Ancestral Asian Source(s) of New World Y-Chromosome Founder Haplotypes

T. M. Karafet,[1,3] S. L. Zegura,[2] O. Posukh,[3] L. Osipova,[3] A. Bergen,[4] J. Long,[4] D. Goldman,[4] W. Klitz,[5] S. Harihara,[6] P. de Knijff,[7] V. Wiebe,[3] R. C. Griffiths,[8] A. R. Templeton,[9] and M. F. Hammer[1, 2]

[1]Laboratory of Molecular Systematics and Evolution and [2]Department of Anthropology, University of Arizona, Tucson; [3]Laboratory of Human Molecular and Evolutionary Genetics, Institute of Cytology and Genetics, Novosibirsk, Russia; [4]Section of Population Genetics and Linkage, National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, Bethesda; [5]Department of Integrative Biology, University of California, Berkeley; [6]Department of Biological Sciences, University of Tokyo, Tokyo; [7]Department of Human Genetics, Leiden University, Leiden; [8]Mathematics Department, Monash University, Clayton, Australia; and [9] Department of Biology, Washington University, St. Louis

Summary

Haplotypes constructed from Y-chromosome markers were used to trace the origins of Native Americans. Our sample consisted of 2,198 males from 60 global populations, including 19 Native American and 15 indigenous North Asian groups. A set of 12 biallelic polymorphisms gave rise to 14 unique Y-chromosome haplotypes that were unevenly distributed among the populations. Combining multiallelic variation at two Y-linked microsatellites (*DYS19* and *DXYS156Y*) with the unique haplotypes results in a total of 95 combination haplotypes. Contra previous findings based on Y-chromosome data, our new results suggest the possibility of more than one Native American paternal founder haplotype. We postulate that, of the nine unique haplotypes found in Native Americans, haplotypes 1C and 1F are the best candidates for major New World founder haplotypes, whereas haplotypes 1B, 1I, and 1U may either be founder haplotypes and/or have arrived in the New World via recent admixture. Two of the other four haplotypes (YAP$^+$ haplotypes 4 and 5) are probably present because of post-Columbian admixture, whereas haplotype 1G may have originated in the New World, and the Old World source of the final New World haplotype (1D) remains unresolved. The contrasting distribution patterns of the two major candidate founder haplotypes in Asia and the New World, as well as the results of a nested cladistic analysis, suggest the possibility of more than one paternal migration from the general region of Lake Baikal to the Americas.

Introduction

The early peopling of the Americas represents one of the more controversial topics in the study of human evolution. It is generally accepted that the ancestors of Native Americans came from Asia; however, the timing, place(s) of origin, and number of "waves" of migration are among the issues debated by geneticists, archaeologists, biological anthropologists, and historical linguists. On the basis of an interdisciplinary synthesis in the 1980s, it was hypothesized that contemporary indigenous New World populations could be classified into three groups reflecting three separate migrations of ancestral Asian populations across the Bering Strait. In their tripartite model, Greenberg et al. (1986) posited that the first migration (beginning ≥12,000 years before present [YBP]) eventually resulted in the spread of Amerind-speakers throughout North, Central, and South America, whereas subsequent migrations brought the ancestors of the NaDene-speakers and Aleut-Eskimo–speakers into North America. Studies of classic genetic markers have tended to support the three-wave model (Cavalli-Sforza et al. 1994), whereas mtDNA results have been interpreted to support a range of models with as many as four major migrational waves (Torroni et al. 1992, 1993*a*, 1994; Horai et al. 1993; Shields et al. 1993; Szathmary 1993; Lorenz and Smith 1994). More-recent comprehensive reviews of the mtDNA evidence have resulted in the conjecture that all Native American populations trace to a single ancestral founder population that lived in the region of Mongolia/North China (Merriwether et al. 1995; Kolman et al. 1996).

Surveys of variation on the nonrecombining portion of the Y chromosome have revealed a single major paternal haplotype at relatively high frequencies in Native American populations from Alaska to Argentina (Santos et al. 1996; Underhill et al. 1996; Bianchi et al. 1997; Karafet et al. 1997; Lell et al. 1997). These findings generally have been interpreted to support a single origin

for the three major New World linguistic groups (Underhill et al. 1996; Karafet et al. 1997). However, because of the paucity of Y-chromosome polymorphisms sampled in Native American and Asian populations, it may be premature to infer a single paternal founder lineage and to conclude that this haploid system unequivocally supports a single-origin model for Native Americans (Karafet et al. 1997). Indeed, our present expanded database leads to a different conclusion. Thus, we now believe that there are at least two major New World paternal founder haplotypes and that these lineages may have arrived in the New World via separate migratory processes.

Subjects, Material, and Methods

DNA Samples

We analyzed a total of 2,198 males from 60 worldwide populations (table 1). The approximate geographic locations of these populations are shown in figure 1. The DNA samples included subsets of the samples examined by Hammer et al. (1997, 1998) and Karafet et al. (1997), although the exact number of subjects for each population occasionally varies among these studies. In addition, we included the following new samples: 62 Inuit Eskimos, 12 Mixe, 29 Mixtecs, 22 Kazakhs, 30 Evenks, and 18 Melanesians, which were collected by the coauthors, whereas 17 Ngöbé, 12 Kuna, 10 Emberá, and 15 Wounan from Panama were provided by E. Bermingham and C. Kolman. All sampling protocols were approved by the Human Subjects Committee at the University of Arizona.

The 15 North Asian populations in our survey were divided into the following subgroups based on geographic, linguistic, and ethnohistorical data. The "Northeast Siberia" subgroup included Siberian Eskimos, Koryaks, Chukchi, and Yukagirs (see populations 27–30 in fig. 1). These populations inhabit the extreme northeastern part of Siberia and are members of three different linguistic families within the Eurasiatic superphylum sensu Greenberg (1996). Buryats, Siberian Evenks, Manchurian Evenks, Oroqens, Evens, and Yakuts were combined into the "Baikal Region" subgroup (see populations 21–26 in fig. 1). Although these populations speak very different Eurasiatic languages and occupy the vast territory from the Yenisey River to the shores of the Sea of Okhotsk and the Bering Strait, their ancestors are thought to have lived in the vicinity of Lake Baikal, from where they moved to the north, east, and west (Levin and Potapov 1964). The "Nentsi/Komi" subgroup comprises the Forest Nentsi, Tundra Nentsi, and Komi (see populations 16–18 in fig. 1). These populations live in the northern part of western Siberia. The Forest and Tundra Nentsi represent closely related groups that speak different dialects of the same language, which belongs to the Samoyed branch of the Uralic family. A group of Finno-Ugric (Uralic)–speaking Komi, a people whose original homeland was to the west of the Urals, now shares territory with the Tundra Nentsi, with whom they have close contacts. The Kets and Selkups were joined as the fourth subgroup ("Kets/Selkups") (see populations 19 and 20 in fig. 1). These two populations currently inhabit the eastern part of western Siberia and the Yenisey River valley; however, it is thought that their ancient homelands are farther south, on the slopes of the Sayan and Altai Mountains (Popov and Dolgikh 1964; Prokof'yeva 1964). Although the Selkups speak one of the Samoyed languages, linguistically and genetically they are quite distant from the Forest Nentsi and Tundra Nentsi (Hajdü 1975; Karafet et al. 1998), whereas the Kets are classified as a linguistic isolate (Cavalli-Sforza et al. 1994). Conversely, the culture and traditions of the Selkups are so close to those of the Khants and Kets that in the eighteenth and nineteenth centuries these three populations were collectively known as "Ostyaks" (Prokof'yeva 1964).

Genotyping of Y-Chromosome Polymorphisms

The polymorphic sites in our survey included 12 biallelic polymorphisms and two microsatellite markers. *DYS287* (YAP) was scored according to the method of Hammer and Horai (1995). Polymorphic nucleotide sites in the YAP region and a C→T transition within *DYS199* were genotyped by sequence-specific oligonucleotide (SSO) hybridization (Hammer et al. 1997; Karafet et al. 1997). The A→G transition (SRY_{10831}), the G→A transition (SRY_{4064}), and the C→T transition (SRY_{9138}), all within the SRY region (Whitfield et al. 1995), as well as a G→A transition within the *DYS257* sequence-tagged site were genotyped according to methods reported by Hammer et al. (1998). We genotyped the M9 C→G transversion (Underhill et al. 1997) and the C→T transition at position 711 within the *RPS4Y* gene (Bergen et al., in press), using SSO hybridization. We genotyped a T→C transition within the single-copy Tat (RBF5) region, using the procedure reported by Zerjal et al. (1997), and two Y-chromosome microsatellite markers (*DYS19* and *DXYS156*) according to methods reported by Hammer et al. (1997) and Karafet et al. (1998). Variation in the length of the poly-A tail at the 3′ end of the YAP element was analyzed as described elsewhere by Hammer et al. (1997).

