



Diversidade Genética do Sistema HLA em Portugal, Cabo Verde e Guiné-Bissau

Hélder Spínola de Freitas

Tese apresentada na Universidade da Madeira
para a obtenção do grau de Doutor

2005



Diversidade Genética do Sistema HLA
em Portugal, Cabo Verde e Guiné-Bissau

Hélder Spínola de Freitas

Tese apresentada na Universidade da Madeira para a obtenção
do grau de Doutor

2005

Diversidade Genética do Sistema HLA em Portugal, Cabo Verde e
Guiné-Bissau

por

Hélder Spínola de Freitas

Tese submetida de acordo com o regulamento em vigor na Universidade da
Madeira para obtenção do grau de

Doutor em Ciências Biológicas, especialidade de Citogenética e Biologia
Molecular

Universidade da Madeira

2005

Diversidade Genética do Sistema HLA em Portugal, Cabo Verde e
Guiné-Bissau

Hélder Spínola de Freitas

Supervisor Científico:

Professor Catedrático António Brehm
Departamento de Biologia, Universidade da Madeira

Universidade da Madeira

A todos aqueles que,
independentemente de alcançarem ou não
os seus objectivos, encontram força
e motivação suficiente para dar o seu melhor.

Índice

Agradecimentos	xv
Abreviaturas	xvi
Resumo	xvii
Abstract	xix
1. Introdução	1
1.1. A descoberta do sistema HLA.....	1
1.2. Nomenclatura do sistema HLA	2
1.3. Técnicas de caracterização do sistema HLA	3
1.4. Estrutura e função do sistema HLA	7
1.5. Evolução e diversificação do sistema HLA	10
1.6. Importância médica do sistema HLA	12
1.7. Aplicação dos genes HLA nas disputas de paternidade.....	16
1.8. Importância dos genes HLA em estudos populacionais	17
1.9. História das populações em estudo	20
1.10. Objectivos da tese	22
2. Material e métodos.....	25
2.1. Amostragem das populações	25
2.2. Genes estudados.....	27
2.3. Métodos utilizados	27
2.4. Análise de dados.....	31
3. Distribution of HLA alleles in Portugal and Cabo Verde. Relationships with the slave trade route.....	35
3.1. Abstract.....	35
3.2. Introduction	36

3.3. Materials and methods	39
3.3.1. Population samples	39
3.3.2. HLA typing and statistics	40
3.4. Results and discussion	41
3.4.1. HLA class I and II allele frequencies	41
3.4.1.1. HLA-A	45
3.4.1.2. HLA-B	46
3.4.1.3. HLA-DRB1	47
3.4.2. Haplotype frequencies	48
3.4.2.1. West European haplotypes	48
3.4.2.2. Mediterranean haplotypes	50
3.4.2.3. Iberian-North African haplotypes	50
3.4.2.4. The sub-Saharan component	50
3.4.2.5. The 'Oriental' component	51
3.4.3. Principal Component Analysis	52
3.4.4. Population comparisons	54
4. HLA genes in Portugal inferred from sequence-based typing: in the crossroad between Europe and Africa	57
4.1. Abstract	57
4.2. Introduction	58
4.3. Materials and methods	60
4.3.1. Population samples and HLA typing	60
4.3.2. Data analysis	62
4.4. Results	63
4.4.1. Data for HLA-A	68
4.4.2. Data for HLA-B	69

4.4.3. Data for HLA-DRB1	71
4.4.4. Haplotype frequencies	72
4.4.5. Phylogenetic analyses	78
4.5. Discussion	82
5. HLA class I and II polymorphisms in Azores show different settlements in Oriental and Central islands	87
5.1. Abstract	87
5.2. Introduction	88
5.3. Materials and methods	90
5.3.1. Population samples and HLA typing	90
5.3.2. Data analysis	91
5.4. Results	93
5.4.1. HLA-A <i>locus</i>	97
5.4.2. HLA-B <i>locus</i>	97
5.4.3. HLA-Cw <i>locus</i>	97
5.4.4. HLA-DRB1 <i>locus</i>	98
5.4.5. HLA-DQA1 <i>locus</i>	98
5.4.6. HLA-DQB1 <i>locus</i>	98
5.4.7. Haplotype frequencies	99
5.4.8. Phylogenetic analyses	107
5.5. Discussion	111
5.5.1. Diversity between Azores islands	114
6. HLA polymorphisms in Cabo Verde and Guiné-Bissau inferred from sequence-based typing	117
6.1. Abstract	117
6.2. Introduction	118

6.3. Materials and methods	120
6.3.1. Population samples and HLA typing	120
6.3.2. Data analysis	121
6.4. Results	123
6.4.1. Guiné-Bissau allele frequencies	128
6.4.2. Cabo Verde allele frequencies	129
6.4.3. Haplotype frequencies	129
6.4.4. Phylogenetic analyses	136
6.5. Discussion	139
7. HLA genes in Madeira island (Portugal) inferred from sequence- based typing: footprints from different origins	143
7.1. Abstract	143
7.2. Introduction	144
7.3. Materials and methods	145
7.3.1. Population samples and HLA typing	145
7.3.2. Data analysis	146
7.4. Results	146
7.4.1. Allele frequencies	146
7.4.2. Haplotype frequencies	148
7.5. Discussion	150
8. Discrepancies on HLA typing by SSOP and SBT techniques: a case study	155
8.1 Abstract	155
8.2 Introduction	156
8.3 Materials and methods	157
8.4 Results and discussion	157

9. Resultados globais	161
9.1. Parâmetros gerais	161
9.2. HLA-A	163
9.3. HLA-B	164
9.4. HLA-DRB1	165
9.5. Frequências haplotípicas	166
9.6. Análises filogenéticas	167
10. Discussão geral	169
10.1. Distribuição dos alelos HLA em Portugal e Cabo Verde. Influências do comércio de escravos	169
10.2. Caracterização dos genes HLA na população Portuguesa por sequenciação: na encruzilhada entre a Europa e África	171
10.3. Polimorfismos dos genes HLA mostram diferenças no povoamento dos grupos Central e Oriental do arquipélago dos Açores	190
10.3.1. Diversidade entre as ilhas Açorianas	192
10.4. Caracterização do sistema HLA em Cabo Verde e Guiné- Bissau por sequenciação nucleotídica	194
10.5. Os genes HLA na ilha da Madeira (Portugal) caracterizados por sequenciação nucleotídica: marcas de diferentes origens	196
10.6. Discrepâncias na caracterização do sistema HLA por SSOP e SBT: um caso de estudo	199
10.7. Discussão global	200
10.8. Perspectivas futuras	204
11. Referências bibliográficas	207
12. Portais na Internet	225

Agradecimentos

Apresento os meus mais sinceros agradecimentos a António Brehm pelo acompanhamento e apoio que dispensou ao longo do desenvolvimento deste trabalho. Agradeço também a Derek Middleton e a todos os membros do Laboratório de Histocompatibilidade e Immunogenética da Irlanda do Norte pelo acolhimento e apoio na abordagem à caracterização dos genes HLA nas populações estudadas.

Agradeço ainda às minhas colegas do Laboratório de Genética Humana pelo apoio prestado. Agradeço a todos os membros do Departamento de Biologia da Universidade da Madeira que directa ou indirectamente prestaram apoio ao trabalho desenvolvido. Agradeço ao Governo Regional da Madeira que, através do Fundo Social Europeu, apoiou financeiramente este projecto de doutoramento e ao programa Interreg III B Madeira- Açores- Canárias que financiou parte da investigação desenvolvida no âmbito deste trabalho.

Por último, agradeço a toda a minha família e amigos pelo apoio moral, e não só, que sempre deram no desenvolvimento deste trabalho, nomeadamente o Paulo Lucas na concepção da capa deste manuscrito. Destaco particularmente a minha esposa pela forma incondicional como sempre me incentivou e amparou, principalmente nos momentos mais difíceis, e os meus pais por quererem sempre o melhor para mim.

Abreviaturas

a.C.- Antes de Cristo

Bp- Base pairs (pares de bases)

d.C.- Depois de Cristo

DC- Doença celíaca (celiac disease)

DNA- Deoxyribonucleic acid (ácido desoxiribonucleíco)

HIV- Human immunodeficiency virus

HLA - Human leukocyte antigens

Kb- Kilobases

Mb- Megabases

MHC - Major histocompatibility complex

PCA- Principal component analysis

PCO- Principal coordinates analysis

PCR- Polymerase chain reaction

SBT - Sequence based typing

RFLP- Restriction fragment length polymorphysm

SSOP- Sequence specific oligonucleotide probes

SSP- Sequence specific priming

STR- Short tandem repeat

Resumo

O polimorfismo dos genes HLA-A, HLA-B e HLA-DRB1 foi estudado em três populações Portuguesas (Portugal continental, Madeira e Açores) e em duas Africanas (Guiné-Bissau e Cabo Verde). Os dados em alta resolução, obtidos por sequenciação (SBT), foram comparados com a caracterização efectuada pelo método SSOP revelando 4,6% de incongruências entre os resultados destes dois métodos.

Os alelos mais frequentes em cada um dos *loci* foram: HLA-A*0201 em todas as populações (13,5-26%); HLA-B*5101 em Portugal continental (12%, o mesmo para o B*440301), Madeira (9,7%) e Açores (9,8%) e B*350101 na Guiné-Bissau (14,4%) e Cabo Verde (13,2%); HLA-DRB1*0701 em Portugal continental (15%), Madeira (15,7%) e Açores (18,3%), DRB1*1304 na Guiné-Bissau (19,6%) e DRB1*110101 em Cabo Verde (10,1%).

Os haplotipos 3-*loci* mais predominantes em cada população foram: A*020101-B*440301-DRB1*070101 em Portugal continental (3,1%), A*020101-B*510101-DRB1*130101 na Madeira (2,7%), A*2902-B*4403-DRB1*0701 nos Açores (2,4%), A*2301-B*1503-DRB1*110101 na Guiné-Bissau (4,6%) e A*3002-B*350101-DRB1*1001 em Cabo Verde (2,8%).

O presente trabalho revela que a população continental Portuguesa tem sido influenciada geneticamente por Europeus e Norte Africanos devido a várias imigrações históricas. O Norte de Portugal parece concentrar, provavelmente devido à pressão da expansão Árabe, um antigo *pool* genético originado pela influência milenar de Europeus e Norte Africanos.

Os dados obtidos nos genes do sistema HLA corroboram as fontes históricas que confirmam que o povoamento dos Açores teve o contributo de outros Europeus, essencialmente Flamengos, para além dos Portugueses. As frequências alélicas e haplotípicas neste arquipélago não apresentam uma distribuição homogénea entre as ilhas do grupo Oriental e Central. O grupo Central revela uma influência clara da Europa Central e uma muito menor afinidade a Portugal continental.

As frequências alélicas e haplotípicas mostram que a ilha da Madeira foi povoada por Europeus, a maioria Portugueses, mas também por sub-Saharianos devido ao comércio de escravos.

Cabo Verde não é uma população tipicamente sub-Sahariana pois revela uma importante influência genética Europeia, para além da base genética Africana. A análise dos haplotipos e dendrogramas mostram uma influência genética Caucasiana no actual *pool* genético Cabo-Verdiano. Os dendrogramas e a análise das coordenadas principais mostram que os Guineenses são mais semelhantes aos Norte Africanos do que qualquer outra população sub-Sahariana já estudada ao nível do sistema HLA, provavelmente devido a contactos históricos com outros povos, nomeadamente Árabes do Este Africano e Berberes.

Abstract

HLA-A, HLA-B, and HLA-DRB1 genetic polymorphism was examined in three Portuguese (mainland Portugal, Madeira and Azores) and two African populations (Guiné-Bissau and Cabo Verde). Data obtained at high-resolution level, using sequence-based typing, were compared with medium resolution typing previously obtained by SSOP, showing 4.6% incongruence between the two methods.

The most frequent allele at each *loci* was: HLA-A*0201 (13.5-26%) in all populations; HLA-B*5101 in mainland Portugal (12%, the same as B*440301), Madeira (9.7%), and Azores (9.8%) and B*350101 in Guiné-Bissau (14.4%) and Cabo Verde (13.2%); HLA-DRB1*0701 in mainland Portugal (15%), Madeira (15.7%), and Azores (18.3%), DRB1*1304 in Guiné-Bissau (19.6%), and DRB1*110101 in Cabo Verde (10.1%).

The predominant three-*loci* haplotype found in each population was: A*020101-B*440301-DRB1*070101 in mainland Portugal (3.1%), A*020101-B*510101-DRB1*130101 in Madeira (2.7%), A*2902-B*4403-DRB1*0701 in Azores (2.4%), A*2301-B*1503-DRB1*110101 in Guiné-Bissau (4.6%), and A*3002-B*350101-DRB1*1001 in Cabo Verde (2.8%).

The present study demonstrates that the mainland Portuguese population has been genetically influenced by Europeans and North Africans, via several historic immigrations. North Portugal seems to concentrate, probably due the pressure of Arab expansion, an ancient genetic pool originated from several North Africans and Europeans influences throughout millenniums.

HLA data corroborates historical sources that say the Azores were populated not only by Portuguese but also by other Europeans, mostly Flemish people. Haplotype and allele frequencies in this archipelago show no homogeneous distribution between Oriental and Central islands. The Central group clearly shows an influence of central Europeans, with much less affinity to mainland Portugal.

HLA allele and haplotype frequencies show that Madeira island were populated by Europeans, mostly Portuguese, but also by sub-Saharan Africans due to slave trade.

Cabo Verde is not a typical sub-Saharan population showing an important European genetic influence besides the African basis. Haplotypes and dendrogram analysis shows a Caucasian genetic influence in today's gene pool of Cabo Verdeans. Dendrograms and principal coordinates analysis show that Guineans are more similar to North Africans than other HLA studied sub-Saharans, probably due to ancient and recent historical links with other peoples, namely East Africans Arabs and Berbers.

Capítulo 1

1. Introdução

1.1. A descoberta do sistema HLA

O Complexo Maior de Histocompatibilidade (MHC - Major Histocompatibility Complex), existente em todos os vertebrados, é constituído, entre outros, por genes com importantes funções imunológicas e foi descoberto em 1937 por Peter Gorer durante o estudo de transplantes em ratos (Gorer, 1937). Mais tarde Jean Dausset (1958) publicou as suas observações e conclusões sobre a capacidade do soro de pessoas submetidas a transfusões sanguíneas aglutinar leucócitos de outros indivíduos. Dausset descobriu assim o primeiro antígeno humano, denominado MAC (agora HLA-A*02), um dos produtos dos genes do sistema humano de antígenos leucocíticos (HLA - Human Leukocyte Antigens), o correspondente humano do MHC. Mais tarde, Jon van Rood e van Leeuwen (1963) publicaram a descoberta de outro gene do sistema HLA a que deram o nome de FOUR (agora HLA-B). Com a invenção da técnica de cultura linfocitária mista em 1964 (Bach e Hirschhorn, 1964; Bain *et al.* 1964) foram-se acumulando resultados que conduziram à

descoberta de novos genes, nomeadamente o HLA-DR na década de 70 do século XX (Yunis e Amos, 1971; Tosi *et al.* 1978).

O envolvimento do MHC na resposta imunitária do organismo só foi confirmada no início da década de setenta do século XX (Shevach *et al.*, 1972) mas actualmente conhece-se já em detalhe a estrutura e função dos genes classe I e classe II do sistema HLA (Bjorkman *et al.*, 1987; Brown *et al.* 1993). Apesar de actualmente desconhecermos ainda o mecanismo exacto que associa determinados alelos e haplotipos (conjunto de alelos de diferentes *loci* transmitidos em bloco de geração em geração) do sistema HLA a um elevado risco de desenvolver determinadas doenças autoimunes, já em 1967 esta associação era conhecida (Amiel, 1967).

Devido às suas múltiplas funções, nomeadamente ao nível da histocompatibilidade e regulação imunitária, a investigação em torno do sistema HLA despertou o interesse de várias equipas de investigação possibilitando a descoberta de largas centenas de alelos nos diferentes genes.

1.2. Nomenclatura do sistema HLA

Esta enorme diversidade genética tornou necessário, desde muito cedo, o estabelecimento de regras claras ao nível da nomenclatura utilizada, tendo sido organizado em 1968 por Bernard Amos o primeiro encontro do Comité de Nomenclatura da Organização Mundial de Saúde (WHO-Nomenclature-Committee, 1968).

