

# Hierarchical Patterns of Global Human Y-Chromosome Diversity

Michael F. Hammer,\*† Tatiana M. Karafet,\* Alan J. Redd,\* Hamdi Jarjanazi,\*  
Silvana Santachiara-Benerecetti,‡ Himla Soodyall,§ and Stephen L. Zegura†

\*Laboratory of Molecular Systematics and Evolution and †Department of Anthropology, University of Arizona; ‡Department of Genetics, Università degli Studi di Pavia, Pavia, Italy; and §SAMIR, University of Witwatersrand, Johannesburg, South Africa

We examined 43 biallelic polymorphisms on the nonrecombining portion of the Y chromosome (NRY) in 50 human populations encompassing a total of 2,858 males to study the geographic structure of Y-chromosome variation. Patterns of NRY diversity varied according to geographic region and method/level of comparison. For example, populations from Central Asia had the highest levels of heterozygosity, while African populations exhibited a higher level of mean pairwise differences among haplotypes. At the global level, 36% of the total variance of NRY haplotypes was attributable to differences among populations (i.e.,  $\Phi_{ST} = 0.36$ ). When a series of AMOVA analyses was performed on different groupings of the 50 populations, high levels of among-groups variance ( $\Phi_{CT}$ ) were found between Africans, Native Americans, and a single group containing all 36 remaining populations. The same three population groupings formed distinct clusters in multidimensional scaling plots. A nested cladistic analysis (NCA) demonstrated that both population structure processes (recurrent gene flow restricted by isolation by distance and long-distance dispersals) and population history events (contiguous range expansions and long-distance colonizations) were instrumental in explaining this tripartite division of global NRY diversity. As in our previous analyses of smaller NRY data sets, the NCA detected a global contiguous range expansion out of Africa at the level of the total cladogram. Our new results support a general scenario in which, after an early out-of-Africa range expansion, global-scale patterns of NRY variation were mainly influenced by migrations out of Asia. Two other notable findings of the NCA were (1) Europe as a “receiver” of intercontinental signals primarily from Asia, and (2) the large number of intracontinental signals within Africa. Our AMOVA analyses also supported the hypothesis that patrilocal effects are evident at local and regional scales, rather than at intercontinental and global levels. Finally, our results underscore the importance of subdivision of the human paternal gene pool and imply that caution should be exercised when using models and experimental strategies based on the assumption of panmixia.

## Introduction

Knowledge of just how genetic variation is partitioned among human populations has important implications for studies of human origins, DNA forensics, and the etiology of human disease. Compared with other species, humans have relatively low levels of genetic diversity. Also, the proportion of that diversity that exists between all levels of human population is correspondingly low (Nei 1987). For example, studies of blood groups and protein polymorphisms have shown that approximately 85% of all human genetic diversity can be found within a single population, while only 5%–10% of the total diversity is partitioned among major geographic regions (Lewontin 1972; Nei and Roychoudhury 1974; Latter 1980). A similar apportionment of genetic diversity has been revealed at the DNA level by genomewide analyses of a variety of different markers (Bowcock and Cavalli-Sforza 1991; Batzer et al. 1994; Deka et al. 1995; Barbujani et al. 1997; Jorde et al. 2000). Concordant results have also been obtained from other human traits, such as cranial morphology (Relethford and Harpending 1994). Overall, these studies have generally underscored the lack of discontinuity among human groups (Lewontin 1972; Barbujani et al. 1997) and the relative homogeneity of the human species. This observation, in turn, may reflect a relatively recent ori-

gin for human between-groups differentiation, relatively high rates of migration among different human groups, or both (Relethford 1995).

Assuming a 1:1 sex-ratio, autosomal and X-linked regions of the genome have four- and threefold higher effective sizes, respectively, than the nonrecombining portion of the Y chromosome (NRY) and the mitochondrial DNA molecule. Consequently, increased levels of population subdivision due to genetic drift are expected for these uniparentally inherited haploid regions of the genome. Until recently, the small number of known NRY polymorphisms has hindered a comprehensive assessment of the global structure of Y-chromosome diversity. Earlier studies indicated that Y-chromosome polymorphisms were geographically restricted and that  $F_{ST}$  values for the NRY were higher than those for mtDNA (Jobling and Tyler-Smith 1995; Cavalli-Sforza and Minch 1997; Underhill et al. 1997; Hammer et al. 1998; Perez-Lezaun et al. 1999). Indeed, the higher observed  $F_{ST}$  for the NRY compared with that for mtDNA led Seielstad, Minch, and Cavalli-Sforza (1998) to propose that females have had an eightfold higher migration rate than males. It is unclear, however, whether the suggested underlying cause of this higher mobility (i.e., local-scale patrilocal, defined anthropologically as the tendency for a wife to move into her husband's natal domicile) would lead to a higher global  $F_{ST}$  for the Y chromosome (Stoneking 1998). On the other hand, contrasting signals in nested cladistic analyses of NRY and mtDNA data sets led Hammer et al. (1998) to hypothesize that male migration rates may have been higher than those for females at the intercontinental level. De-

Key words: subdivision, patrilocal, gene flow, male migrations.

Address for correspondence and reprints: Michael F. Hammer, Laboratory of Molecular Systematics and Evolution, Biosciences West room 239, University of Arizona, Tucson, Arizona 85721. E-mail: mhammer@u.arizona.edu.

*Mol. Biol. Evol.* 18(7):1189–1203, 2001

© 2001 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

**Table 1**  
**NRV Haplotype Frequencies in 10 Regional Groups**

MUTATION NO./NAME	HT <sup>a</sup>	POPULATION GROUP <sup>b</sup>										SUM	REFERENCE <sup>c</sup>	
		SAF	NAF	MEA	EUR	SAS	CAS	NAS	EAS	OCE	AME			
Ancestral haplotype ..	h1	24	1	0	0	0	0	0	0	0	0	0	25	Hammer et al. (1998)
1. M13 .....	h2	2	9	0	0	0	0	0	0	0	0	0	11	Underhill et al. (1997)
2. DYS <sub>108</sub> .....	h3	12	0	0	0	0	0	0	0	0	0	0	12	This paper
3. PN3 .....	d													Hammer (1995)
4. DYS198/M6 .....	d													Underhill et al. (1997)
5. M14 .....	d													Underhill et al. (1997)
6. 16E4 <sub>20642</sub> .....	h4	9	0	0	0	0	0	0	0	0	0	0	9	This paper
7. SRY <sub>10831.1</sub> <sup>e</sup> .....	h5	20	0	0	0	0	0	0	0	0	0	0	20	Hammer et al. (1998)
8. 50f2(P) .....	h6	5	0	0	0	0	0	0	0	0	0	0	5	Jobling et al. (1996)
9. DYS265 <sub>83</sub> .....	h7	7	0	0	0	0	0	0	0	0	0	0	7	This paper
10. DYS194 <sub>450</sub> .....	h8	1	0	0	0	0	0	0	0	0	0	0	1	This paper
11. MSY2.1 <sup>e</sup> .....	h9	5	0	0	0	0	0	0	0	0	0	0	5	Bao et al. (2000)
12. DYS190 <sub>905</sub> .....	h10	3	0	0	0	0	0	0	0	0	0	0	3	This paper
13. 486,O,2 <sub>118453</sub> .....		0	0	0	0	0	0	0	0	0	0	0	0	This paper
14. DYS287 .....	h11	0	0	0	0	0	4	2	58	1	0	0	65	Hammer et al. (1998)
15. M15 .....	h12	0	0	0	0	0	0	2	7	0	0	0	9	Underhill et al. (1997)
16. SRY <sub>4064</sub> .....	h13	21	1	0	1	0	0	0	0	0	0	0	23	Hammer et al. (1998)
17. PN2 .....	h14	16	65	15	53	5	3	0	0	1	1	1	159	Hammer (1995)
18. PN1 .....	h15	102	4	1	0	0	0	0	0	0	0	3	110	Hammer (1995)
19. RPS4Y <sub>711</sub> .....	h16	0	0	1	0	4	40	194	35	73	11	358	358	Karafet et al. (1999)
20. DYS194 <sub>1391</sub> .....	h17	0	0	0	0	0	0	0	0	23	0	0	23	This paper
21. DYS263/M8 .....	h18	0	0	0	0	0	0	0	5	0	0	0	5	Underhill et al. (1997)
22. DYS188 <sub>792</sub> .....	h19	0	2	8	6	54	10	2	3	0	1	86	86	Hammer et al. (2000)
23. DYS190 <sub>922</sub> .....	h20	0	0	5	52	0	2	5	1	1	2	68	68	This paper
24. DYS221 <sub>136</sub> .....	h21	0	0	13	14	2	7	1	0	2	1	40	40	Hammer et al. (2000)
25. p1212 .....	h22	1	32	75	29	19	34	6	3	1	1	201	201	Hammer et al. (2000)
26. DYS260/M12 .....	h23	0	3	1	12	8	1	0	0	0	0	25	25	Underhill et al. (1997)
27. M9 .....	h24	1	6	20	6	17	20	39	25	100	0	234	234	Karafet et al. (1999)
28. M122 .....	h25	0	0	0	0	0	12	27	128	25	0	192	192	Su et al. (1999)
29. LINE-1 .....	h26	0	0	0	0	0	2	8	55	0	0	65	65	Santos et al. (2000)
30. MSY2.2 <sup>e</sup> .....	h27	0	0	0	0	0	2	1	18	13	0	34	34	Bao et al. (2000)
31. Tat .....	h28	0	0	1	5	0	6	65	0	0	1	78	78	Karafet et al. (1999)
32. ARSEP <sub>71227</sub> .....	h29	0	0	0	0	0	0	2	8	0	0	10	10	This paper
33. M95 .....	h30	0	0	0	0	0	1	0	39	0	0	40	40	Su et al. (1999)
34. SRY <sub>465</sub> .....	h31	0	0	0	0	0	0	0	38	1	0	39	39	Shinka et al. (1999)
35. DXY55Y .....	h32	0	0	0	0	0	0	0	23	0	0	23	23	Shinka et al. (1999)
36. DYS234/M4 .....		0	0	0	0	0	0	0	0	0	0	0	0	Underhill et al. (1997)
37. DYS214/M5 .....	h33	0	0	0	0	0	0	0	0	25	0	25	25	Underhill et al. (1997)
38. DYS257 <sub>162</sub> .....	h34	0	0	0	0	0	0	0	0	14	0	14	14	This paper
39. DYS214/M16 .....	h35	0	0	0	0	0	0	0	0	3	0	3	3	Underhill et al. (1997)
40. DYS257 <sub>108</sub> .....	h36	0	1	4	4	28	45	103	10	2	64	261	261	Hammer et al. (1998)
41. DYS194 <sub>469</sub> .....	h37	0	5	21	98	2	23	10	1	18	32	210	210	Hammer et al. (2000)
42. SRY <sub>10831.2</sub> <sup>e</sup> .....	h38	0	2	15	47	57	51	25	4	0	0	201	201	Hammer et al. (1998)
43. DYS199/M3 .....	h39	0	0	0	0	0	0	3	0	0	156	159	159	Karafet et al. (1999)
Regional totals .....		229	131	180	327	196	263	495	461	303	273	2,858		

<sup>a</sup> Haplotypes as defined in figure 1.<sup>b</sup> The composition of each regional group is described in *Materials and Methods*.<sup>c</sup> Reference to genotyping assay protocol and/or first report of mutation.<sup>d</sup> Multiple mutations on same lineage.<sup>e</sup> Site with recurrent mutation.

spite these observations, most human population genetics models assume panmixia. This study was designed to measure the degree of Y-chromosome structure on a global scale (i.e., to test the assumption of panmixia) and to test the global applicability of the patrilocality hypothesis.

