HLA polymorphisms in Forros and Angolares from São Tomé Island (West Africa): Evidence for the Population Origin

Nádia Saldanha[a], Carla Spínola[a], Margarida R. Santos[b], Joaquim P. Simões[b], Jácome Bruges-Armas[b], António Brehm[a], Hélder Spínola[a]

Abstract

São Tomé Island in the West coast of Africa, in the Gulf of Guinea, was discovered uninhabited in 1470 and settled by Portuguese and people of different origins in sub-Saharan Africa, mostly slaves recruited from the Gulf of Guinea, Congo and Angola. During the settlement process, sub-Saharan Africans from different geographic and ethnic origins mixed together on São Tomé Island and, to some extent, also with Portuguese. The main ethnic group of São Tomé Island are the Forros, descendants of liberated slaves who speak a Creole language with mixed Portuguese and Bantu. The Angolares, another ethnic group, are probably descendants of slaves who escaped from plantations and practiced endogamy while maintaining their own Bantu language.

HLA-A, HLA-B, and HLA–DRB1 loci polymorphisms were typed using high-resolution sequence-based typing in these two ethnic groups. Allele frequencies, haplotypes and phylogenetic analysis confirm that the West Coast of Africa is the place of origin of São Tomé Island's main genetic pool. The Forros and Angolares systematically cluster together in phylogenetic analysis and are not statistically different from each other which makes plausible the hypothesis that Angolares are descendants of slaves who escaped from plantations and practiced endogamy.

Introduction

The human leukocyte antigen (HLA) system, the major histocompatibility complex in humans, includes the most highly polymorphic loci in the human genome and is located on the short arm of chromosome 6 (6p21.3), spanning over 4 Mb of DNA (Bodmer, 1987; Klein and Sato, 2000; Naik, 2003). It consists of a closely linked set of genes highly important for medical purposes, namely in transplantation, autoimmune diseases and allergies (Boehncke et al, 1998; Gilbert et al, 2003; Riley and Olerup, 1992). The most polymorphic HLA loci (e.g. HLA-A, HLA-B, HLA-DRB1) have been used in population studies in order to assess gene flow on the basis of allele and haplotype frequencies. HLA loci have been successfully used to analyze populations according to geography which makes them good genetic markers for population studies (Arnaiz-Villena et al, 2002; Cao et al, 2004; Sanchez-Mazas, 2001; Spínola et al, 2002; Spínola et al, 2005a). Other molecular markers, like autosomic and Y-chromosome STR *loci* and mtDNA, are widely used for this type of research as a complement to HLA *loci* (Gonçalves et al, 2002; Rosa et al, 2004; Rosa et al, 2006).

São Tomé and Príncipe islands, located 300 km from the West coast of Africa in the Gulf of Guinea as shown in **Figure 1**, were discovered uninhabited by Portuguese sailors in 1470 (Peres, 1960). The archipelago was settled first by people from different regions of sub-Saharan Africa, mostly slaves from the Gulf of Guinea, Congo, and Angola, brought to work in local plantations, and, to a minor extent, Portuguese involved in the slave trade between Africa and the Americas. In the first centuries after the discovery of São Tomé and Príncipe, beside the Portuguese, other Europeans were involved in the slave trade along the coast of Africa, namely French, Spanish, Dutch, and English. These people could have contributed on a minor scale to the present-day genetic pool of this archipelago (Neves, 1989). During the settlement process, Portuguese males commonly choose female slaves as mates (Garcia, 1966), and there was also mixing of slaves from different geographic and ethnic origins (Tenreiro, 1961).

In the 19th century a new economic cycle based on coffee and cacao plantations brought to the islands a new wave of sub-Saharan African people from Cabo Verde archipelago, Angola and Mozambique (Barata, 1966).

Address for correspondence: Hélder Spínola, hspinola@uma.pt
a University of Madeira, Human Genetics Lab, Funchal, Portugal
b Institute for Molecular and Cell Biology (IBMC), Porto, Portugal
Received: July 13, 2009; accepted: September 24, 2009

