Am. J. Hum. Genet. 61:719–733, 1997

Differential Structuring of Human Populations for Homologous X and Y Microsatellite Loci

Rosaria Scozzari,[1] Fulvio Cruciani,[1] Patrizia Malaspina,[3] Piero Santolamazza,[1] Bianca Maria Ciminelli,[3] Antonio Torroni,[1] David Modiano,[2] Douglas C. Wallace,[4] Kenneth K. Kidd,[5] Antonel Olckers,[6] Pedro Moral,[7] Luciano Terrenato,[3] Nejat Akar,[8] Raheel Qamar,[9] Atika Mansoor,[9] Syed Q. Mehdi,[9] Gianfranco Meloni,[10] Giuseppe Vona,[11] David E. C. Cole,[12] Wangwei Cai,[13] and Andrea Novelletto[3]

[1]Dipartimento di Genetica e Biologia Molecolare, and [2]Fondazione Pasteur Cenci-Bolognetti, Istituto di Parassitologia, Università "La Sapienza," and [3]Dipartimento di Biologia, Università "Tor Vergata," Rome; [4]Department of Genetics and Molecular Medicine, Emory University, Atlanta; [5]Department of Genetics, Yale University School of Medicine, New Haven; [6]Department of Human Genetics, University of Pretoria, Pretoria; [7]Departamento de Biologia Animal, Universitat de Barcelona, Barcelona; [8]Pediatric Molecular Pathology Unit, University of Ankara, Ankara; [9]A. Q. Khan Research Laboratories, Islamabad; [10]Clinica Pediatrica A. Filia, Università degli Studi, Sassari; [11]Dipartimento di Biologia Sperimentale, Sezione di Scienze Antropologiche, Università di Cagliari, Cagliari; [12]Departments of Clinical Biochemistry, Medicine, and Genetics, University of Toronto, Toronto; and [13]Istituto Internazionale di Genetica e Biofisica, CNR, Napoli

Summary

The global pattern of variation at the homologous microsatellite loci DYS413 (Yq11) and DXS8174 and DXS8175 (Xp22) was analyzed by examination of 30 world populations from four continents, accounting for more than 1,100 chromosomes per locus. The data showed discordant patterns of among- and within-population gene diversity for the Y-linked and the X-linked microsatellites. For the Y-linked polymorphism, all groups of populations displayed high F_{ST} values (the correlation between random haplotypes within subpopulations, relative to haplotypes of the total population) and showed a general trend for the haplotypes to cluster in a population-specific way. This was especially true for sub-Saharan African populations. The data also indicated that a large fraction of the variation among populations was due to the accumulation of new variants associated with the radiation process. Europeans exhibited the highest level of within-population haplotype diversity, whereas sub-Saharan Africans showed the lowest. In contrast, data for the two X-linked polymorphisms were concordant in showing lower F_{ST} values, as compared with those for DYS413, but higher within-population variances, for African versus non-African populations. Whereas the results for the X-linked loci agreed with a model of greater antiquity for the African populations, those for DYS413 showed a confounding pattern that is apparently at odds with such a model. Possible factors involved in this differential structuring for homologous X and Y microsatellite polymorphisms are discussed.

Received January 21, 1997; accepted for publication June 30, 1997.

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

0002-9297/97/6103-0030$02.00

Introduction

Human polymorphism at the DNA level, involving nuclear restriction sites (RSPs) (Wainscoat et al. 1986; Bowcock et al. 1991; Kidd et al. 1991), mtDNA sequences (Cann et al. 1987; Vigilant et al. 1991; Horai et al. 1995; Wallace 1995), and the length of mini- and microsatellites (Edwards et al. 1992; Bowcock et al. 1994; Deka et al. 1995*a*, 1995*b*; Armour et al. 1996; Tishkoff et al. 1996), has been used recently to infer genetic relationships among human populations. Initial reports showing the highest heterozygosities for RSPs among Europeans were revisited in terms of an ascertainment bias (Bowcock et al. 1994; Kidd and Kidd 1996; Rogers and Jorde 1996). On the other hand, mtDNA and mini- and microsatellite polymorphisms displayed a higher gene diversity within Africa, as compared with other geographic areas. These findings have been interpreted as lending support to the theory of an African origin of modern humans, although Templeton (1993) and Jorde et al. (1995) have suggested that there is a lack of critical tests of this hypothesis.

Human variation for sex chromosomes is an invaluable tool for population genetic analyses. Loci of the male-specific portion of the human Y chromosome are haploid, paternally transmitted, do not recombine, and are subject to only male-specific mutation rates (Crow 1993). Although they are especially useful for the identification of male lineages, polymorphisms on the human Y chromosome have been reported to be very few, relative to the autosomes and the X chromosome (Jakubiczka et al. 1989; Malaspina et al. 1990; Spurdle and Jenkins 1992; Dorit et al. 1995; Whitfield et al. 1995; Underhill et al. 1996). On the other hand, the X chromosome is, with regard to its nonpseudoautosomal portion, predominantly present in females; it does recombine, and is subject to alternating male-specific and female-

specific mutation rates. Comparative studies of nucleotide substitutions (Shimmin et al. 1993) have shown a slower evolutionary rate for the sequences on the X chromosome than for their homologous Y-linked counterparts.

The complexity of the band pattern associated with the Y-linked microsatellite YCAIII (Mathias et al. 1994) was resolved recently by the determination that these bands constitute a family of six closely related microsatellite-containing loci (called the "CAIII loci"; Malaspina et al. 1997). Of these loci, two were located on Yq11, in the region containing the sulfatase and the Kallmann syndrome pseudogenes of the Y chromosome. The other four loci were shown to be X-linked and to map to Xp22, in a narrow region containing the functional counterparts of the same pseudogenes that are on the Y chromosome. Thus the CAIII loci contribute other sequences to the nonrecombining region of homology between Xp and Yq, described in the evolutionary model of Meroni et al. (1996). All the CAIII loci were shown to contain an internal $(CA)_n$ microsatellite, with flanking sequences (of 111 bp) displaying an average identity value of .88 across loci (Malaspina et al. 1997). Analysis of locus-specific single-nucleotide substitutions indicated that the Y-specific members of this family became isolated on the Y chromosome relatively recently, as compared with the duplication and divergence of the X-linked members. By developing new PCR assays, Malaspina et al. (1997) could amplify specifically the two Y-derived bands (bands 1 and 2; fig. 1*A* and *B*) and each of the X-derived bands. In addition to the polymorphism of the Y loci (hereafter called "DYS413"), described elsewhere (Mathias et al. 1994), an extensive polymorphism was displayed by two of the X loci (hereafter called "DXS8175" and "DXS8174"), corresponding to bands 4 and 6 (fig. 1*A, C,* and *D*). In view of the high degree of similarity among the flanking sequences, this family of loci provides a novel opportunity to investigate human population structuring for sex-specific polymorphisms of the same nature. In fact, the CAIII family represents a model system in which the possible effects of flanking DNA on the mutability of the microsatellite are more likely to be equalized, as compared with evolutionarily unrelated loci. In addition, because detailed information on the molecular variation underlying these polymorphisms is now available, the molecular distances between alleles/haplotypes (Excoffier et al. 1992; Goldstein et al. 1995; Slatkin 1995; Michalakis and Excoffier 1996) could be incorporated into the study of variation, improving our knowledge of the population structure.

In this article, we describe the population distribution of the DYS413, DXS8175, and DXS8174 polymorphisms in a large sample including representatives of four continents. Analysis of molecular variance was performed for each of the microsatellite loci, by use of both the allele/haplotype frequencies and the differences in the number of repeats between alleles/haplotypes. This analysis revealed a substantial divergence among human groups and among populations within groups, for the Y microsatellite polymorphism. The results for the X microsatellite loci were consistent with one another but showed a substantially lower level of divergence than those for the Y microsatellite loci.

Subjects and Methods

Subjects

The sample consisted of 1,419 unrelated individuals (1,116 males and 303 females) belonging to 30 populations from four continents (Europe, Africa, Asia, and the Americas). These populations were aggregated into nine groups, on the basis of a geographic criterion.

