EVOLUTION OF THE ABO BLOOD GROUP LOCUS IN PRE-COLUMBIAN NATIVE AMERICANS

By

FERNANDO ALBERTO VILLANEA

A thesis submitted in partial fulfillment of
the requirements for the degree of

MASTER OF ARTS IN ANTHROPOLOGY

WASHINGTON STATE UNIVERSITY
Department of Anthropology

MAY 2010

To the Faculty of Washington State University:

The members of the Committee appointed to examine the thesis of FERNANDO ALBERTO VILLANEA find it satisfactory and recommend that it be accepted.

_____

Brian M. Kemp, Ph.D., Chair

_____

Jeremiah Busch, Ph.D.

_____

Tim A. Kohler, Ph.D.

EVOLUTION OF THE ABO BLOOD GROUP LOCUS IN PRE-COLUMBIAN NATIVE AMERICANS

Abstract
by Fernando A. Villanea, M.A.
Washington State University
May 2010

Chair: Brian M. Kemp

Among all studied genes in humans, the locus coding for ABO blood group alleles is the single best studied genetic trait. This makes this gene a prime marker to address Native American population history. Across loci, Native Americans have the lowest genetic diversity of all continental populations. While some Native American populations exhibit A, B and O alleles, such as some Eskimo groups, and some North American populations exhibit A and O alleles, all other Native American populations from North, Central, and South America are nearly fixed for the O allele.

Contrastingly, most other human populations outside of America seem to always maintain all three alleles and even orbit around similar allelic frequencies. This has been attributed to frequency dependent selection, mediated by our immune response to bacteria and virus, pressure which is consistent across the world. The high variance in ABO allele frequencies in Native Americans stands as an oddity.

In this study, four hypotheses were tested to explain the low overall Native American diversity and the high within Native American variance of ABO allele frequencies: a) Founder's effect in the original population, b) Post-European contact reduction of diversity, c) Post-European contact natural selection and d) High population structuring following subsequent migrations into the continent.

To address the possibility of ABO allele frequencies changing after European contact, ABO allele diversity data was recovered from pre-contact Muwekma-Ohlone samples. This ancient population, referenced as CA SCL-38, exhibits similar mitochondrial haplogroup and ABO allelic diversity as contemporary Native American populations from the same geographic area. Additionally a model was created to simulate loss of allelic diversity under the effect of genetic drift as a consequence of a reduced effective population size. Modeled data sets a lower limit to the effective population size a group would require in order to maintain diversity.

Contrasted against ABO diversity data from extant populations, my results are consistent with ABO diversity being lost as a result of isolation as populations migrated deeper into the continent and not as a result of a bottleneck in the original founder population.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Background

Anthropologists have always been interested in recording human variation, as marked by their early attempts at categorizing humans into hierarchies, to post-20th century efforts to understand the distribution of genetic variation. In the last century, the single most widely studied human genetic trait has been ABO blood types (Molnar, 2002). While early interest in blood typing individuals worldwide aimed to evaluate hierarchical classifications, this marker has remarkable importance to the medical field. Without a simple system of blood group identification successful blood transfusions, tissue and organ transplants would be impossible (Brace, 2005).

Blood transfusions were performed before the 20th century, but met with limited success (Molnar, 2002). The existence of a blood trait that varies between humans was first reported by Landsteiner (1901), who earned a Nobel Prize for his work on the ABO system in 1930. It was Landsteiner who dubbed the four phenotypes A, B, AB and O (Mourant et al., 1976). ABO blood phenotypes were so well recorded that this system provided some of the first evidence of Mendelian transmission in humans (Daniels, 2002; Loring, 2005; Molnar, 2002), a critical piece of evidence that demonstrated that inheritance among *Homo sapiens* is no different than that recorded for the common pea (*Pisum sativum*) (Mendel, 1866). Over the last century millions of people around the world have had their blood types recorded (Molnar, 2002).

Yet, in light of the medical importance of this genetic locus for tissue transplant by medical investigators, it is surprising how very little attention has been focused on the evolutionary forces that explain the modern distribution of ABO alleles. In the last decade the ABO system has been discovered to be among the most polymorphic genes in humans with

around 100 (mostly rare) alleles known at the sequence level. More is known about the distribution of ABO alleles than for practically any other human biological trait (Yip, 2002).

The ABO gene

The ABO gene is located on chromosome 9 and extending approximately 20,000 base pairs between positions 135,120,384-135,140,451 (Kent et al., 2002). Functionally, the ABO gene codes for a specific glycosyltransferase enzyme. The A and B allele glycosyltranferases transform the precursor H antigen (which lacks a terminal sugar) by adding N-acetyl-galactosamine to produce the A antigen and D-galactose to produce the B antigen. The ABO phenotype is thus determined codominantly by specific transferases coded by different alleles in the ABO locus, also with different efficiencies of H antigen transformation. Heterozygotes exhibit both types of antigens on their red blood cells. An individual homozygote for any O alleles have only H antigens, as O alleles lack glycosyltransferase activity and thus terminal sugars are not added to the precursor antigen H (Sharon and Fibach, 1991) (Table 1).

Table 1 Phenotypes, genotypes and ratios of cell membrane antigens for the common ABO alleles -A1, A2, B, O- (from Sharon and Fibach 1991).

Phenotype	Genotype	Cell membrane antigens	A/H antigen ratio	B/H antigen ratio
A	A1A1	A,A	200	0
A	A1A2	A,A	31	0
A	A1O	A,H	3.1	0
A	A2O	A,H	1.1	0
B	BB	B,B	0	37
B	BO	B,H	0	1.9
AB	A1B	A,B	159	63
AB	A2B	A,B	39	34
O	OO	H,H	0	0

Evolution of the human ABO polymorphism

The ABO allele “A101” is likely the ancestral human allele, probably predating the appearance of our species (Calafell et al., 2008). Around 2.5 mya, a second allele “O01” evolved from A101

by a deletion in position 261 of Exon 6. This mutation "Δ261" creates a premature stop codon resulting in a truncated protein lacking glycosyltransferase function (Gagneux and Varki, 1999). However, O alleles are not deleterious. A second O allele, "O02" evolved from A101 around 1.5 mya by a second, independent deletion at position 261. Phylogenies based on the sequences of ABO alleles confirm that the O01 and O02 alleles arose via convergence. Both O alleles were maintained in the ancestral hominid population for about another million years before the evolution of "O09", a third convergent O allele that also carries the Δ261 mutation.

The first functional variant, allele "A201" evolved approximately 260,000 YBP at about the same time the first B allele appeared. This ancestral B allele is thought to have recombined with A201 at some point, giving rise to the oldest recognized B allele known from human samples, "B101". Because the original B allele has since been lost from human populations, its phylogenetic position is not clear (Calafell et al., 2008; Yamamoto et al., 1990; Yip, 2002) (Fig 1). In addition to the two O alleles mentioned above, it is believed that there are at least three additional O alleles that evolved via convergence. A summary of common alleles can be found in Table 2 (Yip, 2002).

Table 2 Informative sequence, protein and functional differences of common ABO alleles.

Allele	Sequence	Protein	Allele (co	Transferase	Antigen	Associated Antib	Population
A101	reference	reference	A101	ATransferase	A (terminal N-acetylgalactosamine)	Anti-B	Global
A201	nt 467	Pro156Leu	A201	ATransferase	A (terminal N-acetylgalactosamine)	Anti-B	Global
B101	nt 297, 526, 657, 703, 796, 803, 930	Arg176Gly, Gly235Ser, Leu266Met, Gly268Ala	B101	BTransferase	B (D-galactose)	Anti-A	Global
O01	G261-	Stop codon at n352–354	O01	Truncates Atransferase	H (precursor)	Anti-A, Anti-B	Global
O02 (O1v)	G261-, nt 106, 188, 189, 220, 297, 646, 681, 771, 829	Stop codon at n352–354, Val36Phe, Arg63His, Pro74Ser	O02	Truncates Atransferase	H (precursor)	Anti-A, Anti-B	Global
O0210	G261-, G542A	Stop codon at n352–354	O0210	Truncates Atransferase	H (precursor)	Anti-A, Anti-B	Native Americans
O0303	nt 53, 220, 526, 802	Gly268Arg	O0303	Inactivates antibody-sugar binding site	H (precursor)	Anti-A, Anti-B	Caucasians
O0304	800insG, 1060delC	Non-sense frame shift	O0304	Inactivates Atransferase	H (precursor)	Anti-A, Anti-B	Rare

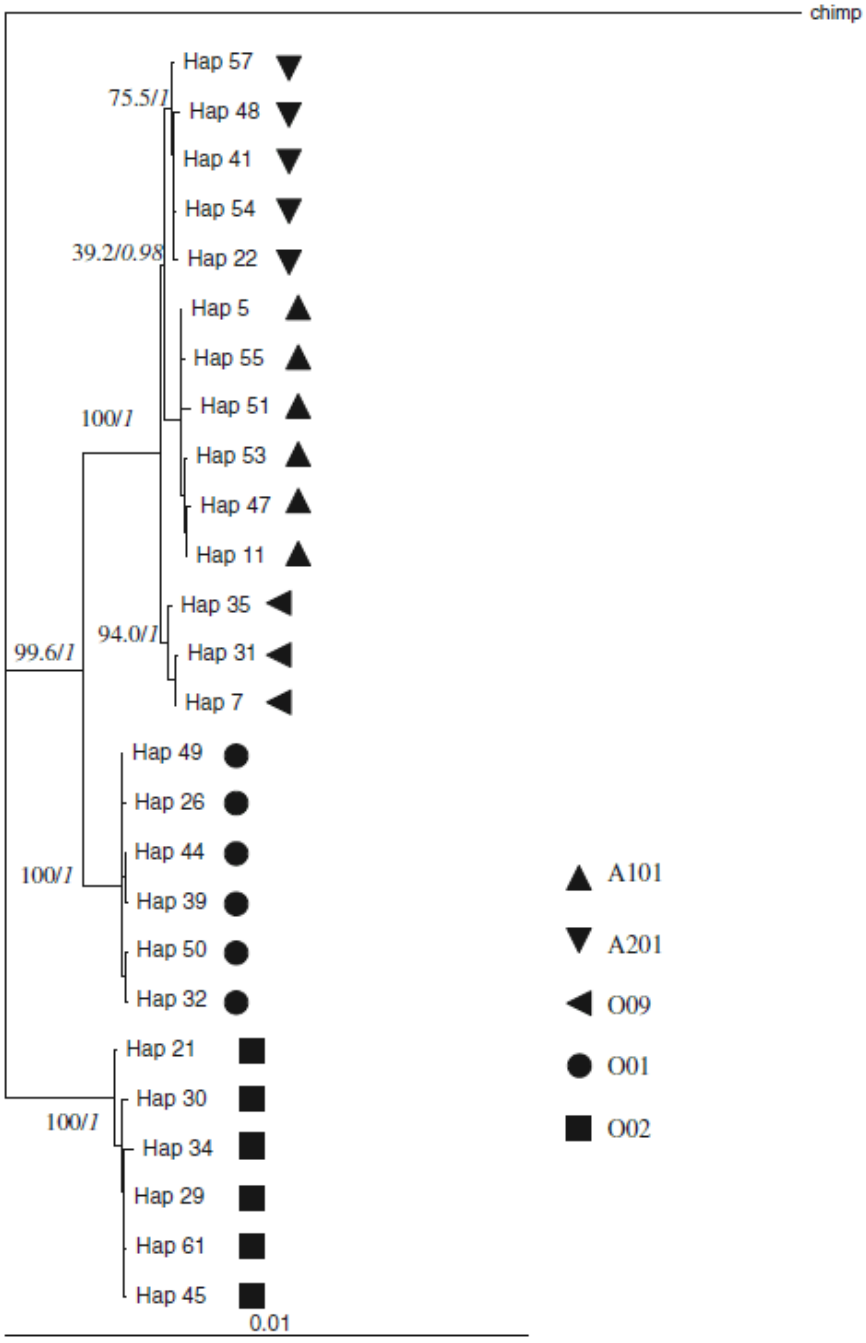

Figure 1 Maximum-likelihood (ML) tree of the presumably non-recombinant ABO haplotypes in the Seattle human SNP database. Lineages are indicated by symbols. Figures indicate percent bootstrap support of the ML tree for main nodes, and, after the *slash*, the posterior probabilities obtained with a Bayesian model (Figure 4 from Calafell et al., 2008).

Although the A101 is the oldest allele, currently the O01 and O02 alleles are the most common (Table 3 and Fig 2) and are widely distributed. With 40 known O alleles (from at least five independent evolutionary events) (Cavalli-Sforza et al., 1992; Yip, 2002), the maintenance of alleles with no glycosyltransferase activity (a potential drawback) over such a long evolutionary time (especially considering the enormous effect of drift in human prehistory) suggests that negative frequency dependent selection has favored the observed patterns of high heterozygosity. In other words the fitness of one individual's genotype is directly related to the presence and abundance of the other genotypes. Under this form of selection, the least common phenotype always has the highest fitness. Moreover, the fact that frequencies of O alleles are relatively homogeneous among contemporary human populations may also be interpreted as the result of geographically homogeneous selective pressures, a surprising phenomenon given the wide geographical and ecological distribution of humans (Gagneux and Varki, 1999; Saitou and Yamamoto, 1997; Seymour et al., 2004).

Table 3 Frequencies for the human ABO alleles in a global sample (from Cavalli-Sforza et al., 1992).

