Worldwide distribution of human Y-chromosome haplotypes.

F R Santos, N O Bianchi and S D Pena

Genome Res. 1996 6: 601-611
Access the most recent version at doi:10.1101/gr.6.7.601

RESEARCH

Worldwide Distribution of Human Y-chromosome Haplotypes

Fabrício R. Santos,[1] Néstor O. Bianchi,[2] and Sérgio D.J. Pena[1,3]

[1]Departamento de Bioquímica, Universidade Federal de Minas Gerais, 31.270-910 Belo Horizonte, Brazil; [2]Instituto Multidisciplinario de Biologia Celular (IMBICE), La Plata, Argentina

We surveyed several human populations worldwide with three PCR-based polymorphisms located in the human Y chromosome: the alphoid heteroduplex (αh) polymorphic system, the *DYS19* microsatellite locus, and a polymorphic *Alu* insertion (YAP). By typing with the former two polymorphisms (αh and *DYS19*) we found 46 different haplotypes in 364 males from several populations worldwide. There were significant geographic differences in the distribution of the haplotypes, several of which were seen in only one population and can be used as populational markers in future surveys. The haplotypic diversity in major ethnic groups revealed high levels in Greater Asians, followed by Africans and Caucasians, and a very low diversity was seen in Amerindians. The discrimination probability of such haplotypes for a random sample of Brazilian Caucasians was 0.82, suggesting great potential usefulness in forensic studies. The parsimonious relationship between different αh types and the addition of the YAP polymorphism data allowed the construction of an informative picture of the origin and evolution of the αh polymorphism. The *DYS19* allele diversity related to each αh type allowed a crude estimation of the antiquity of many αh types. These ancient αh types were present in different populations suggesting a common ancestor that could antedate the first out-of-Africa migrations.

Human genomic diversity can be used to infer relationships between individuals in identity tests or between populations in evolutionary studies (Pena et al. 1995a). Two very special compartments of the human genome, the Y chromosome (nonpseudoautosomal portion) and mitochondrial DNA (mtDNA), are of special interest because they are haploid and free of genetic recombination. Human Y chromosomal polymorphisms have a father-to-son inheritance, thus establishing patrilineages, whereas mtDNA polymorphisms are passed from mother to children, establishing matrilineages. Therefore, they constitute analogous and complementary evolutionary genetic markers that should reveal ancestral haplotypes corresponding to founding males and females. Recently, two studies based on the sequencing of large regions estimated the coalescence time for the human Y chromosomes to be in the range of 51,000–411,000 (Whitfield et al. 1995) and 37,000–49,000 (Hammer 1995) years ago, respectively. These disparate estimates should be clarified with the study of other Y polymorphisms.

[3]Corresponding author.
E-MAIL: spena@dcc.ufmg.br; FAX: (5531)227-3792.

Unfortunately, the human Y chromosome has exhibited a paucity of DNA polymorphisms. Several systematic searches for restriction fragment length polymorphisms (RFLPs) have met with little or no success (Jakubiczka et al. 1989; Malaspina et al. 1990; Spurdle and Jenkins 1992). Moreover, extensive sequencing of two noncoding regions in the Y chromosome from several individuals has revealed a striking monomorphism (Seielstad et al. 1994; Dorit et al. 1995). However, prospecting for size variation has been more profitable (Oakey and Tyler-Smith 1990; Jobling 1994; Mathias et al. 1994) and a few PCR-based polymorphisms were described recently in the human Y chromosome: the tetranucleotide repeat locus *DYS19* containing a $(GATA)_n$ repeat (Roewer et al. 1992; Santos et al. 1993a), a polymorphic *Alu* insertion (YAP; Hammer and Horai 1995), and three polymorphic $(CA)_n$ microsatellites (Mathias et al. 1994). The most variable of these is the *DYS19* microsatellite that has at least nine different alleles (Santos et al. 1993a, 1996a) and has been applied successfully in forensic casework (Roewer and Epplen 1992), paternity testing (Santos et al. 1993b), and population studies (Gomolka et al. 1994; Santos et al. 1996a).

Recently, we described a new type of PCR-

based Y-linked polymorphism, the alphoid heteroduplex (αh) system detected by heteroduplex analysis of divergent alphoid subunits located in the edges of the repetitive array of the human Y chromosome centromere (Santos et al. 1995a). The molecular basis of the polymorphism is complex and is based on the coamplification of alphoid repeats with slight differences in their sizes and/or base sequences. At least two types of loci are amplified simultaneously from each Y chromosome: the locus (αhL) situated in the left edge (proximal to Yp) and a variable number of different loci on the right edge (proximal to Yq) of the alphoid array. The locus αhL (281 bp) apparently is conserved in sequence among different individuals and contains a 4-bp deletion compared to the 285-bp right-edge loci, whose sequences generally differ from each other by single-base changes (Santos et al. 1995a). At the completion of the PCR reaction, heteroduplexes are formed between the DNA strands of the left-side locus with their complements from the right-side loci. There is formation of a pair of heteroduplex molecules for each different right-side locus (or set of loci with identical sequences). These heteroduplexes assume different conformations reflecting their sequence differences and can be resolved on nondenaturing polyacrylamide gels. In our initial study we typed 93 males and observed 12 different heteroduplex pairs (h1–h12) that resulted from the combination of αhL with different right-side loci named αh1 to αh12, respectively. Each individual could be typed according to the presence or absence of each of the heteroduplex pairs in different combinations. Thus, 13 distinct αh types (we will use the denomination type rather than the more specific haplotype to avoid confusion with the multimarker Y chromosome haplotypes) were found in the 93 males, including a pattern without any heteroduplex formation, which was shown to be caused by deletion of the αhL locus. Thus, the αh polymorphism displayed a high level of variation mainly generated by point mutations in the right-side loci of the Y-chromosomal alphoid array. Detailed explanations about the αh polymorphism as well as complete laboratory protocols can be obtained in http://www.bioch.ox.ac.uk/~fabricio/ah.html.