Populations from northern Europe included 20 male subjects from England (Ciminelli et al. 1995; Hammer et al. 1997) and 48 Danish subjects (35 males and 13 females). The southern-European sample consisted of male subjects (20 from Veneto, northern Italy, 46 from southern Italy, 55 from Sardinia, and 22 from Greece) from the populations previously described by Ciminelli et al. (1995) and by Hammer et al. (1997). We also included 75 Spanish Basques (55 males and 20 females) from the Guipuzcoa province, 36 males from southern Spain, 100 individuals (84 males and 16 females) from central Italy, 40 additional individuals (37 males and 3 females) from southern Italy, 55 additional individuals from Sardinia (42 males and 13 females), and 29 additional males from Greece.

The western-Asian group included 20 males from Turkey (Figus et al. 1995) and 32 males from the eastern Arabian peninsula (21 from the United Arab Emirates and 11 from Oman) (Ciminelli et al. 1995; Hammer et al. 1997). A sample of 20 Pakistani males (of the Pathan ethnic group) represented southern Asians, and 48 Egyptian males (Ciminelli et al. 1995; Hammer et al. 1997) constituted the northern-African group.

The western-African group consisted of 10 populations from two nation-states (Cameroon and Burkina Faso), belonging to different linguistic families (Grimes 1992). The Bamileke (52 males) and the Ewondo (31 males and 19 females) are Bantu-speaking peoples living in the Bamileke Plateau (western Cameroon) and at the northern boundary of the equatorial forest (southern Cameroon), respectively. The Ouldeme (23 males and 20 females) and the Daba (18 males and 8 females) both speak the Chadic branch of the Afro-Asiatic language family and live on Mounts Mandara (northern Cameroon). The Fali (39 males and 20 females) and the Tali (15 males and 21 females) speak the Adamaua branch of the Niger-Congo language

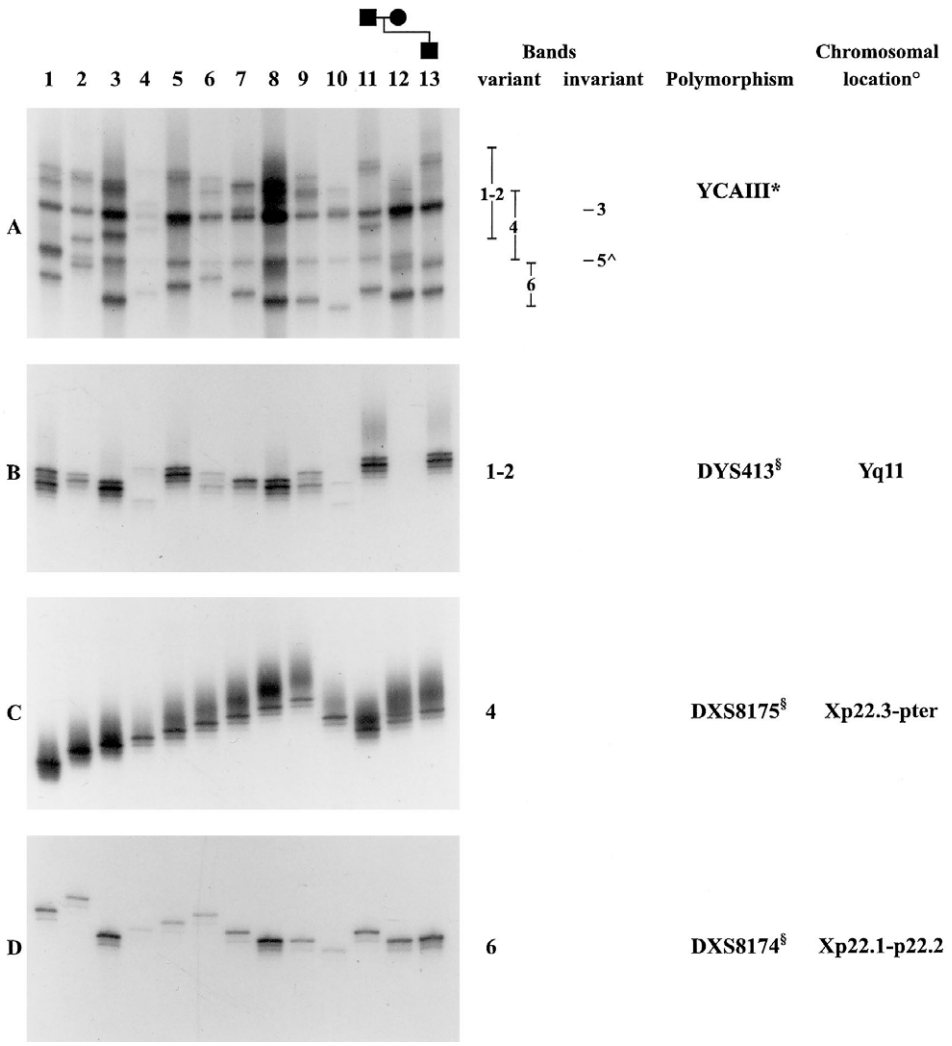

Figure 1 Band patterns obtained for 12 males and one female, by different PCR assays. The migration range of the polymorphic bands and the chromosomal location of the corresponding loci also are shown. *A,* Patterns obtained with the assay of Mathias et al. (1994). The carat symbol (∧) indicates that, for band 5, the single example of variant pattern that was observed, from among 1,243 random individuals, is depicted (lane 13); the segregation pattern (lanes 11, 12, and 13) supported the X-chromosomal assignment of the band 5 locus as determined by physical mapping (Malaspina et al. 1997). The asterisk (*) indicates that the primers and PCR conditions used were as reported by Mathias et al. (1994). *B,* Patterns generated with the DYS413-specific assay (bands 1–2), for the same subjects as in panel *A*. Haplotypes are named in accordance with the repeat number of the corresponding amplified fragments and are as follows: 22/20 (lanes 1, 6, and 9); 22/21 (lane 2); 21/20 (lanes 3 and 8); 23/18 (lane 4); 23/22 (lane 5); 21/21 (lane 7); 20/17 (lane 10); and 24/23 (lanes 11 and 13). *C,* Patterns generated with the DXS8175-specific assay (band 4), for the same subjects as in panel *A*. Phenotypes are named in accordance with the repeat number of the corresponding amplified fragments and are as follows: 10 (lane 1); 12 (lane 2); 13 (lane 3); 14 (lanes 4 and 11); 15 (lane 5); 16 (lanes 6, 10, and 13); 17 (lane 7); 18 (lane 8); 19 (lane 9); and 16/15 (lane 12). *D,* Patterns generated with the DXS8174-specific assay (band 6), for the same subjects as in panel *A*. Phenotypes are named in accordance with the repeat number of the corresponding amplified fragments and are as follows: 13 (lanes 1 and 5); 15 (lane 2); 11 (lanes 3, 8, 9, 12, and 13); 12 (lanes 4, 7, and 11); 14 (lane 6); and 10 (lane 10). Note the absence of Y-derived bands 1 and 2 for the female (lane 12), in panels *A* and *B*. The degree symbol (°) indicates data described in Malaspina et al. (1997). A section symbol (§) indicates that the primers and PCR conditions used were as described by Malaspina et al. (1997).

family and live in northern Cameroon. The Fulbe are a large population (~13,000,000 total) living in an area spanning western and central Africa and speaking Fulfulde, a West Atlantic language of the Niger-Congo language family. Two Fulbe samples were examined—one from Burkina Faso (20 males and 20 females) and the other from northern Cameroon (17 males and 20 females). The Mossi (56 males and 17 females) are the dominant population of Burkina Faso (~4,000,000 total) and speak Mooré, a Gur language of the Niger-Congo language family. The Rimaibe (42 males and 17 females) were the ancient slaves of the Fulbe from Burkina Faso and speak the same Fulfulde language as do the Fulbe.

The southern African group came from Kimberley (Northern Cape Province) and consisted of two Khoisan-speaking populations, the !Kung (64 males and 20 females) and the Khwe (26 males and 6 females). The eastern Asian group included a single sample of 37 Chinese Li (28 males and 9 females) described elsewhere (Cai et al. 1994; Scozzari et al. 1996).

Four samples were included in the group of Native Americans—the Ojibwa sample (19 males and 16 females) from Manitoulin Island (Canada) and the Guahibo sample (30 males) from Estados Amazonas (Venezuela), previously described by Scozzari et al. (1997); the Seminole sample (27 males) described by Huoponen et al. (1997); and the Mayan sample (8 males and 25 females) described by Kidd et al. (1991).