Region	A	B	O
Europe	0.210	0.080	0.710
North Africa	0.220	0.120	0.660
Sub-Saharan Africa	0.160	0.130	0.710
Near East	0.220	0.140	0.640
Central Asia	0.190	0.200	0.610
India	0.190	0.240	0.570
Northern Asia	0.201	0.140	0.659
East Asia	0.270	0.180	0.550
Southeast Asia	0.170	0.190	0.640
Pacific Islands	0.221	0.090	0.689
New Guinea	0.210	0.130	0.660
Australia	0.230	0.020	0.750
North America	0.115	0.030	0.855
South America	0.007	0.004	0.989
Average	0.187	0.121	0.692
S. D.	0.063	0.070	0.114
Variance	0.004	0.005	0.013

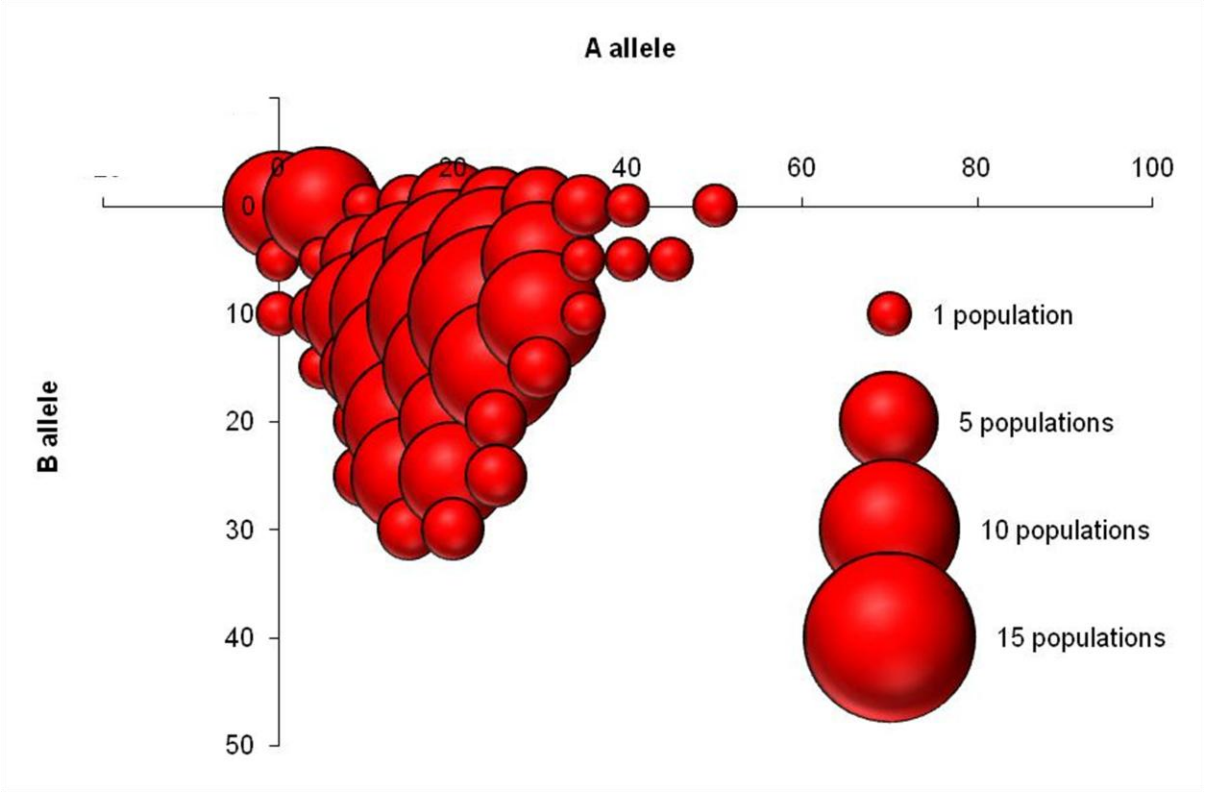

Figure 2 Distribution of A alleles and B alleles frequencies in a sample of 217 of the world's populations (Re-drawn from Molnar, 2002, Figure 3-5). Bubble size represents a number of populations that fall under those frequencies of A and B allele. Notice how the majority of populations clump around A and B frequencies of 20%.

ABO system and Immune response

Related antigen-antibody systems to ABO, such as the Old World primate-specific anti-a-galactosyl antibody (anti-Gal), mediate recognition of alien cells and function to detect cross species viruses (Rother and Squinto, 1996). Similarly, the A and B antigens-antibodies mediate self-recognition of cells. Antibodies of individuals with blood group A present an immune reaction to B antigens and vice versa. Individuals of blood group O carry anti-A and anti-B antibodies, whereas those of blood group AB carry none. This specificity to non-self antigens is probably mediated by spontaneously forming antibodies (NAbs), although it is contested whether antibody production responds to external antigens encountered early in development (Preece et al., 2002). While immune reaction is medically vital for successful blood and organ transfusions,

evolutionarily the likely function of the system is to recognize pathogens that carry non-self antigens of similar structure to that of the A and/or B antigens, genetically coded or acquired from previous hosts, allowing for prompt recognition (Preece et al., 2002).

The H antigen appears to have more limited immune function, at least relative to how A and B antigens work. Evidence indicates O type mothers who secrete antigens into other body fluids have reduced fertility against A or B bearing fathers, suggesting that maintaining O alleles comes at a fitness cost to both O type mothers and A and B type fathers (Molnar, 2002). However the fact that O alleles have evolved multiple times while being enzymatically non-functional and universally maintained at higher levels than A and B alleles may also indicate that the O allele's selective advantage appeared later in evolutionary history than the original selective advantage of the AB antigen system. The presence of anti-A/B antibodies could be the leading selective force, since it reacts to both A and B antigens. The presence of carbohydrate substances on bacteria or other antigenic carriers would then be an important target for such a sensitive antibody system (Preece et al., 2002).

In this sense, bacterial pathogens such as *Vibrio cholerae* (cholera) and *Helicobacter pylori* -that produce their own membrane antigens- are identified by our immune system as foreign objects using the different A, B or H antigens in our own cells as indicators of self. ABO allele frequencies seem to be maintained by the balance between selective pressures driven by being exposed to several different pathogens. This accounts for global maintenance of A, B and O polymorphism as well as some deviations from it. For example, in Glass et al's (1985) case study in the Ganges delta, the geographic origin of cholera, an individual's blood group type was not found to affect risk of having a culture-proven infection with *V. cholerae* but was directly related to the severity of disease. Individuals with the most severe diarrhea were more often of

blood group O compared with those with asymptomatic infection (0.68 of blood group O individuals versus only 0.36 of all others blood groups, $p < 0.01$) and much less often of AB (0 versus 0.7 of all other groups, $p < 0.01$). Not incidentally, today populations in the region also exhibit very low frequencies of O alleles (0.12-0.20 of O genotype), a probable consequence to historical exposure to *V. cholerae* (Glass et al., 1985).

Unlike bacteria, in viruses the glycosylation of surface carbohydrates is determined by the previous host cell, since a virus is unable to provide the complicated enzymatic machinery needed for glycosylation (Preece et al., 2002). In vitro studies using the measles virus (*Morbillivirus* spp.) as a model clearly demonstrates that terminal glycosylation of the host cell determines the terminal glycosylation pattern on measles virus. Viruses from blood group A hosts are coated in A antigens and H antigens (AA or AO) and viruses from blood group B hosts are coated in B antigens and H antigens (BB or BO). The same study also demonstrated that identification and neutralization of invading viruses by host serum is strictly complement-dependent; A serum neutralizes B (and H) coated viruses and vice versa (Preece et al., 2002).

Additionally, since the first appearance of the A and B antigen/antibody system, selection has driven some viruses to either acquire the complete cellular coating of its host (the viral envelope) or incorporate A and B antigens as motifs in their capsid coating (molecular mimicry), making it impossible for the host immune system to detect based solely on A and B antibodies (Calafell et al., 2008; Seymour et al., 2004; Swerdlow et al., 1994). For example all poxviruses retain the viral envelope of their budding host. Counter selection in the human immune system could then explain the rise of multiple O alleles, as O genotypes are hypersensitive to all other human antigens by producing both the anti-A and anti-B antibody.

Fluctuation between rising frequencies of types selected against by specific pathogens in previous generations is thought to keep all alleles present in most populations across the world, with the exception of most Native American and some Native Australian populations (Seymour et al., 2004).

Selection at ABO locus and global patterns of polymorphism

In a broad sense, the human population maintains the A, B and O alleles in high polymorphism with frequencies averaging 0.187 (S.D. 0.063) for A, 0.121 (S.D. 0.070) for B and 0.692 (S.D. 0.114) for O (Table 3). This global polymorphic distribution and frequencies of ABO alleles cannot be easily explained by any simple scenarios of selection driven by a single suit of pathogens. In turn, negative frequency dependent selection between alleles that code for antigens/antibodies, alleles with no enzymatic activity, and all of the above, has been modeled to account for modern ABO distribution (see section on modeling ABO polymorphism, Seymour et al., 2004).

Pathogens evolve much faster than long-lived vertebrates. Within a single host an expanding pathogen may be able to adapt and overcome the host's immune system over the multiple pathogen generations it takes to spread the infection within the host. This process ratchets virulence as pathogens infecting from previous hosts are better adapted to phenotypically identical immune systems, as they do not illicit an antibody reaction. If this dynamic holds true for the ABO locus, maintenance of diversity would be selected via constant higher fitness attributed to those exhibiting the rarest phenotypes (Gagneux and Varki, 1999).

Negative frequency dependent selection has likely maintained ABO polymorphism for at least 2.5 my, the time of the emergence of the first polymorphic variant, and has functioned as the great equalizer, maintaining similar diversity levels even in isolated populations, as long as

they are not too small. However two conditions can be responsible for driving allele frequencies away from the global tendency: 1) episodes of strong selection against one allele over the others, for example Cholera in the Ganges delta (Glass et al., 1985) or smallpox in Iceland (Adalsteinsson, 1985) or 2) genetic drift as variation at any loci is constantly lost by stochastic forces at a rate inversely proportional to the effective population size. *Homo sapiens* are well known as a species for exhibiting a fairly small effective population size (~10 000, Chen and Li, 2001), especially remarkable, given the species tremendous current census size (over 6 billion). The effective population sizes of archaic populations and especially those that moved into previously uncolonized areas after the human diaspora from Africa were likely to be even smaller. However, negative frequency dependent selection may have been responsible for maintaining polymorphism at the ABO locus in the descendents of these human populations across the world (Table 3). Even other well known linguistically isolated populations around the world where drift has presumably shaped their genetic make-up are polymorphic for ABO (Table 4).

While selection against a specific type, such as selection against O alleles by *V. cholera* in the Ganges delta, or isolation of populations such as the Basques, Saami or the Ait Haddidu Berbers can skew frequencies away from the global average, there are very few cases of monomorphic human populations. Some of those few cases are Native American groups, for example the Bororo, Guarani and Ona of South America (Molnar, 2002; Swerdlow et al., 1994).

Table 4 Frequencies for human ABO alleles in well known linguistically isolated populations across the world (from Molnar, 2002).

Isolates	A	B	O
Iceland	0.20	0.05	0.75
Irish	0.17	0.07	0.76
Saami	0.37	0.09	0.54
Basques	0.24	0.02	0.74
Bearnais	0.24	0.04	0.72
Corsicans	0.22	0.03	0.75
Sardinians	0.20	0.07	0.73
Walsers	0.21	0.05	0.73
Bergamasques	0.24	0.06	0.70
Valle Ladine	0.20	0.03	0.77
Svani	0.23	0.07	0.70
Saudi Arabians	0.16	0.11	0.75
Towara Bedouin	0.16	0.09	0.74
Jebeliya Bedouin	0.12	0.26	0.62
Ait Haddidu Berbers	0.07	0.05	0.89
Average	0.20	0.07	0.73
S.D.	0.07	0.06	0.08
Variance	0.00	0.00	0.01

Native American populations as an exceptions to global patterns

All modern Native American populations exhibit high frequencies of O blood groups (Table 5, Fig 3) (Llop et al., 2006; Molnar, 2002; Swerdlow et al., 1994; Szathmary, 1979). While the O allele is not completely fixed, only a few populations exhibit high frequencies of the other two alleles, but most cases of polymorphism have been attributed to European admixture or more simply have not been addressed (as in the case of Estrada-Mena et al., 2009).

The observation that Native Americans exhibit almost exclusively blood group O has traditionally been accepted as the result of founder's effect at the time when humans first entered the Americas (Estrada-Mena et al., 2009). More recently, selection driven by viruses, particularly smallpox, introduced following European contact, has been invoked to explain Native American lack of diversity. In support of the later hypothesis, smallpox-driven selection has been shown to drive O alleles to fixation as was the case in West Bengal and Bihar populations of India in an intense smallpox epidemic in the 1960s (Adalsteinsson, 1985)

Smallpox was initially introduced to the Americas by Europeans and, in fact, over the following centuries subsequent waves of the disease were brought in via sailors (Patterson and Runge, 2002). While smallpox was severe even among Europeans, who had been exposed to it since medieval times, they may have had coevolved resistance mechanisms (Sabeti et al., 2005). Smallpox, however, was devastating to the Native Americans, who had never been exposed to the virus prior to contact. Smallpox ultimately killed more Native Americans in the early centuries than any other disease, or conflict, post-European contact (Patterson and Runge, 2002).

The effect of population decline due to recent selection aside, the low modern diversity observed among Native American populations stands as an exception to an otherwise high pattern of global diversity. Native Americans differ even from Asian populations such as Siberians, with whom they share relatively recent ancestors (Mourant et al., 1976; Tamm et al., 2007; Torroni et al., 1993).

Table 5 Frequencies for human ABO alleles in Native Americans (from Mourant, et al., 1976, Halverson and Bolnick, 2008, Szathmary, 1979)

North America	A	B	O	n
Caddo-Oklahoma	0.162	0	0.838	47
Catawba-SouthCarolina	0.218	0.081	0.701	104
Cherokee-NorthCarolina	0.018	0.009	0.973	166
Choctaw/Chickasaw-Oklahoma	0.044	0	0.956	81
Chippewa-Minnesota	0.064	0	0.936	161
Ojibwa-ManitoulinIsland	0.172	0.044	0.784	105
Ojibwa-Pikangikum	0.225	0	0.775	95
Omaha-Nebraska	0.056	0.003	0.941	168
Pawnee-Oklahoma	0.226	0.012	0.762	80
Penobscot-Maine	0.27	0.002	0.728	249
Seminole-Florida	0.019	0.004	0.977	381
Seminole-Oklahoma	0.039	0.039	0.922	224
Sioux-SouthDakota	0.159	0.015	0.826	260
Wichita-Oklahoma	0.243	0.02	0.737	49
Ho-Chunk-Nebraska	0.238	0	0.762	86
Average	0.1435	0.0153	0.8412	
s.d.	0.0931	0.0230	0.0994	
Variance	0.0087	0.0005	0.0099	

Eskimos	A	B	O	n
Inuit-Nome-Alaska	0.259	0.074	0.667	254
Inupiat-PointBarrow-Alaska	0.291	0.0631	0.6459	329
Unangan-Aleutian-Islands	0.2684	0.0318	0.6998	144
Inuit-LabradorBaffinIslands-Canada	0.2547	0.0034	0.7419	146
Inuit-Angmagssalik-Greenland	0.4069	0.1048	0.4883	569
Inuit-CapeFarewell-Greenland	0.3307	0.0251	0.6442	484
Inuit-CapeYork-Greenland	0.0877	0.0325	0.8798	124
Inuit-Wainwright-Alaska	0.3578	0.09	0.5522	111
Inuit-Copper-Canada	0.2926	0.0333	0.674	320
Inuit-Aivilik-Canada	0.4253	0	0.5747	109
Inuit-Kuujjuaq-Canada	0.211	0.0127	0.7763	278
Inuit-Thule-Greenland	0.0895	0	0.9105	152
Inuit-KodiakIsland-Alaska	0.2767	0.0636	0.6597	306
Inuit-KoniagIsolates-Alaska	0.2115	0.1071	0.6814	257
Inuit-Augpilagtok-Greenland	0.2273	0.1027	0.67	152
Inuit-Ittoqqortoormiit-Greenland	0.3558	0.0143	0.6299	246
Inuit-Angmagssalik-Greenland	0.4391	0.103	0.4579	739
Average	0.2815	0.0507	0.6679	
s.d	0.1012	0.0401	0.1181	
Variance	0.0102	0.0016	0.0140	

Athabaskan+Related	A	B	O	n
Dene-Canada	0.1117	0.0099	0.8784	204
Kwakwaka'wakw+Nuu-chah-nulth +Haida+Salish	0.0654	0.0033	0.9313	300