## Materials and Methods

### Populations Sampled

We analyzed a total of 2,858 males from 50 populations (table 1). The sample was divided into the following 10 major geographic regions: sub-Saharan Africa

(SAF), North Africa (NAF), the Middle East (MEA), Europe (EUR), South Asia (SAS), Central Asia (CAS), North Asia (NAS), East Asia (EAS), Oceania (OCE), and the Americas (AME). Many of the samples analyzed here were also included in our previous studies (Hammer et al. 1997, 1998, 2000; Karafet et al. 1999), although the exact number of subjects reported for each population sometimes differs. The 50 populations were distributed according to geographic region in the following way (see accompanying citation below for populations not described in the aforementioned references): SAF ( $n = 229$ ) = 73 Khoisan, 55 East Bantus, 26 Pygmies, 26 Bagandans, and 49 Gambians; NAF ( $n = 131$ )

= 51 Ethiopians, 50 Egyptians, and 30 Tunisians; MEA ( $n = 180$ ) = 20 Saudi Arabians, 88 Syrians, and 72 Turks; EUR ( $n = 327$ ) = 86 Greeks, 62 Italians, 58 Romanians (this study), 34 Germans, 44 Russians, and 43 British; SAS ( $n = 196$ ) = 75 Sri Lankans, 59 Indians, and 62 Pakistanis (Qamar et al. 1999); CAS ( $n = 263$ ) = 45 Turkmen (this study), 14 Tajiks (this study), 77 Uzbeks (this study), 30 Kazakhs, 29 Altai, and 68 Uyghurs (this study); NAS ( $n = 495$ ) = 148 Mongolians, 81 Buryats, 122 Selkups, 27 Forest Nentsi, 95 Evenks, and 22 Siberian Eskimos; EAS ( $n = 461$ ) = 70 Vietnamese, 58 Miao (this study), 84 Chinese Han, 52 Manchu (this study), 76 Koreans, and 121 Japanese; OCE ( $n = 303$ ) = 60 East Indonesians, 47 Papua New Guineans, 50 Melanesians, 78 Australian Aboriginal people, 18 Micronesians, and 50 Polynesians (this study); and AME ( $n = 273$ ) = 80 Navajos, 45 Cheyenne, 24 Pima, 72 Mayans (this study), 28 Mixtecs, and 24 Wayus. All sampling protocols were approved by the Human Subjects Committee at the University of Arizona.

### Mutation Detection

Mutation detection analysis was performed using single-stranded conformation polymorphism (SSCP) (Sheffield et al. 1993) and denaturing high-performance liquid chromatography (DHPLC) (Underhill et al. 1997). Two panels of DNA samples were employed to ascertain polymorphisms using the above methods. For SSCP,  $n = 20$  (9 sub-Saharan Africans, 3 Asians, 3 Native Americans, 3 Europeans, and 2 Oceanians), and for DHPLC,  $n = 57$  (17 sub-Saharan Africans, 15 Asians, 11 Europeans, 7 Native Americans, and 7 Oceanians).

The SSCP method was used to screen a set of 20 sequence-tagged sites (STSs). The DHPLC method was used to screen for mutations in the following set of three clones that were previously used as probes to detect restriction fragment length polymorphism (RFLP) variation on the NRYs of humans and great apes (Allen and Ostrer 1994): clone 4-1 (*DYS188*), clone 3-11 (*DYS190*), and clone 3-8 (*DYS194*). Mutational variation within four STSs (*DYS221*, *DYS257*, *DYS199*, and *DYS211*) and two clones (3-8 and 4-1) was previously reported (Karafet et al. 1997; Hammer et al. 1998, 2000). Additional variation was found at sites within three STSs (*DYS7*, *DYS265*, and *DYS257b*) and two clones (3-8 and 3-11). Finally, Ya5 Alu elements within the 16E4 and 486,O,2 clones (GenBank accession numbers AC003094 and AC002531, respectively; <http://www.ncbi.nlm.nih.gov/Genbank/index.html>), as well as a 683-bp region of an arylsulfatase pseudogene (*ARSEP*, GenBank accession number AC002992) were screened for polymorphisms using DHPLC.

The *DYS7* (GenBank accession number G12023), *DYS265* (GenBank accession number G12016), and *DYS257* (GenBank accession number G38358) STSs were amplified using the conditions and primers reported by Vollrath et al. (1992). The Y-specific clones 3-8 (*DYS194*) and 3-11 (*DYS190*) (Allen and Ostrer 1994) were sequenced by primer walking (GenBank accession numbers AF257064 and AF337053, respectively), and

the sequence information was used to design primers to amplify shorter fragments for DHPLC analysis. DNA sequencing was performed by standard procedures to identify mutations which altered mobility on SSCP gels or DHPLC chromatograms.

As in previous mutation detection surveys (Underhill et al. 1997, 2000; Hammer et al. 1998; Karafet et al. 1999), we sequenced homologous DNA regions encompassing all sites found to be polymorphic on the human NRY in great ape species (e.g., one common chimpanzee, one bonobo, and one gorilla) to determine ancestral states, as well as the position of the root of the human NRY haplotype tree.

### Allele-Specific Genotyping Assays

A total of 23 segregating sites were discovered using the two mutation detection methods. Of these 23, 15 were chosen for genotyping in the entire sample (10 new and 5 previously published polymorphisms). The other 8 polymorphisms were found to be so rare in a subset of the 2,858 chromosomes that they were excluded from subsequent analyses. After determining the location of a sample with respect to its position on the haplotype tree, no further genotyping was undertaken for that sample. This hierarchical genotyping protocol means that not every individual was typed for every marker, and hence it is possible that some recurrent mutations remained undetected using this strategy (Underhill et al. 2000). Nevertheless, because the homoplasy rate for single-nucleotide polymorphisms (SNPs) on the NRY is so low (Underhill et al. 2000), it is unlikely that undetected multiple "hits" would seriously affect either our phylogenetic or our diversity analyses. The remaining 20 previously published polymorphisms (from the entire battery of 43 polymorphisms) were also genotyped for all 2,858 chromosomes, with the aforementioned caveats.

Variation at all previously unpublished polymorphic sites (table 1: mutations 2, 6, 9, 10, 12, 13, 20, 23, 32, and 38) was genotyped using allele-specific PCR (Sommer, Groszbach, and Bottema 1992). The PCR conditions and primer sequences employed in these allele-specific genotyping assays were deposited in the National Center for Biotechnology Information (NCBI) dbSNP database (<http://www.ncbi.nlm.nih.gov/SNP>). Mutations numbered 3, 7, 14, 16–20, 22, 24, 25, and 40–43 in table 1 were genotyped according to methods reported by Hammer and Horai (1995), Hammer et al. (1998, 2000), and Karafet et al. (1999). Other previously published mutations included mutation 8 (Jobling et al. 1996); mutations 1, 4, 5, 15, 21, 26, 27, 36, 37, and 39 (Underhill et al. 1997); mutation 31 (Zerjal et al. 1997); mutations 34 and 35 (Shinka et al. 1999); mutations 28 and 33 (Su et al. 1999); mutations 11 and 30 (Bao et al. 2000); and mutation 29 (Santos et al. 2000).

### Statistical Analyses

Parsimony analysis of NRY haplotypes was aided by the use of PAUP, version 4.0b4 (Swofford 2000), with outgroup rooting. Measures of haplotype diversity,

including the number of haplotypes ( $k$ ), Nei's (1987) heterozygosity ( $h$ ), and the mean number of pairwise differences among haplotypes ( $p$ ), were calculated using the software package ARLEQUIN (Schneider et al. 1998). We also used ARLEQUIN to perform analysis of molecular variance (AMOVA). AMOVA produces estimates of variance components and  $\Phi$  statistics ( $F$  statistic analogs) reflecting the correlation of haplotypic diversity at different levels of hierarchical subdivision (Excoffier, Smouse, and Quattro 1992). Because the assumptions of random sampling, "pure" genetic drift, and no migration are likely to be violated in all human populations, caution is needed when interpreting  $\Phi$  statistics. Nevertheless, according to Excoffier, Smouse, and Quattro (1992), the resulting variance components can be viewed as convenient summaries of the partitioning of genetic variation within and among populations. We performed multidimensional scaling (MDS) (Kruskal 1964) on the  $\Phi_{ST}$  distances generated in ARLEQUIN using the software package NTSYS (Rohlf 1998). Nested cladistic analyses (NCAs) were carried out using GeoDis, version 2.0 (Posada, Crandall, and Templeton 2000). This novel method attempts to explain statistically significant associations between haplotypes and geography in terms of population history and/or population structure considerations. Population structure processes operate over short time intervals and tend to establish migration-drift equilibria, whereas population history events are considered to be nonrecurrent phenomena that disrupt equilibria. Three conditions underlie the general applicability of NCA and its ability to discriminate among the various population structure processes (i.e., recurrent gene flow restricted by isolation by distance vs. long-distance dispersal) and/or population history events (i.e., contiguous range expansion, long-distance colonization, or fragmentation). These considerations include (1) adequate sampling across the geographic range of the species, (2) temporal polarity of the haplotype network, and (3) mutational resolution in the haplotype tree. Because our cladogram was rooted by outgroup comparisons, we were able to infer the geographical polarity of several of the signals detected by the NCA analysis by considering the distribution of interior and tip clades (e.g., directionality was assumed to go from interior to tip), especially in cases where there was a clear geographic pattern of separation between ancestral and derived haplotypes. For a more extensive explanation of the NCA method, consult Templeton, Routman, and Phillips (1995), Hammer et al. (1998), and Posada, Crandall, and Templeton (2000).

