doi: 10.1111/j.1469-1809.2005.00221.x

Micro-Phylogeographic and Demographic History of Portuguese Male Lineages

Sandra Beleza[1,2*], Leonor Gusmão[1], Alexandra Lopes[1,3], Cíntia Alves[1], Iva Gomes[1], Maria Giouzeli[5], Francesc Calafell[4], Angel Carracedo[2] and António Amorim[1,3]

[1]*IPATIMUP, Instituto de Patologia e Imunologia da Universidade do Porto, R. Dr. Roberto Frias s/n, 4200 Porto, Portugal*
[2]*Unidad de Genética Forense, Instituto de Medicina Legal, Universidad de Santiago de Compostela, E-15705 Santiago de Compostela, Galicia, Spain*
[3]*Faculdade de Ciências da Universidade do Porto, 4050 Porto, Portugal*
[4]*Unitat de Biologia Evolutiva, Facultat de Ciències de la Salut i de la Vida, Universitat Pompeu Fabra, Barcelona 08003, Catalonia, Spain*
[5]*Department of Psychiatry, POWIC SANE Research Centre, University of Oxford, Warneford Hospital, Oxford OX3 7JX, United Kingdom*

Summary

The clinal pattern observed for the distribution of Y-chromosome lineages in Europe is not always reflected at a geographically smaller scale. Six hundred and sixty-three male samples from the 18 administrative districts of Portugal were typed for 25 Y-chromosome biallelic and 15 microsatellite markers, in order to assess the degree of substructuring of male lineage distribution. Haplogroup frequency distributions, Analysis of Molecular Variance (AMOVA) and genetic distance analyses at both Y-SNP and Y-STR levels revealed a general genetic homogeneity of Portuguese sub-populations. The traditional division of the country in north, central and south, which is usually considered in studies addressing questions of the genetic variation distribution in Portugal, was not reflected in the Y-haplotype distribution. Instead, just one sub–region (Alentejo) stood out due to the presence of high diversity levels and a higher number of different lineages, at higher frequencies than in other regions. These results are reconciled with the historical evidence available, assuming that from prehistorical times down to the end of the medieval period this region harboured the most diverse groups of people and, because of economic depression, remained relatively isolated from recent homogenisation movements. The finding of a broadly homogeneous background for the Portuguese population has vast repercussions in forensic, epidemiological and association studies.

Keywords: Y-SNPs, Y-STRs, population structure, phylogeny, Portugal.

Introduction

Portugal, a country situated in the Iberian Peninsula, constitutes the south-western European edge, and faces the Atlantic Ocean to the west and south, just ∼300 km from North Africa. The present Portuguese genetic landscape is the outcome of an old and slow process of gene flow, admixture with many different populations, and local differentiation. These include the expansion from isolated population nuclei in refuges following the Last Glacial Maximum (LGM), the movement of peoples related to the introduction of agriculture, and subsequent Roman and Germanic invaders, which may have influenced the distribution of genetic diversity in the territory. In addition, some admixture events took place in historical epochs that were reported to have left imprints in the Portuguese genetic background, namely of North and sub-Saharan African origin, resulting from the Moslem invasion during the 8th century, and the slave trade, particularly important from the 15th to 18th centuries (Pereira *et al*. 2000a,b; Spinola *et al*. 2002).

*Corresponding author: Sandra Beleza, IPATIMUP, Rua Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal, Tel: +351 22 5570700; Fax: +351 22 5570799. E-mail: sbeleza@ipatimup.pt

Previous analyses of the Portuguese Y-SNP diversity (Pereira *et al*. 2000b) found statistically significant differences between the north and south of Portugal. In that study one haplogroup, E*(xE3a), showed an increasing north-to-south frequency gradient, which was associated with the influx of Berbers during the Islamic period in Iberia in the 8th to 14th centuries. The analysis of mtDNA variability failed to reveal any statistically significant differences (Pereira *et al*. 2000a), although a different study showed opposite frequency gradients of the Paleolithic mtDNA lineage H* (without CRS) and the Neolithic J2 (González *et al*. 2003). These were explained as the result of a higher impact of the Neolithic in the south.

In this work we have conducted a study of the distribution of the Y-chromosome in Portugal, by extending the sampling to all Portuguese districts and by increasing the number of analysed markers. The combined study of slow and fast evolving Y-specific markers provides a suitable approach to analyse the demographic histories of populations on both micro-geographic level and in historical times, at least from a male perspective. In this study we use, for the first time, joint information from both types of markers in order to better understand how the different sequence of historical events interfered with the genetic background of Portugal, and to determine if these events led to a relevant geographic structuring of the male lineages in the territory.

Material and Methods

DNA Samples

A total of 663 blood samples were collected in the 18 districts of Portugal from healthy unrelated individuals born in each area, after informed consent (Figure 1). The samples analysed for Y-STRs (657) and Y-SNPs (658) did not always concur, either because the condition of the sample did not allow further typing or due to the lack of available sample. During the statistical analyses, three categories of samples were considered: either (1) local samples were considered independently, or (2) they were pooled into North, Central and South, taking the rivers Douro and Tagus as barriers, as previously considered by others (Pereira *et al*. 2000a,b; Spinola *et al*. 2002; González *et al*. 2003), or (3) they were assigned to provinces, traditional regions defined according to geography and climate criteria (Amorim Girão, 1941) (Figure 1). Genomic DNA was extracted with the chelex extraction method (Lareu *et al*. 1994).

Marker Typing

Twenty-five Y-chromosome SNP markers were typed in order to define the most frequent male haplogroups in Portugal (Figure 2, plus two SNPs within exon 5 of the *PCDHY* gene). The typing strategy involved several standard methods. The biallelic markers SRY 10831 1/2, YAP, SRY4064, M2, 12f2, M9, 92R7, SRY2627, LLY22g and Tat were typed as previously (Rosser *et al*. 2000). For the remaining 14 Y-SNP markers, the following hierarchical scheme was used (see Table 1 for methods):

- M35, M78, M81 and M123 were tested in individuals that fell within the E*(xE3a);
- M170, M26, M172, M62 and M201 were tested in individuals that fell within haplogroup F*(xK*);

Figure 1 Map of Portugal. Names in the map correspond to the district names. Provinces are individualized by different grey colours. North, centre and south comprise the regions separated by the main Portuguese rivers displayed in the figure, Douro and Tagus.

