The Geographic Distribution of Human *Y* Chromosome Variation

M. F. Hammer,* A. B. Spurdle,† T. Karafet,* M. R. Bonner,* E. T. Wood,* A. Novelletto,‡ P. Malaspina,‡ R. J. Mitchell,§ S. Horai, T. Jenkins† and S. L. Zegura††**

**Laboratory of Molecular Systematics and Evolution and ††Department of Anthropology, University of Arizona, Tucson, Arizona 85721, †Department of Human Genetics, South African Institute for Medical Research and the University of Witwatersrand, Johannesburg 2000, South Africa, ‡Departimento di Biologia, Universita degli Studi "Tor Vergata", Rome 00173, Italy, §Department of Genetics and Human Variation, La Trobe University, Melbourne, Victoria 3038, Australia, and **Department of Human Genetics, National Institute of Genetics, Mishima 411, Japan*

Manuscript received July 7, 1996
Accepted for publication November 27, 1996

ABSTRACT

We examined variation on the nonrecombining portion of the human *Y* chromosome to investigate human evolution during the last 200,000 years. The *Y*-specific polymorphic sites included the *Y* Alu insertional polymorphism or "YAP" element (*DYS287*), the poly(A) tail associated with the YAP element, three point mutations in close association with the YAP insertion site, an A-G polymorphic transition (*DYS271*), and a tetranucleotide microsatellite (*DYS19*). Global variation at the five bi-allelic sites (*DYS271, DYS287,* and the three point mutations) gave rise to five "YAP haplotypes" in 60 populations from Africa, Europe, Asia, Australasia, and the New World (n = 1500). Combining the multi-allelic variation at the microsatellite loci (poly(A) tail and *DYS19*) with the YAP haplotypes resulted in a total of 27 "combination haplotypes". All five of the YAP haplotypes and 21 of the 27 combination haplotypes were found in African populations, which had greater haplotype diversity than did populations from other geographical locations. Only subsets of the five YAP haplotypes were found outside of Africa. Patterns of observed variation were compatible with a variety of hypotheses, including multiple human migrations and range expansions.

GENETIC data have played a key role in current debates about the origin of modern humans. In particular, human mitochondrial DNA data (CANN *et al.* 1987; VIGILANT *et al.* 1991; STONEKING 1993; PENNY *et al.* 1995) have been interpreted to support "Recent Out of Africa" models (MINUGH-PURVIS 1995) rather than the "Multiregional Evolution" model (FRAYER *et al.* 1993). The realization that mtDNA represents a single locus has led to the suggestion that accurate inferences about human population history require the examination of many loci (STONEKING 1993; SZATHMARY 1993; TEMPLETON 1993). Indeed, extensive batteries of multilocus data have been used to portray population affinities within *Homo sapiens* (BATZER *et al.* 1994; BOWCOCK *et al.* 1994; MOUNTAIN and CAVALLI-SFORZA 1994). However, as TEMPLETON (1993, 1996) has pointed out, there have been no critical tests of the two major competing human origin(s) models. Moreover, there are inherent advantages in studying genetic systems without recombination. For instance, they provide opportunities to (1) determine character state polarity, (2) construct trees showing the phylogenetic history of individual haplotypes, and (3) estimate the time frame for the branching events in the trees.

Over the last decade attention has turned to the largest nonrecombining region in the human genome, the male-specific portion of the *Y* chromosome. Although the first *Y*-linked restriction fragment length polymorphisms were identified in 1985 (CASANOVA *et al.* 1985; LUCOTTE and NGO 1985), the paucity of known variation on the *Y* chromosome throughout the late 1980's limited the role of this chromosome in testing hypotheses about the origin and migration patterns of human populations. However, recent methodological advances in the detection of both single nucleotide site and microsatellite variation on the *Y* chromosome (HAMMER 1995; JOBLING and TYLER-SMITH 1995; WHITFIELD *et al.* 1995; UNDERHILL *et al.* 1996) have facilitated studies encompassing such diverse topics as paternal migration at the continental level (TORRONI *et al.* 1994a; HAMMER and HORAI 1995; KARAFET *et al.* 1997; UNDERHILL *et al.* 1996), the demographic histories of single populations and the genetic affinities of closely related groups (ROEWER *et al.* 1996), the number of male founders in isolated populations (NYSTUEN *et al.* 1995), levels of population admixture (HAMMER and HORAI 1995; KARAFET *et al.* 1997), and male and female reproductive patterns (LUCOTTE *et al.* 1994; SALEM *et al.* 1996).

Although our paper is not concerned with testing hypotheses associated with Recent Out of Africa models or Multiregional Evolution, it does address possible early and subsequent human migrations from both global and regional perspectives. Mitochondrial DNA

Corresponding author: Michael Hammer, Department EEB, Biosciences West, University of Arizona, Tucson, AZ 85721.
E-mail: hammer@biosci.arizona.edu

TABLE 1

YAP haplotype and *DYS19* allele frequencies in 60 populations (n = 1500)

Population	n	YAP haplotype					*DYS19* allele					
		1	2	3	4	5	Z	A	B	C	D	E
Khoisan (KHO)												
Sekele	34	32	24	0	12	32	9	15	12	15	37	12
Tsumkwe	31	51	42	0	3	3	16	27	3	19	19	16
Nama	13	62	8	0	15	15	0	23	8	23	15	31
Pygmies (PYG)												
Mbuti	14	50	0	14	0	36	7	29	21	21	7	14
Biaka	14	42	0	0	0	58	0	0	0	21	36	43
West Africans (WAF)												
Gambians[a]	48	15	0	17	10	58	0	8	14	29	29	19
Nigerians[a]	8	0	0	13	0	87	0	0	13	63	13	13
East Africans (EAF)												
Bagandans	29	10	0	10	0	80	0	0	13	38	28	21
Kenyans[a]	18	6	0	11	0	83	0	0	17	44	33	6
Ethiopians[a]	3	67	0	0	33	0	0	0	33	33	33	0
East Bantus (EBA)												
Nguni[a]	41	22	0	15	2	61	0	0	17	49	29	5
Zulu	28	21	0	18	4	57	0	0	22	39	32	7
Ndebele	7	14	0	0	0	86	0	0	0	71	29	0
Xhosa	4	25	0	25	0	50	0	0	25	50	25	0
Swazi	2	50	0	0	0	50	0	0	0	100	0	0
Sotho/Tswana[a]	37	30	0	5	5	60	0	5	8	59	19	9
Sotho	15	47	0	7	13	33	0	13	14	60	13	0
Tswana	13	15	0	8	0	77	0	0	8	61	23	8
Pedi	9	22	0	0	0	78	0	0	0	56	22	22
Tsonga	6	50	0	0	17	33	0	17	0	66	17	0
Venda	2	50	0	0	0	50	0	0	0	50	50	0
West Bantus (WBA)												
Herero/Himba[a]	29	28	0	0	0	72	0	0	14	38	34	14
Herero	27	26	0	0	0	74	0	0	15	34	36	15
Himba	2	50	0	0	0	50	0	0	50	50	0	0
Ambo	26	4	0	8	8	80	0	8	12	53	15	12
Enigmatic Southern Africans (ESA)												
Dama	22	21	5	9	5	60	0	5	36	40	5	14
Lemba	34	67	0	3	6	24	0	6	61	24	9	0
North Africans (NAF)												
Egyptians[a]	93	51	0	1	46	2	1	19	45	21	12	2
West Asians (WAS)												
Saudi Arabians[a]	22	91	0	0	5	5	0	5	55	23	18	0
United Arab Emirates[a]	20	90	0	0	5	5	0	5	65	15	15	0
Omanis[a]	11	100	0	0	0	0	0	0	36	64	0	0
Iranians[a]	5	80	0	0	0	20	0	0	20	80	0	0
East Asians (EAS)												
Japanese[a]	132	58	0	42	0	0	0	8	3	49	25	15
Aomoris	26	61	0	39	0	0	0	7	7	55	24	7
Shizuokans	61	67	0	33	0	0	0	8	2	57	21	12
Okinawans	45	45	0	55	0	0	0	7	2	36	32	23
Tibetans[a]	30	53	0	47	0	0	0	0	23	67	10	0
Koreans[a]	27	100	0	0	0	0	0	0	26	22	41	11
Taiwanese[a]	76	99	0	1	0	0	1	4	28	38	25	4
South Chinese[a]	48	100	0	0	0	0	0	4	10	52	25	8
South Asians (SAS)												
Indians[a]	39	100	0	0	0	0	0	0	23	51	21	5
Southeast Asians[a]	47	100	0	0	0	0	0	2	7	55	23	13
Vietnamese	13	100	0	0	0	0	0	8	8	46	23	15
Philippinos	10	100	0	0	0	0	0	0	10	60	10	20
Malaysians	10	100	0	0	0	0	0	0	0	70	10	20
Laotians	7	100	0	0	0	0	0	0	0	43	57	0
Indonesians	4	100	0	0	0	0	0	0	0	50	50	0
Cambodians	3	100	0	0	0	0	0	0	33	67	0	0

TABLE 1

Continued

Population	*n*	YAP haplotype					*DYS19* allele					
		1	2	3	4	5	Z	A	B	C	D	E
South Europeans (SEU)												
Greeks[a]	83	75	0	0	25	0	0	22	37	27	13	1
Mainland Greeks[a]	42	62	0	0	38	0	0	31	38	17	12	2
Island Greeks[a]	17	82	0	0	18	0	0	12	35	41	12	0
Cretans	24	92	0	0	8	0	0	12	38	33	17	0
Italians[a]	208	86	0	1	13	0	0	14	37	31	10	8
General Italians[a]	52	83	0	0	17	0	0	19	46	23	6	6
Sardinians	55	93	0	0	7	0	0	7	25	33	15	20
Calabrians	29	90	0	0	10	0	0	7	45	41	4	3
Venetians	27	89	0	0	11	0	0	11	41	26	19	3
Puglians	21	81	0	0	19	0	0	24	38	29	9	0
Sicilians	24	79	0	4	17	0	0	21	25	42	4	8
North Europeans (NEU)												
Germans[a]	30	93	0	0	7	0	0	7	50	30	10	3
British[a]	43	98	0	0	2	0	0	12	63	23	2	0
South African Europeans[a]	20	95	0	0	5	0	0	5	60	15	15	5
Australasians (AUS)												
Australian Aboriginal People[a]	43	98	0	0	2	0	0	0	23	35	33	9
Great Sandy Desert	36	100	0	0	0	0	0	0	17	36	36	11
Western Australians[a]	7	86	0	0	14	0	0	0	57	29	14	0
Northern Australians[a]	19	100	0	0	0	0	0	0	32	26	11	32
Papua New Guineans	48	100	0	0	0	0	0	0	33	46	15	6
Native Americans (NAM)												
Navajos	47	100	0	0	0	0	0	57	32	11	0	0

Frequencies were rounded off to the nearest percent. As a result some haplotype rows do not sum to 100%.
[a] Composite samples.

has been used to trace maternally based genetic trails from Asia both to the Pacific (MELTON *et al.* 1995) and to the Americas (MERRIWETHER *et al.* 1995). Likewise, *Y* chromosome (paternal) genetic trails have been used to test hypotheses about the peopling of Japan (HAMMER and HORAI 1995) and to formulate theories about the peopling of the Americas (KARAFET *et al.* 1997; UNDERHILL *et al.* 1996). Unfortunately, there has been only one published study that qualifies as a global survey of human *Y* chromosome haplotype variation, JOBLING and TYLER-SMITH's (1995) excellent review paper, which also presents new compound haplotype data.

HAMMER (1995) performed a sequencing survey of 16 geographically diverse humans for a 2.6-kb region (referred to as the "YAP" region) on the long arm of the human *Y* chromosome and constructed a tree with five *Y* chromosome haplotypes. The mean time to the common *Y* chromosome ancestor was estimated to be ~188 kya (thousand years ago) with a 95% confidence interval from 51 to 411 kya. Other estimates of the age of our common *Y* chromosome ancestor concur in suggesting that analyses of *Y*-linked variation will eventually be useful for testing hypotheses about human evolution during the last 200,000 years (FU and LI 1996; WEISS and VON HAESELER 1996). The present study surveys the aforementioned five YAP haplotypes in a large worldwide sample. In addition, variation at the *DYS271* A-G transition (SEIELSTAD *et al.* 1994) and the *DYS19* microsatellite (ROEWER *et al.* 1992) loci was investigated. The combination of data from these loci with the YAP haplotypes resulted in a total of 27 "combination" haplotypes and formed the basis for our exploration of global *Y* chromosome variation.

MATERIALS AND METHODS

Sample composition: We analyzed a total of 1500 males from 60 populations. These populations were categorized into 15 major groups based on geographic, linguistic, and ethnohistorical criteria (Table 1). A total of 78 samples came from Khoisan-speakers originating from southern Africa (SPURDLE and JENKINS 1992). This grouping was composed of the Khoi (13 Nama from Namibia) and the !Kung San (34 Sekele and 31 Tsumkwe from Namibia). We studied 28 Pygmies (BOWCOCK *et al.* 1994) including 14 Mbuti Pygmies from Zaire and 14 Biaka Pygmies from the Central African Republic. Biaka Pygmies speak languages belonging to the Niger Kordofanian family, whereas, Mbuti Pygmies speak central Sudanic languages belonging to the Nilo-Saharan language family (CAVALLI-SFORZA *et al.* 1994). The West African sample consisted of 48 diverse Gambians (14 Wolof, 13 Mandinka, seven Serere, four Manjago, three Jola, two Fula and five other) (SOODYALL *et al.* 1996), and eight Nigerians (WAINSCOAT *et al.* 1986). The East African sample consisted of 29 Bagandans from Uganda, 18 Kenyans (WAINSCOAT *et al.* 1986; MATHIAS *et al.* 1994), and three Ethiopians. The 141 samples from South African Bantu-speaking groups represented different chiefdoms, each speaking a different Bantu language (SPURDLE and JENKINS 1992).

