Am. J. Hum. Genet. 65:829–846, 1999

Combined Use of Biallelic and Microsatellite Y-Chromosome Polymorphisms to Infer Affinities among African Populations

Rosaria Scozzari,[1] Fulvio Cruciani,[1] Piero Santolamazza,[1] Patrizia Malaspina,[4] Antonio Torroni,[1,5] Daniele Sellitto,[1] Barbara Arredi,[1] Giovanni Destro-Bisol,[2] Gianfranco De Stefano,[4] Olga Rickards,[4] Cristina Martinez-Labarga,[4] David Modiano,[3] Gianfranco Biondi,[6] Pedro Moral,[7] Antonel Olckers,[8] Douglas C. Wallace,[9] and Andrea Novelletto[4,10]

Dipartimenti di [1]Genetica e Biologia Molecolare and [2]Biologia Animale e dell'Uomo and [3]Fondazione Pasteur Cenci-Bolognetti, Istituto di Parassitologia, Università "La Sapienza", and [4]Dipartimento di Biologia, Università "Tor Vergata", Rome; [5]Istituto di Chimica Biologica, Università di Urbino, Urbino, Italy; [6]Dipartimento di Biologia Animale e dell'Uomo, Università di Torino, Turin; [7]Departament de Biologia Animal, Universitat de Barcelona, Barcelona; [8]Department of Human Genetics, University of Pretoria, Pretoria, South Africa; [9]Center for Molecular Medicine, Emory University School of Medicine, Atlanta; and [10]Dipartimento di Biologia Cellulare, Università della Calabria, Arcavacata (Cosenza), Italy

Summary

To define Y-chromosome haplotypes, we studied seven biallelic polymorphic sites. We combined data with those from four dinucleotide-repeat polymorphisms, to establish Y-chromosome compound superhaplotypes. Eight biallelic haplotypes that matched the dendrogram proposed by other investigators were identified in 762 Y chromosomes from 25 African populations. For each biallelic site, coalescence time of lineages carrying the derived allele was estimated and compared with previous estimates. The "ancestral" haplotype (haplotype 1A) was observed among Ethiopians, "Khoisan" (!Kung and Khwe), and populations from northern Cameroon. Microsatellite distributions within this haplotype showed that the Khoisan haplotypes 1A are widely divergent from those of the other two groups. Populations from northern Africa and northern Cameroon share a haplotype (i.e., 1C), which is not observed in other African populations but represents a major Eurasian cluster. Haplotypes 1C of northern Cameroon are clearly distinct from those of Europe, whereas haplotypes 1C of northern African are well intermingled with those of the other two groups. Apportionment of diversity for the Y-chromosomal biallelic haplotypes was calculated after populations were clustered into different configurations. Despite some correspondence between language affiliation and genetic similarity, geographic proximity seems to be a better predictor of genetic affinity.

Received January 22, 1999; accepted for publication June 23, 1999; electronically published August 13, 1999.

Address for correspondence and reprints: Dr. Rosaria Scozzari, Dipartimento di Genetica e Biologia Molecolare, Università "La Sapienza", P.le Aldo Moro 5, 00185, Rome, Italy. E-mail: rscozzar@axcasp.caspur.it

0002-9297/1999/6503-0029$02.00

Introduction

An increasing number of human Y chromosome–specific polymorphisms are being used as markers for human evolutionary studies. The male-specific portion of the Y chromosome is transmitted unchanged from fathers to sons, except for intervening mutations that accumulate through generations. This property allows the reconstruction of phylogenetic relationships, without recombination as a confounding factor (Jobling and Tyler-Smith 1995; Whitfield et al. 1995; Hammer et al. 1997, 1998; Underhill et al. 1997), and is highly valuable for evaluation of male-specific contributions to some relevant aspects of human evolution, such as population migrations, admixture, and drift (Seielstad et al. 1998; Karafet et al. 1999). Population studies of Y chromosome–specific p49a,f/*Taq*I restriction polymorphisms have shown a high level of population structuring (Torroni et al. 1990; Poloni et al. 1997) and a strong correspondence with linguistics (Poloni et al. 1997). Analysis of Y-chromosome short tandem-repeat polymorphisms (Scozzari et al. 1997) has also revealed a high degree of population differentiation, with a fundamental difference between (1) non-Africans, who showed a greater intrapopulation variability, and (2) Africans, who showed the highest interpopulation variability. A recent worldwide study of eight Y chromosome–specific single-nucleotide polymorphisms (SNPs) and an *Alu* insertion (i.e., YAP) showed that the 10 haplotypes observed could be arranged in a simple tree, which was rooted by comparison with orthologous positions in the great apes (Hammer et al. 1998). Age estimates for each branch of this tree were provided, on the basis of the infinite-sites model. In addition, the Khoisan from South Africa were shown to be the main depository of the "ancestral" haplotype, giving support to the "out-of-Africa" theory, but evidence of back migrations from Asia to Africa were also found. However, the study by

Hammer et al. (1998) mainly focused on global patterns of frequency and distribution of the Y chromosome variation, and it is possible that the geographic range used to evaluate regional population affinities within Africa was insufficient and that the overall diversity in this continent was not fully discovered.

The present study reports high-resolution data on the haplotype affiliation of >750 Africans from 25 populations. In addition, the internal diversity of each haplotype was assessed by determination of the allele state at four dinucleotide (i.e., CA) microsatellites. This approach provides a new data set for an independent estimation of the coalescence time of the derived allele at each biallelic site, on the basis of the stepwise-mutation model (Goldstein et al. 1996). Moreover, the combination of biallelic and microsatellite polymorphisms proves to be of great value for the analysis of population structuring (Ciminelli et al. 1995; Scozzari et al. 1997). As an extension of our previous results, we find a correspondence between language affiliation and genetic affinity in some groups of populations, but this relationship is often lost with increasing geographic distance. In other groups of populations, geographic proximity, rather than language affiliation, appears to be the major factor contributing to genetic similarities.

Subjects and Methods

Subjects

A total of 762 male subjects belonging to 25 African populations were studied. Several of these population samples were previously analyzed for some subsets of the polymorphisms here described (see table 1 and references therein). New population samples not previously analyzed for Y-chromosome variation include 48 Amhara and 37 Oromo, from Ethiopia; 34 Bariba, 20 Berba, 36 Fon, and 37 Dendi, from Benin; 12 Bakaka, from southwestern Cameroon; and 1 composite sample, from northern Cameroon (comprising 7 Toupouri, 6 Moundang, 4 Mboum, 1 Massa, 6 Guidar, 2 Guiziga, 2 Haoussa, 2 Lele, 2 Podoko, 1 Mafa, 5 Ngambai, 2 Kanouri, and 2 Sara).

Amhara and Oromo are two major Ethiopian ethnic groups. They speak Semitic and Cushitic languages, respectively, both belonging to the Afro-Asiatic linguistic family. Historical information on these two population groups has been provided by Cavalli-Sforza et al. (1994). The Ethiopian subjects included in the present work are part of a sample described and studied for a number of red-blood-cell and serum polymorphic systems (Scacchi et al. 1994; Tartaglia et al. 1996), as well as for HLA class II alleles (Fort et al. 1998). Among populations from Benin, the Bariba, Berba, and Fon speak related languages of the Niger-Congo linguistic family, whereas the language spoken by the Dendi belongs to the Songhai subdivision of the Nilo-Saharan linguistic family. The population samples from Benin have been presented elsewhere (Rickards and Martinez-Labarga 1994; Biondi et al. 1996). The Bakaka from southwestern Cameroon are Bantu-speaking peoples. The Bakaka subjects examined here are part of a sample previously studied for some protein polymorphisms (Destro-Bisol et al. 1992). Among the populations from northern Cameroon, the Toupuri, Moundang, Mboum, and Massa speak languages of the Adamawa subdivision of the Niger-Congo linguistic family; the Guidar, Guiziga, Haoussa, Lele, Podoko, and Mafa speak one of the Chadic languages of the Afro-Asiatic family, whereas the Ngambai, Kanouri, and Sara speak Nilo-Saharan languages. The subjects from northern Cameroon are part of a sample recently analyzed for a number of protein polymorphisms (Spedini et al., in press).

Molecular Analysis

DNA samples were obtained from blood specimens, by phenol-chloroform extraction and ethanol precipitation. The allele state at the SRY_{4064} site was scored by PAMSA (i.e., PCR amplification of multiple specific alleles (Li et al. 1990; Bottema and Sommer 1993). A sequence encompassing the polymorphic site at position 4064 (Whitfield et al. 1995) was PCR-amplified by means of the following primers: 5 pmol of SRY_{4064}-F (5′-CAGCACATTAGCTGGTATGACA-3′), 20 pmol of SRY_{4064}-R (A specific) (5′-TTCACCCTGTGATCCGCTT-3′), and 0.8 pmol of SRY_{4064}-R (G specific) (5′-AAGGACGCGTAATCGTACGCGTTAGTTCACCCTGTGATCCGCTC-3′). PCR was performed in a 12.5-μl volume containing 50 ng of genomic DNA, 1 × PCR buffer, 80 μM of each dNTP, 1.5 mM $MgCl_2$, and 0.2 U of *Taq* DNA polymerase. The cycling conditions were 94°C for 3 min; 30 cycles of 94°C for 45 s, 62°C for 1 min, and 72°C for 1 min; and a final extension at 72°C for 2 min. The resulting fragments (164 bp for the A allele, 189 bp for the G allele) were electrophoresed on 3% NuSieve 3:1 agarose gel.