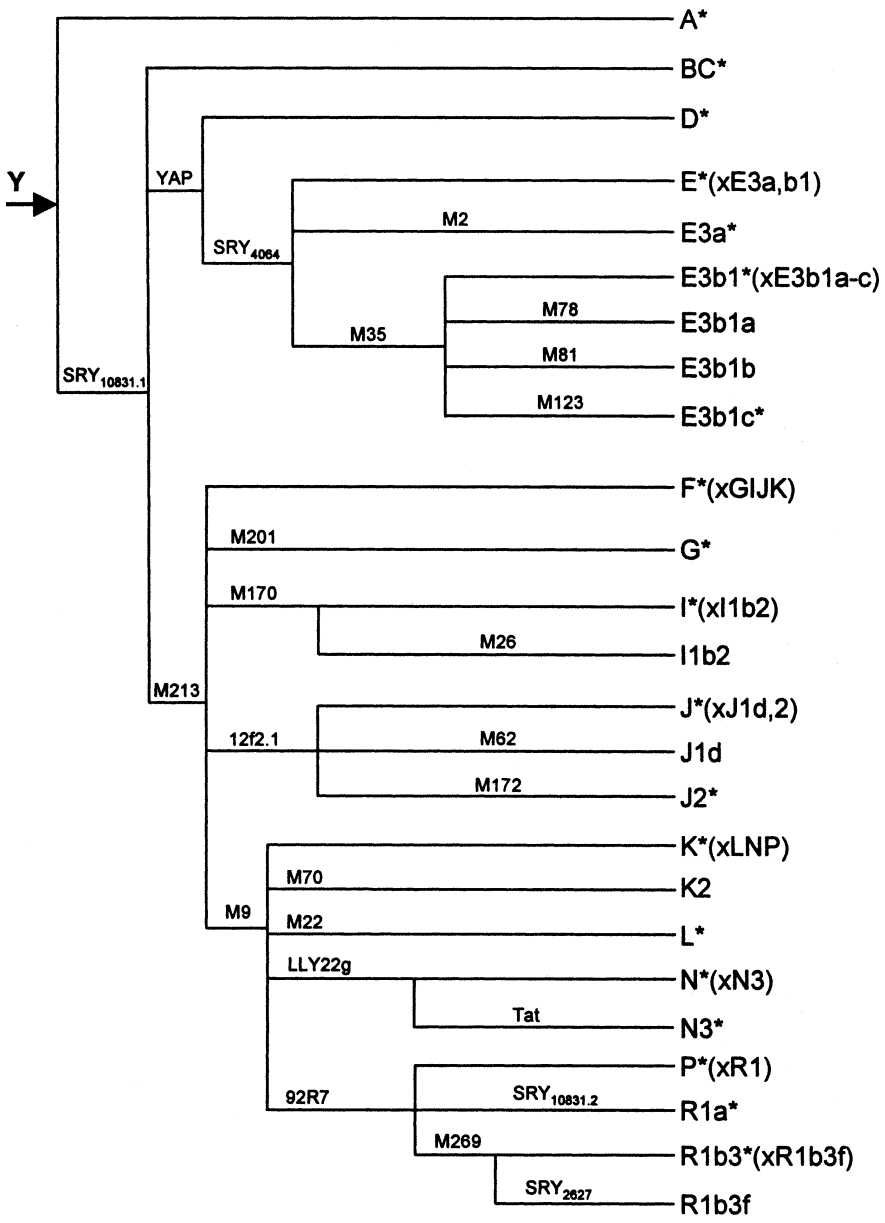

Figure 2 Phylogenetic tree of Y-SNP haplogroups in Portugal. Biallelic markers are displayed in each branch. Note: following the work of Cinnioğlu *et al.* (2004) and Semino *et al.* (2004), "J1d" was attributed to the J*-M62 lineage.

- one BCF* (xGIJK) individual was tested for M213;
- K* (xP) individuals were typed for M70;
- M269 was typed in individuals recognised as R1b* (xR1b3f);
- two individuals that could not be assigned to any of these lineages were later allocated to haplogroup L* due to the to presence of a derived state at M22 (see Table 1);
- finally, two SNPs located within exon 5 of the *PCDHY* gene (NCBI assay IDs: ss5608044 and ss5608045), recently described by Giouzeli *et al.* (2004), were typed according to Lopes *et al.* (2004).

The nomenclature given is according to the YCC (2002) and Jobling & Tyler-Smith (2003). The nomenclature of haplogroups E3b and J was updated according to Cinnioğlu *et al.* (2004), Cruciani *et al.* (2004) and Semino *et al.* (2004).

Fifteen Y-STRs (DYS19, DYS389 I and II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438,

Table 1 Typing strategy and sequence of newly constructed primers used in Y-SNP typing.

Marker	Typing	Ref	Forward and Reverse primers (5′–3′)	Ref	Tm	Enzyme
M35	RFLP	-	GCA TGG TCC CTT TCT ATG GAT	-	60°C	BmrI
			GAG AAT GAA TAG GCA TGG GTT C	-		
M78	RFLP	-	CTT CAG GCA TTA TTT TTT TTG GT	[1]	-	-
			ATA GTG TTC CTT CAC CTT TCC TT	[1]		
M81	RFLP	-	GCA CTA TCA TAC TCA GCT ACA CAT CTC	-	60°C	MaeII
			AAC CAT TGT GTT ACA TGG CCT A	-		
M123	RFLP	-	ATT CAT GCT CTC AGG GGA AA	-	59°C	Sfa NI
			AGC AAA GTT GAG GTT GCA CA	-		
M170	SNaPshot	[2]	Multiplex 2	[2]	-	-
M26	SNaPshot	[2]	Multiplex 2	[2]	-	-
M172	SNaPshot	[2]	Multiplex 2	[2]	-	-
M62	SNaPshot	[2]	Multiplex 2	[2]	-	-
M201	SNaPshot	[2]	Multiplex 2	[2]	-	-
M213	Sequencing	-	GGCCATATAAAAACGCAGCA	[2]	59°C	-
			TGAATGGCAAATTGATTCCA	[2]		
M70	Sequencing	-	TCATAGCCCACTATACTTTGGAC	[2]	59°C	-
			CTGAGGGCTGGACTATAGGG	[2]		
M22	SNaPshot	[2]	Multiplex 1	[2]	-	-
M269	Allele-specific	-	AAT GAT CAG GGT TTG GTT AAA T/C[a,b]	-	59°C	-
	PCR		ACT ATA CTT CTT TTG TGT GCC TTC	-		

[1] Underhill *et al*. 2000; [2] Brion *et al*. 2005.

[a]The two forward primers were labeled with different fluorescent dyes.

[b]Underlined base is a mismatched base, intentionally inserted to prevent the annealing of the non-homologous primer during PCR amplification.

DYS439, DYS460, DYS461, DYS635 (GATA C4), GATA A10 and GATA H4) were typed in two multiplex reactions, using the same primer sequences and amplification and detection conditions as described by Beleza *et al*. (2003). Allele nomenclature was as proposed in the Y Chromosome Haplotype Reference Database (YHRD; www.yhrd.org) and by Gusmão *et al*. (2002).

Statistical Analysis

Haplogroup/haplotype frequencies were estimated by haplogroup/haplotype counting. The heterogeneity in the haplogroup frequencies within the Portuguese territory was assessed by an exact test using a Markov chain procedure, as implemented in the STRUC program included in GENEPOP 3.1c computer package (Raymond & Rousset, 1995). Spatial autocorrelation analysis was done by AIDA (Bertorelle & Barbujani, 1995) for the entire data set. Autocorrelation coefficients (II) were calculated by subdiving population pairs into five up to ten geographic distance classes; in each analysis, classes were defined so that each would contain roughly the same number of population pairs. The same protocol was applied to individual haplogroups, which were subjected to classical autocorrelation analysis (Sokal & Odden, 1978) by means of the PASSAGE software (Rosenberg, 2001).

For both Y-SNP and Y-STR haplotypes, genetic distances, genetic diversities, and the analysis of molecular variance (AMOVA) were calculated with the ARLEQUIN ver 2.000 software (Schneider *et al*. 2000). The distance method used for Y-STRs assumed the stepwise mutational model (R_{ST}; Slatkin, 1995). Another measure for Y-STR diversity, a variance ratio, was also estimated by dividing the variance of each microsatellite in each population by the average variance of the microsatellite across the 18 sub-populations (Jorde *et al*. 2000). A multidimensional scaling (MDS) analysis with pairwise R_{ST} distances was performed using the STATISTICA software package.

