doi: 10.1046/j.1529-8817.2004.00147.x

The Y-chromosomal Heritage of the Azores Islands Population

P. R. Pacheco[1,2], C. C. Branco[1,2], R. Cabral[1,2], S. Costa[1,2], A. L. Araújo[3], B. R. Peixoto[1,2], P. Mendonça[3] and L. Mota-Vieira[1,2*]

[1]*Molecular Genetics and Pathology Unit, Hospital of Divino Espírito Santo, Ponta Delgada, São Miguel Island, Azores, Portugal*
[2]*Instituto Gulbenkian de Ciência, Oeiras, Portugal*
[3]*Hematology Department, Hospital of Divino Espírito Santo, Ponta Delgada, São Miguel Island, Azores, Portugal*

Summary

The Azores, a Portuguese archipelago located in the north Atlantic Ocean, had no native population when the Portuguese first arrived in the 15th century. The islands were populated mainly by the Portuguese, but Jews, Moorish prisoners, African slaves, Flemish, French and Spaniards also contributed to the initial settlement. To understand the paternal origins and diversity of the extant Azorean population, we typed genomic DNA samples from 172 individuals using a combination of 10 Y-biallelic markers (YAP, SRY-1532, SRY-2627, 92R7, M9, sY81, Tat, SRY-8299, 12f2 and LLY22g) and the following Y-chromosomal STR systems: DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393 and DYS385. We identified nine different haplogroups, most of which are frequent in Europe. Haplogroup J* is the second most frequent in the Azores (13.4%), but it is modestly represented in mainland Portugal (6.8%). The other non-European haplogroups, N3 and E3a, which are prevalent in Asia and sub-Saharan Africa, respectively, have been found in the Azores (0.6% and 1.2%, respectively) but not in mainland Portugal. Microsatellite data indicate that the mean gene diversity (D) value for all the loci analysed in our sample set is 0.590, while haplotype diversity is 0.9994. Taken together, our analysis suggests that the current paternal pool of the Azorean population is, to a great extent, of Portuguese descent with significant contributions from people with other genetic backgrounds.

Keywords: Y-chromosome, Y-STR, Y-SNP, Azores Islands

Introduction

The Y-chromosome is a powerful tool to study human evolutionary pathways and to infer major and local male migration movements or patterns (Jobling & Tyler-Smith, 1995). The nonrecombining portion of the Y (NRY) retains a record of the mutational events that occurred along male lineages throughout evolution. Binary polymorphisms are particularly useful to identify stable paternal lineages, traced back in time over thousands of years, because of their low rate of parallel and back mutation (Y Chromosome Consortium 2002). The diversity within these lineages – haplogroups – can be examined using polymorphisms that mutate more rapidly, such as microsatellites, allowing the construction of very detailed Y phylogenies that reveal male-specific aspects of the genetic history (Qamar *et al.* 2002).

Situated in the middle of the north Atlantic Ocean, the Portuguese archipelago of the Azores consists of nine islands (Figure 1) with a total population of 241,763 inhabitants. The Azores was uninhabited at the time of its discovery in the 15th century. Settlement was begun mainly by the Portuguese from the provinces of the south and north mainland Portugal, and Madeirans. Many other early settlers included the Flemish, Spanish, French, Italians, Germans, Scots, Jews, Moorish prisoners and black slaves from Guinea, Cabo Verde and São

*Corresponding author: Luisa Mota-Vieira. Molecular Genetics and Pathology Unit, Hospital of Divino Espírito Santo, 9500-370 Ponta Delgada, São Miguel Island, Azores, Portugal. Phone: +351 296 203 531 Fax: +351 296 203 090. E-mail: lmotavieira@hdes.pt

