

A Back Migration from Asia to Sub-Saharan Africa Is Supported by High-Resolution Analysis of Human Y-Chromosome Haplotypes

Fulvio Cruciani,¹ Piero Santolamazza,¹ Peidong Shen,⁵ Vincent Macaulay,⁶ Pedro Moral,⁷ Antonel Olckers,⁸ David Modiano,² Susan Holmes,⁹ Giovanni Destro-Bisol,³ Valentina Coia,^{3,4} Douglas C. Wallace,¹¹ Peter J. Oefner,⁵ Antonio Torroni,^{1,12} L. Luca Cavalli-Sforza,¹⁰ Rosaria Scozzari,¹ and Peter A. Underhill¹⁰

Dipartimenti di ¹Genetica e Biologia Molecolare, ²Scienze di Sanità Pubblica, Sezione di Parassitologia, and ³Biologia Animale e dell'Uomo, Sezione di Antropologia, Università "La Sapienza," and ⁴Istituto di Medicina Legale, Università Cattolica, Rome; ⁵Stanford Genome Technology Center, Palo Alto, California; ⁶Department of Statistics, University of Oxford, Oxford; ⁷Departament de Biologia Animal, Universitat de Barcelona, Barcelona; ⁸Centre for Genome Research, Potchefstroom University for Christian Higher Education, Pretoria, South Africa; Departments of ⁹Statistics and ¹⁰Genetics, Stanford University, Stanford, California; ¹¹Center for Molecular Medicine, Emory University School of Medicine, Atlanta; and ¹²Dipartimento di Genetica e Microbiologia, Università di Pavia, Pavia, Italy

The variation of 77 biallelic sites located in the nonrecombining portion of the Y chromosome was examined in 608 male subjects from 22 African populations. This survey revealed a total of 37 binary haplotypes, which were combined with microsatellite polymorphism data to evaluate internal diversities and to estimate coalescence ages of the binary haplotypes. The majority of binary haplotypes showed a nonuniform distribution across the continent. Analysis of molecular variance detected a high level of interpopulation diversity ($\Phi_{ST} = 0.342$), which appears to be partially related to the geography ($\Phi_{CT} = 0.230$). In sub-Saharan Africa, the recent spread of a set of haplotypes partially erased pre-existing diversity, but a high level of population ($\Phi_{ST} = 0.332$) and geographic ($\Phi_{CT} = 0.179$) structuring persists. Correspondence analysis shows that three main clusters of populations can be identified: northern, eastern, and sub-Saharan Africans. Among the latter, the Khoisan, the Pygmies, and the northern Cameroonians are clearly distinct from a tight cluster formed by the Niger-Congo-speaking populations from western, central western, and southern Africa. Phylogeographic analyses suggest that a large component of the present Khoisan gene pool is eastern African in origin and that Asia was the source of a back migration to sub-Saharan Africa. Haplogroup IX Y chromosomes appear to have been involved in such a migration, the traces of which can now be observed mostly in northern Cameroon.

Introduction

The sex-specific portion of the human Y chromosome is haploid, is paternally transmitted, and escapes recombination. These features make its DNA sequence variation an invaluable tool for the study of modern human evolution. Haploidy and patrilinearity translate into increased levels of population subdivision compared with the autosomes, and the lack of recombination permits the reconstruction of an unequivocal haplotype phylogeny, which can be related to the geographic distribution of the Y haplotypes, in an approach known as "phylogeography" (Avice et al. 1987; Underhill et al. 2001*b*). Since the discovery of the first polymorphisms in the nonrecombining portion of the Y chromosome (NRY)

~15 years ago (Casanova et al. 1985; Ngo et al. 1986), a large number of studies involving various aspects of human population genetics have been published, but the paucity of usable polymorphic loci on the NRY, which reflects its low level of sequence variation (Jakubiczka et al. 1989; Malaspina et al. 1990; Dorit et al. 1995; Hammer 1995; Whitfield et al. 1995; Shen et al. 2000; Thomson et al. 2000; The International SNP Map Working Group 2001), has hindered progress. Recently, Underhill et al. (1997, 2000, 2001*b*), Rosser et al. (2000), and Hammer et al. (2000, 2001) reported a large number of Y-chromosome biallelic polymorphisms, which has provided a detailed phylogeographic portrait of contemporary global population structure and past population movements and interactions. The availability of these highly geographically structured sets of markers has stimulated the analysis of more restricted areas, leading to important clues about the peopling of Europe (Rosser et al. 2000; Semino et al. 2000; Scozzari et al. 2001), Asia (Su et al. 1999; Capelli et al. 2001; Karafet et al. 2001; Wells et al. 2001), Oceania (Capelli et al. 2001; Kayser et al. 2001; Underhill et al. 2001*a*), and the

Received December 3, 2001; accepted for publication February 7, 2002; electronically published March 21, 2002.

Address for correspondence and reprints: Dr. Fulvio Cruciani, Dipartimento di Genetica e Biologia Molecolare, Università "La Sapienza," Piazzale Aldo Moro 5, 00185, Rome, Italy. E-mail: fulvio.cruciani@uniroma1.it

© 2002 by The American Society of Human Genetics. All rights reserved. 0002-9297/2002/7005-0011\$15.00

Americas (Underhill et al. 1996; Karafet et al. 1999; Santos et al. 1999; Lell et al. 2002).

Africa has had a central role in human evolutionary history. Both genetic and paleoanthropological evidence has accumulated in support of an African origin for our species (Cann et al. 1987; Scozzari et al. 1988; Vigilant et al. 1991; Waddle 1994; Horai et al. 1995; Penny et al. 1995; Hammer et al. 1998; Lahr and Foley 1998; Quintana-Murci et al. 1999; Ingman et al. 2000; Underhill et al. 2000; Walter et al. 2000; Ke et al. 2001). However, few studies specifically dealing with the Y-chromosome diversity of this continent have been published, and these were either based on a small number of polymorphic markers (Seielstad et al. 1994; Scozzari et al. 1999) and/or focused on specific geographic areas inside the continent (Passarino et al. 1998; Thomas et al. 2000; Bosch et al. 2001; Semino et al. 2002).

In the present study, we report the Y-chromosome haplotypes detected by surveying 77 biallelic markers in 22 African populations, representing all major population groups on the continent. In addition, the internal diversity of each binary haplotype was assessed by determination of the allele state at seven STR markers. The observed pattern of NRY variation reveals a profound geographic structuring in Africa, suggestive of several complex demographic episodes involving size fluctuations, migrations, expansions, mergers, and subdivisions.

Subjects and Methods

Subjects

A total of 608 unrelated male subjects belonging to 22 African populations were analyzed. Appropriate informed consent was obtained from all participants. Sample sizes, geographic origin, and linguistic affiliation for each population are reported in table 1.

With the exception of the Ethiopian Jews, all the populations have been previously analyzed for subsets of the Y-chromosome polymorphisms used in the present study (see references in table 1). Slight differences between these studies and the present work in sample sizes for some populations reflect subsequent unavailability of DNA. Mbuti and Biaka Pygmies and the small Lissongo sample were previously reported as “central Africa” in the article by Underhill et al. (2000). Several subjects from Europe (63 Spanish, 20 Italians, 15 Danes, and 4 Poles), and eight subjects from the Middle East have been also analyzed for a specific mutation (M269) and were included for comparison with the African data (see the “Geographic Distribution of Haplotypes” subsection of the “Results” section).

Molecular Analysis

DNA samples were obtained from blood specimens or cultured cells by phenol-chloroform extraction and ethanol precipitation. Markers analyzed include 77 biallelic polymorphisms (fig. 1) and seven STRs (four di- and three tetranucleotide repeats). The majority of the biallelic markers were genotyped according to a hierarchical approach, on the basis of the phylogeny reported by Underhill et al. (2000, 2001*b*). Using this approach, we were able to assign all of the African chromosomes to 37 specific haplotypes of the 131 haplotypes reported by Underhill et al. (2001*b*), with the caveat that, when a terminal branch of the tree proposed by Underhill et al. (2001*b*) was more than one mutation long, only a subset of the markers on that branch was typed in the present study.

The majority of the biallelic markers were typed using PCR-amplified products and subsequent heteroduplex analysis by denaturing high-performance liquid chromatography (DHPLC), as described by Underhill et al. (1997, 2000, 2001*b*). The following markers were analyzed using genotype detection strategies other than DHPLC: M9, M13, and M14 were typed using a PCR/RFLP assay, through use of the primers described by Underhill et al. (1997) and the appropriate restriction enzymes (*Hinf*I, *Dpn*II, and *Afl*III, respectively); markers M12 and M45 (Underhill et al. 1997, 2000) were also typed as RFLPs but through use of primer pairs different from those described in the original references: M12 FOR, 5'-actaaaacaccattagaaacaaagg; M12 REV, 5'-tgagcaacatagtgaccccgat; M45 FOR, 5'-ttggcagtgaaaaattatagcta; M45 REV, 5'-acagttgtgacagtggcacc (the underlined base in one of the primers in each pair indicates a noncomplementary nucleotide that generates *Dpn*II and *Bfa*I restriction sites for M12 and M45, respectively); DYS271 (Seielstad et al. 1994), YAP (Hammer and Horai 1995), PN2 and PN3 (Hammer et al. 1997), and DYS257 (Hammer et al. 1998) were genotyped according to the procedures described in the original references; SRY₁₀₈₃₁ (Whitfield et al. 1995) and M40 (corresponding to SRY₄₀₆₄ in Whitfield et al. 1995) were analyzed according to the procedure described by Scozzari et al. (1999); and p12f2 (Casanova et al. 1985) was genotyped according to the procedure described by Rosser et al. (2000). Three new polymorphic markers (M269, M236, and M288) are here described for the first time. The M269 polymorphism is a C→T transition in the EIF1AY (eukaryotic translation initiation factor 1A, Y chromosome) gene discovered at the Stanford Genome Technology Center. A 379-bp fragment containing the M269 polymorphism was amplified using the primers M269 FOR (5'-ctaagatcagagtatctccctttg) and M269 REV (5'-aaattgtttcaattaccag) and was genotyped by DHPLC.