Mantel tests were performed in the Mantel Nonparametric Test Calculator, version 2.0 for Windows, developed by Adam Liedloff (http://www.sci.qut.edu.au/NRS/mantel.htm), in order to check for correlation between geographical coordinates of each locality and genetic distances.

Relative age of the main Y-chromosome lineages (with a sample size >7) present in Portugal were

estimated using a Bayesian approach incorporated in the program Batwing (Wilson *et al*. 2003). We considered a model of exponential growth from initial constant population size, using the effective population size and the population growth rate priors specified in Weale *et al*. (2001). We considered two different sets of mutation rates as mutation priors: those estimated from pedigree data (Kayser *et al*. 2000) and the effective mutation rate estimated in phylogenetic analyses (Zhivotovsky *et al*. 2004). All other parameters were given flat, uninformative priors to minimise their impact on the results. The median and the 5% and 95% quantiles for the posterior distribution of the time since the most recent common ancestor of the sample (TMRCA or T) were estimated. Calculations were based on 50,000 runs of MCMC estimator after a 10,000 run "burn in" time. Generation time was set at 25 years.

Results

Y-SNP Haplogroup Variation

Of the 25 haplogroups discriminated by the 25 Y-SNPs that were shown to be polymorphic in European populations (Rosser *et al*. 2000; Semino *et al*. 2000; Underhill *et al*. 2000; 2001; Jobling & Tyler-Smith, 2003) only 17 were present in our sample (Figure 2). Haplogroup frequencies observed in each population and their respective diversities are shown in Table 2. Portugal has a typical Western European haplogroup composition, characterized by high levels of haplogroup R1b3*-M269 and smaller levels of haplogroups I*-M170, G*-M201 and E3b1*-M35 (Semino *et al*. 2000; Bosch *et al*. 2001; Cruciani *et al*. 2002; Flores *et al*. 2004; Brion *et al*. 2005). The only exception is the higher frequency of haplogroup E3b1b, which was already reported to be a characteristic of Southwestern European populations, namely Iberia and Sicily (Rosser *et al*. 2000; Cruciani *et al*. 2004; Semino *et al*. 2004). The study of two recently described SNPs within exon 5 of the *PCDHY* gene, 288 bp apart, in a sample of 69 male individuals of European descent (Giouzeli *et al*. 2004) and in a sample of 112 Portuguese male individuals (Lopes *et al*. 2004) showed them to be polymorphic, but presenting only two allelic combinations: one ancestral (TG) and one doubly derived (GT). A rough estimation of these mutation ages suggested a more recent origin for the derived haplotype compared to M173 (Lopes *et al*. 2004). In our sample, we found that only R1a* and R1b3* chromosomes harboured the derivate state at both *PCDHY* SNPs, meaning that these SNPs could be phylogenetically equivalent to either M173 or M207 mutations. However, more lineages within clade R and within P(xR) need to be typed for these *PCDHY* SNPs to be precisely allocated in the Y phylogenetic tree. The most frequent haplogroup was R1b3*(xR1b3f), presenting frequencies higher than 50% in almost all sub-populations except Alentejo. The other haplogroup frequencies remained lower than 10% in the total sample and within the provinces, but exhibited the opposite trend in some districts.

No significant differences in haplogroup frequencies were found between districts ($P = 0.09$, exact test), as well as between north, central and south ($P = 0.64$, exact test), or between provinces ($P = 0.66$, exact test). Furthermore, and contrary to previous findings (Pereira *et al*. 2000b), no correlations were found between the frequency of any haplogroup and latitude or longitude. Spatial autocorrelograms were also computed by means of AIDA (Bertorelle & Barbujani, 1995). Different ways of dividing pairs of populations into distance classes yielded consistently non-significant correlograms (data not shown). The same was true for the frequencies of all haplogroups present in the Portuguese sample.

The haplogroup diversity values ranged between 0.402 and 0.893. No differences in these values were found between the North, Centre and South (Mann-Whitney test, $P > 0.05$). Curiously, the districts of southern Alentejo show higher diversity values than the other provinces, with marginally significant differences from the North and Central coastal provinces Entre Douro e Minho and Beira Litoral (Mann-Whitney test, $P = 0.049$). To exclude a spurious effect due to different sample sizes, we tested to see if the haplogroup diversities and their standard errors were correlated across provinces, which was shown not to be the case (Spearman's rho $= 0.19$, $P = 0.445$).

Y-STR Haplotype Variation

The typing of a 15 Y-STR set allowed the detection of 577 different haplotypes, 36 of them being shared by two sub-populations, nine by three, two by four and one by five sub-populations (STR haplotype

Table 2 Y-SNP frequency distribution and haplogroup diversity per population in Portugal.