In this study we undertook a worldwide populational survey of males studied simultaneously with the αh polymorphic system and *DYS19*, thus establishing Y-chromosome haplotypes. In the sample analyzed we identified 22 distinct αh types (thus recognizing 9 new ones) and 8 *DYS19* alleles, described previously (Santos et al. 1996a). The combined use of αh and *DYS19* revealed 46 distinct Y-chromosome haplotypes that displayed characteristic geographical distributions. In addition, some samples were analyzed for YAP (Hammer 1994; Hammer and Horai 1995) providing interesting insights into the origin of the αh variation.

RESULTS

Y Chromosomes in Different Populations Include New Alphoid Heteroduplexes and Types

The joint analysis of 364 males with αh and *DYS19* revealed 46 distinct Y-chromosome haplotypes (Table 1), which included 22 different αh

Table 1. Y-chromosome Haplotypes (αh and *DYS19*) in 364 Individuals Worldwide

αh	*DYS19**								N**
Type	0	Z	A	B	C	D	E	F	
I		1	8	2	5	1	1		18
II	1		98	74	13	5	1	2	194
III			1	21	21	8	5		56
IV					11	2			13
V			4	6	3				13
VI				1					1
VII				1					1
VIII					1				1
IX				4	8	7	5		24
X					1				1
XI					1				1
XII			4						4
XIII			1						1
XIV				1					1
XV			2						2
XVI			2	1					3
XVII				1					1
XVIII				1	9	10			20
XIX				1					1
XX				6					6
XXII			1						1
XXIII						1			1
N**	1	1	121	120	73	34	12	2	364

(*) *DYS19* alleles are 0 (no amplification of *DYS19*), Z (182 bp), A (186 bp), B (190 bp), C (194 bp), D (198 bp), E (202 bp), F (206 bp).
(**) N = number of individuals.

types, 9 of which were novel (Fig. 1) and 8 *DYS19* alleles described previously (Santos et al. 1996a). The 22 αh types resulted from different combinations of the previously described heteroduplex pairs (Santos et al. 1995a) with the addition of the five newly found heteroduplex pairs described here (h13–h17; Table 2; Fig. 1). Each heteroduplex band indicated in Table 2 corresponds to its respective alphoid right-side (αh#) generator locus, where # can be 1–17 (e.g., h1 corresponds to the presence of the αh1 locus). The problem of heterogeneity of heteroduplex band intensity in some αh types has been discussed previously (Santos et al. 1995a) and can cause haplotyping errors when DNA of poor quality or in small amounts is analyzed. For instance, we had described type αhXXI (containing only heteroduplex pairs h5 and h6) in a single Pygmy male (Pena et al. 1995b). The use of improved PCR conditions (see Methods) on the poor-quality DNA of this individual allowed the appearance of h1 together with h5 and h6, indicating that he really belonged to the previously known type αhV. Hence, we have decided that before assigning new αh types, an unambiguous result must be seen using improved PCR conditions. Furthermore, for complex αh phenotypes, that is, patterns with more than two heteroduplex pairs (Table 2; Fig. 1), experiments of cloning and mixing with PCR products of the αhL clone (Santos et al. 1995a) should be used to confirm the existence of a new right-side heteroduplex generator locus. It will also allow us to determine correctly the heteroduplex pairs that sometimes have one of the bands comigrating with other ones (Santos et al. 1995a).

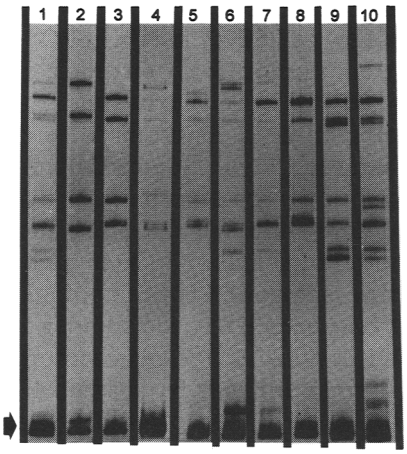

Figure 1 Gel with new αh types. (Lane *1*) The αhIX type commonly found among Africans (Santos et al. 1995a) and used here as a comparative marker; (lane *2*) αhXIV; (lane *3*) αhXV; (lane *4*) αhXVI; (lane *5*) αhXVII; (lane *6*) αhXVIII; (lane *7*) αhXIX; (lane *8*) αhXX; (lane *9*) αhXXII; and (lane *10*) αhXXIII. The main homoduplex alphoid amplification product with 281–285 bp is indicated by a large arrow. The heteroduplex pairs found in each αh type are described in Table 2.