The southeastern and southwestern Bantu-speakers are postulated to have followed different migratory paths from west-central Africa, a general southbound course (southeastern Bantu) and a southwesterly route across central Africa toward the western parts of the continent (southwestern Bantu) (SPURDLE and JENKINS 1992). Eighty-six southeastern Bantu speakers (referred to as "East Bantus") were grouped into 41 Nguni (28 Zulu, seven Ndebele, four Xhosa, and two Swazi), 37 Sotho/Tswana (15 Sotho, 13 Tswana, and nine Pedi), six Tsonga, and two Venda. The 55 southwestern Bantu speakers (referred to as "West Bantus") included 27 from the Herero, 26 from the Ambo, and two from the Himba chiefdoms. Two enigmatic southern African groups were included in this study. The 34 Lemba are eastern Bantu speakers (Venda) with a possible Semitic origin (SPURDLE and JENKINS 1996), and the 22 Dama are from a Khoi-speaking group, who are biologically more closely related to non-Khoisan African groups (SPURDLE and JENKINS 1992). The North African group included 93 Egyptians (31 from the Dakhalia region northeast of Cairo and 62 with paternal birthplaces both north and south of Cairo) (CIMINELLI *et al.* 1995).

The West Asian composite group included 22 Saudi Arabian, 20 United Arab Emirates, 11 Omani, and five Iranian samples (CIMINELLI *et al.* 1995). Additional Asian samples were grouped into East Asians and South Asians based on ethnohistorical criteria (CAVALLI-SFORZA *et al.* 1994). The East Asian group included 132 Japanese (45 Okinawans, 61 Shizuokans, and 26 Aomoris) (HAMMER and HORAI 1995), 27 Koreans, 30 Tibetans, 76 Taiwanese (mainly Hakkas and immigrants from southern China), and 48 southern mainland Chinese (MELTON *et al.* 1995). The South Asian group included 39 Indians [19 from Andra Pradesh and 20 from South Africa who originated mainly from the Gujerat and Bengali provinces (SPURDLE and JENKINS 1996)] and 47 Southeast Asians (13 Vietnamese, 10 Philippinos, 10 Malaysians, seven Laotians, four Indonesians, and three Cambodians).

Populations surveyed from southern Europe included 83 Greeks (24 from Crete and 59 Greek-born males presently residing in Australia, including 17 from various Greek Islands, 19 from Macedonia, nine from Peloponissos, and 14 from other mainland locations (MITCHELL *et al.* 1993; CIMINELLI *et al.* 1995), and 208 Italians (55 from Sardinia, 29 from Calabria, 27 from Veneto, 24 from Sicily, 21 from Puglia, and 52 from other continental Italian locations) (PERSICHETTI *et al.* 1992; MITCHELL *et al.* 1993; CIMINELLI *et al.* 1995). The 40 Italian and 59 Greek samples collected in Melbourne, Australia came from volunteers who filled out a questionnaire in their own language verifying their place of birth. The northern European sample consisted of 30 Germans, 43 English-born males (19 of whom are presently residing in Australia) (PERSICHETTI *et al.* 1992; CIMINELLI *et al.* 1995), and 20 South African Europeans (SPURDLE and JENKINS 1992).

The Australasian sample included 43 Australian Aboriginal People (36 from the northern edge of the Great Sandy Desert and seven from the west coast of Australia) (PERNA *et al.* 1992), 19 indigenous northern Australians, and 48 Papua New Guineans (18 from the coast and 30 from the highlands) (STONEKING and WILSON 1989). A single sample of 47 Navajos from Northern Arizona and New Mexico represented Native Americans.

DNA extraction: DNA samples provided by other investigators were as follows: seven Nigerians and three Kenyans by J. S. WAINSCOAT; 14 Biaka Pygmies, 14 Mbuti Pygmies, three Ethiopians, and three Cambodians by K. KIDD, J. KIDD, and J. ROGERS; 29 Bagandans by L. LOUIE; 21 Egyptians by Y. GAD; 30 Germans by G. RAPPOLD; 27 South Chinese, 43 Australian Aboriginal People, and 48 Papua New Guineans by M. STONEKING; and 48 Gambians by A. V. S. HILL. Blood samples from Taiwan were provided by L. L. HSIEH. Whole blood DNA was extracted using the procedure of LAHIRI and NURNBERGER (1991) and as described in SPURDLE and JENKINS (1992). Many of the DNA samples were prepared from buccal cells collected at the University of Arizona (where the sampling protocol was approved by the Human Subjects Committee). Buccal cell DNA was isolated according to the procedure of RICHARDS *et al.* (1993).

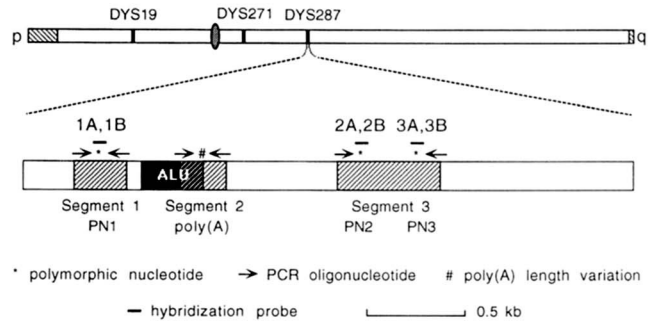

FIGURE 1.—Schematic representation of the *Y* chromosome and the specific genetic regions studied. The short arm (p), the centromere (shaded oval), and the long arm (q) of the *Y* chromosome are indicated in the top illustration. The three dark vertical lines refer to the *DYS19*, *DYS271*, and *DYS287* loci. ▧ at the termini of the chromosome represent the pseudoautosomal regions. The bottom shows an enlargement of *DYS287* (the YAP region). The three segments 1–3 (▨) represent the PCR-amplified regions containing the polymorphic nucleotide sites (*). ■ designates the location of the YAP element when present. # refers to the variable length poly(A) tail at the 3′ end of the YAP element. Arrows refer to oligonucleotide primers used in the PCR reactions (primers used for the amplification of the region encompassing the YAP element are not shown), – refer to hybridization probes, and numbers/letters above the horizontal lines refer to the names of the hybridization probes.

PCR amplification: Three segments within the 2.6-kb YAP region (including *DYS287*) (HAMMER 1995) were amplified (Figure 1). Segment 1 encompasses the C-T transition (PN1), segment 2 encompasses the variable length polydeoxyadenylate tract (poly-A tail) at the 3′ end of the YAP sequence, and segment 3 encompasses the C-T (PN2) and G-A (PN3) transitions. The following primer pairs were used to amplify segments 1–3, respectively: R368 (5′-TCACATAATTTCATTTTCCC-3′) and F591 (5′-TAGTCCTCTCCTTATTAACG-3′), AR237 (5′-CGCCATTGCAGTCCGCAGTC-3′) and F1070 (5′-CAAGTTAGCTGTCCATACTG-3′), and R1962 (5′-GATGCAAATGAGAAAGAACT-3′) and F2476 (5′-CTAAAAACTGGAGGGAGAAA-3′). The AR237 primer is part of the 3′ end of the YAP element and hybridizes to a number of Alu elements in the genome; whereas, the F1070 primer is derived from a flanking single-copy sequence. 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, 2.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-HCl (pH 8.3), and 0.5 U AmpliTaq DNA polymerase (Perkin-Elmer). The cycling conditions were 94° for 2 min, and then 30 cycles of 94° for 1 min, 51° for 1 min, and 72° for 1 min. A fourth segment encompassing the insertion site of the YAP element (not shown in Figure 1) was amplified as described in HAMMER and HORAI (1995).

The *DYS271* and *DYS19* loci were amplified and genotyped according to the procedures of SEIELSTAD *et al.* (1994) and HAMMER and HORAI (1995), respectively. Polymorphism in the length of the poly(A) tail on the 3′ end of the Alu sequence was analyzed in the following manner. PCR products were mixed with loading dye and resolved on 20 × 35 ×

0.15 cm 6% polyacrylamide/0.15% bisacrylamide gels in TBE buffer (50 mM Tris-borate/EDTA pH 8.3) for 8 hr at 300 V. Following electrophoresis, the gel was stained with ethidium bromide and the fragments were visualized by UV light.

Site-specific oligonucleotide (SSO) hybridization: SSO probes were designed according to the methods described by IKUTA *et al.* (1987). Figure 1 shows the location of the hybridization probes to detect the polymorphic base substitutions in segments 1 and 3. The sequences of these probes are as follows: Probe 1A, 5′-GAGAGCCTTTTGTCTTAAA-3′; probe 1B, 5′-TTAAGACAAAAAGCTCTCC-3′; probe 2A, 5′-AAACTAATGCCTTCTCCTC-3′; probe 2B, 5′-GGAGGA-GAAAGCATTAGTT-3′; probe 3A, 5′-GCTTGAAGGGGAGG-TAAAT-3′; and probe 3B, 5′-AATTTACCTCCCCCTCAAG-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, 0.5% SDS) for 30 min at 53°. Labeled SSO probes were then added directly to the hybridization solution to a concentration of 2 pmol/ml, and hybridization was carried out for 2 hr at 53°. Membranes were rinsed in wash solution (2× SSPE, 0.1% SDS) at room temperature, washed at 53 for 30 min, and exposed to film for 2–48 hr.

Data analysis: For each individual, the combination of sequence variants (both point and length mutations) observed across the three YAP segments and the *DYS271* and *DYS19* loci is referred to as a *Y* chromosome haplotype. An unbiased estimate of haplotype diversity (h) equivalent to heterozygosity in each population was calculated using Equation 8.5 of NEI (1987). The variance of h was determined as described in NEI (1978). The diversity values can theoretically range from 0 to 1, with 0 representing the minimum value (*i.e.*, a single haplotype fixed in a population) and 1 representing the maximum possible genetic variability (*i.e.*, an equal frequency of an infinitely large number of haplotypes). The same measure was used to estimate diversity carried by each of the five YAP haplotypes. Fst analyses were performed using the computer program Antana (HARPENDING and ROGERS 1984) for the YAP data. For the microsatellite locus, MICHALAKIS and EXCOFFIER's (1996) PHIst with AMOVA procedure was adopted. These PHIst values are related to SLATKIN's (1995) Rst (GOLDSTEIN *et al.* 1995). Associations between YAP haplotypes and alleles at the *DYS19* locus were investigated using the linkage disequilibrium (D) method of WEIR (1990). Normalized disequilibrium values (D') between −1 and +1 were determined by dividing D by the appropriate denominator given on p. 97 in WEIR (1990).

The PHYLIP Package (FELSENSTEIN 1993) was used to compute genetic distances for use in distance matrix programs and for constructing maximum likelihood phylogeny estimates based on *Y* chromosome haplotype frequencies. Genetic distances were generated using two different methods: CAVALLI-SFORZA's chord distance ($4\bar{D}$) (CAVALLI-SFORZA and BODMER 1971) and REYNOLDS *et al.*'s (1983) genetic distance (Fst). These methods assume that there is no new mutation and that all gene frequency changes are due to genetic drift. Constant and equal population sizes are not assumed. The neighbor-joining (NJ) (SAITOU and NEI 1987) and the Unweighted Pair Group Method Using Arithmetic Averages (UPGMA) (SOKAL and MICHENER 1958) options of the program NEIGHBOR and the Fitch and Margoliash (FM) option (FITCH and MARGOLIASH 1967) of the program FITCH were used to construct branching diagrams from matrices of pairwise distances. The NJ and FM programs fit a tree that has unconstrained branch lengths and assume both additivity and independent errors. The UPGMA option constructs a tree by successive agglomerative mergers using an average-linkage method and assumes that an "evolutionary clock" is valid. Accordingly, the true branch lengths from the root of the tree to each tip are the same and the expected amount of evolution in any lineage is proportional to elapsed time. The CONTML program was used to estimate phylogenies based on restricted maximum likelihood using *Y* chromosome haplotype frequencies (FELSENSTEIN 1981). This program takes haplotype frequencies as input and performs a square-root transformation on the data. Then the Brownian motion model is used on the resulting coordinates. It assumes that each locus evolves independently via genetic drift. The CONSENSE program was used to read the trees produced from the CONTML, NEIGHBOR, and FITCH programs and compute majority-rule consensus trees.

RESULTS

Variation at four polymorphic sites within the YAP region (*DYS287*): The frequency of *Y* chromosomes with the YAP element (YAP$^+$) varied greatly among geographical regions. Consistent with previous surveys (HAMMER 1994; SEIELSTAD *et al.* 1994; SPURDLE *et al.* 1994), YAP$^+$ chromosomes were found at high frequencies in samples throughout the African continent (mean frequency = 61.7%): East Africans (88.0%), West Africans (87.5%), South African Bantu-speakers (76.6%), Pygmies (53.5%), and North Africans (49.5%). The Khoisan sample had the lowest frequency of YAP$^+$ chromosomes (26.9%) of all the African samples. Although YAP$^+$ chromosomes were not found in the South Asian or New World samples, they were found at extremely low frequencies in the Australasian sample (1.0%), at low frequencies in the West Asian (8.6%) and European (13.8%) samples, and at moderate frequencies in the East Asian sample (22.3%).

Direct DNA sequencing of the YAP element in a small sample of YAP$^+$ chromosomes demonstrated that the number of adenine residues at the 3′ end of the Alu sequence (poly(A) tail) is variable (HAMMER 1995). Three poly(A) tail length alleles were identified at this mononucleotide microsatellite site: long (L: 46 bp), short (S: 26 bp), and very short (VS: 19 bp). Oligonucleotide primers were designed to amplify an ~220-bp DNA fragment encompassing the poly(A) tail of the YAP element (Figure 1, segment 2). After PCR amplification of this segment from all 437 YAP$^+$ chromosomes in this survey, we observed the three known length alleles (Figure 2A) as well as a new poly(A) allele intermediate in size (36 bp) between the L and S alleles. We refer to this as the medium (M) allele (not shown in Figure 2A). The L allele was associated with 88.4%, 9.9%, and 2.2% of the East Asian, African, and European YAP$^+$ chromosomes, respectively. The S allele (absent in East Asia) was associated with 97.8% and 90.1% of the European and African YAP$^+$ chromosomes, respectively. The VS and M alleles were identified only in association with East Asian YAP$^+$ chromosomes (7.1% and 4.5%, respectively).