Statistical Analyses

For each individual, the combination of sequence variants observed across the set of 12 biallelic polymorphic sites is referred to as a "Y-chromosome haplotype." Kinship (R) matrix analyses with sample-size correction were performed with the program ANTANA (Harpending and Rogers 1984), according to procedures devel-

Table 1

Y-Chromosome Unique-Haplotype Frequencies in 60 Populations and 8 Geographic Regions

POPULATION (N)	2	1A	1B	1F	3G[a]	3A[a]	4[a]	5[a]	1U	1I	1E	1C	1G	1D
Europeans (EUR):														
1. British (32)	0	0	6	0	0	0	1	0	0	0	0	22	0	3
2. Germans (32)	0	0	11	0	0	0	2	0	1	0	0	15	0	3
3. Italians (39)	0	0	18	0	0	0	5	0	0	0	0	15	0	1
4. Greeks (42)	0	0	19	0	0	0	14	0	0	1	0	6	0	2
5. Russians (30)	0	0	5	0	0	0	1	0	2	1	0	8	0	13
Africans (AFR)														
6. East Bantus (55)	0	2	16	0	0	7	3	27	0	0	0	0	0	0
7. Gambians (48)	0	2	4	0	0	8	5	28	1	0	0	0	0	0
8. Egyptians (36)	0	2	8	0	0	1	14	3	3	0	0	5	0	0
9. East Africans (43)	0	2	0	0	0	5	0	36	0	0	0	0	0	0
10. West Bantus (53)	0	0	3	0	0	2	2	41	0	0	0	4	0	1
11. Khoisan (75)	22	18	13	0	0	0	7	14	0	0	0	1	0	0
12. Pygmies (38)	0	0	17	0	0	2	0	19	0	0	0	0	0	0
Australasians (AUS):														
13. Australian Aboriginal People (31)	0	0	0	16	0	0	0	0	13	0	0	2	0	0
14. Papua New Guineans (47)	0	0	0	5	0	0	0	0	42	0	0	0	0	0
15. Melanesians (55)	0	0	0	14	0	0	0	0	37	0	1	3	0	0
North Asians (NAS):														
16. Komi (8)	0	0	0	0	0	0	0	0	3	5	0	0	0	0
17. Forest Nentsi (27)	0	0	0	0	0	0	0	0	15	11	0	1	0	0
18. Tundra Nentsi (52)	0	0	0	0	0	0	0	0	38	13	0	1	0	0
19. Kets (12)	0	0	0	2	0	0	0	0	0	0	0	10	0	0
20. Selkups (122)	0	0	0	2	0	0	0	0	9	0	0	93	0	18
21. Buryats (81)	0	0	1	49	0	0	0	0	5	23	0	2	0	1
22. Siberian Evenks (95)	0	0	7	65	0	0	0	0	2	16	0	4	0	1
23. Evens (41)	0	0	1	26	0	0	0	0	0	9	0	2	1	2
24. Manchurian Evenks (41)	0	0	0	18	0	0	0	0	17	0	0	4	0	2
25. Oroqens (23)	0	0	0	21	0	0	0	0	1	1	0	0	0	0
26. Yakuts (20)	0	0	0	2	0	0	0	0	1	15	0	2	0	0
27. Siberian Eskimos (22)	0	0	0	0	0	0	0	0	0	11	0	7	3	1
28. Chukchi (4)	0	0	0	1	0	0	0	0	0	1	0	1	1	0
29. Koryaks (12)	0	0	1	4	0	0	0	0	1	4	0	2	0	0
30. Yukagirs (12)	0	0	1	6	0	0	0	0	0	3	0	2	0	0
Central Asians (CAS):														
31. Mongolians (148)	0	0	6	78	4	0	0	0	46	4	0	7	0	3
32. Altais (29)	0	0	1	7	1	0	0	0	2	0	0	2	0	16
33. Tibetans (30)	0	0	0	1	16	0	0	0	12	0	0	0	0	1
34. Kazakhs (22)	0	0	7	5	0	0	0	0	5	3	0	2	0	0
East Asians (EAS):														
35. Japanese (118)	0	0	1	6	55	0	0	0	54	0	0	2	0	0
36. Taiwanese (44)	0	0	0	3	1	0	0	0	39	1	0	0	0	0
37. South Chinese (26)	0	0	0	2	0	0	0	0	23	0	0	1	0	0
38. Koreans (28)	0	0	0	6	2	0	0	0	19	0	0	0	0	1
39. Indonesians (53)	0	0	0	19	0	0	0	0	34	0	0	0	0	0
40. Southeast Asians (38)	0	0	1	0	1	0	0	0	35	0	0	1	0	0
South Asians (SAS):														
41. Indians (60)	0	0	24	2	0	0	0	0	5	0	0	13	0	16
Native Americans (NAM):														
42. Alaskan Eskimos (4)	0	0	0	0	0	0	0	0	0	0	0	1	3	0
43. Inuit Eskimos (62)	0	0	6	0	0	0	0	0	1	0	0	26	27	2
44. Tanana (12)	0	0	1	5	0	0	0	0	0	0	0	1	5	0
45. Navajos (56)	0	0	2	1	0	0	0	0	0	1	0	23	29	0
46. Cheyenne (44)	0	0	1	7	0	0	1	1	0	0	0	27	7	0
47. Havasupai (10)	0	0	0	0	0	0	0	0	0	0	0	5	5	0
48. Pima (24)	0	0	0	0	0	0	0	0	0	0	0	14	10	0
49. Pueblos (18)	0	0	0	0	0	0	0	0	0	0	0	9	9	0
50. Zapotecs (15)	0	0	1	0	0	0	2	0	0	0	0	3	9	0

Table 1 (continued)

| | NO. OF INDIVIDUALS WITH HAPLOTYPE | | | | | | | | | | | | | |
POPULATION (N)	2	1A	1B	1F	3G[a]	3A[a]	4[a]	5[a]	1U	1I	1E	1C	1G	1D
51. Ngöbé 17	0	0	0	0	0	0	0	0	0	0	0	3	14	0
52. Kuna (12)	0	0	1	0	0	0	0	0	0	0	0	0	11	0
53. Emberá 10	0	0	0	0	0	0	0	0	0	0	0	4	6	0
54. Wounan (15)	0	0	1	0	0	0	0	0	0	0	0	3	11	0
54. Wounan (15)	0	0	1	0	0	0	0	0	0	0	0	3	11	0
55. Mixe (12)	0	0	0	0	0	0	0	0	0	0	0	0	12	0
56. Mixtecs (29)	0	0	0	0	0	0	0	0	0	0	0	3	26	0
57. Wayus (26)	0	0	3	2	0	0	0	1	0	0	0	11	9	0
58. Karitianas (2)	0	0	0	0	0	0	0	0	0	0	0	0	3	0
59. Suruis (3)	0	0	0	0	0	0	0	0	0	0	0	0	2	0
60. Chileans (3)	0	0	0	0	0	0	0	0	0	0	0	1	2	0

[a] YAP[+].

oped by Harpending and Jenkins (1973). The genetic maps resulted from standard principal-components analysis (PCA) eigenvector/eigenvalue decompositions of the R matrix, whereby population centroids and haplotypes can be plotted either separately or together. Standard admixture estimates were derived from equations given by Reed (1969).

An estimated Y chromosome–haplotype tree (cladogram) was generated by PAUP version 3.0 (Swofford 1990), with outgroup rooting. The spatial distribution of the Y-chromosome variation was assessed by nested cladistic analysis (Templeton et al. 1995; Hammer et al. 1998), which attempts to explain any statistically significant associations between haplotypes and geography, in terms of population history and/or population structure considerations. The first step in the analysis is the conversion of the estimated haplotype tree into a series of nested branches (clades). Then the geographic data are quantified, with statistical significances determined by random permutation testing. If the null hypothesis of no association between haplotype and geography is rejected, then the analysis continues by generating specific inferences about possible causal mechanisms for the patterns detected by use of the inference key reported in the work of Templeton et al. (1995). Phenomena such as recurrent gene flow restricted by isolation by distance, contiguous-population range expansion, long-distance dispersal (colonization), and past population fragmentation are representative of the biological events and evolutionary processes inferred by application of the nested cladistic analytic approach (for a more detailed explanation of the procedure, see Templeton et al. 1995).

