

## The Central Siberian Origin for Native American Y Chromosomes

Fabrício R. Santos,<sup>1,2</sup> Arpita Pandya,<sup>1</sup> Chris Tyler-Smith,<sup>1</sup> Sérgio D. J. Pena,<sup>2</sup> Moses Schanfield,<sup>3</sup> William R. Leonard,<sup>4</sup> Ludmila Osipova,<sup>5</sup> Michael H. Crawford,<sup>6</sup> and R. John Mitchell<sup>7</sup>

<sup>1</sup>Department of Biochemistry, Oxford University, Oxford; <sup>2</sup>Departamento de Bioquímica, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil; <sup>3</sup>Analytical Genetic Testing Center, Inc., Denver; <sup>4</sup>Department of Anthropology, University of Florida, Miami; <sup>5</sup>Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia; <sup>6</sup>Department of Anthropology, University of Kansas, Lawrence; and <sup>7</sup>School of Genetics and Human Variation, La Trobe University, Bundoora, Australia

### Summary

Y chromosomal DNA polymorphisms were used to investigate Pleistocene male migrations to the American continent. In a worldwide sample of 306 men, we obtained 32 haplotypes constructed with the variation found in 30 distinct polymorphic sites. The major Y haplotype present in most Native Americans was traced back to recent ancestors common with Siberians, namely, the Kets and Altaians from the Yenisey River Basin and Altai Mountains, respectively. Going further back, the next common ancestor gave rise also to Caucasoid Y chromosomes, probably from the central Eurasian region. This study, therefore, suggests a predominantly central Siberian origin for Native American paternal lineages for those who could have migrated to the Americas during the Upper Pleistocene.

### Introduction

The pre-Columbian settlers of the New World, who gave rise to the present-day Native Americans, are commonly believed to have come from Siberia, through the Bering land bridge, in the period 30,000–12,000 years before present (ybp). These conclusions are based on cultural, morphological, and genetic similarities between populations of the New World and Siberia (for a review, see Crawford 1998). Although affinities between Asians and Native Americans have been acknowledged for a long time, no particular population in Siberia, except for some Asian Eskimos and their relatives, has been pointed out as directly descended from ancient groups related to

the American founder populations (Cavalli-Sforza et al. 1994). Morphological studies have suggested a place of origin for Native Americans along the Amur River region (Crawford 1998), and, more recently, the investigation of mtDNA lineages (Kolman et al. 1996; Merriwether et al. 1996), as well as of retrovirus infections (Neel et al. 1994), has suggested that Mongolia, instead of Siberia, is the source of populations that share the more recent ancestors with the founding population of the Americas.

Siberia is an inhospitable place for human settlements, but the first hominids may have arrived as early as 260,000 ybp (Waters et al. 1997). The population density has never been high, and it is still vastly uninhabited. These conditions favor a high degree of population isolation and genetic drift, which could have played an important role since the first migrants left for the Americas. Furthermore, the number of indigenous Siberian populations has been decreasing since the beginning of the Russian territorial expansion in the seventeenth century (Forsyth 1996). Many populations have become extinct, and others—such as the Kets, who are inhabitants of the Yenisey River Basin—are now reduced to <1,000 individuals and are speakers of an isolated language unrelated to any other extant languages (Grimes 1996).

The three-migrations theory (Greenberg et al. 1986) postulates that each of the major Native American groups—namely, the Amerindians, Na-Denes, and Eskimo-Aleuts—came to the Americas in three distinct migratory waves from Siberia, during the period 12,000–6,000 ybp. This theory received support a posteriori from the analysis of mtDNA (Torroni et al. 1993) and several autosomal markers (Cavalli-Sforza et al. 1994), but these data extended the entrance time of these groups to the Americas to 34,000–6,000 ybp. This model has been criticized mainly for the claims of the existence of a genetic homogeneity in present-day Amerindians (Schanfield 1992), who are considered to be the descendants of the first migrants. In addition, further mtDNA analyses have been shown to be consistent with a single migration into the Americas (Merriwether et al.

Received July 7, 1998; accepted for publication December 14, 1998; electronically published February 2, 1999.

Address for correspondence and reprints: Dr. Fabrício R. Santos, Departamento de Biologia Geral, ICB/UFMG, Caixa Postal 486, 31.270-910 Belo Horizonte, MG, Brazil. E-mail: fsantos@mono.icb.ufmg.br

© 1999 by The American Society of Human Genetics. All rights reserved.  
0002-9297/99/6402-0034\$02.00

1995) and with a single migration with further population reexpansion (Forster et al. 1996; Bonnato and Salzano 1997). Some anthropometric studies have also revealed the existence of skeletons not typical of Mongoloids among the oldest hominids in the Americas, suggesting an earlier, non-Mongoloid migration (Neves and Pucciarelli 1991; Lahr 1995).

Recent studies of the human Y chromosome have shown a major haplotype present in >90% of nonadmixed southern and central Amerindians (Pena et al. 1995; Santos et al. 1995a), indicating a genetic homogeneity and a pronounced founder effect in the formation of these populations. Later, some northern Amerindians and also Na-Dene and Eskimo speakers were studied, and, despite their higher level of admixture (Crawford 1998), these populations also displayed the same major haplotype, although not in such a high frequency (Santos et al. 1996b; Underhill et al. 1996). This major Native American haplotype initially was identified as the combination of alphoid heteroduplex ( $\alpha h$ ) type II and the microsatellite *DYS19* A allele (Pena et al. 1995; Santos et al. 1995a), and it subsequently was shown to also be defined by a C→T transition in the *DYS199* locus (Underhill et al. 1996). The study of several Siberian populations (Karafet et al. 1997; Lell et al. 1997) has identified the presence of the *DYS199* T allele only in Asian Eskimos and related tribes from the Beringia region. The presence of the T allele in far northeastern Siberia was explained to be the result of either a back migration of Native American populations bearing the *DYS199* T allele or simply a split of populations inhabiting Beringia, after the glaciation period. This suggests that the *DYS199* T allele is a useful marker for the identification of Y haplotypes originating after the first migration to the Americas or Beringia but that more-informative (ancient) markers are needed to trace the origin of these Y chromosomes within Asia.