In the evaluation of the new αh types we tried to establish possible parsimonious mutational pathways for their conversion to or from known types. For this, we established rules of conversion as follows: (1) Deletions–deletion of one of the right-side loci will cause the disappearance of the corresponding heteroduplex pair. This is probably irreversible, because the reconstitution of the original type would need a locus duplication and further point mutations in specific sites to generate the same heteroduplex, that is, a chain of unlikely events; (2) point mutations in a single right-side locus–a point mutation in one of the right-side loci will cause a shift in migration of the corresponding heteroduplex pair (Santos et al. 1995a), that is, the disappearance of the original heteroduplex and appearance of a new pair; reversibility is also unlikely; (3) point mutations in reiterated right-side loci–some of the right side loci may exist permanently or temporarily in more than one copy. Mutation in only one of these copies will result in the appearance of a new heteroduplex pair. The reversion to the original type is possible by a single deletion step of the newly mutated unit or gene conversion back to the original sequence locus. From these rules emerges a picture for interconversion of αh types, according to which mutations may lead to an increase or a decrease in complexity of the heteroduplex patterns. Increases in complexity should be unique events with reversion possible, whereas decreases in complexity are not necessarily unique but should be essentially irreversible. In addition, to verify the instability of the αh polymorphism we analyzed 97 Brazilian father/son pairs (Santos et al. 1993a). The father-to-son inheritance of the αh type (11 types found in this sample) was observed in all instances.

Many of the αh types were restricted to specific populations or appeared in single males, which probably could be attributed to recent origin. The αh types predominant in specific cohorts can be used as markers of these particular populations (Table 3). Our African sample was limited to only 34 individuals but clearly shows a predominance of type αhIX. Types αhX, αhXXII,

Table 2. Twenty-two αh Haplotypes Scored as Presence or Absence of Heteroduplex Pairs

αh Type	Heteroduplex Pair																
	h 1	h 2	h 3	h 4	h 5	h 6	h 7	h 8	h 9	h 10	h 11	h 12	h 13	h 14	h 15	h 16	h 17
I	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
II	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
III	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-
IV	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-
V	+	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-
VI	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
VII	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
VIII	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-
IX	+	-	-	-	+	+	+	+	-	-	-	-	-	-	-	-	-
X	+	-	-	-	+	+	+	-	+	-	-	-	-	-	-	-	-
XI	+	-	-	-	+	-	-	-	-	+	-	-	-	-	-	-	-
XII	-	-	-	-	+	+	-	-	-	-	+	-	-	-	-	-	-
XIII	-	-	-	-	+	+	-	-	-	-	+	+	-	-	-	-	-
XIV	+	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-
XV	-	-	-	-	+	-	-	-	-	-	+	-	-	-	-	-	-
XVI	-	-	-	-	+	+	-	-	-	-	-	-	+	-	-	-	-
XVII	+	-	-	-	+	-	-	-	-	-	-	-	-	+	-	-	-
XVIII	+	-	-	-	+	+	-	-	-	-	-	-	-	-	+	-	-
XIX	+	-	-	-	+	-	-	-	-	-	-	-	-	-	+	-	-
XX	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	+	-
XXII	+	-	-	-	+	-	+	+	-	-	-	-	-	-	-	-	+
XXIII	+	-	-	-	+	-	+	+	-	-	-	-	-	-	-	-	-

Presence (+) and absence (−) of heteroduplex pairs (h).

and αhXXIII, the latter two found in Pygmies, are variants of type αhIX, generated by point mutations and/or deletions (Table 2; Fig. 1). The Greater Asians constitute a very diverse group (Table 3) with samples from Oceania to Northeast Asia (Nei and Roychoudury 1993). Remarkable is the exclusive presence of type αhXVIII and αhXX in Mongolians (Table 3), the latter being predominant among the Buryat subpopulation (data not shown). Type αhXVIII is similar to type αhV but has an additional heteroduplex band h15 (Table 2). The Mongolian type αhXIX is a simple variant of αhXVIII originated by deletion of αh6 locus, thus missing the heteroduplex band h6 (Table 2; Fig. 1). Four new αh types were found in Brazilians (αhXIV, αhXV, αhXVI, and αhXVII) that may represent new mutations or recent migrations, reflecting the multi-ethnic origin of this population. Types αhXIV and αhXV can be derived from types αhV and αhXII, respectively, by a single deletion step (Table 2; Fig. 2). Type αhXVI can be derived from αhV by point mutations and αhXVII can be derived by point mutations from type αhV or by duplication and point mutations from αhIII (Fig. 2). Overall, Brazilians resembled Caucasians with the predominant αhII type (Table 3) reflecting the major European background of this collected sample from males involved in paternity disputes in Southeastern Brazil (Santos et al. 1993a). The population group with the largest frequency of the αhII type, almost always in association with *DYS19* allele A is that of Amerindians. In our initial study we haplotyped 73 Amerindian individuals from 12 different tribes ranging from Argentina to Mexico (Mapuche, Wichi, Chorote, Chulupi, Toba, Huilliche, Atacameño, Suruí, Karitiana, Quechua, Auca, Maia) and identified the presence of haplotype IIA in 74% of them (Pena et al. 1995b). We then studied 37 additional Amerindians belonging to five tribes from the Amazon Basin and Central Brazil (Waiwai, Gavião, Zoró, Suruí, Xavante). Again, haplotype IIA was present in the great majority (87%) of the individuals (Santos et al. 1995b). By pooling the results from the two studies, we calculate that the haplotype IIA was present in 78% of the Amerindians tested ($n = 110$). If we exclude Mapuches, who have a high degree of admixture, this percentage increases to 90%. We also studied a North Amerindian population, the Mvskokes (Creeks) from Oklahoma. The most frequent haplotype in Mvskokes was again IIA (38% of the individuals tested) which, considering the extensive admixture of this tribe with Caucasians and Blacks, is strong evidence that for this North Amerindian population this is also a major founder haplotype (Santos et al. 1996b). Although further studies will be necessary for confirmation, our data support the notion that haplotype IIA is the major, and perhaps, single founder haplotype in American Natives and, consequently, that North and South Amerindians originated from the same single migration of an ancestral Asian population in the Pleistocene.