The mean and SD of the time to the most recent common ancestral (TMRCA) Y-chromosome sequence, as well as the ages of each of the mutations in our cladogram, were estimated by the program GENETREE (Griffiths and Tavare 1994; Harding et al. 1997). This procedure models the ancestry of the sample sequences by simulating a coalescent process, including time information, conditional on a specified haplotype tree with a given value of θ (the only independent parameter in the model), where $\theta = 2N_m\mu$, N_m is the constant effective male population size, and μ is the estimated total mutation rate per sequence per generation (Hammer et al. 1998). Our simulation results are based on 5 million replicate runs. The model assumes both random mating and constant effective male population size going back in time. Although in the short run these assumptions may not strictly hold, a coalescent model should be appropriate over a long time period. To avoid a bias in mutational age estimates because of overrepresentation of North Asian and Native American males in our global sample, we analyzed a subset of the populations shown in table 1. Thus, a total of 1,336 individuals were included in the present coalescence analysis, geographically distributed as follows: 348 Africans, 307 East Asians, 175 Europeans, 133 Australasians, 116 North Asians, 100 Central Asians, 60 South Asians, and 97 Native Americans.

Similarity between the composition of this sample and that of a previously published analysis (Hammer et al. 1998) allowed us to examine consistency of mutational age and TMRCA estimates, between data sets with overlapping but differing sets of mutational sites. For example, the data set reported by Hammer et al. (1998) contained 9 of the 12 mutational sites described in the present study. Although a range of θ values was investigated in the previous analysis, a θ value of 2.5 was then considered the most appropriate, on the basis of the observed number of segregating sites in DNA-sequencing surveys of the YAP and SRY regions. The TMRCA and mutational ages for our current data set were estimated for $\theta = 2.0$, 2.5, 3.0, 3.5, and 4.0, and the dates in figure 2 are now based on the $\theta = 3.0$ analysis. Our mutation model assumes that all segregating sites in a fixed sequence length are known. However,
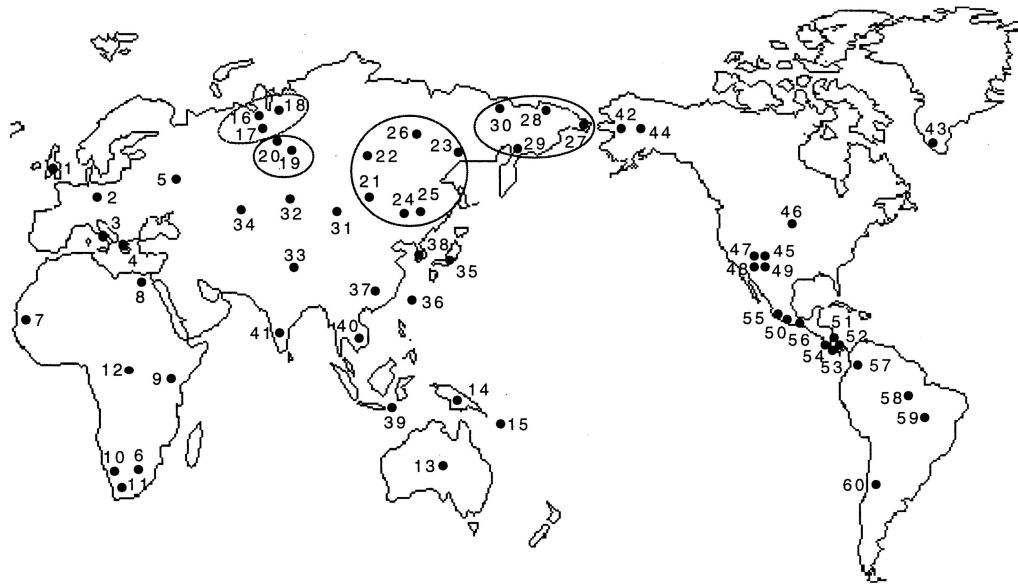
Figure 1 Approximate geographic locations of 60 study populations. Numeric population codes are as defined in table 1. The 15 North Asian populations are allocated to the following four subgroups: Nentsi/Komi (16–18), Kets/Selkups (19 and 20), Baikal Region (21–26), and Northeast Siberians (27–30).

most of our new data come from single nucleotide polymorphism (SNP) fragments. The increase from $\theta = 2.5$ to $\theta = 3.0$ attempts to accommodate the larger number of SNP fragments. The most realistic interpretation of the computed ages is that they are estimated lower bounds for the true values, since, for a fixed θ, additional unobserved segregating sites would have the effect of increasing these values. The effect of incorporating population subdivision in the model, which is also expected to increase mutational age and TMRCA estimates, will be the subject of a subsequent report.

Results

Geographic Distribution of Y-Chromosome Haplotypes

Figure 2 shows the evolutionary relationships among the 14 unique Y-chromosome haplotypes defined by variation at the 12 biallelic sites in our survey. When these 14 haplotypes were combined with variation from two Y-linked microsatellites (*DYS19* and *DXYS156Y*), a total of 95 "combination" haplotypes were produced (data not shown). The 14 haplotypes are unevenly distributed among the 60 populations (table 1). Nine unique haplotypes were found in Native American populations, the most frequent of which was haplotype 1G (53.5%), followed by haplotypes 1C (35.8%), 1B (4.3%), 1F (4.0%), YAP+ 4 (0.8%), YAP+ 5 (0.5%), 1D (0.5%), 1U (0.3%), and 1I (0.3%). The 1G haplotype (defined by the

DYS199-T allele) was identified in all North, Central, and South American groups. This predominantly New World haplotype was absent in Africa, Europe, and Australasia. Figure 3 shows a combination haplotype network relating the *DXYS156Y* and *DYS19* microsatellite variation associated with haplotypes 1C and 1G. Allelic variation at these two microsatellites revealed six 1G combination haplotypes, one of which (1G-12A) accounted for 88% of all 1G haplotypes. Previously we had found haplotype 1G in three Siberian Eskimos, one Chukchi, and one Even from the northern coast of the Sea of Okhotsk (Karafet et al. 1997). We now report that three of these five Siberian 1G chromosomes were 1G-12A (unfortunately, low amounts of DNA did not permit testing of the other two).

Haplotype 1C was also widely distributed in North, Central, and South America. In the Old World the highest frequencies of haplotype 1C were found in the Kets (83.3%) and the Selkups (76.2%); however, in the neighboring Nentsi/Komi subgroup its frequency (2.3%) was ~33-fold lower. Haplotype 1C was present at moderate frequencies (24.0%) in the Northeast Siberia subgroup and at much lower frequencies (4.7%) in the populations that originated in the Baikal Region. Interestingly, the second highest Old World regional frequency (37.7%) of haplotype 1C was found in our composite European sample, in which it demonstrated a north-south trend with a maximum frequency (68.8%) in the British (table
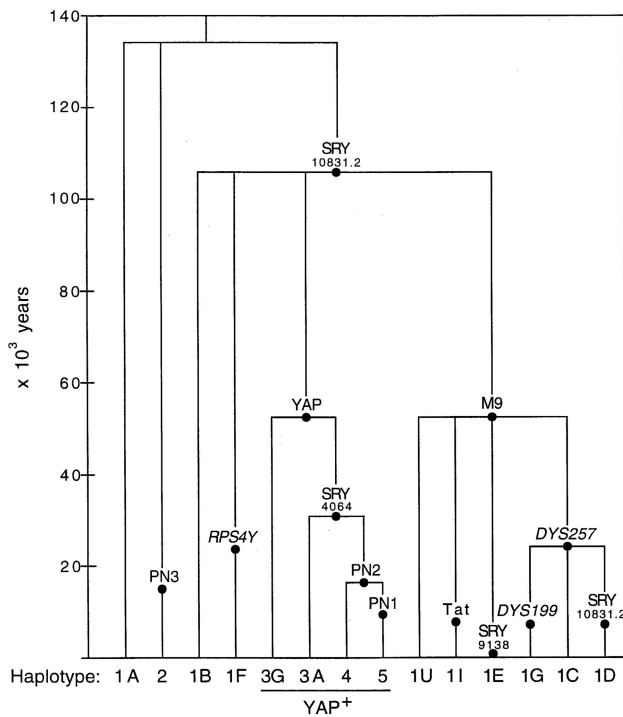
Figure 2 Time-scaled gene tree for 14 unique Y-chromosome haplotypes. Each dot represents a mutational event at one of 12 biallelic sites on the nonrecombining portion of the Y chromosome. Note that two mutational events (10831.1 and 10831.2) occurred at position 10831 in the SRY region and that, in the coalescence analysis, they were treated as separate mutations. Mutational ages (in 1,000s of years) are indicated on the *Y*-axis, and the haplotypes defined by these mutational events are shown on the *X*-axis.

1). Thus, haplotype 1C has a distinctly northern-Eurasian distribution in the Old World.