© University College London 2005

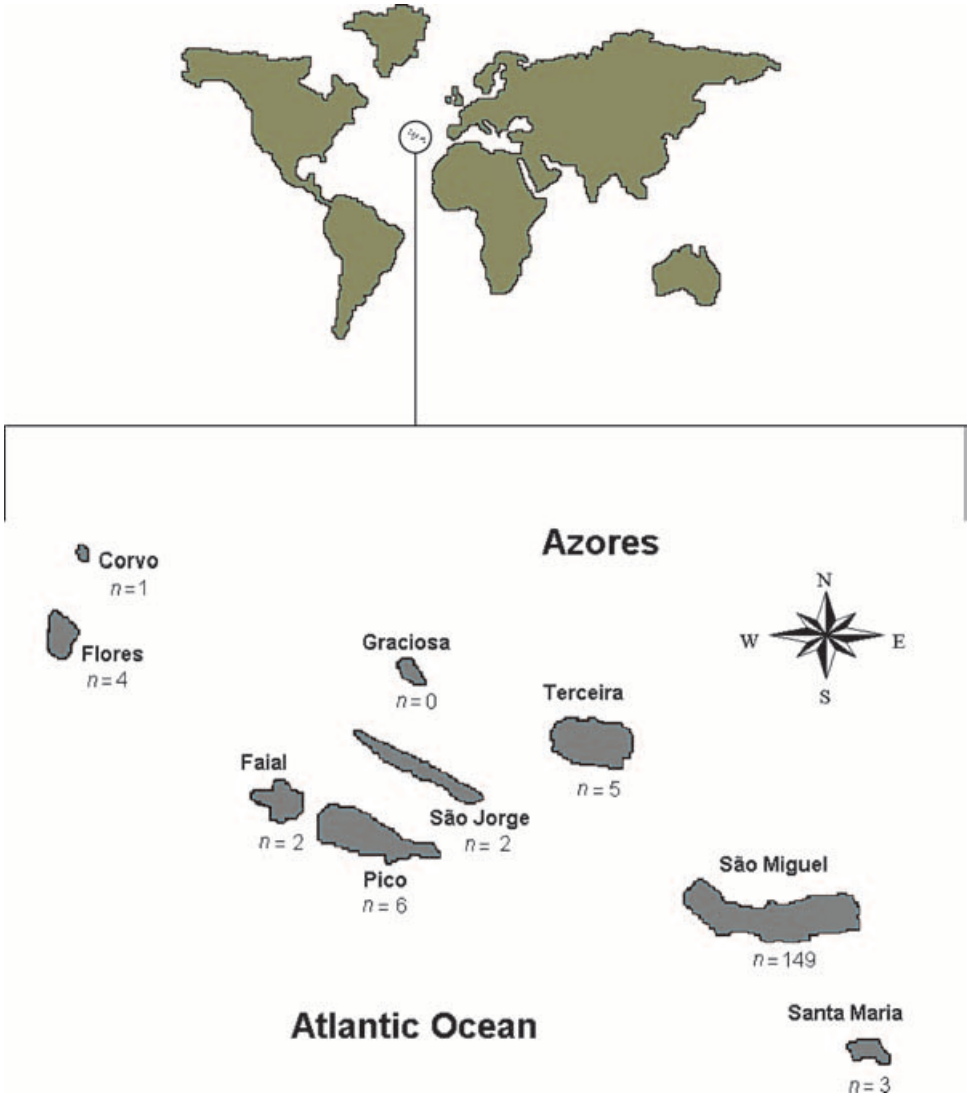

Figure 1 Geographic location of the Azores archipelago (n = number of individuals sampled). Map is not drawn to scale. The islands spread out in the area of the parallel that passes through Lisbon (39°, 43′/39°, 55′, North latitude).

Tomé. The peopling of the Azores Islands was a slow and difficult process, initiated by about 2,000 families leading to the 71,846 families of today (Guill, 1993; Portugal Census, 2001).

Here we report on the diversity of the Y-chromosome of Azorean individuals, using a combination of slowly evolving biallelic loci and rapidly evolving microsatellite loci. This allowed for an assessment of the relative diversity and phylogenetic context of the Azores Islands Y-chromosomal pool. We aimed to address the following questions: (i) how does the Y-chromosomal distribution in the Azores fit in the context of other European populations, and (ii) how did geographical isolation affect the Y-chromosomal distribution in the Azores compared to mainland Portugal.

Material and Methods

Terminology and nomenclature

The terminology and nomenclature used here are those proposed by the Y Chromosome Consortium (YCC NRY tree 2002). The terms "haplogroup" and "haplotype" are used according to de Knijff (2000).

 © University College London 2005

Population Samples

The sample set comprised 172 unrelated healthy blood donors, from the anonymous DNA bank of São Miguel, with signed informed consent (Mota-Vieira *et al*, 2003). The origin of each individual's father was used to sort the samples into: São Miguel (N = 149), Faial (N = 2), Flores (N = 4), Pico (N = 6), Santa Maria (N = 3), São Jorge (N = 2), Terceira (N = 5) and Corvo (N = 1), Figure 1. Due to a disproportionate number of samples, we combined them all into a single group: Azores. Blood samples (7.5 ml) were collected by venipuncture into EDTA tubes. DNA was extracted using the PUREGENE® kit (Gentra Systems Inc.).

PCR Amplification of Y-SNPs and Endonuclease Digestion

A total of 10 Y-biallelic markers were selected, based on the probability of their occurrence in the European populations (Rosser *et al.* 2000, and references therein). The base substitutions were as follows: 92R7 C→T; M9 C→G; SRY-2627 C→T; SRY-1532 A→G →A; sY81 A→G; SRY-8299 G→A; LLY22g C→A and Tat T→C. The LLY22g marker was typed using conditions kindly supplied by C. Tyler-Smith (personal communication). The 12f2 deletion was typed according to Rosser *et al.* (2000). Polymerase Chain Reaction (PCR) amplifications were carried out in a singleplex 20 μl reaction mixture including 1X PCR buffer, 2.5 mM $MgCl_2$, 0.1 mM dNTP mix, 1 μM of forward and reverse amplification primers, 1 U of *Taq* DNA polymerase (PROMEGA) and 40 ng of genomic DNA. PCR was carried out according to the following conditions: an initial denaturation step at 95°C for 2 min., 30 cycles of 94°C for 30 sec., 60°C for 30 sec., 72°C for 1 min. and a final extension step at 72°C for 5 min. For restriction fragment length polymorphism analysis, 1 U of the appropriate restriction enzyme in 2.5 μl of 1 X digestion buffer was added directly to 25 μl of PCR reaction and incubated at the appropriate temperature for 2 hours. Digests were analysed by electrophoresis on polyacrylamide gels (12%) and visualized by ethidium bromide. Analysis of the Y-chromosomal *Alu* repeat insertion (YAP) was carried out by PCR and analysed by agarose gel electrophoresis, as described elsewhere (Hammer & Horai, 1995).

