2 -doMajorMinor 1 -SNP_pval-doGeno 5 -doPost 1 -postCutoff 0.95, -nIND 10, -minMapQ 30 -minQ 30. To call variants the following parameters were used: -GL 1 -doMajor 1 -SNP_pval 1e-6 -minInd 10 -doMaf 2 -minMapQ 30 -minQ 30. The derived allele frequencies were also calculated with ANGSD using the Chimp genome aligned to hg19. The parameters were as follows: -doSaf 1 -GL 1 -minMapQ 30 -minQ 20 -anc -minInd 10.

In order to prevent spurious results due to DNA damage, for all analyses, the final data sets from both populations were filtered for transitions (G→A, C→T) that fell under a frequency of 0.1. Although this prevented detection of certain rare alleles, we felt that this conservative approach was warranted given the excess of low frequency transitions in the ancient dataset, despite

sequence trimming. This excess is reflected in the Transition/Transversion ratios (Ti/Tv) before and after the low frequency transitions were removed in the ancient data set (9.2 vs. 3.2). The Ti/Tv ratio is further exacerbated when examining the allele frequency spectrum below 0.1 without such filtering (Fig. 32). The same false positive effect can be seen in predicting deleterious alleles (Fig. 32). These spurious transitions were amplified in the dataset due to the calling of all 25 ancient individuals together coupled with the low frequency nature of damaged sites.

To call runs of homozygosity we utilized PLINK (http://pngu.mgh.harvard.edu/~purcell/plink). Although other programs are available specifically for calling ROH in exomes, PLINK provided the flexibility to use genotypes generated by ANGSD, as well as site filtering for low frequency transitions below 0.1. In doing so, we were able to continue using the safeguards against ancient DNA damage, which may have spuriously broken up homozygous segments and prevented ROH detection in the ancient population. The following PLINK parameters were used to call ROH in both populations: --homozyg-density 500 --homozyg-gap 100 --homozyg-kb 10 --homozyg-snp 25 --homozyg-window -het 1 --homozyg-window-missing 10 --homozyg-window-snp 50 --homozyg-window-threshold .05. Although PLINK was originally developed to detect ROH from SNP chip data, it has proven effective with exomes, especially for detecting larger segments in excess of 1500kb (Pippucci et al., 2011; 2014).

Relatedenss between the contempary Tsimshian individuals was measured via the Ajk statistics using VCFtools (Danecek, 2011). Two Tsimshian individuals (T057 and T058) showed a relatively high relatedness to each other (score = 0.5) and were removed from the analysis. The remaining pairwise comparisons show an average relatedness score of 0.04, with 0 being non-related and 1 being self.

To determine coding variants, GERP scores, PhyoPhen2 classification, and private alleles (via 1,000 Genomes and dbSNP filters), the annotation suite ANNOVAR was utilized (http://www.openbioinformatics.org/annovar/), using the variant files generated by ANGSD. All population genetic measures were calculated with VCFtools (http://vcftools.sourceforge.net),

using the variant files generated by ANGSD. All graphs were generated in R (http://www.r-project.org) with the ggplot2 library (http://ggplot2.org).

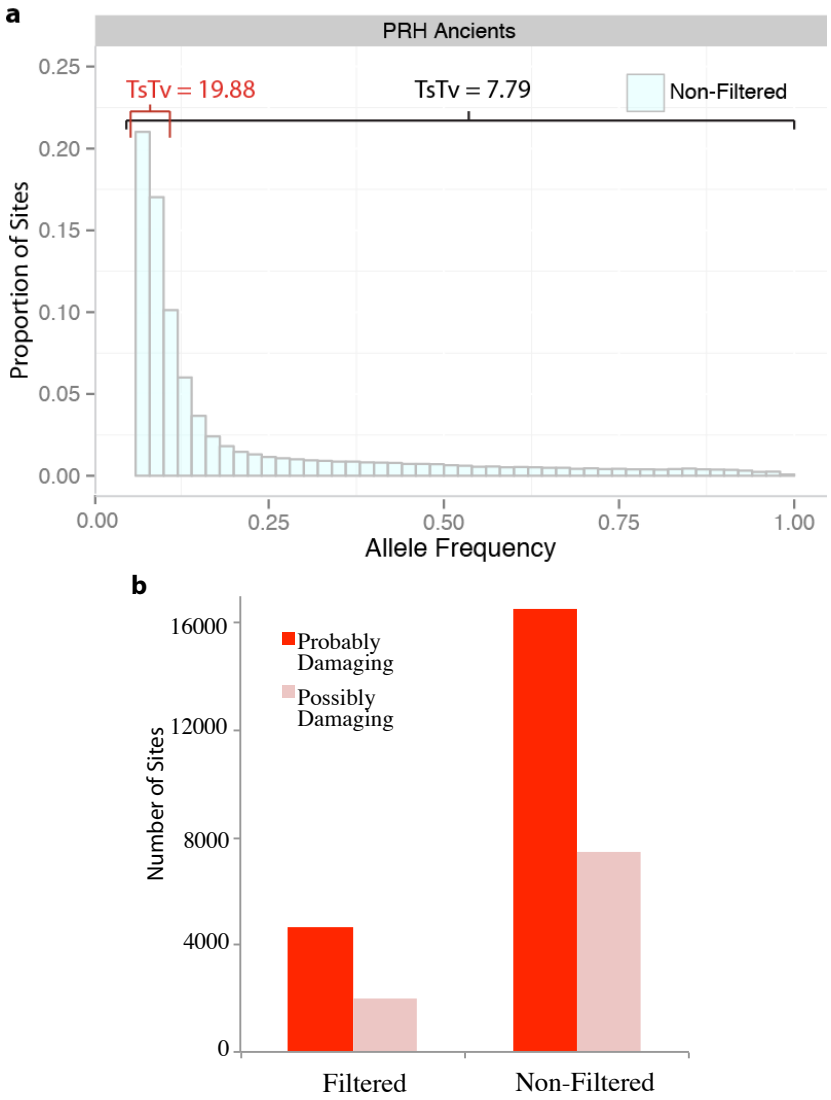

Figure 32. Transition Frequencies below 0.1 and Ancient DNA Damage. a) Allele frequency of the ancient individuals against hg19. Without removing transitions below 0.1, the Transition/Transversion ratio is greatly inflated to 7.79 vs. 3.21 when filtered. The ratio is most disrupted below a frequency of 0.1 due to the concentration of low frequency ancient DNA damage caused by the pooling of all individuals for allele frequency estimation. b) Polyphen2 predictions comparing the filtered and non-filtered ancient dataset. Without filtering, the deleterious predictions nearly quadruple due to the concentration of ancient DNA damaged sites.