Microsatellite Typing

Genomic DNA samples were obtained, by standard techniques, from placentas, fresh blood, or lymphoblastoid cell lines (Danish and Mayan subjects). Microsatellite typing was performed on all subjects, with the assay described by Mathias et al. (1994), except for the Basques, the southern Spaniards, the Ojibwa, and the Guahibo. In order to fully resolve the haplotype/allele diversity at DYS413 and at DXS8175, all subjects were reexamined by the specific assays for these loci (Malaspina et al. 1997). The DXS8174-specific assay was performed on all females, as well as on all males carrying a nucleotide substitution that suppresses PCR for this locus when examined by the method of Mathias et al. (1994). The Basques, the southern Spaniards, the Ojibwa, and the Guahibo were typed only with the specific assays.

Statistical Analysis

Values (h) for the heterozygosity of the X-linked loci and for the computationally equivalent Y-haplotype diversity were estimated as $h = 1 - \Sigma x_i^2$, where x_i is the estimated frequency of the ith allele/haplotype in the system (Nei 1987). The STRUC program (Raymond and Rousset 1995) was used to evaluate the degree of association between alleles at the DXS8175 and the DXS8174 loci. Analysis of molecular variance was performed by use of the program AMOVA, version 1.55 (Excoffier et al. 1992), and by use of the method described in Michalakis and Excoffier (1996). Two hierarchical levels (individuals into populations and populations into groups) were considered. For DYS413, patterns with a single band (fig. 1*A* and *B,* lane 7) were interpreted as overlapping PCR products. Two allelic series were assumed, one producing the larger fragment of each pattern and the other producing the smaller fragment (Mathias et al. 1994; Goldstein et al. 1996). Although some overlapping of allele sizes must exist, that assumption was necessary for the pairwise comparisons preliminary to the estimation of genetic variances. The error thus generated, however, seems unlikely to systematically bias the results. In practice, for the Y-linked polymorphism, the distance matrix used in AMOVA contained minimum estimates of the sum of the squared differences in the repeat numbers over the two series. For DXS8174 and DXS8175 the distance between any two X alleles was simply the squared difference in the number of repeats. This analysis provided estimates of variance components and of *F*-statistic analogs representing the correlation of haplotypes at different levels of 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) (for a thorough discussion, see Excoffier et al. 1992). When no pairwise microsatellite allelic distances were used, the Φ-statistics became the usual multiallelic *F*-statistics (Wright 1965; Long 1986). The significance levels of the components of variance and the corresponding *F*- or Φ-statistics were obtained by comparison of the actual values with the distribution of 1,000 values obtained by randomization. The nonparametric Mann-Whitney test was used to assay the equality of heterozygosities, haplotype diversities, and within-population variances.

Results

DYS413

The screening of 30 human populations revealed many new haplotypes (table 1), as compared with previous analyses (Mathias et al. 1994; Ruiz Linares et al. 1996). In comparison with the original YCAIII assay, the DYS413-specific assay revealed many short bands, previously underscored, since they migrated together with or faster than the PCR products from other loci of the CAIII family. In table 1, haplotypes are reported as the number of repeats in each PCR product. It is worth noting that the overall variability of this system results from two allelic series. However, several males displayed a single band (e.g., see fig. 1*A* and *B,* lane 7). Possible explanations for these haplotypes include the presence of two loci of identical length, the presence of a mutation that suppresses PCR from one of the loci, or the absence of one of the loci. Among these possibilities, the presence of overlapping PCR products is favored, on the basis of a relatively high band intensity.

A total of 13, 27, and 16 haplotypes were found among northern Europeans, southern Europeans, and western Asians, respectively. In each of these groups, the range of PCR sizes was never <9 repeats. Overall, Europeans showed haplotypes spanning 12–26 repeats (28 different patterns). The 22/22-repeat haplotype was observed in all the populations, although with a 10-fold fluctuation in frequency. Europeans shared with western

Asians a set of haplotypes corresponding to low-molecular-weight PCR products. In particular, the 17/17-repeat haplotype commonly was observed in populations from the Mediterranean area.

With regard to individual populations, the Basques showed the highest frequency of the 23/23-repeat haplotype (74.5%) and the lowest haplotype-diversity value, thus confirming that this population is at an extreme of the European frequency distribution of Y-chromosomal haplotypes (Santachiara Benerecetti et al. 1994; Lucotte and Hazout 1996; Santachiara Benerecetti and Semino 1996). With the exception of the Basques (haplotype-diversity value of .43) and of the southern Spaniards (haplotype-diversity value of .66), all European populations showed haplotype-diversity values >.80. A value >.80 also was shown by the western Asians.

The Pakistani and the Egyptian populations shared with the European and the western Asian populations high haplotype numbers and high diversity values. In particular, the Egyptians, with 16 haplotypes and a diversity value of .88, were well separated from the rest of the Africans. Western and southern Africans displayed 17 and 11 haplotypes, respectively, with PCR sizes spanning 6 and 7 repeats, respectively. In the pool of sub-Saharan Africans, a restricted range of 18–24 repeats was found. In most African populations, a single haplotype accounted for a frequency >40%. This was the 22/20-repeat haplotype in the Bantu and the Tali from Cameroon, in the Fulbe, the Mossi, and the Rimaibe from Burkina Faso, and in the Khwe from South Africa; the 23/21-repeat haplotype in the Ouldeme; the 21/20-repeat haplotype in the Fali and the Rimaibe; and the 24/24-repeat haplotype in the Fulbe from Cameroon. The Ouldeme from Cameroon could be clearly differentiated from the other western African populations for the exclusive presence, at a relatively high frequency (~30%), of the 23/19-repeat haplotype. With few exceptions, all of the haplotypes that were very common in sub-Saharan Africans were rare among Europeans. The haplotype-diversity values of sub-Saharan Africans invariably were <.80 and fell below .40 for the Fulbe of Burkina Faso and the Ewondo of Cameroon.

The small group of Chinese Li was distinctive, since this group showed a high frequency (21.4%) of the 20/17-repeat haplotype, which has never been found elsewhere. Native Americans showed 18 haplotypes, spanning 17–26 repeats, with notable frequencies of high-molecular-weight DYS413 bands, which were rare or absent elsewhere. Among Native Americans, the Ojibwa were distinctive because of the relatively high frequency (37.5%) of the 23/23-repeat haplotype, which also was found frequently among Europeans. Similar to the analysis of the DYS199 polymorphism, this finding lends support to the hypothesis that the Europeans' gene flow into the Ojibwa group was mostly male specific (Scozzari et al. 1997). Overall, Native Americans showed relatively high values of haplotype diversity, similar to those found in Europeans. Comparison of sub-Saharan African populations versus the rest of the world populations showed a significantly reduced haplotype diversity in sub-Saharan Africa (Mann-Whitney $P < .001$).

DXS8175

Percent frequencies of DXS8175 alleles are shown in table 2, together with heterozygosity values. Six different alleles were found at DXS8175, among northern Europeans, southern Europeans, and northern Africans, whereas only four alleles were found in both western and southern Asians. Overall, in these groups the maximum range of PCR sizes was 9 repeats, and the overall number of repeats was within a range of 10–18. Heterozygosity in each population was invariably <.69.

Ten alleles, spanning 10–20 repeats, were found among sub-Saharan Africans. In both western and southern Africans, the 14-, 15-, or 16-repeat alleles were again the most common alleles, but seven additional alleles were found. The 13- and 19-repeat alleles, which have never been observed elsewhere, were found in both sub-Saharan African groups. It is worth noting that the DXS8175 alleles of 18–20 repeats could be identified only with the specific assay (Malaspina et al. 1997), since they produced bands that overlapped with DYS413 products in the original PCR assay (Mathias et al. 1994) (see fig. 1*A* and *C*, lanes 8 and 9). Heterozygosity values in sub-Saharan African populations were always >.67 and were >.70 in 10 of 12 populations.

The Chinese Li were characterized by a low heterozygosity value and by the shortest range of allelic sizes (4 repeats). Native Americans showed five alleles, spanning 12–18 repeats, and the lowest heterozygosity values. In contrast with DYS413, the heterozygosity values of the sub-Saharan African populations were significantly greater than the rest of the world populations (Mann-Whitney $P < .0001$).