Population, code	n	H	E* (xE3a, b1)	E3a*	E3b1* (xE3b1-c)	E3b1a	E3b1b	E3b1c*	F* (xGIJK)	G*	I* (xI1b2)	I1b2	J* (xJ1d, 2)	J2*	K2	L*	R1a*	R1b3* (xR1b3f)	R1b3f
Entre Douro e Minho																			
Viana do Castelo, **VC**	59	0.695	1.7	0	3.4	0	6.8	0	0	5.1	8.5	0	3.4	15.3	1.7	0	0	52.5	1.7
Braga, **B**	51	0.705	0	2.0[a]	0	2.0	5.9	2.0	0	7.8	15.7	2.0	3.9	0	3.9	0	3.9	51.0	0
Porto, **P**	118	0.652	0	0	0	4.2	6.8	1.7	0	5.9	3.4	0.9	3.4	8.5	2.5	0.9	1.7	57.6	2.5
Total	***228***	***0.676***	***0.4***	***0.4***	***0.9***	***2.6***	***6.6***	***1.3***	***0***	***6.1***	***7.5***	***0.9***	***3.5***	***8.3***	***2.6***	***0.4***	***1.8***	***55.3***	***1.8***
Trás os Montes																			
Bragança, **BN**	25	0.590	0	0	0	4.0	4.0	4.0	0	4.0	12.0	4.0	0	4.0	0	0	0	64.0	0
Vila Real, **VR**	39	0.673	0	0	0	2.6	2.6	0	0	2.6	7.7	5.1	0	7.7	7.7	0	5.1	56.4	2.6
Total	***64***	***0.636***	***0***	***0***	***0***	***3.1***	***3.1***	***1.6***	***0***	***3.1***	***9.4***	***4.7***	***0***	***6.3***	***4.7***	***0***	***3.1***	***59.4***	***1.6***
Beira Litoral																			
Aveiro, **Av**	66	0.626	0	0	0	4.6	3.0	3.0	0	3.0	3.0	1.5	6.1	4.6	6.1	0	1.5	60.7	3.0
Viseu, **Vi**	30	0.547	0	0	0	0	13.3	0	0	0	3.3	0	0	6.7	3.3	0	3.3	66.7	3.3
Coimbra, **Co**	20	0.742	0	0	5.0	0	0	5.0	0	5.0	5.0	0	15.0	10.0	0	5.0	0	50.0	0
Total	***116***	***0.625***	***0***	***0***	***0.9***	***2.6***	***5.2***	***2.6***	***0***	***2.6***	***3.4***	***0.9***	***6.0***	***6.0***	***4.3***	***0.9***	***1.7***	***60.3***	***2.6***
Beira Interior																			
Guarda, **Gu**	30	0.708	0	0	0	10.0	3.3	0	3.3	10.0	0	0	3.3	6.7	3.3	0	3.3	53.3	3.3
Castelo Branco, **CB**	28	0.489	0	0	0	3.6	7.1	0	0	7.1	7.1	0	0	0	0	0	0	71.4	3.6
Total	***58***	***0.605***	***0***	***0***	***0***	***7.0***	***5.3***	***0***	***1.8***	***8.8***	***3.5***	***0***	***1.8***	***3.5***	***1.8***	***0***	***1.8***	***62.3***	***3.5***
Estremadura																			
Leiria, **Le**	35	0.629	0	0	2.9	5.7	5.7	0	0	8.6	11.4	2.9	0	0	0	0	0	60	2.9
Santarém, **Sa**	8	0.821	0	0	0	25.0	0	0	0	0	0	12.5	0	25.0	0	0	0	37.5	0
Total	***43***	***0.674***	***0***	***0***	***2.3***	***9.3***	***4.6***	***0***	***0***	***7.0***	***9.3***	***4.7***	***0***	***4.7***	***0***	***0***	***0***	***55.8***	***2.3***
Lisboa e Setúbal																			
Lisboa, **Li**	35	0.692	0	0	2.9	2.9	11.4	0	0	5.7	2.9	0	0	8.6	0	0	5.7	54.3	5.7
Setúbal, **Se**	27	0.402	3.7	0	3.7	3.7	0	0	0	3.7	0	0	3.7	3.7	0	0	0	77.8	0
Total	***62***	***0.577***	***1.6***	***0***	***3.2***	***3.2***	***6.5***	***0***	***0***	***4.8***	***1.6***	***0***	***1.6***	***6.5***	***0***	***0***	***3.2***	***64.5***	***3.2***
Alentejo																			
Portalegre, **Po**	28	0.759	0	0	0	14.3	10.7	0	0	3.6	7.1	0	3.6	14.3	0	0	0	44.8	0
Évora, **Ev**	29	0.805	0	0	0	6.9	3.5	0	0	13.8	10.3	3.5	10.4	3.5	3.5	0	0	41.4	3.5
Beja, **Be**	8	0.893	12.5	0	0	0	12.5	12.5	0	0	0	12.5	0	12.5	0	0	0	37.5	0
Total	***65***	***0.790***	***1.5***	***0***	***0***	***9.2***	***7.7***	***1.5***	***0***	***7.7***	***7.7***	***3.1***	***6.2***	***9.2***	***1.5***	***0***	***0***	***43.1***	***1.5***
Algarve																			
Faro, **Fa**	21	0.424	0	0	0	0	0	0	0	4.8	4.8	0	4.8	9.5	0	0	0	76.2	0
Total Portugal	**657**	**0.652**	**0.5**	**0.2**	**0.9**	**4.1**	**5.6**	**1.2**	**0.2**	**5.5**	**6.1**	**1.5**	**3.4**	**7.0**	**1.6**	**0.3**	**2.0**	**57.7**	**2.2**

[a]The only E3a* chromosome found holds a STR-haplotype typical of E3b1b (see Table S1).

Table 3 Y-STR haplotype diversity and average variance ratio per population in Portugal.

Population	Viana do Castelo	Braga	Porto	Bragança	Vila Real	Aveiro	Viseu	Coimbra	Guarda	Castelo Branco	Leiria	Santarém	Lisboa	Setúbal	Portalegre	Évora	Beja	Faro
N	60	53	119	45	39	48	30	21	30	28	31	7	35	26	26	28	8	23
Haplotype diversity	1.0000	0.9993	0.9997	0.9990	1.0000	0.9991	0.9954	1.0000	0.9977	1.0000	0.9957	1.0000	0.9966	1.0000	0.9938	0.9974	1.0000	0.996
Average variance ratio	1.0827	1.0601	0.9851	0.9170	0.9057	0.9023	1.0109	1.0205	1.0234	0.6616	1.1215	1.2802	0.9138	0.8144	1.1186	1.0879	1.1053	0.9889

data within each haplogroup is given in supplementary material, Table S1). Haplotype diversity was over 99% in all districts (Table 3), and shown to be uncorrelated with haplogroup diversity ($r = 0.02$). However, estimating a Y-STR variance ratio for each district (Jorde *et al*. 2000) in order to remove the possible bias arising from the contribution of highly mutating Y-STRs, the overall correlation increased to $r = 0.613$. No differences in the Y-STR diversity values were found between the North, Centre and South (Mann-Whitney test, $P > 0.05$). No differences were also observed between the different provinces, although the samples from Alentejo showed consistently higher diversity values.

Genetic Distances and MDS Analysis

Haplogroup distributions in Portuguese sub-populations did not show a clear differentiation (Figure 3). Nevertheless, since the occurrence of unique mutation events at the biallelic loci is much older than the establishment of the Portuguese population, and Y-STR loci possess a higher evolutionary rate, Y-STR haplotypes are more informative in testing whether some heterogeneity exists among these sub-populations as a consequence of either genetic drift or low gene flow. Obviously, given the non-recombinant nature of the NRY, differentiation between haplogroups and haplotypes is correlated (Mantel test between haplogroup Fst and haplotype Rst: $r = 0.554$, $P < 0.01$). However, this value implies that only $\sim 1/4$ of the STR haplotype variation can be explained by haplogroup differentiation; the remaining variation may be a consequence of drift (affecting STR haplotypes more intensely, each of which tends to be relatively infrequent) and/or mutation locally generating STR variation. R_{ST} pairwise comparison values were low (data not shown), as could be predicted from such a microgeographic study of neighbouring populations. The MDS plot of Figure 3 also shows that there is not clear geographic heterogeneity in the Portuguese Y-chromosomal variation. However, with respect to the Y-STR MDS plot, the samples from Alentejo tend to separate from a common cluster composed of the other Portuguese sub-populations (Figure 3).

Moreover, the Mantel test revealed no correlation between the geographical distance and Y-SNP based

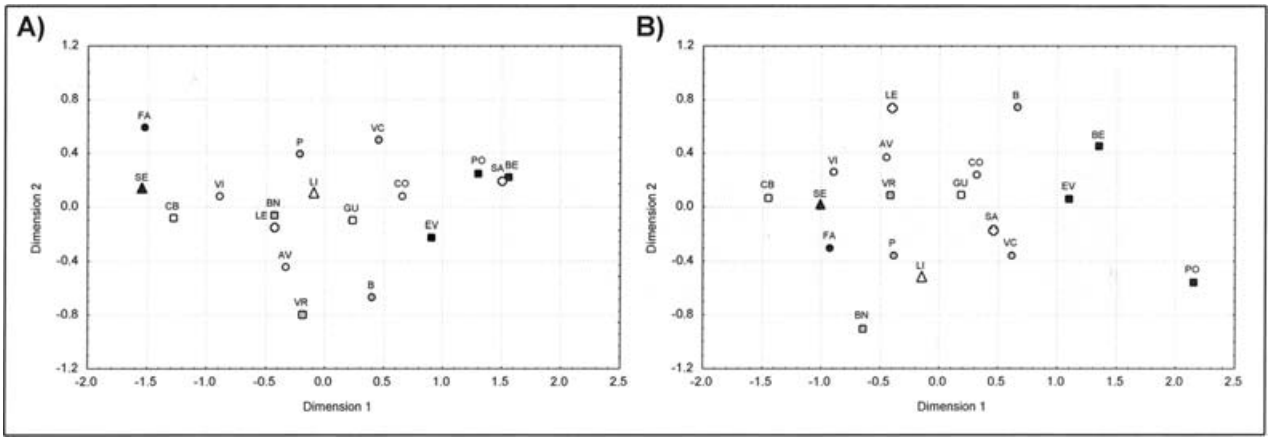

Figure 3 Multidimensional scaling plot of A) F_{ST} and B) Rst pairwise values found when analysing, respectively, Y-SNP and Y-STR variation in Portugal. Stress value = 0.104 in A) and 0.155 in B). Grey, white and black symbols represent northern, central and southern Portuguese districts, respectively. Circles, squares, triangles and crosses represent provinces. District codes and corresponding provinces are indicated in Table 2.