CHAPTER 5

CONCLUSION

Over the preceding millennia before European contact, indigenous populations inhabited nearly all ecological zones of the Americas and evolved complex cultures, languages, and civilizations, in relative isolation from other continental populations (Thornton, 1998; Mulligan et al., 2004). At the same time, genomic features were evolving, as a consequence of demographic factors and possible adaptations to local environments. The rapid environmental change that ensued after contact, however, devastated indigenous groups—mainly due to the pathogens carried by non-natives (Boyd, 1999; Patterson and Runge, 2002; Ramenofsky, 2003). These pathogens, which had been co-evolving with Eurasian populations for centuries, interacted with indigenous peoples adapted to an ancient environment that ceased to exist after their arrival. The results of this project is consistent with ancient adaptations, which may have impacted the virulence of diseases such as smallpox.

The evolutionary history of the indigenous peoples of the Americas, however, is much more expansive than the events which occurred after European contact. This is exemplified by the early history of the Pacific Northwest coast. The populations of this region show a dual ancestral lineage, with genetic signatures from a primary migration wave from Asia, which reached both continents, and that of a subsequent migration with regional impact. A surprising feature of the population history, reveled by measuring effective population size, showed evidence of a possible population decline that began well before European contact. These factors paint an interesting picture regarding changes in genetic diversity through time. It would be far too simplistic to think that indigenous populations simply engaged in a continual population expansion since the initial peopling. Instead, their history is nuanced by the arrival of peoples from subsequent migration waves and discontinuous trends in genetic diversity.

The evolutionary history was significantly impacted after European contact, where the aftermath of epidemics, warfare, and displacement resulted in a tremendous loss of genetic diversity through massive population declines (Thornton et al., 1991; Patterson and Runge, 2002; Ramenofsky, 2003). Theoretical predictions of a population that has undergone such a severe

and recent contraction entail deleterious consequences stemming from the accompanied increases in homozygosity, on both the individual and population levels (Nei et al., 1975; Chakraborty and Nei, 1977). These consequences could limit adaptive ability due to the lack of genetic variation, and increase the chances of rare and complex diseases in contemporary populations (Chakraborty and Nei, 1977; Bouzat, 2010). And yet, this project shows that the observed greatly deviates from the expected. Much like the cultural and spiritual resilience demonstrated by many modern indigenous communities, their genomic patterns do not show the predicted crippling effects of a population decline (Bouzat, 2010). Instead, their genomes show a genetic connectivity with non-native groups, which has increased genetic diversity and mitigated the possible "deleterious" effects. Thus, this work helps to illuminate the dynamics of human adaptation to new environments, in both the context of isolation and rapid merging of populations. It also reflects the genomic aspects of globalization. As human population boundaries begin to fade through global integration, so will the boundaries to gene flow, which will boost genetic diversity and increase the resilience to future epidemics and environmental challenges.

Future Directions

This work explores the various evolutionary facets of an indigenous population from the Americas. While the history of this population has many features that can be considered characteristic of other native groups, there are many other features that are not. Migration histories, environmental factors, and cultural differences all combine to create distinction in genomic patterns. Thus, the scope of the studies presented here should be extended to other indigenous groups of the Americas, especially to those that may have encountered distinct environmental pressures. The Americas are comprised of both moderate and extreme environments, from frozen tundras to tropical rain forests, and its indigenous peoples were able to adapt and thrive without limitation (Thornton, 1998; Mulligan et al., 2004).

This raises several questions for future studies. First, if ancient adaptations to the Americas did have an impact on the disease outcome with respect to European borne pathogens, were these adaptations different in various native groups? Could these adaptations have been different or lacking in some groups due to differential environmental factors? Could ancestral migration

waves, in terms of their original genetic composition, affect later genomic features in different populations? Pathogen loads also vary greatly between ecological zones (Leendertz et al., 2006; Karlsson et al., 2014). Perhaps areas in the tropics, which carry different pathogen vectors, may have resulted in distinct immune-based adaptations.

The timing of a potential adaptive event is also of interest here. If certain alleles rose to high frequency early in the first migration that populated both continents, they may have persisted in many native populations at the time of European contact. This could help explain, from a genetic perspective, the widespread elevated mortality rates of indigenous groups with respect to European-borne disease.

The use of ancient DNA will be key in answering these questions, as it was for this dissertation. Combining ancient DNA with modern sampling greatly amplifies the study of complex biological processes that govern evolutionary adaptations, by permitting the examination of genetic patterns through time. These ancient genetic findings could also be placed into a more complex molecular framework utilizing microbiological techniques to determine cellular function. By combining population genetics and functional genomics, a richer characterization of the evolution of a potential adaptive trait could be achieved. With these future studies, a more comprehensive picture can be formed detailing the evolutionary history of these indigenous peoples as they populated, adapted, and established civilizations in the Americas, thousands of years before the arrival of Europeans.

BIBLIOGRAPHY

Achilli A, Perego UA, Lancioni H, Olivieri A, Gandini F, Hooshiar Kashani B, Battaglia V, Grugni V, Angerhofer N, Rogers MP, Herrera RJ, Woodward SR, Labuda D, Smith DG, Cybulski JS, Semino O, Malhi RS, Torroni A. 2013. Reconciling migration models to the Americas with the variation of North American native mitogenomes. Proceedings of the National Academy of Sciences 110:14308–14313.

Acuña-Alonzo V, Flores-Dorantes T, Kruit JK, Villarreal-Molina T, Arellano-Campos O, Hünemeier T, Moreno-Estrada A, Ortiz-López MG, Villamil-Ramírez H, León-Mimila P, Villalobos-Comparan M, Jacobo-Albavera L, Ramírez-Jiménez S, Sikora M, Zhang L-H, Pape TD, Granados-Silvestre M de A, Montufar-Robles I, Tito-Alvarez AM, Zurita-Salinas C, Bustos-Arriaga J, Cedillo-Barrón L, Gómez-Trejo C, Barquera-Lozano R, Vieira-Filho JP, Granados J, Romero-Hidalgo S, Huertas-Vázquez A, González-Martín A, Gorostiza A, Bonatto SL, Rodríguez-Cruz M, Wang L, Tusié-Luna T, Aguilar-Salinas CA, Lisker R, Moises RS, Menjivar M, Salzano FM, Knowler WC, Bortolini MC, Hayden MR, Baier LJ, Canizales-Quinteros S. 2010. A functional ABCA1 gene variant is associated with low HDL-cholesterol levels and shows evidence of positive selection in Native Americans. Human Molecular Genetics 19:2877–2885.

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Research 19:1655–1664.

Archer D. 2011. The Lucy Island Archaeological Project. Unpublished report on file with the British Columbia Archaeology Branch, Victoria.

Auton A, Bryc K, Boyko AR, Lohmueller KE, Novembre J, Reynolds A, Indap A, Wright MH, Degenhardt JD, Gutenkunst RN, King KS, Nelson MR, Bustamante CD. 2009. Global distribution of genomic diversity underscores rich complex history of continental human populations. Genome Research 19:795–803.

Ben Langmead, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology 10:R25.