Um código de 4 dígitos que distingue os alelos que diferem ao nível das proteínas codificadas foi implementado pela primeira vez no Relatório de Nomenclatura em 1987 em sequência do X *Workshop* de

Histocompatibilidade que decorreu em Nova Iorque nesse mesmo ano (Dupont, 1988). Desta forma foi possível acomodar os diferentes alelos, identificados através de métodos moleculares, que constituem cada grupo serológico. A utilização de um asterisco depois da designação do *locus* permitiu distinguir a caracterização alélica por métodos moleculares da caracterização por métodos serológicos. Em 1990 foi adicionado um quinto dígito para distinguir as sequências que diferem apenas por substituições nucleotídicas sinónimas (mutações silenciosas) (Bodmer *et al.*, 1991). No entanto, este sistema de nomenclatura já está a atingir, para certos grupos de alelos (e.g HLA-A*02 e HLA-B*15), o nível máximo de alelos que consegue comportar pelo que já estão a ser adoptadas novas convenções. Assim, para acomodar novos alelos nos grupos mais polimórficos, assim que estes atinjam 99 alelos, uma segunda série será utilizada para continuação. A série B*95 está assim reservada para acomodar novos alelos quando o B*1599 for atingido, da mesma forma que a série A*92 dará continuidade aos alelos do grupo A*02 (Marsh *et al.*, 2002).

1.3. Técnicas de caracterização do sistema HLA

A evolução das técnicas utilizadas na identificação e caracterização das diferentes formas dos genes que compõem o sistema HLA constituiu um dos factores cruciais no desenvolvimento dos conhecimentos nesta área. Os primeiros passos na descoberta do sistema HLA foram dados com recurso às técnicas de leuco-aglutinação com as quais foi possível definir vários antigénios codificados por este conjunto de genes (Berah e Dausset, 1963). Até meados dos anos 80 do século XX as técnicas serológicas constituíam a única forma de caracterizar a diversidade alélica do sistema HLA.

O teste linfocitotóxico foi amplamente divulgado na abordagem serológica para identificação dos antígenos do sistema HLA, existindo actualmente múltiplas variações desta técnica base. Esta técnica consiste na mistura de linfócitos B com soro contendo anticorpos específicos para determinados antígenos do sistema HLA. Os anticorpos, quando na presença de linfócitos B contendo antígenos complementares, ligam-se provocando, com a adição de um complementar, a destruição membranar, o que permite a sua identificação por coloração (Kippax *et al.*, 1985; Zachary *et al.* 1995).

Devido a problemas de reacção cruzada e indisponibilidade de certos anticorpos, muitos laboratórios adoptaram técnicas de genética molecular para a caracterização do sistema HLA. No entanto, estes métodos serológicos continuam a ser amplamente utilizados em testes de *crossmatching* com o objectivo de validar potenciais dadores seleccionados para transplantes de órgãos e tecidos. Este procedimento é desenvolvido como salvaguarda à ocorrência de rejeição aguda do transplante pois alguns pacientes poderão ter sido previamente sensibilizados contra determinados antígenos HLA. Mesmo que os alelos do dador e do receptor sejam compatíveis, em sequência de transfusões de sangue, da realização de outros transplantes ou de uma gravidez, o paciente poderá ter desenvolvido anticorpos contra alguns dos antígenos do dador levando a uma reacção violenta logo no início do transplante, pondo em causa o seu sucesso (Moore *et al.*, 1997)¹.

Os métodos de restrição enzimática (RFLP- Restriction Fragment Length Polymorphism) permitiram dar os primeiros passos na identificação

¹ <http://www.pharmacy.wisc.edu/courses/718-430/2001presentation/Flanner.pdf> (26-09-2005)
<http://www.hla.ucla.edu/labServices/labServicesABT.htm> (26-09-2005)

genética destes alelos no início da década de 80 do século XX permitindo uma maior precisão comparativamente aos métodos serológicos. Com o advento da PCR (Polymerase Chain Reaction) (Saiki *et al.*, 1985, Mullis e Faloona, 1987), particularmente após a sua automatização, depressa os métodos de caracterização alélica do sistema HLA por restrição enzimática evoluíram e mais tarde foram mesmo ultrapassados por outros procedimentos mais modernos (Howell *et al.*, 1989; Ota *et al.*, 1992). A associação entre a PCR e a restrição enzimática permitiu a adopção de uma técnica de caracterização alélica do sistema HLA mais eficiente e em alta resolução. Após a amplificação dos exões polimórficos do *locus* HLA em estudo, o DNA obtido por PCR é digerido com enzimas de restrição específicas e o perfil dos fragmentos produzidos é analisado após separação por electroforese (Mitsunaga *et al.*, 1998).

Vários métodos foram desenvolvidos com base na PCR. A caracterização alélica por SSOP (Sequence Specific Oligonucleotide Probes) recorre a *primers* (oligonucleotidos) complementares de zonas conservadas para amplificar os exões polimórficos dos *loci* HLA (exões 2 e 3 nos genes HLA classe I e exão 2 nos genes classe II). O DNA amplificado é fixo a uma membrana e é hibridado com sondas oligonucleotídicas marcadas com fluorescência. As sondas complementares às zonas polimórficas do DNA amplificado produzem um padrão de hibridação cuja análise permite identificar os alelos presentes do sistema HLA (Middleton, 2000). Esta técnica permite a caracterização do sistema HLA em baixa ou alta resolução consoante o painel de sondas utilizado e é adequada para grandes quantidades de amostras (Middleton e Williams, 1997).

O teste SSP (Sequence Specific Priming) utiliza directamente a técnica de PCR para identificar os alelos do sistema HLA com recurso a oligonucleotidos complementares a sequências nucleotídicas polimórficas conhecidas (Olerup e Zetterquist, 1992; Olerup e Zetterquist, 1993; Bunce *et al.*, 1995). Esta técnica baseia-se no facto de um *primer* complementar à cadeia de DNA ser mais eficiente na amplificação por PCR do que os *primers* com um ou mais nucleótidos não coincidentes. *Primers* específicos para a amplificação de determinados alelos, com separação e visualização através de electroforese em géis de agarose, permitem caracterizar em baixa ou alta resolução os genes HLA (Middleton e Williams, 1997).

A caracterização alélica dos genes do sistema HLA através da sequenciação nucleotídica (SBT- Sequence Based Typing) é o método molecular actualmente mais fiável, o qual permite inclusive a identificação directa de novos alelos (McGinnis *et al.*, 1995; Middleton e Williams, 1997; Kurz *et al.*, 1999). Esta técnica desenvolve-se com a amplificação, por PCR, dos fragmentos contendo os exões polimórficos a sequenciar. As técnicas e tecnologias mais recentes permitem a sequenciação directa do fragmento sendo que a sua natureza heterozigótica implica a utilização de um programa informático próprio para a discriminação dos alelos presentes em cada amostra.

A escolha do método a utilizar na caracterização alélica do sistema HLA depende das especificidades de cada laboratório e dos objectivos do trabalho. Desde a rapidez necessária até ao número de amostras, passando pelo equipamento existente, orçamento disponível e experiência dos recursos humanos, inúmeras são as variáveis que influenciam a escolha do método. Consoante as necessidades e os objectivos dos laboratórios,

nomeadamente se pretendem baixa ou alta resolução, poderão optar por utilizar mais de um método. No entanto, em qualquer dos casos, qualquer um dos métodos disponíveis necessita de ser continuamente actualizado por forma a integrar as alterações que permitem a evolução das técnicas e considerar os novos alelos que continuamente são descobertos (Middleton e Williams, 1997).

1.4. Estrutura e função do sistema HLA

Os *loci* HLA estão localizados no braço mais pequeno do cromossoma 6 na banda 6p21.3 formando um complexo sistema de genes funcionalmente relacionados entre si (Figura 1) (Bender *et al.* 1983). O sistema HLA é constituído por 224 genes, dos quais 128 são genes funcionais e 96 são pseudogenes (MHC sequencing consortium, 1999).

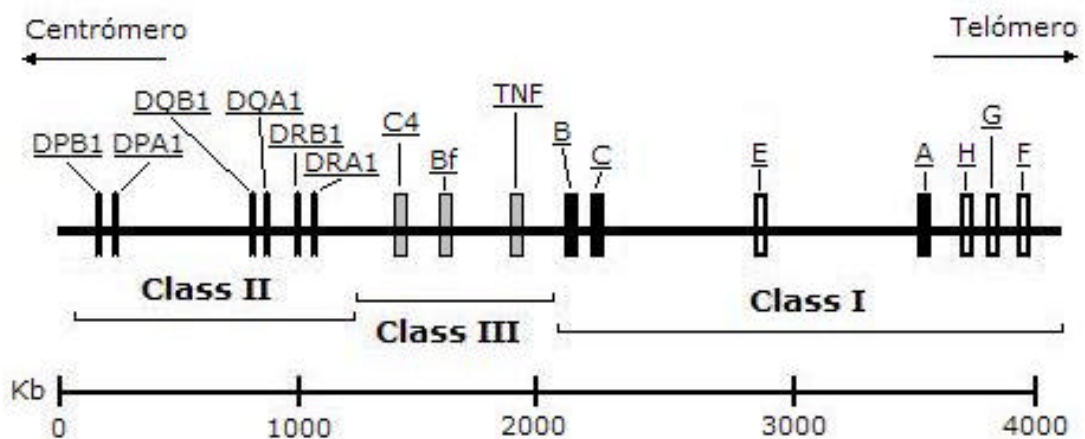


Figura 1. Mapa da região HLA (banda 6p21.3 do cromossoma 6). Os genes e pseudogenes não clássicos classe I estão representados por rectângulos não preenchidos.

Os genes funcionais do sistema HLA são os mais variáveis do genoma humano com capacidade de expressão, 40% dos quais desempenham um importante papel em funções imunológicas. Para além destes genes existem ainda no sistema HLA elementos repetitivos que compreendem quase 50% da sequência do sistema HLA (MHC sequencing consortium, 1999; Garrigan & Hedrick, 2003).

O sistema HLA, com cerca de 4 Mb (megabases) de extensão, está subdividido em três regiões: classe I, classe II e classe III, de acordo com a estrutura e função dos seus genes. Os genes classe I, localizados na porção mais telomérica do sistema HLA, são constituídos por vários exões, intercalados por intrões, que codificam para a formação da cadeia polipeptídica alfa de receptores glicoproteicos da membrana das células nucleadas, constituídos por uma região extracelular de ligação a antigénios, uma região transmembranar e um segmento citoplasmático (MHC sequencing consortium, 1999). Os polimorfismos nos genes classe I localizam-se essencialmente nos exões 2 e 3, os quais codificam para as estruturas moleculares alfa 1 e alfa 2 do receptor membranar onde se localiza o sítio de ligação aos antigénios (Malissen *et al.*, 1982). Estes receptores membranares interagem com linfócitos T CD8 positivos que detectam alterações de expressão nas células, alterações estas que podem ocorrer devido a infecções ou ao desenvolvimento de células tumorais. Desta forma os linfócitos T (CD8+) reconhecem as células que deverão atacar para combater o avanço de uma determinada infecção ou de um tumor (Doherty & Zinkernagel, 1975).

Os genes classe II localizam-se na região mais centromérica do sistema HLA e codificam para as estruturas alfa e beta de receptores membranares

que se expressam nas células que apresentam antígenos, nomeadamente linfócitos B e células dendríticas. O exão 2, que codifica para a estrutura extracelular do receptor membranar, responsável pela ligação aos antígenos, concentra a maior parte dos polimorfismos nestes genes. Estes receptores membranares, constituídos por uma porção extracelular, uma região transmembranar e um segmento citoplasmático, reconhecem a presença de elementos estranhos e estimulam os linfócitos T CD4 positivos, reforçando a resposta imunitária contra agentes infecciosos (Doherty & Zinkernagel, 1975; Gatti & Pierre, 2003).

A região classe III do sistema HLA localiza-se entre os genes classe I e classe II e, apesar de não ser tão polimórfico como estes últimos, constitui o segmento do genoma humano com maior densidade de genes (Xie *et al.*, 2003)¹. Ao contrário da região classe I e classe II onde existem dezenas de pseudogenes, a região classe III possui apenas dois (MHC sequencing consortium, 1999). Os genes presentes nesta região não são, na sua maioria, relacionados entre si nem com os genes da classe I e II e não possuem um padrão comum de expressão. Os genes classes III do sistema HLA possuem funções várias, destacando-se o seu papel na codificação de proteínas solúveis importantes na modelação e regulação da resposta imunitária (Xie *et al.*, 2003)¹.

Os genes clássicos HLA-A, HLA-B, HLA-Cw (classe I), HLA-DRB1, HLA-DPB1 e HLA-DQB1 (classe II), co-dominantes, são os mais polimórficos e mais estudados do sistema HLA (Mizuki *et al.*, 1996). Por outro lado, os genes não clássicos, nomeadamente HLA-E, HLA-F e HLA-G (classe I), com funções ainda mal conhecidas, possuem uma distribuição muito restrita,

¹ www.genome.org (10-10-2005)

baixos níveis de expressão e são pouco polimórficos comparativamente aos genes clássicos, com os quais partilham uma estrutura similar (Le Bouteiller, 1997).

Os genes classe I e II constituem no seu conjunto 1998 alelos diferentes, sendo o mais polimórfico o *locus* HLA-B com 627 alelos, seguindo-se o HLA-DRB1 com 394, o HLA-A com 349, o HLA-Cw com 182 e o HLA-DPB1 com 116 (Marsh *et al.*, 2005).

1.5. Evolução e diversificação do sistema HLA

O MHC, cujo correspondente humano é o sistema HLA, surgiu através de repetidas duplicações e conversões de genes ao longo de milhões de ano no decorrer da evolução dos vertebrados desde a divergência dos ciclóstomos (Kasahara *et al.*, 1996; Kasahara *et al.*, 1997). A análise do MHC de alguns vertebrados mostra diferenças notórias, nomeadamente ao nível da inclusão, ordem e sequências nucleotídicas dos genes mas, por outro lado, revela a existência de uma estrutura comum principalmente entre os mamíferos (Yuhki *et al.* 2003)¹. Ao contrário do que sucede nos mamíferos, noutros vertebrados como aves e peixes, os *loci* MHC nem sempre estão reunidos num único conjunto, dispersando-se no mesmo ou em cromossomas diferentes (Miller *et al.*, 1994; Bingulac-Popovic *et al.*, 1997).

A grande quantidade e diversidade de informação disponível relativamente à variabilidade, estrutura e função dos genes do sistema HLA apresentam padrões de variação que mostram sinais evidentes de selecção. A frequência de homozigóticos em vários *loci* HLA é significativamente inferior ao esperado em condições neutrais em inúmeras populações

¹ www.genome.org (10-10-2005)

humanas. Vários *loci* do sistema HLA submetidos ao teste de neutralidade Ewens-Watterson têm rejeitado a neutralidade em inúmeras populações (Tiercy *et al.*, 1992; Ellis *et al.*, 2000; Mack *et al.*, 2000; Renquin *et al.*, 2001; Piancatelli *et al.*, 2004; Cao *et al.*, 2004). A variação nos *loci* HLA está essencialmente concentrada nas regiões envolvidas na codificação dos sítios de ligação aos antígenos dos receptores membranares, situação que não é susceptível de ter sido originada a partir de uma evolução neutral (Parham *et al.*, 1988; Hedrick *et al.*, 1991; Valdes *et al.*, 1999). A maior acumulação de substituições não sinónimas nas regiões que codificam para os sítios de ligação aos antígenos revela também a actuação da selecção natural, privilegiando o aumento da variação que neste caso é funcionalmente relevante (Nei & Gojobori, 1986; Hughes & Nei, 1988).

O mecanismo que parece ter maiores responsabilidades na manutenção dos elevados níveis de polimorfismo encontrados nos genes do sistema HLA, e que fazem deles os mais variáveis do genoma humano, é a selecção oscilante, decorrente da acção de diferentes parasitas (Hedrick & Thomson, 1983; Hedrick, 1999).

A selecção oscilante refere-se a mecanismos de selecção natural que conduzem à manutenção de polimorfismos genéticos numa população, em oposição à selecção direccionada que favorece um único alelo. Entre os diferentes mecanismos inerentes à selecção oscilante destacam-se como os mais estudados a “vantagem dos heterozigóticos” e a “selecção dependente da frequência”. A “vantagem dos heterozigóticos” é baseada na ideia de que os heterozigóticos são capazes de reconhecer o dobro dos parasitas, um por cada alelo, apresentando uma vantagem adaptativa em relação aos homozigóticos (Spencer & Marks, 1988). A “selecção dependente da

frequência" baseia-se no pressuposto de que os alelos mais raros possuem uma forte vantagem selectiva relativamente aos mais comuns para os quais os parasitas podem ter já desenvolvido resistência (Bodmer, 1972).