In this study, the Y chromosomes of five linguistically distinct Siberian populations, as well as of those of Native Americans, Europeans, Indians, Mongolians, central East Asians, and Africans, were analyzed with a set of seven polymorphic systems identifying 30 variable loci in the nonrecombining portion of the Y chromosome. The worldwide distribution of haplotypes and their evolutionary network shows the recent common ancestry of Caucasoid and Native American Y chromosomes, as well as the identification of intermediate Y haplotypes in Siberian populations from the Altai Mountains and the Yenisey River Basin, namely, the Altaians and Kets, respectively.

## Material and Methods

### DNA Samples

Most of the 306 male samples were obtained as DNA or were extracted from plugs prepared for pulsed-field

gel electrophoresis (Mathias et al. 1994). Samples from Europeans (most were British), Indians (India and Sri Lanka), Africans (Kenyans, Pygmies, and San), central East Asians (Chinese and Japanese), Mongolians (Khalkhs), and Siberians (Buryats, Yakuts, Evenkis, Altaians, and Kets) were subsets of those described elsewhere (Mathias et al. 1994; Zerjal et al. 1997). Ten samples, from south and central Amerindians and a Na-Dene, were purchased from the National Institute of General Medical Science, and an additional 10 Native American samples (not Aleut-Eskimos) came from paternity tests in North America.

### DNA Polymorphisms

The variants of 6-kb and 4.1-kb alphoid units were first identified by hybridization with the pY $\alpha$ l probe (Mathias et al. 1994) and subsequently were checked by *Hind*III digestion of  $\alpha h$  PCR products (Santos et al. 1995b), to avoid confusion because of the comigration of 4.1-kb bands on the gel from both the 6-kb and the 4.1-kb units. For most cases, 92R7 typing was performed by hybridization (Mathias et al. 1994), but PCR was also used to type some individuals (M. Hurles, F. R. Santos, A. Pandya, and C. Tyler-Smith, unpublished data).

Additional systems—namely, *DYS199*, *SRY-1532*, Tat, the Y *Alu* polymorphism (YAP), and  $\alpha h$ —were scored, after PCR, in a MJR PTC200 thermocycler, with a 12.5- $\mu$ l reaction volume, 1  $\mu$ M of each primer, 200  $\mu$ M dNTPs, 1.5 mM MgCl<sub>2</sub>, 1 U *Taq* (Bioline) per tube with 1 × KCl buffer (Bioline), and other changes as follows. The locus *DYS199* (Underhill et al. 1996) was amplified for 30 cycles at 94°C for 20 s, 61°C for 20 s, and 72°C for 30 s, with a modified reverse primer, 5'-AGG TAC CAG CTC TTC CCA ATT-3', containing a G→C base change (underlined) that creates an artificial *Mfe*I restriction site when the *DYS199* C allele is present. This PCR/RFLP protocol using the *Mfe*I enzyme (New England Biolabs), resolved in native polyacrylamide gels stained by silver (Santos et al. 1996a), allowed us to determine without doubt the allele state at this site that had been detected previously by an allele-specific-primer protocol (Underhill et al. 1996). The locus *SRY-1532*, with an A→G mutation (Whitfield et al. 1995), was amplified for 30 cycles at 94°C for 20 s, 60°C for 20 s, and 72°C for 30 s, with the primers *SRY1* (5'-TCC TTA GCA ACC ATT AAT CTG G-3') and *SRY2* (5'-AAA TAG CAA AAA ATG ACA CAA GGC-3') and 0.5 U *Taq* per tube. The G allele was detected by the presence of the *Dra*III (Boehringer) restriction site on a 1.5% agarose gel in Tris/acetate/EDTA 0.5 ×. The YAP system was scored, as described elsewhere (Hammer and Horai 1995), in the PCR conditions described above but with PRO-MEGA enzyme and buffer. The Tat polymorphism (Zerjal et al. 1997) and the  $\alpha h$  system (Santos et al. 1995b,

1996a) were detected and classified as described previously. Most individuals were also typed for the tetranucleotide microsatellite *DYS19* (Santos et al. 1996a).

#### *Population Genetics of Y Chromosomes*

Haplotype frequencies and gene diversities (Nei 1987) were calculated for all populations. A parsimonious network was constructed either manually or by median network analysis (Bandelt et al. 1995), with the knowledge of the molecular mechanisms of the  $\alpha$ h-system mutations (Santos et al. 1996a) and other loci (Mathias et al. 1994; Hammer and Horai 1995; Jobling and Tyler-Smith 1995; Whitfield et al. 1995; Underhill et al. 1996). On the basis of this network, a haplotype distance matrix was constructed by use of the number of mutation steps between each pair of Y haplotypes. The hierarchic distribution of Y chromosome diversity, measured as the variance components among individuals, populations, and geographic groups, was computed by use of analysis of molecular variance (AMOVA) software. Genetic distances between populations were calculated, and their significance was tested by use of a permutation procedure (Excoffier et al. 1992). Population pairwise  $F_{ST}$ 's and other genetic distances (Excoffier et al. 1992) were used to draw neighbor-joining and UPGMA trees, with the PHYLIP package (Felsenstein 1993), that were visualized by use of TreeView software (Page 1996).

