

Distribution of HLA alleles in Portugal and Cabo Verde. Relationships with the slave trade route

H. SPÍNOLA¹, A. BREHM¹, F. WILLIAMS², J. JESUS¹ AND D. MIDDLETON²

¹Human Genetics Laboratory, Centre of Macaronesian Studies, University of Madeira, Campus of Penteada, 9000 Funchal, Portugal

²Northern Ireland Histocompatibility and Immunogenetics Laboratory, City Hospital and University of Ulster, Belfast, N. Ireland

(Received 11.2.02. Accepted 10.5.02)

SUMMARY

HLA-A, *-B*, and *-DR* frequencies were analysed in populations from Portugal and the Madeira and Cabo Verde Archipelagos, aiming to characterize their genetic composition. Portuguese settlers colonized both Archipelagos in the 15th and 16th centuries. Madeira received many sub-Saharan slaves to work in the sugar plantations, and Cabo Verde served as a pivotal market in the Atlantic slave trade and was populated by individuals coming from the Senegambia region of the West African coast. The population of Madeira shows the highest genetic diversity and the presence of alleles and haplotypes usually linked to sub-Saharan populations, the haplotypes accounting for 3.5% of the total. Cabo Verde presents typical markers acknowledged to be of European or Ibero-Mediterranean origin, thus revealing the admixture of European settlers with Sub-Saharan slaves. Altogether the number of European haplotypes reaches 15% of the total. The Portuguese population shows a perceivable and significant heterogeneity both in allele and haplotype frequencies, unveiling a differential input of peoples from different origins. A PCA of the populations studied, plus other relevant ones, clearly shows gene heterogeneity in mainland Portugal as well as the differences and relationships between these populations and Madeira and Cabo Verde.

INTRODUCTION

Human Leukocyte Antigens (HLA) class I loci (*A*, *B* and *C*) and class II (*DRB1*, *DQA1*, *DQB1* and *DPB1*) code for proteins on the surface of white blood cells that are involved in the immune response. These loci belong to a class of genes in the major histocompatibility complex (MHC) region on chromosome 6. Some physically closely related HLA loci show strong linkage disequilibria (Piazza & Lonjou, 1997), and thus are inherited as haplotypes. Many studies have shown that haplotype frequencies are characteristic of particular populations and even certain alleles are exclusively found in some ethnic

groups (Arnaiz-Villena *et al.* 1995, 1999; Clayton *et al.* 1997; Ivanova *et al.* 2001). Based on clinal variation of HLA alleles and haplotypes, associated with the fact that this class of gene system is probably the most polymorphic of the human genome, they have also been used to track relationships and origins of populations, as well as their present day genetic structure (Martinez-Laso *et al.* 1995; Chimge *et al.* 1997; Arnaiz-Villena *et al.* 2001a; Modiano *et al.* 2001; Sanchez-Mazas, 2001). The HLA composition of populations constitutes a complement to the examination of human genetic relationships, which is generally biased towards the maternally or paternally inherited mtDNA or Y haplotypes.

In recent years, HLA, mtDNA and Y typing has been used to characterize the genetic composition of present day Iberian populations

Correspondence: António Brehm. Tel: +351291705383. Fax: +351291705399.
E-mail: brehm@uma.pt

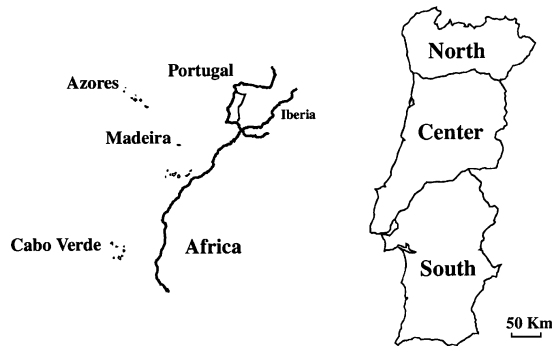


Fig. 1. Iberian Peninsula and the West African coast with the Atlantic archipelagos of Azores, Madeira and Cabo Verde.

(Martinez-Laso *et al.* 1995; Pinto *et al.* 1996; Arnaiz-Villena *et al.* 1997; Rando *et al.* 1998; Pereira *et al.* 2000; Gómez-Casado *et al.* 2001; Scozzari *et al.* 2001). This Atlantic fringe of Europe is particularly interesting because strong evidence is lacking for the origin of its first settlers, and because this region saw, in historic epochs, several admixtures with other peoples, mainly of North and sub-Saharan Africa origin. The Madeira and Cabo Verde Archipelagos were discovered in 1419 and 1456(–60) respectively, and throughout three centuries they acted pivotally in the Atlantic slave trade in which Portugal was involved (as did the Azores, but to a lesser extent). Slaves were captured from the coast of Guinea, a West Atlantic African region that was then known as Senegambia. They were taken to the Cabo Verde Islands, which served as a central outpost in the slave trade and distributed to Madeira, mainland Portugal, and also North America, the Antilles, and later Brazil (Russell-Wood, 1998). In the 16th century, due to the slave import, 10% of the population of south Portugal and Madeira was of sub-Saharan origin (Godinho, 1965). The settlement of the islands of Cabo Verde is itself interesting and peculiar. The first settlers were a few European men of Iberian and North European origin (less than 10%), while African slaves originating from the Guinea coast constituted most of the remaining population (Carreira, 1983). Although some HLA research has been carried out to understand the genetic relatedness of Iberians and North Afri-

cans, or other European populations (Bruges-Armas *et al.* 1999; Sánchez-Velasco *et al.* 1999; Gómez-Casado *et al.* 2001), none has really focused on the HLA genetic imprint left by the admixture between the colonizers and the colonized along the Atlantic islands involved in the slave trade. Also, it is very difficult to find appropriate population genetic HLA data for comparisons; the populations may not be correctly identified, allele nomenclature differs in serologically or DNA typed samples, or the results are not standardized (presenting either gene or allele frequency data). Also serologically typed samples are difficult to use for comparisons with DNA typed samples.

In the present work, we studied HLA frequencies from Madeira and the Cabo Verde Islands, and also mainland Portugal. The aim of this study was to characterize the genetic composition of these populations based on HLA markers and to uncover the genetic imprint left by slaves in mainland Portugal and Madeira, as well as the degree of admixture that has occurred in Cabo Verde. Little HLA data is available for West African sub-Saharan populations so Cabo Verde may partially reflect the genetic composition of the people of the countries from which the settlers came. The data has also been compared with previously published results for the Azores, Portugal and North Africa.

