

## Microsatellites provide evidence for Y chromosome diversity among the founders of the New World

ANDRÉS RUIZ-LINARES\*<sup>‡</sup>, DANIEL ORTÍZ-BARRIENTOS\*<sup>‡</sup>, MAURICIO FIGUEROA\*, NATALIA MESA\*<sup>‡</sup>, JUAN G. MÚNERA<sup>§</sup>, GABRIEL BEDOYA\*, IVÁN D. VÉLEZ\*, LUIS F. GARCÍA\*, ANNA PÉREZ-LEZAUN<sup>¶</sup>, JAUME BERTRANPETIT<sup>¶</sup>, MARCUS W. FELDMAN<sup>||</sup>, AND DAVID B. GOLDSTEIN\*\*<sup>†</sup>

\*Facultad de Medicina and <sup>§</sup>Departamento de Antropología, Universidad de Antioquia, Apartado Aereo 1226 Medellín, Colombia; <sup>‡</sup>Department of Community Genetics, Imperial College School of Medicine, Northwick Park Hospital, Harrow HA1 3UJ, United Kingdom; <sup>¶</sup>Unitat d'Antropologia, Facultat de Biologia, Universitat de Barcelona, 08028 Barcelona, Spain; <sup>||</sup>Department of Biology, Stanford University, Stanford, California 94305; \*\*Department of Zoology, University of Oxford, Oxford OX1 3PS, United Kingdom; and <sup>†</sup>Department of Biology, University College, London NW1 2HE, England

Communicated by L. L. Cavalli-Sforza, Stanford University School of Medicine, Stanford, CA, March 1, 1999 (received for review August 18, 1998)

**ABSTRACT** Recently, Y chromosome markers have begun to be used to study Native American origins. Available data have been interpreted as indicating that the colonizers of the New World carried a single founder haplotype. However, these early studies have been based on a few, mostly complex polymorphisms of insufficient resolution to determine whether observed diversity stems from admixture or diversity among the colonizers. Because the interpretation of Y chromosomal variation in the New World depends on founding diversity, it is important to develop marker systems with finer resolution. Here we evaluate the hypothesis of a single-founder Y haplotype for Amerinds by using 11 Y-specific markers in five Colombian Amerind populations. Two of these markers (DYS271, DYS287) are reliable indicators of admixture and detected three non-Amerind chromosomes in our sample. Two other markers (DYS199, M19) are single-nucleotide polymorphisms mostly restricted to Native Americans. The relatedness of chromosomes defined by these two markers was evaluated by constructing haplotypes with seven microsatellite loci (DYS388 to 394). The microsatellite backgrounds found on the two haplogroups defined by marker DYS199 demonstrate the existence of at least two Amerind founder haplotypes, one of them (carrying allele DYS199 T) largely restricted to Native Americans. The estimated age and distribution of these haplogroups places them among the founders of the New World.

Various aspects of the peopling of the New World engender considerable controversy among archaeologists and geneticists. Although there is agreement that the initial peopling of the American continent involved migration through Beringia from Asia into America, there is considerable debate as to the details of this process (1–3). Particularly contentious are the migratory pattern, its precise timing, and the identity of the founding populations. A model proposed by Greenberg *et al.* (4) posits three migration waves giving rise to the three major Native American linguistic groups into which Greenberg classifies existing languages: Amerind, Na-Dene, and Eskimo-Aleut (4, 5). Most archaeological evidence points to an initial migration around 12,000 years ago (1–3), and this has been taken by Greenberg to be the time of the proto-Amerind migration. More recently, this date has been questioned because seemingly older archaeological sites have been found, particularly in South America. However, there is disagreement among archaeologists as to the reliability of the dating of these pre-Clovis sites (1–3).

Although genetic approaches to the population history of Native Americans have mostly used either autosomal or mtDNA markers, a number of recent analyses have begun to exploit Y

chromosome-specific variation (6–12). These studies have shown a consistent pattern of scarce haplotypic diversity in Native American populations. An initial report found high frequencies of “allele” 18 (a complex restriction fragment length polymorphism pattern defined at locus DYS1 by probe 49a/f) in populations from North and South America (6). Later, a single predominant Y haplotype was detected in North and South Amerinds by combining heteroduplex analysis of sequence variants in repetitive aliphoid subunits ( $\alpha$ h “allele” II) and the tetranucleotide repeat locus DYS394 (allele 186) (7, 8). More recently, Underhill *et al.* (9) detected in Native Americans of the three major linguistic groups a C-to-T nucleotide polymorphism at locus DYS199 for which the T allele usually showed high frequencies. These various polymorphisms have been shown to be associated (10). Thus, alleles DYS1 18;  $\alpha$ h II; DYS394 186; and DYS199 T could define a major, perhaps single, founder haplotype for all Native American populations (11).

Available data indicate that most of the constituent alleles of the putative single Y founder haplotype are either absent or seen at very low frequencies in Asia. Of particular interest, because it is most likely a single event polymorphism, the DYS199 T allele has been detected only in Chuckchan populations of extreme northeastern Siberia (11, 12). The presence of the T allele in these populations could imply that they inherited it from an ancestral population shared with Native Americans. Alternatively, the T allele might be of much more recent New World origin, with migration being responsible for the presence of this polymorphism across the three Native American linguistic groups as well as in the Siberian populations.

