Am. J. Hum. Genet. 63:1852–1861, 1998

mtDNA Haplogroup X: An Ancient Link between Europe/Western Asia and North America?

Michael D. Brown,[1] Seyed H. Hosseini,[1] Antonio Torroni,[2] Hans-Jürgen Bandelt,[3] Jon C. Allen,[1] Theodore G. Schurr,[1] Rosaria Scozzari,[2] Fulvio Cruciani,[2] and Douglas C. Wallace[1]

[1]Center for Molecular Medicine, Emory University School of Medicine, Atlanta; [2]Dipartimento di Genetica e Biologia Molecolare, Universita' di Roma "La Sapienza," Rome; and [3]Mathematisches Seminar, University of Hamburg, Hamburg

Summary

On the basis of comprehensive RFLP analysis, it has been inferred that ~97% of Native American mtDNAs belong to one of four major founding mtDNA lineages, designated haplogroups "A"–"D." It has been proposed that a fifth mtDNA haplogroup (haplogroup X) represents a minor founding lineage in Native Americans. Unlike haplogroups A–D, haplogroup X is also found at low frequencies in modern European populations. To investigate the origins, diversity, and continental relationships of this haplogroup, we performed mtDNA high-resolution RFLP and complete control region (CR) sequence analysis on 22 putative Native American haplogroup X and 14 putative European haplogroup X mtDNAs. The results identified a consensus haplogroup X motif that characterizes our European and Native American samples. Among Native Americans, haplogroup X appears to be essentially restricted to northern Amerindian groups, including the Ojibwa, the Nuu-Chah-Nulth, the Sioux, and the Yakima, although we also observed this haplogroup in the Na-Dene–speaking Navajo. Median network analysis indicated that European and Native American haplogroup X mtDNAs, although distinct, nevertheless are distantly related to each other. Time estimates for the arrival of X in North America are 12,000–36,000 years ago, depending on the number of assumed founders, thus supporting the conclusion that the peoples harboring haplogroup X were among the original founders of Native American populations. To date, haplogroup X has not been unambiguously identified in Asia, raising the possibility that some Native American founders were of Caucasian ancestry.

Received July 13, 1998; accepted for publication October 6, 1998; electronically published November 25, 1998.

Address for correspondence and reprints: Dr. Michael D. Brown, Center for Molecular Medicine, Emory University School of Medicine, 420B Dental School, 1462 Clifton Road, Atlanta, GA. E-mail: mdbrown@gmm.gen.emory.edu

© 1998 by The American Society of Human Genetics. All rights reserved.
0002-9297/98/6306-0031$02.00

Introduction

For Native Americans, extensive RFLP and control region (CR; also known as the "D-loop") sequence analysis has unambiguously identified four major founding mtDNA haplogroups, designated "A"–"D" (Torroni et al. 1992; 1993*a*). Together, these haplogroups account for ~97% of modern Native American mtDNAs surveyed to date (Torroni and Wallace 1994; Merriwether et al. 1995). Apparent non–haplogroup A–D mtDNAs can result from reversion of key A–D markers, recent admixture with non–Native Americans, or represent additional Native American founding mtDNA lineages. A striking example of the presence of non–haplogroup A–D genotypes in Native Americans can be seen in the Ojibwa, an Amerindian population from the Great Lakes region of North America. Using high-resolution RFLP analysis, Torroni et al. (1993*a*) found that 25% of the northern Ojibwa mtDNAs did not belong to haplogroups A–D and that nearly all of these "other" mtDNAs encompassed four distinct but related haplotypes characterized by the RFLP motif −1715 *Dde*I and +16517 *Hae*III. This motif was also present in 4% of the Navajo, but it was not observed in 18 other tribes from North, South, and Central America. The high incidence of this motif in the Ojibwa has been confirmed recently by Scozzari et al. (1997), who reported its presence in 26% of the southeastern Ojibwa from Manitoulin Island, Canada.

A recent survey of European mtDNAs has demonstrated the presence of the same "other" haplotype motif in modern European populations, in which it is called "haplogroup X." Haplogroup X represents ~4% of European mtDNAs and has been found to be further characterized by C→T transitions at nucleotide position (np) 16223 and np 16278 of the CR (Torroni et al. 1996). Thus, haplogroup X mtDNAs are minimally characterized by a combination of RFLP and CR markers including −1715 *Dde*I, +16517 *Hae*III, and the 16223T and 16278T mutations.

On the basis of the presence of the 16223T-and-

16278T motif in nine non–haplogroup A–D mtDNAs from the Nuu-Chah-Nulth and Yakima initially reported by Ward et al. (1991, 1993), Bailliet et al. (1994) suggested the existence of a fifth haplogroup in the Americas. More recently, Forster et al. (1996) analyzed the same data set and proposed that this haplogroup (haplogroup X) constituted an additional founding Native American lineage. However, this conclusion was based on such a limited number of samples that it was not possible to address important issues, such as the origin and diversity of the Native American haplogroup X mtDNAs, as well as the relationship between Old World and Native American haplogroup X mtDNAs.

To investigate these issues, we performed high-resolution RFLP and complete CR sequence analysis for available Native American and Old World mtDNAs and found 22 Native American and 14 European mtDNAs that belonged to haplogroup X. Despite a shared consensus RFLP haplotype, substantial genetic differences exist between the Native American and European mtDNAs. Phylogenetic analysis indicates that the two groups are related—but only distantly—to each other and that considerable genetic substructure exists within both groups. Further, coalescence-age estimates for haplogroup X in the Americas, based on either RFLP or CR sequence data, clearly indicate the antiquity of this haplogroup in the New World. Overall, these data exclude the possibility that the occurrence of haplogroup X in Native Americans is due to recent European admixture and, instead, provide a rigorous demonstration that this haplogroup represents an additional founding mtDNA lineage in Native Americans.

