

2-2004

## Patterns of mtDNA Diversity in Northwestern North America

Ripan S. Malhi  
*University of California, Davis*

Katherine E. Breece  
*University of California, Davis*

Beth A. Schultz Shook  
*University of California, Davis*

Frederika A. Kaestle  
*Indiana University - Bloomington*

James C. Chatters  
*Central Washington University*

*See next page for additional authors*

Follow this and additional works at: <https://digitalcommons.cwu.edu/cotsfac>



Part of the [Archaeological Anthropology Commons](#), and the [Genetics Commons](#)

---

### Recommended Citation

Malhi, R. S., Breece, K. E., Shook, B. A. S., Kaestle, F. A., Chatters, J. C., Hackenberger, S., & Smith, D. G. (2004). Patterns of mtDNA Diversity in Northwestern North America. *Human Biology*, 76(1), 33–54. <https://doi.org/10.1353/hub.2004.0023>

This Article is brought to you for free and open access by the College of the Sciences at ScholarWorks@CWU. It has been accepted for inclusion in All Faculty Scholarship for the College of the Sciences by an authorized administrator of ScholarWorks@CWU. For more information, please contact [scholarworks@cwu.edu](mailto:scholarworks@cwu.edu).

---

**Authors**

Ripan S. Malhi, Katherine E. Breece, Beth A. Schultz Shook, Frederika A. Kaestle, James C. Chatters, Steven Hackenberger, and David Glenn Smith

---

## *Patterns of mtDNA Diversity in Northwestern North America*

RIPAN S. MALHI,<sup>1,2</sup> KATHERINE E. BREECE,<sup>1</sup> BETH A. SCHULTZ SHOOK,<sup>1</sup> FREDERIKA A. KAESTLE,<sup>3</sup> JAMES C. CHATTERS,<sup>4</sup> STEVEN HACKENBERGER,<sup>4</sup> AND DAVID GLENN SMITH<sup>1,5</sup>

*Abstract* The mitochondrial DNA (mtDNA) haplogroups of 54 full-blooded modern and 64 ancient Native Americans from northwestern North America were determined. The control regions of 10 modern and 30 ancient individuals were sequenced and compared. Within the Northwest, the frequency distribution for haplogroup A is geographically structured, with haplogroup A decreasing with distance from the Pacific Coast. The haplogroup A distribution suggests that a prehistoric population intrusion from the subarctic and coastal region occurred on the Columbia Plateau in prehistoric times. Overall, the mtDNA pattern in the Northwest suggests significant amounts of gene flow among Northwest Coast, Columbia Plateau, and Great Basin populations.

As the number and diversity of samples analyzed in North America increase, studies of genetic diversity have moved from a continental to a regional level of analysis (Parr et al. 1996; Carlyle et al. 2000; Malhi et al. 2001, 2003; Kaestle and Smith 2001). Studies of classic genetic markers routinely attribute patterns of genetic variation in the Americas to genetic drift acting on small tribal populations and not to prehistoric population events or common ancestry (Cavalli-Sforza et al. 1994; Long 1991). Recent studies, however, have indicated that patterns of mtDNA variation in North America are geographically structured (Lorenz and Smith 1996; O'Rourke et al. 2000), and patterns of genetic variation in Native American populations are increasingly used to infer population history and genetic relatedness among groups.

Across the North American continent, haplogroup A occurs with high frequencies in the Arctic and Subarctic, haplogroup B reaches high frequencies in the Southwest, haplogroups C and X have high frequencies in the East, and

<sup>1</sup>Department of Anthropology, University of California, Davis, CA 95616.

<sup>2</sup>Trace Genetics LLC, P.O. Box 2010, Davis, CA 95617.

<sup>3</sup>Department of Anthropology, Indiana University, Bloomington, IN 47405.

<sup>4</sup>Tetratech Foster Wheeler Inc., Bothell, WA 98011; and Department of Anthropology, Central Washington University, Ellensburg, WA 98926.

<sup>5</sup>California Regional Primate Research Center (CRPRC), Davis, CA 95616.

*Human Biology*, February 2004, v. 76, no. 1, pp. 33–54.

Copyright © 2004 Wayne State University Press, Detroit, Michigan 48201-1309

KEY WORDS: ANCIENT DNA, AMERINDIANS, NATIVE AMERICANS, PENUTIAN, SALISHAN, MITOCHONDRIAL DNA, GENETIC DIVERSITY

haplogroup D is most common in central California, the Great Basin, and the Columbia Plateau (Lorenz and Smith 1996; O'Rourke et al. 2000; Malhi et al. 2002). Although populations within the same region of North America tend to exhibit similarities in haplogroup frequencies, some populations within a region still exhibit significant differences in haplogroup frequency distributions. The differences among populations within regions could reflect disequilibrium caused by recent population intrusions, structured gene flow, genetic drift, sampling error, or a combination of these factors. For example, in the Southwest most populations exhibit high frequencies of haplogroups B and C, yet these populations cluster into three groups. Pueblo groups exhibit the highest frequencies of haplogroup B in North America, Yuman and Piman groups cluster because of the presence of relatively high frequencies of both haplogroup B and haplogroup C, and Athapaskan groups cluster together as a result of the presence of relatively high frequencies of haplogroup A (Malhi et al. 2003). The differences in haplogroup frequency distributions between Southern Athapaskans and other Southwest groups can be largely attributed to the intrusion of Athapaskan groups into the Southwest in recent prehistory. In contrast, Algonquian populations in northeastern North America all share high frequencies of haplogroups A and C and lower frequencies of haplogroup X. Some Algonquian groups, however, differ significantly in their haplogroup frequency distributions. Because these differences do not correlate with geographic distance, genetic drift or sampling error is probably a major component of these differences. A specific example of sampling error is that of the Micmac, among whom only seven individuals were sampled (Malhi et al. 2001) and four of them belonged to haplogroup X.

In this study we examine the pattern of mtDNA diversity in northwestern North America using both modern and ancient samples. We explain the patterns of haplogroup frequency and haplotype distributions and use these distributions in conjunction with other anthropological data to gain insight into the population prehistory of Native American groups in the Northwest.

## **Previous Research in the Northwest**

Many studies have investigated the pattern of mtDNA diversity in populations from northwestern North America. Ward et al. (1993) were unable to detect significant genetic differentiation among the Bella Coola, Nuu-Chah-Nulth, and Haida populations, despite their linguistic differences. They attributed this lack of genetic differentiation to a recent common ancestry rather than to extensive gene flow among unrelated groups. Ward et al. (1993) also hypothesized that the Haida are substantially younger than the Bella Coola and Nuu-Chah-Nulth, consistent with the popular but controversial hypothesis that the Na-Dene language family arrived in the Americas more recently than the Amerind language family (Greenberg et al. 1986). Shields et al. (1993) argued that circumarctic populations exhibit lower amounts of genetic diversity than Amerind-speaking

groups, such as the Bella Coola, Nuu-Chah-Nulth, and Yakama, and they attributed this difference to a recent common ancestry among circumarctic populations followed by their expansion down both sides of Beringia. This difference was later attributed to the lower haplogroup diversity in circumarctic populations than in Amerind populations, because most circumarctic populations are nearly fixed for haplogroup A (Lorenz and Smith 1997). Lorenz and Smith (1997) identified several shared lineages between the Haida and Bella Coola, but unlike Ward et al. (1993), they argued that gene flow is a more likely explanation for the similarities than recent common ancestry. Both Lorenz and Smith (1997) and Stone and Stoneking (1998) suggested that the ancestors of the Haida probably colonized the Americas at the same time as populations in the Amerindian group, but isolation and genetic drift caused the pattern of reduced diversity within them.