PCR Amplification of Y-STRs

Seven microsatellite loci were typed using fluorescently labelled primers for five tetranucleotide markers (DYS389I, DYS389II, DYS390, DYS391 and DYS393), one trinucleotide repeat loci (DYS392), and one tetranucleotide repeat marker (DYS385). Primer sequences were obtained from the Y-STR haplotype database (www.ystr.org). The PCR protocol used was as follows: an initial denaturation at 95°C for 15 min. to activate HotStarTaq™ DNA polymerase (QIAGEN); 30 cycles of 94°C for 1 min., 51°C for 1 min., 72°C for 1 min. and a final 10 min. extension step at 72°C. Each 25 μl reaction contained 2 U of *Taq* DNA polymerase, 1X PCR buffer, 50 mM KCl, 4 mM $MgCl_2$, 0.25X Q Solution, 0.2 mM each of the four deoxyribonucleotide triphosphates, 0.4 μM of forward and reverse amplification primers and 50 ng of genomic DNA. An aliquot of 1 μl of each PCR product was combined with 0.5 μl CEQ™DNA size standard kit 400, 29 μl deionized formamide (Qbiogene), and run on a CEQ™8000 Genetic Analysis System (Beckman Coulter).

Statistical Analysis

Alleles are designated by their number of repeats. Since the DYS389II product contains DYS389I, we subtracted the corresponding DYS389I repeat length from that of DYS389II, to avoid double-counting the variation at DYS389I (Roewer *et al.* 1996). For DYS385, which is a duplicated Y-STR locus, the allele locus assignment was performed so that for each individual, the shorter allele was assigned to one locus (DYS385a), and the longer to another (DYS385b).

Population differentiation between the Azores and other populations was assessed using haplogroup frequencies included in the Arlequin software package (Schneider *et al.* 2000). Genetic distances, as pairwise Fst, were represented in two-dimensional space using Multi Dimensional Scaling (MDS) analysis included in the SPSS software package (version 10.0).

Results

Y-chromosome Biallelic Polymorphisms

The biallelic loci used in this study divided Azorean Y-chromosomes into twelve clades, which are usually

© University College London 2005

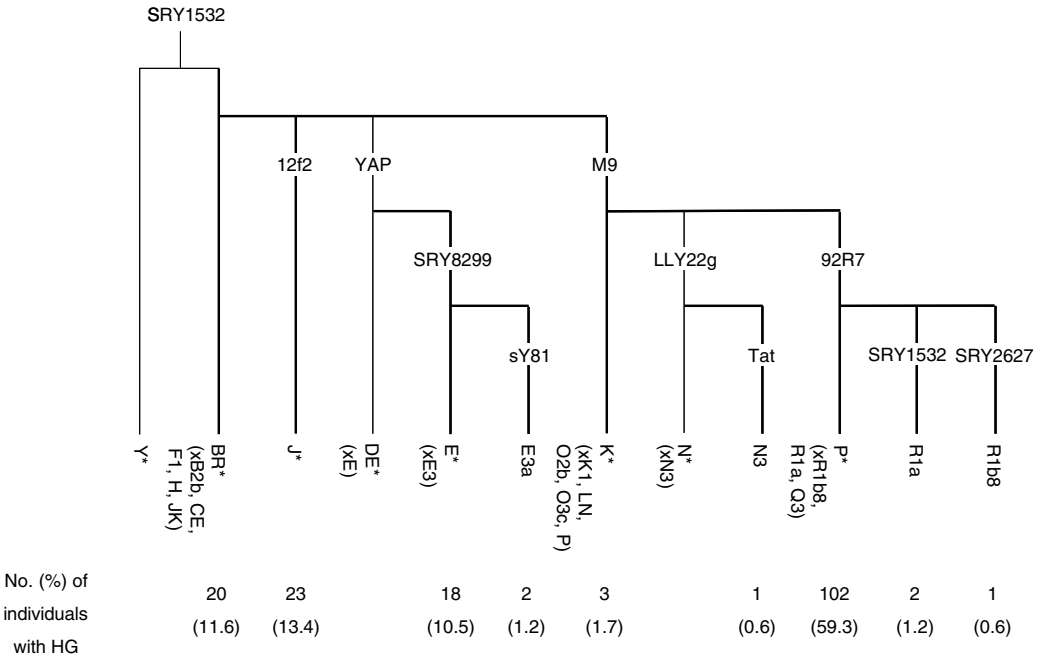

Figure 2 Phylogenetic tree of the Y-chromosome haplogroups and their percent frequencies in the Azorean sample. Bold lines indicates HG present in the Azorean population.

referred to as haplogroups (HGs). A Y-chromosomal HG tree with 10 biallelic markers and HG frequencies is shown in Figure 2. We identified 9 different HGs out of a possible 12, which indicates the degree of information from the markers selected. HG P*(×R1b8, R1a, Q3) was the most frequent, comprising 59.3% of the total sample. Interestingly, our data showed a high frequency of lineage J*, the second most frequent HG in our population, comprising 13.4% of the Y-chromosomes. Lineages BR*(xB2b, CE, F1, H, JK), 11.6%, and E*(xE3), 10.5%, are both frequent in the Azores. Lineage R1a has a frequency of 1.2%, four times higher than that described for northern and southern Portuguese populations (0.3%; Rosser *et al.* 2000). In the Azores, R1b8 accounts for 0.6% of the Y-chromosomes. Albeit at a low frequency (1.2%), we also detected the sub-Saharan HG E3a (Figure 2). In addition, lineage N3, which is primarily found in Asians, was present in the Azores at a frequency of 0.6%.

In order to test the hypothesis of random distribution of HGs among population groups, we computed Fst values using HG frequencies as implemented by Arlequin. HG frequency data for northern and southern Portuguese, Spanish, Basque, east Anglian, Belgian, French, Dutch, Bavarian, German, Sardinian, Italian, Turkish, Greek, Algerian, Canarians, Caboverdean and northern African populations were retrieved from Rosser *et al.* (2000), Flores *et al.* (2003) and Gonçalves *et al.* (2003). Population differentiation between the Azores and those populations listed above was calculated. No significant difference was observed between the Azoreans and the northern and southern Portuguese, Belgian, French or Italian samples ($p = 0.05$), suggesting no population differentiation. In contrast, comparison with the remaining populations revealed a significant difference ($p < 0.05$). The analysis of pairwise genetic distances, represented in two-dimensional space with multidimensional scaling (MDS, Figure 3), revealed that the genetic relationship among populations corresponds tightly to their relative geographical distances.