Subjects and Methods

Population Samples

A total of 36 (22 Native American and 14 European) individuals were available in our collection, as putative haplogroup X samples, and were used for high-resolution RFLP and CR sequence analysis. The 22 Native Americans (designated "NA1"–"NA22") consisted of 7 northern Ojibwa (NA1, NA2, NA7–NA10, and NA20) from the northwestern region of Ontario, 2 southwestern Ojibwa samples (NA3 and NA21) from Wisconsin, 5 southeastern Ojibwa (NA4–NA6, NA11, and NA22) from Manitoulin Island in Lake Huron, 2 Navajo (NA14 and NA15) from New Mexico, 2 Nuu-Chah-Nulth (NA12 and NA13) from Vancouver Island, British Columbia (Torroni et al. 1993*a*; Scozzari et al. 1997), and 4 Navajo (NA16–NA19) collected from New Mexico by Dr. Rene Herrera. The 14 Caucasian-European haplogroup X samples (designated "CE1"–"CE14") included 2 Caucasians of European ancestry (CE1 and CE4) from the United States and 1 French Canadian (CE5), (Torroni et al. 1994*a*), 1 Finn (CE8) (Torroni et al. 1996), 5 Israeli Druze (CE2 and CE11–CE14), and 5 Italians (CE3, CE6, CE7, CE9, and CE10).

mtDNA Analysis

High-resolution RFLP analysis was performed as described elsewhere (Torroni et al. 1996). *Acc*I was also used to detect the haplogroup X–associated *Acc*I site gain at np 14465. RFLP analysis had previously been performed for samples NA1 and NA2 (haplotype AM76), NA7–NA9 (haplotype AM75), NA10 (haplotype AM74), and NA14 and NA15 (haplotype AM29) (Torroni et al. 1993*a*), CE1, CE4, and CE5 (haplotypes 71, 7, and 101, respectively) (Torroni et al. 1994*a*). Partial RFLP analysis was performed on eight mtDNAs (NA3, NA12, NA13, NA16–NA19, and NA21), because of limitations of sample material, and included assays for the haplogroup X RFLP markers (−1715 *Dde*I, −10394 *Dde*I, +14465 *Acc*I, and +16517 *Hae*III), as well as for the restriction-site changes (−1413 *Taq*I, −8150 *Msp*I, −10254 *Mbo*I, and +13367 *Bam*HI) previously found in the high-resolution screening of other Native American haplogroup X samples (see table 1). For the Nuu-Chah-Nulth samples (NA12 and NA13), it has been shown elsewhere that these non–haplogroup A–D mtDNAs are −10394 *Dde*I and +16517 *Hae*III (Torroni et al. 1993*a*). These mtDNAs were further tested for the 1715 *Dde*I and 14465 *Acc*I sites, but a complete haplotype assignment was not possible.

The entire CR was sequenced for all 36 putative haplogroup X samples. The target sequence encompassed 1,349 nucleotides (np 15880–660), which included both CR hypervariable segments—HVS-I and HVS-II—along with some flanking coding region. Sequence analysis was performed by means of standard cycle sequencing with fluorescent dideoxy nucleotides. Both strands of the mtDNA were sequenced, and ambiguities were resolved by repeated sequencing reactions using different deoxyoligonucleotide primers.

Phylogenetic networks, which retain ambiguities in the branching structure by incorporation of reticulations, were constructed by means of the median algorithm of Bandelt et al. (1995), with two different parameter settings. The implicit parameter governing the decision about prediction of recurrent mutations is hidden in equation (5) of that paper: it equals 2 by default, but, instead, we used 3, to capture all most parsimonious (MP) trees here. The elucidation of all MP trees could be done by hand in this case, by means of the mathematical techniques of Foulds et al. (1979), Hendy and Penny (1982), and Bandelt et al. (1995).

Coalescence times for haplogroup X in North America were calculated from HVS-I sequences (with either one