## Results

#### *Worldwide Distribution of Y Haplotypes*

This study comprised a sample of 306 men from populations encompassing distinct linguistic affiliations (table 1). They are representatives of different geographic areas expected to be informative for the Americas settlement study. The analysis with seven polymorphic systems revealed the variation at 30 distinct loci in the nonrecombining portion of the Y chromosome, which allowed the discrimination of 32 haplotypes among 306 men (table 1). The major Amerindian haplotype (Pena et al. 1995; Underhill et al. 1996) is described here as haplotype 31, which is associated with several markers, such as  $\alpha$ h type II, the alphoid 4.1-kb units, the *DYS19* T allele, the 92R7 *HindIII*– allele, and also the *DYS19* microsatellite A allele (data not shown), as well as with the ancestral states of the polymorphisms YAP and Tat. Haplotype 10, differing from haplotype 31 only by the mutation at *DYS19*, was very frequent (30%) in our Native American sample and was found exclusively among the North American Indians; in addition, it was also observed in a Mongolian and four Indians. Haplotype 20, which is similar to haplotypes 10 and 31, was seen in a single North American Indian and in some populations from the central region of Siberia. It was particularly frequent in a sample of the rapidly disap-

pearing Ket population (70%) and also was found in some Altaians (17.4%) and a single Mongolian. Haplotype 23, which is very different from haplotypes 31, 10, and 20, was seen in a single Na-Dene and could be a more recent migrant haplotype from Asia, since it is most frequent in Mongolia (42%) and also is seen in many Siberians (table 1). Haplotype 1, also similar to haplotype 10 and the most frequent in Europe (53%), is also present in India (14.5%) and was found in 20% of the Native Americans, exclusively in the samples collected for paternity tests in North America, but it is absent from Siberia or central East Asia. European ancestry was confirmed for at least one of these Native American samples with haplotype 1 in the paternity-test report. Therefore, the presence of haplotype 1 in North American Indians can be explained as a result of recent admixture with Europeans, whereas haplotypes 10, 20, and 23 cannot be explained in the same way, because they are absent from Europe.

#### *The Y Haplotype Network*

The 32 different Y haplotypes were connected into a parsimonious network (fig. 1) assuming no recombination, because all markers were located in the Y-specific region. Information on the mutation mechanisms that characterize the variability of the  $\alpha$ h system (Santos et al. 1995a, 1995b, 1996a) was used, as well as some additional information about the known ancestral states for *DYS199* (Underhill et al. 1996), YAP (Hammer and Horai 1995), Tat (Zerjal et al. 1997), and *SRY-1532* (Whitfield et al. 1995) and also the inferred ancestral state for 92R7 (Mathias et al. 1994; Jobling and Tyler-Smith 1995; Jobling et al. 1997). Variability of the 6-kb and 4.1-kb alphoid units was partially associated with the  $\alpha$ h system (Santos et al. 1995b), and most of the Y chromosomes seemed to have both the 6-kb and 4.1-kb units (Mathias et al. 1994; Santos et al. 1995b; F. R. Santos, A. Pandya, and C. Tyler-Smith, unpublished data). The deletion of the 6-kb units can generate chromosomes bearing 4.1-kb divergent units only, and further deletion of the 4.1-kb units generates chromosomes containing no divergent units. Since deletion events that cause the loss of many units of alphoid DNA are relatively frequent (Mathias et al. 1994; Santos et al. 1995a, 1995b, 1996a), we allowed these events to be recurrent on the network (fig. 1).

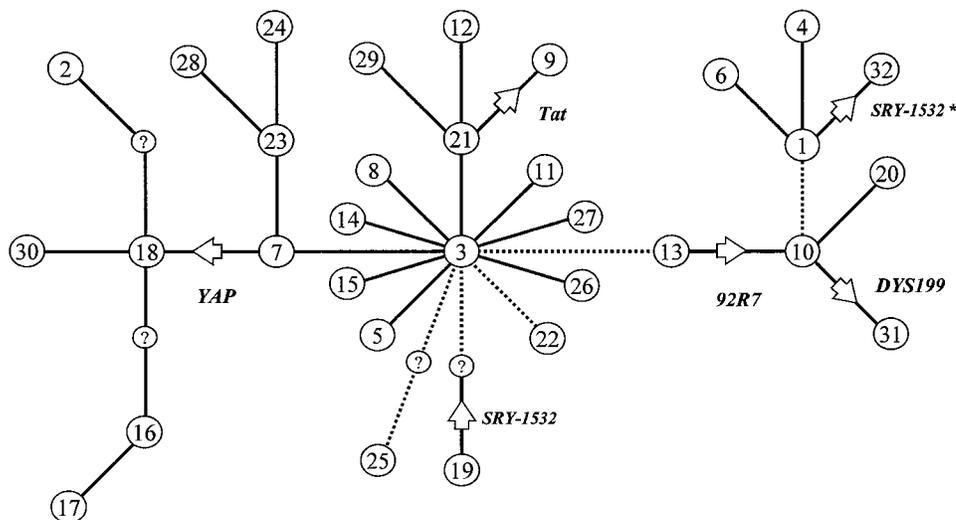
The occurrence of an extra reverse mutation at the *SRY-1532* locus is represented as a recurrent step in this parsimony network, leading to haplotype 32. It is supported by the association with 92R7 alleles, by the specific association with  $\alpha$ h type II, and by analysis with an additional 19 Y markers (F. R. Santos, A. Pandya, and C. Tyler-Smith, unpublished data). In addition, the geographic distribution (table 1) of haplotype 32 is very similar to that of haplotype 1, its deduced immediate