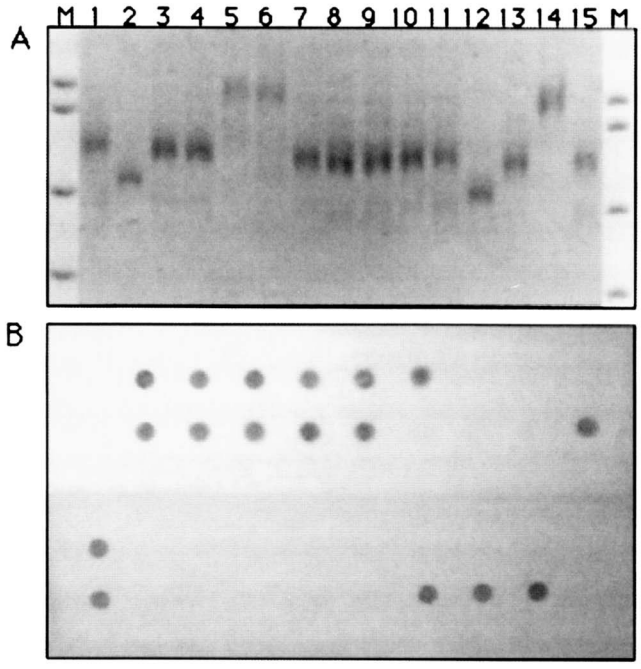

FIGURE 2.—Detection and rapid genotyping of alleles at the poly(A) tail of the YAP element and the three polymorphic nucleotide sites in the YAP region. (A) Polyacrylamide gel electrophoresis of the segment 2 PCR product from 15 YAP$^+$ males. Lanes 2 and 12 contain samples with the poly(A)-VS allele; lanes 1, 3, 4, 7–11, 13, and 15 contain samples with the poly(A)-S allele, and lanes 5, 6, and 14 contain samples with the poly(A)-L allele. Lanes marked with M contain a molecular weight standard (pBR-*Msp* 1: 242, 238, 217, and 201 bp). (B) Autoradiograph of SSO hybridization with probes 2A and 2B. Seventeen DNA samples were hybridized in turn with the 2A and 2B probes. The hybridization signals seen in the top two rows indicate detection of the C allele by probe 2A, while signals in the bottom two rows indicate detection of the T allele by probe 2B.

SSO typing of variation at three polymorphic nucleotide sites (PN1–PN3) in the YAP region (Figure 1) was performed on 16 individuals of known sequence (HAMMER 1995) to verify the accuracy of the method. For each of the three polymorphic sites (PN1, C-T; PN2, C-T; PN3, G-A), specific hybridization with only the corresponding probe was observed (Figure 2B). Subsequently, all 1500 individuals were typed at PN1, PN2, and PN3 to determine the frequencies of the YAP region sequence variants. The PN1-T variant was found in Africa at a frequency of 42.8% and in West Asia at a frequency of 5.2%. The PN2-T variant, absent in the Asian samples, was found at a frequencies of 55.5.% and 13.5% in African and European populations, respectively. The PN3-A variant was found only in Khoisan populations, where it had a frequency of 28.2%, and in a single Dama individual. There was maximal linkage disequilibrium among all of the polymorphic loci in the YAP region, consistent with the hypothesis that each of the polymorphisms originated a single time in human evolution. Variation at the four polymorphic sites resulted in five YAP haplotypes (Figure 3, thick solid lines) (HAMMER 1995). The four length alleles associated with YAP$^+$ chromosomes produced three subtypes of haplotype 3, referred to as 3a, 3b, and 3c (Figure 3). Because the S allele was associated with the PN1-T and PN2-T variants in all cases in this survey, it was not possible to determine the order of occurrence of the poly(A)L-S or PN2 C-T mutational events (Figure 3).

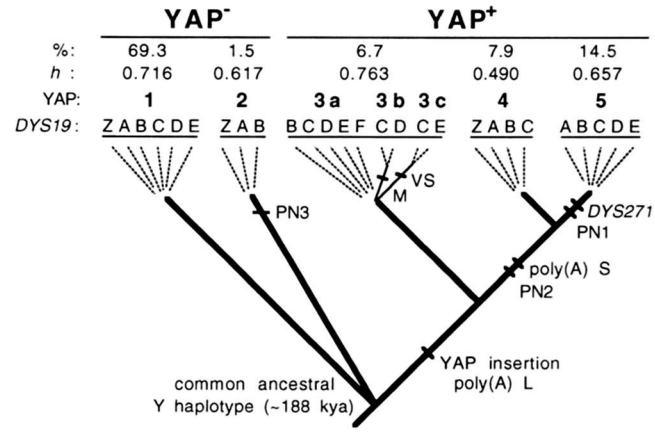

FIGURE 3.—*Y* chromosome haplotype tree based on allelic associations at the YAP, *DYS271*, and *DYS19* loci. Haplotypes carrying the YAP element are referred to as YAP$^+$ and those lacking the YAP element are referred to as YAP$^-$ (the ancestral state). The thick solid lines represent the branching order and relative time of appearance of lineages based on the YAP insertion and single nucleotide mutational changes within the YAP region (HAMMER 1995) and the *DYS271* locus. Thin solid lines refer to lineages created solely by changes in the length of the poly(A) tail associated with the YAP element (M, VS). Mutational changes at the variable nucleotide (PN1–3 and *DYS271*) and poly(A) tail (L, M, S, VS) sites are indicated by short solid bars across branches of the tree. Dotted lines depict lineages with different *DYS19* alleles, and the uncertainty in order and timing of origination of these lineages is illustrated by the use of dotted lines that remain detached from the YAP haplotype lineages. For each YAP haplotype (bold numerals) the following are shown above the tree: frequency (%), diversity (h), and associations with *DYS19* allelic states.

Variation at the *DYS271* locus: The A-G transition at the *DYS271* STS locus was typed in 352 individuals representing all 15 major groups in Table 1. Initially, the *DYS271*-G allele was found to be associated only with YAP haplotype 5; therefore, most of the subsequent typing was performed on haplotypes 4 and 5. All 225 haplotype 5 chromosomes typed had the *DYS271*-G allele. The *DYS271*-A allele was found associated with all haplotype 4 chromosomes ($n = 81$) and all the typed chromosomes with haplotypes 1–3 ($n = 46$). These data suggest that the *DYS271* A-G transition occurred on the lineage leading to YAP haplotype 5, after the occurrence of the PN2 C-T transition (Figure 3). Because we did not find either a haplotype 4 chromosome associated with the *DYS271*-G allele or a haplotype 5 chromosome associated with the *DYS271*-A allele, it was not possible to determine the order of occurrence of the PN1 C-T and *DYS271* mutational events. Although additional surveys may uncover a chromosome with one of these intermediate haplotypes, in the *Y* chromosomes surveyed here the *DSY271* A-G transition did not add a new lineage to the YAP haplotype tree shown in Figure 3.

Variation at the *DYS19* microsatellite locus: Polymorphism at the *DYS19* locus on the short arm of the *Y* chromosome is due to differences in the number of GATA tandem repeats (Roewer *et al.* 1992). Primers designed to unique sequences flanking the repeated region amplify fragments ranging in size from 178 to 206 bp in four-nucleotide increments reflecting eight to 15 copies of the GATA motif (Roewer *et al.* 1992; Ciminelli *et al.* 1995; Santos *et al.* 1996). In previous surveys five alleles (A–E) corresponding to the 186- to 202-bp fragments were found to be common in human populations (Roewer *et al.* 1992; Santos *et al.* 1993; Gomolka *et al.* 1994; Muller *et al.* 1994). Recently, three rare allele classes have been identified: the 178-bp (178-bp allele) (Ciminelli *et al.* 1995), the 182-bp (Z allele) (Santos *et al.* 1996), and the 206-bp (F allele) (Hammer and Horai 1995).

In this survey, we found alleles Z-F at the following worldwide frequencies: 0.7%, 10.0%, 27.1%, 35.5%, 18.1%, 8.4%, and 0.1%, respectively. Although allele profiles for most samples studied were unimodal, the highest frequency allele varied among groups. Similar to previous studies (Gomolka *et al.* 1994; Muller *et al.* 1994; Santos *et al.* 1996), the B allele was found to be the most frequent variant in northern Europe (58.1%) (Table 1). The B allele was also the modal allele in the South European sample (mean frequency = 36.8%); however, when the Italian and Greek samples were broken down into subpopulations there was a great deal of variation among these lower-level units. Some of this *intra*regional variation has been attributed to isolation and genetic drift (Ciminelli *et al.* 1995). In contrast, the C allele was the modal allele in all East and South Asian samples (mean frequencies = 46.3% and 53.5%, respectively) except for the Koreans and Laotians where the D allele was more common. The C allele did, however, predominate in a larger sample of Koreans not reported in this paper (n = 247, C = 38.9%, Wook Kim, personal communication). This predominance of the C allele is similar to the majority of the results from other surveys of Asian populations (Gomolka *et al.* 1994; Muller *et al.* 1994; Hammer and Horai 1995). The C allele was also the modal allele in the Australasian sample (38.2%) and in most African populations (Table 1). The absence of the A allele in the Australasian sample distinguished it from the East and South Asian samples. Consistent with previous studies of Native American populations (Pena *et al.* 1995; Santos *et al.* 1996; Underhill *et al.* 1996), the A allele was found to predominate in the Navajo sample.

YAP haplotype frequencies: The frequencies of the five YAP haplotypes in all 60 populations examined in this study are given in Table 1 and are shown for 24 selected groups in Figure 4 (color). Haplotype 1 (filled-white in the pie chart of Figure 4) was found in all populations and was the most frequent worldwide (mean frequency = 69.2%). Other than a single Western Australian with haplotype 4 (not shown in Figure 4, see below), haplotype 1 was the only YAP haplotype found in the New World, South Asian, and Australasian samples. Furthermore, within East Asia it was the only haplotype found in the Korean, South Chinese, and Taiwanese samples (except for a single Taiwanese individual with haplotype 3). Haplotype 2 (filled-green in Figure 4) was limited to the Khoisan-speaking populations and varied in frequency from 4.5% in the Dama to 41.9% in the Tsumkwe San. Except for the Khoisan, haplotype 3 (filled-red in Figure 4) was found in all African samples (mean African frequency = 6.1%) at frequencies ranging from 3.6% (West Bantus) to 16.1% (West Africans). Haplotype 3 was not found in the European samples (except in a single Sicilian not apparent in Figure 4 due to its low frequency) but was the only YAP^+ haplotype in the East Asian samples (mean frequency = 22.4%). Within East Asia haplotype 3 was found at high frequencies only in the Japanese and Tibetan samples, and in a single Taiwanese male (Table 1). Haplotype 4 (filled-blue in Figure 4) was widespread in Africa (except in the Pygmy samples) and ranged in frequency from 2.0% in East Africa to 46.2% in North Africa (mean African frequency = 12.7%). This haplotype had a distribution pattern opposite to that of hap-

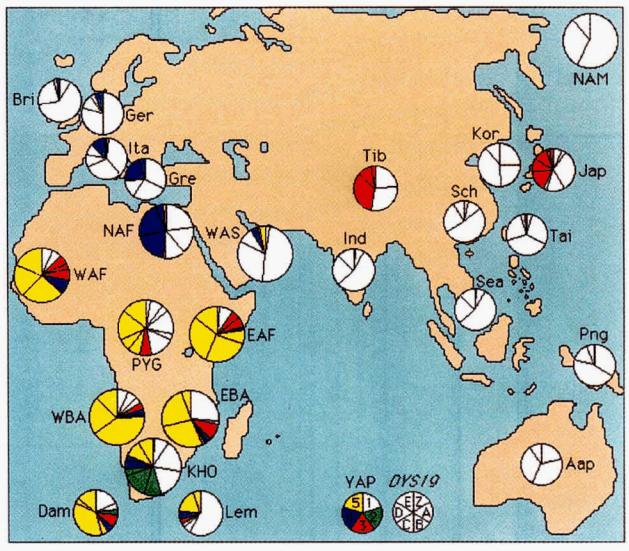

FIGURE 4.—Geographical distribution of *Y* chromosome haplotype frequencies. Nine of the 15 major groups in Table 1 are shown by large circles. The 15 smaller circles are component populations of the other six major groups. The frequencies of the five YAP haplotypes are shown as colored pie charts (see left insert at bottom of figure for color codes). The proportions of the different *DYS19* alleles associated with each YAP haplotype are indicated by solid lines within the colored portions of the pie chart (see right insert for orientation of *DYS19* alleles within each YAP haplotype). When six *DYS19* alleles are present, the order of appearance is as shown in the insert; however, when fewer alleles are present, consult Figure 5A for their identity and order. Three-letter codes not defined in Table 1 include: Dam, Dama; Lem, Lemba; Jap, Japanese; Tib, Tibetans; Kor, Koreans; Tai, Taiwanese; Sch, South Chinese; Ind, Indians; Sea, Southeast Asians; Png, Papua New Guineans; Aap, Australian Aboriginal People; Bri, British; Ger, Germans; Ita, Italians; Gre, Greeks.

lotype 3 outside of Africa: it was present as the major YAP$^+$ haplotype in Europe (13.5%) and was absent in the Asian samples. Haplotype 4 was also present in a single aboriginal male from Western Australia; however, this occurrence is likely due to European paternal admixture. Haplotype 5 (filled-yellow in Figure 4) was limited almost entirely to sub-Saharan Africa where it was the most frequent YAP haplotype (mean frequency = 52.1%). The frequency of haplotype 5 ranged from 17.9% in the Khoisan sample to 76.4% in the West Bantu sample. This haplotype was also found at low frequencies in the West Asian (5.2%) and the North African (2.2%) samples.