Table 4 AMOVA results of Y-chromosome lineages in Portugal.

Grouping and Source of Variation	Y-SNPs (%)	Y-STRs (%)
No Grouping		
Among populations	0.64	0.67
Within populations	99.36	99.33
North/Centre/South		
Among groups	−0.25	0.05
Among populationsWithin groups	0.81	0.64
Within populations	99.44	99.31
Provinces		
Among groups	0.52	0.52
Among populationsWithin groups	0.18	0.21
Within populations	99.29	99.26
Alentejo vs Portugal		
Among groups	2.99*	3.86*
Among populationsWithin groups	0.04	−0.07
Within populations	96.97	96.21

*Statistically significant values after Bonferroni correction for multiple tests.

genetic distances ($r = 0.0751$, $P = 0.23$) or with Y-STR based genetic distances ($r = -0.0053$, $P = 0.46$).

AMOVA Analysis

Analysis of molecular variance (AMOVA) was computed from haplogroup and haplotype frequencies (Table 4). The overall value obtained between the 18 districts with both data sources was low and non-significant, consistent with the values found for other regions in Iberia (Brion *et al*. 2004; Flores *et al*. 2004). Furthermore, when grouping the districts according to latitude (North, Central and South Portugal), the percentage of variation explained by differences among groups was even smaller. A higher between-group component was found when dividing sub-populations according to the administrative region to which they belong (Table 4).

According to all the evidence obtained until now, namely the higher genetic diversities and genetic distances of populations from Alentejo, we sought to find out the proportion of variation present between Alentejo and the remainder of the Portuguese regions. This criterion led to the highest and most significant value found for the "among groups" component, and to the smallest for the "among populations within groups" component (Table 4).

Phylogeographic Analysis

In order to determine the major components that have left their imprint on the Portuguese male gene pool during its demographic history, we estimated the relative age of the major lineages present in the territory; by considering two different mutation rate estimates (Kayser *et al*. 2000; Zhivotovsky *et al*. 2004) we have obtained an upper and a lower boundary for the age of each lineage in Portugal (Table 5).

Previous phylogeographic analyses allowed us to conclude that haplogroup R1*(xR1a)-M173 has been spread throughout Europe since Paleolithic times

Table 5 Y-chromosome haplogroup variance and age estimates in Portugal based on 15 Y-STR loci.

Haplogroup	n	Variance	Posterior Probabilities (in Ky)[a]	
			T_{lower}[1]	T_{upper}[2]
R1b3*(xR1b3f)-M269	335	0.286	14.1 (7.7, 34.3)	38.3 (22.9, 77.4)
R1b3f – SRY_{2627}	14	0.172	3.1 (1.7, 6.9)	9.9 (5.9, 19.2)
R1a*- $SRY_{10831.2}$	13	0.517	13.6 (7.6, 29.1)	40.8 (26.2, 77.2)
K2 – M70	14	0.299	5.6 (3.1,13.1)	17.5 (10.7, 36.5)
J*(xJ1d,2)- 12f2	21	0.329	8.7 (4.5, 21.0)	27.7 (15.7, 59.7)
J2* – M172	44	0.607	11.9 (7.0, 25.2)	37.5 (25.1, 67.0)
I*(xI1b2)-M170	38	0.521	10.3 (6.1, 20.8)	33.5 (22.2, 61.0)
I1b2-M26	10	0.376	6.6 (3.7, 13.9)	20.2 (12.8, 37.0)
G*-M201	36	0.452	8.5 (4.9, 20.1)	27.8 (18.0, 57.8)
E3b1a-M78	26	0.383	5.5 (3.3, 10.5)	17.2 (11.9, 28.8)
E3b1b-M81	37	0.129	2.8 (1.4, 7.1)	8.0 (4.8, 18.1)
E3b1c*-M123	8	0.426	7.8 (4.4, 16.6)	23.4 (14.8, 43.4)

[a]Median (95% equal-tailed intervals).
[1]Estimated with the mutation rate obtained from pedigree data (Kayser *et al*. 2000)
[2]Estimated with the effective mutation rate obtained in phylogenetic analyses (Zhivotovsky *et al*. 2004).

(Semino *et al*. 2000). Later, M269 was found to be present in all European R1*(xR1a)-M173 individuals by Cruciani *et al*. (2002), and STR variance analysis suggested that the distribution of R1b3*-M269 in this continent is more the result of its diffusion from the Iberian refugia after the LGM, during the Late Upper Paleolithic (Cinnioğlu *et al*. 2004). The same analysis of variance made with the Portuguese R1b3*(xR1b3f) individuals corroborates these findings (STR variance of the Portuguese R1b3*(xR1b3f) equals 0.29, a smaller value when compared with that obtained by Cinnioğlu *et al*. [2004] in other Iberian samples [0.32]). Age estimates obtained from the STR variation associated with R1b3*(xR1b3f) chromosomes in Portugal (Table 5) point to a pre-Neolithic age for this haplogroup, and set an upper limit for its introduction to Portugal (Barbujani *et al*. 1998). Another lineage also found to be characteristic of LGM expansions from refuge areas, I*-M170 (Rootsi *et al*. 2004; see also Table 5), was present in our sample at lower levels.

Due to their decreasing frequencies from the Middle East to Europe, haplogroups J, G and E have been associated with the demic diffusion of Neolithic farmers into Europe (Semino *et al*. 2000; Underhill *et al*. 2001). J lineages were found to have a greater impact in the eastern and central part of the Mediterranean basin (Scozzari *et al*. 2001; Di Giacomo *et al*. 2003) than in Iberia and northeastern Europe (Scozzari *et al*. 2001; Flores *et al*. 2004). Later, the refinement of the phylogeny of haplogroups J and E allowed the discrimination of a collection of subclades of Neolithic and post-Neolithic origin, reflecting very different evolutionary histories and processes of migration (Cruciani *et al*. 2004; Semino *et al*. 2004). Indeed, the age estimates of J*(xJ1d,2)-12f2, G*-M201 and E3b1c*-M123 (Table 5) are consistent with an introduction to Portugal no earlier than the Neolithic. J2*-M172 presents higher age estimates (Table 5) but, judging from its wide variance, this must be a consequence of the coexistence of more than one subclade within this haplogroup (mainly J2-M172 and J2f*-M67, both of Neolithic origin according to Semino *et al*. 2004) that was not discriminated with the set of Y-SNP markers typed. This set of post-Paleolithic Y chromosomes comprises 17% of the male Portuguese background, a frequency that is smaller than that observed in Italy or Greece for the same component (Semino *et al*. 2000; Di Giacomo *et al*. 2003) but higher than in northeastern Europe (Semino *et al*. 2000; Wells *et al*. 2001). Within the Portuguese territory, this component is present at 17.5% in the north, 15% in the centre and 20.4% in the south.