1.6. Importância médica do sistema HLA

Inúmeros estudos têm demonstrado que os genes do sistema HLA estão associados a mais de 500 doenças autoimunes ou provocadas por agentes externos. Cerca de 4% da população em países industrializados são afectados por doenças autoimunes, número este com tendência para aumentar. A predisposição para o desenvolvimento destas doenças é muitas vezes o resultado de uma combinação em rede de vários alelos do sistema HLA. Desta forma, a caracterização alélica do sistema HLA pode ser usada para identificar indivíduos com elevado risco de desenvolver determinadas doenças assim como auxiliar no seu diagnóstico. Tendo em conta as evidências actuais do envolvimento directo dos *loci* HLA e os progressos no conhecimento da estrutura e função dos produtos destes genes na resposta imunitária, poderão, a curto prazo, ser melhor compreendidos os mecanismos responsáveis pelo desenvolvimento destas doenças. Assim, com esta informação disponível, será possível actuar de uma forma específica ao nível da prevenção e combate a estas enfermidades (Thorsby, 1997; Thorsby and Lie, 2005).

A espondilite anquilosante, uma doença reumática crónica e inflamatória, muitas vezes progressiva, que afecta a coluna vertebral e as articulações sacroilíacas, está fortemente associada à presença do alelo HLA-B*27. O HLA-B*27 contribui em 16 a 50% para o risco genético de desenvolvimento desta doença autoimune pelo que está também

dependente de outros factores genéticos e ambientais ainda desconhecidos (Reveille *et al.*, 2001; Calin, 2002). A espondilite anquilosante tem uma prevalência de 0,2 a 0,8% em populações Caucasianas e cerca de 90% dos doentes são HLA-B*27 positivos, característica que só acontece em aproximadamente 2% da população Portuguesa (Braun *et al.*, 1998; Marker-Hermann & Hoehler, 1998; Khan, 2000; Gonzalez *et al.*, 2002; Spínola *et al.*, 2005). As propriedades do HLA-B*27 que determinam esta susceptibilidade à doença são ainda desconhecidas mas algumas possibilidades têm sido aventadas, nomeadamente a hipótese do "péptido artrítogénico". Esta hipótese propõe que uma infecção bacteriana provoca uma resposta dos linfócitos T contra um péptido estranho ao organismo mas estruturalmente similar a outro derivado do tecido articular normal, envolvendo os receptores HLA-B*27. Desta forma, o sistema imunitário ataca os tecidos normais do próprio organismo não os reconhecendo como seus, levando ao desenvolvimento da espondilite anquilosante (Benjamin & Parham, 1990; Ramos & Lopez de Castro, 2002).

A artrite reumatóide é uma doença autoimune que pode assumir vários graus de desenvolvimento e que causa inflamações crónicas das articulações, podendo afectar também outros tecidos do organismo (Cornelia & Goronzy, 2000). Apesar de ser ainda desconhecida a causa exacta que origina esta doença, sabe-se que a inflamação da membrana sinovial das articulações ocorre pela reacção autoimune das células T (Cope & Sonderstrup, 1998). Vários alelos do gene HLA-DRB1 (DRB1*0101, DRB1*0102, DRB1*0401, DRB1*0404, DRB1*0405, DRB1*0408, DRB1*1001 e DRB1*1402) estão associados a um risco elevado de desenvolvimento da artrite reumatóide, sendo que a presença de mais do

que um destes alelos eleva ainda mais esse risco, suportando a ideia do envolvimento das células T CD4 positivas, através do reconhecimento de determinado antígeno artritogénico, no desenvolvimento desta doença (Weyand *et al.*, 1992; Schmidt *et al.*, 1996; Gibert *et al.*, 2003).

A diabetes tipo 1, uma doença autoimune específica de apenas um órgão, resulta da destruição das células beta do pâncreas, responsáveis pela produção de insulina. A destruição destas células ocorre pela reacção autoimune das células T CD4 positivas despoletada por mecanismos ainda desconhecidos (Katz *et al.*, 1995; Healey *et al.*, 1995).

A susceptibilidade para o desenvolvimento da diabetes tipo 1 está associada a várias haplotipos do sistema HLA, havendo outros que conferem protecção e outros que aparentemente são neutrais (Singal & Blajchman, 1973; Larsen & Alper, 2004). A susceptibilidade para o desenvolvimento de diabetes tipo 1 é determinada principalmente pela combinação de alguns alelos HLA-DQ e HLA-DR existentes no indivíduo. Os alelos HLA-DQA1*03, HLA-DQB1*0201 e HLA-DRB1*0405 são os que conferem maior susceptibilidade ao desenvolvimento da doença enquanto que os alelos HLA-DQA1*0102, HLA-DQB1*0602, HLA-DRB1*0403 e HLA-DRB1*0406 são os que conferem maior protecção (Thorsby, 1997).

A doença celíaca (DC) é a enteropatia mais comum induzida pela ingestão de alimentos e resulta de uma intolerância permanente às proteínas glúten, presentes nos cereais, que provocam uma reacção autoimune contra as vilosidades da mucosa intestinal. Cerca de 95% dos doentes com DC apresentam o haplotipo HLA-DQA1*0501-HLA-DQB1*0201 que quando na presença do haplotipo HLA-DQA1*0201-DQB1*0202 agrava a predisposição para a doença. Os receptores membranares, codificados por

estes alelos, ligam-se aos péptidos glúten e apresentam-nos às células T induzindo uma resposta inflamatória que causa a doença (Sollid *et al.*, 1989; van de Wal *et al.*, 1996; Vartdal *et al.*, 1996).

A esclerose múltipla é uma doença inflamatória que afecta o sistema nervoso central através da desmielinização dos neurónios, levando a uma deficiência acentuada ao nível sensorial, motor e cognitivo (McFarlin & McFarland, 1982a; McFarlin & McFarland, 1982b).

A presença do haplotipo HLA-DRB1*1501-HLA-DQB1*0602 é um factor de risco no desenvolvimento de esclerose múltipla, embora sejam ainda desconhecidos os mecanismos responsáveis pelo desenvolvimento desta doença autoimune mediada pelas células T (Hillert & Olerup, 1993; Barcellos *et al.*, 2003; Sospedra & Martin, 2005).

Outras doenças têm revelado uma forte associação aos genes do sistema HLA, nomeadamente: o desenvolvimento de psoríase, uma doença inflamatória crónica que afecta a pele, o couro cabeludo, unhas e articulações, e que está fortemente associada à presença do alelo HLA-Cw*0602 (Elder *et al.*, 1994); a narcolepsia, uma doença de perturbação do sono, está associada à presença do alelo HLA-DQB1*0602 (Chabas *et al.*, 2003); a hemocromatose hereditária, uma desordem ao nível do metabolismo que provoca acumulação excessiva de ferro em alguns órgãos, está fortemente associada a uma mutação do gene *HFE* localizado na região classe I do sistema HLA (Lyon & Frank, 2001); e o desenvolvimento de alergias, nomeadamente ao pólen e a alimentos, associadas a alguns alelos de *loci* HLA classe II (Boehncke *et al.*, 1998).

Pelo seu papel central na estrutura e funcionamento do sistema imunitário, os genes do sistema HLA apresentam também uma importância

particular na resistência ou progressão de doenças infecciosas. A título de exemplo, a progressão da infecção por HIV é condicionada pela presença de determinados alelos HLA classe I e classe II e o alelo HLA-B*53 confere protecção ao desenvolvimento da malária em crianças do Oeste Africano (Hill *et al.*, 1991; Al Jabri, 2002).

A compatibilidade nos genes HLA-A, HLA-B e HLA -DRB1 entre o dador e receptor no transplante de órgãos e tecidos é fundamental para o seu sucesso por diminuir significativamente a probabilidade de rejeição (Cicciarelli, 2004). Apesar de ser mais provável encontrar um dador entre indivíduos aparentados nem sempre isso é possível pelo que tem surgido em vários países registos de dadores de órgãos e tecidos, nomeadamente de medula óssea.

1.7. Aplicação dos genes HLA nas disputas de paternidade

Os genes do sistema HLA têm sido também utilizados em testes de paternidade, tendo só por si um poder de exclusão superior a 90% (Ejele & Nwauche, 2004). Por outro lado, a taxa de mutação nos HLA é inferior à que ocorre nos loci STR (Short Tandem Repeat) pelo que a combinação destes dois sistemas na resolução de disputas de paternidade oferece resultados mais fiáveis, conferindo maior informação e diminuição das probabilidades de falsas exclusões devido a mutações (Grubic *et al.*, 2004).

1.8. Importância dos genes HLA em estudos populacionais

O extenso polimorfismo e o *linkage desequilíbrio* existentes entre diferentes *loci* do sistema HLA têm sido amplamente utilizados como marcadores genéticos em estudos antropológicos. As frequências alélicas e haplotípicas dos *loci* HLA variam entre as diferentes populações e grupos étnicos. Esta variabilidade nos genes HLA tem sido amplamente investigada em diferentes populações de forma a clarificar as suas origens e relações evolutivas (Pimthanotai *et al.*, 2001; Luo *et al.*, 2002; Piantatelli *et al.*, 2004; Ayed *et al.*, 2004).

A caracterização da diversidade alélica e haplotípica nos principais genes do sistema HLA constitui um poderoso instrumento em estudos populacionais. Estes genes extraordinariamente polimórficos têm-se revelado, a par dos microsatélites STR e do D-loop do DNA mitocondrial, muito importantes no estudo da história das migrações humanas (Arnaiz-Villena *et al.*, 2001a; Gómez-Casado *et al.*, 2001). Pela sua função imunitária, os genes HLA sofrem ao longo do tempo a influência da selecção natural, predominantemente uma selecção oscilante devido à irregularidade geográfica e temporal das influências dos elementos patogénicos. Por este facto, este conjunto de genes pode não ser tão útil no estudo de eventos demográficos muito antigos ocorridos há algumas dezenas de milhares de anos (Hedrick, 1999).

As frequências alélicas nos genes do sistema HLA têm demonstrado ser mais similares entre populações com origens comuns. Determinados alelos são mesmo característicos de algumas populações humanas com determinadas origens geográficas (Middleton *et al.* 2004). Desta forma, as frequências alélicas e a presença ou ausência de determinados alelos

específicos permitem inferir sobre a origem e relações passadas das populações.

A título de exemplo, frequências elevadas do alelo HLA-A*0201 (acima de 20%) é uma característica muito comum nas populações Caucásicas enquanto que as populações da África sub-Sahariana apresentam outros alelos como os mais frequentes, nomeadamente o A*2301 e A*3001 (Middleton *et al.* 2004; Cao *et al.* 2004). O alelo A*0201 corresponde a mais de 95% dos alelos do grupo A*02 em populações Europeias Caucásicas enquanto que nos Chineses de Singapura corresponde apenas a 25% e menos de 3% em Hindus Indianos (Tiercy, 2002). O A*0211, raro noutras populações mundiais, é o alelo mais frequente do seu grupo no Norte e Oeste da Índia (Shankarkumar *et al.* 2003).

As populações Europeias apresentam no grupo A*68 o alelo A*6801 como o mais predominante enquanto que nas populações sub-Saharianas o A*6802 é o mais prevalente (Middleton *et al.* 2004; Cao *et al.* 2004). Também os alelos A*3001 e A*3002 apresentam frequências muito mais elevadas nas populações Africanas do que nas Europeias (Middleton *et al.* 2004; Cao *et al.* 2004). O alelo A*0206 aparece com frequências elevadas apenas em populações Asiáticas e Ameríndias, podendo aparecer em quantidades residuais noutras populações por influência das primeiras (Middleton *et al.* 2004). Os alelos A*1101, A*2402 e A*3303 aparecem nos Asiáticos com frequências significativamente mais elevadas do que noutras populações. Por outro lado, os alelos B*0801 e B*4402 aparecem nos Caucásicos com frequências muito mais elevadas do que noutras populações (Cao *et al.*, 2001). Entre os Caucásicos o alelo DRB1*0401 é o mais frequente do seu grupo enquanto que nas populações Asiáticas o mais

prevalente é o DRB1*0405 (Jaini *et al.*, 2002). Alguns alelos são considerados específicos do continente Africano (e.g. A*0202, A*0225, B*1503, B*1516, B*2703, B*4202 e B*5703) e a sua presença em populações de outras origens reflecte misturas ocorridas no passado (Cao *et al.* 2004).

O grupo HLA-B*27 constitui mais um exemplo das diferenças na distribuição alélica entre populações. O B*2705 é o alelo que possui uma distribuição mais alargada no globo apresentando frequências mais elevadas nas regiões Circumpolar e sub-Ártica da Eurásia e América do Norte. O B*2702 possui distribuição restrita às populações Caucásicas, sendo particularmente predominante nalgumas populações Caucásicas do Médio Oriente (Judeus) e do Norte de África. O B*2703 é o alelo deste grupo que predomina nas populações sub-Saharianas e o B*2704 o mais frequente nas populações Asiáticas (Blanco-Gelaz *et al.*, 2001; Birinci *et al.*, 2005).

Os alelos dos diferentes genes do sistema HLA são frequentemente transmitidos em bloco de geração em geração e certas combinações (haplotipos) são encontradas mais frequentemente do que o esperado pelas respectivas frequências alélicas. Este fenómeno conhecido como *linkage disequilibrium* resulta de uma localização fisicamente próxima dos *loci* envolvidos e do efeito das forças selectivas (Huttley *et al.* 1999). Têm sido identificados inúmeros haplotipos do sistema HLA cuja distribuição geográfica coincide com a existência de relações passadas entre as populações (Gómez-Casado *et al.*, 2000; Arnaiz-Villena *et al.*, 2001^a; Sanchez-Velasco *et al.*, 2003).

Por exemplo, os haplotipos A*010101-B*0801-DRB1*030101 e A*2902-B*4403-DRB1*0701 são de origem Europeia, considerando-se o primeiro de influência Celta (Muro *et al.*, 2001). O haplotipo A*020101-B*440301-DRB1*070101 tem sido encontrado nas populações do Norte e Oeste Europeu (Sanchez-Velasco *et al.*, 2003). Já os haplotipos A*3301-B*1402-DRB1*010201 e A*020101-B*180101-DRB1*1104 são considerados de origem Mediterrânica (Gómez-Casado *et al.*, 2000). O haplotipo A*020101-B*070201-DRB1*150101 tem sido encontrado em populações Europeias e Norte Africanas e os haplotipos A*020101-B*510101-DRB1*130101, A*010101-B*5801-DRB1*070101 e A*3002-B*180101-DRB1*030101 têm sido apontados como de origem Ibérica e Berbere (Gómez-Casado *et al.*, 2000; Sanchez-Velasco *et al.*, 2003).

Desta forma, as frequências haplotípicas do sistema HLA, associadas às frequências alélicas, permitem identificar a influência genética nas populações humanas das migrações e contactos existentes entre elas ao longo do período histórico e mesmo pré-histórico.

1.9. História das populações em estudo

O território continental Português é habitado desde o Paleolítico, tendo sido o receptáculo de várias migrações de inúmeros povos com diferentes origens: Celtas, Germânicos, Fenícios e uma grande diversidade de povos da bacia Mediterrânica. Árabes e Norte Africanos ocuparam a maior parte deste território durante mais de 700 anos (Arnaiz-Villena *et al.*, 1997). No século XV, devido ao tráfico de escravos, muitos indivíduos da costa da Guiné, entre o Senegal e o golfo da Guiné, foram trazidos para Portugal (Godinho, 1965). Estas populações da costa Oeste Africana constituem o

resultado de uma complexa e contínua rede de migrações, invasões e cruzamento de povos de diferentes origens que se iniciou há 40 mil anos (Alimen, 1987). O arquipélago de Cabo Verde, descoberto por volta de 1460, foi povoado por alguns Europeus e, principalmente, por escravos originários da costa Oeste Africana, onde se situa a actual Guiné-Bissau (Carreira, 1983). Cabo Verde funcionou durante três séculos como um interposto de escravos Africanos que eram enviados para Portugal continental, Madeira, Açores e continente Americano (Russell-Wood, 1998).

A ilha da Madeira foi descoberta pelos Portugueses em 1420 e começou a ser povoada por nobres Portugueses, Judeus, exilados, condenados e escravos do Norte de África. Depois da descoberta de Cabo Verde, a Madeira começou a receber escravos da costa Oeste Africana, os quais chegaram mesmo a constituir, em meados do século XVI, 10% da população (Pereira, 1989). A partir do momento em que a Madeira se transformou num ponto importante para o comércio no Atlântico Norte, inúmeros indivíduos de várias origens Europeias tais como Espanhóis, Italianos, Franceses e Ingleses também se fixaram na ilha (Russell-Wood, 1998).

O arquipélago dos Açores foi descoberto na primeira metade do século XV e oficialmente povoado em 1439. Este arquipélago constituído por nove ilhas desempenhou um papel importante no comércio marítimo entre a América e a Europa durante os séculos XV e XVI (Monod, 1991). A sua posição estratégica atraiu povoadores de diferentes origens, a maior parte Portugueses mas também Judeus, escravos Africanos, Flamengos, Franceses, Italianos, Ingleses e Espanhóis (Guill, 1993). O povoamento do grupo Central do arquipélago dos Açores foi fortemente influenciado por

Flamengos pois já em 1490 estavam instalados nestas ilhas cerca de 2000 indivíduos provenientes de Flandres (Rogers, 1979).