Bianchine PJ, Russo TA. 1992. The Role of Epidemic Infectious Diseases in the Discovery of America. Allergy and Asthma Proceedings 13:225–232.

Black FL. 1975. Infectious diseases in primitive societies. Science 187:515–518.

Black FL. 1992. Why did they die? Science 258:1739–1740.

Bodmer W, Bonilla C. 2008. Common and rare variants in multifactorial susceptibility to common diseases. Nature Genetics 40:695–701.

Bodner M, Perego UA, Huber G, Fendt L, Rock AW, Zimmermann B, Olivieri A, Gomez-Carballa A, Lancioni H, Angerhofer N, Bobillo MC, Corach D, Woodward SR, Salas A, Achilli A, Torroni A, Bandelt HJ, Parson W. 2012. Rapid coastal spread of First Americans: Novel insights from South America's Southern Cone mitochondrial genomes. Genome Research 22:811–820.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

Bolnick DA. 2006. Asymmetric Male and Female Genetic Histories among Native Americans from Eastern North America. Molecular Biology and Evolution 23:2161–2174.

Bouzat JL. 2010. Conservation genetics of population bottlenecks: the role of chance, selection, and history. Conserv Genetics 11:463–478.

Boyd R. 1990. Demograhic History, 1774-1874. In: Handbook of Native Americans. Vol. 7. ed. Sturtevant, W. Washington, D.C.: Smithsonian Institution. p 135–148.

Boyd R. 1999. The Coming of the Spirit of Pestilence. Seattle: University of Washington Press.

Briggs AW, Stenzel U, Johnson PLF, Green RE, Kelso J, Prufer K, Meyer M, Krause J, Ronan MT, Lachmann M, Paabo S. 2007. Patterns of damage in genomic DNA sequences from a Neandertal. Proceedings of the National Academy of Sciences 104:14616–14621.

Caldwell. 2005. Culturally Competent Research with American Indians and Alaska Natives: Findings and Recommendations of the First Symposium of the Work Group on American Indian Research and Program Evaluation Methodology. American Indian and Alaska Native Mental Health Research 12:1–21.

Casals F, Hodgkinson A, Hussin J, Idaghdour Y, Bruat V, de Maillard T, Grenier J-C, Gbeha E, Hamdan FF, Girard S, Spinella J-F, Larivière M, Saillour V, Healy J, Fernández I, Sinnett D, Michaud JL, Rouleau GA, Haddad E, Le Deist F, Awadalla P. 2013. Whole-Exome Sequencing Reveals a Rapid Change in the Frequency of Rare Functional Variants in a Founding Population of Humans. PLoS Genet 9:e1003815.

Cavalli-Sforza LL, Feldman MW. 2003. The application of molecular genetic approaches to the study of human evolution. Nature Genetics 33:266–275.

Census of Canada. 2006. British ColumbiaStatistical Profile of Aboriginal Peoples 2006:1–18.

Chakraborty R, Nei M. 1977. Bottleneck Effects on Average Heterozygosity and Genetic Distance with the Stepwise Mutation Model. Evolution 31:347.

Charlesworth B. 2009. Effective population size and patterns of molecular evolution and variation. Nature Reviews Genetics 10:195–205.

Cibulskis K, McKenna A, Fennell T, Banks E, DePristo M, Getz G. 2011. ContEst: estimating cross-contamination of human samples in next-generation sequencing data. Bioinformatics 27:2601–2602.

Consortium T1GP, author C, committee S, Medicine PGBCO, BGI-Shenzhen, Broad Institute of MIT and Harvard, Illumina, Technologies L, Max Planck Institute for Molecular Genetics, Science RA, Washington University in St Louis, Wellcome Trust Sanger Institute, Technologies AGA, Medicine BCO, College B, Hospital BAW, Cardiff University, The Human Gene Mutation Database, Laboratory CSH, Universities CAS, European Bioinformatics Institute, Laboratory EMB, Johns Hopkins University, Leiden University Medical Center, Louisiana State University, US National Institutes of Health, Oxford University, The Translational Genomics Research Institute, University of California, Santa Cruz, University of Chicago, University of Michigan, University of Montreal, University of Utah, University of Washington, Universities YAS, BGI-Shenzhen SVG, Medicine EPGBCO, Yale University, group SAE, management S, group W. 2010. A map of human genome variation from population-scale sequencing. Nature 467:1061–1073.

Cook ND. 1998. Born to Die: Disease and New World conquest, 1492-1650. Cambridge: Cambridge University Press.

Cooper GM, Stone EA, Asimenos G, NISC Comparative Sequencing Program, Green ED, Batzoglou S, Sidow A. 2005. Distribution and intensity of constraint in mammalian genomic sequence. Genome Research 15:901–913.

Coventry A, Bull-Otterson LM, Liu X, Clark AG, Maxwell TJ, Crosby J, Hixson JE, Rea TJ, Muzny DM, Lewis LR, Wheeler DA, Sabo A, Lusk C, Weiss KG, Akbar H, Cree A, Hawes AC, Newsham I, Varghese RT, Villasana D, Gross S, Joshi V, Santibanez J, Morgan M, Chang K, Hale W IV, Templeton AR, Boerwinkle E, Gibbs R, Sing CF. 2010. Deep resequencing reveals excess rare recent variants consistent with explosive population growth. Nature Communications 1:131.

Creanza N, Ruhlen M, Pemberton TJ, Rosenberg NA, Feldman MW, Ramachandran S. 2015. A comparison of worldwide phonemic and genetic variation in human populations. Proceedings of the National Academy of Sciences of the United States of America PNAS 112:1265-72

Cui Y, Lindo J, Hughes CE, Johnson JW, Hernandez AG, Kemp BM, Ma J, Cunningham R, Petzelt B, Mitchell J, Archer D, Cybulski JS, Malhi RS. 2013. Ancient DNA Analysis of Mid-Holocene Individuals from the Northwest Coast of North America Reveals Different Evolutionary Paths for Mitogenomes. PLoS ONE 8:e66948.

Cybulski JS ed. 2001. Perspectives on Northern Northwest Coast Prehistory. Quebec : Canadian Museum of Civilization.

Charlesworth A, Charlesworth B. 2003. Inbreeding Depression and its Evolutionary Consequences. Annual Reviews Ecology 18:237–268.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R; 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158.

Delgado JC, Baena A, Thim S, Goldfeld AE. 2006. Aspartic acid homozygosity at codon 57 of HLA-DQ beta is associated with susceptibility to pulmonary tuberculosis in Cambodia. The Journal of Immunology 176:1090–1097.

Dillehay TD, Ramirez C, Pino M, Collins MB, Rossen J, Pino-Navarro JD. 2008. Monte Verde: seaweed, food, medicine, and the peopling of South America. Science 320:784–786.