## **Biocultural Context**

The principal language families of the Columbia Plateau are Salishan and Penutian. The Salishan languages occur on the Northwest Coast and extend into the northern regions of the plateau. This language family is believed to have originated on the Northwest Coast near the lower Fraser Valley and spread to the plateau across the Coastal Range (Elmendorf 1965; Suttles and Elmendorf 1963). Linguistic evidence of this homeland is based on the greater linguistic diversity among the Salishan languages located near the coast than among the Salishan languages on the plateau. Kinkade (1991) provided lexical evidence of a coastal origin for the Salish language family by reconstructing proto-Salish words for several Pacific coastal shellfish and plants. The degree of diversity among the Salishan languages suggests that the interior expansion occurred approximately 1500 yr B.P. (Suttles 1987). The Wakashan language family, spoken by the Nuu-Chah-Nulth (formerly known as the Nootka) located on Vancouver Island, has been argued to be distantly related to Salish. Together, both the Salishan and Wakashan language families make up what Swadesh (1953) called the Mosan language superfamily. This relationship, however, has not been proven to the satisfaction of many American historical linguists (Campbell 1997).

Speakers of Penutian languages, which are unrelated to Salish and other language families on the Columbia Plateau, are located primarily in the southern region of the plateau, the Willamette Valley, and California's Central Valley (Delancey and Golla 1997). The greatest language diversity within the Penutian superfamily is located in the southwesternmost Columbia Plateau and Willamette Valley, suggesting that this region is the homeland of Penutian. Living on the boundary between the plateau and Great Basin are speakers of the Numic language, a part of the Uto-Aztecan language family, whose homeland has long been debated (Hill 2001).

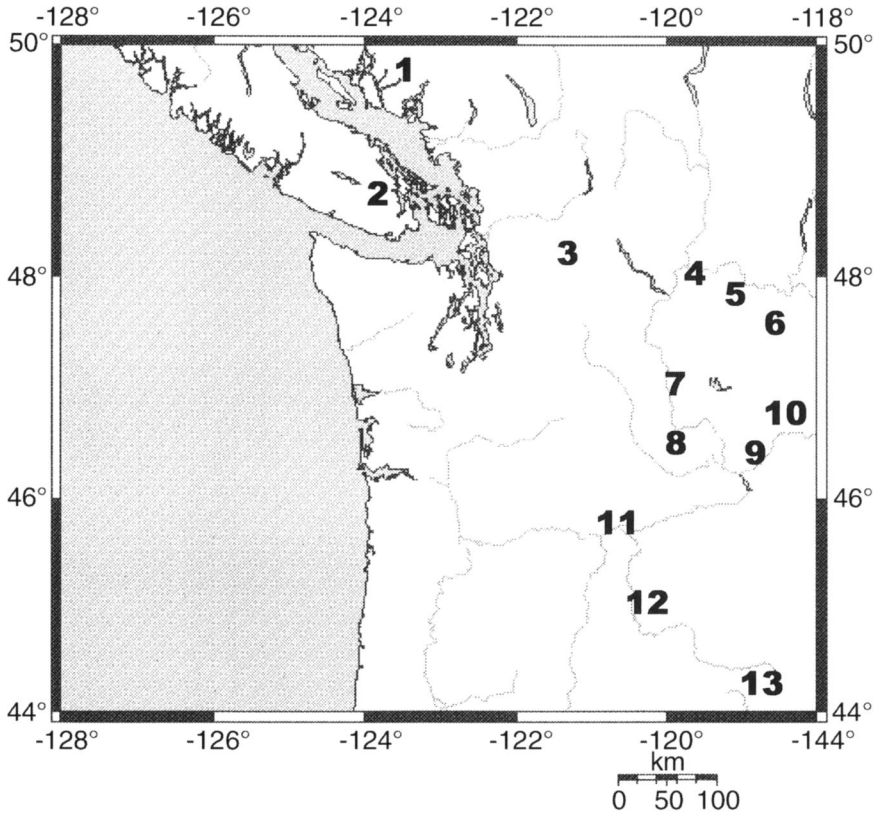
The prehistory of the Northwest shows extensive evidence of high settlement mobility, extensive interregional trade, and widespread conflict. In the early

Holocene Epoch, the Columbia Plateau was occupied by small, extremely mobile bands of hunter-gatherers (Chatters and Pokotylo 1998) associated with the Western Stemmed Tradition, a technological pattern found throughout the Great Basin and into California as well (Willig and Aikens 1988; Moratto 1984). At the same time, populations along the coast of British Columbia were using a microblade technology with apparent links to southeastern Alaska and the Arctic (Carlson 1996; Matson and Copeland 1995). By 8000 yr B.P. obsidian from the northwestern Great Basin and olivella shell beads from the Pacific coast had found their way to Marmes Rockshelter, near the confluence of the Palouse and Snake Rivers, nearly 300 mi upriver from the mouth of the Columbia River (Ames et al. 1998), providing strong evidence of interregional trade that persisted throughout prehistory. The pattern of high mobility and migration persisted until after 5000 yr B.P., when evidence of more sedentary lifeways appears in the archeological record on both the Northwest Coast and the Columbia Plateau (Chatters 1995; Matson and Copeland 1995). Distinctive cultures of the Northwest Coast and the Columbia Plateau that are recognizable in historic times developed between approximately 1500 and 2500 yr B.P. (Chatters and Pokotylo 1998; Matson and Copeland 1995).

Intertribal warfare was common along the Northwest Coast, with Nuu-Chah-Nulth, Haida, Tlingit, and other Athapaskan and Salish groups frequently raiding each other and taking slaves (Ames and Maschner 1999). In addition, archeological evidence of conflict in the Northwest exists from the time of the Kennewick Man (over 8000 yr B.P.) to European contact (Chatters 2000, in press). Most groups in the Northwest retained a more or less mobile hunter-gatherer lifestyle with extensive intergroup interaction up to the time of European contact (Driver and Massey 1957).

## Materials and Methods

**Populations Studied.** The locations of the modern populations and ancient groups analyzed in this study are shown in Figure 1. The sources for samples from the Bella Coola (a coastal Salish group) and the Nuu-Chah-Nulth are described by Smith, Lorenz et al. (2000). The Vantage group is represented by tooth samples from the Middle Columbia River and is radiocarbon dated to 500–1500 yr B.P. (S. Hackenberger, unpublished data, 1999). The protohistoric tooth samples from San Poil, Nespelem, Wenatchee, and Douglas County are approximately 200 years old and were found in regions occupied by speakers of Salish languages in prehistoric times. The protohistoric tooth samples from the Snake River, John Day River, and Palouse River are approximately 200 years old and were collected in regions occupied by Sahaptian speakers in prehistoric times. Five protohistoric tooth samples known only to be from the general Columbia Plateau region were also analyzed. The samples were provided by Central Washington University, Ellensburg. Tooth samples from the protohistoric Memaloose



**Figure 1.** Geographic location of ancient and modern Native American groups analyzed in this study: (1) Bella Coola (Salishan); (2) Nuu-Chah-Nulth (Wakashan); (3) Wenatchee (Salishan); (4) Nespelem (Salishan); (5) San Poil (Salishan); (6) Douglas (Salishan); (7) Vantage (500–100 yr B.P.); (8) Yakama (Sahaptin); (9) Snake River (Sahaptin); (10) Palouse (Sahaptin); (11) Wishram (Chinook, 200 yr B.P.); (12) John Day (Sahaptin); (13) Northern Paiute (Numic).