## Results

During the course of this research, the following 10 previously unpublished Y-specific polymorphisms were discovered: a C→T transition at position 20642 of 16E4; a G→A transition at position 905 of *DYS190*; a T→G transversion at position 922 of *DYS190*; a T→C transition at position 450 of *DYS194*; a T→C transition at position 1391 of *DYS194*; a C→A transversion at position 118453 of 486,O,2; a T→C transition and a T

deletion at positions 71227 and 71228, respectively, of ARSEP; a C→T transition at position 108 of *DYS7*; a G→C transversion at position 83 of *DYS265*; and a G→A transition at position 162 of *DYS257*. Comparisons with the homologous sequences from one common chimpanzee, one bonobo, and one gorilla allowed us to infer the ancestral states at all of these sites except 16E4<sub>20642</sub> and 486,O,2<sub>118453</sub>, which only occurred in human-specific Alu elements.

In addition to these 10 new markers, we surveyed 33 previously published polymorphisms (table 1). Mutational events at two sites were recurrent (*SRY*<sub>10831</sub> and *MSY2*). The character states at all 41 mutational sites give rise to 44 possible NRY haplotypes, of which 39 were present in this survey. The frequencies of these 39 haplotypes (h1–h39) in each regional group are reported in table 1. Figure 1 displays a maximum-parsimony tree showing the evolutionary relationships of all 39 haplotypes. Haplotypes in figure 1 are color-coded by geography. The pie charts represent the frequencies of occurrence of the haplotypes within each of the 10 geographic regions listed in table 1, and the overall size of each circle represents a global haplotype frequency.

The position of the root in figure 1 was determined by outgroup comparisons. In order to confirm the position of the root, a subset of chromosomes was genotyped at four sites (M91, M42, M94, and M139) that mark the two most basal lineages on the maximum-parsimony tree presented by Underhill et al. (2000). Overall, the tree in figure 1 exhibits a high degree of underlying similarity with Underhill et al.'s (2000) maximum-parsimony tree (even though our analysis was based on ~fourfold fewer markers, with only 15 common to both studies). In fact, our figure 1 and the Y-chromosome tree in Underhill et al. (2000) offer remarkably strong mutual confirmation. For instance, Underhill et al.'s (2000) haplogroup I is represented by our haplotypes 1–4. The rest of the correspondences are as follows, again with the Underhill et al. (2000) haplogroup designation first, followed by our matching haplotype number(s): haplogroup II—haplotypes 5–10; haplogroup III—haplotypes 13–15; haplogroup IV—haplotypes 11 and 12; haplogroup V—haplotypes 16–18; haplogroup VI—haplotypes 19–23; haplogroup VII—haplotypes 25–27 and 29–32; haplogroup VIII—haplotypes 24, 28, and 33–35; haplogroup IX—haplotypes 36–38; and haplogroup X—haplotype 39.

## Haplotype Diversity

Diversity statistics for the 10 regional and 5 continental groups are presented in table 2. The number of regional haplotypes ( $k$ ) ranged from 10 in South Asians to 18 in East Asians. Regional haplotype diversity values ( $h$ ) ranged from 0.605 in Native Americans to 0.878 in Central Asians, while the mean number of pairwise differences ( $p$ ) ranged from 1.31 in Native Americans to 3.93 in sub-Saharan Africans. Four distinct patterns appear when two diversity statistic values for the 10 regional groups in table 2 are compared: (1) low  $p$ /low  $h$ , (2) high  $p$ /high  $h$ , (3) high  $p$ /low  $h$ , and (4) moderate

$p$ /high  $h$ . The Americas represent the only region that exhibits the first pattern of concordantly low  $p$  and low  $h$  values. The low  $p$  value occurs here because 90% of Native American Y-chromosome lineages are one-step neighbors restricted to haplotypes 36–39 (magenta in fig. 1), while the low  $h$  value is due to the fact that 57% of the Native Americans in our study have haplotype 39 (table 1). The second pattern, where a high  $p$  value is accompanied by a concordantly high  $h$  value, is seen only in Europeans. This pattern reflects intermediate frequencies of relatively divergent haplotypes found in different parts of the tree (i.e., blue in fig. 1). The third pattern, where a high  $p$  value is discordantly combined with a low  $h$  value, characterizes both African regional groups. In sub-Saharan Africa, the extremely high  $p$  value is influenced by the marked divergence among the dark green haplotypes in figure 1. The contrasting relatively low  $h$  value may occur because 45% of sub-Saharan African Y chromosomes exhibit a single haplotype (h15). Likewise, 50% of the North Africans have a single (but different) haplotype (h14), resulting in a low  $h$  value, while the rather high North African  $p$  value is associated with the occurrence of a diverse set of lineages (light green in fig. 1). Finally, Asian and Oceanian populations exhibit the fourth pattern, moderate  $p$  and high  $h$  values. Although their  $p$  values are moderate, the Central and East Asians have the highest  $h$  values among the 10 regions. These high  $h$  values are probably due to the lack of any predominant Central or East Asian haplotype (red and orange, respectively in fig. 1).

#### Analysis of Molecular Variance

According to Excoffier, Smouse, and Quattro (1992, p. 482),  $\Phi_{ST}$  can be viewed as the correlation of random haplotypes within populations relative to that of random pairs of haplotypes drawn from the whole species, while  $\Phi_{CT}$  is the correlation of random haplotypes within a group of populations relative to that of random pairs of haplotypes drawn from the whole species, and  $\Phi_{SC}$  is the correlation of the molecular diversity of random haplotypes within populations relative to that of random haplotypes drawn from the region. Table 3 presents variance components and  $\Phi$  statistics at five different grouping levels, which summarize the geographic partitioning of NRY diversity. When all 50 populations were combined into a global analysis, 64% of the variance was within populations, and thus  $\Phi_{ST} = 0.36$ . The  $\Phi$  statistics at the 10 regional and 5 continental levels were very similar. However, when the world was partitioned into three groups (Africa/Americas/"rest of the world"), all three  $\Phi$  statistics increased noticeably in value. When each continental grouping was contrasted with the remainder of the world (table 3, analyses 5–9), the African/non-African comparison exhibited the highest  $\Phi_{ST}$  and  $\Phi_{CT}$  values, but the lowest  $\Phi_{SC}$  value. In the American/non-American comparison, the  $\Phi_{ST}$  and  $\Phi_{CT}$  values were second highest, while the  $\Phi_{SC}$  value was second lowest. In contrast, the three  $\Phi$  statistics for the European/non-European, Asian/non-Asian, and Oceanian/non-Oceanian comparisons were nearly iden-

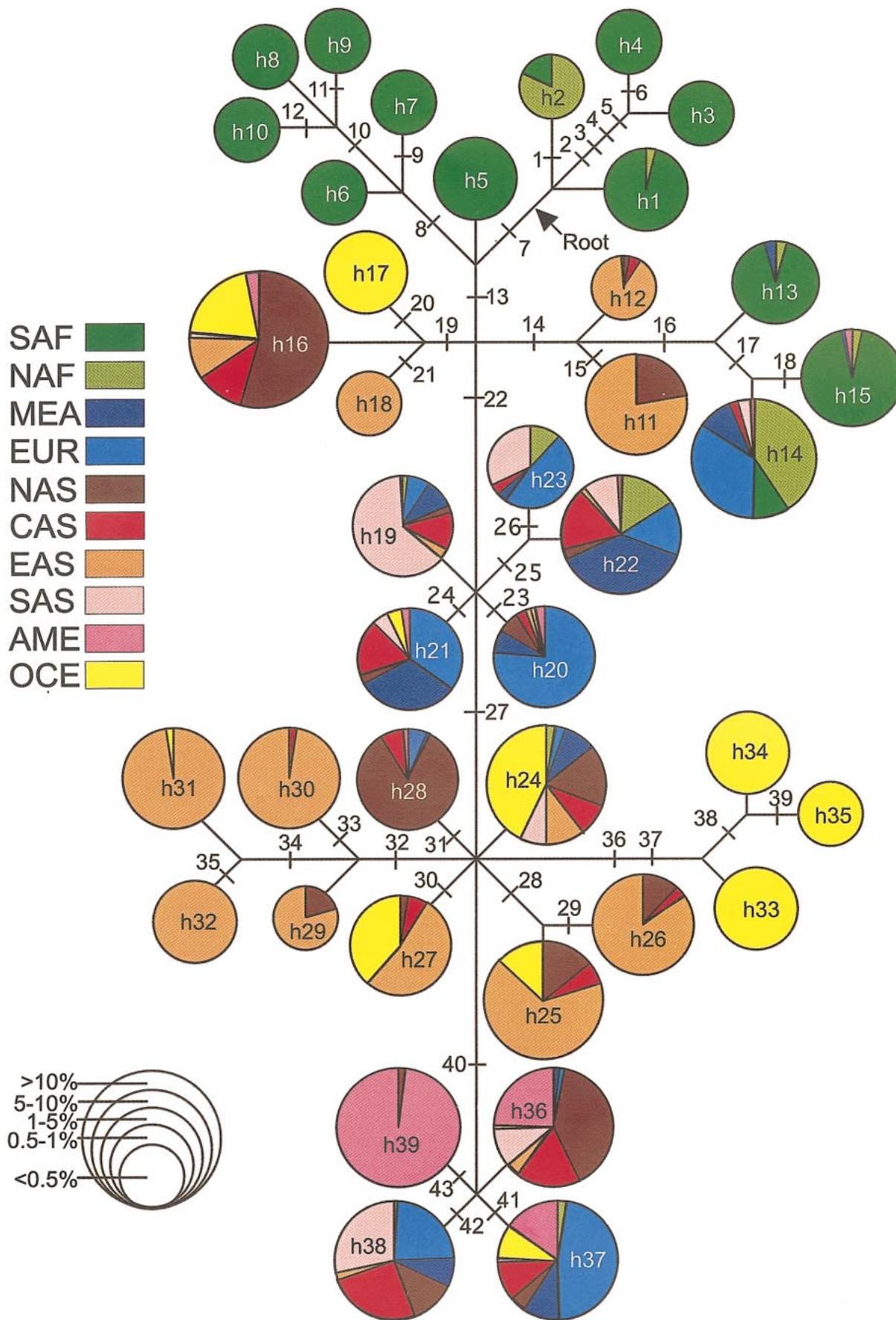
tical. Their  $\Phi_{ST}$  values corresponded closely to the global and 10 regional  $\Phi_{ST}$  values. However, their  $\Phi_{CT}$  values dropped  $\sim 10$ -fold compared with the 10-, 5-, and 3-group partitionings, while their  $\Phi_{SC}$  values were the highest of any the comparisons. The combined results suggest that while African and Native American populations were the most differentiated from the other groups, European, Asian, and Oceanian populations were practically indistinguishable from each other on the basis of all three  $\Phi$ -statistic patterns.