DXS8174

The results of the analysis of the DXS8174-length polymorphism are shown in table 3. Overall, six alleles were found. Of these, the 10-repeat allele was observed only in two European populations, the northern Africans, and the Chinese Li. A common feature of all samples was the presence of the 11-repeat allele at high frequencies, within the range of 65%–100%. The sub-Saharan Africans displayed two large-sized alleles (14 and 15 repeats) not found in any other group. The Chinese Li harbored four alleles and showed the highest heterozygosity value, whereas the Native Americans, with only two alleles (11 and 12 repeats), had the lowest

Table 1

Haplotype Frequencies and Haplotype-Diversity Values for the DYS413 Polymorphism

Population	n[a]	Frequency of Haplotype[b] (%)																	
		17/14	17/16	17/17	18/16	18/17	18/18	20/17	20/18	20/20	21/19	21/20	21/21	22/12	22/17	22/18	22/19	22/20	22/21
Northern European:																			
English	20												5.0						10.0
Danish	35			8.6									5.7						8.6
Southern European:																			
Basque	55			1.8						1.8			1.8						5.5
Southern Spaniard	36									2.8				2.8					5.6
Venetian	20			20.0		5.0							5.0						
Central Italian	84			10.7			1.2			2.4		2.4	4.8					3.6	15.5
Southern Italian	81		1.2	19.8		3.7				1.2		1.2	4.9					1.2	3.7
Sardinian	97			11.3						2.1	1.0		34.0					2.1	7.2
Greek	51			23.5	7.8					2.0	2.0		5.9			2.0			11.8
Western Asian:																			
Turkish	20			10.0			5.0			10.0		5.0							10.0
Eastern Arabian peninsula	32	3.1		3.1					3.1	12.5					3.1			6.3	31.3
Southern Asian:																			
Pakistani	20									10.0			10.0					15.0	5.0
Northern African:																			
Egyptian	48			6.3			2.1			6.3			4.2						25.0
Western African:																			
Bamileke	51									5.9		11.8					3.9	70.6	
Ewondo	31									6.5		3.2						83.9	6.5
Ouldeme	23									4.3									8.7
Daba	18											27.8						5.6	27.8
Fali	39											46.2						28.2	5.1
Tali	15									6.7		33.3						40.0	
Fulbe Cameroon	17											17.6						5.9	
Fulbe Burkina Faso	20											5.0	5.0					80.0	
Mossi	56									7.1		28.6					1.8	58.9	
Rimaibe	42									2.4		52.4						40.5	
Southern African:																			
!Kung	63						4.8			4.8	1.6	7.9	31.7					27.0	3.2
Khwe	26											7.7	15.4					61.5	11.5
Eastern Asian:																			
Chinese Li	28			3.6				21.4				7.1			7.1			53.6	7.1
Native American:																			
Ojibwa	16																		
Seminole	27												3.7						18.5
Mayan	8														12.5			12.5	
Guahibo	28																3.6		
Total	1,107																		

[a] No. of Y chromosomes analyzed.
[b] Haplotypes are indicated by the no. of repeats in the observed bands.
[c] Calculated as $1 - \Sigma x_i^2$, where x_i is the frequency of the ith haplotype.

heterozygosity values. The DXS8174 heterozygosity values for sub-Saharan Africans were not significantly different from those of the rest of the world populations (Mann-Whitney $P = .54$).

As reported elsewhere (Malaspina et al. 1997), several individuals who did not show any PCR product from locus DXS8174 when tested with the original YCAIII assay (Mathias et al. 1994) were shown to display the PCR product when tested with the DXS8174-specific assay. This so-called null allele contributes a further heterogeneity at this locus. By combining the DXS8174-specific primers (Malaspina et al. 1997) with those originally described elsewhere (Mathias et al. 1994), we analyzed alleles of this type that were sampled from among southern Europeans, western Africans, southern Africans, and the Chinese and found that, in all cases, the discrepancy between the two assays was due to a sequence variation in the stretch corresponding to primer YCAIIIA. In order to assess the world distribution of this molecular type, we reexamined all the males carrying the null allele, with the DXS8174-specific assay. Interestingly, all of the X chromosomes with the DXS8174 null allele turned out to be 11 repeats long, producing the frequencies of the 11-null haplotype reported in table 3 as percentages of the entire population sample.

22/22	23/14	23/18	23/19	23/20	23/21	23/22	23/23	24/19	24/20	24/21	24/22	24/23	24/24	25/21	25/22	25/23	25/24	25/25	26/22	26/23	h^c
15.0							35.0			5.0	5.0	10.0		5.0		10.0					.815
5.7					8.6	8.6	31.4					11.4	2.9			8.6					.844
1.8				1.8	1.8	9.1	74.5														.431
13.9					2.8	8.3	55.6				2.8	2.8			2.8						.657
15.0					5.0	10.0	30.0					10.0									.820
8.3					8.3	9.5	22.6			2.4	2.4	4.8				1.2					.882
16.0				1.2	6.2	2.5	22.2			1.2	1.2	8.6	1.2					1.2	1.2		.867
11.3					6.2	3.1	14.4			1.0		4.1	1.0		1.0						.825
19.6					3.9	3.9	9.8					3.9					3.9				.866
15.0				5.0	5.0	10.0	20.0			5.0											.885
21.9						6.3				3.1	6.3										.822
40.0					5.0	5.0	5.0									5.0					.785
14.6	2.1				12.5	4.2	2.1	2.1		2.1	2.1	4.2	2.1	8.3							.884
2.0				2.0									3.9								.481
																					.287
			30.4		43.5		4.3	8.7													.699
22.2					5.6	5.6				5.6											.744
					17.9		2.6														.672
6.7									6.7	6.7											.711
					5.9	11.8						5.9	52.9								.664
5.0							5.0														.350
				1.8					1.8												.565
						4.8															.559
15.9		1.6					1.6														.789
									3.8												.577
																					.651
18.8					6.3	25.0	37.5								12.5						.742
				7.4	3.7	7.4	18.5				11.1	11.1			3.7		3.7	3.7		7.4	.883
				12.5	25.0		12.5				12.5	12.5									.844
3.6							7.1						46.4		10.7	28.6					.684

Analysis of Linkage Disequilibrium between DXS8175 and DXS8174

In order to estimate the amount of disequilibrium, if any, between the two X-linked microsatellite loci and the presence of preferential combinations, we analyzed the joint distribution of the DXS8175 and the DXS8174 alleles in each population. An application of the two-tailed Fisher's exact test to this analysis failed to show any evidence of preferential association between length alleles, in any population examined. The two loci are located in Xp22 and are separated by the discriminant breakpoint BXP21 (Schaefer et al. 1993). In the high-resolution genetic map of Dib et al. (1996), the two closest markers on either side of BXP21 (DXS996 and DXS1223) are 1.9 cM apart. Even when the location of DXS8175 and DXS8174 is assumed to be within this genetic interval, the lack of disequilibrium is compatible with randomization due to recombination.