New variation was also found for the types αhI that are characterized by the absence of heteroduplex bands. These individuals have a deletion of the left or the right-side loci, thus precluding the formation of heteroduplexes. In general, the αh types are determined by variation in the nucleotide sequences of right-side loci. In this

Table 3. Y-chromosome Haplotypes Found in Major Geographical Groups

Major geographical groups (360*)	Populations	αh X *DYS19* haplotypes
Africans (34)	Kenyans	IX-C (6), IX-B (4), IX-D (2), V-B (1), X-C (1)
	Pygmies	IX-D (4), IX-E (4), III-C (3), III-D (2), III-A (1), III-E (1), V-B (1), IX-C (1), XXII-A (1), XXIII-D (1)
	Bushman	I-Z (1)
Caucasians (54)	Europeans**	II-B (24), III-B (7), II-C (2), II-A (1), II-D (1), II-E (1), III-C (1), I-C (1), VI-B (1), VII-B (1), XIII-A (1)
	Indians	II-D (1), II-0 (1), III-E (1), VIII-C (1),
	Iraqi	III-C (1),
	unknown	III-D (2), IV-C (2), II-A (1), III-B (1), III-C (1), V-A (1)
Greater Asians (61)	Australians	II-B (1), IX-C (1), XII-A (1)
	Cambodian	XI-C (1)
	Chinese	III-C (5), III-E (1)
	Japanese	I-C (1), III-C (1), III-D (1)
	Melanesians	I-C (1), III-C (1)
	Mongolians	XVIII-D (10), XVIII-C (9), XX-B (6), III-B (3), III-E (2), II-B (2), II-C (2), II-F (2), I-B (2), I-A (1), II-D (1), III-D (1), IV-C (1), V-B (1), V-C (1), XVIII-B (1), XIX-B (1)
Amerindians (110)	Atacameño	II-A (1)
	Auca	II-A (1)
	Chorote	II-A (6)
	Chulupi	II-A (1)
	Gavião	II-A (12), I-A (5)
	Huilliche	II-B (2), IV-C (1)
	Karitiana	II-A (2)
	Maia	II-A (2)
	Mapuche	II-A (8), II-B (4), III-B (3), III-C (2), II-C (1), V-B (1), XII-A (1)
	Quechua	II-A (1)
	Suruí	II-A (8)
	Toba	II-A (5), IX-E (1)
	Waiwai	II-A (5)
	Wichi	II-A (24), II-B (2), XII-A (1)
	Xavante	II-A (5)
	Zoró	II-A (5)
mixed	Brazilians (101)	II-B (38), II-A (10), II-C (8), IV-C (7), III-B (6), III-C (6), I-A (2), I-C (2), III-D (2), IV-D (2), V-A (2), V-B (2), V-C (2), XV-A (2), XVI-A (2), I-D (1), I-E (1), II-D (1), IX-D (1), XII-A (1), XIV-B (1), XVI-B (1), XVII-B (1)

Numbers of individuals for populations and haplotypes are in parentheses.
(*) Four individuals are from unknown origin (Mathias et al. 1994).
(**) Most are British (Mathias et al. 1994).

sense, the deletion of the left-side locus is epistatic over variation on the right side, generating the unique αhI type independently of whatever loci are present on the right side. We can determine the molecular genotypes by artificially mixing into the PCR products a cloned left-side locus to generate heteroduplexes (Santos et al. 1995a). Eighteen αhI individuals were thus subdivided into five different genotypes (Table 4). Four individuals (three Amerindians and one Mongolian) did not show any heteroduplexes after mixing in of cloned αhL but did show them after addition of cloned right-side loci, thus demonstrating the deletion of all right-side loci. We called this type αhI[L]. The other 14 αhI individuals had deletion of αhL and thus had the presence of different right-side loci. One type found in one Brazilian, one Mongolian, and two Amerindians was called αhI[1], because it had only the right locus αh1 (if the left-side locus were present these males would