Clearly, a more extensive characterization of Y chromosome markers in Native American and Asian populations is needed to further evaluate the number of founding haplogroups, their time of entry into the New World, and their place of origin in Asia. The study of these issues is complicated by the sometimes-extensive admixture of contemporary Amerindian populations. As suggested by Underhill *et al.* (9), an extreme possibility is that all DYS199 C chromosomes currently seen in Native American individuals have been recently introduced by immigrants. This hypothesis can be tested by comparing the haplotypic background of Amerindian and non-Amerindian DYS199 C chromosomes. Additional data on well defined polymorphisms should also allow refined estimation of the age of the C-to-T transition at locus DYS199 and a better definition of founder haplotypes, thus providing a more detailed picture of the founding population(s) to be contrasted with putative relatives in Asia.

Here we report results for 11 such Y chromosome-specific polymorphic markers in five Colombian Amerind populations. When compared with non-Amerind Y chromosomes, our data clearly show that most DYS199 C chromosomes seen in Colombian Amerinds are autochthonous, suggesting that at least two

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

PNAS is available online at [www.pnas.org](http://www.pnas.org).

<sup>‡</sup>To whom reprint requests should be addressed. e-mail: [andres.ruiz@ic.ac.uk](mailto:andres.ruiz@ic.ac.uk).

Table 1. Percentage allele frequencies at four biallelic loci in five Colombian Amerind populations

Population	Locus-allele							
	DYS287 Yap <sup>+</sup>	<i>n</i>	DYS271 G	<i>n</i>	DYS199 T	<i>n</i>	M19A	<i>n</i>
Embera	0	8	0	8	57	7	ND	—
Ingano	11	9	0	9	10	10	0	10
Ticuna	0	37	0	37	77	35	59	35
Wayuu	6	18	6	18	48	21	10	20
Zenu	2	50	2	50	38	56	0	45

founder haplotype existed among the initial Amerind settlers. Assuming a single-founder haplotype each for the DYS199 T and DYS199 C lineages, coalescence times were estimated based on the observed microsatellite diversity. Our data indicate that the DYS199T lineage is about 9,000–11,000 years old. The age of the DYS199 C lineage is more difficult to establish but probably lies in the range of 5,000–18,000 years. One of the polymorphisms examined, M19 A, has so far been detected only in two South American populations. However, we detected considerable haplotypic diversity associated with this marker, suggesting that it is likely to have a wider population distribution in the region.

### MATERIALS AND METHODS

**Populations Studied.** The total number of unrelated male samples available for typing was 137 from five Colombian Amerind populations: 8 Embera, 10 Ingano, 40 Ticuna, 21 Wayuu (or Goajiro), 58 Zenu (or Sinu). The samples for the Zenu and Embera populations were available at Universidad de Antioquia and had been collected for other studies. In three instances (Ticuna, Wayuu, and Ingano), samples were collected from informed consenting individuals.

**Experimental Procedures.** We examined 11 previously reported markers from the nonrecombining portion of the Y chromosome. Not all markers could be examined in every sample due mostly to limited DNA availability. Seven of the markers typed correspond to microsatellite loci [five are tetranucleotide repeats: DYS389, DYS 390, DYS 391, DYS 393, DYS394 (also known as DYS19); two are trinucleotide repeats: DYS388, DYS392] and were typed radioactively (13). Locus DYS389 is duplicated, and we scored only the smaller allele detected. Locus DYS287 is an Alu indel and was PCR-typed as described by Hammer and Horai (14). The A-to-G polymorphism at DYS271 was typed by using a restriction assay as described by Seielstad (15). The polymorphic C-to-T transition at locus DYS199 was typed by allele-specific amplification (9).

To detect the T-to-A transversion at marker M19 (16), an allele-specific amplification assay was developed. The primers used were: M19-R, 5'-TGAACCTACAAATGTGAAACT-3'; M19-FA, 5'-TATTTTTGTGAAGACTGTTGTAA-3'; M19-FT 5'-TATTTTTGTGAAGACTGTTGTAT-3'.

Two PCR reactions were performed for each individual by using the reverse primer and one of the forward allele-specific primers. Each reaction contained 20–100 ng of genomic DNA, 12.5 pmol of each primer, 0.2 mM each dNTP, 50 mM KCl, 10 mM Tris-HCl, pH 9.0 (at 25°C), 0.1% Triton X-100, 1.5 mM MgCl<sub>2</sub>, 10% DMSO, and 1 unit of *Taq* DNA polymerase in a 25- $\mu$ l volume. The mixture was subjected to 30 cycles of: 94°C for 60 s, 48°C for 60 s, and 72°C for 60 s. After amplification, the PCR products were visualized on 2% agarose gels.

**Randomization Test.** Geographic clustering of the haplotype neighbor-joining tree (17) was evaluated by using a randomization test (18). First, the length of the tree obtained from the data was evaluated by using Wagner parsimony. This assumes equal likelihood for changes in binary character states, taking as character state the geographic origin of the sample. Next, geographic labels (continent of origin of sample) were randomized, and the number of character state changes necessary to obtain the tree was counted again. This procedure was repeated 1,000 times.

Tree reconstruction and parsimony analyses used programs in the PHYLIP (Version 3.5) package (19).

