J Mol Evol (1995) 41:966–973

JOURNAL OF MOLECULAR EVOLUTION

Recurrent Simple Tandem Repeat Mutations During Human Y-Chromosome Radiation in Caucasian Subpopulations

B.M. Ciminelli,[1] F. Pompei,[1] P. Malaspina,[1] M. Hammer,[2] F. Persichetti,[1] P.F. Pignatti,[3] A. Palena,[4] N. Anagnou,[4] G. Guanti,[5] C. Jodice,[1] L. Terrenato,[1] A. Novelletto[1]

[1] Dipartimento di Biologia, Università degli studi "Tor Vergata," via E. Carnevale, 00173 Roma, Italy
[2] University of Arizona, Tucson, AZ, USA
[3] Istituto di Scienze Biologiche, Università di Verona, Italy
[4] University of Crete, Iraklion, Greece
[5] Cattedra di Genetica Medica, Università di Bari, Italy

Received: 10 March 1995 / Accepted: 6 April 1995

Abstract. The haplotypes at four polymorphic loci of the Y chromosome were determined in 245 Caucasian males from 12 subpopulations. The data show that haplotype radiation occurred among Caucasians. Haplotype radiation was accompanied by recurrent mutations at STR loci that caused partial randomization of haplotype structure. The present distribution of alleles at short tandem repeats (STRs) can be explained by a mutation pattern similar to those described for autosomal STRs. The degree of variation among groups of subpopulations was assayed by using the Analysis of Molecular Variance. The results confirm a faster divergence of the Y chromosome as compared to the rest of the genome.

Key words: Human Y chromosome — Y polymorphisms — Human radiation — Single tandem repeats

Introduction

The human Y chromosome is transmitted from fathers to sons only as a haploid entity without recombination. The genetic diversity of this chromosome both within and among populations is thus to be considered as a record of the original combinations of mutational events that occurred along male lineages, since these are not reshuffled by genetic recombination. In this context the Y chromosome represents the male counterpart of what mitochondrial DNA (mtDNA) represents for female lineages. These observations make the study of Y chromosome variation highly valuable for describing the evolution of polymorphic sequences in a system in which mutation is the sole source of genetic change. In addition the analysis of such a variation across the extant human populations may add significantly to the understanding of the radiation of human male lineages.

In the recent past the study of human Y-chromosome diversity has been hindered by the relative paucity of common polymorphisms (Jakubiczka et al. 1989; Malaspina et al. 1990; Spurdle and Jenkins 1992), a finding at odds with reports on a higher mutation rate in males than females (Shimmin et al. 1993). In addition, for some of the described polymorphisms precise information on their molecular nature is lacking.

Recently, Santos et al. (1993) and Mathias et al. (1994) identified Y-linked polymorphic short tandem repeats (STR) consisting of reiterated GATA and CA units, respectively. These define genetic loci at which the addition of a discrete number of basic units generates multiallele series and interallele differences can be scaled according to a phenetic metric based on the number of units.

Correspondence to: A. Novelletto

A variety of techniques have been devised that take into account information on the molecular relatedness of different alleles at polymorphic loci to describe the divergence among populations. Excoffier et al. (1992) developed an approach called AMOVA (*A*nalysis of *MO*lecular *VA*riance) in order to obtain an estimate of variation quotas within populations, among populations of the same geographical area and among different areas. They applied this method to human mtDNA haplotypes resulting from the analysis of 62 restriction sites in ten populations representing the major racial groups and obtained estimates for variance components under various assumptions. In particular they considered phenetic equidistance among haplotypes (multiallelic AMOVA) and molecular distances among haplotypes in terms of restriction-site differences (haplotypic AMOVA). Excoffier and Smouse (1994) later showed that the AMOVA approach can also provide additional information to evaluate phylogenetic trees reconstructed relying on the parsimony method.

In this work we generated data from four Y-chromosome polymorphisms in 12 Caucasian subpopulations. All of the polymorphisms have a known molecular basis and include the base substitution in position XY275 (Ellis et al. 1989), the *Alu* insertion polymorphism detected by probe pYAP (Persichetti et al. 1992; Hammer 1994), and the polymorphic short tandem repeats (STRs) at loci Y27H39 (Santos et al. 1993) and YCAII (Mathias et al. 1994). We then applied AMOVA to Y-chromosome data for the first time, taking into account the molecular basis of interallele variation. In a previous paper (Persichetti et al. 1992) we had reported that Y chromosome diverges at a rate faster than the other chromosomes, since we found, within Caucasians, a degree of differentiation comparable to that of autosomes in the entire world population. Here we confirm this finding by analyzing a different set of polymorphisms. In addition we show that di- and tetranucleotide repeated sequences do not display a disequilibrium as strong as other polymorphisms. This denotes a certain degree of haplotype randomization, which can be attributed to recurrent mutation. The peculiar pattern of geographic radiation of Y alleles supports the observations of Weber and Wong (1993) and the model of Di Rienzo et al. (1994) on mutation in STRs.