The SRY_{10831} site (Whitfield et al. 1995) was PCR-amplified by means of the following primers: SRY_{10831}-F (5′-TCTGACTCTTTGGTTCACCA-3′) and SRY_{10831}-R (5′-AAGTGTTGGTTCTCCTGTA-3′). The final 25-μl mix contained 1 × PCR buffer, 0.2 mM of each dNTP, 10 pmol each primer, 2 mM $MgCl_2$, and 0.5 U of *Taq* DNA polymerase. PCR products were digested with the restriction enzyme *Mae*III (Boehringer), under conditions specified by the manufacturer, and the resulting fragments were electrophoresed on 3% NuSieve 3:1 agarose gel. The A→G transition causes the gain of a *Mae*III site, resulting in either a two-fragment (180 bp and 100 bp) or a three-fragment (142 bp, 100 bp, and 38 bp)

Table 1

Populations Sampled

Geographic Region, Country, and Population (Abbreviation)	Linguistic Affiliation[a]		No. of Y Chromosomes	Reference
	Family	Sublevel		
Northern Africa:				
Morocco:				
Arabs (MOR)	Afro-Asiatic	Semitic	56	Malaspina et al. (1998)
Egypt:				
Northern:				
Egyptian (NEG)	Afro-Asiatic	Semitic	25	Persichetti et al. (1992)
Southern:				
Egyptian (SEG)	Afro-Asiatic	Semitic	21	Ciminelli et al. (1995)
Eastern Africa:				
Ethiopia:				
Amhara (AMH)	Afro-Asiatic	Semitic	48	Present study
Oromo (ORO)	Afro-Asiatic	Cushitic	37	Present study
Western Africa:				
Burkina Faso:				
Mossi (MOS)	Niger-Congo	Voltaic	56	Scozzari et al. (1997)
Rimaibe (RIM)	Niger-Congo	West Atlantic	42	Scozzari et al. (1997)
Fulbe (FUB)	Niger-Congo	West Atlantic	20	Scozzari et al. (1997)
Benin:				
Bariba (BAR)	Niger-Congo	Voltaic	34	Present study
Berba (BER)	Niger-Congo	Voltaic	20	Present study
Fon (FON)	Niger-Congo	Kwa	36	Present study
Dendi (DEN)	Nilo-Saharan	Songhai	37	Present study
Central Africa:				
Cameroon:				
Northern:				
Fali (FAL)	Niger-Congo	Adamawa	39	Scozzari et al. (1997)
Tali (TAL)	Niger-Congo	Adamawa	15	Scozzari et al. (1997)
Mixed (OAD)	Niger-Congo	Adamawa	18	Present study
Fulbe (FUC)	Niger-Congo	West Atlantic	17	Scozzari et al. (1997)
Ouldeme (OUL)	Afro-Asiatic	Chadic	23	Scozzari et al. (1997)
Daba (DAB)	Afro-Asiatic	Chadic	18	Scozzari et al. (1997)
Mixed (OCH)	Afro-Asiatic	Chadic	15	Present study
Mixed (ONS)	Nilo-Saharan	Central Sudanic/Saharan	9	Present study
Southern:				
Bamileke (BAM)	Niger-Congo	Benué-Congo, Bantoid	44	Scozzari et al. (1997)
Ewondo (EWO)	Niger-Congo	Benué-Congo, Bantu	30	Scozzari et al. (1997)
Bakaka (BAK)	Niger-Congo	Benué-Congo, Bantu	12	Present study
Southern Africa:				
South Africa:				
!Kung (KUN)	Khoisan	Northern	64	Scozzari et al. (1997)
Khwe (KHW)	Khoisan	Central	26	Scozzari et al. (1997)

[a] According to Grimes (1996) and Greenberg (1987).

pattern. The following polymorphisms were also typed, according to the procedures described in the original references: DYS257 (Hammer et al. 1998), DYS271 (Seielstad et al. 1994), three markers within the YAP region (i.e., the presence/absence of the YAP element, a C→T transition at position 1,682 [PN2], and a G→A transition at position 1,926 [PN3]) (Hammer 1995; Hammer and Horai 1995; Hammer et al. 1997), and the presence/absence of a *Hin*dIII site in centromeric alphoid units (Santos et al. 1995). The allele state for YCAIIa, YCAIIb, DYS413a, and DYS413b microsatellites was scored as described by Mathias et al. (1994) (YCAIIa and YCAIIb), Malaspina et al. (1997), and Scozzari et al. (1997) (DYS413a and DYS413b).

All 762 subjects were typed for the YAP, alphoid/*Hin*dIII, PN2, PN3, YCAIIa, YCAIIb, DYS413a, and DYS413b polymorphisms. The SRY_{10831} polymorphism was studied in all YAP^- and in 367 of 559 YAP^+ subjects. The DYS271 polymorphism was analyzed in all the YAP^+ and in 139 of 203 of the YAP^- subjects. The DYS257 was scored in all the YAP^- subjects and in 24 of 559 YAP^+ subjects. Finally, the SRY_{4064} was scored in all YAP^+ subjects and in 84 of 203 YAP^- subjects. The typing of Moroccan and Egyptian samples for the pres-

ence/absence of the YAP insertion, the presence/absence of the *Hin*dIII site in the alphoid units, and the YCAIIa, YCAIIb, DYS413a, and DYS413b dinucleotide microsatellites has been reported elsewhere (Malaspina et al. 1998). Data for the DYS413a and DYS413b microsatellites of many of the sub-Saharan samples have been previously published by Scozzari et al. (1997).

Statistical Analysis

Indexes of Y-chromosome diversity were obtained by use of the Arlequin package, version 1.1 (Schneider et al. 1997). Haplotype diversity and its sampling variance were calculated as

$$\hat{H} = \frac{n}{n-1}\left(1 - \sum_{i=1}^{k} p_i^2\right)$$

and

$$V(\hat{H}) = \frac{2}{n(n-1)}\left\{2(n-2)\left[\sum_{i=1}^{k} p_i^3 - \left(\sum_{i=1}^{k} p_i^2\right)^2\right] + \sum_{i=1}^{k} p_i^2 - \left(\sum_{i=1}^{k} p_i^2\right)^2\right\},$$

respectively, where p_i is the relative frequency of the ith biallelic haplotype and k is the number of haplotypes.

Arlequin was also used to perform analysis of molecular variance (AMOVA [Excoffier et al. 1992]). Two hierarchical levels (i.e., individuals clustered into populations and populations clustered into groups) were considered. Φ Statistics were then calculated, representing the haplotype correlations at the various levels of the hierarchical groupings—that is, Φ_{CT} (cluster of subpopulations relative to total population), Φ_{SC} (subpopulation relative to cluster of subpopulations), and Φ_{ST} (subpopulation relative to total population). Molecular distances among biallelic haplotypes were computed by counting the number of mutations that separate two haplotypes along the evolutionarily most parsimonious tree (see fig. 1). Significance levels of the components of the variance and of the corresponding Φ statistics were obtained by comparison of the actual values with the distribution of 1,000 values obtained by randomization.

Correspondence analysis was performed with the program ANACOR, version 0.4, included in the SPSS package, version 6.1.3. This method allows one to examine the relationships between two nominal variables in a multidimensional space and is the method of choice for cross-tabulations in which the cells contain frequency counts. We used the matrix of the biallelic haplotype counts with row-principal normalization, since we were primarily interested in the differences and similarities among row categories (populations).

Network Construction and Coalescence Analysis

For each biallelic haplotype (or lineage), networks of "adjacent" (one repeat difference over the four loci) microsatellite haplotypes were constructed. As has been described elsewhere (Cooper et al. 1996; Malaspina et al. 1998), microsatellite haplotypes with the same combined repeat length were placed at the same horizontal level, and the network was constructed by sequential addition of haplotypes differing by a single repeat unit.

For each biallelic site, coalescence time of lineages sharing the derived allele was estimated by calculation of the variance of CA-repeat number averaged over the four microsatellites, with the obtained value then being entered into equation (2) of Goldstein et al. (1996). The relevant haplotypes found in various European and Asian populations, for a total of 472 Y chromosomes (62 Spaniards, 76 continental Italians, 184 Sardinians, 35 Danes, 45 Arabs, 62 Chinese, and 8 Kachari [R. Scozzari, unpublished data]) were also included. Time to the most recent common ancestor was obtained by calculation of the average repeat-number variance for the global sample of 762 African and 472 Eurasian Y chromosomes. The initial variance of $(CA)_n$ was considered to be null. Values of 5.6×10^{-4} (Weber and Wong 1993) and 20 were used for the mutation rate and number of years per generation, respectively. A constant effective population size of 5,000 was used. This figure is close to that obtained by Goldstein et al. (1996) for the male population worldwide. The 95% confidence intervals were calculated by inclusion of the interlocus variance of variances into the equations for V_1 and V_2, reported by Goldstein et al. (1996), followed by recalculation of the corresponding coalescence times.

Results

Six of the SNPs listed in the Subjects and Methods section (SRY_{10831}, SRY_{4064}, PN2, PN3, DYS271, and DYS257), along with the YAP insertion, defined eight Y-chromosome haplotypes (fig. 1). These haplotypes matched the dendrogram proposed by Hammer et al. (1997, 1998), if one assumes that the DYS271 A→G transition identifies the same group of chromosomes as does the C→T transition, at the PN1 site (Hammer 1995; Hammer et al. 1997, 1998). A single instance of a haplotype carrying the DYS257 A and SRY_{10831} A alleles was found among 762 African Y chromosomes; this was identified as haplotype 1D, according to the protocol of Hammer et al. (1998). All haplotypes were also screened

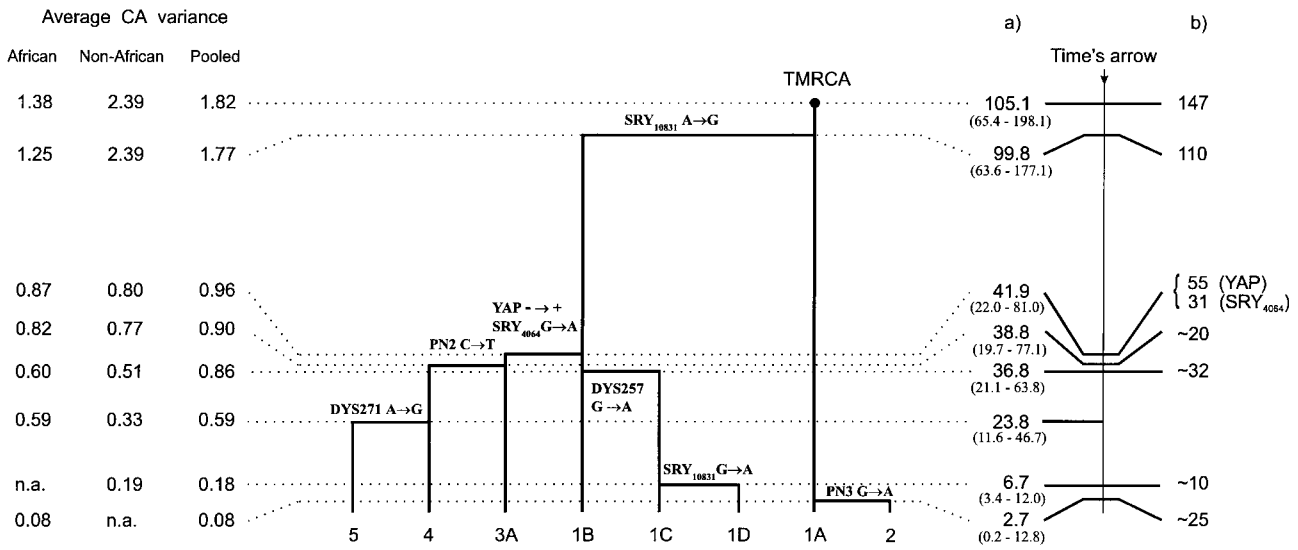

Figure 1 Gene tree of the eight biallelic haplotypes in the sample of 762 African Y chromosomes. Mutations are indicated on the branches, at distances proportional to their estimated age. The values on the left refer to the variance in CA number, averaged over the four microsatellite loci. The figures obtained for the 762 African and the 472 non-African Y chromosomes have been reported separately and then have been pooled. The values on the right (part "a") correspond to the coalescence time (in 1,000s of years), estimated according to the method of Goldstein et al. (1996) (also see the Subjects and Methods section); data in parentheses are 95% confidence intervals. The estimates obtained by Hammer et al. (1998), by a different method, are also indicated (part "b"). n.a. = not applicable.