Kwakwaka'wakw+Nuu-chah-nulth +Haida+Tlingit+Gitishan+Cowichan	0.096	0.0039	0.9001	394
Nuu-chah-nulth	0.0076	0	0.9924	198
Haida	0.1038	0.0092	0.887	437
Tlingit	0.0789	0.0321	0.889	79
NorthernAthapaskan	0.1416	0.0015	0.856	650
Navajo-Arizona	0.1463	0.0022	0.8515	457
Navajo-NewMexico	0.0045	0	0.9955	112
Tinglit-Alaska	0.1292	0	0.8708	120
Average	0.0885	0.0062	0.9052	
s.d	0.0504	0.0098	0.0519	
Variance	0.0025	0.0001	0.0027	

Meso+SouthAmerica	**A**	**B**	**O**	**n**
Natives-Guatemala	0.03	0.01	0.96	237
Maya-Yucatan	0.009	0.0045	0.9865	223
Chaguanco+Chamaco+Chiriguano-Argentina	0	0	1	120
MatoGroso	0	0	1	587
Bororo-MatoGrosso	0	0	1	119
Caiua-MatoGrosso	0	0	1	237
Guarani-MatoGrosso	0	0	1	107
Tariana+Tucano+Macu-Amazonas	0	0	1	237
Pijao-Colombia	0	0	1	281
Cañari-Andes	0.0176	0.0079	0.9745	1726
Cara-Ecuador	0.0187	0.0068	0.9745	1838
Panzaleo-Ecuador	0.0205	0.008	0.9715	2570
Palta-Ecuador	0.0124	0.0079	0.9797	568
Jibaro-Amazonas	0	0	1	111
Macá-Paraguay	0	0	1	108
Andes-Peru	0	0	1	100
Uros-Peru	0	0	1	27
Ona-TierradeFuego	0	0	1	21
Ona+Yamana+Alakaluf-Argentina	0	0	1	34
Average	0.0057	0.0024	0.9919	
s.d.	0.0094	0.0037	0.0131	
Variance	0.0001	0.0000	0.0002	

All Native Americans	**A**	**B**	**O**
Average	0.1294	0.0194	0.8512
s.d.	0.1290	0.0310	0.1499
Variance	0.0166	0.0010	0.0225

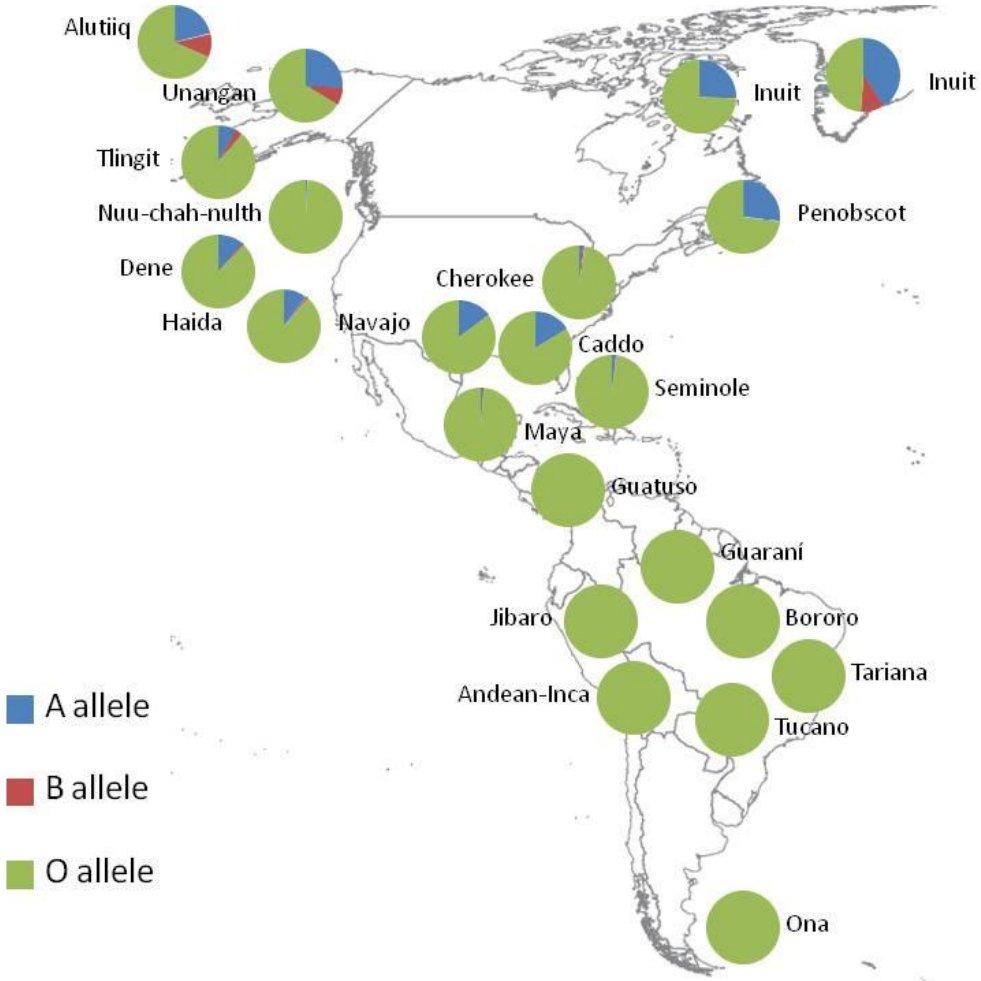

Figure 3 ABO allele frequencies in Native Americans (selected groups from Table 5).

Hypothesis and Expectations

Founder's effect

Proto-Native Americans are thought to have evolved as an ancestral group to all Native Americans in isolation even before entering the empty American continent, during a time termed the "Beringian standstill" period between 30,000 YBP and 15,000 YBP (Tamm et al., 2007). This is one part of a model known as the "Beringian Incubation Model" or "BIM" that has found ample recent support from a number of whole mitochondrial DNA studies (Achilli et al., 2008; Tamm et al., 2007).

Support for this model was also found in the nuclear genome with the discovery of a 275 bp allele at the D9S1120 microsatellite unique to America and observed at high frequencies in all American populations, yet absent from all other world-wide population except for a few across the Bering Strait, has solidified the BIM and the migration of a single proto-American population ancestral to most Native Americans (Schroeder et al., 2009; Schroeder et al., 2007).

To evaluate whether variation at the ABO locus was lost among proto-Native Americans either during the standstill period, before the actual migration, or crossing into America, it is crucial to estimate the effective population size (N_e) of the population, as smaller sizes increase the chances of allele loss. The degree of the bottleneck could range from complete elimination of polymorphism to a partial reduction of the number of founding alleles in proto-Native Americans.

Some contemporary Native American groups do exhibit A and B alleles in considerable frequencies, although the distribution of these alleles is highly structured geographically (Figure 3). This possibly suggests that the bottleneck was not strong enough to completely eliminate diversity. However European admixture has not been ruled out as a possible cause for this pattern. Sequence data from O alleles has provided additional resolution into Native American diversity. Far from being monomorphic, at the sequence levels Native Americans present a variety of O haplotypes, some isolated geographically (such as O05 and Ov7 in the Cayapa, O32 and O33 in the Aymara), some shared with other human populations (O1 and O1v), and one haplotype called "O1v542" which has been found in all Native American populations screened at this level (Nahua, Mazahua, Maya, Mexican Mestizo, Cayapa, Aymara) and may very well be unique to the Americas (Estrada-Mena et al., 2009). This pattern of reduced ABO diversity but

not complete fixation of one allele is less consistent with one strong bottleneck due to founder's effect.

Other DNA markers such as mitochondrial haplogroups have been the source of much debate as of the degree of the bottleneck effect in proto-American population diversity (Kemp et al., 2005; Tamm et al., 2007; Torroni et al., 1993). The only uncontested fact is that the minimum number of females was equal to or higher than the number of founding mitochondrial haplogroups know thus far (possibly>9), but an upper limit to population size has not been agreed on (Tamm et al., 2007).

Post-contact reduction of diversity

European contact had a catastrophic effect in Native American populations ranging from population reduction to complete extinction of some groups and overall diversity loss ranging from 50-90% (Mulligan et al., 2004). In light of this drastic event it is appropriate to expect an associated random reduction in genetic diversity in all loci, including ABO, in addition to any non-random (selection) changes in diversity.

Ancient DNA studies ranging as far as 12 300 BP (Gilbert et al., 2008) and more recently into populations that lived a few hundreds of years before European contact provide a unique look into the genetic diversity of ancient Native American populations. Ancient mitochondrial DNA studies were among the first to consider the possibility of a diversity reduction following contact, as modern Native American present low mitochondrial diversity but at the same time all Native American haplotypes are endemic to America. If ancient Native Americans presented considerably higher mitochondrial diversity, ancient DNA studies would have recorded mitochondrial haplotypes no longer present in contemporary populations (Stone and Stoneking,

1998). So far only one study has recovered one mitochondrial haplotype (M) not currently sampled in extant Native Americans (Malhi et al., 2007).

Two early studies, Parr et al. (1996) and Carlyle et al. (2000) compared the mitochondrial haplogroup diversity of two ancient Southwest populations, the Freemont and the Anasazi with contemporary Southwest groups. Their samples proved to not only belong to recognized Native American founding haplogroups observed in contemporary populations, but to be of statistically similar frequencies to modern Pueblo groups.

Gonzáles-Oliver (2001) conducted a similar study of ancient Mayan mitochondrial diversity, again finding the same haplogroups in frequencies similar to modern Mayan groups. These frequencies were completely congruent with expectations of any given Mesoamerican group. Two studies, Kemp et al. (2005) and De la Cruz et al (2008) on two ancient Aztec populations, recovered frequencies that would make this population statistically indistinguishable from modern Chibcha peoples, ancient Maya from Xcaret, contemporary Maya from the Yucatan and importantly the Nahua, who are the descendants of the Aztecs.

Although these are isolated studies, ancient mitochondrial haplotype diversity seems to be comparable to modern diversity. If diversity in this marker is comparable to autosomal loci diversity, it is likely that the overall pattern of genetic diversity was not significantly altered with population reduction following European contact, even more as autosomal markers have an effective population size 2-4 larger than mitochondrial markers. It is possible, however, that the ABO locus alone could have reduced its diversity by targeted selection, for example after a severe smallpox epidemic.

Post-contact natural selection of O alleles

If ABO alleles among contemporary Native American populations were strongly shaped by intense selection following the smallpox epidemic, it is possible that ABO frequencies predating European contact would be more consistent with the global average. There is only one study so far that has addressed diversity at the ABO locus using ancient pre-contact samples. Halverson and Bolnick (2008) successfully typed 15 ancient Native Americans from three archaeological sites (Table 6): 1) the Pete Klunk mound group (dated to 1,825-675 YBP) in Calhoun County Illinois, 2) the Wright mound (dated to 1,950–1,790 YBP) in Montgomery County Kentucky, and 3) the Hardin Village cemetery (dated to 500–325 YBP) in Greenup County Kentucky. While not representing a single "population," these samples yielded frequencies very similar to contemporary groups of the region: the Caddo, Cherokee, Choctaw, Chippewa, Ojibwa, Omaha, Pawnee, Seminole, Sioux and Wichita (Table 7), which are all characterized by reduced diversity and low frequencies of non-O alleles.

Halverson and Bolnick's (2008) results are consistent with ABO patterns achieving modern low diversity well before European contact. Further sampling is needed to better test this hypothesis, however, to generalize to the evolution of all Native American groups. This is jointly due the low heterozygosity of their results and the fact that their sample is quite scattered geographically and temporally.

Table 6 ABO allele frequency from a precontact population (N=15) and an average of modern Native American populations of the same area (Halverson and Bolnick, 2008).

Allele	Ancient	Modern
A	0.330	0.120
B	0.000	0.014
O (O1+O1v)	0.967	0.865
O1	0.267	N/A
O1v	0.700	N/A

Table 7 Fisher's Exact test comparison of Halverson and Bolnick (2008) precontact population with modern Native American tribes of the same region. Low p values indicate populations significantly distinguishable from the precontact population.

Population	p value
Caddo-Oklahoma	0.18283+-0.0019
Catawba-SouthCarolina	0.03682+-0.0013
Cherokee-NorthCarolina	1.00000+-0.0000
Choctaw/Chickasaw-Oklahoma	1.00000+-0.0000
Chippewa-Minnesota	0.48493+-0.0010
Ojibwa-ManitoulinIsland	0.22577+-0.0038
Ojibwa-Pikangikum	0.06743+-0.0020
Omaha-Nebraska	1.00000+-0.0000
Pawnee-Oklahoma	0.10180+-0.0027
Penobscot-Maine	0.01548+-0.0004
Seminole-Florida	1.00000+-0.0000
Seminole-Oklahoma	1.00000+-0.0000
Sioux-SouthDakota	0.30802+-0.0037
Wichita-Oklahoma	0.06929+-0.0024
Winnebago-Nebraska	0.03628+-0.0013

High population structuring after migration

If one strong bottleneck had shaped ABO frequencies before the various founding groups spread into America, all Native American populations today might be expected to exhibit similar allele frequencies. By contrast, if drift had reduced diversity in independent groups after they became isolated from each other, then given the stochastic nature of genetic drift, different Native American groups would be expected to differ in what alleles became fixed.

While Native Americans as a whole have lower diversity at the ABO locus than other continental populations, variance within the Native American samples (Table 5) is much higher compared with global ABO frequencies (Table 3) or even ABO frequencies in linguistically isolated populations (Table 4) (ANOVA $F= 3.088$. $p=0.030$ for A allele, $F= 48.16 p<0.010$ for B allele, $F= 13.58$, $p<0.001$ for O, Fig 4). The high variance within Native American groups is

evidence of demographic processes occurring independently in isolated groups of Native Americans after entering America.

a)