Materials and Methods

Subjects. Overall, 245 DNAs from Caucasian males were studied. Some of these were previously described (Persichetti et al. 1992): 19 English (out of 21) selected after reviewing the paternal origin; 23 northern and 25 southern Sardinians; and 24 Egyptians (out of 31) with paternal birthplace north of Cairo. In addition we analyzed 5 new northern and 2 southern Sardinians, based on the geographical boundary already described (Persichetti et al. 1992); 20 Egyptian newborns with a parental birthplace south of Cairo collected during a screening for alpha-thalassemia (A. Novelletto, unpublished data); 26 and 20 adult males from Veneto (northeastern Italy) and Puglia (southeastern Italy), respectively, collected as random blood donors; 26 adult males from Calabria (southwestern Italy) collected as unrelated members of families segregating SCA1 (Frontali et al. 1991); 24 Greeks collected during a survey on beta-thalassemia in Crete; and 20, 11, and 5 males from United Arab Emirates (U.A.E.), Oman and Iran respectively, collected as fathers or patients affected with beta-thalassemia (El-Kalla and Mathews 1993).

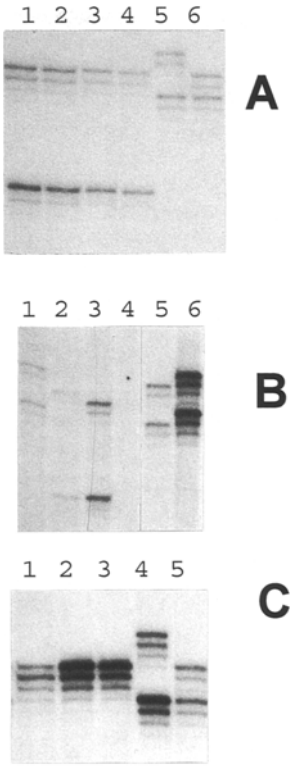

Fig. 1. YCAII patterns obtained in different subpopulations. **A** Sardinians. Lanes 1–4: phenotype a3,b – 8. Lanes 5, 6: control phenotypes a5,b1 and a3,b1. **B** Sardinians. Lanes 1, 6: Control phenotype a5,b1. Lane 2: phenotype a3,b – 8. Lane 3: phenotype a2,b – 8. Lane 4: control female. Lane 5: phenotype a4,b – 1. Lane 5: **C** Oman, U.A.E. and Greece. Lane 1: phenotype a4,b3. Lanes 2,3: Phenotypes a4,b4. Lane 4: phenotype a7,b1. Lane 5: Control phenotype a4,b1.

DNA Assays. Polymorphism for the single base substitution in position XY275 was examined as described (Persichetti et al. 1992).

Polymorphism at locus Y27H39 (DYS19) was analyzed as described (Roewer et al. 1992; Santos et al. 1993) except that primer 1 was end-labeled with ^{32}P by polynucleotide kinase and products were resolved on 6% polyacrylamide gels. PCR products from four subjects were sequenced after cloning into pUC18 (Sure-clone kit, Pharmacia) by the dideoxy-chain-termination method. The sequence of the 194-bp product was deposited under accession No. X77751. YCAII typing was performed as described (Mathias et al. 1994) with the annealing temperature increased to 56°C. Under these conditions one or two stronger bands were observed over a ladder background (see Fig. 1), a condition often seen in STR typing. Phenotypes were determined by comparison with the patterns produced by cell lines OXEN, 853, and 3E7 (kindly provided by C. Tyler-Smith). Allele nomenclature was similar to that of Mathias et al. (1994) except that arabic numerals were used.

The YAP polymorphism was typed by standard PCR using primers complementary to sequences flanking the insertion (Hammer and Horai 1995). Genomic DNA was amplified in 25 µl with 30 cycles of denaturation, annealing, and extension at 94°C 60 s, 51°C 60 s, and 72°C 60 s, respectively. Upon 2% agarose gel electrophoresis, the alleles cor-

responding to the absence (allele 1) and presence (allele 2) of the YAP element could be distinguished as a 150- or 450-bp band, respectively.

