

# Variation in Short Tandem Repeats Is Deeply Structured by Genetic Background on the Human Y Chromosome

Elena Bosch,<sup>1</sup> Francesc Calafell,<sup>1</sup> Fabrício R. Santos,<sup>2,3</sup> Anna Pérez-Lezaun,<sup>1</sup> David Comas,<sup>1</sup> Noufissa Benchemsi,<sup>4</sup> Chris Tyler-Smith,<sup>2</sup> and Jaume Bertranpetit<sup>1</sup>

<sup>1</sup>Unitat de Biologia Evolutiva, Facultat de Ciències de la Salut i de la Vida, Universitat Pompeu Fabra, Barcelona; <sup>2</sup>Cancer Research Campaign Chromosome Molecular Biology Group, Department of Biochemistry, University of Oxford, Oxford; <sup>3</sup>Departamento de Biologia Geral, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil; and <sup>4</sup>Centre National de Transfusion Sanguine, Rabat, Morocco

## Summary

Eleven biallelic polymorphisms and seven short-tandem-repeat (STR) loci mapping on the nonrecombining portion of the human Y chromosome have been typed in men from northwestern Africa. Analysis of the biallelic markers, which represent probable unique events in human evolution, allowed us to characterize the stable backgrounds or haplogroups of Y chromosomes that prevail in this geographic region. Variation in the more rapidly mutating genetic markers (STRs) has been used both to estimate the time to the most recent common ancestor for STR variability within these stable backgrounds and to explore whether STR differentiation among haplogroups still retains information about their phylogeny. When analysis of molecular variance was used to study the apportionment of STR variation among both genetic backgrounds (i.e., those defined by haplogroups) and population backgrounds, we found STR variability to be clearly structured by haplogroups. More than 80% of the genetic variance was found among haplogroups, whereas only 3.72% of the genetic variation could be attributed to differences among populations—that is, genetic variability appears to be much more structured by lineage than by population. This was confirmed when two population samples from the Iberian Peninsula were added to the analysis. The deep structure of the genetic variation in old genealogical units (haplogroups) challenges a population-based perspective in the comprehension of human genome diversity. A population may be better understood as an association of lineages from a deep and population-independent gene genealogy, rather than as a complete evolutionary unit.

Received July 21, 1999; accepted September 8, 1999; electronically published November 23, 1999.

Address for correspondence and reprints: Dr. Jaume Bertranpetit, Unitat de Biologia Evolutiva, Facultat de Ciències de la Salut i de la Vida, Universitat Pompeu Fabra, Doctor Aiguader 80, 08003 Barcelona, Catalonia, Spain. E-mail: jaume.bertranpetit@cexs.upf.es

© 1999 by The American Society of Human Genetics. All rights reserved. 0002-9297/1999/6506-0018\$02.00

## Introduction

Human evolution and population studies based on the Y chromosome have increased notably during recent years (Hammer 1995; Jobling and Tyler-Smith 1995; Santos et al. 1995*b*; Whitfield et al. 1995; Cooper et al. 1996; Deka et al. 1996; Roewer et al. 1996; de Knijff et al. 1997; Hammer et al. 1997, 1998; Pérez-Lezaun et al. 1997, 1999; Hurles et al. 1998). The singular characteristics of the Y chromosome, which include paternal inheritance and absence of recombination through most of its length, make this chromosome a powerful tool for tracing and comparing paternal lineages of human populations in a way similar to the use of mtDNA to study maternal lineages. In spite of the slow initial discovery of polymorphic markers on the Y chromosome, which started in the mid 1980s (Casanova et al. 1985; Lucotte and Ngo 1985), the detection of variation on this chromosome has grown dramatically during recent years (Jobling and Tyler-Smith 1995; Hammer and Zegura 1996; Underhill et al. 1997). The variety of polymorphic markers now available on the nonrecombining portion of the Y chromosome ranges from base substitutions and deletion/insertion polymorphisms, which are rare (probably even unique) events in evolution and which tend to be biallelic, to faster-mutating polymorphisms such as microsatellites—also known as short tandem repeats (STRs)—and the MSY1 minisatellite (Bouzekri et al. 1998; Jobling et al. 1998*a*).

The presence of different types of polymorphisms with different mutational mechanisms and rates strengthens even more the applicability of analysis of the nonrecombining portion of the Y chromosome to the study of human evolution at different geographic and time scales. Some examples of the wide range of subjects to which the study of Y-chromosome polymorphisms has been applied include investigations of both the origin and the dispersal of anatomically modern humans (Hammer et al. 1997, 1998), the first settlement of the Americas (Pena et al. 1995; Santos et al. 1995*a*, 1996, 1999*a*; Underhill et al. 1996; Karafet et al. 1997), the Asian

paternal contribution to northern European populations (Zerjal et al. 1997), the colonization of Polynesia as well as later European admixture (Hurles et al. 1998), the colonization of mountain habitats in central Asia (Pérez-Lezaun et al. 1999), and the study of both the Cohen (Thomas et al. 1998) and the Jefferson (Foster et al. 1998) lineages.

The evolution of each of these markers on the Y chromosome is not independent, since changes to a particular locus always happen on a well-defined background for all the other polymorphisms to which it remains linked because of the absence of recombination. In particular, the absence of recombination has facilitated the inference of the genealogy of haplogroups—that is, groups of chromosomes defined by the combination of alleles at different biallelic polymorphisms. The combination of slow- and fast-mutating systems has added values. The typing of STRs within haplogroups has allowed the investigation of the origin and dispersal of certain haplogroups (Zerjal et al. 1997; Hurles et al. 1998), as well as examination of the population movements associated with those dispersals.

Genetic diversity is often thought of as being structured by population, to the extent that a population perspective has been thought of as a complete description of the genetic diversity. Classic population-genetics theory modeled the dynamics of allele-frequency change among populations, in terms of forces such as drift, selection, and migration. This approach implies that a mere description of variation is fully informative on genetic grounds. However, genetic diversity may be more deeply structured by gene genealogy than by the ethnogenesis process that gave rise to the population. The gene genealogy of a particular genome region can be recognized from the inferred genealogy of its slowly mutating polymorphisms, which constitute the stable background on which the variation of faster-mutating systems took place. This rapidly produced variation may contain information on the evolution of each deep branch of the gene genealogy and may even include a detailed footprint of the full gene genealogy. The population perspective may thus be out of the scale of the genetic processes. For instance, Estivill et al. (1994) found that haplotypes of three STRs in the cystic fibrosis transmembrane conductance regulator (CFTR) gene were clearly different in chromosomes bearing the  $\Delta F508$  mutation (which causes cystic fibrosis) compared with those found in nonaffected chromosomes. Moreover, STR haplotypes from any of a number of nonaffected European populations were much closer to each other than were haplotypes from nonaffected and  $\Delta F508$ -bearing chromosomes belonging to the same population. Therefore, genetic diversity within the

CFTR gene in Europe is more deeply structured by genetic background than by population.

Since the production of biallelic variation is continuous in time, some of that variation—the most ancient—will define haplogroups with wide geographic distributions; however, most of it, being more recent, will be found in more-restricted areas and may have occurred after certain population splits. In fact, there are population-specific Y-chromosome biallelic polymorphisms, such as Tat in northwestern Asians (Zerjal et al. 1997), SRY-2627 in north Iberians (Hurles et al. 1999), and DYS199 in Native Americans (Underhill et al. 1996). The most recent biallelic markers may be population-specific and even family-specific. However, a characteristic distribution of haplogroups will be found in each population, and some of those haplogroups may be much older than the populations in which they are found.

In this study, we have scored 11 Y-specific biallelic markers in men from northwestern Africa; this has allowed the characterization, for the first time, of Y-chromosome haplogroup distribution in this area. We have also explored Y-chromosome variation at seven STRs in the same individuals, to characterize genetic variation by stable genetic background (i.e., by biallelic polymorphism haplogroup) and to compare the apportionment of genetic variation by haplogroup with that apportioned by population. This will allow an alternative perspective to the population approach for the comprehension of human genetic diversity. The time to the most recent common ancestor (TMRCA) for the STR variability within haplogroups and the microsatellite diversity within them were analyzed as well. These analyses were also done in two population samples from the northern Iberian Peninsula. Finally, we explored whether STR differentiation by haplogroup retains information about the haplogroup phylogeny.

## Material and Methods

### *Samples*

Genetic analyses were performed on a sample of Y chromosomes from 129 unrelated healthy men from northwestern Africa. Appropriate informed consent was obtained from all participants in this study, and, in most cases, information about both the geographic origin and native language of each man's four grandparents was recorded. The samples obtained were from 44 Arabs, 42 Tahelhits, and 14 other Moroccan Berbers, as well as from 29 Saharawis. DNA was extracted from fresh blood by use of standard phenol-chloroform protocols. For parts of the data analysis, additional samples (from 51 Basques and 27 Catalans) from northern Iberia were included.