Y-chromosome STR Polymorphisms

A Y-chromosomal haplotype was constructed for each individual using seven loci (see Material and Methods). Overall, 118 different haplotypes were observed in the 172 sample set (68.6% discriminatory capacity). Haplotype diversity was high (0.9994), due to high variability of the Y-STRs. Allele frequencies and gene diversity values are listed in Table 1. The mean gene diversity (D)

 © University College London 2005

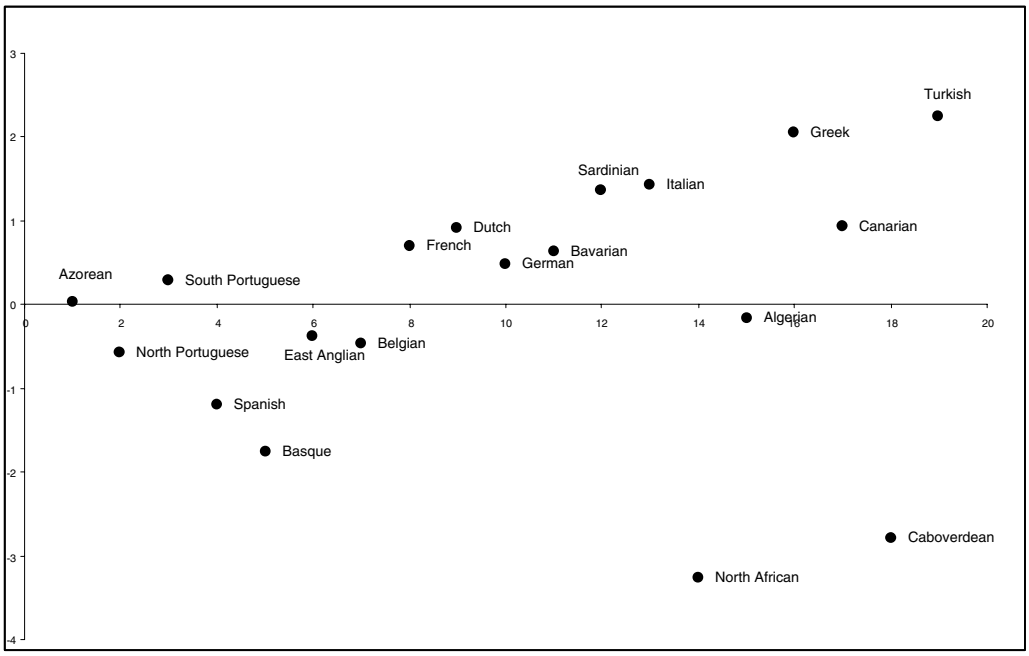

Figure 3 Genetic relationships between populations represented by Multidimensional Scaling. Note the position of the African samples that reflects the major division between the populations.

value for the loci is 0.590 (values range from 0.4592 to 0.8212, Table 1).

To investigate the separation of recently diverged populations, we performed a locus-by-locus analysis between the Azorean population and those populations we assumed to be the closest [e.g., the Madeirans (Fernandes *et al.* 2001), central Portuguese (Carvalho *et al.* 2000) and northern Portuguese (Gonzalez-Neira *et al.* 2000)], using microsatellite analysis (data not shown). Pairwise Fst showed no statistical differences ($p < 0.05$) for the DYS389II, DYS391 and DYS393 loci. However, excepting the DYS389II locus, the other loci show a statistical difference ($p < 0.05$) between the Azoreans and the central Portuguese. The comparison of Azoreans and northern Portuguese shows that difference is found only for the DYS390 locus. Taken together, these data suggest no genetic differentiation between northern Portuguese, Madeirans and Azoreans.

Y-chromosome STR Polymorphism within Haplogroups

When combining the SNPs with the STRs the number of haplotypes increased from 118 (STRs alone) to 123 (SNPs and STRs) and the discriminatory capacity raised from 68.6% to 71.5% (Table 2). The most common haplotypes were found on a P*(xR1b8, R1a, Q3) background. Haplotype H7 (13-16-24-10-13-13-11/14) occurred 10 times (5.8%), H6 (13-16-24-11-13-13-11/14) accounted for 9 individuals (5.2%) and the third most frequent haplotype, H15 (13-16-23-11-13-13-11/14), was found 6 times (3.5%). From the 172 males there were 98 unique haplotypes (56.9%).

The two most common haplotypes in the Azores, H7 (5.8%) and H6 (5.2%), are represented in the YHRD - Y Chromosome Haplotype Reference Database (www.yhrd.org) at 1.49% and 3.42%, respectively. As of July 2004, this database contains 15,545 haplotypes from 114 different European regions. Haplotype 13-16-24-11-13-13-11/14 is recorded at 3.42% in the European database, but at only 0.58% (H79) in the Azores. In addition, our data show a low frequency (17.4%) of population-specific haplotypes. Of the remaining 82.6% non-unique haplotypes, the majority are shared with the mainland Portuguese and Madeirans (51.2%), Germans (64.5%), Spanish (56.3%) and Italians (50%). High numbers of non-unique haplotypes and consequent haplotype sharing indicate a close relationship between populations (Kayser *et al.* 2001). Two haplotypes were shared by two different HG backgrounds (Table 2), one