AMOVA. Analysis of molecular variance (AMOVA) (Excoffier et al. 1992) was performed with the computer program Winamova (version 1.05), kindly provided by the author. In order to replicate the AMOVA scheme used by Excoffier et al. (1992) to analyze mtDNA restriction data, we grouped the 12 subpopulations here studied as follows: (1) English; (2) Veneto; (3) Puglia, Calabria; (4) northern Sardinia, southern Sardinia; (5) Greece; (6) northern Egypt, southern Egypt; and (7) U.A.E., Oman, and Iran. Grouping of the European subpopulations was defined a priori based on previous analyses of Y-chromosome polymorphisms (Persichetti et al. 1992), and of non-Y-linked polymorphisms (Barbujani and Sokal 1990, 1991; Piazza et al. 1988).

The matrices of interallele phenetic distances used in AMOVA were computed according to a Euclidean metric. For each polymorphism the minimal change in molecular weight between alleles (300, 4, and 2 bp for YAP, Y27H39, and YCAII, respectively) was taken as the basic step. Thus, the difference between the two alternative YAP phenotypes was scored as 1, and between any two Y27H39 phenotypes the score was the number of tetranucleotide units separating them. The difference between any two YCAII patterns was computed as the minimum number of dinucleotide changes separating them under the assumption of two polymorphic series (YCAIIa, b). Patterns with a single band were interpreted as overlapping YCAIIa and YCAIIb PCR products (Results). This formalization is conceptually different from that originally used (Excoffier et al. 1992), in which allele differences at a single locus (restriction site) could only be 0 or 1, and higher values were attained by considering restriction haplotypes. In our data set multiple loci could be combined, leading to higher-order interhaplotypic distances. Since only results of AMOVA for individual loci are reported here we did not use the terms "multiallelic" and "haplotypic" AMOVA.

Significance levels of the components of variance and the corresponding Φ statistics (Excoffier et al. 1992) were obtained by comparing the actual values with the distribution of 1,000 values obtained by simulation.

Results

The base substitution in position XY275 was assayed in all subpopulations except Veneto and Greece for a total of 200 subjects. This site was found polymorphic only in Egypt, with frequencies for the lower allele (presence of the *Msp*I site) of 0.125 ± 0.068 and 0.05 ± 0.05 among northern and southern Egyptians, respectively. In the four subjects carrying the *Msp*I site, there was no association with specific alleles at any of the other polymorphic loci, showing that randomization occurred. In view of its low degree of polymorphism, this system was not considered in further calculations.

Both YAP alleles were found in all the subpopulations examined except the English and the small number of Omanites. The north–south variation in allele frequencies (Persichetti et al. 1992) was confirmed, with a null frequency of allele 2 in the U.K., a maximum in Egypt, and intermediate values in southern Italy, Greece, and the Arab peninsula. Heterogeneity among subpopulations was highly significant (contingency chi square = 48.4, 11 d.f., $P \ll 0.001$).

Y27H39 showed six alleles ranging from 178 to 202 bp. This system was highly polymorphic in all subpopulations, without an apparent pattern of geographical variation. For example, the 186-bp allele was the most frequent in northern Egypt (0.375), whereas the 190-bp allele was the commonest in the U.K., southern Egypt, southern Sardinia, Puglia, and U.A.E. (0.47, 0.40, 0.33, 0.40, and 0.65, respectively), and the 194-bp allele in northern Sardinia, Calabria, Oman, and Iran (0.36, 0.46, 0.64, and 0.80, respectively). A single male carrying a previously undescribed 178-bp allele was found among Egyptians. Finally, in Sardinia the uncommon 202-bp allele was found with frequencies of 0.18 and 0.22 in the northern and southern part of the island, respectively. Overall, the heterogeneity among subpopulations was highly significant (contingency chi square = 80.3, 44 d.f., $P < 0.001$ [178-bp allele pooled]).

Twenty-four different YCAII patterns were found, exceeding by far those originally described by Mathias et al. (1994) in a group of 93 DNAs from different geographic groups. In particular we found many patterns characterized by bands with molecular weight greater than level 5 or smaller than level 1. Some of the new patterns are shown in Fig. 1. Mathias et al. (1994) identified two "allelic" series, named YCAIIa (upper series) and YCAIIb (lower series). We retained the proposed nomenclature by omitting level 0 and attributing to low bands the names −1, −2, etc., in descending order. Overall 7 alleles were found for the a series (spanning 7 dinucleotide units) and 8 for the b series (spanning 12 dinucleotide units).