*Biallelic Polymorphism Typing*

We typed eight base substitutions, an Alu insertion, the polymorphic number of adenine residues in its 3' end, and a duplication/deletion polymorphism (see table 1). Polymorphism 92R7, which was originally described as an RFLP by Mathias et al. (1994), was converted to a PCR format by M. E. Hurles, F. R. Santos, and C. Tyler Smith (unpublished data), by amplification of a 55-bp fragment containing the polymorphic site in the 92R7 system. *HindIII* digestion was used to detect the C→T base substitution that destroys a *HindIII* site equivalent to the presence of the 6.7-kb band in the 92R7 Southern blots (Mathias et al. 1994). SRY-2627 (also known as pSRY-373) was analyzed by PCR amplifica-

tion, as reported elsewhere (Bianchi et al. 1997), by use of the pSRY244 (forward) and pSRY634 (reverse) primers. *BanI* digestion was employed to detect the C→T transition at base-pair position 373 (Santos et al. 1999a). SRY-1532 screening was performed with primers SRY-1 and SRY-2 (Santos et al. 1999b), which amplify a 167-bp male-specific fragment spanning the polymorphic position 10,831 of region SRY (Whitfield et al. 1995; Kwok et al. 1996). PCR and cycling conditions were performed as described elsewhere (Santos et al. 1999a). Amplified products were digested with *DraIII* and were incubated with Pronase (Boehringer Mannheim). The Y-chromosome Alu insertion polymorphism (YAP) element at the DYS287 locus was analyzed by PCR amplification, as

**Table 1**  
**Y-Chromosome Biallelic Polymorphisms Analyzed and Their Allele Frequencies in Northwestern Africa**

Polymorphism <sup>a</sup> and Allele	Frequency	Type	Method	References
92R7:				
C	124 (.961)	Base substitution	<i>HindIII</i> site loss, PCR	Mathias et al. (1994)
T	5 (.039)			
SRY-2627:				
C	129 (1)	Base substitution	<i>BanI</i> site loss, PCR	Bianchi et al. (1997); Santos et al. (1999a)
T	0 (0)			
SRY-1532:				
A	0 (0)	Base substitution	<i>DraIII</i> site gain, PCR	Whitfield et al. (1995); Kwok et al. (1996); Santos et al. (1999a)
G	129 (1)			
YAP: <sup>b</sup>				
0	24 (.186)	Alu insertion	PCR	Hammer (1994); Hammer and Horai (1995)
1	105 (.814)			
Poly (A) tail: <sup>c</sup>				
L	1 (.010)	Poly (A) tail-length polymorphism	PCR	Hammer (1995); Hammer et al. (1997)
S	104 (.990)			
SRY-8299:				
G	24 (.186)	Base substitution	<i>BsrBI</i> site loss, PCR	Whitfield et al. (1995); Santos et al. (1999a)
A	105 (.814)			
sY81:				
A	123 (.953)	Base substitution	<i>NlaIII</i> site loss, PCR	Seielstad et al. (1994)
G	6 (.047)			
M9: <sup>d</sup>				
C	123 (.953)	Base substitution	PCR, DHPLC <sup>e</sup>	Underhill et al. (1997)
G	6 (.047)			
12f2: <sup>d</sup>				
10 kb	114 (.884)	Duplication/deletion	<i>TaqI</i> ( <i>EcoRI</i> ), filter hybridization	Casanova et al. (1985)
8 kb	15 (.116)			
50f2 P: <sup>d</sup>				
8.5 kb	129 (1)	Base substitution	<i>TaqI</i> , filter hybridization	Guellaen et al. (1984)
3.1 kb	0 (0)			
50f2 I: <sup>d</sup>				
8.5 kb	129 (1)	Base substitution	<i>TaqI</i> , filter hybridization	Guellaen et al. (1984)
4 kb	0 (0)			

<sup>a</sup> For each biallelic polymorphism, the ancestral state is presented above the derived state.

<sup>b</sup> For the YAP polymorphism, “0” denotes absence of the Alu sequence and “1” denotes presence of the *Alu* sequence.

<sup>c</sup> The poly (A) tail-length polymorphism is found within the YAP element, and four different alleles have been described so far. Of those, we found two—S (small, 26 bp) and L (large, 46 bp)—in our sample. Their frequencies are given with respect to the total number of YAP-positive chromosomes.

<sup>d</sup> Only YAP-negative individuals were tested for these polymorphisms. YAP-positive individuals were presumed to have polymorphisms M9 C, 12f2 10 kb, 50f2 P 8.5 kb, and 50f2 I 8.5 kb.

<sup>e</sup> DHPLC = denaturing high-performance liquid chromatography.

described by Hammer and Horai (1995). Variation in the number of adenine residues at the 3' end of the YAP Alu sequence, also known as the poly (A) tail-length polymorphism (Hammer 1995; Hammer et al. 1997), was typed by resolution of the amplified products of YAP-positive individuals on 20 × 20-cm (1 mm-thick) 6% polyacrylamide gels in 1 × Tris-borate EDTA at 40 mA for 5 h and by visualization with silver staining. The SRY-8299 system was genotyped with the primers and PCR conditions described by Santos et al. (1999a). The amplified fragments, which contained the G→A polymorphic site at position 4,064 of the SRY region (Whitfield et al. 1995), were then digested with *Bsr*BI and were analyzed by electrophoresis (Santos et al. 1999a). Polymorphism sY81 (DYS271) was amplified as described elsewhere (Seielstad et al. 1994). Amplified products were digested with *Hsp*92II (isoschizomer of *Nla*III). The M9 C→G base substitution was PCR amplified and was typed by denaturing high-performance liquid chromatography, as described elsewhere (Underhill et al. 1997). A total of 1 μg of genomic DNA from each YAP-negative individual was digested to completion by use of 20 U of *Taq*I (Boehringer Mannheim); it was then electrophoresed on a 1% agarose gel in 0.5 × Tris-acetate EDTA for 14 h at 25 V and was transferred to Hybond<sup>™</sup>-N+ nylon membranes, by use of standard procedures (Southern 1975). DNA probes 50f2 (DYS7) and 12f2 (DYS11) for Southern blot analysis were radioactively labeled by random priming (Feinberg and Vogelstein 1983, 1984). After prehybridization with salmon sperm DNA, filters were hybridized overnight with probe 12f2 (Casanova et al. 1985) at 68°C in 1% SDS 5 × Denhardt's solution, 10% dextran sulphate, and 0.5 M sodium phosphate; they were then washed, at 65°C, to a stringency of 0.1 × SSC plus 1% SDS. The filter hybridization and washing conditions used, plus further details for probe 50f2, can be found in a report by Jobling (1994). In both cases, bands were visualized by autoradiography done at -70°C with Fuji Rx film. Biallelic polymorphism data for Basques and Catalans were obtained from Semino et al. (1996), Underhill et al. (1997) and Underhill (unpublished results), and Hurles et al. (1999).

#### STR Polymorphism Typing

Two trinucleotide repeat polymorphisms—DYS388 and DHS392—and six tetranucleotide repeat polymorphisms—DYS19, DHS389I, DHS389II, DHS390, DHS391, and DHS393—were typed in all Y chromosomes. A PE Biosystems 9600 thermal cyler was used. PCR reactions were performed in a 10-μl final reaction volume containing 100 ng genomic DNA, 50 mM KCl, 10 mM tris-HCl (pH 8.3), 1.5 mM MgCl<sub>2</sub> (2.5 mM for DHS19), 250 μM each dNTP, 0.2 μM

each primer, and 1 U *Taq* DNA Polymerase (Gibco BRL). Forward primers were fluorescently labeled. The PCR cycling conditions used were those described by Pérez-Lezaun et al. (1999). PCR products were run in an ABI 377<sup>™</sup> sequencer. ABI GS500 TAMRA was used as internal lane standard. The GENESCAN 672<sup>™</sup> and GENOTYPER 1.1<sup>™</sup> software packages were used to collect the data and to analyze fragment sizes. Y-STR alleles were named according to the number of repeat units they contain. The number of repeats was established through the use of sequenced allele ladders and reference samples kindly provided by P. de Knijff. Genome Database primers for the DHS389 locus amplify a partially duplicated region and generate two PCR products, which are referred to as DHS389I (239–263 bp) and DHS389II (353–385 bp). Both fragments are variable in length, and the study of their sequence structure has shown that DHS389II contains DHS389I plus two additional stretches of tetranucleotide repeats (Rolf et al. 1998; Pestoni et al. 1998). Therefore, we have used only the length variability of the shorter fragment in the numerical analysis. The STR haplotypes for northern Iberians were those described by Pérez-Lezaun et al. (1997).

#### Data Analysis

Analysis of molecular variance (AMOVA) was performed for STR allele frequencies among haplogroups, by use of the Arlequin package (Schneider et al. 1997). A simple hierarchical partitioning of haplotypes in haplogroups was tested, without further pooling haplogroups. Genetic dissimilarity among STR haplotypes was weighed by the difference in allele length, which is equivalent to the  $R_{ST}$  measure (Slatkin 1995; Schneider et al. 1997). AMOVA was also performed directly on STR haplotype frequencies; since the results were very similar to those obtained with  $R_{ST}$ , only  $R_{ST}$  results are given. Genetic diversity within each haplogroup was measured by different parameters, such as the mean number of allele differences between pairs of Y-chromosome haplotypes and the mean number of differences in repeat sizes between pairs of Y-chromosome haplotypes. To test whether these parameters were statistically significantly different between haplogroups, we performed a permutation procedure similar to those described in Graven et al. (1995) and in Mateu et al. (1997). In each iteration, chromosomes are permuted across haplogroups, the relevant parameter is recomputed in both haplogroups, and the difference is recorded. In this way, after 10,000 iterations, a distribution of the difference in the parameter between haplogroups is obtained under the null hypothesis of no difference. The probability of obtaining a difference in

**Table 2**

**Y-Chromosome Haplogroups Studied and Their Frequencies in Northwestern Africa**

BIALLELIC POLYMORPHISM	HG	HG	HG	HG	HG	HG	HG	HG	HG	HG	HG	HG	
	7	3	2+	26	1	22	15	6	9	4	21 / 21L	8	
	Status of Allele												
92R7	0	1	0	0	1	1	0	0	0	0	0	0	
SRY-2627	0	0	0	0	0	1	0	0	0	0	0	0	
SRY-1532	0	0	1	1	1	1	1	1	1	1	1	1	
YAP (poly [A] tail) <sup>a</sup>	–	–	–	–	–	–	–	–	–	+	(S)	+	
SRY-8299	0	0	0	0	0	0	0	0	0	0	1	+	
sY81	0	0	0	0	0	0	0	0	0	0	0	1	
M9	0	1	0	1	1	1	0	0	0	0	0 <sup>b</sup>	0 <sup>b</sup>	
12f2	0	0	0	0	0	0	0	0	1	0	0 <sup>b</sup>	0 <sup>b</sup>	
50f2 P	0	0	0	0	0	0	0	1	0	0	0 <sup>b</sup>	0 <sup>b</sup>	
50f2 I	0	0	0	0	0	0	1	0	0	0	0 <sup>b</sup>	0 <sup>b</sup>	
	No. (Frequency [ <i>n</i> = 129])												
	0 (0)	0 (0)	3 (.023)	1 (.008)	5 (.039)	0 (0)	0 (0)	0 (0)	15 (.116)	0 (0)	98 (.759)/1	.008)	6 (.047)

NOTE.—For each biallelic polymorphism, alleles are represented as “0” or “1,” according to the presence of an ancestral or derived state, respectively (see table 1).