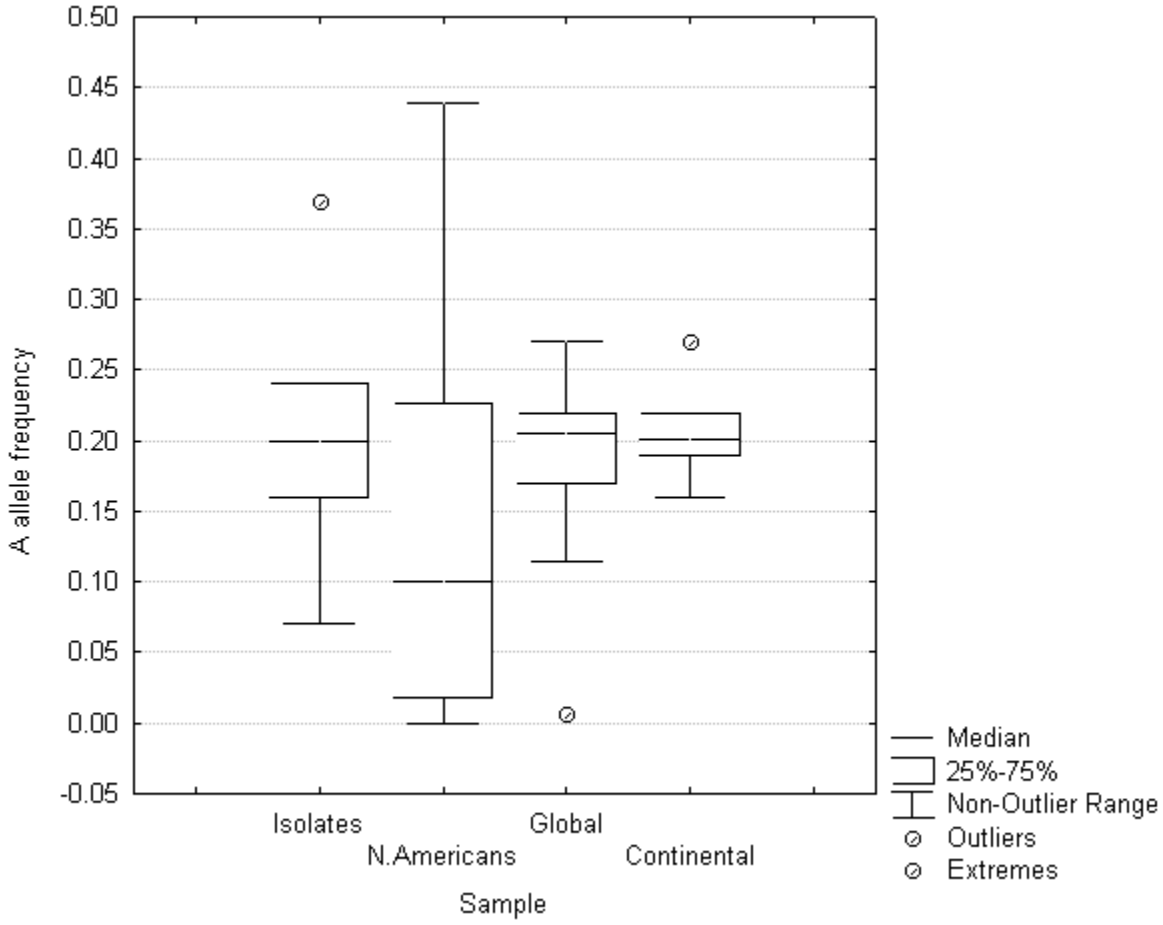

Figure 4 Variance in allele frequencies for a) A allele, b) B allele and c) O allele for 4 population samples: Global (Table 3), Isolates (Table 4) Native Americans (Table 5) and a "Continental" population equal to the Global population without Native Americans. Boxes represent the 25th and 75th quartiles. Analysis of variance is significant at 95% confidence for all 3 alleles.

b)

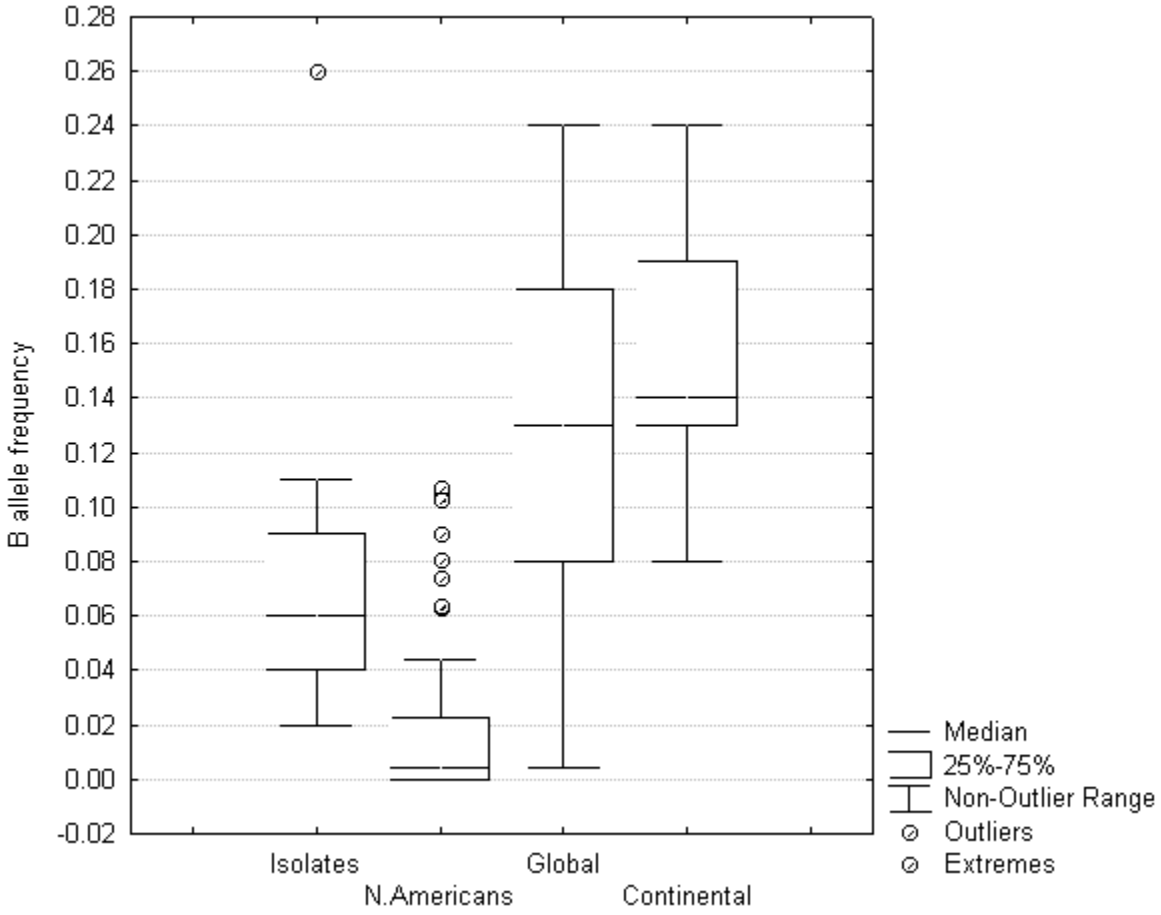

Figure 4 Variance in allele frequencies for a) A allele, b) B allele and c) O allele for 4 population samples: Global (Table 3), Isolates (Table 4) Native Americans (Table 5) and a "Continental" population equal to the Global population without Native Americans. Boxes represent the 25th and 75th quartiles. Analysis of variance is significant at 95% confidence for all 3 alleles.

c)

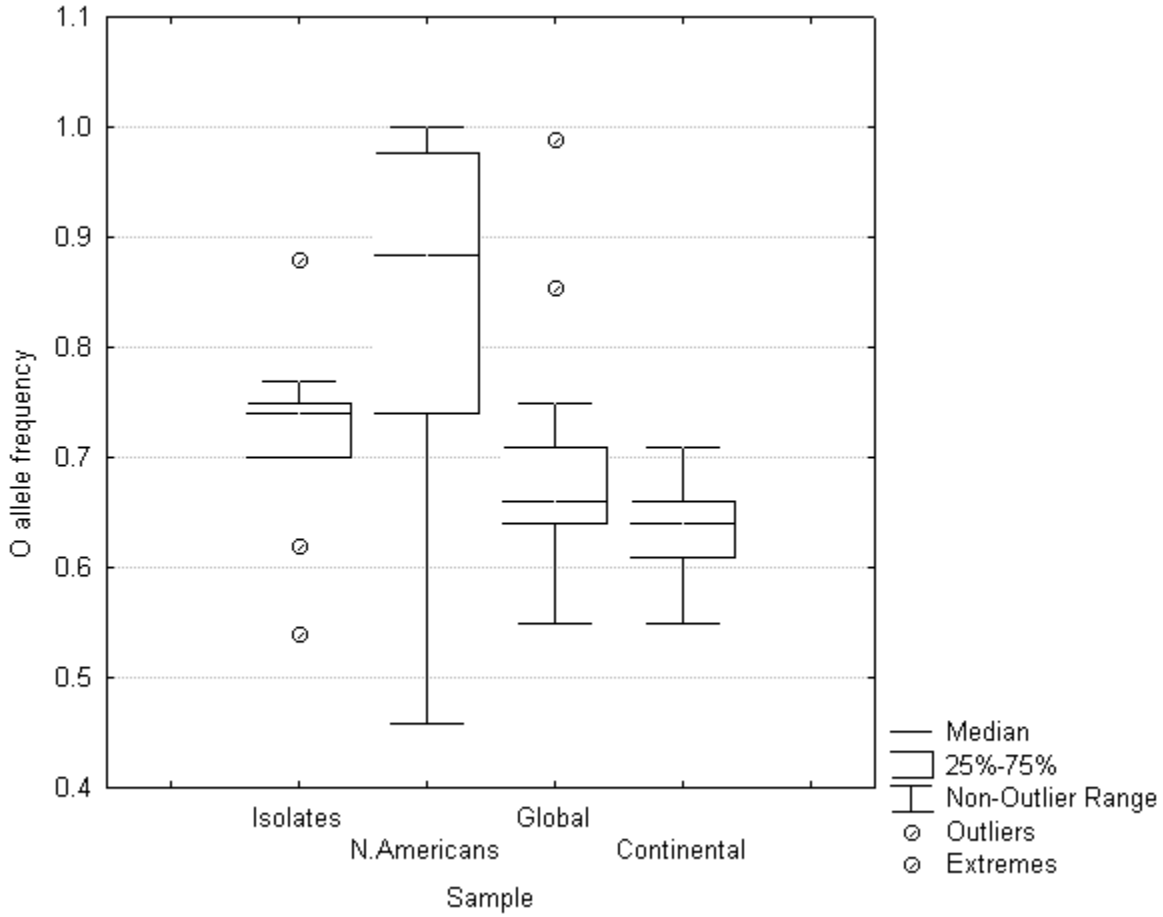

Figure 4 Variance in allele frequencies for a) A allele, b) B allele and c) O allele for 4 population samples: Global (Table 3), Isolates (Table 4) Native Americans (Table 5) and a "Continental" population equal to the Global population without Native Americans. Boxes represent the 25th and 75th quartiles. Analysis of variance is significant at 95% confidence for all 3 alleles.

Congruently, Wang et al.'s (2007) analysis of 678 autosomal microsatellite markers between 24 Native American populations representing North, South and Central America and other world populations. They found that Native American populations are at the same time the least heterozygous continental "population" and the most highly structured (Wang et al 2007). The phylogenetic tree best fitted to the data from the microsatellite panel, has a structure better conciliated with a scenario of one migration into America followed by swift population isolation

via migrations deeper into the continent and isolation of groups splitting up along the way taken subsets of the original gene pool of the founding population. They also concluded that this scenario would produce a set of low diversity subpopulations with different patterns of variations that may then drift away from each other (Wang et al., 2007).

Furthermore, modeled evolution of the Native American private allele D9S1120 emphasizes high population structuring and isolated evolution of diversity after migration into the American continent (Schroeder et al., 2009). Although it has been argued that a similar pattern of the distribution of D9S1120 could also be achieved through high gene flow between Native American and Asian populations, this seems unlikely in light of all other evidence (Ray et al., 2009).

Isolation and drift may have been facilitated by differences between emerging languages. It should be noted that Native Americans possess the highest linguistic diversity compared with other continental "populations," possibly the result of a high degree of isolation promptly following Native American dispersion (Nettle 1999).

Modeling diversity loss in Native American populations

Gagneux and Varki (1999) claim that ABO diversity is maintained globally by frequency dependent selection via higher fitness of rarer variants. While negative frequency dependent selection is a likely explanation for the maintenance of polymorphism globally, a second evolutionary force is necessary to explain why O alleles have been maintained at higher frequencies (Seymour et al., 2004). If there is differential fitness based on recognition of terminal glycosylation from AB transferases, as is the case with viruses that present a viral envelope (or mimicry), individuals with no glycosylation would be sensitive to both A and B antigens present in viruses incubated in A, B or AB phenotype hosts. The interaction of both forces on ABO

frequencies has been modeled by Seymour et al. (2004) over multiple generations (Fig 5). Their model, at equilibrium, maintains the frequency of the O allele as twice or more those of the A and B alleles. A notable feature of their models is that there are no oscillations in allele frequency during the approach to equilibrium. This is likely because their model does not take into consideration sampling error issues arising from population of limited sizes, as their population size is infinite.

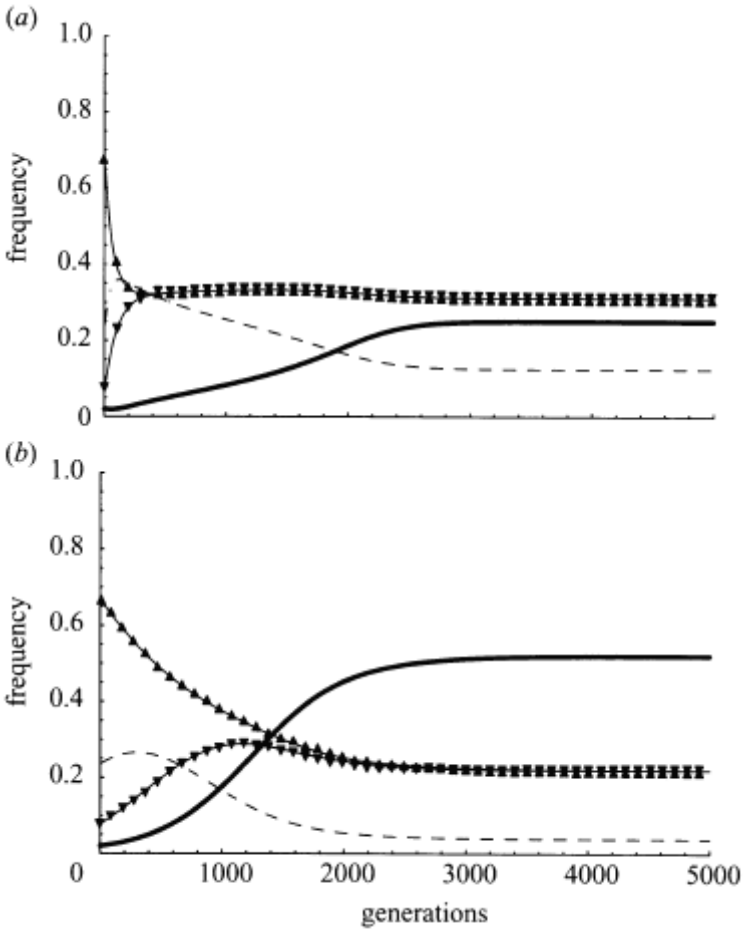

Figure 5 ABO evolution as modeled by Seymour et al. (2004). Evolution of O (solid line), A (up-triangles), B (down-triangles) and AB (dashed line) phenotype frequencies under the influence of (a) bacteria alone and (b) bacteria and virus.

Methods

Ancient DNA

The colonization of the Americas by Europeans had an enormous impact on Native American populations, ranging from extinction of some groups to forced relocation and subsequent admixture among other groups (some previously not in contact with one another) and with European settlers themselves. As such, studying Native American evolution using genetic analysis from contemporary samples can present significant biases. While most research is conducted keeping these biases in mind, ancient DNA provides the opportunity to sample directly from prehistoric populations. Ancient DNA studies, however, present their own set of challenges, making validation of results an extremely slow and laborious process. As few ancient DNA studies exist, each one represents a point of light in the vast darkness of human origins and evolution.

Ancestral Muwekma Ohlone

In this study ancient DNA from the Yukisma site (CA SCL-38), an ancestral Muwekma Ohlone burial site north Santa Clara County, California was analyzed. The site is located approximately 6 miles from the present Southern end of San Francisco Bay (Morley, 1997). This site falls under the Tamyen-speaking territory, part of the Ohlone languages in the Utian family (Bellifemine, 1997). One rib bone from each of the 252 burials was reserved for this genetic analysis prior to remaining skeletal elements being re-interred. With approval from the Muwekma Ohlone the samples were transferred to the Kemp Lab of Molecular Anthropology and Ancient DNA at Washington State University. Radiocarbon dating places deposition from 200 YBP to at least 2200 YBP.