#### Multiple-Dimensional Scaling

Figure 2 shows the results of multidimensional scaling based on  $\Phi_{ST}$  genetic distances, as well as  $\Phi_{CT}$  values for the three intercluster comparisons. The correlation between the original  $\Phi_{ST}$  molecular distance matrix and a Euclidean distance matrix derived from the two-dimensional plot was extremely high ( $r = 0.983$ ). Africa, the Americas, and the 36 remaining populations formed three distinct clusters, paralleling the results of AMOVA.

#### Nested Cladistic Analysis

Figure 3 displays the nested cladogram for the 39 haplotypes (plus the five not found in our survey) generated by applying the nesting rules given in Templeton, Boerwinkle, and Sing (1987) and Templeton and Sing (1993). This nesting methodology produced 19 one-step clades, 9 two-step clades, 3 three-step clades, and a four-step clade that nested the entire cladogram. The NCA indicated highly statistically significant associations between clades and geographic locations for the entire cladogram ( $P < 0.0001$ ). Out of a total of 32 nested clades, 21 exhibited statistically significant associations with geography (data not shown). When the null hypothesis of no association between haplotype and geography is rejected, the analysis continues by generating specific explanatory inferences involving population history and/or structure considerations. With the aid of the key published on the GeoDis 2.0 website (<http://bioag.byu.edu/zoology/crandallLab/geodis.htm>) (Posada, Crandall, and Templeton 2000), we were able to infer the probable causes of these 21 patterns (table 4). Population structure inferences included both recurrent gene flow restricted by isolation by distance (RFG/IBD;  $n = 10$  inferences), as well as long-distance dispersal processes (LDD;  $n = 1$ ). Population history events included contiguous range expansions (CRE;  $n = 6$ ) and long-distance colonizations (LDC;  $n = 4$ ). Of these 21 inferences, 12 involved an intercontinental geographic pattern, while 9 were limited to a single continent. Some of the intercontinental signals affected more than two geographic regions, and some were redundant because they were detected at different nesting levels in the cladogram. As a result, the 12 intercontinental signals were deemed to represent only 11 separate entities (6 range expansions/long-distance colonization events and 5 recurrent gene flow/long-distance dispersal processes), all but 1 of which are illustrated by arrows in figure 4. Similarly, the nine intracontinental signals were deemed



to represent only eight separate signals (shown as arrows within circles in fig. 4). The remaining intercontinental signal involving global gene flow (via isolation by distance) is not depicted in figure 4 because it had no inherent polarity. As in our earlier global nested cladistic analyses (Hammer et al. 1998; Karafet et al. 1999), a contiguous range expansion out of Africa was detected at the level of the total cladogram (denoted by the widest black arrow in fig. 4).

## Discussion

### Evolutionary Tree for NRY Haplotypes

The root of the gene tree in figure 1 fell between two sets of haplotypes (h1–h4 and h5–h10) that are entirely restricted to the continent of Africa (table 1), thereby supporting the hypothesis of an African origin of contemporary NRY lineages (Hammer et al. 1998; Underhill et al. 2000). Haplotype 1 represents the “ancestral” human NRY haplotype (previously designated haplotype 1A in Hammer et al. 1998), which is found at relatively high frequencies in Khoisan populations. Haplotype 2, a one-step neighbor of h1, was found in ~14% of Ethiopian chromosomes in this study, as well as in 42% of Underhill et al.’s (1997) Sudanese samples. The predominantly East African distribution of h1 and h2 (data not shown) lends support to the hypothesis of a remnant archaic gene pool along the African rift and a possible wider East African range for the ancestors of the Khoisan (Scozzari et al. 1999). Other than the two aforementioned sets of haplotypes confined to Africa and a set of YAP<sup>+</sup> haplotypes (h13–h15), all haplotypes in the remainder of the tree were absent or very rare in sub-Saharan African populations.

In contrast to other kinds of genetic data (Przeworski, Hudson, and Di Rienzo 2000), both our present NRY tree and that of Underhill et al. (2000) clearly indicate that haplotypes found outside of Africa are not a subset of those found within Africa. However, the NRY tree does show a branching pattern similar to that seen in the gene trees of several other loci: African-specific branches are found on both sides of the root of the tree and are separated from the remaining sets of African and non-African branches (Labuda, Zietkiewicz, and Yotova 2000).

### Apportionment of NRY Biallelic Diversity

This study represents the most extensive  $\Phi$ -statistic analysis utilizing Y-chromosome biallelic markers to date combining sample size, geographic coverage, and number of markers. Caution should be exercised when comparing variance partitions among studies because results will depend on which populations are sampled,

how the populations are grouped (nested), and what underlying models of population structure are assumed (Urbanek, Goldman, and Long 1996). With this caveat in mind, a sample of previously reported total among-groups variation values (whether measured by  $\Phi_{ST}$ ,  $F_{ST}$ , or  $G_{ST}$ ) for Y-specific RFLPs and/or SNPs was found to range from 0.230 to 0.645 (Hammer et al. 1997; Poloni et al. 1997; Seielstad, Minch, and Cavalli-Sforza 1998; Kittles et al. 1999; Jorde et al. 2000). Our global  $\Phi_{ST}$  value of 0.360 is only slightly lower than the mean value (0.413) calculated from the five studies cited above. Poloni et al. (1997) cautioned that their  $F_{ST}$  value of 0.230 may be an underestimate in part because of recurrent mutation acting on the p49a,f/*TaqI* polymorphic system. It is not clear why the  $F_{ST}$  value of 0.645 reported by Seielstad, Minch, and Cavalli-Sforza (1998) based on the data presented in Underhill et al. (1997) is so much higher than the value reported here. When we analyzed the Underhill et al. (1997) data set using the 10 population groupings provided in their figure 2, we obtained a  $\Phi_{ST}$  value of 0.540 and an  $F_{ST}$  value of 0.414.

In general, the within-populations variance component for Y-chromosome data is much smaller than the values reported for mtDNA (Excoffier, Smouse, and Quattro 1992; Seielstad, Minch, and Cavalli-Sforza 1998; Kittles et al. 1999; Jorde et al. 2000). On the other hand, the among-groups and the among-populations-within-groups component values for Y chromosomes usually exceed those for mtDNA. Seielstad, Minch, and Cavalli-Sforza (1998) appealed to a lower transgenerational migration rate for males as the major explanatory factor for why their Y-chromosome  $F_{ST}$  value (0.645) was so much higher than the mtDNA-based value they recalculated from Excoffier, Smouse, and Quattro’s (1992) data (0.186). Their estimated eightfold higher female migration rate was attributed to patrilocality operating primarily at the local and perhaps regional levels (Seielstad, Minch, and Cavalli-Sforza 1998). Although the majority of human societies practice patrilocality (Murdock 1967), it is unclear whether this effect extends to intercontinental and global levels (Stoneking 1998). In order to investigate this proposition, we analyzed our Y-chromosome data according to the grouping design presented in Excoffier, Smouse, and Quattro (1992) (albeit with different samples), thereby permitting a comparison of the  $\Phi$  statistics generated from our Y-chromosome data with those derived from a global mtDNA RFLP data set. Although the “patrilocality effect” was extremely clear at the interregional level within continents ( $\Phi_{SC}$  ratio = 4.4), it was much less apparent at the intercontinental level ( $\Phi_{CT}$  ratio = 1.4) or at the global level ( $\Phi_{ST}$  ratio = 1.7). It is even possible that the latter two ratios would be closer to one (or less than

←

FIG. 1.—Evolutionary tree for 39 NRY haplotypes. The 43 mutational events listed in table 1 are shown by cross-hatches. Haplotypes numbered h1–h39 in circles correspond to designations in table 1. The root of the genetree is denoted by an arrow. Haplotypes are color-coded by geography (see figure for color-coding key). The pie charts represent the frequencies of occurrence of the haplotypes within each of the 10 geographic regions listed in table 1. The overall size of each pie chart corresponds to one of five frequency classes (see figure for frequency class key) and represents the frequency of that haplotype in the global sample of 2,858 chromosomes.

**Table 2**  
**NRY Haplotype Diversity in 10 Regional and 5 Continental Groups**

Group	No. of Populations	<i>N</i>	<i>k</i> <sup>a</sup>	<i>h</i> <sup>b</sup>	<i>p</i> <sup>c</sup>
Regional					
Sub-Saharan Africans	5	229	15	0.767 ± 0.002	3.93 ± 0.131
Europeans	6	327	12	0.828 ± 0.001	3.44 ± 0.097
North Africans	3	131	12	0.689 ± 0.003	3.31 ± 0.150
East Asians	6	461	18	0.867 ± 0.000	2.88 ± 0.071
Central Asians	6	263	17	0.878 ± 0.001	2.60 ± 0.086
Oceanians <sup>d</sup>	6	303	16	0.809 ± 0.001	2.60 ± 0.080
Middle Easterners	3	180	13	0.782 ± 0.002	2.54 ± 0.102
North Asians	6	495	17	0.775 ± 0.001	2.48 ± 0.060
South Asians	3	196	10	0.803 ± 0.001	2.05 ± 0.082
Native Americans <sup>d</sup>	6	273	11	0.605 ± 0.002	1.31 ± 0.050
Continental					
Africans	8	360	20	0.841 ± 0.001	3.96 ± 0.105
Europeans <sup>e</sup>	9	507	14	0.852 ± 0.000	3.23 ± 0.074
Asians	21	1,415	23	0.904 ± 0.000	2.85 ± 0.040

<sup>a</sup> Number of haplotypes.  
<sup>b</sup> Haplotype diversity ± SE (Nei 1987).  
<sup>c</sup> Mean number of pairwise differences ± SE (Schneider et al. 1998).  
<sup>d</sup> Both a regional and a continental group.  
<sup>e</sup> European and Middle Eastern populations combined.

one) without the high mutation rate and accompanying homoplasy known to affect mtDNA restriction sites (Excoffier, Smouse, and Quattro 1992) and thought to depress *F*<sub>ST</sub> values (Jorde et al. 2000). For instance, the at-least-10-fold-higher mtDNA mutation rate would be expected to increase the within-groups variance component and thus decrease the  $\Phi_{ST}$  values for mtDNA relative to the NRY, thereby artificially inflating the corresponding NRY/mtDNA ratio (Jin and Chakraborty 1995). Numerous possible explanations exist for the discrepancy between the  $\Phi_{SC}$  ratio and the other two values. For instance, increased intercontinental male migration, decreased intercontinental female gene flow, and sex-specific demographic factors may all contribute to the  $\Phi$ -statistic patterns (Stoneking 1998; Fix 1999; Karafet et al. 1999).

