



## SHORT REPORT

# Further characteristics of proto-European Y chromosomes

Lluís Quintana-Murci<sup>1</sup>, Ornella Semino<sup>1</sup>, Eric Minch<sup>2</sup>, Giuseppe Passarimo<sup>3</sup>, Agnese Brega<sup>4</sup> and A Silvana Santachiara-Benerecetti<sup>1</sup>

<sup>1</sup>Dipartimento di Genetica e Microbiologia, Università di Pavia, Pavia, Italy

<sup>2</sup>Department of Genetics, Stanford University Medical Center, Stanford, USA

<sup>3</sup>Dipartimento di Biologia Cellulare, Università della Calabria, Cosenza, Italy

<sup>4</sup>Dipartimento di Biologia e Genetica per le Scienze Mediche, Università di Milano, Milano, Italy

**We examined a set of populations mainly from Europe but also from the Middle East and North Africa for the three Y-linked microsatellites YCAII, DYS19 (about 1300 individuals) and DYS392 (about 350 individuals). Three markers (YCAII a5-b1 Ht, DYS19-190 bp and DYS392-254 bp) show decreasing gradients of frequency from western Europe towards the Middle East which parallel that of the proto-European 49a,f/TaqI Ht 15. Indeed, a strong linkage disequilibrium between these markers and the 49a,f Ht 15 is observed. We therefore suggest that the 49a,f/TaqI Ht 15, YCAII a5-b1 Ht, DYS19-190 bp and DYS392-254 bp Y chromosome could represent a component of the proto-European gene pool. This European specific compound haplotype distinctively characterises western Europeans and its very high incidence in these populations (particularly in the Basques) is discussed.**

**Keywords:** proto-European Y chromosomes; European and Mediterranean populations; STR polymorphisms

The various components of the genetic structure of European populations have recently been debated in several papers.<sup>1–8</sup> The authors argue the relative palaeolithic and neolithic genetic contribution to modern Europeans and the geographic pattern that would have resulted from the various migrations occurring in Europe at different times. These topics have been approached by analysis of both classical and uniparental-inherited markers. The principal component (PC) analysis of classical markers<sup>9–11</sup> detected different population expansions associated with patterns of migrations of palaeolithic and neolithic peoples.

The mtDNA data, however, have given rise to disagreement on the importance of the different contributions to the European gene pool.<sup>3,5–7</sup> Recently Torroni *et al*<sup>8</sup> described a pre-neolithic European-specific mtDNA marker, the haplogroup V, whose homeland is most likely the northern Iberian peninsula/southern France. From this area, together with the older associated haplogroup H of Middle Eastern origin, it would have spread northeastward across Europe according to the second component of the classical markers PC analysis.<sup>10</sup> Thus, these pre-neolithic haplogroups would have substantially contributed to the current central and northern European maternal gene pool.

As to Y chromosome data, in addition to a neolithic component revealed by the 12f2/TaqI-8kb allele (with maximum values in the Middle East and a decreasing frequency gradient towards western Europe), they

Correspondence: AS Santachiara-Benerecetti, Dipartimento di Genetica e Microbiologia, Università di Pavia, Via Abbiategrasso, 207, 27100 Pavia, Italy. Tel: (39) 0382 505542; Fax: (39) 0382 528496; E-mail: [santa@ipvgen.unipv.it](mailto:santa@ipvgen.unipv.it)  
Received 17 September 1998; revised 18 December 1998; accepted 10 February 1999

highlighted an important proto-European component in European populations indicated by the Y-49a,f/*TaqI* haplotype (Ht) 15.<sup>2</sup> This European-specific haplotype displays a gradient of frequencies opposite to that shown by the neolithic 12f2-8 kb allele and maximum frequencies in peoples of western Europe, particularly in the Basques.<sup>2</sup> It is of interest to note that, based on anthropological, linguistic and archaeological evidence, this population is thought to be a relic of ancient palaeolithic and/or mesolithic European populations.<sup>10,12,13</sup>

