

2013

## Brief communication: Evolution of a specific O allele (O1vG542A) supports unique ancestry of Native Americans

Fernando A. Villanea

Deborah A. Bolnick

Cara Monroe

Rosita Worl

Rosemary Cambra, et al.

# Brief Communication: Evolution of a Specific O Allele (O1v<sup>G542A</sup>) Supports Unique Ancestry of Native Americans

Fernando A. Villanea,<sup>1</sup> Deborah Bolnick,<sup>2,3</sup> Cara Monroe,<sup>1,4,5</sup> Rosita Worl,<sup>6</sup> Rosemary Cambra,<sup>7</sup> Alan Leventhal,<sup>8</sup> and Brian M. Kemp<sup>1,4\*</sup>

<sup>1</sup>*School of Biological Sciences, Washington State University, Pullman, WA 99164-4910*

<sup>2</sup>*Department of Anthropology, University of Texas, Austin, TX 78712-1723*

<sup>3</sup>*Population Research Center, University of Texas, Austin, TX 78712-1723*

<sup>4</sup>*Department of Anthropology, Washington State University, Pullman, WA 99164-4236*

<sup>5</sup>*Department of Anthropology, University of California-Santa Barbara, Santa Barbara, CA 93106-3210*

<sup>6</sup>*Sealaska Heritage Institute, One Sealaska Plaza, Suite 400, Juneau, AK 99801*

<sup>7</sup>*Tribal Chairwoman, Muwekma Ohlone Tribe, San Jose, CA 95131*

<sup>8</sup>*University College of Social Science, San Jose State University, San Jose, CA 95112*

**KEY WORDS** ancient DNA; ABO blood group; entrance of humans into the Americas; ancestry informative markers

**ABSTRACT** In this study, we explore the geographic and temporal distribution of a unique variant of the O blood group allele called O1v<sup>G542A</sup>, which has been shown to be shared among Native Americans but is rare in other populations. O1v<sup>G542A</sup> was previously reported in Native American populations in Mesoamerica and South America, and has been proposed as an ancestry informative marker. We investigated whether this allele is also found in the Tlingit and Haida, two contemporary indigenous populations from Alaska, and a pre-Columbian population from California. If O1v<sup>G542A</sup> is present in Na-Dene speakers (i.e., Tlingits), it would indicate that Na-Dene speaking groups share ancestry with other Native American groups and support a Beringian origin of the allele, consistent with the Beringian Incubation Model.

If O1v<sup>G542A</sup> is found in pre-Columbian populations, it would further support a Beringian origin of the allele, rather than a more recent introduction of the allele into the Americas via gene flow from one or more populations which have admixed with Native Americans over the past 5 centuries. We identified this allele in one Na-Dene population at a frequency of 0.11, and one ancient California population at a frequency of 0.20. Our results support a Beringian origin of O1v<sup>G542A</sup>, which is distributed today among all Native American groups that have been genotyped in appreciable numbers at this locus. This result is consistent with the hypothesis that Na-Dene and other Native American populations primarily derive their ancestry from a single source population. *Am J Phys Anthropol* 000:000–000, 2013. © 2013 Wiley Periodicals, Inc.

Native American branches of the human mitochondrial tree are differentiated from their Asian sister branches by a number of unique mutations (Tamm et al., 2007; Achilli et al., 2008; Fagundes et al., 2008b; Perego et al., 2009, 2010). This pattern is consistent with the “Beringian Incubation Model” (or “BIM”), which posits a period of isolation of the proto-Native American population prior to entering the Americas, during which this differentiation accumulated via mutation and genetic drift (Tamm et al., 2007; Fagundes et al., 2008a; Kitchen et al., 2008; Mulligan et al., 2008). While most Native Americans are the direct maternal descendants of a single Asian source population that migrated into the Americas approximately 15,000–20,000 years before present (YBP) [see recent review by Kemp and Schurr (2010)], mitochondrial DNA (mtDNA) evidence also supports more recent migration of humans into North America carrying subhaplogroups D2a and D3, which is largely associated with Aleut and Eskimo prehistory (Tamm et al., 2007; Crawford et al., 2010). The BIM does not specify the number of migrations into the American continents, as multiple waves could have originated from the same Beringian population.

If the BIM is correct, it predicts that a number of Native American ancestry informative markers (AIMs) exist across the genome that evolved prior to the initial entry of humans into the Americas and, thus, should be

widely distributed in both North and South American populations. The 9 repeat allele (9RA) at the D9S1120 locus represents one such marker, as it is has been observed in all Native American populations genotyped at the locus, yet is absent from all other world-wide populations that have been similarly screened, except for the Chukchi and Koryaks across the Bering Strait (Schroeder et al., 2007, 2009). The distribution of the Y-chromosome M3 single nucleotide polymorphism (SNP) (Underhill et al., 1996), which defines the Q1a3a1a branch of the human Y-chromosome tree,

Additional Supporting Information may be found in the online version of this article.

Grant sponsors: Washington State University (to B.M.K.), University of Texas at Austin (to D.A.B.), and Sealaska Heritage Institute.

\*Correspondence to: Brian M. Kemp, Department of Anthropology, Washington State University, Pullman, WA 99164. E-mail: bmkemp@wsu.edu

Received 18 February 2013; accepted 16 April 2013

DOI: 10.1002/ajpa.22292

Published online 00 Month 2013 in Wiley Online Library (wileyonlinelibrary.com).

largely mirrors the geographic pattern of the 9RA (Underhill et al., 1996; Karafet et al., 1997, 1999, 2001; Lell et al., 1997, 2002; Bortolini et al., 2003; Zegura et al., 2004; Dulik et al., 2012a, 2012b). The presence of the 9RA and M3 SNP in Na-Dene and Aleut-Eskimo speakers is also important because it supports a shared origin of these two groups with all other Native American populations. Because multiple waves of migration could have originated from the same ancestral population at different points in time, and because of the random nature of genetic drift, it is possible for populations that have diverged from a common ancestor to exhibit different mitochondrial markers and yet share a vast part of their nuclear genome. The strength of genetic drift is four times greater on the mitochondrial genome than it is on any autosomal DNA locus. Additional Native American AIMs have been identified, such as the ABCA1 (ATP-binding cassette transporter A1), C230 variant (Acuna-Alonzo et al., 2010; Hunemeier et al., 2012), or Albumin types Mexico and Naskapi (Smith et al., 2000), which are very widespread, but these probably evolved after the entrance of humans into the Americas and, thus are not informative to the BIM.