The distribution of E3b1a-M78 (E3b1a) in Western Europe was described as resulting from late demographic expansions from the Balkans along the Mediterranean Sea (Cruciani *et al*. 2004; Semino *et al*. 2004), and E31b1a was considered to be a signature of Greek colonists in south Italy (Semino *et al*. 2004). Its presence in the Portuguese territory might have the same origin (see Table 5). In fact, within Portugal there is a higher frequency of E3b1a in the south, which decreases

toward the north (6.2%, 4.8% and 2.7% in south, centre and north, respectively). Nevertheless, and given the current sample sizes, these values are not significantly different from each other ($P = 0.64$, exact test), and the autocorrelation analysis was not significant (data not shown).

E3b1b-M81 was found at 1.5–11.5% in a number of Spanish populations (Bosch *et al*. 2001; Cruciani *et al*. 2004; Flores *et al*. 2004; Brion *et al*. 2005), being as high as 20%–40% in Pasiegos from Cantabria, Spain (Brion *et al*. 2004; Cruciani *et al*. 2004). The distribution of these E3b1b chromosomes and respective age estimates in Europe (e.g. Table 5) led to the suggestion that its presence in the continent was mediated by the Islamic occupation of the Iberian Peninsula (Bosch *et al*. 2001; Cruciani *et al*. 2004). The degree of North African contribution was found to be highly variable across different Iberian populations (Cruciani *et al*. 2004; Flores *et al*. 2004). The same holds true in Portugal, where no clear pattern can be observed (Table 2). Considering the division of the territory by latitude, we find 5.8% of E3b1b lineages in the north, 6% in the centre and 4.4% in south. Pereira *et al*. (2000*b*) speculated that the observation of a north to south increasing gradient of E* (xE3a) might be associated with the patterns of Islamic occupation in Portugal reported in the historical records (stronger and longer-lasting in the south). Our results do not support this hypothesis, and the higher frequencies of E* (xE3a) in the south observed by the authors may have two non-exclusive explanations. Firstly, they might be a consequence of the higher frequency of E3b1a rather than E3b1b in this region. Secondly, these observations were based on a relatively small sample size for the south compared with the northern one, and the southern samples were mainly from Alentejo, which possesses higher frequencies of both haplogroups. Estimating the frequency of the same lineage in our extended sample we also obtained an increasing gradient of 11.0% in the north, 13.1% in the centre and 14.4% in the south, although this was not significant. One fact worth mentioning is the complete absence of this lineage in the most southern Portuguese region, Algarve, the area that is described as one of those most influenced by the Islamic rule (Torres, 1993), and its presence at 6.6% in the northern region, Entre Douro e Minho. Interestingly, the North African mtDNA counterpart, U6, is also found at 6% in the same northern region, the highest frequency found in Portugal (Pereira *et al*. 2000a; González *et al*. 2003), being rarer in southern areas.

When analysing globally the European Y lineages, we observe that all of them, R1b3* (xR1b3f) excepted, are observed at higher frequencies in Alentejo (Table 2).

Finally, we have detected two L chromosomes, a haplogroup that is characteristic of Southern Indian populations (Kivisild *et al*. 2003) but extremely rare in Europe (Semino *et al*. 2000; Scozzari *et al*. 2001; Flores *et al*. 2004; Brion *et al*. 2005).

Discussion

Y-chromosome population studies on a large geographic scale (Rosser *et al*. 2000; Semino *et al*. 2000) have revealed a significant substructuring of Y-chromosome variation and important clinal patterns for some male lineages in Europe. However, research in more restricted geographical contexts has shown that this pattern is not always reflected on a smaller scale (Di Giacomo *et al*. 2003; Flores *et al*. 2004), revealing important genetic drift and founder effects (Di Giacomo *et al*. 2003). In this work we have analysed for the first time joint information given by Y-SNPs, and Y-STRs in order to assess the degree of micro-geographic substructuring of the Portuguese male lineages. It is known from archaeological data (Ribeiro, 1966) that, throughout the demographic history of Portugal, there was a clear differentiation between cultures that existed to the north of the river Douro and those that influenced the south, beyond the Tagus River. However, our results failed to detect any significant male gene pool heterogeneity between the north, central and south regions. Conversely, our analysis revealed a general genetic homogeneity of Portuguese sub-populations, with the exception of those belonging to Alentejo. The relative differentiation of Alentejo comes from the presence and abundance of a higher number of different lineages of recent ancestry than are found in other sub-populations (Table 2). On the other hand, the lineage showing an opposite clinal distribution to the arrival of new influences in

Western Europe, R1b3*-M269, presents an unusually low frequency in Alentejo: 43.1% vs values higher than 50% in other Western European populations (Rosser *et al*. 2000; Semino *et al*. 2000). This seems rather curious since populations from Alentejo nowadays are the smallest in size and are still decreasing. However, this was not the case up to the end of the 15th century, when they harboured important rural and urban populations.

As in other European populations, almost all Portuguese Y chromosomes belong to haplogroups R1b3*-M269, I*-M170, J*-12f2, G*-M201 and E3b1*-M35.

One haplogroup that deserves special consideration is E3b1b-M81. Historical records show that in Portugal (and in all Iberia) the Muslim influence was more limited to the south, but occupying a vast area that had the Iberian central mountain range as its northern limit (Torres, 1993). However, what we observe today at the genetic level, both in the paternal (Table 2; Brion *et al*. 2004; Cruciani *et al*. 2004; Flores *et al*. 2004) and the maternal gene pool (Pereira *et al*. 2000a; González *et al*. 2003; Maca-Meyer *et al*. 2003), is that this North African influence is well spread all over Iberia. However, the later Christian Reconquest is described (Bartlett, 1993) as a period of extensive migrations all over the Iberian Peninsula, involving the resettlement of Iberia on a vast scale. These might have erased any differential distribution of the Y-chromosome that could have existed in the past, especially for E3b1b; and today we witness a rather homogeneous genetic landscape in Portugal (and probably also in most of Iberia; Flores *et al*. 2004), at least from the male perspective.

Finally, our findings for the Portuguese paternal background and the ones of Pereira *et al*. (2000a; 2000b) for the paternal and maternal components of the Portuguese population do not reconcile with the clines detected by González *et al*. (2003). A more detailed survey of the distribution pattern of female lineages in Portugal is still needed, with a larger sample size and a more in-depth phylogenetic analysis.

In conclusion, we show the usefulness of a detailed Y-chromosome analysis to the understanding of the demographic history of human populations in the recent past, and at a micro-geographic level, a population attribute that has had a great impact on forensic and epidemiological applications. The finding of a rather homogeneous background for the Portuguese Y-chromosome gene pool, and all the strong indications that the same carachteristic is found for its female counterpart, are of great relevance for association studies. Although genetic differentiation follows a random distribution, and certainly some genes of genetic epidemiological interest are bound to show more stratification than Y lineages do, it is nonetheless true that the Y chromosome is the genomic region most sensitive to drift, and thus, with the largest oportunity for stratification. The low levels of differentiation for the Y chromosome in Portugal add additional confidence to the use of general Portuguese samples in case-control designs.

Acknowledgements

This work was supported by POCTI, Programa Operacional Ciência, Tecnologia e Inovação. Sandra Beleza and Alexandra Lopes have a PhD grant (SFRH/BD/860/2000 and SFRH/BD/7006/2001, respectively) from FCT. Two anonymous reviewers made significant contributions to the improvement of this manuscript. We wish to thank local hospitals from the whole country for providing the samples that made this study possible.

