

# Demographic and immune-based selection shifts before and after European contact inferred from 50 ancient and modern exomes from the Northwest Coast of North America

**Authors:** John Lindo<sup>1</sup>, Emilia Huerta-Sánchez<sup>2</sup>, Shigeki Nakagome<sup>1</sup>, Morten Rasmussen<sup>3,4</sup>, Barbara Petzelt<sup>5</sup>, Joycelyn Mitchell<sup>5</sup>, Jerome S Cybulski<sup>6</sup>, Eske Willerslev<sup>3</sup>, Michael DeGiorgio<sup>7,8\*</sup>, Ripan S Malhi<sup>1,9\*</sup>

The susceptibility of Native Americans to infectious disease has been postulated as a major factor for their population decline after European contact. To investigate if a preexisting genetic component contributed to this phenomenon, we analyzed 50 exomes of both ancient and modern individuals from the Northwest Coast of North America, dating from before and after European contact. We confirmed the genetic continuity between the ancient and modern individuals and modeled the population collapse after European contact, inferring a 57% reduction in effective population size. We also identified signatures of positive selection on immune-related genes in the ancient but not the modern group. The strongest selection signal in the ancients came from the human leukocyte antigen (HLA) gene *HLA-DQA1*, with alleles that are close to fixation. The important immune function of *HLA-DQA1* supports an ancient adaptation to the environments of the Americas. The modern individuals show a marked decrease in the frequency of the associated alleles (the most pronounced variant showing a 64% difference). This decrease is likely due to the environmental change associated with European colonization, which resulted in a shift of selection pressures, whereby negative selection may have acted on the same gene after contact. Furthermore, the selection pressure shift could correlate to the European-borne epidemics of the 1800s, suffered in the Northwest Coast region. This is among the first studies to examine a single population through time and exemplifies the power of such studies in uncovering nuanced demographic and adaptive histories.

---

<sup>1</sup>Department of Human Genetics, University of Chicago, 920 E 58<sup>th</sup> St, Chicago, IL 60637, USA. <sup>2</sup>Department of Molecular Cell Biology, University of California, 5200 North Lake Road, Merced, CA 95343, USA. <sup>3</sup>Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, DK-1350 Copenhagen K, Denmark. <sup>4</sup>Department of Genetics, School of Medicine, Stanford University, 291 Campus Drive, Stanford, CA 94305, USA. <sup>5</sup>Metlakatla Treaty Office, PO Box 224, Prince Rupert, BC, Canada V8J 3P6. <sup>6</sup>Canadian Museum of History, 100 Rue Laurier, Gatineau, Quebec K1A 0M8, Canada; University of Western Ontario, London, Ontario N6A 3K7, Canada; Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada. <sup>7</sup>Departments of Biology and Statistics, Pennsylvania State University, 502 Wartik Laboratory, University Park, PA 16802, USA. <sup>8</sup>Institute for Cyberscience, Pennsylvania State University, 502 Wartik Laboratory, University Park, PA 16802, USA. <sup>9</sup>Carl R. Woese Institute for Genomic Biology, University of Illinois, 1206 W Gregory Dr., Urbana, IL 61820, USA. \*To whom correspondence to should be addressed: E-mail: malhi@illinois.edu, mxd60@psu.edu.

The decline of Native American populations after European contact has been linked to several factors including warfare, alterations in social structure, and an overwhelming introduction of European-borne pathogens<sup>1-3</sup>. Although the extent of the population decline remains contentious, European-borne epidemics may have disproportionately contributed to the phenomenon<sup>4,5</sup>. The debate has prompted researchers to explore the possibility of genetic susceptibility, where low genetic variation in HLA genes and immunologically naïve populations are linked to the exacerbated pathogen-associated mortality rates<sup>6-8</sup>. Assumptions of homogeneity among certain immune genes<sup>8</sup>, however, are based on surveys of living Native Americans who represent the surviving members of communities affected by European contact and colonization. Thus, they fail to consider immune-related genetic factors that may have existed before contact.

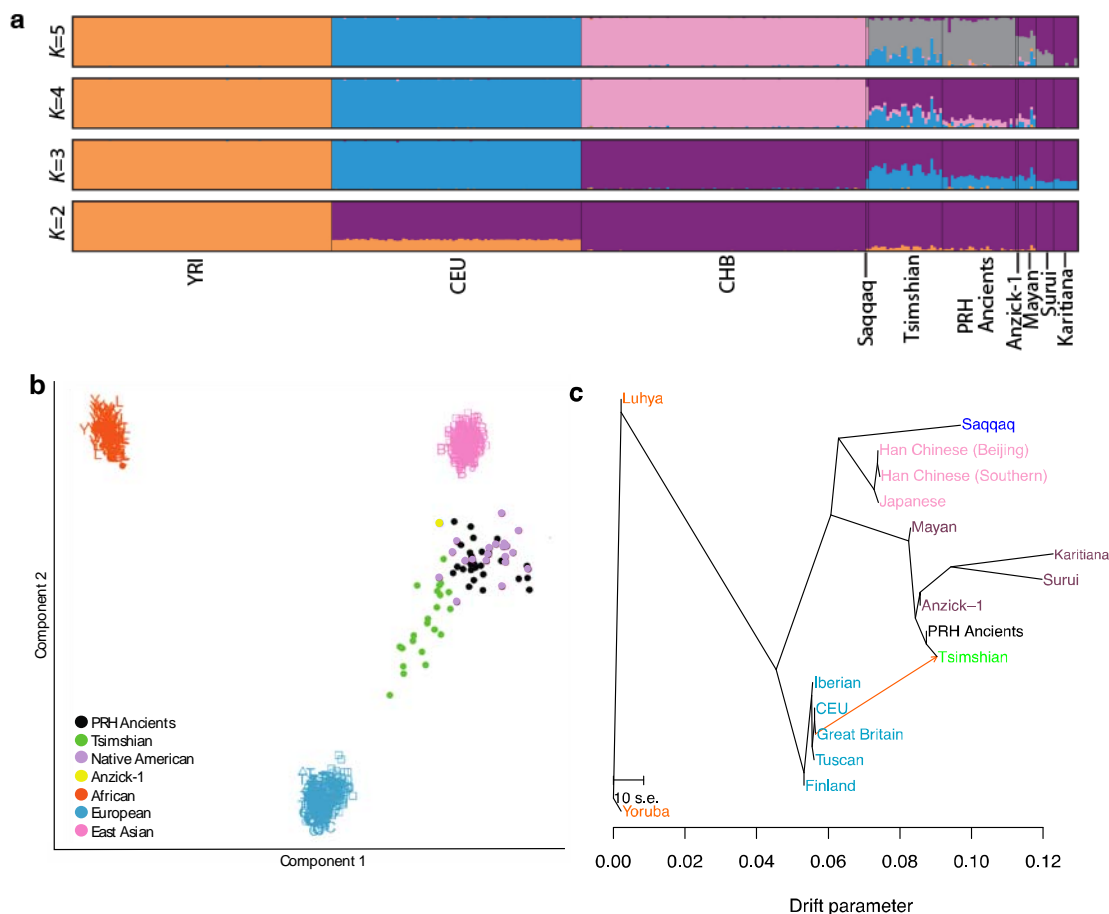
The immunological history of the indigenous peoples of the Americas is undoubtedly complex. As people entered the Americas and expanded into different regions, approximately 15,000 to 20,000 years before present (BP)<sup>9,10</sup>, groups encountered environments with varying ecologies and with relatively little gene flow from other continental populations until European contact<sup>11</sup>. We hypothesize that indigenous peoples adapted to local pathogens, resulting in long-lasting changes to immune-related loci. Ancient immune adaptations are suspected to have occurred throughout human history as populations spread into varying environments across the globe<sup>12</sup>. If the indigenous peoples of the Americas adapted to local pathogens, those adaptations would have proven useful in ancient times but not necessarily after European colonialists altered the environment with their pathogens, some of which may have been novel<sup>13-15</sup>. Existing genetic variation as a result of adaptation prior to European contact could thus have contributed to the indigenous population decline after European contact.

