Out of Africa and Back Again: Nested Cladistic Analysis of Human Y Chromosome Variation

M. F. Hammer, T. Karafet,*‡ A. Rasanayagam,* E. T. Wood,* T. K. Altheide,* T. Jenkins,§ R. C. Griffiths,‖ A. R. Templeton,¶ and S. L. Zegura†*

*Laboratory of Molecular Systematics and Evolution and †Department of Anthropology, University of Arizona; ‡Laboratory of Human Molecular and Evolutionary Genetics, Institute of Cytology and Genetics, Novosibirsk, Russia; §Department of Human Genetics, University of Witwatersrand, Johannesburg, South Africa; ‖Mathematics Department, Monash University, Clayton, Australia; and ¶Department of Biology, Washington University, St. Louis, Missouri

We surveyed nine diallelic polymorphic sites on the Y chromosomes of 1,544 individuals from Africa, Asia, Europe, Oceania, and the New World. Phylogenetic analyses of these nine sites resulted in a tree for 10 distinct Y haplotypes with a coalescence time of ~150,000 years. The 10 haplotypes were unevenly distributed among human populations: 5 were restricted to a particular continent, 2 were shared between Africa and Europe, 1 was present only in the Old World, and 2 were found in all geographic regions surveyed. The ancestral haplotype was limited to African populations. Random permutation procedures revealed statistically significant patterns of geographical structuring of this paternal genetic variation. The results of a nested cladistic analysis indicated that these geographical associations arose through a combination of processes, including restricted, recurrent gene flow (isolation by distance) and range expansions. We inferred that one of the oldest events in the nested cladistic analysis was a range expansion out of Africa which resulted in the complete replacement of Y chromosomes throughout the Old World, a finding consistent with many versions of the Out of Africa Replacement Model. A second and more recent range expansion brought Asian Y chromosomes back to Africa without replacing the indigenous African male gene pool. Thus, the previously observed high levels of Y chromosomal genetic diversity in Africa may be due in part to bidirectional population movements. Finally, a comparison of our results with those from nested cladistic analyses of human mtDNA and β-globin data revealed different patterns of inferences for males and females concerning the relative roles of population history (range expansions) and population structure (recurrent gene flow), thereby adding a new sex-specific component to models of human evolution.

Introduction

A holistic understanding of both the biocultural processes responsible for and the actual trajectory of human evolution ultimately requires the integration of data from genetics, paleontology, archaeology, linguistics, and ethnohistory (Cavalli-Sforza, Menozzi, and Piazza 1994). Within genetics, autosomal and haploid systems can offer complementary, but not necessarily identical, glimpses into our evolutionary history (Jorde et al. 1995). It is also possible that maternal (mtDNA) and paternal (the nonrecombining portion of the Y chromosome) lineages will offer different insights into the global dispersal of *Homo sapiens* (Cavalli-Sforza and Minch 1997). Stoneking (1993) underscored the growing utility of mtDNA data for understanding human maternal evolutionary pathways. Recently, Hammer and Zegura (1996) also presented an extensive review of the Y chromosome literature pertaining to human evolutionary studies.

In a subsequent data-oriented paper, Hammer et al. (1997) proposed that their Y chromosome combination haplotype distributions were compatible with multiple human migrations. The high haplotype diversity values found in sub-Saharan Africa were attributed to great population antiquity and/or large effective population size. Although most of their hypothesized dispersal events originated in Africa, one movement seemed to emanate from Asia. Unfortunately, the number of available Y chromosome haplotypes was deemed to be insufficient to perform Templeton's (1993) nested cladistic analysis, which provides a variety of statistical tests for the existence and causation of associations between haplotype distributions and geography. Since our earlier report, we have doubled the number of informative Y chromosome haplotypes from 5 to 10 and are now able to pursue a model-based testing approach to many of the hypotheses concerning the origins and subsequent migrations of the exclusively paternal component of our genetic heritage.

Key words: Y chromosome haplotypes, human evolution, nested cladistic test, coalescence times, population structure.

Address for correspondence and reprints: Michael Hammer, Department EEB, Biosciences West, University of Arizona, Tucson, Arizona 85721. E-mail: mhammer@u.arizona.edu.

Mol. Biol. Evol. 15(4):427–441. 1998

Materials and Methods

Sample Composition

We analyzed a total of 1,544 males from 35 populations (table 1). The approximate geographic locations of the study populations are shown in figure 1. Many of the samples examined in Hammer et al. (1997) and Karafet et al. (1997) were also used here, although the exact number of subjects from each population occasionally varies among the three studies. In addition, we included the following new samples: 4 West Africans (1 each from Cameroon, Ghana, Mali, and Togo), 20 Russians, 12 Kets, 50 Indonesians (28 from the Moluccas and 22 from the Nusa Tenggaras), 56 South Asian Indians (22 Kotas and 34 from Madras), 83 Sri Lankans (40 Tamils and 43 Sinhalese), 37 Melanesians from Vanuatu, and 14 Amerinds (3 Mayans, 3 Surui, 2 Karatiana, 1 Tohono O'Odham, 1 Porch Creek, 1 Sioux, 1 Shuswap, and 2 miscellaneous).

Table 1
Y Chromosome Haplotype Frequencies (%) in 35 Populations and 8 Geographic Regions ($N = 1,544$)

Population	*N*	Haplotype									
		2	1A	1B	1C	1D	1E	3G	3A	4	5
Sub-Saharan Africans	380	6	6	17	2	0*	0	0	8	5	57
1. Khoisan	70	31	20	17	1	0	0	0	0	10	20
2. Pygmies	38	0	0	45	0	0	0	0	5	0	50
3. West Africans	59	0	3	7	0	0	0	0	15	9	66
4. East Africans	44	0	0	2	0	0	0	0	11	0	86
5. East Bantus	95	0	4	26	0	0	0	0	11	4	55
6. West Bantus	54	0	0	6	7	2	0	0	4	4	78
7. Dama	20	5	5	5	5	0	0	0	10	5	65
Europeans	217	0	0	33	33	12	0	0	0*	22	0
8. British	44	0	0	11	73	9	0	0	0	7	0
9. Germans	21	0	0	38	33	19	0	0	0	10	0
10. Russians	20	0	0	25	20	50	0	0	0	5	0
11. Italians	48	0	0	31	35	2	0	0	2	29	0
12. Greeks	84	0	0	46	13	7	0	0	0	33	0
North Asians	235	0	0	83	15	2	0	0	0	0	0
13. Komi	21	0	0	100	0	0	0	0	0	0	0
14. West Siberians	46	0	0	74	24	2	0	0	0	0	0
15. Kets	12	0	0	8	83	8	0	0	0	0	0
16. Buryats	81	0	0	96	3	1	0	0	0	0	0
17. Evenks	55	0	0	98	0	2	0	0	0	0	0
18. Eskimos	20	0	0	30	65	5	0	0	0	0	0
Central Asians	101	0	0	51	11	20	0	19	0	0	0
19. Altai	31	0	0	32	16	48	0	3	0	0	0
20. Mongolians	43	0	0	74	14	9	0	2	0	0	0
21. Tibetans	27	0	0	33	0	4	0	63	0	0	0
East Asians	296	0	0	82	1	0*	0	17	0	0	0
22. Koreans	29	0	0	90	0	3	0	7	0	0	0
23. Japanese	97	0	0	52	2	0	0	46	0	0	0
24. South Chinese	59	0	0	98	2	0	0	0	0	0	0
25. Taiwanese	20	0	0	95	0	0	0	5	0	0	0
26. Indonesians	53	0	0	100	0	0	0	0	0	0	0
27. Southeast Asians	38	0	0	95	3	0	0	3	0	0	0
South Asians	155	0	0	72	13	15	0	0	0	0	0
28. Indians	72	0	0	68	17	15	0	0	0	0	0
29. Sri Lankans	83	0	0	76	10	15	0	0	0	0	0
Australasians	116	0	0	96	3	0	1	0	0	0	0
30. Papua New Guineans	46	0	0	100	0	0	0	0	0	0	0
31. Melanesians	37	0	0	95	3	0	3	0	0	0	0
32. Australian Ab. People	33	0	0	91	9	0	0	0	0	0	0
Native Americans	44	0	0	18	82	0	0	0	0	0	0
33. Tanana	8	0	0	75	25	0	0	0	0	0	0
34. Navajo	18	0	0	11	89	0	0	0	0	0	0
35. Amerinds	18	0	0	0	100	0	0	0	0	0	0

Note.—Frequencies were rounded off to the nearest integer. As a result, some haplotype rows do not sum to 100%. Zeros with an asterisk indicate the presence of a haplotype at a frequency greater than 0 but less than 0.005.

DNA Extraction

All new DNA samples referred to above were provided by co-authors except the following: 4 West Africans from G. Rappold, 50 Indonesians from M. Stoneking, 56 South Asian Indians from R. Herrera, and 37 Melanesians from J. Martinson. Eleven of the new DNA samples were isolated from cell lines in the Y Chromosome Consortium Repository, while the others were extracted from buccal cells as described in Hammer et al. (1997). All sampling protocols were approved by the Human Subjects Committee at the University of Arizona.