<sup>a</sup> Minus sign (–) = YAP-negative; plus sign (+) = YAP-positive; (S) = short poly (A) tail; and (L) = long poly (A) tail.

<sup>b</sup> Assumed but not typed.

the parameter between two haplogroups that was larger than the observed difference was recorded.

TMRCAs of the STR variability generated within the chromosomes bearing the derived allele of each biallelic polymorphism was estimated from the mean allele-size variance of the seven STRs within all chromosomes bearing that derived allele by use of equation 5 in Di Rienzo et al. (1998):  $V = T\mu\eta_2$ , where  $V$  is the variance in repeat size,  $T$  stands for time in generations after a population expansion,  $\mu$  is the mutation rate, and  $\eta_2$  is the variance in the number of repeats gained or lost at each mutation event. The average mutation rate used was  $1.2 \times 10^{-3}$ , with a 95% confidence interval (CI) of  $4.6 \times 10^{-4}$  to  $2.8 \times 10^{-3}$  (Bianchi et al. 1998). This estimate comprises data from deep-rooting pedigrees (Heyer et al. 1997) and for father-son transmissions (Kayser et al. 1997), as well as from family cell lines from the CEPH. It has been reported that microsatellites tend to accumulate mutations in lymphoblastoid cell lines, which would result in an overestimation of germline mutation rates (Banchs et al. 1994). However, Bianchi et al. (1998) did not find any mutations in the CEPH cell lines they typed, and, therefore, this method does not bias their germline-mutation-rate estimate upward. Since all mutations observed (Heyer et al. 1997; Kayser et al. 1997) consist of the gain or loss of one repeat, mutation-size variance ( $\eta_2$ ) was set to 1; the generation time used was 20 years. The 95% CIs for TMRCAs were estimated by taking into account both the interlocus sampling variance and the 95% CI of the mutation rate estimate.

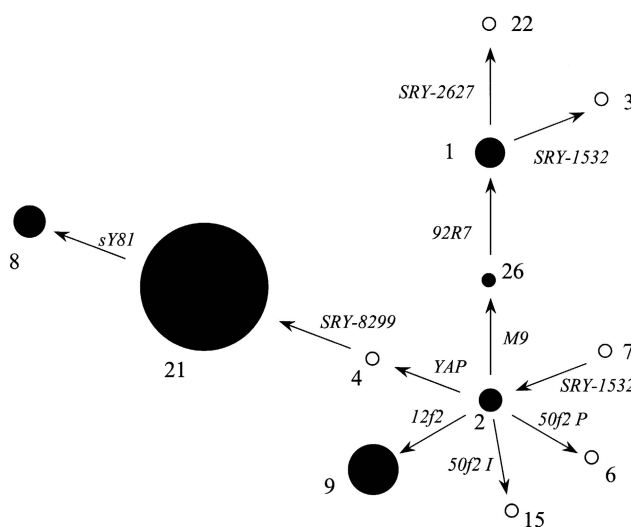
Net genetic distances between haplogroups ( $D_{ij}$ ) were computed, according to the Jensen Difference (Rao 1982), as  $D_{ij} = d_{ij} - ([d_{ii} + d_{jj}]/2)$ , where  $d_{ij}$  is the mean

number of absolute differences in allele length between pairs of chromosomes of haplogroups  $i$  and  $j$  and where  $d_{ii}$  and  $d_{jj}$  are the mean pairwise allele length differences within haplogroups  $i$  and  $j$ , respectively. In its version for DNA-sequence data,  $D_{ij}$  is the most common measure of genetic distance among population samples of mtDNA sequences (see, for example, Bertranpetit et al. [1995]). It represents the average genetic difference between haplotypes belonging to different groups, corrected for the average variability within those groups.

**Results**

*Y-Chromosome Biallelic Polymorphisms*

Allele frequencies of the 11 Y-chromosome biallelic polymorphisms in the total sample of 129 northwestern African men studied are shown in table 1. The combination of their allelic states allowed us to identify 6 of the 12 different haplogroups of Y chromosomes that have been described to date with the same polymorphisms (see table 2). Ancestral and derived alleles for each biallelic polymorphism (92R7 in Jobling 1994 and in Mathias et al. 1994) were inferred or were established by typing of nonhuman primates. These polymorphisms include SRY-2627 (Bianchi et al. 1997; Hurler et al. 1999), SRY-1532 (Whitfield et al. 1995), YAP (Hammer and Horai 1995), poly (A) tail (Hammer et al. 1997), SRY-8299 (Whitfield et al. 1995), sY81 (Seielstad et al. 1994), M9 (Underhill et al. 1997), 12f2 (Casanova et al. 1985), and 50f2 P and 50f2 I (Guellaen et al. 1984). Haplogroup distribution within the northwestern African population



**Figure 1** Parsimony network relating the haplogroups defined by 10 of the unique-event polymorphisms analyzed. Arrowheads indicate the derived states, as inferred or established by typing nonhuman primates (Jobling and Tyler-Smith 1995; Jobling et al. 1996, 1997; Hurles et al. 1998). Blackened circles are proportional to haplogroup frequency, and unblackened circles indicate haplogroups detectable but not found in our set of 129 Y chromosomes from northwestern Africa.

studied is also shown in the parsimony network of figure 1, on the basis of other networks (Jobling and Tyler-Smith 1995; Jobling et al. 1996, 1997; Hurles et al. 1998), except for the consideration of the SRY-8299 polymorphism defining haplogroup (HG) 21, which will be described elsewhere (Vogt et al. 1997). All but one of the Y chromosomes carrying the YAP insertion presented the small allele with respect to the poly (A) tail-length polymorphism of the YAP element (Hammer et al. 1997). The exception carried the large allele and belonged to HG 21. Except where indicated, this chromosome has been included in HG 21 in numerical analyses. The main feature of Y-chromosome-haplogroup distribution within northwestern Africa is the high frequency (76.7%) of HG 21. This haplogroup, which is a subset of YAP groups 3 and 4 (Hammer et al. 1997) and which is equal to the YAP groups later called “3A and 4A” (Altheide and Hammer 1997), has previously been found in Europe (14%), Egypt (47%), and sub-Saharan Africa (12%), but it has never been found at such high frequencies (Altheide and Hammer 1997; Hammer et al. 1997). No Y chromosome belonging to HG 4 (which is characterized by the fact that it carries the ancestral allele of SRY-8299 with respect to HG 21) was found; however, this haplogroup, which, according to the nomenclature used by Altheide and Hammer (1997), is equivalent to YAP haplotype 3G, has so far been found to be restricted to central and east Asian pop-

ulations (Hammer et al. 1998). Six (4.7%) of the 129 Y chromosomes analyzed contained the A→G transition that defines HG 8, which has been found at high frequencies in sub-Saharan African populations, in 2% of Egyptians, and in 5% of west Asians. This transition has not been found in 983 chromosomes from Europe, the rest of Asia, Australasia, and the Americas (Seielstad et al. 1994; Hammer et al. 1997). A sub-Saharan African origin for this haplogroup was suggested (Hammer et al. 1998), and, thus, its presence in northwestern Africa may indicate sub-Saharan admixture in northwestern Africa. No other sub-Saharan-specific haplogroups—such as HG 6, which is present in Pygmies and in San (Jobling 1994; Jobling et al. 1997), or HG 7 (Jobling et al. 1997)—were found in our samples. HG 9, seen here with a frequency of 11.6%, has a Mediterranean distribution, with its highest frequencies occurring in the Near East (Casanova et al. 1985; Brega et al. 1987; Mitchell et al. 1993, 1997; Semino et al. 1995). This pattern was interpreted as being produced either by such colonizing seafaring peoples as the Phoenicians (Mitchell and Hammer 1996; Mitchell et al. 1997) or by neolithic farmers (Gonçalves and Lavinha 1994). On the other hand, HG 1 (3.9%) has been found predominantly in Europe (Mathias et al. 1994; Jobling et al. 1997; Mitchell et al. 1997), and HG 2 (2.3%), in Europe and Asia (Jobling and Tyler-Smith 1995; Jobling et al. 1997). HG 26 was found at a frequency of .8%. No Y chromosome was found to belong either to HG 3, which is present in European populations (Jobling et al. 1997), or to HG 15, which seems to be specific to Indian populations (Pandya et al., in press).

#### Y-Chromosome STR Polymorphisms

Allele-frequency distributions for eight Y-chromosome STRs are given in the last column (designated as “Overall”) of table 3; gene diversities and allele-size variances can also be found in the same table. Of a total of 129 complete haplotypes constructed, considering seven of the Y-chromosome STR polymorphisms studied, 56 distinct Y-chromosome-haplotype configurations were obtained. The most frequent Y-chromosome haplotype, 13-12-11-24-9-11-13 (DYS19-DYS388-DYS389I-DYS390-DYS391-DYS392-DYS393), was found in 33 individuals, whereas 42 haplotypes were observed in unique copies. Haplotype diversity was estimated at  $.93 \pm .02$ .