An alternative hypothesis, originally posited by Greenberg et al. (1986), details separate source populations for “Amerinds,” Na-Denes, and Aleut-Eskimos, groups that are argued to have entered the Americas in that order. A recent genome-wide study by Reich et al. (2012) has found support for at least three streams of gene flow from Asia into the Americas. Most Native Americans were found to descend from a single source population, which Reich et al. (2012) termed “First American.” Reich et al. (2012) found that Na-Dene and Eskimo-Aleut populations derive most their genomes from the First American stream, but each also traces some ancestry to additional streams of gene flow into the Americas (where the Na-Dene Chipewyan derive ~10% of their ancestry from a second stream and Eskimo-Aleuts derive ~43% of their ancestry from a third stream). The First American population may, at first glance, appear nearly synonymous with Greenberg et al.’s (1986) Amerind, a controversial linguistic grouping (Nichols, 1990; Campbell, 1997). However, because the majority of Na-Dene and Eskimo-Aleut ancestry originates from the First American population (contrary to Greenberg et al.’s hypothesis), First American is not equivalent to or distinct from Na-Dene and Eskimo-Aleut (as Amerind is). However, because all Native American populations are largely descended from the same ancestral population, and because they diverged recently from Asian populations, it is expected they carry a majority of shared polymorphisms. This is the case for genome-wide analyses, in which the majority of polymorphisms have little informative power for resolving the specific histories of Native American populations. Focusing on polymorphisms that are unique to the Americas and absent in Asian populations may be more informative for resolving the specific relationships between Native American populations.

Early evidence that Native Americans are genetically distinct from other human populations was found in the ABO blood group system. Many Native American populations are fixed for O alleles, a pattern that is not found in any other human populations (Swerdlow et al., 1994; Molnar, 2002; Llop et al., 2006). While the O allele is not the only allele present in Native American populations, there is a very strong North-South decline in the frequencies of the A and B alleles, which are found

predominantly in populations belonging to the Na-Dene and Aleut-Eskimos linguistic groups (Szathmary, 1979). Indigenous populations of South America are fixed for the O allele (Mourant et al., 1976).

Most O alleles evolved from A alleles by a deletion in position 261 of exon 6 of the ABO gene. This mutation “Δ261” creates a premature stop codon, resulting in a truncated protein lacking glycosyltransferase function (Gagneux and Varki, 1999). Although, O alleles are not deleterious, the shortened protein is no longer functional, and thus the gene accumulates mutations in a neutral fashion. Among all human populations there are at least 40 neutral variants of the O allele (Yip, 2002).

Sequence level data has provided additional resolution of O allele diversity exhibited by Native Americans (Estrada-Mena et al., 2010). Some alleles may be restricted to a single population, such as O05 and Ov7 that have, thus far, been exclusively described in the Cayapa of Ecuador and O32 and O33 in the Aymara of the Bolivian Andes (Estrada-Mena et al., 2010). Other less derived O alleles, for example, O1 and O1v, are shared between all prehistoric and contemporary Native American populations sampled to date (Halverson and Bolnick, 2008; Estrada-Mena et al., 2010; Georges et al., 2012), as well as with Old World populations (Chester and Olsson, 2001). One allele of particular interest, dubbed “O1v<sup>G542A</sup>,” has been found in various Mesoamerican and South American populations (Table 1), and may very well be unique to the Americas because it was not observed in Japanese, Chinese, and Koreans (Estrada-Mena et al., 2010). If O1v<sup>G542A</sup> is truly a unique American allele and is found in all Native American populations, it would support a single ancestral population for those that carry the allele, consistent with the BIM, as argued by Estrada-Mena et al. (2010).

However, two questions must be answered before the O1v<sup>G542A</sup> allele can be considered a Native American AIM or evidence for the BIM. First, how widespread is the O1v<sup>G542A</sup> allele in the Americas? Is it found throughout the Americas, or is it restricted to Mesoamerica and South America? Because indigenous populations north of central Mexico have not yet been screened for this allele, it is possible that the defining mutation is found only in Mesoamerica and South America. If that is the case, then the allele likely originated sometime after the entrance of humans into the Americas, and its frequency in Mesoamerica and South America would reflect the population history of those regions rather than the initial migration of people into the Americas.

Second, was the O1v<sup>G542A</sup> allele present in the Americas in prehistoric times, or was it introduced by admixture after European contact in AD 1492? No pre-contact populations have yet been screened for this allele. The O1v<sup>G542A</sup> mutation has been reported in at least one individual in an European population (Yip, 2000) and one individual in a Middle-Eastern population (Yip et al., 2006), so it is possible the allele may have been introduced from the Old World after AD 1492. Alternatively, the presence of O1v<sup>G542A</sup> in Europe and the Middle East might be due to mutational homoplasy at the ABO locus or even postcontact movement from the Americas to the Old World.

To address these two questions, we screened a large number of Tlingits from southeastern Alaska, a more limited sample of Haida and Inuit from Alaska, and one ancient pre-Columbian population from California for the O1v<sup>G542A</sup> allele. The analysis of these populations

T1

AQ4

TABLE 1. Sample size and counts of total chromosomes, O alleles, O1v alleles, and O1v<sup>G542A</sup> alleles; frequencies of O1v<sup>G542A</sup> alleles relative to O1v alleles, total O alleles, and total alleles sampled in all American populations sampled at the time of this study