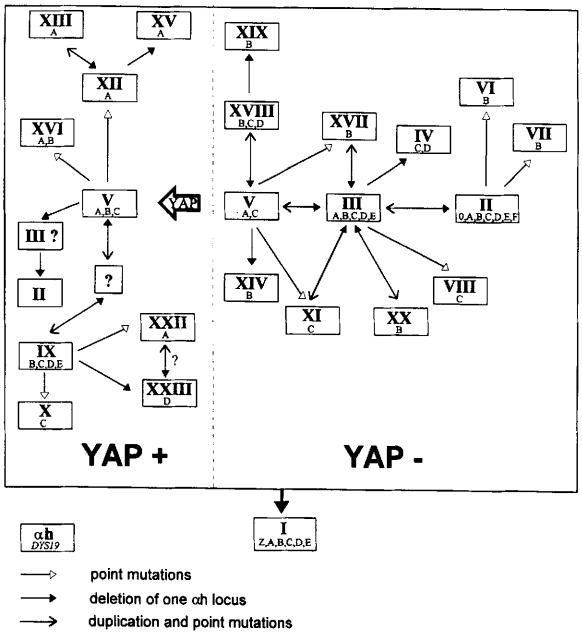

Figure 2 Origin and evolution of αh/YAP/*DYS19* haplotypes. It is assumed that the YAP insertion occurred in an individual carrying the αhV type. From this probable ancestor αhV/YAP$^+$ haplotype, several complex αh types (*left* of the dotted line) were created by mutational molecular events. These αh/YAP$^+$ haplotypes are mainly found in Africans whereas all other major geographic groups, and also some African Pygmies, display αh/YAP$^-$ haplotypes (*right* of the dotted line). Simple αh types, like αhXIV, αhIII, αhIV, αhII, present in YAP$^-$ chromosomes could theoretically be originated de novo by simple deletion events from the αhV/YAP$^+$ haplotype. The origin of the single individual αhII YAP$^+$ is shown in the diagram through a likely intermediate αhIII YAP$^+$ that was not yet found. Any αh type may convert to αhI by deletion of the left-side locus. Underneath each αh type are indicated the *DYS19* alleles seen associated with it worldwide.

be αhII). The αhI[1,5] genotype was even more common, being found in four Brazilians, three Greater Asians, and one Caucasian; if the left-side locus were present these males would be αhIII. Other αhI genotypes were found in single males and displayed in the mixing experiments, heteroduplexes not seen in any other individual (indicated by the prefix *v*): αhI[1,v1,v2] found in the Bushman!Kung (Santos et al. 1995a) and αhI[v3] found in a Brazilian.

Geographic Distribution of the Y-chromosome αh/*DYS19* Haplotypes

The overall worldwide Y-chromosome haplotype diversity was 0.87, but significant heterogeneity was observed among the major geographic groups. Amerindians ($D = 0.38$) presented reduced variability when compared with Caucasians ($D = 0.77$), Africans ($D = 0.88$), and Greater Asians ($D = 0.92$). This is because of the presence in Amerindians of the IIA founder Y-chromosome haplotype (Pena et al. 1995b, Santos et al. 1995b). In all populations, as expected, an obvious linkage disequilibrium is observed between the αh polymorphic system and *DYS19*. We quantified this in a random sample of 100 Brazilian Caucasians and obtained high statistical significance ($\chi h^2 = 310$, $P < 0.001$). The analysis of *DYS19* gene diversity [*D*(*DYS19*)] within each αh haplotype should be informative. *D*[*DYS19*] is highest for αhII (0.59), αhIII (0.69), αhIX (0.73), and αhXVIII (0.68). The simplest explanation for the high diversity seen in αhII and αhIII types would be that they probably do not represent unique mutational events, as discussed above. However, the analysis of 139 individuals of this sample with >10 other Y-linked polymorphisms (Mathias et al. 1994; Jobling and Tyler-Smith 1995; C. Tyler-Smith, pers. comm.) revealed that in most cases the αhII and αhIII types appear to have a single origin, because these two types belong to distinct Y-chromosome groups, that is, if αhII originated recurrently from αhIII (Fig. 2) these new cases of αhII type should be grouped together with αhIII types. Moreover, αhII individuals from Great Britain ($n = 29$) who most likely represent closely related patrilineages, display five different *DYS19* alleles. The de novo origin of simple types as αhII and αhIII must be rare events but are expected to occur as we demonstrated below. Therefore, in our case, the high *D*[*DYS19*] seen in αhII and αhIII may really reflect their old origin. Both αh types are today present in some ethnic groups where they have a different *DYS19* distribution and the sum of these frequencies increased the overall *D*[*DYS19*] in each αh type. It is evident in the case of the αhII type, which in Amerindians ($n = 95$) was associated with the *DYS19* allele A (91%) and in British Caucasians ($n = 32$) with the *DYS19* allele B (75%), respectively. The former represents a founder haplotype of Amerindians (Pena et al. 1995b; Santos et al. 1995b) whereas there is the suggestion that the latter may represent a founder haplotype in Great Britain (Santos et al. 1996a). The other types αhIX and αhXVIII with high *D*[*DYS19*] gene diversities, are complex (Table 2) and thus probably do reflect unique

Table 4. Genotypes Found in Individuals Displaying the αhI Pattern

Genotypes	N	αh loci					
		αh L	αh 1	αh 5	αh v1 *	αh v2 *	αh v3 *
I(L)	4	+					
I(1)	4		+				
I(1,5)	8		+	+			
I(1, v1, v2)	1		+		+	+	
I(v3)	1						+

(*) Variant (v) loci generate heteroduplexes after mixing with αhL (Santos et al. 1995a) that are only seen in these single males.

mutational events. Different than αhII and αhIII, they are found mostly in single geographic groups (type αhIX in Africans and type αhXVIII in Mongolians). The high *DYS19* variability associated with these αh types is probably a reflection of their antiquity and represent exclusively the diversity originated in these single ethnic groups.

The Origin and Evolution of αh Haplotypes and Other Polymorphisms

The YAP represents a unique molecular event (Hammer 1994). We performed the analysis of the YAP polymorphism in 240 of our sampled individuals. YAP typing was not done in most of the Amerindian samples because DNA was scarce and also because that was part of a separate study (see below). In our study only Y-chromosome haplotypes containing alphoid type αhV were found to be shared between YAP^+ and YAP^- chromosomes. Haplotypes $\alpha hV/YAP^+$ appeared in two Africans, six Brazilians, one Mongolian, the 3E7 human cell line, and one Amerindian. The αhV/YAP− appeared in only two individuals: one Caucasian and one Mongolian. In a recent study (N.O. Bianchi, G. Baillet, C.M. Bravi, R.F. Carnese, F. Rothhamer, V.L. Martinez-Marignac, and S.D.J. Pena, in prep.) one single instance of a YAP^+ αhII individual from La Plata (Argentina) was observed. This is very important because it supports our hypothesis that the αhII type can be generated recurrently. The evolutionary pathway for generation of αh haplotypes and their relation to *DYS19* and YAP polymorphisms is depicted in Figure 2.

