High-Resolution SNPs and Microsatellite Haplotypes Point to a Single, Recent Entry of Native American Y Chromosomes into the Americas

Stephen L. Zegura,* Tatiana M. Karafet,†§ Lev A. Zhivotovsky,‡ and Michael F. Hammer*†

*Department of Anthropology, University of Arizona; †Division of Biotechnology, University of Arizona; ‡Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia; §Institute of Cytology and Genetics, Novosibirsk, Russia

A total of 63 binary polymorphisms and 10 short tandem repeats (STRs) were genotyped on a sample of 2,344 Y chromosomes from 18 Native American, 28 Asian, and 5 European populations to investigate the origin(s) of Native American paternal lineages. All three of Greenberg's major linguistic divisions (including 342 Amerind speakers, 186 Na-Dene speakers, and 60 Aleut-Eskimo speakers) were represented in our sample of 588 Native Americans. Single-nucleotide polymorphism (SNP) analysis indicated that three major haplogroups, denoted as C, Q, and R, accounted for nearly 96% of Native American Y chromosomes. Haplogroups C and Q were deemed to represent early Native American founding Y chromosome lineages; however, most haplogroup R lineages present in Native Americans most likely came from recent admixture with Europeans. Although different phylogeographic and STR diversity patterns for the two major founding haplogroups previously led to the inference that they were carried from Asia to the Americas separately, the hypothesis of a single migration of a polymorphic founding population better fits our expanded database. Phylogenetic analyses of STR variation within haplogroups C and Q traced both lineages to a probable ancestral homeland in the vicinity of the Altai Mountains in Southwest Siberia. Divergence dates between the Altai plus North Asians versus the Native American population system ranged from 10,100 to 17,200 years for all lineages, precluding a very early entry into the Americas.

Introduction

The recent publication of highly congruent human Y chromosome trees (Underhill et al. 2000; Hammer et al. 2001) and a standardized nomenclatural system for the resulting binary polymorphism-based consensus tree (YCC 2002) has provided an opportunity to understand paternal population origins, relationships, and dispersals with more phylogenetic and geographic resolution and less terminological ambiguity than was hitherto possible. The inclusion of microsatellite data can facilitate the estimation of population divergence times, which can then be compared (and contrasted) with estimated mutational ages of the polymorphic markers, thereby providing a chronological framework for major population history events. Indeed, the synergistic combination of these two kinds of data also offers a powerful tool with which to assess patterns of migration, admixture, and ancestry, as well as to identify additional microevolutionary processes associated with population structure *sensu lato*.

The last major human continental colonization episode, the settlement of the Americas, is a topic of intense interest and controversy for researchers in numerous scientific fields, including genetics. In 1986 Greenberg, Turner, and Zegura published a widely cited, synthetic, position paper on the early peopling of the Americas that stressed the apparent congruence of the then available data from linguistics, dental morphology, and traditional biparental nuclear genetic systems within the context of the archaeological record. Their major explanatory hypothesis, the "three-wave" or "tripartite" model, was based on the proposition that all indigenous Native American populations could be allocated to three distinct linguistically defined groups (i.e., Amerind, Na-Dene, and

Aleut-Eskimo) that had their origins in three chronologically separate migrations from different geographic areas of Asia (Greenberg, Turner, and Zegura 1986). Although a large number of studies from diverse disciplines have subsequently explored the issues raised in their highly controversial paper, unsolved problems today still include the number and timing of early migration(s) to the Americas, the geographic location of the source populations(s), and the evolutionary processes that have interacted to sculpt the Native American gene pool (Zegura 2002). Our present contribution seeks to provide a more fine-grained paternal genetics perspective for the eventual resolution of these important questions than either our earlier attempts (Karafet et al. 1997, 1999) or those of other research groups (Lell et al. 1997, 2002; Santos et al. 1999; Underhill et al. 1996).

The first two human Y chromosome marker studies appeared in 1985 (Casanova et al. 1985; Lucotte and Ngo 1985). It was not until almost a decade later that Torroni and co-workers (1994*a*) published the first Y chromosome data on Native Americans. Numerous surveys of variation on the non-recombining portion of the Y chromosome (NRY) devoted primarily to Amerind speakers quickly followed (Pena et al. 1995; Santos et al. 1995, 1996; Underhill et al. 1996; Bianchi et al. 1997; Karafet et al. 1997; Lell et al. 1997). These early data were generally interpreted to support a single-origin (and one-wave) model for the members of Greenberg's (1987) three major New World linguistic groups, despite occasional sampling problems wherein one or more of these linguistic groups lacked representation.

Karafet et al. (1999) investigated four migration models for the early paternal peopling of the Americas and presented a visual portrayal of these various models along with their hypothesis as to the geographic source of Native American Y chromosomes, shown as a circle that included the following territory: Lake Baikal (eastward to the Trans-Baikal and southward into northern Mongolia), the Lena River headwaters, the Angara and Yenisey river basins, the
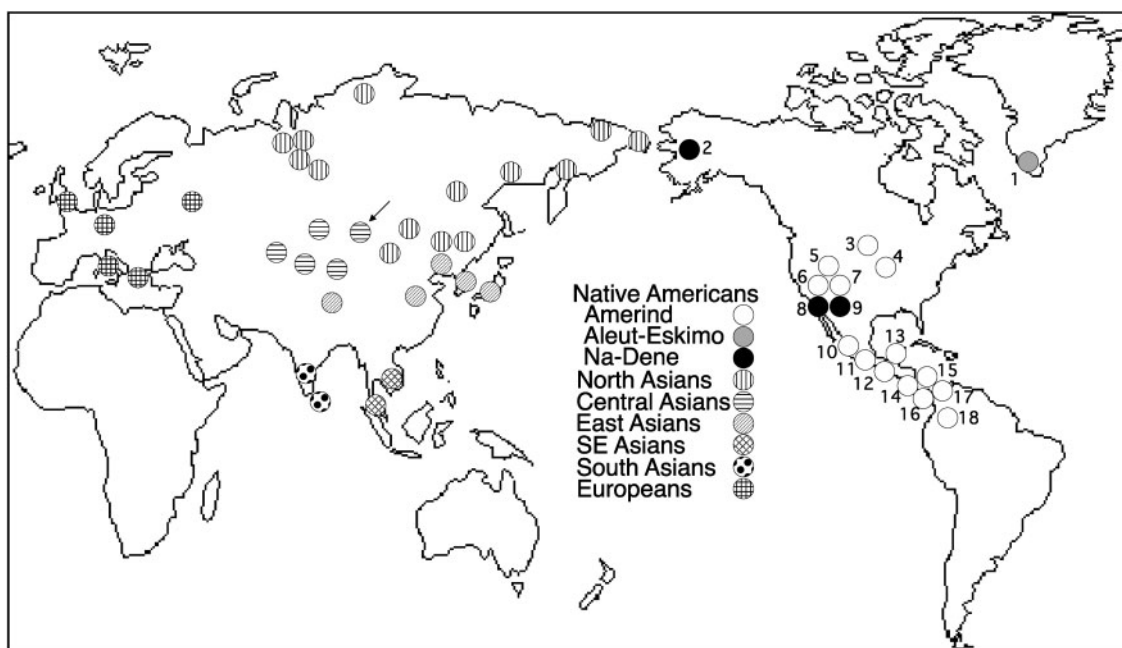
Fig. 1.—Map of 51 sampling localities. The 28 Asian and 5 European populations are subdivided into 5 geographic groupings. The 18 Native American populations are subdivided by language family (Greenberg 1987). See table 1 for Native American population names. Arrow denotes Altai population.

Altai Mountain foothills, and the region south of the Sayan Mountains (including Tuva and western Mongolia). Although both of their proposed major Y chromosome American founding lineages could be traced to possible ultimate dispersal sources within this circle, the authors favored a two-migration scenario, a proposal that has recently been supported by Lell et al. (2002) based on their Y chromosome data. Unfortunately, relatively secure dates for the two migrations (or for the single-migration scenario) based on Y chromosome microsatellites have not been published.

Thus, the major purposes of the present article are to (1) use a larger Y chromosome database that includes many more microsatellite and single-nucleotide polymorphism markers to refine our previous analyses (Karafet et al. 1997, 1999) of founder versus admixture-derived lineages in the Americas; (2) narrow down the most probable area of the postulated Asian geographic source of Native American Y chromosomes; (3) estimate the time of divergence between the Native American population system(s) and various Asian population systems; and (4) address the most likely number of migrations detected so far by Y chromosome data from Native Americans.