Table 1**Sampled Populations**

GEOGRAPHIC AREA, COUNTRY, AND POPULATION	N ^a	LINGUISTIC AFFILIATION ^b		REFERENCE(S)
		Family	Sublevel	
Northern Africa:				
Morocco:				
Arabs	49	Afro-Asiatic	Semitic	Scozzari et al. 1999, 2001
Berbers	64	Afro-Asiatic	Semitic	Scozzari et al. 2001
Eastern Africa:				
Ethiopia:				
Ethiopian Jews	22	Afro-Asiatic	Semitic	Present study
Western Africa:				
Burkina Faso:				
Mossi	49	Niger-Congo	Voltaic	Scozzari et al. 1997, 1999
Rimaibe	37	Niger-Congo	West Atlantic	Scozzari et al. 1997, 1999
Fulbe	20	Niger-Congo	West Atlantic	Scozzari et al. 1997, 1999
Central Western Africa ^c :				
Cameroon (northern):				
Fali	39	Niger-Congo	Adamawa	Scozzari et al. 1997, 1999
Tali	15	Niger-Congo	Adamawa	Scozzari et al. 1997, 1999
Mixed ^d	18	Niger-Congo	Adamawa	Scozzari et al. 1999
Fulbe	17	Niger-Congo	West Atlantic	Scozzari et al. 1997, 1999
Ouldeme	21	Afro-Asiatic	Chadic	Scozzari et al. 1997, 1999
Daba	18	Afro-Asiatic	Chadic	Scozzari et al. 1997, 1999
Mixed ^d	15	Afro-Asiatic	Chadic	Scozzari et al. 1999
Mixed ^d	9	Nilo-Saharan	Central Sudanic/Saharan	Scozzari et al. 1997, 1999
Cameroon (southern):				
Bamileke	48	Niger-Congo	Benué-Congo, Bantoid	Scozzari et al. 1997, 1999
Ewondo	29	Niger-Congo	Benué-Congo, Bantu	Scozzari et al. 1997, 1999
Bakaka	12	Niger-Congo	Benué-Congo, Bantu	Scozzari et al. 1999
Central Africa:				
Central African Republic:				
Biaka Pygmies ^e	20	Niger-Congo	Bantu	Underhill et al. 2000
Lissongo ^e	4	Niger-Congo	Bantu	Underhill et al. 2000
Democratic Republic of Congo:				
Mbuti Pygmies ^e	12	Nilo-Saharan	Central Sudanic	Underhill et al. 2000
Southern Africa:				
South Africa:				
!Kung	64	Khoisan	Northern	Scozzari et al. 1997, 1999
Khwe	26	Khoisan	Central	Scozzari et al. 1997, 1999

^a Number of Y chromosomes analyzed.

^b According to Grimes and Grimes (2000) and Greenberg (1987).

^c See Spedini et al. (1999) and the Laboratory of Molecular Anthropology Web site for anthropological information regarding Cameroonian populations.

^d Composite sample; see Scozzari et al. (1999).

^e Reported as central Africa in the article by Underhill et al. (2000).

The M236 and M288 polymorphisms were discovered by sequencing the Y-specific region in the proximity of the pseudoautosomal boundary (R.S., unpublished data) and are due to transversions (G→C and C→A, respectively) that are 177 bp and 175 bp away from the pseudoautosomal boundary (nucleotide positions 18310 and 18312 in the Y-chromosome cosmid cAMF3.1; GenBank accession number X96421). A PCR/RFLP assay was developed using primers SRY18121 FOR (5'-ccttctg-agctatactgtatgt) and SRY18531 REV (5'-atgatgctcagga-ctcagacct) and the restriction enzymes *Ava*I, *Acc*III, and *Acc*II.

The (CA)_n dinucleotide repeats YCAIIa and YCAIIb (Mathias et al. 1994) and DYS413a and DYS413b (Mathias et al. 1994; Malaspina et al. 1997) were analyzed and their allele sizes scored as described by Malaspina et al. (1998). The (GATA)_n tetranucleotide repeats DYS391 (Roewer et al. 1996; Kayser et al. 1997), A7.2 (White et al. 1999), and DYS439 (Ayub et al. 2000) were coamplified in a single PCR using the primers described in the original references, after [γ -³²P]dATP terminal labeling of one of the primers in each pair, and were electrophoresed on 6% sequencing gel. Sequenced alleles were included in each run, as size-calibration stan-

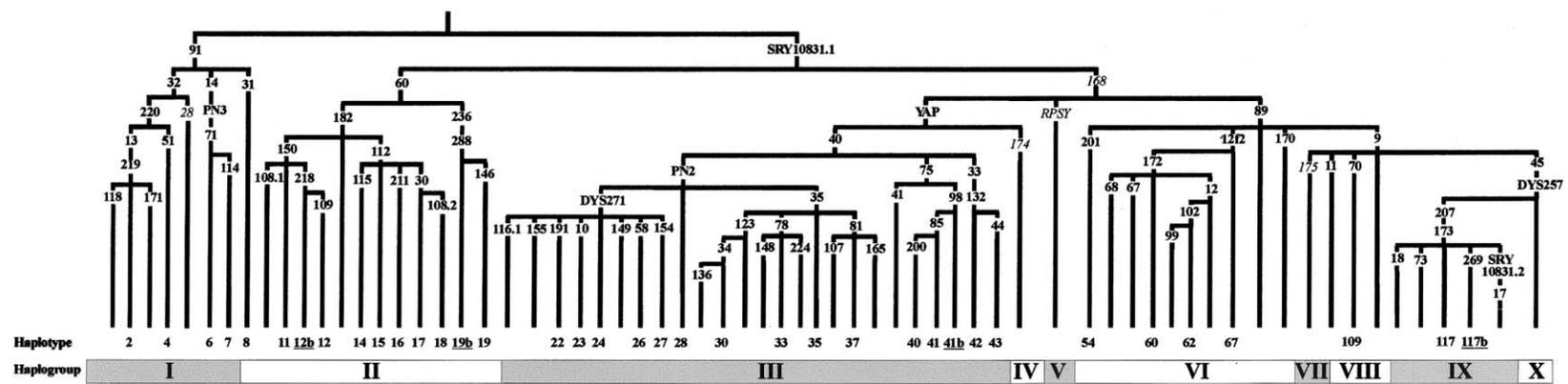


Figure 1 Maximum-parsimony phylogeny of NRY biallelic markers. Markers are indicated on the branches. Italicized markers were not typed in the present study, but their allelic state could be deduced unequivocally from their position in the genealogy given by Underhill et al. (2001*b*). Group IX markers defining the 117b derivative haplotypes 110–112, 114–116, and 118 (Underhill et al. 2001*b*; P.A.U., unpublished data) have not been typed in the present study. Haplogroup and haplotype designations are according to Underhill et al. (2001*b*), with some modifications (see the “Results” section). Only the haplotypes observed in the present study are numbered. Underlined haplotypes are those described here for the first time. Phylogenetic relationships between group IX markers M207 and M173 (Karafet et al. 2001) and M17 and SRY₁₀₈₃₁ (Weale et al. 2001) have recently been resolved.

dards. Note that Gonzalez-Neira et al. (2001) recently reported that DYS439 (Ayub et al. 2000) is likely to be the same marker previously described as GATA A4 by White et al. (1999).

Statistical Analysis

Correspondence analysis was performed using the SPSS package, version 8.0 (SPSS Inc.). Data have been standardized by removing row (populations) and column (binary Y haplotypes) means, and a row principal normalization was also used. In this analysis, 31 African populations have been included: 22 populations from the present study, 5 populations from the study by Underhill et al. (2000) (Mali, Sudan, Ethiopia, southern African Bantu, and Khoisan), and 4 Moroccan populations from the study by Bosch et al. (2001) (Saharawis, southern Moroccan Berbers, Arabs, and north-central Moroccan Berbers). These studies have analyzed a largely overlapping—but not identical—set of markers; thus, population comparisons were performed by use of binary haplotypes defined by only the SNPs that were shared by the three studies.

Analysis of molecular variance (AMOVA; Excoffier et al. 1992) was performed using Arlequin, version 2.000 (Schneider et al. 2000; Arlequin's Home on the Web). Two hierarchical levels (individuals into populations and populations into geographic groups) were considered. Φ statistics and their significance levels were calculated as described by Scozzari et al. (1999). To evaluate the molecular distances between haplotypes, we used, as a reference, the tree described by Underhill et al. (2001b), modified to include a few additional markers (SRY₁₀₈₃₁, DYS257, M269, M236, and M288).

For each binary haplotype, phylogenetic relationships of the haplotypes defined by the seven microsatellites analyzed were depicted by means of median joining (MJ) networks (Bandelt et al. 1999), through use of the Network 2.0e program (Life Sciences and Engineering Technology Solutions Web site) with the epsilon value set to zero.

To estimate the time to the most recent common ancestor (TMRCA) of a binary haplotype cluster, we used the three tetranucleotide loci. The results for the complex dinucleotide microsatellites YCAIIa, YCAIIb, DYS413a, and DYS413b were excluded from the analysis, since an accurate estimate of their mutation rate is not available, owing to the relatively low number of meioses analyzed and mutations observed (1 mutation in 490 meioses; Kayser et al. 2000). TMRCA can be estimated from ASD/μ , where ASD is the arithmetic mean, across loci, of the average squared distance statistic (from each haplotype in the cluster to the putative ancestral haplotype, which was assumed to be the haplotype carrying the most frequent allele at each microsatellite) (Goldstein et

al. 1995; Slatkin 1995) and μ is an estimator of the mutation rate of these loci—specifically, the fraction of meioses that involved mutation at tetranucleotide loci (16/7,292) in the study by Kayser et al. (2000). This provides a point estimate of TMRCA in number of generations. The value of 25 years per generation was used to convert this to years. To obtain a CI, we performed a parametric bootstrap, over both ASD and μ . Specifically, at each of 10,000 simulations, we (i) sampled a mutation rate from binomial (7,292, 16/7,292)/7,292; then (ii) assumed that the genealogy of the cluster was perfectly starlike, with a mutational depth equal to the measured ASD, and simulated a symmetric single-step mutation process at each locus and calculated ASD. For each simulation, we evaluated ASD/μ and, from the distribution of values that resulted, we used, as a 95% CI, the interval that covered the central 95% of values. It is worth noting that (unmodeled) uncertainty in the shape of the genealogy, as well as in the mutation process, means that these CIs are underestimated. Moreover, possible range constraints on the microsatellite allele length may influence the observed variability (Kayser et al. 2000).

Results

Y-Chromosome Phylogeny

Data from the present study and from Underhill et al. (2001b) have been used to augment a phylogenetic tree composed of 77 biallelic polymorphisms (fig. 1). Thirty-seven binary haplotypes were found in the total sample of 608 African subjects. The notation of Underhill et al. (2001b) for these haplotypes was kept, except for four haplotypes that are described here for the first time. These are: (1) haplotype 12b, carrying the M218 mutation but ancestral at M109; (2) haplotype 19b, carrying the derived alleles at M236 and M288 but ancestral at M146; (3) haplotype 41b, carrying the mutated allele at M98 but ancestral at M85; and (4) haplotype 117b, carrying the M269 derived allele on a M207/M173 background.