DISCUSSION

Human Y-linked polymorphisms (in the nonpseudoautosomal portion of the chromosome) have a peculiar genetics. They are haploid, do not undergo recombination, and thus establish patrilineages. In addition, the number of Y chromosomes in any population is one-quarter the number of autosomal chromosomes, rendering Y-linked genes more prone to genetic drift. Moreover, the population dynamics of the male carriers of the Y chromosome is different from females. Gender-specific activities such as war and hunting as well as polygyny may cause even further reductions on the effective number of Y chromosomes in the population, increasing the predisposition to genetic drift (Pena et al. 1995b). Finally, it should be remembered that because of the lack of recombination, any selectively advantageous mutations in Y-linked traits will lead to simultaneous selection by hitchhiking, for all other polymorphic alleles in the haplotype. All these characteristics render Y-linked polymorphisms attractive for study of human evolution and population genetics.

We surveyed several human populations worldwide with three PCR-based polymorphisms located in the human Y chromosome: the αh polymorphic system, the *DYS19* microsatellite locus, and a YAP. By typing with the former two polymorphisms, we found 46 different haplotypes in 364 males from several populations worldwide. In a random sample of the Brazilian population this corresponded to a discrimination power of 0.82. Thus, Y-chromosome haplotyping with these two markers should have immediate practical use in forensic applications such as paternity testing (Santos et al. 1993b) and criminal investigations, especially in rape cases (Roewer and Epplen 1992).

For evolutionary studies it is of paramount importance to know the tempo and the molecular mechanisms underlying the variability of the polymorphism within and between populations. *DYS19* is a tetranucleotide microsatellite and as such should follow a stepwise mutation model

with recurrent and reversible allele generation by replication slippage (Pena et al. 1994). Rare alleles may originate by unique events, such as the cases of locus duplications (Santos et al. 1996b) and triplications (Santos et al. 1996a) that we have observed, but these are unimportant from the population point of view. Microsatellites generally have very high mutation rates, on the order of 10^{-4} to 10^{-5} (Edwards et al. 1992). However, two distinct lines of evidence point to a slower mutation rate for *DYS19*. First, as shown previously (Santos et al. 1996a), the worldwide distribution of allele frequencies of the *DYS19* tetranucleotide polymorphism exhibit a remarkable heterogeneity. Amerindians showed an overwhelming predominance of the A allele, whereas in Caucasians the B allele was modal, and in Greater Asians and Africans the allele C became predominant. Even within these geographic regions there were significant gradients, as exemplified by the decreasing frequency profile of the B allele from Britain over Germany to Slovakia. Second, we have shown that the majority of Amerindians from tribes distributed from Patagonia to the United States and from different linguistic groups belong to the Y-chromosome haplotype IIA (αhII type + *DYS19* A allele), including tribes with considerable admixture (Pena et al. 1995b; Santos et al. 1995b, 1996b). If *DYS19* evolved at the same fast rate as autosomal microsatellites, this strong founder effect should have already dissipated. The slower pace of *DYS19* evolution may be related to its haploidy. Microsatellite mutation may depend on more than simple replication slippage. Other mechanisms such as gene conversions (intra- and interallelic) could play a role, as proposed on the basis of the observation of a significant allele size association in haplotypes of two closely positioned microsatellites in intron 40 of the von Willebrandt gene (Pena et al. 1994).

For the αh system, knowledge of its molecular basis (Santos et al. 1995a) allowed us to propose a model for evolution, as explained in detail above. This model applies to all alphoid types, except type αhI, which represents the convergence of many different molecular mutational events. According to the model, mutations of alphoid right-side loci may lead to an increase or a decrease in complexity of the heteroduplex patterns. Increases in complexity should be unique events with reversion possible, whereas decreases in complexity are not necessarily unique but should be irreversible. This model was strongly supported by our haplotype studies with the polymorphic *Alu* insertion (YAP) in the Y chromosome. This insertion is supposed to have happened in an ancestral human Y chromosome in Africa (Hammer 1994). *Alu* insertions are known to be unique molecular events and there is no known mechanism for reversibility by specific removal (discussed in Pena et al. 1995a). Thus, as an initial analytical step, it makes sense to divide the world's Y chromosomes into two large groups, YAP^+ and YAP^- (Fig. 2). Most of YAP^+ αh types occur in Africans and correspond to complex patterns with at least three heteroduplex pairs (except αhXV; Table 2), and all of them can be derived from αhV. In our studies, only the haplotypes containing the alphoid type αhV (and one single instance of αhII in an individual from Argentina) were found to be shared between YAP^+ and YAP^- chromosomes. If we consider only the heteroduplex pairs, which are the detected genotypes (Table 2; Fig. 2), we also see that they are not shared between YAP^+ and YAP^- chromosomes, except h1, h5, and h6, which are present in the αhV type. Thus it is reasonable to assume that the YAP insertion occurred in an African individual carrying the αhV type. Working from our results with αh/YAP/*DYS19* haplotypes and following the parsimony rules established above for interconversion of alphoid types, we can build the evolutionary model depicted in Figure 2. Note that the *DYS19* data are very consistent with the model because in almost all cases the *DYS19* alleles seen in derivative αh types are present in the predecessor types. Several inferences emerge from this model. Although the analysis of other Y polymorphisms suggests that the recurrent generation of simple αh types is not a common event (see Results above; Santos et al. 1995a), the reductions in heteroduplex number by deletion events seems a likely pathway for the de novo generation of simple patterns derived from the haplotype YAP^+ αhV, such as αhXIV, αhIII, αhII, and αhIV. We expect that further studies will detect these simple αh types generated by deletion events YAP^+ branch, but not those αh types more complex than the αhV type that are seen in the YAP^- branch.