### RESULTS

Table 1 summarizes the results for the four biallelic markers typed. Two of the markers examined (DYS287 and DYS271) are indicators of admixture. Allele G at locus DYS271 is seen only in African populations and is found only in chromosomes bearing the Alu insertion (Yap<sup>+</sup>) at locus DYS287 (15). This Alu

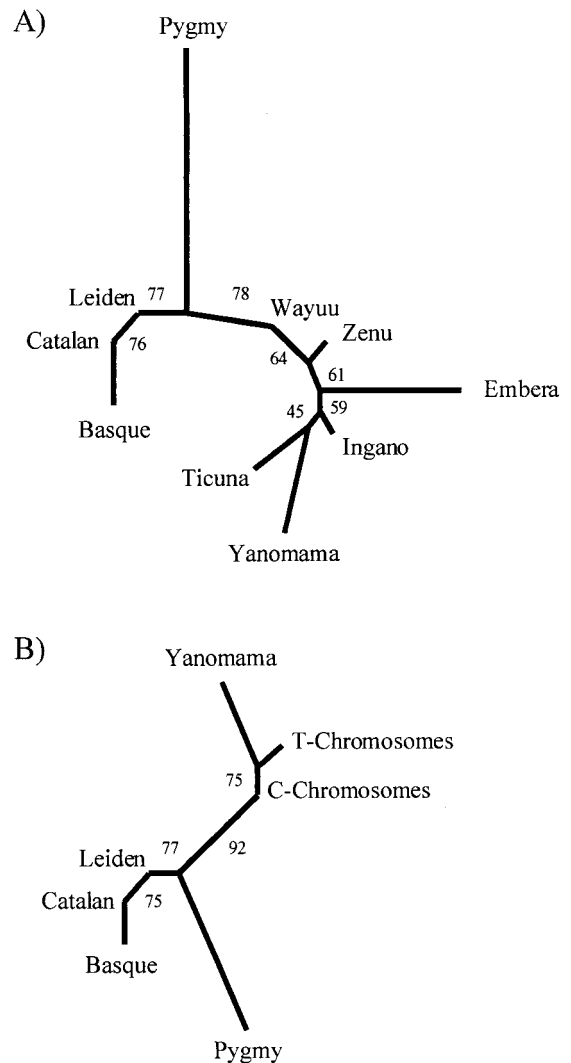


FIG. 1. Neighbor-joining trees (17) relating the five Colombian Amerind populations examined (Embera, Ingano, Ticuna, Wayuu, Zenu) (A) or Y chromosomes bearing a T or a C at locus DYS199 (B) to the Yanomama, Dutch, Basque, Catalan, and Pygmy populations. The trees are based on distance matrices calculated by using Nei's standard genetic distance (21) from allele frequency data for six Y microsatellite loci [data for the non-Colombian populations is from De Kniff *et al.* (22), which does not include information for locus DYS388]. Numbers indicate % bootstrap values for 1,000 bootstraps.

insertion is seen outside of Africa but seems to be characteristic of non-Amerindian populations (20). In our sample, two individuals (one Zenu and one Wayuu) have Yap<sup>+</sup>, DYS271 G Y chromosomes and are thus African in origin. One Ingano chromosome has the Yap<sup>+</sup>, DYS271 A haplotype and is most likely of non-Amerind origin. These three samples were eliminated from most subsequent analyses.

In the populations examined, the frequency of the DYS199 T allele ranged from 10 to 77%, with a frequency in the total sample of 50% (67/133). The M19 A allele, found only on chromosomes bearing the T allele at locus DYS199, occurred at a frequency of 59% among the Amazonian Ticuna and 10% among the Wayuu of the Guajira peninsula (in the extreme north of Colombia). In what follows mention of C and T chromosomes/haplotypes refers to the allele present at locus DYS199 in the Amerind samples examined.

Based on allele frequencies at six Y microsatellite loci, phylogenetic trees relating the Colombian Amerind populations to Yanomama, Basque, Catalan, Dutch, and Pygmy were inferred (Fig. 1A). All Colombian Amerinds group together and with the Yanomama Amerind population in a cluster separate from Pygmies and all Europeans (which are themselves in a separate

cluster). The topology of the tree in Fig. 1A suggests that even in those populations where the DYS199 C allele is predominant (Ingano, Wayuu, and Zenu), most of these chromosomes are not the result of recent admixture with non-Amerindian populations. However, the shortness of the branches leading to these three populations is suggestive of them having some degree of admixture, as confirmed by the presence in them of Yap<sup>+</sup> chromosomes (Table 1).

A comparison of microsatellite allele frequencies between C and T chromosomes irrespective of their population of origin shows a considerable similarity between them (Table 2). Four loci show different modal alleles between T chromosomes and Spanish and Basque chromosomes, of which three have the same modal alleles in C and T chromosomes. A phylogenetic tree depicting the relationship of C and T chromosomes to other world populations, based on microsatellite allele frequencies, is shown in Fig. 1B. Both types of chromosomes cluster together and with the Yanomama Amerind population, confirming that most of the C bearing chromosomes detected are more closely related to the Amerind T bearing chromosomes than they are to non-Amerind chromosomes. Nevertheless, the shortness of the branch leading to C chromosomes is, as above, suggestive of some admixture.