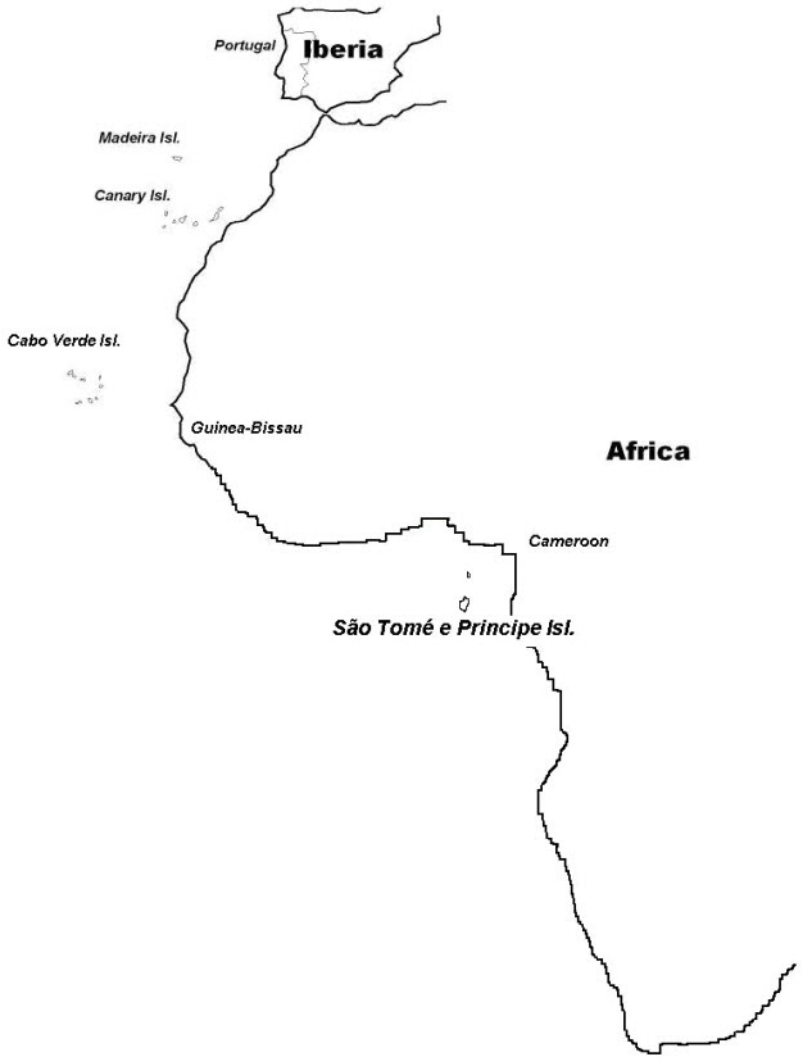


Figure 1. The geographic location of São Tomé e Príncipe archipelago (Gulf of Guinea – West Africa).

Most people on São Tomé speak Forro, a Creole language with mixed Portuguese and Bantu languages used by liberated slaves, known as Forros, considered the first African inhabitants of the archipelago (Henriques, 2000; Tenreiro, 1961). Two other ethnic groups using mixed Portuguese and African Creole languages are the *Mancó,* mostly from Príncipe Island, and the *Tonga*, descendants from people who arrived during the 19th century, after slave abolition. In contrast to the other ethnic groups, the Angolar community inhabiting São Tomé Island has resisted admixture and still maintains its own Bantu language. The origin of the Angolar people remains uncertain but their popular oral traditions say that they are descendants of the survivors of a slave shipwreck in the middle of the 16th century (Henriques, 2000; Romana, 1997). Probably, however, Angolares are just descendants of slaves who escaped from plantations and took take refuge in the most inaccessible forest of the southeastern region of São Tomé (Seibert, 1998).

The diverse origins of the populations of São Tomé and Príncipe archipelagos make them an interesting group in which to study genetic admixture, especially with Afri-

can and European backgrounds. Previous studies on mtDNA revealed that the maternal lineages on São Tomé and Príncipe were almost completely of sub-Saharan origin, clearly belonging to a West African cluster (Mateu et al, 1997; Trovoada et al, 2004).

Several autosomal markers (β-globin haplotypes, APOA1, AT3, FY, LPL, OCA2, RB1, Sb19.3, and GC) have shown that the peopling of São Tomé island combined diverse African contributions and European admixture (10.7%) that emerged from the overseas population relocations promoted by the Atlantic slave trade (Tomas et al 2002). Studies on Y-chromosome markers (STR and SNP) detected European ancestry on São Tomé and Principe and showed differences in the frequency of European haplogroups between the Angolares and Forros populations (Gonçalves et al, 2007; Trovoada et al, 2001).

The main aim of the present work was to analyse the HLA-A, HLA-B, and HLA-DRB1 loci allele and haplotype frequencies on present-day São Tomé and to identify European and African genetic influences. In the study we also searched for genetic differentiation between the Angolares and the Forros, the two main and most ancient ethnic groups of this archipelago.

Materials and Methods

Subjects

The present study population consisted of a total of 98 healthy unrelated males from São Tomé Island (West Africa). Blood samples were collected after informed consent from donors whose parents and grandparents were born in the archipelago. The ethnic group of donors and their ancestors were registered and samples were identified as belonging to two different groups: Forros (n = 66) and Angolares (n = 32). Genomic DNA was isolated from whole blood containing EDTA using a salting-out procedure accordingly to Miller *et al.* (1988), with some modifications.

All subjects were typed using high-resolution sequence-based typing (SBT) for HLA-A and HLA-B according to Kurz *et al.* (1999) and Pozzi *et al.* (1999), and for HLA-DRB1 using specific primers of PCR-Sequence Specific Oligonucleotide Probes (SSOP) typing (Williams et al, 2004), as previously described (Spínola et al, 2005b).

Data Analysis

Basic genetic parameters (allele and haplotype frequencies, gene diversity, and Hardy-Weinberg equilibrium) were estimated with Arlequin v2.000 (Excoffier et al, 2005) at the three HLA loci. In the present study the Ewens-Watterson neutrality test was applied to examine the presence of selective forces influencing allelic diversity at these loci.

An analysis of molecular variance (AMOVA) was performed with Forros and Angolares groups based on Euclidean distances (Excoffier et al, 1992). Variance components were tested for significance by non parametric randomisation tests using 10,000 permutations under the null hypothesis of no population structure. The population genetic software Arlequin v2.000 was employed in all the above analyses.

Comparative analysis of the Forros and Angolares with other populations available in the literature with the same typing resolution was achieved using the software included in the PHYLIP v.3.6 software package (Felsenstein, 2004). The populations used from the literature were: Kenya, Mali, Zambia, Uganda (Cao et al, 2004), Cabo Verde, Guinea Bissau (Spínola et al, 2005c), Portugal (Spínola et al, 2005b) Madeira Island (Spínola et al, 2006) and, from the New Allele Frequency Database (Middleton, et al., 2003) web site, Italy, France, Czech Republic, India, Sudan, Cameroon, Zimbabwe (Shona), Zulu, Tunisia and Morocco. First, SEQBOOT was used to perform a bootstrap analysis from gene frequency data. The program generates multiple data sets re-sampled from the original data. Distance matrices from each replicate data set were generated using GENDIST and used as input to NEIGHBOR to produce neighbour-joining trees. A single consensus bootstrapped tree was obtained with CONSENSUS. The topology was visualized with DrawTree (Felsenstein, 2004). Principal coordinates analysis (PCO) using HLA-A, HLA-B and HLA-DRB1 allele frequencies was carried out on the MultiVariate Statistical Package MVSP3 (Kovach, 2006).