Subjects and Methods
Populations and DNA Samples

We analyzed 63 binary single-nucleotide polymorphisms (SNPs) and 10 short tandem repeats (STRs) on a sample of 2,344 Y chromosomes from 51 populations representing the Americas, Asia, and Europe. The Native American sample (fig. 1) included 588 individuals from 18 populations allocated to Greenberg's (1987) three major Native American language families as follows: 342

Amerind speakers, 186 Na-Dene speakers, and 60 Aleut-Eskimo speakers. Native American linguistic affiliations, sample sizes, SNP haplogroup frequencies and diversity measures, STR haplotype numbers and repeat number variances, and three-letter and numerical population codes are given in table 1, which also contains summary genetic data for six geographically defined Eurasian groupings (Europe, North Asia, Central Asia, South Asia, East Asia, and Southeast Asia). Many of the individuals analyzed here were included in our previous studies (Karafet et al. 1997, 1999, 2001, 2002; Hammer et al. 2001); however, our most recent New World publication (Karafet et al. 1999) contained data from only 12 biallelic polymorphisms and two STRs on a total of 380 Native Americans. New samples collected for this study came from the Apache, Navajo, Sioux, and Maya. All sampling protocols were approved by the Human Subjects Committee at the University of Arizona.

Genetic Markers

The polymorphic sites in our survey included a set of 62 previously published binary NRY markers (Karafet et al. 2001, 2002), most of which have not previously been used to type geographically and linguistically diverse New World populations, and one newly discovered Native American–specific marker (P-39). This new SNP, a G to A mutation at position 60,565 of the arylsulfatase D pseudogene (*ARSDP*), was genotyped by allele-specific polymerase chain reaction (PCR). The following primers were used to amplify the 466 base pair (bp) control band and the 130 bp mutant allele–specific band: P39U (5′ AGAAGGACTGCCTCAGAATGC-3′), P39L (5′-GTTC-GAAAGGGGATCCCTGG-3′, and P39A (5′CCCGGGA-

Table 1
SNP Lineage Frequencies and SNP and STR Diversity Measures for 18 Native American Populations and 6 Eurasian Regions

Population	Linguistic Affiliation[a]	SNP Lineages					SNP Genetic Diversity		STRs		
		n	C	Q	R	Other	Number of Haplogroups	$p \pm SE^b$	n	Number of Haplotypes	Repeat Number Variance
1. Inuit (INU)	Aleut-Eskimo	60	0	48	7	5	7	1.94 ± 1.12	60	28	0.52
2. Tanana (TAN)	Na-Dene	12	5	5	1	1	5	5.47 ± 2.83	11	11	0.77
3. Cheyenne (CHY)	Amerind	44	7	27	7	3	8	4.08 ± 2.08	39	34	0.65
4. Sioux (SIO)	Amerind	44	5	11	22	6	11	4.29 ± 2.17	44	31	0.63
5. Southwest (SOU)	Amerind	10	0	8	2	0	3	1.98 ± 1.22	8	4	0.31
6. Pima (PIM)	Amerind	24	0	21	3	0	3	1.28 ± 0.83	24	19	0.48
7. Pueblo (PUE)	Amerind	18	0	13	5	0	3	2.23 ± 1.29	16	14	0.44
8. Apache (APA)	Na-Dene	96	14	75	5	2	7	3.07 ± 1.61	96	43	0.72
9. Navajo (NAV)	Na-Dene	78	1	72	2	3	8	1.20 ± 0.78	75	42	0.52
10. Mixtec (MXA)	Amerind	28	0	26	2	0	2	0.69 ± 0.54	26	12	0.36
11. Zapotec (ZAP)	Amerind	16	0	12	1	3	5	0.41 ± 0.40	15	12	0.76
12. Mixe (MXE)	Amerind	12	0	12	0	0	1	0.00 ± 0.00	12	6	0.15
13. Maya (MAY)	Amerind	71	0	62	9	0	3	1.16 ± 0.76	68	11	0.27
14. Ngobe (NGO)	Amerind	17	0	17	0	0	2	0.31 ± 0.33	14	13	0.75
15. Kuna (KUN)	Amerind	9	0	9	0	0	1	0.00 ± 0.00	9	6	0.25
16. Waunaan (WAU)	Amerind	14	0	12	2	0	3	3.01 ± 1.66	14	11	0.26
17. Emberra (EMB)	Amerind	10	0	10	0	0	2	0.53 ± 0.48	9	7	0.35
18. Wayu (WAY)	Amerind	25	2	9	11	3	7	4.40 ± 2.25	19	13	0.62
All Native Americans		588	34	449	79	26	17	2.64 ± 1.41	559	266	0.61
Europe		237	0	1	114	122	18	4.78 ± 2.34	231	198	0.79
North Asia		669	187	120	40	322	27	4.79 ± 2.34	623	236	0.91
Central Asia		263	51	25	104	83	30	5.15 ± 2.50	244	168	1.00
South Asia		130	6	5	28	91	15	3.40 ± 1.75	129	102	0.94
East Asia		356	31	2	10	313	24	4.63 ± 2.28	334	270	1.03
Southeast Asia		101	4	4	2	91	21	3.59 ± 1.84	101	84	0.75
Totals		2,344	313	606	377	1,048			2,221		

[a] See Greenberg 1987.
[b] Calculated from data representing all SNP haplogroups.

GGTGGAGGTTATA-3′). The cycling conditions were 94°C for 3 min, followed by 20 touchdown cycles with −0.5°C/cycle increments at 94°C, 68°C, 58°C, 72°C for 30 s, then 15 cycles of standard amplification at 94°C, 58°C, 72°C for 30 s, with the final extension step at 72°C for 2 min. Reactions were run in a final volume of 15 μl containing 10 ng of genomic DNA, 0.2 mM each dNTP, 1 μM each primer, 0.046 μM of TaqStart Antibody (Clontech), 0.0016 μM of Taq DBA polymerase (Eppendorf), and 1.5 mM $MgCl_2$, 75 mM KCl, and 10 mM Tris-HCl (pH 8.3).

For the microsatellite analysis 10 STRs (*DYS19*, *DYS388*, *DYS389I*, *DYS389II*, *DYS390*, *DYS391*, *DYS392*, *DYS393*, *DYS426*, and *DYS439*) were typed in two multiplex PCR reactions. Primer sequences were published in Kayser et al. (1997) and Redd et al. (2002), and PCR conditions were given by Redd et al. (2002). The PCR product was electrophoresed on a 3100 Genetic Analyzer (Applied Biosystems) with a 36 cm array and filter set D. The data were analyzed with Genescan (v. 3.7, Applied Biosystems) and Genotyper (v. 1.1, Applied Biosystems). For all statistical analyses *DYS389*I was subtracted from *DYS389*II because the *DYS389*II PCR product also contains *DYS389*I.

Terminology

We follow the terminological conventions recommended by the Y Chromosome Consortium (YCC 2002) for naming NRY lineages. Capital letters A–R identify the 18 major Y chromosome clades or haplogroups. Lineages not defined on the basis of a derived character state represent interior nodes of the tree and are potentially paraphyletic. Thus, the term *paragroup* (rather than haplogroup) is used to describe these lineages and these paragroups are distinguished by the * (asterisk) symbol. For the sake of simplicity, we will refer to paragroups as *haplogroups* throughout the text. Lineages excluded from a haplogroup are listed in table 2 after an initial "x" symbol within parentheses, after the haplogroup name for the official lineage-based naming system. We opted to omit the "x" notation and parenthetical convention for the short-hand mutation-based names used throughout the text. When no farther downstream markers in the YCC 2002 tree were typed for this study, we considered the most derived typed marker to represent a haplogroup. Table 2 gives a complete list of the lineage-based and mutation-based names of the 42 haplogroups found in this study (Karafet et al. 2002). As suggested by de Knijff (2000), distinct Y chromosomes identified by STRs are designated "haplotypes."