## Results

**Samples and sequencing.** To investigate possible immune-related genes under selection before European contact, we sequenced the exomes of ancient and modern First Nations individuals of the Prince Rupert Harbour (PRH) region of British Columbia, Canada (Supplementary Fig. 1). We then performed genomic scans for positive selection and functional characterization of genes exhibiting the strongest signals. Exomes of 25 modern individuals from two Coast Tsimshian communities, Metlakatla and Lax Kw'alaams (henceforth referred to as “Tsimshian”), were sequenced to a mean depth of 9.66x. The exomes of 25 ancient individuals from archaeological sites in the PRH region (henceforth referred to as “PRH Ancients”; Supplementary Fig. 1) were sequenced to a mean depth of 7.97x. Contamination estimates using the exome-wide data revealed a mean contamination of 0.94% with a 95% CI: 0.83-1.10<sup>16</sup> (Supplementary Table 3). All 25 ancient individuals exhibited patterns of C→T and G→A transitions consistent with deamination due to post-mortem DNA damage<sup>17,18</sup> (Supplementary Fig. 3). Mitochondrial haplogroups were determined for each ancient individual, all showing haplotypes previously identified in Native Americans<sup>19</sup> (Supplementary Table 1).

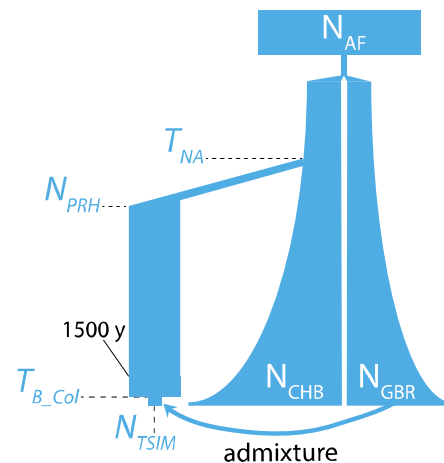
**Examining the genetic relationship between the ancient and modern individuals.** Before proceeding with selection scans we investigated the genetic relationship among the ancient and modern individuals to confirm continuity between the two groups. For these analyses, C/T and G/A polymorphic sites were removed to guard against biases resulting from DNA damage. Multidimensional scaling was performed to assess the genetic relationships of our samples to individuals from the 1,000 Genomes Project<sup>20</sup>, other Native American populations<sup>21</sup>, and two

ancient individuals from the Americas<sup>22,23</sup> (Fig. 1b). The analysis revealed an affinity among the PRH Ancients and the Tsimshian, with the Tsimshian drifting toward Europeans as expected from presumed European admixture<sup>24</sup>. We next employed *ADMIXTURE*<sup>25</sup> to separate our samples and other worldwide populations into clusters. This analysis (at  $K=5$  clusters) suggests that the Tsimshian are a mixture of ancestral components stemming from the PRH Ancients and Europeans (Fig. 1a). Next, the evolutionary relationship among our samples and other worldwide human populations was evaluated via *TreeMix*<sup>26</sup>. With a single migration event, the PRH Ancients exhibit minimal drift and appear ancestral to the Tsimshian with European admixture occurring between the two groups (Fig. 1c). These analyses, combined with local oral histories and evidence from archeology and mitogenomes<sup>19</sup>, allowed for the inference that the ancient and modern individuals represent a single population through time, which includes pre- and post-European contact.



**Figure 1 | Population structure and demographic model of the PRH Ancients and Tsimshian.** (a) *ADMIXTURE* analysis depicting ancestry proportions assuming the number of genetic components,  $K$ , is 2 to 5. The analyses included reference populations from the 1,000 Genomes Project Phase 2 dataset<sup>20</sup>, Native American populations sampled from the Karitiana, Surui, and Maya<sup>21</sup>, and two ancient samples from the Americas—the Saqqaq<sup>22</sup> and Anzick-1<sup>23</sup>. (b) Multi-dimensional scaling plot. The Native American (Surui, Mayan, and Karitiana) populations fall with the PRH Ancients and the Anzick-1 ancient sample from Montana. The modern Tsimshian fall along the gradient leading from the Native Americans and the Europeans, reflecting their admixed history with Europeans. (c) *TreeMix* graph with a single admixture event.

**Demographic model.** We also inferred the population history of the Tsimshian by taking into account the bottleneck that occurred after European contact<sup>1,27,28</sup> (Fig. 2, see Methods). Utilizing both ancient and modern exome-wide data, the demographic parameters were inferred utilizing the joint derived site frequency spectrum of potential synonymous sites with respect to human reference genome hg19. The best-fitting model suggests that a bottleneck occurred approximately 175 years BP (bootstrap 95% confidence interval: 125-225, Table 1) in the ancestors of the modern Tsimshian with an accompanying reduction in effective population size of 57%. The timing of the bottleneck coincides with the documented smallpox epidemics of the 19<sup>th</sup> Century and historical reports of large scale population declines<sup>29,30</sup>. A majority of the European admixture in the population likely occurred after the epidemics<sup>24,29</sup>.