© University College London 2005

Table 1 Allele frequencies and gene diversity values for 7 Y-chromosome STR loci in the Azorean population

Allele	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	Haplotype	DYS385
9				0.0640			9-14	0.0058
10				0.4419		0.0116	9-15	0.0058
11				0.4535	0.3605		10-14	0.0116
12	0.1395			0.0407	0.0058	0.1744	11-11	0.0116
13	0.6395				0.5581	0.7093	11-12	0.0058
14	0.2093	0.0058			0.0756	0.0988	11-13	0.0407
15	0.0116	0.0523				0.0058	11-14	0.4012
16		0.6453					11-15	0.0872
17		0.2326					12-12	0.0233
18		0.0523					12-13	0.0116
19		0.0116					12-14	0.0291
20							12-15	0.0349
21			0.0291				12-16	0.0058
22			0.0756				12-17	0.0058
23			0.3081				12-19	0.0058
24			0.5058				13-13	0.0291
25			0.0640				13-14	0.0349
26			0.0116				13-15	0.0174
27			0.0058				13-16	0.0465
							13-17	0.0291
							13-18	0.0058
							14-14	0.0349
							14-15	0.0174
							15-15	0.0058
							16-16	0.0523
							16-17	0.0058
							16-19	0.0058
							17-17	0.0116
							17-18	0.0174
h	0.5307	0.5269	0.6421	0.5968	0.5560	0.4592		0.8212
D = 0.590								

h = Gene diversity,
D = mean gene diversity.

between P*(xR1b8, R1a, Q3) and J*, and another between BR*(xB2b, CE, F1, H, JK) and E*(xE3). The presence of identical Y-chromosome STR haplotypes found on different SNP HGs is taken as evidence of recurrent mutations, which are likely to occur at STR loci.

Discussion

Prevalent Y-chromosome Lineages in Azores Islands

The non-random distribution of distinctive stable HGs provides patterns of genetic affinity and clues concerning past human movements. Here we investigated the genetic background of the male Azorean population, and discussed the results in light of existing historical records.

HG J*, defined by the 12f2 deletion, is largely confined to Caucasoid populations, with its highest frequencies being found in Middle Eastern populations. It is thought to have originated in the Middle East, where it accounts for over one third of the Y-chromosomes of Jewish, Turkish and Arab populations (Bosh *et al.* 2001; Nebel *et al.* 2001). Our data show that in the Azores this haplogroup is the second most common, with a frequency of 13.4%, twice as high as in mainland Portugal (6.8%; Rosser *et al.* 2000). Using a sampling strategy based on the three geographical groups of the Azores

 © University College London 2005

Table 2 Frequencies of Y-chromosome haplotypes by haplogroup in Azorean population

Haplogroup	H	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS385	Frequency
P* (xR1b8,R1a,Q3)	1	12	16	24	10	13	13	11-15	1
N = 102	2	12	16	24	11	13	13	12-14	1
GD = 0.9986	3	13	16	23	11	13	13	12-13	1
	4	13	16	23	10	13	13	12-14	2
	5	13	16	24	11	13	14	11-13	2
	6	13	16	24	11	13	13	11-14	9
	7	13	16	24	10	13	13	11-14	10
	8	13	16	24	10	14	13	11-14	1
	9	13	16	24	11	13	13	12-15	1
	10	13	16	24	10	14	12	11-14	1
	11	13	17	23	11	13	13	12-14	1
	12	14	16	25	11	13	13	11-15	1
	13	13	16	23	11	13	12	13-16	1
	14	13	16	23	12	13	13	11-14	1
	15	13	16	23	11	13	13	11-14	6
	16	13	16	24	11	13	13	9-14	1
	17	13	17	23	11	13	13	11-13	1
	18	13	17	23	10	13	13	11-14	1
	19	13	16	24	11	13	13	9-15	1
	20	13	16	22	11	13	13	11-14	1
	21	14	16	23	11	13	13	11-14	3
	22	13	16	24	11	13	12	11-15	1
	23	12	14	24	12	14	13	11-14	1
	24	13	16	24	11	13	12	11-14	2
	25	14	16	24	11	14	13	11-15	2
	26	12	17	24	11	14	13	11-14	1
	27	13	17	24	11	13	14	11-15	1
	28	13	17	24	10	13	14	11-15	1
	29	13	16	24	11	13	13	11-15	2
	30	14	15	24	10	13	12	11-11	1
	31	13	16	25	10	13	13	13-13	1
	32	13	17	24	11	13	14	11-14	1
	33	13	18	23	11	13	13	11-14	1
	34	13	15	24	11	13	13	11-14	1
	35	13	16	23	11	14	13	11-14	2
	36	14	16	24	11	13	12	11-14	1
	37	14	15	24	10	13	13	11-14	3
	38	12	17	24	12	13	13	11-14	1
	39	13	17	24	11	13	13	11-15	1
	40	13	16	24	11	13	14	11-14	3
	41	13	16	23	10	14	13	11-15	1
	42	13	17	24	11	13	13	11-12	1
	43	13	17	23	11	13	14	11-15	1
	44	14	16	24	11	14	13	11-14	1
	45	12	16	24	11	13	13	11-14	2
	46	12	17	24	10	13	13	11-13	1
	47	13	16	24	10	13	13	11-11	1
	48	13	16	24	10	13	14	11-14	1
	49	14	16	25	11	13	13	11-14	1
	50	13	16	23	11	13	13	12-14	1
	51	14	16	24	11	13	13	11-14	3
	52	14	17	23	11	13	13	11-14	1
	53	14	16	24	11	13	13	10-14	2
	54	14	16	24	11	13	13	11-15	1