for the alphoid/*Hin*dIII polymorphism. Haplotypes 2 and 1C contained only *Hin*dIII$^-$ chromosomes, and haplotypes 1A and 3A contained only *Hin*dIII$^+$ chromosomes. However, haplotypes 4, 5, and 1B included a large majority of chromosomes with alphoid units containing the *Hin*dIII site but also contained chromosomes (17 of 586) lacking it. These results are more easily interpreted with the occurrence of *Hin*dIII$^+$→*Hin*dIII$^-$ mutations on three independent lineages. No lineages were found in which a minority of *Hin*dIII$^+$ chromosomes were present on a *Hin*dIII$^-$ background. Previously, only two chromosomes explained by homoplasy of the *Hin*dIII$^-$ state had been detected in 240 chromosomes worldwide (Santos et al. 1996) or in >900 Y chromosomes from Eurasian males (Malaspina et al. 1998). Overall, these data favor the hypothesis that the *Hin*dIII mutation is recurrent but unidirectional, with no instances of chromosomes in which the site reverted to the *Hin*dIII$^+$ state. Such an occurrence, in Africa, of chromosomes with homoplasic states convinced us to disregard the alphoid/*Hin*dIII polymorphism in the subsequent analysis of African populations.

Indexes of Y-chromosome diversity are summarized in table 2, where the populations studied have been aggregated according to a geographic criterion (also see table 1). The highest number of different haplotypes was found in the sub-Saharan area. The highest variation, as measured by both the mean number of pairwise inter-haplotype differences and the haplotype-diversity index, was found in southern Africans, whereas the lowest number of different haplotypes was found in western Africans.

Figure 2 illustrates the frequency distribution of the eight haplotypes identified in this study, as well as the partitioning of the populations into the five geographic areas cited in tables 1 and 2. The haplotype defined by the PN3 mutation (i.e., haplotype 2) was found only in the Khoisan, as had been reported elsewhere (Hammer et al. 1998). "Ancestral" haplotype 1A, with an overall frequency of 6.2%, was shared by Khoisan, Ethiopians (with frequencies >20%), and a few other populations from central Africa. Haplotype 1B, with an overall frequency of 10.1%, was present in all groups but predominated in northern and eastern Africans. The only individual with haplotype 1D was found in Egypt. Haplotype 1C was absent in all sub-Saharan populations, with the exception of those from central Africa, where it reached frequencies as high as 95.7% (Ouldeme). The YAP$^+$ allele (haplotypes 3A, 4, and 5) was shared by most of the groups and was always associated with SRY_{4064} A. Populations from sub-Saharan Africa had appreciable frequencies of YAP$^+$/PN2C chromosomes (haplotype 3A), and those from northern and eastern Africa had the highest frequencies (60.8% and 54.1%, respectively) of the YAP$^+$/PN2T chromosomes (haplotype 4). Haplotype 5 (YAP$^+$/PN2T/DYS271G) was shared by populations from western, central, and southern Africa (at frequencies of 85.3%, 53.8%, and 43.3%,

respectively), but it was virtually absent in the rest of the continent.

To assess the level of population structure, various Φ statistics were estimated (table 3 and Appendix A). The Φ values were computed by means of an AMOVA test (Excoffier et al. 1992), by taking into account the interhaplotype molecular differences, as measured on the dendrogram. The overall Φ_{ST} value calculated for the entire African sample, comprising 25 populations, was .388 ($P < .001$), indicating that a large proportion of the overall Y-chromosomal variation resulted from interpopulation differences. Φ Statistics were then calculated to assess the intergroup variation, after the populations were grouped by either linguistic (Greenberg 1987; Grimes 1996) or geographic criteria. To keep the hierarchical groupings comparable across the tests, we used four linguistic families and five main geographic areas (table 3). When the populations were grouped on the basis of the four linguistic families, the Φ_{CT} value was high and statistically significant. However, Φ_{ST} estimates obtained within each language family indicated that a high and statistically significant level of genetic differentiation still persisted among linguistically homogeneous populations. In the case of the Nilo-Saharan–speaking populations, this was mostly due to the high frequency of haplotype 5 in the Dendi from Benin (86.5%), relative to that in other Nilo-Saharan–speaking populations from northern Cameroon (33.3%). In the case of the Afro-Asiatic speaking populations, haplotype 1C was common in all three Chadic-speaking populations, whereas it was found, at much lower frequencies, in only three of the non-Chadic Afro-Asiatic–speaking populations. An even more extreme case is represented by the distribution of haplotype 5, which was found at a very high frequency (90.0%) in the Fulbe from Burkina Faso but was not observed in the Fulbe from northern Cameroon.

When the 25 populations were grouped on the basis of a geographic criterion, the Φ_{CT} was slightly lower than that obtained on the basis of linguistic grouping (Φ_{CT} = .184; $P < .001$). However, Φ_{ST} values for northern, eastern, and western Africa indicated a low population structuring within these groups. In the case of western Africans, only the Rimaibe could be clearly differentiated from all the other populations (mean pairwise Φ_{ST} vs. other populations, .143; also see Appendix A). On the other hand, a very high level of internal population structuring was detected among populations from central Africa (Φ_{ST} = .463; $P < .001$). Since this region includes populations from the entire territory of Cameroon, a separate analysis for southern- and north-

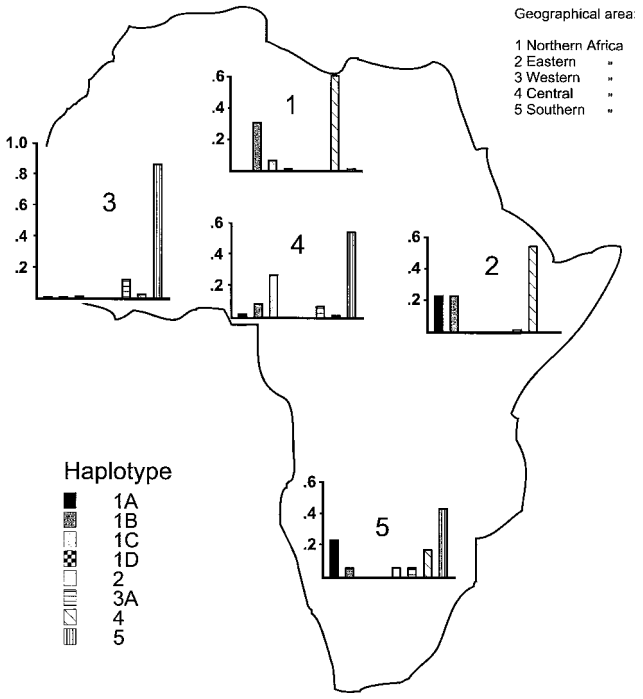

Figure 2 Frequency distribution of the eight Y chromosome biallelic haplotypes, by geographic area. The populations representing the geographic areas are indicated in table 1. Haplotype designations are the same as those used by Hammer et al. (1998), with the assumption that the PN1 C→T and the DYS271 A→G mutations identify the same set of chromosomes.

Table 2

Molecular Indexes of Y-Chromosome Diversity in Africa

Geographic Region of Africa	No. (Designation[a]) of		Mean ± SD of	
	Y Chromosomes	Haplotypes	Pairwise Difference	$\hat{H}$
Northern	102	5 (1B, 1C, 1D, 4, 5)	1.616 ± .964	.539 ± .038
Eastern	85	4 (1A, 1B, 3A, 4)	1.854 ± 1.074	.614 ± .036
Western	245	6 (1A, 1B, 1C, 3A, 4, 5)	.563 ± .461	.261 ± .035
Central	240	6 (1A, 1B, 1C, 3A, 4, 5)	2.384 ± 1.302	.629 ± .025
Southern	90	6 (1A, 1B, 2, 3A, 4, 5)	2.417 ± 1.324	.729 ± .031
Overall	762	8 (1A, 1B, 1C, 1D, 2, 3A, 4, 5)	2.046 ± 1.150	.698 ± .014

[a] Haplotype designation is that of Hammer et al. (1998).

Table 3

Analysis of Molecular Variance for Eight Y Chromosome–Specific Biallelic Haplotypes in Africa

Category (No. of Populations)	Φ_{ST} (*P*)	Φ_{CT} (*P*)	Φ_{SC} (*P*)
Linguistic groups:			
Afro-Asiatic (8)	.248 (.000)		
Nilo-Saharan (2)	.604 (.000)		
Niger-Congo (13)	.318 (.000)		
Khoisan (2)	.135 (.004)		
Overall (25)	.436 (.000)	.229 (.001)	.268 (.000)
Geographic groups:			
Northern Africa (3)	.082 (.017)		
Eastern Africa (2)	.004 (.272)		
Western Africa (7)	.064 (.002)		
Central Africa (11)	.463 (.000)		
Southern Africa (2)	.135 (.004)		
Overall (25)	.411 (.000)	.184 (.000)	.278 (.000)
All Africa (25)	.388 (.000)		

ern-Cameroon population subgroups was also performed. The two subgroups displayed a high degree of intergroup variation (Φ_{CT} = .385; *P* = .018). The homogeneous subgroup of southern Cameroon (Φ_{ST} = .021; *P* = .130) shared, with western African populations, a high frequency of haplotype 5. On the contrary, the subgroup from northern Cameroon retained a high level of internal heterogeneity (Φ_{ST} = .292; *P* < .001). This result was poorly explained by the linguistic fragmentation that characterizes this area; in fact, populations sharing the same language family still displayed large amounts of interpopulation variation (Afro-Asiatic Φ_{ST} = .283 [*P* < .001] and Niger-Congo Φ_{ST} = .219 [*P* < .001]). Thus, it would appear that, at least in this specific geographic and cultural context, language did not represent a major barrier to gene flow.

Correspondence analysis was used to condense the information on the distribution of the eight Y-chromosome haplotypes in Africa. Overall, the first two dimensions capture 71.0% of the total inertia. The first dimension separates northern and eastern Africans from the remaining African populations (fig. 3*A*), because of the high frequency of haplotype 4 (fig. 3*B*) in the former groups (41.8% of the inertia of this dimension) and of haplotype 5 in the latter groups (37.1% of the inertia). The second dimension separates the northern-Cameroon populations from all the rest, because of the distribution of haplotype 1C (87.8% of the inertia).