Table 2. Allele frequencies at seven Y microsatellite loci for Basque, Catalan, DYS-199 C/T, and M19A Colombian Amerind chromosomes

Locus	Allele frequency							<i>n</i>
DYS388	126	129	132	135	138	141	144	
Basque	0.000	<b>0.870</b>	0.074	0.000	0.000	0.000	0.056	54
Catalan	0.034	<b>0.862</b>	0.034	0.034	0.034	0.000	0.000	29
DYS199 C	0.000	<b>0.633</b>	0.233	0.017	0.017	0.033	0.067	60
DYS199 T	0.000	<b>0.714</b>	0.268	0.000	0.000	0.000	0.018	56
M19 A	0.000	<b>0.800</b>	0.200	0.000	0.000	0.000	0.000	25
DYS389	247	251	255	259	263			
Basque	0.143	<b>0.536</b>	0.304	0.000	0.018			56
Catalan	0.121	<b>0.727</b>	0.152	0.000	0.000			33
DYS199 C	0.196	<b>0.510</b>	0.294	0.000	0.000			51
DYS199 T	0.179	<b>0.464</b>	0.286	0.071	0.000			56
M19 A	0.083	<b>0.542</b>	0.208	0.167	0.000			24
DYS390	203	207	211	215	219	223		
Basque	0.000	0.000	0.167	<b>0.778</b>	0.056	0.000		54
Catalan	0.000	0.069	0.172	<b>0.690</b>	0.069	0.000		29
DYS199 C	0.019	0.115	0.269	<b>0.346</b>	0.231	0.019		52
DYS199 T	0.029	0.000	0.118	0.235	<b>0.618</b>	0.000		34
M19 A	0.000	0.000	0.087	0.130	<b>0.783</b>	0.000		23
DYS391	279	283	287	291				
Basque	0.037	0.389	<b>0.537</b>	0.037				54
Catalan	0.033	0.400	<b>0.567</b>	0.000				30
DYS199 C	0.036	<b>0.786</b>	0.179	0.000				28
DYS199 T	0.000	<b>0.789</b>	0.211	0.000				38
M19 A	0.000	<b>0.905</b>	0.095	0.000				21
DYS392	248	251	254	257	260			
Basque	0.264	0.000	<b>0.736</b>	0.000	0.000			53
Catalan	<b>0.545</b>	0.030	0.394	0.030	0.000			33
DYS199 C	0.182	0.091	0.091	<b>0.432</b>	0.205			44
DYS199 T	0.170	0.021	0.000	<b>0.426</b>	0.383			47
M19 A	0.042	0.000	0.000	0.417	<b>0.542</b>			24
DYS393	120	124	128	132	136			
Basque	0.071	<b>0.857</b>	0.071	0.000	0.000			56
Catalan	0.061	<b>0.818</b>	0.121	0.000	0.000			33
DYS199 C	0.107	<b>0.554</b>	0.179	0.161	0.000			56
DYS199 T	0.060	<b>0.580</b>	0.180	0.140	0.040			50
M19 A	0.125	<b>0.583</b>	0.250	0.000	0.042			24
DYS394	182	186	190	194	198	202		
Basque	0.000	0.075	<b>0.792</b>	0.094	0.000	0.038		53
Catalan	0.000	0.067	<b>0.767</b>	0.167	0.000	0.000		30
DYS199 C	0.017	<b>0.525</b>	0.271	0.169	0.017	0.000		59
DYS199 T	0.019	<b>0.750</b>	0.154	0.077	0.000	0.000		52
M19 A	0.000	<b>0.826</b>	0.087	0.087	0.000	0.000		23

Bold type indicates the most frequent allele.



The structure of the trees of Fig. 1 agrees with allele frequency heterogeneity analyses that indicate that the sampled Amerind C chromosomes have significantly different microsatellite allele frequencies when compared with European chromosomes (Fisher's exact test  $P < 0.05$ ) and have lower  $F_{ST}$  values when compared with T chromosomes (0.03) than with Spanish chromosomes (0.18).

The relationship between C and T chromosomes and Y chromosomes of non-Amerind populations was studied further by examining microsatellite haplotypes. Among the 97 Native American samples typed at five or more microsatellite loci, 73 different haplotypes were observed, including 59 singletons. Of the 14 microsatellite haplotypes seen more than once, 10 occurred in more than one haplotypic background (defined by markers DYS199 and M19). Six haplotypes were shared between different Amerind populations and one seven-locus haplotype was identical between an Amerind, a Basque, and a Catalan.

A tree depicting the phylogenetic relationship among Amerind, Basque, Catalan, and available East Asian Y microsatellite haplotypes is shown in Fig. 2. The tree shows a geographic clustering pattern corresponding to each continental population. Also, the Asian chromosomes appear somewhat more closely related to Amerind chromosomes than to European chromosomes. Noticeably, most of the C haplotypes are seen in a cluster with the T haplotypes, suggesting a close evolutionary relationship between most C chromosomes and the T haplotypes. The continental clustering pattern of this tree is highly statistically significant. In 1,000 randomizations of the geographic labels, the minimum number of character changes (i.e., geographic origin) in the tree shown in Fig. 2 is 57. The observed number of character changes is 32 ( $P < 0.001$ ).