© University College London 2005

Table 2 Continued

Haplogroup	H	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS385	Frequency
	55	13	17	24	12	13	13	11-15	1
	56	13	16	25	11	13	13	11-14	1
	57	14	15	24	11	14	12	13-14	1
	58	14	15	25	10	13	13	11-14	1
	59	15	16	25	10	13	13	11-15	1
	60	13	17	25	11	11	13	11-14	1
	61	13	16	23	12	13	12	12-15	1
	62	13	16	23	11	13	12	11-14	1
	63	13	16	24	12	13	13	11-14	1
	64	13	17	24	10	13	13	11-14	1
J*	65	13	16	26	10	11	12	13-13	2
N = 23	66	13	17	23	10	11	12	13-18	1
GD = 0.9872	67	13	15	23	10	11	13	12-12	1
	68	13	16	22	10	11	13	12-19	1
	69	12	16	24	10	11	12	12-17	1
	70	13	16	27	10	11	12	13-13	1
	71	13	16	25	9	11	12	13-17	1
	72	13	16	22	9	11	12	13-17	1
	73	13	16	23	10	11	12	13-16	2
	74	13	18	25	10	11	12	14-14	1
	75	13	16	24	10	13	14	11-14	1
	76	13	18	23	11	11	12	13-16	1
	77	14	16	23	10	11	12	14-14	1
	78	13	16	23	9	11	12	13-16	2
	79	13	16	24	11	13	13	11-14	1
	80	14	17	23	10	11	12	13-15	1
	81	13	16	25	11	11	13	12-16	1
	82	13	16	23	9	11	12	14-15	1
	83	13	17	23	10	11	12	13-17	1
	84	13	17	24	10	11	12	13-17	1
BR*(xB2b,CE,F1,H,JK)	85	12	17	22	10	11	14	14-15	1
N = 20	86	12	16	23	10	11	13	12-13	1
GD = 0.9842	87	15	15	23	10	11	13	12-12	1
	88	12	16	22	10	11	13	13-15	1
	89	12	16	22	10	11	13	13-14	2
	90	12	16	22	10	11	12	13-14	1
	91	13	16	23	10	13	14	14-15	1
	92	12	16	22	10	11	13	12-15	2
	93	14	18	22	10	11	14	13-13	1
	94	12	16	21	11	11	14	11-13	1
	95	12	16	23	10	11	13	13-15	1
	96	12	17	23	10	11	13	14-14	1
	97	12	17	21	10	11	15	13-17	1
	98	12	18	24	10	11	13	16-16	1
	99	14	16	23	10	11	13	12-12	2
	100	12	17	22	12	11	13	13-14	1
	101	13	16	23	11	12	14	15-15	1
E*(xE3)	102	14	17	21	9	11	13	12-15	1
N = 18	103	13	17	22	9	11	13	12-15	1
GD = 0.9815	104	13	17	24	10	11	13	17-18	3
	105	13	17	23	10	11	13	16-16	1
	106	13	19	24	10	11	13	16-16	2
	107	14	16	24	9	11	13	13-14	1
	108	13	16	23	9	11	13	14-14	1

 © University College London 2005

Table 2 Continued

Haplogroup	H	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS385	Frequency
	109	14	16	24	9	11	13	14-14	2
	110	13	18	24	10	11	13	16-17	1
	111	13	18	23	10	11	14	17-17	1
	112	14	17	24	11	11	13	16-16	1
	113	14	17	21	10	11	13	16-16	1
	114	13	17	23	11	11	13	16-19	1
	115	12	18	24	10	11	13	16-16	1
K*(xK1,LN,O2b,O3c,P)	116	13	17	23	10	13	10	12-16	2
N = 3	117	14	18	23	10	14	12	17-17	1
GD = 0.7278									
R1a	118	13	17	24	11	11	13	11-14	1
N = 2	119	13	17	25	11	11	13	11-14	1
GD = 0.5000									
E3a	120	14	17	21	11	11	13	16-16	1
N = 2GD = 0.5000	121	13	17	24	10	11	13	16-16	1
GD = 0.5000									
N3	122	14	16	23	11	14	13	11-13	1
R1b8	123	13	17	24	10	13	13	11-14	1

GD = genotype diversity,
H = Haplotype.

islands, Montiel and colleagues (in this issue) found lineage J at a lower frequency (8.6%) for the whole archipelago, although their study revealed a similar frequency (14.5%) for the islands of the Central group. The high frequency of lineage J raises the question of whether early Jewish settlers left a significant imprint in the genetic pool of the Azorean male population.

The overall northwest (NW) African contribution to the Iberian Y-chromosome pool has been calculated as 7%, with the highest level of contribution (14%) being found in Andalusians from southern Iberia (Bosch *et al.* 2001), a result that is consistent with the population movement associated with Islamic rule in Iberia (Pereira *et al.* 2000). The frequency of the NW African lineage E*(xE3) in mainland Portugal and the Azores (11.7% and 10.5%, respectively) is similar. Montiel and colleagues (in this issue) also found comparable values (13.0%) for the archipelago. The results obtained by us and Montiel *at al.* suggest several hypotheses for the presence of this lineage in the present-day population of the Azores: a direct input of Moorish prisoners, the influence of early Portuguese settlers, or a contribution of both Moorish prisoners and Portuguese.

Lineage E3a, defined by mutation sY81, shows a sub-Saharan distribution pattern. This HG is the most frequent in west African populations, and its presence can be interpreted as resulting from sub-Saharan gene flow. The occurrence of lineage E3a in the Azores is the result of African influence, since it has been detected neither in Europe, nor in Iberian samples (Semino *et al.* 2000; Bosch *et al.* 2001; Rosser *et al.* 2000). The presence of sub-Saharan African slaves in the archipelago since the beginning of its settlement is well documented (Matos 1989). Therefore, we conclude that the 1.2% Y-chromosomes with the E3a background represent the male descendants of black slaves from Guinea, Cabo Verde and São Tomé.

