Genetic Structure and Origin of Peopling in The Azores Islands (Portugal): The View from mtDNA

C. Santos[1,2,*], M. Lima[3], R. Montiel[3,1], N. Angles[1], L. Pires[2], A. Abade[2] and Ma. P. Aluja[1]

[1]*Unity of Anthropology, Department BABVE, Faculty of Sciences, Autonomous University of Barcelona, 08193 Bellaterra (Barcelona), Spain*

[2]*Department of Anthropology, University of Coimbra, 3000 Coimbra, Portugal*

[3]*CIRN and Department of Biology, University of the Azores, 9500 Ponta Delgada, S. Miguel, Azores, Portugal*

Summary

The Azores islands (Portugal), uninhabited when discovered by Portuguese navigators in the fifteenth century, are located in the Atlantic Ocean 1500 km from the European mainland. The archipelago is formed by nine islands of volcanic origin that define three geographical groups: Eastern (S. Miguel and Sta. Maria), Central (Terceira, Faial, Pico, Graciosa and S. Jorge) and Western (Flores and Corvo). To improve the genetic characterisation of the Azorean population, and to clarify some aspects related to the history of settlement, a study of mtDNA was conducted in the population of the archipelago. The HVRI region was sequenced and specific RFLPs were screened in 146 samples obtained from unrelated individuals with Azorean ancestry (50 from the Eastern group, 60 from the Central group, and 37 from the Western group). Samples were classified into haplogroups based on the information obtained from both sequencing and RFLP analysis.

All the analyses performed support the idea that, in the whole group of islands, the majority of mtDNA lineages originated from the Iberian Peninsula, mainly from Portugal (mainland). However contributions from other European populations, especially from Northern Europe, cannot be disregarded. The values obtained for the various diversity parameters in the Azores archipelago indicate that the Azorean population, as a whole, does not exhibit the typical characteristics of an isolated population. The analysis of genetic data by groups of islands showed that the Western group exhibited particular features. The distribution of haplogroups in the Western group is very atypical, being significantly different from what is observed in the Eastern and Central groups. Furthermore, the diversity values are, in general, lower than those observed in other populations used for comparison. African haplogroups were found in all the groups of islands. Therefore the presence of *Moorish* and *African* slaves on the islands, as reported in historical sources, is supported by the mtDNA genetic data, especially in the Eastern group. The presence of Jews in the Central group is also supported by the mtDNA data. Neither historical nor genetic data (phylogeography of mtDNA) supports the idea of a differential settlement history for the Western group; however, it is represented in the phylogenies as an isolated branch. The effect of genetic drift, induced by the reduced population size since peopling occurred, has led to a very atypical distribution of haplogroups/haplotypes in this group of islands.

We cannot ignore the influence of biodemographic and genetic processes, namely founder effect, genetic drift, migration, and even recent mutational events in the mtDNA lineages of the Azorean populations. Nevertheless, a great part of the variation in the Azorean mtDNA can be explained by the settlement history.

*Corresponding author: Cristina Santos, Unity of Anthropology, Department BABVE, Faculty of Sciences, Autonomous University of Barcelona, 08193 Bellaterra (Barcelona), Spain, Fax: +34 935811321, e-mail: Cristina.Santos@uab.es

Introduction

The Azores archipelago (Portugal), located in the Atlantic Ocean, 1500 km from the European mainland, is formed by nine islands of volcanic origin whose relative positions define three geographical groups: Eastern (S. Miguel and Sta. Maria), Central (Terceira, Faial, Pico, Graciosa and S. Jorge) and Western (Flores and Corvo). With an area of 2344 km^2, the Azores presently has a total population of 237315 inhabitants, distributed in a very asymmetric way among the three groups of islands (134885 in the Eastern group, 98101 in the Central group, and only 4329 in the Western group) (INE, 2001).

Historical data indicates that the islands were uninhabited when discovered by Portuguese navigators in the fifteenth century. The peopling of the islands was a slow and difficult process, and was initiated in 1439 in the islands of Sta. Maria and S. Miguel; the small islands of Flores and Corvo were only inhabited towards the end of 16th century (Matos, 1989; Mendonça, 1996). According to historical records, the first settlers came mainly from various regions of mainland Portugal and from Madeira Island. However, people of different origin, such as Spanish, French, Italian, English, German and Flemish (whose important presence is always referred to in accounts of the peopling of the Central group), also made up part of the early settlers (Matos, 1989; Mendonça, 1996). There is clear evidence that Jews also contributed to the peopling of the archipelago. Sephardim Jews expelled from the Iberian Peninsula may have used the Azores as a refuge. The presence of *African* and *Moorish* slaves in the islands is also referred to in historical documents (Matos, 1989; Mendonça, 1996), being well-documented especially for the Western group (Gomes, 1997). However, the specific contribution of these settlers to the different islands is not clear, due to the lack of both historical and genetic data.

The complex process of peopling, the evolution of island populations, always affected by migration along with geographic isolation, has been the main stimulus for carrying out biodemographic and genetic studies on the Azorean population. To improve the genetic characterisation of this population and enhance our knowledge about several aspects related to its origin and dynamics, a research project that includes the analysis of nuclear autosomal markers, Y chromosomal and mitochondrial DNA (mtDNA), is presently being undertaken by our group. Here we present the results of the sequencing of HVRI, and specific RFLP screening of mtDNA. Our main goals were: (i) to describe the variation of mtDNA in the Azorean population; (ii) to gather data in order to contribute to the establishment of the genetic profile of the archipelago; (iii) to shed more light on some aspects related to the history and dynamics of the settlement of the Azores.

Material and Methods

Samples

One hundred and forty-six unrelated individuals, born in the Azores and with Azorean ancestry, were sampled, including 42 previously tested by Pires (1999). The birthplace of the earliest known maternal ancestor was chosen as the sample location. Fifty individuals were originally from the Eastern Group, 60 from the Central Group, and 36 from the Western Group. The samples were taken by buccal swabs and existing samples collected by Pires (1999) by venupuncture. Voluntary donors were asked for informed consent and to fill in an anonymous inquiry concerning the birthplaces of all their known ancestors. The comparative data set consisted of 1493 mtDNA HVRI sequences from the populations listed in Table 1. Populations were selected, when possible, in order to include groups that are mentioned in historical documents as sources of settlers for the Azores islands.

DNA Extraction

Total DNA from buccal cells was extracted using Instagene matrix (BioRad) according to the manufacturer's specifications. DNA from blood samples was previously extracted by Pires (1999) using standard protocols.

MtDNA Analysis

RFLP Analysis

Relevant restriction sites for mtDNA haplogroup determination (Table 2) were selected from those

Table 1 Population (corresponding codes in Figure 2), sample size, references and source of the information used in this study

Population	N	Reference	Source
Azores – Eastern Group	50	This work	This study
Azores – Central Group	60	This work, Pires, 1999	This study, Provided by the author
Azores – Western Group	36	This work	This study
Mainland North Portugal (Pt)	100	Pereira *et al.* 2000	GenBank
Mainland Central Portugal (Pt)	82	Pereira *et al.* 2000	GenBank
Mainland South Portugal (Pt)	59	Pereira *et al.* 2000	GenBank
Spain (Mainland) (Sp)	89	Côrte-Real *et al.* 1996; Pinto *et al.* 1996	HVRbase (Handt *et al.* 1998)[1]
Galicia (Sp)	92	Salas *et al.* 1998	HVRbase (Handt *et al.* 1998)[1]
Basque Country (Sp)	106	Bertranpetit *et al.* 1995; Côrte-Real *et al.* 1996	HVRbase (Handt *et al.* 1998)[1]
France (Fr)	50	Rousselet & Mangin, 1998	HVRbase (Handt *et al.* 1998)[1]
UK (UK)	100	Piercy *et al.* 1993	HVRbase (Handt *et al.* 1998)[1]
Germany (Ger)	67	Hofmann *et al.* 1997	HVRbase (Handt *et al.* 1998)[1]
Austria (Aus)	101	Parson *et al.* 1998	HVRbase (Handt *et al.* 1998)[1]
Italy (It)	49	Francalacci *et al.* 1996	HVRbase (Handt *et al.* 1998)[1]
Israel Arabs/Palestinians (Isr)	117	Di Rienzo & Wilson, 1991; Richards *et al.* 2000	Database of Richards *et al.* 2000[2]
Algeria (Alg)	85	Côrte-Real *et al.* 1996	HVRbase (Handt *et al.* 1998)[1]
Moroccan Jews (MJ)	115	Thomas *et al.* 2002	Provided by the author
Morocco (Berber) (MBer)	60	Rando *et al.* 1998	Provided by the author
Morocco (Moroccans) (Mor)	32	Rando *et al.* 1998	Provided by the author
Senegal (Sen)	119	Graven *et al.* 1995 Rando *et al.* 1998	HVRbase (Handt *et al.* 1998)[1] Provided by the author
Mauritania (Maur)	30	Rando *et al.* 1998	Provided by the author
S. Tome (ST)	50	Mateu *et al.* 1997	Article
Mozambique (Moz)	109	Pereira *et al.* 2001	Article

[1]http://www.hvrbase.de
[2]http://www.stats.ox.ac.uk/∼macaulay/founders2000/tableA.html

proposed by Torroni *et al.* (1993, 1994, 1996), Torroni & Wallace (1994), Chen *et al.* (1995), Macaulay *et al.* (1999), Quintana-Murci *et al.* (1999) and Saillard *et al.* (2000). Some of the specific primers and PCR conditions for each polymorphic site were available from the literature, but others were designed in our laboratory (Montiel, 2001; Santos, 2001).

Six μl of PCR product were used in restriction enzyme digestions and the enzymatic cleavage was performed according to the manufacturer's specifications. Results of restriction analyses were resolved through electrophoresis in agarose gels (2%).

HVRI Amplification and Sequencing

Sequencing was performed for all samples, except those previously sequenced by Pires (1999). A fragment of 443 bp from the HVRI, located between positions 15978 and 16420 (numbered according to Anderson *et al.* 1981), was amplified using primers and PCR conditions described by Pereira *et al.* (2000). PCR products were purified using the QIAquick PCR purification kit (Quiagen). Sequence reactions were carried out using the kit ABI PRISM BigDye Terminator version 3 and run in an ABI3100 sequencer (Perkin Elmer).

Data Analysis

Sequences were aligned and manually checked using BioEdit (Hall, 1999), and all polymorphic positions were confirmed in chromatograms. Sequences without ambiguities were obtained between positions 16050 and 16399. The classification of all the samples by assigning them to haplogroups was performed

Table 2 Positions analysed to define each haplogroup

Position	1715 *Dde*I	3592 *Hpa*I	4216 *Nla*III	4577 *Nla*III	7025 *Alu*I	8994 *Hae*III	9052 *Hae*II	10032 *Alu*I	10394 *Dde*I	10397 *Alu* I	10871 *Mnl*I	11719 *Sma*I	12308 HinfI	13704 *Bst* OI	14465 *Aci*I	14766 *Mse*I	15606 *Alu*I	12705 (C/T)[1]
PreHV			–	+	+	+			–		+	+	–		–	+		
HV or pre*V				+	+				–		+					–		
H					–				–		+					–		
V				–	+				–		+					–		
U			–				+		–		+		+			+		
K			–				–		+		+		+			+		
J			+						+		+			–		+		
T			+						–		+					+		
I	–		–			+		+	+		+		–		+	+		
X			–			+			–		+		–		+	+		
W			–			–			–		+		–			+		
M		–							+	+	–							
L1 or L2		+							+		–							
L3 (?)		–							+/–	–	–							
N*			–			+			+/–		+	–	–		–	+		T
R*			–			+			+/–		+	–	–		–	+		C

[1]The region that encompasses the position 12705 was sequenced using primers L12553 and H13127

using the combined information of RFLPs and HVRI motifs, according to the nomenclature summarized in Richards *et al.* (2000), available at: http://www.stats.ox.ac.uk/∼macaulay/founder2000/motif.html (Macaulay, 2000). Furthermore, to classify African sequences, a comparison with published data (Pereira *et al.*, 2001) was performed. In order to maximise the number of sequences available for population comparison, all analyses based on sequences were restricted to the 302 nucleotides between positions 16069 and 16370.

The haplogroup distribution in the three groups of islands was used to perform an exact test of population differentiation, as described in Raymond & Rousset (1995). The exact test was also used to compare frequencies of each haplogroup in the three groups of islands.

For each population, using the HVRI sequence, standard and molecular diversity indices were estimated using the Arlequin 2.000 software (Schneider *et al.* 2000); namely, gene diversity ($\hat{H}$) (Nei, 1987), number of different haplotypes (K), number of polymorphic sites (S), mean number of pairwise differences (θ_π) (Tajima, 1983), nucleotide diversity (π) (Tajima, 1983; Nei, 1987), and the theta estimator based on the number of different haplotypes (θ_k) (Ewens, 1972). For the major haplogroups found in the Azores (and for comparison in mainland Portugal), $\hat{H}$ (Nei, 1987), θ_π (Tajima, 1983), π (Tajima, 1983; Nei, 1987) and the average number of sites differing between a set of sequences and a specified common ancestor (ρ) (Forster *et al.* 1996) were estimated.