If we are able to determine the mutation rates for generation of all αh types and *DYS19*, we should be able to reach a reasonable estimate for the age of the YAP insertion. Besides, the distributions of some αh types in present-day populations allow the drawing of some evolutionary pictures. Many αh types present in high fre-

quency in single populations, such as αhIX (mainly Africans) and αhXVIII (Mongolians), could conceivably be of recent origin, but the high *DYS19* variability associated with these types rather suggests an ancient origin and relative population isolation. The presence of αhII in most Caucasians and Amerindians and some Greater Asians (Table 3), including Siberians (F.R. Santos, M.H. Crawford, M.S. Schanfield, and S.D.J. Pena, unpubl.), suggests a common ancestor for these three major groups as it was previously suggested for other Y-linked polymorphisms (Mathias et al. 1994; Jobling and Tyler-Smith 1995). The worldwide significant presence of αhIII in all major ethnic groups (Table 3) excepting Amerindians (a very recent major geographic group), suggests its antiquity. Most of the αhIII individuals belong to the group 2 of Y chromosomes (Mathias et al. 1994), which was recently suggested as the best candidate for the ancestral Y haplotype (Jobling and Tyler-Smith 1995).

In conclusion, the present study confirms the evolutionary usefulness of the αh polymorphism, which reflects the variation in many nucleotide sites detected by a single PCR amplification system. In the future, studies of αh typing can be associated with the analysis of other highly variable polymorphisms to answer specific evolutionary issues in their appropriate time scale.

METHODS

Populations Sampled

In this study, 364 males from several populations were used. To group the populations we followed the nomenclature of Nei and Roychoudhury (1993): Four major subdivisions of the phylogenetic tree of human populations are represented by Africans, Caucasians, Greater Asians, and Amerindians. Ninety-one individuals belonging to different geographic groups including Caucasians (Europeans, Indians, and an Iraqi), Greater Asians (Chinese, Japanese, one Cambodian, Australians, and Melanesians), and Africans (Kenyans, Pygmies, and a Bushman Kung) have been described elsewhere in detail (Mathias et al. 1994). One hundred and ten Amerindians from 16 different tribes, 101 Brazilians, 46 Mongolians, and 16 Pygmies had already been typed with both αh and *DYS19* as described by Pena et al. (1995b) and Santos et al. (1995b). An additional sample of 97 individuals (proved sons of the Brazilian individuals; Santos et al. 1993a) was also analyzed for the inheritance of the αh polymorphism. Samples were received as purified DNA or were prepared from agarose plugs as described in Mathias et al. (1994).

PCR and Electrophoresis

DYS19 *Locus*

PCR reactions and electrophoresis were performed as described in Santos et al. (1993a, 1996a); however, when using poor-quality DNA templates, the amount of *Taq* polymerase (a kind gift of Cenbiot, RS, Brazil) increased to 2 units/tube and the reactions underwent 40 cycles in the amplification program.

System αh

PCR conditions were basically those described in Santos et al. (1995a). However, recently we have developed improved PCR conditions as follows: The reaction was performed in a volume of 12.5 μl in 50 mM KCl, 10 mM Tris-HCl (pH 8), 1.5 mM $MgCl_2$, 1 μM of each primer, 1 unit of *Taq* polymerase (Cenbiot, RS, Brazil), and 5–50 ng of genomic DNA. The reaction mix was subjected to 45 cycles of 94°C for 30 sec, 65°C for 30 sec, and 72°C for 60 sec in a PTC150 Thermo-Cycler (MJ Research). When using poor-quality DNA templates the primer concentration was increased to 2 μM, and 2 units/tube of *Taq* were used to allow a better heteroduplex formation. The products (5 μl) were resolved in 16-cm-long (1 mm thick), 1× MDE (AT Biochem) polyacrylamide gel in 1× TBE at 120 V for 15 hr.

YAP

PCR was performed with primers described by Hammer and Horai (1995) in reaction conditions as described above for αh, except that 0.5 units of *Taq* was used per reaction. The reaction mixture was subjected to 30 cycles of 94°C for 60 sec, 55°C for 60 sec, and 72°C for 60 sec in a PTC 150 Thermo-Cycler (MJ Research). Five microliters of the amplified products was loaded on a 1.5% agarose gel in 0.5× TAE or a 5% polyacrylamide gel in 1× TBE and run for ~1 and 2 hr, respectively, at 100 V.

Staining of Gels

Agarose gels were stained with ethidium bromide. For polyacrylamide gels the products were visualized by silver staining as described in Santos et al. (1993a) with minor modifications. The gel was left for 10 min in the fixative solution [10% (vol/vol) ethanol, 0.5% (vol/vol) acetic acid] followed by a 10-min incubation in the silver nitrate solution (150 mg/300 ml ddH_2O). The gel was then submerged in developer solution (1 liter contains 15 g of NaOH and 3 ml of 37% formaldehyde) that is changed after darkening. When the bands became visible the developer solution was replaced with the fixative solution. During this process the gels are stained under agitation, using glassware and clean gloves when replacing the solutions in volumes of ~300 ml.