Dixon EJ, Heaton TH, Lee CM. 2014. Evidence of Maritime Adaptation and Coastal Migration from Southeast Alaska,. In: OWSLEY DW, editor. Kennewick Man The Scientific Investigation of an Ancient American Skeleton. Texas A&M University Press.

Dixon EJ. 1993. Quest for the Origins of the First Americans. University of New Mexico Press.

Dixon EJ. 1999. Bones, Boats, and Bison. Albuquerque: University of New Mexico Press.

Dobyns HF. 1993. Disease transfer at contact. Annual Review of Anthropology 22:273–291.

Dong PDS, Leach S, Stainier DY. 2007. ptf1a determines pancreatic exocrine versus endocrine fates. Developmental Biology 306:416.

Dulik MC, Owings AC, Gaieski JB, Vilar MG, Andre A, Lennie C, Mackenzie MA, Kritsch I, Snowshoe S, Wright R, Martin J, Gibson N, Andrews TD, Schurr TG, The Genographic Consortium, The Genographic Consortium., Adhikarla S, Adler CJ, Balanovska E, Balanovsky O, Bertranpetit J, Clarke AC, Comas D, Cooper A, Sarkissian Der CSI, GaneshPrasad A, Haak W, Haber M, Hobbs A, Javed A, Jin L, Kaplan ME, Li S, Martinez-Cruz B, Matisoo-Smith EA, Mele M, Merchant NC, Mitchell RJ, Parida L, Pitchappan R, Platt DE, Quintana-Murci L, Renfrew C, Lacerda DR, Royyuru AK, Santos FR, Soodyall H, Soria Hernanz DF, Swamikrishnan P, Tyler-Smith C, Santhakumari AV, Vieira PP, Wells RS, Zalloua PA, Ziegle JS, Consortium TG. 2012. Y-chromosome analysis reveals genetic divergence and new founding native lineages in Athapaskan- and Eskimoan-speaking populations. Proceedings of the National Academy of Sciences of the United States of America 109:8471–8476.

Dyke AS, Andrews JT, Clark PU, England JH, Miller GH, Shaw J, Veillette JJ. 2002. The Laurentide and Innuitian ice sheets during the Last Glacial Maximum. Quaternary Science Reviews 21:9–31.

Ehlers CL, Liang T, Gizer IR. 2012. ADH and ALDH Polymorphisms and Alcohol Dependence in Mexican and Native Americans. American Journal of Drug and Alcohol Abuse 38:389–394.

Erlandson JM, Graham MH, Bourque BJ, Corbett D, Estes JA, Steneck RS. 2007. The Kelp Highway Hypothesis: Marine Ecology, the Coastal Migration Theory, and the Peopling of the Americas. The Journal of Island and Coastal Archaeology 2:161–174.

Eshleman JA, Malhi RS, Johnson JRJR, Kaestle FA, Lorenz J, Smith DG. 2004. Mitochondrial DNA and Prehistoric Settlements: Native Migrations on the Western Edge of North America. Human Biology 76:55–75.

Excoffier L, Dupanloup I, Huerta-SÃ nchez E, Sousa VC, Foll M. 2013. Robust Demographic Inference from Genomic and SNP Data. PLoS Genet 9:e1003905.

Fedje DW, Christensen T. 1999. Modeling Paleoshorelines and Locating Early Holocene Coastal Sites in Haida Gwaii. American Antiquity 64:635.

Fenn EA. 2001. Pox Americana: The Great Smallpox Epidemic of 1775-82. New York: Hill and Wang.

Fiedel SJ. 2000. The Peopling of the New World: Present Evidence, New Theories, and Future Directions. Journal of Archaeological Research 8:1–67.

Fix AG. 2005. Rapid deployment of the five founding Amerind mtDNA haplogroups via coastal and riverine colonization. American Journal of Physical Anthropology 128:430–436.

Fladmark K, Ames K, Sutherland P. 1990. Prehistory of the Northern Coast of British Columbia. In: Sturtevant W, editor. Handbook of North American Indians. Vol. 7. Washington, D.C.: Smithsonian Institution.

Fladmark KR. 1979. Routes: Alternate Migration Corridors for Early Man in North America. American Antiquity 44:55.

Fox CW, Scheibly KL, Reed DH. 2008. EXPERIMENTAL EVOLUTION OF THE GENETIC LOAD AND ITS IMPLICATIONS FOR THE GENETIC BASIS OF INBREEDING DEPRESSION. Evolution 62:2236–2249.

Fu W, O'Connor TD, Jun G, Kang HM, Abecasis G, Leal SM, Gabriel S, Altshuler D, Shendure J, Nickerson DA, Bamshad MJ, NHLBI Exome Sequencing Project, Akey JM. 2012. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature 493:216–220.

Fumagalli M, Sironi M, Pozzoli U, Ferrer-Admetlla A, Pattini L, Nielsen R. 2011. PLOS Genetics: Signatures of Environmental Genetic Adaptation Pinpoint Pathogens as the Main Selective Pressure through Human Evolution. PLos Genetics 7(11):e1002355.

Fumagalli M, Vieira FG, Korneliussen TS, Linderoth T, Huerta-Sanchez E, Albrechtsen A, Nielsen R. 2013. Quantifying Population Genetic Differentiation from Next-Generation Sequencing Data. Genetics 195:979–992.

Gibson J, Morton NE, Collins A. 2006. Extended tracts of homozygosity in outbred human populations. Human Molecular Genetics 15:789–795.

Gilad Y, Bustamante C, Lancet D, Paabo S. 2003. Natural Selection on the Olfactory Receptor Gene Family in Humans and Chimpanzees. The American Journal of Human Genetics 73:489–501.

Goebel T. 1999. Pleistocene human colonization of Siberia and peopling of the Americas: An ecological approach. Evolutionary Anthropology: Issues, News, and Reviews 8:208–227.

Gravel S, Henn BM, Gutenkunst RN, Indap AR, Marth GT, Clark AG, Yu F, Gibbs RA, 1000 Genomes Project, Bustamante CD. 2011. Demographic history and rare allele sharing among human populations. Proceedings of the National Academy of Sciences 108:11983–11988.

Gravel S, Zakharia F, Moreno-Estrada A, Byrnes JK, Muzzio M, Rodriguez-Flores JL, Kenny EE, Gignoux CR, Maples BK, Guiblet W, Dutil J, Via M, Sandoval K, Bedoya G, The 1000 Genomes Project, Oleksyk TK, Ruiz-Linares A, Burchard EG, Martinez-Cruzado JC, Bustamante CD. 2013. Reconstructing Native American Migrations from Whole-Genome and Whole-Exome Data. PLoS Genet 9:e1004023.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH-Y, Hansen NF, Durand EY, Malaspinas A-S, Jensen JD, Marquès-Bonet T, Alkan C, Prüfer K, Meyer M, Burbano HA, Good JM, Schultz R, Aximu-Petri A, Butthof A, Höber B, Höffner B, Siegemund M, Weihmann A, Nusbaum C, Lander ES, Russ C, Novod N, Affourtit J, Egholm M, Verna C, Rudan P, Brajković D, Kucan Ž, Gušic I, Doronichev VB, Golovanova LV, Lalueza-Fox C, la Rasilla de 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, Pääbo S. 2010. A draft sequence of the Neandertal genome. Science 328:710–722.