**Figure 2 | Demographic model.** Best fitting model depicting the Tsimshian bottleneck after European contact and the subsequent admixture with Europeans. Fixed demographic parameters are from Gravel et al.<sup>31</sup> and admixture parameters are from Verdu et al.<sup>24</sup>. The parameters in blue were inferred with FastSimCoal2<sup>32</sup>. Population sizes and time splits are not shown to scale.

Parameter	Inferred Value	95% confidence interval
$N_{PRH}$ (Ancients)	13,975	7,532-15,178
$N_{TSIM}$ (Tsimshian)	6,006	4,124-8,048
$T_{NA}$ (CHB/PRH Ancients Split)	15,125	14,105-16,870
$T_{B\_Col}$ (European Contact Bottleneck)	175	125-225

**Table 1 | Parameters estimated for the model displayed in Figure 2 using FastSimCoal2<sup>32</sup>.** Population estimates are for diploid individuals. Time estimates are in years, assuming a generation time of 25 years.

**Scan for positive selection.** To safeguard against false-positive signals of positive selection due to the apparent admixture of Tsimshian individuals with Europeans, we performed an admixture correction (see Methods). The populations were scanned for selection signals, with and without correcting for admixture, utilizing the population branch statistic (PBS)<sup>33</sup>. PBS has proven effective in detecting positively-selected loci amongst high-altitude populations<sup>33,34</sup>. Twenty-five Han individuals from Beijing (CHB), part of the 1,000 Genomes Project<sup>20</sup>, served as the third comparative population. The statistic computes the amount of differentiation at a given locus along a branch leading to a specific population by comparing transformed  $F_{ST}$  values between

each pair of three populations. Figure 3a displays population-specific differentiation for the mean across the exome and for our top candidate gene (*HLA-DQA1*) discussed below.

The genes showing the most extreme and significant PBS values in the PRH Ancients represent strong candidates for positive selection, of which the top candidate, *HLA-DQA1*, is directly involved with immune function (Table 2). 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 3). To assess whether the selection signals were extreme relative to expectations under neutrality, the PBS scores were compared to the distribution of scores based on neutral simulations using our inferred demographic history (Supplementary Fig. 11). Variants from the top candidate with the most pronounced frequency changes were confirmed via Sanger sequencing in all ancient samples reporting data (Supplementary Fig. 10).

Gene	PBS	P-value	SNPs	Description
<i>HLA-DQA1</i>	0.275500845	0.01675	17	Major histocompatibility complex, class II, DQ alpha 1
<i>CELA3A</i>	0.194966437	0.03102	29	Basic leucine zipper nuclear factor 1

**Table 2. Genes with the strongest frequency changes in the PRH Ancient individuals.**

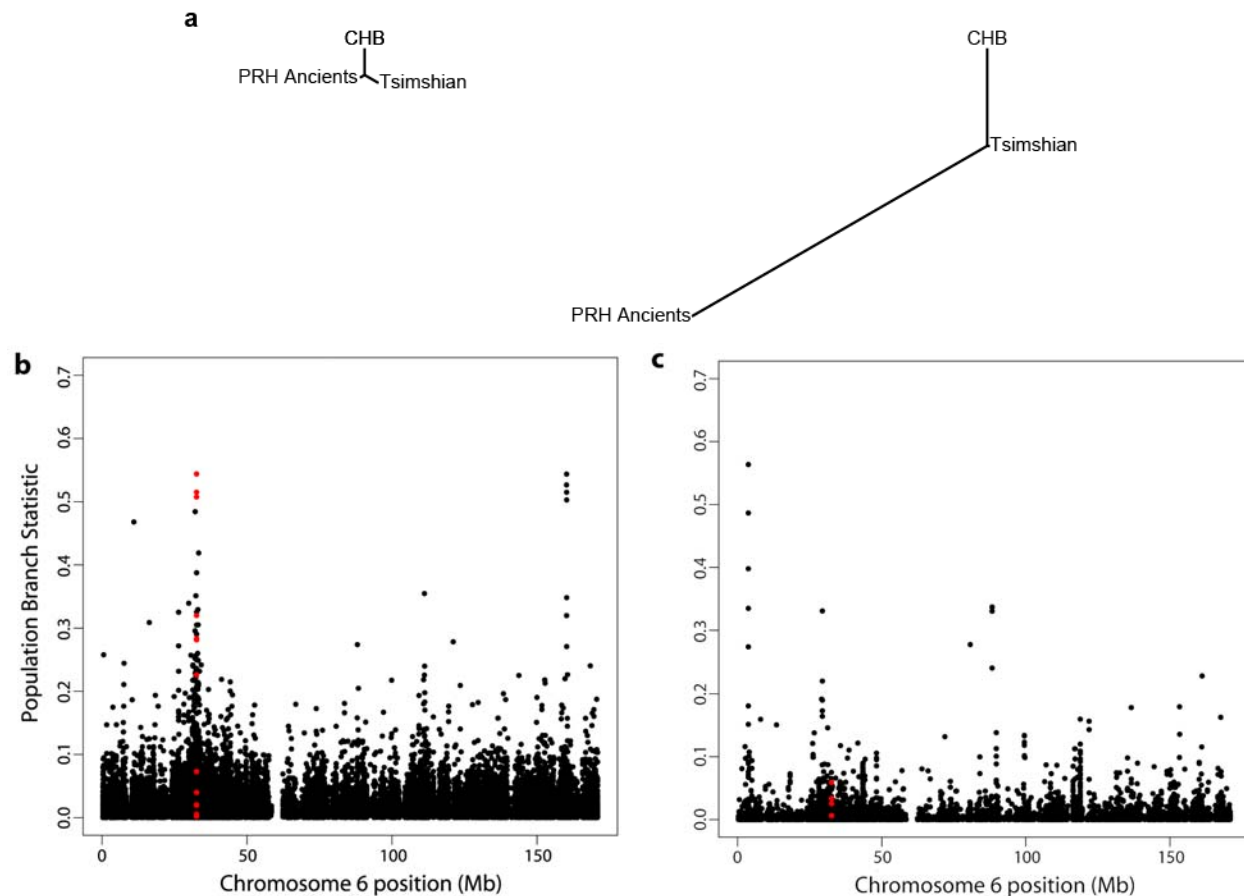
Ontology enrichment term	P value	FDR q value
Peptide antigen binding	$2.61 \times 10^{-8}$	$1.09 \times 10^{-4}$
MHC class II receptor activity	$8.11 \times 10^{-8}$	$1.69 \times 10^{-4}$
G-protein coupled receptor activity	$6.12 \times 10^{-6}$	$8.53 \times 10^{-3}$
Olfactory receptor activity	$7.91 \times 10^{-6}$	$8.26 \times 10^{-3}$
Antigen binding	$1.36 \times 10^{-5}$	$1.14 \times 10^{-2}$

**Table 3. Enriched gene ontologies for the PRH Ancient individuals derived from the PBS selection scan ranked list.** Calculated via GOrilla<sup>35</sup>.