Genotyping Y Chromosome Markers

Whitfield, Sulston, and Goodfellow (1995) identified three polymorphic nucleotide sites within the SRY region on the short arm of the Y chromosome (Yp) and presented a haplotype tree based on these sites. In our paper, we combine data from four markers within the YAP region (Hammer 1995) with information from polymorphic sites within both the SRY region and a previously unpublished Y-linked STS marker (*DYS257*). A schematic representation of the Y chromosome, with marker positions within the SRY and YAP regions as well as the approximate position of the *DYS257* STS on Yp, is shown in figure 2. The SRY region sites include a G→A transition at position 4064 (SRY_{4064}), a C→T transition at position 9138 (SRY_{9138}), and an A→G transition at position 10831 (SRY_{10831}); while the YAP region sites include the presence/absence of the YAP element (YAP^+/YAP^-) between bases 621 and 622, a C→T transition at position 338 (PN1), a C→T transition at position 1682 (PN2), and a G→A transition at position 1926 (PN3) (table 2). These polymorphisms were

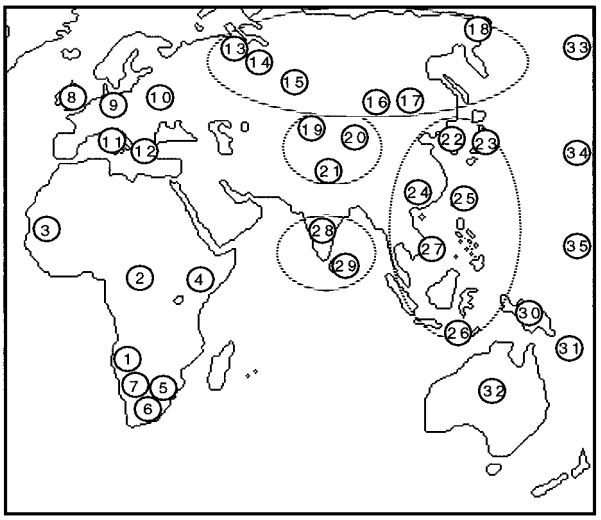

FIG. 1.—Approximate geographic locations of study populations. Numerical population codes within small circles are defined in table 1. The 17 Asian populations are subdivided into four geographic groupings (ovals with dotted lines): North Asians (13–18), Central Asians (19–21), East Asians (22–27), and South Asians (28–29).

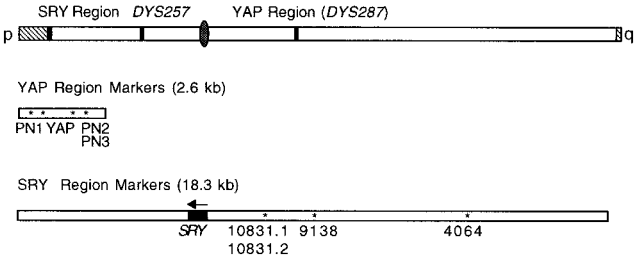

FIG. 2.—Schematic representation of the Y chromosome and marker positions within the specific genetic regions studied. The short arm (p), the centromere (shaded oval), and the long arm (q) are indicated in the top illustration. The three dark vertical lines refer to the SRY region, *DYS257*, and the YAP region, respectively. The hatched boxes at the termini of the chromosome represent the pseudoautosomal regions. The middle and bottom rectangles show scaled enlargements of the 2.6-kb YAP and 18.3-kb SRY regions, respectively. The seven asterisks represent polymorphic nucleotide sites.

initially discovered by sequencing 2.6 kb in the YAP region from 16 Y chromosomes (Hammer 1995) and by sequencing 18.3 kb in the SRY region from 5 Y chromosomes (Whitfield, Sulston, and Goodfellow 1995). Table 2 indicates the three polymorphic nucleotide sites previously sequenced in the SRY region (individuals 1–5) and the four previously sequenced YAP sites (individuals 6–21). In the present study, we genotyped each of the four YAP sites in individuals 1–5 and each of the three SRY sites in individuals 6–21.

The four markers in the YAP region were genotyped according to Hammer et al. (1997). Markers in the SRY region were scored by PCR and site-specific oligonucleotide (SSO) hybridization as follows. Three segments encompassing the polymorphic sites at positions 4064, 9138, and 10831 in the SRY region (Whitfield, Sulston, and Goodfellow 1995) were PCR-amplified using the following primer pairs: SRY_{4064}-R (5′-CATGAGTTCAAATGATTCTT-3′) and SRY_{4064}-F (5′-GTATAATAGGCTGGGTGCTG-3′); SRY_{9138}-R (5′-

Table 2
Nucleotide States at 8 Polymorphic Sites on 21 Human Y Chromosomes

	SRY REGION			YAP REGION					
ORIGIN OF INDIVIDUAL	4064	9138	10831	YAP	PN1	PN2	PN3	*DYS257*	HTYPE[a]
Humans									
1. European (Italian)	G	C	A	−	C	C	G	A	1D
2. Bougainville (Nasioi)	G	T	G	−	C	C	G	G	1E
3. South American (Surui)	G	C	G	−	C	C	G	A	1C
4. Namibian (!Kung)	G	C	A	−	C	C	G	G	1A
5. Zaire (Mbuti Pygmy)	A	C	G	+	C	C	G	G	3A
6. South African (Xhosa)	A	C	G	+	C	C	G	G	3A
7. South African (Zulu)	A	C	G	+	T	T	G	G	5
8. South African (Sotho)	G	C	G	−	C	C	G	G	1B
9. Namibian (!Kung)	G	C	A	−	C	C	A	G	2
10. Zaire (Mbuti Pygmy)	A	C	G	+	T	T	G	G	5
11. Zaire (Mbuti Pygmy)	G	C	G	−	C	C	G	G	1B
12. African American	A	C	G	+	T	T	G	G	5
13. African American	A	C	G	+	T	T	G	G	5
14. Australian Ab. Person	G	C	G	−	C	C	G	G	1B
15. Australian Ab. Person	A	C	G	+	C	T	G	G	4
16. European (Ashkenazi)	G	C	G	−	C	C	G	G	1B
17. European	G	C	G	−	C	C	G	A	1C
18. Japanese	G	C	G	+	C	C	G	G	3G
19. Japanese	G	C	G	−	C	C	G	G	1B
20. Japanese	G	C	G	+	C	C	G	G	3G
21. Iraqi (Jewish)	G	C	G	−	C	C	G	G	1B
Outgroup[b]	G	C	A	−	C	C	G	G[c]	

[a] Haplotype (Htype) refers to the combination of states found on each human Y chromosome.

[b] We examined all seven homologous sites within the SRY and YAP regions in a sample of four common chimpanzees (for site 10831, a total of 25 common chimpanzees was included in the sample), one pygmy chimpanzee, three gorillas, and two orangutans.

[c] *DYS257* STS repeat sequences were obtained from outgroups by cloning and sequencing PCR products from a single male from each of the five hominoid species. Phylogenetic analysis revealed a G at position 108 of the eight great ape STS repeats most closely related to the human sequences.

GACAACCAAGAAGAGGAACC-3′) and SRY_{9138}-F (5′-TTTAACATTGACAGGACCAG-3′); and SRY_{10831}-R (5′-AAAATAGCAAAAAATGACACAAGGC-3′) and SRY_{10831}-F (5′-TCCTTAGCAACCATTAATCTGG-3′). Genomic DNAs were amplified in a 100-μl final volume containing 125 ng genomic DNA, 0.12 μM each primer, 0.2 mM each dNTP, 50 mM KCl, 10 mM Tris-HCl (pH 8.3), 0.5 U AmpliTaq DNA polymerase (Perkin-Elmer), and 3.5 mM, 1.5 mM, and 3.0 mM $MgCl_2$, respectively. The cycling conditions were 94°C for 2 min, and then 30 cycles of 94°C for 1 min, 54°C, 55°C, and 55°C for 1 min (for each segment, respectively), and 72°C for 1 min. The sequences of the site-specific oligonucleotide (SSO) hybridization probes were as follows: probe SRY_{4064}-G, 5′-AGGTCAAGGCGAGCGGATCAC-3′; probe SRY_{4064}-A, 5′-AGGTCAAGGCAAGCGGATCAC-3′; probe SRY_{9138}-C, 5′-ATGCTCTCGGCCTCCCC-3′; probe SRY_{9138}-T, 5′-ATGCTCTCGGTCTCCCC-3′; probe SRY_{10831}-A, 5′-TTCACACAGTATAACATTTTC-3′; and probe SRY_{10831}-G, 5′-TTCACACAGTGTAACATTTTC-3′. The probes were labeled with [γ-^{32}P]-ATP (Amersham) to a specific activity of at least 10^8 cpm/μg DNA. Approximately 200 ng (5 μl) of each amplified DNA was added to denaturation buffer (0.4 NaOH, 25 mM EDTA) and dotted on a nylon membrane (Amersham). The DNA was fixed to the membrane by UV irradiation with a Stratalinker™ UV crosslinker (Stratagene). Membranes were prehybridized in hybridization solution (5 × SSPE, 5 × Denhardt's solution, 0.5% SDS) for 30 min at 53°C. Labeled SSO probes were then added directly to the hybridization solution to a concentration of 2 pmol/ml, and hybridization was carried out overnight at 64°C, 56°C, and 50°C, for probes SRY_{4064}-G/A, SRY_{9138}-C/T, and SRY_{10831}-A/G, respectively. Membranes were rinsed in wash solution (2 × SSPE, 0.1% SDS) at room temperature, washed at the hybridization temperature for 30 min, and exposed to film for 2–48 h.

The *DYS257* sequence-tagged-site (STS) was PCR-amplified using the oligonucleotide primers and amplification conditions given by Vollrath et al. (1992). Although *DYS257* is contained within a single YAC clone mapping to interval 3C2 on the short arm of the human Y chromosome (Foote et al. 1992), we found evidence for at least three copies of the STS (i.e., STS repeats). To genotype the polymorphic G→A transition in one of the STS repeats, *DYS257* PCR products were digested with the restriction enzyme *Ban* I (Stratagene) using conditions specified by the manufacturer, and the resulting fragments were electrophoresed on 2% agarose gels. The G→A transition causes the loss of a *Ban* I restriction site in one of the STS repeats, resulting in either a five-fragment pattern (288, 182, 106, 63, and 43 bp) or a four-fragment pattern (182, 106, 63, and 43 bp). The first pattern is due to the presence of an A at position 108 in one STS repeat, while the second pattern is caused by the presence of a G at position 108 in all STS repeats. The two smallest fragments were difficult to visualize in 2% agarose gels. We also examined the M9 C→G transversion (Underhill et al. 1997) in a subset of 350 Y chromosomes in our sample. The M9 polymorphism was typed as described in Underhill et al. (1997).