#### Y-Chromosome STR Polymorphism within Haplogroups

Y-chromosome STR allele frequency distributions, by haplogroup, are shown in table 3. Clearly, allelic variation at each STR locus shows striking differences among haplogroups (see also table 4, in which the

**Table 3**

**STR Allele Frequencies by Haplogroup and by Total in Northwestern Africa**

Marker and Allele	HG 9 (n = 15)	HG 1 (n = 5)	HG 26 (n = 1)	HG 2+ (n = 3)	HG 21 (n = 99)	HG 8 (n = 6)	Overall (n = 129)
DYS19: <sup>a</sup>							
13	0	0	0	0	.889	0	.681
14	1	.400	1	.667	.071	0	.209
15	0	.200	0	0	.020	.500	.047
16	0	.400	0	0	.010	.500	.047
17	0	0	0	.333	.010	0	.016
DYS388: <sup>b</sup>							
12	0	1	1	0	.960	1	.827
13	0	0	0	.333	.040	0	.039
14	.067	0	0	0	0	0	.008
15	.067	0	0	.333	0	0	.016
16	.133	0	0	0	0	0	.016
17	.600	0	0	.333	0	0	.078
18	.133	0	0	0	0	0	.016
DYS389I: <sup>c</sup>							
9	.067	.200	1	.333	.091	.666	.132
10	.800	.600	0	.667	.172	.167	.271
11	.133	.200	0	0	.717	.167	.581
12	0	0	0	0	.010	0	.008
13	0	0	0	0	.010	0	.008
DYS389II: <sup>d</sup>							
24	0	0	1	0	0	0	.008
25	0	.200	0	.333	0	0	.016
26	.200	.400	0	.667	.152	.167	.178
27	.667	.400	0	0	.626	.500	.596
28	0	0	0	0	.202	.333	.171
29	.133	0	0	0	.010	0	.023
30	0	0	0	0	.010	0	.008
DYS390: <sup>e</sup>							
21	0	0	0	0	0	1	.047
22	0	0	0	0	.040	0	.031
23	.800	0	1	1	.172	0	.256
24	.133	.600	0	0	.606	0	.503
25	.067	.400	0	0	.182	0	.163
DYS391: <sup>f</sup>							
8	0	0	0	0	.010	0	.008
9	0	0	0	0	.778	0	.597
10	.200	.200	1	.667	.182	1	.240
11	.733	.800	0	.333	.030	0	.147
12	.067	0	0	0	0	0	.008
DYS392: <sup>g</sup>							
10	0	0	0	0	.010	0	.008
11	1	0	1	1	.909	1	.891
12	0	0	0	0	.081	0	.062
13	0	1	0	0	0	0	.039
DYS393: <sup>h</sup>							
10	0	0	0	0	.010	0	.008
11	0	0	0	0	0	0	0
12	1	0	0	.667	0	0	.132
13	0	.800	0	.333	.970	.333	.798
14	0	.200	1	0	.020	.667	.062

NOTE.—Both gene diversity (D) and allele-size variance (V) were computed from the total sample.

<sup>a</sup> D = .487; V = .814.

<sup>b</sup> D = .304; V = 2.559.

<sup>c</sup> D = .573; V = .580.

<sup>d</sup> D = .584; V = .643.

<sup>e</sup> D = .654; V = .881.

<sup>f</sup> D = .566; V = .609.

<sup>g</sup> D = .200; V = .209.

<sup>h</sup> D = .342; V = .25.

**Table 4**  
**Most-Frequent STR Alleles within Each Haplogroup Found in Northwestern Africa**

HAPLOGROUP	STR ALLELES <sup>a</sup>						
	DYS19	DYS388	DYS389I	DYS390	DYS391	DYS392	DYS393
HG 9 (n = 15)	<u>14</u> (100%)	17 (60%)	10 (80%)	23 (80%)	11 (73%)	<u>11</u> (100%)	<u>12</u> (100%)
HG 1 (n = 5)	14/16 (40%)	<u>12</u> (100%)	9 (60%)	24 (60%)	11 (80%)	<u>13</u> (100%)	<u>13</u> (80%)
HG 6 (n = 1)	14	<u>12</u>	9	23	10	<u>11</u>	14
HG 2+ (n = 3)	14 (67%)	13-15-17 (33%)	10 (67%)	<u>23</u> (100%)	10 (67%)	<u>11</u> (100%)	12 (67%)
HG 21 (n = 99)	13 (89%)	12 (96%)	11 (72%)	24 (61%)	9 (78%)	<u>11</u> (91%)	13 (97%)
HG S8 (n = 6)	15-16 (50%)	<u>12</u> (100%)	9 (67%)	<u>21</u> (100%)	<u>10</u> (100%)	<u>11</u> (100%)	14 (67%)

<sup>a</sup> Unique alleles are underlined.

most frequent alleles are shown for each haplogroup). Given the variability of a particular STR locus, some haplogroups display several alleles, but, in others, the number of alleles is highly restricted and differentiated. When variation at locus DYS390 is taken as an example, it can easily be seen that allele 21 is present exclusively in HG 8 (where it is the only allele observed), that allele 23 is the only allele found in HG 2+ and is the most frequent allele in HG 9, and that several alleles at intermediate frequencies are present in HG 21 and HG 1. Even when the small number of individuals in some of the haplogroups is taken into account, this obvious microsatellite differentiation among the haplogroups seems to indicate that Y-chromosome genetic variation is strongly structured by haplogroup background. Of the 56 STR-distinct haplotypes found, only one was shared by two different haplogroup backgrounds (HG 2+ and HG 9). The apportionment of Y-chromosome STR-haplotype diversity among and within haplogroups was assessed by AMOVA (table 5): 83.5% ( $P < .0001$ ) of the total genetic variation was attributable to differences between haplogroups. Compared with between-population differentiation, this is an extremely high value. Among the four linguistic subpopulations represented in our sample, the proportion of STR genetic variance explained by between-population difference was 3.72% ( $P = .0088$ ). Of the genetic variability for Y-

chromosome STRs, 3.5% could be apportioned to between-population differences among four European samples (de Knijff et al. 1997), and that fraction was 20.6% among four central Asian samples (Pérez-Lezaun et al. 1999). With the use of 10 Y-chromosome STRs, the apportionment of diversity between three African groups of populations was estimated to be 2.52% (Seielstad et al. 1998).

A number of diversity parameters were computed for each haplogroup, to characterize more extensively the Y-chromosome STR allelic variation within them (see table 6). HG 2+, despite having a low frequency, was found to be the most diverse, as is shown by the average gene diversity, the mean number of different alleles, the mean number of differences in repeat size, and the mean repeat-size variance. However, given the small sample size for HG 2+, the difference, in those parameters, between HG 2+ and any other haplogroup was statistically significant only for the difference in repeat length among haplotypes against HG 21 ( $P = .042$ , by permutation test). As discussed below, HG 2+ is probably the oldest haplogroup among those found in this study, and, thus, it may contain distinct, old lineages with heterogeneous STR-haplotype variation. On the other hand, HG 1 seems to occupy an intermediate position between HG 2+ and the rest of haplogroups, which were found to be more compact. Again, these differences were not statisti-

**Table 5**  
**Percent Fractions of the Genetic Variation That Can Be Attributed to Differences among Haplogroups and to Differences among Populations in Two Different Geographic Areas**

DIFFERENCE IN GENETIC VARIATION	PERCENT FRACTIONS <sup>a</sup> OF GENETIC VARIATION IN TWO DIFFERENT GEOGRAPHIC AREAS STUDIED		
	Northwestern Africa	Iberian Peninsula	Northwestern Africa and the Iberian Peninsula
Among haplogroups	83.5 ( $P < .0001$ )	23.4 ( $P = .0029$ )	66.4 ( $P < .0001$ )
Among populations	3.7 ( $P = .0088$ )	2.2 ( $P = .0890$ )	19.2 ( $P < .0001$ )

<sup>a</sup> As measured by AMOVA.



**Table 6**

**Diversity Parameters in Northwestern Africa by Haplogroup, Except for HG 26, Which Was Found in Only One Individual**

Diversity Parameters	HG 9 (n = 15)	HG 1 (n = 5)	HG 2+ (n = 3)	HG 21 (n = 99)	HG 8 (n = 6)
No. of polymorphic sites <sup>a</sup>	4	5	5	7	3
No. of different STR haplotypes	9	5	3	33	5
Haplotype diversity <sup>b</sup>	.80 ± .11	1.00 ± .13	1.00 ± .27	.88 ± .03	.93 ± .12
Mean gene diversity <sup>c</sup>	.26 ± .18	.41 ± .30	.52 ± .45	.27 ± .17	.25 ± .19
Mean no. of allele differences <sup>d</sup>	1.81 ± 1.10	2.90 ± 1.82	3.67 ± 2.52	1.90 ± 1.10	1.73 ± 1.17
Mean no. of differences in repeat size <sup>e</sup>	2.38 ± 1.85	3.40 ± 3.06	6.67 ± 4.24	2.44 (2.31) <sup>f</sup> ± 1.85	2.00 ± 1.97

<sup>a</sup> No. of STR loci found to be variable within each haplogroup in this sample.

<sup>b</sup> Gene diversity computed from STR haplotype frequencies.

<sup>c</sup> Mean gene diversity at each STR locus within each haplogroup.

<sup>d</sup> Pairwise average of the number of different STR alleles for all pairs of chromosomes within each haplogroup.