The different relative contribution of the two sexes to the European gene geography has also been a matter of debate and, as Cavalli-Sforza and Minch<sup>3</sup> and Richards *et al*<sup>5</sup> have pointed out, further information on male-mediated genetic variation in Europe would be helpful at this stage. We have therefore examined a sample of the previously studied populations<sup>2</sup> for three Y-linked microsatellites in order better to identify the European male-specific features. More than 1300 individuals were analysed for the dinucleotide system YCAII<sup>14</sup> and the tetranucleotide DYS19,<sup>15</sup> and about 350 individuals for the trinucleotide DYS392,<sup>16</sup> following procedures reported by the authors who first described these polymorphisms.<sup>14-16</sup>

Three markers, the YCAII a5-b1 haplotype, the DYS19-190 bp allele and the DYS392-254 bp allele, have turned out to be very useful. Table 1 shows their frequencies in the Europeans examined and in some other populations of interest. Table 1 and the maps of the interpolated frequencies (Figure 1A-C) present all three markers showing their maximum values in peoples of western Europe and a decreasing gradient of frequencies from western to southeastern European populations. The only exceptions to the observed gradients are the unexpectedly high frequencies of YCAII Ht a5-b1 in Hungary (Figure 1A) and the DYS19-190 bp allele in the Middle East (Figure 1B), which are most probably due to recurrent mutations, as also indicated by their being in a different Y haplotype background in these areas (Passarino, G. personal communication 1998).

It is worth noting that, with the above exceptions, the observed gradients parallel well the distribution (Figure 1D) of the more stable<sup>2,26</sup> (Santachiara-Benerecetti A.S., unpublished data 1998) proto-European 49a,f Ht 15.<sup>2</sup> Indeed, a strong linkage disequilibrium between the 49a,f Ht 15 and the three Y microsatellite markers is observed: 81% of the 202 chromosomes 49a,f Ht 15 typed for the YCAII were a5-b1, 89% of the 191 typed

for the DYS19 were 190 bp, and 72% out of the 177 typed for both systems showed the association of the three Y markers. As to the DYS392 STR, 79% out of the 28 chromosomes 49a,f Ht 15 typed for this microsatellite were 254 bp.

We attempted to estimate the age of the Ht 15 Y chromosome variation by analysis of seven microsatellites (YCAIIa, YCAIIb, DYS19, DYS389-I, DYS390, DYS391 and DYS392)<sup>14-16</sup> in 28 Ht 15 Y chromosomes, following the methods of Bertranpetit and Calafell,<sup>27</sup> and Goldstein *et al*,<sup>28</sup> assuming generation spans of 20 years and, for Goldstein's method,<sup>28</sup> an effective population size of 5000 individuals. Different mutation rates ( $1.2 \times 10^{-4}$ ,  $4.8 \times 10^{-4}$ ,  $1 \times 10^{-3}$  and  $2.1 \times 10^{-3}$ )<sup>29-31</sup> were used, and coalescence times of 58 666, 14 668, 7040 and 3352 years (first method), and 51 080, 24 800, 11 186 and 5173 years (second method) were obtained. However, the highest mutation rate ( $2.1 \times 10^{-3}$ ) we used, which gives the most critical values in this context, produces some discrepancies. Although it should provide valid information on up to 40 000 to 50 000 years,<sup>31</sup> it gave a period for the pre-Columbian peopling of the Americas of 2147 BP years which is clearly an underestimate.<sup>32</sup>

Even though such uncertainty in the mutation rates does not allow us to assess safely the date of the origin of the Ht 15 variation, other evidence strongly supports an ancient origin of this haplotype. Its geographic distribution appears to be similar to that of other markers, such as the synthetic map of Europe based on the fifth principal component from nuclear data,<sup>10</sup> the pre-neolithic mtDNA haplogroup V,<sup>8</sup> and the cystic fibrosis  $\Delta$  F-508 allele,<sup>10,33,34</sup> which, it has been suggested precede the spread of agriculture in Europe.

These reasons lead us to consider, therefore, the European 49a,f Ht 15, YCAII a5-b1, DYS19-190 bp and DYS392-254 bp Y chromosome compound haplotype as a most likely component of an ancient proto-European gene pool. It distinctively characterises the populations of western Europe where it presumably originated and where the neolithic people arrived later (6-7000 years ago)<sup>10,35</sup>. Its very high incidence in these populations (particularly in the Basques) and its low level of genetic diversity most likely resulted from genetic drift which occurred during the last glacial period, when population sizes were small, and from subsequent population expansion at the end of that period when climatic conditions became more favourable. Basques, in their longer isolation, appear to have compounded the original genetic idiosyncrasies of western Europeans.