YAP/*DYS19* combination haplotype frequencies: The association of the five YAP haplotypes (including poly(A) tail subtypes) with alleles at *DYS19* produced 27 *Y*-chromosome combination haplotypes (Figure 3). Histogram representations of the 27 combination haplotype frequencies are given for the 15 major groups in Figure 5A. YAP$^-$ haplotypes 1C, 1B, and 1D were the most common worldwide with frequencies of 25.9%, 22.9%, and 11.6%, respectively. The most common YAP$^+$ haplotypes in this survey were haplotypes 5C (6.5%), 4A (5.2%), and 5D (5.0%). All other haplotypes were relatively rare with frequencies under 5%. There was a nearly continuous series of *DYS19* allele sizes associated with each YAP haplotype (Figure 3, dotted lines), while *DYS19* alleles C and D, and C and E were associated with haplotypes 3b and 3c, respectively (Figure 3, thin solid lines). Alleles at the variable nucleotide and the poly(A) tail sites appear to have originated a single time on the haplotype tree in Figure 3 (short solid bars). However, several alleles at the *DYS19* locus probably originated independently multiple times. For example, the Z allele, associated with haplotypes 1, 2, and 4, probably arose from the A allele on at least three occasions. A minimum of 21 changes in allele size of one repeat unit (15 parallel changes) are required to account for the distribution of *DYS19* alleles on the haplotype tree. This is consistent with the (single) stepwise-mutation model and high rate of evolution of short tandem repeats (WEBER and WONG 1993).

Despite the convergent pattern of evolution at the *DYS19* locus, there were several examples of strong associations between YAP haplotypes and *DYS19* alleles. In sub-Saharan Africa, haplotype 1 was negatively associated with *DYS19* alleles Z and A ($D' = -0.63$ and -0.79, respectively), haplotype 2 was positively associated with *DYS19*-Z ($D' = 0.88$), and haplotype 4 was positively associated with *DYS19*-A ($D' = 0.71$). Haplotype 3aB was widely distributed in Africa (Pygmies, West Africans, East Bantus, West Bantus, Dama, and Lemba) (Figure 5A). Haplotype 3a was associated with other *DYS19* alleles only in West Africa (C and D) and East Africa (C). In contrast, haplotype 3 in East Asia was never found with the *DYS19* B allele, and haplotype 1 was strongly associated with alleles A and B ($D' = 1.0$). In Europe there were strong associations between haplotype 1 and alleles C ($D' = 0.930$), D ($D' = 1.0$), and E ($D' = 1.0$), as well as between haplotype 4 and allele A ($D' = 0.82$).

Two African groups, the Dama and Lemba, were termed "enigmatic" because of unclear origins and possible extensive admixture. Their combination haplotype distributions (Figure 5A) were indicative of paternal admixture. The Khoi-speaking Dama shared haplotype 2 with the Khoisan and haplotype 3 with non-Khoisan African groups. Moreover, their overall profiles for haplotypes 1 and 5 were more similar to those of the non-Khoisan populations substantiating other biological evidence for the kinship of the Dama with these non-Khoisan groups (SPURDLE and JENKINS 1992). The Venda-speaking Lemba exhibited an extremely high frequency of haplotype 1B (58.8%), which was also found at high frequencies in West Asia (48.3%) and North Africa (22.6%), but was absent in other East Bantu speakers. This pattern may reflect the retention of Semitic *Y* chromosome haplotypes by this East Bantu-speaking population characterized by its rudimentary Semitic culture (SPURDLE and JENKINS 1992, 1996). Haplotype 5 in the Lemba, however, clearly had a sub-Saharan African origin supporting the hypothesis of a significant male contribution from neighboring South African Bantu-speaking populations.

***DYS19* diversity:** The regional differences in patterns of variation at the *DYS19* locus were reflected in different diversity values ($h \pm$ standard error) among the 15 major groups (data not shown). Low diversity values were found in the North European sample due to the overwhelming prevalence of the B allele (0.600 ± 0.005). Other diversity values were 0.620 ± 0.005, 0.644 ± 0.004, 0.695 ± 0.001, 0.715 ± 0.003, 0.718 ± 0.002, and 0.736 ± 0.001 for the West Asian, South Asian, East Asian, North African, Australasian, and South European samples, respectively. The lowest diversity values in this survey were found in the Navajo sample (0.569 ± 0.007) and in the enigmatic southern African group, the Lemba (0.569 ± 0.013).

This study represents the first large-scale survey of variation at the *DYS19* locus in African populations. The *DYS19* allelic diversity was greater in sub-Saharan African populations (0.763 ± 0.001) than in Australasian (0.718 ± 0.002), European (0.712 ± 0.001), or Asian (0.695 ± 0.001) populations (Bonferroni Test, $P < 0.001$). Within Africa, diversity values were 0.631 ± 0.005, 0.708 ± 0.006, 0.719 ± 0.005, and 0.769 ± 0.004 in the East Bantu, West Bantu, East African, and West African samples, respectively.

The most unusual profiles were seen in the Khoisan and Pygmy populations where there were roughly bimodal distributions of alleles of widely different sizes. This pattern of widespread allelic variation was reflected in higher diversity values than those seen in the rest of the world (0.820 ± 0.002 and 0.823 ± 0.006, respectively, Bonferroni Test, $P < 0.001$). The Khoisan were found to have a relatively high frequency of the unusual Z allele. This allele, not observed in worldwide

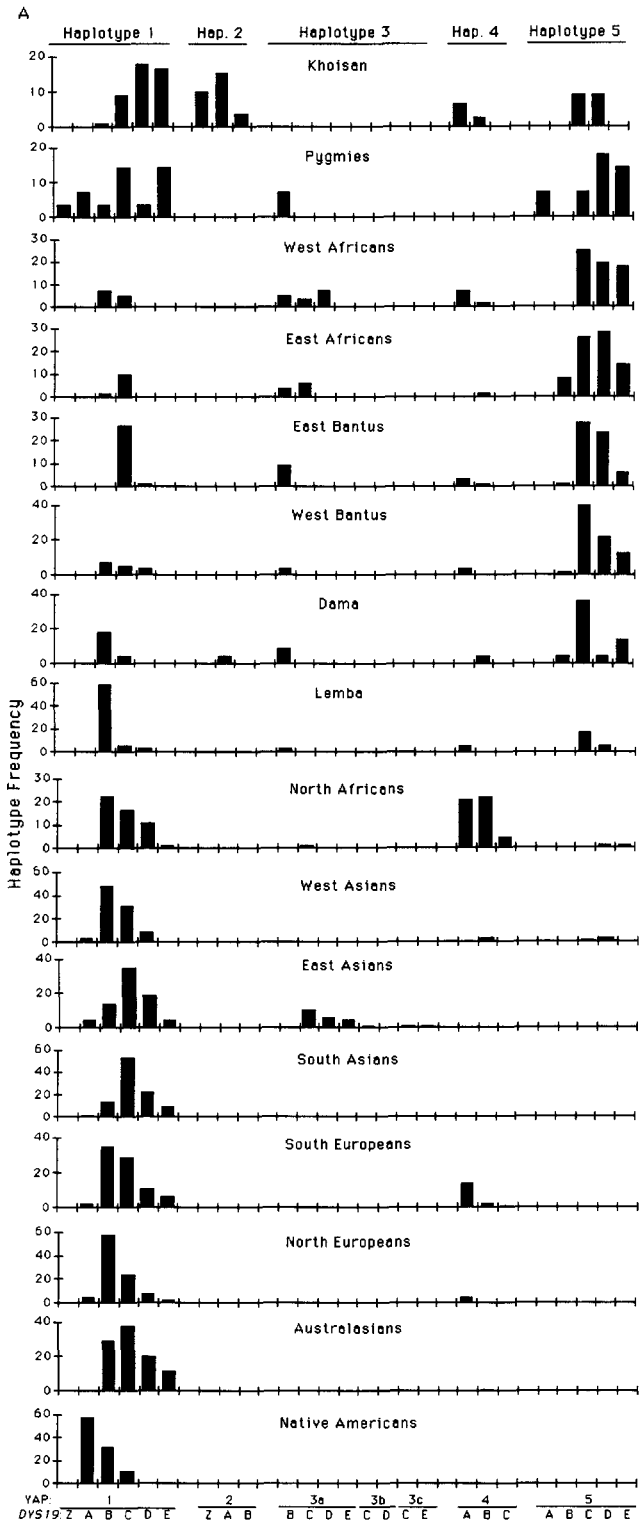

FIGURE 5.—Histogram representations of YAP/*DYS19* combination haplotype frequencies. (A) Twenty-five combination haplotype frequencies for the 15 major groups (the two remaining combination haplotypes are not shown: haplotype 3aF found in a single Japanese male and haplotype 4Z found in a single Egyptian male. (B) Combination haplotype frequencies for the five component populations of East Asia.

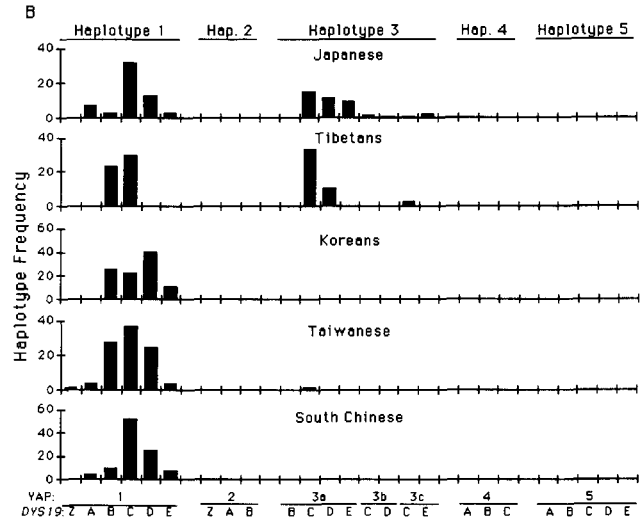

FIGURE 5.—*Continued*

populations studied previously, was recently reported in a single !Kung cell line (SANTOS *et al.* 1996). We identified the Z allele in eight out of 78 (10.3%) Khoisan surveyed, as well as in single Pygmy, Egyptian, and Taiwanese males.

YAP haplotype diversity: The numbers of the five YAP haplotypes found in each major population group ranged from one in South Asia and North America to four in all African populations except the Pygmies (Table 1 and Figure 4 color). Only one population, the possibly admixed enigmatic Dama, was found to contain all five YAP haplotypes. Most non-African populations had just two haplotypes (East Asians, North Europeans, and Australasians), although the South European and West Asian populations had low frequencies of a third haplotype. These patterns of variation in YAP haplotype number and frequency were reflected in significantly greater haplotype diversity in sub-Saharan Africa (0.630 ± 0.001) compared with Asia (0.283 ± 0.001) and Europe (0.239 ± 0.001) (data not shown; Bonferroni Test, $P < 0.001$).

YAP/*DYS19* combination haplotype diversity: In Figure 4 the proportions of the different *DYS19* alleles associated with each YAP haplotype are indicated by solid lines within the colored portions of the pie chart. When considering the 27 combination haplotypes, the number found in each major group ranged from three in the Native American sample to 14 in the East Asian sample (Figure 5A). When the populations making up the East Asian sample were examined separately, the maximum number of haplotypes ranged from four in the Koreans to 13 in the Japanese (Figure 5B). In Africa the number ranged from nine in the Bantu-speaking populations to 11 in the Pygmy, Khoisan, and North African populations. Nine combination haplotypes were found in South Europe, seven in West Asia, six in North Europe, and five in South Asia and Australasia.

Haplotype diversities determined on the basis of the number and relative abundance of each of the 27 combination haplotypes are shown in Figure 6A for the individual populations while those for the major groupings are shown in Figure 6B. For the total sample of