We found 38 males displaying a single band, a proportion higher than reported before. This finding is suggestive of the presence of "null alleles" at either series. However no "null pattern" resulting from "null alleles" at both YCAIIa and b series was found. Although a strong disequilibrium between "null alleles" at one series and "non-null alleles" at the other may mask the presence of the former ones, we considered the patterns with a single band the result of two PCR fragments of the same size, as suggested by Mathias et al. (1994).

The two most common patterns were a4,b1 or a5,b1 in all populations except in Sardinia and U.A.E. In Sardinia the pattern a3,b − 8 was the commonest and reached a frequency of 0.45. This pattern was not found anywhere else and is thus entirely population-specific. In the U.A.E. the commonest pattern was a4,b4 with a frequency of 0.45. However, this pattern did not show the same population specificity, being found in Egypt, Greece, Oman, and northern Sardinia. Overall, heterogeneity among subpopulations for YCAII patterns was highly significant (contingency chi square = 464.2, 253 d.f., $P < 0.001$).

Randomization of Haplotype Structure

Table 1 shows the distribution of Y-chromosome haplotypes in the 12 subpopulations studied. Contingency

analysis between loci to evaluate pairwise distribution of alleles produced highly significant chi-square values: 80.1 (23 d.f., $P \ll 0.001$) for YAP vs YCAII, 131.9 (5 d.f., $P \ll 0.001$) for YAP vs Y27H39 and 278.5 (115 d.f., $P \ll 0.001$) for YCAII vs Y27H39, indicating that alleles at the three loci are indeed in disequilibrium. However, this disequilibrium is different from the near-to-complete disequilibrium that we observed (Persichetti et al. 1992) when using YAP and two other RFLPs. In that paper alleles at YAP and the alphoid polymorphism detected by probe pYα1 (Tyler-Smith and Brown 1987) defined three frames in which the pattern obtained with p49f (Ngo et al. 1986) almost invariably fell, indicating a relative stability of these polymorphisms on a evolutionary time scale. Seielstad et al. (1994) recently described another case of complete disequilibrium between YAP and a base-substitution polymorphism in DYS271, at a distance spanning a large part of Yq. The finding of the YAP insertion at the same site in several individuals and the lack of evidence for a mechanism of specific removal of *Alu* elements led Hammer (1994) to conclude that YAP allele 2 arose only once, before the divergence of modern human groups. In this study a full range of alleles at both the YCAII and Y27H39 loci were associated with each allele at the YAP locus. For example, the more recent YAP allele 2 (presence of the *Alu* insert) was associated with five, three, and five alleles at YCAIIa, YCAIIb and Y27H39 respectively—namely, a great part of the total variation at each of the polymorphic systems. This reveals that haplotype randomization attributable to recurrent mutation at each of the STR occurred along the lineages, leading to the examined types.

Radiation of Haplotypes

This observation opens the way to understanding the mutability of each STR during radiation of Y haplotypes. Reconstruction of phylogenetic trees of Y chromosome is beyond the scope of this work. However, the analysis of population-specific STR allele distributions associated with each YAP allele can provide useful information.

Recent gene flow between neighboring subpopulations can enlarge the number of alleles and widen the distribution found in a given sample. We thus considered the groups described in the Materials and Methods section and computed the minimal number of shifts in allele size required to generate the observed diversity given a fixed YAP frame. For Y27H39, 23 shifts of a single tetranucleotide unit and one shift of two units (to account for the 178-bp allele among Egyptians [Table 1, line 32 vs 19, 33, 38, 42, 51]) could explain the entire variation in all subpopulations. For YCAIIa the entire variation could be explained by 34 shifts of a single dinucleotide unit and two shifts of two units (to account for allele 7 in Greece [Table 1, line 58 vs 45–47] and allele 1 in the Arab peninsula [line 4 vs 21]). For YCAIIb the entire variation could be explained by 26 shifts of a single dinucleotide unit, 2 shifts of 2 units (to account for allele 3 in Veneto [Table 1, line 25 vs 8, 16–18, 30, 45–47, 56] and allele 4 in Sardinia [line 39 vs 11, 22]), and 1 shift of 8 units (to account for allele −8 in Sardinia [lines 14, 26–28 vs 7, 16, 30, 31, 45–47, 55, 56]).