Population	Sample size	Count of total chromosomes	Count of O1v <sup>G542A</sup> chromosomes	Count of O1v chromosomes	Count of O chromosomes	Freq. of O1v <sup>G542A</sup> of O1v chromosomes	Freq. of O1v <sup>G542A</sup> of all O chromosomes	Freq. of O1v <sup>G542A</sup> of total chromosomes	Reference
Arara	15	30	11	29	30	0.39	0.38	0.38	Olsson et al., 1998
Aymara	84	168	20	109	59	0.18	0.34	0.12	Llop et al., 2006
Aymaras	63	126	15	90	126	0.17	0.12	0.12	Estrada-Mena et al., 2012
Aymaras	63	126	12	62	126	0.19	0.10	0.10	Roubinet et al., 2002
Cayapa	35	70	3	31	70	0.10	0.04	0.04	Estrada-Mena et al., 2011
Cayapas	37	74	4	43	74	0.09	0.05	0.05	Roubinet et al., 2001
Huilliche	67	134	93	125	134	0.74	0.69	0.69	Llop et al., 2006
Kayapo	16	32	11	27	32	0.39	0.33	0.33	Olsson et al., 1999
Maya	50	100	28	80	100	0.35	0.28	0.28	Estrada-Mena et al., 2010
Mazahua	50	100	28	68	100	0.41	0.28	0.28	Estrada-Mena et al., 2009
Nahua	37	74	18	59	74	0.31	0.24	0.24	Estrada-Mena et al., 2008
Ohlone	15	30	2	10	27	0.20	0.07	0.07	This study
Parakana	62	124	27	81	124	0.33	0.22	0.22	Barjas-Castro et al., 2003
Tlingit	72	144	16	77	129	0.21	0.13	0.11	This study
Yanomama	17	34	12	31	34	0.39	0.36	0.36	Olsson et al., 2000

allowed us to determine if O1v<sup>G542A</sup> is found in more widespread Native American populations than previously described and if the allele was present in the Americas prior to European contact.

If the O1v<sup>G542A</sup> allele is found in these populations, it would suggest that the allele originated before their divergence, with all North and South American populations inheriting this allele from a common ancestral population. It is possible, however, that this allele is common to all of these populations due to gene flow between the separate “waves” of migrants that introduced genetic variation to the American continents, following the observations of Reich et al. (2012). In this case, a simple migration model can be used to estimate the level of gene flow between populations that is required to explain the modern frequencies of O1v<sup>G542A</sup>, and whether this level of migration is feasible based on our current understanding of Native American prehistory.

## METHODS

### Indigenous Alaskan samples

All Tlingit, Haida, and Inuit samples were collected by Brian M. Kemp and Kari B. Schroeder during Celebration 2008 in Juneau, Alaska. This study was approved by the Institutional Review Board (IRB # 10379) at Washington State University. After a thorough review of the purposes and use of the DNA study and samples, Sealaska Heritage Institute agreed to support the study in its Celebration.

The Tlingit language is probably related to Eyak and Athapaskan languages (Campbell, 1997; Vajda, 2010), which together represent a linguistic macrogroup that is nearly synonymous with “Na-Dene” as envisioned by Greenberg et al. (1986). However, the inclusion of Haida in Na-Dene is a controversial addition (Campbell, 1997). Mitochondrial and Y-chromosomal DNA variation exhibited by the Tlingit and Haida provide no compelling evidence for very recent common ancestry between these two populations (Schurr et al., 2012). In the study of Reich et al. (2012), the Chipewyan, a population belonging to the Athapaskan language family, was used as a representative of Na-Dene. For comparison, the Tlingit are considered Na-Dene in this study, but with the above caveat about the validity of this linguistic grouping in mind. The Haida are not considered to be a Na-Dene population here, but they are informative about how widespread the O1v<sup>G542A</sup> allele is in southeast Alaska. The Inuit belong to the Aleut-Eskimo linguistic group (Campbell, 1997).

All DNA was extracted from saliva using the Norgen Saliva DNA Isolation Kit (Cat# RU45400). Due to polymerase chain reaction (PCR) inhibitors coextracted with these samples, DNA extracts were diluted 1:10. Seventy-two Tlingit samples, four Haida, and three Inuit were typed for ABO alleles using a modified PCR-RFLP method following Hummel et al. (2002). These samples represent individuals for which Tlingit, Haida, and Inuit ancestry was self-reported (Supporting Information Table S1). Although participants were asked for their tribal affiliation for up to two generations, in some cases this information was incomplete. To minimize the inclusion of individuals with non-Native American ancestry, individual who explicitly reported non-Native American ancestry on either parental line were excluded from analysis.

AQ5



To amplify exons 6 and 7 of the ABO locus, we performed 15  $\mu$ L PCRs containing 0.32 mM dNTPs, 1 $\times$  PCR Buffer, 1.5 mM  $MgCl_2$ , 0.24  $\mu$ M primers, 0.3 U of platinum Taq (Invitrogen), and 3  $\mu$ L template DNA. We used the primers described by Hummel et al. (2002). Negative controls (PCR reactions to which no DNA template was added) accompanied every set of PCRs to monitor the presence of contaminating DNA. PCR conditions were as follows: 94°C for 3 min, 60 cycles of 15 s holds at 94°C, 56°C (for Exon 6) and 53°C (for Exon 3), and 72°C, followed by a final 3 min extension period at 72°C. Amplification success was determined by visualizing 5–6  $\mu$ L of the amplicons on a 6% polyacrylamide gel stained with ethidium bromide under UV light.

Enzyme digestion for *RsaI* and *HpyCH4IV* (Exon 6) and *NlaIII* and *MnII* (Exon 7) was performed at approximately 36°C for >3 h. Digested amplicons were visualized by separating 4–5  $\mu$ L on a 6% polyacrylamide gel stained with ethidium bromide under UV light. For individuals exhibiting the O1v allele, an additional PCR fragment was amplified and sequenced using forward primer O1v542F (GGCCACCGTGTCCACTACTA) and reverse primer O1v542R (GAACAGCGGAGTCAGATCT), under the above described PCR conditions with an annealing temperature of 58°C. Amplicons were prepared for sequencing following Kemp et al. (2007) and sequenced in both directions at Elim BioPharm (Hayward, CA).

### Ancient Ohlone samples

Ancient DNA (aDNA) was analyzed from burials excavated from the Yukisma site (CA SCL-38), an ancestral Muwekma Ohlone burial site located in northern Santa Clara County, California. The site is located approximately 6 miles southeast of the southern portion of San Francisco Bay (Leventhal et al., 1993; Morley, 1997; Jurmain, 2001; Gardner et al., 2011). This site falls under the Tamyen-speaking territory, part of the Ohlone languages in the Utian family (Bellifemine, 1997; Jurmain, 2001). One or more skeletal elements from each of the 252 burials were reserved for genetic analysis prior to remaining skeletal elements being reinterred. With approval from the Muwekma Ohlone, the samples were transferred to the Kemp Lab of Molecular Anthropology and aDNA at Washington State University. Radiocarbon dating places deposition from 250 YBP to at least 2,200 YBP, but the burials are predominately from the beginning to middle of the Late Period (740–230 YBP) (Leventhal et al., 1993; Bellifemine, 1997).