Phylogenetic networks (Bandelt *et al.* 1999) among haplotypes were constructed using the program Network 3.1 (www.fluxus-engineering.com). Positions of HVRI were weighted as proposed by Richards *et al.* (1998).

Genetic distance matrices, according to Reynolds *et al.* (1983), were computed in Arlequin 2.000 software (Schneider *et al.* 2000) and were used as in-files in the Neighbour program from the Phylip ver 3.57c package (Felsenstein, 1995), to construct Neighbour-Joining trees (Saitou & Nei, 1987). Neighbour-Joining trees were drawn with TreeView (Page, 1996).

The analysis of non-parametric correlations between the percentage of private lineages and θ_k was performed using the statistical package SPSS 9.0.0 (SPSS inc., 1989–1999).

Results and Discussion

Sequencing and RFLP Results

Data obtained for each sample from HVRI sequencing and RFLP analysis, as well as the haplogroup determined from the RFLP and HVRI motifs, are presented in the appendix. It was possible, using the combined information provided for both systems of mtDNA analysis, to classify all samples into one of the haplogroups described in the literature. This work confirms the previously established correlation between RFLPs and HVRI data (Torroni *et al.* 1996; Francalacci *et al.* 1999; Macaulay *et al.* 1999), and supports the idea that a correct and complete classification of mtDNA is only possible when both methods of analysis are used simultaneously.

Haplogroup Frequencies

The haplogroup frequencies observed in the total sample of the Azores, as well as the partial frequencies observed in the Eastern, Central and Western groups, are summarised in Table 3. The 146 samples were distributed among 31 different haplogroups/sub-haplogroups. Distinct numbers of clusters were found in the Eastern (21), Central (19) and Western (11) groups. The comparison of the haplogroup distribution in the three groups of islands, using the exact test of population differentiation, revealed that there is a similarity between the Eastern and Central groups ($p = 0.29955$). The Western group showed significant differences when compared with the other two groups (Western/Eastern: $p < 0.00001$; Western/Central: $p = 0.00135$).

A comparison of the frequencies of each haplogroup in the three groups revealed that haplogroups J, U and V contribute to significantly differentiate the three groups of islands, and largely to differentiate the Western group (significant differences: J-Western/Central/Eastern: $p = 0.04084$, Western/Eastern: $p = 0.038$, Western/Central: $p = 0.023$; U-Western/Central/Eastern: $p = 0.00056$, Eastern/Central: $p < 0.0001$; V-Western/Central/Eastern: $p < 0.0001$, Western/Eastern: $p < 0.0001$, Western/Central: $p < 0.0001$).

In the total sample of the Azores, all European haplogroups (Torroni *et al.* 1996) were found, as well as the African clusters U6, M1, L1, L2 and L3 (Rando *et al.* 1998; Quintana-Murci *et al.* 1999; Chen *et al.* 1995, 2000; Watson *et al.* 1997) and the Near Eastern N1b cluster (Richards *et al.* 2000). This result provides evidence for a mixed composition of mtDNAs in the Azores islands, that is supported by historical data referring to the contribution of individuals from multiple origins to the peopling of the archipelago.

The analysis by groups (Table 3) revealed that in the Central group, all the main European haplogroups (Torroni *et al.* 1996) were found. A very low frequency of haplogroup U was observed when compared to values observed in European populations. The frequency of J is closer to that observed in North European populations. The frequencies of the remaining haplogroups were similar to those observed in European populations. African (M1, L1) and Near Eastern (N1b) sequences were found in low frequencies in the Central group.

In the Eastern Group, and when considering European haplogroups, only I is absent. The most interesting features are the low frequency of haplogroup H (30%), with respect to European populations, and the higher number of sub-clusters of haplogroup U relative to the Western and Central groups of the archipelago. In the Eastern group, high frequencies of non-European sequences were observed, with members of the U6, M1, L2 and L3 African clusters and N1b Near Eastern cluster detected.

The Western group exhibited a very atypical haplogroup distribution, where the I, J and W haplogroups are absent. Haplogroup U (excluding K) is underrepresented, and the few samples belonging to this cluster are members of the U2 and U6b sub-haplogroups, which are uncommon in European populations but more frequent in Near Eastern and North African populations, respectively (Macaulay *et al.* 1999). The most frequent haplogroup is not H, as it is in European populations, but V, the frequency of which was higher than 33%. In all the results published so far, frequencies of V higher than 10% were only found in Catalonia (12.5%) (Montiel *et al.* 2001), the Basque Country (12.4%) and the Skolt Saami (52%) (Torroni *et al.* 2001). Members of the East African M1 cluster were also found in this group of islands.

Table 3 mtDNA haplogroup distribution (number of individuals and percentage) in the total sample of the Azores, and in Eastern, Central and Western groups

	Azores (N = 146)		Eastern Group (N = 50)		Central Group (N = 60)		Western Group (N = 36)	
Haplogroup	f_I	%	f_i	%	f_i	%	f_i	%
H	47	32.19	15	30	24	40	8	22.22
I	3	2.05			3	5		
J	8	5.48	5	10	3	5		
J1	1	0.68	1	2				
J1a	2	1.37	1	2	1	1.67		
J1b1	3	2.05			3	5		
J2	1	0.68			1	1.67		
L1c	1	0.68			1	1.67		
L2	2	1.37	2	4				
L3e2b	1	0.68	1	2				
M1	4	2.74	1	2	1	1.67	2	5.56
N1b	2	1.37	1	2	1	1.67		
T	3	2.05			3	5		
T1	3	2.05	1	2			2	5.56
T2	9	6.16	3	6	4	6.67	2	5.56
T3	6	4.11	1	2	2	3.33	3	8.33
U								
U2	1	0.68					1	2.78
U3	2	1.37	2	4				
U5	2	1.37	2	4				
U5a/b	3	2.05	3	6				
U5a1a	2	1.37	1	2	1	1.67		
U5b	1	0.68	1	2				
U6a	3	2.05	3	6				
U6b	2	1.37					2	5.56
K	8	5.48	3	6	4	6.67	1	2.78
pre-HV	1	0.68			1	1.67		
pre*V	2	1.37					2	5.56
V	15	10.27	1	2	2	3.33	12	33.33
W	3	2.05	1	2	2	3.33		
X	4	2.74	1	2	2	3.33	1	2.78
R*	1	0.68			1	1.67		

HVRI Diversity

Table 4 presents diversity parameters for the Azorean population and populations cited in Table 1, obtained using HVRI sequences between positions 16069–16370. In the total Azorean sample, 80 different HVRI lineages, characterised by 74 polymorphic sites, were observed. The highest values of diversity estimators were obtained for the Central and Eastern group, with similar values, and the lowest values were observed in the Western group. Compared to other populations, the values for mean numbers of pairwise differences (θ_π) and nucleotide diversity (π) were higher than the majority of European populations in the Azores sample (total or by groups of islands). A median value of gene diversity was observed in the Azores. However, the value registered for the Western group appears to be one of the lowest when the three groups of islands were analysed separately. The values of diversity decreased in the three groups of islands when African sequences (L1, L2, L3 and M1) were excluded (Table 4).

The analysis of diversity by haplogroups in the Azorean sample compared to mainland Portugal (Table 5) showed high values of diversity for the haplogroups U and J, while the diversity of the remaining

Table 4 mtDNA HVRI diversity (from nucleotide positions 16069 to 16370) in the Azorean populations including and excluding African sequences (L1, L2, L3 and M1), and in populations used for comparison

Population	N	*K* (% *K*)	*S* (% *S*)	transi/transv/ indel	$\hat{H} \pm$ sd	$\theta\pi \pm$ sd	$\pi \pm$ sd	θk	Private lineages (%)
Azores	146	80 (54.79)	74 (24.50)	64/13/2	0.9603 ± 0.0107	5.4339 ± 2.6322	0.0179 ± 0.0096	71.7723	41 (51.25)
Eastern Group	50	33 (66.00)	50 (16.56)	41/8/2	0.9486 ± 0.0236	5.3168 ± 2.6102	0.0175 ± 0.0095	41.1428	17 (51.52)
Central Group	60	40 (66.67)	52 (17.22)	47/6/0	0.9565 ± 0.0190	5.4937 ± 2.6799	0.0182 ± 0.0098	51.2801	16 (40.00)
Western Group	36	16 (44.44)	34 (11.26)	30/4/1	0.9159 ± 0.0276	5.1745 ± 2.5649	0.0171 ± 0.0094	10.4725	6 (37.50)
Mainland Portugal	241	142 (58.92)	94 (31.13)	89/12/0	0.9588 ± 0.0096	5.1369 ± 2.4982	0.0170 ± 0.0092	144.2037	84 (59.15)
North Portugal	100	67 (67.00)	69 (22.85)	64/7/0	0.9533 ± 0.0162	5.1406 ± 2.5120	0.0170 ± 0.0092	87.7000	35 (52.22)
Central Portugal	82	61 (74.39)	63 (20.86)	59/7/0	0.9765 ± 0.0107	5.2758 ± 2.5753	0.0175 ± 0.0094	106.3182	26 (42.62)
South Portugal	59	41 (69.49)	54 (17.88)	50/4/0	0.9433 ± 0.0246	4.9621 ± 2.4488	0.0164 ± 0.0090	58.2382	18 (43.90)
Spain	89	71 (79.78)	68 (22.52)	65/5/0	0.9837 ± 0.0081	5.4974 ± 2.6693	0.0182 ± 0.0098	160.6716	37 (52.11)
Galicia	92	51 (55.43)	55 (18.21)	53/3/1	0.9233 ± 0.0239	3.3215 ± 1.7218	0.0110 ± 0.0063	46.3011	21 (41.18)
Basque Country	106	53 (50.00)	53 (17.55)	49/5/0	0.9364 ± 0.0181	3.1821 ± 1.6587	0.0105 ± 0.0060	41.5205	23 (43.40)
France	50	42 (84.00)	52 (17.22)	47/7/0	0.9878 ± 0.0086	4.85720 ± 2.4092	0.0161 ± 0.0089	121.0537	22 (52.38)
UK	100	68 (68.00)	67 (22.19)	62/5/2	0.9752 ± 0.0092	4.7785 ± 2.3549	0.0157 ± 0.0086	92.2527	36 (52.94)
Germany	67	52 (77.61)	56 (18.54)	52/3/1	0.9724 ± 0.0139	4.1728 ± 2.1004	0.0138 ± 0.0077	104.8412	30 (57.69)
Austria	101	64 (63.37)	72 (23.84)	68/7/0	0.9505 ± 0.0170	4.9565 ± 2.4320	0.0164 ± 0.0089	74.0609	36 (56.25)
Italy	49	40 (81.63)	54 (17.88)	52/2/0	0.9685 ± 0.0189	5.5298 ± 2.7042	0.0183 ± 0.0099	99.3853	28 (70.00)
Israel	117	101 (86.32)	93 (30.79)	90/12/3	0.9945 ± 0.0031	6.6033 ± 3.1412	0.0219 ± 0.0115	348.3041	66 (65.35)
Algeria	85	30 (35.29)	37 (12.25)	37/0/0	0.9431 ± 0.0103	5.2725 ± 2.5729	0.0175 ± 0.0094	16.0942	15 (50.00)
Moroccan Jews	115	45 (39.13)	51 (16.89)	49/3/0	0.9077 ± 0.0224	4.0584 ± 2.0398	0.0134 ± 0.0075	26.7323	25 (55.56)
Morocco (Berber)	60	38 (63.33)	50 (16.56)	46/5/1	0.9633 ± 0.0147	5.0957 ± 2.5064	0.0169 ± 0.0092	43.5081	18 (47.37)
Morocco (Moroccans)	32	29 (90.63)	47 (15.56)	44/3/0	0.9879 ± 0.0141	6.8612 ± 3.3158	0.0227 ± 0.0122	144.6747	12 (41.38)
Mauritania	30	23 (76.67)	31 (10.26)	31/0/0	0.9747 ± 0.0172	6.8674 ± 3.3245	0.0227 ± 0.0123	43.3056	8 (34.78)
Senegal	240	124 (51.67)	80 (26.49)	77/6/1	0.9858 ± 0.0025	7.3792 ± 3.4633	0.0249 ± 0.0129	102.4841	104 (83.87)
S. Tome	50	32 (64.00)	50 (16.56)	46/6/0	0.9731 ± 0.0110	9.3350 ± 4.3606	0.0309 ± 0.0160	37.2763	21 (65.62)
Mozambique	109	50 (45.87)	62 (20.53)	58/7/2	0.9611 ± 0.0080	9.8708 ± 4.5518	0.0327 ± 0.0167	35.1729	43 (86.00)
Excluding African sequences (L1, L2, L3 and M1)									
Azores	138	73 (52.90)	67 (22.19)	59/10/2	0.9557 ± 0.0012	4.9439 ± 2.4211	0.0163 ± 0.0088	62.0844	33 (45.21)
Eastern Group	46	29 (60.04)	43 (14.24)	36/5/2	0.9391 ± 0.0274	4.9383 ± 2.4483	0.0162 ± 0.0089	32.6527	13 (44.83)
Central Group	58	38 (65.52)	45 (14.90)	43/3/0	0.9534 ± 0.0202	4.9690 ± 2.4524	0.0165 ± 0.0090	46.7723	14 (36.84)
Western Group	34	15 (44.12)	32 (10.60)	28/4/1	0.9073 ± 0.0304	4.5447 ± 2.2907	0.0150 ± 0.0084	9.7030	5 (33.33)