Island cemetery, near the Dalles, were provided by Yale University and were also analyzed as part of this study; they are estimated to date to 200 yr B.P. Individuals from this population were probably related to the Chinookan-speaking Wishram, who lived nearby on the north side of the Columbia River.

The haplogroups of 54 full-blooded modern and 64 ancient Native Americans were determined by restriction fragment length polymorphism (RFLP) analysis. A subset of 10 modern and 30 ancient samples was sequenced (from np 16,055 to np 16,548 for modern samples and from np 16,055 to np 16,356 for ancient samples).

**DNA Extraction and Typing.** Modern DNA was extracted and typed as described by Malhi et al. (2001). All ancient DNA was extracted, and the PCR

setup was conducted inside an ultraviolet-irradiated glove box in a laboratory with filtered ventilation and positive air pressure; this setup was designed for use with only ancient materials. All equipment and surfaces in this facility were routinely decontaminated with bleach, and all reagents and solutions were routinely checked for DNA contamination by PCR amplification using human primers specific to the mtDNA control region. Laboratory coats (restricted to use only in the ancient DNA laboratory), latex gloves, hairnets, booties, and face masks were worn while working with all ancient materials. Access to the ancient DNA facilities was restricted to lab personnel who had not worked with amplified DNA within the preceding 24-hr period.

The ancient teeth were decontaminated by cleaning their surfaces with 0.6% sodium hypochlorite. The teeth were lightly cracked, demineralized in approximately 1.5 ml of 0.5 M EDTA for 24–72 hr with mild rocking, then digested overnight with 2 mg/ml of Proteinase K (Qiagen) at 55°C with mild rocking. DNA was extracted from digested samples using a three-step phenol/chloroform extraction method: Two extractions using phenol:chloroform (1:1) were followed by an extraction with chloroform:isoamyl alcohol (24:1). The DNA solution was concentrated by precipitation using two volumes of cold 100% ethanol and one-half volume cold 5 M ammonium acetate. This solution was stored in the freezer for up to 3 hr, then centrifuged at high speed. After removing and discarding the supernatant, the DNA pellet was dried and resuspended in 100  $\mu$ l ddH<sub>2</sub>O. To remove co-extracted PCR inhibitors, the DNA solution was purified using the Promega Wizard PCR Preps DNA Purification Kit, as directed by the manufacturer, and stored at –20°C. DNA samples were diluted (one-half to one-twentieth) as needed before setting up PCR reactions.

PCR amplification reactions were carried out in 25- $\mu$ l volumes with 3- $\mu$ l DNA template, 20  $\mu$ M of each primer, 10 $\times$  Buffer [50  $\mu$ M Tris (pH 8.4), 1.5  $\mu$ M MgCl<sub>2</sub>, 20  $\mu$ M NaCl, 500 mg/ml BSA], 1.5 units Platinum Taq (Gibco), 200 mM of each dNTP, and 13.3  $\mu$ l ddH<sub>2</sub>O. After an initial 4-min denaturation at 95°C, 37 cycles were performed, consisting of 30 sec of denaturation at 95°C, 30 sec of annealing at 52°–55°C, and 30 sec of extension at 72°C, followed by a final 3-min extension at 72°C. A 5- $\mu$ l portion of amplification product was electrophoresed on a 6% polyacrylamide gel and stained with ethidium bromide to confirm the presence of PCR product of an appropriate size. To assess the presence or absence of diagnostic restriction sites, restriction enzymes were added to the remaining 20  $\mu$ l and incubated overnight at 37°C. The primers used for amplification of these segments were those described by Smith et al. (1999). A negative control was run for each DNA extraction and PCR amplification. Haplogroups were confirmed by at least one additional extraction from each prehistoric tooth sample and by at least two PCR amplifications of each extraction.

A 301-bp region of HVSI was sequenced for ancient samples using two sets of overlapping primers. Coordinates for these primers, labeled L16055, H16218, L16209, and H16356, have been described by Kaestle (1998). The PCR products were filtered using a Microcon 100 filter unit (Millipore) and then were



submitted for sequencing to the DBS Automated DNA sequencing facility at the University of California, Davis. Both the heavy and light strands were sequenced to preclude undetected sequencing errors. All sequences generated in this study, including contaminant modern DNA sequences, are reported in Appendix A.

**DNA Haplogroup and Sequence Analysis.** Sequences from the Mouse database (<http://www.gen-epi.de/mouse/>), Shields et al. (1993), and Kaestle (1998) were used in addition to the sequences generated in this study to construct haplotype networks. Sequences from other sources include Athapaskan ( $N = 7$ ) (Shields et al. 1993), Bella Coola ( $N = 13$ ) (Ward et al. 1993), Haida ( $N = 7$ ) (Ward et al. 1993), Inuit ( $N = 6$ ) (Shields et al. 1993), Northern Paiute ( $N = 17$ ) (Kaestle 1998), Nuu-Chah-Nulth ( $N = 20$ ) (Ward et al. 1993), Pyramid Lake ( $N = 16$ ) (Kaestle 1998), Shoshoni ( $N = 2$ ) (Kaestle 1998), Tlingit ( $N = 1$ ) (Torroni et al. 1993), Yakama ( $N = 6$ ) (Shields et al. 1993), and Yurok ( $N = 2$ ) (Lorenz and Smith 1997).

Pairwise comparisons and tests for homogeneity of haplogroup frequency distributions were made between all populations and groups using Fisher's exact probability test (Raymond and Rousset 2000). Genetic distances were calculated between all pairs of populations using Nei's distance (1987) and the Cavalli-Sforza chord distance in Gendist, and phylogenetic trees were constructed using the neighbor-joining method in Neighbor and Drawtree in the PHYLIP, version 3.572, software package (Felsenstein 1993). Nucleotide positions 16,182–16,183 were excluded from the analysis of haplogroup B haplotypes because polymorphism at these sites appears to be hypervariable and therefore neither informative of phylogenetic relationships nor reported in a consistent manner by different investigators. Haplotype statistical parsimony networks were constructed using the TCS Network Program (Clement et al. 2000).

$\theta_s$  was calculated using the Arlequin package (Schneider et al. 1997) and Microsoft Excel, as

$$\theta_s = \frac{S}{\sum_{i=1}^{n-1} \frac{1}{i}}$$

(Watterson 1975).

$\theta_s$  estimates gene diversity within a population independent of haplotype frequencies, where  $S$  is the number of segregating sites and  $n$  is the number of sequences used in the study. Therefore,  $\theta_s$  reflects the diversity in a population resulting from long-term history and is less influenced by generational and sampling effects than nucleotide diversity. Pairwise  $F_{ST}$  estimates were calculated for haplotypes using the Arlequin package (Schneider et al. 1997). Mantel tests (Mantel 1967) were conducted to compare genetic and geographic distances using the Matrix program (Smouse and Long 1992).

## Results

The haplogroup frequencies of the samples studied are listed in Table 1 and in Appendix B. The modern Northwest Coast groups are clearly separated from the modern Plateau–Great Basin groups in the neighbor-joining tree of ancient and modern groups in the Northwest (Figure 2a). The ancient Vantage group clusters with the Nuu-Chah-Nulth of the Northwest Coast rather than with the Plateau–Great Basin groups.