Results

Table 1 shows HLA-A, HLA-B, and HLA-DRB1 allele frequencies in Forros and Angolares. A total of 29 and 23 HLA-A, 38 and 23 HLA-B, and 29 and 28 HLA-DRB1 alleles were found in Forros and Angolares, respectively. Forros (HLA-A 0.93, HLA-B 0.93, and HLA-DRB1 0.94) and Angolares (HLA-A 0.95, HLA-B 0.92, and HLA-DRB1 0.94) presented high values of heterozygosity and yielded non-significant results in the Ewens-Watterson neutrality test, except HLA-A in Angolares (P=0.04). Except for HLA-A in Angolares (P=0.001) and HLA-B in Forros (P=0.02), all three loci showed Hardy-Weinberg equilibrium in both groups. The exact test of population differentiation, performed by Arlequin, shows no significant differences between Forros and Angolares (P=0.37). However, all both groups are significantly different from all other populations included in the comparison (P<0.001).

Allele Frequencies

The most frequent HLA-A alleles found in Forros were A*0201, A*2301, and A*6802 (13% each). With a similar frequency (14%), HLA-A*6802 was also the

Table 1
HLA-A, HLA-B and HLA-DRB1 Allele Frequencies in the Forros and Angolares from São Tomé Island.

HLA-A	Forros N=66	Angolares N=32	HLA-B	Forros N=66	Angolares N=32	HLA-DRB1	Forros N=66	Angolares N=32
A*0101	0.030	0.047	B*0702	0.098	0.063	DRB1*0101	0.008	0.047
A*0201	0.128	0.063	B*0708	0.015	0	DRB1*0102	0.053	0.031
A*0202	0.045	0.063	B*0801	0.045	0.063	DRB1*0301	0.120	0.181
A*0205	0.038	0.016	B*0803	0.023	0	DRB1*0302	0.053	0.109
A*0217	0.008	0	B*1302	0.008	0	DRB1*0303	0.008	0
A*0301	0.060	0.031	B*1402	0.008	0	DRB1*0305	0.023	0.016
A*1101	0	0.031	B*1501	0.015	0.016	DRB1*0307	0.008	0
A*2301	0.128	0.092	B*1503	0.045	0.063	DRB1*0401	0.023	0.016
A*2304	0.008	0	B*1510	0.076	0.031	DRB1*0405	0.008	0
A*2305	0.008	0	B*1517	0	0.016	DRB1*0701	0.091	0.016
A*2402	0.015	0.047	B*1518	0.008	0	DRB1*0704	0	0.016
A*2403	0	0.016	B*1523	0.008	0	DRB1*0801	0.015	0
A*2601	0.008	0	B*1801	0.015	0.031	DRB1*0804	0.008	0.063
A*2901	0.008	0	B*1804	0	0.016	DRB1*0901	0.030	0.016
A*2902	0.030	0.016	B*2705	0.008	0	DRB1*0902	0	0.016
A*2903	0	0.016	B*3501	0.053	0.063	DRB1*1001	0.023	0.016
A*3001	0.060	0.016	B*3801	0.008	0	DRB1*1101	0.083	0.047
A*3002	0.008	0.031	B*4102	0.008	0.016	DRB1*1102	0.023	0.016
A*3004	0.023	0.031	B*4201	0.023	0.031	DRB1*1104	0.008	0
A*3101	0.023	0	B*4402	0.03	0	DRB1*1107	0	0.016
A*3201	0.008	0	B*4403	0.015	0	DRB1*1201	0.03	0.031
A*3301	0.023	0.047	B*4501	0.008	0.016	DRB1*1202	0	0.016
A*3303	0.045	0.047	B*4504	0.008	0	DRB1*1301	0.068	0.016
A*3402	0.038	0.047	B*4901	0.008	0.047	DRB1*1302	0.023	0
A*3601	0.030	0.063	B*4903	0.008	0	DRB1*1303	0.076	0.063
A*6601	0.008	0.031	B*5101	0.061	0.063	DRB1*1304	0	0.016
A*6602	0.008	0.016	B*5104	0.015	0	DRB1*1307	0	0.016
A*6801	0.038	0	B*5106	0.015	0.031	DRB1*1310	0.008	0
A*6802	0.128	0.139	B*5201	0.008	0.016	DRB1*1312	0.008	0
A*6901	0.008	0	B*5301	0.192	0.244	DRB1*1405	0.008	0
A*7401	0.030	0.063	B*5302	0.008	0.016	DRB1*1407	0	0.016
A*7403	0	0.031	B*5303	0.023	0	DRB1*1408	0	0.016
A*8001	0.008	0	B*5304	0	0.016	DRB1*1501	0.023	0.031
			B*5305	0.008	0	DRB1*1503	0.110	0.109
			B*5307	0.008	0	DRB1*1505	0	0.016
			B*5308	0.008	0.016	DRB1*1507	0.008	0
			B*5703	0.030	0.016	DRB1*1601	0.008	0.016
			B*5801	0.030	0.047	DRB1*1602	0.045	0
			B*5802	0.030	0.063	DRB1*1608	0	0.016
			B*6701	0.008	0			
			B*8101	0.015	0			

most frequent allele in Angolares, followed by A*2301 (9.2%). HLA*0201 tends to show lower frequencies in sub-Saharans than in Europeans (Middleton et al., 2003), which is in agreement with frequencies found in São Tomé Island. A*2301 and A*6802 are two typical sub-Saharan high-frequency alleles. In Forros and Angolares, A*2301 shows frequencies intermediate between West (15-23%) and East Africans (6-8%) (Middleton et al., 2003). The higher frequency of HLA-A*2301 in Forros, compared to Angolares, could denote a greater genetic influence of populations from the Gulf of Guinea and the Northwest coast of Africa, where this allele reaches high frequencies, and from where slaves were brought to the archipelago. HLA-A*6802 in sub-Saharans appears at higher frequencies than HLA-A*6801, as was also found in both groups of São Tomé Island, but the opposite is found in Europeans.