The high frequencies of haplotype 1C-12A with respect to the number of 1C haplotypes in both North Asia (104 of 131) and the New World (58 of 134), as well as the very high frequency of haplotype 1G-12A with respect to the number of 1G haplotypes in the New World (176 of 200), support the hypothesis that the *DYS199* C→T transition occurred on a Y chromosome with the 1C-12A haplotype (fig. 3). Thus, we may be able to trace the origin of the New World 1G haplotype by investigating the sources of combination haplotype 1C-12A in Asia. Similar to the distributional pattern of haplotype 1C, the highest Old World frequency (64.9%) of haplotype 1C-12A with respect to all Y chromosomes in a given subgroup was in the Kets/Selkups subgroup. However, unlike the rank order for haplotype 1C, haplotype 1C-12A was found at its second highest frequency (12.5%) in the Northeast Siberia subgroup, whereas it was only found at low frequency (2.3%) in our composite European sample. Haplotype 1C-12A was present at low frequencies (≤3%) in the Nentsi/Komi subgroup

and populations from the Baikal Region and was found in only four other populations in our survey: the Altais (6.9%), Kazakhs (4.5%), Mongolians (2.0%), and Japanese (0.8%).

Haplotype 1F has a pattern of distribution that is different than that of 1C, in both the Old World and the New World. In the New World it was only found in four Native American populations (the Tanana, Navajo, Cheyenne, and Wayu), whereas in the Old World haplotype 1F has a north-south distribution involving Asia and Australasia. Haplotype 1F reached its highest frequencies (60.1% and 52.7%, respectively) in populations from the Baikal Region and Mongolia. It was also found at moderate frequencies in many other Asian and Australasian populations, including the Northeast Siberia subgroup (22.0%) (table 1). Most (87%) Native American 1F haplotypes were combination haplotype 1F-11C (data not shown). The highest Old World frequency of this combination haplotype in continental Asia was found in the Baikal Region (24.6%), followed by Mongolia (15.5%) and Northeast Siberia (8.3%) (data not shown).

Eight Native American populations were characterized by low frequencies of haplotype 1B (table 1). This haplotype was found at higher frequencies in Europe
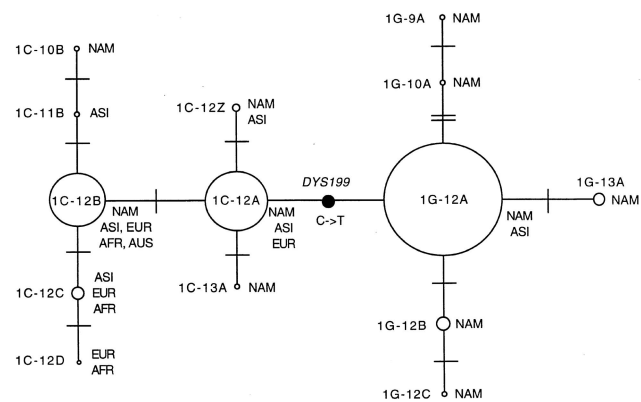

Figure 3 Combination-haplotype network showing the *DXYS156Y* and *DYS19* microsatellite variation associated with haplotypes 1C and 1G. For frequencies >40%, haplotype designations are given within the circles, and, for frequencies <10%, haplotype designations are given near the circles, where the number after the hyphen refers to *DXYS156Y* alleles 9–13 and where the letter following that number refers to *DYS19* alleles Z–D. Circled areas reflect relative global haplotype frequencies. The network was constructed by starting with the most frequent combination haplotype and then adding adjacent haplotypes that differed by a single mutational step. Each crosshatch indicates a single microsatellite mutation, and the blackened circle represents the *DYS199* C→T transition. Three-letter codes to the right of each circle denote the geographic locations of populations possessing the combination haplotype: NAM = Native Americans; ASI = Asians; AFR = Africans; EUR = Europeans; AUS = Australasians.

(33.7%) and Africa (17.5%) than in Asia (1.9%). Two of the 17 combination haplotypes associated with haplotype 1B (i.e., 1B-12B and 1B-12C) accounted for 14 of the 16 New World 1B haplotypes in our survey. These two combination haplotypes also accounted for all of the 1B chromosomes in North Asia and for 12 of the 14 1B haplotypes in Central Asia (data not shown).

The five remaining New World unique haplotypes (1D, 1U, 1I, YAP⁺ 4, and YAP⁺ 5) are very rare and together account for only 2.4% of all Native American Y chromosomes. Haplotype 1I was identified in a single Navajo, whereas haplotypes 1D and 1U were limited to two and one Inuit males, respectively. Haplotype 1I, defined by the Tat T→C transition (Zerjal et al. 1997), has a northern-Eurasian distribution similar to that of haplotype 1C, but the actual populations carrying it at the highest frequencies differed from those showing a high frequency of haplotype 1C (Karafet et al. 1998). Haplotype 1U, defined by the M9 C→G transversion (Underhill et al. 1997), was found at high frequencies in Asia and Australia, although in Europe and Africa this haplotype was very rare. Haplotype 1D was widespread in Europe, as well as in North and Central Asia.

We found the YAP element at very low frequencies in North, Central, and South America. Only two of four YAP⁺ haplotypes (i.e., 4 and 5) were present in our survey of Native Americans. Because these haplotypes are limited almost entirely to Africa and Europe, the presence of YAP⁺ haplotypes in the New World is most likely due to admixture between Native Americans and people of African and European descent (Karafet et al. 1997).

Principal-Components (R-Matrix) Analysis

Figure 4a compares eight composite populations and provides a plot of the distribution of unique haplotypes that are responsible for the dispersion among the populations. Africans and Native Americans are clearly distinguished from the other populations, along the second axis (which accounts for 23% of the total variation), on the basis of the high frequencies of haplotypes YAP⁺ 5 and 1G, respectively. The first axis (representing 25% of the variation) also highlights the distinctiveness of Africans. In addition, the first axis separates the Australasians, as well as the East, North, and Central Asians, from the other groups, because of high prevalences of
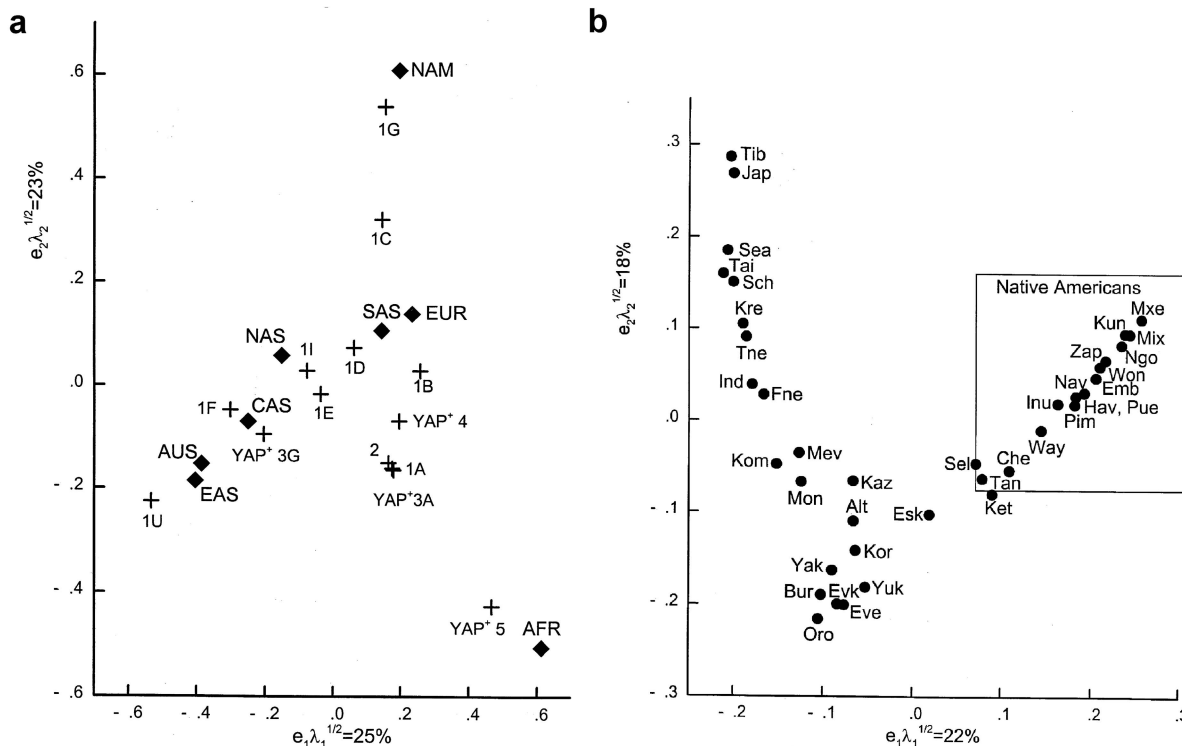

Figure 4   Genetic maps based on a principal-components decomposition of the R matrix, derived from 14 unique haplotypes. *a,* Genetic map of composite populations from eight geographic regions (*blackened diamonds*) and 14 unique haplotypes (*crosses*). The populations included within each geographic region, as well as the geographic codes, are listed in table 1. *b,* Genetic map of 38 Native American and Asian populations. The three-letter population codes are those shown in table 1, except for the following: Esk = Siberian Eskimos; Evk = Siberian Evenks; Fne = Forest Nentsi; Kre = Koreans; Mev = Manchurian Evenks; Mxe = Mixe; Sch = South Chinese; Sea = Southeast Asian; Tne = Tundra Nentsi.

haplotypes 1U, 1F, and 1I. Figure 4*b* reveals the genetic structure of 38 populations from the New World and Asia, on the basis of the 14 unique haplotypes. Although the first two axes contain only 40% of the total variation, Native American groups show a great deal of genetic similarity. The lower-left portion of the elongated Native American cluster is characterized by a concentration of populations exhibiting haplotype 1F (see box in fig. 4*b*). Two of the five populations from western Siberia (the Kets and the Selkups) have close genetic affinities to the New World populations, despite the enormous geographic expanses separating these regions. Except for the Siberian Eskimos, most of the other Siberian populations differ considerably from Native Americans and have greater affinity to Central and East Asians.