The mtDNA haplotype networks were constructed with sample sizes of 43, 24, 22, and 40 for haplogroups A, B, C, and D, respectively (Figure 3). The haplogroup A network was constructed in two steps to avoid excessive reticulation. Many peripheral haplotypes in the network are shared between Northwest Coast and Columbia Plateau individuals. In addition, as previously demonstrated by Forster et al. (1996), a Bella Coola haplotype occupies a clade defined by a transition at np 16,355 that predominantly includes Haida individuals. An ancient Vantage haplotype (radiocarbon dated to  $1230 \pm 40$  yr B.P.) clusters with haplotypes from the Inuit because of a transition at np 16,212. As originally noticed by Forster et al. (1996), a modern Yakama haplotype clustered with a predominantly Athapaskan group defined by transitions at np 16,192, np 16,233, and np 16,331.

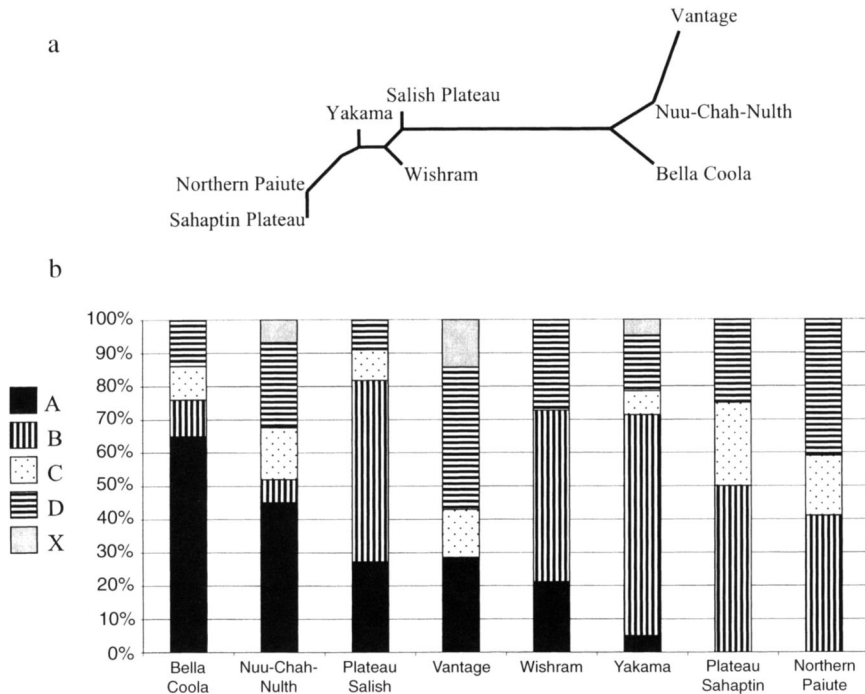
The haplogroup B network was constructed in three steps and exhibits a high frequency of the basal haplotype. A transition at np 16,319 defines a clade of solely Columbia Plateau haplotypes, whereas ancient Great Basin individuals share haplotypes with modern Northwest Coast and Northern California individuals. The three-step network for haplogroup C demonstrates a restricted distribution of the Northern Paiute haplotypes. In addition, tribal-specific markers occurred at np 16,242 for the Nuu-Chah-Nulth and at np 16,343 for the Yakama. An ancient Vantage haplotype (radiocarbon dated to  $950 \pm 40$  yr B.P.) shares a transition at np 16,086 with two Bella Coola individuals. The three-step network for haplogroup D, like that for haplogroup B, exhibits a high frequency of the basal haplotype. This network includes a peripheral haplotype that is shared between a Bella Coola and a Nuu-Chah-Nulth individual. In addition, a clade defined by a transition at np 16,274 defines a primarily Wishram-specific cluster, which includes one Nuu-Chah-Nulth individual, and another clade defined by a transition at np 16,294 is represented by two ancient Great Basin, two Columbia Plateau, and one Northwest Coast haplotypes. Of three Northwest haplotypes assigned to haplogroup X, one is shared among the Yakama, Nuu-Chah-Nulth, and Vantage groups, and the others are each shared by one Nuu-Chah-Nulth and one protohistoric sample from the Columbia Plateau.

As illustrated in Table 2, the Northern Paiute and Bella Coola exhibit unusually low diversity for haplogroup D compared to other groups in their regions, and the Northwest Coast and Columbia Plateau groups exhibit low diversity in haplogroup B. However, the ancient samples from the Great Basin displayed diversity similar to that seen in modern Great Basin and Southwest groups.

When all haplogroups were used to construct a genetic distance (Nei's distance based on haplogroup frequency distributions), the Mantel test exhibited a

**Table 1.** Haplogroup Frequency Distribution of Ancient and Modern Native American Groups in Northwestern North America

<i>Group</i>	<i>Age</i>	<i>Language Family</i>	<i>Geographic Location</i>	<i>N</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>X</i>	<i>Reference</i>
Plateau Salish	200 yr B.P.	Salishan	Northern Plateau	11	0.273	0.545	0.091	0.091	0.000	This study
Plateau Sahaptian	200 yr B.P.	Sahaptian	Southern Plateau	8	0.000	0.500	0.250	0.250	0.000	This study
Yakama	Modern	Sahaptian	Southern Plateau	42	0.050	0.664	0.071	0.167	0.048	Shields et al. (1993)
Wishram	200 yr B.P.	Chinook	Southern Plateau	33	0.212	0.515	0.000	0.273	0.000	This study
Vantage	500–1500 yr B.P.	—	Northern Plateau	7	0.286	0.000	0.143	0.429	0.143	This study
Bella Coola	Modern	Salishan	Coast	84	0.650	0.110	0.100	0.140	0.000	Torroni et al. (1993), Ward et al. (1993), this study ( <i>n</i> = 26)
Nuu-Chah-Nulth	Modern	Wakashan	Coast	102	0.451	0.069	0.157	0.255	0.069	Ward et al. (1991, 1993), this study ( <i>n</i> = 28)
Northern Paiute	Modern	Numic	Great Basin	98	0.000	0.410	0.180	0.410	0.000	Kaestle and Smith (2001)