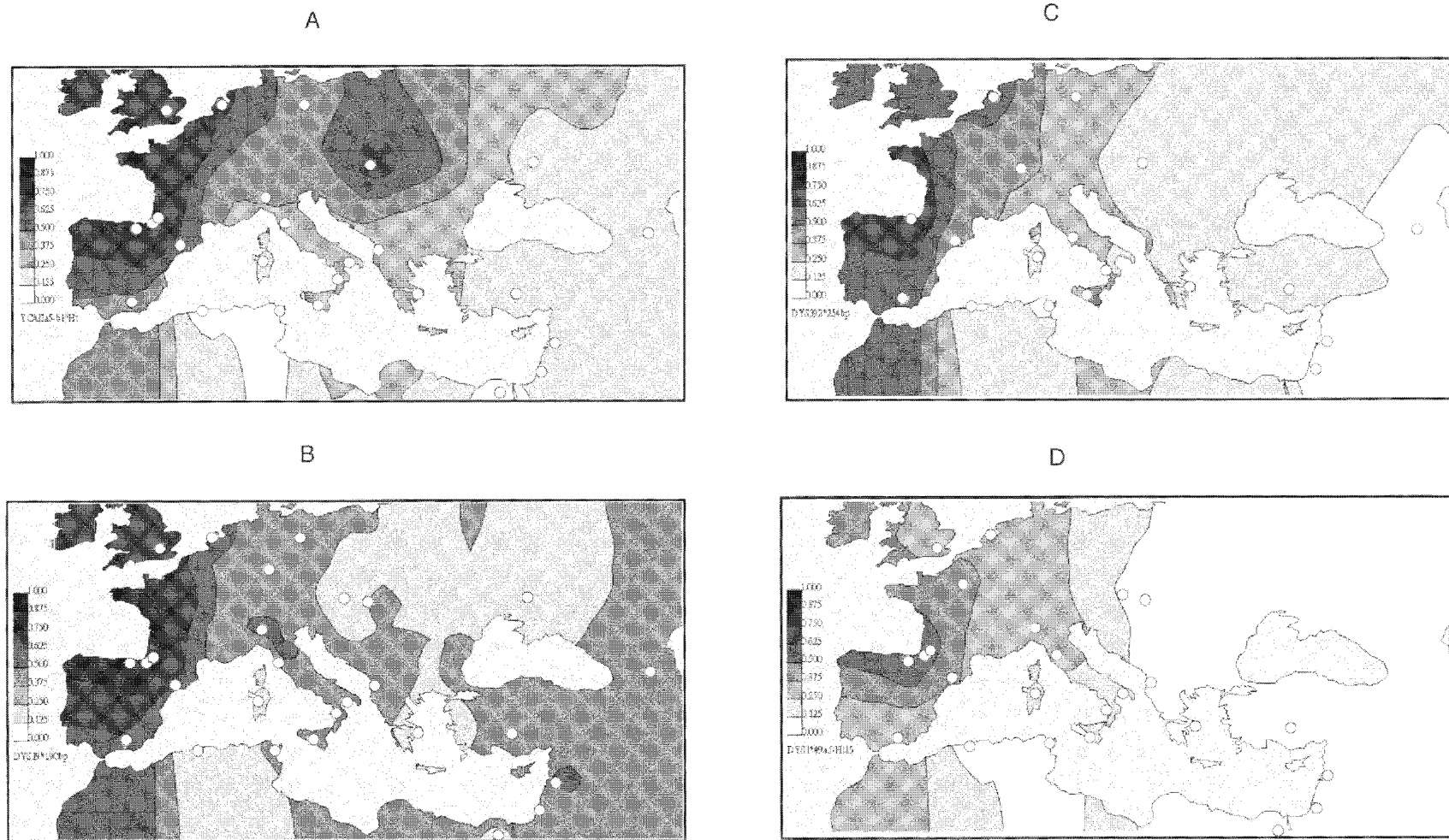
**Table 1** Frequencies in percentage of the YCAII a5-b1 haplotype, DYS19-190 bp allele and DYS392-254 bp allele in different populations