The archaeological project was initially conducted by Ohlone Families Consulting Services (OFCS), cultural resource management arm of the Muwekma Ohlone Tribe. After an initial testing program, which included eight 1m by 1m excavated units and 16 auger bore holes, conducted during April 1993, 243 discrete burial features were exposed, documented and their contents removed from the project area between August and October 1993. It is believed the site is a large prehistoric earth mound cemetery. From this burial assemblage, 244 individuals associated with 3200 beads and artifacts were analyzed (Bellifemine, 1997).

The valley floor is mainly composed of sedimentary rocks, including fluvial, aeolian, weathered and estuarine deposits. The valley climate is described as Mediterranean, with characteristically wet winters and dry summers (Bellifemine, 1997). With over 250 possible aDNA samples, CA SCL-38 represents one of the closest examples of a complete precontact Native American population, comparable only to the 260 preserved skeletons of the Norris Farm 36 cemetery (Stone and Stoneking, 1998).

DNA Extraction

Approximately 0.5 g of material was removed from each rib sample. Each sample was submerged in 6% sodium hypochlorite (full strength household bleach) for 15 min (Kemp and Smith, 2005). Each sample was independently rinsed 1-2 times with DNA free water (Gibco) to remove the bleach. Individual samples were transferred to a 15 mL polypropylene conical tube and submerged in 3 mL of molecular grade (DNA free) 0.5 M EDTA, pH 8.0 (Gibco) for at least 48 hours. An extraction control, to which no bone is added, accompanied each extraction batch of seven samples and was subjected to all of the steps that follow.

Three mg of Proteinase K (Invitrogen, Fungal Proteinase K) were added and the samples were incubated at 60-65°C for approximately 3 hours. A three-step phenol/chloroform extraction

method was used: two extractions adding an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) to the EDTA, followed by one extraction with an equal volume of chloroform:isoamyl alcohol (24:1).

To aid in the removal of co-extracted polymerase chain reaction (PCR) inhibitors, DNA was precipitated out of solution with isopropanol (Hänni et al., 1995). This was performed by adding one half volume of room temperature 5 M ammonium acetate and, to this combined volume, one volume of room temperature absolute isopropanol, then storing the solution overnight at room temperature.

Samples were centrifuged for 30 min at 3100 rpm to pellet the DNA. Then the liquid was discarded and the samples air-dried for 15 minutes. The pellet was washed with 1 mL of 80% ethanol by vortexing for about 30 seconds (making sure to dislodge the pellet, if visible, from the side of the tube). Tubes were centrifuged again for 30 min at 3100 to pellet the DNA. The ethanol was discarded and the samples were allowed to air-dry for 15 min. To further remove co-extracted PCR inhibitors, DNA was re-suspended in 300 μL of 60-65°C DNA-free ddH_2O and incubated at 60-65°C for 20 minutes and vortexed three times before silica extraction (Glass et al., 1985; Kemp et al., 2006) using the Wizard PCR Preps DNA Purification System (Promega), following the manufacturer's instructions (modified by eluting the DNA in the final step with two aliquots of 50 μL of 60-65°C ddH_2O as recommended by Dr. Jodi Barta). DNA extracts were stored at -20°C.

PCR Amplification

Mitochondrial aDNA haplogrouping

Each sample was first screened for the markers definitive of Native American mtDNA haplogroups A, B, C, D (Torroni et al., 1993) to help ensure samples were not contaminated by

modern non-Native American DNA. As the ratio of mtDNA to nuclear DNA in ancient samples is expected to be approximately 1000:1 (Schwarz et al., 2009), samples that do not amplify for mitochondrial DNA are unlikely to amplify for nuclear DNA. Screening for mitochondrial haplogroups also works as the first step for screening samples with sufficient aDNA for ABO gene amplification.

Fifteen microliter PCR amplification reactions contained: 0.32 mM dNTPs, 1X PCR Buffer, 1.5 mM $MgCl_2$, 2.4 mM primers, 0.3 U of platinum Taq (Invitrogen), and 1.5 μL template DNA. Negative controls (PCR reactions to which no DNA template was added) accompanied every set of PCR reactions to monitor the presence of contaminating DNA.

PCR conditions were as follows: 94°C for 3 min, 60 cycles of 15 second holds at 94°C, 55°C, and 72°C, followed by a final three minute extension period at 72°C. Amplification success was determined by visualizing ~5-6 μL of the amplicons on a 6% polyacrylamide gel stained with ethidium bromide under UV light. This was preformed both to confirm successful amplification and for later restriction enzyme digest. In the case of B haplotypes scoring was performed from the presence or absence of a 9-bp deletion. Primers are described in Kemp et al. (2007).

ABO blood group testing

ABO blood grouping was scored using a modified Hummel at al. (2002) PCR-RFLP method. PCR amplification reactions contained 7.26 μL of DNA-free ddH_2O (Gibco), 0.32 mM dNTPs, 1X PCR Buffer, 1.5 mM $MgCl_2$, 2.4 mM primers, 0.3 U of platinum Taq (Invitrogen), and 3 μL template DNA. Negative controls (PCR reactions to which no DNA template was added) accompanied every set of PCR reactions to monitor the presence of contaminating DNA.

PCR conditions were as follows: 94°C for 3 min, 60 cycles of 15 second holds at 94°C, 56°C (for Exon 6) and 53°C (for Exon 3), and 72°C, followed by a final three minute extension period at 74°C. Amplification success was confirmed as just described. Primers are described in Hummel et al. (2002).

Enzyme digestion for *Rsa*I and *Hpy*CH4IV (Exon 6) and *Nla*III and *Mnl*I (Exon 7) was performed at approximately 36°C for over 3 hours. Digestion was visualized by running 4-5 μL of the amplicons on a 6% polyacrylamide gel stained with ethidium bromide under UV light.

To rule out allelic dropout (see below), positive ABO result were attempted to be replicated three independent times following Halverson and Bolnick's (2008) recommendation. To rule out contamination from the researcher, each sample was replicated one additional time independently by Dr. Brian M. Kemp.

Data Analysis

Paired Fisher exact tests were performed using the program Arlequin version 3.0 (Excoffier, 2005) to compare the CA SCL-38 mitochondrial haplotype frequencies to modern California Native America populations as well as other western North American groups (Eshleman et al., 2004; Kaestle and Smith, 2001; Lorenz and Smith, 1996; Malhi et al., 2003; Smith et al., 2000). The program was set at 100 000 Markov Chain steps and the alpha value was set at 0.05. A similar paired Fisher's exact test was performed to compare the CA-SCL-38 ABO frequencies to other Native American populations. Fst distances were also calculated in Arlequin and the Fst distance matrix was processed in a Principal Coordinate Analysis using DistPCoA (Legendre and Anderson, 1998) to visualize structuring within Native American groups. A scatter plot was created based on the first two coordinates. As a comparison point, Table 3 non-Native American populations were also included in the plot.

ABO diversity evolution model

This study modeled frequency dependent selection at the ABO locus under similar principles as Seymour et al (2004, Fig 5) and incorporated variable effective population sizes. Comparisons of the results yielded by the model with known modern ABO diversity in Native Americans can be used to estimate the effective population size of a group at the moment of diversity loss.

Natural selection was modeled by assuming that the fitness ($w_{i(n)}$) of individuals is reduced by the frequency of other individuals that carry the same antigens as they are all potential hosts for pathogens. Pathogens coming from these "same-phenotype" hosts are already adapted to bypass the ABO identification system:

$w_{i(n)}$= 1-freq(ii)-freq(i(n))

where homozygotes ii and heterozygotes i(n) both carry the same antigens (antigen H is present in low quantities even in AA or BB individuals; see Table 1) (Sharon and Fibach, 1991).

Differential transmission of viruses coated in A and B antigens (molecular mimicry or viral envelope) lowers the fitness of A and B bearing genotypes (homozygous or heterozygous) by a number "x" and increases the fitness of the genotype that does not carry A or B (OO) by the sum of all "x":

$w_{i(n)}$= 1-freq(ii)-freq(i(n))-x if i= A or B

w_{ii}= 1-freq(ii)-freq(i(n))+x if i= O

For the case of the model presented here, "x" was assigned arbitrarily as 0.025, to match the ABO frequencies of Seymour et al.'s model (2004). Genotype fitness (w) was then calculated for each zygote as:

w_{AA}= 1-$freq_{AA}$-$freq_{AO}$-0.025

w_{AO}= 1-$freq_{AA}$-$freq_{AO}$-0.025

w_{BB}= 1-$freq_{BB}$-$freq_{BO}$-0.025

w_{BO}= 1-$freq_{BB}$-$freq_{BO}$-0.025

w_{AB}= 1-$freq_{AB}$-0.025-0.025

w_{OO}= 1-$freq_{OO}$+0.15

The mean fitness of the population can then be estimated as:

$$w_{(bar)} = p^2(w_{AA})+2pq(w_{AB})+q^2(w_{BB})+2pr(w_{AO})+2qr(w_{BO})+r^2(w_{OO})$$

These equations can be used to estimate how negative frequency dependent selection influences the frequencies of ABO alleles (p, q, r) on each generation based on its fitness relative to the mean fitness of the population:

$$p' = \frac{p^2(w_{AA})+pq(w_{AB})+pr(w_{AO})}{w_{(bar)}},$$

$$q' = \frac{q^2(w_{BB})+qr(w_{BO})+qp(w_{AB})}{w_{(bar)}},$$

$$r' = \frac{r^2(w_{OO})+rp(w_{AO})+rq(w_{BO})}{w_{(bar)}},$$

This system allows polymorphism to be maintained unless sampling error (i.e. genetic drift) removes all but one allele from the population.

If Native American loss of ABO diversity is the result of drift acting because of their small effective population size as bands became isolated, this can be modeled by randomly generating gametes carrying p, q, r alleles. The effective population size is represented as the number of gametes generated. Genetic drift can then be simulated, scoring the number of generated gametes that carry p, q, r alleles based on the distribution of p_0, q_0, r_0 in the previous generation. Then new zygote frequencies are generated and fed into the selection component of

the model. Fluctuations of allele frequencies can be traced through multiple generations by repeating each step multiple times.

For this model, the interplay between selection and genetic drift as the effective population size decreases was measured by generating 500, 100, 75, 50, 25, and 10 gametes per generation and measuring loss of diversity after 35 generation (roughly 1000 years with 30 year generations) (Tremblay and Vézina, 2000). All equations were run in Microsoft Excel. Resulting ABO frequencies were recorded after 100 replications of each 35 generation cycle.

Results

CA-SCL-38 DNA extraction and amplification

DNA extraction was attempted for 68 individuals from the CA-CSL-38 archaeological samples. Out of those 68 samples only 41 samples amplified for mitochondrial DNA haplogroups, either because there was not sufficient DNA for amplification or because silica purification was not successful at fully removing PCR inhibitors. Table 8 lists mitochondrial haplogroup frequencies for all 41 samples.

Table 8 Mitochondrial haplogroup frequencies for the CA SCL-38 population.

Haplogroup	Frequency
Haplogroup A	5.3
Haplogroup B	23.7
Haplogroup C	18.4
Haplogroup D	52.6
Other	0

Fisher's exact test-pairwise comparisons between mitochondrial haplogroups frequencies of the CA-SCL-38 population and contemporary western Native American groups are found in Figure 6. CA-SCL-38 is significantly indistinguishable from Cahuilla, Costanoan, Modern Great

Basin (multiple groups), Sahaptain, Serano/Vanyume, Sierra Miwok, Tubat and Yokuts populations as well as Stanford and Cook site ancient populations.

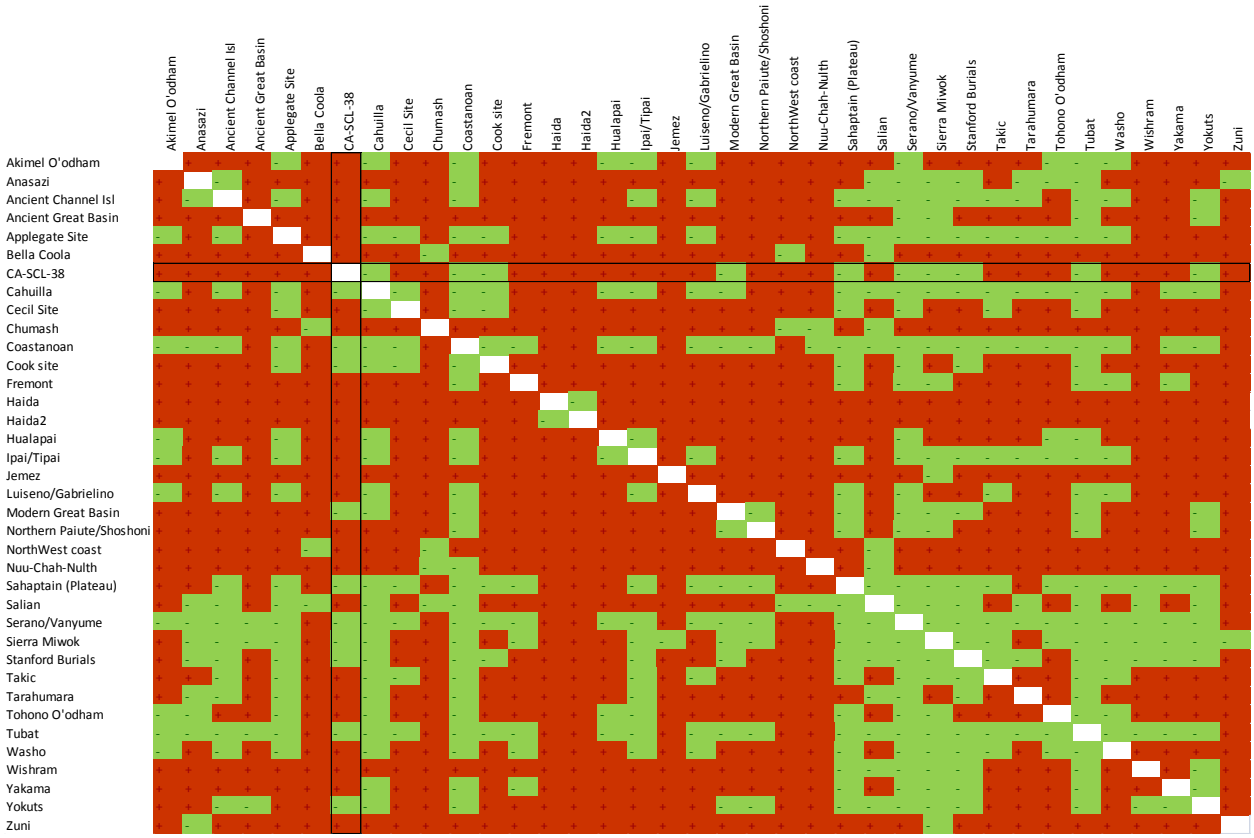

Figure 6 Pairwise Fisher's Exact test significance matrix for mitochondrial haplogroups. Red cells indicate significantly distinguishable populations at the 0.05 level of probability, green cells indicate significantly indistinguishable populations. CA-SCL-38 is delimited by dark bars.