**Relevance of the *HLA-DQA1* gene.** The most extreme PBS score belonged to the *HLA-DQA1* gene, 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 PRH Ancients (100%) and the Tsimshian (36%) falls in the 5' untranslated region (Table 4). This region may be indicative of selection acting on the regulation of the gene, as the associated alleles exhibit evidence of chromatin alterations and eQTL hits in a variety of cells—including Monocytes-CD14+ and Primary T helper 17 cells<sup>36</sup> (Supplementary Table 10). The chromosomal region where the gene is located also shows strong differentiation along the branch leading to the PRH Ancients (Fig. 3).

HLA-DQ is one of the three main types of MHC class II molecules, along with DR and DP, and is mainly expressed on antigen presenting cells<sup>37</sup>. 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<sup>38</sup>. The molecules are known to be highly polymorphic, mainly due to sequence differences corresponding to the binding domain of the molecule, which can impact binding affinities<sup>39</sup>. 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<sup>38</sup>. The polymorphic nature of these molecules across

*different* populations, however, would not explain the heightened differentiation in the PRH Ancients with respect to their presumed descendants, the Tsimshian.



**Figure 3 | Population branch statistic (PBS) of the *HLA-DQA1* gene.** (a) Trees based on exome-wide data, with the left tree showing a small branch length of the PRH Ancients relative to Tsimshian and CHB and the right tree showing strong differentiation along the PRH Ancients branch at the *HLA-DQA1* gene. (b,c) Manhattan plots of the PBS score as a function of SNP position on chromosome 6 for the PRH Ancients (b) and the modern Tsimshian (c). The region highlighted in red shows PBS scores for SNPs within 10kb of the *HLA-DQA1* gene.

**Haplotype structure and local ancestry of *HLA-DQA1*.** The top candidate for selection in the ancient population, *HLA-DQA1*, showed large allele frequency changes in the UTR5 region of the modern population (Table 4). Although there is a slight reference bias in the ancient samples due to mapping and possibly the design of our capture probes (Supplementary Fig. 14), the high frequency cannot be attributed to this feature since the derived alleles putatively under selection are for the *alternate* allele. To assess whether the frequency change was due to European admixture, we examined the haplotype structure among populations. To visualize the haplotypes in the *HLA-DQA1* region, we phased the ancient and modern samples using Beagle 4.1<sup>40</sup>. We took a randomly chosen haplotype from ancient sample PRH 125, and computed the number of pairwise differences to this haplotype for each haplotype in the modern and ancient samples as well as the Great Britain (GBR) samples from the 1,000 Genomes Phase 3 data<sup>41</sup>. We then ordered the haplotypes based on their number of pairwise differences to this arbitrarily chosen haplotype from sample PRH 125, and grouped them by population. Supplementary Figure 15a



shows very similar haplotypes between the ancient and modern individuals, while those of the European population are distinct.

We next explored the local ancestry of the *HLA-DQA1* gene. We used RFMix<sup>42</sup> to infer ancestry along chromosome 6 in the modern Tsimshian population. We utilized the PopPhased program, which tries to correct for phasing errors, and a window size of 0.2cM, four generations since the admixture event between the Tsimshian and Europeans, and 100 trees generated per random forest. For the reference panel, we used Phase 3 data from the 1,000 Genomes Project<sup>41</sup>. We used 25 individuals from the GBR (European panel), CHB (East Asian panel), and PEL (Native American Panel). The PEL chosen showed little to no admixture (see Methods). Supplementary Figure 15b indicates that only one haplotype could be attributed to European ancestry, while the remaining 49 are attributed to Native American ancestry.

Gene segment	SNP ID	Ancestral allele	Derived allele	Exonic function	PRH		
					Ancients frequency	Tsimshian frequency	CHB frequency
UTR5	rs9272426	A	G	N/A	1	0.37	0.46
UTR5	rs3207966	G	A	N/A	1	0.41	0.16
UTR5	rs3187964	G	C	N/A	0	0.55	0.49
UTR5	rs1047985	G	A	N/A	1	0.43	0.11
Exonic	rs1047989	C	A	nonsynonymous	1	0.55	0.51
Exonic	rs1047993	C	T	synonymous	1	0.42	0.12
Exonic	rs10093	G	C	synonymous	0.89	0.67	0.79
Exonic	rs1129765	T	C	synonymous	0.84	1	1
Exonic	rs1142323	A	G	nonsynonymous	0.77	0.27	0.14

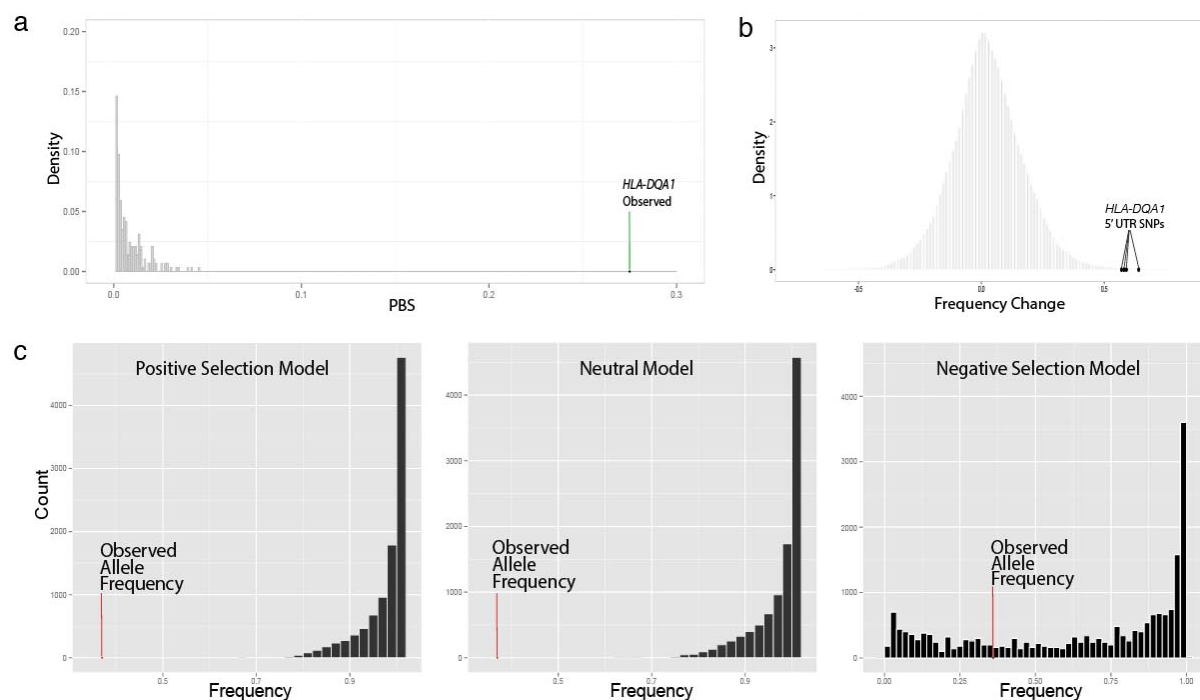
**Table 4 | Population frequencies for the *HLA-DQA1* SNPs on chromosome 6.** The population frequencies are for the derived allele with respect to the chimpanzee reference.