Geographic Distribution of Haplotypes

Haplotype frequencies in 22 African populations are given in table 2. Following Underhill et al. (2000, 2001b), haplotypes have been partitioned into haplogroups according to either the presence or the absence of alleles located in the interior of the phylogeny. Six groups (I–III, VI, VIII, and IX) have been observed among the African populations analyzed. Group III (15 different haplotypes) accounts for the majority (73%) of the chromosomes and is widespread across the African continent. The African-specific groups I and II (Underhill et al. 2000, 2001b), which account for 7% and

Table 2**Y-Chromosome Haplotype Frequencies (%) in 22 African Populations**

POPULATION (N ^a)	HAPLOTYPE FREQUENCY ^b (%)																																					
	I				II						III						VI				VIII	IX																
	2	4	6	7	8	11	12b	12	14	15	16	17	18	19	19b	22	23	24	26	27	28	30	33	35	37	40	41	41b	42	43	54	60	62	67	109	117	117b	
Arabs (49)																						43	33						2	10		10					2	
Berbers (64)					3												5					11	69				2		5		6							
Ethiopian Jews (22)	41																			18	14	9	9								5			5		5		
Mossi (49)													2			22	67				2		2				4											
Rimaibe (37)																8	51	5		3						27			5									
Fulbe (Burkina Faso) (20)																		90											10									
Fali (39)							18									33	26																				23	
Tali (15)																20	7	47										20									7	
Mixed Adamawa (18)	6						6	6									11	17																			56	
Fulbe (Cameroon) (17)	12																			6										53					18	12		
Ouldeme (21)							5																															95
Daba (18)																		28				6				22											44	
Mixed Chadic (15)	7						7									7	7					7														67		
Mixed Nilo-Saharan (9)	22											22				22	11								11												11	
Bamileke (48)														4		56	25		15																			
Ewondo (29)							10									21	66																					
Bakaka (12)																25	67		8																			
Biaka Pygmies (20)							5		5	20	5					40	25																					
Lissongo (4)																25	25	25																				
Mbuti Pygmies (12)							8		8	17						33	8								25													
!Kung (64)		28	5	3					8							16	23						11				6											
Khwe (26)		12															50		4				31			4												

^a Number of Y chromosomes analyzed.^b Haplotype numbers are according to the nomenclature of Underhill et al. (2001b), except for four haplotypes (12b, 19b, 41b, and 117b) that are reported here for the first time.

6% of the total sample, respectively, are scattered across the continent, with lowest frequency among northwestern African populations. Group VI chromosomes are present at a low frequency in northern Africa and Ethiopia and are completely absent in samples from the sub-Saharan region. Only four group VIII chromosomes have been found: one in Ethiopia and three in the Fulbe from Cameroon. Group IX chromosomes are restricted to Cameroon (26%), with a single instance in Morocco.

Most of the binary haplotypes showed a strict regional distribution within the African continent (table 2 and fig. 2; also see fig. 1). This finding was particularly pronounced for the African-specific group I haplotypes. Haplotype 2, which harbors the derived allele at M13 and M219, was found in eastern Africa (41%) and, at lower frequencies, among the northern Cameroonians (4%). Haplotype 4, harboring the M51 mutation, and haplotypes 6 and 7, which share PN3 and several other mutations, were found exclusively in the Khoisan from southern Africa. Interestingly, the ancient lineage defined by the M31 mutation (haplotype 8), previously found in one subject from Mali (Underhill et al. 2000), was present in two Berbers from Morocco. Haplotypes 1, 3, and 5, previously identified in a few subjects from eastern Africa (Underhill et al. 2000), were not found in the present study.

Regarding the group II haplotypes, a clear-cut difference was observed between Pygmies and Khoisan, on one side, and all the other African populations, on the other: the former have mainly group II haplotypes sharing the derived allele at M112 (haplotypes 14–18), whereas in other African populations, group II chromosomes display mainly the M150 mutation (haplotypes 9–12 and 12b). Among these, haplotype 12, which carries the M109 mutation, is the most frequent and was found in several populations from Cameroon. The same haplotype was also previously observed in some eastern Africans and Bantu speakers from southern Africa (Underhill et al. 2000). The new markers M236 and M288 define an old group II clade (haplotypes 19 and 19b), which is restricted to few sub-Saharan subjects.

As previously mentioned, group III chromosomes are widespread all over Africa, but the distribution of the numerous distinctive haplotypes is not homogeneous across the continent. Chromosomes carrying the DYS271 (M2) mutation (haplotypes 20–27 in fig. 1, and defined as haplotype 5 in the article by Hammer et al. [1997]) are mainly limited to sub-Saharan populations and reach frequencies >65% in some populations (Burkina Faso and southern Cameroon; see also Scozzari et al. 1999). Several sublineages carrying the DYS271 mutation have been observed; one of these is haplotype 22, which is defined by the M191 mutation and accounts for 33% of the DYS271 chromosomes. In the African sample that we have analyzed, this haplotype has the same distribution as the hap-

lotype carrying the DYS271 mutation alone (haplotype 24). The group of chromosomes sharing the derived allele at PN2 and the ancestral state at DYS271, previously defined as haplotype 4 by Hammer et al. (1997), includes many haplotypes having different geographic distributions. The least-resolved haplotypes (28 and its derivative 35) at present were observed in populations from both eastern and sub-Saharan Africa (Underhill et al. 2000; present study). Among the most differentiated lineages carrying the M35 mutation, haplotype 30 (M34) is present in eastern Africa, as is haplotype 33 (M78), which is also found at high frequency in northern Africa, whereas haplotype 37 (M81) is found at high frequencies only in northern Africa. Haplotypes 30, 33, and 37 have also been found in populations from Europe and the Middle East (Underhill et al. 2000; Bosch et al. 2001; R.S., unpublished data). Haplotypes carrying the mutations M75 and M33 (haplotypes 39–43, previously identified as haplotype 3A by Hammer et al. [1997]) are present at low frequencies across the entire continent but with different individual distributions (table 2 and Underhill et al. 2000).

Group VI (haplotype 54, which carries the M201 mutation, and haplotypes 60, 62, and 67, which carry the p12f2 mutation) represents 16% and 5% of the northern and eastern African chromosomes, respectively, but is absent in sub-Saharan populations. Haplotypes carrying the M201 or p12f2 mutations are common in southern central Europe and the Middle East (Semino et al. 1996, 2000; Rosser et al. 2000; Malaspina et al. 2001; Scozzari et al. 2001), and their distribution has been associated with the Neolithic expansion from the Middle East (Semino et al. 1996, 2000; Bosch et al. 2001). It is worth noting that the frequency of group VI chromosomes in the Ethiopian Jews (just one chromosome out of 22) is similar to that reported for the p12f2 chromosomes in the Oromo from Ethiopia (4%) and is considerably lower than the frequency reported for the Amhara of the same region (33%), for whom a strong Middle Eastern genetic component has been reported (Semino et al. 2002). These data, together with those reported elsewhere (Ritte et al. 1993a, 1993b; Hammer et al. 2000) suggest that the Ethiopian Jews acquired their religion without substantial genetic admixture from Middle Eastern peoples and that they can be considered an ethnic group with essentially a continental African genetic composition.

The four group VIII chromosomes listed in table 2 all carry the M70 mutation that defines haplotype 109. This haplotype was also found at low frequencies across Europe (Semino et al. 2000; R.S., unpublished data).

With the exception of a single Y chromosome from Morocco with the M269 mutation (haplotype 117b), all group IX African chromosomes are characterized by the presence of the M173 and M207 derived alleles and the