To disclose the internal diversity of haplotypes and to further investigate their informativeness in relation to the African population structure, we constructed Y-chromosome compound superhaplotypes, on the basis of the seven biallelic polymorphic sites described above and the four dinucleotide microsatellites (YCAIIa, YCAIIb, DYS413a, and DYS413b). Their relative frequencies are shown in Appendix B. In agreement with our previous results from an independent data set (Malaspina et al. 1998), microsatellite haplotype homoplasy was relatively low. In fact, of the 141 microsatellite patterns, 111 (79%) were unique to a single biallelic haplotype, whereas 26 and 4 were found on two and three haplotypes, respectively. None was found on more than three haplotypes. Also in this series, whenever a microsatellite pattern was found on more than one haplotype,

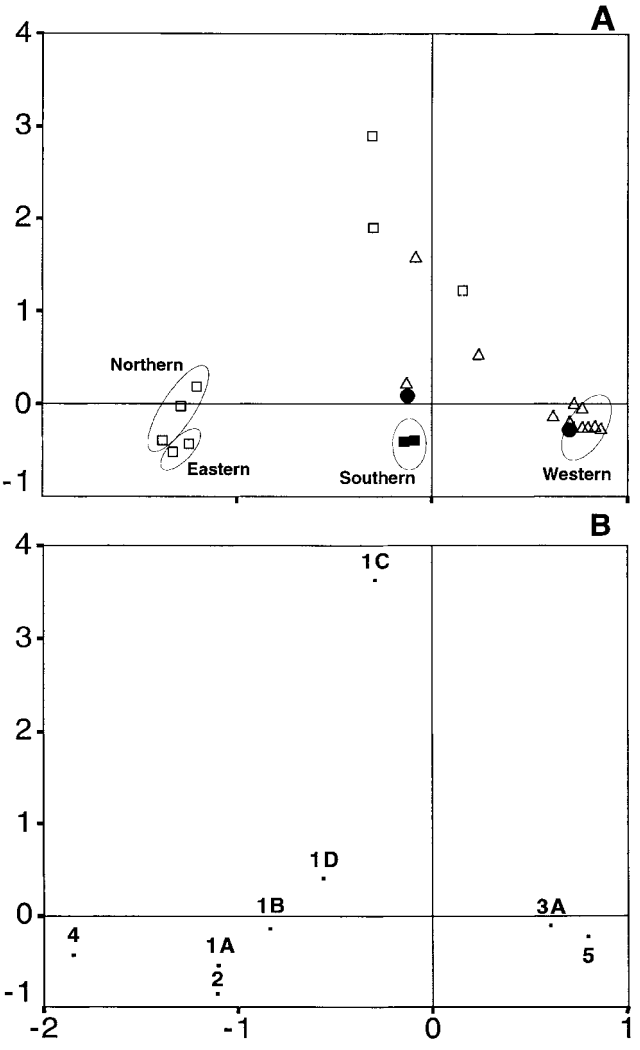

Figure 3 Plot of correspondence analysis scores. *A*, Plot of populations, in the space of the first two dimensions: first dimension (*X*-axis, 40.7% of the total inertia) vs. second dimension (*Y*-axis, 30.3% of the total inertia). Symbols used to identify language families are as follows: unblackened squares, Afro-Asiatic; unblackened triangles, Niger-Congo; blackened circles, Nilo-Saharan; blackened squares, Khoisan. Four of the five geographic groups (see table 1) also are circled. The fifth group (not circled) is widely scattered and is composed of the populations from central Africa. *B*, Plot of haplotype scores. Haplotype nomenclature is that of Hammer et al. (1998).

it either occurred at low frequency on both haplotypes or showed a moderate-to-high frequency on one haplotype only. The single exception was the pattern 22-19-22-22, observed on 13 haplotypes 4 and on 11 haplotypes 1B.

The average variance of CA-repeat number at the four dinucleotide microsatellites was used to calculate coalescence times (see the Subjects and Methods section). In all cases the resulting variance value was well below the estimated asymptotic mutation-drift equilibrium value of 2.52, thus making it feasible to use dinucleotide diversity for such a calculation (Goldstein et al. 1996). The accumulated CA variance and the corresponding estimated coalescence time of the new mutation at each of the seven polymorphic sites are shown in figure 1. The same figure also reports the corresponding coalescence times obtained by Hammer et al. (1998), who used a different method.

For each of the biallelic haplotypes, microsatellite data were fitted onto a network of adjacent haplotypes, as has been discussed elsewhere (Malaspina et al. 1998). The most informative of these networks are illustrated in figures 4–6. Figure 4 shows the network display of ancestral haplotypes 1A. There is a sharp distinction between the Ethiopian and central-African gene pool, on one hand, and the Khoisan gene pool, on the other hand, which are encompassed by two entirely separated subnetworks.

Figure 5 illustrates the network obtained when all haplotypes 1C detected in Africa (mostly in northern and central Africa; see Appendix B) are connected with those found in Europe (see the Subjects and Methods section). The European haplotypes included 127 chromosomes belonging to network 3.1 (YAP-/alphoid-*Hin*dIII-) (Malaspina et al. 1998), which, for the purpose of the present work (R. Scozzari, unpublished data), have also been confirmed to carry PN2C, SRY_{10831} G, SRY_{4064} G, DYS257 A, and DYS271 A (haplotype 1C). Figure 5 shows a pattern comprising two almost-separate sections—one for Europe and the other for central Africa—linked by a single haplotype that is found in both groups and that is placed at the top of the network. Haplotypes of both clusters were found in northern Africans. Figure 6, which illustrates the network for haplotype 5, shows a dramatically different picture, in which almost all major haplotypes of the network are shared by all three sub-Saharan population groups.

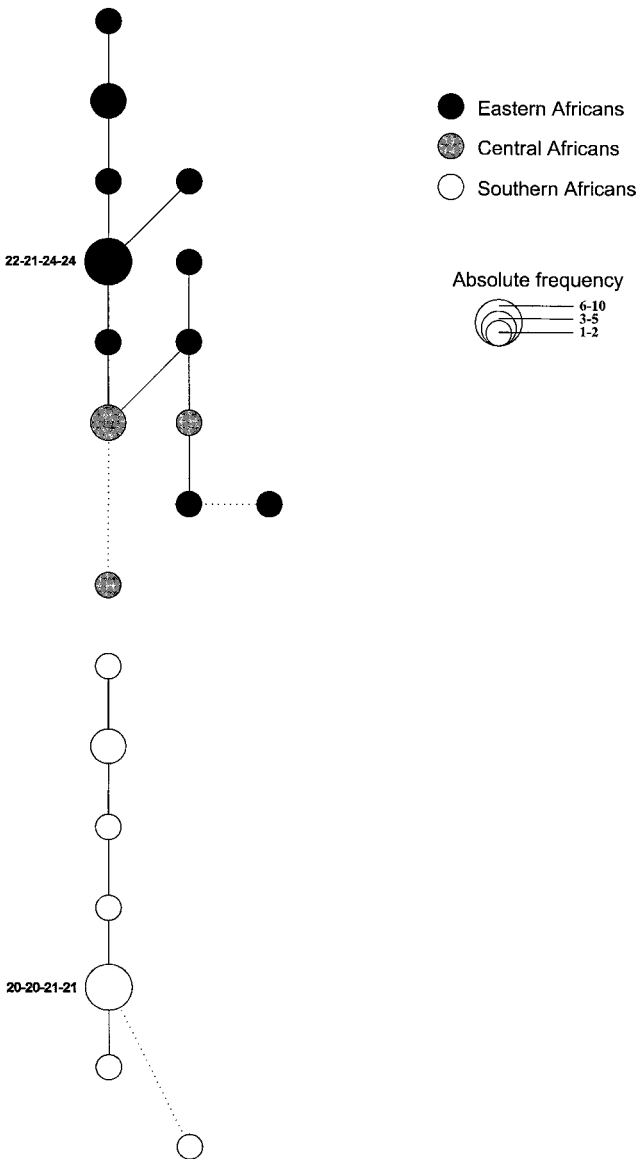

Figure 4 Network of adjacent microsatellite haplotypes within lineage 1A. Two haplotypes could not be placed within either of the two subnetworks. All possible adjacent relationships (see the Subjects and Methods section) are indicated by unbroken connecting lines; dashed lines denote a two-repeat difference. The circled area is proportional to the absolute frequency of the sampled haplotype. Population origin is indicated by degree of gray shading. Allele size (in no. of CA repeats) is reported for selected haplotypes.

Discussion

An increasing number of SNPs and insertion/deletion polymorphisms have been reported for the human Y chromosome (Hammer 1995; Whitfield et al. 1995; Hammer et al. 1997, 1998; Underhill et al. 1997; Karafet et al. 1999; Santos et al. 1999), thus allowing a detailed reconstruction of its sequence evolution and dispersal. We have used seven of these informative polymorphisms to investigate the genetic affinities among African populations. A comparison of our results with those of previous studies shows both congruities and differences. Our data confirm that the mutational events can be arranged, without ambiguity, within the tree first proposed by Hammer et al. (1997, 1998). In our series a single

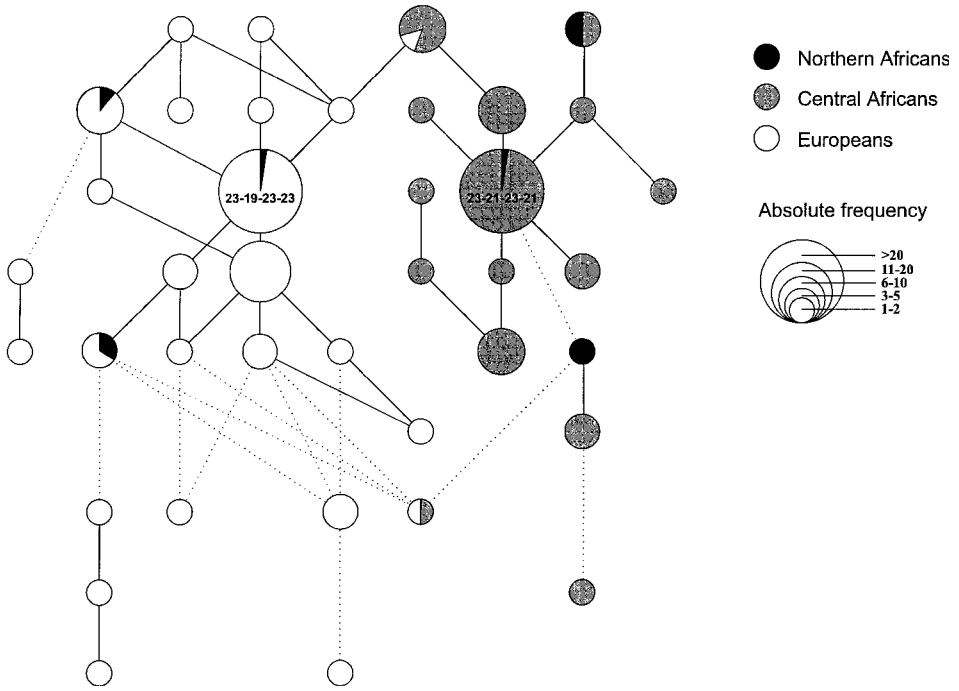

Figure 5 Network of adjacent microsatellite haplotypes within lineage 1C. Three haplotypes could not be placed within this network (also see the legend to fig. 4).

subject showed the SRY_{10831} A allele, which required us to postulate a reversion event at this site. We also have confirmed that the YAP^+ chromosomes found in Africa represent only a subset of the global YAP^+ chromosomes. In fact, no instances of the YAP insertion on chromosomes with the ancestral state at SRY_{4064} (allele G) were found. This observation is compatible with the hypothesis that, if one presumes an Asiatic origin, then either such chromosomes never migrated to Africa or, alternatively, African populations bearing them are extinct or still escape detection (Altheide and Hammer 1997).