MATERIALS AND METHODS

Population samples

The study population consisted of a total of three hundred and ninety four unrelated males from mainland Portugal ($n = 145$), and from the Archipelagos of Madeira (MA, $n = 185$) and Cabo Verde (CV, $n = 64$). The individuals from mainland Portugal were further subdivided into three regions here regarded as sub-populations: North (NP, $n = 46$), Central (CP, $n = 50$) and South Portugal (SP, $n = 49$) (Fig. 1). All individuals had been subjected to an interview in order to select those that could unambiguously certify that all relatives for three generations were from

Table 1. *Populations used in the present study and for comparisons*

Population codes	Country/population	<i>n</i>	Reference(s)
NP	North Portugal	46	Present study
CP	Centre Portugal	50	Present study
SP	South Portugal	49	Present study
MA	Madeira Island	185	Present study
CV	Cabo Verde Archipelago	64	Present study
AZ	Azores Archipelago	174	Bruges-Armas <i>et al.</i> (1999)
PO	Portugal	240	Martinho <i>et al.</i> (1995), Santos <i>et al.</i> (1997)
NI	Northern Ireland	1000	Williams <i>et al.</i> (1999), Middleton <i>et al.</i> (2000)
MS	Murcia-Spain	173	Muro <i>et al.</i> (2001)
BA	Spain/Basque	82	Martinez-Laso <i>et al.</i> (1995)
FR	France	102	Bera <i>et al.</i> (2001)
IT	Italy	159311	Rendine <i>et al.</i> (1998)
DU	Netherlands/Dutch	98	Moghaddam <i>et al.</i> (1998)
MR	Morocco	98	Gómez-Casado <i>et al.</i> (2001)
AL	Algeria	106	Arnaiz-Villena <i>et al.</i> (1995)
CA	Cameroon	92	Ellis <i>et al.</i> (2000)
MO	Burkina Faso (Mossi)	53	Modiano <i>et al.</i> (2001)
RI	Burkina Faso (Rimaibé)	47	Modiano <i>et al.</i> (2001)
FU	Burkina Faso (Fulani)	49	Modiano <i>et al.</i> (2001)
MD	Senegal/Mandenka	191	Dard <i>et al.</i> (1992), Tiercy <i>et al.</i> (1992)

n = max. number of individuals analysed for the 3 loci.

the same region. Blood samples were collected from donors after informed consent. These populations were compared with others from the literature. In order to perform comparisons with our data, we have selected from the literature a few populations for which data tables are consistent. In this case we have chosen populations where allele frequencies would be directly available or could be inferred from the data presented. Table 1 shows the origin and total number of samples included in the analysis.

HLA typing and statistics

Genomic DNA was isolated from whole blood containing EDTA, using a Chelex standard method (Lareu *et al.* 1994). HLA Class I (*A* and *B*) and Class II (*DRB1*) allelic lineages were DNA-typed by a medium resolution PCR-SSCP method as previously described (Middleton, 2000). For purposes of brevity we have referred to these allele lineages as ‘alleles’ in this report. Basic genetic parameters (allele frequencies, Hardy–Weinberg Equilibrium (HWE)), among populations at the three loci were estimated with Arlequin v.2.000 (Schneider *et al.* 2000).

Estimation of haplotypes from genotypic data using maximum-likelihood has proved to be the

best approach to the deductive method using pedigree information (Schipper *et al.* 1998). Few software programs exist to estimate maximum-likelihood haplotypes from genotypic data. PHASE (Stephens *et al.* 2001) and Arlequin suit this purpose and we have tried both. There is at present no possibility to objectively choose between both outputs; once we compared our results with other published results, all estimated with Arlequin, we decided to choose this program. An analysis of molecular variance (AMOVA) restricted to our five populations (NP, CP, SP, MA, and CV), but including the already published Azores (AZ, Bruges-Armas *et al.* 1999), was performed based on Euclidean distances between all pairs of haplotypes (Excoffier *et al.* 1992). The total genetic variation between the populations was partitioned into hierarchical levels of grouping, and variance components were tested for significance by nonparametric randomisation tests using 10000 permutations under the null hypothesis of no population structure. Whenever it was necessary to compare single allele frequencies between two populations, we used the test to calculate population proportions between two populations proposed by Daniel (1987). Allelic frequencies among populations were log transformed in order to

Table 2. *Frequencies of HLA-A, -B and -DR alleles in the Portuguese and Cabo Verde population. Mandenka and Cameroon populations are also included for comparison as typical West African data. Allelic designations represent allele lineages and are symbolized by an asterisk following the locus name (the 3 highest values in each population in bold)*