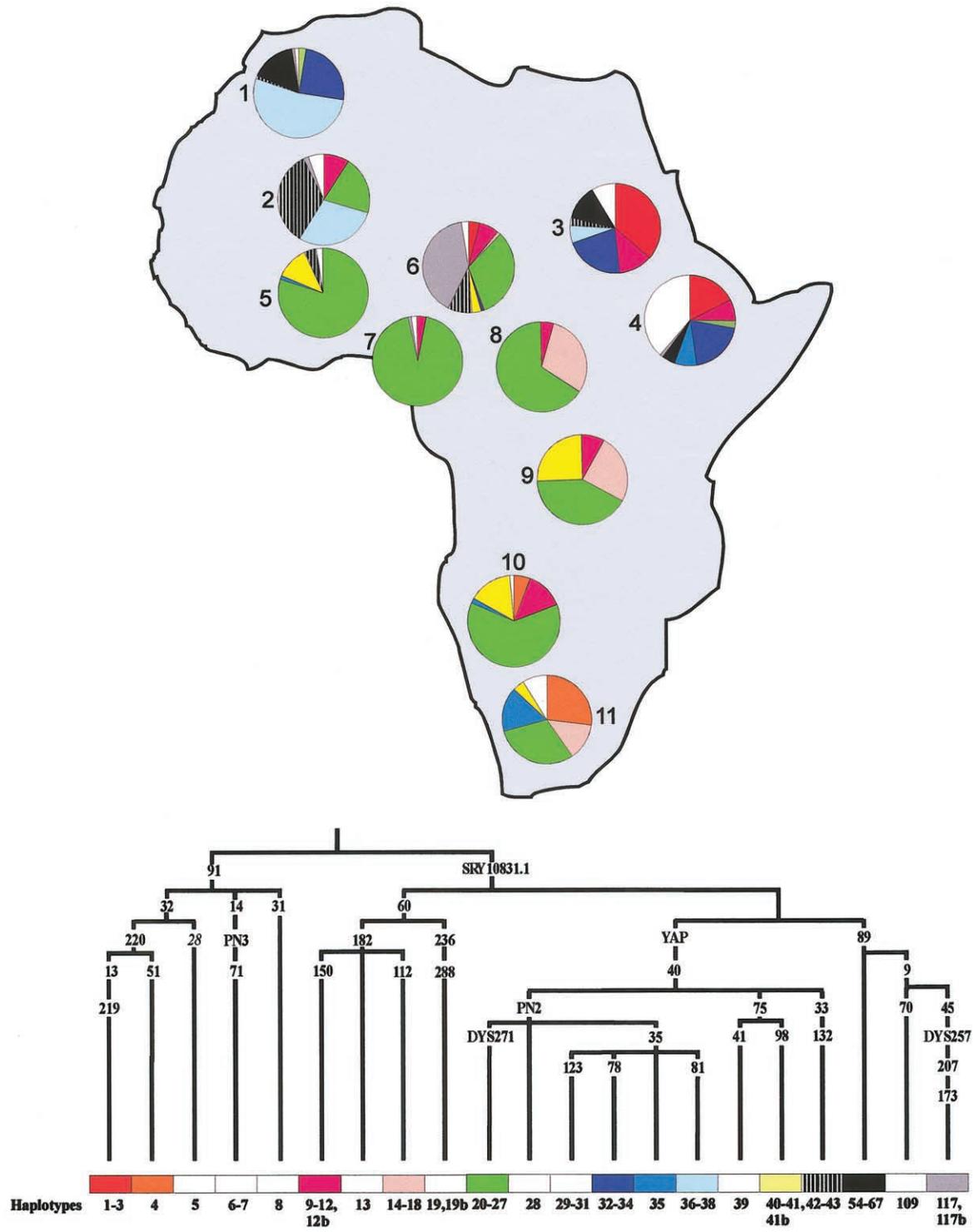


Figure 2 Distribution of major Y-chromosome haplotypes in 11 geographic areas of Africa. Some evolutionarily related haplotypes were pooled according to the simplified phylogeny shown at the bottom of the figure. The 12 most common haplotypes or groups of related haplotypes are represented by different colors. The remaining haplotypes are represented in the white sector of the relevant pie charts. Geographic groups are as follows: 1, Morocco; 2, Mali; 3, Sudan; 4, Ethiopia; 5, Burkina Faso; 6, northern Cameroon; 7, southern Cameroon; 8, Central African Republic (Biaka Pygmies); 9, Democratic Republic of Congo (Mbuti Pygmies); 10, southern Africa Bantu; 11, southern Africa Khoisan. Data for Morocco, Burkina Faso, Cameroon, Central African Republic, and Democratic Republic of Congo are from the present study; the frequencies for Ethiopia and the southern Africa Khoisan have been obtained by pooling data from the present study and the study by Underhill et al. (2000); data for Mali, Sudan, and southern Africa Bantu are from Underhill et al. (2000).

absence of the downstream mutations (haplotype 117). Haplotype 117 was found only in Cameroon, where it accounts for 26% of the chromosomes (40% in northern Cameroon). Chromosomes from Cameroon with this haplotype are the same as those reported in a previous article as belonging to haplotype 1C (Scozzari et al. 1999). Since, so far, no population data have been published for the M269 mutation, in the present study 102 European and 8 Middle Eastern Y chromosomes were analyzed for this marker. These chromosomes had been previously classified as haplotype 1 by Scozzari et al. (2001) (DYS257 A/ SRY₁₀₈₃₁ G chromosomes, corresponding to haplotypes 110–118 and 123 [group IX], and 124–131 [group X] of Underhill et al. [2001b]). In contrast to the group IX chromosomes from Cameroon, all western Eurasian chromosomes were found to carry the M269 derived allele.

Correspondence Analysis

In a first run of the correspondence analysis (fig. 3A), we analyzed 22 African populations from the present study together with 9 African populations taken from the studies by Underhill et al. (2000) and Bosch et al. (2001) (see the “Statistical Analysis” subsection of the “Subjects and Methods” section). The first and the second dimensions capture 17% and 13% of the inertia, respectively. On the basis of the first two dimensions scores, populations are grouped into three major clusters corresponding to sub-Saharan, northwestern, and eastern Africans. The main determinants of these clusterings are haplotype 37 (common in northern Africa), haplotype 2 (common in eastern Africa), and haplotype 22/24 (common in sub-Saharan Africa). Two populations stand apart in the plot: the Fulbe from Cameroon, because of the high frequency of haplotype 43, and the Mali sample, which is located between the major sub-Saharan cluster and the northwestern cluster, because of the intermediate frequency of haplotypes 37 and 22/24. To achieve a better resolution of the genetic relationships among sub-Saharan populations, a second correspondence analysis was run by including only populations with negative values on the first dimension of the previous correspondence analysis. As is shown in figure 3B, four population groups were identified: the Khoisan from southern Africa, the northern Cameroonians, the Pygmies, and the Niger-Congo-speaking populations from various regions (Burkina Faso, southern Cameroon, and southern Africa). Unlike the other two population clusters, both the Khoisan and the northern Cameroonians show a high degree of interpopulation diversity, which is illustrated by the more scattered distribution of the points in the plot. The Khwe sample in figure 3B occupies an intermediate position between the other Khoisan and the Niger-Congo-speaking popula-

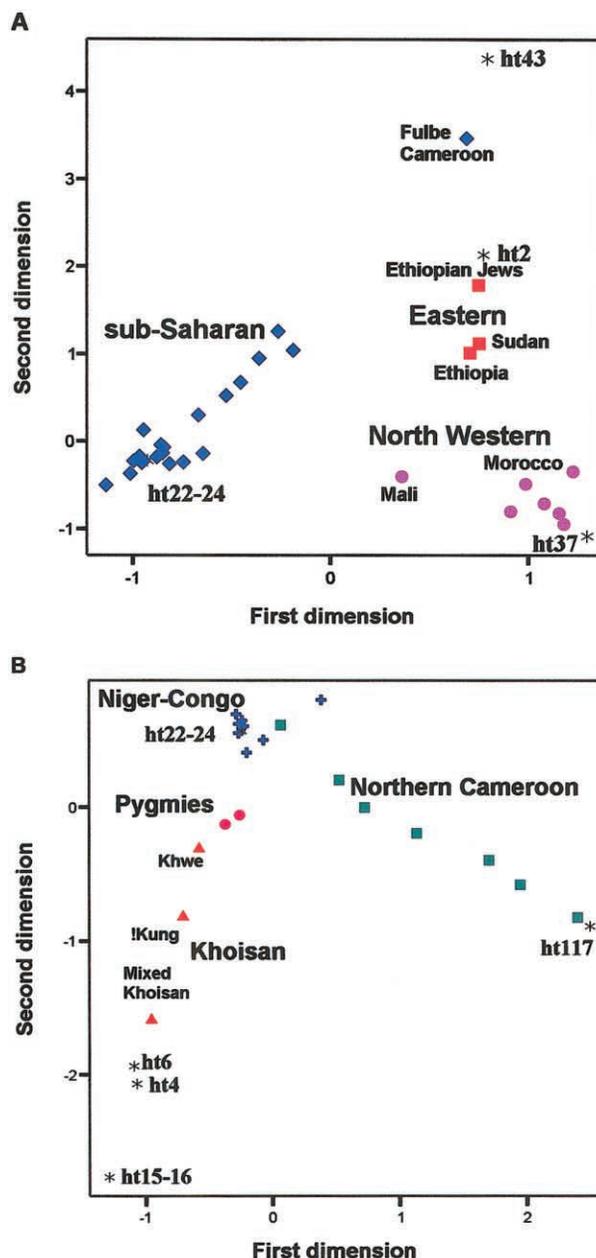


Figure 3 Correspondence analysis scores; plot of populations and haplotypes in the space of the first (X-axis) and second (Y-axis) dimension. Only the haplotypes (represented by asterisks [*]) most contributing to the inertia of the first and/or second dimension are represented. *A*, 31 African populations, including the 22 populations listed in table 1, 5 populations from Underhill et al. (2000), and 4 Moroccan populations from Bosch et al. (2001). Symbols used to identify groups of populations are as follows: diamonds, sub-Saharan Africans; circles, northwestern Africans; squares, eastern Africans. *B*, Correspondence analysis for the 20 African populations having negative scores for the first dimension in the previous analysis. The following symbols have been used to identify groups of populations: triangles, Khoisan; circles, Pygmies; squares, northern Cameroonians; crosses, other sub-Saharan Africans (populations from Burkina Faso and Bantu-speaking populations from southern Cameroon and southern Africa).

tions, most likely because of the relatively high frequencies of haplotypes carrying the DYS271 mutation.

AMOVA Analysis

To assess the level of population structure, we estimated various Φ statistics (table 3) by use of AMOVA (Excoffier et al. 1992) and by taking into account interhaplotype molecular differences (see the “Statistical Analysis” subsection of the “Subjects and Methods” section). The overall Φ_{ST} value calculated for the entire African sample, comprising 22 populations without groupings, was 0.342 (significantly greater than zero, $P < 10^{-4}$), indicating that a large proportion of the overall Y-chromosome variation resulted from interpopulation differences. When the populations were partitioned into seven groups according to a geographic criterion, a high degree of both inter- and intragroup variability ($\Phi_{CT} = 0.230$, $P < 10^{-4}$; $\Phi_{SC} = 0.170$, $P < 10^{-4}$) was observed. The high Φ_{SC} value was mainly contributed by the populations from northern Cameroon, which showed a high degree of differentiation ($\Phi_{ST} = 0.249$) (for a discussion, see Scozzari et al. 1999).

Network Analysis

The variation at seven STR loci was used to reconstruct an MJ network for those binary haplotypes that showed a wide geographic distribution within the continent. The aim of this analysis was to evaluate whether the observed geographic distribution for each binary haplotype was attributable either to recent or more ancient gene flow. MJ networks for haplotypes 2, 15, 22, 24, 35, and 41 are displayed in figure 4 (A–F, respectively). Networks for haplotypes 2, 15, and 35 (fig. 4A, B, and E) depict both a clear nonuniform geographic distribution of STR haplotypes and no instances of haplotype sharing across geographic regions (with a single exception in the haplotype 35 network). Two distinct

clusters of haplotypes are present in the haplotype 2 network (fig. 4A), corresponding to the Cameroonian and the Ethiopian chromosomes. In the haplotype 35 network (fig. 4E), Ethiopian chromosomes are well differentiated from southern African chromosomes, and the single western African haplotype differs by four repeats from the closest chromosomes in the network. A more extreme situation is observed for the haplotype 15 network (fig. 4B), where a high degree of inter- and intra-population structuring is present, with the Pygmy chromosomes being ≥ 7 mutational steps apart from Khoisan chromosomes, which, in turn, seem to have a high degree of internal diversity. Networks for haplotypes 22, 24, and 41 (fig. 4C, D, and F) show a reduced degree of geographic structuring, with several instances of STR haplotype sharing among different regional groups, suggesting a more recent common ancestry.