Table 1

RFLP and CR Sequence Variation of Haplogroup X mtDNAs

Sample (Population)	RFLP Haplotype[a]	CR Sequence[b]
	111111111	1111111111111111111
	11223678888003334566	56666666666666666666
	47586351355230374925	90111111112222222335011112222222233334555
	11332875979593606221	29244688891122557251734590002222600128022
	35714300122441645567	73658323933737458079313350475678399549212
	lckgeejieejjcgmtsiae	a aba
Cambridge Reference Sequence	++---+-+++++-+-+-+--	GTTGCAAAT-GTCAAGCCTTATGATATGGTAGA---CTCAC
NA1 (Northern Ojibwa)	.-.............+.++	.C.....CCC..TG..T.CCG.AGCGC....AGC.C.....
NA2 (Northern Ojibwa)	.-.............+.++	.C.....CCC..TG..T.CCG.A.CGC....AGC.C.....
NA3 (Southwestern Ojibwa)	.-.............+..+	.C.....CCCA.T...T..CG.AGCG......GCCC.....
NA4 (Southeastern Ojibwa)	.-.............+..+	.C.....CCCA.T...T..CG.AGCG......GC.C.....
NA5 (Southeastern Ojibwa)	.-.............+..+	.C.....CCCA.T...T..CG.AGCG......GCCC.....
NA6 (Southeastern Ojibwa)	.-.............+..+	.C.....CCCA.T...T..CG.AGCG......GC.C.....
NA7 (Northern Ojibwa)	.-.............+..+	.C.....CCCA.T...T..CG.AGCG......GC.C...--
NA8 (Northern Ojibwa)	.-.............+..+	.C.....CCCA.T...T..CG.AGCG......G..C...--
NA9 (Northern Ojibwa)	.-.............+..+	.C.....CCCA.T...T..CG.AGCG......GCCCG....
NA10 (Northern Ojibwa)	.-.....-........+..+	.C.....CCCA.T...T..CG.AGCG......GC.C.....
NA11 (Southeastern Ojibwa)	.-.....-........+..+	.C.....CCCA.T...T..CG..GCG......GCCC.....
NA12 (Nuu−Chah−Nulth)	u-uuuuuuuuuuuuuu+uu+	......-CCCA.T...T..CG..GCG..A...GCCC.....
NA13 (Nuu−Chah−Nulth)	u-uuuuuuuuuuuuuu+uu+	......-CCCA.T...T..CG..GCG..A...GCCC.....
NA14 (Navajo)	--.........-....+..+	...A...CCCA.T.C.T..CG..GCG..A...GC.CG....
NA15 (Navajo)	--.........-....+..+	.......CCCA.T.C.T..CG..GCG..A...GC.C.....
NA16 (Navajo)	--.........-....+..+	.......CCCA.T.C.T..CG..GCG..A...GCCC.....
NA17 (Navajo)	--.........-....+..+	.......CCCA.T.C.T..CG..GCG..A...GC.C.....
NA18 (Navajo)	--.........-....+..+	.......CCCA.T.C.T..CG..GCG..A...GC.C.....
NA19 (Navajo)	--.........-....+..+	.......CCCA.T.C.T..CG..GCG..A...GC.C.....
NA20 (Northern Ojibwa)	--.........-....+..+	.......CCCA.T.C.T..CG..GCG..A...GC.CG....
NA21 (Southwestern Ojibwa)	--.........-....+..+	.......CCCA.T.C.T..CG..GCG..A...GC.C.....
NA22 (Southeastern Ojibwa)	.-...........+.+..+	........CC..T...T..CG...C...A...GC.C.....
CE1 (European)	.-......-......-+-.+	A......CCC..T...T..CG..GC...AC..GCCC.....
CE2 (Druze)	.-...-..-......-+-.+	A.C....CCC..T...T..CG..GC...AC..GC.C.....
CE3 (Italian)	.-...........-+-.+	A.......C...T...T..CG..GC...AC..GC.C..T..
CE4 (European)	.-....+.-......-+-.+	A...T..CCC..T...TT.CG..GC...AC..GC.C.....
CE5 (French Canadian)	.-..+...-.-....-+-.+	A.......CC..T...T..CG...C...AC..GC.C.....
CE6 (Italian)	.-+...........+..+	.......CCC..T...T..CGC.GC.......GC.C.G...
CE7 (Italian)	.-..........-..+..+	........C...T...T..CG..GC...A...GC.C.....
CE8 (Finnish)	.-.+..........+..+	......CCCC..T..AT..CG..GC...A.G.G..C.....
CE9 (Italian)	.-.......-.....-+..+	.....-.CCC..T...T..CG...C.CA....GC.C.....
CE10 (Italian)	.-.............+..+	........C..CT...T..CG.A.C...AC..GCCC.....
CE11 (Druze)	.-.............+..+	........C.......T..CG..GC...A...G..C.....
CE12 (Druze)	.-.............+..+	........C.......T..CG..GC...A...G..C.....
CE13 (Druze)	.-.............+..+	........C.......T..CG..GC...A...G..C.....
CE14 (Druze)	.-.........+...+..+	........C.......T..CG..GC...A...G..C.....

[a] Restriction-endonuclease sites are indicated as follows: a = *Alu*I, b = *Ava*II, c = *Dde*I, e = *Hae*III, g = *Hinf*I, i = *Msp*I, j = *Mbo*I, k = *Rsa*I, l = *Taq*I, m = *Bam*HI, s = *Acc*I, t = *Bst*OI. A dot (.) denotes identity with the Cambridge reference sequence (CRS; Anderson et al. 1981); a plus sign (+), which denotes a site gain, or a minus sign (−), which denotes a site loss, indicates deviation from the CRS; "u" denotes that status is unknown or that no data were available. We were able to perform only partial RFLP haplotype analysis on samples NA3, NA16–NA19, and NA21; for these samples, we tested for all restriction sites that were found to be variable in the Native American haplogroup X high-resolution RFLP screening, including 1413 *Taq*I, 1715 *Dde*I, 8150 *Msp*I, 10254 *Mbo*I, 10394 *Dde*I, 13367 *Mbo*I, and 14465 *Acc*I. For NA12 and NA13, we were able to test only for the haplogroup X–associated markers 1715 *Dde*I and 14465 *Acc*I. Torroni et al. (1993*a*) had previously shown that NA12 and NA13 were −10394 *Dde*I and +16517 *Hae*III. The +13366m site is linked with −13367b and +13367j site changes, and the −8391e site is linked to the +8391b site change.

[b] The entire CR (np 15880–660) was sequenced. A dot (.) indicates identity with the CRS, and a dash (–) indicates nucleotide insertion/deletion. Note that the -15925i site loss is due to the 15927G mutation, the +16226a site gain is due to the 16227G mutation, and the +16517e site gain is due to the 16519C mutation.

or two founder sequences being assumed) and also were estimated from RFLP haplotypes. For HVS-I sequence data, coalescence ages were estimated by means of $\rho_{\text{HVS-I}}$, the mean transitional distance from inferred founder root sequence(s), in which one transition within the np 16090–16365 sequence was estimated to correspond to 20,180 years (Forster et al. 1996, Watson et al. 1997). To estimate the sampling error for ρ, we assumed a Poisson distribution for the observed starlike network, with SD calculated as $\sqrt{\rho/n}$, where n denotes the sample size. Transversions, insertions/deletions, and variation in the number of C's in the unstable region np 16182–16193 were not included in the calculations of $\rho_{\text{HVS-I}}$ values. For RFLP data, we used the estimated mean distance, ρ_{RFLP}, from the inferred root sequence, with one restriction-site change being estimated to correspond to 24,420 years (for ρ_{RFLP} calibration, see Torroni et al. 1998). Major insertion/deletions, as well as variation at the hypervariable +16517 *Hae*III site, were not considered in this analysis. This calibration, which was based on a direct comparison of Siberian and Native American haplogroups A–D RFLP data and HVS-I data, would correspond to an mtDNA evolutionary rate of ~2.7%/million years (Myr), which fits with the 2.2%–2.9%/Myr estimate (Torroni et al. 1994*c*) used for divergence-based time estimates. The average nucleotide sequence divergence was calculated from complete RFLP haplotypes, by means of the maximum-likelihood procedure of Nei and Tajima (1983).

Results

RFLP and sequencing results confirm that all 22 Native American and 14 European mtDNAs analyzed belong to haplogroup X (table 1). Figure 1 illustrates the reduced median network encompassing both the RFLP and the CR variation observed in the Native American and European haplogroup X mtDNAs. For easy reference, we have highlighted clusters (subclusters I–V) of closely related mtDNAs from Native American populations. The network suggests that European and Native American haplogroup X mtDNAs are separated into two major branches. Nearly all Native American mtDNAs are encompassed by the branch harboring the 16213A and 200G variants, whereas all European mtDNAs are included within the branch that lacks these mutations (table 1 and fig. 1).