1.10. Objectivos da tese

Esta tese teve por objectivo caracterizar em alta resolução a população Portuguesa relativamente aos principais genes do sistema HLA (HLA-A, HLA-B e HLA-DRB1), distinguindo entre o Norte, Centro e Sul de Portugal continental, Açores e Madeira. No sentido de obter a informação necessária para melhor distinguir a influência decorrente do comércio de escravos em que Portugal esteve envolvido nos séculos XV a XVII, foram também caracterizadas geneticamente ao mesmo nível duas populações Africanas, Guiné-Bissau e Cabo Verde.

Este trabalho pretende contribuir para a clarificação da origem e influências genéticas da população Portuguesa, em particular das sub-populações consideradas, e de Cabo Verde e Guiné-Bissau. Através do estudo do extenso polimorfismo existente nos genes do sistema HLA e do *linkage desequilibria* que ocorre entre os seus *loci*, pretende-se identificar marcadores genéticos (frequências alélicas e haplotípicas) que clarifiquem as suas origens e relações evolutivas. É ainda objectivo deste estudo relacionar os marcadores genéticos encontrados com a influência das migrações humanas que envolveram ou afectaram as populações consideradas.

Assim, a informação agora obtida para os *loci* HLA-A, HLA-B e HLA-DRB1 pretende constituir um contributo importante, a par dos dados disponíveis ao nível dos microsatélites STR e do D-loop do DNA mitocondrial (Brehm *et al.*, 2002, 2003; González *et al.*, 2003; Fernandes & Brehm, 2003;

Fernandes *et al.*, 2003; Gonçalves *et al.*, 2003; Rosa *et al.*, 2004; Gonçalves *et al.*, 2005), para o estudo e caracterização genética das populações Portuguesa, Cabo-Verdiana e Guineense.

Pelo seu importante papel ao nível do sistema imunitário, os genes HLA estão sujeitos à influência da selecção natural (Hedrick, 1999), pelo que, no âmbito deste trabalho, foram procurados sinais da sua actuação.

Os dados obtidos pretendem ainda constituir uma base de informação relevante para o desenvolvimento de futuros estudos nas populações Portuguesa, Cabo-Verdiana e Guineense relativamente à associação dos genes HLA a múltiplas doenças genéticas. No entanto, desde já, os dados obtidos pretendem também proporcionar informação sobre a prevalência na população dos alelos e haplotipos considerados associados a essas doenças (Thorsby & Lie, 2005) e sobre a maior ou menor probabilidade de encontrar dadores compatíveis com determinados fenótipos para transplantes de órgãos e tecidos.

É também propósito deste trabalho comparar a eficiência de duas técnicas de caracterização genética dos *loci* HLA (SSOP e SBT) de forma a aumentar a informação disponível sobre a sua fiabilidade.

Capítulo 2

2. Material e métodos

2.1. Amostragem das populações

A caracterização do sistema HLA na população Portuguesa foi desenvolvida numa amostra de 432 indivíduos não relacionados entre si, saudáveis e do sexo masculino cujos pais e avós nasceram e viveram na mesma região, subdividindo-se de acordo com essa informação em Portugal continental (n=145), arquipélago dos Açores (n=102) e ilha da Madeira (n=185) (Figura 2). Portugal continental foi subdividido em três diferentes regiões, com limites coincidentes com as barreiras naturais formadas pelos rios Tejo e Douro; Norte de Portugal (n=46), Centro de Portugal (n=50) e Sul de Portugal (n=49). A amostragem no arquipélago dos Açores considerou a necessidade de garantir a representatividade das diferentes ilhas, principalmente as do grupo Central e Oriental. Na análise do sistema HLA na população Açoriana foram também incluídos dados de uma amostra de 129 indivíduos da ilha Terceira anteriormente publicados (Armas *et al.*, 2004).

De forma a obter informação, não disponível anteriormente, essencial na análise comparativa da população Portuguesa, foram incluídos neste estudo 189 indivíduos saudáveis do sexo masculino de dois países Africanos; Guiné-Bissau e Cabo Verde (Figura 2). A amostra da Guiné-Bissau consistiu em 62 indivíduos pertencentes a sete diferentes grupos étnicos; balanta (n=10), papel (n=11), mandinga (n=9), felupe (n=5), bijagós (n=10), fula (n=10) e mancanha (n=10). A amostra do arquipélago de Cabo Verde reuniu um total de 124 indivíduos cujos pais e avós nasceram e viveram na mesma ilha. De acordo com esta informação a amostra foi subdividida em duas; indivíduos das ilhas do Barlavento (n=62) e indivíduos das ilhas do Sotavento (n=62).

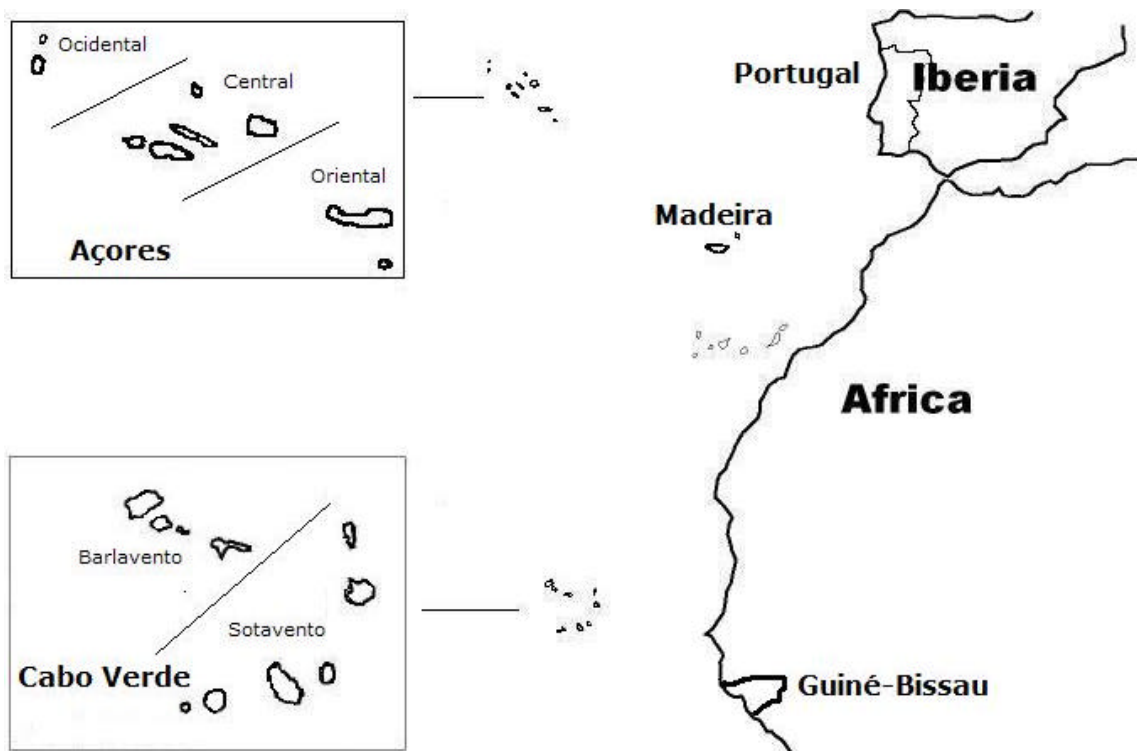


Figura 2. Mapa de localização das cinco populações estudadas: Portugal continental, Madeira, Açores, Cabo Verde e Guiné-Bissau.

2.2. Genes estudados

Com o propósito de possibilitar a caracterização do sistema HLA e inferir sobre as influências genéticas e origens da população Portuguesa, foram escolhidos para este estudo os genes HLA-A, HLA-B e HLA-DRB1. A escolha destes três *loci* prende-se com o facto de serem os mais polimórficos do sistema HLA e dos mais diversos no genoma humano, características que lhes conferem enormes potencialidades em estudos antropológicos (Mizuki *et al.* 1996; Marsh *et al.* 2005). Apesar das evidências de que estes genes estão sujeitos à pressão selectiva, vários estudos em inúmeras populações mundiais têm permitido identificar as influências genéticas resultantes de migrações e contactos entre populações ocorridos há vários milhares de anos (Arnaiz-Villena *et al.*, 1997, 1999; Gomez-Casado *et al.*, 2001; Sánchez-Velasco *et al.*, 2003).

Por outro lado, torna-se importante a caracterização destes três genes do sistema HLA por serem cruciais para o transplante de órgãos e tecidos, estarem associados ao desenvolvimento de inúmeras doenças do foro imunológico e possuírem grandes potencialidades forenses (Boehncke *et al.*, 1998; Gibert *et al.*, 2003; Cicciarelli, 2004; Ejele & Nwauche, 2004).

2.3. Métodos utilizados

A caracterização dos genes HLA-A, HLA-B e HLA-DRB1 nas populações em estudo foi desenvolvida, numa primeira fase, em média resolução com recurso à técnica SSOP (Sequence Specific Oligonucleotide Probes) (Middleton, 2000). A extracção do DNA foi concretizada através de um método clássico utilizando Chelex a partir de amostras de sangue total

contendo EDTA (Walsh *et al.*, 1991). O Chelex é uma resina que se agrega aos diversos componentes celulares, excepto o material genético, permitindo a extracção de DNA de forma rápida e com boa qualidade para amplificação por PCR (Vigilant, 1999). Foram amplificados os exões 2 e 3 dos *loci* HLA-A e HLA-B e o exão 2 do *locus* HLA-DRB1 através da técnica PCR utilizando oligonucleotídeos específicos. O DNA amplificado foi utilizado, com recurso à técnica SSOP, para identificar os grupos de alelos do sistema HLA. O DNA amplificado foi imobilizado em membranas e posteriormente hibridado com sondas oligonucleotídicas marcadas complementares a sequências específicas (Middleton, 2000).

A caracterização em alta resolução dos *loci* HLA-A, HLA-B e HLA-DRB1 em todas as populações em estudo foi desenvolvida através de sequenciação nucleotídica (SBT- Sequence Based Typing), o método actualmente mais fiável que inclusive permite a identificação directa de novos alelos. A extracção do DNA para os procedimentos de sequenciação foi efectuada através de uma metodologia clássica utilizando fenol-clorofórmio a partir de amostras de sangue total contendo EDTA. O fenol e o clorofórmio são solventes orgânicos que possibilitam a remoção de proteínas e estruturas celulares da solução aquosa deixando apenas o DNA (Sambrook *et al.*, 1989).

Os exões 2 e 3 dos *loci* HLA-A e HLA-B e o exão 2 do *locus* HLA-DRB1 foram amplificados por PCR recorrendo a oligonucleotídeos específicos (Kurz *et al.*, 1999; Williams *et al.*, 2004) (Tabela 1). Os fragmentos amplificados foram purificados e sequenciados em ambas as direcções numa ABI PRISM 310 GENETIC ANALYZER (Applied Biosystems), fazendo uso do kit ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction (Applied

Loci		Primers	F. (bp)	P.
HLA-A	Exão 2	<i>Sense</i> AI5 : 5' GAG GGT CGG GC(G/A) GGT CTC AGC CA 3' <i>Antisense</i> 3AIIn2SQR2: 5' CTC GGA CCC GGA GAC TGT 3'	390	(1)
	Exão 3	<i>Sense</i> 3AIIn2SQF2: 5' TTA GGC CAA AAA TCC CCC 3' <i>Antisense</i> 3AIIn-66 : 5' TGT TGG TCC CAA TTG TCT CCC CTC 3'	419	(1)
HLA-B	Exão 2	<i>Sense</i> Bx1: 5' GGG AGG AGC GAG GGG ACC (G/C)CA G 3' <i>Antisense</i> Bex2R: 5' GGT CAC TCA CCG (G/T)C CTC G 3'	363	(2)
	Exão 3	<i>Sense</i> Bex3F: 5' GGG GCC AGG GTC TCA CA 3' <i>Antisense</i> Bint3: 5' GGA GGC CAT CCC CGG CGA CCT AT 3'	344	(2)
HLA-DRB1	DRB1*01	<i>Sense</i> AMP1: 5' TTC TTG TGG CAG CTT AAG TT 3' <i>Antisense</i> AMPB(n): 5' CCG CTG CAC (T/C)GT GAA (G/T)CT CT 3'	261	(3)
	DRB1*02	<i>Sense</i> DR2: 5' TTC CTG TGG CAG CCT AAG AGG 3' <i>Antisense</i> AMPB(n): 5' CCG CTG CAC (T/C)GT GAA (G/T)CT CT 3'	261	(3)
	DRB1*03 ^a	<i>Sense</i> 3/11/6GF: 5' GT TTC TTG GAG TAC TCT ACG TC 3' <i>Antisense</i> DR3R: 5' GTA GTT GTG TCT GCA GTA (G/A)T 3'	232	(4)
	DRB1*04 ^b	<i>Sense</i> AMP4: 5' GT TTC TTG GAG CAG GTT AAA C 3' <i>Antisense</i> AMPB CCG CTG CAC TGT GAA GCT CT 3'	263	(3)
	DRB1*07	<i>Sense</i> DR7: 5' A CGT TTC CTG TGG CAG GG 3' <i>Antisense</i> AMPB(n): 5' CCG CTG CAC (T/C)GT GAA (G/T)CT CT 3'	265	(3)
	DRB1*08/12 ^c	<i>Sense</i> 8/12 GF: 5' GT TTC TTG GAG TAC TCT ACG GG 3' <i>Antisense</i> AMPB(n): 5' CCG CTG CAC (T/C)GT GAA (G/T)CT CT 3'	263	(4)
	DRB1*09	<i>Sense</i> DR9: 5' CGT TTC TTG AAG CAG GAT AAG TT 3' <i>Antisense</i> AMPB(n): 5' CCG CTG CAC (T/C)GT GAA (G/T)CT CT 3'	262	(3)
	DRB1*10	<i>Sense</i> DR10 : 5' ACC AGA CCA CGT TTC TTG GAG G 3' <i>Antisense</i> AMPB(n): 5' CCG CTG CAC (T/C)GT GAA (G/T)CT CT 3'	259	(3)
	DRB1*11/13 ^d	<i>Sense</i> 3/11/6GF: 5' GT TTC TTG GAG TAC TCT ACG TC 3' <i>Antisense</i> 11/6R: 5' GTA GTT GTG TCT GCA GTA GG 3'	232	(5)
	DRB1*14 ^e	<i>Sense</i> 3/11/6GF: 5' GT TTC TTG GAG TAC TCT ACG TC 3' <i>Antisense</i> DR14R: 5' CC GTA GTT GTG TCT GCA A 3'	234	(6)
<p>Amplifica também os alelos ^aDRB1*1107, DRB1*1127, DRB1*1333; ^bDRB1*1122, DRB1*1410; ^cDRB1*1105, DRB1*1317, DRB1*14(04/11/15/28/31); ^dDRB1*14 (alelos não amplificados pelos outros <i>primers</i>); ^eDRB1*1113, DRB1*1117. F. (bp)= tamanho do fragmento em pares de bases. P= programas de PCR:</p> <p>(1) 94°C/20 s; 66°C/50 s; 72°C/22 s (10 ciclos) e 94°C/20 s; 65°C/50 s; 72°C/22s (22 ciclos).</p> <p>(2) 96°C/20 s; 68°C/30 s; 72°C/30 s (35 ciclos).</p> <p>(3) 96°C/1 m; 58°C/1 m; 72°C/1 m (38 ciclos).</p> <p>(4) 96°C/1 m; 64°C/1 m; 72°C/1 m (15 ciclos) e 96°C/1 m; 56°C/1 m; 72°C/1 m (25 ciclos).</p> <p>(5) 96°C/1 m; 65°C/1 m; 72°C/1 m (15 ciclos) e 96°C/1 m; 62°C/30 s; 72°C/1 m (25 ciclos).</p> <p>(6) 96°C/1 m; 60°C/1 m; 72°C/1 m (38 ciclos).</p>				

Tabela 1. *Primers* e respectivos programas de PCR utilizados na amplificação e sequenciação dos *loci* HLA-A, HLA-B e HLA-DRB1.

Biosystems, Foster City, CA, USA) e dos *primers* utilizados na amplificação. Antes da sua utilização na reacção de sequenciação, o produto de PCR foi convenientemente purificado por centrifugação em colunas de Sephadex G-50 (Pharmacia) para remoção do excesso de *primers* e nucleótidos. A reacção de sequenciação, através da técnica de PCR, permitiu a obtenção de inúmeros fragmentos em cadeia simples com vários tamanhos, tantos tamanhos quantos nucleótidos existentes na sequência. Estes fragmentos foram obtidos através da utilização de nucleótidos marcados com fluorescência (uma marcação diferente para cada um dos quatro nucleótidos do código genético), os quais terminam a polimerização da nova cadeia sempre que são incluídos. Após remoção do excesso de nucleótidos marcados, a utilização do sequenciador automático para a separação dos fragmentos e a identificação do nucleótido inserido no fim de cada um deles permite a obtenção da sequência nucleotídica do fragmento analisado¹.