Nested Cladistic Analysis of Y-Chromosome Haplotypes

Figure 5 depicts the nested cladistic design for the 14 unique Y chromosome haplotypes in our survey. The nested cladistic analysis indicated that highly statistically significant associations exist between clades and geographic location for 9 of 10 nesting clades (i.e., at 5 of 6 one-step clades, all 3 two-step clades, and the entire cladogram). Although the overall permutational χ^2 statistic was not statistically significant for the 1-1 clade, all the associated interior-tip distance comparisons were significant at the 5% level. Thus, this clade was also
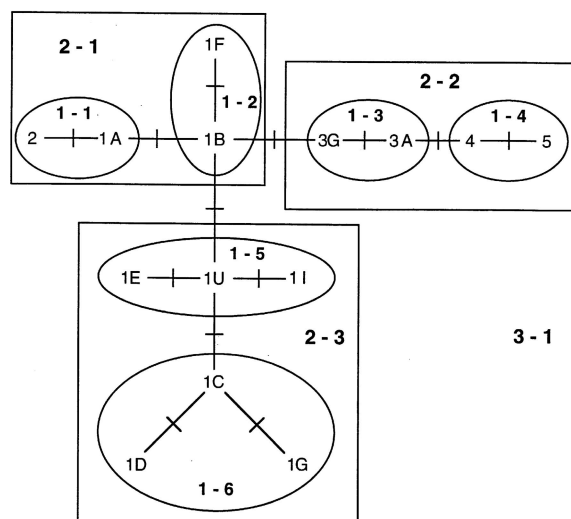

Figure 5 Nested cladistic design for 14 unique Y-chromosome haplotypes. Ovals contain one-step clades, which are designated "1-1"–"1-6." Rectangles contain two-step clades, which are designated "2-1"–"2-3." A single three-step clade, 3-1, encompasses the entire cladogram. Cross-hatches refer to the mutational events shown in fig. 2.

Table 2

Main Inferences from Results of Nested Cladistic Analysis

Clade	Inference
Haplotypes nested in 1-1	Restricted gene flow with isolation by distance (within Africa)
Haplotypes nested in 1-2	Contiguous range expansion (Asia into New World and Australasia)
Haplotypes nested in 1-3	Range expansion (Asia into Africa)
Haplotypes nested in 1-4	Range expansion (within Africa)
Haplotypes nested in 1-5	Contiguous range expansion (into North Asia/Australasia)
Haplotypes nested in 1-6	Restricted gene flow with some long distance dispersal (Europe and Asia, North Asia into New World)
1-step clades nested in 2-1	Contiguous range expansion (out of Africa)
1-step clades nested in 2-2	Restricted gene flow with isolation by distance (Old World)
1-step clades nested in 2-3	Contiguous range expansion (Old World to New World)
2-step clades nested in 3-1	Contiguous range expansion (out of Africa)

retained for subsequent inferential analysis. Table 2 summarizes the main inferences drawn from our haplotype data. Similar to those in a previous analysis of a subset of these haplotypes (Hammer et al. 1998), the results of the present study indicate that the global distribution of human Y chromosomes has been influenced by both population structure (restricted gene flow with isolation by distance) and population history (range expansions and long-distance dispersals). Three episodes of restricted gene flow with isolation by distance were indicated: within Africa, associated with the 1-1 clade (haplotypes 1A and 2); throughout the entire Old World, associated with the 2-2 clade (haplotypes 3G, 3A, YAP⁺ 4, and YAP⁺ 5); and between and within Asia and Europe and North Asia and the New World, associated with the 1-6 clade (haplotypes 1C, 1D, and 1G). In the last inference, the only one of the three that was not apparent in our earlier analysis, haplotypes 1C and 1D gave evidence for restricted gene flow primarily between Europe and Asia, whereas haplotypes 1C and 1G indicated long-distance dispersal involving North Asia and the New World.

There was also evidence for several range expansions at different levels in the cladogram. Two inferences were similar to those in our previous analysis (Hammer et al. 1998): (*a*) an early out-of-Africa range expansion detected both in the 2-1 clade (haplotypes 2, 1A, 1B, and 1F) and in the entire cladogram and (*b*) an Asia-to-Africa range expansion associated with clade 1-3 (haplotypes 3G and 3A). In addition, there was evidence for four new range expansions: one within Africa, associated

with clade 1-4 (YAP$^+$ haplotypes 4 and 5); one within Asia, associated with clade 1-5 (haplotypes 1U, 1I, and 1E); and two involving Asia and the New World. The first of these was a contiguous range expansion from Asia into the New World and Australasia, associated with clade 1-2 (haplotypes 1B and 1F). The second was a contiguous range expansion from the Old World to the New World, associated with clade 2-3 (haplotypes 1U, 1I, 1E, 1C, 1G, and 1D). Because the aforementioned inference of long-distance dispersal associated with clade 1-6 is nested within clade 2-3, this long-range colonization event may be a secondary consequence of an earlier contiguous range expansion from the Old World to the New World.

Coalescence Analysis

Figure 2 presents the mean age estimates for the 13 mutational events in the Y chromosome–haplotype tree, which are based on $\theta = 3.0$. This value of θ produced results that were in general agreement with a previously reported analysis that used a subset of these mutational sites and $\theta = 2.5$ (Hammer et al. 1998). In particular, there was general concordance of mutational ages defining the origins of five haplotypes (1C, 1B, 1D, YAP$^+$ 4, and YAP$^+$ 5) common to these two analyses. The ages of four additional mutations (defining haplotypes 1G, 1F, 1U, and 1I) are also shown in figure 2. Published age estimates exist for two of these four haplotypes: depending on the approach used, the $DYS199$ C→T transition producing 1G occurred either 2,100 or 30,000 YBP (Underhill et al. 1996), and the date for the Tat T→C transition producing 1I occurred ~2,000–4,000 YBP (Zerjal et al. 1997). Our estimated age for the $DYS199$ mutational event (~7,600 ± 5,000 YBP) is much closer to the more recent estimate by Underhill et al. (1996), and our estimated confidence interval for the age of the Tat mutational event (~8,400 ± 7,000 YBP) encompasses Zerjal et al.'s (1997) dates (fig. 2). Our newly computed TMRCA value of 134,250 ± 44,980 years is slightly more recent than our previously reported estimate of 147,000 ± 51,000 years (Hammer et al. 1998).

Discussion

New World Y-Chromosome Founder Haplotypes

Haplotype 1C fits widely used criteria for a New World founding lineage because it is widespread in both Asia and the New World. Haplotypes 1C and 1G are the most common New World haplotypes, accounting for ~90% of the Native American Y chromosomes examined. Previously, we had hypothesized that haplotype 1G may have originated in the New World/eastern Beringia and that its presence in five males from the far

northeasterly regions of Siberia was most compatible with westerly gene flow from Alaska to the Siberian side of the Bering Strait (Karafet et al. 1997). Subsequently, Lell et al. (1997) confirmed the presence of the $DYS199$-T allele in both the Siberian Eskimo population (7 of 34) and the Chukchi population (4 of 24) and hypothesized a general Beringian origin for this mutation. Although our mean mutational age of the 1G haplotype (~7,600 YBP) favors a New World origin, the large SD (~5,000 years) encompasses the final stages of the land bridge, and thus we cannot definitively rule out a Beringian origin for this haplotype.

Haplotypes 1F and 1B were found at frequencies of ~4% each in our Native American sample. Although the total frequencies of these haplotypes are similar, their patterns of distribution in both the Old World and the New World are very different. Haplotype 1F was found in only four of our New World populations but was found at relatively high frequencies (41.7% and 16.7%, respectively) in the Tanana and Cheyenne. Because haplotype 1F is widely distributed in Asia and is absent in both Africa and Europe (i.e., in the major source regions for New World admixture), we propose that haplotype 1F may well be a second New World founder haplotype. Haplotype 1B was present at low frequencies in eight Native American populations. Given its relatively high frequency (33.7%) in European populations, we cannot completely exclude the possibility that the presence of haplotype 1B in the New World is, at least in part, the result of recent contact with Europeans. Because of these ambiguities, we will defer further discussion of the status of haplotype 1B until we have more-definitive data.