<sup>e</sup> Pairwise average of the cumulative absolute difference in STR allele length for all pairs of chromosomes within each haplogroup.

<sup>f</sup> Without large poly (A) tail individual.

cally significant. In table 6, for HG 21, the number shown in parentheses corresponds to the mean number of differences in repeat size obtained without taking into account the individual with the large poly (A) tail, who seems to contribute disproportionately to the parameter. In fact, this individual would be classified as having YAP haplotype 3A, according to Hammer et al. (1998), whereas all other individuals in HG 21 would be classified as having YAP haplotype 4. This individual may belong to a related, but different, evolutionary branch in the Y-chromosome genealogy.

For the STR variability generated within the chromosomes bearing derived states in the biallelic markers that characterize the haplogroups found in northwestern Africa, TMRCA were estimated from the mean allele length variance of the seven STRs considered (table 7). Mean allele length variances ranged from .181 to .844. The most ancient TMRCA, estimated at 14,067 years ago (ya) (95% CI 757–68,782 ya), was for SRY-1532 A→G, which is found in all 129 northwestern African Y chromosomes analyzed. The TMRCA for STR variability within substitution SRY-8299 G→A, which defines haplogroups 21 and 8, was estimated to be 6,483 ya (95% CI 493–30,782 ya). The TMRCA for mutation sY81 A→G, which is found solely on chromosomes bearing allele SRY-8299 A and which defines HG 8, was estimated to be younger (3,017 ya, with a 95% CI 0–18,565 ya), and that for the 8-kb allele at the 12f2 polymorphism was estimated to be 4,583 ya (95% CI 0–27,609 ya). Finally, the TMRCA for mutations M9 C→G and 92R7 C→T were found to be ~7,867 ya (95% CI 1,264–33,347 ya) and 5,233 ya (95% CI 0–27,696 ya), respectively.

The mean repeat-number size difference was also computed for pairs of haplogroups, and it was found to be much larger than the within-haplogroup means (see the numbers below the diagonal in table 8). With the use

of a stepwise mutation model, the difference in repeat size between a pair of alleles is expected to grow with time (Goldstein et al. 1995a, 1995b). The mean number of differences between pairs of haplogroups was converted to genetic distances (which are indicated by numbers above the diagonal in table 8), by means of the Jensen Difference (Rao 1982), which takes into account variability within each haplogroup. Thus, a measure of differentiation between haplogroups, which is closely related to  $D_{sw}$  (Shriver et al. 1995), a measure of genetic distance for independent STRs, was obtained. HG 26 was not included either in the calculation of the mean repeat-number size difference for pairs of haplogroups or in the reconstruction of the haplogroup genealogy, since it contained only one chromosome. In accordance with AMOVA results, the average difference in repeat lengths between pairs of haplotypes belonging to different haplogroups was always larger than the average differences within the haplogroups being compared. With a few exceptions (e.g., the average difference between HG 1 and HG 2+ compared with the internal differences in either haplogroup or the average difference between HG 2+ and HG 9 compared with the internal average difference in HG 2+), this pattern was statistically significant by permutation test ( $P$  values .039–.001). The largest distances were found between HG 8 and HG 9 (9.68) and between HG 1 and HG 9 (8.10), a finding that is in accordance with the parsimony network (see fig. 1), where these haplogroups occupy peripheral positions. On the contrary, the shortest distance was found between HG 2+ and HG 9, which are separated by only one unique-mutation event. The discordant genetic distances obtained between HG 8 and HG 21 (a large genetic distance for a single point mutation) and between HG 8 and HG 2+ (a rather low value for haplogroups three mutational events apart), in relation to those that were expected, could be attributed

Table 7

## TMRCA Computed from STR Allele-Size Variance in Northwestern Africa

	MUTATION					
	SRY-1532 (A→G)	12f2 (10Kb→8Kb)	M9 (C→G)	92R7 (C→T)	SRY-8299 (G→A)	sY81 (A→G)
Haplogroups containing the derived allele	All ( <i>n</i> = 129)	9 ( <i>n</i> = 15)	1 + 26 ( <i>n</i> = 6)	1 ( <i>n</i> = 5)	21 + 8 ( <i>n</i> = 105)	8 ( <i>n</i> = 6)
Mean STR allele-length variance <sup>a</sup>	.844	.275	.472	.314	.389	.181
Expansion time (ya) <sup>b</sup>	14,067	4,583	7,867	5,233	6,483	3,017
(95% CI) <sup>c</sup>	757–68,782	0–27,609	1,264–33,347	0–27,696	493–30,782	0–18,565

<sup>a</sup> Mean STR allele length variance is the average over STRs of allele length variance.

<sup>b</sup> Times estimates were computed on the basis of equation 5 in Di Rienzo et al. (1998; see also the Material and Methods section).

<sup>c</sup> CI was computed by taking into account variance interlocus sampling error and mutation rate estimate error.

to the small sample size of HG 8 (*n* = 6) and HG 2+ (*n* = 3), to the high internal diversity within HG 2+, or to stochastic processes in the generation of STR variation within haplogroups, which may have played an important role, as discussed below. A minimum spanning tree constructed from the distance matrix showed all other haplogroups stemming from HG 2+, which matches the known haplogroup genealogy (fig. 1), with the exception of the position of HG 8, which derives from HG 21. The different levels of internal diversity within haplogroups could bias the genetic distance estimates among them (Charlesworth 1998); however, if the correction for internal diversity was not applied, the haplogroup phylogeny reconstructed had little resemblance to the actual known phylogeny.

## Replication Analysis

Given the small sample sizes in some of the haplogroups, we sought to confirm our results by adding Y-chromosome data for northern Iberians (51 Basques and 27 Catalans). The same STRs had been typed (Pérez-Lezaun et al. 1997) in those samples, and it was possible to allocate the chromosomes to the haplogroups defined by biallelic polymorphisms (Semino et al. 1996; Underhill et al. 1997; Hurles et al. 1999; P. Underhill, unpublished data). This sample had a different, partly overlapping haplogroup composition with respect to northwestern Africa: 55 (70.5%) chromosomes belonged to HG 1, eight (10.3%) to HG 2+, one (1.3%) to HG 9, two (2.6%) to HG 21, and 12 (15.4%) to HG 22, which has a geographic distribution that is almost restricted to northern Iberia and that is suggested to have sprung quite recently from HG 1 (Hurles et al. 1999).

AMOVA performed on STR-haplotype variability among haplogroups in northern Iberia showed that 23.4% (*P* = .0029) of the genetic variation could be apportioned among haplogroups. This result contrasts with the 83.5% result found among haplogroups in northwestern Africa (table 5). This may be the result of the presence, in northern Iberia, of HG 22, which is

suggested to have recently originated from HG 1 in northern Iberia (Hurles et al. 1999); in fact, all but one of the Y-chromosome STR haplotypes in HG 22 were found in HG 1. The fraction of Y-chromosome STR genetic variation among Basques and Catalans is 2.2%, which is not significantly different from zero (*P* = .0890) and which is still much smaller than the fraction of genetic variation among haplogroups. When we pooled the samples from northwestern Africa and northern Iberia, the fraction of genetic variation attributable to haplogroups was 66.4%, whereas it was 19.2% among populations (table 5). Thus, although the results in the pooled sample were not as extreme as those seen when samples from northern Africa alone were considered, it still holds that genetic variation is more deeply structured by lineage than by population.

If we pooled the samples from northwestern Africa and northern Iberia, we reduced the number of haplogroups with small sample sizes and achieved a greater statistical power. The pattern of genetic diversity within haplogroups was confirmed in the pooled sample: HG 2+ (*n* = 11) had the largest in-

Table 8

Genetic Distances among Haplogroups (with One Exception<sup>a</sup>) in Northwestern Africa

HAPLOGROUP	GENETIC DISTANCES AMONG HGs				
	1	2+	21	8	9
HG 1	<u>3.40</u>	4.77	4.44	5.9	8.1
HG 2+	9.80	<u>6.67</u>	4.40	4.56	.25
HG 21	7.29	8.95	<u>2.31</u>	6.13	7.76
HG 8	8.60	8.89	8.29	<u>2</u>	9.68
HG 9	10.99	4.78	10.1	11.87	<u>2.38</u>

NOTE.—Distances below the diagonal of underlined numbers represent the mean number of repeat differences among all pairs of STR haplotypes between two haplogroups. Distances along the diagonal of underlined numbers denote the mean number of repeat differences within haplogroups. Distances above the diagonal of underlined numbers denote genetic distances among hap

<sup>a</sup> HG 26, which was found in only one chromosome, was not included in the determination of genetic distances.

ternal diversity (average difference in repeat length among STR haplotypes 6.67), when compared with HG 1 ( $n = 60$ ; 3.22), HG 9 ( $n = 16$ ; 2.85), HG 21 ( $n = 101$ ; 2.42), HG 22 ( $n = 12$ ; 2.41), and HG 8 ( $n = 6$ ; 2.00). Using a permutation procedure, we tested whether this parameter was significantly different between pairs of haplogroups, and we found that it was significantly different between HG 2+ and any other haplogroup ( $P$  values between .0001 and .0112), and between HG 1 and HG 21 ( $P = .0021$ ).

Mutation ages were also estimated from the pooled sample. Although sample sizes were greatly increased for the sets of chromosomes bearing some mutations, mutation ages did not change noticeably. The age of SRY-1532 A→G was estimated, in the pooled sample ( $n = 207$ ), to be 13,550 ya (95% CI 2,279–56,826 ya), whereas its estimated age in northwestern Africa was 14,067 ya (table 7). The age of 12f2 10 kb→8 kb ( $n = 16$ ) was estimated to be 5,516 ya (95% CI 0–31, 130 ya). The age estimates for M9 C→G ( $n = 73$ ) and for 92R7 C→T ( $n = 72$ ) were 6,617 (95% CI 1,236–27,000) and 6,300 (95% CI 1,079–26,304) ya, respectively. On the YAP-positive branch, the age estimate for SRY-8299 ( $n = 107$ ) was 6,383 ya (95% CI 493–30,304 ya), and that for sY81 remains unchanged, since no northern Iberian Y chromosome was found in HG 8. The haplogroup phylogeny reconstructed from STR haplotypes was the same with the pooled samples as it was with northwestern African chromosomes alone, with the addition of the correct link between HG 1 and HG 22.