Greenberg JH, II CGT, Zegura SL. 1986. The Settlement of the Americas: A Comparison of the Linguistic, Dental, and Genetic Evidence. Current Anthropology 27:477.

Grueber Ce, Wallis Gp, Jamieson Ig. 2008. Heterozygosity–fitness correlations and their relevance to studies on inbreeding depression in threatened species. Molecular Ecology 17:3978–3984.

Havlicek J, Roberts SC. 2009. MHC-correlated mate choice in humans: A review. Psychoneuroendocrinology 34:497–512.

Heaton TH, Grady F. 2003. The late Wisconsin vertebrate history of Prince of Wales Island, Southeast Alaska. In: Schubert BW, editor. Ice Age Cave Faunas of North America. +++

Hodge FS, Weinmann S, Roubideaux Y. 2000. Recruitment of American Indians and Alaska Natives Into Clinical Trials. Annals of Epidemiology 10:S41–S48.

Howrigan DP, Simonson MA, Keller MC. 2011. Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. BMC Genomics 12:460.

Huerta-Sanchez E, DeGiorgio M, Pagani L, Tarekegn A, Ekong R, Antao T, Cardona A, Montgomery HE, Cavalleri GL, Robbins PA, Weale ME, Bradman N, Bekele E, Kivisild T, Tyler-Smith C, Nielsen R. 2013. Genetic Signatures Reveal High-Altitude Adaptation in a Set of Ethiopian Populations. Molecular Biology and Evolution 30:1877–1888.

Humphreys K, Grankvist A, Leu M, Hall P, Liu J, Ripatti S, Rehnström K, Groop L, Klareskog L, Ding B, Grönberg H, Xu J, Pedersen NL, Lichtenstein P, Mattingsdal M, Andreassen OA, O'Dushlaine C, Purcell SM, Sklar P, Sullivan PF, Hultman CM, Palmgren J, Magnusson PKE. 2011. The Genetic Structure of the Swedish Population. PLoS ONE 6:e22547.

Hunley K, Healy M. 2011. The impact of founder effects, gene flow, and European admixture on native American genetic diversity. American Journal of Physical Anthropology 146:530–538.

Haynes C. 1964. Fluted Projectile Points: Their Age and Dispersion: Stratigraphically controlled radiocarbon dating provides new evidence on peopling of the New World. Science 145:1408-13.

Jacob S, McClintock MK, Zelano B, Ober C. 2002. Paternally inherited HLA alleles are associated with women's choice of male odor. Nature Genetics 30:175–179.

Jones EY, Fugger L, Strominger JL, Siebold C. 2006. MHC class II proteins and disease: a structural perspective. Nature Reviews Immunology 6:271–282.

Jonsson H, Ginolhac A, Schubert M, Johnson PLF. 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics 29:1682–1684.

Kaestle F and Smith D. 2001. Ancient mitochondrial DNA evidence for prehistoric population movement: The numic expansion. American Journal of Physical Anthropology 115:1–12.

Karlsson EK, Kwiatkowski DP, Sabeti PC. 2014. Natural selection and infectious disease in human populations. Nature 15:379–393.

Kawaguchi Y, Cooper B, Gannon M, Ray M, MacDonald R, Wright C. 2002. The role of the transcriptional regulator Ptf1a in converting intestinal to pancreatic progenitors. Nature Genetics 32:128–134.

Keinan A, Clark AG. 2012. Recent Explosive Human Population Growth Has Resulted in an Excess of Rare Genetic Variants. Science 336:740–743.

Keller A, Zhuang H, Chi Q, Vosshall LB, Matsunami H. 2007. Genetic variation in a human odorant receptor alters odour perception. Nature 449:468–472.

Kemp BM, Malhi RS, McDonough J, Bolnick DA, Eshleman JA, Rickards O, Martinez-Labarga C, Johnson JR, Lorenz JG, Dixon EJ, Fifield TE, Heaton TH, Worl R, Smith DG. 2007. Genetic analysis of early holocene skeletal remains from Alaska and its implications for the settlement of the Americas. Am J Phys Anthropol 132:605–621.

Kim HS, Park MH, Song EY, Park H, Kwon SY, Han SK, Shim Y-S. 2005. Association of HLA-DR and HLA-DQ genes with susceptibility to pulmonary tuberculosis in Koreans: preliminary evidence of associations with drug resistance, disease severity, and disease recurrence. Hum Immunol 66:1074–1081.

Kirin M, McQuillan R, Franklin CS, Campbell H, McKeigue PM, Wilson JF. 2010. Genomic Runs of Homozygosity Record Population History and Consanguinity. PLoS ONE 5:e13996.

Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics 15:356.

Kryukov GV, Pennacchio LA, Sunyaev SR. 2007. Most Rare Missense Alleles Are Deleterious in Humans: Implications for Complex Disease and Association Studies. The American Journal of Human Genetics 80:727–739.

Lahren L and Bonnichsen R. 1974. Bone foreshafts from a clovis burial in southwestern Montana. Science 186:147–150.

Lander ES, Botstein D. 1987. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. Science 236:1567–1570.

Langmead B and Salzberg S. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9:357–359.

Larsen CS. 1994. In the wake of Columbus: Native population biology in the postcontact Americas. Am J Phys Anthropol 37:109–154.

Lee MH, Lu K, Hazard S, Yu H, Shulenin S, Hidaka H, Kojima H, Allikmets R, Sakuma N, Pegoraro R, Srivastava AK, Salen G, Dean M, Patel SB. 2001. Identification of a gene, ABCG5, important in the regulation of dietary cholesterol absorption. Nat Genet 27:79–83.

Leendertz F, Pauli G, Maetz-Rensing K, Boardman W, Nunn C, Ellerbrok H, Jensen S, Junglen S, Christophe B. 2006. Pathogens as drivers of population declines: the importance of systematic monitoring in great apes and other threatened mammals. Biological Conservation 131:325–337.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup 1GPDP. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL, Myers RM. 2008. Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation. Science 319:1100–1104.

Linnen CR, Hoekstra HE. 2010. Measuring Natural Selection on Genotypes and Phenotypes in the Wild. Cold Spring Harbor Symposia on Quantitative Biology 74:155–168.

Lohmueller KE, Indap AR, Schmidt S, Boyko AR, Hernandez RD, Hubisz MJ, Sninsky JJ, White TJ, Sunyaev SR, Nielsen R, Clark AG, Bustamante CD. 2008. Proportionally more deleterious genetic variation in European than in African populations. Nature 451:994–997.