Analysis of Molecular Variance

In order to assess the proportion of the total variability accounted for by variation among populations and among groups of populations, we applied AMOVA to each of the polymorphisms examined. In

Table 2

Allele Frequencies and Heterozygosity Values for the DXS8175 Polymorphism

Population	n[a]	Frequency of Allele[b] (%)										h[c]
		10	12	13	14	15	16	17	18	19	20	
Northern European:												
English	20				40.0	30.0	25.0		5.0			.685
Danish	61		1.6		21.3	34.4	41.0	1.6				.668
Southern European:												
Basque	95				16.8	44.2	35.8	3.2				.647
Southern Spaniard	36				19.4	38.9	38.9	2.8				.659
Venetian	20		5.0		10.0	45.0	40.0					.625
Central Italian	110		2.7		18.2	43.6	33.6	1.8				.662
Southern Italian	89		1.1		21.3	37.1	40.4					.653
Sardinian	121		1.7		15.7	47.9	34.7					.625
Greek	51	2.0			15.7	43.1	37.3	2.0				.650
Western Asian:												
Turkish	19				5.3	52.6	42.1					.543
Eastern Arabian peninsula	32				15.6	40.6	40.6	3.1				.645
Southern Asian:												
Pakistani	20		15.0		5.0	35.0	45.0					.650
Northern African:												
Egyptian	48		2.1		10.4	27.1	45.8	12.5	2.1			.689
Western African:												
Bamileke	52	5.8	15.4	3.8	23.1	7.7	30.8	11.5	1.9			.804
Ewondo	69	5.8	8.7		26.1	10.1	29.0	10.1	5.8	4.3		.811
Ouldeme	63	4.8	1.6		15.9	12.7	50.8	12.7	1.6			.682
Daba	34	17.6	5.9		14.7	14.7	23.5	20.6		2.9		.824
Fali	79	2.5	3.8		20.3	8.9	46.8	12.7	2.5	2.5		.712
Tali	57	3.5	10.5		15.8	17.5	35.1	15.8	1.8			.784
Fulbe Cameroon	56	5.4		1.8	32.1	14.3	37.5	8.9				.724
Fulbe Burkina Faso	58		3.4		6.9	36.2	36.2	17.2				.702
Mossi	83	6.0	2.4	1.2	19.3	14.5	37.3	16.9	1.2		1.2	.769
Rimaibe	71	14.1	2.8		19.7	8.5	32.4	16.9	1.4	2.8	1.4	.799
Southern African:												
!Kung	104		33.7	1.0	26.0	1.0	37.5	1.0				.678
Khwe	38		21.1		10.5	5.3	42.1	5.3	13.2	2.6		.744
Eastern Asian:												
Chinese Li	46				4.3	41.3	50.0	4.3				.576
Native American:												
Ojibwa	47				10.6	19.1	70.2					.459
Seminole	27		3.7			44.4	48.1		3.7			.568
Mayan	58		1.7		1.7	44.8	51.7					.531
Guahibo	30					86.7	13.3					.231
Total	1,694											

[a] No. of X chromosomes analyzed.
[b] Alleles were named in accordance with the repeat number.
[c] Calculated as $1 - \Sigma x_i^2$, where x_i is the frequency of the ith allele.

a first run, the nine main groups reported in tables 1–3 were considered (see the "All Groups" column in table 4). Application of AMOVA to DYS413 showed that an appreciable (~10%) and significant amount of the total variance is attributable to differences among human groups and among populations within groups, with or without consideration of molecular distances among haplotypes. The overall worldwide F_{ST} and Φ_{ST} values, on the basis of the 30 populations, were .21 and .22, respectively. AMOVA, as applied to DXS8175 and DXS8174, produced lower estimates for the among-groups and the among-populations/within-groups proportions of total variance, both with and without consideration of molecular differences between alleles. The F_{ST} estimates were concordant with expectations based on a lower effective population size for DYS413 relative to both DXS8175 and DXS8174. When allele/haplotype molecular dif-

Table 3

Allele Frequencies and Heterozygosity Values for the DXS8174 Polymorphism

Population	n[a]	Frequency of Allele[b] (%)						h[c]	Frequency of 11-Null Haplotype[d] (%)
		10	11	12	13	14	15		
Northern European:									
English	20		65.0	35.0				.455	20.0
Danish	35	2.9	82.9	14.3				.292	11.4
Southern European:									
Basque	55	3.6	80.0	14.5	1.8			.337	nt
Southern Spaniard	36		66.7	27.8	5.6			.475	nt
Venetian	20		90.0	10.0				.180	45.0
Central Italian	77		84.4	14.3	1.3			.267	21.8
Southern Italian	83		84.3	13.3	2.4			.271	17.1
Sardinian	91		85.7	12.1	2.2			.250	17.6
Greek	37		67.6	32.4				.438	13.6
Western Asian:									
Turkish	20		75.0	25.0				.375	0
Eastern Arabian peninsula	32		90.6	9.4				.170	6.3
Southern Asian:									
Pakistani	20		100.0					.000	0
Northern African:									
Egyptian	48	2.1	77.1	14.6	6.3			.380	25.0
Western African:									
Bamileke	51		74.5	5.9	19.6			.403	5.9
Ewondo	31		83.9	12.9	3.2			.279	6.5
Ouldeme	31		74.2	9.7	16.1			.414	17.4
Daba	32		71.9	12.5	12.5		3.1	.451	0
Fali	39		71.8	10.3	17.9			.442	7.7
Tali	31		83.9		16.1			.271	6.7
Fulbe Cameroon	36		88.9		11.1			.198	6.3
Fulbe Burkina Faso	44		88.6		11.4			.201	27.8
Mossi	49		81.6	6.1	12.2			.315	14.3
Rimaibe	38		81.6		18.4			.301	18.4
Southern African:									
!Kung	64		90.6	1.6	1.6		6.3	.174	0
Khwe	34		73.5	5.9	5.9	8.8	5.9	.441	11.5
Eastern Asian:									
Chinese Li	28	7.1	71.4	3.6	17.9			.452	7.1
Native American:									
Ojibwa	42		90.5	9.5				.172	nt
Seminole	27		85.2	14.8				.252	0
Mayan	58		87.9	12.1				.212	0
Guahibo	21		81.0	19.0				.308	nt
Total	1,230								

[a] No. of X chromosomes analyzed.
[b] Alleles were named in accordance with the repeat number.
[c] Calculated as $1 - \Sigma x_i^2$, where x_i is the frequency of the ith allele.
[d] See Results section in text. nt = not tested.

ferences were considered, the Φ_{ST} for DYS413 became ~4- and 7-fold larger than the Φ_{ST} for the two X loci.

In order to examine the role of population subdivision, for Africans and non-Africans, we applied AMOVA separately to Europeans (two groups—northern and southern Europeans) and to sub-Saharan Africans (two groups—western and southern Africans) (table 4). For DYS413, it is worth noting that in both cases the proportion of variation among populations/within groups largely exceeded that among groups. These results were replicated when the groups of western Asians, southern Asians, and northern Africans were considered together with the European groups (five Caucasoid groups) (F_{SC} = .08 vs. F_{CT} = .01; Φ_{SC} = .09 vs. Φ_{CT} = −.01), when sub-Saharan Africans were grouped according to a national criterion (F_{SC} = .20 vs. F_{CT} = −.01; Φ_{SC} = .30 vs. Φ_{CT} = −.02) or to a linguistic

Table 4

Hierarchical Analysis of Molecular Variance for the DYS413, DXS8175, and DXS8174 Polymorphisms

Polymorphism and Type of Comparison	All Groups		Europeans		Sub-Saharan Africans	
	Variance[a] (%)	*F*-statistic	Variance[a] (%)	*F*-statistic	Variance[a] (%)	*F*-statistic
Without interallele/haplotype distances:						
DYS413:						
Among groups	9.7	F_{CT} = .10***	−2.5	F_{CT} = −.03	1.0	F_{CT} = .01
Among populations/within groups	10.9	F_{SC} = .12***	8.2	F_{SC} = .08***	18.6	F_{SC} = .19***
Within populations	79.4 [.37]	F_{ST} = .21***	94.3 [.40]	F_{ST} = .06***	80.3 [.31]	F_{ST} = .20***
DXS8175:						
Among groups	6.1	F_{CT} = .06***	.9	F_{CT} = .01***	5.1	F_{CT} = .05***
Among populations/within groups	2.0	F_{SC} = .02***	−.7	F_{SC} = −.01	1.4	F_{SC} = .02***
Within populations	92.0 [.34]	F_{ST} = .08***	99.8 [.33]	F_{ST} = .00	93.5 [.38]	F_{ST} = .07***
DXS8174:						
Among groups	2.5	F_{CT} = .03***	−.9	F_{CT} = −.01	2.1	F_{CT} = .02*
Among populations/within groups	1.2	F_{SC} = .01	2.3	F_{SC} = .02*	.7	F_{SC} = .01
Within populations	96.3 [.15]	F_{ST} = .04***	98.6 [.16]	F_{ST} = .01*	97.3 [.16]	F_{ST} = .03
		Φ-statistic		Φ-statistic		Φ-statistic
With interallele/haplotype distances:						
DYS413:						
Among groups	11.7	Φ_{CT} = .12**	5.7	Φ_{CT} = .06	−5.6	Φ_{CT} = −.06
Among populations/within groups	10.6	Φ_{SC} = .12***	9.0	Φ_{SC} = .10***	31.7	Φ_{SC} = .30***
Within populations	77.7 [4.75]	Φ_{ST} = .22***	85.4 [7.82]	Φ_{ST} = .15***	73.9 [1.22]	Φ_{ST} = .26***
DXS8175:						
Among groups	2.7	Φ_{CT} = .03	−.3	Φ_{CT} = .00	4.8	Φ_{CT} = .05
Among populations/within groups	2.4	Φ_{SC} = .03**	−1.0	Φ_{SC} = −.01	2.7	Φ_{SC} = .03**
Within populations	94.9 [1.98]	Φ_{ST} = .05***	101.3 [.76]	Φ_{ST} = −.01	92.6 [3.50]	Φ_{ST} = .07***
DXS8174:						
Among groups	2.1	Φ_{CT} = .02*	−1.9	Φ_{CT} = −.02	−.2	Φ_{CT} = .00
Among populations/within groups	.7	Φ_{SC} = .01	2.3	Φ_{SC} = .02*	.1	Φ_{SC} = .00
Within populations	97.2 [.39]	Φ_{ST} = .03***	99.5 [.20]	Φ_{ST} = .01	100.1 [.68]	Φ_{ST} = .00