Statistical Analyses

Population genetic structure indices (molecular variances and Φ statistics) and the mean number of pairwise differences among haplogroups (p) were estimated by ARLEQUIN 2.000 software (Schneider et al. 2000). The relationships among genetic, geographic, and linguistic structure were assessed by the Mantel test, also employ-

Table 2
Lineage-Based and Mutation-Based Names of the 42 Haplogroups in the 51 Populations

Lineage-Based Name	Mutation-Based Name[a]	Derived State at	Ancestral state at
C*	C-RPS4Y$_{711}$*	RPS4Y$_{711}$	M8, M38, M217
C1	C-M8	M8	
C3*(xC3b,C3c)	C-M217*	M217	P39, M86
C3b	C-P39	P39	
C3c	C-M86	M86	
D*(xD1)	D-M174*	M174	M15
D1	D-M15	M15	
E*(xE3)	E-SRY$_{4064}$*	SRY$_{4064}$	P2
E3*(xE3a)	E-P2*	P2	P1
E3a	E-P1	P1	
F*(xG, H1, I, J, K)	F-P14*	P14	M201, M52, P19, 12f2, M9
G*(xG2)	G-M201*	M201	P15
G2	G-P15	P15	
H1[b]	H-M52[b]	M52	
I*(xI1b)	I-P19*	P19	P37.2
I1b[b]	I-P37.2[b]	P37.2	
J*(xJ2)	J-12f2.1*	12f2.1	M172
J2*(xJ2e)	J-M172*	M172	M12
J2e	J-M12	M12	
K*(xL, M, N, O, P)	K-M9*	M9	LLY22g, M20, M4, M175, P27
L	L-M20	M20	
M	M-M4	M4, M5	
N*	N-LLY22g*	LLY22g	P43, Tat, M128
N1	N-M128	M128	
N2	N-P43	P43	
N3a	N-M178	M178	
O*	O-M175*	M175	M119, P31, M122
O1	O-M119	M119	
O2*	O-P31*	P31	M95, SRY$_{+465}$
O2a	O-M95	M95	
O2b*(xO2b1a)	O-SRY$_{+465}$*	SRY$_{+465}$	47z
O2b1a	O-47z	47z	
O3*(xO3c, O3e)	O-M122*	M122	LINE-1, M134
O3c	O-LINE-1	LINE-1	
O3e	O-M134	M134	
P*	P-P27*	P27	P36, M207
Q*(xQ3)	Q-P36*	P36	M3
Q3[b]	Q-M3[b]	M3	
R*(xR1)	R-M207*	M207	M173
R1*	R-M173*	M173	SRY$_{10831b}$, P25
R1a[b]	R-SRY$_{10831b}$[b]	SRY$_{10831b}$	
R1b[b]	R-P25[b]	P25	

[a] Abbreviated without parenthetical system.
[b] No farther downstream markers were typed, so we refer to this lineage as a haplogroup.

ing ARLEQUIN 2.000 software. Geographic distances were calculated between populations from latitude and longitude data for the sample sites. The matrix of pairwise linguistic distances among populations was constructed according to the method described by Excoffier, Harding, and Sokal (1991) and Poloni et al. (1997). Language classifications were adopted from Greenberg (1987). Populations related within a linguistic family were set to distances from 0 to 4. Distances of 5 or 6 were assigned to pairs of populations belonging to different language macrofamilies. We performed nonmetric multidimensional scaling (MDS) (Kruskal 1964) on Slatkin's linearized Φ_{ST} distances using the software package NTSYS (Rohlf 1998). Median-joining networks (Bandelt, Forster, and Rohl 1999) were constructed using the NETWORK 2.0c program. For network calculations, microsatellite loci were weighted according to their variances such that higher weights were assigned to the least variable loci. The reduced median output was used as input for the median-joining network. This procedure reduces the ability of the median-joining algorithm to produce large reticulations within the network (Hurles et al. 2002). Divergence times were estimated using the microsatellite-based procedures $(\delta\mu)^2$ and T_D devised by Goldstein et al. (1995) and Zhivotovsky (2001), respectively.

Results
NRY Haplogroup Distribution in Native Americans

Figure 2 presents an evolutionary tree for the 17 Native American haplogroups found in this study, subdivided by language family (Greenberg 1987), with frequencies corrected for unequal sample sizes in the three linguistic groupings. This tree reflects the newly standardized Y Chromosome Consortium hierarchical nomenclature system (YCC 2002). The 17 haplogroups present in the 18 Native Americans populations fall into 9 of the 18 major Y chromosome haplogroup divisions. Table 1

presents frequency data for haplogroups C, Q, and R that together account for 95.6% of the 588 Native American Y chromosomes. The following minor haplogroup data are most probably due to admixture (Karafet et al. 1999) and are not presented in table 1: haplogroups E (n = 7), F* (n = 2), G (n = 3), I (n = 10), and J (n = 3). A single haplogroup N-M178 individual is also omitted from table 1. Although European/African admixture is less likely to have occurred here, this was the most frequent haplogroup (22.7%) in our Siberian study (Karafet et al. 2002) and may represent a very rare additional founding haplogroup or Asian admixture.

Only three major haplogroups were present at frequencies greater than 5% in Native Americans (Q = 76.4%, R = 13.4%, and C = 5.8%). As will be discussed below, at least 76 of the 79 haplogroup R individuals are probably due to post-1492 European admixture. The most frequent single haplogroup (Q) was present in 76.4% of the samples as follows: Q-P36* = 23.8% and Q-M3 = 52.6%. The Q lineage occurred in all 18 Native American populations, whereas the much less numerous C lineage (5.8%) was restricted to the following 6 populations: Tanana (n = 5), Navajo (n = 1), Apache (n = 14), Cheyenne (n = 7), Sioux (n = 5), and Wayu (n = 2).

Y Chromosome Diversity

The number of haplogroups in the various Native American populations ranged from 1 in the monomorphic Mixe and Kuna to 11 in the Sioux. Although only 17 haplogroups were found in Native Americans, the much smaller Central Asian sample contained 30 different haplogroups (table 1). The mean number of pairwise differences among haplogroups (p) ranged from 0, again in the Mixe and Kuna, to 5.47 in the Tanana, with an overall Native American value of 2.64. The corresponding Asian p values ranged from 3.40 in South Asia to 5.15 in Central Asia. Both SNP diversity measures exhibit the same pattern: on average Native American Y chromosome diversity is much reduced when compared with Asian diversity. For the microsatellite data, the mean number of pairwise differences also indicates a moderate reduction in genetic diversity for the Native Americans (5.13) compared with Asian values ranging from 5.66 in Southeast Asia to 6.10 in Central Asia (data not shown).

The STR data reflect a general reduction in number of haplotypes and variance in repeat number for the Native American data compared with the Asian data (table 1). The variance in repeat number value for Native Americans was 0.61, whereas the Asian variances ranged from 0.75 in Southeast Asia to 1.03 in East Asia (table 1). Thus, the overall trend in the STR data is, once again, toward a reduction in genetic diversity/variation for the Native American data set.

Median-Joining Microsatellite Networks

Figure 3 displays a median-joining network (Bandelt, Forster, and Rohl 1999) for haplogroup Q-P36* in Asia and the Americas, noting the position of the Q-M3 lineage (see cross-hatch). The ancestral node leading to Q-M3 has
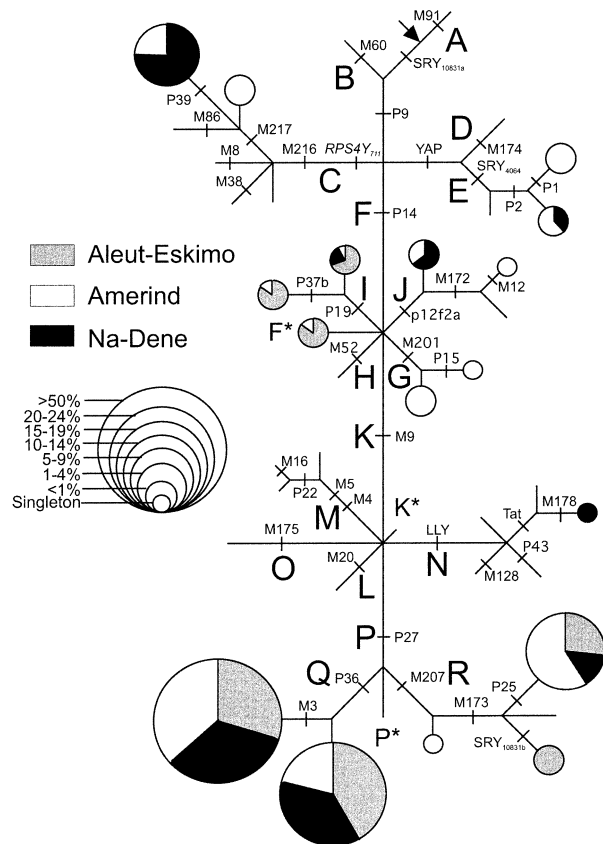