As sequências obtidas foram analisadas com o programa Matchtools Allele Identification (Applied Biosystems) de forma a identificar os alelos presentes. Os casos em que os polimorfismos no codão 86 do exão 2 do *locus* HLA-DRB1 resultaram em ambiguidades na identificação dos alelos, foram utilizados os *primers antisense* Dr86GR (5' CTG CAC TGT GAA GCT CTC AC 3') e Dr86VR (5' CTG CAC TGT GAA GCT CTC CA 3') juntamente com os *primers sense* (Tabela 1) em duas amplificações separadas. Nestes casos o programa de PCR utilizado foi o seguinte: 96°C/1 m; 66°C/1 m; 72°C/1 m (15 ciclos) e 96°C/1 m; 58°C/1 m; 72°C/1 m (25 ciclos) (Williams *et al.*, 2004).

¹ http://www.ashi-hla.org/publicationfiles/ASHI_Quarterly/25_2_2001/highthrusbt3.htm (27-09-2005)

2.4. Análise de dados

Vários parâmetros genéticos foram calculados ou estimados através da aplicação do programa Arlequin v2.000 (Schneider *et al.*, 2000) aos resultados obtidos: frequências alélicas, diversidade genética, equilíbrio Hardy-Weinberg, frequências haplotípicas, *linkage desequilibrium*, testes de neutralidade selectiva (Ewes-Watterson, Slatkin e Chakraborty) e análise de variância molecular (AMOVA).

As frequências alélicas são calculadas pela divisão do número de vezes que cada alelo for encontrado pelo total existente (equivalente ao número de cromossomas analisados). Com uma amostra adequada, as frequências alélicas calculadas permitem uma boa aproximação à real prevalência dos diferentes alelos na população. Esta informação é fundamental para a caracterização das populações em estudo e desenvolvimento de análises comparativas. As frequências alélicas nos genes em estudo não possuem uma distribuição homogénea nas populações humanas, sendo muito dependente da sua origem geográfica transforma-se numa característica específica das populações (Middleton *et al.*, 2004).

A diversidade genética, equivalente à heterozigotia, corresponde à probabilidade de dois alelos escolhidos ao acaso, num determinado *locus*, não estarem presentes no mesmo indivíduo (Nei, 1987). Com o desenvolvimento e aplicação das técnicas moleculares na caracterização alélica dos genes do sistema HLA é cada vez mais improvável a identificação errada de indivíduos homozigóticos, tornando este parâmetro de variabilidade genética cada vez mais robusto.

O equilíbrio de Hardy-Weinberg estipula que uma população em determinadas condições, nomeadamente na ausência de selecção, migração

e mutação, atinge um estado de equilíbrio no decurso de uma geração. Este equilíbrio traduz-se na equação $p^2 + 2pq + q^2 = 1$, em que p e q são as frequências dos alelos de um determinado *locus*. Os princípios do equilíbrio de Hardy-Weinberg são uma referência para identificar alterações às frequências alélicas e conseqüentemente a ocorrência de evolução (Hardy, 1908).

O termo haplotipo pode ser definido como a combinação dos alelos de dois ou mais genes localizados no mesmo cromossoma e portanto transmitidos por apenas um dos progenitores. Os haplotipos presentes num indivíduo só podem ser determinados com exactidão através do estudo dos seus ascendentes, particularmente pais e irmãos. No entanto, é possível estimar os haplotipos mais prováveis e as suas frequências aplicando um algoritmo de expectativa-maximização (Dempster *et al.*, 1977; Excoffier & Slatkin, 1995; Schneider *et al.*, 2000).

Devido à estreita relação entre os genes do sistema HLA e ao efeito das forças selectivas, muitas combinações haplotípicas possuem uma prevalência muito superior à esperada de acordo com a frequência dos alelos que as compõem. Este fenómeno designado por *linkage disequilibrium* consiste numa associação não casual de dois ou mais alelos de diferentes *loci* no mesmo cromossoma (Weir, 1996). Os padrões de *linkage disequilibrium* observados nas populações resultam da conjugação de vários factores genéticos e eventos demográficos. A recombinação tende a reduzir o *linkage disequilibrium*, situação relativamente rara entre *loci* muito próximos. O *linkage disequilibrium* é de primordial importância nos estudos genéticos populacionais pelas implicações biológicas e médicas que possui, sendo de particular interesse no mapeamento de genes

responsáveis por doenças em humanos (Li & Stephens, 2003). O Delta (D) é a medida absoluta do *linkage disequilibrium* e é definido pela diferença entre a frequência haplotípica observada e a frequência haplotípica esperada tendo em conta a prevalência dos alelos que compõem o haplotipo (Weir, 1996).

Os testes de neutralidade selectiva Ewens-Watterson, Ewens-Watterson-Slatkin e Chakraborty baseiam-se no modelo do alelo infinito e constituem uma forma de examinar a presença de forças selectivas que influenciam a diversidade alélica em cada *locus* (Ewens, 1972; Watterson, 1978; Chakraborty, 1990; Slatkin, 1994; Slatkin, 1996). Estes testes de neutralidade, incluídos no programa Arlequin, comparam as frequências alélicas observadas com as esperadas num modelo neutral, onde actua apenas a mutação e a deriva genética, detectando os desvios ao modelo neutral e o défice ou excesso de diversidade genética.

As populações em estudo foram comparadas entre si e com outras disponíveis com o mesmo nível de informação (Ellis *et al.*, 2000; Pimthanotai *et al.* 2001; Muro *et al.*, 2001; Layrisse *et al.*, 2001; Luo *et al.*, 2002; Tsuneto *et al.*, 2003; Sanchez-Velasco *et al.*, 2003; Middleton *et al.*, 2003, 2004; Piancatelli *et al.*, 2004; Ayed *et al.*, 2004; Leffell *et al.*, 2004; Uyar *et al.*, 2004). Foi desenvolvida uma análise da variância molecular (AMOVA) com base nas distâncias Euclidianas entre os diferentes pares de alelos. AMOVA é um método que estima e testa a diferenciação entre populações a partir de dados moleculares (Excoffier *et al.*, 1992). A variação genética entre as diferentes populações e sub-populações em estudo foi estimada e os correspondentes índices de fixação *Fst* foram utilizados na avaliação da existência de diferenças significativas entre elas.

A análise comparativa das populações para a construção de dendrogramas foi desenvolvida com recurso ao conjunto de programas incluídos no pacote de *software* PHYLIP v.3.6 (Felsenstein, 2004). A partir das frequências alélicas dos genes HLA-A, HLA-B e HLA-DRB1 o programa SEQBOOT do PHYLIP gerou múltiplas amostragens a partir dos dados iniciais. Para cada uma destas amostragens foi gerada uma matriz de distâncias através do programa GENDIST, as quais foram utilizadas no programa NEIGHBOR para a construção das árvores filogenéticas. Uma única árvore de consenso foi construída a partir dos vários dendrogramas com recurso ao programa CONSENSE. As populações foram também comparadas através de uma análise das coordenadas principais (PCO-Principal coordinates analysis) utilizando o MultiVariate Statistical Package MVSP3 (Kovach Computing Services, Anglesey, Wales, UK)¹.

Os resultados obtidos na caracterização alélica dos genes HLA-A, HLA-B e HLA-DRB1, nos 1242 cromossomas em estudo, efectuada através de duas técnicas distintas, SSOP e SBT, foram comparados entre si de forma a identificar as discrepâncias existentes.

¹ <http://www.kovcomp.com/mvsp> (25-04-2003)

Capítulo 3

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

Hélder Spínola, António Brehm, Fionnoula Williams, José Jesus and
Derek Middleton

Reprinted with minor modifications from *Annals of Human Genetics*
2002: 66: 285-296.

3.1. Abstract

HLA-A, HLA-B, and HLA-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.

3.2. 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 cells and 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 disequilibrium* (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; Arnaiz-Villena *et al.*, 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 (Arnaiz-Villena *et al.*, 2001a; Arnaiz-Villena *et al.*, 2001b; Martinez-Laso *et al.*, 1995; Chingé *et al.*, 1997; 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 (Arnaiz-Villena *et al.*, 1997; Martinez-Laso *et al.*, 1995; Pinto *et al.*, 1996; 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 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 Africans or other European populations (Gómez-Casado *et al.*, 2001; Bruges-Armas *et al.*, 1999; Sánchez-Velasco *et al.*, 1999), 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.

3.3. Materials and methods

3.3.1. Population samples

The study population consisted of a total of 394 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), Centre (CP, n=50) and South Portugal (SP, n=49). 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 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 2 shows the origin and total number of samples included in the analysis.

Pop. codes	Country/population	n	References
NP	North Portugal	46	Present study
CP	Centre Portugal	50	Present study
SP	South Portugal	49	Present study
MA	Madeira	185	Present study
CV	Cabo Verde	64	Present study
AZ	Azores	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	Spain/Murcian	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/Rimaibe	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

Table 2. Populations used in the present study and for comparisons. n= maximum number of individuals analysed for the 3 *loci*.

3.3.2. HLA typing and statistics

Genomic DNA was isolated from whole blood containing EDTA, using a Chelex standard method (Walsh *et al.*, 1991). HLA Class I (A and B) and Class II (DRB1) allelic lineages were DNA-typed by a medium resolution PCR-SSOP 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.2000 (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 (Bruges-Armas *et al.*, 1999) (AZ), was performed based on Euclidean distances between all pairs of haplotypes (Excoffier *et al.*, 1992). The total genetic variation between the populations was portioned 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 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.

3.4. Results and discussion

3.4.1. HLA class I and II allele frequencies

All populations studied are in Hardy-Weinberg equilibrium. 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 ($F_{ct} = 0.0047$) and between populations F_{st} values were non significant except when each one was compared with CV (minimum value of F_{st} pairwise comparisons was 0.013, in all four cases, $p < 0.0001$). Tables 4, 5, and 6 present the HLA-A, HLA-B, and HLA-DRB1 allele frequencies respectively.

		NP	CP	SP	MA	CV
N° alleles	HLA-A	13	15	14	18	17
	HLA-B	20	22	22	22	19
	HLA-DR	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)
Heterozygosity	HLA-A	0.913	0.940	0.877	0.821	0.921
	HLA-B	0.913	0.880	0.877	0.870	0.937
	HLA-DR	0.891	0.880	0.836	0.827	0.906

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

Alleles	NP	CP	SP	MA	CV	MD ^a	CA ^a
A*01	7.6	13	5.1	7.0	5.5	4.7	1.1
A*02	31.6	28	24.5	27.6	17.2	17.9	19.0
A*03	6.5	11	8.2	7.0	7.0	5.5	8.3
A*11	5.4	5.0	14.3	5.9	2.3	0	0
A*23	10.9	1.0	6.1	6.8	13.3	18.5	18.6
A*24	5.4	8.0	10.2	14.1	6.3	0.3	0.5
A*25	1.1	0	0	0.8	0	0	0
A*26	5.4	1.0	2.0	3.5	4.7	4.5	2.3
A*29	3.3	8.0	7.1	6.5	4.7	4.0	10.4
A*30	2.2	4.0	5.1	1.6	15.4	16.8	14.3
A*31	6.5	4.0	4.1	2.7	1.6	1.2	1.1
A*32	5.4	4.0	3.1	5.4	3.1	1.9	1.1
A*33	8.7	3.0	4.1	3.0	5.5	13.1	2.8
A*34	0	0	0	0.3	0.8	2.5	2.2
A*36	0	0	0	0.3	0	0	2.2
A*66	0	2	1	1.1	0	0.3	7.1
A*68	0	7	5.1	5.7	5.5	8.2	3.4
A*69	0	1	0	0	1.6	0	0.6
A*74	0	0	0	0.7	4.7	0.6	5
A*80	0	0	0	0	0.8	0	0

Table 4. HLA-A allele frequencies in the Portuguese and Cabo Verde populations. Mandenka and Cameroon populations are also included for comparison as typical West African data (the 3 highest values in each population in bold). ^aData from Dard *et al.* (1992) and Ellis *et al.* (2000).

Alleles	NP	CP	SP	MA	CV	MD ^a	CA ^a
B*07	3.3	8.0	1.0	5.1	6.3	6.1	6.0
B*08	6.5	4.0	2.0	7.6	8.6	7.3	5.4
B*13	1.1	1.0	1.0	0.5	0	0	2.7
B*14	6.5	4.0	5.1	7.3	8.6	2.8	3.2
B*15	5.4	8.0	2.0	6.5	13.3	11.9	9.3
B*18	7.6	9.0	11.2	6.5	3.1	3.6	2.7
B*27	4.3	1.0	2.0	1.9	3.1	1.5	0.5
B*35	9.8	14.0	21.7	9.7	12.3	13.3	7.1
B*37	0	2.0	0	0.5	0	0.6	0.5
B*38	0	0	1.0	3.8	0.8	0	0
B*39	4.3	1.0	1.0	1.9	0	1.9	0.5
B*40	5.4	3.0	1.0	4.9	3.1	0	1.1
B*41	2.2	0	1.0	1.1	0.8	0.6	0.5
B*42	0	1.0	0	0	1.6	2.8	4.9
B*44	15	15.0	20.5	15.6	6.3	6.8	8.7
B*45	1.1	5.0	1.0	1.6	1.6	0.9	3.3
B*47	0	1.0	0	0	0	1.5	1.6
B*49	7.6	2.0	4.1	6.2	2.3	5.3	5.4
B*50	2.2	0	4.1	4.1	0.8	0.3	0.5
B*51	13.3	13.0	11.2	10.0	7.8	3.8	1.6
B*52	0	1.0	0	0.3	0	5.3	0
B*53	1.1	1.0	0	1.1	8.6	4.0	10.9
B*55	1.1	1.0	0	0	0	0	0
B*56	0	0	1.0	0	0	3.1	0
B*57	0	3.0	4.1	2.2	5.5	1.2	2.7
B*58	2.2	2.0	2.0	2.2	5.5	8.2	16.5
B*67	0	0	1.0	0	0	0	0
B*78	0	0	0	0	0	7.2	0
B*81	0	0	1.0	0	0	0	4.4

Table 5. HLA-B allele frequencies in the Portuguese and Cabo Verde populations. Mandenka and Cameroon populations are also included for comparison as typical West African data (the 3 highest values in each population in bold). ^aData from Dard *et al.* (1992) and Ellis *et al.* (2000).

Alleles	NP	CP	SP	MA	CV	MD ^a
DRB1*01	15.2	4.0	11.2	8.9	12.5	1.1
DRB1*02	12.0	14.0	5.1	11.6	3.9	2.9
DRB1*03	13.0	16.0	14.3	12.2	10.2	13.4
DRB1*04	9.8	14.0	13.3	9.7	10.2	0.8
DRB1*07	17.7	14.0	15.3	16.0	7.0	6.8
DRB1*08	4.0	7.0	3.1	4.1	3.0	12.4
DRB1*09	1.1	2.0	1.0	1.1	6.3	1.0
DRB1*10	0	0	1.0	0.8	5.5	3.8
DRB1*11	9.8	11.0	14.3	13.0	16.0	18.2
DRB1*12	1.1	0	3.1	1.1	2.0	0.3
DRB1*13	13.0	16.0	17.3	19.6	23.4	38.5
DRB1*14	3.3	2.0	1.0	1.9	0	0.8

Table 6. HLA-DRB1 allele frequencies in the Portuguese and Cabo Verde populations. Mandenka population are also included for comparison as typical West African data (the 3 highest values in each population in bold). ^aData from Dard *et al.* (1992).

3.4.1.1. 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-Saharan (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.

3.4.1.2. 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.*, 2001a; Arnaiz-Villena *et al.*, 1995; Arnaiz-Villena 2001b). HLA-B*35 is 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.

3.4.1.3. HLA-DRB1

HLA Class II DRB1*13 can be found in sub-Saharan populations at 20%, rising to 48.8% in a Gambian population (Hill *et al.*, 1991; Tiercy *et al.*, 1992) (Table 6). 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 DRB1*01, DRB1*04 and DRB1*07 are all low and the frequencies of DRB1*13 and DRB1*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, DRB1*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.

3.4.2. Haplotype frequencies

3.4.2.1. West European haplotypes

A*01-B*08-DRB1*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%) (Table 7), 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-DRB1*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 (Bruges-Armas *et al.*, 1999) (2%) and in the present study only in SP (3%). A*02-B*44-DRB1*07 is widespread over Europe but the present study 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).