Out of the 41 samples successfully amplified for Native American mitochondrial haplogroups, 15 samples (36.6%) successfully amplified for ABO markers. ABO genotypes over all amplifications, as well as the consensus genotype (assuming allelic drop out is common, heterozygous were assumed to be the correct genotype in case of discrepancy), are shown in Table 9. Table 10 lists observed genotypic and allelic frequencies for A, B, O1 and O1v for CA-SCL-38. Heterozygosity of observed genotype counts is not significantly different from expected at Hardy-Weinberg equilibrium under a Fisher's exact test (p=0.329).

Table 9 Mitochondrial haplogroups and ABO genotypes for the CA SCL-38 population

Sample	mt Haplogroup	Replicas	Rep 1	Rep 2	Rep 3	Rep 4	Rep 5	Rep 6	Consensus
B001	A	1	X						
B003	B	1	X						
B004	C	5	AO1	O1O1	AO1	O1O1	O1O1		**AO1**
B005	B	1	X						
B006	B	4	O1O1v	X	X	O1O1v			**O1O1v**
B008	D	1	X						
B009	D	1	X						
B010	D	4	O1O1v	X	X	O1O1v			**O1O1v**
B011	D	1	X						
B012	B	1	X						
B015	D	1	X						
B016	A	1	X						
B018	D	1	X						
B019	B	1	X						
B020	C	1	X						
B023	B	3	O1O1	O1O1v	O1O1				**O1O1**
B024	D	1	X						
B025	C	1	X						
B026	D	1	X						
B027	D	1	X						
B038	D	4	O1O1v	O1vO1v	O1O1v	O1O1v			**O1O1v**
B043	D	5	X	O1O1	X	O1O1	O1O1		**O1O1**
B048	D	6	O1O1	O1O1	O1O1	O1O1	O1O1	O1O1	**O1O1**
B050	D	1	O1O1v						**O1O1v**
B056	D	6	O1vO1v	X	X	X	X	O1vO1v	**O1vO1v**
B080	D	5	O1O1	X	X	AA	AA		**AA**
B097	N/A								
B110	N/A								
B111	N/A								
B112	N/A								
B113	N/A								
B114	B	1	X						
B115	N/A								
B116	N/A								
B117	C	1	X						
B118	C	1	X						
B119	N/A								
B130	N/A								
B131	N/A								
B134	D	5	O1O1	X	O1O1	O1vO1v	O1O1		**O1O1**
B136	N/A								
B138	N/A								
B141	N/A								
B143	N/A								
B144	N/A								

B145	N/A								
B146	N/A								
B161	N/A								
B162	N/A								
B164	N/A								
B165	N/A								
B168	D	1	X	X					
B170	D	1	O1O1						**O1O1**
B171	D	7	O1O1v	AO1v	O1O1v	AO1v	O1O1v	AO1v	**AO1v**
B173	N/A								
B174	N/A								
B177	D	1	X						
B178	N/A								
B190	N/A								
B191	N/A								
B192	C	1	X						
B193	D	2	O1O1	O1O1					**O1O1**
B194	C	1	X						
B195	B	1	X						
B196	B	1	X						
B206	N/A								
B217	D	1	X						
B233	C	4	O1O1	O1O1	O1O1	O1O1			**O1O1**

Table 10 ABO genotype and allele frequencies for the CA SCL-38 population. Expected genotype frequencies at Hardy-Weinberg equilibrium are shown next to observed genotypic frequencies

Genotypes	Observed	Expected HW
O1O1	0.467	0.401
O1vO1v	0.067	0.054
O1O1v	0.267	0.296
AO1	0.067	0.169
AO1v	0.067	0.062
AA	0.067	0.018

Alleles	Frequencies
O1	0.633
O1v	0.233
A	0.133

Cross populations comparisons

Results of the Fisher's exact test-pairwise comparisons between A, B and O frequencies of the CA-SCL-38 population and a panel of 61 contemporary Native American populations grouped as Athabaskan, Aleut-Eskimos, North American and South/Meso-American as well as Halverson and Bolnick's (2008) ancient population are shown in Figure 7. CA-SCL-38 is significantly distinguishable from some Eskimo populations (both Angmagssalik Eskimo populations, Aivilik Eskimos and Wainwright Eskimos), only Navaho-New Mexico from the Athabaskan group, and most of the South/Meso-American populations. CA-SCL-38 is significantly indistinguishable from all North American groups and from Halverson and Bolnick's (2008) ancient population.

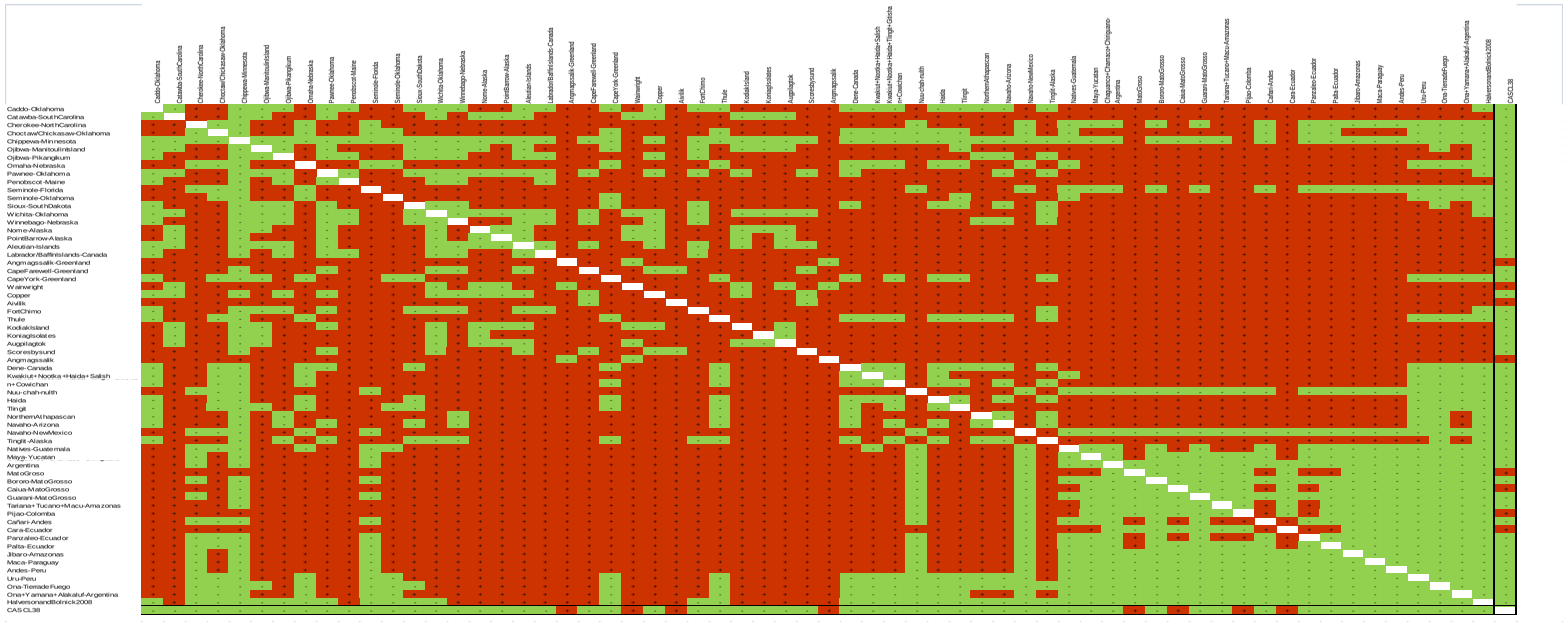

Figure 7 Pairwise Fisher's Exact test significance matrix for ABO frequencies. Red cells indicate significantly distinguishable populations, green cells indicate significantly indistinguishable populations. CA-SCL-38 is delimited by dark bars.

The Athabaskan, Aleut-Eskimos, North American and South American Native American populations seem to spread apart from global populations in a Principal Coordinate Analysis based on Fst distances calculated from ABO allele frequencies (Fig 8). CA-SCL-38 falls closer to other North American groups on the first coordinate, but it is also the most extreme negative value for the second coordinate.

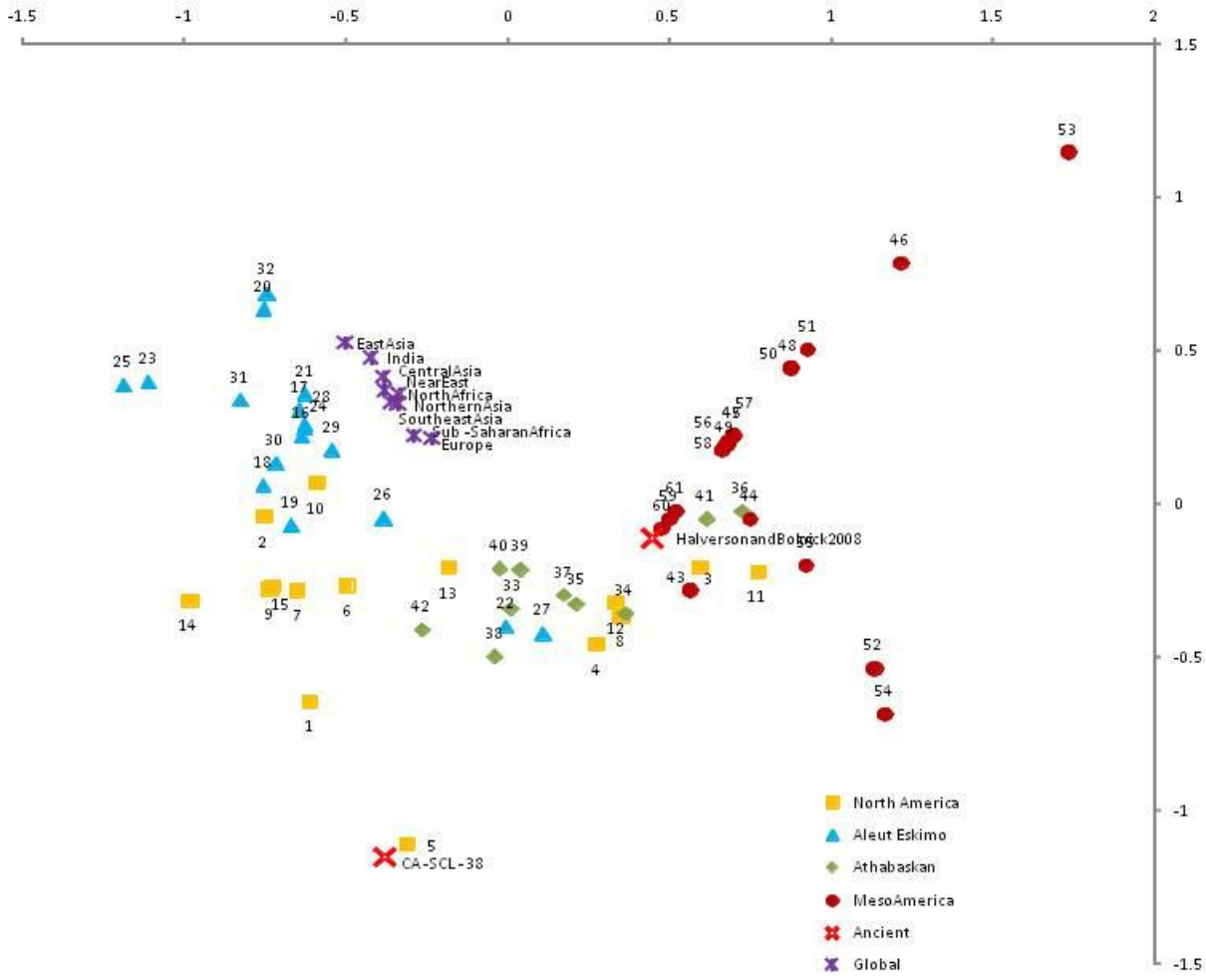

Figure 8 Principal Coordinate Analysis based on F_{st} distances calculated from ABO allele frequencies for 61 extant Native American populations as well as Halverson and Bolnick (2008) and CA-SCL-38 ancient populations. The first component explain 8.3% of the variance, the second component explains 3.2% of the variance.

Figure 8 Principal Coordinate Analysis based on F_{st} distances calculated from ABO allele frequencies for 61 extant Native American populations as well as Halverson and Bolnick (2008) and CA-SCL-38 ancient populations. The first component explain 8.3% of the variance, the second component explains 3.2% of the variance.

Modeling ABO diversity loss

Negative frequency dependent selection as modeled in this study is extremely robust in its ability to maintain variation at the ABO locus. At a population size of Ne>500 an equilibrium state is met after 10-20 generations (Fig 9). Following expected results, as the effective population size decreases the chance of allele loss increases. At effective population size 50, 16% of all simulations had led to loss of one allele (A or B in all cases). At effective population size 25, 82% of all simulations had lost at least one allele (A or B) and 8% resulted in the loss of

both the A and B alleles. At effective population sizes of 10 (and presumably lower) diversity was lost by either one (76%) or both alleles (23%) in nearly all replications (Fig 10).

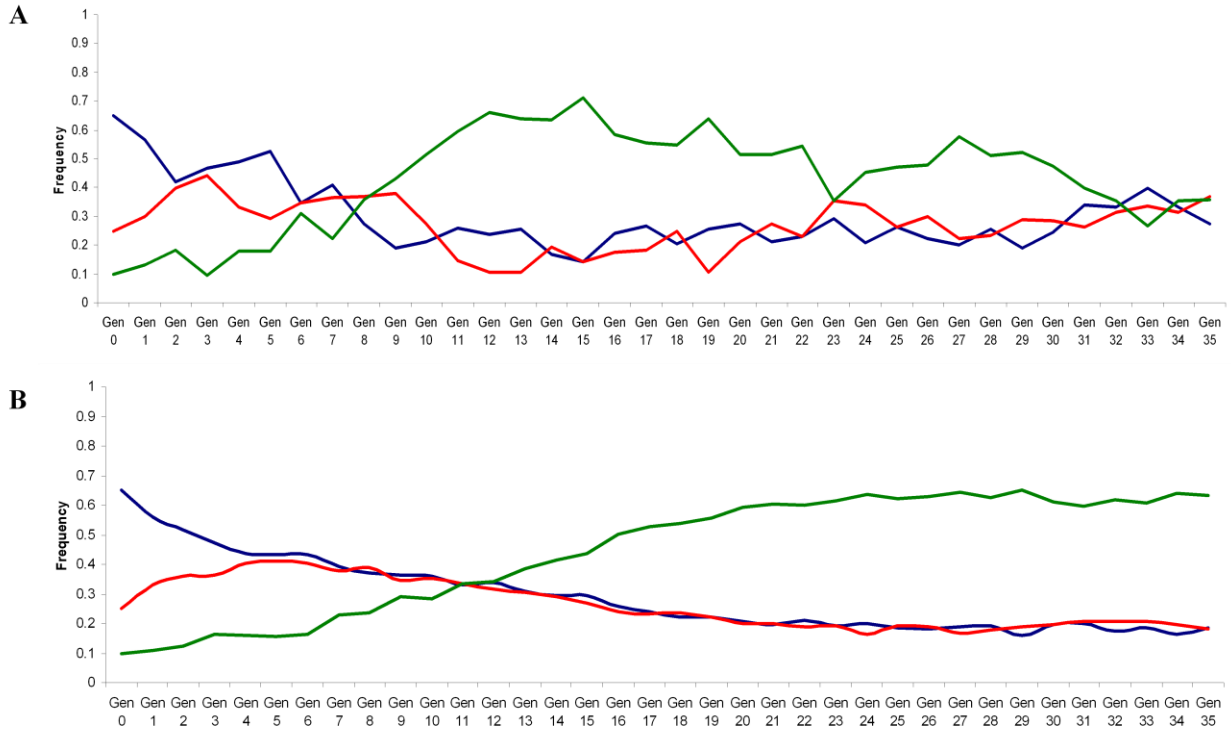

Figure 9 ABO evolution modeled under negative frequency dependent selection. Evolution of O allele (green), A allele (blue) and B allele (red) frequencies on a small population Ne= 50 (A) and a large population Ne= 500 (B). Notice how fluctuation between generations is much larger than on a small population.