FIG. 2.—Evolutionary tree for the 18 major NRY haplogroups (denoted by capital letters [A–R] according to YCC [2002] recommendations). The root of the tree is indicated by an arrow; cross-hatches represent mutational events. Mutational events defining the 42 haplogroups found in the study are labeled by mutation name (table 2). Haplogroups are coded in white, black, and gray for Native Americans according to language family (see key; Greenberg 1987). The pie charts represent frequency of occurrence of a haplogroup within each of Greenberg's (1987) three Native American language families (weighted by language group sample size). The overall size of each pie chart corresponds to one of eight frequency classes (see insert) and represents the frequency of that haplogroup in the total sample of 588 Native American Y chromosomes.

haplotype (DYS19 = 13; DYS388 = 12; DYS389I = 13; DYS389II = 30; DYS390 = 23; DYS391 = 10; DYS392 = 14; DYS393 = 13; DYS426 = 12; and DYS439 = 12) and was present in 3 Altai, 1 Ket, and 1 Selkup. The vast majority of the close neighbors of this node were also confined to the Altai, Ket, and Selkup populations.

Figure 4 shows the C lineage network for Asia and the Americas, noting the position of the C-P39 mutation (see cross-hatch in right oval). All of the Native Americans are clustered on the left side of the diagram on the C-P39 branch (see left oval) except the two Wayu (denoted by an arrow). The ancestral node leading to C-P39 has haplotype (15–13–13–29–24–9–11–13–11–11) and was present in 2 Altai. This ancestral node is also connected to a one-step neighbor (DYS19 = 16) below it in the network that was found in 11 Altai. The first node after the C-P39 mutation differs from the ancestral node only at DYS390 (23 versus 24 repeats) and was found in a single Cheyenne individual. The one-step neighbor (DYS393 = 12) to the left of
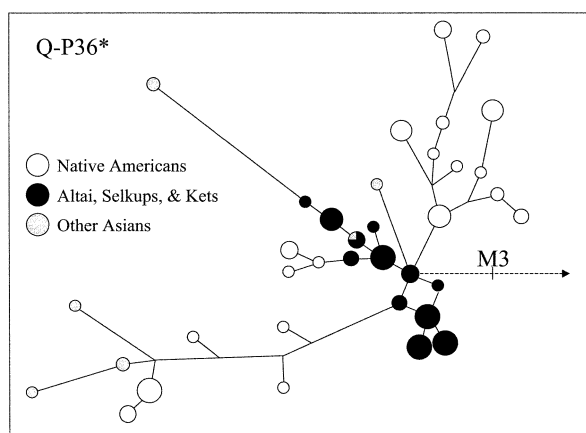
FIG. 3.—Median-joining microsatellite network for paragroup Q-P36* with position of the M3 mutation denoted by a cross-hatch. Microsatellite haplotypes are represented by circles with area proportional to the number of individuals with that haplotype. Only those haplotypes with a count of two or more are shown. Branch lengths are proportional to the number of one-repeat mutations separating the two haplotypes. Haplotypes are coded in white, black, or gray by population system (see key), with haplotype sharing indicated by pie chart divisions.

this node leads to a mixed Amerind and Na-Dene lineage, whereas the two-step neighbor ($DYS389$II $= 28$; $DYS391 = 10$) below it leads to an exclusively southwestern Na-Dene branch present in 14 Apache and 1 Navajo. The haplotype for the 2 Wayu (15–13–13–30–25–10–11–13–11–11) exhibited 6 mutational step differences from the C-P39 modal haplotype (15–13–13–28–23–9–11–12–11–11), reflecting its marked divergence from the predominant Native American C-haplogroup.
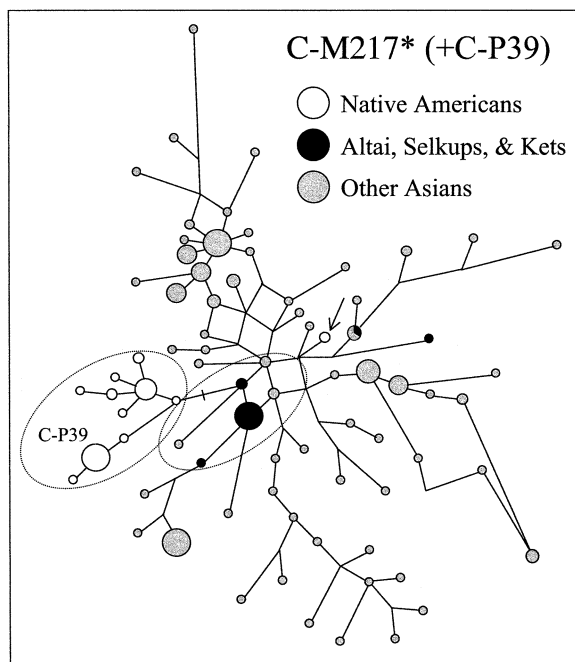

FIG. 4.—Median-joining microsatellite network for paragroup C-M217* and haplogroup C-P39 (see oval, far left) with the position of the P39 mutation denoted by a cross-hatch within the right-most oval. The two Wayu individuals outside the C-P39 cluster are indicated by an arrow. See figure 3 legend for additional explanatory material.
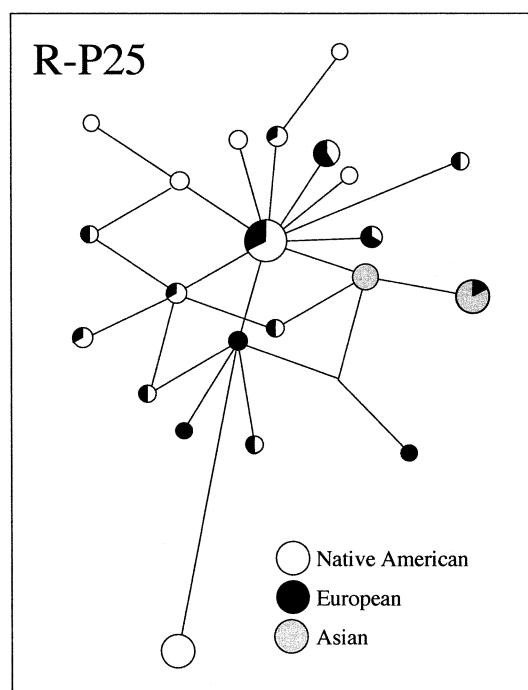

FIG. 5.—Median-joining microsatellite network for haplogroup R-P25. The large, centrally located haplotype is the Native American and European modal haplotype. See figure 3 legend for additional explanatory material.

Figure 5 gives the network for haplogroup R-P25 in Europe, Asia, and the Americas. The large central node represents 12 individuals (4 Sioux, 2 Mixtec, 1 Cheyenne, 1 Wayu, 1 Greek, 1 Italian, 1 Russian, and 1 Britain) deriving from four Native American and four European populations and exhibiting haplotype (14–12–13–29–24–11–13–13–12–12), which is identical to the R-P25 modal haplotype for both Native Americans and Europeans. In contrast, this modal haplotype differs from the Asian modal haplotype at two positions ($DYS390 = 23$; $DYS393 = 12$). Extensive sharing of haplotypes between Native Americans and Europeans is evident throughout the network.