Quantification of Genetic Variation

Gene diversity was calculated for each population and ma-

jor ethnic group (Nei and Roychoudury 1993) according to the formula $D = 1 - \Sigma\, i^2$, where i is the allelic or haplotypic frequency.

ACKNOWLEDGMENTS

We thank Dr. Chris Tyler-Smith for providing unpublished information, advice, and comments on the manuscript. This work was supported by grants from Conselho Nacional de Pesquisas (CNPq), Financiadora de Estudos e Projetos (FINEP), Fundaĉão de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) and UNESCO.

REFERENCES

Dorit, R.L., H. Akashi, and W. Gilbert. 1995. Absence of polymorphism at the ZFY locus on the human Y chromosome. *Science* **268:** 1183–1185.

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

Gomolka, M., J. Hundrieser, P. Nurnberg, L. Roewer, J.T. Epplen, and C. Epplen. 1994. Selected di- and tetranucleotide microsatellites from chromosomes 7, 12, 14 and Y in various Eurasian populations. *Hum. Genet.* **93:** 592–596.

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

———. 1995. A recent common ancestry for human Y chromosomes. *Nature* **378:** 376–378.

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

Jakubiczka, S., J. Arnemann, H.J. Cooke, M. Krawczak, and J. Schmidtke. 1989. A search for restriction fragment length polymorphism on the human Y chromosome. *Hum. Genet.* **84:** 86–88.

Jobling, M.A. 1994. A survey of long-range DNA polymorphisms on the human Y chromosome. *Hum. Mol. Genet.* **3:** 107–114.

Jobling, M.A. and C. Tyler-Smith. 1995. Fathers and sons: The Y chromosome and human evolution. *Trends Genet.* **11:** 449–456.

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

Mathias, N., M. Bayés, and C. Tyler-Smith. 1994. Highly informative compound haplotypes for the human Y chromosome. *Hum. Mol. Genet.* **3:** 115–124.

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

Oakey, R. and C. Tyler-Smith. 1990. Y chromosome DNA haplotyping suggests that most european and asian men are descended from one of two males. *Genomics* **7:** 325–330.

Pena, S.D.J., K.T. Souza, M. Andrade, and R. Chakraborty. 1994. Allele associations of two polymorphic microsatellites in intron 40 of the human von Willebrandt factor gene. *Proc. Natl. Acad. Sci.* **91:** 723–727.

Pena, S.D.J., V.F. Prado, and J.T. Epplen. 1995a. Diagnosis of human genetic individuality. *J. Mol. Med.* **73:** 555–564.

Pena, S.D.J., F.R. Santos, N. Bianchi, C.M. Bravi, F.R. Carnese, F. Rothhammer, T. Gerelsaikhan, B. Munkhtuja, and T. Oyunsuren. 1995b. Identification of a major founder Y-chromosome haplotype in Amerindians. *Nature Genet.* **11:** 15–16.

Roewer, L. and J.T. Epplen. 1992. Rapid and sensitive typing of forensic stains by PCR amplification of polymorphic simple repeat sequences in case work. *Forensic Sci. Int.* **53:** 163–171.

Roewer, L., J. Arnemann, N.K. Spurr, K.-H. Grzeschik, and J.T. Epplen. 1992. Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts. *Hum. Genet.* **88:** 389–394.

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

Santos, F.R., J.T. Epplen, and S.D.J. Pena. 1993b. Testing deficiency paternity cases with a Y-linked tetranucleotide repeat polymorphism. In *DNA fingerprinting: State of the science* (ed. S.D.J. Pena, R. Chakraborty, J.T. Epplen, and A.J. Jeffreys), pp. 261–265. Birkhauser Verlag, Basel, Switzerland.

Santos, F.R., S.D.J. Pena, and C. Tyler-Smith. 1995a. PCR haplotypes for the human Y chromosome based on alphoid satellite variants and heteroduplex analysis. *Gene* **165:** 191–198.

Santos, F.R., M. Hutz, C.E.A. Coimbra, R.V. Santos, F.M. Salzano, and S.D.J. Pena. 1995b. Further evidence for the existence of a major founder Y chromosome haplotype in Amerindians. *Braz. J. Genet.* **18:** 669–672.

Santos, F.R., T. Gerelsaikan, B. Munkhtuja, T. Oyunsuren, J.T. Epplen, and S.D.J. Pena. 1996a. Geographic differences in the allele frequencies of the human Y-linked tetranucleotide polymorphism *DYS19*. *Hum. Genet.* **97:** 309–313.

Santos, F.R., L. Rodriguez-Delfin, S.D.J. Pena, J. Moore,

and K.M. Weiss. 1996b. North and South Amerindians may have the same major founder Y chromosome haplotype. *Am. J. Hum. Genet.* **58:** 1369–1370.

Seielstad, M.T., J.M. Herbert, A.A. Lin, P.A. Underhill, M. Ibrahim, D. Vollrath, and L.L. Cavalli-Sforza. 1994. Construction of human Y-chromosomal haplotypes using a new polymorphic A to G transition. *Hum. Mol. Genet.* **3:** 2159–2161.

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

Whitfield, L.S., J.E. Sulston, and P.N. Goodfellow. 1995. Sequence variation of the human Y chromosome. *Nature* **378:** 379–380.

Received February 26, 1996; accepted in revised form May 9, 1996.