Analysis of the alphoid/*Hin*dIII polymorphism revealed an unexpected rate of recurrent mutation. This is partially attributable to the fact that we have analyzed an extensive number of haplotypes in which the site loss can be detected. Nevertheless, this polymorphism can be useful to mark local subsets of chromosomes, provided that a more evolutionarily stable background is defined.

Age of Y-Chromosome Biallelic Lineages

Overall, our estimates of the coalescent times for lineages carrying the new allele at each of the biallelic sites (fig. 1) are consistent with those that Hammer et al. (1998) obtained by a different method; the observed discrepancies deserve discussion. First, both methods can be affected by sampling biases due to population structuring (Goldstein et al. 1995, 1996; Hammer et al. 1998), yielding data sets with different information content for both biallelic and microsatellite polymorphisms. This may be particularly relevant in Africa, the continent with the greatest Y-chromosomal structuring (Scozzari et al. 1997). Second, the estimate for recent mutations, such as PN3, may suffer from founding effects and large random fluctuations in the process of accumulation of repeat variance in microsatellites. Third, the demographic changes accompanying the spread of each mutation may affect inferences of mutation antiquity. The method used here has poor sensitivity to the assumed population size, for repeat variances of the order of $\leqslant .1$; on the other hand, estimates for mutations with large associated CA variances are heavily dependent on population size (Malaspina et al. 1998). The A→G mutation at DYS271 is considered to mark the Bantu expansion 2,000–3,000 years ago (Passarino et al. 1998). Taken at face value, our estimate of the antiquity of this mutation would suggest that it existed long before the spread of the Bantu languages (Hiernaux 1968; van Bakel 1981). Expansion of the male effective size to >5,000 would correspond to a lowering of values of this estimate. Moreover, Goldstein et al. (1999), examining dinucleotide diversity associated with mutant FXI alleles and assuming a range of growth rates, have shown that confidence intervals are strongly dependent on the precise shape of the gene genealogy. In keeping with the notion

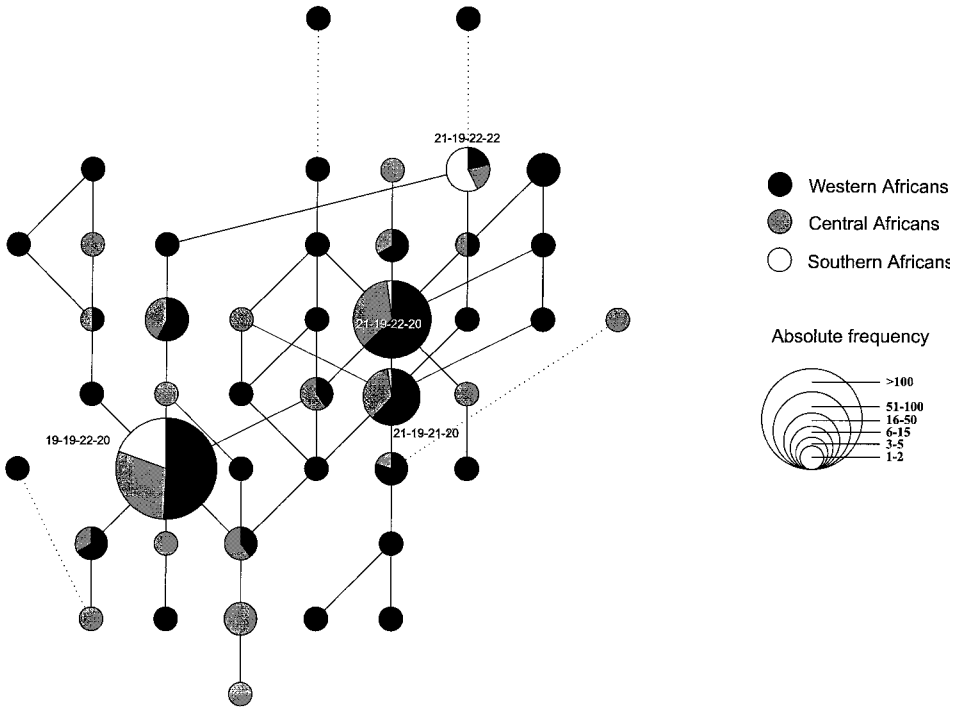

Figure 6 Network of adjacent microsatellite haplotypes within lineage 5. Four haplotypes could not be placed within this network (also see the legend to fig. 4).

that the Bantu expansion was a relatively recent phenomenon involving haplotype 5 is the observation of a very low population specificity of microsatellite types within this haplotype (fig. 6). In the present work, the PN1 C→T and the DYS271 A→G mutations were assumed to identify the same set of chromosomes as initially had been reported by Hammer et al. (1997). However, the dating of the PN1 mutation (Hammer et al. 1998) gives a more recent estimate than does the dating of the DYS271 mutation (fig. 1). This could imply the existence of Y chromosomes carrying the DYS271 G–derived—but not the PN1-derived—allele, a condition that has not yet been detected.

Geographic and Linguistic Affinities

Summarizing worldwide population data on the Y chromosome p49a,f/*Taq*I restriction polymorphisms (including data from 29 African population samples belonging to three different linguistic families), Poloni et al. (1997) have suggested that linguistic affiliation has strongly influenced the distribution of Y chromosomes. According to their data, Niger-Congo–speaking populations show a very low level of differentiation, followed by the Afro-Asiatic–speaking populations (including the Ethiopians), who were not found to be statistically different from each other. By contrast, the Khoisan populations were found to be statistically different from each other as well as from all other populations. The present paper provides data on Y-chromosome variation in a comparable number of African populations, giving a more detailed picture of the genetics of the African continent and including populations from the Nilo-Saharan linguistic family. The inferred pattern of interpopulation genetic relationships is somehow similar to that derived from p49a,f/*Taq*I markers, in that a correspondence with the linguistic families is found. However, as illustrated by Φ statistics (table 3) and correspondence analysis (fig. 3), geography seems to be a better criterion by which to identify genetically homogeneous groups. We have observed both a sharp differentiation between geographically separated populations of the same linguistic family (for instance, the Dendi from Benin relative to other Nilo-Saharan–speaking populations from northern Cameroon) and some close similarities among linguistically different populations living in the same area (e.g., northern Cameroon). Also, the Φ_{ST} estimates within the four linguistic families remain high, indicating a high level of internal differentiation. Only if populations from northern Cameroon are excluded from the analysis do the Φ_{ST} values of the Niger-Congo and the Afro-Asiatic linguistic groups drop to .064 ($P < .001$) and .052 ($P = .017$), respectively, values that are as low as those observed by Poloni et al. (1997). The genetics of northern Cameroon is consistent with written records and oral

traditions that depict this area as a region where important kingdoms were created and where different peoples with different histories of extensive trading, raiding, and migration have settled at different times (Mveng 1963; Faraut 1981; Mbuagbaw et al. 1990; Maceachern 1993).

A considerable amount of haplotype diversity is evident among the Khoisan, who show a haplotype repertoire almost as differentiated as that of the entire African continent; because of the high frequency of haplotype 5, they are genetically similar to western Africans, and this similarity is higher for the Khwe than for the !Kung. The Khoisan, on the whole, are nevertheless peculiar, in showing relatively high frequencies of both haplotype 4, which is also found in northern and eastern Africans, and the ancestral haplotype (haplotype 1A), which they share only with Ethiopians and a few central Africans. The mutations defining these haplotypes have very different ages. Although the Khoisan are reported to have occupied a wider region of sub-Saharan Africa in the past (McEvedy 1980; Cavalli-Sforza et al. 1994) and, possibly, to have experienced a Paleolithic expansion (Reich and Goldstein 1998), it is unlikely that all these mutations arose in the Khoisan and then spread to the entire continent. Instead, our findings are more easily interpreted if it is assumed that the Khoisan Y-chromosomal gene pool is the result of one or more recent admixture event(s) with an ancient background, and this process seems to be more prominent in the Khwe (see also Hiernaux 1974). This conclusion is supported by the finding that the microsatellite patterns associated with haplotype 5 of the Khoisan are identical to those of the main stock of chromosomes 5 found elsewhere (fig. 6 and Appendix B).

The Amhara and Oromo from Ethiopia, both of which speak languages of the Afro-Asiatic linguistic family, appear to constitute a rather homogeneous group. Having relatively high frequencies of haplotype 1B, they are similar to northern Africans, in showing caucasoid features. However, the presence and the frequency of haplotypes 4 and 1A make them also similar to the Khoisan. These eastern African–Khoisan affinities have been highlighted by the second principal component in the analysis by Cavalli-Sforza et al. (1994). Recently, Passarino et al. (1998) have compared the distribution of some Y-chromosome polymorphisms in Senegalese versus Ethiopians and have discussed it in terms of similarities and differences among western Africans, Ethiopians, and Khoisan. They have proposed haplotype YAP^+/DYS271A as a representative of pre-Bantu African Y chromosomes. PN2 analysis was not included in that study. We have reported here that the YAP^+/DYS271A chromosomes, also carrying the PN2T mutation (haplotype 4), are commonly found in northern and eastern Africans and in the Khoisan. Therefore, we suggest that the chromosomes YAP^+/PN2T/DYS271A (haplotype 4) represent a component of the pre-Bantu gene pool shared by all these three groups. A northern-African origin has been proposed for this haplotype (Hammer et al. 1998); however, the home range of the precursor haplotype 3A does not support this conclusion, since the latter is absent in northern Africa but is found in western, central, and southern Africa, albeit at low frequencies. In this context, different measures of the microsatellite variation associated with haplotype 4 lead to contrasting conclusions: haplotype diversity is higher in northern and eastern Africa than in western and central Africa ($.91 \pm .01$ vs. $.79 \pm .08$), whereas CA variance is greater in the latter region (.77 vs. .86).