**Table 1**

**Worldwide Distribution of Human Y Chromosome Haplotypes**

HAPLOTYPE	NO. OF INDIVIDUALS, BY POPULATION																	Total (%)	
	92R7	DYS199	SRY-1532	Tat	YAP	αh	pYαl	Native					Central						
								Altai	Buryat	Evenki	Ket	Yakut	American	European	Mongolian	East Asian	Indian		African
1	1	0	1	0	0	II	0						4	28			8	40 (13.1)	
2	0	0	1	0	1	XIII	6							1				1 (.3)	
3	0	0	1	0	0	III	6				1			16	6	39	24	2 88 (28.8)	
4	1	0	1	0	0	VI	0							1				1 (.3)	
5	0	0	1	0	0	VIII	6									1		1 (.3)	
6	1	0	1	0	0	VII	0							1				1 (.3)	
7	0	0	1	0	0	V	6			18					3			21 (6.9)	
8	0	0	1	0	0	IV	6							2	1		1	4 (1.3)	
9	0	0	1	1	0	XX	6		3			5		2	1			11 (3.6)	
10	1	0	1	0	0	II	4.1						6		1		4	11 (3.6)	
11	0	0	1	0	0	XXVI	6										1	1 (.3)	
12	0	0	1	0	0	XXVII	6									4	1	5 (1.6)	
13	0	0	1	0	0	II	4.1				1				2	2	1	6 (2.0)	
14	0	0	1	0	0	I (1 + 5)	6								1	1	1	3 (1.0)	
15	0	0	1	0	0	XXVIII	6										2	2 (.7)	
16	0	0	1	0	1	IX	6											12 (3.9)	
17	0	0	1	0	1	X	6										1	1 (.3)	
18	0	0	1	0	1	V	6							1	1		2	4 (1.3)	
19	0	0	0	0	0	I (K)	0										1	1 (.3)	
20	1	0	1	0	0	I (1)	4.1	4				7		1		1		13 (4.2)	
21	0	0	1	0	0	XX	6	1			2					2		5 (1.6)	
22	0	0	1	0	0	I (L)	0									1		1 (.3)	
23	0	0	1	0	0	XVIII	6	4	1	1				1		19		26 (8.5)	
24	0	0	1	0	0	XIX	6									1		1 (.3)	
25	0	0	1	0	0	0	0									1		1 (.3)	
26	0	0	1	0	0	XI	6									1		1 (.3)	
27	0	0	1	0	0	XXX	6									1		1 (.3)	
28	0	0	1	0	0	XXXI	6									1		1 (.3)	
29	0	0	1	0	0	XXIV	6	1										1 (.3)	
30	0	0	1	0	1	XXV	6	1										1 (.3)	
31	1	1	1	0	0	II	4.1						8					8 (2.6)	
32	1	1	0	0	0	II	0	$\frac{12}{23}$	$\frac{4}{4}$	$\frac{1}{23}$	$\frac{2}{10}$	$\frac{5}{5}$	$\frac{8}{20}$	$\frac{1}{53}$	$\frac{4}{45}$	$\frac{12}{50}$	$\frac{12}{55}$	$\frac{18}{18}$	$\frac{32}{306}$ (10.5)
Total	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
D	...	...	...	...	...	...	...	.661	.375	.374	.460	.000	.705	.625	.783	.382	.732	.525	...

NOTE.—The absolute haplotype frequencies, total number of individuals, and Nei diversity values (*D*) are shown for each population. For the point mutations and YAP, alleles were classified as “0” or “1” according to the presence of an ancestral or derived state, respectively. The αh types were classified as published previously (Santos et al. 1995b, 1996a), and for the pYαl probe (Mathias et al. 1994; Jobling and Tyler-Smith 1995) alleles were classified as “6” (6-kb unit), “4.1” (4.1-kb unit), or “0” (absence of divergent unit, detected by *Hind*III). For the other loci, the alleles classified as “1” correspond to the absence of a *Hind*III site, for 92R7; the T allele, for *DYS199*; the G allele, for *SRY-1532*; the C allele, for *Tat*; or the presence of the *Alu* insert, for YAP.



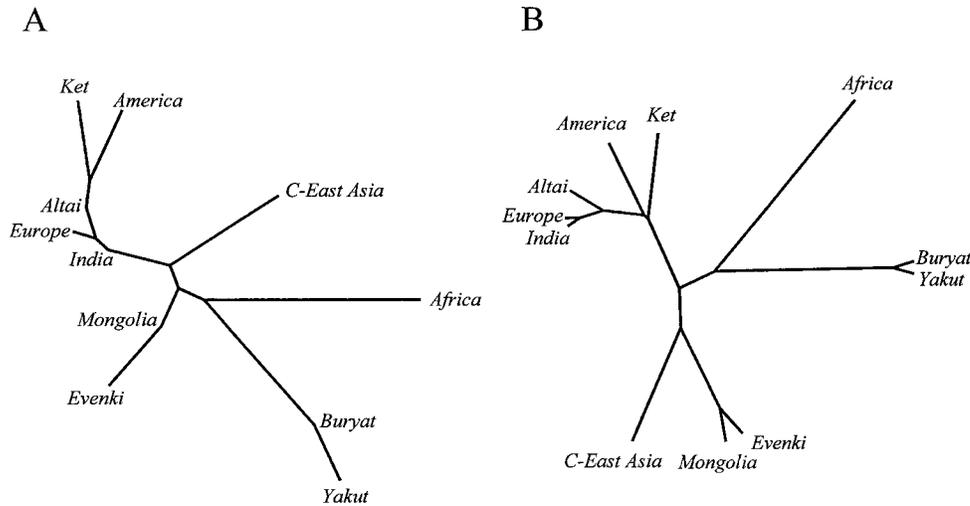
**Figure 1** Network of Y chromosome haplotypes. The haplotypes are connected through mutation steps: solid lines indicate aliphoid mutation events (Santos et al. 1996a), dotted lines indicate aliphoid variation involving the 6-kb and 4.1-kb units, and arrows indicate the change from an ancestral to a derived allele state in each Y locus (see table 1). Some intermediate haplotypes not observed in this study are indicated by a question mark (?). The direction of mutations in the aliphoid variation was not depicted but can be inferred by comparison with the mutations with known ancestral states. This network assumes a reverse mutation (indicated by an asterisk [\*]) of *SRY-1532* in the 92R7 lineage leading to haplotype 32. When haplotype 19 is considered the root, the first chromosome found outside Africa that is ancestral to the lineage leading to haplotype 31 is haplotype 3, a very frequent chromosome (28.8%; table 1) found in all continents, except the Americas. The rare haplotype 13, where the 4.1-kb units first appeared, is derived from haplotype 3. Haplotype 10 is derived from haplotype 13, by a mutation at the 92R7 locus that finally acquired the mutation at the *DYS199* locus, leading to haplotype 31. Haplotype 10 also originated haplotypes 20 and 1, the latter being the ancestor of haplotype 32.