HLA-B*5301 was the most frequent HLA-B allele in Forros (19%) and Angolares (24%). HLA-B*5301 is common in sub-Saharans, reaching the highest frequencies in Burkina Faso (22.3%) and Mali (16%), but is rare or absent in Europeans and Asians. The HLA-B*5802 allele, one of the next most frequent in Angolares (6.3%) and with 3% in Forros, is a typical allele from sub-Saharans, for which the highest frequencies have been found in Cameroon (14.3%) and Kenya (12.5%), and is rare or absent in Europeans and Asians. HLA-B*0702 was the other most common allele in Forros (9.8%) and Angolares (6.3%), reaching such high frequencies in sub-Saharans only in Cameroons (8%).

The most frequent HLA-DRB1 allele in the São Tomé Island population was HLA-DRB1*0301 (Forros 12% and Angolares 18%) and HLA-DRB1*1503 (11% in both Forros and Angolares). The HLA-DRB1*1503 allele is almost absent in other world populations, as opposed to sub-Saharans where it reaches frequencies as high as 29% in Cameroon or 17% in Rwanda. HLA-DRB1*0302 is a typical sub-Saharan allele with frequencies ranging from 3 to 9% but reaching 11% in Angolares.

Haplotype Frequencies

The exact test of linkage disequilibrium between the three pairs of loci, an extended Fisher's Exact Test performed with Arlequin v2.000 (Excoffier et al, 2005), was statistically significant only between HLA-A and HLA-B (P=0.016), and HLA-B and HLA-DRB1 (P=0.02) in the Forros.

The most representative three- and two-loci haplotypes with statistically significant linkage disequilibrium in the Forros and Angolares are listed in **Table 2**. The complete list of two- and three-loci haplotypes found in the Forros and Angolares is available the **Supplementary Data** file.

A*6802-B*0702-DRB1*1301 in Forros (3%) and A*6802-B*5301-DRB1*0804 in Angolares (4.7%) were the most frequent three-loci haplotypes found in each group. These two haplotypes were specific to each group and we didn't find them in other sub-Saharan populations, probably due to the very small number of African populations typed on the three loci considered. However, the related two-loci haplotype, A*6802-B*0702, was also present in Kenya (1.1%), Zulu (2%), Uganda (1.2%) and Zambia (2.3%), and the A*6802-B*5301 was found in Kenya (1.6%) and Mali (1.8%) (Middleton et al., 2003; Cao et al, 2004).

The second most frequent haplotype in Forros was A*0201-B*5101-DRB1*0701 with 2.3%. This haplotype was also found in the North of Portugal (1.1%), and in the oriental Azores islands (2.6%) (Middleton et al., 2003; Cao et al, 2004).

The most frequent two-loci haplotypes in Forros were A*6802-B*5301 (5.9%), also present in Kenya (1.6%), A*2301-DRB1*0301 (3.8%) and B*5301-DRB1*1101 (4%) (Middleton et al., 2003; Cao et al, 2004). Angolares also show some frequent two-loci haplotypes common to Forros, namely the A*6802-B*5301 (7.8%) and A*2301-DRB1*0301 (4.7%) haplotypes. The other most frequent two-loci haplotypes in Angolares were absent in Forros.

Phylogenetic Analyses

A dendrogram constructed with HLA-A, HLA-B and HLA-DRB1 is shown in **Figure 2a**, or just with class I (HLA-A and HLA-B) allele frequencies in **Figure 2b.** These figures show a close relationship between the São Tomé Island population and sub-Saharans, particularly with the geographically nearby populations from the West coast of Africa. Forros and Angolares cluster with each other and not far from Guinea-Bissau, Mali, and Cameroon. A dendrogram constructed with low resolution HLA-A and HLA-B (data not shown), in order to include some sub-Saharan populations with no high resolution typing, reveals that Forros and Angolares ethnic groups cluster to Cameroon, Podokwo, and Uldeme (unpublished data), to Mali and to Burkina Fasso Rimaibe and Mossi ethnic groups. A Principal Coordinate Analysis is shown in **Figure 3** and is consistent with the dendrograms plotting Forros and Angolares not far from Guinea-Bissau and Cameroon.

Discussion

São Tomé and Príncipe archipelago was settled after the year 1470 by people from different origins, primarily sub-Saharan Africa, and, to a minor extent, Europe, mostly Portuguese (Neves, 1989). The present-day population of São Tomé and Príncipe consists of Forros (the

Table 2
Most Frequent Haplotypes in Forros and Angolares from São Tomé Population Estimated by Maximum Likelihood, with the Respective Frequencies (Freq), Value of Linkage Disequilibrium (D) and their Level of Significance (P). Only haplotypes with higher frequencies with statistical significance linkage disequilibrium are shown.