Discussion

Pattern of Diversity in Contemporary African Populations

The high degree of interpopulation NRY haplotype diversity that we observed in Africa agrees with previous reports based on the analysis of autosomal, Y chromosome, and mtDNA markers (Melton et al. 1997; Scozzari et al. 1997, 1999; Watson et al. 1997; Jorde et al. 2000; Torroni et al. 2001). The observed NRY Φ_{ST} value (0.342; $P < 10^{-4}$) is the highest among those so far reported for other continents (Hammer et al. 2001; Karafet et al. 2001) and appears to be highly related to geography—that is, the seven geographic clusters listed in table 3 ($\Phi_{CT} = 0.230$; $P < 10^{-4}$). As shown in figure 3A, three main clusters of populations from northern, eastern, and sub-Saharan Africa could be identified. Haplotype sharing among these three population groups

Table 3

AMOVA for Y-Chromosome-Specific Binary Haplotypes in Africa

Geographic Region ^a	No. of Populations	No. of Groups	Φ_{ST} (P^b)	Φ_{CT} (P^b)	Φ_{SC} (P^b)
Northwestern Africa	2	1	.059 (.014)		
Eastern Africa	1	1	...		
Western Africa	3	1	.134 (.000)		
Central western Africa	11	2	.453 (.000)	.269 (.042)	.252 (.000)
Northern Cameroon	8	1	.249 (.000)		
Southern Cameroon	3	1	.057 (.039)		
Central Africa	3	1	.000 (.529)		
Southern Africa	2	1	.093 (.008)		
Overall	22	7	.361 (.000)	.230 (.000)	.170 (.000)
Sub-Saharan Africa	19	5	.332 (.000)	.179 (.006)	.185 (.000)
All of Africa	22	1	.342 (.000)		

^a Geographic groups are defined as in table 1. Northern and southern Cameroon have been kept separated.

^b P value indicates the fraction of cases in which a Φ value greater than the quoted value is obtained in a permutation test of samples across populations and therefore tests the null hypothesis of no population structure.

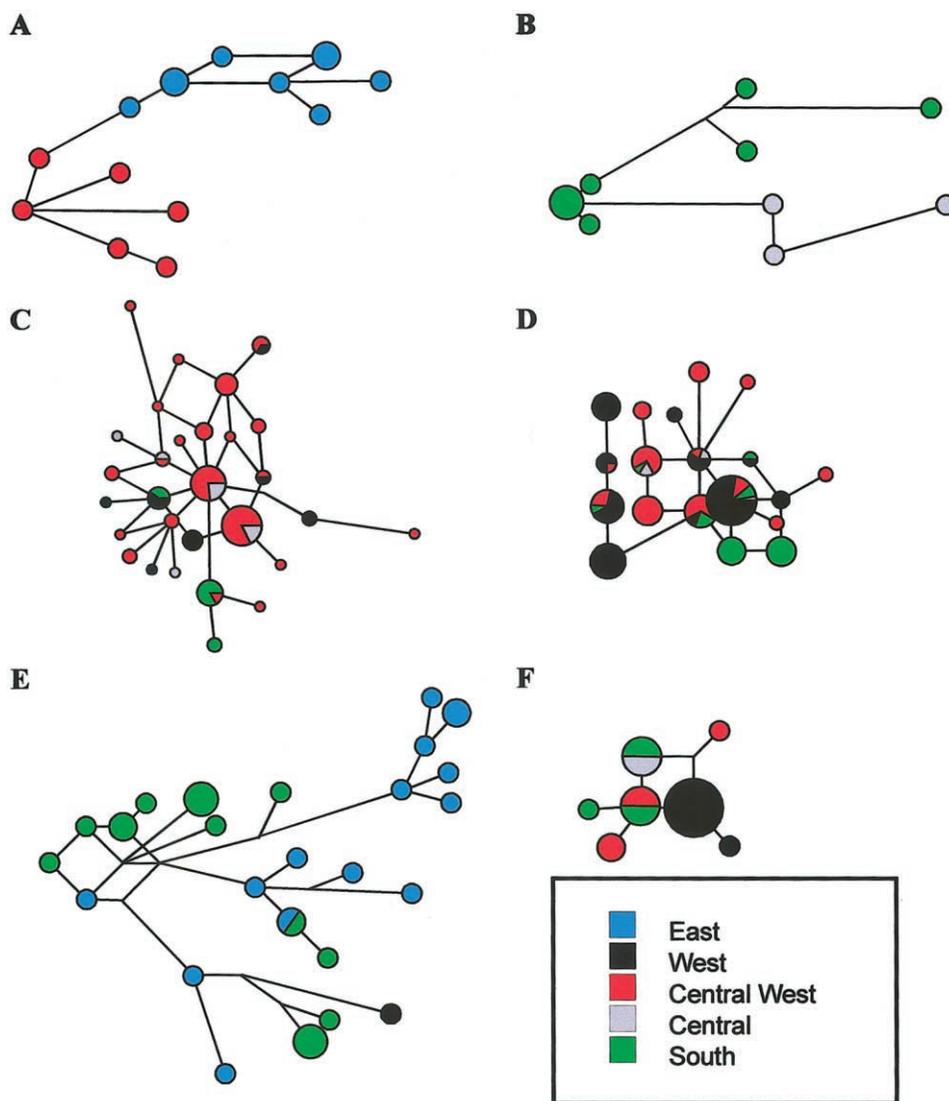


Figure 4 MJ microsatellite networks of six different binary haplotypes. Microsatellite haplotypes are represented by circles, with areas proportional to the number of individuals harboring the haplotype. A, haplotype 2 (M13); B, haplotype 15 (M112); C, haplotype 22 (M191); D, haplotype 24 (DYS271); E, haplotype 35 (M35); and F, haplotype 41 (M85). For each network, the smallest circles represent a count of one individual, with the exception of the network depicted in D (haplotype 24), where only haplotypes having an absolute count of two or more are represented. Branch lengths are proportional to the number of one-repeat mutations separating two haplotypes. Branch length of the network B (haplotype 15) is one-half that in the other networks. Network B also includes three haplotype 15 chromosomes reported by Underhill et al. (2000). Network E (haplotype 35) also includes 16 haplotype 35 chromosomes from southern Ethiopia (R.S., unpublished data). Pygmy chromosomes in network F also carry the M200 derived allele (haplotype 40).

is very limited, with several instances of haplotypes common in one area but rare or absent in other regions. The only notable exceptions are represented by the haplotypes bearing the M78 (common in northern and eastern Africa) and the M35 (common in eastern and southern Africa) mutations. The geographic structuring that we observed agrees quite well with a previous analysis of genetic variation in Africa, based on autosomal protein markers (Cavalli-Sforza et al. 1994), where northern, eastern, and sub-Saharan Africans formed distinct clus-

ters in both the population genetic tree and principal component analysis. The strong differentiation observed here between northern and eastern Africans was not apparent in a previous study dealing with NRY variation in African populations (Scozzari et al. 1999), mainly because of the lack of resolution within the group of chromosomes bearing the PN2 T and DYS271 A alleles (haplotype 4 in the study by Scozzari et al. [1999]). These chromosomes are very frequent in both northern and eastern Africa and have now been dissected into

several haplotypes (haplotypes 28–37 in fig. 1), the majority of which show a high degree of geographic structuring (table 2 and fig. 2). For example, haplotype 37, defined together with two rare haplotypes (36 and 38) by mutation M81, was observed at high frequencies (27%–76%) in several populations from northern Africa (Bosch et al. 2001; present study), but it was found to be rare or absent in eastern Africa (Underhill et al. 2000; present study). A similar frequency distribution has been reported for the Y haplogroup HG 25.2 (Scozzari et al. 2001), which carries the XY275 G allele on a YAP⁺ PN2 T background because of a rare recombinational event between the pseudoautosomal marker XY275 and the NRY/pseudoautosomal region boundary. In the present study, all of the HG 25.2 chromosomes were also found to carry the M81 mutation, and vice-versa, suggesting that the M81 mutation and the XY275 recombination event are phylogenetically equivalent. The coalescence age of the M81-bearing chromosomes, estimated from the variation observed at the three microsatellite loci, was only 1,995 years (95% CI 1,068–4,005 years), a value similar to that estimated for HG 25.2 (1,400 years) through use of four dinucleotide repeats (Scozzari et al. 2001). Thus, both the age and the high frequency of the M81 haplotypes suggest that a demographic expansion has occurred in northwestern Africa about 2,000 years ago.

In sub-Saharan Africa, the spread of two haplotypes that are related through the presence of the DYS271 derived allele (haplotypes 22 and 24) seems to have partially erased pre-existing genetic differences among different geographic regions. These haplotypes were found in all sub-Saharan areas analyzed, with a cumulative frequency of ~80% in western Africa and southern Cameroon. Nevertheless, a high level of population ($\Phi_{ST} = 0.332$, $P < 10^{-4}$) and geographic ($\Phi_{CT} = 0.179$, $P = .006$) structuring can be detected south of the Sahara, the result of both a different local impact of these haplotypes and the presence at high frequency of population-specific haplotypes (M51 and M14 lineages in Khoisan, M112 lineages in Khoisan and Pygmies, haplotype 117 in northern Cameroonians). These findings are clearly depicted by the correspondence analysis (fig. 3B), which identifies four clusters of populations corresponding to Khoisan, Pygmies, northern Cameroonians, and a group of poorly differentiated Niger-Congo-speaking populations from western, central western, and southern Africa.

Recent Population Expansions in Sub-Saharan Africa and the Impact on Pre-Existing Hunter-Gatherer Communities

Among the most common haplotypes, haplotypes 22, 24, and 41 have been found to be present in all of the

sub-Saharan regions analyzed here (table 2 and fig. 2). Network analysis of the STRs associated with these haplotypes showed a low degree of geographic differentiation (fig. 4C, D, and F), suggesting that their present-day distribution is the consequence of relatively recent range expansion(s). Previous studies associated the diffusion of Bantu-speaking peoples from central western Africa toward southern Africa with the presence of DYS271-bearing chromosomes (haplotypes 22, 24, and derivatives) (Passarino et al. 1998; Scozzari et al. 1999; Underhill et al. 2001b). Here we propose that haplotype 41, which reaches a frequency of 15% in the southern African Bantu (Underhill et al. 2000), might also have been involved in such an expansion event. This hypothesis is supported by the fact that the mean variance of STR alleles of haplotype 41 chromosomes is higher in the central western Africans than in the southern Khoisan (0.14 and 0.07, respectively), a finding which also holds true for DYS271 haplotypes 22 and 24.

The impact of the Bantu expansion on pre-existing hunter-gatherer communities was also appreciable. The contribution of Bantu-speaking peoples to the male-specific gene pool of the Pygmies is >50%, and a similar degree of admixture is detected also in the Khoisan-speaking !Kung (45%) and Khwe (58%). These Y-chromosome data agree with mtDNA data showing a higher “Bantu component” in the Khwe than in the !Kung (Chen et al. 2000), and they also correlate with the physical appearance of the former (Hiernaux 1974). However, the impact of the Bantu on the hunter-gatherer communities could have been less extreme in other southern African regions, as is possibly indicated by the 17% of Bantu chromosomes observed in the composite Khoisan sample analyzed by Underhill et al. (2000).