Three additional haplotypes—1I, 1U, and 1D—may represent New World founding haplotypes. Unfortunately, they are so rare in Native Americans that we cannot determine whether these haplotypes represent minor founder haplotypes or are the result of recent admixture. The best cases for minor founder haplotypes are 1I, present almost exclusively (119 of 122 chromosomes) in North and Central Asia, and 1U, which is 98.5% (459 of 466 chromosomes) Asian. In contrast to the situation for haplotypes 1U and 1I, ~25% of the haplotype 1D chromosomes in our survey are European.

The best cases for New World haplotypes that originated through recent contact are YAP$^+$ haplotypes 4 and 5, because these haplotypes are predominantly found in European and African populations (Hammer et al. 1997). In our sample of 374 Native American Y chromosomes, there were three individuals (two Zapotecs and a Cheyenne) with YAP$^+$ haplotype 4 and two individuals (one Cheyenne and one Wayu) with YAP$^+$ haplotype 5. The frequency of Native American YAP$^+$ haplotype 5 chromosomes gives a rough indication of the level of gene flow from African American males, because this haplotype accounts for ~63% of West Af-

rican Y chromosomes and is not found in European populations (Hammer et al. 1997). The microsatellite alleles associated with YAP$^+$ haplotype 4 chromosomes help to pinpoint the Old World origins of these haplotypes in Native American populations. YAP$^+$ haplotype 4 in association with the *DXYS156*-Y12 allele (in one Zapotec) probably represents admixture with American males of European descent, because this combination haplotype accounts for ~70% of European YAP$^+$ haplotype 4 chromosomes, although it has not been found in West African populations (E. T. Wood and M. F. Hammer, unpublished data). Our combined estimates for paternal admixture (Reed 1969), based on the frequencies of YAP$^+$ haplotypes 4 and 5 in New World populations, ranged from 7% to 13% (depending on assumptions about the European versus African origins of particular YAP$^+$ haplotype 4 chromosomes). Estimates of Native American admixture specifically involving African American Y chromosomes ranged from 1% to 11%, whereas estimates of the European-derived component of the Native American paternal gene pool ranged from 2% to 6%.

In summary, we postulate that haplotypes 1C and 1F, with their markedly contrasting geographic distributions, are the best candidates for New World founder haplotypes, whereas haplotypes 1B, 1I, and 1U may be founder haplotypes, and/or they may have arrived in the New World via recent admixture. Two of the other four haplotypes (YAP$^+$ haplotypes 4 and 5) are most probably present because of post-Columbian admixture, whereas haplotype 1G may have originated in the New World, and the Old World source of the final New World haplotype (1D) remains unresolved.

Candidate Source Region(s)/Population(s) and Number of Migrations

Our attention now turns to the identification of candidate Asian source population(s) for Native American founding haplotypes. In particular, we will be concerned with determining the putative geographic homeland(s) of combination haplotype 1C-12A (the precursor to haplotype 1G) and combination haplotype 1F-11C. Our initial approach will be to identify contemporary Asian populations that have the highest frequencies of these candidate New World founding haplotypes and/or those Asian groups that are most closely related to New World populations. Then, we will compare four competing models for the number of paternal migrations to the New World: (1) two source populations/two migrations, (2) one source population/two migrations, (3) one source population/one migration, and (4) those involving more than two source populations and migrations.

The candidate source populations for New World 1C haplotypes include the ancestors of the Selkups and the

Kets. These two populations have the highest frequencies in Asia of both haplotype 1C (76.9%) and combination haplotype 1C-12A (64.9%) (data not shown). Although Europeans have a high frequency (37.7%) of haplotype 1C, they are less likely as a major candidate source population, because they have a very low frequency (2.3%) of combination haplotype 1C-12A (data not shown). The moderately high frequency (24.0%) of 1C haplotypes in the Northeast Siberia subgroup is consistent with the hypothesis of a genetic trail of haplotype 1C from western Siberia through northeastern Siberia en route to the New World. Also, according to our PCA map (fig. 4*b*), the Selkups and Kets (along with the Siberian Eskimo) are the Asian populations genetically closest to the New World cluster.

The candidate source populations for haplotype 1F include the populations originating in the region to the north and east of Lake Baikal, as well as Mongolians. These populations had the highest frequencies of both haplotype 1F and the specific combination haplotype, 1F-11C, that characterizes the majority (87%) of New World 1F chromosomes (data not shown). Similar to the case of haplotype 1C, the Northeast Siberia subgroup also exhibits a moderately high frequency (22.0%) of 1F haplotypes, possibly reflecting a second trail of Y-chromosome haplotypes through this part of Siberia.

The simplest explanation for the contrasting distributions of haplotypes 1C and 1F in both the Old World and the New World is that they arrived in the New World via separate migrations from different source populations (model 1). Under this scenario the ancestors of populations now living in West Siberia (for 1C) and the Baikal Region (for 1F) are the most likely sources for these two migratory waves (fig. 6). The logic of this model is similar to the hypothesis posited by Torroni et al. (1993*b*), whereby New World mtDNA founder haplogroup B represents a migratory event that is separate from the migration that brought haplogroups A, C, and D to the Americas.

Given that the ancestors of the populations now living in the regions depicted in figure 6 most probably did not live in these same locations when geographic expansion led to the peopling of the New World, another interpretative model involving a single source population combined with two migrations (model 2) should be considered. This scenario is actually closer in spirit to both Forster et al.'s (1996) maternally based one-migration/reexpansion model and the proposal that NaDene and Eskimo-Aleut populations originated from an ancestral Asiatic source population that also gave rise to the Amerind mtDNA gene pool (Shields et al. 1993). Thus, it is possible that a diverse population system occupying the general area including Lake Baikal (eastward to the Trans-Baikal and southward into northern Mongolia), the Lena River headwaters, the Angara and Yenisey
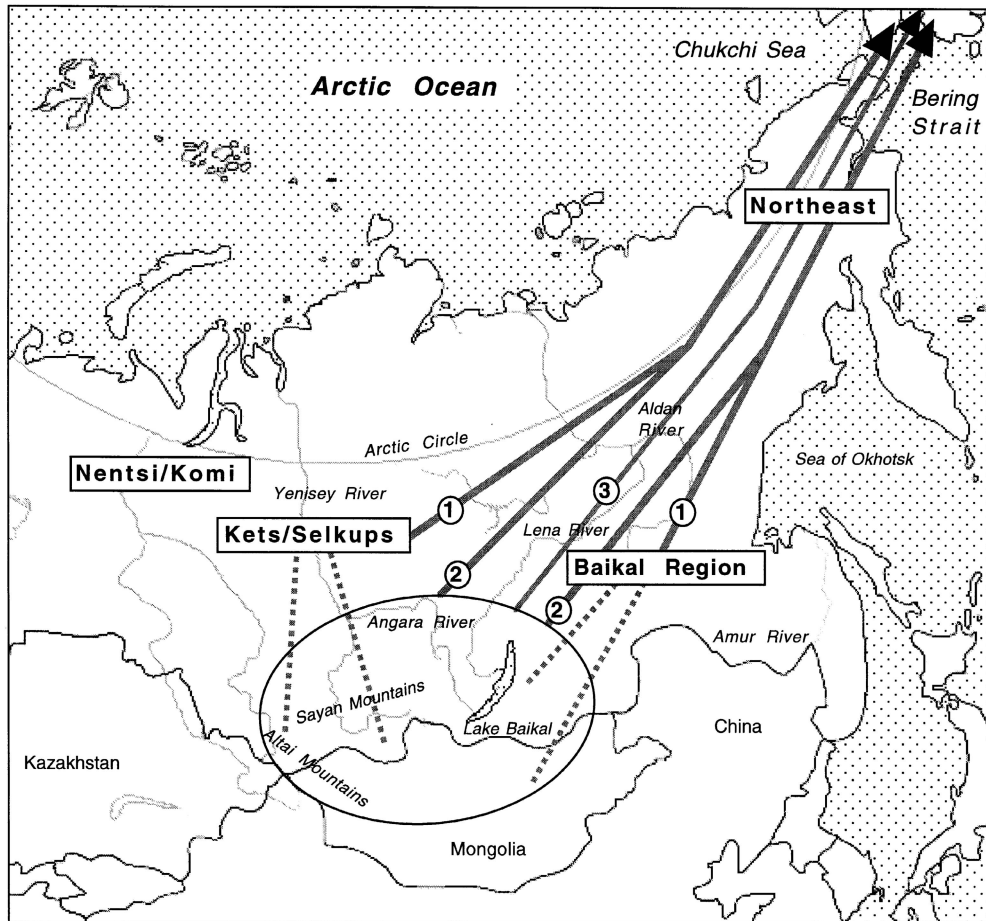
Figure 6 Alternative models for the origin of New World founder haplotypes. Numbers in circles on the solid-lined arrows refer to models 1–3 as defined in the text. These illustrative arrows do not denote precisely defined geographic routes. The dotted lines extending into the large circle indicate the possible ancestral homeland of the Kets/Selkups (haplotype 1C-12A) and of the Baikal Region populations (haplotype 1F-11C).