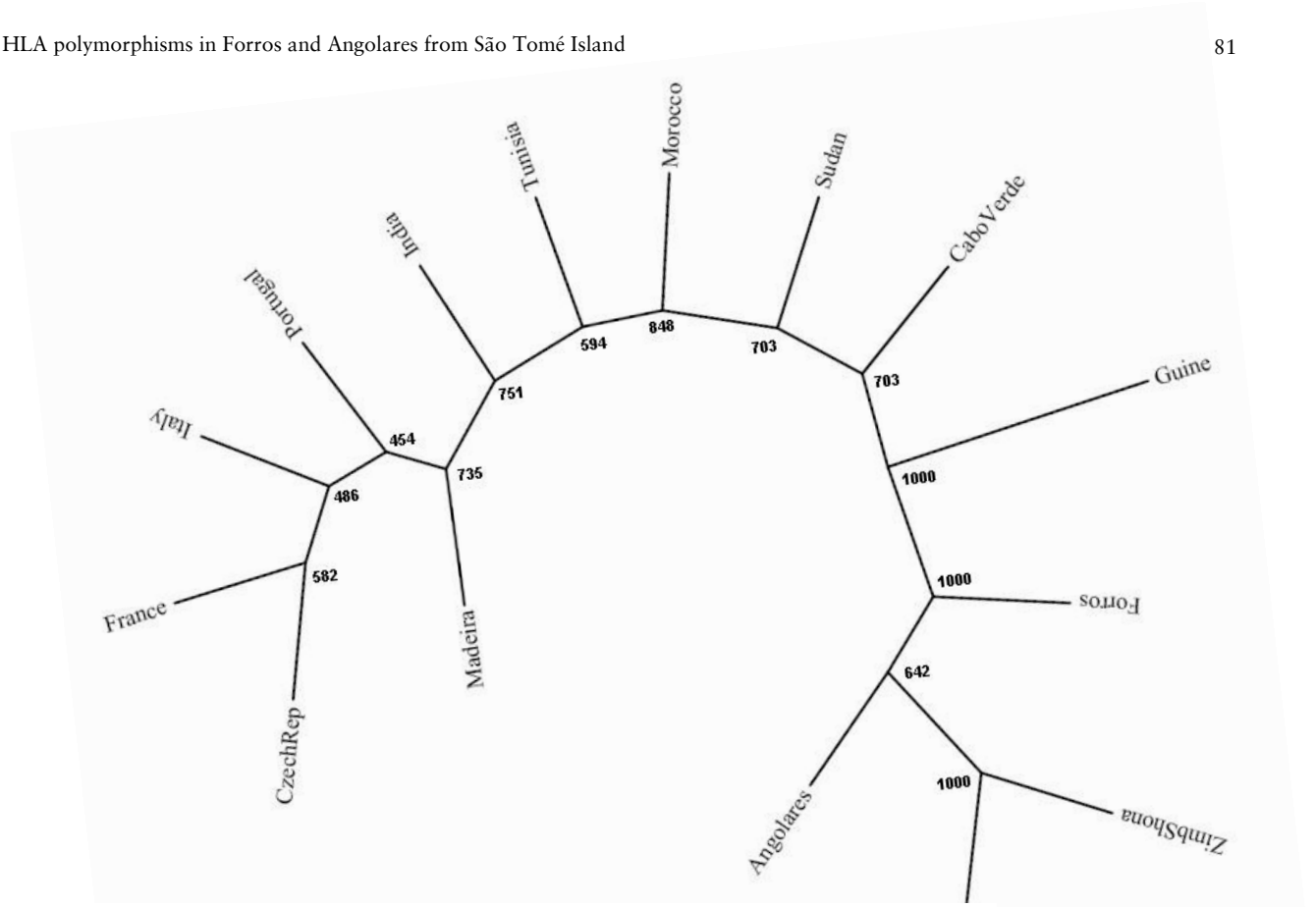
Haplotypes	Freq	D	P
A-B-DRB1 Forros > 1%			
A*0201-B*0702-DRB1*0301	0.015	0.016	P<0.005
A*0201-B*5101-DRB1*0701	0.023	0.018	P<0.005
A*0205-B*5801-DRB1*0302	0.015	0.013	P<0.005
A*3001-B*0803-DRB1*0302	0.015	0.014	P<0.005
A*3001-B*5301-DRB1*0701	0.015	0.014	P<0.005
A*3101-B*5301-DRB1*0102	0.015	0.012	P<0.005
A*3601-B*5301-DRB1*1101	0.015	0.012	P<0.005
A*6802-B*5301-DRB1*1303	0.015	0.009	P<0.02
A*6802-B*0702-DRB1*1301	0.030	0.026	P<0.005
A-B Forros >2%			
A*0101-B*5301	0.023	0.020	P<0.005
A*0201-B*5101	0.038	0.030	P<0.005
A*0205-B*5801	0.023	0.020	P<0.005
A*2301-B*0702	0.028	0.020	P=0.04
A*2301-B*3501	0.028	0.020	P=0.01
A*3402-B*0702	0.030	0.030	P<0.005
A*6802-B*5301	0.059	0.030	P=0.02
A-DRB1 Forros >2%			
A*0201-DRB1*0102	0.023	0.020	P=0.01
A*0201-DRB1*0701	0.030	0.020	P=0.03
A*2301-DRB1*0301	0.038	0.020	P=0.02
A*3303-DRB1*1303	0.023	0.020	P<0.005
A*6802-DRB1*1301	0.030	0.020	P<0.005
B-DRB1 Forros >2%			
B*0702-DRB1*1301	0.030	0.020	P<0.005
B*0801-DRB1*0301	0.023	0.020	P<0.005
B*1503-DRB1*1101	0.021	0.020	P<0.005
B*1510-DRB1*0701	0.038	0.020	P<0.005
B*4402-DRB1*0301	0.023	0.020	P<0.005
B*5301-DRB1*1101	0.040	0.020	P=0.02
B*5301-DRB1*1303	0.038	0.020	P=0.01

Haplotypes	Freq	D	P
A-B-DRB1 Angolares > 1%			
A*0201-B*1503-DRB1*0301	0.031	0.026	P<0.005
A*3002-B*5802-DRB1*0302	0.031	0.025	P<0.005
A*3303-B*1801-DRB1*1201	0.031	0.028	P<0.005
A*3601-B*5301-DRB1*1503	0.031	0.022	P<0.005
A*6802-B*5301-DRB1*0804	0.047	0.033	P<0.005
A-B Angolares >3%			
A*2301-B*0801	0.031	0.030	P<0.005
A*3002-B*5802	0.031	0.030	P<0.005
A*3303-B*1801	0.031	0.030	P<0.005
A*3402-B*0702	0.031	0.030	P=0.02
A*3601-B*5301	0.047	0.030	P=0.02
A*6802-B*5301	0.078	0.040	P=0.02
A*7401-B*3501	0.031	0.030	P<0.005
A-DRB1 Angolares >3%			
A*0201-DRB1*0301	0.047	0.040	P<0.005
A*1101-DRB1*0301	0.031	0.030	P<0.005
A*2301-DRB1*0301	0.047	0.030	P=0.04
A*2402-DRB1*0302	0.031	0.030	P<0.005
A*3002-DRB1*0302	0.031	0.030	P<0.005
A*3303-DRB1*1201	0.031	0.030	P<0.005
A*3601-DRB1*1503	0.031	0.020	P=0.01
A*6802-DRB1*0804	0.047	0.020	P=0.03
B-DRB1 Angolares >3%			
B*1503-DRB1*1503	0.031	0.020	P=0.01
B*1801-DRB1*1201	0.031	0.030	P<0.005
B*3501-DRB1*1101	0.031	0.030	P<0.005
B*4901-DRB1*0301	0.047	0.040	P<0.005
B*5301-DRB1*0804	0.047	0.030	P=0.02
B*5801-DRB1*0301	0.031	0.020	P=0.03
B*5802-DRB1*0302	0.031	0.020	P=0.01