All extractions and PCR preparations were conducted in a dedicated aDNA facility, with amplification and post-PCR processing conducted in a laboratory located in a separate building. Approximately 0.5 g of material was carefully removed from each of 68 ribs. Each sample was submerged in 6% sodium hypochlorite (full strength household bleach) for 15 min (Kemp and Smith, 2005). Each sample was rinsed 1–2 times with DNA free water (Gibco) to remove the bleach. DNA was extracted following the method described by Kemp et al. (2007) with the following modifications. Following isopropanol precipitation, the pelleted DNA was resuspended in 300  $\mu$ L of 60–65°C DNA-free ddH<sub>2</sub>O and incubated at 60–65°C for 20 min and vortexed three times before silica extraction using the Wizard PCR Preps DNA Purification System (Promega).

Each sample was first screened for the markers diagnostic of Native American mtDNA haplogroups A, B, C, and D following Kemp et al. (2007). To confirm that the screened individuals conform to haplogroup frequencies expected in Native Americans from the California and Great Basin areas, a Fisher's exact test for haplogroup frequencies was performed between the ancient Ohlone population and the Cahuilla, Costanoan, Modern Great Basin (multiple groups), Sahaptain (200YBP, protohistoric/historic) (Malhi et al., 2004), Serano/Vanyume, Sierra Miwok, Tubat, and Yokuts modern populations (Johnson and Lorenz, 2006). Also compared were two prehistoric populations: CA-SCL-287/SMA-263 (1176–1954 YBP), which is considered to be an ancestral Muwekma Ohlone Site and is located on the border of Santa Clara and San Mateo Counties, CA, within the vicinity of CA-SCL-38 (Monroe et al., 2009); and CA-SOL-270 (ca. 2000 YBP), which is located in Solano, County, CA east of the San Francisco Bay and also thought to be part of the historic Ohlone tribal area (Eshleman, 2002).

Samples that yielded Native American mtDNA types were screened for ABO variation as described above for modern samples. This was replicated three times by F.A.V. (Table 2), and once more by B.M.K., following recommendations by Halverson and Bolnick (2008). Results were screened against the mitochondrial and blood types of the researchers to eliminate possible contamination originating from them. None of the researchers belong to any of the mitochondrial haplogroups found in the samples. While one researcher (F.A.V.) has the genotype O1vO1v<sup>G542A</sup>, this genotype was not found in any of the ancient samples. To settle differences in typing between replications, this study assumes that allelic drop-out is the most likely explanation. Note that each allele is determined by a combination of four restriction enzymes, and so allelic drop-out in one PCR translates to two genotypes (Table 2, Hummel et al., 2002). For example, in sample B171, allelic drop-out explains the shift from the genotype AO1v (+/–, +/–) into O1O1v (+/+, +/–) as these two genotypes are differentiated by heterozygosity at only one locus.

Ancient samples that tested positive for both mtDNA and the ABO locus were sent to the aDNA laboratory at the University of Texas at Austin for independent confirmation (Table 2). Samples were extracted following the protocol described by Bolnick and Smith (2007). The mtDNA haplogroup of each sample was determined, and ABO genotyping was performed 1–5 times per sample.

### Gene flow

To estimate the degree of gene flow between Na-Dene and First American groups sufficient to explain the presence of O1v<sup>G542A</sup> in both groups, we performed a simple estimation of gene flow following Schroeder et al. (2007). Specifically, if the Tlingit descend from a separate founding population in which O1v<sup>G542A</sup> was not present, the amount of gene flow ( $m$ ) required to bring O1v<sup>G542A</sup> to the frequencies observed today is calculated as  $m = (p_h - p_2) / (p_1 - p_2)$ , where  $p_h$  is the observed average frequency of O1v<sup>G542A</sup> in the Na-Dene,  $p_1$  is the observed average frequency (with a minimum and maximum range) of O1v<sup>G542A</sup> in the other (non-Na-Dene) Native American groups; and  $p_2$  is the frequency of O1v<sup>G542A</sup> in the Na-Dene prior to gene flow, which was set to zero.

AQ1

UNIQUE O ALLELE (O1v<sup>G542A</sup>) IN NATIVE AMERICANS

5

TABLE 2. Mitochondrial haplogroups, and ABO genotypes for 15 ancient Ohlone samples.

Sample ID Author	mt Haplogroup		ABO Blood type										O1v <sup>G542A</sup> FAV
	FAV	DAB	FAV 1	FAV 2	FAV 3	BMK	DAB 1	DAB 2	DAB 3	DAB 4	DAB 5	Consensus	
B004	C	C	AO1	O1O1	AO1	O1O1	O1O1	AO1	O1O1	NA	NA	AO1	-
B006	B	B	O1O1v	NA	NA	NA	O1O1	O1O1v	NA	NA	NA	O1O1v	NA
B010	D	D	O1O1v	O1O1v	O1O1v	O1O1v	O1vO1v	NA	O1O1	O1O1	O1O1v	O1O1v	Yes/No
B023	B	B	O1O1	O1O1v	O1O1	NA	O1O1	NA	NA	NA	NA	O1O1v	NA
B038	D	D	O1O1v	O1vO1v	O1O1v	O1O1v	O1O1v	O1O1v	NA	NA	NA	O1O1v	NA
B043	D	D	O1O1	O1O1	O1O1	NA	O1vO1v	NA	O1vO1v	O1O1v	O1O1v	O1O1v	NA
B048	D	NA	O1O1	O1O1	O1O1	O1O1	NA	NA	NA	NA	NA	O1O1	No/No
B050	D	D	O1O1v	NA	NA	NA	O1O1v	NA	NA	NA	NA	O1O1v	NA
B056	D	NA	O1vO1v	O1vO1v	NA	NA	NA	NA	NA	NA	NA	O1vO1v	NA
B080	D	D	O1O1	AA	NA	NA	O1O1	O1O1	NA	NA	NA	AO1	-
B134	D	D	O1O1	O1O1	O1O1v	O1O1v	O1O1v	O1vO1v	O1O1	O1O1v	NA	O1O1v	Yes/No
B170	D	D	O1O1	O1O1	NA	NA	O1O1	NA	NA	NA	NA	O1O1	-
B171	D	D	AO1v	O1O1v	AO1v	O1O1v	O1vO1v	O1O1	O1O1	NA	NA	AO1v	NA
B193	D	D	O1O1	O1O1	NA	NA	O1vO1v	O1O1	O1vO1v	NA	NA	O1O1v	NA
B233	C	C	O1O1	O1O1	O1O1	O1O1	O1O1	NA	NA	NA	NA	O1O1	-