River basins, the Altai Mountain foothills, and the region south of the Sayan Mountains (including Tuva and western Mongolia) was the ultimate source for the dispersals of both haplotype 1F and haplotype 1C (see circle in fig. 6). Although these two dispersals have distinct distributional signatures, they may represent either concurrent or sequential resampling phenomena from the same ancestral population system.

A third model that should be considered is a single-source/single-migration model (model 3). This model has been proposed on the basis of the supposed presence of all four mtDNA haplogroups (A–D) in Native American populations throughout the Americas (Merriwether et al. 1995; Kolman et al. 1996; Stone and Stoneking 1998). The scarcity of the New World founding mtDNA haplogroups in Asian populations has also been used as evidence against multiple migrations from different source populations, whereas their circumscribed presence in Mongolian, Tibetan, Korean, and Taiwanese/ Han Chinese populations has been used to support the theory of a single migration of New World mtDNA lineages from Central Asia (Merriwether et al. 1995).

Although fewer polymorphic sites on the Y chromosome have been surveyed, the available data present a picture different than that portrayed by the mtDNA distributions: there is an extremely irregular distribution of haplotype 1F in Native American populations, and haplotypes 1C and 1F are not rare in Asia (although their frequencies vary considerably in different populations). Other than populations from the Northeast Siberia subgroup, there is no single Asian population that has even moderate frequencies of both candidate founder haplotypes. The frequencies of the 1C-12A/1F-11C haplotypes in the four candidate source populations (Mongolian, Tibetan, Korean, and Taiwanese/Han Chinese) for New World mtDNA lineages are 2.0%/5.5%, 0.0%/ 3.3%, 0.0%/7.1%, and 0.0%/6.8%, respectively (data not shown). Also, the frequency of haplotype 1U, the

ancestor of haplotype 1C and the oldest (fig. 2) and most widespread Asian paternal haplotype, ranges from 30% to 90% in these four maternal candidate source populations, yet it is nearly absent in Native American populations (i.e., it is present only in a single Inuit). Similar to the study by Karafet et al. (1997), in which Tibet was excluded as a major source of New World Y chromosome haplotypes, our new data do not support the hypothesis that either Mongolia, Korea, or Taiwan was the sole source of Native American paternal lineages (although in the case of Mongolia more demes should be examined, because our sample was composed mainly of Khalkhs).

Additional support for more than one migration comes from the nested cladistic analysis. Two contiguous range expansions from Asia to the New World were inferred, the first involving nesting clade 1-2 and the second involving 2-3 (fig. 5). One reasonable implication of these results is that haplotype 1F spread from Asia to the New World via a microevolutionary event *different* than that which brought haplotype 1C from Asia to the New World (i.e., via a separate migration). Interestingly, our Y-chromosome analysis parallels the situation seen in a nested cladistic analysis of mtDNA. Three distinct range expansion/colonization events were revealed in Templeton's (1998b) analysis of Torroni et al.'s (1993a, 1993b) Siberian and New World mtDNA data. First, a contiguous range expansion was detected throughout Siberia. Next, similar to our results based on study of Y chromosomes, several clades in two separate nested series indicated possible range expansion/ colonization events from Siberia to the Americas. The third event was a contiguous range expansion throughout the Americas. Superimposed on all of these events and occurring in both Siberia and the Americas was restricted gene flow, primarily through isolation by distance but with occasional long-distance dispersal in the Americas. Templeton (1998a) concluded that mtDNA expansion into the New World either involved more than one colonization event (i.e., model 1, 2, or 4), a sufficiently large colonizing population to carry over many of the haplotype lineages from the ancestral range (i.e., a version of model 3), or an extended colonization (i.e., a model similar in spirit to that of Forster et al. 1996). Because it is unlikely that a single migrating deme was large enough to carry all founding mtDNA and Y chromosome–haplotype lineages, a model that invokes a single migration of a relatively small population is less compatible with the entire corpus of haploid data.

How well do our new Y-chromosome data correspond to the tenets of Greenberg et al.'s (1986) tripartite model (one possible version of model 4)? Table 3 shows the frequency data for seven of the nine New World haplotypes (European- and African-derived YAP$^+$ haplotypes 4 and 5 have been omitted) grouped according to Greenberg's (1987) three major Native American linguistic families, as well as Bonferroni pairwise tests for each of the three two-way comparisons for each haplotype. Only 2 of the 21 Bonferroni tests were statistically significant: the 1F Eskimo-Aleut versus NaDene comparison ($P < .05$) and the 1D Eskimo-Aleut versus Amerind comparison ($P < .01$). For the five other haplotypes, including 1C and 1G, which together constitute ~90% of the New World haplotypes, none of the pairwise Bonferroni tests reached statistical significance. Thus, as for our earlier data set (Karafet et al. 1997), the new Y-chromosome data still provide no statistical support for the linguistically based three-separate-migrations model. Of course, additional Y-chromosome data may eventually uncover the trails of additional migrations. However, even if a three-wave model is eventually supported by Y-chromosome evidence, it is not clear how these migratory events would relate to the linguistic structure of New World populations, given the

Table 3

Frequencies of Seven Y-Chromosome Haplotypes in Major New World Linguistic Divisions

	DATA ON HAPLOTYPE						
	1B	1F	1U	1I	1C	1G	1D
	Frequency (Standard Error)[a]						
Linguistic groups (*N*):							
A. Eskimo-Aleut (66)	9.1 (3.5)	...	1.5 (1.5)	...	40.9 (6.1)	45.5 (6.1)	3.0 (2.1)
B. NaDene (68)	4.4 (2.5)	8.8 (3.4)	...	1.5 (1.5)	35.3 (5.8)	50.0 (6.1)	...
C. Amerind (235)	3.0 (1.1)	3.8 (1.3)	...	...	35.3 (3.1)	57.9 (3.2)	...
	Bonferroni *P* Value[b]						
Comparisons:							
A vs. B	NS	<.05	NS	NS	NS	NS	NS
A vs. C	NS	NS	NS	NS	NS	NS	<.01
B vs. C	NS	NS	NS	NS	NS	NS	NS

[a] An ellipsis denotes that the haplotype is absent.

[b] NS = not significant.

present lack of concordance between these two bodies of data.

In summary, on the basis of our distributional data, the two haploid nested cladistic analyses, PCA results, and demographic considerations, we have constructed the following scenario: although one migration brought haplotype 1C to the New World, another migration was responsible for bringing haplotype 1F to the Americas. Thus, our inferences of separate dispersals for haplotypes 1C and 1F imply a model with at least two waves for the peopling of the New World. Despite the fact that our present data do not allow us to refute model 3, the totality of the present evidence leads us to prefer a scenario more similar to either model 1 or model 2.

Further Considerations and Caveats

How do we reconcile differences between our two-migration scenario and the recent maternally based hypothesis of a single migrational wave? First, it is not clear that one should expect complete concordance between inferences based on maternally inherited genetic systems and those based on paternally inherited genetic systems. It is becoming evident that there are significant differences in the geographic patterning of mtDNA diversity versus Y-chromosome diversity, differences that may be related to different male and female demographic histories (Bamshad et al. 1998; Hammer et al. 1998; Seielstad et al. 1998). In general, a greater fraction of the total genetic variance is distributed among populations for the Y-chromosome data than for mtDNA data. One explanation for this pattern at the local level is a higher female migration rate via patrilocality (defined anthropologically as the tendency for a wife to move into her husband's natal domicile) (Seielstad et al. 1998). However, additional factors—such as occasional long-distance migrations of males, accompanied by intergenerational genetic drift and founder effects—are likely to play a role in the processes that generate the observed global patterns (Hammer et al. 1998). Thus, both the increased mobility of mtDNA at the local and regional levels *and* the patrilocality of Y chromosomes may lead to different inferences about the number of migrations to the New World. According to these premises, source populations from geographically separated regions in Asia would exhibit greater mtDNA homogeneity, whereas their paternal genomes would show greater amomg-group genetic differentiation. Because of these demographic considerations, differentiation between a one-wave model and models that imply multiple migrations may be more difficult when mtDNA data are used than when data on Y-chromosome haplotypes are used. Finally, the vexing bottom line is that, as more data are collected, sex-specific microevolutionary scenarios may actually diverge rather than converge, because there were real differences in male and female population structure and demographic histories through time.