[a] For the within-population variance, the absolute value is given in brackets.
* $P < .05$.
** $P < .01$.
*** $P < .001$.

criterion (F_{SC} = .15 vs. F_{CT} = .05; Φ_{SC} = .39 vs. Φ_{CT} = −.18). Moreover, population subdivisions were more apparent when a measure of molecular distances among haplotypes was introduced (for the Europeans, F_{ST} = .06 vs. Φ_{ST} = .15; for sub-Saharan Africans, F_{ST} = .20 vs. Φ_{ST} = .26). A similar result also was obtained for the four Native American populations (F_{ST} = .14 vs. Φ_{ST} = .27). This indicates that a substantial fraction of the DYS413 variability among populations is due to divergent molecular types, denoting an ongoing radiation process.

For DXS8175 and DXS8174, all population groupings reported above produced lower F_{SC} and Φ_{SC} values, as compared with DYS413. Native Americans showed an F_{ST} of .19 for DXS8175, which mainly is attributable to the dramatic variation of the 15- and 16-repeat allele frequencies among the four populations (also see table 2).

Comparative Within-Population Variation of Homologous X and Y Microsatellites

The world pattern of within-population variation of DYS413, DXS8175, and DXS8174 was analyzed by plotting of the variances within each population, for the three possible pairs of loci (fig. 2). Within-population variances were computed by division of the within-population sum of squares, inclusive of inter-haplotype or inter-allele distances, by the appropriate df. Variances at the two X-linked loci (fig. 2*A*) were correlated positively (r = .63; $P < .01$), showing a coherent accumulation of variation. In view of the lack of linkage disequilibrium between DXS8174 and DXS8175, the observed trend supports evidence of a common mechanism for the accumulation of variation for both loci that produces high values of within-population variance in sub-Saharan African populations. This also clearly emerges

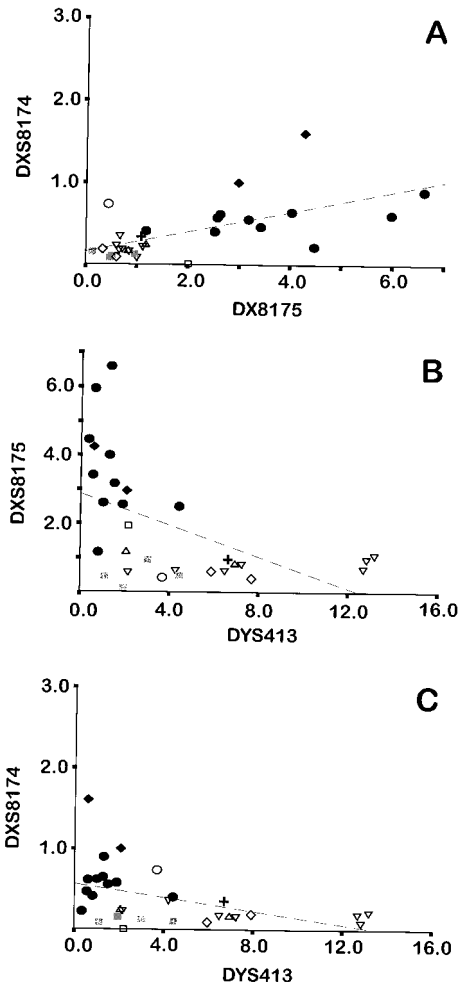

Figure 2 Plots of within-population variances for the reported loci. *A*, DXS8174 versus DXS8175. *B*, DXS8175 versus DYS413. *C*, DXS8174 versus DYS413. Unblackened triangles denote northern Europeans; inverted unblackened triangles denote southern Europeans; unblackened circles denote eastern Asians; unblackened diamonds denote western Asians; unblackened squares denote southern Asians; boldface plus signs (+) denote northern Africans; blackened circles denote western Africans; blackened diamonds denote southern Africans; and grey squares denote Native Americans. The dashed line represents the interpolated regression line. Note the scale difference on the vertical axes in panels *B* and *C*.

from the absolute values of the within-population molecular variances, reported in table 4.

The plot of DYS413 versus DXS8175 within-population variances (fig. 2*B*) showed a clear negative correlation ($r = -.50$; $P < .01$) that well discriminates sub-Saharan African populations from the rest of the world populations. All variances for the sub-Saharan African populations are plotted at the left edge of the graph (low Y diversity and medium-to-high X diversity); those for the Native Americans at the bottom-left corner (low Y diversity and low X diversity); and those for the Caucasoids at the bottom edge (low X diversity and medium-to-high Y diversity). Finally, the plot of DYS413 versus DXS8174 within-population variances (fig. 2*C*) showed a moderately negative correlation ($r = -.42$; $P = .02$), with the southern-African populations outlying the overall trend, owing to their minimal variance for DYS413, as opposed to their high variance for DXS8174.

Sub-Saharan Africans and Europeans constituted the vast majority of populations analyzed. It is worth noting that all correlations were lost when these two groups of populations were examined separately from each other. This indicates that the overall correlations were the result of two major clusters characterized by widely discrepant values of variances. A lower within-population variance at DYS413, for sub-Saharan Africans as compared with Europeans, is also apparent in table 4.

We used the nonparametric Mann-Whitney test to assay the equality of within-population variances for sub-Saharan Africans versus the rest of the world populations. For DYS413 sub-Saharan Africans exhibited significantly lower within-population variances ($P = .0001$), whereas for both DXS8175 and DXS8174 they exhibited significantly greater within-population variances ($P < .0001$ and $P = .0002$, respectively).

Discussion

We previously had found that the primers designed by Mathias et al. (1994) to amplify a Y-linked polymorphic locus (YCAIII) indeed are able to detect a family (CAIII) of CA repeat-containing sequences with homologous members on the X and Y chromosomes (Malaspina et al. 1997). In addition to the Y members, two of the X-linked sequences also showed length polymorphism, thus offering a novel opportunity to address questions regarding both the accumulation of variation for loci with different transmission patterns and the relative evolutionary rates of the two chromosomes. In fact, the number and size ranges of alleles/haplotypes observed in extant populations can be taken as the net result of the opposing forces of mutation and fixation in increasing and reducing, respectively, genetic variability. The present survey showed an overall higher allelic diversity for the Y-linked loci than for the X-linked loci. When it is considered that the antiquity of DYS413 is not greater than that of the two X-linked members, our world data suggest that the accumulation of variants on the Y chromosome proceeded at a rate faster than that for the X chromosome. Shimmin et al. (1993) produced data in favor of a higher mutation rate for the Y chromosome, owing to its exclusive male transmission. Additional locus-specific factors might affect DYS413 mutability. One or more of the few substitutions in the flanking regions, which distinguish DYS413 from the other CAIII-family members, might influence the

mutation rate, as has been observed for minisatellites (Monckton et al. 1994; Andreassen et al. 1996), or the event that isolated DYS413 on the Y chromosome could have involved a long CA-repeat allele that is more prone to further changes (Primmer et al. 1996). As to the X-linked microsatellites, the present data do not support a high mutation rate, at least for DXS8174. In fact, a worldwide complete association between a length allele and a sequence-variant allele was observed. In addition, constraints on the overall length of X-linked microsatellites cannot be ruled out. We analyzed the 80 X-linked CA repeats and a representative sample (10%) of the autosomal CA repeats reported by Gyapay et al. (1994) and found shorter size ranges for the former group (Mann-Whitney $P < .02$). On the other hand, specific limiting factors for DXS8175 and DXS8174, in the gene-dense region where these two loci are located (Ferrero et al. 1995), can hardly be reconciled with the lack of linkage disequilibrium between the two loci and with their coherent accumulation of variation (fig. 2*A*). Only the unlikely occurrence of independent stabilizing selective factors for each of the two loci would be compatible with these observations.