The reduced median network further suggests that np 153 mutated more than twice (possibly even five times). This mutation has indeed arisen multiple times in human mtDNA, being found in Caucasians (Stoneking et al. 1991; Piercy et al. 1993) and Asians (Stoneking et al. 1991; Lee et al. 1997). The question of whether either np 225, np 16213, and np 200 underwent recurrent mutation cannot be unequivocally decided. However, np 225 appears to be stable, since this variant has never been seen in populations from Asia (Stoneking et al. 1991; Redd et al. 1995; Lee et al. 1997), Central and South America (Ginther et al. 1993; Batista et al. 1995; Easton et al. 1996), or Africa (Stoneking et al. 1991; Graven et al. 1995; Y.-S. Chen and D. C. Wallace, unpublished data). It has been detected in 4/167 European mtDNAs (Stoneking et al. 1991; Piercy et al. 1993; Torroni et al. 1996), each time on a haplogroup X background including the 16223T, 16278T, 153G, and 195C mutations, which renders 225A a marker for a major part of haplogroup X.

The consensus motif of our total sample of 36 haplogroup X mtDNAs is −1715 *Dde*I, +14465 *Acc*I, +16517 *Hae*III/16519C, 16189C, 16223T, 16278T, 73G, 153G, 195C, 225A, and 263G, relative to the human reference sequence (Anderson et al. 1981). However, it is not clear whether the root haplotype for haplogroup X would also have 153G, and/or 225A, since these polymorphisms are found in the majority, but not all, of the haplogroup X mtDNAs. Overall, the sequence data and phylogenetic analysis suggest that the Native American and the European haplogroup X mtDNAs share a common maternal ancestor but also suggest that they diverged from each other long ago.

Previously, both Ward et al. (1991) and Shields et al. (1993) reported CR sequences for seven Nuu-Chah-Nulth and two Yakima that contained the 16223T and 16278T mutations and did not belong to haplogroups A–D. Hence, these mtDNAs likely belong to haplogroup X. Because these data encompass only HVS-I, these samples could not be included within figure 1. Five of these Nuu-Chah-Nulth and both Yakima mtDNAs also harbored 16213A, thus indicating that they probably belong to the 16213A/200G branch. However, none of these harbored either the 16093C variant of subclusters III and IV or the 16254C variant of subcluster V, indicating that these are subcluster II mtDNAs, which also encompasses our two Nuu-Chah-Nulth samples. The remaining two Nuu-Chah-Nulth reported by Ward et al. (1991) did not harbor either the 16213A variant or any of the other HVS-I variants observed in the Native American 16213A/200G branch. Thus, they appear to be very similar to our Ojibwa sample NA22 (subcluster I). One of these mtDNAs most closely represents the consensus haplogroup X motif (indicated by an asterisk in fig. 1), whereas the other differs from the first only by the presence of the variant 16362C.

The coalescence time for haplogroup X mtDNAs in the New World was calculated by means of either RFLP or CR sequence data. RFLP data suggest only one possible founder haplotype, defined by −1715 *Dde*I, +14465 *Acc*I, and +16517 *Hae*III, which is shared by Native Americans and Europeans (table 1) and is central