Our AMOVA results support two tentative conclusions: (1) patrilocality effects are evident at local and regional scales rather than at the intercontinental and

global levels of analysis, and (2) sole reliance on *F*<sub>ST</sub> values based on Wright's (1969) island model of population structure may result in distorted pictures of the geographic extent of the patrilocality effect and of the possible markedly different sex-specific migration rates noted above.

Tripartite Division of Global NRY Variation

The multidimensional scaling plot in figure 2 underscores the distinctiveness of Native American and African populations with respect to Eurasian and Oceanian populations seen in the AMOVA results (table 3). The pattern of low Native American NRY diversity is also concordant with the position of the Americas as an outlier in figure 2. These results fit a scenario whereby a combination of relatively recent colonization, repeated founder effects, small population sizes, and extensive intergenerational genetic drift is responsible for the dis-

**Table 3**  
**Structure of Y-Chromosome Haplotype Variation for 50 Worldwide Populations**

GROUPING	VARIANCE COMPONENTS (%)			$\Phi$ STATISTICS <sup>a</sup>		
	Within Populations	Among Groups	Among Populations Within Groups	$\Phi_{ST}$	$\Phi_{CT}$	$\Phi_{SC}$
	1. One global	64.0			0.360	
2. Ten regional	62.6	23.7	13.7	0.374	0.237	0.180
3. Five continental	60.6	23.6	15.7	0.394	0.236	0.206
4. Africa/Americas/"rest of the world"	50.8	33.2	16.1	0.492	0.332	0.240
5. African/non-African	45.9	36.5	17.6	0.541	0.365	0.278
6. European <sup>b</sup> /non-European	62.6	3.1	34.4	0.374	0.031	0.354
7. Asian/non-Asian	62.7	4.2	33.2	0.373	0.042	0.346
8. Oceanian/non-Oceanian	62.6	2.5	34.8	0.374	0.025	0.357
9. American/non-American	52.1	22.5	25.4	0.479	0.225	0.328

<sup>a</sup> All values were statistically significant at *P* < 0.0001 except  $\Phi_{CT}$  for comparison 7, for which *P* = 0.007, and for comparisons 6 and 8, for which *P* was not significant.  
<sup>b</sup> European and Middle Eastern populations combined.

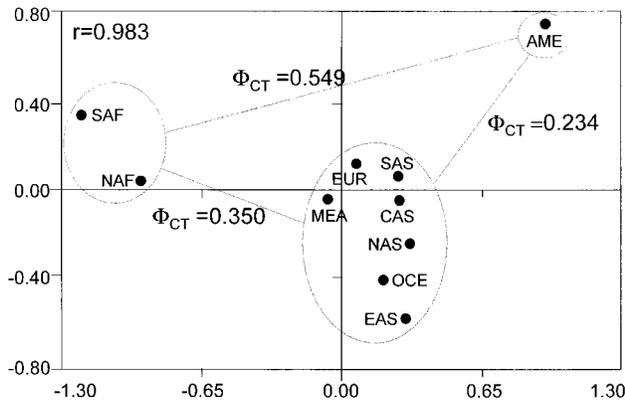


FIG. 2.—MDS plot of 10 regional populations based on  $\Phi_{ST}$  genetic distances, and  $\Phi_{CT}$  values for three intercluster comparisons. For three-letter population codes, see *Materials and Methods*.

tinctiveness of Native Americans with respect to their Asian forebears, as well as the remainder of the world (Karafet et al. 1999). As in many other genetic studies (Vigilant et al. 1991; Nei and Roychoudhury 1993; Cavalli-Sforza, Menozzi, and Piazza 1994), African populations occupy a distinct region of the multidimensional space in figure 2. However, unlike those of the Americas, sub-Saharan African populations are characterized by a diverse set of ancient haplotypes that are not shared globally (e.g., basal haplotypes h1–h10 in fig. 1) in combination with a set of more derived haplotypes that are widely shared within Africa and, again, are not shared globally (e.g., h13 and h15). Therefore, the distinctiveness of African populations better fits a scenario of African-specific lineage admixture (Labuda, Zietkiewicz, and Yotova 2000).

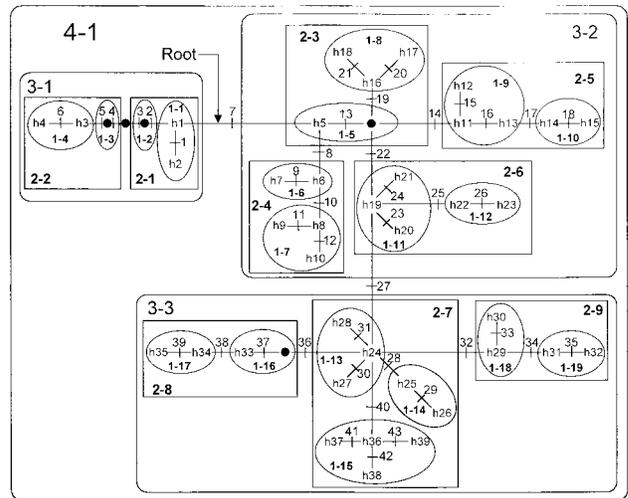


FIG. 3.—Nested cladistic design for 39 NRY haplotypes. The 43 mutational events listed in table 1 are shown by cross-hatches. Haplotypes h1–h39 are named as shown in table 1. The root of the cladogram is denoted by an arrow. Filled circles represent haplotypes that were missing in this sample of Y chromosomes. Ovals contain one-step clades which are designated 1-1 through 1-19. Rectangles contain two-step clades which are designated 2-1 through 2-9. Rounded rectangles contain three-step clades which are designated 3-1 through 3-3. A single four-step clade (4-1) encompasses the entire cladogram.

Although North Africa occupies a position relatively close to sub-Saharan Africa in figure 2, when traditional  $F_{ST}$  and CHORD distance statistics were employed (data not shown), North Africa moved closer to the Middle Eastern and European portion of the central cluster, as might be expected from ethnohistoric connections between North Africa and the Middle East

**Table 4**  
Main Inferences from Results of Nested Cladistic Analysis

Clade <sup>a</sup>	Inference Chain <sup>b</sup>	Inference <sup>c</sup>	Geography <sup>d</sup> /Population
1-1	1 → 2 → 11 → 12 No	CRE	Within Africa
1-6	1 → 2 → 3 → 4 No	RGF/IBD	Within Pygmy/Khoisan
1-7	1 → 2 → 3 → 4 No	RGF/IBD	Within Pygmy/Khoisan
1-8	1 → 2 → 3 → 5 → 6 → 13 Yes	LDC	From Asia to OCE (Polynesia)
1-9	1 → 2 → 11 → 12 → 13 Yes	LDC	Asia to Africa
1-10	1 → 2 → 11 → 12 No	CRE	NAF to sub-Saharan Africa
1-11	1 → 2 → 11 → 12 No	CRE	SAS to NAF, EUR/MEA
1-13	1 → 2 → 3 → 4 No	RGF/IBD	Within Asia and OCE
1-14	1 → 2 → 3 → 4 No	RGF/IBD	Within EAS
1-15	1 → 2 → 3 → 5 → 6 → 7 Yes	RGF/LDD	Within Asia and from Asia to EUR/AME
1-18	1 → 2 → 3 → 4 No	RGF/IBD	Within Asia
2-3	1 → 2 → 11 → 12 No	CRE	Africa to Asia, Asia to AME and OCE
2-5	1 → 2 → 3 → 4 No	RGF/IBD	Asia to Africa, Africa to EUR/MEA
2-6	1 → 2 → 3 → 4 No	RGF/IBD	Within SAS, NAF, and EUR/MEA
2-7	1 → 2 → 11 → 12 → 13 Yes	LDC	Out of Asia to EUR and AME
2-8	1 → 2 → 3 → 4 No	RGF/IBD	Within OCE (Melanesia)
2-9	1 → 2 → 11 → 12 No	CRE	Within EAS
3-1	1 → 2 → 3 → 4 No	RGF/IBD	Within Africa
3-2	1 → 2 → 3 → 4 No	RGF/IBD	Global
3-3	1 → 2 → 3 → 5 → 6 → 13 Yes	LDC	Within Asia and from Asia to OCE
Total	1 → 2 → 11 → 12 No	CRE	Out of Africa

<sup>a</sup> Only clades resulting in the rejection of the null hypothesis are included.

<sup>b</sup> See [http://bioag.byu.edu/zoology/crandall\\_lab/geodis.htm](http://bioag.byu.edu/zoology/crandall_lab/geodis.htm) for inference key.

<sup>c</sup> CRE = contiguous range expansion; RGF/IBD = recurrent gene flow restricted by isolation by distance; LDC = long-distance colonization; LDD = long-distance dispersal.

<sup>d</sup> Three-letter geographic codes are defined in *Materials and Methods*.

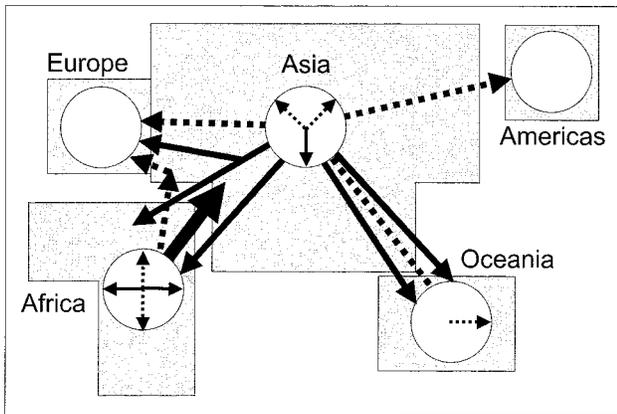


FIG. 4.—Inferences from nested cladistic analysis of Y-chromosome data. Intercontinental signals are indicated by arrows between continent ideograms (note: arrows are not meant to indicate routes of migration), and intracontinental signals are shown by arrows within circles (empty circles for Europe and the Americas denote the absence of intracontinental signals). Solid arrows represent population history events (contiguous range expansions and long-distance colonizations), while population structure processes (recurrent gene flow restricted by isolation by distance and long-distance dispersals) are indicated by dashed arrows (and, in one instance, a dashed line between Asia and Oceania where no polarity could be inferred). The widest solid arrow denotes early range expansion out of Africa at the level of the total cladogram.

(Cavalli-Sforza, Menozzi, and Piazza 1994), thereby producing a pattern similar to that depicted in the maximum-likelihood network of Underhill et al. (2000).