"NA" denotes reactions which failed to amplify, while "-" denote analysis not conducted

## RESULTS

## Modern Alaskan samples

Of 144 chromosomes representing 72 Tlingit samples typed for ABO using PCR-RFLPs, 129 exhibited O alleles and 77 of these were typed as O1v alleles (Table 1). Sequencing revealed 16 of the 144 alleles to be O1v<sup>G542A</sup> (11%). The frequency of the O1v<sup>G542A</sup> allele among the Tlingit falls within the range of frequencies reported for other Native American populations, ranging from 4 to 69% (Estrada-Mena et al., 2010). The relative frequency of the O1v<sup>G542A</sup> allele of all O1v alleles among the Tlingit is 21%, which also falls in the range of Native American frequencies (9–74%, Table 1, Fig. 1). One of the four Haida exhibited one O1v<sup>G542A</sup> allele and none of the three Inuit exhibited this allele.

F1

## Ancient Ohlone samples

Forty-one of the sixty-eight (60%) ancient samples were typed as belonging to one of the four major Native American mtDNA haplogroups A-D. A Fisher's exact test shows that the ancient Ohlone population is not significantly distinguishable at the 0.05 level of probability from Cahuilla, Costanoan, Modern Great Basin (multiple groups), Sahaptain (200YBP, protohistoric/historic) (Malhi et al., 2004), Serano/Vanyume, Sierra Miwok, Tubat, and Yokuts populations (Johnson and Lorenz, 2006) as well as the CA-SCL-287/SMA-263 (1176–1954 YBP) (Monroe et al., 2009) and CA-SOL-270 (ca. 2,000 YBP) (Eshleman, 2002) prehistoric populations. These results suggest that the mtDNA results from these samples are authentic, as the data make phylogenetic sense given where the population was sampled.

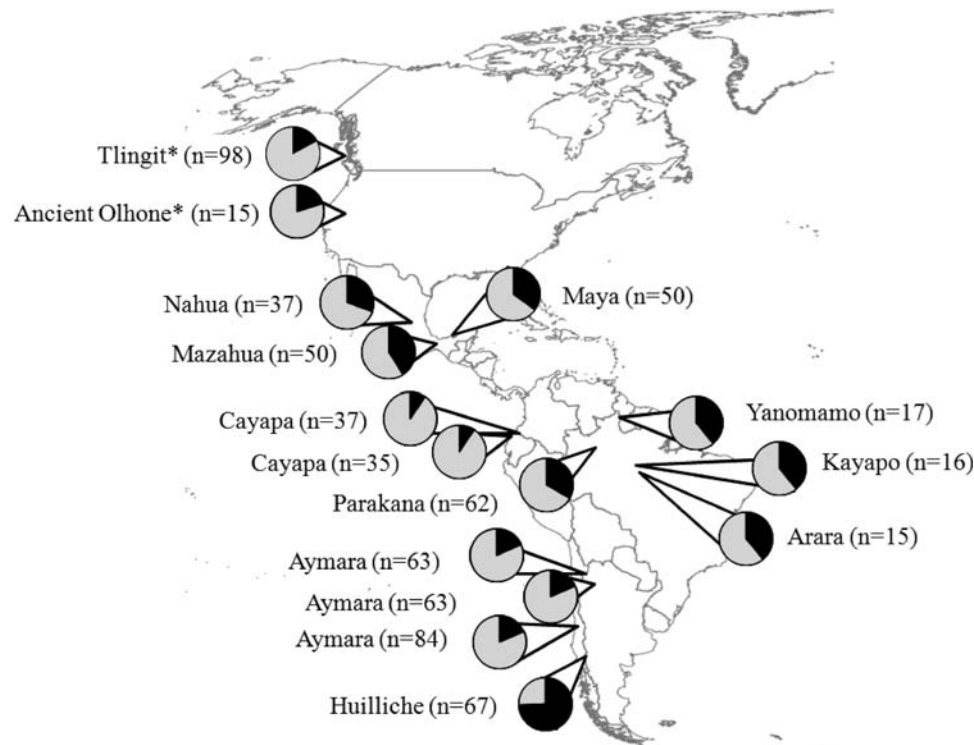
Out of these 41 samples, 15 produced data for the ABO locus (Table 2). Ten of the 30 chromosomes exhibited O1v alleles, of which two were typed as O1v<sup>G542A</sup>. Six other samples carrying the O1v allele failed to yield readable sequence information. The frequency of the O1v<sup>G542A</sup> mutation in O1v chromosomes is, therefore, 0.20 or higher. In addition, we identified three individuals carrying the A allele, which is rare in North American Native Americans. The importance of precontact presence of A alleles has been discussed elsewhere (see Halverson and Bolnick, 2008).

## Gene flow

If the presence of O1v<sup>G542A</sup> in the Na-Dene (i.e., Tlingit) is the result of admixture with other Native American groups, the frequency of O1v<sup>G542A</sup> observed today in Na-Dene would require gene exchange of 48% (17% minimum and 100% maximum, using the lowest and highest frequencies, respectively, of O1v<sup>G542A</sup> observed among a non-Na-Dene population) with First Americans.

## DISCUSSION

Finding the O1v<sup>G542A</sup> variant in the Tlingit and Haida, as well as in a pre-Colombian North American population, allows several conclusions to be drawn. First, the distribution of O1v<sup>G542A</sup> has now been widened by over 3,000 kilometers north, across all Native American populations sampled at the time of this study, including a Na-Dene population. This finding indicates close ancestry among all groups sampled for the ABO locus. The distribution of O1v<sup>G542A</sup> is now more consistent with the previously reported distribution of the 9RA



**Fig. 1.** Distribution and relative frequency of the O1v<sup>G542A</sup> variant (black) out of all O1v alleles (gray). Populations marked with a star (\*) were screened in this study.

at the D9S1120 microsatellite, which also includes Aleut-Eskimos. It is possible that O1v<sup>G542A</sup> will be detected in these populations as well, although this will require further work. Here, the allele was not detected among the Inuit, but only three samples were included in this study.