Lineage N3, defined by a Tat biallelic polymorphism, is specific to Asians and northern Europeans, and has not been found in the Iberian peninsula or in other European countries (Rosser *et al.* 2000; Helgason *et al.* 2000). This mutation probably arose in the Mongolia/China area, and its present distribution stretches from Japan to Norway (Zerjal *et al.* 1997). The presence of this lineage in the Azores (0.6%) is intriguing. Historical records of the presence of Asians or Mongolians in the archipelago are not known, but Bruges-Armas and colleagues (1999) have recently described the presence of Mongolian HLA genes at a high frequency in the Terceira Island population (Azores). Thus, it is possible that the N3 Lineage

© University College London 2005

may have been introduced during the expansion of trade navigation between Europe, America and Asia, in the XVI and XVII centuries, when the Azores had a strategic role due to its geographic position (Russel-Wood, 1998).

Lineage R1b8, defined by a C→T base substitution at the SRY-2627, arose recently in Iberia. This lineage has its highest frequency in Basques (11%) and Catalans (22%), but in other regions these chromosomes are rare or absent (Hurles *et al.* 1999). In the Azores, its frequency is marginal (0.6%), probably reflecting the descendants of Spaniards who came to the islands during the reign of Spain over Portugal, from 1580 to 1640 (Matos, 1989).

Lineage R1a is most frequent in central eastern Europe, comprising approximately half of the chromosomes in Russian, Polish and Slovakian samples. In contrast, frequencies in southeast and southwest Europe are low. In our sample set, R1a is four times more common than in mainland Portugal (Rosser *et al.* 2000), which may be explained by the following reasons: (i) this chromosome only arrived with Portuguese settlers, and subsequently increased in frequency; (ii) some chromosomes came with Portuguese settlers, while others came directly from central eastern Europeans; and (iii) they are an exclusive contribution from central eastern Europe. Records and papers exploring historical settlement show that some Europeans (e.g. Flemish) contributed to the peopling of the Azores, so we believe that all of the above hypotheses are possible.

Variability of Y-chromosome STRs in Azores Islands

Comparisons of allelic frequencies between our sample set and those found in central mainland Portugal (Carvalho *et al.* 2000) show differences. Indeed, historical records show that the first Portuguese settlers were mainly from north and south Portugal. The mean gene diversity value across loci in the Azorean sample ($D = 0.590$) is higher than the value reported for northern Portugal ($D = 0.517$), from which Azoreans are believed to be partially derived (Guill, 1993). It is also higher than that observed for Europeans ($D = 0.503$). Likewise, haplotype diversity value in the Azores (0.9994) is higher than in northern Portugal (0.980) and Europe (0.985). We conclude that the diversity found in Azorean Y-chromosomes is derived from the admixture of Portuguese with other populations.

One advantage of Y-chromosome markers compared with mtDNA is that Y-chromosome polymorphisms seem to show a higher degree of population specificity (Seielstad *et al.* 1998), making them more informative for tracing population relationships. Comparisons between the paternal Y-chromosomes (present study and Montiel *et al.* in this issue) and the maternal mtDNA (Santos *et al.* 2003) show some evidence of differential sex-specific influences. Here, the paternal Middle East influence was estimated at 13.4%, which is higher than the 7.5% obtained by Santos *et al.* (2003). Another difference was the smaller contribution from Africans. We estimated a clear African Y-chromosome contribution of 1.2%, whereas they identify an 11.3% contribution of African mtDNA. The Y-chromosome and mtDNA results are, in general, concordant; they both indicate the same history for the peopling of the Azores and suggest that there was some gender differentiation in the population pathways.

Concluding Remarks

The presence of HGs that have a widespread distribution in Europe, in combination with others of clear sub-Saharan, Asian and Middle Eastern origin, reflects the diverse patterns defining the extant Azorean Y-chromosomal pool. We conclude that the current paternal Y-chromosomal pool in the Azores is of Portuguese descent, with a considerable contribution from individuals from multiple origins.

Acknowledgements

We are grateful to all the blood donors who agreed to participate in the present study and to the Hematology Department of Hospital do Divino Espírito Santo, explicitly Maria L. Almeida, Júlio Carvalho, Marta Loura and Catarina Matos. We also would like to express our gratitude to Karyne Hyde and Patrice Eydoux for helpful comments on the manuscript. We also thank Mónica Carvalho and Maria Conceição Vide for their helpful discussions. This work was supported by a grant from the Direcção Regional da Ciência

 © University College London 2005

e Tecnologia, Azores, Portugal. C.C.B. is founded by a grant (ref. SFRH/BD/12254/2003) from Fundação para a Ciência e a Tecnologia – Operational Programs Science Technology and Innovation (POCTI), Portugal.

References

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.

Bruges-Armas, J., Martinez-Laso, J., Martins, B., Allende, L., Gomez-Casado, E., Longas, J., Varela, P., Castro, M. J. & Arnaiz-Villena, A. (1999) HLA in the Azores Archipelago: possible presence of Mongoloid genes. *Tissue Antigens* **54**, 349–359.