## Discussion

Although biallelic polymorphisms can be regarded as unique mutational events (or as events of very low probability) and can allow us to identify deep lineages of Y chromosomes (called, in the present study, “haplogroups”), STR polymorphisms exhibit a faster mutation rate (0.0012 per locus per generation) (Heyer et al. 1997; Kayser et al. 1997; Bianchi et al. 1998), producing highly informative markers for studies of recent evolutionary (or historical) events (Roewer et al. 1996; Pérez-Lezaun et al. 1997). As a result of their differential rate of evolution, the combination of the data obtained from both types of markers on the nonrecombining portion of the Y chromosome provides an interesting perspective on different levels of resolution in the phylogeny of Y chromosomes.

The six different haplogroups found in northwestern Africa may offer clues as to which groups of Y chromosomes contributed to the composition of the present-day population. As we have described, HG 21 is the major haplogroup that characterizes Y-chromosome diversity in northwestern Africa. Its high fre-

quency—the highest ever reported for this haplogroup—can be interpreted as the result of a specific founder effect or drift process in this geographic region. On the contrary, HG 8 and HG 9 may have been introduced by gene flow from sub-Saharan Africa and from the eastern Mediterranean basin, respectively (Jobling and Tyler-Smith 1995; Jobling et al. 1996, 1997).

## Compartmentalization of Genetic Variance

Since no recombination occurs between haplogroups, we can consider that they behave like independent units (or subpopulations) without migration. Therefore, it would be expected that the compartmentalization of the genetic information they carry would be complete. AMOVA showed that >80% of the genetic variance was found among haplogroups in northwestern Africa and that >65% was found in northwestern Africa plus northern Iberia; these levels of compartmentalization are very high, but they are not complete. There are two complementary explanations for this result.

First, the origins of haplogroups are not independent. Each haplogroup arose when a rare mutation occurred on a given Y chromosome that belonged to a certain haplogroup and that carried a microsatellite haplotype. Genetic variation in the new haplogroup was reset to zero, but the founding haplotype was either similar or altogether identical to other haplotypes in the parental haplogroups. Thus, immediately after the origin of a haplogroup, its genetic background was likely to be closely related to that of the parental haplogroup, until mutation and drift differentiated them. Sometimes, it may be possible to identify this founder haplotype from the most common haplotype of the new haplogroup. The signal will generally decrease with time. This is clearly exemplified by HG 22 chromosomes in northern Iberia, which contain STR haplotypes that are also found in its parental HG 1.

Second, the recurrent nature of microsatellite mutation implies that haplotypes that are identical by state may not be identical by descent. In the same way that microsatellite haplotypes in each haplogroup will differentiate with time from the common ancestor from which they derive, haplotypes belonging to different haplogroups could occasionally converge with time as well. In fact, in our northwestern African sample, two Y chromosomes belonging to HG 2+ and HG 9 shared an STR haplotype.

We have analyzed 11 biallelic polymorphisms that define haplogroups with known phylogeny (Jobling and Tyler-Smith 1995; Jobling et al. 1996, 1997; Hurles et al. 1998, 1999) and with only one instance of back-mutation (SRY-1532; see fig. 1), which does not affect the present analysis, since neither HG 3 nor HG 7 has

been found. On that well-defined background, we have considered seven tri- and tetranucleotide STRs, and, as noted above, we have observed an almost complete compartmentalization. Malaspina et al. (1998) typed four dinucleotide STRs and found that 36 of 179 STR haplotypes were shared among the four haplogroups (“frames” in Malaspina et al. [1998]) defined by two biallelic polymorphisms. The much smaller number of STR haplotypes shared among haplogroups in our study may reflect the larger number of biallelic and STR loci typed.

It could be argued that, if more polymorphic sites were included and if the haplogroups were further subdivided, the apportionment of genetic variation among haplogroups would be lower. Preliminary results (E. Bosch, unpublished data) show that our sample of HG 21 Y chromosomes from northwestern Africa can be subdivided by typing three additional biallelic polymorphisms, thereby resulting in three different subhaplogroups. When AMOVA was repeated in the whole set of Y chromosomes from northwestern Africa, and when the three subhaplogroups of HG 21 were considered, the fraction of genetic variation among haplogroups was 84.3%, which is very similar to (and is, in fact, slightly higher than) that found without splitting HG 21.

We have also shown that genomic and population perspectives on STR genetic variation give very different results. When we defined groups of chromosomes according to the biallelic variants they carry, the fraction of STR genetic variation found among them was much higher than that found among groups of chromosomes defined by their population origin. This may happen if the origin of the populations (the ethnogenesis process) is more recent than that of most of the biallelic polymorphisms. The compartmentalization of STR genetic variation is expected to be tighter for haplogroups than it is for populations, since a haplogroup goes through a strict bottleneck with a size of one chromosome, whereas population bottlenecks have not been so narrow, and since there is gene flow between populations but there is no gene exchange between the nonrecombining portion of the Y chromosome. In summary, genetic background determines the STR genetic variation to a much greater extent than does the population background.

#### *Variance in Repeat Number: STR Variation and Mutation Age*

The analysis of STR repeat-size variance within lineages of Y chromosomes bearing a particular derived allele at a biallelic locus has been used to estimate TMRCA (table 7). That is, we have tried to estimate the time to which the observed STR variation within each Y-chromosome lineage coalesces. As we will dis-

cuss, this time is expected to be correlated to the actual age of the mutation that defines each lineage. Thus, we found both an old mutation (SYR-1532), which was borne by all the chromosomes in our sample, and five younger mutations. These estimated TMRCA (see table 7) match the known haplogroup genealogy (fig. 1), in which SRY-8299 precedes sY81 and M9 precedes 92R7.

Hammer et al. (1998) typed a worldwide sample of Y chromosomes for a set of biallelic markers that partly overlaps with our set. The authors estimated mutation ages by use of coalescence analysis (Griffiths and Tavaré 1994), which is done on the basis of both the shape of the gene tree and the number of chromosomes bearing each combination of mutations. The ages for those biallelic mutations that overlap with our study were 110,000 ± 36,000 ya for the polymorphic site at position 10,831.1 of the SRY region (synonymous to SRY-1532), ~31,000 ya for the polymorphic site at position 4,064 of the SRY region (synonymous to SRY-8299), ~30,000 ya for DYS257 (which seems to have happened in the same phylogenetic branch as 92R7, as they appear to be phylogenetically equivalent; Jobling et al. 1998b), and ~11,000 ya for PN1 (which cosegregates with sY81; Hammer et al. [1997]). All of them are clearly older than our estimates for the TMRCA through STR variation, although both age sets are correlated and have broad, overlapping 95% CIs. The largest discrepancy is found for SRY-1532, a deep mutation in the phylogeny, from which many different haplogroups—most of which have not been found in this study—derive. Since most (>80%) of the chromosomes in our sample belong to a single recent branch of the genealogy (that bearing SRY-8299 and sY81) and since other branches bearing SRY-1532 are found at low frequency or are altogether absent, this could lead to an underestimation of the STR allele length variance within SRY-1532 and, therefore, of its TMRCA. However, when we added Y chromosomes from northern Iberia, which contributed other branches derived from SRY-1532, the age estimate through STR variation did not increase. In a study of Y chromosomes from Polynesia and Melanesia (Hurles et al. 1998), the authors also refrained from dating the TMRCA for SRY-1532 from STR data, on the basis that mutation/drift equilibrium renders such dating methods unable to resolve suitably far back in time. Apart from this particular case, several reasons could account for the general lag between our TMRCA estimates through STR variation and mutation ages obtained by coalescence analysis in Hammer et al. (1998):

1. *Mutation Age Precedes TMRCA by Definition.* We have estimated the time to the current observed STR variation. The observed STR haplotypes may coalesce to a time that is more recent than the actual mutation age, because of the extinction of lineages that appeared

in the first generations after the mutation arose and that we are not able to detect. Moreover, the variance in repeat size, which we have used to estimate the TMRCA, is zero until the first mutation in a STR. The method we used (Di Rienzo et al. 1998) dates, in fact, an expansion that should have happened, obviously, at some unknown time after the mutation arose. The distributions of pairwise differences in repeat number found within the most frequent haplogroups (not shown) are smooth and bell-shaped, as is expected for a lineage that underwent an expansion (Shriver et al. 1997). This justifies the use of the equation given by Di Rienzo et al. (1998). Coalescence prior to the actual mutation, the onset of STR mutations, and the expansion detected by the method suggested in Di Rienzo et al. (1998) all may contribute to the lag between mutation age and our estimate of TMRCA.

2. *Population Biases.* We have northwestern African Y chromosomes, which contain a particular set of haplogroups as a result of their population history (founder events, gene flow, and admixture), and this fact could lead to an underestimation of haplogroup variability and mutation ages. To test for this possible bias, we used coalescence analysis on northwestern Africa haplogroup phylogeny and frequencies, according to the same methods and parameters used in Hammer et al. (1998), and we were able to reproduce their mutation age estimates, with minor discrepancies. Another test for this possible bias results from the addition of chromosomes from other populations. When we added two population samples from northern Iberia, the age estimates through STR variation of the mutations present in NW Africa did not change. Therefore, it does not seem that population bias plays a major role in the discrepant differences between mutation ages, as seen elsewhere (Hammer et al. (1998), and our STR-based TMRCA estimates.