Ages (coalescent times) of DYS199 C, DYS199 T, and M19 A haplogroups were estimated in two ways. The first approach requires identification of the ancestral haplotype within each haplogroup. To estimate the age of haplogroups we used the ASD distance, defined as the average (across loci) of the squared difference in microsatellite repeat numbers between two haplotypes (24, 25). Under the strict stepwise mutation model, the expected value of ASD calculated between the ancestral haplotype and its descendants is  $\mu\tau$  ( $\mu$  = mutation rate;  $\tau$  = genera-

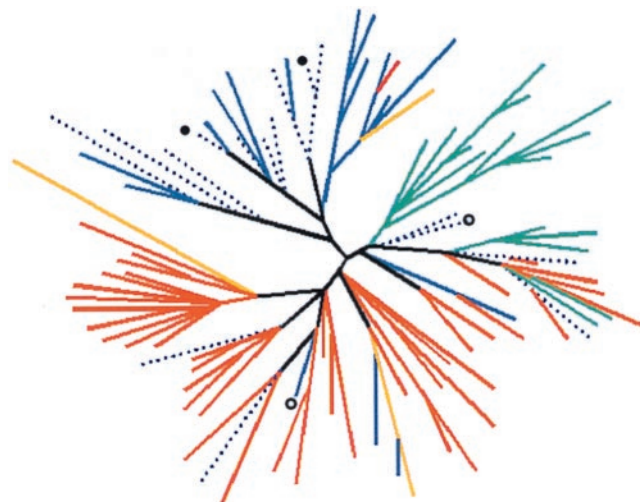


FIG. 2. Neighbor-joining tree (17) relating six locus Y microsatellite haplotypes found in Amerinds, Europeans (Basque and Catalan), and East Asians (mostly Yakut from Siberia and Buryat from Mongolia; ref. 23). Amerinds are shown in blue (C chromosomes with dotted lines), Europeans in red, Asians in green, and (Yap<sup>+</sup>) chromosomes in yellow. Black dots indicate two microsatellite haplotypes shared by DYS199T and C chromosomes. Empty dots indicate haplotypes shared between Amerinds and Europeans. The tree is based on allele-sharing distances between haplotypes (24). No data for locus DYS388 is available in non-Amerinds.

tions) (26). The ages of the DYS199 T, DYS199 C, and M19 A lineages can then be estimated from the mean ASD between the putative ancestral haplotype of each lineage and all observed DYS199 T, DYS199 C, or M19 A chromosomes, respectively. Considering that the time elapsed has been sufficient to introduce a considerable number of mutations in founder chromosomes, we cannot use the individual haplotype frequencies to identify the ancestral founder haplotype [as done by Thomas *et al.* (26) for a much shorter time scale]. We identify the ancestral haplotype by using the constituent microsatellite allele frequencies, haplotype relationships, and geographic distribution. Among the DYS199 T chromosomes, only haplotypes 3, 4, and 8 are observed in more than one population (Table 3). Haplotype 4 falls in the center of a minimum spanning network (27) of T chromosomes (data not shown) and its constituent alleles are modal (Table 2). Similarly, haplotype 3 falls within the center of a minimum spanning network of haplotypes carrying the A allele at locus M19 (data not shown), and each of its constituent alleles is modal among M19 A chromosomes (Table 2). For these reasons, we identify haplotypes 4 and 3 as the ancestors for haplogroups DYS199 T and M19 A, respectively. Because we are interested in the age of founder haplogroups in the New World, in the case of C chromosomes we restrict attention to those haplotypes included in the Amerind cluster seen in the tree of Fig. 2 (these haplotypes are shown in Table 3). For this cluster, haplotype 3 has modal alleles at all loci and falls in the interior of a minimum spanning network (data not shown) and is therefore identified as ancestral. The mean ASD between each of the putative ancestral haplotypes and the observed haplotypes in its lineage produces estimates of  $\tau$  for each lineage. Notice that the estimates of coalescence times for lineages carrying DYS199 T and M19 A are not estimates of the age of the C-to-T and T-to-A mutations themselves, which must predate these coalescence times.

An alternative method of estimating the coalescent time of a lineage assumes a single founder ancestral haplotype and rapid population growth after entry into the New World or appearance of the novel mutation. In this case, we may use the result  $V = \mu\tau$ , in which V is the variance of repeat scores observed (averaged over loci) within a lineage,  $\mu$  is the stepwise mutation rate, and  $\tau$  is the coalescent time for a single-step mutation model (25). In our case, this variance was calculated separately by using all available DYS199 T or M19 A chromosomes or the DYS199C haplotypes shown in Table 3.

Assuming an STR mutation rate of  $2.1 \times 10^{-3}$  (28, 29) and a generation time of 27 years (30), estimated coalescent times based on ASD and the putative ancestral haplotype are 5,657 (3,672–8,181), 11,456 (9,423–13,797), and 7,251 (5,292–10,017) for chromosomes carrying alleles DYS199 C, DYS199 T, and M19 A, respectively (95% confidence interval calculated as in ref. 26). By using the variance approach the numbers are 5,233, 9,334, and 6,094. We note that the age calculated for the DYS 199-C lineage is underestimated by the exclusion of those haplotypes that do not cluster with T chromosomes. For comparison, consideration of all the Amerind C chromosomes detected (excluding the clearly admixed types) leads to ASD and variance age estimates of 18,642 and 13,114 years, respectively.