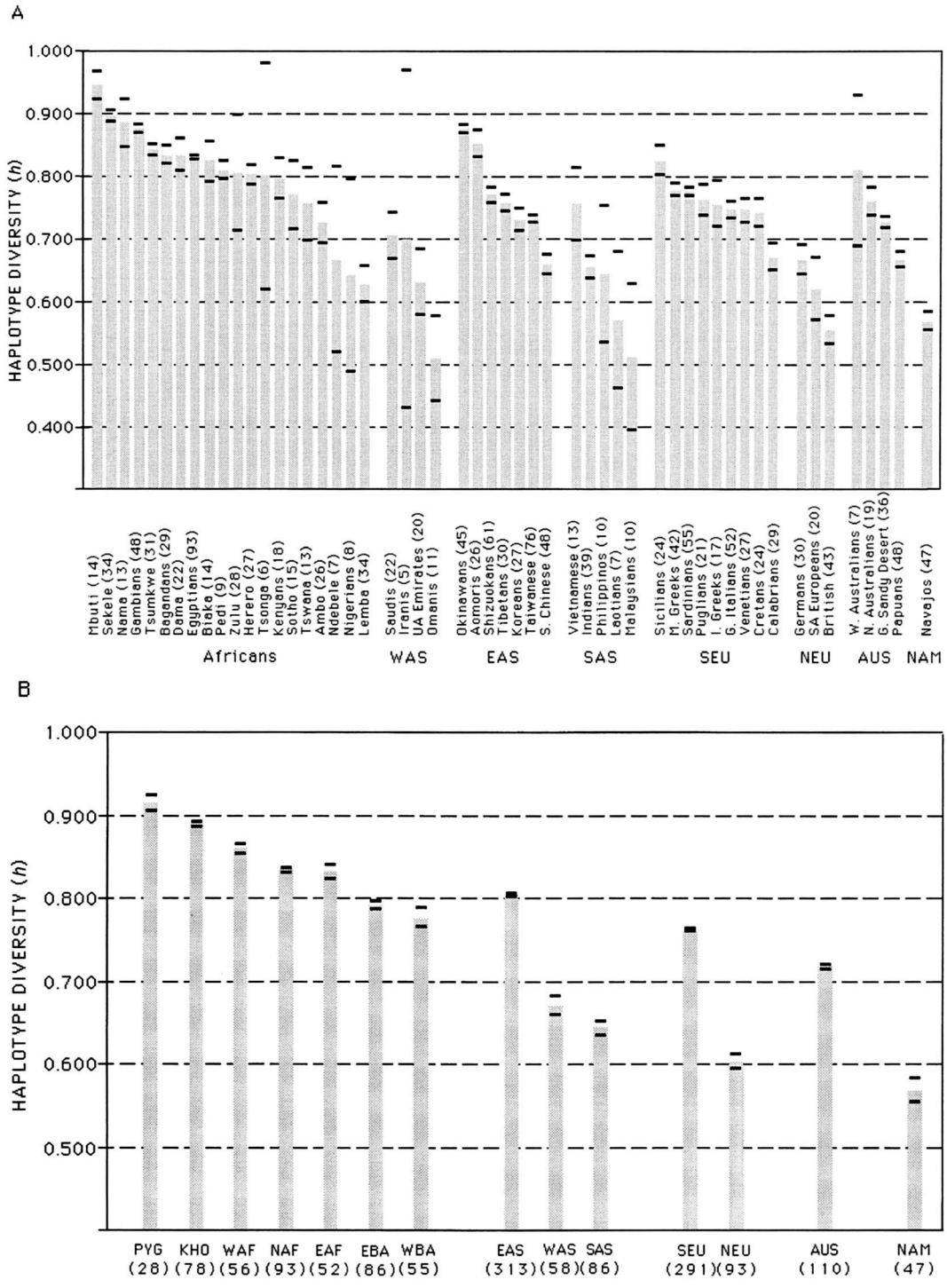

FIGURE 6.—Combination haplotype diversities. Haplotype diversities (h) are given on the vertical (Y) axis and population names, codes, and sample sizes are shown on the horizontal (X) axis. For each population the top of the gray bar represents the sample mean and the two horizontal lines represent the 95% confidence interval. When the standard error is especially small, these symbols occasionally become indistinct. (A) The 53 populations in Table 1 with $n > 5$. (B) The 15 major groups without the enigmatic southern Africans.

1500 individuals the worldwide haplotype diversity equaled 0.852 ± 0.000. The mean combination haplotype diversity was significantly higher for sub-Saharan Africans (0.878 ± 0.000) than for Asians (0.779 ± 0.001), Europeans (0.736 ± 0.001), or Australasians (0.718 ± 0.002) (Bonferroni Test, $P < 0.001$). There were no statistically significant correlations between sample size (Y) and haplotype diversity (X) for the 53 populations listed in Table 1 where $n > 5$ ($r = 0.133$; slope = 23.1; Y-intercept = 11.0) or for the 15 major groupings excluding the enigmatic southern Africans ($r = 0.022$; slope = 17.7; Y-intercept = 89.0). Levels of Y chromosome haplotype diversity ranged from 0.509 ± 0.031 in the Omanis to 0.945 ± 0.010 in the Mbuti Pygmies (Figure 6A). Regions with the highest haplotype diversities ($h > 0.700$) were Africa, East Asia, South Europe, and Australasia (Figure 6B). The New World and North Europe were characterized by low levels of haplotype diversity. South Asia and West Asia had intermediate levels of haplotype diversity. Within Africa the Pygmies and Khoisan had the highest diversity levels and the East Bantus and West Bantus had the lowest.

TABLE 2

Diversity of five YAP haplotypes in six geographic regions

	YAP haplotype				
	1	2	3	4	5
SAF	0.730 ± 0.002 (122)	0.617 ± 0.013 (23)	0.493 ± 0.017 (29)	0.416 ± 0.019 (22)	0.656 ± 0.001 (213)
NAF	0.667 ± 0.005 (47)	—[a]	—[b]	0.614 ± 0.006 (43)	—[b]
ASI	0.639 ± 0.001 (382)	—[a]	0.715 ± 0.004 (70)	—[b]	—[b]
EUR	0.662 ± 0.001 (331)	—[a]	—[b]	0.275 ± 0.011 (52)	—[a]
AUS	0.718 ± 0.002 (109)	—[a]	—[a]	—[b]	—[a]
NAM	0.569 ± 0.007 (47)	—[a]	—[a]	—[a]	—[a]

Values are $h \pm$ SE, with sample size in parentheses. SAF, sub-Saharan Africa; NAF, North Africa; ASI, Asia; EUR, Europe; AUS, Australasia; NAM, North America.

[a] Haplotype not present.

[b] Diversity values not indicated when haplotype is present in fewer than three individuals.

The high levels of haplotype diversity in East Asia were mainly due to the high haplotype diversity of the Japanese (0.835 ± 0.002). This diversity value is equivalent to the North African and East African values. It is interesting that the island of Okinawa had one of the highest haplotype diversities in the world (0.876 ± 0.003), lower only than the values for the Mbuti Pygmies and Sekele San (*t*-tests, $P < 0.0001$). The haplotype diversity of the East Asian sample excluding the Japanese was 0.752 ± 0.001. This value lies between those from South Europe and Australasia.

Regional variation in diversity by YAP haplotype: The genetic diversity carried by each of the five YAP haplotypes was determined based on the number and relative abundance of *DYS19* alleles associated with each haplotype in six global regions (Table 2). Haplotype 1 diversity was greatest in sub-Saharan Africa followed by Australasia, Asia, North Africa, Europe, and North America (all pairwise comparisons were statistically significant at the 0.001 level except the North Africa/Europe comparison where $P > 0.05$). Sub-Saharan African populations were characterized by widely divergent diversity levels (data not shown). These values ranged from 0.083 ± 0.015 in the East Bantus (the lowest of all the 15 major groups in Table 1) to 0.772 ± 0.032 in the West Bantus and 0.833 ± 0.020 in the Pygmies (the two highest values of the 15 major groups). Although haplotype 3 diversity was greater in Asians than in sub-Saharan Africans (Bonferroni Test, $P < 0.001$) (Table 2), the East Asian (0.715 ± 0.004) and West African (0.722 ± 0.032) diversity values did not differ significantly from each other (Bonferroni Test, $P > 0.05$). Haplotype 4, present in three of the six geographic regions, reached its highest diversity value in North Africa. Although haplotype 2 was confined to the Khoisan, its diversity value was relatively high. Likewise, haplotype 5 had a relatively high diversity value but was more widespread within sub-Saharan Africa and only occurred at low frequencies in North Africa and West Asia.

Fst and PHIst analyses: The overall worldwide Fst value based on the 15 major groups in Table 1 (excluding the enigmatic southern Africans) was 0.302 for the YAP data. This Fst estimate is concordant with expectations based on the fourfold lower effective population size of the *Y* chromosome relative to autosomes (HAMMER 1995). Worldwide Fst values for a variety of autosomal genetic data sources as well as for craniometric data sets cluster around 0.1 (RELETHFORD 1995). In contrast, the worldwide PHIst value for the *DYS19* data was only 0.163. This lower than expected figure may be explained by higher mutation and convergence rates for the *DYS19* locus relative to the YAP nucleotide substitutional sites. CIMINELLI *et al.* (1995) found similar results in "Caucasoid" populations and concluded that convergent evolution of *DYS19* alleles in different populations may result in the underestimation of the among-group variance for microsatellite loci.

Population trees: Maximum likelihood, neighbor-joining, Fitch-Margoliash, and UPGMA methods were employed to analyze the frequencies of the 27 YAP/*DYS19* combination haplotypes. Neighbor-joining (NJ), Fitch-Margoliash (FM), and UPGMA analyses were based on genetic distances generated from both the Chord ($4\bar{D}$) and REYNOLDS *et al.*'s (Fst) equations. For heuristic purposes, a majority-rule consensus tree based on the seven possible combinations of tree-building and distance measures (ML, NJ/$4\bar{D}$, NJ/Fst, FM/$4\bar{D}$, FM/Fst, UPGMA/$4\bar{D}$, and UPGMA/Fst) is shown in Figure 7. All seven methods consistently grouped the North African/non-African populations into a single cluster (1), and all sub-Saharan African populations (except the Khoisan)

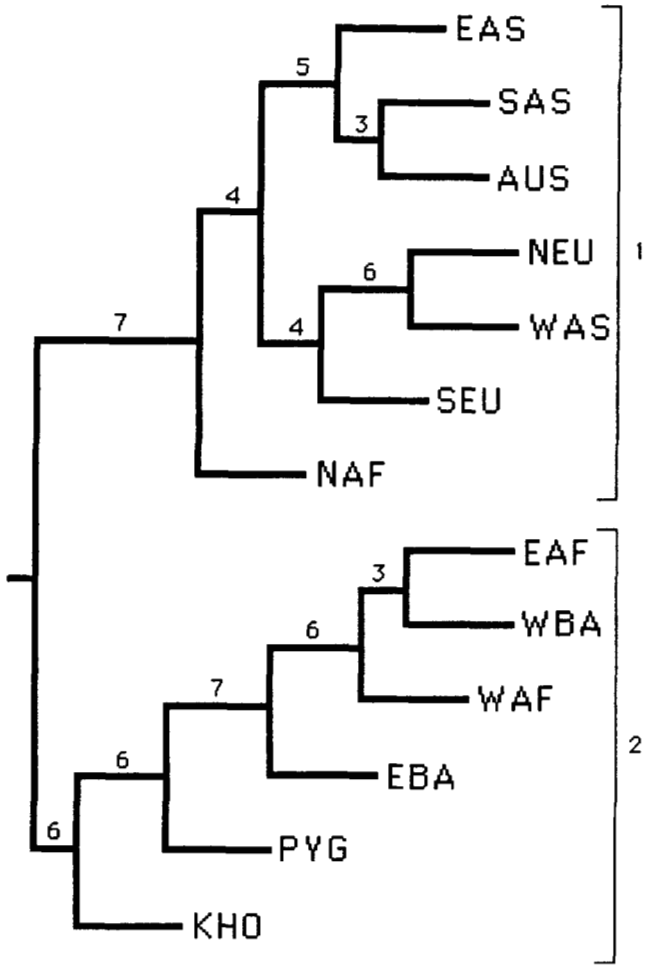

FIGURE 7.—Majority-rule consensus tree for 15 major groups (without enigmatic southern Africans and Native Americans) based on the frequencies for the 27 combination haplotypes. Seven trees were generated using combinations of tree-building and distance measures. The numbers at the nodes indicate the number of methods in which the group consisting of the composite populations to the right of that node occurred. Numbers to the right of the tree represent the two primary clusters.

into another cluster (2). Only the UPGMA/4$\bar{D}$ method placed the Khoisan in cluster 1. When data from the five YAP haplotypes were analyzed separately, the consensus tree topology was similar to that of Figure 7 (tree not shown). The occasional placement of the Khoisan with the rest of the world replicates the results of SPURDLE and JENKINS (1992). These authors postulated small sample size and genetic drift as reasons for the lack of consistent clustering of the Khoisan with the rest of sub-Saharan Africa based on *Y* chromosomal polymorphisms, while CAVALLI-SFORZA *et al.* (1988) pointed to admixture with "Caucasoids" as a possible cause for this phenomenon based on autosomal data.

DISCUSSION

Global patterns of *Y* chromosome variation: The global patterns of human *Y* chromosome haplotype variation presented here have several features in common with other genetic data sets. The Fst value at the YAP locus in the sample of 14 world population groupings of 0.302, equivalent to an Fst of 0.098 for nuclear genes, is similar to estimates from classical genetic polymorphisms (0.09–0.11: NEI and ROYCHOUDHURY 1982), 100 DNA polymorphisms (0.14: BOWCOCK *et al.* 1991), and dinucleotide microsatellites (0.11: DEKA *et al.* 1995). Population trees based on *Y*-chromosome combination haplotypes show that sub-Saharan African populations are clearly differentiated from the rest of the world (Figure 7). This finding is similar to results from classical genetic markers (CAVALLI-SFORZA *et al.* 1988; NEI and ROYCHOUDHURY 1993), mtDNA (CANN *et al.* 1987; VIGILANT *et al.* 1991), and microsatellites (BOWCOCK *et al.* 1994). Furthermore, sub-Saharan African populations exhibit the highest within-group diversity levels for the YAP system, the *DYS19* locus, and the combination haplotype analyses (Figure 6B). Finally, the global analysis of *Y*-chromosome haplotypes shows that paternally inherited variation outside of Africa is a subset of the variation existing within Africa (Figure 4). This parallels results from surveys at other loci (CANN *et al.* 1987; CHEN *et al.* 1995; TISHKOFF *et al.* 1996).

These within-group and among-group patterns of genetic variation have commonly been interpreted as supporting a model in which an (older) African population gave rise to a founder population(s) that carried subsets of genetic diversity to Asia and Europe (RELETHFORD 1995). The higher diversity in Africa is claimed to support this model because it is assumed that greater within-group variation reflects a larger accumulation of mutations in Africa, and therefore greater time depth (STONEKING and CANN 1989). Another critical assumption for this view is that newly formed daughter populations must experience a bottleneck to reduce levels of within-group variation outside of Africa (ROGERS and JORDE 1995). In addition, proponents of Recent Out of Africa models generally argue that the genetic distinctiveness of African populations reflects a simple branching process in which African populations diverged before non-African populations. On the whole, these propositions underscore the difficulty in disentangling the effects of population history on global diversity patterns from influences due to population structure components.