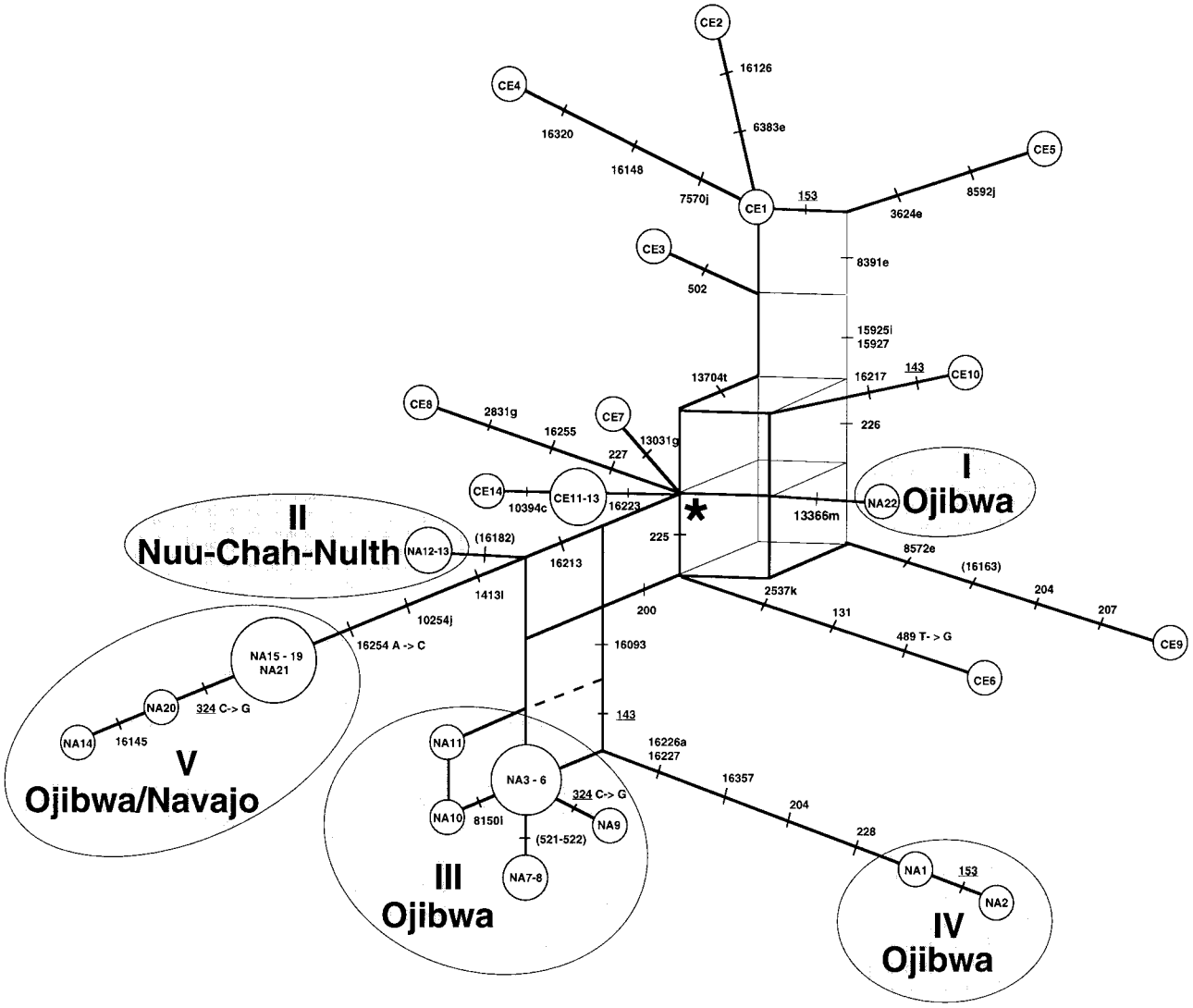

Figure 1 Reduced median network of 22 Native American (NA) and 14 Caucasian (CE) haplogroup X mtDNAs as defined by both RFLP and CR variation. This network represents all of the MP trees. The boldface lines indicate the more plausible MP trees, and the dashed link is not found in any MP tree. mtDNA types are represented by circles, with areas proportional to number of individuals. Lines are labeled by restriction site changes and CR mutations, with restriction site changes indicated by their enzyme letter code listed in table 1. CR nucleotide variants correspond to transitions, while transversions are further specified, and insertions/deletions are indicated, in terms of np, within parentheses. Variations of C's in the unstable regions between np 16182–16193 and np 309–315 were not included. The node marked with a large asterisk matches the haplogroup X basic motif: −1715c, +14465s, 16189C, 16223T, 16278T, +16517e/16519C, 153A, 195C, and 225A. The order of mutations on a path not interrupted by branches or nodes is arbitrary. Underlining indicates nucleotide positions which have mutated more than once, and reticulations indicate ambiguity in the topology. Parallel lines in a reticulation represent the same mutation. Ellipses indicate Native American haplogroup X subdivided into clusters (I–V) of closely related individuals.

in the phylogeny of Native American haplogroup X mtDNAs. Using RFLP data, we calculated the mean distance from the putative founder haplotype (ρ), for all Native American mtDNAs, using both our high-resolution RFLP haplotype data set ($n = 14$ samples) and our total RFLP data set ($n = 22$ samples) (table 2). Using the calibration of $\rho = 1$ as corresponding to 24,420 years, for RFLP data (Torroni et al. 1998), we estimate that the age of haplogroup X is ⩾23,000 years. For CR HVS-I data, the presence, in both Europe and the Americas, of the basic HVS-I motif of 16189C, 16223T, and 16278T (subcluster I in fig. 1) suggests that this is the founder motif that originated in the Old World and moved to the Americas. When this motif was used as the root sequence, the coalescence time was estimated to be 31,000–36,000 years ago (table 2).

Table 2

Coalescence-Time Estimates for Haplogroup X in the Americas

	$\rho \pm \sqrt{\rho/n}$ (Coalescence Time[a]), by	
n	RFLP	HVS-I
14[b]	.93 ± .26 (23,000 ± 6,000 years ago)	1.79 ± .36 (36,000 ± 7,000 years ago)
22[c]	⩾.95 ± .21 (⩾23,000 ± 5,000 years ago)	1.59 ± .27 (32,000 ± 5,000 years ago)
31[d]	…	1.52 ± .22 (31,000 ± 4,000 years ago)

[a] Calculated by calibrating $\rho = 1$ to either 24,420 years, for RFLP data, or 20,180 years, for HVS-I data (see Subjects and Methods). Average nucleotide sequence–diversity calculation, when complete RFLP data and an mtDNA sequence evolution rate of 2.2%–2.9%/Myr are used, gives a coalescence time of 13,000–17,000 years ago (see Discussion).

[b] Includes all completely RFLP-haplotyped Native American haplogroup X mtDNAs.

[c] Includes all Native American haplogroup X mtDNAs (8 partial and 14 complete RFLP haplotypes).

[d] Includes all Native American haplogroup X mtDNAs (8 partial and 14 complete RFLP haplotypes), as well as 7 Nuu-Chah-Nulth (Ward et al. 1991) and 2 Yakima (Shields et al. 1993).

Discussion

Our analysis confirmed that haplogroup X is present in both modern Native American and European populations. For the Native Americans, this haplogroup encompasses ~25% of the Ojibwa, 15% of the Sioux, 11%–13% of the Nuu-Chah-Nulth, 7% of the Navajo, and 5% of the Yakima (table 3). Thus, with the exception (see below) of the Na-Dene–speaking Navajo, the distribution of this haplogroup among the Native Americans appears to be restricted to northern Amerindian populations.

In studies of Native American mtDNA diversity, the co-occurrence of the same haplogroup at significant frequencies in both the modern Native American and European populations is unique. Recent European genetic admixture cannot explain the presence of haplogroup X in the Amerindians. First, if the occurrence of haplogroup X were the result of female gene flow from Europeans, then other, more common European mtDNA haplogroups should also be present in the northern Native Americans, and they are not. Second, the Native American and European mtDNAs are very different and are connected only through an ancient common ancestor. Hence, Native American and European haplogroup X mtDNAs diverged long ago. Finally, Native American haplogroup X mtDNAs encompass substantial continent-specific diversity, implying an ancient arrival in America. Thus, haplogroup X represents a fifth founding mtDNA haplogroup for the Native Americans, supporting the conclusions of Bailliet et al. (1994), Forster et al. (1996), and Scozzari et al. (1997).