PCR-Cloning and Sequencing of *DYS257* STS Repeats

The *DYS257* STS amplification products from males representing humans and four species of great apes (*Pan troglodytes*, *Pan paniscus*, *Gorilla gorilla*, and *Pongo pygmaeus*) were ligated into the pCR II-TOPO vector (Invitrogen) according to the manufacturer's protocol. Single clones were grown and plasmid DNA was isolated using standard procedures. DNA sequences were determined with universal primers to the plasmid vector using the Sequenase kit (U.S. Biochemical) according to manufacturer's specifications. The DNA sequences of several clones (>10) from each species were determined. Sequence alignments were carried out using the method of Feng and Doolittle (1987). A more detailed description of *DYS257* STS repeat sequences from humans and great apes, as well as the identification of the G→A transition at position 108 in one of the human STS repeats, will be presented elsewhere. The DYS257 sequence data can be found at the world wide web site http://www-shgc.stanford.edu/websearch.html by typing DYS257 into the query box and selecting SHGC-5459.

Phylogenetic Analyses

Haplotypes were inferred from typing all, or a subset, of the eight polymorphic sites in table 2. Parsimony analyses of human Y chromosome haplotypes were aided by the use of PAUP 3.0 for the Macintosh computer (Swofford 1990). Two rounds of branch-and-bound analyses were performed. The first round included the 10 haplotypes and 9 mutational characters in table 2. It should be noted that there are two mutational events associated with SRY_{10831} (see below). The second run included these nine sites as well as the M9 mutational character (Underhill et al. 1997). Parsimony analyses of aligned hominoid *DYS257* STS sequences were also carried out using the branch-and-bound option of PAUP 3.0.

Nested Cladistic Analysis of Geographical Distances

The nested cladistic analysis of geographical distances as described in Templeton, Routman, and Phillips (1995) is a method for analyzing the spatial distributions of the genetic variation in a phylogenetic framework. The first step in this analysis is to convert the estimated haplotype tree into a series of nested branches (clades) by using the nesting rules given in Templeton, Boerwinkle, and Sing (1987) and Templeton and Sing (1993). Briefly, these nesting rules start at the tips of the haplotype tree and move one mutational step into the interior, uniting all haplotypes that are connected by this procedure into a ''one-step clade.'' After pruning off the initial one-step clades from the tips, this procedure is repeated on the more interior portions of the haplotype tree as needed until all haplotypes have been placed into one-step clades. The next level of nesting uses the one-step clades as its units, rather than individual haplotypes. The nesting rules are the same; however, ''two-step

clades'' are now formed. This nesting procedure is repeated until a nesting level is reached such that the next higher nesting level would result in only a single category spanning the entire original haplotype network. The resulting nested clades are designated by ''C-N'' where ''C'' is the nesting level of the clade and ''N'' is the number of a particular clade at a given nesting level.

Once the haplotype tree has been converted into a nested statistical design, the geographical data are quantified by forming two distance statistics (Templeton, Routman, and Phillips 1995): (1) the clade distance, D_c, which measures the geographical range of a particular clade; and (2) the nested clade distance, D_n, which measures how a particular clade is geographically distributed relative to its closest evolutionary sister clades (i.e., clades in the same next-higher-level nesting category). Distance contrasts between older and younger clades are important in discriminating the potential causes of geographical structuring of the genetic variation (Templeton, Routman, and Phillips 1995). In our analysis, temporal polarity is determined by outgroup comparisons. The statistical significances of the different distance measures and the old–young (interior-tip) contrasts are determined by random permutation testing. This procedure simulates the null hypothesis of a random geographical distribution for all clades within a nesting category given the marginal clade frequencies and sample sizes per locality. There are two major reasons for failing to reject this null hypothesis: (1) the samples are inadequate to detect geographical structuring even though it exists, and (2) the population is panmictic over the sampled area such that any haplotype frequency differences are due only to sampling or drift effects that do not result in a geographic pattern. As there is no way of discriminating between these alternatives with the existing data, biological inference is confined to those cases in which the null hypothesis is rejected.

If the null hypothesis is rejected, the analysis continues by seeking the biological causes of these statistically significant (i.e., nonrandom) haplotype–geography associations in terms of population structure and/or population history considerations. Templeton, Routman, and Phillips (1995) consider three major biological factors that can cause a significant spatial/phylogenetic association of haplotype variation: (1) recurrent genetic drift coupled with restricted gene flow, particularly gene flow restricted by isolation by distance; (2) past fragmentation events; and (3) population range expansion. The detailed expected impacts of these types of evolutionary forces and events on the nested patterns of geographical distances are given in Templeton, Routman, and Phillips (1995), and an empirical validation is given in Templeton (1998*a*). In order to make this pattern evaluation explicit and consistent, a detailed inference key is provided as an appendix to Templeton, Routman, and Phillips (1995). The nested clade analysis does not treat these patterns as mutually exclusive, but instead searches for multiple overlaying patterns within the same data set. The aforementioned key also incorporates the types of pattern artifacts that can arise from inadequate sampling. As a consequence, even though statistical significance may have been detected, the inference key can sometimes result in no definitive biological inference. In this manner, the key identifies the deficiencies in the current sample that must be corrected before strong biological inference can result.

Coalescence Analyses

The coalescent process (Kingman 1982) was used to model the ancestry of the sample sequences. Our model assumes both random mating and a constant effective male population size (N_m) going back in time, and ignores the effect of population subdivision. The timescale is in units of N_m generations, and while in the short run the assumptions may not strictly hold, the coalescent model should be appropriate over a long time period. In the coalescent timescale, mutations occur according to a Poisson process with rate $\theta/2$ along the ancestral lineages, where

$$\theta = 2N_m\mu \tag{1}$$

and μ is the total mutation rate per sequence per generation. The mean and standard deviation of the time to the most recent common ancestor (TMRCA) and the ages of each of the mutations can be found conditional on the full information of the pattern of mutations in the sequences, or equivalently, from the gene tree deduced from the data assuming an infinitely-many-sites model of mutation (Griffiths and Tavaré 1994). Thus, the approach taken here (Griffiths and Tavaré 1994; Harding et al. 1997) simulates a coalescent process including time information conditional on a specified haplotype tree with a given value of θ, the only independent parameter in the model. Specifically, an advanced simulation technique was used such that each simulation run generated an estimated likelihood, TMRCA, and mutational ages. If there are r simulation runs generating (TMRCA, likelihood) pairs $(t_1, l_1), \ldots, (t_r, l_r)$, then an empirical distribution of the TMRCA has (TMRCA, probability) pairs (t_i, p_i), where

$$p_i = l_i \Big/ \sum_{j=1}^{r} l_j, \qquad i = 1, \ldots, r. \tag{2}$$

The estimated means and standard deviations come from this empirical distribution. Ages of mutations were treated similarly. Our simulation results are based on 1,000,000 replicate runs. Although computationally intensive, this approach preserves more information than do pairwise DNA sequence analyses (Harding et al. 1997; Tavaré et al. 1997).

The empirical distributions of the TMRCA and mutational ages depend only on θ and not on rate variation along the sequences. Estimation of θ, TMRCA, and ages from the data is incomplete, because sequences were only examined at sites known to be segregating from earlier samples (Hammer 1995; Whitfield, Sulston, and Goodfellow 1995). In the individual samples where all sites were examined, estimates of $\hat{\theta} \pm$ SD (Watterson 1975) are $\hat{\theta}_{YAP} = 1.21 \pm 0.57$ based on 16 sequences with four segregating sites and $\hat{\theta}_{SRY} = 1.44 \pm 1.17$ based on five sequences with three segregating sites.

Therefore, an estimate for the combined region based on seven segregating sites is $\hat{\theta}_s = 2.65 \pm 1.39$. Theta estimates using the full gene tree structure for these samples are $\hat{\theta}_{YAP} = 1.31$ and $\hat{\theta}_{SRY} = 1.65$ with a combined $\hat{\theta}_S$ estimate of 2.96 (Griffiths 1989; Griffiths and Tavaré 1994). Our preference is to concentrate on $\theta = 2.5$ (the approximate small sample estimate) for interpretive purposes, but other values of θ are also plausible. The TMRCA and mutational ages for our data set were estimated for $\theta = 1.0$, 2.0, 2.5, 3.0, and 4.0 using the program TIMESIM (Griffiths and Tavaré 1994; Harding et al. 1997). The YAP insertion was modeled as a mutation. Mutations 10831.1 and 10831.2 at the same site are incompatible with a mutation occurring only once at a site and were treated as separate mutations. Standard deviations for the TMRCA and mutational ages are large, with values typically one half of their respective means. The most realistic interpretation of the computed TMRCA and mutational ages is that they are estimated lower bounds for the true values, since for a fixed θ, additional unobserved segregating sites would have the effect of increasing these values. Also, incorporating population subdivision into the model may increase mutational age and TMRCA estimates.