N-sample size; *K*-number of different haplotypes; *S*-number of polymorphic sites; $\hat{H}$-gene diversity; θ_π-mean number of pairwise differences; π-nucleotide diversity; θ_k-theta estimator based on the number of different haplotypes; Private lineages are defined as those found only in one of the populations included in the present study

Table 5 Diversity within major haplogroups in the Azores (Az) and Mainland Portugal (Pt)

Haplogroups	N		K		$\hat{H} \pm$ sd		$\theta_\pi \pm$ sd		$\pi \pm$ sd		ρ	
	Az	Pt	Az	Pt	Az	Pt	Az	Pt	Az	Pt	Az	Pt
U (including K)	24	54	19	39	0.9819 ± 0.0164	0.9762 ± 0.0113	5.5696 ± 2.7719	5.3857 ± 2.6370	0.018382 ± 0.010198	0.017834 ± 0.009687	27/24 = 1.125	59/54 = 1.093
J	15	17	10	10	0.9333 ± 0.0449	0.9265 ± 0.0417	3.9738 ± 2.1058	3.3984 ± 1,8290	0.013158 ± 0.007818	0.011253 ± 0.006778	14/15 = 0.933	15/17 = 0.882
T	21	26	11	16	0.9286 ± 0.0306	0.9323 ± 0.0337	4.3704 ± 2.2489	4.6208 ± 2.3424	0.014472 ± 0.008312	0.015301 ± 0.008641	12/21 = 0.571	22/26 = 0.846
H	47	99	19	44	0.6966 ± 0.0768	0.8023 ± 0.0431	1.4868 ± 0.9145	1.9230 ± 1,1031	0.004891 ± 0.003339	0.006368 ± 0.004046	20/47 = 0.426	43/99 = 0.434
V	15	17	3	10	0.5905 ± 0.0771	0.7941 ± 0.1035	0.6632 ± 0.5394	1.6799 ± 1.0353	0.002196 ± 0.002003	0.005562 ± 0.003836	2/15 = 0.133	12/17 = 0.706
L1+L2+L3+M1	8	18	7	17	0.9643 ± 0.0772	0.9935 ± 0.0210	8.5883 ± 4.4445	9.1597 ± 4.4202	0.028344 ± 0.016709	0.030330 ± 0.016367	17/8 = 2.125	38/18 = 2.110

N-sample size; $\hat{H}$-gene diversity; θ_π-mean number of pairwise differences; π-nucleotide diversity. Major haplogroups are defined as those with a total frequency higher than 10%; ρ-average number of sites differing between a set of sequences and a specified common ancestor (between positions 16051–16399)

haplogroups is lower in the Azores. In the Azorean sample, H and V haplogroups present the lowest values of diversity. The African haplogroups have high values of θ_π, π and ρ, followed by the U haplogroup. However, in the Azores, the highest value for gene diversity is displayed by haplogroup U.

Relating gene diversity and nucleotide diversity in the three groups of islands (Table 4) to intra-haplogroup variation (Table 5), it is possible to infer that the higher nucleotide diversity observed in the Azores, as compared to European populations, is mainly the result of the presence of African sequences in the three groups of islands. Nevertheless, the influence of the variability of some specific haplogroups, especially U, also has some influence on the increase in both nucleotide and gene diversity.

According to Helgason *et al.* (2000), the theta estimator based on the number of different haplotypes (θ_k) is likely to provide a more reliable evaluation of current and historical female effective-population size than θ_π. Of all the populations studied, the Western group presents the lowest value of θ_k, and the Central and Eastern groups also presented relatively low values (Table 4). This indicates that the Azores, and especially the Western group, have recently had a relatively small female effective-population size.

In general, the proportion of private lineages sampled from geographical proximate populations should increase as a function of θ_k, as this parameter reflects the probability of new lineages arising by mutation (Helgason *et al.* 2001). High values of private lineages associated with low values of θ_k, found in some populations used in the comparison such as Mozambique, are the result of the non-inclusion of a sufficient number of geographically related populations in the comparative database to detect gene flow. Considering the analysis of European populations alone, a significant positive correlation is observed between the percentage of private lineages and the values of θ_k ($r_{sp} = 0.596$; $p = 0.009$). The relationship between the percentage of private lineages and the values of θ_k is presented in Figure 1, and varies greatly among the three groups of islands. The Eastern group presents an excess of private lineages in contrast to the Central group, which tends to a scarcity of private lineages. The western group, and the total sample of the Azores, have a

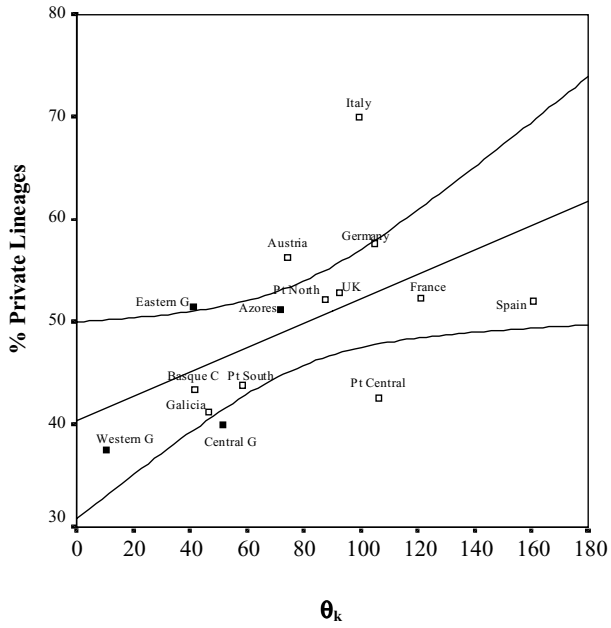

Figure 1 Scatterplot of θ_k values and percentage of private lineages. The least-squares regression line is shown. The curved lines show the 95% confidence region around the regression line.

proportion of private lineages that can be expected by their θ_k.

The excess of private lineages observed in the Eastern group could have different causes. An intuitive explanation is that the isolation of the islands can hinder the migratory flow of new lineages to and from neighbouring populations. However, six of the private lineages found in the Eastern group belong to African clusters (U6a, M1, L2 and L3). Consequently, it is more likely that a sufficient number of African populations was not included in the comparative database.

Within the Azores, the Central group shares the highest number of lineages with the other two groups of islands. Therefore, the relative scarcity of private lineages observed in the Central group could be indicative of either a very high level of emigration (where few lineages remain private for long, because of rapid outward gene flow to neighbouring populations), or of immigration (where new lineages arriving into the population would increase θ_k but not the proportion of private lineages).

The values of diversity parameters, reported for the archipelago, allow us to infer that the Azores population as a whole does not present the typical characteristics of an isolated population. Thus, the values recorded are similar to those observed in the majority of European populations used in the comparative analysis, and higher than those observed in isolated populations like the Basque country and Galicia. Furthermore, neither an excess nor a scarcity of private lineages is observed, in relation to the respective θ_k value, suggesting that values of diversity in the Azores are adjusted to the effective population size of females.

These results are in accordance with studies of autosomal STRs (Lima, personal communication), classical markers (Amorim *et al.* 1979) and those reported in biodemographic studies, which indicate that the Azorean archipelago, when considered as a whole, does not exhibit levels of endogamy and consanguinity higher than those reported for some populations from mainland Portugal (Cunha, 1987; Lima, 1991; Lima & Soares, 1992; Smith *et al.* 1992).

When analyses were carried out by groups of islands, the Western group appears to have different features. The haplogroup distribution is very atypical, diversity values are in general lower than those observed in the other groups of islands, and the value of gene diversity is one of the lowest registered in all of the populations used for comparison. However, the relationship between the percentage of private lineages and values of θ_k indicate that this group of islands does not undergo a higher level of isolation than that observed in the other groups of islands. This leads to the conclusion that the observed haplogroup distribution and lower diversity are strictly related to the reduced number of females present in the island since peopling, and makes the roles of founder effect and genetic drift in small populations evident.

Phylogeography of Azores mtDNA

Figure 2 (a, b, c, d, e) represents phylogenetic networks of haplotypes found in the Azores and populations used for comparison. To simplify networks, haplotypes absent in the Azores were only represented if they were phylogenetically related to haplotypes present in the Azorean population. Examining the networks, from the 80 different haplotypes (based on HVRI information) found in the Azores, only six are present in more than one group of islands. Two haplotypes are present in the three groups, three are common to the Central and Eastern

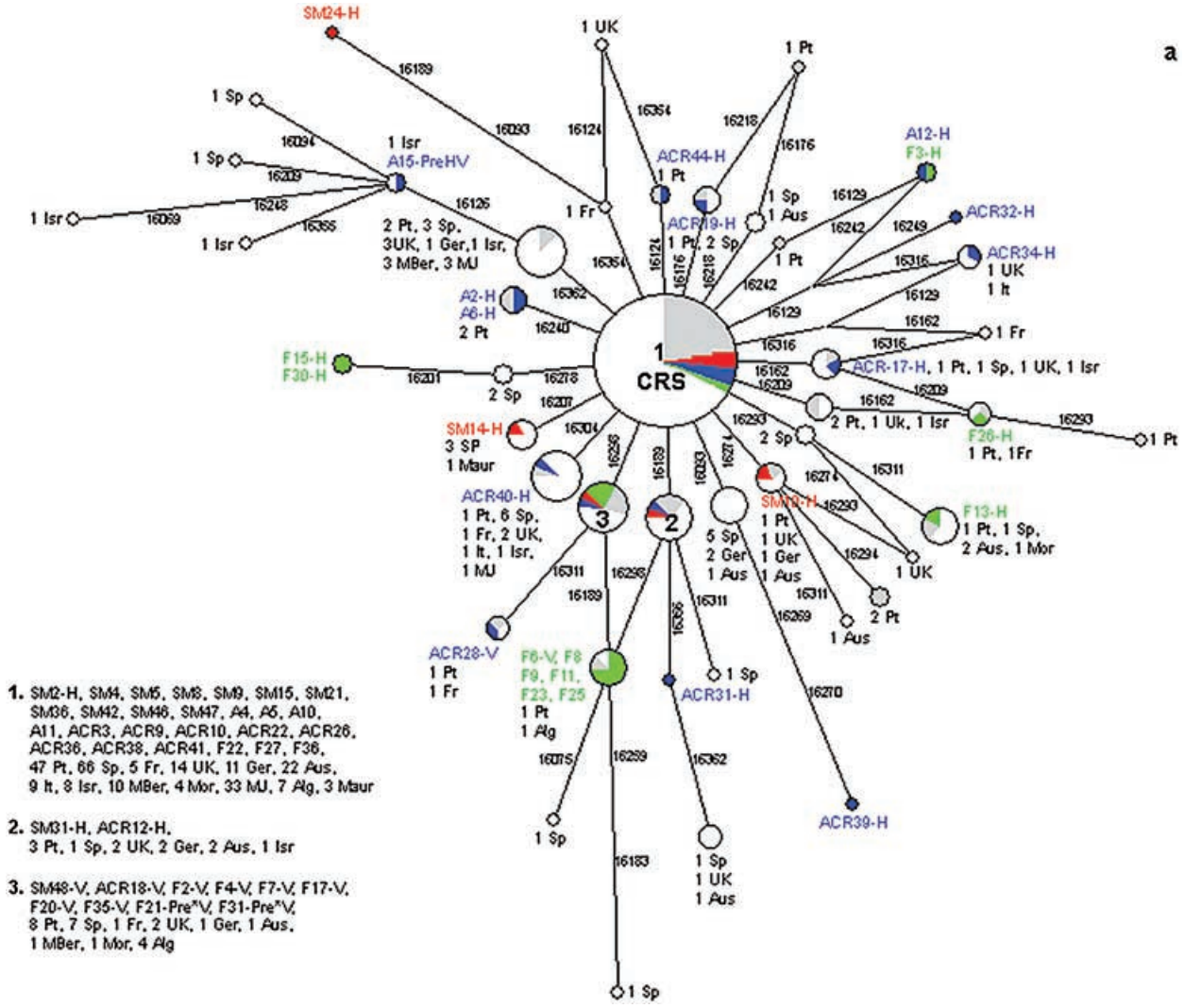

Figure 2 Phylogenetic Networks of (a) H, V, HV, Pre-HV, Pre*V; (b) U (including K); (c) J, T, JT; (d) N, I, W, X; (e) L1, L2, L3, M. Circle sizes are proportional to the haplotype frequency. (Red - Eastern group; Blue - Central group; Green - Western group; Gray - Mainland Portugal; White - other populations).

groups, and one to the Central and Western groups. The Central group is the one that shares the highest number of haplotypes with the other two. This result indicates limited migratory movements among the three groups of islands, until at least the beginning of the 19th century (since the geographical location of the sample was based on the origin of the great-grand-parents of the donor).