"Ancestral" haplotype 1A was first reported, by Hammer et al. (1998), in the Khoisan. We find this haplotype at high frequencies both in the Khoisan and in eastern Africans. Its presence in these two groups attests to a remnant archaic gene pool pointing to eastern Africa as the most likely birthplace of modern humans, if one accepts the notion that the Khoisan were once present in an area (including eastern Africa) much wider than that which they presently occupy (for a discussion, see Cavalli-Sforza et al. 1994) However, as a result of the microsatellite analyses (fig. 4), it has also become clear that the haplotype 1A repertoire of the Khoisan is clearly distinct from that of the Ethiopians, thus attesting to a long-lasting independent evolution of these two groups.

Affinities between African and Non-African Populations

Also very interesting is the distribution of haplotype 1C, which is common in Egypt (13.0%) and very common in northern Cameroon (40.9%) but which is found at a very low frequency in the rest of Africa (Hammer et al. 1998; present study). In view of the strong correlation between DYS257 A and the alphoid/*Hin*dIII$^-$, this haplotype can be expected to overlap largely with haplotypes contributing to network 3.1 reported by Malaspina et al. (1998). This expectation has indeed been verified in a subset of chromosomes (see the Results section), and a concentration of these haplotypes is seen in western Europe. Figure 5 illustrates how, via the microsatellite-allele states, the haplotypes 1C found in Europe could be related to those from northern Cameroon. Two almost-separate subnetworks have been identified that are linked to each other through a single haplotype, attesting to a long history of independent evolution. The following two hypotheses are compatible with our observations: (1) ancient episodes of gene flow transferred the founders of haplotype 1C subnetworks from central Africa to Europe, or vice versa, through northern Africa, and (2) a third gene pool containing haplotype 1C, from which the European and African haplotypes 1C were

derived, exists. The recent data reported by Karafet et al. (1999) cast light on this issue and favor the second hypothesis, describing the existence, in central Asia, of high frequencies of haplotype 1C, together with its immediate ancestor, haplotype 1U.

In summary, the following three main conclusions can be drawn from this work: (1) Whereas a time of the order of 20,000 years has been estimated as being necessary for haplotype 5 to reach the level of microsatellite differentiation that we have observed, the present distribution of the molecular subtypes of this haplotype among the sub-Saharan populations is fully compatible with a recent dispersion associated with the diffusion of the Bantu languages. (2) A group of archaic Y-chromosome haplotypes link the Khoisan to the Ethiopians, thus supporting the notion of a proto-Khoisan presence in eastern Africa ⩾10,000 years ago (McEvedy 1980; Cavalli-Sforza et al. 1994); however, the two groups or their ancestors have been cut off from each other for so long that even the shared-by-descent portion of their gene pool has now completely diverged. Validation of the theory that associates the Khoisan with the root of all humans (for a discussion, see Cavalli-Sforza et al. 1994) requires that more data be gathered. (3) A group of Y-chromosome haplotypes specifically relate populations from northern Cameroon to a main western European cluster. Molecular analyses of Asian populations are beginning to elucidate the geographic origin and time dispersal(s) of this group of haplotypes (Karafet et al. 1999).

Acknowledgments

We would like to express our gratitude to all aboriginal peoples, for their helpful collaboration. We are deeply indebted to Prof. Massimo Cresta (Dipartimento di Biologia Animale e dell'Uomo, Università "La Sapienza", Rome) and the "Gruppo Laici del Terzo Mondo" (Naples), for helping with the sample collection in Benin, and to Prof. Gabriella Spedini (Dipartimento di Biologia Animale e dell'Uomo, Università "La Sapienza", Rome), the coordinator of a long-term genetic and epidemiological survey (1985–97) in Cameroon, for providing blood samples from this region. This work has been supported by funds from the Ministry of the University, Progetti Ricerca Interesse Nazionale 1997, and P.F. Beni Culturali contract 97.00702.PF36 (all to R.S.) and by Consiglio Nazionale delle Ricerche contracts 97.04211.CT04 and 98.00485.CT04 (both to A.N.)

Appendix A

Table A1

Pairwise Φ_{ST} Values

	Pairwise Φ_{ST} Value																							
	MOR	NEG	SEG	AMH	ORO	MOS	RIM	FUB	BAR	BER	FON	DEN	FAL	TAL	OAD	FUC	OUL	DAB	OCH	ONS	BAM	EWO	BAK	KUN
NEG	.01																							
SEG	.18	.04																						
AMH	.09	.02	.03																					
ORO	.03	.01	.10	.00																				
MOS	.52	.60	.70	.58	.55																			
RIM	.32	.37	.50	.40	.35	.16																		
FUB	.47	.51	.61	.51	.47	−.03	.11																	
BAR	.41	.45	.55	.47	.43	.00	.07	−.02																
BER	.55	.60	.69	.57	.55	.02	.25	.05	.06															
FON	.49	.55	.65	.54	.50	−.02	.14	−.03	−.01	.01														
DEN	.49	.56	.66	.54	.51	−.02	.12	−.04	.00	.05	−.02													
FAL	.18	.12	.15	.17	.16	.31	.18	.23	.18	.30	.26	.28												
TAL	.30	.31	.42	.36	.31	.08	−.02	.03	−.02	.19	.06	.05	.09											
OAD	.36	.20	.06	.19	.26	.71	.54	.61	.56	.68	.65	.67	.15	.43										
FUC	.24	.12	.06	.07	.13	.74	.50	.66	.59	.75	.69	.71	.21	.45	.17									
OUL	.71	.65	.51	.56	.64	.94	.86	.95	.88	.99	.93	.94	.55	.87	.28	.66								
DAB	.20	.07	.04	.12	.14	.59	.36	.47	.41	.56	.52	.54	.05	.26	.02	.07	.47							
OCH	.45	.28	.11	.26	.34	.80	.65	.73	.67	.79	.76	.77	.25	.56	−.04	.24	.19	.09						
ONS	.19	.06	.00	.01	.06	.64	.40	.53	.45	.65	.58	.60	.03	.28	.03	.04	.63	−.01	.13					
BAM	.54	.60	.69	.58	.55	.00	.21	−.01	.02	.00	−.02	.01	.30	.12	.70	.74	.95	.58	.79	.64				
EWO	.38	.41	.51	.43	.39	.03	.09	.01	−.02	.07	.01	.03	.14	−.01	.51	.55	.86	.37	.63	.39	.02			
BAK	.51	.55	.63	.53	.50	.00	.21	.02	.03	.00	−.01	.03	.26	.13	.62	.70	.99	.49	.74	.56	−.02	.04		
KUN	.14	.10	.12	.06	.06	.36	.23	.29	.26	.35	.32	.33	.07	.17	.18	.12	.51	.10	.26	−.03	.36	.23	.32	
KHW	.19	.23	.38	.28	.21	.17	.04	.12	.07	.23	.14	.13	.11	.00	.43	.41	.81	.27	.54	.25	.20	.06	.18	.13

Note.—Abbreviations are as defined in table 1. Underlined values are significant at the $P < .05$ level.

Appendix B

Table B1

Frequencies of Y-Chromosome Superhaplotypes in 25 African Populations

Superhaplotype[a]		Frequency in[b]																								
		MOR (56)	NEG (25)	SEG (21)	AMH (48)	ORO (37)	MOS (56)	RIM (42)	FUB (20)	BAR (34)	BER (20)	FON (36)	DEN (37)	FAL (39)	TAL (15)	OAD (18)	FUC (17)	OUL (23)	DAB (18)	OCH (15)	ONS (9)	BAM (44)	EWO (30)	BAK (12)	KUN (64)	KHW (26)
1A:																										
20 19	21 21	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	1.6	...
20 20	21 19	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	1.6	...
20 20	21 21	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	14.1	...
21 18	26 23	...	...	...	2.1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
21 19	24 24	...	...	...	...	2.7	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
21 20	21 21	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	1.6	...
21 20	24 24	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	6.7	...	...	...	...	...	...
21 21	23 22	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	5.9	...	...	...	...	...	...	...	...	...
21 21	24 23	...	...	...	...	...	...	...	...	...	...	...	...	...	...	5.6	5.9	...	...	...	22.2	...	...	...	...	...
21 21	24 24	...	...	...	...	2.7	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
21 21	25 23	...	...	...	...	...	...	...	...	...	...	2.8	...	...	...	...	...	...	...	...	...	...	...	...	...	...
21 21	25 24	...	...	...	2.1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
22 19	24 23	...	...	...	2.1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
22 20	21 21	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	3.1	...
22 20	22 21	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	3.1	11.5
22 20	22 22	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	1.6	...
22 21	24 23	...	...	...	2.1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
22 21	24 24	...	...	...	8.3	8.1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
22 21	25 24	...	...	...	...	2.7	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
22 21	26 24	...	...	...	4.2	2.7	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
22 21	27 24	...	...	...	...	2.7	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
22 22	24 24	...	...	...	...	2.7	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
23 22	23 23	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	1.6	...
1B:																										
20 18	22 21	...	...	4.8	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
20 19	22 21	...	...	...	...	2.7	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
20 20	21 21	...	...	4.8	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
20 20	22 21	3.6	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
21 18	22 21	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	22.2	...	...	...	...	...
21 19	18 18	...	...	4.8	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
21 20	21 21	...	4.0	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
21 21	21 20	...	...	...	...	...	1.8	...	...	...	...	...	...	...	...	5.6	...	...	...	...	...	...	...	...	...	...
21 21	23 18	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	1.6	...
22 19	17 17	...	12.0	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
22 19	21 21	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	1.6	...
22 19	22 21	1.8	...	4.8	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
22 19	22 22	7.1	...	...	10.4	5.4	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
22 21	18 18	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	4.7	...
22 21	20 20	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	4.3	...	...	...	...	6.7	...	...	...
22 21	21 20	...	...	...	...	...	...	...	...	...	...	...	...	17.9	...	...	...	...	...	...	...	4.5	3.3	...	...	...
22 21	21 21	...	...	...	...	2.7	...	...	...	...	...	...	...	...	...	5.6	...	...	...	6.7	...	...	...	...	...	...