Although haplotypes 22, 24, and 41 were probably all involved in the Bantu expansion, the processes that determined the current distribution of these haplotypes in the Sudanese belt (a region south of the Sahara extending from western to central Africa) seem to have been more complex and perhaps involved a separate expansion. In particular, haplotype 24 and its derivative, haplotype 22, harbor opposite clinal distributions in the region, a finding that is at odds with the hypothesis of a parallel dispersion of these two lineages in the area. Haplotype 22 has a frequency of 23% in Cameroon (where it represents 42% of haplotypes carrying the DYS271 mutation), 13% in Burkina Faso (16% of haplotypes carrying the DYS271 mutation) and only 1% in Senegal (Semino et al. 2002), whereas haplotype 24 reaches its highest frequency (81%) in Senegal (Semino et al. 2002). A possible explanation might be that haplotype 24 chromosomes were already present across the Sudanese belt when the M191 mutation, which defines haplotype 22, arose in central western Africa. Only then would a later demic expansion have brought haplotype

22 chromosomes from central western to western Africa, giving rise to the opposite clinal distributions of haplotypes 22 and 24.

Khoisan Origins and Genetic Affinities

The Khoisan people exhibit a number of characteristics, such as light skin color, female steatopygia and macronymphia, and the presence of “click” sounds in their language, that make them markedly distinct from neighboring Bantu groups and other sub-Saharan Africans (Hiernaux 1974; Cavalli-Sforza et al. 1994). Genetic analyses based on autosomal (Excoffier et al. 1987; Cavalli-Sforza et al. 1994) and mtDNA (Watson et al. 1996; Chen et al. 2000) markers also showed that the Khoisan represent an outlier group in the context of sub-Saharan Africa, a finding confirmed by our correspondence analysis (fig. 3B). If we exclude the recent introduction of Bantu chromosomes (see above), the current Khoisan Y chromosomes fall into only four distinct subsets: those carrying the M51 (haplotype 4) and M14 (haplotypes 6 and 7) mutations, which are Khoisan-specific, and those characterized by the M112 (haplotype 15) and M35 (haplotype 35) mutations, which are shared with Pygmies and Ethiopians, respectively. These haplotypes are not closely related and coalesce to the root of the Y-chromosome phylogeny, suggesting that, as previously noted (Scozzari et al. 1999), the gene pool of modern Khoisan people is the result of several admixture events, ending with Bantu population-mediated gene flow. Of note is the observation that the M51 lineage found in the Khoisan, which represents 26% of the Khoisan chromosomes analyzed to date, and the M13 lineage, which is found at a high frequency in Ethiopia (41%, including data from Underhill et al. [2000]), are united by M220, which is indicative of a shared common paternal ancestry. These findings, along with the sharing of haplotype 35 (table 2 and fig. 2), suggest a certain degree of ancient genetic affinity between Khoisan and Ethiopians. Hypotheses about the presence of some ancestors of modern Khoisan in eastern Africa have been made on the basis of Khoisan-like skeletal materials found in eastern and northeastern Africa (Bräuer 1978; Tobias 1978) and on the basis of linguistic affinities with some modern eastern African populations that also use “clicks” in their languages (Greenberg 1963). Such a scenario is reinforced by the observation that haplotype 28, from which haplotype 35 is derived, and haplotype 5, which is phylogenetically the closest to haplotypes 2 and 4, have been found in Ethiopia but not in southern Africa. Although these data support the sharing of an ancestral gene pool between Khoisan and Ethiopians, the high divergence of haplotypes carrying the M13 and M51 mutations (fig. 1, and the large number of mutations observed in a more complete tree than that in fig.

1 [L.L.C.-S., unpublished data]) and the extensive interpopulation STR diversity observed within haplotype 35 (fig. 4E) indicate that the Ethiopian and Khoisan Y-chromosome components have been separated for a considerable period of time.

Group IX Chromosomes in Sub-Saharan Africa: An Asian Origin?

In sub-Saharan Africa, the majority of the haplotypes fall within one of three groups (groups I–III in fig. 1) sharing the ancestral allelic state at the M89 locus. In contrast, the majority of non-African chromosomes carry the derived allele at this locus (groups VI–X in fig. 1). The only notable exception in sub-Saharan Africa is represented by a set of chromosomes that harbors the M207 and M173 mutations (haplotype 117 in fig. 1) and is found in different linguistic groups of northern Cameroon, at an average frequency of 40%. These two mutations define all members of group IX (haplotypes 110–123 in the study by Underhill et al. [2001b]), a haplogroup that shares the M9 mutation with haplogroups VII, VIII, and X (haplotypes 81–96, 97–109, and 124–131, respectively, in the study by Underhill et al. [2001b]). So far, all group IX chromosomes from Europe and the Middle East that we analyzed were found to carry either the M269 mutation (haplotype 117b and derivatives) or the SRY₁₀₈₃₁ and M17 mutations (haplotypes 119–122). Both of these groups of haplotypes are very common in western Eurasia but harbor opposite frequency clines (R.S., unpublished data; P.A.U., unpublished data). Haplotypes carrying the SRY₁₀₈₃₁ mutation are not restricted to western Eurasia but are also common in central, northern, and southern Asia (Hammer et al. 2001; Karafet et al. 2001). Exclusive to Asia is a third group IX lineage, characterized by the M73 mutation (haplotype 123). The fourth and last group IX lineage identified so far is represented by the Cameroonian haplotype 117, which lacks any known downstream mutations.

How can the presence of Group IX chromosomes at considerable frequency in Cameroon be explained? A priori, we can envision three possibilities. First, group IX chromosomes in Cameroon are due to rather recent male gene flow from Europe or the Near East. Second, the entire M9 superclade (haplogroups VII–X) has an African origin. Third, group IX chromosomes in Cameroon represent a footprint of a male back migration from Asia to Africa. The first scenario seems to be very unlikely, because only derived haplotypes, carrying the M269 or M17/SRY₁₀₈₃₁ mutations, have been detected in western Eurasia. The second hypothesis, an African origin of the M9 superclade that includes haplotype 117, would imply a subsequent impressive extinction of derivative lineages in sub-Saharan Africa, since no other

haplotypes carrying the M9 mutation (haplogroups VII–X) have been observed in this region (the only exception being represented by a few haplotype 109 chromosomes found in the Fulbe from Cameroon). The last scenario, that of a back migration from Asia to Africa, currently appears to be by far the most plausible. This is because most of the M9 haplotypes (the majority of group VII and VIII lineages, as well as some group IX and X lineages reported by Underhill et al. [2000]) have been observed only in Asia. Moreover, this possibility appears to be further supported by the recent finding of the UTY2+/M173– intermediate haplotype (Karafet et al. 2001) in central and northeastern Asia (the UTY2 marker in the study by Karafet et al. [2001] corresponds to M207 in the present study).

On the basis of phylogeographic Y-haplotype analyses, Asia has been regarded as the source of several old migrations leading to the peopling of America, Oceania, and Europe (Karafet et al. 1999; Santos et al. 1999; Hammer et al. 2001; Underhill et al. 2001b; Wells et al. 2001; Lell et al. 2002). In particular, M173-bearing chromosomes in Europe are considered to delineate an ancient expansion from Asia during the Upper Paleolithic, ~30,000 years ago (Semino et al. 2000; Underhill et al. 2001b; Wells et al. 2001). It is quite reasonable to hypothesize that an ancient Asian gene pool was the source of both the European (haplotype 117b) and Cameroonian (haplotype 117) M173 chromosomes. The fact that haplotype 117 is rare or absent in Asia (P.A.U., unpublished data) or the Middle East (present study), suggests that a large portion of its microsatellite diversity in Cameroon accumulated within the African continent after the proposed back-migration event, probably as a consequence of a population expansion. The coalescence age of the African haplotype 117, which we estimated as 4,100 years (95% CI 2,400–8,060 years), could thus represent a date for such an expansion and a lower limit for the time of entry into Africa. The occurrence of the latter event may not necessarily be recent. Although anthropological evidence indicates recent movements between western Asia and Africa by pastoralists (Cavalli-Sforza et al. 1994), the phylogeography and diversity patterns of M173-associated lineages suggest an earlier demographic history. The absence in northern Cameroon of Y haplotypes affiliated with the recolonization of Europe following the Last Glacial Maximum, as well as the subsequent Neolithic transition (Semino et al. 2000), is consistent with this interpretation. Interestingly, phylogenetic analysis of primate T-cell lymphotropic viruses type 1 indicate a putative Asian origin (Vandamme et al. 1998) followed by a simian- or human-mediated introduction to Africa $\geq 20,000$ years ago (Van Dooren et al. 2001).

An ancient human back migration from Asia to Africa had already been proposed by Altheide and Hammer

(1997) and Hammer et al. (1998, 2001), on the basis of nested cladistic analysis of Y-chromosome data. They suggested that the presence of YAP⁺ chromosomes in Africa was due to such an event, but this has recently been questioned by Underhill et al. (2001b) and Underhill and Roseman (2001), primarily on the basis of the Asian-specific YAP⁺ subclade that neutralizes the previous phylogenetic inferences. Thus, the only evidence of a migration from Asia to sub-Saharan Africa that is fully supported by Y-chromosome data relies, at least for the moment, on the finding of haplogroup IX chromosomes in Cameroon.

Interestingly, a frequency of 13% has been previously reported in an Egyptian sample for a group of chromosomes defined as haplotype 1C (Scozzari et al. 1999) and closely related to the M173 chromosomes. Unfortunately, this sample was not available for the present study. Although we cannot define more precisely the haplotype of the Egyptian 1C Y chromosomes, it is worth noting that four of six of these chromosomes showed dinucleotide microsatellite haplotypes that matched or were one-step neighbors of the M173 chromosomes found in Cameroon.

The genetic uniqueness of the northern Cameroon populations outlined here is based entirely on Y-chromosome evidence. It is desirable that additional markers are examined to provide a complement to the Y-chromosome data. In particular, an mtDNA analysis might help to evaluate possible sex-specific differences in migratory behavior.