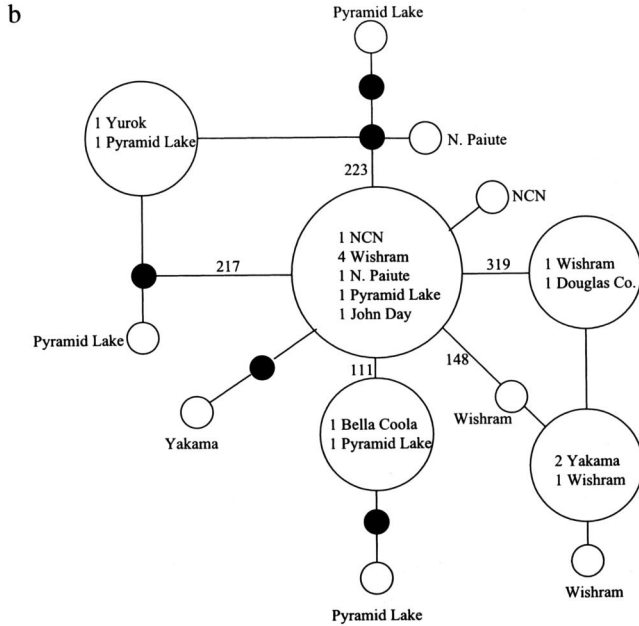
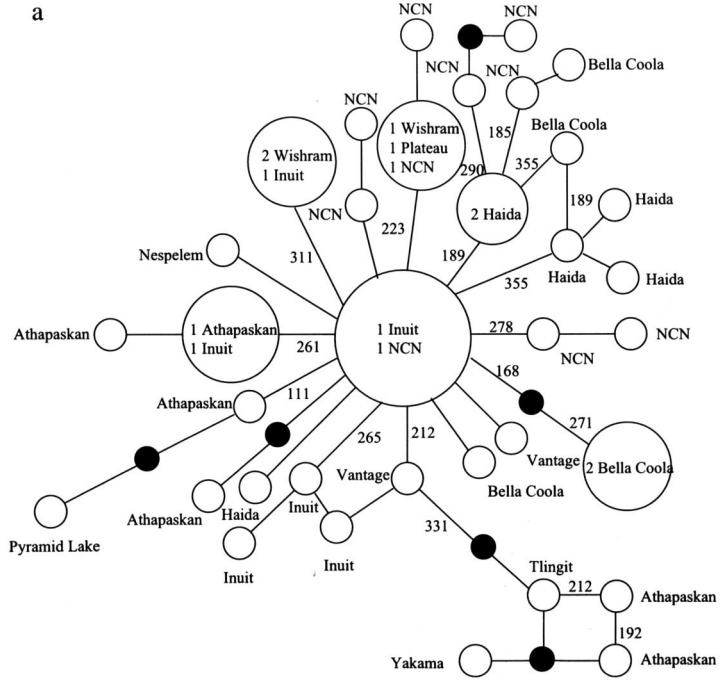
**Figure 2.** (a) Neighbor-joining tree constructed using the Cavalli-Sforza chord distance measure of ancient and modern populations from the Northwest Coast, Columbia Plateau, and Great Basin. (b) Histogram of haplogroup frequency distributions.

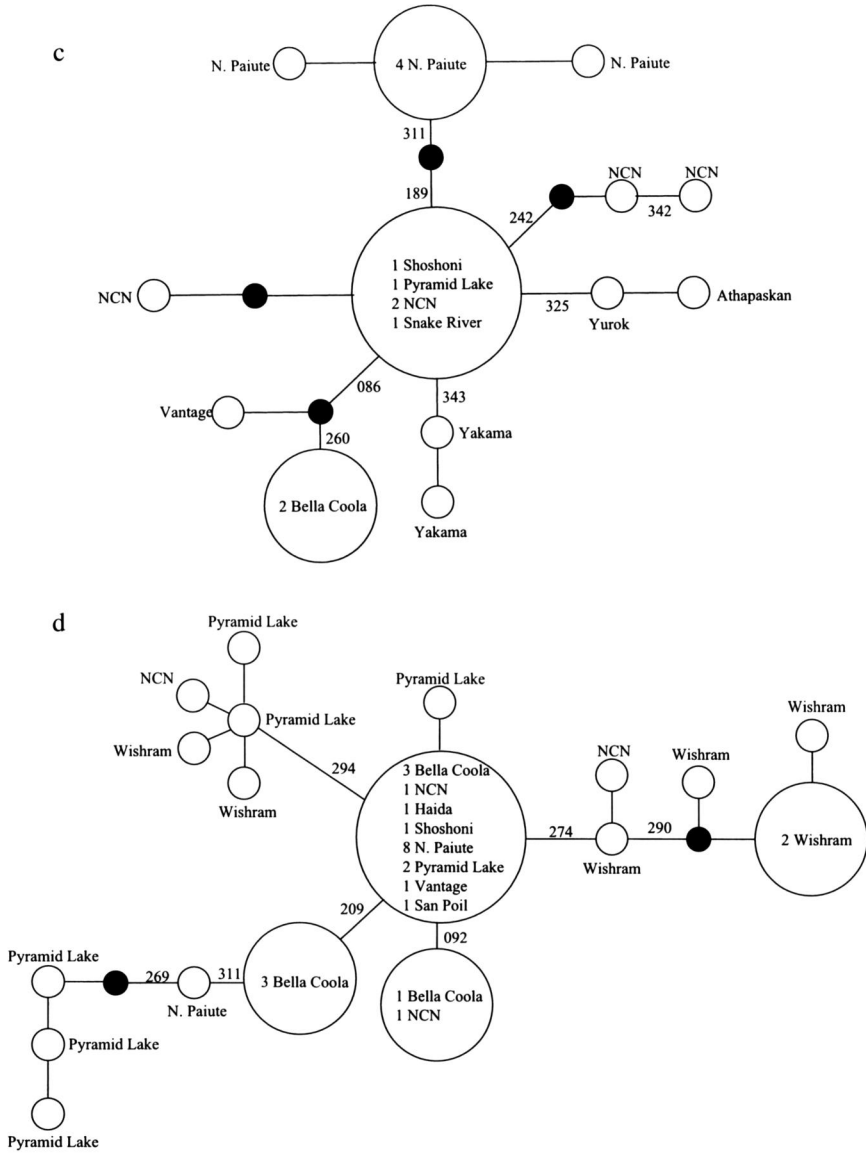
significant correlation with geographic distance. However, only distances constructed from haplogroup A exhibited a correlation with geographic distance, when distances were constructed from each of the independent haplogroups. This same pattern persisted when groups of  $N < 30$  were excluded to reduce the confounding factor of sampling error (Table 3).

## Discussion

Modern populations on the Northwest Coast and Columbia Plateau, like those in the American Southwest (Malhi et al. 2003), Central Mexico (Kemp et al., in press), and northeastern North America (Malhi et al. 2001), display a general pattern of regional continuity in haplogroup frequency distributions. Columbia Plateau groups exhibit high frequencies of haplogroups B and D, whereas coastal groups exhibit high frequencies of haplogroup A along with moderate frequencies of haplogroups B, C, and D. A strong pattern of geographic structuring occurs with haplogroup A, which decreases in frequency with distance from the coast (Figure 4). The Vantage group (the ancient Columbia Plateau group),

*mtDNA Diversity in Northwestern North America / 43*





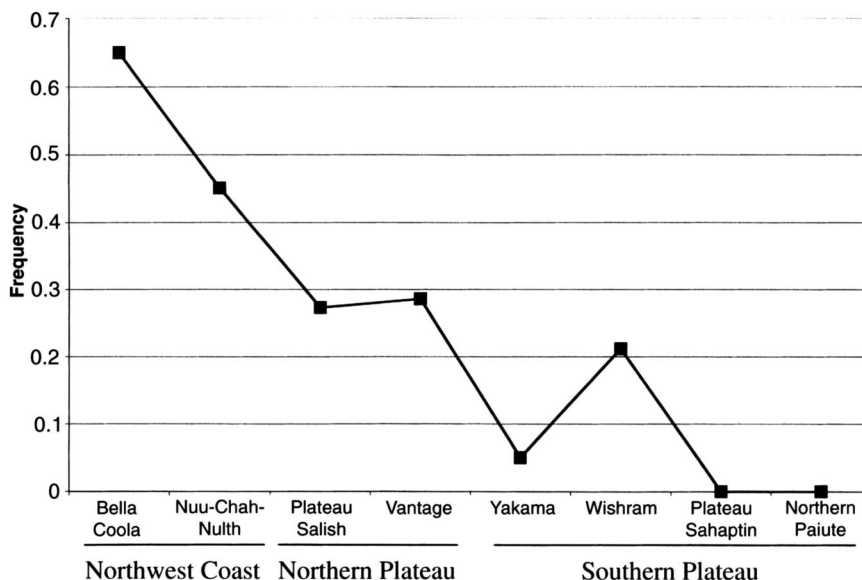
**Figure 3.** (a) Haplogroup A network; (b) haplogroup B network; (c) haplogroup C network; (d) haplogroup D network. Numbers correspond to the last three digits of the nucleotide position and indicate defining mutations for each clade. Size of circle and numbers preceding names correspond to the number of individuals found with that haplotype. Solid circles represent hypothetical haplotypes not found in our sample. NCN, Nuu-Chah-Nulth.