(*continued*)

Table B1 (continued)

Superhaplotype[a]		Frequency in[b]																								
		MOR (56)	NEG (25)	SEG (21)	AMH (48)	ORO (37)	MOS (56)	RIM (42)	FUB (20)	BAR (34)	BER (20)	FON (36)	DEN (37)	FAL (39)	TAL (15)	OAD (18)	FUC (17)	OUL (23)	DAB (18)	OCH (15)	ONS (9)	BAM (44)	EWO (30)	BAK (12)	KUN (64)	KHW (26)
1B:																										
22 21	22 21	…	…	…	…	2.7	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
22 22	22 21	7.1	4.0	14.3	6.3	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
22 22	22 22	3.6	…	4.8	4.2	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
22 22	23 21	1.8	4.0	…	2.1	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
23 22	20 20	…	4.0	…	2.1	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
23 23	20 20	…	…	…	2.1	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
24 22	20 20	…	…	…	2.1	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
24 23	20 20	…	…	4.8	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
24 23	21 20	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	17.6	…	…	…	…	…	…	…	…	…
1C:																										
21 19	22 21	…	…	…	…	…	…	…	…	2.9	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
21 19	23 21	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	5.9	…	…	…	…	…	…	…	…	…
21 19	23 23	1.8	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
21 21	21 20	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	5.6	…	…	…	…	…	…	…
21 21	22 21	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	27.8	…	…	…	…	…	…	…
21 21	23 21	…	4.0	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
22 21	24 21	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	5.6	…	…	…	…	…	…	…
23 19	23 23	…	4.0	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
23 19	24 23	…	4.0	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
23 21	22 21	…	…	…	…	…	…	…	…	…	…	…	…	2.6	…	…	…	8.7	…	…	…	…	3.3	…	…	…
23 21	22 22	…	…	…	…	…	…	…	…	2.9	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
23 21	23 19	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	30.4	…	…	…	…	…	…	…	…
23 21	23 20	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	6.7	…	…	…	…	…	…
23 21	23 21	…	…	4.8	…	…	…	…	…	…	…	…	…	15.4	…	5.6	…	43.5	5.6	40.0	11.1	…	…	…	…	…
23 21	23 22	…	…	…	…	…	…	…	…	…	…	…	…	…	…	33.3	5.9	…	…	…	…	…	…	…	…	…
23 21	23 23	…	…	…	…	…	…	…	…	…	…	…	…	2.6	…	…	…	4.3	…	20.0	…	…	…	…	…	…
23 21	24 19	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	4.3	…	…	…	…	…	…	…	…
23 21	24 21	…	…	…	…	…	…	…	…	…	…	…	…	…	6.7	5.6	…	…	…	…	…	…	…	…	…	…
23 21	25 21	…	…	9.5	…	…	…	…	…	…	…	…	…	…	…	11.1	…	…	…	…	…	…	…	…	…	…
23 22	23 21	…	…	…	…	…	…	…	…	…	…	…	…	2.6	…	…	…	…	…	…	…	…	…	…	…	…
23 22	24 19	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	4.3	…	…	…	…	…	…	…	…
1D:																										
23 19	22 22	…	…	4.8	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
2:																										
18 18	20 20	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	4.7	…
18 18	21 20	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	3.1	…
3A:																										
18 17	21 21	…	…	…	2.1	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
19 19	21 20	…	…	…	…	…	…	2.4	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
20 19	21 20	…	…	…	…	…	…	26.2	…	…	…	…	2.7	…	…	…	…	…	22.2	…	11.1	…	…	…	4.7	3.8
20 19	21 21	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	1.6	…
21 18	21 21	…	…	…	…	…	…	…	5.0	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
21 18	22 22	…	…	…	…	…	…	…	5.0	2.9	…	5.6	2.7	…	…	…	…	…	…	…	…	…	…	…	…	…
21 18	23 22	…	…	…	…	…	…	…	…	2.9	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
21 18	24 24	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	52.9	…	…	…	…	…	…	…	…	…
21 19	20 20	…	…	…	…	…	5.4	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…

	21 19	21 20	...	...	...	...	...	...	...	...	...	...	...	...	...	20.0	...	...	...	...	...	...	...	...	...	...	...	...	...
	21 21	22 22	...	...	...	...	...	...	...	...	2.9	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	22 18	23 22	...	...	...	...	...	...	4.8	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	22 19	21 21	...	...	...	...	...	...	...	...	...	...	...	2.7	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
4:																													
	19 19	22 20	...	...	...	...	...	1.8	...	...	...	...	...	...	...	...	...	5.9	...	...	...	...	...	...	...	...	...	...	...
	19 19	22 22	...	...	...	...	2.7	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	20 19	22 22	...	4.0	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	21 18	23 22	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	5.6	...	...	...	...	...	...	...	...	...
	21 19	21 20	...	...	...	...	...	...	2.4	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	21 19	21 21	...	...	...	2.1	5.4	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	21 19	22 18	...	...	...	2.1	2.7	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	21 19	22 21	...	...	...	4.2	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	21 19	22 22	...	...	...	4.2	2.7	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	21 19	23 14	...	...	4.8	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	21 19	23 20	1.8	...	...	2.1	2.7	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	21 19	23 21	32.1	...	4.8	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	21 19	23 22	...	...	...	4.2	...	...	...	...	...	...	...	...	...	...	...	...	...	...	6.7	...	...	...	...	...	...	...	...
	21 19	24 20	...	...	...	...	...	1.8	...	...	...	...	...	2.7	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	21 19	24 23	...	...	4.8	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	21 19	25 20	...	...	...	...	2.7	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	21 21	22 18	...	...	...	2.1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	22 18	23 20	...	...	...	2.1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	22 18	24 22	...	4.0	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	22 19	21 21	3.6	...	...	...	8.1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	9.4	11.5	
	22 19	22 20	...	...	...	2.1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	11.5	
	22.19	22 21	23.2	16.0	4.8	2.1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	22 19	22 22	5.4	8.0	9.5	...	10.8	...	...	...	...	...	...	2.7	...	...	...	...	...	...	...	...	...	...	...	...	1.6	...	
	22 19	23 21	...	4.0	...	2.1	2.7	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...		
	22 19	23 22	1.8	8.0	...	4.2	2.7	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	22 19	24 19	...	4.0	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	22 19	24 20	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	3.8	
	22 19	24 21	...	4.0	...	...	2.7	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	22 19	24.22	1.8	...		...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	22 19	25 21	...	8.0	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	22.20	21 21	...	...	...	...	10.8	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	22.20	24 21	...	...	...	4.2	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	22 21	22 22	...	...	...	2.1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	22 21	24 24	...	...	4.8	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	22 21	25 23	...	...	...	4.2	2.7	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	22 22	21 21	...	...	...	2.1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	23 19	21 21	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	3.8	
	23 19	23 20	...	...	...	2.1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	23 21	22 21	...	...	4.8	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	24 19	22 22	...	...	...	...	2.7	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
5:																													
	19 17	22 20	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	8.3	...	...			
	19 18	20 20	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	2.3	...	...	...	...			
	19 18	22 20	...	...	...	...	...	3.6	...	...	...	...	...	...	...	...	...	...	...	...	...	2.3	...	...	...	...			
	19 19	20 20	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	4.5	...	8.3	...	...			
	19 19	21 20	...	...	...	...	...	1.8	...	...	2.9	...	...	...	...	...	...	...	...	...	...	6.8	...	...	...	...			
	19 19	21 21	...	...	...	...	...	...	...	...	...	...	...	5.4	...	...	...	...	...	...	...	...	...	...	...	...			

(*continued*)

Table B1 (continued)

Superhaplotype[a]		Frequency in[b]																								
		MOR (56)	NEG (25)	SEG (21)	AMH (48)	ORO (37)	MOS (56)	RIM (42)	FUB (20)	BAR (34)	BER (20)	FON (36)	DEN (37)	FAL (39)	TAL (15)	OAD (18)	FUC (17)	OUL (23)	DAB (18)	OCH (15)	ONS (9)	BAM (44)	EWO (30)	BAK (12)	KUN (64)	KHW (26)
5:																										
19 19	22 13	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	16.7	…	…
19 19	22 18	…	…	…	…	…	…	…	…	…	5.0	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
19 19	22 19	…	…	…	…	…	…	…	…	…	…	…	…	…	…	5.6	…	…	…	…	…	…	…	…	…	…
19 19	22 20	…	…	…	…	…	33.9	33.3	10.0	32.4	25.0	41.7	18.9	17.9	20.0	11.1	…	…	5.6	6.7	…	18.2	66.7	8.3	23.4	50.0
19 19	22 21	…	…	…	…	…	…	…	…	…	…	…	…	2.6	…	…	…	…	…	…	…	…	…	8.3	…	…
19 19	22 22	…	…	…	…	…	…	…	…	2.9	…	2.8	2.7	…	6.7	…	…	…	16.7	…	…	…	…	…	…	…
19 19	23 20	…	…	…	…	…	…	…	…	…	…	…	2.7	…	…	…	…	…	…	…	…	…	…	…	…	…
19 19	24 20	…	…	…	…	…	…	…	…	…	…	…	2.7	…	6.7	…	…	…	…	…	…	…	…	…	…	…
19 19	24 21	…	…	…	…	…	…	…	…	…	…	2.8	…	…	…	…	…	…	…	…	…	…	…	…	…	…
19 19	25 20	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	8.3	…	…
19 19	25 21	…	…	…	…	…	…	…	…	…	…	…	2.7	…	…	…	…	…	…	…	…	…	…	…	…	…
20 19	20 19	…	…	…	…	…	…	…	…	…	…	…	2.7	…	…	…	…	…	…	…	…	…	…	…	…	…
20 19	21 20	…	…	…	…	…	…	…	…	…	…	2.8	…	…	…	…	…	…	…	…	…	…	…	…	…	…
20 19	22 20	…	…	…	…	…	…	…	…	5.9	…	…	…	…	…	…	…	…	…	…	…	4.5	…	8.3	…	…
20 19	22 22	…	…	…	…	…	…	…	…	…	…	…	2.7	…	…	…	…	…	…	…	…	…	…	…	…	…
20 20	21 20	…	…	…	…	…	…	…	…	…	5.0	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
20 20	22 20	…	…	…	…	…	…	…	…	…	…	…	2.7	…	…	…	…	…	…	…	…	…	…	…	…	…
21 12	22 19	…	…	…	…	…	1.8	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
21 17	22 20	…	…	…	…	…	…	…	…	…	…	2.8	…	…	…	…	…	…	…	…	…	…	…	…	…	…
21 18	20 19	…	…	…	…	…	…	…	…	2.9	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
21 19	20 19	…	…	…	…	…	…	…	…	…	5.0	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
21 19	20 20	…	…	…	…	…	1.8	2.4	…	…	…	2.8	2.7	…	…	…	…	…	…	…	…	…	…	8.3	…	…
21 19	21 20	…	…	…	…	…	25.0	21.4	5.0	…	…	…	16.2	28.2	13.3	5.6	…	…	…	6.7	11.1	2.3	…	8.3	…	3.8
21 19	21 21	…	…	…	…	…	…	…	…	…	…	5.6	…	…	…	…	…	…	…	…	…	…	…	…	…	…
21 19	22 18	…	…	…	…	…	…	…	…	5.9	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
21 19	22 19	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	4.5	…	…	…	…
21 19	22 20	…	…	…	…	…	19.6	7.1	70.0	23.5	30.0	27.8	10.8	10.3	20.0	…	…	…	…	…	…	40.9	16.7	16.7	3.1	…
21 19	22 21	…	…	…	…	…	…	…	…	…	…	…	2.7	…	…	…	…	…	…	…	…	…	3.3	…	…	…
21 19	22 22	…	…	…	…	…	…	…	…	2.9	…	2.8	2.7	…	…	…	…	…	5.6	…	11.1	2.3	…	…	12.5	…
21 19	23 20	…	…	…	…	…	1.8	…	…	…	…	…	2.7	…	…	…	…	…	…	…	…	2.3	…	…	…	…
21 19	23 23	…	…	…	…	…	…	…	5.0	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
21 19	24 20	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	11.1	…	…	…	…	…
21 19	24 24	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	4.5	…	…	…	…
21 20	21 20	1.8	…	…	…	…	…	…	…	…	…	…	…	…	…	5.6	…	…	…	…	…	…	…	…	…	…
21 20	22 20	…	…	…	…	…	…	…	…	…	…	…	2.7	…	…	…	…	…	…	…	…	…	…	…	…	…
21 21	20 20	…	…	…	…	…	…	…	…	…	…	…	…	…	6.7	…	…	…	…	…	…	…	…	…	…	…
21 21	22 20	…	…	…	…	…	…	…	…	…	10.0	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
22 19	21 20	…	…	…	…	…	…	…	…	…	…	…	2.7	…	…	…	…	…	…	…	…	…	…	…	…	…
22 19	22 20	…	…	…	…	…	…	…	…	…	5.0	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
22 19	22 21	…	…	…	…	…	…	…	…	…	15.0	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
23 19	24 23	…	…	…	…	…	…	…	…	2.9	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
23 21	22 20	…	…	…	…	…	…	…	…	2.9	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…