Population	Geographic location	YCAII a5-b1 Ht		DYS19-190 bp		DYS392-254 bp		Ref.
		No.	%	No.	%	No.	%	
Spanish Basque	43°15'N, 2°56'W	48	81.3	52	78.8	–	–	present study
French Basque	43°30'N, 1°30'W	43	74.4	42	83.3	–	–	present study
Béarnais	43°30'N, 1°00'W	27	70.4	24	75.0	–	–	present study
Basque 2	43°15'N, 2°56'W	–	–	52	78.9	52	75.0	17
Catalan	41°25'N, 2°10'E	32	59.4	26	69.2	–	–	present study
Catalan 2	41°25'N, 2°10'E	–	–	25	80.0	32	37.5	17
Andalusian	37°00'N, 3°30'W	55	49.1	48	62.5	30	53.3	present study
Dutch	52°30'N, 4°50'E	23	65.2	29	58.6	–	–	present study
Dutch 2	52°30'N, 4°50'E	–	–	–	–	88	54.0	18
British	51°30'N, 0°10'W	19	63.2	60	70.0	–	–	19, 20
Finnish	60°10'N, 24°58'E	–	–	–	–	280	6.1	21
German	50°00'N, 10°00'E	–	–	306	44.0	–	–	19
Slovak	48°06'N, 17°00'E	–	–	81	21.0	–	–	19
German (Berlin)	52°20'N, 13°00'E	77	48.0	233	39.0	70	26.0	18
Swiss (Berne)	46°57'N, 7°30'E	–	–	100	56.0	64	39.0	18
Hungarian	47°30'N, 19°05'E	53	69.8	52	44.2	30	20.0	present study
North Italian	45°28'N, 9°12'E	35	45.7	64	54.7	–	–	present study
Central Italian	43°20'N, 10°50'E	113	38.0	89	55.1	–	–	present study
Sardinian	40°30'N, 9°00'E	133	12.8	239	27.2	27	22.2	present study
Roman	41°53'N, 12°30'E	–	–	–	–	100	26.0	18
Calabrian	39°20'N, 15°39'E	70	32.9	105	47.6	29	20.7	present study
Apulian	40°20'N, 16°52'E	67	35.8	83	39.8	–	–	present study
Sicilian	37°20'N, 14°00'E	47	36.2	55	50.9	32	46.9	present study
Albanian	41°20'N, 19°49'E	40	27.5	40	40.0	–	–	present study
Greek	38°00'N, 23°44'E	87	26.4	42	28.6	30	23.3	present study
Adygei	48°00'N, 34°00'E	21	19.0	20	25.0	–	–	present study
Georgian	42°30'N, 45°00'E	64	15.6	55	41.8	30	10.0	present study
Turkish	37°51'N, 32°30'E	84	20.2	43	39.5	36	16.7	present study
Sephardi Jewish <sup>a</sup>	32°00'N, 35°00'E	42	16.7	42	40.5	–	–	present study
Ashkenazi Jewish <sup>a</sup>	32°00'N, 35°00'E	41	17.1	41	46.3	27	11.1	present study
Lebanese	33°52'N, 35°30'E	52	15.4	77	51.9	29	10.3	present study
Algerian	36°50'N, 3°00'E	35	2.9	38	23.7	14	7.1	present study
Tunisian	36°50'N, 10°13'E	64	4.7	72	41.7	42	7.1	present study
Egyptian	30°03'N, 31°15'E	44	6.8	44	36.4	–	–	20

<sup>a</sup>The individuals belonging to these two samples were collected in Tel Aviv and were mainly from Tunisia and Morocco (Sephardi) and from Poland and Russia (Ashkenazi). However, they were both located in Israel because of their Middle Eastern origin, as testified by autosomal and uniparental marker studies.<sup>22-25</sup>

At present, in comparison with mtDNA analysis, the Y chromosome seems to display, at least in Europe, a greater geographic clustering and more defined frequency clines. In any case, one must not forget that mtDNA and Y chromosome variations record two separate genetic histories of a population and supply information on related but different aspects of the same reality.