Thus, a straightforward interpretation of genetic diversity patterns may be overly simplistic. As demonstrated by RELETHFORD (1995), the four lines of genetic evidence traditionally used to support Recent Out of Africa models do not necessarily do so. Specifically, neither the relatively low amount of among-group variation within our species, nor the relatively high levels of within-group variation in sub-Saharan Africa, nor the marked genetic divergence of sub-Saharan African populations confirm the recent African origin models entailing a migration of a small population out of Africa (RELETHFORD 1995). As postulated by the version known as the "Weak Garden of Eden" model, the ensu-

ing expansions in effective population size may have occurred much later than this origin (HARPENDING *et al.* 1993). Also, the recent age of our common mitochondrial DNA and *Y* chromosome ancestors may indicate a small effective population size throughout the Late Middle Pleistocene, rather than the time of origin of anatomically modern humans. Even within the context of a recent African origin model, a single emigration from a homogeneous African source population is seen as being too simplistic. For example, as discussed by LAHR and FOLEY (1995), human behavioral, linguistic, morphological, and genetic diversity patterns can also be interpreted in terms of multiple dispersals from a variable population in Africa. These multiple dispersals probably originated from different parts of Africa at different times and left Africa through at least two independent routes.

Numerous explanations are possible for our global *Y*-chromosome haplotype patterns, some of which are not mutually exclusive (TAKAHATA 1994). In addition to the recent African origin model with synchronic or delayed population size expansion and the multiple African dispersal hypothesis mentioned above, other models deserve serious consideration. For instance, scenarios based on the following evolutionary processes are all quite plausible: (1) restricted, recurrent gene flow via isolation by distance, (2) long-distance colonization followed by long-range gene flow, or (3) one or more migrations followed by extensive genetic drift due to founder effects and small population sizes leading to random haplotype extinctions. Although we cannot yet exclude any of these hypotheses based on rigorous inferential statistical tests, we propose that a multiple migration explanation appears to be eminently compatible with our data.

In the aggregate, the above models and scenarios employ a continuum for the amounts of gene flow and genetic drift responsible for global diversity patterns. Our picture of global *Y*-chromosome variation may result from multiple evolutionary processes including both short range (1) and long range (2) migration, as well as genetic drift and extinction (3). The difficulty in distinguishing these evolutionary processes and their effects is underscored by clear-cut examples in which genetic drift can obscure population history (RELETHFORD 1996) or where branching and migration models can produce the same pattern of genetic distances (RELETHFORD 1995). The potentially complicating phenomenon of recurrent mutation involving single nucleotide and Alu insertion polymorphisms is less likely to be involved; nevertheless, it cannot be ignored.

Origins of the YAP haplotypes: We now consider the genetic and geographic connections associated with each of the five YAP haplotypes. Figure 4 shows that all five YAP haplotypes are found in sub-Saharan Africa. Only subsets of these five haplotypes are found outside of Africa. The combination of information about the relative age of each of the major YAP haplotypes with distributional data forms the basis for our hypotheses about migrations and range expansions. The ancestral states for the polymorphic sites in the YAP region allowed HAMMER (1995) to determine the relative ages of four of the five YAP haplotypes: haplotype 1 with its basal position on the haplotype tree represents the ancestral YAP haplotype (~188 kya), and haplotype 5 is the most derived or recent haplotype. The intermediate positions of haplotypes 3 and 4 in Figure 3 reflect their intermediate times of origin. We also received the following approximate absolute dates for the mutational events in Figure 3 from R. C. GRIFFITHS (personal communication) calculated by coalescence-based methods explained in GRIFFITHS and TAVARE (1994): YAP insertional event (~135 kya), PN2 (~90 kya), PN1 (~30 kya), and PN3 (~25 kya). In the context of these absolute ages it is important to recall that the time of origin of polymorphisms is generally thought to precede the dates of population splits and any subsequent expansions (LI and GRAUR 1991; HARPENDING *et al.* 1993; STONEKING 1993). The global distribution of haplotype 1 *vs.* the primarily sub-Saharan African distribution of haplotype 5 are consistent with their postulated relative ages. In accord with their intermediate temporal positions in Figure 3, haplotype 3 is found in both Africa and Asia (north of the Himalayas), while haplotype 4 occurs in Africa and West Asia, as well as throughout Europe. Haplotype 2 may be a private polymorphic haplotype limited to the Khoisan. Although it is not useful for discussing intercontinental migration patterns, it may turn out to be useful for measuring levels of admixture between the Khoisan and neighboring populations (*e.g.*, the Dama).

Because haplotype 1 is the oldest haplotype and is distributed globally, we cannot determine whether this pattern reflects long-term migration drift and/or range expansions. We postulate (but cannot prove) that haplotype 1 traces a genetic trail of the first modern humans out of Africa. A southern Asian expansion from Africa resulting in the occupation of Australia and Papua New Guinea 50–60 kya is thought to precede expansions to Europe and North Asia (NEI and ROYCHOUDHURY 1993; CAVALLI-SFORZA *et al.* 1994; LAHR and FOLEY 1995). Conversely, haplotype 5 is the most derived haplotype in Figure 3 and its distribution may reflect a recent out of Africa event by a southern route very similar to the one just postulated for haplotype 1: going first to West Asia and, according to new data, continuing all the way to India (A. PANDYA, L. SINGH, C. TYLER-SMITH and M. F. HAMMER, unpublished results). The distribution of haplotype 4 exhibits a frequency gradient from North Africa (46.2%) to Northwest Europe (4.3%). The implications of this pattern may be twofold. First, because of its stochastic nature, genetic drift is not expected to produce clinal distributions. Thus, haplotype 4 may provide suggestive evidence arguing against the random haplotype extinction model (3). Second, this cline may chronicle any and/or all of

the following population movements to and through Europe: (1) the Levantine expansion of anatomically modern humans (*i.e.*, ~40 kya) involving the eventual demise of the Neanderthals (STRINGER and GAMBLE 1993), (2) the slightly later pre-agricultural migration (*i.e.*, ~25 kya) detected by mtDNA data (RICHARDS *et al.* 1996), or (3) the even more recent Neolithic expansion (*i.e.*, ~10 kya) associated with the spread of agriculture as postulated by AMMERMAN and CAVALLI-SFORZA's (1984) demic diffusion model.

In each of the preceding cases, an argument could be made for migrations that originated within Africa at different times and places, and that exited Africa via the two routes proposed by LAHR and FOLEY (1995): a southern route through the horn of Africa to West Asia, South Asia, and eventually to Australasia, and a northerly route through North Africa, the Levant, and eventually to Europe. The diversity of each YAP haplotype may also be used to bolster hypotheses regarding the origins and ages of the migrational events discussed above. Haplotype 1 diversity was greatest in sub-Saharan Africa, which can provisionally be interpreted to support an African origin model (Table 2). Australasia exhibits the second highest level of haplotype 1 diversity; thus, an argument can be made that Australasia represents the terminus of an early out of Africa migration. Haplotype 2 was confined to the Khoisan and its high diversity value may reflect either great antiquity in Africa, a wider geographic distribution (CAVALLI-SFORZA *et al.* 1994), and/or a larger effective population size for the ancestral Khoisan. Haplotype 4 reached its highest diversity level in North Africa. Although haplotype 5 occurred at low frequencies in West Asia and North Africa, it attained higher frequencies, greater diversity, and was more widespread in sub-Saharan Africa. Although the present distributions of haplotypes 2, 4, and 5 are compatible with African origin models, the haplotype 3 data are indicative of a possible Asian origin (see below) while the global occurrence of haplotype 1 is consistent with any Old World geographic origin.

African *vs.* Asian origin of YAP haplotype 3: As in the cases of haplotypes 4 and 5, the world distribution of haplotype 3 (*i.e.*, widespread in Africa and present in localized regions outside of Africa) can be interpreted to fit a sub-Saharan African origin model. However, unlike the cases of haplotypes 4 and 5, the diversity data in Table 2 lead to more complicated scenarios. The statistically significant higher level of diversity associated with haplotype 3 in Asia compared with sub-Saharan Africa (Bonferroni Test, $P < 0.001$) may be interpreted in favor of an Asian origin of haplotype 3. Because haplotypes 4 and 5 are derived from haplotype 3 and account for the majority (55.8%) of *Y* chromosomes in Africa, this hypothesis implies that a major component of African diversity is derived from Asia and contradicts the African origin hypothesis. Also, the frequency of haplotype 3 in East Asia (22.4%) is three times higher than its average frequency (7.1%) in sub-Saharan Africa. One of the consequences of a putative Asian origin for haplotype 3 is that all YAP^+ chromosomes (Figure 3, see YAP insertion leading to haplotypes 3–5) also must have an Asian origin *contra* HAMMER (1994). New data from a different locus on the *Y* chromosome in which the Asian form of haplotype 3 is shown to be ancestral to the African form of haplotype 3 also support this Asian origin hypothesis (T. ALTHEIDE and M. F. HAMMER, unpublished data).

An alternative to an Asian origin of haplotype 3 is an African origin hypothesis coupled with a subsequent reduction in the diversity associated with haplotype 3 in Africa. This hypothesis is supported by the finding of equivalent levels of haplotype 3 diversity in West Africa and East Asia, and much lower levels of diversity associated with haplotype 3 in other surveyed regions of sub-Saharan Africa (Figure 5A). The pattern of higher diversity in West Africa compared with the Bantu-speaking populations in southern and eastern Africa (Figure 6B) may be due to founder effects associated with the Bantu expansion (CAVALLI-SFORZA *et al.* 1994). The lack of close clustering of the East and West Bantus as well as the inconsistent placement of the West Bantus in Figure 7 could also be due to the effects of genetic drift.

YAP haplotype 3 and East Asian population affinities: We initially grouped the Japanese, Tibetans, Taiwanese, Koreans, and South Chinese into a single major East Asian group based on geographic and ethnohistorical criteria. The *Y* chromosome data presented here provide evidence for significant differences among populations within this grouping. For example, the highest world frequencies of YAP haplotype 3 are found in Japan (41.7%) and Tibet (46.7%); whereas, the frequency of this haplotype is <2% in all other East Asian populations (Table 1). Also, when the frequencies of the combination haplotypes are viewed separately for these five populations (Figure 5B), it can be seen that the Japanese and Tibetans share several haplotypes that are not present in the other groups (except for a single Taiwanese with haplotype 3aC). The distinctiveness of the Japanese and Tibetans with respect to the other East Asian groups is also apparent in dendrograms based on the 27 combination haplotypes (data not shown). The similarity between the Tibetan and Japanese populations makes sense in the context of other studies showing a northern Asian (*i.e.*, Mongolian) origin for Tibetans (NEI and ROYCHOUDHURY 1993; TORRONI *et al.* 1994b).

CAVALLI-SFORZA *et al.* (1994) suggest that populations between the Yangtse and Yellow Rivers in China form the dividing line between Northeast Asians and Southeast Asians. The Koreans (who are situated north of the Yangtse River) are generally considered to be a North Asian group (NEI and ROYCHOUDHURY 1993; CAVALLI-SFORZA *et al.* 1994). This would lead us to expect that the Koreans would be more similar to the Japanese and Tibetans than to the South Chinese and Taiwanese in *Y* chromosome haplotype frequencies. Also, the Yayoi

migration entered Japan through the Korean peninsula 2300 years ago (HAMMER and HORAI 1995). However, both the haplotype 3 and combination haplotype data place the Koreans with the South Chinese and Taiwanese. One explanation for this unexpected affinity is that a recent range expansion from South China has resulted in the replacement of haplotype 3 by haplotype 1 in Korea.

Global patterns of *Y* chromosome diversity: The highest values for *Y* chromosome combination haplotype diversity are generally found in sub-Saharan Africa (Figure 6B). This pattern is similar to results from other kinds of genetic data as mentioned above and can be attributed to great population antiquity and/or large effective population size in sub-Saharan Africa. Extensive admixture can also lead to high diversity levels as in the case of the North African sample that comes from a fairly wide area around Cairo. Urban populations from North African countries are thought to be admixed with contributions from Berbers and Bedouins, as well as from Mediterranean and sub-Saharan African populations (CAVALLI-SFORZA *et al.* 1994). High levels of diversity outside of Africa are concentrated in East Asia, South Europe, and Australasia. Although a claim can be made that the Australasian diversity levels reflect remnants of an old population system, the South Europe diversity levels are more likely a product of extensive admixture as exemplified by the Sicilians who have the highest diversity value of all European populations (Figure 6A).

The case of East Asia is even more complex. For instance, the high Japanese diversity levels may be due to a combination of an ancient population (Jomon) subjected to recent admixture (Yayoi) as discussed in HAMMER and HORAI (1995). This mixing effect is further illustrated by the finding of higher diversity levels in the extreme southern and northern regions of the Japanese archipelago, where remnants of the ancient Jomon gene pool are thought to survive. Because diversity levels may not have reached equilibrium in populations that have undergone large expansions in size within the last 10,000 years, some rather large populations may exhibit lower than expected within-group variation (ROGERS and JORDE 1995). Thus, the unexpectedly low diversity level seen in the South Chinese sample may be a product of a homogenizing influence associated with the agriculturally based Holocene population expansions originating in Southeast Asia (CAVALLI-SFORZA *et al.* 1994). This trend toward lower diversity also includes Southeast Asia and India.

The lowest diversity value for any major region in the Old World is found in North Europe (Figure 6B), and within that region, the British sample surprisingly has the third lowest diversity value of the 53 populations in Figure 6A. Both the low diversity values and the clinal distribution of haplotype 4 are concordant with AMMERMAN and CAVALLI-SFORZA's (1984) demic diffusion model involving the relatively recent Neolithic expansion associated with the spread of agriculture across Europe toward the British Isles. Other examples of extremely low diversity in Figure 6A probably have different explanations [*i.e.*, the Omanis and Malaysians are characterized by especially small sample sizes of 11 and 10, respectively, while suspected recent founder events/population fissions could explain the Navajo (GREENBERG *et al.* 1986) and Lemba values]. Lemba males are thought to have a Semitic origin (SPURDLE and JENKINS 1996) while mtDNA analyses point to African ancestry for Lemba females (SOODYALL *et al.* 1996). Overall, our data imply a small founding population of males from the Middle East consistent with SPURDLE and JENKINS' (1996) conclusions.