Results

Haplotype Designations

The character states at all eight polymorphic sites in the 21 Y chromosomes presented in table 2 produced 10 distinct SRY/YAP/*DYS257* combination haplotypes. Haplotypes previously designated as haplotype 1 based on polymorphism only in the YAP region (Hammer 1995) were subdivided into five SRY/YAP/*DYS257* haplotypes. Haplotype 1A, present in a single individual (!Kung, number 4), contains the ancestral state at each of the seven polymorphic sites in the YAP and SRY regions as determined by comparison with the homologous sites on Y chromosomes of great apes. Y chromosomes differing from haplotype 1A only by a substitution at SRY_{10831} (i.e., presence of a G) were named haplotype 1B. Haplotype 1B was found in six of the 21 Y chromosomes and exhibited the widest geographical distribution (e.g., present in individuals from Africa, Australia, Europe, Japan, and West Asia). Haplotype 1C, differing from 1B by a substitution (A) at *DYS257*, was found in two individuals (an Amerind and a European). Haplotype 1D, also found in a European, is unusual in that it has the *DYS257*-A allele (like haplotype 1C) but does not have the SRY_{10831}-G substitution. Haplotype 1E, found in a single Melanesian, is derived from 1B by a C→T transition at SRY_{9138}.

Y chromosomes containing the YAP element (YAP^+) were previously designated as YAP haplotypes 3–5 (Hammer 1995; Hammer et al. 1997). The G→A transition at SRY_{4064} resulted in the subdivision of YAP haplotype 3 into two haplotypes (Altheide and Hammer 1997): haplotype 3G differs from 1B only by the presence of the YAP element, while haplotype 3A also contains the A at SRY_{4064}. Haplotypes 4 and 5 are further differentiated from haplotype 3A by C→T transitions at PN2 and PN1.

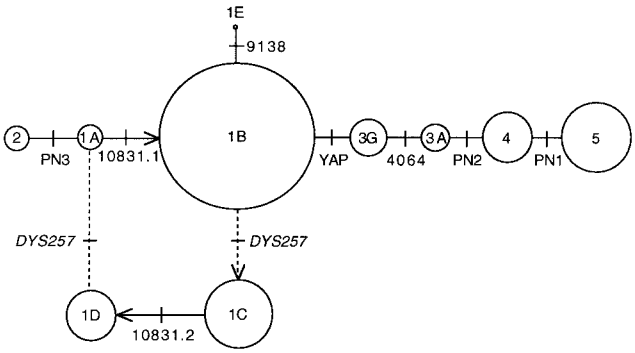

FIG. 3.—Y chromosome haplotype tree based on nine polymorphic markers. Each crossbar represents a single point mutational event. Circle areas reflect relative, rather than exact, global haplotype frequency proportionalities. Dotted lines represent alternative tree topologies resulting from homoplasy in the data set (see text for explanation). The arrows on the lineages connecting haplotypes 1A, 1B, 1C, and 1D represent the resolved tree topology after a parsimony analysis including a 10th mutational character, M9.

Y Chromosome Haplotype Tree

Initially, we performed a parsimony analysis of the 10 haplotypes using the 8 polymorphic sites listed in table 2. This exploratory phylogenetic analysis resulted in four equally likely most-parsimonious networks (unrooted trees). Two of these involved homoplasy at *DYS257*, and the other two involved homoplasy at SRY_{10831}. A haplotype tree for these 10 haplotypes is given in figure 3. Ambiguity resulting from the aforementioned homoplasy is designated by reticulation in the tree. In order to root the haplotype network, we used outgroup comparisons for the sites in the SRY, YAP, and *DYS257* regions. Our most extensive outgroup data were for the 10831 site in the SRY region (table 2), and this site alone identifies three possible candidates for the root: haplotypes 1A, 1D, and 2. The outgroup state of site PN3 eliminates haplotype 2 (fig. 3), leaving only haplotypes 1A and 1D as candidates. The outgroup state of *DYS257* was needed to choose conclusively between 1A and 1D as the root. By comparing homologous *DYS257* STS sequences from four species of great apes, we found that the *DYS257*-G allele is the most likely ancestral state for the human polymorphic STS. Because 1D is associated with the derived *DYS257*-A allele, this haplotype is eliminated as a candidate for the root of the Y haplotype tree.

We confirmed haplotype 1A as the root of the tree by genotyping the recently published M9 C→G transversion polymorphism (Underhill et al. 1997) in a subset of the Y chromosomes in our sample (data not shown). We found complete associations both between the M9-C allele (ancestral state) and haplotypes 1A, 2, 3G, 3A, 4, and 5 and between the M9-G allele (derived state) and haplotypes 1C, 1D, and 1E. Haplotype 1B was found to be associated with both the M9-C and M9-G alleles. A parsimony analysis incorporating the M9 mutational character resulted in a single 10-step tree with

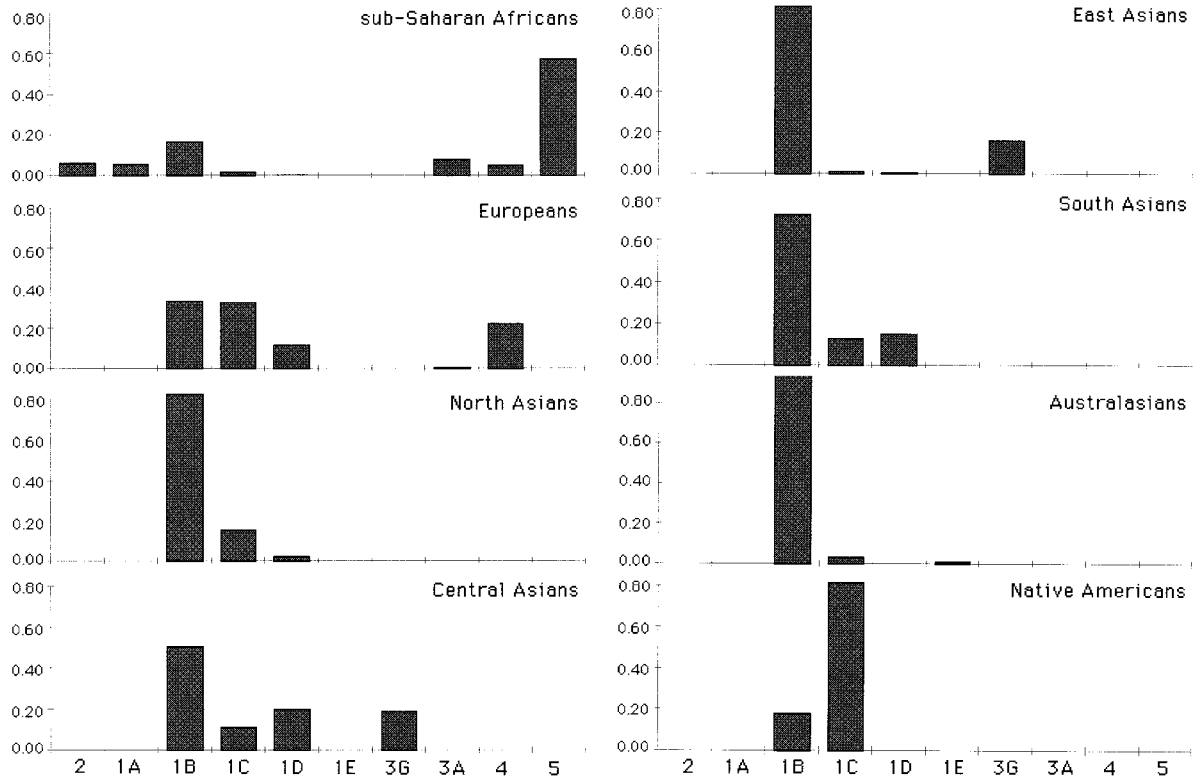

FIG. 4.—Histogram representations of 10 Y chromosome haplotype frequencies for 8 geographic regions. The haplotype frequency for 1B in Australasians (96%) is not drawn to scale.

the following directionality: haplotype 1A → [1B/M9-C → 1B/M9-G] → 1C → 1D.

In sum, the evidence points to the following: (1) an ancestral G allele at the *DYS257* STS, (2) a single origin for the *DYS257*-A allele in human evolution, and (3) the occurrence of both an A→G transition and a G→A reversion at the SRY_{10831} site. This information conclusively demonstrates that the root of the tree is haplotype 1A and resolves the reticulation in the haplotype network.

Geographic Distribution of Y Chromosome Haplotypes

Haplotype 1B was the most frequent haplotype in the global sample of 1,544 individuals (55.2%). Haplotype 5 was the second most frequent haplotype (14.1%), followed by haplotypes 1C (12.2%), 1D (4.9%), 3G (4.4%), 4 (4.3%), 3A (2.0%), 2 (1.5%), and 1A (1.4%). Haplotype 1E was identified in only a single individual from Melanesia. Figure 4 displays the frequencies of each of these 10 haplotypes in 8 geographical regions.

Eight of the 10 haplotypes were present in sub-Saharan African populations. Of these eight, three (haplotypes 1A, 2, and 5) were limited to African males in this survey. Ancestral haplotype 1A, present in only 5.5% of the African sample, was relatively rare. On the other hand, haplotype 5 was the most common haplotype (57.1%) in Africa as well as the most derived (fig. 3). Haplotypes 1C (1.6%) and 1D (0.3%) were rare in Africa, although they were relatively more common in Europe and Asia.

Four haplotypes occurred in Asian populations. Of these, haplotype 1B was the most prevalent (76.1%), while the other three (1C, 1D, and 3G) were found at much lower frequencies, ranging between 6% and 9%. Haplotype 3G (8.6%), the most ancestral YAP^+ haplotype (Altheide and Hammer 1997), was restricted to Asia and occurred only in central (18.8%) and east (16.6%) Asian populations.

European populations were characterized by intermediate frequencies of four haplotypes: 1B (33.2%), 1C (32.7%), 4 (22.1%), and 1D (11.5%). Haplotype 3A was found only in a single Sicilian individual. The Y chromosomes of Australasian males were predominantly haplotype 1B (95.7%), while most of the remaining Y chromosomes were 1C (3.4%). Native American Y chromosomes were predominantly haplotype 1C (81.8%). Those of the remaining Native American males were haplotype 1B (18.2%).