Forty-one haplotypes (represented in the networks as single colour circles: red to the Eastern group, blue to the Central group and green to the Western group) appear to be exclusive to the Azores, and thirty-nine are shared with other populations. By analysing the shared haplotypes, thirty-three were found to be common to the Iberian Peninsula (mainly to mainland Portugal). However, only eleven are shared exclusively with the Iberian Peninsula, the remaining twenty-two also being present in one or more of the other populations. Specifically, sixteen are found in other European populations, ten in Israel, five in Moroccan Jews, eleven in North Africa and four in West Africa. Considering that these haplotypes belong to European haplogroups (H, J, T U, K, V and X) (Torroni *et al.* 1996) their most probable origins are the Iberian Peninsula or other European populations. However, a possible North African origin cannot be excluded for some sequences. The work of

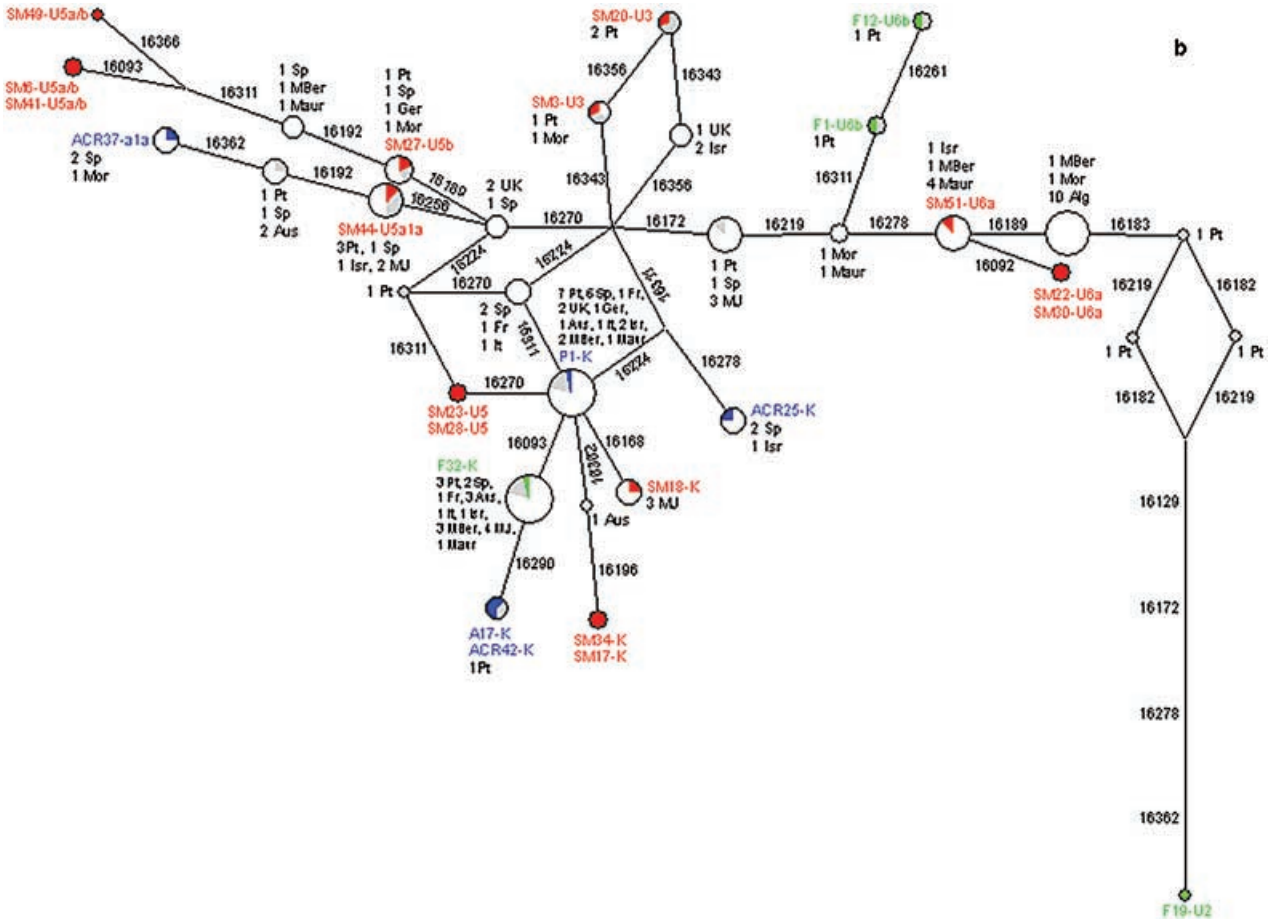

Figure 2 (*continued*)

Rando *et al.* (1998) attests to the fact that there is strong evidence for some degree of European genetic input into North Africa. There is, for example, evidence that haplogroups U5 (Richards *et al.* 1998) and V (Torroni *et al.* 1998) are represented in North Africa. Only one haplotype (ACR34-H) is shared exclusively with non-Iberian European populations (Fig. 2a).

Haplotypes represented by A15-pre-HV (Fig. 2a) and SM11-N1b (Fig. 2d) are shared exclusively with the Israeli population, and SM18-K (Fig. 2b) with Moroccan Jews. Taking into account the fact that the Israel Arab/Palestinian sample represents an extant population living in the region to which Jews trace their origin (Thomas *et al.* 2002), and that Moroccan Jews probably included descendants of the Sephardim Jews that were expelled from the Iberian Peninsula, the three haplotypes (represented by A15, SM11 and SM18) previously mentioned could indicate the presence of Sephardim Jews in the Azores. The haplotype represented by SM51 (Fig. 2b), belonging to sub-haplogroup U6a, is shared with Israeli and North African populations. This haplotype is probably originally from North Africa, since the U6 haplogroup has been described as a North African cluster (Rando *et al.* 1998).

U6 sequences are represented in both the Eastern and Western groups; however, in the Western group (samples F1, F12) they belong to sub-cluster U6b, which, according to Plaza *et al.* (2001), is restricted to the Iberian Peninsula. In the Eastern group (samples SM22, SM30 and SM51) they are assigned to U6a, found in both the Iberian Peninsula and North Africa (Plaza *et al.* 2001; Rando *et al.* 1998). Thus, only lineages found in the Eastern group could plausibly constitute evidence of the influence of *Moorish* slaves, which is in accordance with historical data.

The haplotype represented by A16 (Fig. 2d), a member of the European haplogroup I, is present in non-Iberian European populations. In S. Tome, the same

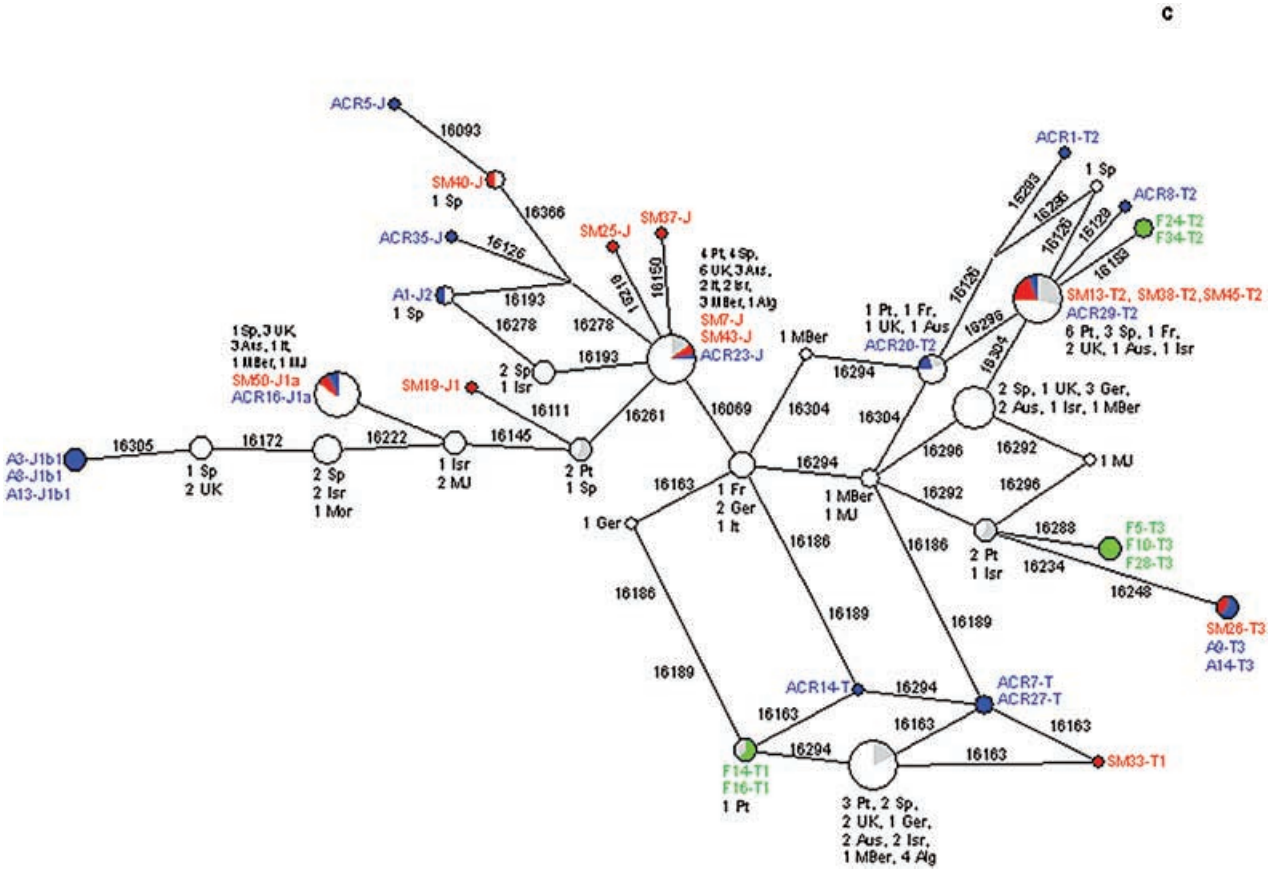

Figure 2 (*continued*)

haplotype was found (Mateu *et al*. 1997); however that sample probably belongs to haplogroup L3 (Calafell, Comas & Mateu, personal communication).

In summary, the analysis of shared haplotypes indicates that the majority of haplotypes is shared with Iberian Peninsula populations, principally with mainland Portugal. A small influence from other European (especially North European), Near East/Jewish and North African populations can also be confirmed.

The 41 private lineages found could have been introduced into the Azores both by European and by African individuals belonging to populations that are not included in the comparison, for example, populations from Flanders and Madeira Island, for which data are not available. Alternatively, as suggested by Helgason *et al*. (2000), they may arise from populations included in the analysis in which these lineages have not yet been sampled, or have been lost since the inhabitation of the Azores.

On the other hand, the hypothesis of the origin of some private lineages in recent mutational events occurring after the settlement of the islands cannot be disregarded.

The possible origin of the non-shared haplotypes, to be discussed later, was inferred using the partial networks presented in Figure 2 and two mtDNA databases, namely the database of Richards *et al*. (2000) and the HVRbase (Handt *et al*. 1998), both available online.