**Table 2.** Haplotype Diversity of Native American Groups for Haplogroups D and B

<i>Population</i>	<i>Region</i>	<i>Individuals</i>	<i>Number of Haplotypes</i>	$\theta_s$	<i>Reference</i>
<b>Haplogroup D</b>					
Bella Coola	Northwest Coast	6	3	1.314	Torroni et al. (1993), Ward et al. (1993), this study ( $n = 4$ )
Northern Paiute	Great Basin	12	4	1.656	Kaestle (1998)
Creek	Southeast	6	5	3.066	Bolnick and Smith (2003)
Wishram	Columbia Plateau	7	7	3.266	This study
Nuu-Chah-Nulth	Northwest Coast	4	4	3.273	Ward et al. (1991, 1993)
Pyramid Lake	Great Basin	8	7	3.471	Kaestle (1998)
<b>Haplogroup B</b>					
Nuu-Chah-Nulth	Northwest Coast	5	4	1.440	Torroni et al. (1993), Ward et al. (1991)
Wishram	Columbia Plateau	5	3	1.440	This study
Yakama	Columbia Plateau	15	4	1.540	Shields et al. (1993)
Pueblo	Southwest	15	7	2.150	Malhi (2001)
Piman	Southwest	8	6	3.090	Malhi (2001)
Pyramid Lake	Great Basin	6	6	3.504	Kaestle (1998)
Yuman	Southwest	11	8	4.100	Malhi (2001)
Washo	Great Basin	5	4	4.320	Kaestle (1998)

**Table 3.** Correlation Between Genetic Distance and Geographic Distance

<i>Haplogroup</i>	<i>Coefficient of Correlation</i>	$R^2$	<i>P Value</i>
All	0.644	0.415	0.005
All ( $N > 30$ )	0.546	0.298	0.034
A	0.810	0.656	0.001
B	0.211	0.045	0.501
C	-0.204	0.042	0.591
D	0.434	0.189	0.179
X	-0.169	0.029	0.743
<b>With Vantage ancient group</b>			
All	0.321	0.103	0.084
A	0.802	0.644	0.001
B	-0.112	0.012	0.637
C	-0.292	0.085	0.350
D	0.107	0.012	0.691
X	-0.235	0.055	0.439



**Figure 4.** Frequency of haplogroup A in populations from the Northwest Coast to the northern Columbia Plateau to the southern Columbia Plateau.

however, exhibits a similarity in haplogroup frequency distribution to the coastal Nuu-Chah-Nulth ( $p = 0.659$ ), although it is not significantly different from the Salish and Sahaptian plateau groups ( $p = 0.060$  and  $0.110$ , respectively). Although a similarity of the haplogroup frequency distribution of the Vantage group and the Nuu-Chah-Nulth could likely result from sampling error, an ancient connection to the Northwest Coast is also supported by haplotype data. The Vantage group exhibits a haplotype of haplogroup A defined by a transition at np 16,212 that is found among the coastal Inuit and a Tlingit individual (Torrioni et al. 1993). In addition, the inland Yakama have a haplotype of haplogroup A defined by transitions at np 16,192, np 16,233, and np 16,331 that is also found in high frequency in Athapaskan populations. Haplogroup and haplotype A distributions among the Northwest populations suggest that a Coastal/Subarctic population intruded into the northern Columbia Plateau before 1500 yr B.P., perhaps represented by the hypothesized Salishan expansion.

The decreased diversity of haplogroup D, in addition to the high frequency of haplogroup A, in the Bella Coola compared to other Northwest populations is not unexpected, considering the extensive admixture between the Bella Coola and groups to the north, such as the Haida, who exhibit a high frequency of haplogroup A and the near absence of haplogroup D. The similarity in diversity of haplogroups B and D between the Nuu-Chah-Nulth and the Wishram suggests that these groups experienced similar population histories. This evidence, along with the similarities in haplogroup frequencies between northern Columbia Plateau groups and the Nuu-Chah-Nulth, suggests that the Nuu-Chah-Nulth experienced greater contact with plateau groups than the Bella Coola. Increased contact



between the Nuu-Chah-Nulth and plateau groups is also supported by the appearance of *Dentalium* shells from Vancouver Island on the Columbia Plateau approximately 3500 yr B.P. (Ames et al. 1998).

Other than haplogroup A, haplogroup frequencies are not structured according to geographic distance, suggesting that genetic drift played a significant role in the patterns differentiating these groups, at least during more recent prehistory. However, additional shared haplotypes of haplogroups B, C, D, and X are observed between Coastal and Columbia Plateau populations. The ancient Vantage group, the Nuu-Chah-Nulth, and the Yakama share a haplogroup X haplotype. In addition, a haplogroup C haplotype from the Vantage group clusters with two Bella Coola haplotypes that share a transition at np 16,086. The significant number of shared derived mutations and haplotypes among Northwest populations suggests that these groups probably maintained some interregional gene flow during their prehistory.

On the Columbia Plateau the difference in haplogroup frequency distributions between the Wishram (Chinook language) and both the Yakama and the Sahaptian plateau populations is statistically significant at or near the 0.05 level of probability ( $p = 0.048$  and  $0.056$ , respectively). This variation might be explained by genetic drift and sampling error, or the Wishram's high frequency of haplogroup A could have resulted from admixture with Coastal populations. The Chinook were the middlemen in the Northwest slave trade between these two regions (Ames and Maschner 1999), providing one cultural mechanism for this admixture. In addition, the Chinook language family extends from the middle Columbia River to the Pacific Coast, suggesting that Chinook speakers migrated up and down the Columbia River, maintaining contact with both Northwest Coast and Columbia Plateau peoples.

Overall, the genetic data presented here suggest that a significant population intrusion from the Coastal/Subarctic region occurred in prehistoric times, possibly co-occurring with the spread of the Salishan language family. The additional sharing of non-A haplotypes suggests either that other earlier expansions might have occurred or that a substantial amount of gene flow occurred among Northwestern groups. Warfare, along with the abundant interregional trade networks, including the slave trade, probably contributed to the constant flow of migrants that were exchanged among groups in the Northwest Coast, Columbia Plateau, and Great Basin, which in turn influenced the mtDNA pattern of populations in the Northwest.

*Acknowledgments* This research was funded by the National Science Foundation through Dissertation Improvement Grant BCS-0002187 and by the Wenner Gren Foundation through a research grant to R.S. Malhi. We would like to thank Dianah Ngo, Max Pavesic, and Robert Yohe for their advice.