Second, finding O1v<sup>G542A</sup> in a pre-Columbian Ohlone population demonstrates that the allele was present in the Native American population prior to European contact. While this observation may be unsurprising given the distribution of the O1v<sup>G542A</sup> allele among contemporary Native Americans, it is an important direct observation. Taken together with the observation of O1v<sup>G542A</sup> in all Native American populations tested for this allele, and the rarity of the allele in non-Native American populations, the most likely explanation is that O1v<sup>G542A</sup> originated in the population ancestral to all Native Americans and became widespread as this population migrated and spread into the continent, consistent with the BIM proposed by Tamm et al. (2007). In this case, the few alleles previously reported outside of America are most likely either the consequence of gene flow into Europe and the Middle East from the Americas or mutational homoplasy.

An alternative hypothesis accounts for the presence of alleles such as O1v<sup>G542A</sup> and 9RA as the result of secondary gene flow between Native American groups descended from different source populations, more closely following the three wave migration model originally posited by Greenberg et al. (1986). While it is possible that O1v<sup>G542A</sup> originated in the Amerinds and was absent in a secondary wave that contributed to Na-Dene ancestry, this scenario requires a high level of gene flow (48%) to explain the frequency of this allele in the

Tlingit. This result is consistent with the estimated level of gene flow necessary to explain the shared presence of 9RA (91.9–92.8%, Schroeder et al., 2007). Based on a nuclear genome-wide survey of variation, Reich et al. (2012) argued for at least three streams of population movement into the Americas, followed by extensive gene flow between them. For example, in their study, they found that the Chipewyan, a Na-Dene population, derives 90% of its ancestry from the First Americans population, with the remaining 10% of their ancestry from a genetically distinct source population. Yet, the most parsimonious explanation for the presence of O1v<sup>G542A</sup> and 9RA in such high frequencies among both First American and Na-Dene populations is shared ancestry. According to the study of Reich et al. (2012), the Na-Dene are indeed unique among Native American populations, but the fact remains that they are still 90% First Americans. While these genetic data cannot inform us if the remaining 10% of their ancestry was associated with the introduction of Na-Dene languages into the Americas, if that were true, it would provide very weak support for the model originally proposed by Greenberg et al. (1986), which argued for biologically distinct origins of Amerind, Na-Dene, and Eskimo populations. In a strict sense, Greenberg et al.'s model assumed discrete biological groups, which are not consistent with the high degree of gene flow inferred from Reich et al.'s (2012) data.

Finally, taken together, this new evidence from a locus unlinked with D9S1120, the Y-chromosome M3 marker, or the mitochondrial genome further supports the shared ancestry of all modern Native Americans, which are descended from a population that evolved in isolation for enough time to allow for the rise of unique variants. In



this case, it will be particularly informative to screen additional Na-Dene populations, as well as Aleut-Eskimos and northeastern Siberian groups, to take advantage of expanding studies of Native American ancestry into relatively unexplored nuclear markers.

### ACKNOWLEDGMENTS

The authors are especially grateful to the individuals who participated in this study at Celebration 2008—Gunalchéesh. They also thank the Muwekma Ohlone Tribe for allowing them to study the remains of their ancestors. They thank Kari Schroeder for her participation in collection of the Tlingit, Haida, and Inuit samples analyzed in this study. They also thank Drew Kitchen for useful discussions about the implications of recent full genomic analysis of Native Americans.