This analysis does not represent an attempt to reconstruct the mutational history of Y haplotypes among Caucasians and it is perhaps still biased by recent admixture. Nonetheless, it shows that the observed variation in a set of STR allele distributions in the absence of recombination can be explained by a majority of mutational events involving small changes in allele size and only a few events involving large changes in size. These results are in agreement with the results obtained by Weber and Wong (1993) and Di Rienzo et al. (1994) on autosomal STR polymorphisms.

The case of Sardinia is paradigmatic. The most likely scenario for this isolated subpopulation is that an aboriginal mutation generated the unusual haplotype containing YCAIIb −8 fragment (Fig. 1), in association with YCAIIa allele 3 (Table 1, lines 26–28). The population specificity of this pattern can only be explained by a mutational event rather than inward gene flow. The spread of this chromosome within the island was then accompanied by multiple events of smaller magnitude, 1 affecting YCAIIa and producing allele 2 (line 14) and at least 2 affecting Y27H39. One of these produced the unusual 202-bp allele, the frequency of which is also increased in Sardinia. The alternative hypothesis must assume that YCAIIb allele −8 has arisen several times independently through recurrent large-scale mutations. We have already highlighted the relatively high frequency of p49f haplotype XII (Persichetti et al. 1992) in Sardinia and attributed it to a founder effect. The observation that 18 out of 19 carriers of 49f haplotype XII also carry the YCAII pattern a3,b − 8 shows that the increased frequency of 49f haplotype XII is better explained by an aboriginal mutation followed by selection and/or drift.

We further explored the possibility of tracing mutation events at Y27H39. In fact the PCR product corresponding to Y27H39 has the structure of an imperfect repeat with more than a single uninterrupted $(GATA)_n$ stretch. Different length changes may thus affect different segments of the sequence. By sequencing PCR products from four subjects with haplotypes 14, 28, 32, and 35 (Table 1) we confirmed (Roewer et al. 1992) that allele differences consist in variation of the longest $(GATA)_n$ stretch starting at position 95, with n values of 8, 10, and 12 for allele sizes of 178, 186, and 194, respectively.

AMOVA

In order to evaluate the degree of divergence among the examined subpopulations and to examine the relevance of haplotype radiation, we applied AMOVA to the data of Table 1 after grouping the subpopulations as described

Table 1. Distribution of Y-chromosome haplotypes in 12 Caucasian subpopulations (for allele nomenclature see Materials and Methods)[a]

	Haplotype			Subpopulation											
	YCAII	YAP	Y27H39	UK	VE	NS	SS	PU	CA	GR	NE	SE	UA	OM	IR
1	a1,b1	1	186	1									1		
2	a1,b1	1	190										2	1	
3	a1,b1	1	202						1						
4	a1,b1	2	194												1
5	a1,b − 2	1	190										1		
6	a2,b1	1	190					1		1				1	
7	a2,b1	1	194			1	1	2	2					1	1
8	a2,b1	1	198		1										
9	a2,b1	2	190								1				
10	a2,b2	1	190												1
11	a2,b2	1	194			1	1			1		1			
12	a2,b2	1	198	1											
13	a2,b − 1	1	194					1		1		1			
14	a2,b − 8	1	194			1	1								
15	a3,b1	1	186	1											
16	a3,b1	1	190	2	1		1	1		2		1		1	
17	a3,b1	1	194		1					1				1	
18	a3,b1	1	198		1										
19	a3,b1	2	186		2	2		4	2	1		1			
20	a3,b1	2	190						1			1			
21	a3,b1	2	198										1		
22	a3,b2	1	194			1	1		1					1	
23	a3,b2	1	198								1				
24	a3,b3	1	194							1	1				
25	a3,b3	1	198	1	1			1	1	2					
26	a3,b − 8	1	194			5	2								
27	a3,b − 8	1	198			4	3								
28	a3,b − 8	1	202			5	6								
29	a4,b1	1	186							1					
30	a4,b1	1	190		2	2	3	1	2	5	2	1			
31	a4,b1	1	194				1	2	4	1	1			2	
32	a4,b1	2	178								1				
33	a4,b1	2	186				1		1	1	8				
34	a4,b1	2	190								1	2			
35	a4,b1	2	194								2	1			
36	a4,b2	1	190						2						
37	a4,b3	1	194	1										1	1
38	a4,b3	2	186									1			
39	a4,b4	1	190			1					2	2	8	1	
40	a4,b4	1	194							1		1	1		
41	a4,b − 1	1	190		1										
42	a4,b − 1	2	186			1					1				
43	a4,b − 1	2	190										1		
44	a5,b1	1	186	1											
45	a5,b1	1	190	7	5	1	5	5	4	1	2		1		
46	a5,b1	1	194	4	2	1		1	4	1			1	1	1
47	a5,b1	1	198		2	1		1		2		1	2		
48	a5,b2	1	198									1			
49	a5,b3	1	194									1			
50	a5,b3	1	198									2			
51	a5,b3	2	186									1			
52	a5,b4	1	194								1				
53	a5,b − 1	1	194						1				1		
54	a5,b − 2	1	194		1										
55	a6,b1	1	190			1									
56	a6,b1	1	194		1		1								
57	a6,b5	1	190									1			
58	a7,b1	1	194							1					
Total				19	21	28	27	20	26	24	24	20	20	11	5