*Received 4 December 2002; revision received 29 July 2003.*

## Literature Cited

- Ames, K.M., D.E. Dumond, J.R. Galm et al. 1998. Prehistory of the Southern Plateau. In *Handbook of North American Indians*, W.C. Sturtevant, ed. Washington, DC: Smithsonian Institution, v. 12, 103–119.
- Ames, K.M., and H.D.G. Maschner. 1999. *Peoples of the Northwest Coast: Their Archeology and Prehistory*. New York: Thames and Hudson.
- Bandelt, H.J., P. Forster, and A. Rohl. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molec. Biol. Evol.* 16:37–48.
- Bolnick, D.A.W., and D.G. Smith. 2003. Unexpected patterns of mitochondrial DNA variation among Native Americans from the Southeastern United States. *Am. J. Phys. Anthropol.*, published online on May 5, 2003.
- Brown, M.D., S.H. Hosseini, A. Torroni et al. 1998. MtDNA haplogroup X: An ancient link between Europe/Western Asia and North America? *Am. J. Hum. Genet.* 63:1852–1861.
- Campbell, L. 1997. *American Indian Languages: The Historical Linguistics of Native America*. New York: Oxford University Press.
- Carlson, R.L. 1996. Introduction to early human occupation in British Columbia. In *Early Human Occupation in British Columbia*, R.L. Carlson and L. Dalla Bona, eds. Vancouver: UBC Press, 3–10.
- Carlyle, S.W., R.L. Parr, M.G. Hayes et al. 2000. Context of maternal lineages in the greater Southwest. *Am. J. Phys. Anthropol.* 11:85–101.
- Cavalli-Sforza, L.L., and A.W.F. Edwards. 1967. Phylogenetic analysis models and estimation procedures. *Am. J. Hum. Genet.* 19:223–257.
- Cavalli-Sforza, L.L., P. Menozzi, and A. Piazza. 1994. *The History and Geography of Human Genes*. Princeton, NJ: Princeton University Press.
- Chatters, J.C. 1995. Population growth, climatic cooling, and the development of collector strategies on the Southern Plateau, western North America. *J. World Prehist.* 9:341–400.
- Chatters, J.C. 2000. The discovery and first analysis of an early Holocene human skeleton from Kennewick, Washington. *Am. Anthropol.* 65:291–316.
- Chatters, J.C. In press. Safety in numbers: The influence of the bow and arrow on village formation on the Southern Plateau. In *Complex Hunter-Gatherers: Evolution and Organization of Prehistoric Communities on the Plateau of Northwestern North America*. W.C. Prentiss and I. Kuijt, eds. Salt Lake City: University of Utah Press.
- Chatters, J.C., and D. Pokotylo. 1998. Prehistory: Introduction. In *Handbook of North American Indians*, W.C. Sturtevant, ed. Washington, DC: Smithsonian Institution, v. 12, 73–80.
- Clement, M., D. Posada, and K.M. Crandall. 2000. TCS: A computer program to estimate gene genealogies. *Molec. Ecol.* 9:1657–1659.
- Delancey, S., and V. Golla. 1997. The Penutian hypothesis: Retrospect and prospect. *Int. J. Am. Linguist.* 63:171–202.
- Derenko, M.V., T. Grzybowski, B.A. Malyarchuk et al. 2001. The presence of mitochondrial haplogroup X in Altaians from South Siberia. *Am. J. Hum. Genet.* 69:237–241.
- Driver, H.E., and W.C. Massey. 1957. Comparative studies of North American Indians. *Trans. Am. Phil. Soc.* 47:165–456.
- Elmendorf, W.W. 1965. Linguistic and geographic relations in the Northern Plateau Area. *SW J. Anthropol.* 17:365–382.
- Excoffier, L., and A. Langaney. 1989. Origin and differentiation of human mitochondrial DNA. *Am. J. Hum. Genet.* 44:73–85.
- Felsenstein, J. 1993. *PHYLIP (Phylogeny Inference Package)*. Seattle: Department of Genetics, University of Washington.
- Forster, P., R. Harding, A. Torroni et al. 1996. Origin and evolution of Native American mtDNA variation: A reappraisal. *Am. J. Hum. Genet.* 59:935–945.

- Foster, M.K. 1996. Language and the culture history of North America. In *Handbook of North American Indians*, W.C. Sturtevant, ed. Washington, DC: Smithsonian Institution, v. 17, 64–110.
- Greenberg, J.H., C.G.L. Turner, and S.L. Zegura. 1986. The settlement of the Americas: A comparison of linguistic, dental, and genetic evidence. *Curr. Anthropol.* 27:477–497.
- Hauswirth, W.W., C.D. Dickel, D.J. Rowold et al. 1994. Inter- and intrapopulation studies of ancient humans. *Experient* 50:585–591.
- Hill, J.H. 2001. Proto-Uto-Aztecan: A community of cultivators in central Mexico? *Am. Anthropol.* 103:913–934.
- Kaestle, F.A. 1998. Molecular evidence for prehistoric Native American population movement: The Numic expansion. Ph.D. dissertation, University of California, Davis.
- Kaestle, F.A., and D.G. Smith. 2001. Ancient mitochondrial DNA evidence for a prehistoric population movement: The Numic expansion. *Am. J. Phys. Anthropol.* 115:1–12.
- Kemp, B.M., A. Resendez, J.A. Roman Berrelleza et al. In press. An analysis of ancient mtDNA from Tlateloco: Pre-Columbian relations and the spread of Uto-Aztecan ancient DNA and archeology. University of Southern Illinois, Carbondale.
- Kinkade, M.D. 1991. Prehistory of the native languages of the Northwest Coast. In *Proceedings of the Great Ocean Conference*, v. 1, *The North Pacific to 1600*. Portland, OR: Oregon Historical Society Press, 137–158.
- Long, J.C. 1991. The genetic structure of admixed populations. *Genetics* 127:417–428.
- Lorenz, J.G., and D.G. Smith. 1996. Distribution of four founding mtDNA haplogroups among Native North Americans. *Am. J. Phys. Anthropol.* 101:307–323.
- Lorenz, J.G., and D.G. Smith. 1997. Distribution of sequence variation in the mtDNA control region of native North Americans. *Hum. Biol.* 69:749–776.
- Malhi, R.S. 2001. Investigating prehistoric population movements in North America with ancient and modern mtDNA. Ph.D. dissertation, on file at the Department of Anthropology, University of California, Davis.
- Malhi, R.S., J.A. Eshleman, J.A. Greenberg et al. 2002. The structure of diversity within New World mitochondrial DNA haplogroups: Implications for the prehistory of North America. *Am. J. Hum. Genet.* 70:905–919.
- Malhi, R.S., H.M. Mortensen, J.A. Eshleman et al. 2003. MtDNA Prehistory in the American Southwest. *Am. J. Phys. Anthropol.* 120:108–124.
- Malhi, R.S., B.A. Schultz, and D.G. Smith. 2001. Distribution of mitochondrial DNA lineages among Native American tribes of northeastern North America. *Hum. Biol.* 73:17–55.
- Malhi, R.S., and D.G. Smith. 2002. Haplogroup X confirmed in prehistoric North America [brief communication]. *Am. J. Phys. Anthropol.* 119:84–86.
- Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27:209–220.
- Matson, R.G., and G. Copeland. 1995. *Prehistory of the Northwest Coast*. New York: Plenum.
- Moratto, M.J. 1984. *Archeology of California*. New York: Academic Press.
- Nei, M. 1987. *Molecular Evolutionary Genetics*. New York: Columbia University Press.
- O'Rourke, D.H., M.G. Hayes, and S.W. Carlyle. 2000. Spatial and temporal stability of mtDNA haplogroup frequencies in native North America. *Hum. Biol.* 72:15–34.
- Parr, R.L., S.W. Carlyle, and D. O'Rourke. 1996. Ancient DNA analysis of Fremont Amerindians of the Great Salt Lake wetlands. *Am. J. Phys. Anthropol.* 99:507–518.
- Raymond, M., and F. Rousset. 2000. Genepop on the Web. Population Differentiation Option. Available at <http://wbiomed.curtin.edu.au/genepop/>.
- Schneider, S., D. Roessli, and D.L. Excoffier. 1997. *Arlequin: A Software for Population Genetic Data Analysis*. Geneva: University of Geneva.
- Shields, G.F., A.M. Schmeichen, B.L. Frazier et al. 1993. MtDNA sequences suggest a recent evolutionary divergence for Beringian and northern North American populations. *Am. J. Hum. Genet.* 53:549–562.
- Smith, D.G., J. Lorenz, B.K. Rolfs et al. 2000. Implications of the distribution of albumin Naskapi and albumin Mexico for New World prehistory. *Am. J. Phys. Anthropol.* 111:557–572.