Divergence Time Estimates

Table 3 presents divergence time estimates for the Altai plus North Asians versus the Native American population system for both the Q and C lineages using three different microsatellite-based procedures. All estimates range between 10,100 and 17,200 years; moreover, there is extensive overlap for the dates derived from the two lineages, given that all standard errors exceed 3,200 years. The Upper Bound T_D dates assume that the variance in the number of repeats at the beginning of population separation (V_o) equals zero, whereas for the Lower Bound calculations V_o is a predicted value of the within-population variance in repeat scores (V) prior to population split, assuming a linear approximation of V as a function of time (Zhivotovsky, unpublished data). Because the T_D statistic does not assume mutation-drift equilibrium, and because it is independent of population dynamics and is

Table 3
Divergence Time Estimates with Standard Errors in Years

Comparison	Upper Bound by T_D with $V_o = 0$	Lower Bound by T_D with V_o from Data[a]	From $(\delta\mu)^2$
Altai + North Asia versus Native Americans			
Q lineage	17,200 ± 4,600	14,700 ± 5,700	10,143 ± 3,714
C lineage	13,900 ± 3,200	13,600 ± 4,100	10,500 ± 4,929
Trichotomy for North, Central, and South Americans			
Entire Q lineage	15,600 ± 3,700	11,500 ± 5,900	
Only Q-M3	14,900 ± 3,800	10,100 ± 4,500	
Trichotomy for Aleut-Eskimo, Na-Dene, and Amerind Language families			
Entire Q lineage	15,400 ± 3,600	10,900 ± 6,000	
Only Q-M3	14,800 ± 3,900	10,100 ± 4,300	

NOTE.—The generation time was assumed to be 25 years with an effective Y-STR mutation rate = 0.0007 per generation; standard errors of estimates obtained by jack-knifing over loci (Weir 1996).

[a] V_o is a predicted value of the within-population variance in repeat scores (V) prior to population split assuming a linear approximation of V as a function of time.

robust to weak gene flow, it corrects for many of the complicating factors that have led to underestimates for the time of divergence based on $(\delta\mu)^2$ (Goldstein et al. 1995; Zhivotovsky 2001). For both the Q and C lineages, dates based on $(\delta\mu)^2$ are 3,000 to 7,000 years younger than the corresponding T_D dates. The separation times for the Native American trichotomy (i.e., North, Central, and South) range from a Lower Bound of 10,100 years for the Q-M3 haplogroup to an Upper Bound of 15,600 years for the entire Q lineage, a date only slightly younger than the published mutational age for the marker defining the Q lineage of 17,700 ± 4,820 years calculated by the program GeneTree (Hammer and Zegura 2002). When subdivided by language family, the dates for Q-M3 and the entire Q-lineage are remarkably similar to those based on geography.

AMOVA and Mantel Tests

The Φ_{ST} value for the sample of 486 Native Americans without suspected European/African admixture (i.e., when all haplogroup E, F, G, I, J, R-M207*, and R-P25 samples were removed) was 0.19 (table 4), a value that vividly contrasts with our recently reported Φ_{ST} of 0.41 for 18 Siberian populations (Karafet et al. 2002). When the 18 Native American populations were divided into three geographic groupings (see footnote in table 4), the Φ_{ST} value was 0.21. For the three Greenberg (1987) language family analysis, the corresponding Φ_{ST} value was 0.17. However, the among-populations within-groups (Φ_{SC}) measure exhibited slightly higher values for the language family analysis (0.21) than for the geographic-based analysis (0.14). These results did not differ appreciably from those based on analyses of molecular variance (AMOVAs) that included suspected admixed Y chromosomes, or from those based on analyses of STR variation (data not shown).

Mantel tests were used to calculate correlation and partial correlation coefficients between genetic, geographic, and linguistic distances based on both SNP and STR data (results not shown). The Native American genetic data exhibited only two statistically significant values: the SNP-based genetics–geography correlation 0.255 ($P = 0.045$) and the genetics–geography, with language held constant, partial correlation 0.254 ($P = 0.039$). Thus, both AMOVA and correlation analyses demonstrate that language affiliation is a poor predictor of paternal genetic affinities among Native American populations.

Discussion
Founder Haplogroups and Genetic Diversity

According to our present data, the Native American population system had two major founding haplogroups (Q = 76.4%; C = 5.8%), which together account for 82.2% of the 588 Native American Y chromosomes in this study. Probable admixture (see below) accounts for 17.3% of our sample, with 0.5% (i.e., three individuals) still unresolved as to ancestry. Although haplogroup Q is found at high frequencies throughout the Americas and in all three of Greenberg's (1987) linguistic groups, haplogroup C has a much more patchy distribution, with most of the C-P39 chromosomes in our sample concentrated in the three Na-Dene populations. Interestingly, haplogroup C had never been discovered in any Aleut-Eskimo sample (including ours) until very recently, when Bosch et al. (2003) reported finding two Greenland Inuit with this haplogroup in their sample of 69 males from six Greenlandic Inuit settlements. The two haplogroup C individuals came from the relatively isolated Ittoqqortoormiit, the only eastern Greenlandic settlement in their study (sample size = 15). This unexpected finding of haplogroup C in members of the Aleut-Eskimo language family means that both founding haplogroups are present in all three of Greenberg's major Native American linguistic groupings and underscores the possibility that genetic drift in small populations might be responsible for the differing frequencies of the relatively low frequency haplogroup C in the different geographical and linguistic components of the Native American population system. Thus, rather than the Q and C founding lineages representing two distinct founding events (Karafet et al. 1999; Lell et al. 2002), it is quite possible that they represent a major and minor component of the Y chromosomes in a single

Table 4
Analysis of Molecular Variance (AMOVA)

Group	n	No. of Populations	No. of Groups	Φ_{ST}[c]	Φ_{SC}[c]	Φ_{CT}[c]
Native Americans	486	18	1	0.19		
Geographic groups[a]	486	18	3	0.21	0.14	0.08
Linguistic groups[b]	486	18	3	0.17	0.21	−0.05

Note.—All analyses are based on SNPs and on samples without suspected non-Native American admixture.

[a] Group 1: INU and TAN; Group 2: CHY, SIO, NAV, APA, PIM, PUE, and SOU; Group 3: ZAP, NGO, KUN, EMB, WAU, MXA, MXE, MAY, and WAY.

[b] Aleut-Eskimo: INU; Na-Dene: TAN, NAV, and APA; Amerind: Remaining 14 populations.

[c] All Φ-statistic P values are <0.01, except the geographic groups Φ_{CT} value where $P = 0.037$ and the nonsignificant linguistic groups Φ_{CT} value.

polymorphic founding population (Tarazona-Santos and Santos 2002).

Our diversity results (table 1) also underscore the potential role of genetic drift on the Native American population system. Both SNP diversity measures showed a reduction in Native American paternal genetic diversity compared with values from Asian populations. Likewise, the overall trend in the STR data was toward a reduction in genetic diversity/variation for the Native American system, although to a lesser degree.

Asian Source Region

Earlier studies based on a much smaller number of markers led to the hypothesis of a Central Asian/South Siberian source for Native American Y chromosomes (Karafet et al. 1999; Santos et al. 1999). Our new SNP and microsatellite data have the potential to permit a finer geographic resolution than was previously possible. Both Native American founder haplogroups are present at moderately high frequencies in our sample of 98 southern Altai (Q = 17%; C = 22%); however, it is the STR data that proved to be of critical import for narrowing down the presumptive Asian source region. The ancestral nodes leading to both Q-M3 (fig. 3) and C-P39 (fig. 4), the two Native American–specific haplogroups, were present in the southern Altai individuals. Although the Kets and Sekups currently inhabit the eastern part of Western Siberia and the Yenisey River Valley, according to Russian ethnographers, their ancient homelands are thought to lie farther south, on the slopes of the Sayan and Altai mountains (Popov and Dolgikh 1964; Prokof'yeva 1964; Karafet et al. 1999). Thus, our present data support the hypothesis that the Altai Mountain region is the principal candidate for the geographic source of the founding Native American Y chromosomes.

This hypothesis is concordant with the recent results of Derenko et al. (2000, 2001), involving a candidate Asian source region for all five major Native American mtDNA founder haplogroups (i.e., haplogroups A, B, C, D, and X). Until their 2001 report, the enigmatic minor founder lineage, haplogroup X, had never been discovered anywhere in East, Central, or North Asia, although it is present in both Europeans and Native Americans. Derenko et al. (2001) found that not only did both northern and

southern Altaians have haplogroups A, B, C, and D, but 3.5% of the 202 Altai surveyed were actually haplogroup X. It should be noted, however, that all of our Altai Y chromosomes were derived from southern Altai populations and represent different samples than those used in the Derenko et al. (2001) study. Also, the Altaian haplogroup X mtDNAs are not identical to Native American haplogroup X mtDNAs (Derenko et al. 2001). Nevertheless, as far as we are aware, only the Altai region possesses all of the major Native American Y chromosome and mtDNA founding haplogroups, thereby making it the best available candidate for the ancestral source region for the Native American population system. As a caveat, we must note that it is, of course, possible that a population moved into the Altai Mountain region (presumably from the southwest) and that part of this population remained in the vicinity of the Altai and Sayan Mountains while their relatives continued moving to the northeast, eventually crossing Beringia to the Americas after the Last Glacial Maximum. In fact, by running the arrow of time backward to 100,000 years ago, all Native Americans can ultimately be traced to a dispersal from Africa; however, our main intent was to try to locate those Asian populations that are genetically the closest paternal relatives of Native Americans and who may have shared a common source with today's Native American population system. Unfortunately, without numerous chronologically secure and geographically appropriate ancient DNA samples, we may never be able to prove conclusively that modern Native Americans actually *came* from the Altai Mountains.