ancestor, and is quite distinct from that of haplotype 19, which shares with haplotype 32 the same allele A at *SRY-1532*. The same network was obtained with the procedure of median network analysis (Bandelt et al. 1995), which allows the resolution of networks containing such recurrent markers. Our unique network (fig. 1) shows the sequential accumulation of mutations used to trace several Y lineages from the ancestral haplotype. The most likely root for this network is haplotype 19, seen only in an African San, because it bears the ancestral states for all the loci for which this information is known or inferred (fig. 1). A recent study with several new biallelic Y markers (Underhill et al. 1997) also supports the conclusion that the San haplotype 19 is the most ancestral Y chromosome. Following from the probable root, haplotypes 3, 13, and 10 are direct ancestors of the Native American haplotype 31, and other related haplotypes, such as haplotypes 20 and 1, share with haplotype 31 the common ancestor haplotype 10. The frequent and widespread haplotype 3 is probably very old, because it gave rise to most, if not all, of the Y chromosomes found outside Africa. Haplotype 13 also is probably old but is very rare, which could indicate that the 92R7 mutation happened quite soon after the origin of haplotype 13, producing haplotype 10. However, a simple deletion of the 6-kb aliphoid units in an

individual with haplotype 3 can lead to a recurrent haplotype 13. By using other Y markers (F. R. Santos, A. Pandya, and C. Tyler-Smith, unpublished data), we found that the two Chinese individuals with haplotype 13 (table 1) are recurrent types, because of a de novo deletion of the 6-kb units. Thus, interpretation of the distribution of the few haplotype 13 individuals should be made with care. Fortunately, haplotype 10, the immediate ancestor of haplotype 31, is defined by the point mutation at 92R7 and, together with all chromosomes derived from it, makes up the 92R7 lineage that is important for the tracing of the major migrant Y chromosome to the Americas.

#### Global Diversity of Y Chromosome Haplotypes

Our study used markers defining many branches of the 92R7 lineage, with a consequent bias toward a better resolution of populations containing this lineage, despite the fact that most variation was found in the  $\alpha h$  polymorphism, with no such apparent bias (Santos et al. 1995b, 1996a). For this reason, the gene diversity (Nei 1987) for each population (table 1) usually was increased when different 92R7-lineage derivatives were present, which also could have increased slightly the within-population variance calculated with AMOVA



**Figure 2** Trees of Y chromosome populations. The population pairwise  $F_{ST}$ 's calculated by AMOVA (Excoffier et al. 1992) were used to draw neighbor-joining (A) and UPGMA (B) trees (Felsenstein 1993; Page 1996). The cluster of Caucasoids/Kets/Altaians/Native Americans is evident in both trees. Evenki Y chromosomes cluster with related Mongolians and central East Asians ("C-East Asia"), but Yakuts and Buryats form another distinct cluster, as expected because of the high frequency of haplotype 9, which contains the Tat mutation (Zerjal et al. 1997).

(discussed below). Therefore, this study was oriented toward a detailed investigation of Y lineages that are interesting with regard to the peopling of the Americas and should not be considered a broad and unbiased description of all worldwide Y lineages.

The genetic structure of this Y chromosome data was analyzed in detail by use of AMOVA (Excoffier et al. 1992), which allows an estimation of the relative distribution of genetic diversity in three hierarchic levels: among individuals, among populations, and among geographic groups. The AMOVA resulted in the values 59%, 25%, and 16%, respectively, for worldwide Y chromosome diversity. The value of 59% of the total Y chromosome variability found among individuals is relatively low, compared with the value of 85% obtained for other autosomal DNA polymorphisms (Barbujani et al. 1997). Thus, the higher degree of interpopulation and geographic diversity (41%) of Y chromosomes observed in this study emphasizes the usefulness of Y chromosome haplotype analysis, to discriminate between populations and to elucidate past male migrations within and across continents (Jobling and Tyler-Smith 1995; Santos and Tyler-Smith 1996).

#### Y Chromosome Population Trees

AMOVA (Excoffier et al. 1992) also generated a matrix of  $F_{ST}$  analogs between populations of Y chromosomes. This procedure avoids the use of allele frequencies to calculate genetic distances, since the loci in the nonrecombining portion of the Y chromosome are not independent. Otherwise, if we used only the haplotype

frequencies (considering the Y chromosome as a single locus, as expected), we would lose the information of shared ancestry. When running AMOVA, we computed the similarity of populations sharing haplotypes related by descent, by taking into account both haplotype frequencies and molecular differences between haplotypes (see Material and Methods). These pairwise  $F_{ST}$ 's computed by AMOVA were used to draw neighbor-joining and UPGMA gene trees (Felsenstein 1993), to display the relationship of Y chromosome populations (fig. 2). Similar trees were obtained with two other distances (data not shown). By use of a nonparametric permutation test (Excoffier et al. 1992), the calculated  $F_{ST}$  distances were shown to be significant ( $P < .05$ ), except between the Buryats and Yakuts, the closest groups on the trees. In all the trees, the Native American Y chromosomes clustered with Kets, Altaians, and Caucasoids (Europeans and Indians). European admixture cannot explain this cluster, because if we exclude in the analysis all haplotypes present in Siberians and Amerindians that are also found in Europe (such as haplotype 1, which appears in four Native Americans), the tree remains very similar (data not shown). This tree structure also did not change when we used the inferred frequency  $\sim 87\%$  for haplotype 31 among 132 Native Americans, published previously (Pena et al. 1995; Santos et al. 1995a; Underhill et al. 1996).