[a] Subpopulations: UK, English; VE, Veneto; NS, northern Sardinia; SS, southern Sardinia; PU, Puglia; CA, Calabria; GR, Greece; NE, northern Egypt; SE, southern Egypt; UA, United Arab Emirates; OM, Oman; IR, Iran

Table 2. Hyerarchical analysis of variance on three Y-chromosome polymorphisms in Caucasians

Variance component	Variance	% of total	Φ statistics	*P*
Without interallele differences				
YAP				
Among groups	0.023	16.33	Φct = 0.163	.124
Among subpopulations/groups	0.003	1.98	Φsc = 0.024	.379
Within subpopulations	0.115	81.69	Φst = 0.183	.001
YCAII				
Among groups	0.040	9.05	Φct = 0.090	.011
Among subpopulations/groups	0.015	3.37	Φsc = 0.037	.043
Within subpopulations	0.390	87.58	Φst = 0.124	.001
Y27H39				
Among groups	–0.003	–0.81	Φct = –0.008	.577
Among subpopulations/groups	0.013	3.51	Φsc = 0.035	.076
Within subpopulations	0.354	97.31	Φst = 0.027	.035
With interallele differences[a]				
YCAII				
Among groups	0.446	23.72	Φct = 0.237	.028
Among subpopulations/groups	0.027	1.46	Φsc = 0.019	.253
Within subpopulations	1.408	74.82	Φst = 0.252	.001
Y27H39				
Among groups	0.022	3.91	Φct = 0.039	.153
Among subpopulations/groups	0.011	1.91	Φsc = 0.020	.141
Within subpopulations	0.537	94.18	Φst = 0.058	.002

[a] Not applicable to YAP, due to the presence of only two alleles in this system

in Materials and Methods. The results obtained for each polymorphism are reported in Table 2.

In the simplest AMOVA scheme the raw data consist of allele frequencies obtained in the 12 subpopulations, with no additional information on allele structure (Table 2, upper part). Both YAP and YCAII revealed a relatively high proportion (16 and 9%, respectively) of total variance accounted for by variation among the seven groups of subpopulations (U.K., Veneto, Sardinia as a whole, southern Italy, Greece, Egypt as a whole, and the Arab Peninsula plus Iran). In the case of YCAII this proportion is statistically significant. However, Y27H39 provided a very low value for the among-groups variance proportion, in line with the observation of poorly geographically structured allele frequency distributions. For all of the polymorphisms the within groups/among subpopulations proportion of variance is of the order of few percent and significant only for YCAII.

Information on molecular differences among alleles was then entered in the AMOVA scheme for YCAII and Y27H39 in order to evaluate the effect of radiation of different types in different subpopulations on the overall divergence among groups. YCAII revealed a 2.5-fold increase of the among-group component (statistically significant), accompanied by a decrease of the within-group component. This indicates that the evolution of different allelic frequencies in the different groups was also accompanied to some extent by a geographically specific radiation of types. An increase in the among-group component (3.9%) was also seen for Y27H39, again accompanied by a decrease of the within-group component. These low values are in agreement with the observation that a great part of the total variation at this locus is still present in each of the subpopulations examined here.