A worldwide search for the 41 private haplotypes found in the Azores was performed in mtDNA databases, and eight matches with samples that were not included in the comparison were found. Haplotypes represented by ACR31-H (Fig. 2a), F24-T (Fig. 2c), ACR30-W (Fig. 2d) and SM49-U5a/b (Fig. 2b) are shared with Northern European populations that were not included in the analysis; ACR8-T2 (Fig. 2c), SM12-X (Fig. 2d) and ACR7-T (Fig. 2c) match with one sequence each, found, respectively, in Rome (Italy), Romany and Armenia; SM22 (Fig. 2b), which

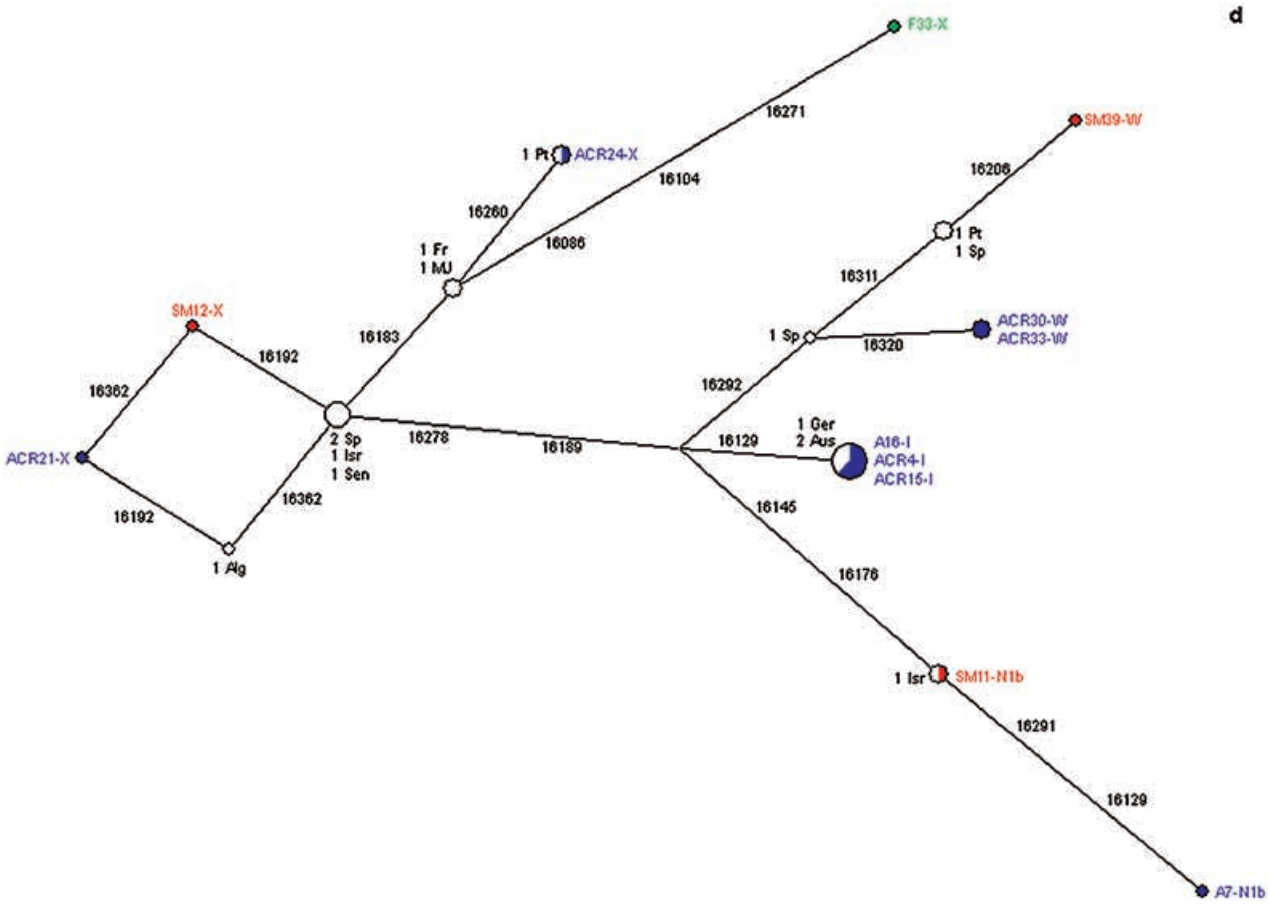

Figure 2 (*continued*)

belongs to the North African cluster U6a (Rando *et al.* 1998), is not shared with populations used for comparison, but was found in the Canary Islands (Rando *et al.* 1999) and in Moroccan Berbers (Anglés, personal communication).

Haplotypes classified as J can provide valuable information for the inference of the origin of the Azorean population. This haplogroup is present at high frequencies in Northern European populations, and some specific clusters, such as J1b1, can be assigned to a very restricted geographic area. Haplogroup J is absent in the Western group, but its frequency is close to 14% in both the Eastern and Central groups, higher than that observed in the Iberian Peninsula and similar to that observed in Northern European populations. Among the ten J haplotypes found in the Azores, six are private lineages (Fig. 2c).

The haplotype represented by ACR5 could have been derived from the haplotype 16069, 16126, 16278, 16366, found in mainland Portugal and in the Eastern group (SM40). SM19 could have been derived from the haplotype 16069, 16126, 16261, found in mainland Portugal. Haplotypes represented by SM25 and SM37 could be derived from haplotype 16069, 16126 found in the Azores (SM7, SM43 and ACR23) and at high frequencies in Europe, especially in United Kingdom. Haplotype A3 was possibly derived from the haplotype 16069, 16126, 16145, 16172, 16222, 16261, found in the United Kingdom and Spain.

Nevertheless, other origins can be postulated for the previously mentioned J haplotypes. Importantly, the influence of the population from the region of Flanders that was not included in the analysis is always mentioned in the peopling of the Central group. It is expected that this population, like the surrounding Northern European populations, would have high frequencies of haplogroup J, and some of the non shared haplotypes found in the Azores could match with those which,

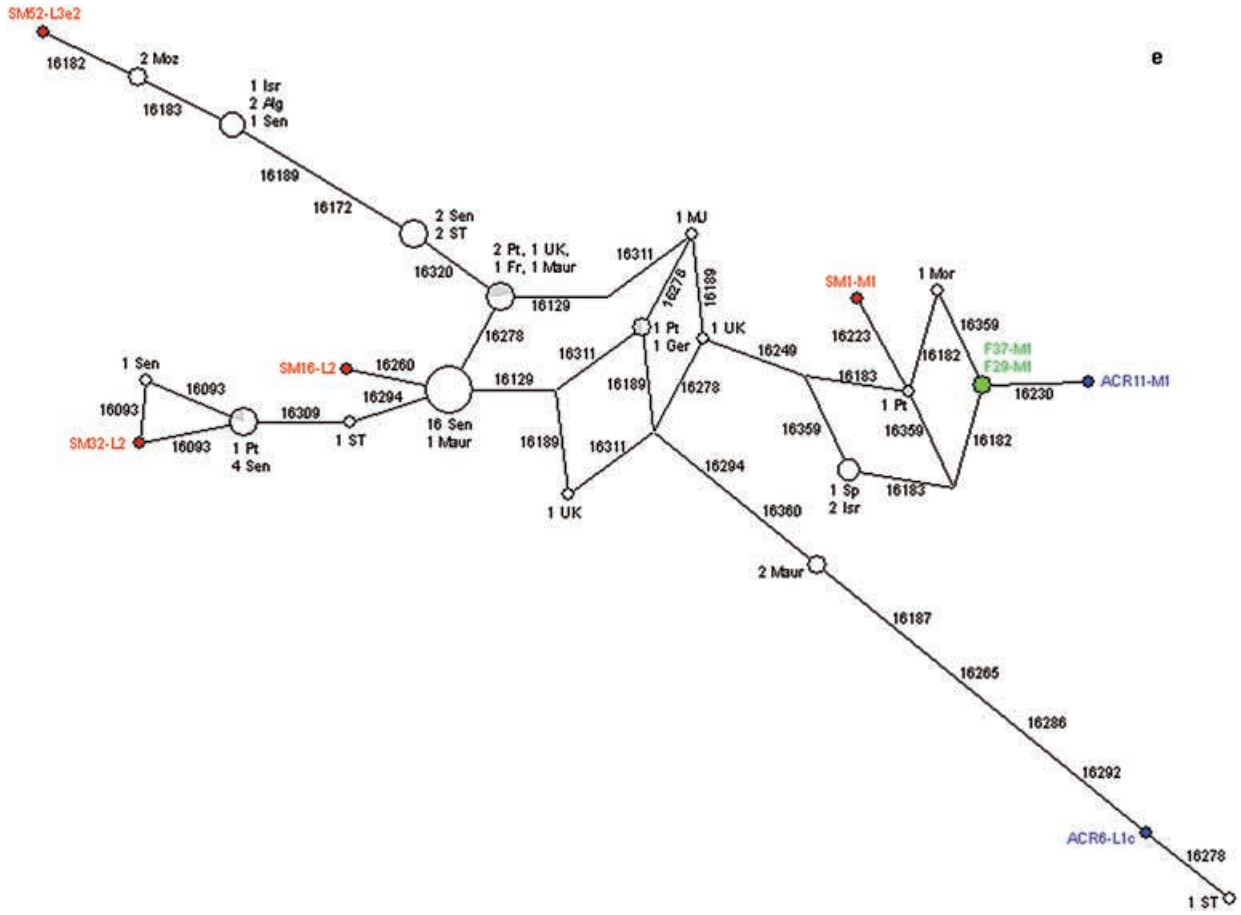

Figure 2 (*continued*)

hypothetically, would be found if the mtDNA of Flemish populations was analysed.

East African M1 sequences (SM1, ACR11, F29 and F37) (Fig. 2e) do not match any sequence present in the mtDNA databases, probably because African populations are less studied. The most similar sequence that was found is a North African sequence with the motif 16129, 16182C, 16183C, 16189, 16223, 16249, 16311. All the M1 sequences found in the Azores have a very similar motif, and probably trace their origin to the same geographical region in East Africa. With respect to L1, L2 and L3 sequences (Fig. 2e), they are located in the network near to sequences from various sub-Saharan populations and from mainland Portugal African sequences. This indicates a common origin for both African sequences present in mainland Portugal and the Azores islands, which is consistent with the historical context of the 16th century, when Portugal was actively involved in the slave trade (Thomas, 1998).

The analysis of non-shared haplotypes indicates that the majority of the European private lineages have their probable origin in Northern European and Near Eastern populations. Evidence of influences from both North Africa and sub-Saharan Africa were found in the Azores; however, a strict geographic origin for African sequences is difficult to define.

The terminal position in the networks of several Azorean haplotypes, and their strict relation with other Azorean haplotypes, seem to indicate that some of the private lineages could have originated in the Azores.

Estimates of Admixture

To estimate the proportion of admixture, the information concerning the possible origin of both shared and non-shared haplotypes was summarised. Three major groups were defined for the origin of mtDNA in the Azores (Europe, Near East and Jews, and Africa).

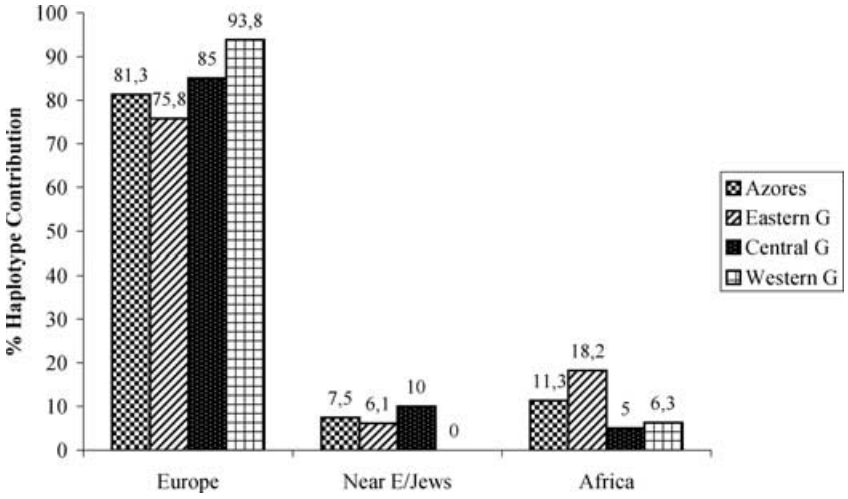

Figure 3 Bar chart showing the pattern of haplotype contribution from parental populations to the Azores.

Whenever a haplotype is shared with a European population, it is admitted that its most probable origin is Europe. Thus, it is possible to infer that probably: a minimum of 11.25% of the haplotypes found in the Azores are originally from African populations; a minimum of 7.5% are derived from Near Eastern and Jewish populations; and a maximum of 81.25% are from European populations.

Figure 3 illustrates the observation that mtDNA gene pools of all the groups of islands have their major contribution from European populations. However, there are some differences among the three groups of islands. The Eastern group exhibits the highest contribution from non-European populations (close to 25%), mainly African (18.2%), revealing the influence of slaves in the peopling of this group of islands.

In the Central group, the non-European influence is approximately 15%, 10% from Near Eastern/Jewish populations, and 5% from African populations. The Western group exhibits the lowest contribution of non-European lineages (6.25%), which is exclusively African. According to historical information the contribution of slaves to the peopling of the islands, mainly to the Eastern and Western groups, was very important, being well-documented for the Western group (for review see Gomes, 1997). One of the headmen of Flores island was also leader of Santo Antão island (Cabo Verde), and a great number of the slaves who arrived at the Western group were originally from Cabo Verde. Surprisingly, no L1, L2 or L3 mtDNA lineages were found in the Western group. This indicates that factors such as differential mortality may have affected the fitness of slaves, limiting the transmission of their genes until now.