Macaulay AC. 1994. Ethics of research in Native communities. Canadian Family Physician 40:1888-97.

Malhi RS, Breece KE, Shook BAS, Kaestle FA, Chatters JC, Hackenberger S, Smith DG. 2004. Patterns of mtDNA Diversity in Northwestern North America. Human Biology 76:33–54.

Malhi RS, Mortensen HM, Eshleman JA, Kemp BM, Lorenz JG. 2003. Native American mtDNA prehistory in the American Southwest. American Journal of Physical Anthropology 120:108–124.

Mancall P and Merrell J 2000. American Encounters. New York: Psychology Press.

Martindale A, Marsden S. 2003. DEFINING THE MIDDLE PERIOD *(3500 BP to 1500 BP) in Tsimshian History through a Comparison of Archaeological and Oral Records*. BC Studies 138:13–50.

McInnes R, 2011. 2010 Presidential Address: Culture: The Silent Language Geneticists Must Learn— Genetic Research with Indigenous Populations. American Journal of Human Genetics 88:254–261.

McMichael AJ. 2013. Globalization, Climate Change, and Human Health. N Engl J Med 368:1335–1343.

Merbs CF. 1992. A new world of infectious disease. Am J Phys Anthropol 35:3–42.

Meyer M, Kircher M. 2010. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. Cold Spring Harbor Protocols doi:10.1101/pdb.prot5448.

Milinski M, Croy I, Hummel T, Boehm T. 2013. Major histocompatibility complex peptide ligands as olfactory cues in human body odour assessment. Proceedings of the Royal Society B: Biological Sciences 280:20122889–20122889.

Moreno-Estrada A, Gravel S, Zakharia F, McCauley JL, Byrnes JK, Gignoux CR, Ortiz-Tello PA, Martínez RJ, Hedges DJ, Morris RW, Eng C, Sandoval K, Acevedo-Acevedo S, Norman PJ, Layrisse Z, Parham P, Martinez-Cruzado JC, Burchard EG, Cuccaro ML, Martin ER, Bustamante CD. 2013. Reconstructing the Population Genetic History of the Caribbean. PLoS Genet 9:e1003925.

Moss WJ, Griffin DE. 2006. Global measles elimination. Nat Rev Micro 4:900–908.

Motulsky AG. 1989. Metabolic polymorphisms and the role of infectious diseases in human evolution. Human Biology:835–869.

Mulligan CJ, Hunley K, Cole S, Long JC. 2004. Population genetics, history, and health patterns in native americans. Annual Review of Genomics and Human Genetics 5:295–315.

Mulligan CJ, Kitchen A, Miyamoto MM. 2008. Updated Three-Stage Model for the Peopling of the Americas. PLoS ONE 3:e3199.

Mulligan CJ, Robin RW, Osier MV, Sambuughin N, Goldfarb LG, Kittles RA, Hesselbrock D, Goldman D, Long JC. 2003. Allelic variation at alcohol metabolism genes (ADH1B, ADH1C, ALDH2) and alcohol dependence in an American Indian population. Hum Genet 113:325–336.

Nei M, Maruyama T, Chakraborty R. 1975. The Bottleneck Effect and Genetic Variability in Populations. Evolution 29:1.

Nelson MR, Wegmann D, Ehm MG, Kessner D, Jean PS, Verzilli C, Shen J, Tang Z, Bacanu S-A, Fraser D, Warren L, Aponte J, Zawistowski M, Liu X, Zhang H, Zhang Y, Li J, Li Y, Li L, Woollard P, Topp S, Hall MD, Nangle K, Wang J, Abecasis G, Cardon LR, Zöllner S, Whittaker JC, Chissoe SL, Novembre J, Mooser V. 2012. An Abundance of Rare Functional Variants in 202 Drug Target Genes Sequenced in 14,002 People. Science 337:100–104.

Nguyen TM, Sawyer JK, Kelley KL, Davis MA, Kent CR, Rudel LL. 2012. ACAT2 and ABCG5/G8 are both required for efficient cholesterol absorption in mice: evidence from thoracic lymph duct cannulation. J Lipid Res 53:1598–1609.

Nielsen R, Paul JS, Albrechtsen A, Song YS. 2011. Genotype and SNP calling from next-generation sequencing data. Nature 12:443–451.

Nielsen R, Korneliussen T, Albrechtsen A, Li Y, Wang J. 2012. SNP Calling, Genotype Calling, and Sample Allele Frequency Estimation from New-Generation Sequencing Data. PLos One 7(7):e37558.

O'Fallon BD, Fehren-Schmitz L. 2011. Native Americans experienced a strong population bottleneck coincident with European contact. Proceedings of the National Academy of Sciences of the United States of America 108:20444–20448.

Ovsyannikova IG, Pankratz VS, Salk HM, Kennedy RB, Poland GA. 2014. HLA alleles associated with the adaptive immune response to smallpox vaccine: a replication study. Hum Genet 133:1083–1092.

Ovsyannikova IG, Vierkant RA, Pankratz VS, Jacobson RM, Poland GA. 2011. Human Leukocyte Antigen Genotypes in the Genetic Control of Adaptive Immune Responses to Smallpox Vaccine. Journal of Infectious Diseases 203:1546–1555.

Ovsyannikova IG, Vierkant RA, Poland GA. 2006. Importance of HLA-DQ and HLA-DP polymorphisms in cytokine responses to naturally processed HLA-DR-derived measles virus peptides. Vaccine 24:5381–5389.

Palstra F, Ruzzante D. 2008. Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild population persistence? Molecular Ecology 17:3428–3447.

Patterson KB, Runge T. 2002. Smallpox and the Native American. The American journal of the medical sciences 323:216.

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. 2012. Ancient Admixture in Human History. Genetics 192:1065–1093.

Pause BM. 2011. Processing of Body Odor Signals by the Human Brain. Chemosensory Perception 5:55–63.

Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg NA, Li JZ. 2012. Genomic Patterns of Homozygosity in Worldwide Human Populations. The American Journal of Human Genetics 91:275–292.

Perego UA, Achilli A, Angerhofer N, Accetturo M, Pala M, Olivieri A, Kashani BH, Ritchie KH, Scozzari R, Kong Q-P, Myres NM, Salas A, Semino O, Bandelt H-J, Woodward SR, Torroni A. 2009. Distinctive Paleo-Indian Migration Routes from Beringia Marked by Two Rare mtDNA Haplogroups. Current Biology 19:1–8.

Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet 8:e1002967.

Pippucci T, Benelli M, Magi A, Martelli PL, Magini P, Torricelli F, Casadio R, Seri M, Romeo G. 2011. EX-HOM (EXome HOMozygosity): A Proof of Principle. Hum Hered 72:45–53.

Pippucci T, Magi A, Gialluisi A, Romeo G. 2014. Detection of runs of homozygosity from whole exome sequencing data: state of the art and perspectives for clinical, population and epidemiological studies. Hum Hered 77:63–72.