This survey included populations from nine human groups from four continents and showed distinctive features of the X-linked versus the Y-linked microsatellite variation. First, different quotas of among-population variation were found for loci on the X vs. the Y chromosome, and between groups of populations from different continents, for the Y chromosome. Second, different within-population variances were found between the two X loci, between the X and the Y loci, and between groups of populations. These differences, in part, were due to different ranges of allele size, across loci and across groups of populations.

As to among-population variation, it is clear that, in all major groups and, in particular, among sub-Saharan Africans, there was a trend for the DYS413 haplotypes to cluster in a population-specific fashion. Similar to those obtained with the YCAII microsatellite polymorphism (Ciminelli et al. 1995), our results indicate that a large fraction of among-population/within-group variation is the result of both the variation of haplotype frequencies and the accumulation of peculiar molecular types during the radiation of different populations. Our data show that in sub-Saharan Africa this process led to notable divergence among populations geographically close to each other and/or sharing linguistic affinities. Hammer (1995) and Hammer et al. (1997) observed that high Y-chromosomal F_{ST} values are expected, on the basis of the lower effective population size of this chromosome relative to the other chromosomes. Demographic factors such as male-specific bottlenecks, high variance of male reproductive success, and a reduced male-specific gene flow may contribute high F_{ST} and Φ_{ST} values for the Y chromosome, and their role may be stronger in sub-Saharan African societies. A sharply different picture emerges for the X-linked loci DXS8175 and DXS8174. These two loci have a limited power in revealing interpopulation differences. The data indicate a low level of geographic or ethnic structuring of populations, for DXS8175, and an even lower level for DXS8174, with F_{ST} values below the lowest boundary of the F_{ST} range reported for other microsatellites (Deka et al. 1995*a*).

In summary, the loci examined in this study showed an opposite relation, as compared with the data of Bowcock et al. (1994), who reported an overall negative correlation between F_{ST} and allelic diversity, by examining several autosomal microsatellite loci. They attributed this to the confounding effect of recurrent mutations at loci with elevated mutation rates. We argue that the positive correlation observed in our data set is to be attributed to the male specificity of DYS413. In particular, a fast progression towards fixation of the Y chromosome can explain the rapid accumulation of divergent types, also in the presence of elevated mutation rates.

With regard to within-population variation, for DXS8175 and DXS8174 the highest heterozygosities and variances were seen in sub-Saharan Africans (tables 2–4 and fig. 2). A similar pattern has been seen with mtDNA (Vigilant et al. 1991) and mini- and microsatellite loci (Bowcock et al. 1994; Armour et al. 1996; Tishkoff et al. 1996) and has been interpreted as an indication of an African origin of modern *Homo sapiens*. For the Y-linked polymorphism, sub-Saharan Africans exhibited the lowest levels of haplotype diversity and variance, as opposed to the levels for the X-linked loci. Moreover, on the whole, a shorter range of allele sizes was observed in the sub-Saharan Africans, relative to the European populations. A similar pattern has been seen in other studies of Y chromosome–specific polymorphisms and might reflect a small effective population size of the Y chromosome, in African societies (Torroni et al. 1990; Spurdle et al. 1994; Jobling and Tyler-Smith 1995), possibly as a result of male reproductive habits (Konotey-Ahulu 1980). In a large survey, Hammer et al. (1997) used the tetranucleotide microsatellite DYS19 and showed that African populations do not display such a reduced diversity. The presence of only seven alleles at the DYS19 system and their recurrent origin on different haplotypic backgrounds may not reveal the full spectrum of variability in non-African populations. In this context, the higher number of DYS413 haplotypes seems more permissive for extremely short and extremely long sizes.

Studies of additional Y-chromosomal polymorphisms and human populations will be necessary to confirm the inverse relationship between X and Y within-population variances observed in this study. If this is the case, mod-

els of increasing complexity can be used to explain such a relationship. Under the simplistic assumption that within-group variation directly reflects divergence time, a reduced antiquity of sub-Saharan African populations, limited to the male-specific lineage, can be hypothesized, possibly as a result of a male-specific bottleneck or a selective sweep that occurred after the African/non-African split. However, a selective sweep contrasts with the findings of Hammer (1995) and Hammer et al. (1997). Following these authors' extensive discussions, additional hypotheses need to integrate the simple model of accumulation of variants in a population of constant size with information on demography and population history. In addition, the specific pattern of microsatellite mutation (for a review, see Freimer and Slatkin 1996) must be taken into account. The branching pattern of Caucasoid mtDNA types analyzed by Di Rienzo and Wilson (1991) revealed a so-called "star phylogeny" and was in agreement with the model of a recent and fast demographic expansion of non-African populations. The consideration of such a model has important implications for the coalescent theory, since, under these circumstances, the variation that is produced constantly has a higher probability of being retained in the expanded population. In the case of microsatellites, this would imply a higher probability of the retention of alleles with the most extreme sizes, since these alleles are produced in low amounts, as compared with intermediate sizes. This mode of production of new variants especially would affect microsatellites of the Y chromosome, for which the stepwise mutational model has been favored (Ciminelli et al. 1995; Cooper et al. 1996). Thus, populations in which variability is retained more may display a wider range of Y-chromosomal microsatellite alleles, which may not be necessarily attributed to a greater antiquity. Moreover, in this context, we cannot exclude that a demographic transition involving a rapid change in population size affected the male and female lineages to a different extent (for an example, see Terrenato et al. 1979). Finally, contingent factors of the European history must not be underestimated (Cavalli-Sforza et al. 1994). Zerjal et al. (1997) recently have described a Y-specific northern-Asian contribution to the northern-European gene pool, and Semino et al. (1996) have produced Y chromosome data favoring a neolithic demic diffusion into Europe. The variation of DYS413 in regions of Africa other than the three nation-states of the west and the south, as well as in large parts of Asia, is still unexplored, and the possibility that the higher variation of this locus in Europe is the result of admixture is still open.

In brief, the results of this study support the following conclusions. First, a higher evolutionary rate appears to characterize the Y-linked microsatellite sequences relative to the X-linked sequences. Second, a marked rate of divergence for the Y-linked microsatellite loci characterizes all human groups, especially African populations. Third, the higher within-population variability of the X chromosome, as compared with the Y chromosome, in sub-Saharan Africa and of the Y chromosome, as compared with the X chromosome, in Europe prompts caution in the interpretation that levels of within-population variation directly reflect divergence time. Not only can demographic factors influence the level of within-population variation, but also the possibility that they could act in a sex-specific manner has to be taken into account.

Acknowledgments

We are indebted to all the aboriginal peoples for their helpful collaboration, which has made this study possible. We are very grateful to Prof. G. Spedini for providing the blood samples from Cameroon. We also thank Profs. P. F. Pignatti, G. Guanti, and N. Anagnou for providing subsets of the DNA samples. We thank Dr. C. Tyler-Smith for sharing data with us prior to publication. This work was supported by P. F. Beni Culturali contracts 96.01182.PF36 (to R.S.) and 96.01192.PF36 (to L.T.).