## DISCUSSION

The identification of founder haplotypes is a key element in the use of genetic data to explore the peopling of the New World. Properties of the distributions of these haplotypes should help evaluate the number of migratory waves and their time of occurrence, and could pinpoint the places in Asia where the migrant populations might have originated. Four major founding Native American mtDNA haplotypes have been identified with a widespread geographic distribution in the American continent (31–33). Recent diversity analyses of these mtDNA haplogroups agree with a single migration scenario, as similar age estimates were obtained across haplogroups and across the three main linguistic families (34, 35).

Table 3. Counts of DYS199 T and C haplotypes typed at five or more microsatellite loci

Haplotype	DYS388	DYS389	DYS390	DYS391	DYS392	DYS393	DYS394	<i>n</i>
DYS199 T								
1	129	255	215	283	260	124	186	4
2	132	255	211	283	257	—	186	3
<b>3</b>	<b>129</b>	<b>251</b>	<b>219</b>	<b>283</b>	<b>260</b>	<b>124</b>	<b>186</b>	<b>3</b>
<b>4</b>	<b>129</b>	<b>251</b>	<b>219</b>	<b>283</b>	<b>257</b>	<b>124</b>	<b>186</b>	<b>2</b>
5	132	251	219	283	257	124	186	2
6	129	255	219	283	260	124	186	2
7	129	259	219	287	260	124	186	2
8	129	247	219	—	257	124	186	2
9	129	259	—	283	260	124	186	2
10	129	251	219	287	248	124	186	1
11	132	251	215	283	248	124	186	1
12	129	255	215	283	257	120	186	1
13	129	251	219	283	257	136	190	1
14	132	251	219	283	257	128	194	1
15	132	251	219	283	257	128	190	1
16	129	251	219	283	260	120	186	1
17	129	251	219	287	260	124	186	1
18	129	255	219	283	257	124	182	1
19	129	259	219	283	257	124	186	1
20	129	259	219	283	260	128	186	1
21	129	255	211	283	257	124	186	1
22	129	247	215	283	257	120	186	1
23	129	251	219	283	248	128	194	1
24	132	255	215	287	248	—	186	1
25	129	251	215	287	248	—	186	1
26	129	251	—	283	257	128	186	1
27	132	251	219	—	257	132	186	1
28	129	251	—	283	257	132	186	1
29	132	251	219	—	257	128	186	1
30	132	247	219	283	257	—	186	1
31	129	251	219	—	260	128	—	1
32	132	255	219	—	—	132	190	1
DYS199 C								
1	129	255	215	283	257	124	186	2
2	132	255	223	287	257	120	186	1
<b>3</b>	<b>129</b>	<b>251</b>	<b>215</b>	<b>283</b>	<b>257</b>	<b>124</b>	<b>186</b>	<b>2</b>
4	129	255	215	283	260	124	186	1
5	129	255	215	283	260	128	186	1
6	132	251	219	283	257	124	186	1
7	129	255	219	283	260	124	194	1
8	129	251	207	283	257	124	186	1
9	129	251	211	283	257	120	194	1
10	129	247	211	283	260	120	186	1

DYS199 T haplotype 3 is shared between Wayuu and Zenu, haplotype 4 between Wayuu and Ticuna, and haplotype 8 between Ticuna and Zenu. DYS199 C haplotypes shown are only those that cluster with DYS199 T haplotypes in the tree of Fig. 2. Putative ancestral haplotypes are shown in **bold**. Missing data is indicated by —.

Previous Y chromosome studies of Native Americans have indicated the possibility that a single founding Y haplotype (carrying the DYS199 T allele) existed among the initial settlers of the New World, DYS199 C chromosomes present in contemporary populations possibly representing recent admixture (9, 11). However, the few markers typed allowed neither precise testing of the admixture hypothesis nor a more detailed analysis of the genetic diversity of the putative single-founder haplotype.

Recent work has demonstrated the value for human population studies of examining genetic variation in Y chromosome haplotypes including slowly evolving markers and rapidly evolving microsatellites (23, 26, 36). The slowly evolving markers allow one to define particular haplotype sets whose diversity can then be evaluated with microsatellite markers. In the case of Native Americans, the single-nucleotide polymorphism at locus DYS199 enabled us to identify the putative previously described founder haplotype. Microsatellites have allowed us to evaluate the diver-

sity of DYS199 T and C chromosomes to examine their population origin and make an approximation to their age.

As expected, some of the C chromosomes detected in our sample are not Amerind in origin. However, further analyses clearly show that most of the C chromosomes present in Colombian Amerinds are not of recent introduction. This conclusion is supported by phylogenetic analyses of populations and haplotypes that reveal a closer relationship between most C and T chromosomes than between C and non-Amerind chromosomes. The similarity in allele frequencies between C and T chromosomes is suggestive of a low level of Y chromosome variation at the time of the C-to-T mutation, perhaps occurring around the time of entry into the American continent. This is in agreement with the presence of the DYS199 T allele in the three main Native American linguistic groups and its restricted distribution in Asia, and suggests that T and C chromosomes were present among the founders of the New World (9, 11, 12). Consistent with this

scenario, recent analyses of mtDNA diversity suggest that a considerable genetic differentiation might have occurred in the ancestors of Native Americans while in Beringia before their dispersal in the American continent (34, 35).