Timing of Entry into the Americas

A variety of genetic dating techniques, including mutational ages, mismatch distribution expansion dates, coalescence ages, and population divergence dates, have been employed to estimate the date of colonization. Cavalli-Sforza, Menozzi, and Piazza (1994) used autosomal data as a basis for their estimate of 32,000 years ago for the divergence of the Native American population system. Stone and Stoneking (1998) discussed a number of studies based on mtDNA that favored colonization dates before 20,000 BP and presented their own evidence for a 23,000–37,000 population expansion. Although mtDNA haplogroup lineages A, C, and D have generally yielded dates earlier than 20,000 BP, Schurr (2000) gave a restriction fragment length polymorphism (RFLP)-derived date of 17,700–13,500 years ago for haplogroup B, an estimate consistent with the earlier claims of Torroni et al. (1994b) and Wallace (1997) that haplogroup B in the Americas was considerably younger than the other three lineages. A possible dispersal from Eurasia to the Americas has also been dated based on haplogroup X. Brown et al. (1998) proposed that this range expansion took place either between 36,000 and 23,000 BP or 17,000 and 12,000 BP.

Earlier dating attempts using Y chromosome data have lacked precision. For instance, the origin of the Q-M3 Native American–specific lineage has been dated at either 30,000 years ago or as recently as 2,100 years ago (Underhill et al. 1996), 11,000–9,000 years ago (Ruiz-

Linares et al. 1999), 7,650 ± 5,000 years BP (Karafet et al. 1999), and 5,820 ± 2,330 years BP (Karafet, unpublished data). The mutational age of Q-P36*, the marker defining the entire Q lineage, is 17,700 ± 4,820 years BP (Hammer and Zegura 2002), whereas age estimates for the entire C lineage and the Native American-specific C-P39 are 27,500 ± 10,100 and 2,550 ± 1910 years BP, respectively (Hammer and Zegura 2002; Karafet, unpublished data). Bianchi et al. (1998) estimated that their major founder compound haplotype (based on Q-M3, a Y-specific alphoid system, and 7 microsatellites) had an average age of 22,770 years.

In contrast, all of our divergence time estimates range from 10,100 to 17,200 years ago irrespective of statistical method, population comparison, or haplogroup employed, and standard errors range from 3,200 to 6,000 years (table 3). Especially noteworthy is the general lack of temporal separation between the divergence dates based on the Q and C lineages, with only the Upper Bound T_D date hinting at an earlier separation for the Q lineage. Our divergence dates are most compatible with the late entry (<20,000 BP) school championed by most American archaeologists (Meltzer 1993, 1997; West 1996; Fiedel 2000). Indeed, the earliest generally accepted archaeological site in the Americas is Monte Verde, Chile, at 14,500 years BP (calibrated) (Meltzer 1997), and there are no securely dated skeletal remains older than 12,000 years BP (uncalibrated) anywhere in the Americas (Powell and Neves 1999; Zegura 2002). It should also be remembered that genetic evidence is expected to provide *maximum* age estimates for the peopling of the Americas, whereas archaeology only provides *minimum* estimates (unless we are fortunate enough to find the very first site in the Americas; Meltzer 2004). Likewise, Nettle's (1999) recent language-based analysis argues for a 13,000–14,000 BP entry date. In sum, the paternal genetic data lead to the conclusion that a relatively late entry date is more likely than the mtDNA-based early entry (>20,000 years ago) scenario. Moreover, mtDNA lineage expansions could have taken place in Asia rather than the Americas (Stone and Stoneking 1998), and it should be remembered that polymorphism (caused by a new mutation) generally precedes polytypy (i.e., population divergence) in evolution (Pamilo and Nei 1988), so that many of the proposed early colonization dates may not, in fact, date the population divergence associated with the actual formation of the Native American population system.

Number of Migrations

Only the synthetic work on traditional serogenetic and protein-coding loci by Cavalli-Sforza, Menozzi, and Piazza (1994) and some of the early mtDNA work (Torroni et al. 1992) have supported the Greenberg, Turner, and Zegura (1986) tripartite model. At present, the preferred explanation for many mtDNA workers is a single migration (Merriwether, Rothhammer, and Ferrell 1995; Kolman, Sambuughin, and Bermingham 1996; Bonatto and Salzano 1997a, 1997b; Stone and Stoneking 1998), although a four-migration scheme is preferred by Torroni et al. (1994a, 1994b) and Wallace (1997), also based on

mtDNA data. In fact, numerous independent data sets from linguistics (Nichols 1994/1995), immunology (Schanfield 1992), skeletal biology (Neves et al. 1999), and archaeology (Roosevelt et al. 1996) have been interpreted to support a four-migration model, often by grafting an earlier Pre-Clovis entry onto the Greenberg, Turner, and Zegura (1986) scenario (Neves and Pucciarelli 1991; Powell and Neves 1999; Zegura 2002). As mentioned earlier, the initial Y chromosome data led to a single-origin model (Pena et al. 1995; Santos et al. 1995, 1996, 1999; Underhill et al. 1996; Bianchi et al. 1997; Karafet et al. 1997; Lell et al. 1997), whereas later studies supported a two-wave model (Karafet et al. 1999; Lell et al. 2002), and Forster et al. (1996) presented a single-migration (followed by a re-expansion) model based on mtDNA evidence that can be interpreted as a two-wave scenario.

Our new data and analyses are most consistent with the single-migration alternative. For instance, (1) the divergence dates for the Q and C lineages were generally quite similar (table 3), (2) both of these lineages seem to have originated in the Altai Mountain region (figs. 3 and 4), (3) the AMOVA Φ_{CT} values for Greenberg's three linguistic groups were not statistically significant (table 4), (4) and genetics and language were uncorrelated in Mantel tests. Therefore, we have no compelling data that would refute Laughlin's (1986: 490) contention that a "single small migration some 16,000 years ago appears most parsimonious."

Thus, despite distributional differences for the Q and C lineages on both sides of the Bering Strait (Karafet et al. 1999, 2002), we cannot dismiss the parsimonious conjecture that the initial founding population possessed both lineages with somewhat unequal frequencies, and that these initial frequency differences were increased through time by successive episodes of intragenerational and intergenerational genetic drift. Subsequently, this single polymorphic Beringian population became subdivided geographically and linguistically (Szathmary 1993). This scenario is concordant with theoretical expectation for a population fragment event (Knowles and Maddison 2002). Moreover, it is consistent with the population fragmentation signal between Asia and the Americas detected in mtDNA data by Templeton (1998, 2002).

Admixture and Haplogroup R

The combination of new haplogroup and microsatellite data from haploid systems with phylogeographic information allows inferences about admixture that can resolve earlier controversies in the literature. For instance, Lell et al. (2002) and Lell, Sukernik, and Wallace (2002) proposed a two-migration model wherein an unresolved lineage, equivalent to haplogroup R, joined haplogroup C as chief binary polymorphism markers for their second migration, which they derived from the Lower Amur River/Sea of Okhotsk region. Their earlier migration supposedly brought the Q lineage to the Americas from southern Middle Siberia. Four microsatellite markers were also used to characterize these two hypothesized migrations. Tarazona-Santos and Santos (2002) questioned the

validity of Lell et al.'s (2002) second migration and proposed that the presence of what are now known to be haplogroup R individuals in the Americas was due to admixture.