Genetic Distances and Phylogenetic Reconstruction

To examine the relationship between all pairs of populations used in the comparison, genetic distances according to Reynolds *et al.* (1983) were calculated, and used to build a phylogenetic tree (not shown). In contrast to European populations, African populations present higher genetic distances among themselves. This justifies the existence of an identifiable gradient observed in the N-J tree from sub-Saharan to North African populations, and the aggregation of the European populations in one extreme of the tree. It is impossible to establish any relation among European populations. Furthermore, the three groups of islands are located in this assemblage.

To try to clarify the position of the Azores with respect to other populations, various N-J trees were built. The position of the various populations was very similar in all the trees. In Figure 4, a tree that excludes Algerian and Western and Eastern African populations is shown, where the African populations that show less genetic distance relative to the Azores are maintained.

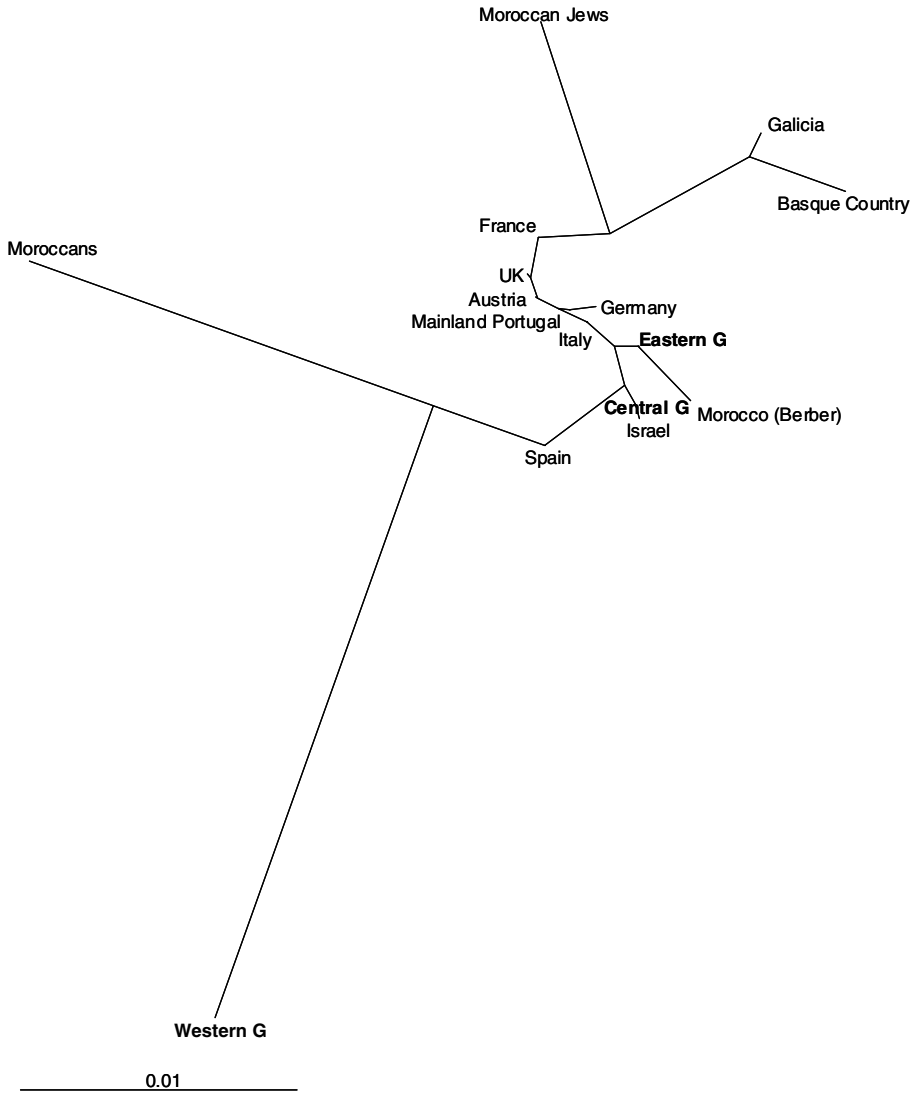

Figure 4 Neighbour-Joining tree (Saitou & Nei, 1987) constructed from Reynolds *et al.* (1983) genetic distances.

The analysis of Figure 4 reveals that the geographically and/or culturally isolated populations, such as those of the Basque Country, Galicia and Moroccan Jews, are separated from the remaining European populations. Moroccans also appear to be separated, because this population is the most differentiated relative to European populations. The three groups of islands are located in different branches of the tree that emerge from a group of European populations. However, the Western group appears in a very profound and isolated branch, reinforcing the previous results that indicate a differentiation of this group of islands. The Central group is located in the same branch as Israel, while the Eastern group is positioned in the same branch as the Moroccan Berbers. The position of the Eastern group corroborates previous results, indicating some affinities with African populations. In a similar manner the position of the Central group indicates that there is some relationship with Near Eastern populations.

The analysis of phylogeography of Azores mtDNA, estimates of admixture, genetic distances and phylogenetic reconstruction allow us to infer a possible origin for the mtDNA gene pool of the Azores. All the analyses support the idea that, in all the groups of islands, the majority of mtDNA lineages come from the Iberian Peninsula, mainly from Portugal, but the contribution

of other European populations, especially from Northern Europe, cannot be disregarded. The presence of *Moorish* and *African* slaves in the islands is also supported by mtDNA genetic data, especially in the Eastern group, as well as the presence of Jews, mainly in the Central group.

Neither historical nor genetic data supports the idea of a differential settlement for the Western group; however, it appears in the phylogenies in an isolated branch. With 4329 inhabitants (INE, 2001) today, the Western group has reduced its population size by half in less than one century. Thus, this "bottleneck", combined with a small population size since peopling, has led to a very atypical distribution of haplogroups/haplotypes that influences all the results obtained. Future genetic and biodemographic investigations in the Western group will certainly help to clarify its position relative to the rest of the archipelago, and to other populations.

Conclusions

This work confirms that studies of large population samples in restricted geographical contexts can produce valuable insights concerning specific genetic and demographic features that would remain undetectable in broad scale surveys.

The distinct analyses allow us to infer that the Azorean population does not present the typical characteristics of an isolated population, as is usually postulated for islands. However, the Azores archipelago has a recent demographic history, was peopled by individuals of multiple origins, which naturally leads to a population with high diversity, and it is probable that the effect of isolation, if it exists, remains undetectable at this time.

Separate analyses of the three groups of islands appeared to be valuable for detecting the specific features of the Western group, which presents some differentiation with respect to the other groups of islands. This differentiation should be the result of genetic drift processes, induced by the small population size since peopling, and eventually the result of the recent reduction in population size.

The analysis of lineage sharing supports the idea that mainland Portugal and Northern European populations could have contributed the majority of the mtDNA lineages that are currently observed in the Azorean mtDNA gene pool. However, many lineages are still private to the Azores, and their origins are sometimes difficult to infer, since historically related populations such as the population of Flanders are not used for comparison, because they remain uncharacterised for mtDNA.

Contrary to studies using autosomal markers (Lima *et al*. personal communication), it was possible to infer the influence of African and Near East/Jewish individuals in the islands, pointing out the effectiveness of non-recombining markers to reconstruct the history of populations. It was even possible to detect differential contributions of non-European populations to the peopling of the three groups of islands, mainly the contribution of Africans to the Eastern group, and the contribution of Near East/Jewish to the Central group, both in accordance with historical data.

Part of the variation in the Azorean mtDNA can be explained by the settlement history of the archipelago. Furthermore, it has been necessary to bear in mind that the variation of mtDNA was affected during the last 500 years, by biodemographic and genetic processes like founder effect, genetic drift, migration, and even recent mutational events. Thus, the results presented here arise from the combination of the history of peopling and the evolution of populations, with all their biodemographic and genetic interactions.

Acknowledgments

This work was supported by a grant (ref. SFRH/BD/723/2000) from Fundação para a Ciência e a Tecnologia - Operational programs Science Technology and Innovation (POCTI) and Society of Information (POSI) financed by UE founds and national founds of MCES.

Appendix HVRI region sequences and RFLP specific screening in the Azorean population

Sample Name	HVRI (16050–16399)	Hapl. Seq.	*Mnl*I 10871	*Dde*I 10394	*Mse*I 14766	*Hin*fI 12308	*Alu*I 7025	*Nla*III 4216	*Nla*III 4577	*Hae*I 9052	*Hae*III 8994	*Dde*I 1715	*Alu*I 10032	*Acc*I 14465	*Sma*I 11719	*Bst*0I 13705	*Alu*I 15606	*Alu*I 10397	*Hpa*I 3592	12705 C/T	Hapl. RFLP	Hapl. Seq/RFLP
Eastern group																						
SM1	129, 183C, 189, 189insC(1), 249, 311	M1?	−	+			+											+	−		M	M1?
SM2	CRS	H or HV or U or R	+	−	−		−														H	H
SM3	343, 390	U3	+	−	+	+	+	−		+											U	U3
SM4	CRS	H or HV or U or R	+	−	−		−		+												H	H
SM5	CRS	H or HV or U or R	+	−	−		−														H	H
SM6	93, 189, 189insC(1), 192, 270, 311	U5a or U5b	+	−	+	+	+	−	+	+	+	+			−						U	U5a/b
SM7	69, 126	J	+	+	+		+	+								−					J	J
SM8	CRS	H or HV or U or R	+	−	−		−									+					H	H
SM9	CRS	H or HV or U or R	+	−	−		−									+					H	H
SM10	274	H or HV or U or R	+	−	−		−														H	H
SM11	145, 176G, 223, 390	N1b	+	−	+	−	+	−				−	−	−							Other	N1b
SM12	189, 192, 223, 278	X	+	−	+	−	+	−	+		+	+		+	−						X	X
SM13	126, 294, 296, 304	T2	+	−	+		+	+									+				T	T2
SM14	207	H or HV or U or R	+	−	−		−														H	H
SM15	CRS	H or HV or U or R	+	−	−		−														H	H
SM16	223, 260, 278, 390, 399	L2	−	+			+												+		L1 or L2	L2
SM17	196, 224, 311, 362	K	+	+	+	+	+	−		−											K	K
SM18	168, 224, 311	K	+	+	+	+	+	−		−											K	K
SM19	69, 111A, 126, 261	J1	+	+	+		+	+								−					J	J1
SM20	343, 356, 390	U3 or U4	+	−	+	+	+	−		+											U	U3
SM21	CRS	H or HV or U or R	+	−	−		−														H	H
SM22	92, 172, 219, 278	U6a	+	−	+	+	+	−		+											U	U6a
SM23	224, 270, 311	U5 or K	+	−	+	+	+	−		+											U	U5
SM24	93, 189, 354	H or HV or U or R	+	−	−		−														H	H
SM25	126, 218A	J?	+	+	+		+	+								−					J	J
SM26	126, 234, 248, 292, 294	T3	+	−	+		+	+									+				T	T3
SM27	189, 270	U5b	+	−	+	+	+	−		+											U	U5b
SM28	224, 270, 311	U5 or K	+	−	+	+	+	−		+											U	U5
SM30	92, 172, 219, 278	U6a	+	−	+	+	+	−		+											U	U6a
SM31	189, 189insC(2)	H or HV or U or R	+	−	−		−														H	H
SM32	93A, 223, 278, 294, 309, 390	L2	−	+			+												+		L1 or L2	L2
SM33	126, 163A, 186, 189, 294	T1	+	−	+		+	+									+				T	T1
SM34	196, 224, 311, 362	K	+	+	+	+	+	−		−											K	K
SM36	CRS	H or HV or U or R	+	−	−		−														H	H
SM37	69, 126, 150	J	+	+	+		+	+								−					J	J
SM38	126, 294, 296, 304	T2	+	−	+		+	+									+				T	T2
SM39	206C, 223, 292, 311	W	+	−	+	−	+	−	+		−	+									W	W
SM40	69, 126, 278, 366	J	+	+	+		+	+								−					J	J
SM41	93, 189, 189insC(1), 192, 270, 311	U5a or U5b	+	−	+	+	+	−		+											U	U5a/b
SM42	CRS	H or HV or U or R	+	−	−		−														H	H
SM43	69, 126	J	+	+	+		+	+								−					J	J
SM44	256, 270, 399	U5a1a	+	−	+	+	+	−		+											U	U5a1a
SM45	126, 294, 296, 304	T2	+	−	+		+	+									+				T	T2
SM46	CRS	H or HV or U or R	+	−	−	−	−	−	+		+	+									H	H
SM47	CRS	H or HV or U or R	+	−	−		−														H	H
SM48	298	Pre*V or V	+	−	−	−	+	−	−			+									V	V
SM49	189, 192, 270, 311, 336	U5a or U5b	+	−	+	+	+	−		+											U	U5a/b