- Smith, D.G., R. Malhi, J.A. Eshleman et al. 2000. *Report on DNA Analysis of the Remains of "Kennewick Man" from Columbia Park*. Washington, DC: National Park Service.
- Smith, D.G., R.S. Malhi, J. Eshleman et al. 1999. Distribution of mtDNA haplogroup X among Native North Americans. *Am. J. Phys. Anthropol.* 110:271–284.
- Smouse, P.E., and J.C. Long. 1992. Matrix correlation analysis in anthropology and genetics. *Yrbk. Phys. Anthropol.* 35:187–213.
- Stone, A.C., and M. Stoneking. 1998. MtDNA analysis of a prehistoric Oneonta population: Implications for the peopling of the New World. *Am. J. Hum. Genet.* 62:1153–1170.
- Suttles, W. 1987. *Coast Salish Essays*. Vancouver: Talonbooks.
- Suttles, W., and W.W. Elmendorf. 1963. Linguistic evidence for Salish prehistory. In *Proceedings of the 1962 Annual Spring Meeting of the American Ethnological Society, Symposium on Language and Culture*, V. Garfield and W. Chafe, eds. Seattle: University of Washington Press, 42–52.
- Swadesh, M. 1953. Mosan I: A problem of remote common origin. *Int. J. Anthropol. Linguist.* 19:26–44.
- Torroni, A., T.G. Schurr, M.F. Cabell et al. 1993. Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am. J. Hum. Genet.* 53:563–590.
- Ward, R.H., B.L. Frazier, K. Dew-Jager et al. 1991. Extensive mitochondrial diversity within a single Amerindian tribe. *Proc. Natl. Acad. Sci. USA* 88:8720–8724.
- Ward, R.H., A. Redd, D. Valencia et al. 1993. Genetic and linguistic differentiation in the Americas. *Proc. Natl. Acad. Sci. USA* 90:10,663–10,667.
- Watterson, G. 1975. On the number of segregation sites in the genetical models without recombination. *Theor. Popul. Biol.* 7:256–276.
- Whistler, K.A. 1979. Linguistic prehistory in the Northwest California culture area. In *A Study of Cultural Resources in Redwood National Park*, P.M. Bicket, ed. Denver: National Park Service, 18–29.
- Willig, J.A., and C.M. Aikens. 1988. The Clovis–Archaic interface in far western North America. *Nevada State Mus. Anthropol. Papers* 21:1–40.

**Appendix A. HVS1 DNA Sequence**

	16111	16129	16168	16185	16212	16223	16230	16250	16271	16290	16311	16319	16362	N
<i>Haplogroup A</i>														
CRS	C	G	C	C	A	C	A	C	T	C	T	G	T	1
Nuu-Chah-Nulth A	T	A	.	T	.	T	.	.	.	T	.	A	.	1
Bella Coola A	T	.	T	.	T	.	.	C	.	T	.	A	C	1
Wishram A	T	.	.	.	T	.	.	.	C	T	C	A	?	2
Wishram A	T	.	.	.	.	.	.	.	.	T	.	A	?	1
Plateau A	T	.	.	.	.	.	.	.	.	T	.	A	?	1
Nespelem A	T	.	.	.	T	.	G	.	.	T	.	A	?	1
Vantage A	T	.	.	.	T	.	.	T	.	T	.	A	?	1
Vantage A	T	.	.	.	G	T	.	.	.	T	.	A	?	1
<i>Haplogroup B</i>														
CRS	C	A	T	T	C	T	G							
Northern Paiute B	.	.	C	C	.	.	.	I						
Northern Paiute B	.	.	C	C	T	C	.	I						
Wishram B	.	.	C	C	.	.	A	I						
Wishram B	.	.	C	C	.	.	.	I						
Wishram B	.	C	C	C	.	.	.	I						
Wishram B	T	C	C	C	.	.	A	I						
Wishram B	T	C	C	C	.	.	A	I						
Wishram B	.	C	C	C	.	.	.	I						
John Day B	.	C	C	C	.	.	.	I						
Douglas B	.	C	C	C	.	.	A	I						



Haplogroup X	16189	16213	16223	16255	16278	N
CRS	T	G	C	G	C	
Plateau X	?	A	T	A	T	1
Vantage X	C	A	T	.	T	1

Modern Contaminants Likely Source	16222	16234	16235	16240	16242	16256	16261	16270	16291	16292	16293	16311
CRS	C	C	A	A	C	C	C	C	C	C	A	T
JAE, BSS, DGS <sup>a</sup>	.	.	.	.	.	.	.	.	.	.	.	.
RSM <sup>a</sup>	.	T	.	C	G	.	.	.	.	.	.	C
JGL <sup>a</sup>	.	.	.	.	.	T	.	T	.	.	.	.
Unknown	.	.	G	.	.	.	.	.	T	.	G	.
Unknown	T	.	.	.	.	.	T	.	.	.	.	.

CRS, Cambridge Reference Sequence.

a. Past and present researchers in the laboratory include JAE (Jason A. Eshleman), BSS (Beth S. Shook), DGS (David G. Smith), RSM (Ripan S. Malhi), and JGL (Joseph G. Lorenz).

**Appendix B.** Haplogroup Assignment of Protohistoric Samples from the Columbia Plateau

<i>Sample</i>	<i>Location</i>	<i>Haplogroup</i>	<i>Probable Language Group</i>
GO 66	San Poil	B	Salishan
GO 55	San Poil	B	Salishan
GO 52	San Poil	B	Salishan
GO 61	San Poil	D	Salishan
GO 122	Douglas	B	Salishan
GO 126	Wenatchee	B	Salishan
GO 11	Rock Island (Wenatchee)	B	Salishan
GO 178	Nespelem	A	Salishan
GO 83	Nespelem	A	Salishan
GO 84	Nespelem	A	Salishan
GO 92	Nespelem	C	Salishan
GO 300	John Day	D	Sahaptian
GO 152	John Day	B	Sahaptian
GO 210	John Day	B	Sahaptian
GO 171	John Day	B	Sahaptian
GO 163	Palouse	B	Sahaptian
GO 15	Snake River	D	Sahaptian
GO 199	Snake River	C	Sahaptian
GO 216	Snake River	C	Sahaptian
GO 22	unknown	D	unknown
GO 11B	unknown	D	unknown
GO 89	unknown	A	unknown
GO 301	unknown	C	unknown
GO 19	unknown	X	unknown