References

Amorim Girão, A. (1941) *Geografia de Portugal*, p. 330. Portucalense Editora, Porto.

Barbujani, G., Bertorelle, G. & Chikhi, L. (1998) Evidence for Paleolithic and Neolithic gene flow in Europe. *Am J of Hum Genet* **62**, 488–492.

Bartlett, R. (1993) *The Making of Europe: Conquest, Colonization, and Cultural Change, 950–1350*. New Jersey: Princeton University Press.

Beleza, S., Alves, C., González-Neira, A., Lareu, M., Amorim, A., Carracedo, A. & Gusmão, L. (2003) Extending STR markers in Y chromosome haplotypes. *Int J Legal Med* **117**, 27–33.

Bertorelle, G. & Barbujani, G. (1995) Analysis of DNA diversity by spatial autocorrelation. *Genetics* **140**, 811–819.

Bosch, E., Calafell, F., Comas, D., Oefner, P. J., Underhill, P. A. & Bertranpetit, J. (2001) High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between northwestern Africa and the Iberian Peninsula. *Am J Hum Genet* **68**, 1019–1029.

Brion, M., Quintans, B., Zarrabeitia, M., Gonzalez-Neira, A., Salas, A., Lareu, V., Tyler-Smith, C. & Carracedo,

A. (2004) Micro-geographical differentiation in Northern Iberia revealed by Y-chromosomal DNA analysis. *Gene* **329**, 17–25.

Brion, M., Sobrino, B., Blanco-Verea, A., Lareu, M. V. & Carracedo, A. (2005) Hierarchical analysis of 30 Y-chromosome SNPs in European populations. *Int J Legal Med* **119**, 10–15.

Cinnioğlu, C., King, R., Kivisild, T., Kalfoglu, E., Atasoy, S., Cavalleri, G. L., Lillie, A. S., Roseman, C. C., Lin, A. A., Prince, K., Oefner, P. J., Shen, P., Semino, O., Cavalli-Sforza, L. L. & Underhill, P. A. (2004) Excavating Y-chromosome haplotype strata in Anatolia. *Hum Genet* **114**, 127–148.

Cruciani, F., Santolamazza, P., Shen, P., Macaulay, V., Moral, P., Olckers, A., Modiano, D., Holmes, S., Destro-Bisol, G., Coia, V., Wallace, D. C., Oefner, P. J., Torroni, A., Cavalli-Sforza, L. L., Scozzari, R. & Underhill, P. A. (2002) A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. *Am J Hum Genet* **70**, 1197–1214.

Cruciani, F., La Fratta, R., Santolamazza, P., Sellitto, D., Pascone, R., Moral, P., Watson, E., Guida, V., Colomb, E. B., Zaharova, B., Lavinha, J., Vona, G., Aman, R., Cali, F., Akar, N., Richards, M., Torroni, A., Novelletto, A. & Scozzari, R. (2004) Phylogeographic analysis of haplogroup E3b (E-M215) y chromosomes reveals multiple migratory events within and out of Africa. *Am J Hum Genet* **74**, 1014–1022.

Di Giacomo, F., Luca, F., Anagnou, N., Ciavarella, G., Corbo, R. M., Cresta, M., Cucci, F., Di Stasi, L., Agostiano, V., Giparaki, M., Loutradis, A., Mammi', C., Michalodimitrakis, E. N., Papola, F., Pedicini, G., Plata, E., Terrenato, L., Tofanelli, S., Malaspina, P. & Novelletto, A. (2003) Clinal patterns of human Y chromosomal diversity in continental Italy and Greece are dominated by drift and founder effects. *Mol Phylogenet Evol* **28**, 387–395.

Flores, C., Maca-Meyer, N., Gonzalez, A. M., Oefner, P. J., Shen, P., Perez, J. A., Rojas, A., Larruga, J. M. & Underhill, P. A. (2004) Reduced genetic structure of the Iberian peninsula revealed by Y-chromosome analysis: implications for population demography. *Eur J Hum Genet* **12**, 855–863.

Giouzeli, M., Williams, N. A., Lonie, L. J., DeLisi, L. E. & Crow, T. J. (2004) ProtocadherinX/Y, a candidate gene-pair for schizophrenia and schizoaffective disorder: a DHPLC investigation of genomic sequence. *Am J Med Genet* **129B**, 1–9.

González, A. M., Brehm, A., Perez, J. A., Maca-Meyer, N., Flores, C. & Cabrera, V. M. (2003) Mitochondrial DNA affinities at the Atlantic fringe of Europe. *Am J Phys Anthropol* **120**, 391–404.

Gusmão, L., González-Neira, A., Lareu, M., Costa, S., Amorim, A. & Carracedo, A. (2002) Chimpanzee homologous of human Y specific STRs: a comparative study and a proposal for nomenclature. *Forensic Sci Int* **126**, 129–136.

Jobling, M. & Tyler–Smith, C. (2003) The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* **4**, 598–612.

Jorde, L. B., Watkins, W. S., Bamshad, M. J., Dixon, M. E., Ricker, C. E., Seielstad, M. T. & Batzer, M. A. (2000) The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *Am J Hum Genet* **66**, 979–988.

Kayser, M., Roewer, L., Hedman, M., Henke, L., Henke, J., Brauer, S., Kruger, C., Krawczak, M., Nagy, M., Dobosz, T., Szibor, R., de Knijff, P., Stoneking, M. & Sajantila, A. (2000) Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. *Am J Hum Genet* **66**, 1580–1588.

Kivisild, T., Rootsi, S., Metspalu, M., Mastana, S., Kaldma, K., Parik, J., Metspalu, E., Adojaan, M., Tolk, H. V., Stepanov, V., Golge, M., Usanga, E., Papiha, S. S., Cinnioglu, C., King, R., Cavalli-Sforza, L., Underhill, P. A. & Villems, R. (2003) The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am J Hum Genet* **72**, 313–332.

Lareu, M. V., Phillips, C. P., Carracedo, A., Lincoln, P. J., Syndercombe Court, D. & Thomson, J. Á. (1994) Investigation of the STR locus HUMTH01 using PCR and two electrophoresis formats: UK and Galician Caucasian population surveys and usefulness in paternity investigations. *Forensic Sci Int* **66**, 41–52.

Lopes, A. M., Calafell, F. & Amorim, A. (2004) Microsatellite Variation and Evolutionary History of PCDHX/Y Gene Pair Within the Xq21.3/Yp11.2 Hominid-Specific Homology Block. *Mol Biol Evol* **21**, 2092–2101.

Maca-Meyer, N., Sanchez-Velasco, P., Flores, C., Larruga, J. M., Gonzalez, A. M., Oterino, A. & Leyva-Cobian, F. (2003) Y chromosome and mitochondrial DNA characterization of Pasiegos, a human isolate from Cantabria (Spain). *Ann Hum Genet* **67**, 329–339.

Pereira, L., Prata, M. J. & Amorim, A. (2000a) Diversity of mtDNA lineages in Portugal: not a genetic edge of European variation. *Ann Hum Genet* **64**, 491–506.

Pereira, L., Prata, M. J., Brion, M., Jobling, M. A., Carracedo, A. & Amorim, A. (2000b) Clinal variation of YAP+ Y-chromosome frequencies in Western Iberia. *Hum Biol* **72**, 937–944.

Raymond, M. & Rousset, F. (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J Heredity* **86**, 248–249.