Alleles	NP	CP	SP	MA	CV	MD ^a	CA ^a
<i>HLA-A*</i>							
01	7.6	13	5.1	7.0	5.5	4.7	1.1
02	31.6	28	24.5	27.6	17.2	17.9	19.0
03	6.5	11	8.2	7.0	7.0	5.5	8.3
11	5.4	5.0	14.3	5.9	2.3	0	0
23	10.9	1.0	6.1	6.8	13.3	18.5	18.6
24	5.4	8.0	10.2	14.1	6.3	0.3	0.5
25	1.1	0	0	0.8	0	0	0
26	5.4	1.0	2.0	3.5	4.7	4.5	2.3
29	3.3	8.0	7.1	6.5	4.7	4.0	10.4
30	2.2	4.0	5.1	1.6	15.4	16.8	14.3
31	6.5	4.0	4.1	2.7	1.6	1.2	1.1
32	5.4	4.0	3.1	5.4	3.1	1.9	1.1
33	8.7	3.0	4.1	3.0	5.5	13.1	2.8
34	0	0	0	0.3	0.8	2.5	2.2
36	0	0	0	0.3	0	0	2.2
66	0	2.0	1.0	1.1	0	0.3	7.1
68	0	7.0	5.1	5.7	5.5	8.2	3.4
69	0	1.0	0	0	1.6	0	0.6
74	0	0	0	0.7	4.7	0.6	5.0
80	0	0	0	0	0.8	0	0
<i>HLA-B*</i>							
07	3.3	8.0	1.0	5.1	6.3	6.1	6.0
08	6.5	4.0	2.0	7.6	8.6	7.3	5.4
13	1.1	1.0	1.0	0.5	0	0	2.7
14	6.5	4.0	5.1	7.3	8.6	2.8	3.2
15	5.4	8.0	2.0	6.5	13.3	11.9	9.3
18	7.6	9.0	11.2	6.5	3.1	3.6	2.7
27	4.3	1.0	2.0	1.9	3.1	1.5	0.5
35	9.8	14.0	21.7	9.7	12.3	13.3	7.1
37	0	2.0	0	0.5	0	0.6	0.5
38	0	0	1.0	3.8	0.8	0	0
39	4.3	1.0	1.0	1.9	0	1.9	0.5
40	5.4	3.0	1.0	4.9	3.1	0	1.1
41	2.2	0	1.0	1.1	0.8	0.6	0.5
42	0	1.0	0	0	1.6	2.8	4.9
44	15	15.0	20.5	15.6	6.3	6.8	8.7
45	1.1	5.0	1.0	1.6	1.6	0.9	3.3
47	0	1.0	0	0	0	1.5	1.6
49	7.6	2.0	4.1	6.2	2.3	5.3	5.4
50	2.2	0	4.1	4.1	0.8	0.3	0.5
51	13.3	13.0	11.2	10.0	7.8	3.8	1.6
52	0	1.0	0	0.3	0	5.3	0
53	1.1	1.0	0	1.1	8.6	4.0	10.9
55	1.1	1.0	0	0	0	0	0
56	0	0	1.0	0	0	3.1	0
57	0	3.0	4.1	2.2	5.5	1.2	2.7
58	2.2	2.0	2.0	2.2	5.5	8.2	16.5
67	0	0	1.0	0	0	0	0
78	0	0	0	0	0	7.2	0
81	0	0	1.0	0	0	0	4.4
<i>HLA-DRB1*</i>							
01	15.2	4.0	11.2	8.9	12.5	1.1	na
02	12.0	14.0	5.1	11.6	3.9	2.9	na
03	13.0	16.0	14.3	12.2	10.2	13.4	na
04	9.8	14.0	13.3	9.7	10.2	0.8	na
07	17.7	14.0	15.3	16.0	7.0	6.8	na

Alleles	NP	CP	SP	MA	CV	MD ^a	CA ^a
08	4.0	7.0	3.1	4.1	3.0	12.4	na
09	1.1	2.0	1.0	1.1	6.3	1.0	na
10	0	0	1.0	0.8	5.5	3.8	na
11	9.8	11.0	14.3	13.0	16.0	18.2	na
12	1.1	0	3.1	1.1	2.0	0.3	na
13	13.0	16.0	17.3	19.6	23.4	38.5	na
14	3.3	2.0	1.0	1.9	0	0.8	na

^a Data from Dard *et al.* (1992), Tiercy *et al.* (1992) and Ellis *et al.* (2000); na (data not available).

better interpret the analysis and used in the MVSP v.3.12 statistical package, to perform Principal Component Analysis (PCA) displaying the population's position in two dimensions and thus assessing the relationships among them.

RESULTS AND DISCUSSION

HLA class I and II allele frequencies

All populations studied are in Hardy–Weinberg equilibrium. Table 2 presents the allele frequencies for the three loci. Overall gene diversity (GD) across all loci for each population, observed heterozygosity, and the number of alleles for each locus and number of combined haplotypes, are shown in Table 3. In all cases the observed heterozygosity values were not statistically different from the expected. The AMOVA results indicate that only 0.47% of the total genetic divergence was due to differences among the three regions (Portugal, MA and CV). Most of the divergence was attributed to variation within each population (99.3%). Overall differentiation among regions is small (FCT = 0.0047) and between populations FST values were non significant except when each one was compared with CV (minimum value of FST pairwise comparisons was 0.013, in all four cases, $p < 0.0001$).

*HLA-A**

Regarding *HLA-A* distribution, *HLA-A*02* shows a clear decreasing north-south cline in mainland Portugal, and intermediate frequencies in MA. It is the most frequent allele in CV. *HLA-A*11* and *HLA-A*24* are usually linked to Mediterranean populations with frequencies around 4–6%, and are supposedly absent in sub-

Saharans (Arnaiz-Villena *et al.* 2001a; Gómez-Casado *et al.* 2001; Muro *et al.* 2001). Both are much more frequent in SP, compared to CP and NP (one-sided test, $p < 0.01$). Their presence in CV suggests that they were probably introduced either by white colonizers (as happened with Martinicans (Bera *et al.* 2001)) or slaves, with a genetic background originating in North Africa, as it is believed that no direct North-African input to Cabo Verde took place. *HLA-A*23* exhibits bizarre behaviour in the sense that it reaches the highest frequency in NP and CV (11 and 13%), 6% in SP and MA but only 1% in CP (one sided test, $p < 0.03$). This allele is present in European populations at frequencies around 4% (Ivanova *et al.* 2001; Rendine *et al.* 1998). *HLA-A*30*, found widespread in sub-Saharan populations (Modiano *et al.* 2001; Bera *et al.* 2001; Uko *et al.* 1997) and also in North Africa (Gómez-Casado *et al.* 2001), is the second most frequent allele in CV. *HLA-A*33* attains the highest value in NP (8.7%) being confined to percentages varying from 3–4% in the remaining regions of the country. Strangely, in Nigeria it is the second most frequent (Uko *et al.* 1997) but its frequency in CV is no different from that of mainland Portugal. Finally, if alleles *HLA-A*74* and *HLA-A*80* (found in relatively high frequencies in CV) are of sub-Saharan origin, the presence of *HLA-A*74* in MA could be due to slave introduction.

*HLA-B**

Within *HLA-B* allele distribution, *HLA-B*35* shows a clear increasing north-south cline in Portuguese mainland populations. High frequencies of this allele have been attributed to West African (Modiano *et al.* 2001; Dard *et al.* 1992) and Mediterranean populations (Arnaiz-Villena *et al.* 1995, 2001a, b). *HLA-B*35* is

Table 3. *Basic genetic indices for the 3 loci in the Portuguese and Cabo Verde populations*