What distinguishes our results in figure 2 from autosomal global genetic analyses is the particular subdivision pattern that emerges, wherein Africans and Native Americans occupy opposite ends of the plot, while populations from Europe, Asia, and Oceania form a large, central cluster. For instance, Cavalli-Sforza, Menozzi, and Piazza's (1994, p. 82) principal-components ordination shows Africa clearly differentiated from the rest of the world; however, the Americas fall within the northern Eurasian portion of their map, which is separated from a southern Asian/Oceanian cluster. In other surveys, Africa and Oceania are frequently positioned as the outliers (Nei and Roychoudhury 1993; Stoneking et al. 1997). One possible reason for the distinctiveness of the haploid NRY pattern compared with diploid autosomal patterns is the stronger effect of genetic drift because of the smaller effective population size of the NRY (i.e.,  $NRY N_e = 1/4$  autosomal  $N_e$ ).

The very similar and extremely low  $\Phi_{CT}$  values for the European/non-European, Asian/non-Asian, and Oceanian/non-Oceanian comparisons in table 3, as well as the high  $\Phi_{ST}$  and  $\Phi_{CT}$  values for the African/non-African, American/non-American, and Africa/Americas/"rest of the world" comparisons, coincide with the observed pattern of global Y-chromosome diversity portrayed in figure 2. When  $\Phi_{CT}$  values were calculated for the three intercluster comparisons in figure 2, the African/Native American comparison showed the largest between-groups differentiation ( $\Phi_{CT} = 0.549$ ), while the central cluster was less differentiated from the Americas ( $\Phi_{CT} = 0.234$ ) than from Africa ( $\Phi_{CT} = 0.350$ ), in ac-

cord with ethnohistoric evidence (Cavalli-Sforza, Menozzi, and Piazza 1994; Hammer and Zegura 1996; Crawford 1998; Karafet et al. 1999; Cavalli-Sforza 2000).

#### Nested Cladistic Analysis as a Synthetic Explanatory Tool

In order to understand the causal mechanisms underlying the pattern of NRY variation reflected in the MDS plot and AMOVA results, the spatial distribution of our global NRY database was investigated by NCA (Templeton, Routman, and Phillips 1995; Hammer et al. 1998). In figure 4, inter- and intracontinental population history events (contiguous range expansions and long-distance colonizations) are depicted by solid arrows, while population structure processes (recurrent gene flow restricted by isolation by distance and long-distance dispersals) are indicated by dashed arrows (and one dashed line between Asia and Oceania, for which polarity could not be inferred). It is clear from figure 4 that both population structure and history have played important roles in shaping patterns of global NRY variation.

One of the most notable findings from the NCA analysis was the predominance of intercontinental signals detected emanating from Asia (fig. 4). These multiple out-of-Asia signals included gene flow episodes to Europe and the Americas, along with range expansions to Oceania, Africa, and Europe. In contrast, the NCA only detected two out-of-Africa signals. These NCA inferences help to explain the MDS plot (i.e., Asia's membership in the central cluster), the AMOVA results (i.e., lack of significant differentiation of Eurasian and Oceanian populations), and the diversity statistics (i.e., similar patterns of diversity in Asia and Oceania). Contrary to previously published studies of mtDNA (Redd et al. 1999) and autosomal markers (Harding et al. 1997; Stoneking et al. 1997), the NRY results suggest a strong affinity between mainland Asian and Oceanian populations. This different pattern may be due to either ascertainment bias in our NRY database or higher rates of male migration between Asia and Oceania. Support for the latter conjecture comes from the two long-distance colonization events, as well as a gene flow signal detected between Asia and Oceania in the NCA.

The fact that Europe is primarily a receiver rather than a sender of signals in figure 4 underscores the importance of gene flow/population movements into this continent. It also helps to explain Europe's central position in the MDS plot, its high  $h$  and  $p$  diversity statistic values, the concordant pattern of the three  $\Phi$  statistics for the European/non-European and Asian/non-Asian comparisons, and the observation that Europe has the lowest continental  $\Phi_{ST}$  value (see below). Interestingly, all incoming signals to Europe came from Asia. Two of these signals appear to have originated in Asia, one being a long-distance dispersal (from within nested clade 1-15) and the other being a contiguous range expansion (from within nested clade 1-11). The third was a gene flow signal that may have actually originated in Africa before moving to the Levant and eventually to Europe.

This latter signal was postulated to result from the Neolithic demic diffusion of Levantine farmers into Europe (Hammer et al. 1998) and corresponds to Semino et al.'s (2000) Eu4 lineage. The former two signals may well correspond to the two proposed Paleolithic migratory episodes that contributed a major portion of the modern European paternal gene pool (Semino et al. 2000).

After an early out-of-Africa range expansion (widest arrow in fig. 4), the majority of signals involving Africa were intracontinental events and processes. To explain this in the context of the two different sets of NRY haplotypes in sub-Saharan Africa (i.e., an ancient set of haplotypes overlaid by derived shared haplotypes), one probably needs a layered temporal framework whereby, for instance, early subdivision with inferred gene flow between the Khoisan and Pygmies (e.g., table 4) is combined with later, more extensive gene flow and historical events such as the Bantu expansion (Cavalli-Sforza, Menozzi, and Piazza 1994). These inferences are compatible with the model put forward by Labuda, Zietkiewicz, and Yotova (2000) in which the gene pool of sub-Saharan Africans is seen to be composed of two clades that evolved separately and then eventually underwent hybridization.

#### Comparative Framework: NRY and mtDNA Patterns in Sub-Saharan Africa

Our results may help to inform the debate concerning the conflicting patterns observed in sub-Saharan African mitochondrial and nuclear DNA. A basic inconsistency has been noted concerning the relative branch lengths in population trees for sub-Saharan African and non-African populations (Jorde et al. 1995). In contrast to non-African populations, sub-Saharan African populations appear to be well differentiated in mtDNA-based trees, suggesting that they have been subdivided for an extended period (Mountain 1998). Population trees based on nuclear polymorphisms do not show this pattern: for example, sub-Saharan African populations appear to be more closely related, and non-African and sub-Saharan African branches are more comparable in length. This discrepancy has not been satisfactorily explained by models incorporating an ascertainment bias in nuclear polymorphisms, a higher substitution (and homoplasy) rate for mtDNA, limited sample sizes, or various population-level factors (e.g., size changes) (Jorde et al. 1995; Mountain 1998). Additional explanatory factors have been suggested, including lack of selective neutrality in mtDNA and differences in male versus female migration rates and/or effective sizes within sub-Saharan Africa (Jorde et al. 1995).

We examined this problem from the perspective of the NRY by undertaking two new analyses. We wanted to know (1) whether NRY-based genetic distances within sub-Saharan Africa were smaller than those for non-African locales, and (2) if NRY data showed less differentiation among sub-Saharan African populations than did mtDNA data. First,  $\Phi_{ST}$  genetic distances were calculated for each continent separately. Asia had the highest  $\Phi_{ST}$  value (0.271), followed by Africa (0.222),

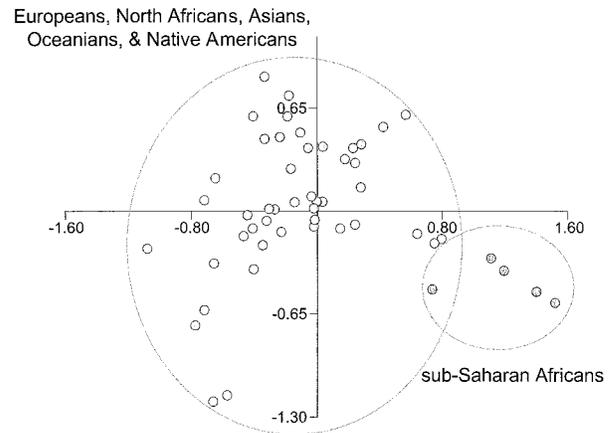


FIG. 5.—MDS plot of 50 populations based on  $\Phi_{ST}$  genetic distances.

the Americas (0.188), Oceania (0.133), and Europe (0.128). Thus, at least one non-African locale (Asia) had larger genetic distances than Africa. When sub-Saharan Africa was analyzed separately, its  $\Phi_{ST}$  of 0.251 was still smaller than that of Asia. In contrast, mtDNA data typically show much greater among-groups variation for sub-Saharan African populations than for non-African groups. For instance, Melton et al. (1997) reported a  $\Phi_{ST}$  value of 0.339 for sub-Saharan Africa, compared with values of 0.045 and 0.007 for Asian and European populations, respectively. The second new analysis consisted of an MDS plot for all 50 populations (fig. 5). Here, sub-Saharan African populations were more tightly clustered than they were in Excoffier et al.'s (1996) mtDNA-based plot, and the sub-Saharan African populations were also more tightly clustered than non-African populations, the exact opposite of the mtDNA pattern. Moreover, the overall pattern of sub-Saharan African NRY phylogeography is closer to other nuclear system results (Cavalli-Sforza, Menozzi, and Piazza 1994; Stoneking et al. 1997) than to those for its haploid mtDNA counterpart.

We favor an explanation for these contradictory mtDNA and NRY results that involves, at least in part, a higher male-versus-female migration rate in sub-Saharan Africa. Obviously, this suggestion conflicts with our previous statements about possible local and regional patterns of greater female migration for the rest of the world; however, it is concordant with the recent results of Carvajal-Carmona et al. (2000) and Mesa et al. (2000) on sex-specific gene flow patterns over the last few hundred years in Colombia, South America. The discordant NRY/mtDNA patterns also fit a model involving male-biased gene flow during the Bantu expansion, which in turn could have produced the widespread sub-Saharan African distribution of a single haplotype (h15) in clade 1-10 (previously called YAP<sup>+</sup> haplotype 5 in Hammer et al. [1998]). The importance of African male migration is also underscored by the observation that four of the eight intracontinental signals detected in the NCA involve Africa. These gene flow episodes and two range expansions certainly add to the impression that male dispersal rates are quite high within Africa. It

is also possible that other factors (such as selection; see next section) are responsible for producing the discordant picture of mtDNA and NRY results (Hey 1997).