In the interpretation of our results, the following caveats should be kept in mind. First, gene trees such as our Y-chromosome scaled coalescent tree (fig. 2), the numerous mtDNA trees in the literature (Cann et al. 1987), and the recent global β-globin–analysis tree based on autosomal sequence data (Harding et al. 1997) are not equivalent to population trees. Inferences about population relationships derived from gene trees must be made very cautiously, especially since each gene has its own evolutionary history (Harpending et al. 1998). Second, the standard errors of the estimated ages of the Y haplotypes are high and often encompass a time span too long to allow discrimination among competing hypotheses concerning the timing of the peopling of the New World. This caveat also applies to many of the age estimates for mtDNA lineages in the literature. Third, because there is no recombination for mtDNA, the entire molecule acts as a single locus. This is also true for most of the Y chromosome (except for the two terminal pseudoautosomal regions). Thus, in terms of evolutionary behavior, the 14 Y-chromosome markers used here should be considered as a single locus (Hammer and Zegura 1996). Finally, we wish to underscore the real possibility that extensive gene flow in Asia, on Beringia, and subsequently in the New World (Karafet et al. 1997) may have complicated the population history/structure of Native Americans to the extent that any candidate phylogenetic signals must be interpreted with great caution.

One of the main contributions of this study is that it makes specific predictions about the sources of paternal Native American lineages in Asia. This information alone cannot give us definitive answers about the origin(s) of Native American populations; it is only through comparisons with other genetic systems, as well as with data from fields such as archaeology, skeletal biology, linguistics, etc., that a coherent picture of the early peopling of the Americas may emerge. In this holistic context our inferences may best serve as heuristics for these cognate disciplines to see whether similar relationships exist between candidate Asian source populations and Native American populations.

These data may also help explain a puzzling observation resulting from the discovery of "Kennewick man," an ~9,000-year-old skeleton with so-called caucasoid characteristics that recently was found in Washington State (Morell 1998). Our inference of a southwestern-Siberian/Central Asian origin for Native American 1C haplotypes suggests the hypothesis that males expanded in two directions from the region within the circle in figure 6: one pool of 1C haplotypes moved northeastward toward the New World, while a different pool of 1C haplotypes moved westward toward Europe.

Moreover, we have recently analyzed unpublished Y-chromosome SNP data that support this bidirectional-movement hypothesis. Thus, the conjecture of shared paternal ancestry could explain the observed morphological similarities between Europeans and the Kennewick specimen, without the need to invoke more-simplistic models in which Europeans are seen as having participated in the early colonization of the New World.

Acknowledgments

References

Bamshad MJ, Watkins WS, Dixon ME, Jorde LB, Rao BB, Naidu JM, Prasad BV, et al (1998) Female gene flow stratifies Hindu castes. Nature 395:651–652

Bergen A, Wang C-Y, Jefferson K, Dey C, Smith KD, Park S-C, Tsai S-J, et al. An Asian-Native American paternal lineage identified by RPS4Y resequencing and by microsatellite haplotyping. Ann Hum Genet (in press)

Bianchi NO, Bailliet G, Bravi CM, Carnese RF, Rothhammer F, Martinez-Marignac VL, Pena SDJ (1997) Origin of Amerindian Y-chromosomes as inferred by the analysis of six polymorphic markers. Am J Phys Anthropol 102:79–89

Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. Nature 325:31–36

Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton

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

Greenberg JH (1987) Language in the Americas. Stanford University Press, Stanford
——— (1996) The linguistic evidence. In: West FH (ed) American beginnings: the prehistory and palaeoecology of Beringia. University of Chicago Press, Chicago, pp 525–536

Greenberg JH, Turner CG, Zegura SL (1986) The settlement of the Americas: a comparison of the linguistic, dental, and genetic evidence. Curr Anthropol 27:477–497

Griffiths RC, Tavare S (1994) Ancestral inference in population genetics. Stat Sci 9:307–319

Hajdü P (1975) Finno-Ugrian languages and peoples. Deutsch, London

Hammer MF, Horai S (1995) Y chromosomal DNA variation and the peopling of Japan. Am J Hum Genet 56:951–962

Hammer MF, Karafet T, Rasanayagam A, Wood ET, Altheide TK, Jenkins T, Griffiths RC, et al (1998) Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. Mol Biol Evol 15:427–441

Hammer MF, Spurdle AB, Karafet T, Bonner MR, Wood ET, Novelletto A, Malaspina P, et al (1997) The geographic distribution of human Y chromosome variation. Genetics 145:787–805

Hammer MF, Zegura SL (1996) The role of the Y chromosome in human evolutionary studies. Evol Anthropol 5:116–134

Harding RM, Fullerton SM, Griffiths RC, Bond J, Cox MJ, Schneider JA, Moulin DS, et al (1997) Archaic African *and* Asian lineages in the genetic ancestry of modern humans. Am J Hum Genet 60:772–789

Harpending HC, Batzer MA, Gurven M, Jorde LB, Rogers AR, Sherry ST (1998) Genetic traces of ancient demography. Proc Natl Acad Sci USA 95:1961–1967

Harpending H, Jenkins T (1973) Genetic distance among southern African populations. In: Crawford MH, Workman PL (eds) Methods and theories of anthropological genetics. University of New Mexico Press, Albuquerque, pp 177–199

Harpending H, Rogers A (1984) Antana: a package for multivariate data analysis. Department of Anthropology, University of Utah, Salt Lake City

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

Karafet T, de Knijff P, Wood E, Ragland J, Clarke A, Hammer MF (1998) Different patterns of variation at the X- and Y-chromosome linked microsatellite loci *DXYS156X* and *DXYS156Y* in human populations. Hum Biol 70:979–992

Karafet T, Osipova LP, Posukh OL, Wiebe V, Hammer MF (1999) Y chromosome microsatellite haplotypes and the history of Samoyed-speaking populations in Northwest Siberia. In: Goldstein DB, Schlötterer C (eds) Microsatellites: evolution and applications. Oxford University Press, Oxford (in press)

Karafet T, Zegura SL, Vuturo-Brady J, Posukh O, Osipova L, Weibe V, Romero F, et al (1997) Y chromosome markers and trans-Bering Strait dispersals. Am J Phys Anthropol 102:301–314

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

Lell JT, Brown MD, Schurr TG, Sukernik RI, Starikovskaya YB, Torroni A, Moore LG, et al (1997) Y chromosome polymorphisms in native American and Siberian populations: identification of native American Y chromosome haplotypes. Hum Genet 100:536–543

Levin MG, Potapov LP (1964) The peoples of Siberia. University of Chicago Press, Chicago

Lorenz JG, Smith DG (1994) Distribution of the 9-bp mitochondrial DNA region V deletion among North American Indians. Hum Biol 66:777–788

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

Morell V (1998) Kennewick man's trials continue. Science 280:190–192

Popov AA, Dolgikh BO (1964) The Kets. In: Levin MG, Po-

tapov LP (eds) The peoples of Siberia. University of Chicago Press, Chicago, pp 607–619

Prokof'yeva ED (1964) The Selkups. In: Levin MG, Potapov LP (eds) The peoples of Siberia. University of Chicago Press, Chicago, pp 587–606

Reed TE (1969) Caucasian genes in American Negroes. Science 165:762–768

Santos FR, Rodriguez-Delfin L, Pena SDJ, Moore J, Weiss KM (1996) North and South Amerindians may have the same major founder Y chromosome haplotype. Am J Hum Genet 58:1369–1370

Seielstad MT, Minch E, Cavalli-Sforza LL (1998) Genetic evidence for a higher female migration rate in humans. Nat Genet 20:278–80

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

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

Swofford D (1990) PAUP: phylogenetic analysis using parsimony. Illinois Natural History Survey, University of Illinois, Champaign-Urbana

Szathmary EJE (1993) mtDNA and the peopling of the Americas. Am J Hum Genet 53:793–799

Templeton AR (1998a) Human races: a genetic and evolutionary perspective. In: Sussman RW (ed) The biological basis of human behavior: a critical review. Washington University Press, St Louis, pp 180–192

——— (1998b) Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. Mol Ecol 7:381–397

Templeton AR, Routman E, Phillips CA (1995) Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. Genetics 140:767–782

Torroni A, Neel JV, Barrantes R, Schurr TG, Wallace DC (1994) 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 (1993a) 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, Schanfield 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 (1993b) mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. Am J Hum Genet 53:591–608

Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, et al (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. Genome Res 7:996–1005

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

Whitfield LS, Sulston JE, Goodfellow PN (1995) Sequence variation of the human Y chromosome. Nature 378:379–380

Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos FR, Schiefenhövel W, et al (1997) Genetic relationships of Asians and northern Europeans, revealed by Y-chromosomal DNA analysis. Am J Hum Genet 60:1174–1183