[a] The allele state at seven biallelic polymorphic sites (see text) was combined with that at four dinucleotide microsatellites. The nomenclature of biallelic haplotypes (first column) is that of Hammer et al. (1998). Microsatellite haplotypes are indicated by the repeat number at YCAIIa, YCAIIb, DYS413a, and DYS413b, in that order.

[b] Abbreviations are as in table 1; in the column headings, nos. in parentheses are no. of Y chromosomes analyzed.

References

Altheide TK, Hammer MF (1997) Evidence for a possible Asian origin of YAP$^+$ Y chromosomes. Am J Hum Genet 61:462–466

Biondi G, Rickards O, Martinez-Labarga C, Taraborelli T, Ciminelli B, Gruppioni G (1996) Biodemography and genetics of the Berba of Benin. Am J Phys Anthropol 99: 519–535

Bottema CDK, Sommer SS (1993) PCR amplification of specific alleles: rapid detection of known mutations and polymorphisms. Mutat Res 288:93–102

Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton, NJ

Ciminelli BM, Pompei F, Malaspina P, Hammer M, Persichetti F, Pignatti PF, Palena A, et al (1995) Recurrent simple tandem repeat mutations during human Y- chromosome radiation in Caucasian subpopulations. J Mol Evol 41:966–973

Cooper G, Amos W, Hoffman D, Rubinsztein DC (1996) Network analysis of human Y microsatellite haplotypes. Hum Mol Genet 5:1759–1766

Destro-Bisol G, Battaggia C, Macchiarelli R, Bailly C, Scozzari R, Spedini G (1992) A bio-anthropological study on Bakakas of Cameroon. Ann Hum Biol 19:185–195

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

Faraut F (1981) Les Mboum. In: Tardits C (ed) Colloques Internationaux du C.N.R.S. 551: contribution de la recherche ethnologique a l'histoire des civilisations du Cameroun. Vol 1. Éditions du Centre National de la Recherche Scientifique, Paris, pp 159–169

Fort M, De Stefano GF, Cambon-Thomsen A, Giraldo-Alvarez P, Dugoujon J-M, Ohayon E, Scano G, et al (1998) HLA class II allele and haplotype frequencies in Ethiopian Amhara and Oromo populations. Tissue Antigens 51:327–336

Goldstein DB, Reich DE, Bradman N, Usher S, Seligsohn U, Peretz H (1999) Age estimates of two common mutations causing factor XI deficiency: recent genetic drift is not necessary for elevated disease incidence among Ashkenazi Jews. Am J Hum Genet 64:1071–1075

Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995) Genetic absolute dating based on microsatellites and the origin of modem humans. Proc Natl Acad Sci USA 92:6723–6727

Goldstein DB, Zhivotovsky LA, Nayar K, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1996) Statistical properties of the variation at linked microsatellite loci: implications for the history of human Y chromosomes. Mol Biol Evol 13:1213–1218 (erratum: Mol Biol Evol 14:354 [1997])

Greenberg JH (1987) Classificazione delle lingue africane. In: Ki-Zerbo J (ed) Storia generale dell'Africa. Vol 1: Metodologia e preistoria dell' Africa. Jaca Book, Milan, pp 313–332

Grimes BF (1996) Ethnologue: languages of the world, 13th 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, et al (1998) Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. Mol Biol Evol 15:427–441

Hammer MF, Spurdle AB, Karafet T, Bonner MR, Wood ET, Novelletto A, Malaspina P, et al (1997) The geographic distribution of human Y chromosome variation. Genetics 145: 787–805

Hiernaux J (1968) Bantu expansion: the evidence from physical anthropology confronted with linguistic and archaeological evidence. J Afr Hist 9:505–515

Hiernaux J (1974) The people of Africa. Weidenfeld & Nicolson, London

Jobling MA, Tyler-Smith C (1995) Fathers and sons: the Y chromosome and human evolution. Trends Genet 11: 449–456

Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, et al (1999) Ancestral Asian source(s) of New World Y-chromosome founder haplotypes. Am J Hum Genet 64:817–831

Li H, Cui X, Arnheim N (1990) Direct electrophoretic detection of the allelic state of single DNA molecules in human sperm by using the polymerase chain reaction. Proc Natl Acad Sci USA 87:4580–4584

Maceachern S (1993) Selling the iron for their shackles: Wandala-Montagnard interactions in northern Cameroon. J Afr Hist 34:247–270

Malaspina P, Ciminelli BM, Viggiano L, Jodice C, Cruciani F, Santolamazza P, Sellitto D, et al (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, et al (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

Mathias N, Bayés M, Tyler-Smith C (1994) Highly informative compound haplotypes for the human Y chromosome. Hum Mol Genet 3:115–123

Mbuagbaw TE, Brain R, Palmer R (1990) A history of the Cameroon. Longman Group UK, Harrow, England

McEvedy C (1980) The Penguin atlas of African history. Penguin Books, Harmondsworth, England

Mveng E (1963) Histoire du Cameroun. Éditions Présence Africaine, Paris

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

Persichetti F, Blasi P, Hammer M, Malaspina P, Jodice C, Terrenato L, Novelletto A (1992) Disequilibrium of multiple DNA markers on the human Y chromosome. Ann Hum Genet 56:303–310

Poloni ES, Semino O, Passarino G, Santachiara-Benerecetti AS, Dupanloup I, Langaney A, Excoffier L (1997) Human genetic affinities for Y-chromosome p49a,f/*Taq*I haplotypes

show strong correspondence with linguistics. Am J Hum Genet 61:1015–1035

Reich DE, Goldstein DB (1998) Genetic evidence for a Paleolithic human population expansion in Africa. Proc Natl Acad Sci USA 95:8119–8123

Rickards O, Martinez-Labarga C (1994) Genetic variation in the peoples of Benin (West Africa). In: Argano R, Cirotto C, Grassi-Milano E, Mastrolia L (eds) Contributions to animal biology. Halocynthia Association, Palermo, pp 355–366

Santos FR, Bianchi NO, Pena SDJ (1996) Worldwide distribution of human Y-chromosome haplotypes. Genome Res 6:601–611

Santos FR, Pandya A, Tyler-Smith C, Pena SDJ, Schanfield M, Leonard WR, Osipova L, et al (1999) The central Siberian origin for Native American Y chromosomes. Am J Hum Genet 64:619–628

Santos FR, Pena SDJ, Tyler-Smith C (1995) PCR haplotypes for the human Y chromosome based on alphoid satellite DNA variants and heteroduplex analysis. Gene 165: 191–198

Scacchi R, Corbo RM, Rickards O, De Stefano GF (1994) Survey of seven plasma protein polymorphisms in the Amhara and Oromo populations of Ethiopia. Am J Hum Biol 6:773–781

Schneider S, Kueffer J-M, Roessli D, Exeoffier L (1997) Arlequin version 1.1: a software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Geneva

Scozzari R, Cruciani F, Malaspina P, Santolamazza P, Ciminelli BM, Torroni A, Modiano D, et al (1997) Differential structuring of human populations for homologous X and Y microsatellite loci. Am J Hum Genet 61:719–733

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

Seielstad MT, Minch E, Cavalli-Sforza LL (1998) Genetic evidence for a higher female migration rate in humans. Nat Genet 20:278–280

Spedini G, Destrol-Bisol G, Mondovi S, Kaptué L, Taglioli L, Paoli G. The peopling of sub-Saharan Africa: the case study of Cameroon. Am J Phys Anthropol (in press)

Tartaglia M, Scano G, De Stefano GF (1996) An anthropogenetic study on the Oromo and Amhara of central Ethiopia. Am J Hum Biol 8:505–516

Torroni A, Semino O, Scozzari R, Sirugo G, Spedini G, Abbas N, Fellous M, et al (1990) Y chromosome DNA polymorphisms in human populations: differences between Caucasoids and Africans detected by 49a and 49f probes. Ann Hum Genet 54:287–296

Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, et al (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. Genome Res 7:996–1005

van Bakel MA (1981) The "Bantu" expansion: demographic models. Curr Anthropol 22:688–691

Weber JL, Wong C (1993) Mutation of human short tandem repeats. Hum Mol Genet 2:1123–1128

Whitfield LS, Sulston JE, Goodfellow PN (1995) Sequence variation of the human Y chromosome. Nature 378:379–380