Assuming a single-founder haplotype for the DYS199 T and DYS199 C lineages, estimated coalescence times should approximate their time of entry into the New World. Based on the observed level of microsatellite variability, we have estimated the age of the DYS199 T haplogroup at 9,334–11,456 years by using two different approaches. The coalescence time of Amerind DYS199 C chromosomes is more difficult to establish because of the possibility of some level of admixture in these chromosomes. We have made two estimates, using in the first case all available C chromosomes with no direct evidence for admixture and in the second case only those C chromosomes with a higher probability of being of Amerind origin, based on phylogenetic analysis (Fig. 2). The first estimate is likely to be inflated by the presence in this sample of non-Amerind chromosomes. On the other hand, the second estimate is likely to be biased downward because not all Amerind haplotypes are within a single cluster (Fig. 2). These two approximations, giving ASD-based estimates of 18,462 and 5,675, thus represent a range for the age of the C haplogroup. Excluding the possibilities of back-mutation and differential selective pressures and considering that DYS199 C represents the ancestral allelic state, it seems more likely that the age of this haplogroup is closer to the older estimate. The age estimated for the DYS199 C and T lineages is lower than that obtained for the four mtDNA founder haplotypes (about 30,000–40,000 years ago; refs. 34 and 35). However, apart from uncertainties in the mutation process (37), our estimates could be biased downward, particularly considering that we have examined only populations from Colombia.

The phylogenetic analysis of microsatellite haplotypes is complicated by the possibility of recurrent mutation. The occurrence in our sample of several identical microsatellite haplotypes in C and T chromosomes is mostly indicative of homoplasmy. However, the probability of homoplasmy diminishes with the time elapsed since haplotype separation, as evidenced by the lower haplotype sharing between Amerind and non-Amerind populations. At that level, extensive haplotype sharing is most likely the result of recent admixture, and microsatellites can help define haplotypes of such an admixed origin. With these caveats in mind, it is interesting to see the substantial level of geographic structure detected when we compared six-locus microsatellite haplotypes (Fig. 2). This small number of markers already allows detection of Y haplotype clustering at the continental level consistent with an important regional apportionment of Y chromosome diversity (38, 39). Precise evaluation of phylogenetic relatedness of Native American founder haplotypes will require more microsatellites as well as additional slowly evolving biallelic markers. In particular, the reconstruction of microsatellite haplotypes in C chromosomes of Na-Dene and Eskimo populations should help define whether they result from admixture or, as we have shown for Colombian Amerinds, they are mostly Native American in origin. This approach should be highly informative for evaluating the relatedness of T chromosomes and estimating the age of haplogroups across major Native American linguistic groups, thus allowing independent testing of the three-migration-wave hypothesis.

The M19 A allele was initially identified in the Ticuna population of the upper Amazon (16). In our sample, we detected two Wayuu chromosomes bearing this allele. Observed microsatellite diversity in M19 A chromosomes results in an age estimate for this haplogroup of about 5,000–10,000 years. It therefore seems likely that M19 A chromosomes should also be present in other South American populations. To date, these have not been detected in the Surui and Karitiana, also from the Amazon basin, nor in four other South Amerind populations (16), but a more exhaustive search is warranted and should be informative for reconstructing ancient population movements within South America.

Special thanks are offered to the members of the communities that collaborated in this study and the many people that helped during field work, including: M. Florez, A. Ramirez, M. F. Ramirez, J. F. Ramirez, G. Velandia, Secretarías de Salud, and Organización Zonal Indígena del Putumayo. M. C. Bortolini, D. Labuda, and three anonymous reviewers made useful comments on an earlier version of the manuscript. P. Underhill provided sequence information for the typing of marker M19. This work was funded by Colciencias (Grant 1115-05-132-94 to A.R.L.). D.O.B. was partially supported by the Northwick Park Institute for Medical Research. M.W.F. is supported by National Institutes of Health Grant 28428, and D.B.G. by the Biotechnology and Biological Sciences Research Council.