References

Andreassen R, Egeland T, Olaisen B (1996) Mutation rate in the hypervariable VNTR g3 (D7S22) is affected by allele length and a flanking DNA sequence polymorphism near the repeat array. Am J Hum Genet 59:360–367

Armour JAL, Anttinen T, May CA, Vega EE, Sajantila A, Kidd JR, Kidd KK, et al (1996) Minisatellite diversity supports a recent African origin for modern humans. Nat Genet 13: 154–160

Bowcock AM, Kidd JR, Mountain JL, Hebert JM, Carotenuto L, Kidd KK, Cavalli-Sforza LL (1991) Drift, admixture, and selection in human evolution: a study with DNA polymorphisms. Proc Natl Acad Sci USA 88:839–843

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368:455–457

Cai W, Filosa S, Martini G (1994) DNA haplotypes in the G6PD gene cluster studied in the Chinese Li population and their relationship to $G6PD^{Canton}$. Hum Hered 44:279–286

Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. Nature 325:31–36

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

Crow JF (1993) How much do we know about spontaneous human mutation rates? Environ Mol Mutagen 21:122–129

Deka R, Jin L, Shriver MD, Yu LM, DeCroo S, Hundrieser J, Bunker CH, et al (1995*a*) Population genetics of dinucleotide $(dC\text{-}dA)_n \cdot (dG\text{-}dT)_n$ polymorphisms in world populations. Am J Hum Genet 56:461–474

Deka R, Shriver MD, Yu LM, Ferrell RE, Chakraborty R (1995*b*) Intra- and inter-population diversity at short tandem repeat loci in diverse populations of the world. Electrophoresis 16:1659–1664

Di Rienzo A, Wilson AC (1991) Branching pattern in the evolutionary tree for human mitochondrial DNA. Proc Natl Acad Sci USA 88:1597–1601

Dib C, Fauré S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, et al (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 380:152–154

Dorit RL, Akashi H, Gilbert W (1995) Absence of polymorphism at the ZFY locus on the human Y chromosome. Science 268:1183–1185

Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R (1992) Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 12:241–253

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

Ferrero GB, Franco B, Roth EJ, Firulli BA, Borsani G, Delmas-Mata J, Weissenbach J, et al (1995) An integrated physical and genetic map of a 35 Mb region on chromosome Xp22.3-Xp21.3. Hum Mol Genet 4:1821–1827

Figus A, Angius A, Loudianos G, Bertini C, Dessì V, Loi A, Deiana M, et al (1995) Molecular pathology and haplotype analysis of Wilson disease in Mediterranean populations. Am J Hum Genet 57:1318–1324

Freimer NB, Slatkin M (1996) Microsatellites: evolution and mutational processes. In: Wiley J (ed) Variation in the human genome. Ciba Foundation, Chichester, England, pp 51–72

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

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

Grimes BF (1992) Ethnologue: languages of the world, 12th ed. Summer Institute of Linguistics, Dallas

Gyapay G, Morissette J, Vignal A, Dib C, Fizames C, Millasseau P, Marc S, et al (1994) The 1993–94 Généthon human genetic linkage map. Nat Genet 7:246–339

Hammer MF (1995) A recent common ancestry for human Y chromosomes. Nature 378:376–378

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

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

Huoponen K, Torroni A, Wickman PR, Sellitto D, Gurley DS, Scozzari R, Wallace DC (1997) Mitochondrial DNA and Y chromosome–specific polymorphisms in the Seminole tribe of Florida. Eur J Hum Genet 5:25–34

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

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

Jorde LB, Bamshad MJ, Watkins WS, Zenger R, Fraley AE, Krakowiak PA, Carpenter KD, et al (1995) Origins and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data. Am J Hum Genet 57:523–538

Kidd JR, Black FL, Weiss KM, Balazs I, Kidd KK (1991) Studies of three Amerindian populations using nuclear DNA polymorphisms. Hum Biol 63:775–794

Kidd KK, Kidd JR (1996) A nuclear perspective on human evolution. In: Boyce AJ, Mascie-Taylor CGN (eds) Molecular biology and human diversity. Cambridge University Press, Cambridge, pp 242–264

Konotey-Ahulu FID (1980) Male procreative superiority index (MPSI): the missing coefficient in African anthropogenetics. Br Med J 281:1700–1702

Long JC (1986) The allelic correlation structure of Gainj- and Kalam-speaking people. I. The estimation and interpretation of Wright's *F*-statistics. Genetics 112:629–647

Lucotte G, Hazout S (1996) Y-chromosome DNA haplotypes in Basques. J Mol Evol 42:472–475

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, Persichetti F, Novelletto A, Iodice C, Terrenato L, Wolfe J, Ferraro M, et al (1990) The human Y chromosome shows a low level of DNA polymorphism. Ann Hum Genet 54:297–305

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

Meroni G, Franco B, Archidiacono N, Messali S, Andolfi G, Rocchi M, Ballabio A (1996) Characterization of a cluster of sulfatase genes on Xp22.3 suggests gene duplications in an ancestral pseudoautosomal region. Hum Mol Genet 5: 423–431

Michalakis Y, Excoffier L (1996) A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. Genetics 142:1061–1064

Monckton DG, Neumann R, Guram T, Fretwell N, Tamaki K, MacLeod A, Jeffreys AJ (1994) Minisatellite mutation rate variation associated with a flanking DNA sequence polymorphism. Nat Genet 8:162–170

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Primmer CR, Ellegren H, Saino N, Pape Møller A (1996)

Directional evolution in germline microsatellite mutations. Nat Genet 13:391–393

Raymond M, Rousset F (1995) An exact test for population differentiation. Evolution 49:1280–1283

Rogers AR, Jorde LB (1996) Ascertainment bias in estimates of average heterozygosity. Am J Hum Genet 58:1033–1041

Ruiz Linares A, Nayar K, Goldstein DB, Hebert JM, Seielstad MT, Underhill PA, Lin AA, et al (1996) Geographic clustering of human Y-chromosome haplotypes. Ann Hum Genet 60:401–408

Santachiara Benerecetti AS, Semino O (1996) Y-chromosome polymorphisms and history of populations. Cell Pharmacol 3:15–17

Santachiara Benerecetti AS, Semino O, Passarino G, Bertranpetit J, Fellous M (1994) The genetic peculiarity of the Basques shown by some Y-specific polymorphisms. Am J Hum Genet Suppl 55:A164

Schaefer L, Ferrero GB, Grillo A, Bassi MT, Roth EJ, Wapenaar MC, van Ommen G-JB, et al (1993) A high resolution deletion map of human chromosome Xp22. Nat Genet 4: 272–279

Scozzari R, Cruciani F, Santolamazza P, Parra E, Sellitto D, Modiano D, Meloni G, et al (1996) Novel tetranucleotide repeat polymorphism in the human adenosine deaminase gene: interethnic comparison of three major human groups. Hum Biol 68:315–320

Scozzari R, Cruciani F, Santolamazza P, Sellitto D, Cole DEC, Rubin LA, Labuda D, et al (1997) mtDNA and Y chromosome–specific polymorphisms in modern Ojibwa: implications about the origin of their gene pool. Am J Hum Genet 60:241–244

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

Shimmin LC, Chang BH-J, Li W-H (1993) Male-driven evolution of DNA sequences. Nature 362:745–747

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457–462

Spurdle A, Jenkins T (1992) The search for Y chromosome polymorphism is extended to Negroids. Hum Mol Genet 1: 169–170

Spurdle AB, Hammer MF, Jenkins T (1994) The Y *Alu* polymorphism in southern African populations and its relationship to other Y-specific polymorphisms. Am J Hum Genet 54:319–330

Templeton AR (1993) The "Eve" hypotheses: a genetic critique and reanalysis. Am Anthropol 95:51–72

Terrenato L, Ulizzi L, San Martini A (1979) The effects of demographic transition on the opportunity for selection: changes during the last century in Italy. Ann Hum Genet 42:391–399

Tishkoff SA, Dietzsch E, Speed W, Pakstis AJ, Kidd JR, Cheung K, Bonné-Tamir B, et al (1996) Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. Science 271:1380–1387

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, 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

Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. Science 253:1503–1507

Wainscoat JS, Hill AVS, Boyce AL, Flint J, Hernandez M, Thein SL, Old JM, et al (1986) Evolutionary relationships of human populations from an analysis of nuclear DNA polymorphisms. Nature 319:491–493

Wallace DC (1995) Mitochondrial DNA variation in human evolution, degenerative disease, and aging. Am J Hum Genet 57:201–223

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

Wright S (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution 19:395–420

Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos FR, Schiefenhovel W, et al (1997) Genetic relationships of Asians and northern Europeans revealed by Y-chromosomal DNA analysis. Am J Hum Genet 60:1174–1183