Pitblado BL. 2011. A Tale of Two Migrations: Reconciling Recent Biological and Archaeological Evidence for the Pleistocene Peopling of the Americas. J Archaeol Res 19:327–375.

Plat J, Mensink RP. 2002. Increased intestinal ABCA1 expression contributes to the decrease in cholesterol absorption after plant stanol consumption. FASEB J 16:1248–1253.

Raghavan M, Skoglund P, Graf K, Metspalu M, Albrechtsen A, Moltke I, Rasmussen S, Stafford T, Orlando L, Metspalu E, Karmin M, Tambets K, Rootsi S, Mägi R, Campos P, Balanovska E, Balanovsky O, Khusnutdinova E, Litvinov S, Osipova L, Fedorova S, Voevoda M, DeGiorgio M, Sicheritz-Ponten T, Brunak S, Demeshchenko S, Kivisild T, Villems R, Nielsen R, Jakobsson M, Willerslev E. 2013. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature 505:87–91.

Raghavan M, DeGiorgio M, Albrechtsen A, Moltke I, Skoglund P, Korneliussen TS, Grønnow B, Appelt M, Gulløv HC, Friesen TM, Fitzhugh W, Malmström H, Rasmussen S, Olsen J, Melchior L, Fuller BT, Fahrni SM, Stafford T, Grimes V, Renouf MAP, Cybulski J, Lynnerup N, Lahr MM, Britton K, Knecht R, Arneborg J, Metspalu M, Cornejo OE, Malaspinas A-S, Wang Y, Rasmussen M, Raghavan V, Hansen TVO, Khusnutdinova E, Pierre T, Dneprovsky K, Andreasen C, Lange H, Hayes MG, Coltrain J, Spitsyn VA, Gotherstrom A, Orlando L, Kivisild T, Villems R, Crawford MH, Nielsen FC, Dissing J, Heinemeier J, Meldgaard M, Bustamante C, O'Rourke DH, Jakobsson M, Gilbert MTP, Nielsen R, Willerslev E. 2014. The genetic prehistory of the New World Arctic. Science 345:1255832–1255832.

Raghavan M, steinrucken M, Harris K, Schifels S, Rasmussen S, Albrechtserr A, Valdiosera C, Avila-Arcos MC, Malaspinas A-S, DeGiorgio M. 2015. Early Peopling of the Americas. Manuscript Submitted.

Ramchandani VA. 2013. Genetics of Alcohol Metabolism. In: Alcohol, Nutrition, and Health Consequences. Totowa, NJ: Humana Press. p 15–25.

Ramenofsky A. 2003. Native American disease history: past, present and future directions. World Archaeology 35:241–257.

Rasmussen M, Anzick S, Waters M, Skoglund P, DeGiorgio M, Stafford T, Rasmussen S, Moltke I, Albrechtsen A, Doyle S, Poznik G, Gudmundsdottir V, Yadav R, Malaspinas A, White S, Allentoft ME, Cornejo OE, Tambets K, Eriksson A, Heintzman PD, Karmin M, Korneliussen TS, Meltzer DJ, Pierre TL, Stenderup J, Saag L, Warmuth VM, Lopes MC, Malhi RS, Brunak S, Sicheritz-Ponten T, Barnes I, Collins M, Orlando L, Balloux F, Manica A, Gupta R, Metspalu M, Bustamante C, Jakobsson M, Nielsen R, Willerslev E. 2014. The genome of a Late Pleistocene human from a Clovis burial site in western Montana. Nature 506:225–229.

Rasmussen M, Guo X, Wang Y, Lohmueller KE, Rasmussen S, Albrechtsen A, Skotte L, Lindgreen S, Metspalu M, Jombart T, Kivisild T, Zhai W, Eriksson A, Manica A, Orlando L, La Vega De FM, Tridico S, Metspalu E, Nielsen K, Avila-Arcos MC, Moreno-Mayar JV, Muller C, Dortch J, Gilbert MTP, Lund O, Wesolowska A, Karmin M, Weinert LA, Wang B, Li J, Tai S, Xiao F, Hanihara T, van Driem G, Jha AR, Ricaut FX, de Knijff P, Migliano AB, Gallego Romero I, Kristiansen K, Lambert DM, Brunak S, Forster P, Brinkmann B, Nehlich O, Bunce M, Richards M, Gupta R, Bustamante CD, Krogh A, Foley RA, Lahr MM, Balloux F, Sicheritz-Ponten T, Villems R, Nielsen R, Wang J, Willerslev E. 2011. An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia. Science 334:94–98.

Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, Moltke I, Metspalu M, Metspalu E, Kivisild T, Gupta R, Bertalan M, Nielsen K, Gilbert MTP, Wang Y, Raghavan M, Campos PF, Kamp HM, Wilson AS, Gledhill A, Tridico S, Bunce M, Lorenzen ED, Binladen J, Guo X, Zhao J, Zhang X, Zhang H, Li Z, Chen M, Orlando L, Kristiansen K, Bak M, Tommerup N, Bendixen C, Pierre TL, Grønnow B, Meldgaard M, Andreasen C, Fedorova SA, Osipova LP, Higham TFG, Ramsey CB, O Hansen von T, Nielsen FC, Crawford MH, Brunak S, Sicheritz-Pontén T, Villems R, Nielsen R, Krogh A, Wang J, Willerslev E. 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature 463:757–762.

Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, Parra MV, Rojas W, Duque C, Mesa N, García LF, Triana O, Blair S, Maestre A, Dib JC, Bravi CM, Bailliet G, Corach D, Hünemeier T, Bortolini MC, Salzano FM, Petzl-Erler ML, Acuña-Alonzo V, Aguilar-Salinas C, Canizales-Quinteros S, Tusié-Luna T, Riba L, Rodríguez-Cruz M, Lopez-Alarcón M, Coral-Vazquez R, Canto-Cetina T, Silva-Zolezzi I, Fernandez-Lopez JC, Contreras AV, Jimenez-Sanchez G, Gómez-Vázquez MJ, Molina J, Carracedo Á, 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 J-M, 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 488:1–6.

Reynolds J, Weir B, Cockerham C, 1983. Estimation Of The Coancestry Coefficient: Basis For A Short-Term Genetic Distance. Genetics 105:767.

Roberts L. 1989. Disease and death in the New World. Science 246:1245–1247.

Roche PA, Furuta K. 2015. The ins and outs of MHC class II-mediated antigen processing and presentation. Nature Reviews Immunology.

Salmela E, Lappalainen T, Fransson I, Andersen PM, Dahlman-Wright K, Fiebig A, Sistonen P, Savontaus M-L, Schreiber S, Kere J, Lahermo P. 2008. Genome-Wide Analysis of Single Nucleotide Polymorphisms Uncovers Population Structure in Northern Europe. PLoS ONE 3:e3519.