An ancient arrival of haplogroup X in the Americas could be corroborated by the presence of haplogroup X in pre-Columbian human remains. Two studies on mtDNA variation in pre-Columbian samples have reported partial CR sequences that include the 16223T-and-16278T motif (Hauswirth et al. 1994; Ribeiro-Dos-Santos et al. 1996). However, in the absence of either more-complete CR sequence or RFLP data, it is not possible to definitively assign these pre-Columbian mtDNAs to haplogroup X. In a third study, Stone and Stoneking (1993, 1998) analyzed mtDNA variation in skeletal remains of 52 individuals from an Amerindian (Oneota culture) population, from the Illinois River valley, which dates to 1300 A.D. Two of these mtDNAs were not subsumed within Native American haplogroups A–D, and partial CR sequence (essentially HVS-I) analysis indicated that these samples contained not only the 16223T and 16278T mutations but also the 16093C, 16189C, 16227G, and 16357C mutations (Stone and Stoneking 1998). These sequences are essentially identical to those of Ojibwa mtDNAs NA1 and NA2 (table 1 and subcluster IV in fig. 1), demonstrating the presence of haplogroup X in the Americas prior to European introgression.

Since haplogroup X appears to be a pre-Columbian, founding Native American mtDNA lineage, the question remains: Where did this haplogroup originate? Thus far, haplogroup X has not been detected in numerous Asian/Siberian populations analyzed by high-resolution restriction analysis or CR sequencing (table 3). Three (2.9%) of 103 Mongolians harbored both the 16223T and 16278T mutations, but they also contained the +10394 *Dde*I and +10397 *Alu*I markers (Kolman et al. 1996), placing them in Asian superhaplogroup M (Ballinger et al. 1992) and excluding them from haplogroup X. Among Koreans, analysis of CR HVS-I and II sequences showed that 13 (4.3%) of 306 contained the 16223T and 16278T mutations (Lee et al. 1997), but none harbored the consensus haplogroup X motif, suggesting that they are also members of superhaplogroup M. Finally, partial CR sequences of a large group of eastern Asians (Han from Taiwan, Japanese, Ainu, Ryukyans, and Ko-

Table 3

Frequency of Haplogroup X mtDNAs in Native American and Asian Populations

Population (*n*)	Frequency of Haplogroup X (%)	Reference(s)
Native Americans		
Na-Dene		
Haida (41)	…	Ward et al. (1993)
Dogrib (30)	…	Torroni et al. (1993*a*)
Apache (25)	…	Torroni et al. (1993*a*)
Navajo (92)	6.5	Torroni et al. (1993*a*), present study
Amerindians:		
North America:		
Northern Ojibwa (28)	25.0	Torroni et al. (1993*a*), present study
Southern Ojibwa (35)	25.7	Scozzari et al. (1997), present study
Sioux (41)	14.6	Bianchi and Bailliet (1997)
Nuu-Chah-Nulth (15)	13.3	Torroni et al. (1993*a*), present study
Nuu-Chah-Nulth (63)	11.1[a]	Ward et al. (1991)
Yakima (42)	4.8[a]	Shields et al. (1993)
Bella Coola (40)	…	Ward et al. (1993)
Pima (30)	…	Torroni et al. (1993*a*)
Seminole (37)	…	Huoponen et al. (1997)
Central America (243)	…	Torroni et al. (1993*a*, 1994*a*, 1994*c*), Kolman et al. (1995)
South America (451)	…	Ginther et al. (1993), Horai et al. (1993), Torroni et al. (1993*a*), Easton et al. (1996), Ribeiro-Dos-Santos et al. (1996)
American Eskimos (22)	…	Shields et al. (1993)
Asians		
Siberian:		
Eskimos (157)	…	Ivanova (1993), Shields et al. (1993), Torroni et al. (1993*b*), Starikovskaya et al. (1998)
Chukchi (66)	…	Starikovskaya et al. (1998)
Koryak (107)	…	T. Schurr and D. C. Wallace (unpublished data)
Nivkhs (57)	…	Torroni et al. (1993*b*)
Evenks (51)	…	Torroni et al. (1993*b*)
Udegeys (45)	…	Torroni et al. (1993*b*)
Tibetans (54)	…	Torroni et al. (1994*b*)
Mongolians (103)	…	Kolman et al. (1996)
Koreans (319)	…	Ballinger et al. (1992), Lee et al. (1997)
Malaysian Chinese (14)	…	Ballinger et al. (1992)
Taiwanese Han (20)	…	Ballinger et al. (1992)
Vietnamese (28)	…	Ballinger et al. (1992)
Malays (14)	…	Ballinger et al. (1992)
Malay Aborigines (32)	…	Ballinger et al. (1992)

[a] Putative and obtained from literature report—only partial CR sequence was available. These samples have the 16223T and 16278T CR mutations and do not belong to Native American mtDNA haplogroups A–D; thus it is probable that they are members of haplogroup X.

reans) revealed that 14 (4.8%) of 293 had both the 16223T and 16278T mutations, but none of these contained the 16093C, 16213A, or 16254C mutations (Horai et al. 1996). However, since only part of the CR was analyzed and no RFLP data were obtained, it is not possible to definitively classify these mtDNAs as members of either haplogroup X or superhaplogroup M. Thus, to date, haplogroup X mtDNAs have not been unambiguously identified in Asians.

The apparent absence of haplogroup X in northern Asia parallels, in part, the situation already seen for Native American mtDNA haplogroup B. However, unlike haplogroup X, haplogroup B is present in Tibetans, Koreans, Japanese, and Mongolians. Moreover, haplogroup B is widely distributed among North, Central, and South American Amerindian populations, whereas haplogroup X is restricted to some of the most-northern Amerindian populations. Its presence in the Navajo but not in other Na-Dene populations suggests that, in a manner similar to that characterizing some nuclear-gene markers (Schell and Blumberg 1988), the Navajo have acquired haplogroup X through admixture with northern Amerindian populations. This could have occurred during or after the recent migration (1,000 years ago)

that brought the ancestors of the Navajo from the Athapaskan homeland (Alaska and western Canada) to the southwestern United States (Haskell 1987). An Amerindian origin of the Navajo haplogroup X mtDNAs is also supported by the fact that the Navajo sequences are very similar, if not identical, to those observed in some Ojibwa (fig. 1). Also, the homogeneity of the Navajo sequences (table 1 and fig. 1) suggests that the Navajo acquired haplogroup X very recently.