first African inhabitants descended from liberated slaves), Angolares (more resistant to mixing with other groups and probably descendants of slaves who escaped from plantations), Mancó (who live on Príncipe island) and Tonga (descendants of people who arrived in the 19th century after slave abolition including immigrants from Cabo Verde archipelago). Forros, Mancó and Tonga speak a Creole language, a mixture of Portuguese and Bantu languages, but Angolares maintain their own Bantu language (Barata, 1966; Henrique, 2000; Seibert, 1998; Tenreiro, 1961).

Forros and Angolares show no significant HLA differentiation from each other, but both groups are significant-

a



b

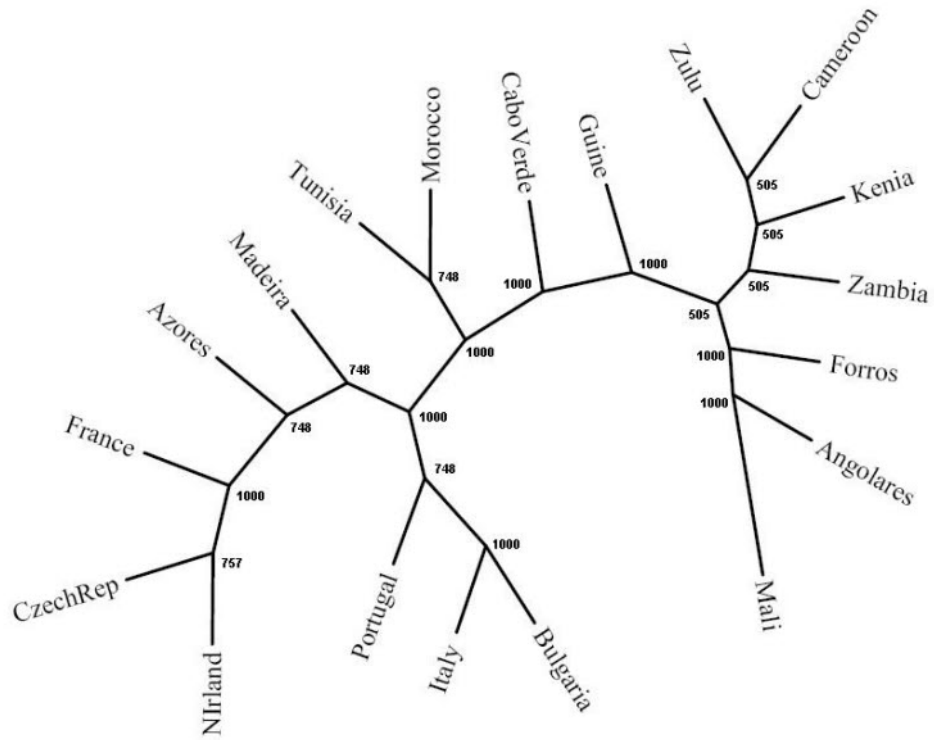


Figure 2. Neighbor-joining (NJ) dendrograms showing the comparative position of São Tomé populations (Forros and Angolares) with others typed with similar resolution. Standard genetic distances among populations were calculated using (a) HLA-A, HLA-B and HLA–DRB1 and b) HLA-A and HLA-B allele frequencies. Numbers above branches are node support (out of 1000 trees) after the bootstrap technique implemented in PHYLIP package program BOOT. For references of the populations used see *Data Analysis*.

ly different from all other populations included in the comparison. Our results are consistent with previous studies on mtDNA and Y-chromosome that point to the West coast of Africa as the place of origin of the São Tomé and Príncipe population's main genetic pool (Gonçalves et al, 2007; Mateu et al, 1997; Trovoada et al, 2004).

Although some references in the literature consider the Gulf of Guinea, Congo and Angola the specific places of origin of slaves that were brought by Portuguese to the archipelago in the first centuries of peopling (Neves, 1989), due to the few West African populations available for comparisons, especially south to Gulf of Guinea, our results only confirm West Africa as the origin of the main genetic pool of Forros and Angolares. The position of Forros and Angolares in the dendrograms and PCO near Guinea-Bissau, Mali and Cameroon supports this hypothesis.

The lack of West African populations typed for the most polymorphic HLA loci makes it difficult to understand the specific origin of the most frequent haplotypes found in São Tomé Island. However, the alleles involved and the related two-loci haplotypes show that they have a clear provenance from sub-Saharans.

Forros are not statistically different from Angolares (P=0.03) and cluster together in phylogenetic analysis. This could mean that Angolares have a common origin with Forros, which makes plausible the hypothesis that Angolares are descendants of slaves that escaped and remained genetically isolated. In fact, founder effects, genetic drift and no admixture could explain the small differences that Angolares have with Forros on allele and haplotype frequencies. The higher European genetic input in Forros than in Angolares, as demonstrated previously on Y-chromosome studies (Gonçalves et al, 2007), could also explain the differences between them despite the hypothesis of a common origin.