**Simulations of the *HLA-DQA1* allele trajectories.** To explore whether the allele frequencies differences between the two time periods could be explained by long-term balancing selection, drift, or changes in selection pressures, we conducted a series of simulations based on our demographic model. First, we examined whether long-term balancing selection under heterozygote advantage could explain our data. The parameters inferred from our demographic model were implemented in the forward-time simulator SLiM<sup>43</sup>. A *de novo* mutation was introduced 5 million years in the past (assuming a generation time of 25 years) that evolved under heterozygote advantage (per-generation selection coefficient  $s=0.1$ , and dominance parameter  $h=100$ ) until the present. The distribution of the resulting PBS scores can be seen contrasted with the observed data in Figure 4a. Because the distribution of the PBS scores under long-term balancing selection is shifted toward small values compared to neutrality (Supplemental Figure 11), the data are inconsistent with long-term balancing selection under heterozygote advantage.

To model evolutionary forces acting on the *HLA-DQA1* derived alleles after European contact, we chose the frequency of the allele showing the greatest change of 0.67. We utilized a simulation-based approach, described in detail in Supplementary Note 7, to evaluate models under positive, neutral, and negative selection. We also used the same approach to obtain estimates of the correlation between the time of environmental change ( $t$ ) and the selection coefficient ( $s$ ) (Supplementary Fig. 12). Figure 4c shows that neither a neutral or positive

selection scheme could fit our data, with none of the simulations reaching the observed frequency in the modern population. However, the negative selection model was compatible, where 26% of the simulations either reached or surpassed the observed frequency.

We also investigated if the observed allele frequency in the ancient population could be better explained by drift rather than selection. Using the same general method, we simulated the initial allele frequencies at the time Native Americans split off from East Asian populations by randomly sampling allele frequencies from a backward simulation conditioned on the modern CHB frequency. We then simulated allele frequencies at 60 generations ago—the time at which the ancient population was sampled. The resulting distribution in Supplementary Figure 7 shows that a neutral scenario is not a good fit for our data.



**Figure 4 | Selection scenarios before and after European Contact.** a) Distribution of PBS scores for long-term balancing selection (heterozygote advantage) simulations involving a genomic region of the same length as the *HLA-DQA1* transcript. Given the observed PBS score in the ancient population, this form of balancing selection is inconsistent with our data. (b) Empirical distribution of frequency changes for all derived alleles in the exome between the ancient and modern populations, showing the *HLA-DQA1* UTR5 SNPs as outliers. (c) Simulations of different selection schemes and the allele frequency of the *HLA-DQA1* SNPs observed in the modern Tsimshian. The allele frequency changes between the ancient and modern populations could not be explained by our demographic model (shift due to neutrality after European contact) or one that included only positive selection (no shift from positive selection after European contact). None of the simulations in which the originally positively-selected allele is neutral or still under positive selection after the time of European contact reached the observed allele frequency in the modern population. A model that incorporated negative selection after European contact was a better fit to our data, where ~26% of the simulations reached the observed frequency.

## Discussion

Our unique data set has allowed us to examine the demography of a single Native American population through three distinct time frames. We first examined the population from a time span



of 5,000 years leading to European contact. Selection scans on the ancient individuals from this period revealed a top candidate for positive selection, *HLA-DQA1*, giving the inference of an immune related adaptive event. We next inferred the severity of the population collapse after European contact, which correlates with historical population declines associated with regional smallpox epidemics<sup>44,45</sup>, as well as general estimates of Native American population declines based on mitochondrial DNA diversity<sup>27,46</sup>. During the contact period, previous long standing positive selection on the *HLA-DQA1* gene may also have been significant. The HLA-DQ receptor has been associated with a variety of colonization era infectious disease, including measles<sup>47,48</sup>, tuberculosis<sup>49,50</sup>, and with the adaptive immune response to the vaccinia virus, which is an attenuated form of smallpox<sup>51,52</sup>. Further studies are needed to investigate if the ancient alleles putatively under positive selection may pose a differential disease outcome with respect to European-borne pathogens, as well as their effect on downstream target genes.

However, when examining the population post-contact and into contemporary times, variants of the *HLA-DQA1* gene experience a marked frequency change. This change presents a more complex scenario when taking into account all three time frames. First, scans for positive selection in the modern Tsimshian, with and without correcting for European admixture, revealed no statistically significant selection on immune-related genes (Supplementary Tables 6 and 7). The gene ontology enrichment analyses also did not suggest a correlation with immune function (Supplementary Table 9). Second, demography alone was unable to explain the large frequency change in the *HLA-DQA1* alleles between the ancient and modern groups based on simulations (Fig. 4c). European admixture in the modern individuals also did not account for the frequency changes since the haplotypes in this region can be attributed to Native American ancestry (Supplementary Fig. 15). Furthermore, *HLA-DQA1* remained a top PBS hit in scans involving both a European admixture correction (Supplementary Table 4) and with an additional scan involving unadmixed Native American individuals from a different modern population (suggesting a regional adaptive event) (Supplementary Table 8; ranked fourth best candidate, with the top three functionally uncharacterized).

We therefore explored alternative explanations for the observed frequency change of the *HLA-DQA1* alleles in the time after contact. Since HLA genes have been previously postulated to be under balancing selection in humans<sup>53,54</sup>, we examined the possibility that long-term balancing selection could explain our data by simulating under a model of heterozygote advantage conditional on our inferred demographic model. We found that this specific type of balancing selection is a poor fit to the data, whereby the *HLA-DQA1* gene is still in an extreme outlier relative to the simulation results (Fig. 4a). Next, we used a forward simulation based approach to trace the *HLA-DQA1* allele trajectories under different selection models after the point of European contact. We found that simulations under our demographic model, which was modified to not include European admixture given our local ancestry results (Supplementary Fig. 15b), was insufficient to explain the frequency change in the modern population—with none of the 10<sup>4</sup> simulations reaching the observed frequency (Fig. 4c). However, upon applying a model of negative selection at the time of contact, we found that simulated allele frequencies were compatible with the observed frequencies in the modern population (Fig. 4c). Although we were unable to precisely identify the selection coefficient necessary to drive the allele frequency change (since the likelihood surface is relatively flat, Supplementary Fig. 12), it is likely that relatively strong negative selection occurred. Such strength would be expected under a time

frame of less than 7 generations and correlates with the high mortality rates associated with the regional smallpox epidemics of the 1800s, which reached upwards of 70%<sup>44,45</sup>.