The time of entry of haplogroup X into the Americas, calculated from both RFLP and CR HVS-I sequence data and on the assumption that there is a single founder root for Native American mtDNAs, yielded a coalescence age, in the New World, of 23,000–36,000 years ago. However, it is possible that there was more than one founder motif. A possible additional founder mtDNA is represented by the HVS-I motif 16189C, 16213A, 16223T, and 16278T, which is supported by the variant 200A in HVS-II. This motif represents 86% of our Native American haplogroup X samples and is found in all of the tribes studied. If two haplogroup X founders are proposed for our 22 Native Americans, then the $\rho_{\text{HVS-I}}$ value would be .818, corresponding to a coalescence time of 17,000 ± 4,000 years ago. If the seven Nuu-Chah-Nulth and two Yakima sequences from the literature are added to our data, a $\rho_{\text{HVS-I}}$ value of .613 is obtained, corresponding to a coalescence time of 12,000 ± 3,000 years ago. A similar coalescence time, 13,000–17,000 years ago, is obtained when the average nucleotide sequence diversity is calculated from complete RFLP data (π = .0377%) and a mtDNA sequence evolution rate of 2.2%–2.9%/Myr (Torroni et al. 1994*c*).

A coalescence time of 23,000–36,000 years ago would suggest that haplogroup X arrived in the Americas during the initial major Amerindian migration 20,000–30,000 years ago. A coalescence time of 12,000–17,000 years ago could be interpreted as a rapid reexpansion of haplogroup X mtDNAs near the time of the Na-Dene expansion, or, alternatively, as an independent and late arrival of haplogroup X mtDNAs into the Americas.

Given the apparent absence of haplogroup X in modern eastern and northern Asia, it is difficult to define a source population for haplogroup X in the Americas. The similarity between the western Asian/European and Native American haplogroup X mtDNAs appears to indicate a western Asia origin of this haplogroup. Indeed, on the basis of limited RFLP data, the coalescence time for haplogroup X in Caucasians is estimated to be 30,000–40,000 years ago (data not shown), compatible with both a Near Eastern origin of haplogroup X and its subsequent spread, probably at a low frequency, into Europe and Asia. If this is the case, then it is possible that this mtDNA was brought to Beringia/America by the eastward migration of an ancestral Caucasian population, of which no trace has so far been found in the mtDNA gene pool of modern Siberian/eastern Asian populations.

In conclusion, we have described the occurrence, variation within, and population distribution of haplogroup X mtDNAs in Native Americans. This haplogroup appears, on the basis of archaeological data, to be pre-Columbian and may have arrived in the Americas either 12,000–17,000 years ago or 23,000–36,000 years ago. Haplogroup X is remarkable in that it has not been found in Asians, including Siberians, suggesting that it may have come to the Americas via a Eurasian migration. However, a more extensive survey of Asian mtDNAs, as well as additional characterization of European and Native American haplogroup X mtDNAs, will be necessary to fully deduce the origin of haplogroup X in North America.

Acknowledgments

The authors would like to thank the following individuals for providing samples for this study: Drs. R. Herrera, K. Kidd, J. Kidd, B. Bonne-Tamir, D. Cole, D. Labuda, R. Fourney, C. Fregeau, G. Troup, D. Smith, M.-L. Savontaus, and R. Sukernik. This work was supported by National Institutes of Health (NIH) awards NS21328, HL45572, AG10130, and AG13154 (all to D.C.W.); NIH Clinical Research Center grant RR00039 (to Emory University School of Medicine); P.F. Beni Culturali contract 96.01182.PF36 (to R.S.); and Italian Consiglio Nazionale delle Ricerche grant 97.04297.CT04 and Ministero dell Università e Ricerca Scientifica e Tecnologica (40%) (to A.T.).

References

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, et al (1981) Sequence and organization of the human mitochondrial genome. Nature 290: 457–465

Bailliet G, Rothhammer F, Carnese FR, Bravi CM, Bianchi NO (1994) Founder mitochondrial haplotypes in Amerindian populations. Am J Hum Genet 55:27–33

Ballinger SW, Schurr TG, Torroni A, Gan YY, Hodge JA, Hassan K, Chen K-H, et al (1992) Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient Mongoloid migrations. Genetics 130:139–152

Bandelt H-J, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations using median networks. Genetics 141:743–753

Batista O, Kolman CJ, Bermingham E (1995) Mitochondrial DNA diversity in the Kuna Amerinds of Panama. Hum Mol Genet 4:921–929

Bianchi NO, Bailliet G (1997) Further comments on the characterization of founder Amerindian mitochondrial haplotypes. Am J Hum Genet 61:244–246

Easton RD, Merriwether DA, Crews DE, Ferrell RE (1996)

mtDNA variation in the Yanomami: evidence for additional New World founding lineages. Am J Hum Genet 59: 213–225

Forster P, Harding R, Torroni A, Bandelt H-J (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. Am J Hum Genet 59:935–945

Foulds LR, Hendy MD, Penny D (1979) A graph theoretic approach to the development of minimal phylogenetic trees. J Mol Evol 13:127–149

Ginther C, Corach D, Penacino GA, Rey JA, Carnese FR, Hutz MH, Anderson A, et al (1993) Genetic variation among the Mapuche Indians from the Patagonian region of Argentina: mitochondrial DNA sequence variation and allele frequencies of several nuclear genes. In: Pena SDJ, Chakraborty R, Epplen JT, Jeffreys AJ (eds) DNA fingerprinting: state of the science. Birkhauser, Boston, pp 211–219

Graven L, Passarino G, Semino O, Boursot P, Santichiarra-Benerecetti AS, Langaney A, Excoffier L (1995) Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka population. Mol Biol Evol 12: 334–345

Haskell JL (1987) Southern Athapaskan migration A.D. 200–1750. Navajo Community College Press, Tucson

Hauswirth WW, Dickel CD, Lawlor DA (1994) DNA analysis of the Windover population. In: Herrmann B, Hummel S (eds). Ancient DNA. Springer, New York, pp 104–121

Hendy MD, Penny D (1982) Branch and bound algorithms to determine minimal evolutionary trees. Math Biosci 59: 277–290