In order to address specific questions AMOVA was also used by varying the grouping of subpopulations. When European vs non-European populations were examined, a drop of the among-group component of variance was observed for all loci (11.14%, 3.28%, –0.45% for YAP, YCAII, and Y27H39, respectively, without interallele differences; 6.76%, 1.56% for YCAII and Y27H39, respectively, with interallele differences), paralleled by a sharp increase of the within-group component (10.87%, 10.01%, 2.98% for YAP, YCAII, and Y27H39, respectively, without interallele differences; 19.67%, 4.73% for YCAII and Y27H39, respectively, with interallele differences; $P < 0.05$ in all cases). The rise of the within-group component of variance caused by considering interallele differences is in line with a radiation of types for YCAII and Y27H39 that occurred within each group, i.e., after the European/non-European split. This analysis shows that the European and non-European groups are themselves highly structured. The linguistic distinction between Arab- and non-Arab-speaking subpopulations is accompanied by a degree of Y-chromosome diversity not greater than that seen within each group.

Discussion

We report here on four Y-chromosome polymorphisms typed in 12 Caucasian subpopulations. We confirm that the low allele at position XY275 is present only among

Egyptians, most probably a result of admixture with African populations in which it is found at higher frequencies (Ellis et al. 1990; Spurdle et al. 1994). Although this allele is found among Asian Indians (Spurdle et al. 1994) we did not find it in males from either the U.A.E. or Oman, two populations showing considerable admixture with Asian Indians (El-Kalla and Mathews 1993; De Leo et al. 1995). We confirmed the randomization of the *Msp*I site with respect to the other polymorphisms, in agreement with the hypothesis of Spurdle et al. (1994) that recombination occurred in the proximal part of the pseudoautosomal region.

Our analysis of three polymorphisms of the male-specific portion of the Y chromosome shows that partial haplotype randomization is a common finding when examining STR sequences. Mathias et al. (1994) have already observed that the same tandem repeat allele is often found on different Y-chromosome backgrounds. This has to be attributed to recurrent mutation during Y-chromosome radiation. In particular our data show that the time elapsed since the separation of the groups here examined allowed several such mutations to occur and be spread. Weber and Wong (1993) examined the pattern of mutability during parent–child transmissions of autosomal STRs and concluded that changes in allele sizes of one repeat unit occur more frequently than larger changes, and tetranucleotide repeats mutate at a rate higher than dinucleotide repeats. Di Rienzo et al. (1994) showed that the allelic distributions at several autosomal STRs fit a model that assumes that small changes are more likely to occur than large ones. Our qualitative analysis of allele distributions shows that the same pattern of mutability can also explain Y-chromosome STR variation. In addition our haplotype distribution is in line with a mutation rate for the tetranucleotide repeat at Y27H39 higher than the dinucleotide repeat at YCAII.

In order to evaluate the divergence among the examined groups we used here for the first time the Analysis of Molecular Variance (AMOVA) on Y-chromosome data. We adapted the AMOVA scheme originally developed (Excoffier et al. 1992) for an array of restriction-enzyme sites to STR multiallele series, each of which thus provided a greater amount of information than a single restriction site. We empirically used the phenetic similarity between alleles as a measure of their relatedness. The observations reported here on mutability of STRs support the view that alleles differing for one repeated unit are on average more evolutionarily related than alleles differing for a higher number of units. This, while granted in the case of arrays characterized by the presence/absence of multiple sites (haplotypes), is an essential prerequisite for the proper application of AMOVA to true multiallele series such as STRs.

Upon AMOVA analysis the YAP polymorphism revealed a strong divergence among Caucasians, as already reported by us (Persichetti et al. 1992). Spurdle et al. (1994) and Seielstad et al. (1994) reported striking differences in the incidence of the YAP insertion between Caucasians and African black populations but showed a large degree of divergence within the latter, too. Hammer (1994) obtained an F_{ST} value of 0.58 by examining populations from Africa, Europe, Asia, and Oceania, a value fourfold higher than the average obtained on 100 non-Y DNA polymorphisms (Bowcock et al. 1991). These data, taken together, reveal that world populations are indeed largely divergent for this polymorphic system.

AMOVA as applied to YCAII polymorphism showed that evolution of different allele frequencies was accompanied by a radiation of types in different Caucasian subpopulations. The data originally reported by Mathias et al. (1994) do not allow a discrimination among Caucasians. However, the possibility of a large-scale recurrent generation of alleles in more distantly related populations is open. This is still to be investigated and would reduce the value of this marker in evolutionary studies.

With AMOVA the Y27H39 polymorphism showed a lower degree of variation both among and within Caucasian groups, although it necessarily took part in the same radiation process as YAP and YCAII. This is most likely explained by frequent recurrent mutation, as revealed by the presence of multiple Y27H39 alleles associated with the unusual YCAII patterns a3,b – 8, a4,b4, a5,b3, and a6,b1.