- Fiedel, S. J. (1992) *Prehistory of the Americas*. (Cambridge Univ. Press, Cambridge, U.K.), 2nd Ed.
- Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. (1994) *History and Geography of Human Genes* (Princeton Univ. Press, Princeton).
- Crawford, M. H. (1998) *The Origins of Native Americans* (Cambridge Univ. Press, Cambridge, U.K.).
- Greenberg, J. H., Turner, C. J. & Zegura, S. L. (1986) *Curr. Anthropol.* **27**, 477–497.
- Ruhlen, M. (1991) *A Guide to the World's Languages: Classification* (Stanford Univ. Press, Palo Alto, CA), Vol. 1.
- Torroni, A., Chen, Y.-S., Semino, O., Santachiara-Beneceretti, A. S., Scott, C. R., Lott, M. T., Winter, M. & Wallace, D. G. C. (1994) *Am. J. Hum. Genet.* **54**, 303–318.
- Pena, S. D., Santos, F. R., Bianchi, N. O., Bravi, C. M., Carnese, F. R. & Rothhammer, F. (1995) *Nat. Genet.* **11**, 15–16.
- Santos, F. R., Rodriguez-Delfin, L., Pena, S. D., Moore, J. & Weiss, K. M. (1996) *Am. J. Hum. Genet.* **58**, 1369–1370.
- Underhill, P. A., Jin, L., Zeng, R., Oefner, P. J. & Cavalli-Sforza, L. L. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 196–200.
- Bianchi, N., Bailliet, G., Bravi, C. M., Pena, S. D. & Rothhammer, F. (1997) *Am. J. Phys. Anthropol.* **102**, 79–89.
- Karafet, T., Zegura, S. L., Vuturo-Brady, J., Posukh, O., Osipova, L., Wiebe, V., Romero, F., Long, J. C., Harihara, S., Jin, F., *et al.* (1997) *Am. J. Phys. Anthropol.* **102**, 301–314.
- Lell, J. T., Brown, M. D., Schurr, T. G., Sukernik, R. I., Starikovskaya, Y. B., Torroni, A., Moore, L. G., Troup, G. M. & Wallace, D. C. (1997) *Hum. Genet.* **100**, 536–543.
- Kaysner, M., Caglia, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., Heidorn, F., Herrman, S., Herzog, B., Hidding, F., *et al.* (1997) *Int. J. Legal. Med.* **110**, 125–133.
- Hammer, M. F. & Horai, S. (1995) *Am. J. Hum. Genet.* **56**, 951–962.
- Seielstad, M. T., Hebert, J. M., Lin, A. A., Underhill, P. A., Ibrahim, M., Vollrath, D. & Cavalli-Sforza, L. L. (1994) *Hum. Mol. Genet.* **3**, 2159–2161.
- Underhill, P. A., Jin, L., Lin, A. A., Qasim Medhi, S., Jenkins, T., Vollrath, D., Davis, R. W., Cavalli-Sforza, L. L. & Oefner, P. J. (1997) *Genome Res.* **7**, 996–1005.
- Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4**, 406–425.
- Maddison W. & Slatkin, M. (1991) *Evolution* **45**, 1184–1197.
- Felsenstein, J. (1993) PHYLIP (Phylogeny Inference Package) (Univ. of Washington, Seattle), Ver. 3.5c.
- Hammer, M. F., Spurdle, A. B., Karafet, T., Bonner, M. R., Wood, E. T., Novelletto, A., Malaspina, P., Mitchell, R. J., Horai, S., Jenkins, T., *et al.* (1997) *Genetics* **145**, 787–805.
- Nei, M. (1987) *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York).
- Knijff, P., Kaysner, M., Caglia, A., Corach, D., Fretwell, N., Herzog, B., Hidding, M., Honda, K., Jobling, M., Krawczak, M., *et al.* (1997) *Int. J. Legal. Med.* **110**, 134–140.
- Zerjal, T., Dashnyam, B., Pandya, A., Kaysner, M., Roewer, L., Santos, F. R., Schiefelhövel, W., Fretwell, N., Jobling, M. A., Harihara, S. *et al.* (1997) *Am. J. Hum. Genet.* **60**, 1174–1183.
- Goldstein, D. B., Ruiz-Linares, A., Cavalli-Sforza, L. L. & Feldman, M. W. (1995) *Genetics* **139**, 463–471.
- Slatkin, M. (1995) *Genetics* **139**, 457–462.
- Thomas, M. G., Skorecki, K., Ben-Ami, H., Parfitt, T., Bradman, N. & Goldstein, D. B. (1998) *Nature (London)* **394**, 138–140.
- Excoffier, L., Smouse, P. & Quattro, J. (1992) *Genetics* **131**, 479–491.
- Weber, J. L. & Wong, C. (1993) *Hum. Mol. Genet.* **2**, 1123–1128.
- Heyer, E., Puymirat, J., Dieltjes, P., Bakker, E. & de-Knijff, P. (1997) *Hum. Mol. Genet.* **6**, 799–803.
- Weiss, K. (1973) *Am. Antiquity* **38**, 1–186.
- Merrifether, A. D. & Ferrel, R. E. (1996) *Mol. Phyl. Evol.* **5**, 241–246.
- Forster, P., Harding, R., Torroni, A. & Bandelt, H. (1996) *Am. J. Hum. Genet.* **59**, 935–945.
- Stone, A. C. & Stoneking, M. (1998) *Am. J. Hum. Genet.* **62**, 1153–1170.
- Bonatto, S. L. & Salzano, F. M. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 1886–1871.
- Bonatto, S. L. & Salzano, F. M. (1997) *Am. J. Hum. Genet.* **61**, 1413–1423.
- Kittles, R. A., Perola, M., Peltonen, L., Bergen, A. W., Aragon, R. A., Virkkunen, M., Linnoila, M., Goldman, D. & Long, J. C. (1998) *Am. J. Hum. Genet.* **62**, 1171–1179.
- Di Rienzo, A., Donnelly, P., Toomajian, C., Sisk, B., Hill, A., Petzl-Erler, M. L., Haines, G. K. & Barch, D. H. (1998) *Genetics* **148**, 1269–1284.
- Ruiz-Linares, A., Nayar, K., Goldstein, D. B., Hebert, J. M., Seielstad, M. T., Underhill, P. A., Lin, A. A., Feldman, M. W. & Cavalli-Sforza, L. L. (1996) *Ann. Hum. Genet.* **60**, 401–408.
- Seielstad, M. T., Minch, E. & Cavalli-Sforza, L. L. (1998) *Nat. Genet.* **3**, 278–280.