Carvalho, M., Anjos, M. J., Andrade, L., Coxinho, C., Corte-Real, F., Gamero, J. J., Vieira, D. N. & Vide, M. C. (2000) Y Chromosome polymorphisms: a comparison between Azores and Continental Portuguese sample. In: *Progress in Forensic Genetics 8*. (eds. Sensabaugh, G. F., Lincoln, P. J. & Olaisen, B.) pp. 302–303. Elsevier Science.

de Knijff, P. (2000) Y chromosomes shared by descent or by state. In: *Archaeogenetics: DNA and the population prehistory of Europe*. (eds. Renfrew, C. & Boyle, K.), pp 301–304. Cambridge: The McDonald Institute.

Fernandes, A. T., Brehm, A., Gusmão, L. & Amorim, A. (2001) Y-chromosome STR haplotypes in the Madeira archipelago population. *Forensic Sci Int* **122**, 178–180.

Flores, C., Maca-Meyer, N., Perez, J.A., Gonzalez, A. M., Larruga, J. M. & Cabrera, V. M. (2003) A predominant European ancestry of paternal lineages from Canary Islanders. *Ann Hum Genet* **67**, 138–152.

Gonçalves, R., Rosa, A., Freitas, A., Fernandes, A., Kivisild, T., Villems, R. & Brehm, A. (2003) Y-chromosome lineages in Cabo Verde Islands witness the diverse geographic origin of its first male settlers. *Hum Genet* **113**, 467–472.

Gonzalez-Neira, A., Gusmão, L., Brion, M., Lareu, M. V., Amorim, A. & Carracedo, A. (2000) Distribution of Y-chromosome STR defined haplotypes in Iberia. *Forensic Sci Int* **110**, 117–126.

Guill, J. H. (1993) *A history of the Azores islands*, Vol. 5. California: Division of Golden Shield International Publications, Cooperation.

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

Helgason, A., Siguroardóttir, S., Nicholson, J., Sykes, B., Hill, E. W., Bradley, D. G., Bosnes, V., Gulcher, J. R., Ward, R. & Stefansson, K. (2000) Estimating Scandinavian and Gaelic ancestry in the male settlers of Iceland. *Am J Hum Genet* **67**, 697–717.

Hurles, M. E., Veitia, R., Arroyo, E., Armenteros, M., Bertranpetit, J., Perez-Lezaun, A., Bosch, E., Shlumukova, M., Cambon-Thomsen, A., McElreavey, K., Lopez De Munain, A., Rohl, A., Wilson, I. J., Singh, L., Pandya, A., Santos, F. R., Tyler-Smith, C. & Jobling, M. A. (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.

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

Kayser, M., Krawczak, M., Excoffier, L., Dieltjes, P., Corach, D., Pascali, V., Gehrig, C., Bernini, L. F., Jespersen, J., Bakker, E., Roewer, L. & de Knijff, P. (2001) An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. *Am J Hum Genet* **68**, 990–1018.

Matos, A. (1989) Povoamento e colonização dos Açores. In: *Portugal no Mundo*. (eds. Albuquerque, L.) pp. 176–188. Lisboa: Publicações Alfa.

Montiel, R., Bettencourt, C., Silva, C., Santos, C., Prata, M. J. & Lima, M. (2004) Analysis of Y-chromosome variability and its comparison with mtDNA variability reveals different demographic histories between islands in the Azores Archipelago (Portugal). *Ann Hum Genet* (in this issue).

Mota-Vieira, L., Pacheco, P. R., Almeida, M.L., Cabral, R., Carvalho, J., Costa, S., Rego, P., Branco, C. C., Loura, M., Peixoto, B.R., Araujo, A. L. & Mendonça, P. (2004) Human DNA bank in São Miguel Island, Azores: Assembly and analysis. *Eur J Hum Genet* **12th**, (Suppl 1) P0995.

Nebel, A., Filon, D., Brinkmann, B., Majumder, P. P., Faerman, M. & Oppenheim, A. (2001) The Y chromosome pool of Jews as part of the genetic landscape of the Middle East. *Am J Hum Genet* **69**, 1095–1112.

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

Qamar, R., Ayub, Q., Mohyuddin, A., Helgason, A., Mazhar, K., Mansoor, A., Zerjal, T., Tyler-Smith, C. & Mehdi, S. Q. (2002) Y-chromosomal DNA variation in Pakistan. *Am J Hum Genet* **70**, 1107–1124.

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.,

© University College London 2005

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.

Roewer, L., Kayser, M., Dieltjes, P., Nagy, M., Bakker, E., Krawczak, M. & 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.

Russel-Wood, A. J. R. (1998) A disseminação das gentes. In: *História da expansão portuguesa: a formação do Império*. Vol. 1. (eds. Bettencourt, F. & Chaudhuri, K.) Navarra: Circulo de Leitores.

Santos, C., Lima, M., Montiel, R., Angles, N., Pires, L., Abade, A. & Aluja, M. P. (2003) Genetic structure and origin of peopling in the Azores islands (Portugal): the view from mtDNA. *Ann Hum Genet* **67**, 433–456.

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

Seielstad, M. T., Minch, E. & Cavalli-Sforza, L. L. (1998) Genetic evidence for a higher female migration rate in humans. *Nature* **20**, 278–280.

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.

SPSS: Statistical Package for Social Sciences. http://www.spss.com

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

Zerjal, T., Dashnyam, B., Pandya, A., Kayser, M., Roewer, L., Santos,Schiefenhovel, W., Fretwell, N., Jobling, M. A., Harihara, S., Shimizu, K., Semjidmaa, D., Sajantila, A., Salo, P., Crawford, M. H., Ginter, E. K., Evgrafov, O. V. & Tyler-Smith, C. (1997) Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. *Am J Hum Genet* **60**, 1174–1183.

Received: 19 April 2004
Accepted: 30 September 2004

 © University College London 2005