Ribeiro, O. (1966) Portugal, A Formação de. In: *Dicionário da História de Portugal* (eds Iniciativas Editoriais), Vol. V, pp. 130–149. Lisboa, Portugal.

Rootsi, S., Magri, C., Kivisild, T., Benuzzi, G., Help, H., Bermisheva, M., Kutuev, I., Barac, L., Pericic, M., Balanovsky, O., Pshenichnov, A., Dion, D., Grobei, M., Zhivotovsky, L. A., Battaglia, V., Achilli, A., Al-Zahery, N., Parik, J., King, R., Cinnioglu, C., Khusnutdinova, E., Rudan, P., Balanovska, E., Scheffrahn, W., Simonescu, M., Brehm, A., Goncalves, R., Rosa, A., Moisan, J. P., Chaventre, A., Ferak, V., Furedi, S., Oefner, P. J., Shen, P., Beckman, L., Mikerezi, I., Terzic, R., Primorac, D., Cambon-Thomsen, A., Krumina, A., Torroni, A., Underhill, P. A., Santachiara-Benerecetti, A. S., Villems, R. & Semino, O. (2004) Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in europe. *Am J Hum Genet* **75**, 128–137.

Rosenberg, M. S. (2001) *PASSAGE. Pattern Analysis, Spatial Statistics, and GeographicExegesis. Version 1.1.* Tempe, AZ: Department of Biology, Arizona State University.

Rosser, Z. H., Zerjal, T., Hurles, M. E., Adojaan, M., Alavantic, D., Amorim, A., Amos, W., Armenteros, M., Arroyo, E., Barbujani, G., Beckman, G., Beckman, L., Bertranpetit, J., Bosch, E., Bradley, D. G., Brede, G., Cooper, G., Corte-Real, H. B., de Knijff, P., Decorte, R., Dubrova, Y. E., Evgrafov, O., Gilissen, A., Glisic, S., Golge, M., Hill, E. W., Jeziorowska, A., Kalaydjieva, L., Kayser, M., Kivisild, T., Kravchenko, S. A., Krumina, A., Kucinskas, V., Lavinha, J., Livshits, L. A., Malaspina, P., Maria, S., McElreavey, K., Meitinger, T. A., Mikelsaar, A. V., Mitchell, R. J., Nafa, K., Nicholson, J., Norby, S., Pandya, A., Parik, J., Patsalis, P. C., Pereira, L., Peterlin, B., Pielberg, G., Prata, M. J., Previdere, C., Roewer, L., Rootsi, S., Rubinsztein, D. C., Saillard, J., Santos, F. R., Stefanescu, G., Sykes, B. C., Tolun, A., Villems, R., Tyler-Smith, C. & Jobling, M. A. (2000) Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am J Hum Genet* **67**, 1526–1543.

Schneider, S., Roessli, D. & Excoffier, L. (2000) *Arlequin: A software for population genetics data analysis*. Ver 2.000. University of Geneva.

Scozzari, R., Cruciani, F., Pangrazio, A., Santolamazza, P., Vona, G., Moral, P., Latini, V., Varesi, L., Memmi, M. M., Romano, V., De Leo, G., Gennarelli, M., Jaruzelska, J., Villems, R., Parik, J., Macaulay, V. & Torroni, A. (2001) Human Y-chromosome variation in the western Mediterranean area: implications for the peopling of the region. *Hum Immunol* **62**, 871–884.

Semino, O., Magri, C., Benuzzi, G., Lin, A. A., Al-Zahery, N., Battaglia, V., Maccioni, L., Triantaphyllidis, C., Shen, P., Oefner, P. J., Zhivotovsky, L. A., King, R., Torroni, A., Cavalli-Sforza, L. L., Underhill, P. A. & Santachiara-Benerecetti, A. S. (2004) Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. *Am J Hum Genet* **74**, 1023–1034.

Semino, O., Passarino, G., Oefner, P. J., Lin, A. A., Arbuzova, S., Beckman, L. E., De Benedictis, G., Francalacci, P., Kouvatsi, A., Limborska, S., Marcikiae, M., Mika, A., Mika, B., Primorac, D., Santachiara-Benerecetti, A. S., Cavalli-Sforza, L. L. & Underhill, P. A. (2000) The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science* **290**, 1155–1159.

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

Spinola, H., Brehm, A., Williams, F., Jesus, J. & Middleton, D. (2002) Distribution of HLA alleles in Portugal and Cabo Verde. Relationships with the slave trade route. *Ann Hum Genet* **66**, 285–296.

Sokal, R. R. & Oden, N. L. (1978) Spatial autocorrelation in biology 1. *Methodology* **10**, 199–228.

Torres, C. (1993) O Garb-Al-Andaluz. In: *História de Portugal* (ed J. Mattoso), Vol. I, pp. 361–439. Lisboa, Portugal: Círculo dos Leitores.

Underhill, P. A., Shen, P., Lin, A. A., Jin, L., Passarino, G., Yang, W. H., Kauffman, E., Bonne-Tamir, B., Bertranpetit, J., Francalacci, P., Ibrahim, M., Jenkins, T., Kidd, J. R., Mehdi, S. Q., Seielstad, M. T., Wells, R. S., Piazza, A., Davis, R. W., Feldman, M. W., Cavalli-Sforza, L. L. & Oefner, P. J. (2000) Y chromosome sequence variation and the history of human populations. *Nat Genet* **26**, 358–361.

Underhill, P. A., Passarino, G., Lin, A. A., Shen, P., Mirazon Lahr, M., Foley, R. A., Oefner, P. J. & Cavalli-Sforza, L. L. (2001) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann Hum Genet* **65**, 43–62.

Weale, M. E., Yepiskoposyan, L., Jager, R. F., Hovhannisyan, N., Khudoyan, A., Burbage-Hall, O., Bradman, N. & Thomas, M. G. (2001) Armenian Y chromosome haplotypes reveal strong regional structure within a single ethnonational group. *Hum Genet* **109**, 659–674.

Wells, R. S., Yuldasheva, N., Ruzibakiev, R., Underhill, P. A., Evseeva, I., Blue-Smith, J., Jin, L., Su, B., Pitchappan, R., Shanmugalakshmi, S., Balakrishnan, K., Read, M., Pearson, N. M., Zerjal, T., Webster, M. T., Zholoshvili, I., Jamarjashvili, E., Gambarov, S., Nikbin, B., Dostiev, A., Aknazarov, O., Zalloua, P., Tsoy, I., Kitaev, M., Mirrakhimov, M., Chariev, A. & Bodmer, W. F. (2001) The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc Natl Acad Sci U S A* **98**, 10244–10249.

Wilson, I., Balding, D. & Weale, M. (2003) Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities. *J R Stat Soc Ser A* **166**, 155–188.

Y Chromosome Consortium (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* **12**, 339–348.

Zhivotovsky, L. A., Underhill, P. A., Cinnioglu, C., Kayser, M., Morar, B., Kivisild, T., Scozzari, R., Cruciani, F., Destro-Bisol, G., Spedini, G., Chambers, G. K., Herrera, R. J., Yong, K. K., Gresham, D., Tournev, I., Feldman, M. W. & Kalaydjieva, L. (2004) The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am J Hum Genet* **74**, 50–61.

Received: 25 January 2005
Accepted: 14 June 2005

Supplementary Material

The following material is available for this article online:
Table S1. Y-chromosome halpotypes in 657 individuals from Portugal.