Haplotypes	Frequencies (%)					Haplotypes	Frequencies (%)				
	NP	CP	SP	MA	CV		NP	CP	SP	MA	CV
A*01-B*08-DRB1*03		3.0		2.2	0.8	A*03-B*58-DRB1*04	1.1				
A*01-B*08-DRB1*13	2.2	1.0				A*11-B*18-DRB1*11			2.2		
A*01-B*35-DRB1*07		2.8	2.0			A*11-B*35-DRB1*01	1.1		3.0	1.4	0.8
A*02-B*08-DRB1*03				1.9		A*11-B*35-DRB1*11			2.9		
A*02-B*14-DRB1*01			2.0	0.3		A*11-B*44-DRB1*13		1.0	2.0	0.3	
A*02-B*14-DRB1*13					2.3	A*23-B*07-DRB1*02					1.6
A*02-B*15-DRB1*04	2.2	1.0	1.0			A*23-B*08-DRB1*03			2.0	0.8	
A*02-B*18-DRB1*02	2.2					A*23-B*15-DRB1*11					3.1
A*02-B*18-DRB1*03		2.0	4.0			A*23-B*49-DRB1*13					3.3
A*02-B*18-DRB1*08		2.0				A*24-B*18-DRB1*14					2.2
A*02-B*18-DRB1*11				2.4		A*24-B*35-DRB1*11		2.0	1.0	0.3	1.6
A*02-B*27-DRB1*04					2.3	A*24-B*44-DRB1*02				1.4	
A*02-B*35-DRB1*01	3.3			1.1		A*24-B*44-DRB1*11				0.8	
A*02-B*35-DRB1*11					2.3	A*24-B*44-DRB1*13			4.0		
A*02-B*35-DRB1*13			2.0		0.8	A*24-B*51-DRB1*07			3.0		
A*02-B*39-DRB1*11	2.2					A*26-B*38-DRB1*13				1.4	
A*02-B*40-DRB1*13		2.0				A*29-B*14-DRB1*01					2.3
A*02-B*41-DRB1*13	2.2					A*29-B*14-DRB1*13			1.0	1.9	
A*02-B*44-DRB1*01		2.0		0.6		A*29-B*44-DRB1*07		2.0	2.0	1.5	0.8
A*02-B*44-DRB1*02		4.9				A*29-B*44-DRB1*11	1.1			1.9	
A*02-B*44-DRB1*07	5.4			0.6		A*29-B*44-DRB1*13			1.0		
A*02-B*44-DRB1*04			3.0			A*30-B*08-DRB1*04					1.6
A*02-B*44-DRB1*13				2.1		A*30-B*18-DRB1*03			3.8		
A*02-B*50-DRB1*07				2.1		A*30-B*18-DRB1*13				0.5	2.3
A*02-B*51-DRB1*04	1.1		3.0	2.2		A*30-B*35-DRB1*03			1.2	0.3	0.8
A*02-B*51-DRB1*11	3.3	2.0		0.6		A*30-B*35-DRB1*10					2.3
A*02-B*51-DRB1*13		3.0	2.0	1.9	1.6	A*30-B*44-DRB1*07	2.2				
A*02-B*53-DRB1*13					2.3	A*30-B*57-DRB1*07					0.8
A*03-B*35-DRB1*01			3.0			A*30-B*57-DRB1*13					0.8
A*03-B*35-DRB1*02		2.0				A*32-B*08-DRB1*03	2.2				
A*03-B*35-DRB1*04		2.0				A*33-B*14-DRB1*01	4.3				1.3
A*03-B*44-DRB1*02	2.2					A*33-B*14-DRB1*13		2.0			
A*03-B*44-DRB1*11			2.0			A*68-B*07-DRB1*02		2.0			
A*03-B*49-DRB1*04			2.0			A*68-B*45-DRB1*04		3.0			

Table 7. 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.

3.4.2.2. 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%.

3.4.2.3. 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 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%).

3.4.2.4. The sub-Saharan component

Few haplotypes have been unambiguously linked to sub-Saharans. 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 7). 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.

3.4.2.5. 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.

3.4.3. Principal Component Analysis

The PCA depicted in Figure 3 is based on HLA-A and HLA-B and HLA-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* HLA-A and HLA-B for all populations, and the result was similar to that obtained when HLA-DR values were

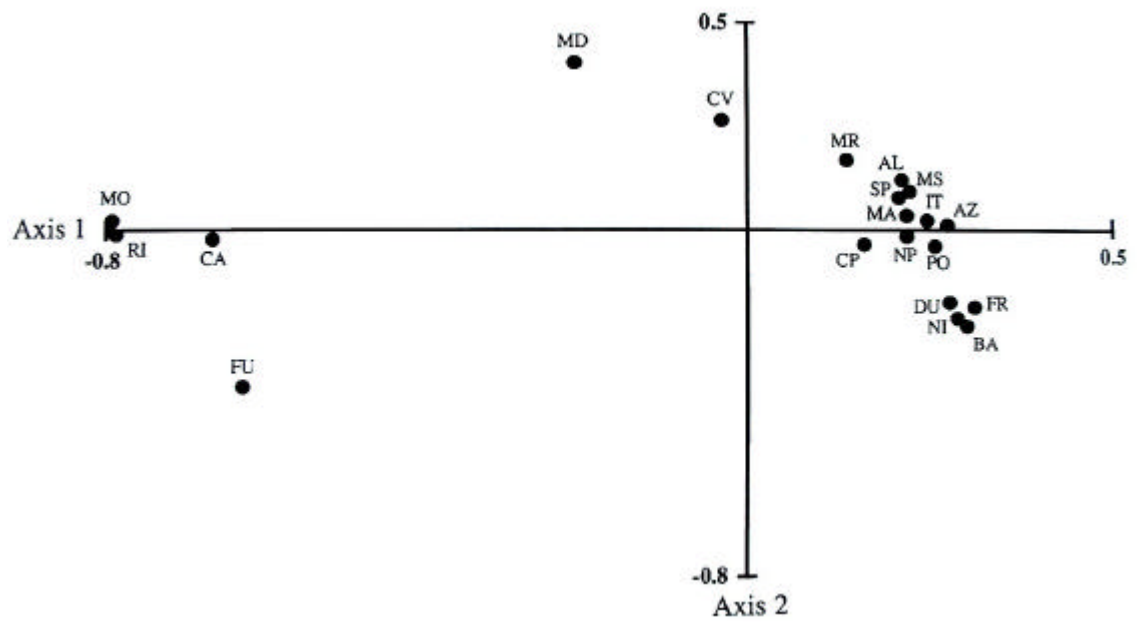


Figure 3. Principal Component Analysis (PCA) showing the relationships among the populations described in Table 2 on the basis of their HLA allele frequencies. First axis extracted 50.52% of the total variance, second axis 10.34%.

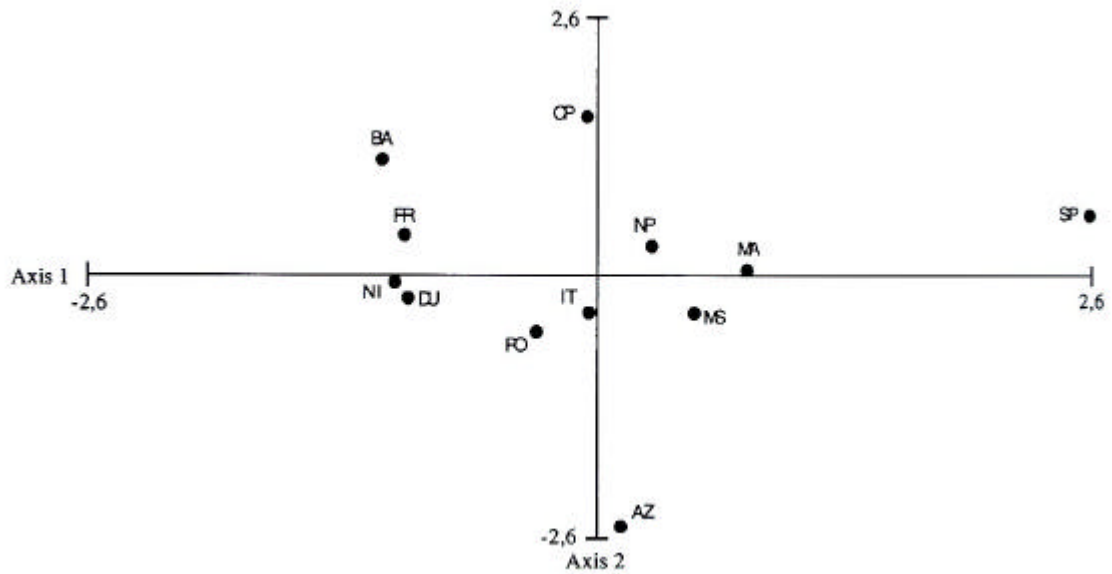


Figure 4. Principal Component Analysis (PCA) restricted to the European samples in order to clarify the relationships involving the five populations included in the present study. The plot is based on data from the three *loci*. First axis extracted 20.71% on the total variation, second axis 19.02% and the third 14.8% (not shown).

included. This prompted us not to exclude the HLA-DR data from 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 4). The Azores and SP

appear separated from the remaining Portuguese populations, as seen in the plot in Figure 4, a fact supported by the high number of haplotypes unique to these populations (see Table 7).

3.4.4. 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 *et al.*, 2003). 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 haplotypes, enabling populations to be discriminated. Table 7 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 4 and 5, HLA-A*01, HLA-B*14, HLA-B*40 and HLA-B*51). The same is true for 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 the Neolith Iberic markers (A*30-B*18) (Martinez-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.

Capítulo 4

4. HLA genes in Portugal inferred from sequence-based typing: in the crossroad between Europe and Africa.

Hélder Spínola, Derek Middleton, António Brehm.

Reprinted with minor modifications from *Tissue Antigens* 2005: 66: 26-36.

4.1. Abstract

HLA-A, HLA-B, and HLA-DRB1 polymorphism was examined in the Portuguese population, discriminating between North, Centre and South inhabitants. All data were obtained at high-resolution level, using sequence-based typing. The most frequent allele at each *loci* was: A*020101 (26%), B*440301 and B*510101 (12% each), and DRB1*070101 (15%). The predominant three-*loci* haplotype was A*020101-B*440301-DRB1*070101 (3.1%), highly frequent in North Portugal (5.4%), lower in Centre (2%) and absent in the South.

The present study demonstrates that the Portuguese population has been genetically influenced by Europeans and North Africans, via several

historic immigrations. North Portugal seems to concentrate, probably due the pressure of Arab expansion, an ancient genetic pool originated from several North Africans and Europeans influences throughout millenniums. South Portugal shows a North African genetic influence, probably of recent origin by means of Berbers accompanying Arab expansion. We found that Centre Portugal is the distribution limit of some alleles and haplotypes that characterize the North or the South of the country. Despite North, Centre and South Portugal not being significantly different in allele frequencies, this study shows that HLA allele and haplotype frequencies are not homogeneous in the country. North and South Portugal shows more similarity to North Africans in opposition to Centre which appears closer to Europeans.

4.2. Introduction

The Human Leukocyte Antigen (HLA) genetic system is perhaps one of the most polymorphic in humans. It consists of a closely linked set of genes highly important in transplantation, anthropological and forensic fields (Riley & Olerup, 1992). In the last years a considerable number of studies linking specific HLA alleles to particular disease conditions, has appeared in the literature. The general HLA profile of a population can be helpful in several fields including tracing its origins relationships with neighbouring populations. Few studies have been performed to characterize the Portuguese at the population biology level. There are a few studies focused on HLA-A, -B, and -DRB1 genes of Portuguese, but these are medium-resolution surveys (Arnaiz-Villena *et al.*, 1997; Spínola *et al.*, 2002).

With recent developments in DNA typing, HLA has been studied at a high resolution level, particularly full sequencing of information which is important in comparisons of populations and disease studies (Piancatelli *et al.*, 2004; Ayed *et al.*, 2004; Anjos & Polychronakos, 2004). Unfortunately as high resolution HLA allele typing is still scarce comparison to suitable data available is difficult.

The Iberian Peninsula, and particularly Portugal, is situated in the southern edge of Europe. Inhabited since the Palaeolithic, the territory has been the receptacle of various human movements from peoples of different origins: Celts, Germanics, Phoenicians and a myriad of peoples with origin on the Mediterranean basin. In historical times the Arabs and North Africans settled most of the country for more than 700 years (Arnaiz-Villena *et al.*, 1997). In the 15th century there was a considerable input of sub-Saharan, brought as slaves mostly from the West African coasts particularly from the Senegal region to the Gulf of Guinea (Godinho, 1965). The impact of all these peoples has been studied with various kinds of DNA markers such as mtDNA (González *et al.*, 2003), microsatellites (Fernandes & Brehm, 2003), Y-chromosome biallelic markers (Gonçalves *et al.*, 2005) as well as HLA surveys (Arnaiz-Villena *et al.*, 1997; Spínola *et al.*, 2002). Recently, mtDNA (González *et al.*, 2003) and Y-chromosome binary markers as well as Y-STRs (Short Tandem Repeat) (Gonçalves *et al.*, 2005) have shed a new hypothesis on the presence of Berber markers in Portugal. According to these authors, a high and almost exclusively prevalence of these markers in the North of the country most probably points to pre-Neolithic movements to Iberia and are not due to recent slave trade or the Arab settlement.

The main goal of the present work is to present a clear genetic profile of the Portuguese based on high resolution typing of the HLA-A, -B and –DRB1 *loci*. Allele and haplotype frequencies of three Portuguese sub-populations and their relatedness to other European and circum-Mediterranean populations were evaluated. The comparison of these frequencies with other populations may also contribute to clarify the origin of the genetic heterogeneity already found in the Portuguese (Spínola *et al.*, 2002).

4.3. Materials and methods

4.3.1. Population samples and HLA typing

The study population consisted of a total of 145 healthy unrelated males randomly distributed in North (NP, n=46), Centre (CP, n=50) and South (SP, n=49) Portugal. EDTA Blood samples were collected after informed consent, from donors whose parents and grandfathers were born and living in the same area. Genomic DNA was isolated from whole blood containing EDTA using a phenol-chloroform procedure and frozen at -20°C until use. All subjects were typed for HLA-A and HLA-B class I *loci* and HLA-DRB1 class II *locus*. Sequence based typing (SBT) of the HLA-A and –B *loci* was performed on exon 2 and exon 3 accordingly to Kurz *et al.* (1999) and Pozzi *et al.* (1999) with minor modifications. DNA fragments amplified by polymerase chain reaction (PCR) were purified and sequenced using ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kits (Applied Biosystems, Foster City, CA, USA) in an ABI PRISM 310 GENETIC ANALYZER (Applied Biosystems), according to the manufacturer's instructions. HLA-DRB1 SBT was performed on exon 2, amplified and sequenced with group

specific primers used in SSOP typing (Williams *et al.*, 2004). All samples were PCR amplified with each of the group specific pair of primers. The positive reactions were sequenced with the amplification primers according to the above description for HLA-A and -B. Sequencing was always performed in forward and reverse directions and processed using the Matchtools Allele Identification packet (Applied Biosystems) which assess the typing by an updated HLA-sequence library with alleles defined up to June 2001. In the present study we did not test for polymorphisms outside exons 2 and 3 in HLA-A and -B and outside exon 2 in HLA-DRB1. Therefore, a few groups of alleles could not be distinguished. A summary of the unresolved ambiguities and the assignment label used in each of these cases for frequency analysis is given in Table 8. Likewise, some heterozygous samples could have more than one allele's combination. In these situations, for allele frequencies and haplotype calculations we used a maximum likelihood estimation assignment, choosing the combination of the most frequent single alleles.

<i>Locus</i>	Ambiguous alleles	assignment label
HLA-A	A*010101/A*0104N	A*010101
	A*020101/A*020108/A*0209/A*0243N	A*020101
	A*2301/A*2307N	A*2301
	A*240201/A*240203/A*2409N/A*2411N	A*240201
	A*680102/A*6811N	A*680102
HLA-B	B*0705/B*0706	B*0705/06
	B*180101/B*1817N	B*180101
	B*2705/B*2713	B*2705
	B*350101/B*3540N/B*3542	B*350101
	B*440201/B*4419N/B*4427	B*440201
	B*510101/B*5111N/B*5130/B*5132	B*510101
HLA-	DRB1*120101/DRB1*1206	DRB1*120101

Table 8. Ambiguous HLA alleles and their respective assignment labels under which they appear in the text.

4.3.2. Data analysis

Basic genetic parameters (allele and haplotype frequencies, gene diversity, and Hardy-Weinberg equilibrium) at the three *loci* were estimated with Arlequin v2.000 (Schneider *et al.*, 2000). In the present study haplotypes were inferred using a maximum likelihood approach. *Linkage disequilibrium* (D) and relative *linkage disequilibrium* (D') between two alleles at two different loci and their level of significance (P) was calculated according to Weir (1996). Ewes-Watterson's, Slatkin's and Chakraborty's selective neutrality statistical tests were applied to examine the presence of selective forces influencing allelic diversity at each *locus*.