In conclusion, even though recurrent mutations have certainly occurred in STR variation, the three Y-linked microsatellites of this study turned out to be very informative. Indeed, each of them identifies geographic clines in Europe and gives evidence of old genetic characteristics which, in spite of subsequent migrations, still persist in European populations. Moreover, the results of this study find support in archaeological



**Figure 1** Maps of interpolated frequencies for **A** YCAII a5-b1 Ht; **B** DYS19-190 bp allele; **C** DYS392-254 bp allele and **D** 49a,f/TaqI Ht 15.<sup>2</sup> Points were linearly interpolated by the Genography program on a grid with a resolution of 0.5° latitude and 1.0° longitude, using inverse square distance weighted values of Voronoi neighbours (Cavalli-Sforza *et al*, p 46).<sup>10</sup> The circles indicate locations of data samples. (Finland is not represented on these maps.)

observations which locate important cave art records from the early palaeolithic in southwestern France and northern Spain,<sup>10</sup> corresponding with the Basque region.

## Acknowledgements

We are grateful to all the colleagues who have provided us with population samples: J Bertranpetit, for Spanish Basques and Catalans; A Cambon-Thomsen for French Basques and Béarnais; J Garcia-Puche for Andalusians; J Béres for Hungarians; the AVIS centre of Volterra for central Italians; A Grasso and P Pignatelli for Apulians; M Maccioni and G Cossu for Sardinians; G Passarino for Calabrians, Albanians and Greeks; J Kidd and K Kidd for Adygei; I Nasidze for Georgians; M Fellous for Algerians and Tunisians; and G Modiano for Ashkenazi and Sephardi Jews. We also thank A Arslan for help in collection of the Turkish samples. A Liu and M van Gijn are also acknowledged for their help in some STR typing. This work has been supported by funds from the Ministry of University (MURST ex 40%) awarded to ASS-B and AB and the CNR PF 'Beni Culturali' to ASS-B; LQ-M is a Marie Curie Fellow of the EC (ERBFMBICT-961830).