	NP	CP	SP	MA	CV
N° alleles					
<i>HLA-A</i>	13	15	14	18	17
<i>HLA-B</i>	20	22	22	22	19
<i>HLA-DR</i>	11	10	12	12	11
N° haplotypes	70	75	68	199	94
Gene diversity (SD)	0.997(0.002)	0.997(0.002)	0.997(0.002)	0.996(0.000)	0.998(0.001)
Obs. heterozygosity					
<i>HLA-A</i>	0.913	0.940	0.877	0.821	0.921
<i>HLA-B</i>	0.913	0.880	0.877	0.870	0.937
<i>HLA-DR</i>	0.891	0.880	0.836	0.827	0.906

extremely frequent in Algerians, suggesting that this cline in Portugal represents an autosomic input of the Arab settlement in the country, which lasted for several centuries. The presence of a high frequency of this allele in CV suggests migration of settlers from Portugal, as previously suggested for *HLA-A*11*. On the other hand, the high frequency of this allele in the Mandenka is probably the result of North African influence in this ethnic group, as seen in mtDNA (Brehm *et al.* 2002). It is worth noting the high frequency of the putative African *HLA-B*15* allele (Bera *et al.* 2001) in CV (13.3%), otherwise present in SP at 2% and in the remaining populations with figures ranging from 5–8%. This allele has also been found in Moroccans at 5% (Gómez-Casado *et al.* 2001). *HLA-B*53* is an allele showing a big variation in African populations, a fact attributed to historical population backgrounds (Ellis *et al.* 2000). It is present in CV at a lower percentage (8.6%) than that found in the nearest mainland African population (13.9% in Serer and Mandenka from Senegal, or 14–17% in Gambia). As expected it is present at a very low frequency in Portugal. *HLA-B*57* appears at 5.5% in CV but is also found with relatively equal frequencies in CP and SP (3–4%) but not in NP, suggesting that its presence in the Centre and South of Portugal is a result of the sub-Saharan slaves that comprised 10% of the population of these regions.

*HLA-DRBI**

HLA Class II *DRBI*13* can be found in sub-Saharan populations at 20%, rising to 48.8% in a Gambian population (Table 2) (Hill *et al.* 1992;

Tiercy *et al.* 1992). NP shows frequencies similar to a North European population (Ronningen *et al.* 1990), but this allele also shows a North–South increasing cline in Portugal. The high frequency of this allele in SP, and especially MA, could suggest an additional introduction of sub-Saharan import. When the frequencies of *DRBI*1*, *DRBI*4* and *DRBI*7* are all low and the frequencies of *DRBI*13* and *DRBI*14* are both high, this is considered to be typical of sub-Saharan Africans (Tiercy *et al.* 1992). However, in CV this is not the pattern. Alternatively, *DRBI*10* could be considered to be of sub-Saharan origin since it exists in CV but is absent, or present at very low levels, in Portugal.

Haplotype frequencies

West European haplotypes

*A*01-B*08-DRBI*03* is thought of as being of Indo-European Celtic origin (Arnaiz-Villena *et al.* 1997) but is probably just West-European, since no evidence was provided that only Celts carried it. It is only found in CP (3.0%), which is rather strange if it was really Indo-European, and its presence in MA and CV suggests a colonizing event from people originating from the Centre of Portugal. *A*02-B*44-DRBI*04* is common in West Europe (Martinez-Laso *et al.* 1995) and North Africa (Arnaiz-Villena *et al.* 1995). It has been found in the Azores (2%, Bruges-Armas *et al.* 1999) and in the present study only in SP (3%). *A*02-B*44-DRBI*07* is widespread over Europe but the present study

Table 4. Three loci haplotype frequencies (HLA-A, -B and -DR) in the Portuguese and Cabo Verde populations (only those haplotypes greater than 2% in any population, or mentioned in the text are shown). Haplotype frequencies are estimate values according to the methodology explained in the Materials and Methods section

Haplotypes (HLA-A, -B, -DR)	Haplotype frequencies (%)				
	NP	CP	SP	MA	CV
A*01-B*08-DRB1*03		3.0		2.2	0.8
A*01-B*08-DRB1*13	2.2	1.0			
A*01-B*35-DRB1*07		2.8	2.0		
A*02-B*08-DRB1*03				1.9	
A*02-B*14-DRB1*01			2.0	0.3	
A*02-B*14-DRB1*13					2.3
A*02-B*15-DRB1*04	2.2	1.0	1.0		
A*02-B*18-DRB1*02	2.2				
A*02-B*18-DRB1*03		2.0	4.0		
A*02-B*18-DRB1*08		2.0			
A*02-B*18-DRB1*11				2.4	
A*02-B*27-DRB1*04					2.3
A*02-B*35-DRB1*01	3.3			1.1	
A*02-B*35-DRB1*11					2.3
A*02-B*35-DRB1*13			2.0		0.8
A*02-B*39-DRB1*11	2.2				
A*02-B*40-DRB1*13		2.0			
A*02-B*41-DRB1*13	2.2				
A*02-B*44-DRB1*01		2.0		0.6	
A*02-B*44-DRB1*02		4.9			
A*02-B*44-DRB1*07	5.4			0.6	
A*02-B*44-DRB1*04			3.0		
A*02-B*44-DRB1*13				2.1	
A*02-B*50-DRB1*07				2.1	
A*02-B*51-DRB1*04	1.1		3.0	2.2	
A*02-B*51-DRB1*11	3.3	2.0		0.6	
A*02-B*51-DRB1*13		3.0	2.0	1.9	1.6
A*02-B*53-DRB1*13					2.3
A*03-B*35-DRB1*01			3.0		
A*03-B*35-DRB1*02		2.0			
A*03-B*35-DRB1*04		2.0			
A*03-B*44-DRB1*02	2.2				
A*03-B*44-DRB1*11			2.0		
A*03-B*49-DRB1*04			2.0		
A*03-B*58-DRB1*04	1.1				
A*11-B*18-DRB1*11			2.2		
A*11-B*35-DRB1*01	1.1		3.0	1.4	0.8
A*11-B*35-DRB1*11			2.9		
A*11-B*44-DRB1*13		1.0	2.0	0.3	
A*23-B*07-DRB1*02					1.6
A*23-B*08-DRB1*03			2.0	0.8	
A*23-B*15-DRB1*11					3.1
A*23-B*49-DRB1*13	3.3				
A*24-B*18-DRB1*14	2.2				
A*24-B*35-DRB1*11		2.0	1.0	0.3	1.6
A*24-B*44-DRB1*02				1.4	
A*24-B*44-DRB1*11				0.8	
A*24-B*44-DRB1*13			4.0		
A*24-B*51-DRB1*07			3.0		
A*26-B*38-DRB1*13				1.4	
A*29-B*14-DRB1*01					2.3
A*29-B*14-DRB1*13		1.0	1.9		