On the basis of our new data and analyses, 76 of 79 Native American R lineage chromosomes belong to haplogroup R-P25. The median-joining network for R-P25 (fig. 5) exhibits extensive sharing of microsatellite haplotypes between Europeans and Native Americans, unlike the case for Asians. Also, the European and Native American modal haplotypes are identical for haplogroup R-P25, whereas the Asian modal haplotype differs at two positions. To investigate the hypothesis of European–Native American admixture for haplogroup R-P25 further, we performed a 16-population non-metric multidimensional scaling (MDS) analysis on the R-P25 microsatellite data (data not shown). The five European populations formed a distinct cluster with five of the seven Native American groups. In contrast, none of the four Asian populations were part of the European–Native American cluster.

In sum, our evidence supports the admixture hypothesis for the presence of R-P25 individuals in Native American populations and concurs with the recent findings of Bosch et al. (2003), who concluded that all 18 of their haplogroup R Greenlandic Inuit ($n = 69$) are the result of European admixture. Their overall admixture estimate for their Greenlandic Inuit sample was $58 \pm 6\%$ and they conjectured that the Greenlandic Inuit sample in Karafet et al. (1999) from Nanortalik ($n = 62$), one of their minor sampling locations ($n = 5$), exhibited between 15% and 56% European admixture, depending on the outcome of further typing to resolve haplogroup status according to the YCC (2002) system. Our new results yielded an admixture estimate of $17\% \pm 5\%$ for this Inuit (i.e., Aleut-Eskimo–speaking) sample, whereas the admixture estimate for our entire Native American sample was $17\% \pm 2\%$, following the procedures in Bosch et al. (2003).

Summary

In conclusion, like recent mtDNA studies, we find Y chromosome support for a single-migration model, with a potential common source for all major Native American Y chromosome and mtDNA founding lineages in the Altai Mountains of Southwest Siberia. Unlike the majority of these mtDNA studies, however, because none of our population divergence date estimates exceed 17, 200 years, we favor a late entry model (i.e., <20,000 BP) that postdates the Last Glacial Maximum (now calibrated at 21,000–25,000 calendar years BP). Finally, it has primarily been the interaction of genetic drift and gene flow both on Beringia and in the Americas that has produced the suite of contemporary Native American Y chromosome haplogroup frequencies that we found in our survey.

Acknowledgments

Literature Cited

Bandelt, H. J., P. Forster, and A. Rohl. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. **16**:37–48.

Bianchi, N. O., G. Bailliet, C. M. Bravi, R. F. Carnese, F. Rothhammer, V. L. Martinez-Marignac, and S. D. Pena. 1997. Origin of Amerindian Y-chromosomes as inferred by the analysis of six polymorphic markers. Am. J. Phys. Anthropol. **102**:79–89.

Bianchi, N. O., C. T. Catanesi, G. Baillet, V. L. Martinez-Marignac, C. M. Bravi, L. B. Vidal-Rioja, R. J. Herrera, and J. S. Lopez-Camelo. 1998. Characterization of ancestral and derived Y-chromosome haplotypes of New World native populations. Am. J. Hum. Genet. **63**:1862–1871.

Bonatto, S. L., and F. M. Salzano. 1997*a*. Diversity and age of the four major mtDNA haplogroups, and their implications for the peopling of the New World. Am. J. Hum. Genet. **61**:1413–1423.

———. 1997*b*. A single and early migration for the peopling of the Americas supported by mitochondrial DNA sequence data. Proc. Natl. Acad. Sci. USA **94**:1866–1871.

Bosch, E., F. Calafell, Z. H. Rosser, S. Norby, N. Linnerup, M. E. Hurles, and M. A. Jobling. 2003. High level of male-biased Scandinavian admixture in Greenlandic Inuit shown by Y-chromosomal analysis. Hum. Genet. **112**:353–363.

Brown, M. D., S. H. Hosseini, A. Torroni, H. J. Bandelt, J. C. Allen, T. G. Schurr, R. Scozzari, F. Cruciani, and D. C. Wallace. 1998. mtDNA haplogroup X: an ancient link between Europe/Western Asia and North America? Am. J. Hum. Genet. **63**:1852–1861.

Casanova, M., P. Leroy, C. Boucekkine, J. Weissenbach, C. Bishop, M. Fellous, M. Purrello, G. Fiori, and M. Siniscalco. 1985. A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance. Science **230**:1403–1406.

Cavalli-Sforza, L. L., P. Menozzi, and A. Piazza. 1994. The history and geography of human genes. Princeton University Press, Princeton, N.J.

de Knijff, P. 2000. Messages through bottlenecks: on the combined use of slow and fast evolving polymorphic markers on the human Y chromosome. Am. J. Hum. Genet. **67**:1055–1061.

Derenko, M. V., T. Grzybowski, B. A. Malyarchuk, J. Czarny, D. Mishcicka-Shliwka, and I. A. Zakharov. 2001. The presence of mitochondrial haplogroup X in Altaians from South Siberia. Am. J. Hum. Genet. **69**:237–241.

Derenko, M. V., B. A. Malyarchuk, I. K. Dambueva, G. O. Shaikhaev, C. M. Dorzhu, D. D. Nimaev, and I. A. Zakharov. 2000. Mitochondrial DNA variation in two South Siberian Aboriginal populations: implications for the genetic history of North Asia. Hum. Biol. **72**:945–973.

Excoffier, L., R. M. Harding, and R. R. Sokal. 1991. Spatial differentiation of RH and GM haplotype frequencies in sub-Saharan Africa and its relation to linguistic affinities. Hum. Biol. **63**:273–307.

Fiedel, S. J. 2000. The peopling of the New World: present evidence, new theories, and future directions. J. Archaeol. Theory **8**:39–103.

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

Goldstein, D. B., A. R. Linares, L. L. Cavalli-Sforza, and M. W. Feldman. 1995. Genetic absolute dating based on microsatellites and the origin of modern humans. Proc. Natl. Acad. Sci. USA 92:6723–6727.

Greenberg, J. H. 1987. Language in the Americas. Stanford University Press, Stanford, Calif.

Greenberg, J. H., C. G. Turner, and S. L. Zegura. 1986. The settlement of the Americas: a comparison of the linguistic, dental, and genetic evidence. Curr. Anthropol. 27:477–497.

Hammer, M. F., T. M. Karafet, A. J. Redd, H. Jarjanazi, S. Santachiara-Benerecetti, H. Soodyall, and S. L. Zegura. 2001. Hierarchical patterns of global human Y-chromosome diversity. Mol. Biol. Evol. 18:1189–1203.

Hammer, M. F., and S. L. Zegura. 2002. The human Y chromosome haplogroup tree: nomenclature and phylogeography of its major divisions. Annu. Rev. Anthropol. 31:303–321.

Hurles, M. E., J. Nicholson, E. Bosch, C. Renfrew, B. C. Sykes, and M. A. Jobling. 2002. Y chromosomal evidence for the origins of Oceanic-speaking peoples. Genetics 160:289–303.

Karafet, T. M., L. P. Osipova, M. A. Gubina et al. (14 coauthors). 2002. High levels of Y-chromosome differentiation among Native Siberian populations and the genetic signature of a boreal hunter-gatherer way of life. Hum. Biol. 74:761–789.

Karafet, T., L. Xu, R. Du, W. Wang, S. Feng, R. S. Wells, A. J. Redd, S. L. Zegura, and M. F. Hammer. 2001. Paternal population history of East Asia: sources, patterns, and microevolutionary processes. Am. J. Hum. Genet. 69:615–628.

Karafet, T. M., S. L. Zegura, O. Posukh et al. (14 co-authors). 1999. Ancestral Asian source(s) of New World Y-chromosome founder haplotypes. Am. J. Hum. Genet. 64:817–831.

Karafet, T., S. L. Zegura, J. Vuturo-Brady et al. (14 co-authors). 1997. Y chromosome markers and trans-Bering Strait dispersals. Am. J. Phys. Anthropol. 102:301–314.

Kayser, M., P. de Knijff, P. Dieltjes, M. Krawczak, M. Nagy, T. Zerjal, A. Pandya, C. Tyler-Smith, and L. Roewer. 1997. Applications of microsatellite-based Y chromosome haplotyping. Electrophoresis 18:1602–1607.

Knowles, L. L., and W. P. Maddison. 2002. Statistical phylogeography. Mol. Ecol. 11:2623–2635.

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

Kruskal, J. B. 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Pyschometrika 29:1–27.

Laughlin, W. S. 1986. Comments on: The settlement of the Americas: a comparison of the linguistic, dental, and genetic evidence. Curr. Anthropol. 27:489–490.