Schroeder KB, Malhi RS, Smith DG. 2006. Opinion: Demystifying Native American genetic opposition to research. Evolutionary Anthropology: Issues, News, and Reviews 15:88–92.

Schurr TG, Sherry ST. 2004. Mitochondrial DNA and Y chromosome diversity and the peopling of the Americas: Evolutionary and demographic evidence. American Journal of Human Biology 16:420–439.

Seguin-Orlando A, Schubert M, Clary J, Stagegaard J. 2013. PLOS ONE: Ligation Bias in Illumina Next-Generation DNA Libraries: Implications for Sequencing Ancient Genomes. PLoS ONE.

Sequin M, Halpin M. 1990. Tsimshian Peoples: Southern Tsimshian, Coast Tsimshian, Nishga, and Gitksan. In: Handbook of North American Indians. Vol. 7. p 1–18.

Shan G. 2005. A Note on Exact Tests of Hardy-Weinberg Equilibrium. Human Heredity 76:887–893.

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.

Skoglund P, Gotherstrom A, Jakobsson M. 2013. Accurate sex identification of ancient human remains using DNA shotgun sequencing. 40:4477–4482.

Smith K, Sax D, Gaines S, Guernier V, Guégan J. 2007. Globalization Of Human Infectious Disease. Ecology 88:1903–1910.

Szpiech ZA, Xu J, Pemberton TJ, Peng W, Zöllner S, Rosenberg NA, Li JZ. 2013. Long Runs of Homozygosity Are Enriched for Deleterious Variation. The American Journal of Human Genetics 93:90–102.

Takahata N, Nei M. 1990. Allelic Genealogy Under Overdominant and Frequency-Dependent Selectionand Polymorphism. Genetics 124:967–978.

Tamm E, Kivisild T, Reidla M, Metspalu M, Smith DG, Mulligan CJ, Bravi CM, Rickards O, Martinez-Labarga C, Khusnutdinova EK, Fedorova SA, Golubenko MV, Stepanov VA, Gubina MA, Zhadanov SI, Ossipova LP, Damba L, Voevoda MI, Dipierri JE, Villems R, Malhi RS. 2007. Beringian Standstill and Spread of Native American Founders. PLoS ONE 2:e829.

Tani T, Ohsumi J, Mita K, Takiguchi Y. 1988. Identification of a novel class of elastase isozyme, human pancreatic elastase III, by cDNA and genomic gene cloning. Journal of Biological Chemistry 263:1231–1239.

Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, McGee S, Do R, Liu X, Jun G, Kang HM, Jordan D, Leal SM, Gabriel S, Rieder MJ, Abecasis G, Altshuler D, Nickerson DA, Boerwinkle E, Sunyaev S, Bustamante CD, Bamshad MJ, Akey JM, Broad GO, Seattle GO, NHLBI Exome Sequencing Project. 2012. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science 337:64–69.

Thornton, R. 1987. American Indian Holocaust and Survival. Norman: University of Oklahoma Press.

Thornton R, Miller T, Warren J. 1991. American Indian Population Recovery Following Smallpox Epidemics. American Anthropologist 93:28–45.

Thornton R. 1997. Aboriginal North American Population and Rates of Decline, ca. a.d. 1500-19001. Current Anthropology 38:310–315.

Thornton R. 1998. Studying Native America. Madison: University of Wisconsin Press.

Torkamani A, Pham P, Libiger O, Bansal V, Zhang G, Scott-Van Zeeland AA, Tewhey R, Topol EJ, Schork NJ. 2012. Clinical Implications of Human Population Differences in Genome-Wide Rates of Functional Genotypes. Frontiers in Genetics 3:1-19.

Verdu P, Pemberton TJ, Laurent R, Kemp BM, Gonzalez-Oliver A, Gorodezky C, Hughes CE, Shattuck MR, Petzelt B, Mitchell J, Harry H, William T, Worl R, Cybulski JS, Rosenberg NA, Malhi RS. 2014. Patterns of Admixture and Population Structure in Native Populations of Northwest North America. PLoS Genetics 10:e1004530.

Wall JD, Jiang R, Gignoux C, Chen GK, Eng C, Huntsman S, Marjoram P. 2011. Genetic Variation in Native Americans, Inferred from Latino SNP and Resequencing Data. Molecular Biology and Evolution 28:2231–2237.

Wang P, Sidney J, Dow C, Mothé B, Sette A, Peters B. 2008. A Systematic Assessment of MHC Class II Peptide Binding Predictions and Evaluation of a Consensus Approach. PLoS Computational Biology 4:e1000048.

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.

Waters MR, Stafford TW. 2007. Redefining the age of Clovis: implications for the peopling of the Americas. Science 315:1122–1126.

Whitcomb C and Lowe M. 2007. Human Pancreatic Digestive Enzymes. Digestive Diseases and Sciences 52:1–17.

Willi Y, Van Buskirk J, Hoffmann AA. 2006. Limits to the Adaptive Potential of Small Populations. Annu. Rev. Ecol. Evol. Syst. 37:433–458.

Woods CG, Cox J, Springell K, Hampshire DJ, Mohamed MD, McKibbin M, Stern R, Raymond FL, Sandford R, Malik Sharif S, Karbani G, Ahmed M, Bond J, Clayton D, Inglehearn CF. 2006. Quantification of Homozygosity in Consanguineous Individuals with Autosomal Recessive Disease. The American Journal of Human Genetics 78:889–896.

Wright W. 1938. Size of a population and breeding structure in relation to evolution. Science 87:430–431.

Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, Xu X, Jiang H, Vinckenbosch N, Korneliussen TS, Zheng H, Liu T, He W, Li K, Luo R, Nie X, Wu H, Zhao M, Cao H, Zou J, Shan Y, Li S, Yang Q, Asan, Ni P, Tian G, Xu J, Liu X, Jiang T, Wu R, Zhou G, Tang M, Qin J, Wang T, Feng S, Li G, Huasang, Luosang J, Wang W, Chen F, Wang Y, Zheng X, Li Z, Bianba Z, Yang G, Wang X, Tang S, Gao G, Chen Y, Luo Z, Gusang L, Cao Z, Zhang Q, Ouyang W, Ren X, Liang H, Huang Y, Li J, Bolund L, Kristiansen K, Li Y, Zhang Y, Zhang X, Li R, Yang H, Nielsen R, Wang J. 2010. Sequencing of 50 Human Exomes Reveals Adaptation to High Altitude. Science 329:75–78.

Yu L, Li-Hawkins J, Hammer RE, Berge KE, Horton JD, Cohen JC, Hobbs HH. 2002. Overexpression of ABCG5 and ABCG8 promotes biliary cholesterol secretion and reduces fractional absorption of dietary cholesterol. J Clin Invest 110:671–680.

Ziegler A, Kentenich H, Uchanska-Ziegler B. 2005. Female choice and the MHC. Trends in Immunology 26:496–502.