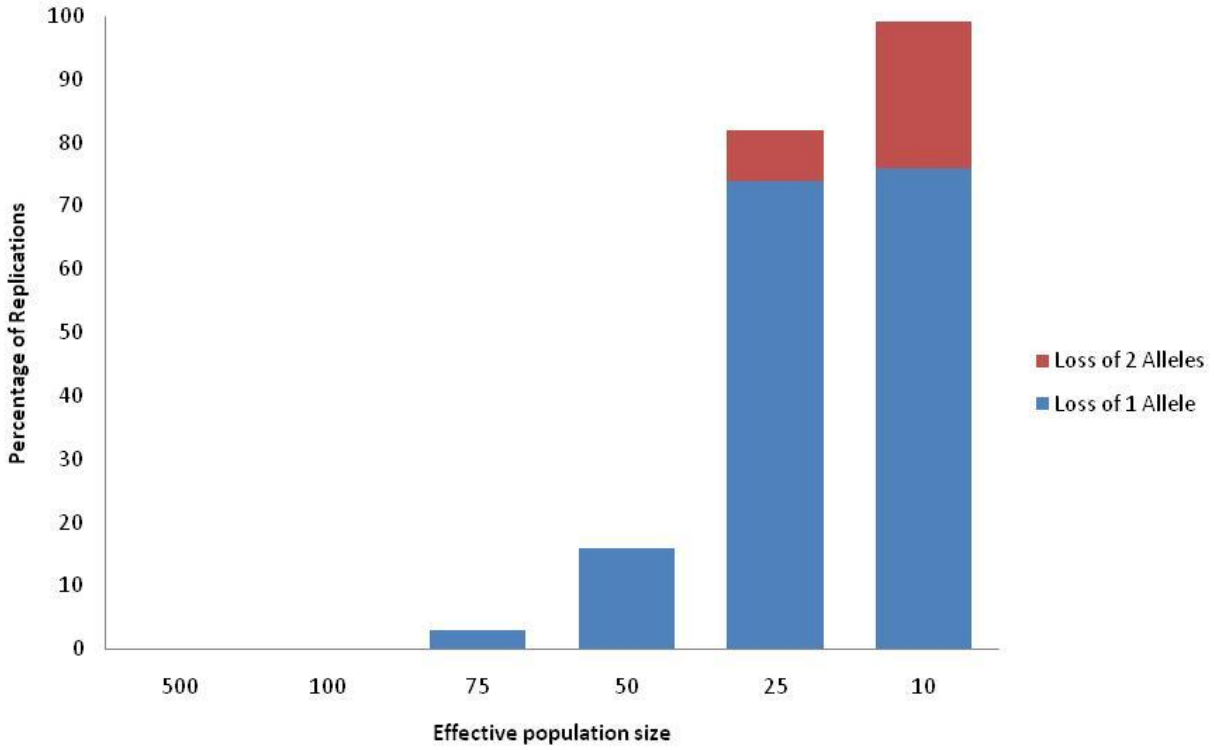

Figure 10 Loss of ABO alleles as modeled by decreasing the effective population size of a population under frequency dependent negative selection.

Discussion

Trouble-shooting ancient DNA

Ancient DNA has the potential to answer evolutionary questions by looking directly at how populations change over time. An abundance of available archaeological samples facilitates this process, thus highlighting the importance of the CA SCL-38 burial site.

When working with aDNA only a few samples will meet the conditions necessary to provide high-quality extractions, and even fewer will preserve high-quality nuclear DNA. Using quantitative PCR, Schwarz et al. (2009) estimated the ratio of mitochondrial to nuclear DNA preserved in ancient mammoth bone samples to range from 245:1 to 17,369:1 copies, while a

modern elephant sample would contain a ratio of around 150:1. The precise reasoning of why mtDNA is preferentially preserved during diagenesis is not known, but it is of enormous importance to those where research aims to study variation in ancient nuclear genes, such as ABO. Additional controls and multiple replications of results are then necessary when working with ancient autosomal markers.

Laboratory controls

In an effort to narrow down the number of CA SCL-38 samples capable of yielding good quality copies of the ABO gene, samples were first screened for preserved mtDNA. Successful identification of a sample to a Native American haplogroup was used as an indicator of high quality DNA present and at the same time as additional control indicating the results originated from endogenous aDNA as opposed to exogenous modern contamination.

As a good measure of control, all ancient DNA researchers in the Kemp laboratory have been haplotyped. It should be noted that my mitochondrial DNA belongs to subhaplogroup A2, part of the A haplogroup. To ensure all data is from an ancient source, negative and positive controls were added to all PCR reactions. Table 9 lists the frequency of mitochondrial haplotypes from the CA SCL-38 population, demonstrating a notorious low frequency of haplogroup A (5.3%). If modern laboratory contamination was the source of these results, many non-Native American haplogroups as well as higher frequency of A haplotypes would be expected. A low frequency of haplogroup A is expected, as shown by figure 6: CA-SCL-38 is significantly indistinguishable from other Californian and Great Basin populations based on mitochondrial haplogroups frequencies, specifically Cahuilla, Costanoan, Sahaptain, Serano/Vanyume, Sierra Miwok, Tubat and Yokuts, further confirming the validity of the data.

Allelic Drop-Out

An additional problem when working with nuclear autosomal loci in ancient or modern samples is allelic dropout (one allele of a heterozygous individual is not amplified during a positive polymerase chain reaction). This is a fairly recurrent problem in modern samples with low DNA yields as well, such as samples collected from stool or hair (Broquet and Petit, 2004). Given the nature of aDNA, and in particular the low copy number of nuclear aDNA, allelic dropout becomes a significant problem. For example the ancient population studied by Halverson and Bolnick (2008) for variation at the ABO locus was found not to be in Hardy-Weinberg (HW) equilibrium because of an excess of homozygotes, which might be explained by allelic dropout. However it should be noted that Halverson and Bolnick (2008) replicated each result four independent times, as I have done here to help control for allelic dropout. While demographic phenomena such as inbreeding can reduce heterozygosity, in their case observed low heterozygosity is probably due to their limited sample size as their "population" is a construct from three isolated sites as opposed to CA-SCL-38.

For genes behaving neutrally, such as the various O alleles in the ABO locus, homozygote and heterozygote frequencies are expected to follow a Hardy-Weinberg distribution (Griffiths et al., 2002). The ABO locus violates Hardy-Weinberg postulates as it is a gene under selection; however, testing against expected values of Hardy-Weinberg is a simple test that can signal artificial loss of heterozygotes by allelic dropout. Heterozygosity at the ABO locus in the CA SCL-38 site is not significantly different from the expected HW distribution using a Fisher's exact test of genotypic counts ($p=0.329$). As heterozygosity is not significantly different than expected at the CA SCL-38 site, genotypic polymorphism at the ABO locus seems to have reliable recovered making allelic dropout unlikely.

Comparison of ancient ABO frequencies with contemporary Native American frequencies

The observation that most Native Americans exhibit almost exclusively blood group O and that variation from this pattern is the result of European admixture is not supported by this study. Based on the results of the pairwise Fisher's Exact tests (Fig 7), it has been demonstrated that a) some Eskimo populations are significantly distinguishable from most other Native American populations for their higher ABO allele diversity b) most South and Mesoamerican populations are significantly distinguishable to other North Native American populations for their lower ABO allele diversity, and c) although CA-SCL-38 presents a low frequency of A alleles, it is still significantly indistinguishable from North American populations (including Athabaskan groups) populations. This suggests ABO patterns of diversity may have been similar across America before European contact, at least in North America.

While CA-SCL-38 groups together with other North American populations on the first coordinate of the PCoA plot, it is also the most extreme value for the second coordinate (Figure 8). However the first and second coordinates only explain 8.3% and 3.2% of the variance respectively, an overall low value.

CA-SCL-38 remains significantly indistinguishable from all North American groups based on the Fisher's exact test. Furthermore the highly structured pattern of ABO polymorphism seen in figure 8 between Athabaskan, Aleut-Eskimos, North American and South/Mesoamerican Native American populations suggests isolation by distance as proto-Native Americans migrated deeper into the American continent. This also accounts for the much higher variance when referenced against global averaged samples.

These patterns taken together are suggestive of demographic factors playing an important role as Native Americans expanded into the continent away from the Northern hemisphere

subsequent to any bottlenecks in the original proto-American population from Beringia. It is highly likely that drift within colonizing bands would have reduced ABO diversity independently from each other as groups diverged, but in contrast, ABO diversity was not lost during the initial incubation period. The same pattern of structuring by isolation is thought to account for the high linguistic diversity within Native Americans (Nettle, 1999).

Modeling diversity loss in Native American populations

If ABO allele diversity is constantly being maintained by selection, genetic drift can potentially move a population away from equilibrium frequencies. Because the effect of genetic drift works at a rate inversely proportional to the effective population size, a measure of genetic diversity loss from the expected frequencies based on frequency dependent selection can be used to predict the effective population size when that population experienced diversity loss.

Negative frequency dependent selection as modeled in this study is extremely effective. It will maintain similar ABO frequencies at a high polymorphic equilibrium in large populations ($Ne>500$) and even in somewhat smaller ones ($Ne<75$) (Fig 11). But as the effective population size becomes lower ($Ne=50$), 16% of all replications had lost one allele (A or B in all cases). At even smaller sizes ($Ne=25$), 82% of all replications had lost at least one allele (A or B) and 8% had lost both the A and B alleles. At extremely small sizes ($Ne=10$ and lower) populations are expected to lose either one (76%) or both alleles (23%) in nearly all cases.

In previous studies, effective population size has been used as a parameter to estimate date of entry into the Americas. While various models have used small population sizes ($Ne<1500$) for the overall population, this number would likely reflect the effective population at the Beringian incubation stage (Mulligan et al., 2004). Based on the ABO model, because Eskimo populations retained the A, B and O alleles and most North American groups retain at

least A and O alleles, the effective population size of the founder population during the Beringian incubation stage would have been great enough to maintain A, B and O alleles (Ne>75).

If allele loss in individual Native American groups resulted from isolation after the initial entrance of humans into the Americas however, individual founding groups would have lost diversity independently. Compared to the model, populations of an effective population size of 50 or more would have been able to maintain diversity in 84% of cases. This suggests as populations moved south into the continent and became more isolated their effective population sizes dropped below 50.

A similar model of "isolation with migration" (Hey, 2005) indicates similar small effective population sizes (Ne= 70) for geographically isolated populations, a size congruent with this model given that a constantly larger than 50 effective population size in isolated populations, say migrants into South America for example, would be sufficient to maintain ABO diversity of three (A, B, O) or at least two alleles (A, O).

Finally, the high interpopulation variance among Native American populations at the ABO locus reflects independent loss-of-diversity events in different founding groups. As such, the variance in final frequencies of model replications provides a good comparison against the variance exhibited by modern Native American groups (Fig 10).

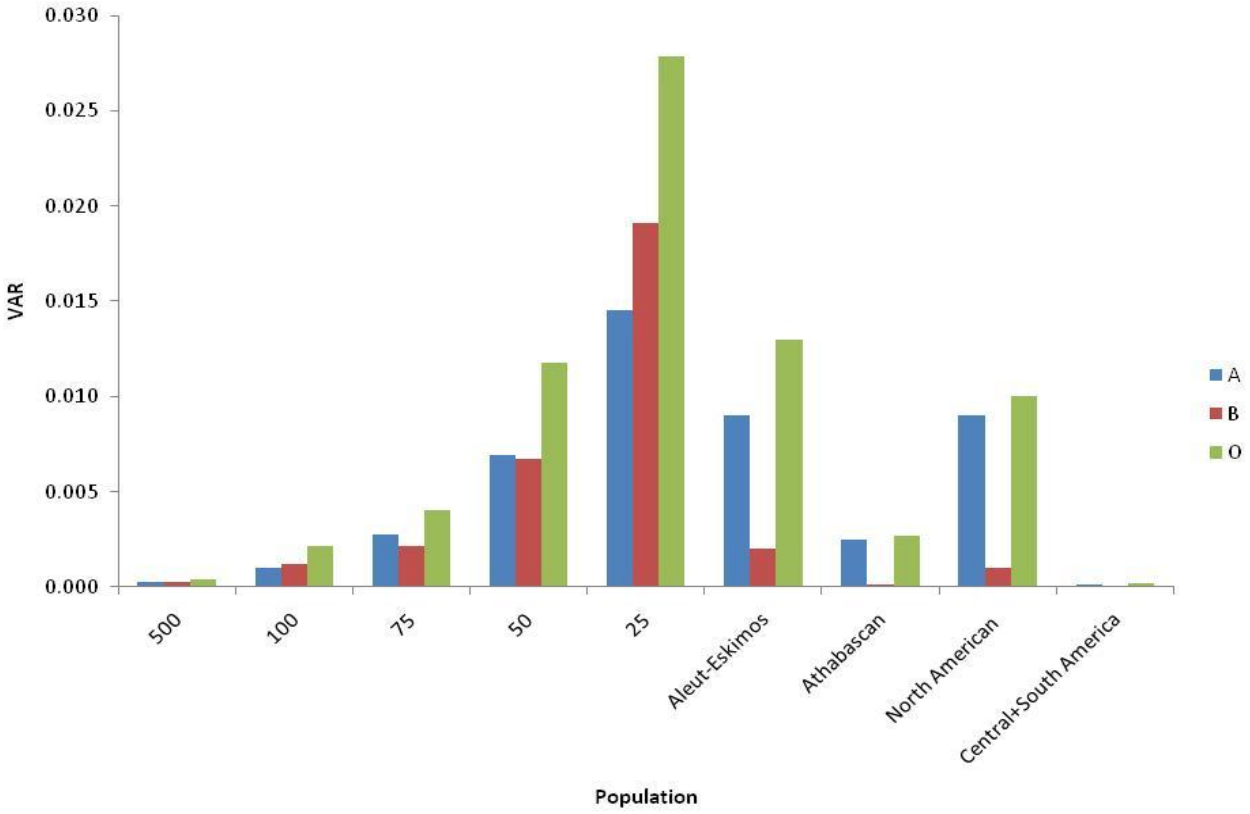

Figure 11 Variance of ABO frequencies between 100 replications of the modeled ABO evolution and within variance in modern Native American groups.