Appendix (*continued*)

Sample Name	HVRI (16050–16399)	Hapl. Seq.	*Mnl*I 10871	*Dde*I 10394	*Mse*I 14766	*Hin*fI 12308	*Alu*I 7025	*Nla*III 4216	*Nla*III 4577	*Hae*II 9052	*Hae*III 8994	*Dde*I 1715	*Alu*I 10032	*Acc*I 14465	*Sma*I 11719	*Bst*0I 13705	*Alu*I 15606	*Alu*I 10397	*Hpa*I 3592	12705 C/T	Hapl. RFLP	Hapl. Seq/RFLP
SM50	69, 126, 145, 231, 261	J1a	+	+	+		+	+								−					J	J1a
SM51	172, 219, 278	U6a	+	−	+	+	+	−		+											U	U6a
SM52	172, 182C, 183C, 189, 223, 320, 390C	L3e2b	−	+			+											−	−		L3?	L3e2b
Central group																						
A1	69, 126, 193, 278	J2	+	+	+		+	+								−					J	J2
A2	240	H or HV or U or R	+	−	−		−														H	H
A3	69, 126, 145, 172, 222, 261, 305T	J1b1	+	+	+		+	+								−					J	J1b1
A4	CRS	H or HV or U or R	+	−	−		−														H	H
A5	CRS	H or HV or U or R	+	−	−		−													C	H	H
A6	240	H or HV or U or R	+	−	−		−														H	H
A7	129, 145, 176G, 223, 291, 390	N1b	+	−	+	−	+	−				−	−	−						T	N*	N1b
A8	69, 126, 145, 172, 222, 261, 305T	J1b1	+	+	+		+	+								−					J	J1b1
A9	126, 234, 248, 292, 294	T3	+	−	+		+	+									+				T	T3
A10	CRS	H or HV or U or R	+	−	−		−														H	H
A11	CRS	H or HV or U or R	+	−	−		−														H	H
A12	129, 242	H or HV or U or R	+	−	−		−														H	H
A13	69, 126, 145, 172, 222, 261, 305T	J1b1	+	+	+		+	+								−					J	J1b1
A14	126, 234, 248, 292, 294	T3	+	−	+		+	+									+				T	T3
A15	126, 362	pre-HV	+	−	+	−	+	−	+		+	+	−	−	+						pre-HV	pre-HV
A16	129, 223, 391	I	+	+	+	−	+	−			+	−	+								I	I
A17	93, 224, 290, 311	K	+	+	+	+	+	−		−											K	K
P1	224, 311	K	+	+	+	+	+	−		−											K	K
ACR1	293, 294, 304	T2?	+	−	+		+	+									+				T	T2?
ACR3	CRS	H or HV or U or R	+	−	−		−														H	H
ACR4	129, 223, 391	I	+	+	+	−	+	−			+	−	+								I	I
ACR5	69, 93, 126, 278, 366	J	+	+	+		+	+								−					J	J
ACR6	129, 187, 189, 223, 265C, 278, 286A, 292, 294, 311, 360	L1c	−	+															+		L1 or L2	L1c
ACR7	126, 186, 189, 294	T	+	−	+		+	+									+				T	T
ACR8	126, 129, 294, 296, 304	T2	+	−	+	−	+	+				+					+				T	T2
ACR9	CRS	H or HV or U or R	+	−	−		−														H	H
ACR10	CRS	H or HV or U or R	+	−	−		−														H	H
ACR11	129, 182C, 183C, 189, 223, 230, 249, 311, 359	M1	−	+														+	−		M	M1
ACR12	189	H or HV or U or R	+	−	−		−														H	H
ACR14	126, 186, 189	T?	+	−	+		+	+									+				T	T
ACR15	129, 223, 391	I	+	+	+	−	+	−			+	−	+								I	I
ACR16	69, 126, 145, 231, 261	J1a	+	+	+		+	+								−					J	J1a
ACR17	162	H or HV or U or R	+	−	−		−														H	H
ACR18	298	Pre* V or V	+	−	−	−	+	−	−			+									V	V
ACR19	176	H or HV or U or R	+	−	−		−														H	H
ACR20	126, 294, 304	T2	+	−	+		+	+									+				T	T2
ACR21	189, 192, 223, 278, 362	X	+	−	+	−	+	−			+	−	−	+							X	X
ACR22	CRS	H or HV or U or R	+	−	−		−														H	H
ACR23	69, 126	J	+	+	+		+	+								−					J	J

Appendix (*continued*)

Sample Name	HVRI (16050–16399)	Hapl. Seq.	*Mnl*I 10871	*Dde*I 10394	*Mse*I 14766	*Hin*fI 12308	*Alu*I 7025	*Nla*III 4216	*Nla*III 4577	*Hae*II 9052	*Hae*III 8994	*Dde*I 1715	*Alu*I 10032	*Acc*I 14465	*Sma*I 11719	*Bst*01 13705	*Alu*I 15606	*Alu*I 10397	*Hpa*I 3592	12705 C/T	Hapl. RFLP	Hapl. Seq/RFLP
ACR24	183C, 189, 223, 260, 278	X	+	−	+	−	+	−			+	−	−	+							X	X
ACR25	278, 311	R1	+	+	+	+	+	−		−											K	K
ACR26	CRS	H or HV or U or R	+	−	−		−														H	H
ACR27	126, 186, 189, 294	T	+	−	+		+	+									+				T	T
ACR28	298, 311	Pre∗V or V	+	−	−	−	+	−	−			+									V	V
ACR29	126, 294, 296, 304	T2	+	−	+		+	+									+				T	T2
ACR30	223, 292, 320	W	+	−	+	−	+	−			−	+									W	W
ACR31	189, 356	H or HV or U or R	+	−	−		−														H	H
ACR32	129, 249	H or HV or U or R	+	−	−		−														H	H
ACR33	223, 292, 320	W	+	−	+	−	+	−			−	+									W	W
ACR34	129, 316	H or HV or U or R	+	−	−		−														H	H
ACR35	69, 278	J?	+	+	+		+	+								−					J	J
ACR36	CRS	H or HV or U or R	+	−	−		−														H	H
ACR37	192, 256, 270, 362, 399	U5a1a	+	−	+	+	+	−		+											U	U5a1a
ACR38	CRS	H or HV or U or R	+	−	−		−														H	H
ACR39	93, 269, 270	H or HV or U or R	+	−	−		−														H	H
ACR40	304	H or HV or U or R	+	−	−		−														H	H
ACR41	CRS	H or HV or U or R	+	−	−		−														H	H
ACR42	93, 224, 290, 311	K	+	−	+	+	+	−		−											K	K
ACR44	124	H or HV or U or R	+	−	−		−														H	H
ACR45	288, 311	H or HV or U or R	+	−	+	−	+	−			+	+			−			−		C	R*	R*
Western group																						
F1	51, 172, 219, 311	U6b	+	−	+	+	+	−		+											U	U6b
F2	298	Pre*V or V	+	−	−	−	+	−	−												V	V
F3	129, 242	H or HV or U or R	+	−	−		−														H	H
F4	298	Pre*V or V	+	−	−	−	+	−	−												V	V
F5	126, 288A, 292, 294	T3	+	−	+	−	+	+									+				T	T3
F6	189, 298	Pre*V or V	+	−	−	−	+	−	−												V	V
F7	298	Pre*V or V	+	−	−	−	+	−	−												V	V
F8	189, 298	Pre*V or V	+	−	−	−	+	−	−												V	V
F9	189, 298	Pre*V or V	+	−	−	−	+	−	−												V	V
F10	126, 288A, 292, 294	T3	+	−	+	−	+	+									+				T	T3
F11	189, 298	Pre*V or V	+	−	−	−	+	−	−												V	V
F12	172, 219, 261, 311	U6b	+	−	+	+		−		+											U	U6b
F13	293, 311	H or HV or U or R	+	−	−		−														H	H
F14	126, 163, 186, 189	T?	+	−	+	−	+	+									+				T	T
F15	201, 278	H or HV or U or R	+	−	−		−														H	H
F16	163, 186, 189	T?	+	−	+	−	+	+									+				T	T
F17	298	Pre*V or V	+	−	−	−	+	−	−												V	V
F19	51, 129C, 182C, 183C, 189, 189insC(1), 362, 399	U2	+	−	+	+	+	−		+											U	U2
F20	298	Pre*V or V	+	−	−	−	+	−	−												V	V
F21	298	Pre*V or V	+	−	−	−	+	−	+												Pre*V or HV	pre*V
F22	CRS	H or HV or U or R	+	−	−		−														H	H
F23	189, 298	Pre*V or V	+	−	−	−	+	−	−												V	V
F24	126, 193, 294, 296, 304	T	+	−	+	−	+	+									+				T	T
F25	189, 298	Pre*V or V	+	−	−	−	+	−	−												V	V
F26	162, 209	H or HV or U or R	+	−	−		−														H	H
F27	CRS	H or HV or U or R	+	−	−		−														H	H

Appendix (*continued*)

Sample Name	HVR I (16050–16399)	Hapl. Seq.	*Mnl*I 10871	*Ddi*I 10394	*Mse*I 14766	*Hin*fI 12308	*Alu*I 7025	*Nla*III 4216	*Nla*III 4577	*Hae*II 9052	*Hae*III 8994	*Ddi*I 1715	*Alu*I 10032	*Acc*I 14465	*Sma*I 11719	*Bst*0I 13705	*Alu*I 15606	*Alu*I 10397	*Hpa*I 3592	12705 C/T	Hapl. RFLP	Hapl. Seq/RFLP
F28	126, 288A, 292, 294	T3	+	–	+	–		+									+				T	T3
F29	129, 182C, 183C, 189, 223, 249, 311, 359	M1	–	+			+											+	–		M	M1
F30	201, 278	H or HV or U or R	+	–	–		–						–								H	H
F31	298	Pre* V or V	+	–	–	–	+	–	+			+									Pre* V or HV	pre* V
F32	93, 224, 311	K	+	+	+	+	+			–											K	K
F33	86, 104, 183C, 189, 189insC(1), 223, 271, 278	X	+	–	+	–	+	–			+		–	+							X	X
F34	126, 193, 294, 296, 304	T2	+	–	+	–	+	+									+				T	T2
F35	298	Pre* V or V	+	–	–	–	+	–	–												V	V
F36	CRS	H or HV or U or R	+	–	–		–														H	H
F37	129, 182C, 183C, 189, 223, 249, 311, 359	M1	–	+			+											+	–		M	M1

Variant positions from the CRS are shown between 16050 and 16399 in HVRI (minus 16000). Substitutions are transitions unless the base change is explicitly indicated. Insertions of one and two cytosines are shown by 'insC(1)' and 'insC(2)' respectively. The results of specific RFLP screening are also presented. + and – indicates, respectively, the presence and absence of the restriction site. Classification into haplogroups is presented using individual and combined information of sequences and RFLPs. Haplogroups presenting an ambiguous classification are assigned with a question mark (?).

References

Amorim, A., Faria, A. S. N. & Pereira, E. L. (1979) Red Cell Acid Phosphatase (ACP_1) Polymorphism in Flores Island (Azores). *Trabalhos do Instituto de Antropologia "Dr. Mendes Corrêa"*. **35**, 1–4.

Anderson, S., Bankier, A., Barrell, M., Bruijn, M. H. L., Coulson, A. R., Drouin, J., Epcron, I. C., Nicrlich, D. P., Roc, B. A., Sanger, F., Chreier, P. H., Smith, A. J. H., Standen, R. & Young, I. G. (1981) Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457–464.

Bandelt, H.-J., Forster, P. & Röhl, A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* **16**, 37–48.

Bertranpetit, J., Sala, J., Calafell, F., Underhill, P. A., Moral, P. & Comas, D. (1995) Human mitochondrial DNA variation and the origin of Basques. *Ann Hum Genet* **59**, 63–81.

Chen, Y. S., Olckers, A., Schurr, T., Kogelnik, A. M., Huoponen, K. & Wallace, D. C. (2000) MtDNA variation in the south African Kung and Khwe and their genetic relationships to other African populations. *Am J Hum Genet* **66**, 1362–1383.

Chen, Y. S., Torroni, A., Excoffier, L., Santachiara-Benerecetti, A. S. & Wallace, D. C. (1995) Analysis of mtDNA variation in African population reveals the most ancient of all human continent-specific haplogroups. *Am J Hum Genet* **57**, 133–149.

Côrte-Real, H., Macaulay, V., Richards, M., Hariti, G., Issad, M. S., Cambon-Thomsen, A., Papiha, S., Bertranpetit, J. & Sykes, B. (1996) Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. *Ann Hum Genet* **60**, 331–350.

Cunha, E. (1987) Consanguinity of the Azores Islands (1979–1985). *Antropologia Portuguesa* **4/5**, 67–79.