3. *STR Saturation.* Given the fast mutation rate of STRs, it is possible that they have reached mutation-drift equilibrium and that, therefore, genetic variation in STRs would remain constant in time. However, if STRs have reached a complete saturation, we would not have observed a correlation between TMRCA and the relative mutation ages from the gene genealogy.

4. *STR Mutation Rate Overestimation.* If STR mutation rates had been overestimated by a factor of 5–6, our TMRCA and the age estimates of Hammer et al. (1998) would match. Note that mutation rates are not homogeneous across STR loci (Carvalho-Silva et al. 1999). Then, mutation in genealogies will tend to be observed preferentially in the STRs with the fastest mutation rates, and the mean mutation rate will be overestimated. A similar situation is found in the mtDNA control region, where the mutation rate estimates per nucleotide are roughly 20 times higher when derived from mother-child transmission studies than when es-

timated from phylogenetic comparisons with nonhuman primate sequences (Jazin et al. 1998).

In sum, the discrepancy between the mutation age estimates provided elsewhere (Hammer et al. [1998]) and our estimates of the TMRCA may be mainly the result of the actual lag between the mutation event and the lineage expansion we dated and of possible overestimation of STR mutation rate, which is impossible to check, given current knowledge.

#### *Haplogroup Phylogeny and STRs*

In the two previous sections of this Discussion, we have seen that STR haplotypes in a haplogroup can preserve an amount of phylogenetic information about the parental haplogroup that declines with time and that the relative time of origin of a haplogroup can be estimated from STR allele-size variance. We can combine both types of information, reconstruct a haplogroup phylogeny, and compare it with the actual phylogeny constructed from the biallelic polymorphisms that define the haplogroups. The results of this exercise have shown that STRs within each haplogroup do trace most of the haplogroup phylogeny. In spite of the small sample sizes of some of the haplogroups, the phylogeny obtained from the STRs matches the known haplogroup phylogeny, with only one discrepancy: the position of HG 8, which stems from HG 2+ in our reconstruction but which is actually derived from HG 21. A similar analysis was performed (Rocha et al. 1997) with the allele frequencies of an STR located at the 5' end of the autosomal  $\alpha$ 1-antitrypsin gene within the electrophoretic variants of protein  $\alpha$ 1-antitrypsin.

We have shown that genetic background prevails over population background in the determination of STR genetic variation on the human Y chromosome. We have also used STR variation within lineages to infer the TMRCA and have shown that the genetic differentiation that is left among haplogroups contains still phylogenetic information but that the overall STR variation is mainly driven by the haplogroup composition of a specific population. The genetic composition of humans may thus be better understood in terms of evolutionarily related genealogical units rather than in demographic terms.

#### **Acknowledgments**

We express our appreciation to the original DNA donors who made this study possible. We especially thank Arpita Pandya for her help and advice concerning the typing of some of the biallelic polymorphisms. The warm reception E.B. received from Cancer Research Campaign Chromosome Molecular Biology Group staff during her stay in the Department of Biochemistry, Oxford University, is also acknowledged with grat-

itude. We appreciate the technical assistance offered by the Unitat de Seqüenciació, Servei Científic-Tècnic, Universitat de Barcelona. We also thank Peter de Knijff, for providing us with allelic ladders, and Peter A. Underhill, for typing YAP-negative individuals for M9 and for sharing with us unpublished individual data on Basques and Catalans. Two anonymous reviewers made suggestions that improved significantly this manuscript. This research was supported by Dirección General de Investigación Científica y Técnica in Spain (PB95-0267-CO2-01), by Direcció General de Recerca, Generalitat de Catalunya (1998SGR00009), and by Institut d'Estudis Catalans. Comissionat per a Universitats i Recerca, Generalitat de Catalunya supported E.B. (FI/96-1153); F.R.S. was supported by the Leverhulme Trust; and C.T.-S. was supported by the Cancer Research Campaign.

## Electronic-Database Information

URLs for data in this article are as follows:

CEPH, <http://www.cephb.fr/> (for data from CEPH-family cell lines)

Genome Database, The, <http://gdbwww.gdb.org/> (for primers for the DYS389 locus)

Arlequin: A Software for Population Genetic Data Analysis, <http://anthropologie.unige.ch/arlequin/>

## References

- Altheide TK, Hammer MF (1997) Evidence for a possible Asian origin of YAP+ Y chromosomes. *Am J Hum Genet* 61:462–466
- Banchs I, Bosch A, Guimerà J, Lázaro C, Puig A, Estivill X (1994) New alleles at microsatellite loci in CEPH families mainly arise from somatic mutations in the lymphoblastoid cell lines. *Hum Mutat* 3:365–372
- Bertranpetit J, Sala J, Calafell F, Underhill PA, Moral P, Comas D (1995) Human mitochondrial DNA variation and the origin of Basques. *Ann Hum Genet* 59:63–81
- Bianchi NO, Bailliet G, Bravi CM, Carnese RF, Rothhammer F, Martínez-Marignac VL, Pena SDJ (1997) Origin of Amerindian Y-chromosomes as inferred by the analysis of six polymorphic markers. *Am J Phys Anthropol* 102:79–89
- Bianchi NO, Catanesi CI, Bailliet G, Martínez-Marignac VL, Bravi CM, Vidal-Rioja LB, Herrera RJ, et al (1998) Characterization of ancestral and derived Y-chromosome haplotypes of New World native populations. *Am J Hum Genet* 63:1862–1871
- Bouzekri N, Taylor PG, Hammer MF, Jobling MA (1998) Novel mutation processes in the evolution of a haploid minisatellite, MSY1: array homogenization without homogenization. *Hum Mol Genet* 7:655–659
- Brega A, Torroni A, Semino O, Maccioni L, Casanova M, Scozzari R, Fellous M, et al (1987) The p12f2/*TaqI* Y-specific polymorphism in three groups of Italians and in a sample of Senegalese. *Gene Geogr* 1:201–206
- Carvalho-Silva DR, Santos FR, Hutz MH, Salzano FM, Pena SD (1999) Divergent human Y-chromosome microsatellite evolution rates. *J Mol Evol* 49:204–214
- Casanova M, Leroy P, Boucekkine C, Weissenbach J, Bishop C, Fellous M, Purrello M, et al (1985) A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance. *Science* 230:1403–1406
- Charlesworth B (1998) Measures of divergence between populations and the effect of forces that reduce variability. *Mol Biol Evol* 15:538–543
- Cooper G, Amos W, Hoffman D, Rubinsztein DC (1996) Network analysis of human Y microsatellite haplotypes. *Hum Mol Genet* 5:1759–1766
- Deka R, Jin L, Shriver MD, Yu LM, Saha N, Barrantes R, Chakraborty R, et al (1996) Dispersion of human Y chromosome haplotypes based on five microsatellites in global populations. *Genome Res* 6:1177–1184
- de Knijff P, Kayser M, Caglia A, Corach D, Fretwell N, Gehrige C, Graziosi G, et al (1997) Chromosome Y microsatellites: population genetic and evolutionary aspects. *Int J Legal Med* 110:134–140
- Di Rienzo A, Donnelly P, Toomajian C, Sisk B, Hill A, Petzl-Erler ML, Haines GK, et al (1998) Heterogeneity of microsatellite mutations within and between loci, and implications for human demographic histories. *Genetics* 148:1269–1284
- Estivill X, Morral N, Bertranpetit J (1994) Reply to Kaplan et al. *Nat Genet* 8:216–218
- Feinberg AP, Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem* 132:6–13
- (1984) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem* 137:266–267
- Foster EA, Jobling MA, Taylor PG, Donnelly P, de Knijff P, Mieremet R, Zerjal T, et al (1998) Jefferson fathered slave's last child. *Nature* 396:27–28
- Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995a) An evaluation of genetic distances for use with microsatellite loci. *Genetics* 139:463–471
- (1995b) Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc Natl Acad Sci USA* 92:6723–6727
- Gonçalves J, Lavinha J (1994) The Y-associated XY275G (low) allele is common among the Portuguese. *Am J Hum Genet* 55:583–584
- Graven L, Passarino G, Semino O, Boursot P, Santachiara-Benerecetti S, Langaney A, Excoffier L (1995) Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. *Mol Biol Evol* 12:334–345
- Griffiths RC, Tavaré S (1994) Ancestral inference in population genetics. *Stat Sci* 9:307–319
- Guellaen G, Casanova M, Bishop C, Geldwerth D, Andre G, Fellous M, Weissenbach J (1984) Human XX males with Y single-copy DNA fragments. *Nature* 307:172–173
- 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* 11:749–761
- (1995) A recent common ancestry for human Y chromosomes. *Nature* 378:376–378
- Hammer MF, Horai S (1995) Y chromosomal DNA variation and the peopling of Japan. *Am J Hum Genet* 56:951–962
- Hammer MF, Karafet T, Rasanayagam A, Wood ET, Altheide TK, Jenkins T, Griffiths RC, et al (1998) Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. *Mol Biol Evol* 15:427–441