### Conclusions, Caveats, and Future Directions

Previous global nested cladistic analyses of human NRY variation (Hammer et al. 1998; Karafet et al. 1999) have demonstrated patterns of diversity unlike those provided by mtDNA (Templeton 1993, 1997, 1999) or autosomal systems (Harding et al. 1997). For instance, Templeton's (1993, 1997, 1999) nested cladistic analyses of human mtDNA data are all highlighted at the deepest level by pervasive gene flow restricted by isolation by distance throughout Africa and southern Eurasia for the entire time to the most recent common ancestor (TMRCA) of mtDNA. A similar extensive worldwide Late Pleistocene gene flow signal was detected at the  $\beta$ -globin locus (Harding et al. 1997; Templeton 1999), the only global autosomal data set analyzed by Templeton, Routman, and Phillips' (1995) nested cladistic procedures. In contrast, all three of our nested cladistic analyses detected a global contiguous range expansion out of Africa at the level of the entire cladogram. In the present NCA, the two deepest gene flow signals were only at the three-step level: one occurred globally, while the other was restricted to the continent of Africa. Our new results support a general scenario in which, after an early out-of-Africa range expansion, global-scale patterns of NRY variation were mainly influenced by migrations out of Asia. Moreover, the greater degree of contact detected by the NCA among Asia, Europe, and Oceania (via both population structure processes and population history events) helps to explain the observed pattern of global NRY diversity.

A major conclusion of the present work is that global human NRY variation is structured, with a significant amount of intergroup variation partitioned among African, Native American, and Eurasian/Oceanian populations. There was also a significant degree of among-populations variation at the intracontinental level; the degree of structure at lower levels of population subdivision remains to be determined.

It should be noted that the pattern of subdivision detected here could also be explained by models that involve natural selection or a combination of microevolutionary forces including selection, migration, genetic drift, and mutation. Additionally, various human social processes, such as polygyny and kin-structured migration, may affect variation on the NRY (Fix 1999). Support for a model involving selection comes from recent findings demonstrating an excess of rare alleles at sites on the NRY (Underhill et al. 1997; Pritchard et al. 1999; Shen et al. 2000; Thomson et al. 2000). Indeed, our mutation screening at the *DYS188*, *DYS190*, and *DYS194* sites on a panel of 58 Y chromosomes from worldwide samples also yielded a significant excess of singleton polymorphisms. This excess of singletons (>twofold more than expected under the hypothesis of constant population size) resulted in a significantly negative Fu and Li's (1997)  $F^*$  statistic of  $-2.67$  ( $P <$

$0.05$ ). These findings are consistent with models based on positive directional selection, expansion from a small population size, and/or ascertainment bias resulting from poor sampling of a subdivided population system. Obviously, more research needs to be focused on distinguishing the possible causes and implications of population subdivision in the human paternal gene pool.

It is important to note the limitations of the different methods employed in this study, as well as their complementary nature for inferring the underlying forces shaping NRY variation in human populations. The NCA is stronger at making inferences toward the interior of a cladogram and weaker at inferring processes/events at the tips. Therefore, as more polymorphisms are discovered (see, e.g., Underhill et al. 2000) and the NRY tree becomes more resolved, more inferences concerning regional variation will emerge. It is also possible that some of the inferences made here will change as more data are collected. Consequently, we have mainly focused on general patterns and have not tried to explain all of the specific signals detected by the NCA. Finally, these methods do not distinguish selection from demographic forces in shaping patterns of diversity.

Standard approaches for the description of population structure based on Wright's (1969) island model and/or  $F$  statistics (e.g., AMOVA) do not attempt to disentangle past events from contemporary processes and thus can be considered nonhistorical (Turner et al. 2000). The combination of nested cladistic and coalescence analyses can theoretically provide the temporal framework for making these crucial distinctions (Schaal and Olsen 2000). For instance, coalescence analysis could provide the missing dates needed to clarify the relative chronology of the many signals in figure 4. Our two previous coalescence analyses of Y-chromosome data (Hammer et al. 1998; Karafet et al. 1999) were performed without population growth or subdivision parameters in the model. Growth has been shown to decrease TMRCA estimates (Pritchard et al. 1999; Thomson et al. 2000). We are presently collaborating with R. C. Griffiths who is developing coalescence analyses incorporating both population growth and subdivision. Population growth should decrease our previously published mutational ages and TMRCA estimates (Hammer et al. 1998; Karafet et al. 1999), while population subdivision should have the opposite effect.

### Acknowledgments

We gratefully acknowledge the excellent laboratory assistance of Roxane Bonner, Matthew Kaplan, Agnish Chakravarti, Christine Ponder, Ji Park, Abdel-Halim Salem, Hwayong Park, and Ammon Corl. We thank Elizabeth Wood, Tasha Altheide, Christopher Tillquist, Robert Griffiths, Alan Templeton, and Michael Nachman for helpful discussions and comments on earlier versions of the manuscript. We also thank our generous collaborators who provided DNA samples. This publication was made possible by grant GM-53566 from the National Institute of General Medical Sciences and grant OPP-9806759 from the National Science Foundation (to

M.F.H.). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH or the NSF.

## LITERATURE CITED

- ALLEN, B. S., and H. OSTRER. 1994. Conservation of human Y chromosome sequences among male great apes: implications for the evolution of Y chromosomes. *J. Mol. Evol.* **39**:13–21.
- BAO, W., S. ZHU, A. PANDYA, T. ZERJAL, J. XU, Q. SHU, R. DU, H. YANG, and C. TYLER-SMITH. 2000. MSY2: a slowly evolving minisatellite on the human Y chromosome which provides a useful polymorphic marker in Chinese populations. *Gene* **244**:29–33.
- BARBUJANI, G., A. MAGAGNI, E. MINCH, and L. L. CAVALLI-SFORZA. 1997. An apportionment of human DNA diversity. *Proc. Natl. Acad. Sci. USA* **94**:4516–4519.
- BATZER, M. A., M. STONEKING, M. ALEGRIA-HARTMAN et al. (11 co-authors). 1994. African origin of human-specific polymorphic Alu insertions. *Proc. Natl. Acad. Sci. USA* **91**:12288–12292.
- BOWCOCK, A., and L. L. CAVALLI-SFORZA. 1991. The study of variation in the human genome. *Genomics* **11**:491–498.
- CARVAJAL-CARMONA, L. G., I. D. SOTO, N. PINEDA et al. (11 co-authors). 2000. Strong Amerind/White sex bias and a possible Sephardic contribution among the founders of a population in northwest Colombia. *Am. J. Hum. Genet.* **67**:1287–1295.
- CAVALLI-SFORZA, L. L. 2000. *Genes, peoples, and languages*. North Point Press, New York.
- CAVALLI-SFORZA, L. L., P. MENOZZI, and A. PIAZZA. 1994. *The history and geography of human genes*. Princeton University Press, Princeton, N.J.
- CAVALLI-SFORZA, L. L., and E. MINCH. 1997. Paleolithic and Neolithic lineages in the European mitochondrial gene pool. *Am. J. Hum. Genet.* **61**:247–254.
- CRAWFORD, M. H. 1998. *The origins of Native Americans: evidence from anthropological genetics*. Cambridge University Press, Cambridge, England.
- DEKA, R., M. D. SHRIVER, L. M. YU, R. E. FERRELL, and R. CHAKRABORTY. 1995. Intra- and inter-population diversity at short tandem repeat loci in diverse populations of the world. *Electrophoresis* **16**:1659–1664.
- EXCOFFIER, L., E. S. POLONI, S. SANTACHIARA-BENERECETTI, O. SEMINO, and A. LANGANEY. 1996. The molecular diversity of the Niokholo Mandenka from Eastern Senegal: an insight into West Africa genetic history. Pp. 141–155 in A. J. BOYCE and C. G. N. MASCIE-TAYLORS, eds. *Molecular biology and human diversity*. Cambridge University Press, Cambridge, England.
- EXCOFFIER, L., P. E. SMOUSE, and J. M. QUATTRO. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**:479–491.
- FIX, A. 1999. *Migration and colonization in human microevolution*. Cambridge University Press, New York.
- FU, Y., and W.-H. LI. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**:915–925.
- HAMMER, M. F. 1995. A recent common ancestry for human Y-chromosomes. *Nature* **378**:376–378.
- HAMMER, M. F., and S. HORAI. 1995. Y chromosomal DNA variation and the peopling of Japan. *Am. J. Hum. Genet.* **56**:951–962.
- HAMMER, M. F., T. KARAFET, A. RASANAYAGAM, E. T. WOOD, T. K. ALTHEIDE, T. JENKINS, R. C. GRIFFITHS, A. R. TEMPLETON, and S. L. ZEGURA. 1998. Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. *Mol. Biol. Evol.* **15**:427–441.
- HAMMER, M. F., A. J. REDD, E. T. WOOD et al. (12 co-authors). 2000. Jewish and middle eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes. *Proc. Natl. Acad. Sci. USA* **97**:6769–6774.
- HAMMER, M. F., A. B. SPURDLE, T. KARAFET et al. (11 co-authors). 1997. The geographic distribution of human Y chromosome variation. *Genetics* **145**:787–805.
- HAMMER, M. F., and S. L. ZEGURA. 1996. The role of the Y chromosome in human evolutionary studies. *Evol. Anthropol.* **5**:116–134.
- HARDING, R. M., S. M. FULLERTON, R. C. GRIFFITHS, J. BOND, M. J. COX, J. A. SCHNEIDER, D. S. MOULIN, and J. B. CLEGG. 1997. Archaic African and Asian lineages in the genetic ancestry of modern humans. *Am. J. Hum. Genet.* **60**:772–789.
- HEY, J. 1997. Mitochondrial and nuclear genes present conflicting portraits of human origins. *Mol. Biol. Evol.* **14**:166–172.
- JIN, L., and R. CHAKRABORTY. 1995. Population structure, stepwise mutations, heterozygote deficiency and their implications in DNA forensics. *Heredity* **74**:274–285.
- JOBLING, M. A., V. SAMARA, A. PANDYA et al. (16 co-authors). 1996. Recurrent duplication and deletion polymorphisms on the long arm of the Y chromosome in normal males. *Hum. Mol. Genet.* **5**:1767–1775.
- JOBLING, M. A., and C. TYLER-SMITH. 1995. Fathers and sons: the Y chromosome and human evolution. *Trends Genet.* **11**:449–456.
- JORDE, L. B., M. J. BAMSHAD, W. S. WATKINS, R. ZENGER, A. E. FRALEY, P. A. KRAKOWIAK, K. D. CARPENTER, H. SOODYALL, T. JENKINS, and A. R. ROGERS. 1995. Origins and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data. *Am. J. Hum. Genet.* **57**:523–538.
- JORDE, L. B., W. S. WATKINS, M. J. BAMSHAD, M. E. DIXON, C. E. RICKER, M. T. SEIELSTAD, and M. A. BATZER. 2000. The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *Am. J. Hum. Genet.* **66**:979–988.
- KARAFET, T. M., S. L. ZEGURA, O. POSUKH et al. (14 co-authors). 1999. Ancestral Asian source(s) of New World Y-chromosome founder haplotypes. *Am. J. Hum. Genet.* **64**:817–831.
- KARAFET, T., S. L. ZEGURA, J. VUTURO-BRADY et al. (14 co-authors). 1997. Y chromosome markers and trans-Bering Strait dispersals. *Am. J. Phys. Anthropol.* **102**:301–314.
- KITTLES, R. A., A. W. BERGEN, M. URBANEK, M. VIRKKUNEN, M. LINNOILA, D. GOLDMAN, and J. C. LONG. 1999. Autosomal, mitochondrial, and Y chromosome DNA variation in Finland: evidence for a male-specific bottleneck. *Am. J. Phys. Anthropol.* **108**:381–399.
- KRUSKAL, J. B. 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **29**:1–27.
- LABUDA, D., E. ZIETKIEWICZ, and V. YOTOVA. 2000. Archaic lineages in the history of modern humans. *Genetics* **156**:799–808.
- LATTER, B. D. H. 1980. Genetic differences within and between populations of the major human subgroups. *Am. Nat.* **116**:220–237.
- LEWONTIN, R. C. 1972. The apportionment of human diversity. *Evol. Biol.* **6**:381–398.
- MELTON, T., C. GINTHER, G. SENSABAUGH, H. SOODYALL, and M. STONEKING. 1997. Extent of heterogeneity in mitochon-