A*29-B*44-DRB1*07		2.0	2.0	1.5	0.8
A*29-B*44-DRB1*11	1.1			1.9	
A*29-B*44-DRB1*13			1.0		
A*30-B*08-DRB1*04					1.6
A*30-B*18-DRB1*03				3.8	
A*30-B*18-DRB1*13					0.5
A*30-B*35-DRB1*03				1.2	0.3
A*30-B*35-DRB1*10					2.3
A*30-B*44-DRB1*07	2.2				
A*30-B*57-DRB1*07					0.8
A*30-B*57-DRB1*13					0.8
A*32-B*08-DRB1*03	2.2				
A*33-B*14-DRB1*01	4.3				1.3
A*33-B*14-DRB1*13				2.0	
A*68-B*07-DRB1*02				2.0	
A*68-B*45-DRB1*04				3.0	

The complete list of haplotypes is available from the authors upon request.

shows it is the most frequent haplotype in NP (5.4%) but absent in CP and SP. *A*29-B*44-DRB1*07* was given as the second most common haplotype in Portugal (Arnaiz-Villena *et al.* 1997), but our data does not confirm this. Other haplotypes thought to be West-European were found in NP but not in other parts of Portugal (*A*30-B*44-DRB1*07*, *A*32-B*08-DRB1*03*).

Mediterranean haplotypes

Several have been given as characteristic of the circum-Mediterranean region, such as *A*33-B*14-DRB1*01* (Arnaiz-Villena *et al.* 1999; Gómez-Casado *et al.* 2001; Muro *et al.* 2001). Previously, the highest frequency has been attributed to Armenians (3.6%), but we have found the frequency in NP to be 4.3%.

Iberian-North African haplotypes

Among the many recognized haplotypes of Iberian-North African extract, *A*30-B*18-DRB1*03* has been described as an Iberian Paleo-North African marker (particularly the bi-loci *A*30-B*18* (Arnaiz-Villena *et al.* 1995; Gómez-Casado *et al.* 2001; Muro *et al.* 2001)). It was only found in the present study in SP (3.8%). The marker *A*02-B*51-DRB1*13* has been suggested to be characteristic of ‘Ibero-Berbers’ (Gómez-Casado *et al.* 2001). However, it is known historically and by the study of mtDNA (Pereira *et al.* 2000) that Berbers settled in NP, rather than CP and SP, and it is strange that we

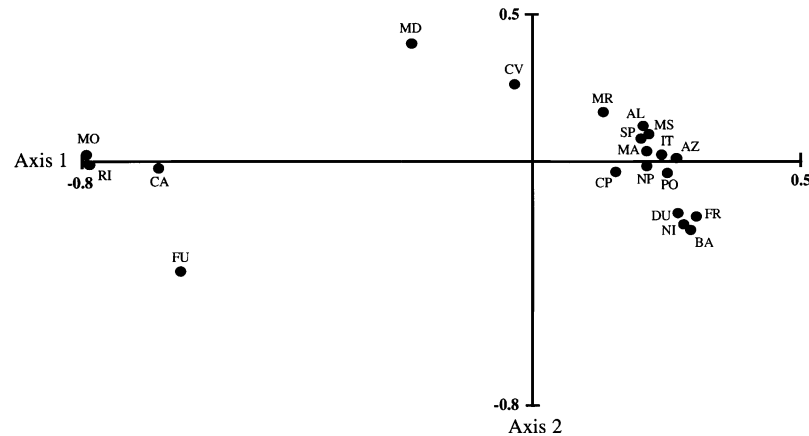


Fig. 2. Principal Component Analysis (PCA) showing the relationships among the populations described in Table 1 on the basis of their HLA allele frequencies. First axis extracted 50.52% of the total variance, second axis 10.34%.

find this haplotype in all populations but NP. Haplotype $A^*26-B^*38-DRB1^*13$, defined as specific to the Portuguese and probably present in the first Western Iberians (Arnaiz-Villena *et al.* 1997; Gómez-Casado *et al.* 2001), was not found in mainland Portugal in the present study, although it appears in MA (1.4%).

The sub-Saharan component

Few haplotypes have been unambiguously linked to sub-Saharan. Haplotype $A^*24-B^*44-DRB1^*11$ is considered such a marker (Bruges-Armas *et al.* 1999) but we did not find it in CV, despite it existing in MA (0.8%). Haplotype $A^*24-B^*44-DRB1^*02$ also exists only in Madeira, with a frequency of 1.4%, and is probably also of African origin. The San people have been characterized by possessing $A^*03-B^*58-DRB1^*04$ at a frequency of 1.1%, the same percentage as found in NP. Haplotype $A^*30-B^*08-DRB1^*04$ exists in 3.3% of the San and 1.6% of CV. Zimbabweans present haplotypes $A^*30-B^*57-DRB1^*07$ and $A^*30-B^*57-DRB1^*13$ (2% each), which are also found in CV (0.8% each). Finally, we found in CV and MA haplotypes that previously have been reported only in Afro-Americans: $A^*30-B^*35-DRB1^*03$ (0.7%); $A^*02-B^*53-DRB1^*13$ (0.9%); $A^*23-B^*07-DRB1^*02$ (2.6%). Several haplotypes could be considered typical sub-Saharan markers since they only appear in CV (Table 4). The haplotype

$A^*30-B^*18-DRB1^*13$, which exists in MA at 0.5% and CV at 2.3% but not in mainland Portugal, was most probably introduced by slaves.

The 'Oriental' component

Several haplotypes that are possibly 'oriental' in their origin have been described in the Azores (Bruges-Armas *et al.* 1999). This led to the hypothesis that a Mongoloid population existed in the Azores prior to discovery by the Portuguese. However, one of these haplotypes ($A^*02-B^*50-DRB1^*07$) was present in 2% of MA and another, $A^*24-B^*44-DRB1^*13$, that is found in 1–3% of the Western Caucasian population, appears in 4% of SP. The previous assumption that the presence in the Azores of these oriental haplotypes is due to habitation by American Indians before the Portuguese colonization in 1434 would now appear to be unsustainable.