## References

- Richards M, Côrte-Real H, Forster P *et al*: Paleolithic and neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet* 1996; **59**: 185–203.
- Semino O, Passarino G, Brega A, Fellous M, Santachiara-Benerecetti AS: A view of the neolithic demic diffusion in Europe through two Y chromosome-specific markers. *Am J Hum Genet* 1996; **59**: 964–968.
- Cavalli-Sforza LL, Minch E: Paleolithic and neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet* 1995; **61**: 247–251.
- Comas D, Calafell F, Mateu E, Pérez-Lezaun A, Bosch E, Bertranpetit J: Mitochondrial DNA variation and the origin of the Europeans. *Hum Genet* 1997; **99**: 443–449.
- Richards M, Macaulay V, Sykes B *et al*: Reply to Cavalli-Sforza and Minch. *Am J Hum Genet* 1997; **61**: 251–254.
- Barbujani G, Bertorelle G, Chikhi L: Evidence for paleolithic and neolithic gene flow in Europe. *Am J Hum Genet* 1998; **62**: 488–491.
- Richards M, Sykes B: Reply to Barbujani *et al*. *Am J Hum Genet* 1998; **62**: 491–492.
- Torroni A, Bandelt HJ, D'Urbano L *et al*: MtDNA analysis reveals a major late paleolithic population expansion from southwestern to northwestern Europe. *Am J Hum Genet* 1998; **62**: 1137–1152.
- Menozi P, Piazza A, Cavalli-Sforza LL: Synthetic maps of human gene frequencies in Europe. *Science* 1978; **201**: 786–792.
- Cavalli-Sforza LL, Menozzi P, Piazza A: *The History and Geography of Human Genes* Princeton University Press: Princeton, 1994.
- Piazza A, Rendine S, Minch E, Menozzi P, Mountain J, Cavalli-Sforza LL: Genetics and the origin of European languages. *Proc Natl Acad Sci USA* 1995; **92**: 5836–5840.
- Aranzadi T: El pueblo Euskalduna, Diputacion de Guipuzcoa. San Sebastian, 1889.
- Collins R: *The Basques*. Blackwell: Oxford, 1986.
- Mathias N, Bayés M, Tyler-Smith C: Highly informative compound haplotypes for the human Y chromosome. *Hum Mol Genet* 1994; **3**: 115–123.
- Roewer L, Arnemann J, Spurr NK, Grzeschik KH, Epplen JT: Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts. *Hum Genet* 1992; **89**: 389–394.
- Roewer L, Kayser M, Dieltjes P *et al*: Analysis of molecular variance (AMOVA) of Y-chromosome-specific microsatellites in two closely related human populations. *Hum Mol Genet* 1996; **5**: 1029–1033.
- Pérez-Lezaun A, Calafell F, Seielstad M *et al*: Population genetics of Y-chromosome short tandem repeats in humans. *J Mol Evol* 1997; **45**: 265–270.
- Kayser M, Caglià A, Corach D *et al*: Evaluation of Y-chromosomal STRs: a multicenter study. *Int J Legal Med* 1997; **110**: 125–133.
- Santos FR, Gerelsaikhon T, Munkhtuja B, Oyunsuren T, Epplen JT, Pena SDJ: Geographic differences in the allele frequencies of the human Y-linked tetranucleotide polymorphism DYS19. *Hum Genet* 1996; **97**: 309–313.
- Ciminelli BM, Pompei F, Malaspina P *et al*: Recurrent simple tandem repeat mutations during human Y-chromosome radiation in Caucasian subpopulations. *J Mol Evol* 1995; **41**: 966–973.
- Kittles RA, Perola M, Peltonen L *et al*: Dual origin of Finns revealed by Y chromosome haplotype variation. *Am J Hum Genet* 1998; **62**: 1171–1179.
- Livshits G, Sokal RR, Kobylansky E: Genetic affinities of Jewish populations. *Am J Hum Genet* 1991; **49**: 131–146.
- Ritte U, Neufeld E, Prager EM *et al*: Mitochondrial DNA affinity of several Jewish communities. *Hum Biol* 1993; **65**: 359–385.
- Ritte U, Neufeld E, Broit M, Shavit D, Motro U: The differences among Jewish communities – Maternal and Paternal contributions. *J Mol Evol* 1993; **37**: 435–440.
- Santachiara-Benerecetti AS, Semino O, Passarino G *et al*: The common, Near-Eastern origin of Ashkenazi and Sephardi Jews supported by Y-chromosome similarity. *Ann Hum Genet* 1993; **57**: 55–64.
- Persichetti F, Blasi P, Hammer M *et al*: Disequilibrium of multiple DNA markers on the human Y chromosome. *Ann Hum Genet* 1992; **56**: 303–310.
- Bertranpetit J, Calafell F: Genetic and geographical variability in cystic fibrosis: evolutionary considerations. In Chadwick DJ (ed). *Variation in the Human Genome*. Ciba Foundation Symposium 197. Wiley: New York, 1996.
- Goldstein DB, Ruiz-Linares AR, Cavalli-Sforza LL, Feldman MW: Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc Natl Acad Sci USA* 1995; **92**: 6723–6727.
- Cooper G, Amos W, Hoffman D, Rubinsztein DC: Network analysis of human Y microsatellite haplotypes. *Hum Mol Genet* 1996; **5**: 1759–1766.
- Heyer E, Puymirat J, Dieltjes P, Bakker E, de Knijff P: Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. *Hum Mol Genet* 1997; **6**: 799–803.
- de Knijff P, Kayser M, Caglià A *et al*: Chromosome Y microsatellites: population genetic and evolutionary aspects. *Int J Legal Med* 1997; **110**: 134–140.



- 32 Underhill PA, Jin L, Zemans R *et al*: A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. *Proc Natl Acad Sci USA* 1996; **93**: 196–200.
- 33 Devoto M, Romeo G, Serre JL *et al*: Gradient of distribution in Europe of the world CF mutation and of its associated haplotypes. *Hum Genet* 1990; **85**: 436–441.
- 34 Slatkin M, Rannala B: Estimating the age of alleles by use of intraallelic variability. *Am J Hum Genet* 1997; **60**: 447–458.
- 35 Bertranpetit J, Cavalli-Sforza LL: A genetic reconstruction of the history of the population of the Iberian Peninsula. *Ann Hum Genet* 1991; **55**: 51–67.