Horai S, Kondo R, Nakagawa-Hattori Y, Hayashi S, Sonoda S, Tajima K (1993) Peopling of the Americas founded by four major lineages of mitochondrial DNA. Mol Biol Evol 10:23–47

Horai S, Murayama K, Hayasaka K, Matsubayashi S, Hattori Y, Fucharoen G, Harihara S, et al (1996) mtDNA polymorphism in east Asian populations, with special reference to the peopling of Japan. Am J Hum Genet 59:579- 590

Huoponen K, Torroni A, Wickman PR, Sellitto D, Gurley DS, Scozzari R, Wallace DC (1997) Mitochondrial DNA and Y chromosome-specific polymorphisms in the Seminole tribe of Florida. Eur J Hum Genet 5:25–34

Ivanova AV (1993) Mitochondrial DNA polymorphism in native inhabitants of Chukotka. PhD thesis, University of Novosibirsk, Novosibirsk

Kolman CJ, Bermingham E, Cooke R, Ward RH, Arias TD, Guionneau-Sinclair F (1995) Reduced mtDNA diversity in the Ngobe Amerinds of Panama. Genetics 140:275–283

Kolman CJ, Sambuughin N, Bermingham E (1996) Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. Genetics 142:1321–1334

Lee SD, Shin CH, Kim KB, Lee YS, Lee JB (1997) Sequence variation of mitochondrial DNA control region in Koreans. Forensic Sci Int 87:99–116

Merriwether DA, Rothhammer F, Ferrell RE (1995) Distribution of the four- founding lineage haplotypes in Native Americans suggests a single wave of migration for the New World. Am J Phys Anthropol 98:411–430

Nei M, Tajima F (1983) Maximum likelihood estimation of the number of nucleotide substitutions from restriction site data. Genetics 105:207–217

Piercy R, Sullivan DM, Benson N, Gill P (1993) The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. Int J Legal Med 106:85–90

Redd AJ, Takezaki N, Sherry ST, McGarvey ST, Sofro ASM, Stoneking M (1995) Evolutionary history of the COII/tRNALys intergenic 9 base pair deletion in human mitochondrial DNA from the Pacific. Mol Biol Evol 12:604–615

Ribeiro-Dos-Santos AKC, Santos SEB, Machado AL, Guapindaia V, Zago MA (1996) Heterogeneity of mitochondrial DNA haplotypes in pre-Columbian natives of the Amazon region. Am J Phys Anthropol 101:29–37

Schell LM, Blumberg BS (1988) Alloalbuminemia and the migration of Native Americans. Yearbook Phys Anthropol 31: 1–14

Scozzari R, Cruciani F, Santolamazza P, Sellitto D, Cole DEC, Rubin LA, Labuda D, et al (1997) mtDNA and Y chromosome–specific polymorphisms in modern Ojibwa: implications about the origin of their gene pool. Am J Hum Genet 60:241–244

Shields GF, Schmiechen AM, Frazier BL, Redd A, Voevoda MI, Reed JK, Ward RH (1993) mtDNA sequences suggest a recent evolutionary divergence for Beringian and northern North American populations. Am J Hum Genet 53:549–562

Starikovskaya YB, Sukernik RI, Schurr TG, Kogelnik AM, Wallace DC (1998) Mitochondrial DNA diversity in Chukchi and Siberian Eskimos: implications for the genetic history of ancient Beringia and the peopling of the New World. Am J Hum Genet 63 (in press)

Stone AC, Stoneking M (1993) Ancient DNA from a Pre-Columbian Amerindian population. Am J Phys Anthropol 92:463–471

Stone AC, Stoneking M (1998) mtDNA analysis of a prehistoric Oneota population: implications for the peopling of the New World. Am J Hum Genet 62:1153–1170

Stoneking M, Hedgecock D, Higuchi RG, Vigilant L, Erlich HA (1991) Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. Am J Hum Genet 48:370–382

Torroni A, Bandelt H-J, D'Urbano L, Lahermo P, Moral P, Sellitto D, Rengo C, et al (1998) mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. Am J Hum Genet 62: 1137–1152

Torroni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, Obinu D, et al (1996) Classification of European mtDNAs from an analysis of three European populations. Genetics 144:1835–1850

Torroni A, Lott MT, Cabell MF, Chen Y-S, Lavergne L, Wallace DC (1994*a*) mtDNA and the origin of Caucasians: identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. Am J Hum Genet 55:760–776

Torroni A, Miller JA, Moore LG, Zamudio S, Zhuang J, Droma T, Wallace DC (1994*b*) Mitochondrial DNA analysis in Tibet: implications for the origin of the Tibetan population and its adaptation to high altitude. Am J Phys Anthropol 93:189–199

Torroni A, Neel JV, Barrantes R, Schurr TG, Wallace DC

(1994*c*) A mitochondrial DNA "clock" for the Amerinds and its implications for timing their entry into North America. Proc Natl Acad Sci USA 91:1158–1162

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

Torroni A, Schurr TG, Yang C-C, Szathmary EJE, Williams RC, Shanfield MS, Troup GA, et al (1992) Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. Genetics 130:153–162

Torroni A, Sukernik RI, Schurr TG, Starikovskaya YB, Cabell MF, Crawford MH, Comuzzie AG, et al (1993*b*) mtDNA variation in aboriginal Siberians reveals distinct genetic affinities with Native Americans. Am J Hum Genet 53: 591–608

Torroni A, Wallace DC (1994) Mitochondrial DNA variation in human populations and implications for detection of mitochondrial DNA mutations of pathological significance. J Bioenerg Biomembr 26:251–261

Ward RH, Frazier BL, Dew-Jager K, Paabo S (1991) Extensive mitochondrial diversity within a single Amerindian tribe. Proc Natl Acad Sci USA 88:8720–8724

Ward RH, Reed A, Valencia D, Frazier B, Paabo S (1993) Genetic and linguistic differentiation in the Americas. Proc Natl Acad Sci USA 90:10663–10667

Watson E, Forster P, Richards M, Bandelt H-J (1997) Mitochondrial footprints of human expansions in Africa. Am J Hum Genet 61:691–704