Although some Siberian and Native American Y chromosomes show remarkably close association with Caucasoid Y chromosomes, other Siberian populations are very distinct, clustering with other Asians. A particular

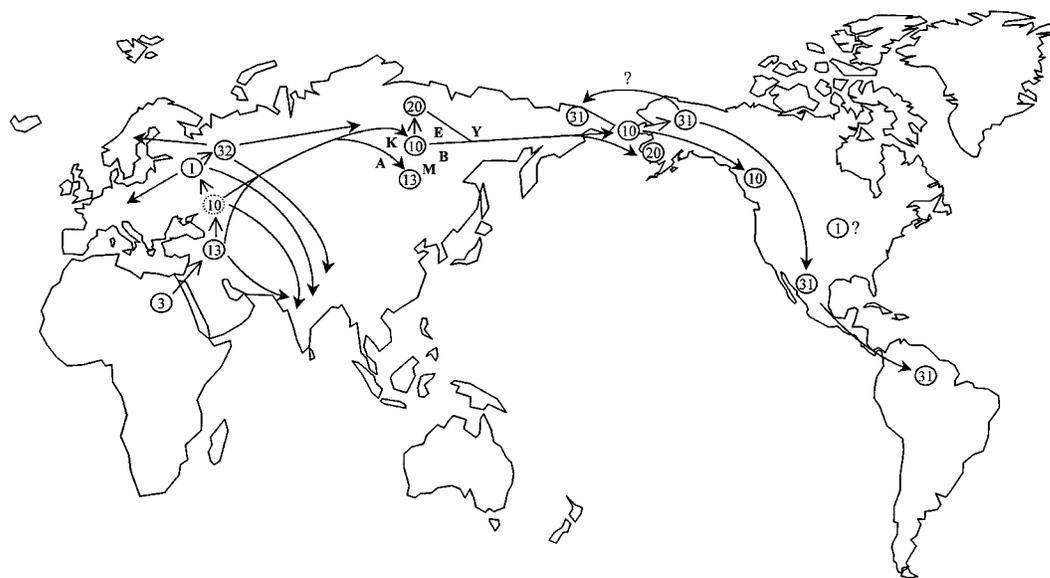
Siberian cluster is formed by Buryat and Yakut Y chromosomes, mainly because of the common origin of most of their Y chromosomes, with the high frequency of the Tat mutation (table 1; Zerjal et al. 1997). Thus, our findings from the haplotype distribution (table 1), the haplotype network (fig. 1), and the Y chromosome population trees (fig. 2) suggest a high interpopulation differentiation in Siberia, probably because of distinct founder populations and subsequent genetic drift. In addition, these data identify the group of Ket and Altaian Y chromosomes that are related to those among Native Americans and Caucasoids, whereas the Evenki Y chromosomes are related to those of Mongolians. Yakut and Buryat Y chromosomes, which previously were shown to have a common origin with Uralic Y chromosomes (Zerjal et al. 1997), form another distinct cluster.

**Discussion**

A human Y chromosome phylogeny was used to trace the origins of the major founder haplotype of the Americas, haplotype 31. The worldwide distribution of Y haplotypes (table 1) associated with the information of sequential mutations, displayed in the network (fig. 1), allowed the construction of a map (fig. 3) showing the

likely pathway of Y chromosomes migrating to the Americas. The present-day distribution of haplotypes related to haplotype 31 can be explained by a radiation from central Eurasia through a northern migration route to the Americas and a southern route to the Indian subcontinent. This dichotomy is supported by the absence of haplotypes 1, 10, 20, and 32 in China and Japan (table 1), including in the analysis of another 56 Chinese and 138 Japanese samples (F. R. Santos and C. Tyler-Smith, unpublished data). A migration route from central Eurasia to northeastern Siberia during the Upper Pleistocene was suggested recently (Lahr and Foley 1994), and the occurrence of several Caucasoid lineages in the Indian subcontinent can be explained by the immigration of Indo-European speakers from central Eurasia after 5,000 ybp (Cavalli-Sforza et al. 1994).

The major Native American haplotype 31 is present on both sides of Beringia, most likely because of an American or Beringian origin of the mutation in the *DYS199* locus (Karafet et al. 1997; Lell et al. 1997). Its immediate ancestor, haplotype 10, is a rare haplotype (11 of 306 individuals) seen only in North America ( $n = 6$ ), India ( $n = 4$ ), and Mongolia ( $n = 1$ ). An old population bearing haplotype 10, a Native American/



**Figure 3** Tracing of Y lineages colonizing the Americas. Thick arrows indicate haplotype displacements, whereas thin arrows indicate haplotype mutations. This world map traces the most likely pathway of the Y lineages leading to the major haplotype of the Americas (haplotype 31). This lineage comprises haplotypes 13 and 10, which are direct ancestors of haplotype 31, and haplotypes 1, 32, and 20, derived from haplotype 10. The sequential accumulation of mutations is evident when moving toward America. The presence of haplotype 10 in the Americas could indicate an origin for the *DYS199* T mutation in that continent, but its presence, as well as that of haplotype 20, likely could represent another migration from Asia. This recent migration could also be responsible for the presence in a Na-Dene of haplotype 23, the most frequent haplotype in Mongolians. Haplotype 10, depicted in central Eurasia (dotted circle), was not observed, but its presence was deduced from the distribution of derived haplotypes in north and south Eurasia. The presence in the Americas of haplotype 1, the most common Caucasoid chromosome, is most probably because of recent admixture. The approximate locations of the sampled populations in northeast Asia are shown on the map. A = Altaian, B = Buryat, E = Evenki, K = Ket, M = Mongolian, and Y = Yakut.

Siberian/Caucasoid common ancestor, has been placed somewhere in central Eurasia (fig. 3). Haplotypes 1 (Caucasoid), 20 (Siberian and Native American), and 31 (Native American) are derived from this ancestor. The most common European chromosome, haplotype 1, appeared in four Native American samples from paternity tests in North America; thus, they very likely could be due to recent admixture. Haplotype 20, another descendant of haplotype 10 by a simple alphoid locus-deletion step (table 1), is very frequent in Kets and was found in some Altaians, all of whom were shown to also have the *DYS19* A allele (data not shown), which is also present in most individuals with haplotype 31 (Pena et al. 1995; Santos et al. 1995a; Underhill et al. 1996). Recently, the Ket language was suggested to be closely related to the Na-Dene language (Greenberg 1996), and the resemblance of Kets to Native Americans and Caucasoids, with regard to physical appearance (Forsyth 1996) and Y chromosomes (this study), makes them the most likely central Siberian population to share the same recent ancestors. The Altaians, a common denomination for seven formerly distinct Turkic populations, exhibit very diverse Y haplotypes (table 1) and could have acquired their Y chromosomes from many neighboring tribes, including the Kets (Forsyth 1996).