### LITERATURE CITED

- Achilli A, Perego UA, Bravi CM, Coble MD, Kong QP, Woodward SR, Salas AT, Torroni A, Bandelt HJ. 2008. The phylogeny of the four pan-American MtDNA haplogroups: implications for evolutionary and disease studies. *PLoS One* 3:e1764.
- Acuna-Alonso V, Flores-Dorantes T, Kruit JK, Villarreal-Molina T, Arellano-Campos O, Hunemeier T, Moreno-Estrada A, Ortiz-Lopez MG, Villamil-Ramirez H, Leon-Mimila P, Villalobos-Comparan M, Jacobo-Albavera L, Ramirez-Jimenez S, Sikora M, Zhang LH, Pape TD, Granados-Silvestre Mde A, Montufar-Robles I, Tito-Alvarez AM, Zurita-Salinas C, Bustos-Arriaga J, Cedillo-Barron L, Gomez-Trejo C, Barquera-Lozano R, Vieira-Filho JP, Granados J, Romero-Hidalgo S, Huertas-Vazquez A, Gonzalez-Martin A, Gorostiza A, Bonatto SL, Rodriguez-Cruz M, Wang L, Tusie-Luna T, Aguilar-Salinas CA, Lisker R, Moises RS, Menjivar M, Salzano FM, Knowler WC, Bortolini MC, Hayden MR, Baier LJ, Canizales-Quinteros S. 2010. A functional ABCA1 gene variant is associated with low HDL-cholesterol levels and shows evidence of positive selection in Native Americans. *Hum Mol Genet* 19:2877–2885.
- 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.
- Bolnick DA, Smith DG. 2007. Migration and social structure among the Hopewell: evidence from ancient DNA. *Am Antiq* 72:627–644.
- Bortolini MC, Salzano FM, Thomas MG, Stuart S, Nasanen SP, Bau CH, Hutz MH, Layrisse Z, Petzl-Erler ML, Tsuneto LT, Hill K, Hurtado AM, Castro-de-Guerra D, Torres MM, Groot H, Michalski R, Nymadawa P, Bedoya G, Bradman N, Labuda D, Ruiz-Linares A. 2003. Y-chromosome evidence for differing ancient demographic histories in the Americas. *Am J Hum Genet* 73:524–539.
- Campbell L. 1997. American Indian languages: the historical linguistics of Native America. New York: Oxford University Press.
- Chester MA, Olsson ML. 2001. The ABO blood group gene: a locus of considerable genetic diversity. *Transfus Med Rev* 15:177–200.
- Crawford MH, Rubicz RC, Zlojutro M. 2010. Origins of Aleuts and the genetic structure of populations of the Archipelago: molecular and archaeological perspectives. *Hum Biol* 82: 695–717.
- Dulik MC, Owings AC, Gaieski JB, Vilar MG, Andre A, Lennie C, Mackenzie MA, Kritsch I, Snowshoe S, Wright R, Martin J, Gibson N, Andrews TD, Schurr TG. 2012a. Y-chromosome analysis reveals genetic divergence and new founding native lineages in Athapaskan- and Eskimoan-speaking populations. *Proc Natl Acad Sci USA* 109:8471–8476.
- Dulik MC, Zhadanov SI, Osipova LP, Askapuli A, Gau L, Gokcumen O, Rubinstein S, Schurr TG. 2012b. Mitochondrial DNA and Y-chromosome variation provides evidence for a recent common ancestry between Native Americans and indigenous Altaians. *Am J Hum Genet* 90:229–246.
- Eshleman JA. 2002. Mitochondrial DNA and prehistoric population movements in Western North America. Dissertation, University of California-Davis.
- Estrada-Mena B, Estrada FJ, Ulloa-Arvizu R, Guido M, Méndez R, Coral R, Canto T, Granados J, Rubí-Castellanos R, Rangel-Villalobos H, García-Carrancá A. 2010. Blood group O alleles in Native Americans: implications in the peopling of the Americas. *Am J Phys Anthropol* 142:85–94.
- Fagundes NJR, Kanitz R, Bonatto SL. 2008a. A reevaluation of the Native American MtDNA genome diversity and its bearing on the models of early colonization of Beringia. *PLoS One* 3:e3157.
- Fagundes NJR, Kanitz R, Eckert R, Valls ACS, Bogo MR, Salzano FM, Smith DG, Silva WA, Zago MA, Ribeiro-dos-Santos AK, Santos SEB, Petzl-Erler ML, Bonatto SL. 2008b. Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *Am J Hum Genet* 82:583–592.
- Gagneux P, Varki A. 1999. Evolutionary considerations in relating oligosaccharide diversity to biological function. *Glycobiology* 9:747–755.
- Gardner KS, Leventhal A, Cambra R, Tribe MO, Bartelink EJ, Martinez A. 2011. Mothers and infants in the prehistoric Santa Clara Valley: what stable isotopes tell us about ancestral Ohlone weaning practices. *SCA (Society of California Archaeology) Proc* 25:1–14.
- Georges L, Seidenberg V, Hummel S, Fehren-Schmitz L. 2012. Molecular characterization of ABO blood group frequencies in pre-Columbian Peruvian highlanders. *Am J Phys Anthropol* 149:242–249.
- Greenberg JH, Turner CG, Zegura SL, Campbell L, Fox JA, Laughlin WS, Szathmary EJE, Weiss KM, Woolford E. 1986. The settlement of the Americas: a comparison of the linguistic, dental, and genetic evidence [and Comments and Reply]. *Curr Anthropol* 27:477–497.
- Halverson MS, Bolnick DA. 2008. An ancient DNA test of a founder effect in Native American ABO blood group frequencies. *Am J Phys Anthropol* 137:342–347.
- Hummel S, Schmidt D, Kahle M, Herrmann B. 2002. ABO blood group genotyping of ancient DNA by PCR-RFLP. *Int J Legal Med* 116:327–333.
- Hunemeier T, Amorim CE, Azevedo S, Contini V, Acuna-Alonso V, Rothhammer F, Dugoujon JM, Mazieres S, Barrantes R, Villarreal-Molina MT, Paixao-Cortes VR, Salzano FM, Canizales-Quinteros S, Ruiz-Linares A, Bortolini MC. 2012. Evolutionary responses to a constructed niche: ancient Mesoamericans as a model of gene-culture coevolution. *PLoS One* 7:e38862.
- Johnson JR, Lorenz JG. 2006. Genetics, linguistics, and prehistoric migrations: an analysis of California Indian mitochondrial DNA lineages. *J Calif Great Basin Anthropol* 26:31–62.
- Jurmain R. 2001. Paleoepidemiological patterns of trauma in a prehistoric population from central California. *Am J Phys Anthropol* 115:13–23.
- Karafet T, Xu L, Du R, Wang W, Feng S, Wells RS, Redd AJ, Zegura SL, Hammer MF. 2001. Paternal population history of East Asia: sources, patterns, and microevolutionary processes. *Am J Hum Genet* 69:615–628.
- Karafet T, Zegura SL, Vuturo-Brady J, Posukh O, Osipova L, Wiebe V, Romero F, Long JC, Harihara S, Jin F, Dashnyam B, Gerelsaikhan T, Omoto K, Hammer MF. 1997. Y chromosome markers and trans-Bering Strait dispersals. *Am J Phys Anthropol* 102:301–314.
- Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, Harihara S, de Knijff P, Wiebe V, Griffiths RC, Templeton AR, Hammer MF. 1999. Ancestral Asian source(s) of New World Y-chromosome founder haplotypes. *Am J Hum Genet* 64:817–831.
- Kemp BM, Malhi RS, McDonough J, Bolnick DA, Eshleman JA, Rickards O, Martinez-Labarga C, Johnson JR, Lorenz JG, Dixon EJ, Fifield TE, Heaton TH, Worl R, Smith DG. 2007.