The Portuguese data was compared to several populations which are available at the same typing resolution (Ellis *et al.*, 2000; Pimtanonthai *et al.*, 2001; Luo *et al.*, 2002; Middleton *et al.*, 2003; Piancatelli *et al.*, 2004; Ayed *et al.*, 2004). An analysis of molecular variance (AMOVA) was performed with these populations based on Euclidean distances between all pairs of haplotypes (Excoffier *et al.*, 1992). The total genetic variation between the 3 Portuguese sub-populations (SP, CP and NP) was estimated and the correspondent *Fst* value was used to evaluate if there was significant difference between them. Variance components were tested for significance by nonparametric randomisation tests using 10000 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 Portuguese data set with other populations available in the literature was achieved using the software included in the PHYLIP v.3.6 package (Felsenstein, 2004). First, SEQBOOT was used to

perform a bootstrap analysis from gene frequency data. The program generates multiple data sets resampled 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 CONSENSE. The topology was visualized with TreeView (Page, 1996). Dendrograms were based on HLA-A, -B and -DRB1 or on HLA-A and -B, in order to include some relevant populations with no high resolution typing for HLA-DRB1 *locus*.

Fst values were used to plot a principal coordinate analysis with a restricted number of populations using the statistical package NTSYS (Rohlf, 1997).

4.4. Results

Tables 9-12 show the allele frequencies of HLA-A, HLA-B and HLA-DRB1 *loci* in the Portuguese population. A total of 27 HLA-A, 43 HLA-B and 36 HLA-DRB1 alleles were found. The population was in Hardy-Weinberg equilibrium for each studied *locus*. The heterozygosity was high for all examined *loci* (HLA-A: 0.89; HLA-B: 0.94; HLA-DRB1: 0.93). Ewes-Watterson, Slatkin's and Chakraborty's tests of selective neutrality yielded non-significant results at the 3 HLA *loci* (all $P > 0.05$), suggesting lack of selection. An AMOVA between the three sub-populations (SP, CP and NP) shows that only 0.03% of the total genetic variation found can be attributed to differences among populations, the remaining is due to within population differences. The exact test of population differentiation performed by Arlequin shows non significant results (SP/CP: $P = 0.23$; SP/NP: $P = 0.056$; CP/NP: $P = 0.055$).

Portugal allele frequencies					
HLA-A	Frequencies	HLA-B	Frequencies	HLA-DRB1	Frequencies
A*010101	0.086	B*070201	0.038	DRB1*0101	0.045
A*020101	0.261	B*0705/06	0.004	DRB1*010201	0.037
A*0202	0.004	B*0801	0.040	DRB1*0103	0.020
A*0205	0.010	B*1302	0.010	DRB1*030101	0.133
A*022001	0.004	B*1401	0.007	DRB1*030201	0.004
A*030101	0.079	B*1402	0.045	DRB1*030501	0.004
A*0302	0.007	B*150101	0.040	DRB1*0315	0.004
A*110101	0.079	B*1503	0.004	DRB1*040101	0.045
A*1110	0.004	B*15170101	0.007	DRB1*0402	0.014
A*2301	0.058	B*1518	0.004	DRB1*040301	0.010
A*240201	0.079	B*180101	0.085	DRB1*0404	0.007
A*2501	0.004	B*1805	0.007	DRB1*040501	0.027
A*2601	0.024	B*2705	0.020	DRB1*040701	0.017
A*2608	0.004	B*2708	0.004	DRB1*0408	0.004
A*290101	0.004	B*350101	0.082	DRB1*070101	0.150
A*2902	0.059	B*3502	0.048	DRB1*0704	0.004
A*3001	0.010	B*3503	0.017	DRB1*080101	0.035
A*3002	0.028	B*3504	0.004	DRB1*080302	0.007
A*310102	0.048	B*3701	0.007	DRB1*080401	0.004
A*3201	0.041	B*3801	0.004	DRB1*090102	0.014
A*3301	0.048	B*3901	0.014	DRB1*1001	0.004
A*3303	0.004	B*390602	0.007	DRB1*110101	0.066
A*6601	0.010	B*4001	0.017	DRB1*1102	0.020
A*680101	0.024	B*4002	0.014	DRB1*110401	0.027
A*680102	0.007	B*4102	0.010	DRB1*1113	0.004
A*6802	0.010	B*4201	0.004	DRB1*120101	0.014
A*6901	0.004	B*440201	0.048	DRB1*130101	0.079
		B*440301	0.120	DRB1*130201	0.045
		B*4501	0.024	DRB1*1303	0.024
		B*4701	0.004	DRB1*1304	0.004
		B*4901	0.045	DRB1*1305	0.004
		B*5001	0.014	DRB1*140101	0.020
		B*5002	0.007	DRB1*150101	0.076
		B*510101	0.120	DRB1*150201	0.004
		B*520101	0.004	DRB1*160101	0.020
		B*5301	0.007	DRB1*160202	0.004
		B*5501	0.007		
		B*5601	0.004		
		B*570101	0.020		
		B*570301	0.004		
		B*5801	0.021		
		B*670102	0.004		
		B*8101	0.004		

Table 9. HLA-A, -B and -DRB1 allele frequencies in the Portuguese population.

South Portugal allele Frequencies					
HLA-A	Frequencies	HLA-B	Frequencies	HLA-DRB1	Frequencies
A*010101	0.05102	B*070201	0.01020	DRB1*0101	0.05102
A*020101	0.23469	B*0705/06	0	DRB1*010201	0.03061
A*0202	0	B*0801	0.02041	DRB1*0103	0.03061
A*0205	0.01020	B*1302	0.01020	DRB1*030101	0.14286
A*022001	0	B*1401	0.0204	DRB1*030201	0
A*030101	0.08163	B*1402	0.03061	DRB1*030501	0
A*0302	0	B*15010101	0.02041	DRB1*0315	0
A*110101	0.13265	B*1503	0	DRB1*040101	0.05102
A*1110	0.01020	B*15170101	0	DRB1*0402	0.02041
A*2301	0.06122	B*1518	0	DRB1*040301	0.02041
A*240201	0.10204	B*180101	0.11224	DRB1*0404	0.01020
A*2501	0	B*1805	0	DRB1*040501	0.02041
A*2601	0.02041	B*2705	0.02041	DRB1*040701	0.01020
A*2608	0	B*2708	0	DRB1*0408	0
A*290101	0	B*350101	0.13265	DRB1*070101	0.14286
A*2902	0.07143	B*3502	0.06122	DRB1*0704	0.01020
A*3001	0	B*3503	0.01020	DRB1*080101	0.02041
A*3002	0.05102	B*3504	0.01020	DRB1*080302	0
A*310102	0.04082	B*3701	0	DRB1*080401	0.01020
A*3201	0.03061	B*3801	0.01020	DRB1*090102	0.01020
A*3301	0.04082	B*390101	0.01020	DRB1*1001	0.02040
A*3303	0	B*390602	0	DRB1*110101	0.07143
A*6601	0.01020	B*400101	0.01020	DRB1*1102	0.02041
A*680101	0.02041	B*4002	0	DRB1*110401	0.04082
A*680102	0.01020	B*4102	0.01020	DRB1*1113	0
A*6802	0.02041	B*4201	0	DRB1*120101	0.03061
A*6901	0	B*440201	0.09184	DRB1*130101	0.08163
		B*440301	0.11224	DRB1*130201	0.05102
		B*4501	0.01020	DRB1*1303	0.04081
		B*4701	0	DRB1*1304	0
		B*4901	0.04082	DRB1*1305	0
		B*5001	0.03061	DRB1*140101	0.01020
		B*5002	0.01020	DRB1*150101	0.02041
		B*510101	0.11224	DRB1*150201	0
		B*520101	0	DRB1*160101	0.03061
		B*5301	0	DRB1*160202	0
		B*5501	0		
		B*5601	0.01020		
		B*570101	0.03061		
		B*570301	0.01020		
		B*5801	0.02041		
		B*670102	0.01020		
		B*8101	0.01020		

Table 10. HLA-A, -B and -DRB1 allele frequencies in South Portugal.

Centre Portugal allele frequencies					
HLA-A	Frequencies	HLA-B	Frequencies	HLA-DRB1	Frequencies
A*010101	0.13	B*070201	0.07	DRB1*0101	0.02
A*020101	0.26	B*0705/06	0.01	DRB1*010201	0.02
A*0202	0	B*0801	0.04	DRB1*0103	0
A*0205	0.01	B*1302	0.01	DRB1*030101	0.14
A*022001	0.01	B*1401	0	DRB1*030201	0.01
A*030101	0.09	B*1402	0.04	DRB1*030501	0.01
A*0302	0.02	B*15010101	0.06	DRB1*0315	0
A*110101	0.05	B*1503	0	DRB1*040101	0.02
A*1110	0	B*15170101	0.01	DRB1*0402	0.02
A*2301	0.01	B*1518	0.01	DRB1*040301	0.01
A*240201	0.08	B*180101	0.08	DRB1*0404	0.01
A*2501	0	B*1805	0.02	DRB1*040501	0.05
A*2601	0.01	B*2705	0.01	DRB1*040701	0.02
A*2608	0	B*2708	0	DRB1*0408	0.01
A*290101	0.01	B*350101	0.05	DRB1*070101	0.14
A*2902	0.07	B*3502	0.07	DRB1*0704	0
A*3001	0.02	B*3503	0.02	DRB1*080101	0.05
A*3002	0.02	B*3504	0	DRB1*080302	0.02
A*310102	0.04	B*3701	0.02	DRB1*080401	0
A*3201	0.04	B*3801	0	DRB1*090102	0.02
A*3301	0.03	B*390101	0	DRB1*1001	0
A*3303	0	B*390602	0.01	DRB1*110101	0.06
A*6601	0.02	B*400101	0.02	DRB1*1102	0.02
A*680101	0.05	B*4002	0.01	DRB1*110401	0.03
A*680102	0.01	B*4102	0	DRB1*1113	0
A*6802	0.01	B*4201	0.01	DRB1*120101	0
A*6901	0.01	B*440201	0.05	DRB1*130101	0.10
		B*440301	0.11	130201	0.03
		B*4501	0.05	DRB1*1303	0.02
		B*4701	0.01	DRB1*1304	0.01
		B*4901	0.01	DRB1*1305	0
		B*5001	0	DRB1*140101	0.02
		B*5002	0	DRB1*150101	0.10
		B*510101	0.12	DRB1*150201	0.01
		B*520101	0.01	DRB1*160101	0.02
		B*5301	0.01	DRB1*160202	0.01
		B*5501	0.01		
		B*5601	0		
		B*570101	0.03		
		B*570301	0		
		B*5801	0.02		
		B*670102	0		
		B*8101	0		

Table 11. HLA-A, -B and -DRB1 allele frequencies in Centre Portugal.

North Portugal allele frequencies					
HLA-A	Frequencies	HLA-B	Frequencies	HLA-DRB1	Frequencies
A*010101	0.07609	B*070201	0.03261	DRB1*0101	0.06522
A*020101	0.29348	B*0705/06	0	DRB1*010201	0.06522
A*0202	0.01087	B*0801	0.06522	DRB1*0103	0.03261
A*0205	0.01087	B*1302	0.01087	DRB1*030101	0.11957
A*022001	0	B*1401	0	DRB1*030201	0
A*030101	0.06522	B*1402	0.06522	DRB1*030501	0
A*0302	0	B*15010101	0.01087	DRB1*0315	0.01087
A*110101	0.05435	B*1503	0.01087	DRB1*040101	0.06522
A*1110	0	B*15170101	0.01087	DRB1*0402	0
A*2301	0.10870	B*1518	0	DRB1*040301	0
A*240201	0.05435	B*180101	0.06522	DRB1*0404	0
A*2501	0.01087	B*1805	0	DRB1*040501	0.01087
A*2601	0.04348	B*2705	0.03261	DRB1*040701	0.02174
A*2608	0.01087	B*2708	0.01087	DRB1*0408	0
A*290101	0	B*350101	0.06522	DRB1*070101	0.17391
A*2902	0.03261	B*3502	0.01087	DRB1*0704	0
A*3001	0.01087	B*3503	0.02174	DRB1*080101	0.03261
A*3002	0.01087	B*3504	0	DRB1*080302	0
A*310102	0.06522	B*3701	0	DRB1*080401	0
A*3201	0.05435	B*3801	0	DRB1*090102	0.01087
A*3301	0.07609	B*390101	0.03261	DRB1*1001	0
A*3303	0.01087	B*390602	0.01087	DRB1*110101	0.05435
A*6601	0	B*400101	0.02174	DRB1*1102	0.02174
A*680101	0	B*4002	0.03261	DRB1*110401	0.01087
A*680102	0	B*4102	0.02174	DRB1*1113	0.01087
A*6802	0	B*4201	0	DRB1*120101	0.01087
A*6901	0	B*440201	0	DRB1*130101	0.05435
		B*440301	0.14130	DRB1*130201	0.05435
		B*4501	0.01087	DRB1*1303	0.01087
		B*4701	0	DRB1*1304	0
		B*4901	0.08696	DRB1*1305	0.01087
		B*5001	0.01087	DRB1*140101	0.03261
		B*5002	0.01087	DRB1*150101	0.10870
		B*510101	0.13043	DRB1*150201	0
		B*520101	0	DRB1*160101	0.01087
		B*5301	0.01087	DRB1*160202	0
		B*5501	0.01087		
		B*5601	0		
		B*570101	0		
		B*570301	0		
		B*5801	0.02174		
		B*670102	0		
		B*8101	0		

Table 12. HLA-A, -B and -DRB1 allele frequencies in North Portugal.

4.4.1. Data for HLA-A

At HLA-A *locus* 10 alleles (37% of all alleles found) comprise more than 80% of observed cumulative frequencies. The most frequent of the 27 HLA-A alleles found in the Portuguese is by far A*020101 found in 26% of the samples, a common characteristic among many populations especially Caucasians. Four alleles follow with frequencies ranging between 7-9% (A*010101, A*030101, A*110101 and A*240201). HLA-A*010101 and A*030101 appears in Portugal at lower frequencies than other European populations but higher than sub-Saharan. HLA-A*010101 and A*030101 allele frequencies in Portugal is higher and lower to North African populations respectively. HLA-A*110101 in Portugal have similar frequencies to other European populations but its distribution is not homogeneous throughout the country (SP: 13%, CP and NP: 5%). The SP A*110101 allele frequency is only similar to Oman (Middleton *et al.*, 2003) and Baloch of Iran (Farjadian *et al.*, 2004). HLA-A*240201 frequency on Portugal is similar to Europeans and North Africans. This allele presents a decreasing gradient from SP (10%) to NP (5%).

HLA-A*2301, with 6% in Portugal, has an intermediate frequency between Europeans (1-2%) and Africans (8-23%). This allele has a uneven distribution along the Portuguese territory with the highest value in NP (11%) and lowest in CP (1%), with 6% in SP.

The most diverse HLA-A groups were A*02 with four different alleles (HLA-A*020101, A*0202, A*0205, and A*022001), in which A*0202 is considered to be a specific African allele (Cao *et al.*, 2004), and A*68 with three (HLA-A*680101, A*680102, and A*6802). HLA-A*020101 allele

account's for 94% of its group showing no balanced frequencies. The frequencies on HLA-A*68 have a much equilibrated distribution with A*680101 accounting for 58.5%, A*6802 for 24.4% and A*680102 for 17.1% of this entire group, a typical European distribution contrasting with sub-Saharanans where A*6802 is more prevalent than A*6801. Other polymorphic groups have no balanced frequencies except HLA-A*30 (A*3001 with 27.3% and A*3002 with 72.7%), a group with higher frequencies in Africans than Europeans.

4.4.2. Data for HLA-B

Fifteen of the 43 observed HLA-B alleles (35% of all alleles found) are responsible for more than 80% of cumulative frequencies. HLA-B*440301 and B*510101 alleles were the most frequent. HLA-B*440301 (12%) appear at similar frequencies in Morocco (12.8%), Italy (10%) and South African Zulu (10.5%) but has a lower prevalence in other Europeans (France 5% and Northern Ireland 7%) and Africans (Tunisia 8%, Cameroon 7% and Zambia 2%). HLA-B*510101 (12%) have higher predominance in Northern Italy (23%) and Bulgaria (21%) and is also present to a lower extent in other Europeans (France 6%; Northern Ireland 3%), North Africans (Morocco 4%; Tunisia 8%) and sub-Saharanans (Cameroon 2%; Zambia 6%). At this *locus* alleles HLA-B*180101 and B*350101 have each a frequency of 8% and all the others appear with much lower or negligible values. The allele HLA-B*180101 can be found in similar frequencies in Tunisia, Bulgaria and Czech Republic, in lower frequencies in Europeans (3%-4%) and Africans (2%-3%) and at double the frequency in Italy. HLA-B*350101 frequency in Portugal is higher than several Europeans and

African populations except Italy (13%) and Mali (13%), the same frequency found in SP.