Our study can be compared to current research on the peopling of the Americas. Recent archeological and anthropological studies of the first settlement of the Americas are revealing many alternative migration routes and older dates for settlement (Roosevelt et al. 1996), as well as raising doubts about the Mongolian origins of the first migrants (Neves and Pucciarelli 1991; Lahr 1995). The multiple-dispersals model, suggested recently by paleoanthropologists (Lahr and Foley 1994), claims that the first migrants to the Americas were from a Southeast Asian stock, whereas our Y chromosome data suggest a northern Eurasian route of migration. However, they also proposed two distinct dispersals from central Eurasia to northeastern Siberia and Europe, one in the middle Upper Pleistocene (50,000–15,000 ybp) and another in the late Upper Pleistocene (15,000 ybp to present). The former could be the source of the Y chromosomes in those who first migrated to the Americas through Siberia, as well as the source of the Y chromosomes in those colonizing Europe in the Paleolithic.

The very recent find of the 9,400-year-old skeleton of the Kennewick man, which displays some Caucasoid characteristics, and his contemporary, the Spirit Cave mummy, suggests that the earliest migrants could be distinct from present-day populations (Morell 1998b). Possible genetic relationships between Eurasians and Native Americans are suggested by the presence of the rare mtDNA haplogroup X in both population groups, which apparently is absent in Siberia (Morell 1998a). Alternatively, in our study, the Y chromosome data reveal a

common ancestor (haplotype 10) between Native Americans and Europeans, who left some rare descendants in Siberia, among the Kets and Altaians. However, the presence of the most common European haplotype 1 in the Americas can be explained as a recent European admixture more likely than as a remnant of a pre-Columbian migrant. Our Y chromosome data, when compared with morphological and mtDNA data, could imply another migration of typically Mongoloid people, who would have left phenotypic traces in their Native American descendants without contributing many of their Y chromosomes. This pattern of unequal paternal and maternal contributions in the gene pools of several populations has been characterized and discussed in detail by Poloni et al. (1997).

The major Native American Y haplotype occurs in high frequencies among Amerindians, Na-Denes, and Aleut-Eskimos (Pena et al. 1995; Santos et al. 1995a, 1996b; Underhill et al. 1996; Karafet et al. 1997; Lell et al. 1997; Rodriguez-Delfin et al. 1997; Underhill et al. 1997). It represented ~90% of 90 nonadmixed South American Indians in our previous studies (Pena et al. 1995; Santos et al. 1995a) and 60% of 412 Native American Y chromosomes analyzed by other groups (Underhill et al. 1996; Karafet et al. 1997; Lell et al. 1997; Rodriguez-Delfin et al. 1997; Underhill et al. 1997), including Y chromosomes from tribes with a very high level of admixture, especially in North America (Santos et al. 1996b; Crawford 1998). The presence of this founder Y haplotype in the Americas suggests a single major migration and is compatible with a settlement model incorporating a population differentiation of all Native Americans in Beringia, as suggested by recent mtDNA studies (Forster et al. 1996; Bonatto and Salzano 1997). The first migrants bearing a proto-Caucasoid Y chromosome (haplotype 10) would have come from the region of central Siberia to Beringia ~30,000 ybp (Cavalli-Sforza et al. 1994; Underhill et al. 1996). The mutation in the *DYS199* locus, which produced haplotype 31, could have happened in this Pleistocene Beringian population, which would have experienced an expansion and migrated south to the Americas, through the Alberta ice-free corridor. Subsequently, the collapse of this corridor 20,000–14,000 ybp (reviewed in Bonatto and Salzano 1997) would have isolated the population that was still in Beringia from the recent migrants in the Americas, who, after a major founder effect, would give rise to the Amerindians. During this time of isolation, new minor Siberian migrants could have come to Beringia, and, at the end of the glaciation (12,000–10,000 ybp), these Beringians finally could have migrated to the Americas, originating the Na-Dene and Eskimo-Aleut speakers, with both still retaining the major haplotype 31 (Underhill et al. 1996; Karafet et al. 1997; Lell et al. 1997) and other Y chromosome lineages in frequencies

higher than those in Amerindians, exemplified by haplotypes 10, 20, and 23 in North American Indians (table 1). These chromosomes could represent later migrations from central Siberia or Mongolia, despite the possibility that present-day individuals with haplotype 10 could be descendants of the first migrants prior to the acquisition of the *DYS199* mutation. Other scenarios, involving earlier dates (<15,000 ypb), for the first settlement of the Americas are likely, but it is difficult to explain a single major migration with further differentiation for at least three major Native American groups.

This study traces the major Native American Y chromosome haplotype to the immediate ancestor shared with present-day Siberians and to an older common ancestor shared with Caucasoids (Europeans and Indians). This common ancestry of Native Americans and Caucasoids could explain the existence of non-Mongoloid skeletons, such as the Kennewick man. Despite the fact that the Y chromosome represents only 1 of 46 in the human male genome, in numeric terms, its exclusive father-to-son inheritance allows us to study patrilineages that reflect the past male migrations but that may not reflect the global history of populations. However, the Y lineage is the largest of many genomic lineages that compose the population history of modern *Homo sapiens*, and it is the counterpart to mtDNA lineage studies. Furthermore, the human Y chromosome seems to display an association with linguistics and geography that is higher than that for mtDNA (Poloni et al. 1997), and our data concur with some current views on the settlement of the Americas. Further analysis of all Y lineages present in the Americas that uses microsatellites (Santos and Tyler-Smith 1996; Zerjal et al. 1997) will be very useful in the detailed study of all trans-Bering Strait migrating lineages, as well as to the more precise determination of their entry time into the Americas.

## Acknowledgments

We thank D. R. Carvalho-Silva and E. Tarazona-Santos for comments on the manuscript. This work was supported by grants from Conselho Nacional de Desenvolvimento Científico e Tecnológico and Fundação de Amparo à Pesquisa do Estado de Minas Gerais, Brazil, and from the Leverhulme Trust, United Kingdom.