Principal Component Analysis

The PCA depicted in Figure 2 is based on *HLA-A* and *-B* and *-DR* loci, although Burkina Faso and Cameroon populations lacked *-DR* data. In this particular case the program converts missing values into zero. We also performed a PCA based on just loci *-A* and *-B* for all populations, and the result was similar to that obtained when *-DR* values were included. This prompted us not to exclude the *-DR* data from

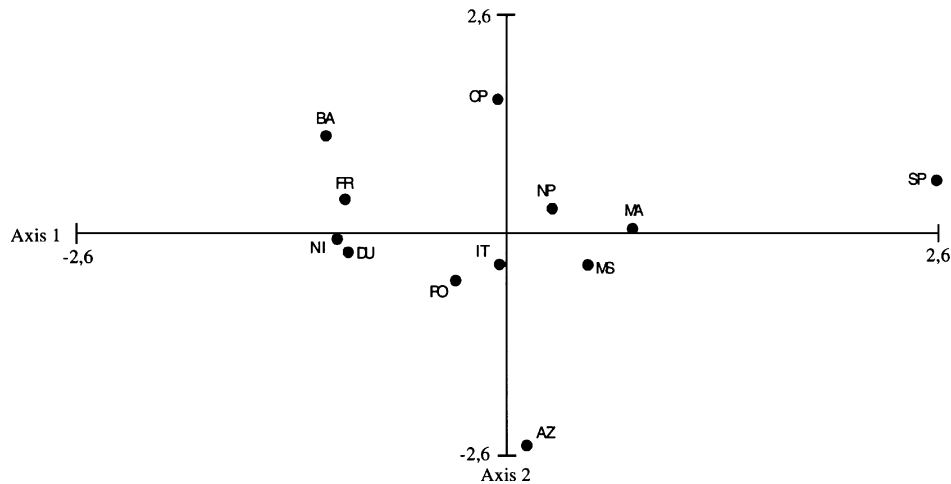


Fig. 3. Principal Component Analysis (PCA) restricted to the European samples in order to clarify the relationships involving the five populations included in the present study. The plot is based on data from the three loci. First axis extracted 20.71% of the total variation, second axis 19.02% and the third 14.8% (not shown).

the analysis. In the plot it is clear that the European populations form a tight cluster and are different from the sub-Saharan populations. The latter form a separate group that is much more diverse than the European one. The position of CV and MD is peculiar, as both are plotted outside the African group, almost joining the European pool. MD is a sub-Saharan population from Senegal, the most closely related region to the Cabo Verde Islands, and the plot clearly reveals this close relationship. CV and the European populations are placed close by, reflecting a similitude in frequencies of the most common alleles. Another intriguing aspect is the high heterogeneity among the four Portuguese samples when compared with the tight cluster formed by North-European populations (Figure 3). The Azores and SP appear separated from the remaining Portuguese populations, as seen in the plot in Figure 3, a fact supported by the high number of haplotypes unique to these populations (see Table 3).

Population comparisons

Two methods are at our disposal to study the genetic composition of a population and eventually unveil the origin of its present day structure: variation of allele frequencies, and haplotype distribution. Following the discovery of Madeira and Cabo Verde, settlement took place with

colonizers mainly originating from Portugal. In 1670 Madeira Island already had one fifth of today's population. Apart from Iberians and north Europeans (Flemish), other peoples contributed to the settlement: moors (white-slaves from Mauritania), sub-Saharan slaves introduced directly from Cabo Verde (at certain times they comprised 10% of the population), Jews, slaves from India and a few natives from the Canary Islands (Guanches) (Pereira, 1989). It is possible to track the genetic imprint left by some of these peoples. The population of Madeira presents rare and African like alleles (*HLA-A*34*, *HLA-A*36* and *HLA-A*74*), distinguishing it from other Portuguese populations. Also, 3.5% of its haplotypes are of African origin (data not shown). MtDNA data also shows the extension of typical African haplogroups (L and M1) in Madeira, 20.4% of the total compared to only 6.5% in the Azores (Brehm, unpublished data). The same applies for STR microsatellite distribution (Fernandes *et al.* 2001). It is more difficult to track the North African input due to the slave trade. The three populations from mainland Portugal show remarkable differences, which may not be apparent through the allele frequencies comparison alone. For example, although alleles *A*02* and *B*44* are common in all five populations studied, they combine with different *DRB1* alleles to form different haplo-

types, enabling populations to be discriminated. Table 4 presents several examples of such cases of haplotypes that are extremely frequent but unique to each population (*A*02-B*44-DRB1*02*, *A*02-B*44-DRB1*04*, *A*02-B*44-DRB1*07*). The heterogeneity of the three sampled populations from mainland Portugal reveals that, contrary to what has been found previously (Arnaiz-Villena *et al.* 1997), there is a strong Mediterranean and North African genetic background in mainland Portugal (as well in Madeira and Cabo Verde), as suggested by high frequencies of haplotypes *A*33-B*14* and *A*30-B*18*. Nevertheless, this is not enough to support the hypothesis of a pre-Neolithic migration of peoples from North Africa to the Iberian Peninsula (Arnaiz-Villena *et al.* 1995), since two of the most common North African haplotypes (*A*03-B*49-DRB1*11* and *A*26-B*44-DRB1*15*) were not found in our samples, which should be expected to occur under such an emigrational model. The existence of a higher concentration of these 'Ibero Berber' markers in the north of Portugal is in accordance with mtDNA markers (Pereira *et al.* 2000). In fact, it is well documented that Berbers were confined to the North by non-Berber Arab armies during the fight for the control of Central and South Iberia.

European presence in Cabo Verde, although never more than 10% of the population, is perceivable at the HLA level. Cabo Verde has alleles found in relatively high percentages in Portuguese populations but not in typical West African Mandenka or Cameroon populations (Table 2, *HLA-A*01*, *HLA-B*14*, *HLA-B*40* and *HLA-B*51*), and the more frequent Ibero-Mediterranean haplotypes (*A*29-B*44-DRB1*07* and *A*01-B*08-DRB1*03*, Arnaiz-Villena *et al.* 1997) or even Neolith Iberic markers (*A*30-B*18*, Martínez-Laso *et al.* 1995). We can only be sure that these haplotypes are originally from the Iberia Peninsula, and not carried with the slaves themselves, after a complete survey is performed in the West African coast Guinean population. All together the non sub-Saharan haplotypes in Cabo Verde comprise 15% of the total (data not shown). The presence of these markers show

that the present day population of Cabo Verde is atypical when compared with other African ones, reflecting how its origin was due to admixture of an African substrate with European genetic input.

The authors are grateful to the Portuguese Chief of Army, and the Chairman of the Joint Chiefs of Staff from the Republic of Cabo Verde, for allowing collecting blood samples. We thank also Matthew Stephen for his help with PHASE calculations. This research was partially supported by a Ph.D. grant attributed to H.S. (FSE/Regional Government of Madeira, Portugal). We acknowledge the support of ICCTI (Ministry of Science and Technology, Lisbon) to A.B. and Northern Ireland Histocompatibility and Immunogenetics Charity Trust Fund to F.W.