- drial DNA of sub-Saharan African populations. *J. Forensic Sci.* **42**:582–592.
- MESA, N. R., M. C. MONDRAGON, I. D. SOTO et al. (13 co-authors). 2000. Autosomal, mtDNA, and Y-chromosome diversity in Amerinds: pre- and post-Columbian patterns of gene flow in South America. *Am. J. Hum. Genet.* **67**:1277–1286.
- MOUNTAIN, J. L. 1998. Molecular evolution and modern human origins. *Evol. Anthropol.* **7**:21–37.
- MURDOCK, G. P. 1967. *Ethnographic atlas*. University of Pittsburgh Press, Pittsburgh, Pa.
- NEI, M. 1987. *Molecular evolutionary genetics*. Columbia University Press, New York.
- NEI, M., and A. K. ROYCHOUDHURY. 1974. Genic variation within and between the three major races of man, Caucasoids, Negroids, and Mongoloids. *Am. J. Hum. Genet.* **26**:421–443.
- . 1993. Evolutionary relationships of human populations on a global scale. *Mol. Biol. Evol.* **10**:927–943.
- PEREZ-LEZAUN, A., F. CALAFELL, D. COMAS et al. (12 co-authors). 1999. Sex-specific migration patterns in Central Asian populations, revealed by analysis of Y-chromosome short tandem repeats and mtDNA. *Am. J. Hum. Genet.* **65**:208–219.
- POLONI, E. S., O. SEMINO, G. PASSARINO, A. S. SANTACHIARA-BENERECETTI, I. DUPANLOUP, A. LANGANEY, and L. EXCOFFIER. 1997. Human genetic affinities for Y-chromosome P49a,f/*TaqI* haplotypes show strong correspondence with linguistics. *Am. J. Hum. Genet.* **61**:1015–1035.
- POSADA, D., K. A. CRANDALL, and A. R. TEMPLETON. 2000. GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Mol. Ecol.* **9**:487–488.
- PRITCHARD, J. K., M. T. SEIELSTAD, A. PEREZ-LEZAUN, and M. W. FELDMAN. 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol. Biol. Evol.* **16**:1791–1798.
- PRZEWORSKI, M., R. R. HUDSON, and A. DI RIENZO. 2000. Adjusting the focus on human variation. *Trends Genet.* **16**:296–302.
- QAMAR, R., Q. AYUB, S. KHALIQ, A. MANSOOR, T. KARAFET, S. Q. MEHDI, and M. F. HAMMER. 1999. African and Levantine origins of Pakistani YAP<sup>+</sup> Y chromosomes. *Hum. Biol.* **71**:745–755.
- REDD, A. J., and M. STONEKING. 1999. Peopling of Sahul: mtDNA variation in aboriginal Australian and Papua New Guinean populations. *Am. J. Hum. Genet.* **65**:808–828.
- RELETHFORD, J. H. 1995. Genetics and modern human origins. *Evol. Anthropol.* **4**:53–63.
- RELETHFORD, J. H., and H. C. HARPENDING. 1994. Craniometric variation, genetic theory, and modern human origins. *Am. J. Phys. Anthropol.* **95**:249–270.
- ROHLF, F. J. 1998. NTSYS-pc: numerical taxonomy and multivariate analysis system. Release 2.02H. Exeter Software, Setauket, N.Y.
- SANTOS, F. R., A. PANDYA, M. KAYSER et al. (13 co-authors). 2000. A polymorphic L1 retroposon insertion in the centromere of the human Y chromosome. *Hum. Mol. Genet.* **9**:421–430.
- SCHAAL, B. A., and K. M. OLSEN. 2000. Gene genealogies and population variation in plants. *Proc. Natl. Acad. Sci. USA* **97**:7024–7029.
- SCHNEIDER, S., J.-M. KUEFFER, D. ROESSLI, and L. EXCOFFIER. 1998. Arlequin: a software for population genetic analysis. Release 1.1. Genetics and Biometry Laboratory, University of Geneva, Geneva, Switzerland.
- SCOZZARI, R., F. CRUCIANI, P. SANTOLAMAZZA et al. (17 co-authors). 1999. Combined use of biallelic and microsatellite Y-chromosome polymorphisms to infer affinities among African populations. *Am. J. Hum. Genet.* **65**:829–846.
- SEIELSTAD, M. T., E. MINCH, and L. L. CAVALLI-SFORZA. 1998. Genetic evidence for a higher female migration rate in humans. *Nat. Genet.* **20**:278–280.
- SEMINO, O., G. PASSARINO, P. J. OEFNER et al. (17 co-authors). 2000. The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. *Science* **290**:1155–1159.
- SHEFIELD, V. C., J. S. BECK, A. E. KWITEK, D. W. SANDSTROM, and E. M. STONE. 1993. The sensitivity of single-strand conformation polymorphism analysis for the detection of single base changes. *Genomics* **16**:325–332.
- SHEN, P., F. WANG, P. A. UNDERHILL et al. (13 co-authors). 2000. Population genetic implications from sequence variation in four Y chromosome genes. *Proc. Natl. Acad. Sci. USA* **97**:7354–7359.
- SHINKA, T., K. TOMITA, T. TODA, S. E. KOTLIAROVA, J. LEE, Y. KUROKI, D. K. JIN, K. TOKUNAGA, H. NAKAMURA, and Y. NAKAHORI. 1999. Genetic variations on the Y chromosome in the Japanese population and implications for modern human Y chromosome lineage. *J. Hum. Genet.* **44**:240–245.
- SOMMER, S. S., A. R. GROSZBACH, and C. D. BOTTEMA. 1992. PCR amplification of specific alleles (PASA) is a general method for rapidly detecting known single-base changes. *Biotechniques* **12**:82–87.
- STONEKING, M. 1998. Women on the move. *Nat. Genet.* **20**:219–220.
- STONEKING, M., J. J. FONTIUS, S. L. CLIFFORD, H. SOODYALL, S. S. ARCOT, N. SAHA, T. JENKINS, M. A. TAHIR, P. L. DEININGER, and M. A. BATZER. 1997. Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. *Genome Res.* **7**:1061–1071.
- SU, B., J. XIAO, P. UNDERHILL et al. (21 co-authors). 1999. Y-Chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age. *Am. J. Hum. Genet.* **65**:1718–1724.
- SWOFFORD, D. 2000. PAUP: phylogenetic analysis using parsimony. Release 4.0b4. Sinauer, Sunderland, Mass.
- TEMPLETON, A. R. 1993. The “Eve” hypothesis: a genetic critique and reanalysis. *Am. Anthropol.* **95**:51–72.
- . 1997. Testing the out-of-Africa replacement hypothesis with mitochondrial DNA data. Pp. 329–360 in G. A. CLARK and C. WILLERMET, eds. *Conceptual issues in modern human origins research*. Aldine de Gruyter, Amsterdam.
- . 1999. Human races: a genetic and evolutionary perspective. *Am. Anthropol.* **100**:632–650.
- TEMPLETON, A. R., E. BOERWINKLE, and C. F. SING. 1987. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in *Drosophila*. *Genetics* **117**:343–351.
- TEMPLETON, A. R., E. ROUTMAN, and C. A. PHILLIPS. 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.
- TEMPLETON, A. R., and C. F. SING. 1993. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses with cladogram uncertainty and recombination. *Genetics* **134**:659–669.
- THOMSON, R., J. K. PRITCHARD, P. SHEN, P. J. OEFNER, and M. W. FELDMAN. 2000. Recent common ancestry of human Y

- chromosomes: evidence from DNA sequence data. *Proc. Natl. Acad. Sci. USA* **97**:7360–7365.
- TURNER, T. F., J. C. TREXLER, J. L. HARRIS, and J. L. HAYNES. 2000. Nested cladistic analysis indicates population fragmentation shapes genetic diversity in a freshwater mussel. *Genetics* **154**:777–785.
- UNDERHILL, P. A., L. JIN, A. A. LIN, S. Q. MEHDI, T. JENKINS, D. VOLLRATH, R. W. DAVIS, L. L. CAVALLI-SFORZA, and P. J. OEFNER. 1997. Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res.* **7**:996–1005.
- UNDERHILL, P. A., P. SHEN, A. A. LIN et al. (21 co-authors). 2000. Y chromosome sequence variation and the history of human populations. *Nat. Genet.* **26**:358–361.
- URBANEK, M., D. GOLDMAN, and J. C. LONG. 1996. The apportionment of dinucleotide repeat diversity in Native Americans and Europeans: a new approach to measuring gene identity reveals asymmetric patterns of divergence. *Mol. Biol. Evol.* **13**:943–953.
- VIGILANT, L., M. STONEKING, H. HARPENDING, K. HAWKES, and A. C. WILSON. 1991. African populations and the evolution of human mitochondrial DNA. *Science* **253**:1503–1507.
- VOLLRATH, D., S. FOOTE, A. HILTON, L. G. BROWN, P. BEER-ROMERO, J. S. BOGAN, and D. C. PAGE. 1992. The human Y chromosome: a 43-interval map based on naturally occurring deletions. *Science* **258**:52–59.
- WRIGHT, S. 1969. *Evolution and the genetics of populations 2. the theory of gene frequencies.* University of Chicago Press, Chicago.
- ZERJAL, T., B. DASHNYAM, A. PANDYA et al. (18 co-authors). 1997. Genetic relationships of Asians and northern Europeans, revealed by Y-chromosome DNA analysis. *Am. J. Hum. Genet.* **60**:1174–1183.

KEITH CRANDALL, reviewing editor

Accepted March 2, 2001