## References

- Bandelt HJ, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations using median networks. *Genetics* 141:743–753
- Barbujani G, Magagni A, Minch E, Cavalli-Sforza LL (1997) An apportionment of human DNA diversity. *Proc Natl Acad Sci USA* 94:4516–4519
- Bonato SL, Salzano FM (1997) 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
- Cavalli-Sforza LL, Menozzi P, Piazza A (1994) *The history and geography of human genes*. Princeton University Press, Princeton, NJ
- Crawford MH (1998) *The origins of Native Americans: evidence from anthropological genetics*. Cambridge University Press, Cambridge
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA data. *Genetics* 131:479–491
- Felsenstein J (1993) *Phylogeny inference package (PHYLIP)*, version 3.5. University of Washington, Seattle
- 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
- Forsyth J (1996) *A history of the peoples of Siberia*. Cambridge University Press, Cambridge
- Greenberg JH (1996) The “Greenberg hypothesis.” *Science* 274:1447
- Greenberg JH, Turner CG, Zegura SL (1986) The settlement of the Americas: a comparison of the linguistic, dental and genetic evidence. *Curr Anthropol* 27:477–498
- Grimes BF [ed] (1996) *Ethnologue: languages of the world*, 13th ed. Summer Institute of Linguistics, Dallas
- Hammer MF, Horai S (1995) Y chromosomal DNA variation and the peopling of Japan. *Am J Hum Genet* 56:951–962
- Jobling MA, Pandya A, Tyler-Smith C (1997) The Y chromosome in forensic analysis and paternity testing. *Int J Legal Med* 110:118–124
- Jobling MA, Tyler-Smith C (1995) Fathers and sons: the Y chromosome and human evolution. *Trends Genet* 11:449–456
- Karafet T, Zegura SL, Vuturo-Brady J, Posukh O, Osipova L, Wiebe 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
- Lahr MM (1995) Patterns of modern human diversification: implications for Amerindian origins. *Yearbook Phys Anthropol* 38:163–198
- Lahr MM, Foley R (1994) Multiple dispersals and modern human origins. *Evol Anthropol* 3:48–60
- 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
- Mathias N, Bayés M, Tyler-Smith C (1994) Highly informative compound haplotypes for the human Y chromosome. *Hum Mol Genet* 3:115–124
- Merriwether DA, Hall WW, Vahlne A, Ferrell RE (1996) mtDNA variation indicates Mongolia may have been the source for the founding population for the New World. *Am J Hum Genet* 59:204–212
- Merriwether DA, Rothhammer F, Ferrell RE (1995) Distri-

- bution 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 (1998a) Genes may link ancient Eurasians, Native Americans. *Science* 280:520
- (1998b) Kennewick man's trials continue. *Science* 280:190–192
- Neel JV, Biggar RJ, Sukernik RI (1994) Virologic and genetic studies relate Amerind origins to the indigenous people of the Mongolia/Manchuria/southeastern Siberia region. *Proc Natl Acad Sci USA* 91:10737–10741
- Nei M (1987) *Molecular evolutionary genetics*. Columbia University Press, New York
- Neves WA, Pucciarelli HM (1991) Morphological affinities of the first Americans: an exploratory analysis based on early South-American human remains. *J Hum Evol* 21:261–273
- Page RDM (1996) TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* 12:357–358
- Pena SDJ, Santos FR, Bianchi N, Bravi CM, Carnese FR, Rothhammer F, Gerelsaikhan T, et al (1995) Identification of a major founder Y-chromosome haplotype in Amerindians. *Nat Genet* 11:15–16
- Poloni ES, Semino O, Passarino G, Santachiara-Benerecetti AS, Dupanloup I, Langaney A, Excoffier L (1997) Human genetic affinities for Y-chromosome P49a,f/*TaqI* haplotypes show strong correspondence with linguistics. *Am J Hum Genet* 61:1015–1035
- Rodriguez-Delfin L, Santos SE, Zago MA (1997) Diversity of the human Y chromosome of South American Amerindians: a comparison with blacks, whites, and Japanese from Brazil. *Ann Hum Genet* 61:439–448
- Roosevelt AC, Lima-da-Costa M, Machado CL, Michab M, Mercier N, Valladas H, Feathers J, et al (1996) Paleoindian cave dwellers in the Amazon: the peopling of the Americas. *Science* 272:373–384
- Santos FR, Bianchi NO, Pena SDJ (1996a) Worldwide distribution of human Y chromosome haplotypes. *Genome Res* 6:601–611
- Santos FR, Hutz M, Coimbra CEA, Santos RV, Salzano FM, Pena SDJ (1995a) Further evidence for the existence of a major founder Y chromosome haplotype in Amerindians. *Braz J Genet* 18:669–672
- Santos FR, Pena SDJ, Tyler-Smith C (1995b) PCR haplotypes for the human Y chromosome based on alphoid satellite variants and heteroduplex analysis. *Gene* 165:191–198
- Santos FR, Rodriguez-Delfin L, Pena SDJ, Moore J, Weiss KM (1996b) North and south Amerindians may have the same major founder Y chromosome haplotype. *Am J Hum Genet* 58:1369–1370
- Santos FR, Tyler-Smith C (1996) Reading the human Y chromosome: the emerging DNA markers and human genetic history. *Braz J Genet* 19:665–670
- Schanfield MS (1992) Immunoglobulin allotypes (GM and KM) indicate multiple founding populations of Native Americans: at least four migrations to the New World. *Hum Biol* 64:381–402
- Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, et al (1993) Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am J Hum Genet* 53:563–590
- Underhill PA, Jin L, Lin AA, Mehdi Q, 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, Zeman 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
- Waters MR, Forman SL, Pierson JM (1997) Diring Yuriakh: a Lower Paleolithic site in central Siberia. *Science* 275:1281–1284
- 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, Schiefenhover W, et al (1997) Genetic relationships of Asians and northern Europeans revealed by Y-chromosomal DNA analysis. *Am J Hum Genet* 60:1174–1183