Considering previous studies (Gonçalves et al, 2007; Tomas et al, 2002; Trovoada et al, 2001) and present data, Forros and Angolares from São Tomé Island show a clear sub-Saharan origin with higher similarities to the closest populations on the West Coast of Africa. In the future, other west sub-Saharan populations typed to the most polymorphic HLA loci will be helpful in re-analysing present data in order to allow us to conclude more about the specific geographic sub-Saharan origin of the São Tomé population.

Acknowledgments

This work was supported by the Portuguese Foundation for Science and Technology and the European Community through the project Nº POCI/BIA-BCM/60440/2004. BIOFORMA and SANTOQUEIJO are sponsors of the University of Madeira Human Genetics Laboratory.

Supplementary Data

The HLA haplotypes for all subjects are included in the Supplementary Data file.

Web Resources

New Allele Frequency Database
http://www.allelefrequencies.net

Arlequin Software for Population Genetics
http://cmpg.unibe.ch/software/arlequin3/

PHYLIP: Phylogeny Inference Package
http://bioweb.uwlax.edu/GenWeb/Evol_Pop/Phylogenetics/Phylip/phylip.htm

MVSP: MVSP—A Multivariate Statistical Package
http://www.kovcomp.co.uk/mvsp/index.html

References

Arnaiz-Villena A, Gómez-Casado E, Martínez-Laso J (2002) Population genetic relationships between Mediterranean populations determined by HLA allele distribution and a historic perspective. *Tissue Antigens*, 60:111-121.

Barata OS (1966) O povoamento de Cabo Verde, Guiné e São Tomé e Príncipe. In: Cabo Verde, Guiné e São Tomé e Príncipe. Curso de extensão universitária. Ano lectivo de 1965-1966. ISCSPU, Lisboa.

Bodmer WF (1987) The HLA system: structure and function. *J Clin Pathol*, 40:948-958.

Boehncke WH, Loeliger C, Kuehnl P, Kalbacher H, Bohm BO, Gall H (1998) Identification of HLA-DR and -DQ alleles conferring susceptibility to pollen allergy and pollen associated food allergy. *Clin Exp Allergy*, 28:434-441.

Cao K, Moormann AM, Lyke KE, Masaberg C, Sumba OP, Doumbo OK, Koech D, Lancaster A, Nelson M, Meyer D, Single R, Hartzman RJ, Plowe CV, Kazura J, Mann DL, Sztein MB, Thomson G, Fernández-Vina MA (2004) Differentiation between African populations is evidenced by the diversity of alleles and haplotypes of HLA class I loci. *Tissue Antigens*, 63:293-325.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, 131:479-491.

Excoffier L, Laval G, Schneider S (2005) Arlequin version 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, 1:47-150.

Felsenstein J (2004) PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author, Department of Genome Sciences, University of Washington, Seattle.

Garcia A (1966) A ilha de São Tomé como centro experimental do comportamento do luso nos trópicos. Separata de DTVDIA 19, Lisboa.

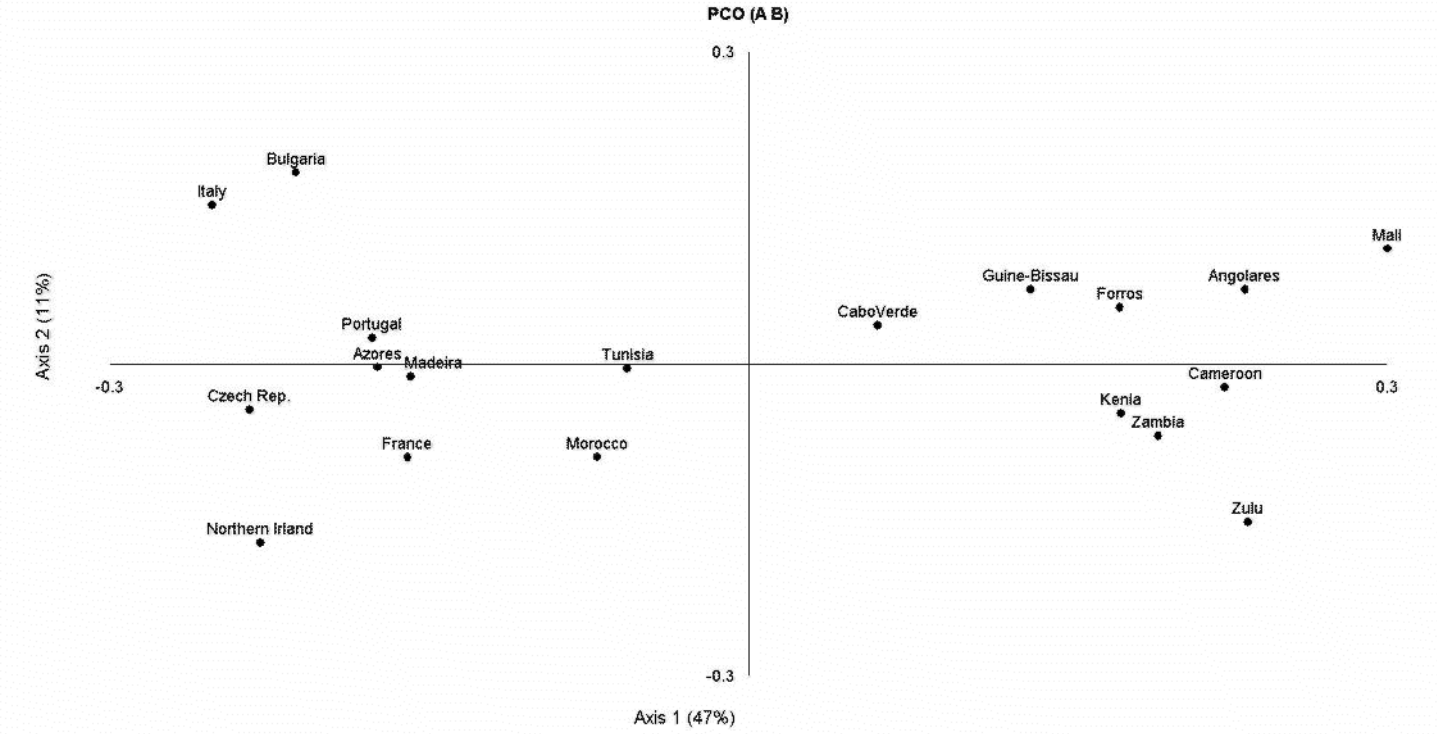


Figure 3. Principal coordinate analysis using HLA-A and HLA-B allele frequencies.