- Genetic analysis of early Holocene skeletal remains from Alaska and its implications for the settlement of the Americas. *Am J Phys Anthropol* 132:605–621.
- Kemp BM, Schurr TG. 2010. Ancient and modern genetic variation in the Americas. In: Auerbach B, editor. *Human variation in the Americas: the integration of archaeology and biological anthropology*. Carbondale: Southern Illinois University. p 12–50.
- Kemp BM, Smith DG. 2005. Use of bleach to eliminate contaminating DNA from the surface of bones and teeth. *Forensic Sci Int* 154:53–61.
- Kitchen A, Miyamoto MM, Mulligan CJ. 2008. A three-stage colonization model for the peopling of the Americas. *PLoS One* 3:e1596.
- Lell JT, Brown MD, Schurr TG, Sukernik RI, Starikovskaya YB, Torroni A, Moore LG, Troup GM, Wallace DC. 1997. Y chromosome polymorphisms in Native American and Siberian populations: identification of Native American Y chromosome haplotypes. *Hum Genet* 100:536–543.
- Lell JT, Sukernik RI, Starikovskaya YB, Su B, Jin L, Schurr TG, Underhill PA, Wallace DC. 2002. The dual origin and Siberian affinities of Native American Y chromosomes. *Am J Hum Genet* 70:192–206.
- Leventhal AM, Jones L, Cambra R, Sanchez N. 1993. Results from a presence/absence subsurface archaeological test excavation program on a portion of prehistoric site: CA-SCL-38 (Yukisma Mound) for the proposed construction of housing unit barracks M8 located within the Elmwood Correctional Facility city of Milpitas, Santa Clara County, California: Ohlone Family Consulting Inc.
- Llop E, Henriquez H, Moraga M, Castro M, Rothhammer F. 2006. Brief communication: molecular characterization of O alleles at the ABO locus in Chilean Aymara and Huilliche Indians. *Am J Phys Anthropol* 131:535–538.
- Malhi RS, Breece KE, Shook BAS, Kaestle FA, Chatters JC, Hackenberger S, Smith DG. 2004. Patterns of mtDNA diversity in northwestern North America. *Hum Biol* 76:33–54.
- Molnar S. 2002. *Human variation*. Upper Saddle River, NJ: Prentice Hall.
- Monroe C, Alan L, Rosemary C, Kemp BM. 2009. Ancient human DNA analysis from CA-SCL-287/CA-SMA-263 burials: new insights regarding California prehistory. In: 43rd Annual Meeting of the Society of California Archaeology. Modesto, CA.
- 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, Kazimiera D-S. 1976. *The distribution of the human blood groups and other polymorphisms*. London: Oxford University Press.
- Mulligan CJ, Kitchen A, Miyamoto MM. 2008. Updated three-stage model for the peopling of the Americas. *PLoS One* 3:e3199.
- Nichols J. 1990. Linguistic diversity and the first settlement of the New World. *Language* 66:475–521.
- Perego UA, Achilli A, Angerhofer N, Accetturo M, Pala M, Olivieri A, Kashani BH, Ritchie KH, Scozzari R, Kong Q-P, Myres NM, Salas A, Semino O, Bandelt H-J, Woodward SR, Torroni A. 2009. Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. *Curr Biol* 19:1–8.
- Perego UA, Angerhofer N, Pala M, Olivieri A, Lancioni H, Kashani BH, Carossa V, Ekins JE, Gomez-Carballea A, Huber G, Zimmermann B, Corach D, Babudri N, Panara F, Myres NM, Parson W, Semino O, Salas A, Woodward SR, Achilli A, Torroni A. 2010. The initial peopling of the Americas: a growing number of founding mitochondrial genomes from Beringia. *Genome Res* 20:1174–1179.
- Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, Parra MV, Rojas W, Duque C, Mesa N, Garcia LF, Triana O, Blair S, Maestre A, Dib JC, Bravi CM, Bailliet G, Corach D, Hunemeier T, Bortolini MC, Salzano FM, Petzl-Erler ML, Acuna-Alonzo V, Aguilar-Salinas C, Canizales-Quinteros S, Tusie-Luna T, Riba L, Rodriguez-Cruz M, Lopez-Alarcon M, Coral-Vazquez R, Canto-Cetina T, Silva-Zolezzi I, Fernandez-Lopez JC, Contreras AV, Jimenez-Sanchez G, Gomez-Vazquez MJ, Molina J, Carracedo A, Salas A, Gallo C, Poletti G, Witonsky DB, Alkorta-Aranburu G, Sukernik RI, Osipova L, Fedorova SA, Vasquez R, Villena M, Moreau C, Barrantes R, Pauls D, Excoffier L, Bedoya G, Rothhammer F, Dugoujon J-M, Larrouy G, Klitz W, Labuda D, Kidd J, Kidd K, Di Rienzo A, Freimer NB, Price AL, Ruiz-Linares A. 2012. Reconstructing Native American population history. *Nature* 488:370–374.
- 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, 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, Smith DG. 2007. A private allele ubiquitous in the Americas. *Biol Lett* 3.
- Schurr TG, Dulik MC, Owings AC, Zhadanov SI, Gaieski JB, Vilar MG, Ramos J, Moss MB, Natkong F. 2012. Clan, language, and migration history has shaped genetic diversity in Haida and Tlingit populations from southeast Alaska. *Am J Phys Anthropol* 148:422–435.
- Smith DG, Lorenz JG, Rolfs BK, Bettinger RL, Green B, Eshleman JA, Schultz B, Malhi RS. 2000. Implications of the distributions of Albinism Naskapi and Albinism Mexico for New World prehistory. *Am J Phys Anthropol* 111:557–572.
- Swerdlow DL, Mintz ED, Rodriguez M, Tejada E, Ocampo C, Espejo L, Barrett TJ, Petzelt J, Bean NH, Seminario L, Tauxe RV. 1994. Severe life-threatening cholera associated with blood group O in Peru: implications for the Latin American epidemic. *J Infect Dis* 170:468–472.
- Szathmari EJE. 1979. Blood groups of Siberians, Eskimos, and subarctic and Northwest Coast Indians: the problem of origins and genetic relationships. In: Laughlin WS, Harper AB, editors. *The First Americans: origins, affinities, and adaptations*. New York: G. Fischer. p 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, Golubenkov MV, Stepanov VA, Gubina MA, Zhadanov SI, Ossipova LP, Damba L, Voevoda MI, Dipierri JE, Vilems R, Malhi RS. 2007. Beringian standstill and spread of Native American founders. *PLoS One* 9:e829 (1–6).
- Underhill PA, Jin L, Zemans R, Oefner PJ, Cavalli-Sforza LL. 1996. A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. *Proc Natl Acad Sci USA* 93:196–200.
- Vajda E. 2010. A Siberian link with Na-Dene languages. *Archaeol Pap Univ Alaska* 5:33–99.
- Yip SP. 2000. Single-tube multiplex PCR-SSCP analysis distinguishes 7 common ABO alleles and readily identifies new alleles. *Blood* 95:1487–1492.
- Yip SP. 2002. Sequence variation at the human ABO locus. *Ann Hum Genet* 66:1–27.
- Yip SP, Choi PS, Lee SY, Leung KH, El-Zawahri Mokhtar M, Luqmani Yunus A. 2006. ABO blood group in Kuwaitis: detailed allele frequency distribution and identification of novel alleles. *Transfusion* 46:773–779.
- Zegura SL, Karafet TM, Zhivotovsky LA, Hammer MF. 2004. High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas. *Mol Biol Evol* 21:164–175.

AQ1: Please confirm whether the given short title is OK.

AQ2: Please confirm that all author names are OK and are set with first name first, surname last.

AQ3: Please confirm whether the authors and their corresponding affiliations are OK as given.

AQ4: Please note that the quotes were deleted after the first instance per style. Please confirm whether this is OK.

AQ5: Please spell out RFLP during the first instance.

AQ6: Please note that the references “Olsson et al., 1998, 1999, 2000; Estrada-Mena et al., 2008, 2009, 2011, 2012; Roubinet et al., 2001, 2002; Barjas-Castro et al., 2003” are provided in the table but not detailed in the list. Please provide the details in the list.