- Hammer MF, Spurdle AB, Karafet T, Bonner MR, Wood ET, Novelletto A, Malaspina P, et al (1997) The geographic distribution of human Y chromosome variation. *Genetics* 145: 785–805
- Hammer MF, Zegura SL (1996) The role of the Y chromosome in human evolutionary studies. *Evol Anthropol* 5:116–134
- Heyer E, Puymirat J, Dieltjes P, Bakker E, de Knijff P (1997) Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. *Hum Mol Genet* 6:799–803
- Hurles ME, Irvén C, Nicholson J, Taylor PG, Santos FR, Loughlin J, Jobling MA, et al (1998) European Y-chromosomal lineages in Polynesians: a contrast to the population structure revealed by mtDNA. *Am J Hum Genet* 63:1793–1806
- Hurles ME, Veitia R, Arroyo E, Armenteros M, Bertranpetit J, Pérez-Lezaun E, Bosch E, et al (1999) Recent male-mediated gene flow over a linguistic barrier in Iberia suggested by analysis of a Y-chromosomal DNA polymorphism. *Am J Hum Genet* 65:1437–1448
- Jazin E, Soodyall H, Jalonon P, Lindholm P, Stoneking M, Gyllenstein U (1998) Mitochondrial mutation rate revisited: hot spots and polymorphism. *Nat Genet* 18:109–110
- Jobling MA (1994) A survey of long-range DNA polymorphisms on the human Y chromosome. *Hum Mol Genet* 3: 107–114
- Jobling MA, Bouzekri N, Taylor PG (1998a) Hypervariable digital DNA codes for human paternal lineages: MVR-PCR at the Y-specific minisatellite, MSY1 (DYF155S1). *Hum Mol Genet* 7:643–653
- Jobling MA, Pandya A, Tyler-Smith C (1997) The Y chromosome in forensic analysis and paternity testing. *Int J Legal Med* 110:118–124
- Jobling MA, Samara V, Pandya A, Fretwell N, Bernasconi B, Mitchell RJ, Gerelsaikhlan T, et al (1996) Recurrent duplication and deletion polymorphisms on the long arm of the Y chromosome in normal males. *Hum Mol Genet* 5:1767–1775
- Jobling MA, Tyler-Smith C (1995) Fathers and sons: the Y chromosome and human evolution. *Trends Genet* 11:449–456
- Jobling MA, Williams G, Schiebel K, Pandya A, McElreavey K, Salas L, Rappold GA, et al (1998b) A selective difference between human Y-chromosomal DNA haplotypes. *Curr Biol* 8:1391–1394
- Karafet T, Zegura SL, Vuturo-Brady J, Posukh O, Osipova L, Wiebe V, Romero F, et al (1997) Y chromosome markers and trans-Bering Strait dispersals. *Am J Phys Anthropol* 102:301–314
- Kayser M, Caglià A, Corach D, Fretwell N, Gehrig C, Graziosi G, Heidorn F, et al (1997) Evaluation of Y-chromosomal STRs: a multicenter study. *Int J Legal Med* 110:125–133
- Kwok C, Tyler-Smith C, Mendoza BB, Hughes I, Berkovitz GD, Goodfellow PN, Hawkins JR (1996) Mutation analysis of the 2Kb 5' to SRY in XY females and XY intersex subjects. *J Med Genet* 33:465–468
- Lucotte G, Ngo NY (1985) p49F, a highly polymorphic probe, that detects *TaqI* RFLPs on the human Y chromosome. *Nucleic Acids Res* 13:8285
- Malaspina P, Cruciani F, Ciminelli BM, Terrenato L, Santolamazza P, Alonso A, Banyko J, et al. (1998) Network analyses of Y-chromosomal types in Europe, northern Africa, and western Asia reveal specific patterns of geographic distribution. *Am J Hum Genet* 63:847–860
- Mateu E, Comas D, Calafell F, Pérez-Lezaun A, Abade A, Bertranpetit J (1997) A tale of two islands: population history and mitochondrial DNA sequence variation of Bioko and São Tomé, Gulf of Guinea. *Ann Hum Genet* 61:507–518
- Mathias N, Bayés M, Tyler-Smith C (1994) Highly informative compound haplotypes for the human Y chromosome. *Hum Mol Genet* 3:115–123
- Mitchell RJ, Earl L, Fricke B (1997) Y-chromosome specific alleles and haplotypes in European and Asians populations: linkage disequilibrium and geographic diversity. *Am J Phys Anthropol* 104:167–176
- Mitchell RJ, Earl L, Williams J (1993) Two Y-chromosome-specific restriction fragment length polymorphisms (DYS11 and DYZ8) in Italian and Greek migrants to Australia. *Hum Biol* 65:387–399
- Mitchell RJ, Hammer MF (1996) Human evolution and the Y chromosome. *Curr Opin Genet Dev* 6:737–742
- Pandya A, King TE, Santos FR, Taylor PG, Thangaraj K, Singh L, Jobling MA, et al
- Pena SDJ, Santos FR, Bianchi NO, Bravi CM, Carnese FR, Rothhammer F, Gerelsaikhlan T, et al (1995) A major founder Y-chromosome haplotype in Amerindians. *Nat Genet* 11: 15–16
- Pérez-Lezaun A, Calafell F, Comas D, Mateu E, Bosch E, Martínez-Arias R, Clarimón J, et al (1999) Gender-specific migration in Central Asian populations revealed by the analysis of Y-chromosome STRs and mtDNA. *Am J Hum Genet* 65: 208–219
- Pérez-Lezaun A, Calafell F, Seielstad MT, Mateu E, Comas D, Bosch E, Bertranpetit J (1997) Population genetics of Y chromosome short tandem repeats in humans. *J Mol Evol* 45: 265–270
- Pestoni C, Cal ML, Lareu MV, Rodríguez-Calvo MS, Carracedo A (1998) Y chromosome STR haplotypes: genetic and sequencing data of the Galician population (NW Spain). *Int J Legal Med* 112:15–21
- Rao CR (1982) Diversity and dissimilarity coefficients: a unified approach. *Theor Pop Biol* 21:24–43
- Rocha J, Pinto D, Santos MT, Amorim A, Amil-Dias J, Cardoso-Rodrigues F, Aguiar A (1997) Analysis of the allelic diversity of a (CA)<sub>n</sub> repeat polymorphism among  $\alpha$ -1-antitrypsin gene products from northern Portugal. *Hum Genet* 99:194–198
- Roewer L, Kayser M, Dieltjes P, Nagy M, Bakker E, Krawczak M, de Knijff P (1996) Analysis of molecular variance (AMOVA) of Y-chromosome-specific microsatellites in two closely related human populations. *Hum Mol Genet* 5:1029–1033
- Rolf B, Meyer E, Brinkmann B, de Knijff P (1998) Polymorphism at the tetranucleotide repeat locus DYS389 in 10 populations reveals strong geographical clustering. *Eur J Hum Genet* 6:583–588
- Santos FR, Carvalho-Silva DR, Pena SDJ (1999a) PCR-based DNA profiling of human Y chromosomes. In: Epplen JT, Lubjuhn T (eds) *Methods and tools in biosciences and medicine*. Birkhauser Verlag, Basel, Switzerland, pp 133–152
- Santos FR, Hutz M, Coimbra CEA, Santos RV, Salzano FM, Pena SDJ (1995a) Further evidence for the existence of a

- major founder Y chromosome haplotype in Amerindians. *Braz J Genet* 18:669–672
- Santos FR, Pandya A, Tyler-Smith C, Pena SDJ, Schanfield M, Leonard WR, Osipova L, et al (1999b) The central Siberian origin for Native Americans' Y chromosomes. *Am J Hum Genet* 64:619–628
- Santos FR, Pena SDJ, Tyler-Smith C (1995b) PCR haplotypes for the human Y chromosome based on alphoid satellite DNA variants and heteroduplex analysis. *Gene* 165:191–198
- Santos FR, Rodriguez-Delfin L, Pena SDJ, Moore J, Weiss KM (1996) North and South Amerindians may have the same major founder Y chromosome haplotype. *Am J Hum Genet* 58:1369–1370
- Schneider S, Kueffer J-M, Roessli D, Excoffier L (1997) Arlequin ver 1.1: a software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland
- Seielstad MT, Hebert JM, Lin AA, Underhill PA, Ibrahim M, Vollrath D, Cavalli-Sforza LL (1994) Construction of human Y-chromosomal haplotypes using a new polymorphic A to G transition. *Hum Mol Genet* 3:2159–2161
- Seielstad MT, Minch E, Cavalli-Sforza LL (1998) Genetic evidence for a higher female migration rate in humans. *Nat Genet* 20:278–280
- Semino O, Passarino G, Brega A, Fellous M, Santachiara-Benerecetti A (1996) A view of the neolithic demic diffusion in Europe through two Y chromosome-specific markers. *Am J Hum Genet* 59:964–968
- Semino O, Passarino G, Liu A, Brega A, Fellous M, Santachiara-Benerecetti AS (1995) Three Y-specific polymorphisms in populations of different ethnic and geographic origin. *Y Chromosome Newsletter* 2:5–6
- Shriver MD, Jin L, Boerwinkle E, Deka R, Ferrell RE, Chakraborty R (1995) A novel measure of genetic distance for highly polymorphic tandem repeat loci. *Mol Biol Evol* 12:914–920
- Shriver MD, Jin L, Ferrell RE, Deka R (1997) Microsatellite data support an early population expansion in Africa. *Genome Res* 7:586–591
- Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139:457–462
- Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 98:503–517
- Thomas MG, Skorecki K, Ben-Ami H, Parfitt T, Bradman N, Goldstein DB (1998) Origins of Old Testament priests. *Nature* 394:138–140
- Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, et al (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res* 7:996–1005
- Underhill P, Jin L, Zeman 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
- Vogt PH, Affara N, Davey P, Hammer M, Jobling MA, Lau YF-C, Mitchell M, et al (1997) Report of the Third International Workshop on Y Chromosome Mapping 1997: Heidelberg, Germany, April 13–16, 1997. *Cytogenet Cell Genet* 79:2–16
- Whitfield LS, Sulston JE, Goodfellow PN (1995) Sequence variation of the human Y chromosome. *Nature* 378:379–380
- Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos FR, Schiefenhövel W, et al (1997) Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. *Am J Hum Genet* 60:1174–1183