Figure 6, A and B, are based upon combination haplotypes; however, the components of these combination haplotypes differ in their tempo and mode of evolution. The YAP sites represent rare single mutational events and fit the infinite sites model (LI 1977). In contrast, the *DYS19* microsatellite system fits the stepwise mutation model, evolves rapidly, and is subject to widespread convergence leading to homoplasy (SHRIVER *et al.* 1993; ESTOUP *et al.* 1995; GOLDSTEIN *et al.* 1995). The legitimacy of combining these two different types of genetic data is open to question. As a result we undertook diversity analyses of each system separately (data not shown). As in the combination haplotype analysis, sub-Saharan Africa had the highest diversity levels in both the YAP and *DYS19* individual analyses. Then the YAP diversity values (Y) were plotted against *DYS19* diversity values (X) for 14 major groups in Table 1 (the enigmatic southern Africans were omitted). A statistically significant correlation ($r = 0.697$, $P < 0.01$) was obtained between these two sets of diversity values. Interestingly, the three most obvious outliers were the Australasians, South Europeans, and East Bantus. The first two groups are characterized by high *DYS19* diversity values with respect to YAP values and it is this microsatellite diversity that causes their relatively high combination haplotype diversity values in Figure 6B. On the other hand, the East Bantus exhibit relatively low *DYS19* diversity resulting in a lower combination value. A different situation obtains for the low West Bantu combination value that reflects concordantly low YAP and *DYS19* values. In fact, the two Bantu groups exhibit the lowest levels of *Y*-chromosome diversity within Africa consistent with the hypothesis of a comparatively recent expansion of Bantu-speaking populations (EXCOFFIER *et al.* 1987; CAVALLI-SFORZA *et al.* 1994). Also, because these results conflict with those derived from mtDNA and autosomal data from some of the same populations, SPURDLE and JENKINS (1993) hypothesized that *Y*-specific diversity values may have been reduced due to the widespread practice of polygamy in some African populations (HAMMER and ZEGURA 1996).

Theoretical considerations: At this juncture one should recall RELETHFORD's (1995) important caveat that among-group patterns of variation, rather than re-

flecting the branching history and age of human populations, can be explained by alternative demographic parameters, especially effective population size. Because of the effects of differing population sizes on the magnitude of genetic drift, often one cannot distinguish between patterns that reflect population structure and those contingent on population history (RELETHFORD 1996; RELETHFORD and HARPENDING 1994), nor are interpretations of within-group patterns of genetic diversity straightforward. For example, the finding of greater within-group variation in sub-Saharan Africa has been used to claim that sub-Saharan African populations are older (CANN *et al.* 1987; VIGILANT *et al.* 1991; TISHKOFF *et al.* 1996). As pointed out by several authors, divergence time is not the only factor that can affect within-group variation (EXCOFFIER and LANGANEY 1989; HARPENDING *et al.* 1993; STONEKING 1993; TEMPLETON 1993; RELETHFORD and HARPENDING 1994; MARTINSON *et al.* 1995; RELETHFORD 1995; ROGERS and JORDE 1995; AOKI and SHIDA 1996). Demographic considerations that influence levels of within-population variation include population bottlenecks and size expansions, migration patterns and reticulation, and degree of population subdivision. Other factors that could specifically influence *Y*-chromosome variation include selective sweeps and genetic hitchhiking (MAYNARD-SMITH 1990), background selection (CHARLESWORTH *et al.* 1993), and high variance in male reproductive success (LUCOTTE *et al.* 1994; PENA *et al.* 1995). An additional complication is the admonition that one cannot use standard measures of within-group population variation (*i.e.*, diversity values) to test geographic origin hypotheses because the effects of mutation rates, inbreeding effective population size, and population age are all conflated with past and present gene flow in the diversity values, themselves (TEMPLETON 1993). The difficulty in empirical studies is determining which combination of the above factors is responsible for differences in the levels of within- and among-group variation. This dilemma can sometimes be addressed by supplemental demographic and ethnohistorical data (RELETHFORD 1996), as well by innovative population genetic analyses such as those used for the comparison of mtDNA mismatch and intermatch distributions that often demonstrate a temporal lag between population origin and expansion/bottleneck (HARPENDING *et al.* 1993; ROGERS and JORDE 1995).

Methodological considerations: The most promising analytical technique for trying to resolve some of these theoretical issues is that of TEMPLETON (1993). His method that attempts to disentangle the effects of population structure and population history on genetic variation is based on a nested cladistic analysis of the geographic distribution of mtDNA haplotype data. The null hypothesis is that there is no geographic association of clades. This hypothesis is tested using a random permutation procedure. If the null hypothesis is rejected, TEMPLETON (1993) seeks to clarify the causes for these nonrandom associations between haplotypes and geography. His two principal explanatory categories are population structure and population history. In this context, population structure includes restricted, recurrent gene flow via isolation by distance, while range expansions reflect population history. The first human data set analyzed using TEMPLETON's (1993) approach was EXCOFFIER's (1990) mtDNA data from 12 Old World populations ($n = 1218$). This analysis did not lead to a dichotomous separation between explanations derived from population history *vs.* population structure considerations. Within higher-order clades, restricted and recurrent gene flow was the principal explanation for the data patterns. At the lowest level in the hierarchy, the data led to the detection of two regional (but no global) range expansions. One of these expansion events involved Arabs and sub-Saharan Africans, while the other involved an expansion across Europe. As TEMPLETON (1993) pointed out, the current distribution of mtDNA variation that he analyzed was influenced both by restricted gene flow and by population expansion. The critical point is his explicit statement that his analysis does not treat population expansion or restricted gene flow as mutually exclusive alternatives. Indeed, TEMPLETON *et al.* (1995) recently found evidence for restricted gene flow via isolation by distance as well as for range expansions in mtDNA data from the tiger salamander, *Ambystoma tigrinum*. From these examples, it is clear that even portions of the same data set can show the differential import of factors associated with population history and population structure.

The graphical display of the geographic distribution of our *Y* chromosome data in Figure 4 is similar to methods used in intraspecific phylogeography (AVISE 1994) and especially in contemporary human mtDNA research (MELTON *et al.* 1995). TEMPLETON et al. (1995, p. 779) provide a cogent caveat about such displays when they state: "Such pictorial representations are an excellent exploratory tool for formulating hypotheses, but are inappropriate inference tools." Our paper is about hypothesis compatibility *sensu* STONEKING (1994) rather than about hypothesis testing *sensu* TEMPLETON (1994). We employed exploratory graphical analyses to generate hypotheses to explain the geographic distribution of our *Y* chromosome data. TEMPLETON (1993) was fortunate to have 76 human mtDNA haplotypes distributed over 18 geographic locations, while TEMPLETON *et al.* (1995) used 23 mtDNA haplotypes distributed over 53 geographic regions in the salamander analysis. We, however, only have five human *Y* chromosome haplotypes distributed over 60 geographic locations. It is doubtful that this represents enough genetic variation to carry out a nested cladistic haplotype test (A. R. TEMPLETON, personal communication).

Conclusions: An important feature of our results is the number of polymorphisms that appear to have originated in Africa (*i.e.*, PN1, PN2, PN3, *DYS271* A–G, and the poly(A) S allele). This may reflect a greater African population size and/or an African ascertainment bias

in the polymorphic sites surveyed. On the other hand, the YAP element insertion and the poly(A) M and VS alleles may have originated in Asia. Overall, our combination haplotype data are compatible with either an African origin, an extremely heterogeneous source population in Africa, the aforementioned larger effective African population size, and/or various gene flow and genetic drift-based scenarios. Our contribution differs from previous genetic studies that have supported a model based on a single migration out of Africa (CANN *et al.* 1987; VIGILANT *et al.* 1991; TISHKOFF *et al.* 1996). For instance, in our opinion the TISHKOFF *et al.* (1996) microsatellite data may have traced a more recent migration as exemplified by haplotypes 4 and 5 in our *Y* chromosome tree (Figure 3). Separate analyses of the patterns of genetic variation associated with each of the five YAP haplotypes point to the conjecture that our data contain evidence consistent with multiple dispersals, some emanating from Africa and one perhaps coming from Asia. It is also possible that our microsatellite data reflect comparatively recent demographic parameters and microevolutionary events, while the rest of our data reflect much earlier population dynamics.

Although we have treated our polymorphic sites as different loci, in actuality the nonrecombining portion of the human *Y* chromosome represents a single locus. The same stricture holds for mtDNA. Although more information can be potentially obtained from multilocus data, different questions can be addressed utilizing single locus haploid systems. Differences between maternal and paternal histories may be more informative for disentangling multiple migration events that have left palimpsest-like genetic and morphological patterns. In addition, the ability to construct haplotype trees with determinable ancestral and derived states and to estimate a time frame for the branching events in the trees provides a heuristic framework for the future testing of hypotheses about modern human origins. Additional *Y* chromosome data from a greater number of polymorphic sites and human populations will be necessary to perform model-based tests of these hypotheses (TEMPLETON 1993, 1996, personal communication).

We thank JOHN CONAWAY, JAMES TUGGLE, JARED RAGLAND, JENNIFER VUTURO-BRADY, AI-WEI HO, and CHARLES HOEFFER for excellent technical assistance. Special thanks go to one of the anonymous reviewers for his/her comments that were both very thoughtful and helpful. Also, we thank LESLIE LOUIE and CHRIS TYLER-SMITH for contributing DNA samples. This work was supported by National Science Foundation grant 0PP-9423429 and National Institutes of Health grant GM-53566 to M.F.H.

LITERATURE CITED

AMMERMAN, A. J., and L. L. CAVALLI-SFORZA, 1984 *Neolithic Transition and the Genetics of Populations in Europe.* Princeton University Press, Princeton.

AOKI, K., and M. SHIDA, 1996 A Monte Carlo simulation study of coalescence times in a successive colonization model with migration, pp. 66–71 in *Prehistoric Mongoloid Dispersals,* edited by T. AKAZAWA and E. J. E. SZATHMARY. Oxford University Press, Oxford.

AVISE, J. C., 1994 *Molecular Markers, Natural History and Evolution.* Chapman & Hall, New York.

BATZER, M. A., M. STONEKING, M. ALEGRIA-HARTMAN, H. BAZAN, D. H. KASS *et al.,* 1994 African origin of human-specific polymorphic *Alu* insertions. Proc. Natl. Acad. Sci. USA **91:** 12288–12292.

BOWCOCK, A. M., J. R. KIDD, J. L. MOUNTAIN, J. M. HERBERT, L. CAROTENUTO *et al.,* 1991 Drift, admixture, and selection in human evolution: a study with DNA polymorphisms. Proc. Natl. Acad. Sci. USA **88:** 839–843.

BOWCOCK, A. M., A. RUIZ-LINARES, J. TOMFOHRDE, E. MINCH, J. R. KIDD *et al.,* 1994 High resolution of human evolutionary trees with polymorphic microsatellites. Nature **368:** 455–457.

CANN, R. L., M. STONEKING and A. C. WILSON, 1987 Mitochondrial DNA and human evolution. Nature **325:** 31–36.

CASANOVA M., P. LEROY, C. BOUCEKKINE, J. WEISSENBACH, C. BISHOP *et al.,* 1985 A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance. Science **230:** 1403–1406.

CAVALLI-SFORZA, L. L., and W. F. BODMER, 1971 *The Genetics of Human Populations.* W. H. Freeman, San Francisco.

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.

CAVALLI-SFORZA, L. L., P. MENOZZI and A. PIAZZA, 1994 *The History and Geography of Human Genes.* Princeton University Press, Princeton.

CHARLESWORTH, B., M. T. MORGAN and D. CHARLESWORTH, 1993 The effect of deleterious mutations on neutral molecular variation. Genetics **134:** 1289–1303.

CHEN, Y.-S., A. TORRONI, L. EXCOFFIER, A. S. SANTACHIARA-BENERECETTI and D. C. WALLACE, 1995 Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. Am. J. Hum. Genet. **57:** 133–149.

CIMINELLI, B., F. POMPEI, P. MALASPINA, M. F. HAMMER, F. PERSICHETTI *et al.,* 1995 Recurrent simple tandem repeat mutations during human Y chromosome radiation in Caucasian subpopulations. J. Mol. Evol. **41:** 966–976.

DEKA, R., L. JIN, M. D. SHRIVER, L. M. YU, S. DECROO *et al.,* 1995 Population genetics of dinucleotide $(dC\text{-}dA)_n\text{-}(dG\text{-}dT)_n$ polymorphisms in world populations. Am. J. Hum. Genet. **56:** 461–474.

ESTOUP, A., L. GARNERY, M. SOLIGNAC and J.-M. CORNUET, 1995 Microsatellite variation in Honey Bee (*Apis Mellifera L.*) populations: hierarchical genetic structure and test of the infinite allele and stepwise mutation models. Genetics **140:** 679–695.

EXCOFFIER, L., 1990 Evolution of human mitochondrial DNA: evidence for departure from a pure neutral model of populations at equilibrium. J. Mol. Evol. **30:** 125–139.

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

EXCOFFIER, L., B. PELLEGRINI, A. SANCHEZ-MAZAS, C. SIMON and A. LANGANEY, 1987 Genetics and history of sub-Saharan Africa. Yrbk. Phys. Anthropol. **30:** 151–194.

FELSENSTEIN, J., 1981 Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. **17:** 368–376.

FELSENSTEIN, J., 1993 PHYLIP: Phylogeny Inference Package, Version 3.5p. Joseph Felsenstein and the University of Washington, Seattle.

FITCH, W. M., and E. MARGOLIASH, 1967 Construction of phylogenetic trees. Science **155:** 279–284.

FRAYER, D. W., M. H. WOLPOFF, A. G. THORNE, F. H. SMITH and G. G. POPE, 1993 Theories of modern human origins: the paleontological test. Am. Anthropol. **95:** 14–50.

FU, Y.-X., and W.-H. LI, 1996 Estimating the age of the common ancestor of men from the *ZFY* intron. Science **272:** 1356–1357.