The results presented here reveal an evolutionary history that spans thousands of years. The immune-related alleles that exhibit strong signals of positive selection in the ancient Native Americans from the Northwest Coast, likely correlate to an adaptation to pathogens that were present in the ancient environments of the region. Our results also suggest that these indigenous peoples may have experienced negative selection on the same immune-related genetic component after European contact and the ensuing population collapse. The shift may represent a form of balancing selection due to fluctuating environments<sup>55</sup>. This inference was only made possible through our examination of a single population through time, revealing nuanced demographic events and the utility of such studies. Furthermore, the evolutionary history detailed here helps to better understand the experiences of Native Americans with disease, in both ancient and colonial periods, by demonstrating a shift in immune-related selection pressures associated with the environmental impact of European contact.

## Methods

**Ethics approval and Community Engagement.** RSM and JSC visited the partnering First Nations prior to collecting samples for this study. During and after community visits and extensive consultation, a research protocol and informed consent documents—agreed upon by the indigenous communities and the non-indigenous researchers—was approved by the University of Illinois Institutional Review Board (#10538). RSM, JSC and JL visited the community annually during the study to report the latest results and continue to visit the First Nations to report on this and related studies.

**DNA Extraction and Library Preparation.** We prepared DNA extracts from 25 ancient individuals from the Prince Rupert region of British Columbia (Supplementary Table 1) and prepared DNA sequencing libraries in a clean room facility. The 25 modern DNA samples underwent similar procedures in a separate facility designated for modern DNA only (Supplementary Note 2).

**Exome Capture.** We captured 4 libraries from each ancient sample utilizing the Illumina TruSeq exome enrichment kit, with protocol modifications suited for ancient DNA (*SI Appendix*). The 4 captured libraries from each individual were pooled and sequenced (single-end) on one lane of the Illumina HiSeq 2000. Two libraries from each modern individual were enriched with either the Illumina TruSeq or Nextera exome enrichment kits (Supplementary Table 2) and pooled with an additional modern sample for single-end sequencing on the Illumina HiSeq 2000.

**Contamination Estimates.** To estimate contamination across the genome-wide data, we used the ContEst tool<sup>16</sup>. The tool 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<sup>16</sup>. HapMap\_3.3 global population frequencies for each SNP, mapped to b37, were used for the estimates. All ancient samples demonstrated contamination below 1%, except for PRH Ancient 163. The estimates are shown on Supplementary Table 3.

**Variant Discovery.** Reads below a read length of 35 were filtered out before mapping to hg19. For analyses requiring genotype calls (e.g., *TreeMix*, *ADMIXTURE*), SAMtools-1.1<sup>56</sup> was utilized with a minimum mapping quality of 30, a minimum base quality of 20, a minimum read depth of 6, and a max read depth of 80. Sites were also filtered for violation of a one-tailed test for Hardy-Weinberg Equilibrium at a  $p$ -value  $< 10^{-4}$ <sup>57</sup>. Due to the low average read depth of the PRH Ancients, genotypes were not called directly for the selection scan. Instead, the program suite ANGSD<sup>58</sup> 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  $10^{-6}$ .

**PBS Selection Scan.** To detect regions under positive selection in both the PRH Ancient and the Tsimshian, the population branch statistics (PBS)<sup>33</sup> was utilized. The population branch statistics (PBS)<sup>33</sup> has proven powerful in detecting hypoxia adaptation in high altitude populations<sup>33,34</sup>. It employs a set of three populations (call them  $X$ ,  $Y$ , and  $Z$ ), and assumes that they have the rooted relationship  $((X,Y),Z)$ . In actuality, the calculation for the PBS does not require a rooted tree, and so their specific rooted relationship does not matter.

An analogous statistic can be calculated for populations  $Y$  and  $Z$ . In this study, we are concerned with the situation in which  $X$  is the PRH Ancients,  $Y$  is the modern Tsimshian, and  $Z$  is the Han Chinese (CHB from the 1,000 Genomes Project). We therefore are interested in computing:

$$PBS_{PRH} = \frac{T_{PRH,Tsimshian} + T_{PRH,CHB} - T_{Tsimshian,CHB}}{2}.$$

Because the *ADMIXTURE* and *TreeMix* analyses (Figs 1a and 1c) indicate a likely admixture event between the modern Tsimshian and Europeans, the allele frequencies in the modern Tsimshian were corrected for admixture using the method described by Huerta-Sánchez et al.<sup>34</sup>. Let  $f_{Tsimshian}^*$  and  $f_{Tsimshian}$  represent the allele frequency at a locus in the Tsimshian population pre and post admixture, respectively. Further, assuming that we use Europeans as a proxy, we let  $f_{European}$  be the allele frequency at the same locus in a reference European population (we used the Great Britain GBR population from the 1,000 Genomes Project). Assuming that the proportion of ancestry derived from Europeans at the locus is  $\alpha$ , under a model of instantaneous admixture, the allele frequency in the Tsimshian post admixture would be

$$f_{Tsimshian} = (1 - \alpha)f_{Tsimshian}^* + \alpha f_{European}.$$

Rearranging, we can solve for the allele frequency prior to admixture as

$$f_{Tsimshian}^* = \frac{f_{Tsimshian} - \alpha f_{European}}{1 - \alpha}.$$

We estimated  $\alpha$  at each locus by choosing an  $\alpha$  value that minimized the  $F_{ST}$  between the admixture-corrected Tsimshian and the Han Chinese (CHB) outgroup population. We used the admixture-corrected allele frequencies for the scans for positive selection.

We used ANGSD<sup>58</sup> to compute allele frequencies for the modern Tsimshian, the PRH Ancient, the Han Chinese (CHB), and the Great Britain (GBR) populations directly from the raw sequencing reads, accounting for the uncertainty in genotype calling. Allele frequencies were

based on 25 individuals from each population. We required that reads had a map quality of at least 30 and each nucleotide had a quality of at least 20. We also only called allele frequencies at sites in which data from at least five individuals was not completely missing. To additionally guard against post-mortem deamination, we trimmed the first and last five nucleotides of each read in the PRH Ancient samples. The allele frequencies in the modern Tsimshian were subsequently corrected for potential European ancestry (see procedure in the directly preceding paragraph).