Lell, J. T., M. D. Brown, T. G. Schurr, R. I. Sukernik, Y. B. Starikovskaya, A. Torroni, L. G. Moore, G. M. Troup, and D. C. Wallace. 1997. Y chromosome polymorphisms in Native American and Siberian populations: identification of Native American Y chromosome haplotypes. Hum. Genet. 100:536–543.

Lell, J. T., R. I. Sukernik, Y. B. Starikovskaya, B. Su, L. Jin, T. G. Schurr, P. A. Underhill, and D. C. Wallace. 2002. The dual origin and Siberian affinities of Native American Y chromosomes. Am. J. Hum. Genet. 70:192–206.

Lell, J. T., R. I. Sukernik, and D. C. Wallace. 2002. Reply to Tarazona-Santos and Santos. Am. J. Hum. Genet. 70:1380–1381.

Lucotte, G., and N. Y. Ngo. 1985. p49f, a highly polymorphic probe, that detects Taq1 RFLPs on the human Y chromosome. Nucleic Acids Res. 13:8285.

Meltzer, D. J. 1993. Pleistocene peopling of the Americas. Evol. Anthropol. 1:157–169.

———. 1997. Monte Verde and the Pleistocene peopling of the Americas. Science 276:754–755.

———. 2004. On possibilities, prospecting, and patterns: thinking about a pre-LGM human presence in the Americas. In D. B. Madsen, ed. Entering America: Northeast Asia and Beringia before the Last Glacial Maximum, University of Utah Press, Salt Lake City, Utah (in press).

Merriwether, D. A., F. Rothhammer, and R. E. Ferrell. 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.

Nettle, D. 1999. Linguistic diversity of the Americas can be reconciled with a recent colonization. Proc. Natl. Acad. Sci. USA 96:3325–3329.

Neves, W. A., D. Munford, M. C. Zanini, and H. M. Pucciarelli. 1999. Cranial morphological variation in South America and the colonization of the New World: towards a four migration model? Ciencia e Cultura 51:151–165.

Neves, W. A., and H. M. Pucciarelli. 1991. Morphological affinities of the first Americans: an exploratory analysis based on early South American human remains. J. Hum. Evol. 21:261–273.

Nichols, J. 1994/1995. The spread of language around the Pacific rim. Evol. Anthropol. 3:206–215.

Pamilo, P., and M. Nei. 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5:568–583.

Pena, S. D. J., F. R. Santos, N. O. Bianchi, C. M. Bravi, F. R. Carnese, F. Rothhammer, T. Gerelsaikhan, B. Munkhtuja, and T. Oyunsuren. 1995. A major founder Y-chromosome haplotype in Amerindians. Nat. Genet. 11:15–16.

Poloni, E. S., O. Semino, G. Passarino, A. S. Santachiara-Benerecetti, I. Dupanloup, A. Langaney, and L. Excoffier. 1997. Human genetic affinities for Y-chromosome P49a,f/TaqI haplotypes show strong correspondence with linguistics. Am. J. Hum. Genet. 61:1015–1035.

Popov, A. A., and B. O. Dolgikh. 1964. The Kets. Pp. 607–619 in M. G. Levin and L. P. Potapov, eds. The peoples of Siberia. University of Chicago Press, Chicago, Ill.

Powell, J. F., and W. A. Neves. 1999. Craniofacial morphology of the first Americans: pattern and process in the peopling of the New World. Yrbk. Phys. Anthropol. 42:153–188.

Prokof'yeva, E. D. 1964. The Selkups. Pp. 587–606 in M. G. Levin and L. P. Potapov, eds. The peoples of Siberia. University of Chicago Press, Chicago, Ill.

Redd, A. J., A. B. Agellon, V. A. Kearney, V. A. Contreras, T. Karafet, H. Park, P. de Knijff, J. M. Butler, and M. F. Hammer. 2002. Forensic value of 14 novel STRs on the human Y chromosome. Forensic Sci. Int. 130:97–111.

Rohlf, F. J. 1998. NTSYS-pc: Numerical taxonomy and multivariate analysis system. Exeter Software, Setauket, N.Y.

Roosevelt, A. C., M. L. daCosta, C. L. Machado et al. (16 co-authors). 1996. Paleoindian cave dwellers in the Amazon: the peopling of the Americas. Science 272:373–384.

Ruiz-Linares, A., D. Ortiz-Barrientos, M. Figueroa et al. (12 co-authors) 1999. Microsatellites provide evidence for Y chromosome diversity among the founders of the New World. Proc. Natl. Acad. Sci. USA 96:6312–6317.

Santos, F. R., M. Hutz, C. E. A. Coimbra, R. V. Santos, F. M. Salzano, and S. D. J. Pena. 1995. Further evidence for the

existence of a major founder Y chromosome haplotype in Amerindians. Braz. J. Genet. **18**:669–672.

Santos, F. R., A. Pandya, C. Tyler-Smith, S. D. Pena, M. Schanfield, W. R. Leonard, L. Osipova, M. H. Crawford, and R. J. Mitchell. 1999. The Central Siberian origin for Native American Y chromosomes. Am. J. Hum. Genet. **64**:619–628.

Santos, F. R., L. Rodriguez-Delfin, S. D. J. Pena, J. Moore, and K. M. Weiss. 1996. North and South Amerindians may have the same major founder Y chromosome. Am. J. Hum. Genet. **58**:1369–1370.

Schanfield, M. S. 1992. Immunoglobulin allotypes (GM and KM) indicate multiple founding populations of Native Americans: evidence of at least four migrations to the New World. Hum. Biol. **64**:381–402.

Schneider, S., J.-M. Kueffler, D. Roessli, and L. Excoffier. 2000. ARLEQUIN: a software for population genetic analysis. Genetics and Biometry Laboratory, University of Geneva, Geneva, Switzerland.

Schurr, T. G. 2000. Mitochondrial DNA and the peopling of the New World. Am. Sci. **88**:246–253.

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.

Szathmary, E. J. E. 1993. Genetics of aboriginal North Americans. Evol. Anthropol. **1**:202–220.

Tarazona-Santos, E., and F. R. Santos. 2002. The peopling to the Americas: a second major migration? Am. J. Hum. Genet. **70**:1377–1380

Templeton, A. R. 1998. Human races: a genetic and evolutionary perspective. Am. Anthropol. **100**:632–650.

———. 2002. Out of Africa again and again. Nature **416**:45–51.

Torroni, A., Y-S. Chen, O. Semino, A. S. Santachiara-Benerecetti, C. Scott, M. T. Lott, M. Winter, and D. C. Wallace. 1994a. mtDNA and Y chromosome polymorphisms in four Native American populations from southern Mexico. Am. J. Hum. Genet. **54**:303–318.

Torroni, A., J. V. Neel, R. Barrantes, T. G. Schurr, and D. C. Wallace. 1994b. 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., T. G. Schurr, C-C. Yang et al. (11 co-authors). 1992. Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. Genetics **130**:153–162.

Underhill, P. A., L. Jin, R. Zemans, P. J. Oefner, and L. L. Cavalli-Sforza. 1996. A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. Proc. Natl. Acad. Sci. USA **93**:196–200.

Underhill, P. A., P. Shen, A. A. Lin et al. (21 co-authors). 2000. Y chromosome sequence variation and the history of human populations. Nat. Genet. **26**:358–361.

Wallace, D. C. 1997. Mitochondrial DNA in aging and disease. Sci. Am. **277**:40–47.

Weir, B. S. 1996. Genetic data analysis 11. Sinauer Associates, Sunderland, Mass.

West, F. H. 1996. American beginnings: the prehistory and paleoecology of Beringia. University of Chicago Press, Chicago, Ill.

YCC (The Y Chromosome Consortium). 2002. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. Genome Res. **12**:339–348.

Zegura, S. L. 2002. Y-chromosomes and the trinity. Pp. 347–359 in B. Frohlich, A. B. Harper, and R. Gilberg, eds. An Aleutian journey: a collection of essays in honor of William S. Laughlin. Publications of the National Museum of Denmark, Ethnographic Series, Copenhagen, Denmark.

Zhivotovsky, L. A. 2001. Estimating divergence time with the use of microsatellite genetic distances: impacts of population growth and gene flow. Mol. Biol. Evol. **18**:700–709.