The results of AMOVA analysis are compatible with the occurrence of a detectable radiation not only among Caucasians as a whole but also within the European and North African/Arabic subgroups. However, the AMOVA results reported here as well as finer analyses (not shown) indicate that only a small amount of total variance is accounted for by differences among neighboring portions of the Sardinian, southern Italian, Egyptian, and Arab subpopulations. In fact the geographical criteria used here may not reflect barriers to gene flow in recent times, and significant amounts of admixture may have occurred.

The polyphyletic origin of STR alleles in different populations may result in the underestimation of the among-group and within-group components of variance. The values here reported for Y-chromosome STR polymorphisms among Caucasians are thus to be considered minimal. Nevertheless, the among-group and within-group components of total variance obtained here for YCAII are in agreement with those obtained for YAP and both are in the range obtained by Excoffier et al. (1992) on mtDNA data from ten world populations representing the major racial groups. This reinforces our previous finding of a faster divergence of the Y chromosome as compared to the rest of the genome.

Acknowledgments. We gratefully acknowledge Drs. L. Excoffier and C. Tyler-Smith for advice and critical reviewing of a first draft of this paper. This work was partially supported by grant No. 00038. PF99 of Consiglio Nazionale delle Ricerche to A.N.

References

Barbujani G, Sokal RR (1990) Zones of sharp genetic change in Europe are also linguistic boundaries. Proc Natl Acad Sci USA 87:1816–1819

Barbujani G, Sokal RR (1991) Genetic population structure of Italy II. Physical and cultural barriers to gene flow. Am J Hum Genet 48: 398–411

Bowcock AM, Kidd JR, Mountain JL, Herbert 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

De Leo R, Deidda G, Novelletto A, El-Kalla S, Mathews A, Felicetti L (1995) Analysis of beta-thalassemia mutations in the United Arab Emirates provides evidence for recurrent origin of the IVS1 nt 5 (G-C) mutation. Hum Mutat (in press)

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994) Mutational processes of simple sequence repeat loci in human populations. Proc Natl Acad Sci USA 91:3166–3170

Ellis N, Taylor A, Bengtsson BO, Kidd J, Rogers J, Goodfellow P (1989) Population structure of the pseudoautosomal boundary. Nature 344:663–665

Ellis N, Kidd J, Goodfellow PJ, Kidd J, Goodfellow PN (1990) Strong linkage disequilibrium between the XY274 polymorphism and the pseudoautosomal boundary. Am J Hum Genet 46:950–955

El-Kalla S, Mathews AR (1993) Molecular characterization of beta-thalassemia in the United Arab Emirates. Hemoglobin 17:355–362

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

Excoffier L, Smouse PE (1994) Using allele frequencies and geographic subdivision to reconstruct gene trees within a species: molecular variance parsimony. Genetics 136:343–359

Frontali M, Iodice C, Lulli P, Spadaro M, Cappellacci S, Giunti P, Malaspina P, Morellini M, Morocutti C, Novelletto A, Persichetti F, Trabace S, Anastasi R, Terrenato L (1991) Spinocerebellar ataxia SCAI in two large Italian kindreds: evidence in favour of a locus position distal to GLOI and the HLA cluster. Ann Hum Genet 55:7–15

Hammer MF (1994) A recent insertion of an Alu element on the Y chromosome is a useful marker for human population studies. Mol Biol Evol 11:7–15

Hammer MF, Horai S (1995) Y chromosomal DNA variation and the peopling of Japan. Am J Hum Genet 56:951–962

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

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

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

Ngo KY, Vergnaud G, Johnsson C, Lucotte C, Weissenbach J (1986) A DNA probe detecting multiple haplotypes on the human Y chromosome. Am J Hum Genet 38:407–418

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

Piazza A, Cappello N, Olivetti E, Rendine S (1988) A genetic history of Italy. Ann Hum Genet 52:203–213

Roewer L, Arnemann J, Spurr NK, Grzeschik KH, Epplen JT (1992) Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts. Hum Genet 89:389–394

Santos FR, Peña SDJ, Epplen JT (1993) Genetic and population study of a Y-linked tetranucleotide repeat DNA polymorphism with a simple non-isotopic technique. Hum Genet 90:655–656

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

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

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

Spurdle AB, Jenkins T (1992) The Y chromosome as a tool for studying human evolution. Curr Opin Genet Dev 2:487–491

Tyler-Smith C, Brown WRA (1987) Structure of the major alphoid satellite DNA on the human Y chromosome. J Mol Biol 195:457–470

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