Di Rienzo, A. & Wilson, A. C. (1991) Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proc Natl Acad Sci USA* **88**, 1597–1601.

Ewens, W. J. (1972) The sampling theory of selectively neutral alleles. *Theor Popul Biol* **3**, 87–112.

Felsenstein, J. (1995) *PHYLIP (Phylogeny Inference Package) ver 3.57c*. Department of Genetics. University of Washington, Seattle.

Forster, P., Harding, R., Torroni, A. & Bandelt, H.-J. (1996) Origin and evolution of Native Americans mtDNA variation: a reappraisal. *Am J Hum Genet* **59**, 935–945.

Francalacci, P., Bertranpetit, J., Calafell, F. & Underhill, P. A. (1996) Sequence diversity of the control region of mitochondrial DNA in Tuscany and its implications for the peopling of Europe. *Am J Phys Anthropol* **100**(4), 443–460.

Francalacci, P., Montiel, R. & Malgosa, A. (1999) A mitochondrial DNA database – applications to problems of nomenclature and population genetics. In: *Genomic*

Diversity: Applications in Human population genetics (eds. S. S. Papiha R. Deka & R. Chakraborty), pp. 103–119. Kluwer Academic/Plenum Publishers, New York.

Gomes, F. (1997) A ilha das Flores: da redescoberta à actualidade (subsídios para a sua história). Câmara Municipal de Lajes das Flores, Azores, Portugal.

Graven, L., Passarino, G., Semino, O., Boursot, P., Santachiara-Benerecetti, S., Langaney, A. & Excoffier, L. (1995) Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka population. *Mol Biol Evol* **12**, 334–345.

Hall, T. A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* **41**, 95–98.

Handt, O., Meyer, S. & Haeseler, A. (1998) Compilation of Human mtDNA control region sequences. *Nucleic Acids Res* **26**(1), 126–129.

Helgason, A., Hickey, E., Coodacre, S., Vidar, B., Stefánsson, K., Ward, R. & Sykes, B. (2001) mtDNA and the islands of north Atlantic: estimating the proportions of Norse and Gaelic ancestry. *Am J Hum Genet* **68**, 723–737.

Helgason, A., Siguroardóttir, S., Gulcher, J., Ward, R., Stefánsson, K. (2000) mtDNA and the origin of the Icelanders: deciphering signals of recent population history. *Am J Hum Genet.* **66**, 999–1016.

Hofmann, S., Jaksch, M., Bezold, R., Mertens, S., Aholt, S., Paprotta, A. & Gerbitz, K. D. (1997) Population genetics and disease susceptibility: characterization of central European haplogroups by myDNA gene mutations, correlation with D loop variants and association with disease. *Hum Mol Genet* **6**(11), 1835–1846.

INE (2001) *Estimativas de população residente 2001 aferidas para os resultados provisórios dos Censos 2001* Instituto Nacional de Estatística, Portugal.

Lima, M. (1991) *Contribuição para o estudo biodemográfico da freguesia de Rabo de Peixe* Provas de aptidão pedagógica e capacidade científica. University of the Azores, Portugal.

Lima, M. & Soares, A. O. (1992) Níveis de consanguinidade e endogamia da população de Furnas. *Antropologia Portuguesa* **9–10**, 33–42.

Macaulay, V. (2000) *Haplogroup motifs.* Retrieved October 8, 2001 from the World Wide Web: http://www.stats.ox.ac.uk/~macaulay/founder2000/motif.html

Macaulay, V., Richards, M., Hickey, E., Veja, E., Cruciani, F., Guida, V., Scozzari, R., Bonné-Tamir, B., Sykes, B. & Torroni, A. (1999)). The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am J Hum Genet* **64**, 232–249.

Mateu, E., Comas. D., Calafell, F., Pérez-Lazaun, A., Abade, A. & Bertranptit, J. (1997) A tale of two islands: population history and mitochondrial DNA sequence variation of Bioko and São Tomé, Gulf of Guinea. *Ann Hum Genet* **61**, 507–518.

Matos, A. (1989) Povoamento e Colonização dos Açores *in Portugal no Mundo* Publicações Alfa, Lisboa, Portugal

Mendonça, L. (1996) *História dos Açores – Visão geral (sécs. XV–XIX)* Centro de Apoio Tecnológico à Educação, Ponta Delgada, Azores, Portugal

Montiel, R. (2001) *Estudio diacrónico de la variabilidad del DNA mitocondrial en población Catalana* PhD dissertation. Autonomous University of Barcelona, Barcelona, Spain. (http://www.tdcat.cesca.es/TDCat-0726101-095837)

Montiel, R., Malgosa, A. & Francalacci, P. (2001) Authenticating ancient human mitochondrial DNA. *Hum Biol* **73**(5), 689–713.

Nei, M. (1987) *Molecular evolutionary genetics* Columbia University Press, New York.

Page, R. (1996) Treeview: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* **12**, 357–358.

Parson, W., Parsons, T., Scheithauer, R. & Holland, M. M. (1998) Population data for 101 Austrian caucasian mitochondrial DNA d-loop sequences: Application of mtDNA sequence analysis to a forensic case. *Int J Legal Med* **111**, 124–132.

Pereira, L., Macaulay, V., Torroni, A., Scozzari, R., Prata, M. J. & Amorim, A. (2001). Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. *Ann Hum Genet* **65**, 439–458.

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

Piercy, R., Sullivan, K., Benson, N. & Gill, P. (1993) The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. *Int J Leg Med* **106**, 85–90.

Pinto, F., González, A., Hernández, M., Larruga, J. M. & Cabrera, V. M. (1996) Genetic relationship between the Canary islanders and their African and Spanish ancestors inferred from mitochondrial DNA sequences. *Ann Hum Genet* **60**, 321–330.

Pires, L. (1999) *A Origem da população dos Açores com base em estudos de DNA mitocondrial* BSc dissertation. University of Coimbra, Coimbra, Portugal

Plaza, S., Calafell, F., Bertranpetit, J. & Comas, D. (2001) *Mitochondrial genome analysis in Human populations from the western Mediterranean* Poster presented in the Euroconference "Inherited Disorders and their Genes in Different European Populations, Exploring Human Genetic Variability." Sant Feliu de Guíxols, Spain, 15–19 November.

Quintana-Murci, L., Semino, O., Bandelt, H.-J., Passarino, G., McElreavery, K. & Santachiara-Benerecetti, A. S. (1999) Genetic evidence of an early exit of *Homo sapiens*

sapiens from Africa through eastern Africa. *Nature Genetics* **23**, 437–147.

Rando, J. C., Cabrera, V. M., Larruga, J. M., Hernández, M., González, A. M., Pinto, F. & Bandelt, H.-J. (1999) Phylogeographic patterns of mtDNA reflecting the colonization of the Canary islands. *Ann Hum Genet* **63**, 413–428.

Rando, J. C., Pinto, F., González, A. M., Hernández, M., Larruga, J. M, Cabrera, V. M. & Bandelt, H.-J. (1998) Mitochondrial DNA analysis of Northwest African populations reveals genetic exchanges with European, Near-Eastern, and sub-Saharan populations. *Ann Hum Genet* **62**, 531–550.

Raymond, M. & Rousset, F. (1995) An exact test for population differentiation. *Evolution* **49**, 1280–1283.

Reynolds, J., Weir, B. & Cockerham, C. (1983) Estimation of the coancestry coefficient: basis for short-term genetic distance. *Genetics* **105**, 767–779.

Richards, M. & Macaulay, V. (2000) Genetic data and the colonization of Europe: genealogies and founders. In: *Archaeogenetics: DNA and the population prehistory of Europe* (eds. C. Renfrew & K. Boyle), pp. 139–151. McDonald Institute for Archaeological Research, Cambridge

Richards, M., Macaulay, V., Bandelt, H. J. & Sykes, B. C. (1998) Phylogeography of mitochondrial DNA in western Europe. *Ann Hum Genet* **62**, 241–260.

Richards, M., Macaulay, V., Hickey, E., Veja, E., Sykes, B., Guida, V., Rengo, C., Sellitto, D., Cruciani, F., Kivisild, T., Villems, R., Thomas, M., Rychkov, S., Rychkov, O., Rychkov, Y., Golge, M., Dimitrov, D., Hill, E., Bradley, D., Romano, V., Cali, F., Vona, G., Demaine, A., Papiha, S., Triantaphyllidis, C., Stefanescu, G., Hatina, J., Belledi, M., Di Rienzo, A., Novelletto, A., Oppenheim, A., Norby, S., Al-Zaheri, N., Santachiara-Benerecetti, S., Scozzari, R., Torroni, A. & Bandelt, H.-J. (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* **67**, 1251–1276.

Rousselet, F. & Mangin, P. (1998) Mitochondrial DNA polymorphisms: a study of 50 French Caucasians individuals and application to forensic casework. *Int J Legal Med* **111**(6), 292–298.

Saillard, J., Magalhaes, P. J., Schwartz, M., Rosenberg, T. & Norby, S. (2000) Mitochondrial DNA variant 11719G is a marker for the mtDNA haplogroup cluster HV. *Hum Biol* **72**(6), 1065–1068.

Saitou, N. & Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**(4), 406–425.

Salas, A., Comas, D., Laureu, M., Bertranpetit, J. & Carracedo, A. (1998) mtDNA analysis of the Galician population: a genetic edge of European variation. *Eur J Hum Genet* **6**, 365–375.

Santos, C. (2001) *Haplogrupos del DNAmt en la población de las Azores (Portugal)* MSc dissertation. University Autónoma of Barcelona, Barcelona, Spain.

Schneider, S., Roessli, D. & Excoffier, L. (2000) *ARLEQUIN ver 2.000: A software for population genetics data analysis* Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Smith, M., Abade, A. & Cunha, E. (1992) Genetic structure of the Azores: marriage ad inbreeding in Flores. *Ann Hum Biol* **19**(6), 595–601.

SPSS Inc., 1989–1999. *SPSS for Windows – Release 9.0.0* Chicago.

Tajima, F. (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**, 437–460.

Thomas, H. (1998) *The slave trade – the history of the Atlantic slave trade 1440–1870*. London: Macmillan Publishers Ltd.

Thomas, M., Weale, M., Jones, A., Richards, M., Smith, A., Redhead, N., Toroni, A., Scozzari, R., Gratrix, F., Tarekegn, A., Wilson, J., Capelli, C., Bradman, N. & Goldstein, D. (2002) Founding mothers of Jewish communities: geographically separated Jewish groups were independently founded by very few female ancestors. *Am J Hum Genet* **70**, 1411–1420.

Torroni, A. & Wallace, D. C. (1994) Mithochondrial DNA variation in human populations and implications for detection of mitochondrial DNA mutations of pathological significance. *J Bioenerg Biomembr* **26**, 261–271.

Torroni, A., Huoponen, K., Francalacci, P., Petrozzi, M., Morelli, L., Scozzari, R., Obinu, D., Savontaus, M. L. & Wallace, D. C. (1996) Classification of European mtDNAs from an analysis of three European populations. *Genetics* **144**, 1835–1850.

Torroni, A., Bandelt, H.-J., d'Urbano, L., Lahermo, P., Moral, P., Sellitto, D., Rengo, C., Forster, P., Savontaus, M. L., Bonne-Tamir, B. & Scozzari, R. (1998) mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *Am J Hum Genet* **62**, 1137–1152.

Torroni, A., Bandelt, H.-J., Macaulay, V., Richards, M., Cruciani, F., Rengo, C., Martinez-Cabrera, V., Villems, R., Kivisild, T., Metspalu, E., Parik, J., Tolk, H. V., Tambets, K., Forster, P., Karger, B., Francalacci, P., Rudan, P., Janicijevic, B., Rickards, O., Savontaus, M. L., Huoponen, K., Laitinen, V., Koivumaki, S., Sykes, B., Hickey, E., Novelletto, A., Moral, P., Sellitto, D., Coppa, A., Al-Zaheri, N., Santachiara-Benerecetti, A. S., Semino, O. & Scozzari, R. (2001) A signal, from human mtDNA, of postglacial recolonization in Europe. *Am J Hum Genet* **69**, 844–852.

Torroni, A., Lott, M., Cabell, M., Chen, Y. S., Lavergne, L. & Wallace, D. C. (1994) mtDNA and the origin of Caucasians. Identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. *Am J Hum Genet* **55**, 760–776.

Torroni, A., Schurr, T., Cabell, M., Brown, M. D., Neel, J. V., Larsen, M., Smith, D. G., Vullo, C. M. & Wallace, D. C. (1993) Asian affinities and continental radiation of the four founding Native Americans mtDNAs. *Am J Hum Genet* **53**, 563–590.

Watson, E., Foster, P., Richards, M. & Bandelt, H.-J. (1997) Mitochondrial footprints of human expansion in Africa. *Am J Hum Genet* **61**, 691–704.

Received: 2 October 2002
Accepted: 13 December 2002