Conclusions: Reconstructing Native American Origins

Negative frequency-dependent selection driven by pathogens of different infection strategies has homogenized and maintained variation at the ABO locus in most human populations. Although it is possible to skew this frequency by intense epidemics of a specific pathogen, allele loss is uncommon. The current highly monomorphic frequencies of some Native Americans, and therefore the breakdown of the negative frequency dependent selection system, as well as the higher variance among Native Americans compared to other world populations is most likely to reflect the small effective population size (Ne= 25-50) of founder groups as they swiftly expanded through a new and uncolonized continent and the huge role drift played in shaping their genetic make-up.

While as a whole, Native Americans exhibit a lower diversity across all loci (Wang et al., 2007) and at the ABO locus, their within groups variance is tremendous. This is again a result of evolution in isolation from each other, which may also be reflected in their enriched linguistic and cultural diversity (Nettle, 1999). Far from a simple evolutionary history, Native Americans have undergone dynamic diversification in a short time span.

It is interesting that Aleut-Eskimo groups exhibit high polymorphism of A, B and O alleles, while at the same time exhibiting the same high within group variance characteristic of the other Native American groups. On the same note, most but not all Athabaskan and related Northwest Coast groups present high frequencies of the A allele but not B. This overall pattern fits the hypothesis that loss of diversity occurred at the individual population level as a result of increased isolation by expanding deeper into the American continent, as Northwest groups possibly settled shortly after their ancestors left Beringia. Alternatively, it has been suggested that Eskimos may have entered the Americas after the original Paleo-Indian migration and as so, both populations may have carried different diversity of ABO alleles. Both scenarios however are indistinguishable based on the available ABO data for this study.

The introduction of smallpox has been argued to have reduced diversity at the ABO locus in Native Americans by driving A alleles to be lost. The CA SCL-38 pre-European contact population demonstrates that ABO diversity was reduced even previous to European contact. Furthermore, since ABO mediates immune response through high diversity, it appears that it was only by chance, in the end, that Native American populations lost variation that could have offered some resistance the introduction of pathogens following European contact.

Negative density dependent selection of the ABO locus failed to maintain polymorphism in Native Americans. While genetic drift is the most likely cause, other explanations such as the

absence of pathogens (or the specific pathogens that reduce O allele fitness) in the newly colonized areas of America or very strong selection against A/B antigens in proto-Americans while still in Asia may have also disrupted selection to maintain diversity. Testing such hypotheses is beyond the scope of this study.

Works Cited

Achilli A, Perego UA, Bravi CM, Coble MD, Kong QP, Woodward SR, Salas AT, A. , and Bandelt HJ (2008) The phylogeny of the four pan-American MtDNA haplogroups: Implications for evolutionary and disease studies. PLoS ONE *3:*e1764.

Adalsteinsson S (1985) Possible changes in the frequency of the human ABO blood groups in Iceland due to smallpox epidemics selection. Annals of Human Genetics *49:*275-281.

Bellifemine V (1997) Mortuary variability in prehistoric Central California: a statistical study of the yukisma site, CA SCL-38. Master of Arts thesis, San Jose University, San Jose.

Brace CL (2005) Race is a four letter word. New York: Oxford University Press.

Broquet T, and Petit E (2004) Blackwell Publishing, Ltd. Quantifying genotyping errors in noninvasive population genetics. Molecular Ecology *13:*3601–3608.

Calafell F, Roubinet F, Ramírez-Soriano A, Saitou N, Bertranpetit J, and Blancher A (2008) Evolutionary dynamics of the human ABO gene. Human Genetics *124:*123-135.

Carlyle SW, Parr RL, Hayes MG, and O'Rourke DH (2000) Context of Maternal Lineages in the Greater Southwest. Am. J. Phys Anthropology *113:*85-101.

Cavalli-Sforza LL, Menozzi P, and Piazza A (1992) The History and Geography of Human Genes. New Jersey: Princeton University Press.

Chen FC, and Li WH (2001) Genomic Divergences between Humans and other Hominoids and the Effective Population Size of the Common Ancestor of Humans and Chimpanzees. American Journal of Human Genetics *68*.

Daniels G (2002) Human blood groups. Oxford: Blackwell.

De La Cruz I, Gonzalez-Oliver A, Kemp Brian M, Roman Juan A, Smith David G, and Torre-Blanco A (2008) Sex Identification of Children Sacrificed to the Ancient Aztec Rain Gods in Tlatelolco. Current Anthropology *49:*519-526.

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

Estrada-Mena B, Estrada FJ, Ulloa-Arvizu R, Guido M, Méndez R, Coral R, Canto T, Granados J, Rubí-Castellanos R, Rangel-Villalobos H, and García-Carrancá A (2009) Blood group O alleles in Native Americans: Implications in the peopling of the Americas. American Journal of Physical Anthropology *9999:*NA.

Excoffier LGL, and S. Schneider (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online *1:*47-50.

Gagneux P, and Varki A (1999) Evolutionary considerations in relating oligosaccharide diversity to biological function. Glycobiology *9:*747-755.

Gilbert MTP, Jenkins DL, Gotherstrom A, Naveran N, Sanchez JJ, Hofreiter M, Thomsen PF, Binladen J, Higham TFG, Yohe RM, II, Parr R, Cummings LS, and Willerslev E (2008) DNA from Pre-Clovis Human Coprolites in Oregon, North America. Science *320:*786-789.

Glass RI, Holmgren JAN, Haley CE, Khan MR, Svennerholm A, Stoll BJ, Hossain KMB, Black RE, Yunus M, and Barua D (1985) Predisposition for cholera of individuals with o blood group possible evolutionary significance. Am. J. Epidemiol. *121:*791-796.

Gonzalez-Oliver A, Marquez-Morfin L, Jimenez JC, and Torre-Blanco A (2001) Founding Amerindian Mitochondrial DNA Lineages in Ancient Maya From Xcaret, Quintana Roo. Am. J. Phys Anthropology *116:*230-235.

Griffiths AJF, Gelbart WM, Lewontin RC, and Miller JH (2002) Modern Genetic Analysis. New York: W. H. Freeman and Company.

Halverson MS, and Bolnick DA (2008) An Ancient DNA Test of a Founder Effect in Native American ABO Blood Group Frequencies. American Journal of Physical Anthropology *137:*342-347.

Hänni C, Brousseau T, Laudet V, and Stehelin D (1995) Isopropanol Precipitation Removes PCR Inhibitors from Ancient Bone Extracts. Nucleic Acids Research *23:*881-882.

Hey J (2005) On the number of New World founders: a population genetic portrait of the peopling of the Americas. PLoS Biol *3*.

Hummel S, Schmidt D, Kahle M, and Herrmann B (2002) AB0 blood group genotyping of ancient DNA by PCR-RFLP. International Journal of Legal Medicine *116:*327-333.

Kaestle F, and Smith D (2001) Ancient mitochondrial DNA evidence for prehistoric population movement. American Journal of Physical Anthropology *115:*1-12.

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, and Smith DG (2007) Genetic analysis of early holocene skeletal remains from Alaska and its implications for the settlement of the Americas. American Journal of Physical Anthropology *132:*605-621.

Kemp BM, Monroe C, and Smith DG (2006) Repeat silica extraction: a simple technique for the removal of PCR inhibitors from DNA extracts. Journal of Archaeological Science *33:*1680-1689.

Kemp BM, Resendez A, Roman-Berrelleza JA, Malhi RS, and Smith DG (2005) An Analysis of Ancient Aztec mtDNA from Tlatelolco: Pre-Columbian Relations and the Spread of Uto-Aztecan. In DM Reed (ed.): Biomolecular archaeology. Genetic approaches to the past. Carbondale, IL: Center for archaeological investigations, Southern Illinois University, pp. 22-46.

Kemp BM, and Smith DG (2005) Use of bleach to eliminate contaminating DNA from the surface of bones and teeth. Forensic Science International *154:*53-61.

Kent W, Sugnet C, Furey T, Roskin K, Pringle T, Zahler A, and Haussler D (2002) The human genome browser at UCSC. Genome Res *12* 996-1006.

Landsteiner K (1901) Über Agglutinationserscheinungen normalen menschlichen Blutes. Wien Klin Wochenschr *14:*1132–1134.

Legendre P, and Anderson MJ (1998) Program DistPCoA: Département de sciences biologiques, Université de Montréal, pp. 10.

Llop E, Henríquez H, Moraga M, Castro M, and Rothhammer F (2006) Brief communication: Molecular characterization of O alleles at the ABO locus in Chilean Aymara and Huilliche Indians. American Journal of Physical Anthropology *131:*535-538.

Lorenz JG, and Smith DG (1996) Distribution of four founding mtDNA haplogroups among Native North Americans. American Journal of Physical Anthropology *101:*307-323.

Loring BC (2005) Race is a four letter word: The genesis of the concept. New York, New York: Oxford University Press.

Malhi RS, Kemp BM, Eshleman JA, Cybulski J, Smith DG, Cousins S, and Harry H (2007) Mitochondrial haplogroup M discovered in prehistoric North Americans. Journal of Archaeological Science *34:*642-648.

Malhi RS, Mortenson HM, Eshleman JA, Kemp BM, Lorenz JG, Kaestle FA, Johnson JR, Gorodezky C, and Smith DG (2003) Native American mtDNA prehistory in the American Southwest. American Journal of Physical Anthropology *120:*108-124.

Mendel G (1866) Versuche über Plflanzen-hybriden. Verhandlungen des naturforschenden Vereines in Brünn, Bd. IV für das Jahr 1865 *Abhand-lungen, 3–47*.

Molnar S (2002) Human Variation. Upper Saddle River, New Jersey: Prentice Hall.

Morley S (1997) The Paleodemography of the Yukisma site, CA SCL-38. Master of Arts thesis, University of San Jose, San Jose.

Mourant AE, Kopec AC, and K. D-S (1976) The distribution of the human blood groups and other polymorphisms. London: Oxford University Press.

Mulligan CJ, Hunley K, Cole S, and Long JC (2004) Population genetics, history, and health patterns in Native Americans. Annual Review of Genomics and Human Genetics *5:*295-315.

Nettle D (1999) Linguistic diversity of the Americas can be reconciled with a recent colonization. Proc. Natl. Acad. Sci *96:*3325-3329.

Parr RL, Carlyle SW, and O'Rourke DH (1996) Ancient DNA analysis of Fremont Amerindians of the Great Salt Lake Wetlands. American Journal of Physical Anthropology *99:*507-518.

Patterson KB, and Runge T (2002) Smallpox and the Native American. The American Journal of the Medical Sciences *323:*216-222.

Preece AF, Strahan KM, Devitt J, Yamamoto F-i, and Gustafsson K (2002) Expression of ABO or related antigenic carbohydrates on viral envelopes leads to neutralization in the presence of serum containing specific natural antibodies and complement. Blood *99:*2477-2482.

Ray N, Wegmann D, Fagundes NJR, Wang S, Ruiz-Linares A, and Excoffier L (2009) A statistical evaluation of models for the initial settlement of the American continent emphasizes the importance of gene flow with Asia. Mol Biol Evol*:*msp238.

Rother RP, and Squinto SP (1996) The [alpha]-Galactosyl Epitope: A Sugar Coating That Makes Viruses and Cells Unpalatable. Cell *86:*185-188.

Sabeti PC, Walsh E, Schaffner SF, Varilly P, Fry B, Hutcheson HB, Cullen M, Mikkelsen TS, Roy J, Patterson N, Cooper R, Reich D, Altshuler D, O'Brien S, and Lander ES (2005) The Case for Selection at *CCR5*-Δ*32*. PLoS Biol *3:*e378.

Saitou N, and Yamamoto F (1997) Evolution of primate ABO blood group genes and their homologous genes. Mol Biol Evol *14:*399-411.

Schroeder KB, Jakobsson M, Crawford MH, Schurr TG, Boca SM, Conrad DF, Tito RY, Osipova LP, Tarskaia LA, Zhadanov SI, Wall JD, Pritchard JK, Malhi RS, Smith DG, and Rosenberg NA (2009) Haplotypic background of a private allele at high frequency in the Americas. MBE *Advanced Access February 12, 2009*.

Schroeder KB, Schurr TG, Long JC, Rosenberg NA, Crawford MH, Tarskaia LA, Osipova LP, Zhadanov SI, and Smith DG (2007) A private allele ubiquitous in the Americas. Biol Lett *3*.

Schwarz C, Debruyne R, Kuch M, McNally E, Schwarcz H, Aubrey AD, Bada J, and Poinar H (2009) New insights from old bones: DNA preservation and degradation in permafrost preserved mammoth remains. Nucl. Acids Res.*:*gkp159.

Seymour RM, Allan MJ, Pomiankowski A, and Gustafsson K (2004) Evolution of the Human ABO Polymorphism by Two Complementary Selective Pressures. Proceedings: Biological Sciences *271:*1065-1072.

Sharon R, and Fibach E (1991) Quantitative flow cytometric analysis of ABO red cell antigens. Cytometry *12:*545-549.

Smith DG, Lorenz JG, Rolfs BK, Bettinger RL, Green B, Eshleman JA, Schultz B, and Malhi RS (2000) Implications of the distributions of Albimin Naskapi and Albumin Mexico for new world prehistory. American Journal of Physical Anthropology *111:*557-572.

Stone AC, and Stoneking M (1998) mtDNA Analysis of a Prehistoric Oneota Population: Implications for the Peopling of the New World. The American Journal of Human Genetics *62:*1153-1170.

Swerdlow DL, Mintz ED, Rodriguez M, Tejada E, Ocampo C, Espejo L, Barrett TJ, Petzelt J, Bean NH, Seminario L, and Tauxe RV (1994) Severe Life-Threatening Cholera Associated with Blood Group O in Peru: Implications for the Latin American Epidemic. The Journal of Infectious Diseases *170:*468-472.

Szathmary EJE (1979) Blood Groups of Siberians, Eskimos Subarctic and Northwest Coast Indians: The problem of origins and genetic relationships. In WS Laughlin and AB Harper (eds.): The First Americans: Origins, Affinities, and Adaptations, pp. 185-209.

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, and Malhi RS (2007) Beringian Standstill and Spread of Native American Founders. PLOS One *9:*e829 (1-6).

Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM, and Wallace DC (1993) Asian affinities and continental radiation of the four founding Native American mtDNAs. Am J Hum Genet *53:*563-590.

Tremblay M, and Vézina H (2000) New Estimates of Intergenerational Time Intervals for the Calculation of Age and Origins of Mutations. The American Journal of Human Genetics *66:*651-658.

Wang S, Jr CML, Jakobsson M, Ramachandran S, Ray N, Bedoya G, Rojas W, Parra MV, Molina JA, Gallo C, Mazzotti G, Poletti G, Hill K, Hurtado AM, Labudo D, Klitz W,

Barrantes R, Bortolini MCt, Salzano FM, Petzl-Erler ML, Luiza T. Tsuneto16 EL, Rothhammer F, Excoffier L, Feldman MW, Rosenberg NA, and Ruiz-Linares As (2007) Genetic Variation and Population Structure in Native Americans. PLOS Genetics *3:*2049-2067.

Yamamoto F-i, Clausen H, White T, Marken J, and Hakomori S-i (1990) Molecular genetic basis of the histo-blood group ABO system. Nature *345:*229-233.

Yip SP (2002) Sequence variation at the human ABO locus. Annals of Human Genetics *66:*1-27.