Further Considerations and Future Research Directions

The present study reports the most extensive survey of Y-chromosome diversity in Africa, in terms of number of markers and populations, and has allowed us to initiate the disentangling of some of the emerging patterns of its complex variation. However, several areas of the continent have not been yet covered, including large portions of the Saharan/Sahelian belt (Niger and Chad), northern Africa (from Tunisia to Egypt), and eastern Africa south of Ethiopia (Kenya and neighboring nation states). The analysis of these areas is necessary for an understanding of the origin and the present distribution of several interesting lineages, for which we have only partial information, to date. In particular, the analysis of areas around northern Cameroon could help to better define both the origin of group IX haplotype 117 and the distribution of the M13-bearing lineages. The presence of M13 chromosomes in central western Africa seems to indicate a migration from eastern Africa to central western Africa that, on the basis of STR data, could be rather ancient (see network in fig. 4A). Also, the analysis of additional populations from eastern and northern Africa could shed some light on the origin and

dispersion of the M78 chromosomes observed at high frequencies in both regions. In sub-Saharan Africa, some haplotypes observed at a low frequency could represent important signatures of pre-Neolithic settlements that have been overwhelmed by the strong demographic impact of farmers. The significance of these low-frequency haplotypes can be evaluated only by surveying much more consistent population samples. At this stage, we can only tentatively propose that group II haplotypes carrying the M236/288 mutations (haplotypes 19 and 19b), as well as the chromosomes carrying the M33 mutation (haplotypes 42 and 43), might represent some of these pre-Neolithic candidates. Finally, although the current Y-chromosome phylogeny fully supports the “out of Africa” model (Underhill et al. 2000, 2001b; Hammer et al. 2001), the possibility of more precisely localizing the geographic origin of modern humans within the continent awaits more genetic—as well as nongenetic—information.

Acknowledgments

We are indebted to Gabriella Spedini (University of Rome “La Sapienza,” Department of Human and Animal Biology), the coordinator of a long-term genetic and epidemiological survey in Cameroon for the samples from this region. We gratefully acknowledge the National Laboratory of Israeli Populations for the Ethiopian Jewish sample. This research was supported by Progetto Finalizzato “Beni Culturali” (Cultural Heritage, Consiglio Nazionale delle Ricerche, Italy) and Grandi Progetti Ateneo Università di Roma “La Sapienza” (both to R.S.), Consiglio Nazionale delle Ricerche grant 99.02620.CT04 (to A.T.), a Fulbright fellowship from the U.S. Department of State and the Italian Ministry of Foreign Affairs (to F.C.), Italian Ministry of the University Cofin grant MM05038334 (to G.D.B. and V.C.), and a Fogerty National Institutes of Health grant (to A.O. and D.C.W). P.A.U. was supported by National Institutes of Health grant 28428 (to L.L.C.-S.). V.M. is a Wellcome Trust Research Career Development Fellow.

Electronic-Database Information

The accession number and URLs for data in this article are as follows:

Arlequin’s Home on the Web, <http://anthro.unige.ch/arlequin/> (for Arlequin version 2.000 software)
 GenBank, <http://www.ncbi.nlm.nih.gov/Genbank/> (for cosmid cAMF3.1 [accession number X96421.1])
 Laboratory of Molecular Anthropology, <http://www.scienzemfn.uniroma1.it/labantro/index.html>
 Life Sciences and Engineering Technology Solutions, <http://www.fluxus-engineering.com/> (for Network version 2e software)

References

- Altheide TK, Hammer MF (1997) Evidence for a possible Asian origin of YAP⁺ Y chromosomes. *Am J Hum Genet* 61:462–466
- Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annu Rev Ecol Syst* 18:489–522
- Ayub Q, Mohyuddin A, Qamar R, Mazhar K, Zerjal T, Mehdi SQ, Tyler-Smith C (2000) Identification and characterisation of novel human Y-chromosomal microsatellites from sequence database information. *Nucleic Acids Res* 28:e8
- Bandelt H-J, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37–48
- Bosch E, Calafell F, Comas D, Oefner PJ, Underhill PA, Bertranpetit J (2001) High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between north-western Africa and the Iberian Peninsula. *Am J Hum Genet* 68:1019–1029
- Bräuer G (1978) The morphological differentiation of anatomically modern man in Africa, with special regard to recent finds from East Africa. *Z Morphol Anthropol* 69: 266–292
- Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. *Nature* 325:31–36
- Capelli C, Wilson JF, Richards M, Stumpf MPH, Gratrix F, Oppenheimer S, Underhill P, Pascali VL, Ko T-M, Goldstein DB (2001) A predominantly indigenous paternal heritage for the Austronesian-speaking peoples of insular Southeast Asia and Oceania. *Am J Hum Genet* 68:432–443
- Casanova M, Leroy P, Boucekkinne C, Weissenbach J, Bishop C, Fellous M, Purrello M, Fiori G, Siniscalco M (1985) A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance. *Science* 230: 1403–1406
- Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton, NJ
- Chen Y-S, Olckers A, Schurr TG, Kogelnik AM, Huoponen K, Wallace DC (2000) mtDNA variation in the South African !Kung and Khwe—and their genetic relationships to other African populations. *Am J Hum Genet* 66:1362–1383
- Dorit RL, Akashi H, Gilbert W (1995) Absence of polymorphism at the ZFY locus on the human Y chromosome. *Science* 268:1183–1185
- Excoffier L, Pellegrini B, Sanchez-Mazas A, Simon C, Langaney A (1987) Genetics and history of sub-Saharan Africa. *Yearb Phys Anthropol* 30:151–194
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479–491
- Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. *Genetics* 139:463–471
- González-Neira A, Elmoznino M, Lareu MV, Sánchez-Diz P, Gusmão L, Prinz M, Carracedo A (2001) Sequence structure

- of 12 novel Y chromosome microsatellites and PCR amplification strategies. *Forensic Sci Int* 122:19–26
- Greenberg JH (1963) *The languages of Africa*. Indiana University, Bloomington
- (1987) *Classificazione delle lingue Africane*. In Kizzerbo J (ed) *Storia generale dell’Africa*. Vol 1: Metodologia e preistoria dell’Africa. Jaca Book, Milan, pp 313–332
- Grimes BF, Grimes JE (2000) *Ethnologue: languages of the world*, 14th ed. Summer Institute of Linguistics, Dallas
- Hammer MF (1995) A recent common ancestry for human Y chromosomes. *Nature* 378:376–378
- Hammer MF, Horai S (1995) Y chromosomal DNA variation and the peopling of Japan. *Am J Hum Genet* 56: 951–962
- Hammer MF, Karafet T, Rasanayagam A, Wood ET, Altheide TK, Jenkins T, Griffiths RC, Templeton AR, Zegura SL (1998) Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. *Mol Biol Evol* 15: 427–441
- Hammer MF, Karafet TM, Redd AJ, Jarjanazi H, Santachiara-Benerecetti S, Soodyall H, Zegura SL (2001) Hierarchical patterns of global human Y-chromosome diversity. *Mol Biol Evol* 18:1189–1203
- Hammer MF, Redd AJ, Wood ET, Bonner MR, Jarjanazi H, Karafet T, Santachiara-Benerecetti S, Oppenheim A, Jobling MA, Jenkins T, Ostrer H, Bonn -Tamir B (2000) Jewish and Middle Eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes. *Proc Natl Acad Sci USA* 97:6769–6774
- Hammer MF, Spurdle AB, Karafet T, Bonner MR, Wood ET, Novelletto A, Malaspina P, Mitchell RJ, Horai S, Jenkins T, Zegura SL (1997) The geographic distribution of human Y chromosome variation. *Genetics* 145:787–805
- Hiernaux J (1974) *The people of Africa*. Weidenfeld and Nicolson, London
- Horai S, Hayasaka K, Kondo R, Tsugane K, Takahata N (1995) Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proc Natl Acad Sci USA* 92:532–536
- Ingman M, Kaessmann H, P  bo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. *Nature* 408:708–713
- International SNP Map Working Group, The (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409:928–933
- Jakubiczka S, Arnemann J, Cooke HJ, Krawczak M, Schmidtke J (1989) A search for restriction fragment length polymorphism on the human Y chromosome. *Hum Genet* 84:86–88
- Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Seielstad MT, Batzer MA (2000) The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *Am J Hum Genet* 66:979–988
- Karafet T, Xu L, Du R, Wang W, Feng S, Wells RS, Redd AJ, Zegura SL, Hammer MF (2001) Paternal population history of East Asia: sources, patterns, and microevolutionary processes. *Am J Hum Genet* 69:615–628
- Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, Harihara S, de Knijff P, Wiebe V, Griffiths RC, Templeton AR, Hammer MF (1999) Ancestral Asian source(s) of new world Y-chromosome founder haplotypes. *Am J Hum Genet* 64:817–831
- Kayser M, Brauer S, Weiss G, Schiefenh  vel W, Underhill PA, Stoneking M (2001) Independent histories of human Y chromosomes from Melanesia and Australia. *Am J Hum Genet* 68:173–190
- Kayser M, Cagli  A, Corach D, Fretwell N, Gehrig C, Graziosi G, Heidorn F, et al (1997) Evaluation of Y-chromosomal STRs: a multicenter study. *Int J Legal Med* 110:125–133
- Kayser M, Roewer L, Hedman M, Henke L, Henke J, Brauer S, Kr  ger C, Krawczak M, Nagy M, Dobosz T, Szibor R, de Knijff P, Stoneking M, Sajantila A (2000) Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. *Am J Hum Genet* 66:1580–1588
- Ke Y, Su B, Song X, Lu D, Chen L, Li H, Qi C, Marzuki S, Deka R, Underhill P, Xiao C, Shriver M, Lell J, Wallace D, Wells RS, Seielstad M, Oefner P, Zhu D, Jin J, Huang W, Chakraborty R, Chen Z, Jin L (2001) African origin of modern humans in East Asia: a tale of 12,000 Y chromosomes. *Science* 292:1151–1153
- Lahr MM, Foley RA (1998) Towards a theory of modern human origins: geography, demography, and diversity in recent human evolution. *Yearb Phys Anthropol* 41:137–176
- Lell JT, Sukernik RI, Starikovskaya YB, Su B, Jin L, Schurr TG, Underhill PA, Wallace DC (2002) The dual origin and Siberian affinities of Native American Y chromosomes. *Am J Hum Genet* 70:192–206
- Malaspina P, Ciminelli BM, Viggiano L, Jodice C, Cruciani F, Santolamazza P, Sellitto D, Scozzari R, Terrenato L, Rocchi M, Novelletto A (1997) Characterization of a small family (CAIII) of microsatellite-containing sequences with X-Y homology. *J Mol Evol* 44:652–659
- Malaspina P, Cruciani F, Ciminelli BM, Terrenato L, Santolamazza P, Alonso A, Banyko J, Brdicka R, Garc a O, Gaudiano C, Guanti G, Kidd KK, Lavinha J, Avila M, Mandich P, Moral P, Qamar R, Mehdi SQ, Ragusa A, Stefanescu G, Caraghin M, Tyler-Smith C, Scozzari R, Novelletto A (1998) Network analyses of Y-chromosomal types in Europe, northern Africa, and western Asia reveal specific patterns of geographic distribution. *Am J Hum Genet* 63:847–860
- Malaspina P, Persichetti F, Novelletto A, Iodice C, Terrenato L, Wolfe J, Ferraro M, Pranita G (1990) The human Y chromosome shows a low level of DNA polymorphism. *Ann Hum Genet* 54:297–305
- Malaspina P, Tsopanomichalou M, Duman T, Stefan M, Silvestri A, Rinaldi B, Garcia O, Giparaki M, Plata E, Kozlov AI, Barbujani G, Vernesi C, Papola F, Ciavarella G, Kovatchev D, Kerimova MG, Anagnou N, Gavrila L, Veneziano L, Akar N, Loutradis A, Michalodimitrakis EN, Terrenato L, Novelletto A (2001) A multistep process for the dispersal of a Y chromosomal lineage in the Mediterranean area. *Ann Hum Genet* 65:339–349
- Mathias N, Bay s M, Tyler-Smith C (1994) Highly informative compound haplotypes for the human Y chromosome. *Hum Mol Genet* 3:115–123
- Melton T, Ginther C, Sensabaugh G, Soodyall H, Stoneking M (1997) Extent of heterogeneity in mitochondrial DNA of sub-Saharan African populations. *J Forensic Sci* 42:582–592