GOLDSTEIN, D. B., A. R. LINARES, L. L. CAVALLI-SFORZA and M. W. FELDMAN, 1995 An evaluation of genetic distances for use with microsatellite loci. Genetics **139:** 463–471.

GOMOLKA, M., J. HUNDRIESER, P. NURNBERG, L. ROEWER, J. T. EPPLEN *et al.,* 1994 Selected di- and tetranucleotide microsatellites from chromosomes 7, 12, 14, and Y in various Eurasian populations. Hum. Genet. **93:** 592–596.

GREENBERG, J. H., C. G. TURNER and S. L. ZEGURA, 1986 The settlement of the Americas: a comparison of the linguistic, dental, and genetic evidence. Current Anthropol. **27:** 477–497.

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

HAMMER, M. F., 1994 A recent insertion of an Alu element on the Y chromosome is a useful marker for human population studies. Mol. Biol. Evol. **11:** 749–761.

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., and S. L. ZEGURA, 1997 The role of the Y chromosome in human evolutionary studies. Evol. Anthropol. (*in press*).

HARPENDING, H., and A. ROGERS, 1984 *Antana: A Package for Multivariate Data Analysis.* Harpending & Rogers, Bosque Farms.

HARPENDING, H. C., S. T. SHERRY, A. R. ROGERS and M. STONEKING, 1993 The genetic structure of ancient human populations. Curr. Anthropol. **34:** 483–496.

IKUTA, S., K. TAKAGI, R. B. WALLACE and K. ITAKURA, 1987 Dissociation kinetics of 19 base paired oligonucleotide-DNA duplexes containing different single mismatched base pairs. Nucleic Acids Res. **15:** 797-811.

JOBLING, M. A., and C. TYLER-SMITH, 1995 Fathers and sons: the Y chromosome and human evolution. Trends Genet. **11:** 449–456.

KARAFET, T., S. L. ZEGURA, J. VUTURO-BRADY, O. POSUKH, L. OSIPOVA *et al.*, 1997 Y chromosome markers and trans-Bering Strait dispersals. Am. J. Phys. Anthropol. (*in press*).

LAHIRI, D. K., and J. I. NURNBERGER, 1991 A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. Nucleic Acids Res. **19:** 5444.

LAHR, M. M., and R. FOLEY, 1995 Multiple dispersals and modern human origins. Evol. Anthropol. **3:** 48–60.

LI, W.-H., 1977 Distribution of nucleotide differences between two randomly chosen cistrons in a finite population. Genetics **85:** 331–337.

LI, W.-L., and D. GRAUR, 1991 *Fundamentals of Molecular Evolution.* Sinauer Associates, Inc., Sunderland, MA.

LUCOTTE, G., and N. Y. NGO, 1985 p49f, a highly polymorphic probe, that detects *Taq1* RFLPs on the human Y chromosome. Nucleic Acids Res. **13:** 82–85.

LUCOTTE, G., N. GERARD, R. KRISHNAMOORTHY, F. DAVID, A. AOUIZERATE *et al.*, 1994 Reduced variability in Y-chromosome-specific haplotypes for some Central African populations. Hum. Biol. **66:** 519–526.

MARTINSON, J. J., L. EXCOFFIER, C. SWINBURN, A. J. BOYCE, R. M. HARDING *et al.*, 1995 High diversity of α-globin haplotypes in a Senegalese population, including many previously unreported variants. Am. J. Hum. Genet. **57:** 1186–1198.

MATHIAS, N., M. BAYES and C. TYLER-SMITH, 1994 Highly informative compound haplotypes for the human Y chromosome. Hum. Mol. Genet. **3:** 115–123.

MAYNARD-SMITH, J., 1990 The Y of human relationships. Nature **344:** 591–592.

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

MERRIWETHER, D. A., F. ROTHHAMMER and R. E. FERRELL, 1995 Distribution of the four founding lineage haplotypes in Native Americans suggests a single wave of migration for the New World. Am. J. Phys. Anthropol. **98:** 411–430.

MICHALAKIS, Y., and L. EXCOFFIER, 1996 A generic estimation of population subdivision using distances between alleles with special references for microsatellite loci. Genetics **142:** 1061–1064.

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

MITCHELL, R. J., L. EARL and J. W. WILLIAMS, 1993 Two Y-chromosome-specific restriction fragment length polymorphisms (*DYS11* and *DYZ8*) in Italian and Greek migrants to Australia. Hum. Biol. **65:** 387–399.

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.

MULLER, S., M. GOMOLKA and H. WALTER, 1994 The Y-specific SSLP of the locus *DYS19* in four different European samples. Hum. Hered. **44:** 298–300.

NEI, M., 1978 Estimation of average heterozygosity and genetic distance from small number of individuals. Genetics **89:** 583–590.

NEI, M., 1987 *Molecular Evolutionary Genetics.* Columbia University Press, New York.

NEI, M., and A. K. ROYCHOUDHURY, 1982 Genetic relationships and evolution of human races. Evol. Biol. **14:** 1–59.

NEI, M., and A. K. ROYCHOUDHURY, 1993 Evolutionary relationships of human populations on a global scale. Mol. Biol. Evol. **10:** 927–943.

NYSTUEN, A., T. BUSINGA, R. CARMI, G. M. DUYK and V. C. SHEFFIELD, 1995 Identification of polymorphic markers for the Y chromosome and the use of these markers to determine the number of male founders in isolated populations. Am. J. Hum. Genet. **57** (Suppl.): A26.

PENA, S. D. J., F. R. SANTOS, N. O. BIANCHI, C. M. BRAVI, F. R. CARNESE *et al.*, 1995 A major founder Y-chromosome haplotype in Amerindians. Nature Genet. **11:** 15–16.

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.

PERNA, N. T., M. A. BATZER, P. L. DEININGER and M. STONEKING, 1992 Alu insertion polymorphism: a new type of marker for human population studies. Hum. Biol. **64:** 641–648.

PERSICHETTI, F., P. BLASI, M. HAMMER, P. MALASPINA, C. IODICE *et al.*, 1992 Disequilibrium of multiple markers on the human Y chromosome. Ann. Hum. Genet. **56:** 303–310.

RELETHFORD, J. H., 1995 Genetics and modern human origins. Evol. Anthropol. **4:** 53–63.

RELETHFORD, J. H., 1996 Genetic drift can obscure population history: problem and solution. Hum. Biol. **68:** 29–44.

RELETHFORD, J. H., and H. C. HARPENDING, 1994 Craniometric variation, genetic theory, and modern human origins. Am. J. Phys. Anthropol. **95:** 249–270.

RELETHFORD, J. H., and H. C. HARPENDING, 1995 Ancient differences in population size can mimic a recent African origin of modern humans. Curr. Anthropol. **36:** 667–674.

REYNOLDS, J., B. S. WEIR and C. C. COCKERHAM, 1983 Estimation of coancestry coefficient: Basis for short-term genetic distance. Genetics **105:** 767–779.

RICHARDS, B., J. SKOLETSKY, A. P. SHUBER, R. BALFOUR, R. C. STERN *et al.*, 1993 Multiplex PCR amplification from the *CFTR* gene using DNA prepared from buccal brushes/swabs. Hum. Mol. Genet. **2:** 159–163.

RICHARDS, M., H. CORTE-REAL, P. FORSTER, V. MACAULAY, H. WILKINSON-HERBOTS *et al.*, 1996 Paleolithic and Neolithic lineages in the European mitochondrial gene pool. Am. J. Hum. Genet. **59:** 185–203.

ROEWER, L., J. ARNEMANN, N. K. SPURR, K.-H. GRZESCHIK and J. T. EPPLEN, 1992 Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts. Hum. Genet. **89:** 389–394.

ROEWER, L., M. KAYSER, P. DIELTJES, M. NAGY, E. BAKKER *et al.*, 1996 Analysis of molecular variance (AMOVA) of Y-chromosome-specific microsatellites in two closely related human populations. Hum. Mol. Genet. **5:** 1029–1033.

ROGERS, A. R., and L. B. JORDE, 1995 Genetic evidence on modern human origins. Hum. Biol. **67:** 1–36.

SAITOU, N., and M. NEI, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425.

SALEM, A.-H., F. M. BADR, M. F. GABALLAH and S. PAABO, 1996 The genetics of traditional living: Y chromosomal and mitochondrial lineages in the Sinai Peninsula. Am. J. Hum. Genet. **59:** 741–743.

SANTOS, F. R., S. D. J. PENA and J. T. EPPLEN, 1993 Genetic and population study of a Y-linked tetranucleotide repeat DNA polymorphism with a simple non-isotopic technique. Hum. Genet. **90:** 655–656.

SANTOS, F. R., T. GERELSAIKHAN, B. MUNKHTUJA, T. OYUNSUREN, J. E. EPPLEN *et al.*, 1996 Geographic differences in the allele frequencies of the human Y-linked tetranucleotide polymorphism *DYS19.* Hum. Genet. **97:** 309–313.

SEIELSTAD, M. T., J. M. HEBERT, A. A. LIN, P. A. UNDERHILL, M. IBRAHIM *et al.*, 1994 Construction of human Y-chromosomal haplotypes using a new polymorphic A to G transition. Hum. Mol. Genet. **3:** 2159–2161.

SHRIVER, M. D., L. JIN, R. CHAKRABORTY and E. BOERWINKLE, 1993

VNTR allele frequency distributions under the stepwise mutation model: a computer simulation approach. Genetics **134:** 983–993.

SLATKIN, M., 1995 A measure of population subdivision based on microsatellite allele frequencies. Genetics **139:** 457–462.

SOKAL, R. R., and C. D. MICHENER, 1958 A statistical method for evaluating systematic relationships. Univ. Kansas Sci. Bull. **28:** 1409–1438.

SOODYALL, H., L. VIGILANT, A. V. HILL, M. STONEKING and T. JENKINS, 1996 mtDNA control-region sequence variation suggests multiple independent origins of an "Asian-specific" 9-bp deletion in sub-Saharan Africans. Am. J. Hum. Genet. **58:** 595–608.

SPURDLE, A. B., and T. JENKINS, 1992 Y chromosome probe p49a detects complex *Pvu* II haplotypes and many new *Taq* I haplotypes in southern African populations. Am. J. Hum. Genet. **50:** 107–125.

SPURDLE, A. B., and T. JENKINS, 1993 Complex polymorphisms are revealed by Y chromosome probe 49a with *Bgl*II, *Hind* III, *Pst*I and *Sst*I. Ann. Hum. Genet. **57:** 41–53.

SPURDLE, A. B., and T. JENKINS, 1996 The origins of the Lemba "Black Jews" of southern Africa: evidence from *p12f2* and other Y chromosome markers. Am. J. Hum. Genet. **59:** 1126–1133.

SPURDLE, A., M. F. HAMMER and T. JENKINS, 1994 The Y Alu polymorphism in Southern African populations, and its relationship to other Y-specific polymorphisms. Am. J. Hum. Genet. **54:** 319–330.

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

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

STONEKING, M., and R. L. CANN, 1989 African origin of human mitochondrial DNA, pp. 17–30 in *The Human Revolution: Behavioural and Biological Perspectives on the Origins of Modern Humans,* edited by P. MELLARS and C. STRINGER. Princeton University Press, Princeton.

STONEKING, M., and A. C. WILSON, 1989 Mitochondrial DNA, pp. 215–245 in *The Colonization of the Pacific: A Genetic Trail,* edited by A. V. S. HILL and S. W. SERJEANTSON. Oxford University Press, New York.

STRINGER, C., and C. GAMBLE, 1993 *In Search of the Neanderthals: Solving the Puzzle of Human Origins.* Thames and Hudson, Inc., New York.

SZATHMARY, E. J. E., 1993 Genetics of aboriginal North Americans. Evol. Anthropol. **1:** 202–220.

TAKAHATA, N., 1994 Repeated failures that led to the eventual success in human evolution. Mol. Biol. Evol. **11:** 803–805.

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

TEMPLETON, A. R., 1994 "Eve": hypothesis compatibility *vs.* hypothesis testing. Am. Anthropol. **96:** 141–147.

TEMPLETON, A. R., 1996 Gene lineages and human evolution. Science **272:** 1363.

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 haplotypes in the tiger salamander, *Ambystoma tigrinum.* Genetics **140:** 762–782.

TISHKOFF, S. A., E. DIETZSCH, W. SPEED, A. J. PAKSTIS, J. R. KIDD *et al.,* 1996 Global patterns of linkage disequilibrium at the *CD4* locus and modern human origins. Science **271:** 1380-1387.

TORRONI, A., Y.-S. CHEN, O. SEMINO, A. S. SANTACHIARA-BENECERETTI, C. R. SCOTT *et al.,* 1994a mtDNA and Y-chromosome polymorphisms in four Native American populations from southern Mexico. Am. J. Hum. Genet. **54:** 303–318.

TORRONI, A., J. MILLER, L. G. MOORE, S. ZAMUDIO, J. ZHUANG *et al.,* 1994b Mitochondrial DNA analysis in Tibet: implications for the origin of the Tibetan population and its adaptation to high altitude. Am. J. Phys. Anthropol. **93:** 189–199.

UNDERHILL, P. A., L. JIN, R. ZEMANS, P. J. OEFNER and L. L. CAVALLI-SFORZA, 1996 A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. Proc. Natl. Acad. Sci. USA **93:** 196–200.

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.

WAINSCOAT, J. S., A. V. S. HILL, A. L. BOYCE, J. FLINT, M. HERNANDEZ *et al.,* 1986 Evolutionary relationships of human populations from an analysis of nuclear DNA polymorphisms. Nature **319:** 491–493.

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

WEBER, J. L., and C. WONG, 1993 Mutation in human short tandem repeats. Hum. Mol. Genet. **2:** 1123–1128.

WEIR, B. S., 1990 *Genetic Data Analysis.* Sinauer Associates, Inc., Sunderland, MA.

Communicating editor: A. G. CLARK