Nested Cladistic Analysis of Y Chromosome Haplotypes

Figure 5 depicts the nested cladistic design for the 10 Y chromosome haplotypes. Table 3 presents the results of a nested contingency analysis for these haplotypes in the 35 populations in our survey. This analysis indicates that highly statistically significant associations exist between clades and geographical locations for the entire cladogram ($P < 0.01$). Three one-step clades and two two-step clades also exhibited similar statistically significant associations ($P < 0.01$), while clade 1-1 showed an association at the 5% level. The null hy-

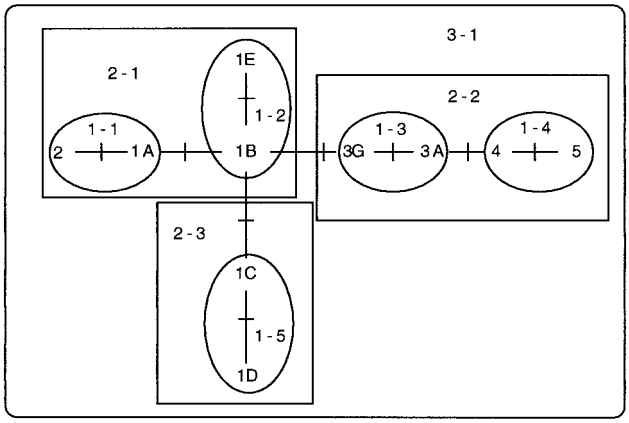

FIG. 5.—Y chromosome nested cladistic design. Ovals contain one-step clades which are designated 1-1 through 1-5. Rectangles contain two-step clades which are designated 2-1 through 2-3. A single three-step clade (3-1) encompasses the entire cladogram.

pothesis of no geographical association is rejected in every nested clade except 1-2 and 2-3.

Figure 6 presents the results of the nested cladistic analysis, and table 4 summarizes the main inferences drawn from our haplotype data. Both population structure (restricted gene flow with isolation by distance) and population history (range expansions) must be considered in the explanation of the global distribution of human Y chromosomes. Four episodes of restricted gene flow with isolation by distance were indicated: (1) within Africa, associated with the 1-1 clade (haplotypes 1A and 2); (2) between and within Africa and Europe, associated with the 1-4 clade (haplotypes 4 and 5); (3) global, associated with the 1-5 clade (haplotypes 1C and 1D); and (4) global, associated with the 2-2 clade (haplotypes 3G, 3A, 4, and 5). Three range expansions were also indicated: (1) out of Africa, associated with the 2-1 clade (haplotypes 2, 1A, 1B, and 1E); (2) Asia to Africa, associated with the 1-3 clade (haplotypes 3G and 3A); and (3) contiguous range expansion on a global scale, associated with the entire cladogram (all haplotypes).

Coalescence Analysis

Figure 7 presents the estimated age of the most recent common ancestral Y chromosome (TMRCA) as well as age estimates for the nine mutational events depicted in figure 3. An effective population size of 5,000 males and a 20-year generation length were assumed. For the $\theta = 2.5$ analysis, the estimated mean for the TMRCA equaled 147,000 years, with a standard deviation of 51,000 years. The 95% confidence interval for this TMRCA estimate is 68,000–258,000 years. The estimated age of the most ancient mutation in the Y cladogram (at the 10831 site) was 110,000 ± 36,000 years. This mutational event defines the origin of haplotype 1B from its precursor, 1A. The second (G→A reversion) mutational event at this site, representing the evolution of haplotype 1D from 1C, did not occur until more than 100,000 years later. The estimated age of the YAP insertional event marking the origin of the YAP^+ clade was 55,000 ± 19,000 years. Subsequent mutations on this lineage occurred between approximately 11,000 and 31,000 years ago.

See figure 8 for TMRCAs and mutational ages for five different values of θ. Note that the TMRCAs and ages of the two oldest mutations in the cladogram are most affected by varying the value of θ. If the actual value of θ lies between 2.0 and 3.0, then the most likely estimate for the mean TMRCA would be between 118,000 and 155,000 years.

Table 3
Nested Contingency Analysis of Geographic Associations

Clade[a]	Permutational Chi-Square Statistic	Probability[b]
1-1	7.70	0.042
1-3	99.00	0.000
1-4	196.62	0.000
1-5	94.02	0.000
2-1	587.00	0.000
2-2	243.36	0.000
Entire cladogram	1,470.14	0.000

NOTE.—Associations are for the human Y chromosome haplotype data.

[a] The clade column includes the nesting clades. Clades without genetic and/or geographic variation are not listed.

[b] The permutational chi-square probability was calculated according to Templeton, Routman, and Phillips (1995).

Discussion and Conclusions

Most previous attempts to reconstruct the evolutionary history of *H. sapiens* using genetic data have relied on hypothesis compatibility *sensu* Stoneking (1994). Data consistency with predictions from models, hypotheses, and theories and concordance of results from disparate data sets rather than rigorous inferential statistical tests *sensu* Templeton (1994) characterize the aims and methodological procedures of this general approach (Relethford 1995; Cann 1997). Even the graphical displays of haplotype geographic distributions advocated by proponents of intraspecific phylogeography (Avise 1994) and used extensively for human mtDNA data (Melton et al. 1995) and, more recently, for human Y chromosome data (Hammer et al. 1997; Zerjal et al. 1997) result in statements about hypothesis compatibility and in the generation of new hypotheses, but not in actual hypothesis testing.

Recently, Hammer et al. (1997) conducted a survey of five YAP haplotypes in 60 worldwide populations. Their study resulted in numerous empirical statements and led to several hypotheses concerning the origin and subsequent spread of these five haplotypes in global populations. The wide geographic distribution of the most ancestral haplotype (1), one of two YAP^- haplotypes, was consistent with an origin anywhere in the world. However, higher levels of diversity associated with this haplotype in Africa led to the hypothesis of an African origin. Haplotype 2, the other YAP^- haplotype, was limited to Khoisan populations in southern Africa. Three of the five haplotypes contained the YAP element (YAP^+) and were referred to as YAP haplotypes 3–5. Hammer et al. (1997) hypothesized that haplotype 4 originated in North Africa and spread first to the Levant

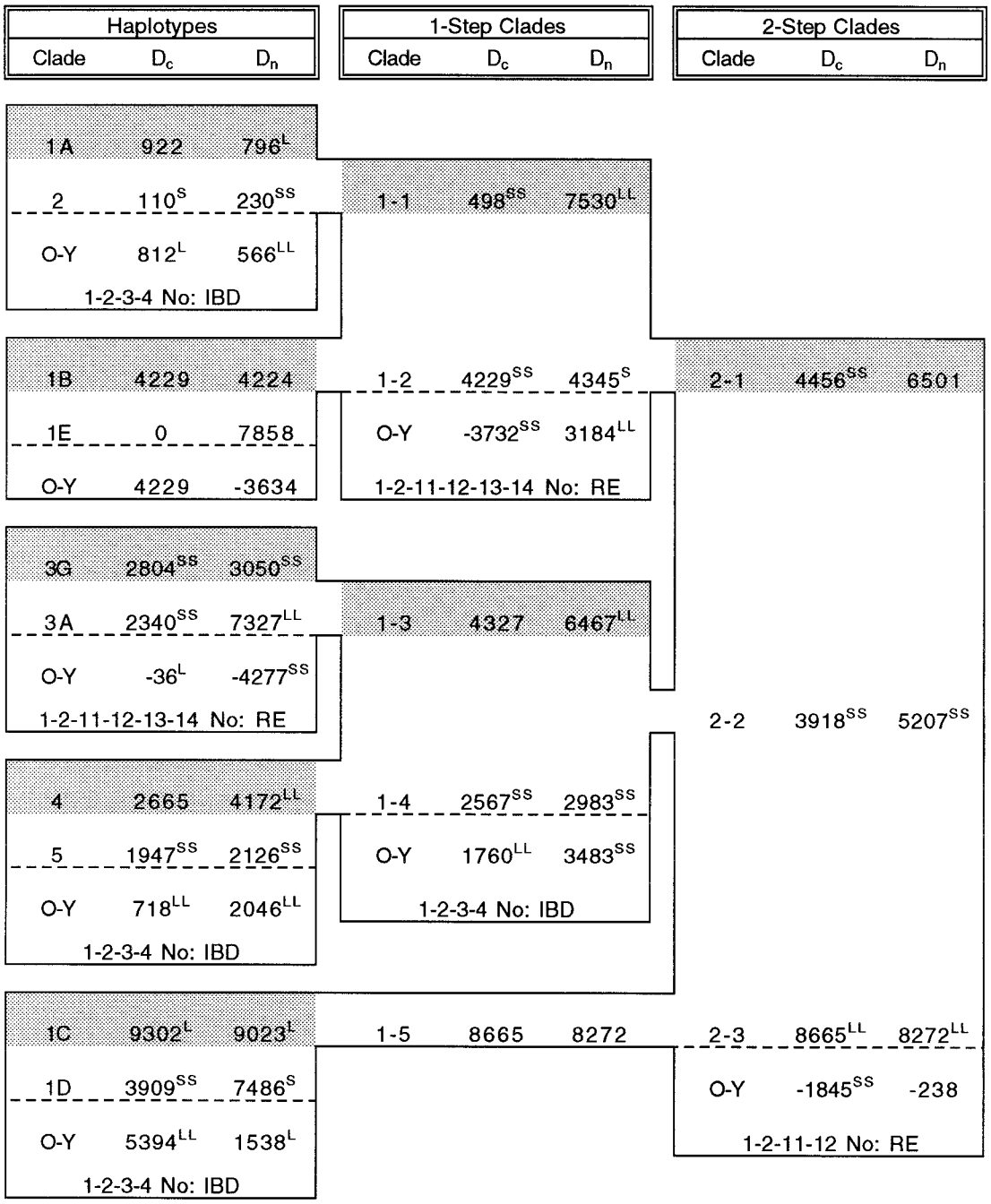

FIG. 6.—Results of the nested geographic analysis of the human Y chromosome haplotypes. The nested design is given in figure 5, as are the haplotype and clade designations. Following the name or number of any given clade are the clade (D_c) and nested clade (D_n) great circle distances. The oldest clade within the nested group is indicated by shading. The average difference between the oldest versus younger clades within a nesting category (with haplotype 1A as the root) for both distance measures is given in the row labeled "O-Y" below a dashed line. A single superscript "S" or "L" designates a significantly small or large distance at the 5% level, respectively, and double letters indicate significance at the 1% level. Below the O-Y row is the inference key chain, wherein the numbers refer to the sequence of questions in the interpretative key given in the appendix in Templeton, Routman, and Phillips (1995). Following the sequence of numbers is the answer to the final question in this interpretive key (yes/no) along with the final biological inference (IBD = isolation by distance; RE = range expansion). For any clade level, the inference results are given at the bottom of the preceding column in the rectangle connected to the next higher nesting level. For the entire cladogram (3-1), the inference result is given at the bottom of the last column, labeled "2-Step Clades."