- Ngo KY, Vergnaud G, Johnsson C, Lucotte G, Weissenbach J (1986) A DNA probe detecting multiple haplotypes of the human Y chromosome. *Am J Hum Genet* 38:407–418
- Passarino G, Semino O, Quintana-Murci L, Excoffier L, Hammer M, Santachiara-Benerecetti AS (1998) Different genetic components in the Ethiopian population, identified by mtDNA and Y-chromosome polymorphisms. *Am J Hum Genet* 62:420–434
- Penny D, Steel M, Waddell PJ, Hendy MD (1995) Improved analyses of human mtDNA sequences support a recent African origin for *Homo sapiens*. *Mol Biol Evol* 12:863–882
- Quintana-Murci L, Semino O, Bandelt H-J, Passarino G, McElreavey K, Santachiara-Benerecetti AS (1999) Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. *Nat Genet* 23:437–441
- Ritte U, Neufeld E, Broit M, Shavit D, Motro U (1993a) The differences among Jewish communities—maternal and paternal contributions. *J Mol Evol* 37:435–440
- Ritte U, Neufeld E, Prager EM, Gross M, Hakim I, Khatib A, Bonn -Tamir B (1993b) Mitochondrial DNA affinity of several Jewish communities. *Hum Biol* 65:359–385
- Roewer L, Kayser M, Dieltjes P, Nagy M, Bakker E, Krawczak M, de Knijff P (1996) Analysis of molecular variance (AMOVA) of Y-chromosome-specific microsatellites in two closely related human populations. *Hum Mol Genet* 5:1029–1033
- Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, Amorim A, Amos W, et al (2000) Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am J Hum Genet* 67:1526–1543
- Santos FR, Pandya A, Tyler-Smith C, Pena SDJ, Schanfield M, Leonard WR, Osipova L, Crawford MH, Mitchell RJ (1999) The central Siberian origin for native American Y chromosomes. *Am J Hum Genet* 64:619–628
- Schneider S, Roessli D, Excoffier L (2000) Arlequin ver.2.000: a software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland
- Scozzari R, Cruciani F, Malaspina P, Santolamazza P, Ciminelli BM, Torroni A, Modiano D, Wallace DC, Kidd KK, Olckers A, Moral P, Terrenato L, Akar N, Qamar R, Mansoor A, Mehdi SQ, Meloni G, Vona G, Cole DEC, Cai W, Novelletto A (1997) Differential structuring of human populations for homologous X and Y microsatellite loci. *Am J Hum Genet* 61:719–733
- Scozzari R, Cruciani F, Pangrazio A, Santolamazza P, Vona G, Moral P, Latini V, Varesi L, Memmi MM, Romano V, De Leo G, Gennarelli M, Jaruzelska J, Villems R, Parik J, Macaulay V, Torroni A (2001) Human Y-chromosome variation in the western Mediterranean area: implications for the peopling of the region. *Hum Immunol* 62:871–884
- Scozzari R, Cruciani F, Santolamazza P, Malaspina P, Torroni A, Sellitto D, Arredi B, Destro-Bisol G, De Stefano G, Rickards O, Martinez-Labarga C, Modiano D, Biondi G, Moral P, Olckers A, Wallace DC, Novelletto A (1999) Combined use of biallelic and microsatellite Y-chromosome polymorphisms to infer affinities among African populations. *Am J Hum Genet* 65:829–846 (erratum: 66:346)
- Scozzari R, Torroni A, Semino O, Sirugo G, Brega A, Santachiara-Benerecetti AS (1988) Genetic studies on the Senegal population. I. Mitochondrial DNA polymorphisms. *Am J Hum Genet* 43:534–544
- Seielstad MT, Hebert JM, Lin AA, Underhill PA, Ibrahim M, Vollrath D, Cavalli-Sforza LL (1994) Construction of human Y-chromosomal haplotypes using a new polymorphic A to G transition. *Hum Mol Genet* 3:2159–2161
- Semino O, Passarino G, Brega A, Fellous M, Santachiara-Benerecetti AS (1996) A view of the Neolithic demic diffusion in Europe through two Y chromosome-specific markers. *Am J Hum Genet* 59:964–968
- Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcik  M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill PA (2000) The genetic legacy of paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. *Science* 290:1155–1159
- Semino O, Santachiara-Benerecetti AS, Falaschi F, Cavalli-Sforza LL, Underhill PA (2002) Ethiopians and Khoisan share the deepest clade of the human Y-chromosome phylogeny. *Am J Hum Genet* 70:265–268
- Shen P, Wang F, Underhill PA, Franco C, Yang W-H, Roxas A, Sung R, Lin AA, Hyman RW, Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ (2000) Population genetic implications from sequence variation in four Y chromosome genes. *Proc Natl Acad Sci USA* 97:7354–7359
- Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139:457–462
- Spedini G, Destro-Bisol G, Mondov  S, Kaptu  L, Taglioli L, Paoli G (1999) The peopling of sub-Saharan Africa: the case study of Cameroon. *Am J Phys Anthropol* 110:143–162
- Su B, Xiao J, Underhill P, Deka R, Zhang W, Akey J, Huang W, Shen D, Lu D, Luo J, Chu J, Tan J, Shen P, Davis R, Cavalli-Sforza L, Chakraborty R, Xiong M, Du R, Oefner P, Chen Z, Jin L (1999) Y-chromosome evidence for a northward migration of modern humans into Eastern Asia during the Last Ice Age. *Am J Hum Genet* 65:1718–1724
- Thomas MG, Parfitt T, Weiss DA, Skorecki K, Wilson JF, le Roux M, Bradman N, Goldstein DB (2000) Y chromosomes traveling south: the Cohen modal haplotype and the origins of the Lemba—the “Black Jews of Southern Africa.” *Am J Hum Genet* 66:674–686
- Thomson R, Pritchard JK, Shen P, Oefner PJ, Feldman MW (2000) Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. *Proc Natl Acad Sci USA* 97:7360–7365
- Tobias PV (1978) *The Bushmen: San hunters and herders of Southern Africa*. Human and Rousseau, Cape Town
- Torroni A, Rengo C, Guida V, Cruciani F, Sellitto D, Coppa A, Calderon FL, Simionati B, Valle G, Richards M, Macaulay V, Scozzari R (2001) Do the four clades of the mtDNA haplogroup L2 evolve at different rates? *Am J Hum Genet* 69:1348–1356
- Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res* 7:996–1005
- Underhill PA, Jin L, Zemans R, Oefner PJ, Cavalli-Sforza LL (1996) A pre-Columbian Y chromosome-specific transition

- and its implications for human evolutionary history. *Proc Natl Acad Sci USA* 93:196–200
- Underhill PA, Passarino G, Lin AA, Marzuki S, Oefner PJ, Cavalli-Sforza LL, Chambers GK (2001a) Maori origins, Y-chromosome haplotypes and implications for human history in the Pacific. *Hum Mutat* 17:271–280
- Underhill PA, Passarino G, Lin AA, Shen P, Lahr MM, Foley RA, Oefner PJ, Cavalli-Sforza LL (2001b) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann Hum Genet* 65:43–62
- Underhill PA, Roseman CC (2001) The case for an African rather than an Asian origin of the human Y-chromosome YAP insertion. In: Jin L, Seielstad M, Xiao C (eds) *Recent advances in human biology, vol. 8: genetic, linguistic and archaeological perspectives on human diversity in Southeast Asia*. World Scientific Publishing, New Jersey, pp 43–56
- Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonn -Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ (2000) Y chromosome sequence variation and the history of human populations. *Nat Genet* 26:358–361
- Van Dooren S, Salemi M, Vandamme A-M (2001) Dating the origin of the African human T-cell lymphotropic virus type-I (HTLV-I) subtypes. *Mol Biol Evol* 18:661–671
- Vandamme A-M, Salemi M, Desmyter J (1998) The simian origins of the pathogenic human T-cell lymphotropic virus type I. *Trends Microbiol* 6:477–483
- Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. *Science* 253:1503–1507
- Waddle DM (1994) Matrix correlation tests support a single origin for modern humans. *Nature* 368:452–454
- Walter RC, Buffler RT, Bruggemann JH, Guillaume MMM, Berhe SM, Negassi B, Libsekal Y, Cheng H, Edwards RL, von Cosel R, N raudeau D, Gagnon M (2000) Early human occupation of the Red Sea coast of Eritrea during the last interglacial. *Nature* 405:65–69
- Watson E, Bauer K, Aman R, Weiss G, von Haeseler A, P abo S (1996) mtDNA sequence diversity in Africa. *Am J Hum Genet* 59:437–444
- Watson E, Forster P, Richards M, Bandelt H-J (1997) Mitochondrial footprints of human expansions in Africa. *Am J Hum Genet* 61:691–704
- Weale ME, Yepiskoposyan L, Jager RF, Hovhannisyan N, Khudoyan A, Burbage-Hall O, Bradman N, Thomas MG (2001) Armenian Y chromosome haplotypes reveal strong regional structure within a single ethno-national group. *Hum Genet* 109:659–674
- Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseva I, Blue-Smith J, Jin L, et al (2001) The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc Natl Acad Sci USA* 98:10244–10249
- White PS, Tatum OL, Deaven LL, Longmire JL (1999) New, male-specific microsatellite markers from the human Y chromosome. *Genomics* 57:433–437
- Whitfield LS, Sulston JE, Goodfellow PN (1995) Sequence variation of the human Y chromosome. *Nature* 378:379–380