REFERENCES

- Arnaiz-Villena, A., Benmamar, D., Alvarez, M., Diaz-Campos, N., Varela, P., Gomez-Casado, E. & Martinez-Laso, J. (1995). HLA allele and haplotype frequencies in Algerians. Relatedness to Spaniards and Basques. *Hum. Immunol.* **43**, 259–268.
- Arnaiz-Villena, A., Dimitroski, K., Pacho, A., Moscoso, J., Gomez-Casado, E., Silvera-Redondo, C., Varela, P., Blagoevska, M., Zdravkovska, V. & Martinez-Laso, M. (2001a). HLA in Macedonians and the sub-Saharan origin of the Greeks. *Tissue Antigens* **57**, 118–127.
- Arnaiz-Villena, A., Iliakis, P., Gonzalez-Hevilla, M., Longas, J., Gomez-Casado, E., Sfyridaki, K., Trapaga, J., Silvera-Redondo, C., Matsouka, C. & Martinez-Laso, J. (1999). The origin of Cretan populations as determined by characterization of HLA alleles. *Tissue Antigens* **53**, 213–226.
- Arnaiz-Villena, A., Martínez-Laso, J., Gómez-Casado, E., Díaz-Campos, N., Santo, P., Martinho, A. & Breda-Coimbra, H. (1997). Relatedness among Basques, Portuguese, Spaniards, and Algerians studied by HLA allelic frequencies and haplotypes. *Immunogenetics* **47**, 37–43.
- Arnaiz-Villena, A., Karin, M., Bendikuzze, N., Gomez-Casado, E., Moscoso, J., Silvera, C., Oguz, F. S., Sarper Diler, A., Pacho, A., Allende, L., Guillen, J. & Martinez Laso, L. (2001b). HLA alleles and haplotypes in the Turkish population: relatedness to Kurds, Armenians and other Mediterraneans. *Tissue Antigens* **57**, 308–317.
- Bera, O., Cesaire, R., Quelvennec, E., Quillivic, F., Chavigny, V., Ribal, C. & Semana, G. (2001). HLA class I and class II allele and haplotype diversity in Martinicans. *Tissue Antigens* **57**, 200–207.
- Bruges-Armas, J., Martinez-Laso, J., Martins, B., Allende, L., Gomez-Casado, E., Longas, J., Varela, P., Gonzalez-Granado, M. & Arnaiz-Villena, A. (1999). HLA in the Azores Archipelago: possible presence of Mongoloid genes. *Tissue Antigens* **54**, 349–359.
- Carreira, A. (1983). *Migrações nas Ilhas de Cabo Verde*. Instituto Caboverdeano do Livro (ed.), 2nd ed., Lisboa.