before continuing its migration northwest across Europe. Haplotype 5, the most derived YAP haplotype, supposedly had an African origin and recently migrated through the horn of Africa to west Asia. In contrast, the haplotype representing the most ancestral YAP$^+$ haplotype (3) was thought to have had an Asian origin. Because haplotypes 4 and 5 evolved from haplotype 3 and accounted for the majority of African Y chromosomes, the implication of this hypothesis was that a large portion of African paternal diversity had its roots in Asia.

Table 4
Main Inferences from Results of Nested Cladistic Analysis

Clade	Inference
Haplotypes nested in 1-1	Restricted gene flow with isolation by distance within Africa
Haplotypes nested in 1-3	Range expansion (Asia into Africa)
Haplotypes nested in 1-4	Restricted gene flow with isolation by distance (Africa and Europe)
Haplotypes nested in 1-5	Global restricted gene flow with isolation by distance
One-step clades nested in 2–1	Range expansion out of Africa
One-step clades nested in 2-2	Restricted gene flow with isolation by distance (Old World)
Two-step clades nested in entire cladogram (3-1)	Contiguous range expansion; however, source is ambiguous because of widespread distribution of oldest clade (2-1)

NOTE.—Only clades resulting in the rejection of the null hypothesis are included above.

Although Hammer et al. (1997) compared the geographic distribution of Y chromosome haplotypes with a phylogenetic tree to generate these multiple migration hypotheses, they were unable to test them. Nor were they able to gain insight into the evolutionary processes responsible for the Old World distribution of Y chromosomes. On the other hand, the present study uses population genetics models to: (1) test for statistically significant geographic structuring of Y haplotypes, (2) identify the possible causes of significant associations between haplotypes and geography, and (3) provide estimated dates for a number of human microevolutionary events.

The major implications of our present analysis for the structure and evolutionary history of the Y chromosome haplotype tree (figs. 3 and 7) are the following: (1) haplotype 1 of Hammer et al. (1997) is now subdivided into five (1A–1E) haplotypes; (2) 1A is the ancestral haplotype and has an African origin; (3) 1B originated ~110,000 years ago and is the most frequent and geographically widespread Y chromosome haplotype in the world; (4) the dispersal of haplotype 1B from Africa throughout the Old World marked the complete replacement of Eurasian Y chromosomes by African Y chromosomes; (5) 1E is the most recently evolved haplotype in our tree; (6) both an A→G transition and a G→A reversion occurred at nucleotide site SRY_{10831}, with more than 100,000 years separating these mutational events; (7) the G→A transition responsible for the polymorphism in the *DYS257* STS occurred on the lineage leading to haplotype 1C before the second mutation at SRY_{10831}; (8) *DYS257* identifies a major Eurasian clade containing haplotypes 1C and 1D; (9) haplotype 1C is both older and more geographically widespread than

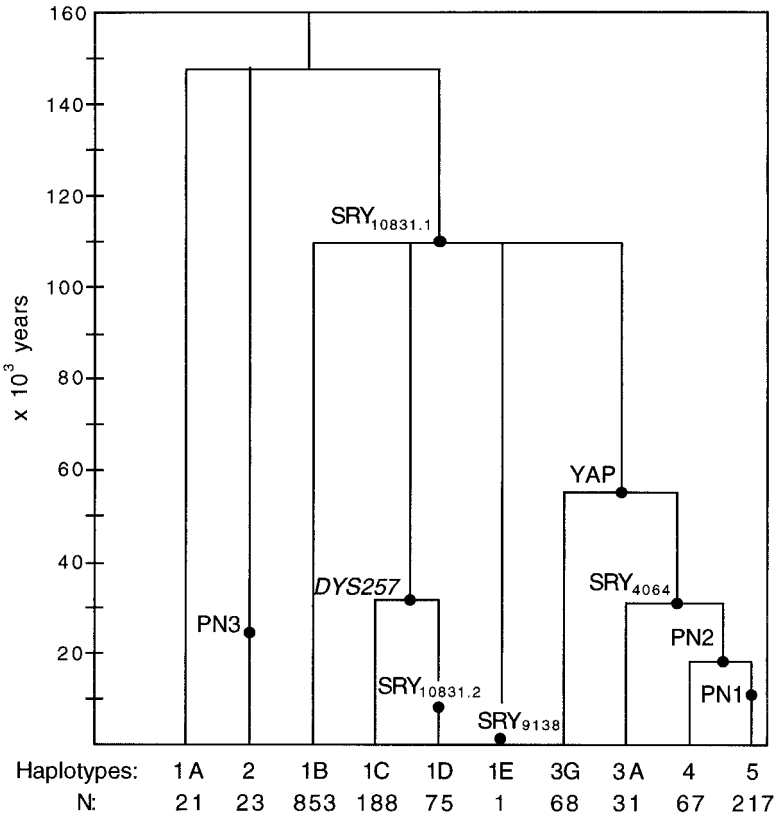

FIG. 7.—Scaled coalescent tree for Y chromosome haplotypes showing ages of mutations estimated from the world set of 1,544 samples when $\theta = 2.5$ (timescale in units of 10^3 years).

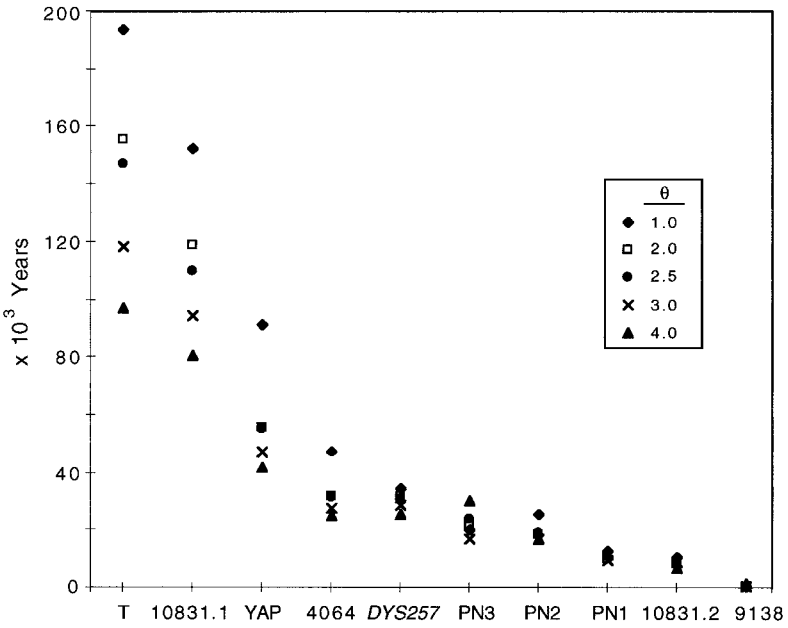

FIG. 8.—Mean TMRCA (T) and mutational ages for each of nine mutations when θ varies from 1.0 to 4.0. The inset shows symbols used for θ values of 1.0, 2.0, 2.5, 3.0, and 4.0.

1D; (10) haplotype 3G had an Asian origin; and (11) the mean TMRCA probably lies between 118,000 and 155,000 years ago.

The African root for our Y chromosome tree is consistent with data from both autosomes (Mountain and Cavalli-Sforza 1994; Nei and Takazaki 1996; Harding et al. 1997) and mtDNA (Horai et al. 1995; Penny et al. 1995; Krings et al. 1997). Although restricted to Africa, the ancestral haplotype (1A) does not exhibit uniform frequencies throughout sub-Saharan Africa (table 1). This haplotype is absent in three of the seven sub-Saharan African populations, is rare in three others, and reaches a maximum of 20% in the Khoisan. The association of the ancestral Y haplotype with the Khoisan is reminiscent of the clustering of mtDNA lineages from members of this ethnolinguistic group near the proposed root of the mtDNA tree (Vigilant et al. 1991; Penny et al. 1995). The range for our estimates for the mean TMRCA is also consistent with other estimates based on Y chromosome DNA sequences (Weiss and von Haeseler 1996; Fu and Li 1997; Tavaré et al. 1997; Underhill et al. 1997) and mtDNA data (Stoneking 1993; Horai et al. 1995), as well as with an approximately fourfold greater TMRCA for the β-globin locus (Harding et al. 1997). However, consult Brookfield (1997) for a cautionary note regarding the validity of the many assumptions involved in the calculation of the TMRCA.