We performed two scans for positive selection. The first was a per-gene scan, in which we calculated PBS for the PRH Ancient population ( $PBS_{PRH}$ ) using all data at given gene. Gene annotations were derived from RefSeq, utilizing the longest transcript for a given gene. The transcript length was taken as the transcription start to the transcription stop, and included both introns and exons.  $F_{ST}$  between each pair of populations was calculated using all SNPs that fell between the transcription start and stop of the gene, as well as 10 kilobases upstream of the transcription start and 10 kilobases downstream of the transcription stop (similar to how it was performed in Huerta-Sánchez et al.<sup>34</sup>). We then ranked each gene in the genome with decreasing PBS. The top two candidates are displayed in Table 2.

The second scan was a per-SNP scan, in which we calculated PBS for the PRH Ancient population ( $PBS_{PRH}$ ) and the modern Tsimshian population ( $PBS_{Tsimshian}$ ) at each SNP. That is,  $F_{ST}$  between each pair of populations was calculated for a given SNP, and this set of  $F_{ST}$  values was used to calculate PBS for that SNP. We then created Manhattan plots, and highlight chromosome 6 in Figure 3.

We also performed an analogous scan in which we substituted 25 mostly unadmixed Peruvian samples from the 1,000 Genomes Project Phase 3<sup>41</sup> for the 25 modern Tsimshian individuals. We identified the individuals showing little to no admixture by running *ADMIXTURE*<sup>25</sup>. The individuals identified and used for this and subsequent analyses were: HG01572, HG01923, HG01926, HG01927, HG01941, HG01951, HG01953, HG01954, HG02008, HG02102, HG02105, HG02146, HG02147, HG02150, HG02259, HG02260, HG02266, HG02271, HG02272, HG02275, HG02278, HG02291, HG02292, HG02304, HG02348. Because the individuals appeared mostly unadmixed, we did not correct allele frequencies for admixture. Results from this scan are highlighted in Supplementary Table 10. Using his different reference sister Native American population still suggests that *HLA-DQA1* is a reasonable candidate, as it is ranked fourth in the scan—with the three genes ranked above it devoid of functional characterization—using Peruvians rather than modern Tsimshian as a reference sister population.

**Demographic History Model.** Parameters for the demographic model (Fig. 2) were inferred with FastSimCoal2<sup>32</sup>. The fixed parameters were implemented from Gravel et al.<sup>31</sup> and were as follows: out of Africa bottleneck ( $N=1861$ ,  $T=51\text{kya}$ )<sup>31</sup>, split between the CHB and GBR (serving as the ghost population) ( $N_{GBR}=1032$ ,  $N_{CHB}=550$ ;  $T=23\text{kya}$ )<sup>31</sup>. Admixture between the GBR and Tsimshian ( $T=100$  years, admixture fraction=0.33) were taken from Verdu et al.<sup>24</sup>. One hundred optimizations were run for the inferred values, taking the best likelihood parameters from each of the 100 sets. The data was simulated with an effective sequence length of 7.4 Mb and per-base per-generation mutation and recombination rates of  $2.5 \times 10^{-8}$ . The optimizations utilized joint derived site frequency spectra (SFS) for the CHB, PRH Ancients, and

Tsimshian. The European population (Great Britain denoted by GBR) served as a ghost population in the model. This SFS contained 7.4 Mb of monomorphic and polymorphic sites based on hg19 potential synonymous sites, where data was reported for each individual. A parametric bootstrapping approach was used to construct the 95% confidence intervals. The inferred parameters and confidence intervals are listed in Table 1 and the model residuals are depicted in Supplementary Figure 13.



## References

1. Thornton, R. Aboriginal North American Population and Rates of Decline, ca. a.d. 1500-1901. *Current Anthropology* **38**, 310–315 (1997).
2. Thornton, R. *American Indian Holocaust and Survival*. (University of Oklahoma Press, 1987).
3. Patterson, K. B. & Runge, T. Smallpox and the Native American. *The American journal of the medical sciences* **323**, 216 (2002).
4. Boyd, R. T. *The Coming of the Spirit of Pestilence*. (University of Washington Press, 1999).
5. Dobyns, H. F. Disease transfer at contact. *Annual Review of Anthropology* **22**, 273–291 (1993).
6. Motulsky, A. G. Metabolic polymorphisms and the role of infectious diseases in human evolution. *Human Biology* **61**, 835–869 (1989).
7. Black, F. L. Infectious diseases in primitive societies. *Science* **187**, 515–518 (1975).
8. Black, F. L. Why did they die? *Science* **258**, 1739–1740 (1992).
9. Tamm, E. *et al.* Beringian Standstill and Spread of Native American Founders. *PLoS One* **2**, e829 (2007).
10. Gravel, S. *et al.* Reconstructing Native American Migrations from Whole-Genome and Whole-Exome Data. *PLoS Genetics* **9**, e1004023 (2013).
11. Fagan, B. M. *Ancient North America*. (Thames & Hudson, 2005).
12. Fumagalli, M. *et al.* Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLoS Genetics* **7**, e1002355 (2011).
13. Roberts, L. Disease and death in the New World. *Science* **246**, 1245–1247 (1989).
14. Merbs, C. F. A new world of infectious disease. *Am. J. Phys. Anthropol.* **35**, 3–42 (1992).
15. Riley, J. C. Smallpox and American Indians revisited. *J Hist Med Allied Sci* **65**, 445–477 (2010).
16. Cibulskis, K. *et al.* ContEst: estimating cross-contamination of human samples in next-generation sequencing data. *Bioinformatics* **27**, 2601–2602 (2011).
17. Briggs, A. W. *et al.* Patterns of damage in genomic DNA sequences from a Neandertal. *Proceedings of the National Academy of Sciences* **104**, 14616–14621 (2007).
18. Seguin-Orlando, A. *et al.* Ligation Bias in Illumina Next-Generation DNA Libraries: Implications for Sequencing Ancient Genomes. *PLoS One* **8**, e78575 (2013).
19. Cui, Y. *et al.* Ancient DNA Analysis of Mid-Holocene Individuals from the Northwest Coast of North America Reveals Different Evolutionary Paths for Mitogenomes. *PLoS One* **8**, e66948 (2013).
20. 1000 Genomes Project Consortium *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
21. Szpiech, Z. A. *et al.* Long Runs of Homozygosity Are Enriched for Deleterious Variation. *The American Journal of Human Genetics* **93**, 90–102 (2013).
22. Rasmussen, M. *et al.* Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757–762 (2010).
23. Rasmussen, M. *et al.* The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* **506**, 225–229 (2014).
24. Verdu, P. *et al.* Patterns of Admixture and Population Structure in Native Populations of Northwest North America. *PLoS Genetics* **10**, e1004530 (2014).
25. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* **19**, 1655–1664 (2009).
26. Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics* **8**, e1002967 (2012).
27. O'Fallon, B. D. & Fehren-Schmitz, L. 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 (2011).
28. Larsen, C. S. In the wake of Columbus: Native population biology in the postcontact Americas. *Am. J. Phys. Anthropol.* **37**, 109–154 (1994).
29. Sequin, M. & Halpin, M. in *Handbook of North American Indians* (ed. Sturtevant, W. C.) **7**, 267–284 (1990).
30. Boyd, R. T. in *Handbook of North American Indians* (ed. Sturtevant, W. C.) **7**, 135–148 (Smithsonian Institution, 1990).
31. Gravel, S. *et al.* Demographic history and rare allele sharing among human populations. *Proceedings of the National Academy of Sciences* **108**, 11983–11988 (2011).



32. Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C. & Foll, M. Robust Demographic Inference from Genomic and SNP Data. *PLoS Genetics* **9**, e1003905 (2013).
33. Yi, X. *et al.* Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* **329**, 75–78 (2010).
34. Huerta-Sanchez, E. *et al.* Genetic Signatures Reveal High-Altitude Adaptation in a Set of Ethiopian Populations. *Molecular Biology and Evolution* **30**, 1877–1888 (2013).
35. Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48 (2009).
36. Consortium, T. G. *et al.* The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
37. Jones, E. Y., Fugger, L., Strominger, J. L. & Siebold, C. MHC class II proteins and disease: a structural perspective. *Nature Reviews Immunology* **6**, 271–282 (2006).
38. Roche, P. A. & Furuta, K. The ins and outs of MHC class II-mediated antigen processing and presentation. *Nature Reviews Immunology* **15**, 203–216 (2015).
39. Wang, P. *et al.* A Systematic Assessment of MHC Class II Peptide Binding Predictions and Evaluation of a Consensus Approach. *PLoS Computational Biology* **4**, e1000048 (2008).
40. Browning, S. R. & Browning, B. L. Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. *The American Journal of Human Genetics* **81**, 1084–1097 (2007).
41. Altshuler, D. M. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
42. Maples, B. K., Gravel, S., Kenny, E. E. & Bustamante, C. D. RFMix: A Discriminative Modeling Approach for Rapid and Robust Local-Ancestry Inference. *The American Journal of Human Genetics* **93**, 278–288 (2013).
43. Messer, P. W. SLiM: Simulating Evolution with Selection and Linkage. *Genetics* **194**, 1037–1039 (2013).
44. Novembre, J., Galvani, A. P. & Slatkin, M. The Geographic Spread of the CCR5  $\Delta$ 32 HIV-Resistance Allele. *PLoS Biol* **3**, e339–9 (2005).
45. Boyd, R. Demographic History 1774–1874, in *Handbook of North American Indians*, Volume 7, 135–148 (Smithsonian Institution, 1990).
46. Llamas, B. *et al.* Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. *Science Advances* **2**, e1501385–e1501385 (2016).
47. Ovsyannikova, I. G., Vierkant, R. A. & Poland, G. A. Importance of HLA-DQ and HLA-DP polymorphisms in cytokine responses to naturally processed HLA-DR-derived measles virus peptides. *Vaccine* **24**, 5381–5389 (2006).
48. Moss, W. J. & Griffin, D. E. Global measles elimination. *Nat Rev Micro* **4**, 900–908 (2006).
49. Kim, H. S. *et al.* 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 (2005).
50. Delgado, J. C., Baena, A., Thim, S. & Goldfeld, A. E. 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 (2006).
51. Ovsyannikova, I. G., Vierkant, R. A., Pankratz, V. S., Jacobson, R. M. & Poland, G. A. Human Leukocyte Antigen Genotypes in the Genetic Control of Adaptive Immune Responses to Smallpox Vaccine. *Journal of Infectious Diseases* **203**, 1546–1555 (2011).
52. Ovsyannikova, I. G., Pankratz, V. S., Salk, H. M., Kennedy, R. B. & Poland, G. A. HLA alleles associated with the adaptive immune response to smallpox vaccine: a replication study. *Hum Genet* **133**, 1083–1092 (2014).
53. Loisel, D. A., Rockman, M. V., Wray, G. A., Altmann, J. & Alberts, S. C. Ancient polymorphism and functional variation in the primate MHC-DQA1 5' cis-regulatory region. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 16331–16336 (2006).
54. Solberg, O. D. *et al.* Balancing selection and heterogeneity across the classical human leukocyte antigen loci: A meta-analytic review of 497 population studies. *Hum. Immunol.* **69**, 443–464 (2008).
55. Huerta-Sanchez, E., Durrett, R. & Bustamante, C. D. Population Genetics of Polymorphism and Divergence Under Fluctuating Selection. *Genetics* **178**, 325–337 (2008).
56. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
57. Wigginton, J., Cutler, D. & Abecasis, G. A note on exact tests of Hardy-Weinberg equilibrium. *The American Journal of Human Genetics* **76**, 887–893 (2005).

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

## Acknowledgments

This project was made possible through the active collaboration of the Lax Kw'alaams and Metlakatla First Nations. We would also like to thank Jun Li for furnishing comparative exome data from contemporary Native American populations. The research was funded by the National Science Foundation (#BCS-1413551) and by the Office of the Vice Chancellor of Research, University of Illinois at Urbana-Champaign, by the Canadian Museum of History, Gatineau, Quebec, Canada, and by Pennsylvania State University startup funds. Portions of this research were conducted with the Advanced CyberInfrastructure computational resources provided by The Institute for CyberScience at Pennsylvania State University.

**Author Contributions:** Conceived and designed the study RSM, JL, MD. Performed the experiments: JL and MR. Analyzed the data: JL, MD, EHS, SN. Contributed reagents/materials/analysis tools: RSM, JL, MD, EW. Wrote the paper: JL, RSM, MD, JSC with contributions from all authors. Community engagement: RSM, JSC, BP, JM, JL. Discussed and interpreted results: JL, MD, RSM, EHS.

## Additional information

**Accession codes:** The ancient data have NCBI Short Read Archive accession no. 288803. The data from modern individuals are available via a data access agreement with RSM at the University of Illinois.

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/naturecommunications](http://www.nature.com/naturecommunications).