- Chimge, N., Tanaka, H., Kashiwase, K., Ayush, D., Tokunaga, K., Saji, H., Akaza, T., Batsuuri, J. & Juji, T. (1997). HLA system in the population of Mongolia. *Tissue Antigens* **49**, 477–484.
- Clayton, J., Lonjou, C. & Whittle, D. (1997). Allele and Haplotype frequencies for HLA loci in various ethnic groups. In *Genetic diversity of HLA. Functional and medical implications*. (ed. Charron, D.), pp. 665–820. Paris: EDK.
- Daniel, W. W. (1987). *Biostatistics: a foundation for analysis in the health sciences*. London: John Wiley & Sons.
- Dard, P., Schreiber, Y., Excoffier, L., Sanchez-Mazas, A., Shi-Isaac, X., Epelbouin, A., Langaney, A. & Jeannet, M. (1992). Polymorphism of HLA class I loci HLA-A, -B and -C in a Mandenka population from Eastern Senegal. *C.R. Acad. Sci. Paris* **314**, 573–578.
- Ellis, J. M., Mack, S. J., Leke, R. F. G., Quakyi, I., Johnson, A. H. & Hurley, C. K. (2000). Diversity is demonstrated in class I HLA-A and HLA-B alleles in Cameroon, Africa: description of HLA-A*03012, *2612, *3006 and HLA-B1403, *4016, *4703. *Tissue Antigens* **56**, 291–302.
- Excoffier, L., Smouse, P. E. & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491.
- Fernandes, A. T., Brehm, A., Alves, C., Gusmão, L. & Amorim, A. (2001). Genetic profile of the Madeira Archipelago population using the new PowerPlex® 16 System Kit. *Forensic Sci. Int.* **3228**, 1–3.
- Godinho, V. M. (1965). *Os descobrimentos e a economia mundial*. Arcádia ed., Vol. 1–2, Lisboa.
- Gómez-Casado, E., del Moral, P., Martínez-Laso, J., García-Gómez, A., Allende, L., Silvera-Redondo, C., Longas, J., González-Hevilla, M., Kandil, M., Zamora, J. & Arnaiz-Villena, A. (2001). HLA genes in Arabic-speaking Moroccans: close relatedness to Berbers and Iberians. *Tissue Antigens* **55**, 239–249.
- Hill, A. V. S., Allsopp, C. E. M., Kwiatkowski, D., Taylor, T. E., Yates, S. N. R., Anstey, N. M., Wirima, J. J., Brewster, D. R., McMichael, A. J., Molyneux, M. E. & Greenwood, B. M. (1992). Extensive genetic diversity in the HLA class II region of Africans, with a focally predominant allele, DRB1*1304. *Proc. Natl. Acad. Sci. USA* **89**, 2277–2281.
- Ivanova, M., Spassova, P., Michailova, A. & Naumova, E. (2001). Distributions of HLA class I alleles and haplotypes in Bulgarians – contribution to understanding the origin of the population. *Tissue Antigens* **57**, 208–215.
- Lareu, M. U., Phillips, C. P., Carracedo, A., Lincoln, A. J., Syndercomb, E., Court, D. & Thomson, J. A. (1994). Investigation of the STR locus HUMTH01 using PCR and two electrophoresis formats: UK and Galician Caucasian population surveys and usefulness in paternity investigations. *Forensic Sci. Int.* **66**, 41–52.
- Martinez-Laso, J., De Juan, D., Martinez-Quiles, N., Gomez-Casado, E., Cuarado, E. & Arnaiz-Villena, A. (1995). The contribution of the HLA-A, -B, -C and -DR, -DQ DNA typing to the study of the origin of Spaniards and Basques. *Tissue Antigens* **45**, 237–245.
- Martinho, A., Simões, O., Santos, P., Ferreira, R., Paiva, A., Mateus, R., Ferreira, A., Gomes, J. & Brêda-Coimbra, H. (1995). Analysis of HLA-DR, DQA1 and DQB1 polymorphism in a Portuguese population. *Eur. J. Immunogenetics* **22**, 125.
- Middleton, D. (2000). PCR-SSOP Class I and and Class II (DRB1). In *ASHI Laboratory Manual, 4th edition*. (eds. G. A. Land, R. M. Strothman & A. B. Hahn). Lenexa, KS: American Society for Histocompatibility and Immunogenetics.
- Middleton, D., Williams, F., Hamill, M. A. & Meenagh, A. (2000). Frequency of HLA-B alleles in a Caucasoid population determined by a two-stage PCR-SSOP typing strategy. *Hum. Imm.* **61**, 1285–1297.
- Modiano, D., Luoni, G., Petrarca, V., Sodiomon Sirima, B., De Luca, M., Simporé, J., Coluzzi, M., Bodmer, J. G. & Modiano, G. (2001). HLA class I in three West African ethnic groups: genetic distances from sub-Saharan and Caucasoid populations. *Tissue Antigens* **57**, 128–137.
- Moghaddam, P. H., de Knijff, P., Schipper, R. F., Kazemi, M., Lambooi, S. L. E., van der Meer, R., Naipal, A., Pesonen, N., van der Slik, A. & Gliphart, M. J. (1998). Selective co-evolution of the D6STNfa microsatellite region with HLA class I and II loci. *Tissue Antigens* **52**, 213–219.
- Muro, M., Marín, L., Torío, A., Moya-Quiles, M., Minguela, A., Rosique-Roman, J., Sanchis, M., Garcia-Calatayud, M., Garcia-Alonso, A. & Alvarz-López, M. (2001). HLA polymorphism in the Murcia population (Spain): in the cradle of the archeological Iberians. *Hum. Immunology* **62**, 910–921.
- Pereira, E. C. N. (1989). *Ilhas de Zarco* (4th edition, Vol. II). Funchal: Câmara Municipal Funchal.
- 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.
- Piazza, A. & Lonjou, C. (1997). HLA in Europe and other Mediterranean countries. In *Genetic diversity of HLA. Functional and medical implications*. Vol I. (ed. D. Charron), pp. 374–386. Paris: EDK.
- Pinto, F., González, A. M., Hernández, M., Larruga, J. & Cabrera, V. (1996). Genetic relationship between the Canary Islanders and their African and Spanish ancestors inferred from mtDNA sequences. *Ann. Hum. Genet.* **60**, 321–330.
- Rando, J. C., Pinto, F., González, A. M., Hernández, M., Larruga, J. M., Cabrera, V. M. & Bandelt, H.-J. (1998). Mitochondrial DNA analysis of North-west African populations reveals genetic exchanges with European, Near-Eastern, and Sub-Saharan populations. *Ann. Hum. Genet.* **62**, 531–550.
- Rendine, S., Borelli, I., Barbanti, M., Sacchi, N., Roggero, S. & Curtioni, E. S. (1998). HLA polymorphisms in Italian bone marrow donors: a regional analysis. *Tissue Antigens* **52**, 135–146.
- Russell-Wood, A. J. (1998). *The Portuguese empire, 1415–1808. A world on the move*. Baltimore, London: Johns Hopkins Univ. Press.
- Ronningen, K. S., Spurkland, A., Markussen, G., Iwe, T., Vartdal, F. & Thorsby, E. (1990). Distribution of HLA class II alleles among Norwegian Caucasians. *Hum. Immunol.* **29**, 275–281.
- Sanchez-Mazas, A. (2001). African diversity from the HLA point of view: influence of genetic drift, geography, linguistics, and natural selection. *Hum. Immunology* **62**, 937–948.

- Sánchez-Velasco, P., Escribano de Diego, J., Paz-Miguel, J. E., Oejo-Vinyals, G. & Leyva-Cobián, F. (1999). HLA-DR, DQ nucleotide sequence polymorphisms in the Pasiegos (Pas valleys, Northern Spain) and comparison of the allelic and haplotypic frequencies with those of other European populations. *Tissue Antigens* **53**, 65–73.
- Santos, P., Fonseca, A. M., Martinho, A., Ferreira, R. V., Simões, O., Mateus, M. R., Gomes, J., Silva, J. M., Freitas, A., Paiva, A. & Brêda-Coimbra, H. (1997). HLA-A,B,C polymorphism in Portuguese Caucasoids by ARMS-PCR SSP based typing. *Eur. J. Immunogenetics* **24**, 15.
- Schipper, R. F., D'Amaro, J., de Lange, P., Schreuder, G. M., van Rood, J. J. & Oudshoorn, M. (1998). Validation of haplotype frequency estimation methods. *Hum Immunol.* **59**, 518–523.
- Schneider, S., Kueffer, J.-M., Roessli, D. & Excoffier, L. (2000). Arlequin v2.000: a software for population genetic data analysis. Geneva, Switzerland: University of Geneva.
- Scozzari, R., Cruciani, F., Pangrazio, A., Santolamazza, P., Vona, G., Moral, P., Latini, V., Varesi, L., Memmi, M. M., Romano, V., De Leo, G., Gennarelli, M., Jaruzelska, J., Villems, R., Parik, J., Macaulay, V. & Torroni, A. (2001). Human Y-chromosome variation in the western Mediterranean area: implications for the peopling of the region. *Hum Immunol.* **62**, 871–884.
- Stephens, M., Smith, N. J. & Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* **68**, 978–989.
- Tiercy, J.-M., Sanchez-Mazas, A., Excoffier, L., Shi-Isaac, X., Jeannet, M., Mach, B. & Langaney, A. (1992). HLA-DR polymorphism in a Senegalese Mandenka population: DNA oligotyping and population genetics of DRB1 specificities. *Am. J. Hum. Genet.* **51**, 592–608.
- Uko, G. P., Asuquo, M. A., Dawkins, R. L., Udim, E. R., Banjoko, S. O. & Jaja, O. (1997). HLA class I phenotype and gene frequencies in parts of southern Nigeria. *Disease Markers* **13**, 123–130.
- Williams, F., Meenagh, A., Maxwell, A. P. & Middleton, D. (1999). Allele resolution of HLA-A using oligonucleotide probes in a two stage typing strategy. *Tissue Antigens* **54**, 59–68.