Our Y chromosome data provided statistically significant associations among haplotypes and geography for the entire cladogram as well as for six nested clades. Contained within the four episodes of gene flow and three range expansions (table 4) are some potentially significant events in the history of paternal human evolution. A scenario consistent with the results of our nested cladistic analysis would involve the following historical events and evolutionary processes: (1) an initial range expansion of Y chromosomes out of Africa, (2) restricted gene flow with isolation by distance within Africa, (3) a subsequent dispersal throughout the Old World via gene flow with isolation by distance, (4) a range expansion back from Asia to Africa (implying bidirectional movement between Africa and Asia), and (5) subsequent contiguous range expansion(s) and episodes of short-range gene flow involving the entire globe.

The initial range expansion out of Africa associated with haplotype 1B seems to have involved a complete replacement of Eurasian Y chromosomes by African Y chromosomes. This scenario is consistent with genetic predictions based on a family of African replacement models derived from both fossil and genetic data (Minugh-Purvis 1995; Hammer and Zegura 1996). We also have firm evidence of an actual range expansion of Y chromosomes from Asia back to Africa contained within nested clade 1-3 (i.e., haplotypes 3G and 3A). Our statistically significant result corroborates the population movement first hypothesized in Hammer et al. (1997) and subsequently supported by Altheide and Hammer (1997). This back-to-Africa migration did not involve a complete replacement of African Y chromosomes by Eurasian Y chromosomes, because the haplotypes ancestral to haplotype 3G are still present in Africa. Our coalescence results provide a possible time frame for this noteworthy range expansion. For example, the YAP insertion probably occurred on an Asian Y chromosome as long ago as ~55,000 years (fig. 7). The SRY_{4064} G→A transition marks the Y chromosome lineage that is postulated to have migrated to Africa and that eventually gave rise to the major portion of contemporary African Y chromosomes (haplotypes 3–5). This mutational event took place ~31,000 years ago, perhaps 10,000 years before the recalibrated dates for the last

glacial maximum of 21,000–22,000 BP (Lowell et al. 1995).

Harding et al. (1997), in a paper on global β-globin sequence variation, also presented evidence that some of the genetic diversity in our species has Asian roots. Their geographical pattern of ancestral-descendant lineages is similar to ours, and in a recently completed nested cladistic analysis of their data, Templeton (1998*b*) detected a statistically significant range expansion of Asian β-globin sequences from southeast Asia into Africa. Although the Asian-specific β-globin lineages reach over 200,000 years into the past, the African sublineages descended from these Asian-specific lineages span a much shorter time period (i.e., originating sometime between 25,000 and 75,000 years ago). Despite the fact that the time frames derived from nuclear and Y chromosomal data have only a minimum overlap, it is still plausible that they chronicle the same migratory event. Nonetheless, what these two data sets do underscore is the real possibility of multidirectional population movements between Africa and Eurasia throughout the evolutionary history of our species.

Turning to more recent human microevolutionary events, the clinal distribution of haplotype 4 noted in Hammer et al. (1997) may indeed reflect Ammerman and Cavalli-Sforza's (1984) demic diffusion process whereby between ~10,000 and 5,000 years ago, early farmers slowly expanded from the Middle East to and through Europe at a rate of about 1 km/year (Cavalli-Sforza, Menozzi, and Piazza 1994). Possible confirmatory evidence for this gradual movement of genes (i.e., farmers) comes from the results of our nested cladistic analysis. Specifically, we inferred a process of restricted, recurrent gene flow with isolation by distance associated with haplotypes 4 and 5 nested within clade 1-4 (rather than long-range gene flow or a range expansion event). The timing and geographical positioning of haplotype 4, originating ~20,000 years ago (fig. 7) in North Africa (unpublished data), are concordant with a scenario in which this haplotype remained in North Africa and the Middle East for approximately 10,000 years and then gradually spread toward Europe starting around 10,000 years ago. This date for the origin of haplotype 4 makes one of the other possible explanations for the distribution of this haplotype in Hammer et al. (1997) (i.e., the Levantine expansion of anatomically modern humans ~40,000 years ago) less likely. On the other hand, more recent gene flow may also be implicated by the distribution of haplotype 4 because of the numerous population movements between the Levant and Europe postulated to have occurred over the last 10,000 years (Richards et al. 1996, 1997; Cavalli-Sforza and Minch 1997).

Interestingly, the only statistically significant range expansion detected in Templeton's (1997) nested cladistic analysis of the Old World human mtDNA data subset from Excoffier and Langaney (1989) was limited to a recent expansion within Europe. Thus, Europe has witnessed two relatively recent microevolutionary phenomena detected by the nested cladistic method: (1) restricted gene flow with isolation by distance for males and (2) a range expansion for females. It is not known whether these phenomena were temporally concurrent or if differences in male and female demographic and life history variables contributed to these contrasting inferences (Cavalli-Sforza and Minch 1997).

Both Templeton's (1993) original nested cladistic analysis of human mtDNA and his methodologically more rigorous reanalysis (Templeton 1997) are highlighted by recurrent gene flow restricted by isolation by distance throughout the Old World for the entire time period encompassing the mtDNA TMRCA. This short-range gene flow is pervasive at all levels of analysis and underscores the paramount influence of population structure on the dynamics of human maternal genome evolution. No intercontinental range expansions similar to the three postulated on the basis of our Y chromosome data are detectable in global mtDNA data. Thus, the effects of population history seem to have left a much clearer intercontinental imprint on our paternal-specific genome than the regional signals left in our mtDNA. One possible explanation for this pattern is that males disperse more than females during long-range intercontinental population movements, while females may disperse more than males during short-range intracontinental migrations.

If males and females do, indeed, exhibit major differences in their ancient population structure and demographic histories, then we might expect traces of these differences to be preserved in the autosomal DNA record. Templeton's (1998*b*) reanalysis of Harding et al.'s (1997) β-globin data represents the only nested cladistic analysis of a human autosomal data set. The deepest clade in the β-globin cladogram showed an out-of-Africa expansion; however, the 800,000-year coalescence time for the β-globin gene tree makes it unlikely that this range expansion had anything to do with the out-of-Africa event detected by our Y chromosome data. On the other hand, this time frame is more concordant with the sudden appearance of the possibly African-derived *Homo antecessor* in Spain sometime before 780,000 years ago (Bermúdez de Castro et al. 1997). Moving to less deep structures, all the midlevel (two-step) clades gave strong evidence for gene flow restricted via isolation by distance occurring more than 200,000 years ago throughout the Old World. Finally, two range expansions were detected at the one-step clade level: (1) the aforementioned expansion from southeast Asia back to Africa, and (2) an out-of-Africa expansion that involved the oldest haplotypes by outgroup rooting, making the temporal framework of this expansion unclear (i.e., it may be a recent expansion or the same one detected at higher levels in the cladogram).

This out-of-Africa expansion was not a replacement event, because it was nested within a two-step clade characterized by gene flow restricted via isolation by distance. In order to equate this out-of-Africa event with the one detected in our Y chromosome data, one would have to argue that perhaps Eurasian males were replaced but females were not. This is consistent with the demographic picture from the nested cladistic analyses of mtDNA data (Templeton 1993, 1997), in which

females show no sign of replacement and gene flow rather than range expansion is the oldest inference. Therefore, the β-globin locus integrates aspects of both the mtDNA and Y chromosome analyses and provides support for the hypotheses of contrasting male and female population structure and demographic histories. Because there is evidence for restricted, recurrent gene flow throughout the Old World during the entire history of anatomically modern humans, as well as for range expansions out of Africa >100,000 years ago, the nested cladistic analysis results from these three types of data conform with genetic predictions based on human origin(s) models characterized by interbreeding between migrating and resident populations. Thus, the combined data add a new sex-specific component to the conceptual framework of both Brauer's (1989) African Hybridization and Replacement model and Smith, Falsetti, and Donnelly's (1989) Assimilation model: the possibility that the Old World female genetic complement was preserved by hybridization, whereas the Eurasian male component was replaced by African Y chromosomes.

Future Directions

One of the clear advantages of a nested cladistic approach using haplotype trees is that spatial and temporal patterns of genetic variation can be examined concurrently. Notice, however, that natural selection is not directly addressed in our present application of the nested cladistic design. Therefore, studies of variation at other nonrecombining segments of the genome are needed to confirm our inferences as well as to clarify the biocultural and demographic contexts responsible for the differing evolutionary trajectories exhibited by maternal- and paternal-specific data. In the short term we need to examine as many single-gene trees covering as many parts of the genome as is logistically feasible. Concordant gene trees may then lead to the construction of robust population trees.

In detecting recurrent and historical events through nested cladistic analysis, recall that the timescale is determined by the coalescence time of the DNA region under examination. It is important to keep this consideration in mind when comparing data from independently segregating segments of the genome. Moreover, a nested cladistic analysis can detect geographical associations within a clade only when mutations have occurred to create variation within the clade. Therefore, the rate of mutation also limits the level of resolution possible with a nested cladistic analysis. For example, even though mtDNA and Y chromosomal DNA appear to have similar coalescence times, the mtDNA data analyzed by Templeton (1997, 1998*a*) have many more mutations and much finer genetic resolution than do the Y chromosome data analyzed here. In particular, the finer genetic resolution of mtDNA makes it easier to identify episodes of genetic exchange as ''recurrent.''