Gilbert M, Balandraud N, Touinssi M, Mercier P, Roudier J, Reviron D (2003) Functional Categorization of HLA-DRB1 Alleles in Rheumatoid Arthritis: The Protective Effect. *Hum Immun*, 64:930–935.

Gonçalves R, Jesus J, Fernandes AT, Brehm A (2002) Genetic profile of a multi-ethnic population from Guiné-Bissau (West African coast) using the new PowerPlex® 16 System kit. *Forensic Sci Int*, 129:78-80.

Gonçalves R, Spínola H, Brehm A (2007) Y-chromosome lineages in São Tomé e Príncipe islands: evidence of European influence. *Am J Hum Biol*, 19:422-8.

Henriques IC (2000) *São Tomé e Príncipe—A invenção de uma sociedade.* Vega e Autor.

Klein J, Sato A (2000) The HLA system; First of two parts. *New Engl J Med*, 343:702-709.

Kovach WL (2006) MVSP—A Multivariate Statistical Package for Windows, version 3.6. Kovach Computing Services, Pentraeth, Wales. See Web Resources.

Kurz B, Steiert I, Heuchert G, Müller CA (1999) New high resolution typing strategy for HLA-A locus alleles based on dye terminator sequencing of haplotypic group-specific PCR-amplicons of exon 2 and exon 3. *Tissue Antigens*, 53:81-96.

Mateu E, Comas D, Calafell F, Pe A, Rez-Lezaun, Abade A, Bertranpetit 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.

Middleton D, Menchaca L, Rood H, Komerofsky R (2003) New allele frequency database. *Tissue Antigens*, 61:403-407. See New Allele Frequency Database under Web Resources.

Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res*, 16:1215.

Naik S (2003) The human HLA system. *J Indian Rheumatol Assoc*, 11:79-83.

Neves CA (1989) São Tomé e Príncipe na segunda metade do séc. XVIII. *Colecção Memórias 2,* Centro de Estudos de História do Atlântico. 1ª Edição, Funchal.

Peres D (1960) *História dos descobrimentos Portugueses.* Coimbra. 2ª Edição.

Pozzi S, Longo A, Ferrara GB (1999) HLA-B locus sequence-based typing. *Tissue Antigens*, 53:275-281.

Riley E, Olerup O (1992) HLA polymorphism and evolution. *Immunol Today*, 13:333-335.

Romana H (1997) São Tomé e Príncipe: Elementos para uma análise antropológica das suas vulnerabilidades e potencialidades. Instituto Superior de Ciência Sociais e Políticas. Universidade Técnica de Lisboa, Lisboa.

Rosa A, Brehm A, Kivisild T, Metspalu E, Villems R (2004) MtDNA profile of West Africa Guineans: Towards a better understanding of the Senegambia region. *Ann Hum Genet*, 68:340-352.

Rosa A, Ornelas C, Brehm A, Villems R (2006) Population data on 11 Y-chromosome STRs from Guiné-Bissau. *Forensic Sci Int*, 157:210-217.

Sanchez-Mazas A (2001) African diversity from the HLA point of view: Influence of genetic drift, geography, linguistics, and natural selection. *Hum Immunol*, 62:937-948.

Seibert G (1998) A Questão da Origem dos Angolares de São Tomé. *Brief Papers*, 5/98. CESA, Lisboa.

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

Spínola H, Brehm A, Bettencourt B, Middleton D, Bruges-Armas J (2005a) HLA class I and II polymorphisms in Azores show different settlements in Oriental and Central Islands. *Tissue Antigens*, 66:217-230.

Spínola H, Middleton D, Brehm A (2005b) HLA genes in Portugal inferred from sequence-based typing: in the crossroad between Europe and Africa. *Tissue Antigens*, 66:26-36.

Spínola H, Bruges-Armas J, Middleton D, Brehm A (2005c) HLA polymorphisms in Cabo Verde and Guiné-Bissau inferred from sequence-based typing. *Hum Immunol*, 66:1082–1092.

Spínola H, Bruges-Armas J, Mora MG, Middleton D, Brehm A (2006) HLA genes in Madeira Island (Portugal) inferred from sequence-based typing: Footprints from different origins. *Hum Immunol*, 43:1726-1728.

Tenreiro F (1961) A ilha de São Tomé; Memórias da Junta de Investigação do Ultramar. 2ª Edição 24, Lisboa.

Tomas G, Seco L, Seixas S, Faustino P, Lavinha J, Rocha J (2002) The peopling of Sao Tome (Gulf of Guinea): origins of slave settlers and admixture with the Portuguese. *Hum Biol*, 74:397-411.

Trovoada MJ, Alves C, Gusmão L, Abade A, Amorim A, Prata MJ, (2001) Evidence for population sub-structuring in São Tomé e Príncipe as inferred from Y-chromosome STR analysis. *Ann Hum Genet*, 65:271-283.

Trovoada MJ, Pereira L, Gusmão L, Abade A, Amorim A, Prata MJ (2004) Pattern of mtDNA variation in three populations from São Tome e Principe. *Ann Hum Genet*, 68:40-54.

Williams F, Meenagh A, Single, R, McNally M, Kelly P, Nelson M, Meyer D, Lancaster A, Thomson G, Middleton D (2004) High resolution HLA-DRB1 identification of a Caucasian population. *Hum Immun*, 65:66-77.