In addition, it is imperative to look for concordance among genetic, linguistic, ethnohistoric, fossil, and/or archeological data sets to lend support to particular hypotheses about human population history, because genetic data alone cannot inform us completely about that history. Ultimately, we need to develop model-based tests applicable to different kinds of data that have the power to discriminate among alternative hypotheses and, specifically, to distinguish among the various competing models of human population origins. Promising initial methodological developments involving this kind of unification wherein different types of data sets can be analyzed by the same model(s) include (1) the expansion of *F* statistics to encompass quantitative morphological data sets (Williams-Blangero and Blangero 1989; Konigsberg 1990; Konigsberg and Blangero 1993; Relethford and Harpending 1994) and (2) the joint analysis of linguistic, ethnohistoric, and genetic data sets using population genetics and multivariate statistics models (Cavalli-Sforza et al. 1988; Cavalli-Sforza, Menozzi, and Piazza 1994; Chen, Sokal, and Ruhlen 1995; Sokal et al. 1996).

In addition to the aforementioned indirect, inferential procedures for reconstructing human evolutionary history, the portentous replicated recovery of Neanderthal mtDNA detailed in Krings et al. (1997) has ushered in a much more direct kind of evidence with the potential for actually distinguishing among competing evolutionary scenarios (Ward and Stringer 1997). This is, indeed, an exciting time for students of human evolution!

Acknowledgments

We thank James Tuggle, Jared Ragland, Roxane Bonner, Ammon Corl, Agnish Chakravarti, and Matt Kaplan for excellent technical assistance, and Jody Hey and an anonymous reviewer for helpful comments. We also thank Peter Underhill for making available unpublished polymorphism information. This publication was made possible by grant GM-53566 from the National Institute of General Medical Sciences (to M.F.H.). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. This work was also supported by grant OPP-9423429 from the National Science Foundation (to M.F.H.).

LITERATURE CITED

ALTHEIDE, T. K., and M. F. HAMMER. 1997. Evidence for a possible Asian origin of YAP^+ Y chromosomes. Am. J. Hum. Genet. **61**:462–466.

AMMERMAN, A. J., and L. L. CAVALLI-SFORZA. 1984. Neolithic transition and the genetics of populations in Europe. Princeton University Press, Princeton, N.J.

AVISE, J. C. 1994. Molecular markers, natural history and evolution. Chapman and Hall, New York.

BERMÚDEZ DE CASTRO, J. M., J. L. ARSUAGA, E. CARBONELL, A. ROSAS, I. MARTÍNEZ, and M. MOSQUERA. 1997. A hominid from the Lower Pleistocene from Atapuerca, Spain: possible ancestor to Neandertals and modern humans. Science **276**:1392–1395.

BRAUER, G. 1989. The evolution of modern humans: a comparison of the African and non-African evidence. Pp. 123–154 *in* P. MELLARS and C. STRINGERS, eds. The human revolution: behavioural and biological perspectives on the or-

igins of modern humans. Edinburgh University Press, Edinburgh.

BROOKFIELD, J. F. Y. 1997. Importance of ancestral DNA ages. Nature **388**:134.

CANN, R. L. 1997. Phylogenetic estimation in humans and neck riddles. Am. J. Hum. Genet. **60**:755–757.

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**:247–251.

CAVALLI-SFORZA, L. L., A. PIAZZA, P. MENOZZI, and J. MOUNTAIN. 1988. Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. Proc. Natl. Acad. Sci. USA **85**:6002–6006.

CHEN, J., R. R. SOKAL, and M. RUHLEN. 1995. Worldwide analysis of genetic and linguistic relationships of human populations. Hum. Biol. **67**:595–612.

EXCOFFIER, L., and A. LANGANEY. 1989. Origin and differentiation of human mitochondrial DNA. Am. J. Hum. Genet. **44**:73–85.

FENG, D.-F., and R. F. DOOLITTLE. 1987. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J. Mol. Evol. **25**:351–360.

FOOTE, S., D. VOLLRATH, A. HILTON, and D. C. PAGE. 1992. The human Y chromosome: overlapping DNA clones spanning the euchromatic region. Science **258**:60–66.

FU, Y.-X., and W.-H. LI. 1997. Estimating the age of the common ancestor of a sample of DNA sequences. Mol. Biol. Evol. **14**:195–199.

GRIFFITHS, R. C. 1989. Genealogical tree probabilities in the infinitely-many-sites model. J. Math. Biol. **27**:667–680.

GRIFFITHS, R. C., and S. TAVARÉ. 1994. Ancestral inference in population genetics. Stat. Sci. **9**:307–319.

HAMMER, M. F. 1995. A recent common ancestry for human Y chromosomes. Nature **378**:376–378.

HAMMER, M. F., A. B. SPURDLE, T. KARAFET et al. (11 coauthors). 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.

HORAI, S., K. HAYASAKA, R. KONDO, K. TSUGANE, and N. TAKAHATA. 1995. Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. Proc. Natl. Acad. Sci. USA **92**:532–536.

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.

KARAFET, T., S. L. ZEGURA, J. VUTURO-BRADY et al. (14 coauthors). 1997. Y chromosome markers and trans-Bering Strait dispersals. Am. J. Phys. Anthropol. **102**:301–314.

KINGMAN, J. F. C. 1982. The coalescent. Stoch. Proc. Appl. **13**:235–248.

KONIGSBERG, L. W. 1990. Analysis of prehistoric biological variation under a model of isolation by geographic and temporal distance. Hum. Biol. **62**:49–70.

KONIGSBERG, L. W., and J. BLANGERO. 1993. Multivariate quantitative genetic simulations in anthropology with an example from the South Pacific. Hum. Biol. **65**:897–915.

KRINGS, M., A. STONE, R. W. SCHMITZ, H. KRAINITZKI, M. STONEKING, and S. PAABO. 1997. Neandertal DNA sequences and the origin of modern humans. Cell **90**:19–30.

LOWELL, T. V., C. J. HEUSSER, B. G. ANDERSEN, P. I. MORENO, A. HAUSER, L. E. HEUSSER, C. SCHLUCHTER, D. R. MARCHANT, and G. H. DENTON. 1995. Interhemispheric correlation of Late Pleistocene glacial events. Science **269**:1541–1549.

MELTON, T., R. PETERSON, A. J. REDD, N. SAHA, A. S. M. SOFRO, J. MARTINSON, and M. STONEKING. 1995. Polynesian genetic affinities with southeast Asian populations as identified by mtDNA analysis. Am. J. Hum. Genet. **57**:403–414.

MINUGH-PURVIS, N. 1995. The modern human origins controversy: 1984–1994. Evol. Anthropol. **4**:140–147.

MOUNTAIN, J. L., and L. L. CAVALLI-SFORZA. 1994. Inference of human evolution through cladistic analysis of nuclear DNA restriction polymorphisms. Proc. Natl. Acad. Sci. USA **91**:6515–6519.

NEI, M., and N. TAKEZAKI. 1996. The root of the phylogenetic tree of human populations. Mol. Biol. Evol. **13**:170–177.

PENNY, D., M. STEEL, P. J. WADDELL, and M. D. HENDY. 1995. Improved analyses of human mitochondrial DNA sequences support a recent African origin for *Homo sapiens*. Mol. Biol. Evol. **12**:863–882.

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.

RICHARDS, M., H. CORTE-REAL, P. FORSTER, V. MACAULAY, H. WILKINSON-HERBOTS, A. DEMAINE, S. PAPIHA, R. HEDGES, H. J. BANDELT, and B. SYKES. 1996. Paleolithic and Neolithic lineages in the European mitochondrial gene pool. Am. J. Hum. Genet. **59**:185–203.

RICHARDS, M., V. MACAULAY, B. SYKES, P. PETTITT, R. HEDGES, P. FORSTER, and H.-J. BANDELT. 1997. Paleolithic and Neolithic lineages in the European mitochondrial gene pool: reply to Cavalli-Sforza and Minch. Am. J. Hum. Genet. **61**:251–254.

SMITH, F. H., A. B. FALSETTI, and S. M. DONNELLY. 1989. Modern human origins. Yearb. Phys. Anthropol. **32**:35–68.

SOKAL, R. R., N. L. ODEN, J. WALKER, D. DI GIOVANNI, and B. A. THOMSON. 1996. Historical population movements in Europe influence genetic relationships in modern samples. Hum. Biol. **68**:873–898.

STONEKING, M. 1993. DNA and recent human evolution. Evol. Anthropol. **2**:60–73.

———. 1994. In defense of ''Eve'': a response to Templeton's critique. Am. Anthropol. **96**:131–141.

SWOFFORD, D. L. 1990. PAUP: phylogenetic analysis using parsimony. Version 3.0. Illinois Natural History Survey, University of Illinois, Champaign-Urbana.

TAVARÉ, S., D. J. BALDING, R. C. GRIFFITHS, and P. DONNELLY. 1997. Inferring coalescence times from DNA sequence data. Genetics **145**:505–518.

TEMPLETON, A. R. 1993. The ''Eve'' hypothesis: a genetic critique and reanalysis. Am. Anthropol. **95**:51–72.

———. 1994. ''Eve'': hypothesis compatibility versus hypothesis testing. Am. Anthropol. **96**:141–147.

———. 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.

———. 1998*a*. Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. Mol. Ecol. (in press).

———. 1998*b*. Human races: a genetic and evolutionary perspective. Am. Anthropol. (in press).

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.

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 (DHPLC). Genome Res. **7**:996–1005.

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.

WARD, R., and C. STRINGER. 1997. A molecular handle on the Neanderthals. Nature **388**:225–226.

WATTERSON, G. A. 1975. On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7**:256–276.

WEISS, G., and A. VON HAESELER. 1996. Estimating the age of the common ancestor of men from the *ZFY* intron. Science **272**:1359–1360.

WHITFIELD, L. S., J. E. SULSTON, and P. N. GOODFELLOW. 1995. Sequence variation of the human Y chromosome. Nature **378**:379–380.

WILLIAMS-BLANGERO, S., and J. BLANGERO. 1989. Anthropometric variation and the genetic structure of the Jirels of Nepal. Hum. Biol. **61**:1–12.

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.

SIMON EASTEAL, reviewing editor

Accepted December 29, 1997