HLA-B*070201, with 4% in Portugal, has a lower frequency than Europeans (7-17%) and sub-Saharan (5-6%) and higher than North Africans (2%), this last frequency much more similar to SP (1%). HLA-B*0801 differs from above in North Africa where it shows higher frequencies than in Portugal. HLA-B*440201 (5%), appearing generally at lower values than other European populations and higher than Africans, shows a strong decreasing gradient South-North in Portugal (SP: 9.2%, CP: 5% and NP: 0%). HLA-B*4901 (5%) is similar to some African populations (Morocco and Cameroon) and higher than Europeans (2-3%). HLA-B*4501 (2%) is higher than in Europeans, except French, and lower than Africans. HLA-B*5002 (0.7%) is a North African Berber allele (4%) that have been found in residual frequencies in USA Hispanics but is absent in CP. The Portuguese present the world highest prevalence of HLA-B*3502 allele (5%) in our knowledge reported to date, the previous highest being 2% in Bulgaria (Middleton *et al.*, 2003). Allele HLA-B*2708 appeared in just one individual, is very rare and was found before in Azoreans and British (Blanco-Gelaz *et al.*, 2001).

The highest level of heterogeneity in HLA-B was observed in B*15 and B*35 groups, each one with four different alleles (percentage of each allele in the group: HLA-B*150101; 73%, B*1503; 7%, B*151701; 13%, and B*1518; 7% and HLA-B*350101; 54.5%, B*3502; 31.5%, B*3503; 11%, B*3504: 3%), showing some equilibrated frequencies specially between B*350101 and B*3502. The others polymorphic groups have only one

highly dominate allele except HLA-B*40 (B*4001 55% and B*4002 45%) and HLA-B*44 (B*440201 28% and B*440301 72%).

4.4.3. Data for HLA-DRB1

At HLA-DRB1 *locus* 14 alleles (39% of all alleles found) comprise more than 80% of observed cumulative frequencies. Alleles HLA-DRB1*070101 and DRB1*030101 are the most common (15 and 13%), with frequencies typical of Mediterranean and other European populations. Alleles HLA-DRB1*110101, DRB1*130101 and DRB1*150101 appear in 6-8% of the samples, with similar frequencies found in European and African (DRB1*110101), European (DRB1*130101) and Mediterranean (DRB1*150101) populations. HLA-DRB1*0101, with 5% in Portugal, has an intermediate frequency between Europeans (8-10%) and Africans (0-2%). HLA-DRB1*040101 and *130201 (5% each), almost absent and with higher frequencies in Africa respectively, have similar frequencies to neighbouring European populations. HLA-DRB1*040501 (3%) has intermediate frequency between European (0-2%) and North Africans (3-6%). HLA-DRB1*110401 (3%) is similar to North Africa and lower than neighbouring European populations (5-7%).

The most diverse HLA-DRB1 groups were DRB1*04 with 7 different alleles (percentage of each allele in the group: DRB1*040101: 38%, DRB1*0402: 11%, DRB1*040301: 7%, DRB1*0404 5%, DRB1*040501 23%, DRB1*040701: 13%, and DRB1**0408: 3%), and HLA-DRB1*13 with 5 alleles (percentage of each allele in the group: DRB1*130101: 50%, DRB1*130201: 29%, DRB1*1303: 15%, DRB1*1304: 3%, and DRB1*1305: 3%), both groups showing balanced frequencies between

some alleles. Other polymorphic groups have no balanced frequencies except HLA-DRB1*01 and DRB1*11. HLA-DRB1*010201 (4%) has similar frequency to Mediterranean and DRB1*0103 (2%), an allele absent in Africa and vestigial in Mediterranean Europe, is similar to frequency in Murcia (South Spain) (Muro *et al.*, 2001). HLA-DRB1*1102 (2%) has a similar frequency to Murcia and North Africa, and HLA-*110401 (3%) is lower than other Europeans (5-7%) and similar to North Africa. HLA-DRB1*0101/DRB1*010201 ratio is more equilibrated in Portugal (as well as other Iberia populations) in a intermediate position between European and African populations. HLA-DRB1*110101/DRB1*110401 ratio in Portugal is similar to other Mediterranean except some Spain populations (Basques and Murcia) where DRB1*110401 is more prevalent than DRB1*110101.

4.4.4. Haplotype frequencies

The exact test of *linkage disequilibrium* between the 3 pairs of *loci* was statistically significant for the pairs HLA-A,-B ($P < 0.0001$) and HLA-B,-DRB1 ($P < 0.0001$). The most representative 2-*loci* (frequency 2%) and 3-*loci* (frequency 1%) haplotypes are listed in Tables 13-19. The calculation of LD, RLD, and χ^2 excluded some non-significant combination among the most frequent two-*loci* haplotypic combinations.

A*020101-B*440301-DRB1*070101 (3.1%), the most frequent three-*loci* haplotype, is highly frequent in NP (5.4%), has a 2% prevalence in CP and is absent in SP. It could be included in A*02-B*44-DRB1*0701 which is considered a North-Western European haplotype (Sanchez-Velasco *et al.*, 2003) and is present in Spaniards at 2%. Tunisia presents the same high resolution haplotype at the same frequency as Portugal (Ayed *et al.*, 2004).

Haplotypes	Frequency	s.d.	D	D'	P
A*010101-B*0801	0.027	0.007	0.024	0.644	**
A*020101-B*440201	0.024	0.005	0.012	0.308	*
A*020101-B*510101	0.068	0.011	0.037	1	**
A*030101-B*070201	0.010	0.006	0.007	0.197	*
A*030101-B*350101	0.026	0.016	0.020	0.548	**
A*030101-B*4901	0.010	0.006	0.007	0.183	*
A*110101-B*350101	0.027	0.009	0.021	0.552	**
A*2301-B*4901	0.014	0.004	0.011	0.29	**
A*240201-B*070201	0.010	0.004	0.007	0.197	*
A*2902-B*440301	0.035	0.009	0.027	0.735	**
A*3002-B*180101	0.020	0.010	0.018	0.486	**
A*3201-B*4901	0.010	0.004	0.008	0.223	**
A*3301-B*1402	0.023	0.008	0.021	0.564	**
A*680101-B*4501	0.014	0.003	0.013	0.355	**

Table 13. Most common HLA-A, -B two-*loci* haplotypes in the Portuguese population as estimated by maximum likelihood (only frequencies above 0.01 with statistical significance *linkage disequilibrium* are shown). *P<0.05, ** P<0.001.

Haplotypes	Frequency	s.d.	D	D'	P
A*010101-DRB1*070101	0.037	0.010	0.024	1	**
A*020101-DRB1*080101	0.021	0.011	0.012	0.498	*
A*020101-DRB1*130101	0.037	0.012	0.016	0.698	*
A*030101-DRB1*0101	0.017	0.005	0.014	0.583	**
A*030101-DRB1*040101	0.010	0.007	0.007	0.289	*
A*110101-DRB1*140101	0.010	0.005	0.009	0.37	**
A*2301-DRB1*070101	0.028	0.011	0.019	0.796	**
A*2301-DRB1*130201	0.017	0.010	0.015	0.621	**
A*2902-DRB1*130101	0.013	0.006	0.009	0.366	*
A*3002-DRB1*030101	0.015	0.005	0.011	0.481	**
A*3201-DRB1*030101	0.017	0.008	0.012	0.498	**
A*3301-DRB1*010201	0.024	0.011	0.022	0.94	**
A*680101-DRB1*040501	0.010	0.006	0.010	0.413	**

Table 14. Most common HLA-A, -DRB1 two-*loci* haplotypes in the Portuguese population as estimated by maximum likelihood (only frequencies above 0.01 with statistical significance *linkage disequilibrium* are shown). *P<0.05, ** P<0.001.

Haplotypes	Frequency	s.d.	D	D'	P
B*070201-DRB1*150101	0.021	0.006	0.017	0.25	**
B*0801-DRB1*030101	0.033	0.011	0.027	0.388	**
B*1402-DRB1*010201	0.028	0.008	0.025	0.364	**
B*1402-DRB1*1303	0.010	0.005	0.009	0.131	**
B*180101-DRB1*030101	0.044	0.009	0.032	0.459	**
B*180101-DRB1*110401	0.010	0.004	0.007	0.108	**
B*350101-DRB1*0101	0.034	0.009	0.031	0.433	**
B*3502-DRB1*110401	0.014	0.009	0.012	0.173	**
B*440201-DRB1*130101	0.020	0.010	0.016	0.228	**
B*440301-DRB1*070101	0.089	0.009	0.071	1	**
B*4501-DRB1*040501	0.010	0.005	0.010	0.136	**
B*4901-DRB1*040101	0.010	0.006	0.008	0.117	**
B*4901-DRB1*130201	0.017	0.005	0.015	0.214	**
B*510101-DRB1*080101	0.024	0.009	0.02	0.281	**
B*510101-DRB1*110101	0.021	0.008	0.013	0.18	**
B*570101-DRB1*070101	0.014	0.009	0.011	0.151	**

Table 15. Most common HLA-B, -DRB1 two-*loci* haplotypes in the Portuguese population as estimated by maximum likelihood (only frequencies above 0.01 with statistical significance *linkage disequilibrium* are shown). *P<0.05, ** P<0.001.

Portugal haplotypes ≥ 0.01	Frequency
A*010101-B*0801-DRB1*030101	0.024
A*010101-B*440301-DRB1*070101	0.024
A*020101-B*070201-DRB1*150101	0.010
A*020101-B*180101-DRB1*030101	0.021
A*020101-B*440301-DRB1*070101	0.031
A*020101-B*510101-DRB1*040101	0.010
A*020101-B*510101-DRB1*110101	0.021
A*020101-B*510101-DRB1*130101	0.014
A*030101-B*350101-DRB1*0101	0.017
A*110101-B*440201-DRB1*130101	0.010
A*2301-B*440301-DRB1*070101	0.010
A*2301-B*4901-DRB1*130201	0.017
A*2902-B*440301-DRB1*070101	0.017
A*3002-B*180101-DRB1*030101	0.017
A*3301-B*1402-DRB1*010201	0.021
A*680101-B*4501-DRB1*040501	0.010

Table 16. Most common (≥ 0.01) HLA-A, -B, -DRB1 three-*loci* haplotypes in the Portuguese population as estimated by maximum likelihood.

SP haplotypes ≥ 0.02	Frequency
A*010101-B*440301-DRB1*070101	0.02
A*020101-B*1402-DRB1*010201	0.02
A*020101-B*180101-DRB1*030101	0.041
A*020101-B*440201-DRB1*0402	0.02
A*020101-B*510101-DRB1*040101	0.02
A*030101-B*350101-DRB1*0101	0.02
A*030101-B*4901-DRB1*040101	0.02
A*110101-B*180101-DRB1*110401	0.02
A*110101-B*350101-DRB1*0101	0.02
A*110101-B*350101-DRB1*110101	0.02
A*110101-B*440201-DRB1*130101	0.02
A*2301-B*0801-DRB1*030101	0.02
A*2301-B*440301-DRB1*1303	0.02
A*240201-B*510101-DRB1*070101	0.031
A*2902-B*440301-DRB1*070101	0.02
A*3002-B*180101-DRB1*030101	0.041

Table 17. Most common (≥ 0.02) HLA-A, -B, -DRB1 three-*loci* haplotypes in South Portugal as estimated by maximum likelihood.

CP haplotypes ≥ 0.02	Frequency
A*010101-B*0801-DRB1*030101	0.03
A*010101-B*180101-DRB1*110101	0.02
A*010101-B*440301-DRB1*070101	0.045
A*020101-B*070201-DRB1*150101	0.03
A*020101-B*15010101-DRB1*130101	0.02
A*020101-B*180101-DRB1*030101	0.02
A*020101-B*1805-DRB1*080302	0.02
A*020101-B*440301-DRB1*070101	0.02
A*020101-B*510101-DRB1*080101	0.03
A*310102-B*510101-DRB1*130101	0.02
A*680101-B*4501-DRB1*040501	0.03

Table 18. Most common (≥ 0.02) HLA-A, -B, -DRB1 three-*loci* haplotypes in Centre Portugal as estimated by maximum likelihood.

NP haplotypes ≥ 0.02	Frequency
A*010101-B*0801-DRB1*030101	0.022
A*020101-B*350101-DRB1*0101	0.022
A*020101-B*440301-DRB1*070101	0.054
A*020101-B*510101-DRB1*110101	0.033
A*030101-B*350101-DRB1*0101	0.022
A*110101-B*0801-DRB1*030101	0.022
A*2301-B*4901-DRB1*130201	0.044
A*240201-B*180101-DRB1*140101	0.022
A*2902-B*440301-DRB1*070101	0.022
A*3201-B*0801-DRB1*030101	0.022

Table 19. Most common (≥ 0.02) HLA-A, -B, -DRB1 three-*loci* haplotypes in North Portugal as estimated by maximum likelihood.

Three related haplotypes were found in the Portuguese: A*010101-B*440301-DRB1*070101 (2.4%), A*2902-B*440301-DRB1*070101 (1.7%) and A*2301-B*440301-DRB1*070101 (1%). The first one, which is most frequent in CP (4.5%), appear at 1% in SP and is absent in NP, has not yet been described in other populations. The HLA-A*2902 haplotype appears in Tunisia (4%) and Northern Ireland (2.1%) (Ayed *et al.*, 2004). Its corresponding low resolution haplotype is considered by Sanchez-Velasco *et al.* (2003) as Western European and is present in Spaniards at 5%. The HLA-A*2301 haplotype has not yet been described in other populations.

A*010101-B*0801-DRB1*030101 (2.4%), whose corresponding low resolution haplotype is considered Indo-European-Celtic origin (Arnaiz-Villena *et al.*, 2001a), is present in CP (3%) and NP (2.2%), and was found in a high frequency (9%) in Northern Ireland (Middleton *et al.*, 2003).

A*3301-B*1402-DRB1*010201 (2.1%), only present in NP (4.3%), was also found in Tunisians at 4% (Ayed *et al.*, 2004). Gómez-Casado *et al.*

(2000) define A*33-B*14-DRB1*0102 as a Mediterranean haplotype and in fact it is present in several populations in this region (Imanishi *et al.*, 1992).

A*3002-B*180101-DR*030101 (1.7%) is exclusively from SP (4.1%), being present in Tunisians (2%) (Ayed *et al.*, 2004). Its low resolution typing was considered by Sanchez-Velasco *et al.* (2003) as an Iberian/paleo-North African haplotype. The related A*020101-B*180101-DRB1*030101 (2.1%) is present in SP (4.1%) and CP (2%), was also described in Iran (1.7%) (Farjadian *et al.*, 2004) and in its low resolution form, in Sardinia Italy (2.2%) (Rendine *et al.*, 1998).

A*020101-B*070201-DRB1*150101 (1%), classified as European-North African (Arnaiz-Villena *et al.*, 1999), was only found in the CP (3%), is present in Northern Ireland population in a frequency of 3% (Middleton *et al.*, 2003).

Other HLA haplotypes, with unknown origin, have also a differential distribution along Portugal as for example A*2301-B*4901-DRB1*130201 which is exclusive found in NP (4.4%), A*680101-B*4501-DRB1*040501 from CP (3%), and A*240201-B*510101-DRB1*070101 found only in SP (3.1%).

Regarding HLA-A, -B haplotypes, A*020101-B*510101 (6.8%) deserve a special attention. This 2 *loci* haplotype has a similar prevalence in SP, CP and NP and is associated in the Portuguese populations with DRB1*110101 (2.1%), DRB1*130101 (1.4%) and DRB1*040101 (1%). Other world populations show either a lower prevalence of this bi-*loci* haplotype (Cuban, Northern Ireland, Kenya, Zambia, Uganda and Taiwan with 1-2%), or higher frequency (Oman 5.6% and Tunisia 4.5%) (Middleton *et al.*, 2003; Ayed *et al.*, 2004) which could point to a possible Arab influence in Portugal.

4.4.5. Phylogenetic analyses

A phylogenetic tree constructed with both class I (HLA-A and -B) and Class II (HLA-DRB1) (Figure 5), clusters Portugal to North African populations. In contrast, a dendrogram based on Class I *loci* only (Figure 6) grasps Portugal with Europeans. The apparent contradiction of these two analyses reflects the intermediate position occupied by Portugal between North Africans and Europeans. A phylogenetic tree based on HLA-A, -B and -DRB1 data but with the Portuguese population sub divided into their 3 main sub-populations (Figure 7) unveils the reason for the discrepancy observed. In fact, NP and SP cluster together with Africans and CP with Europeans. A principal coordinate analysis (Figure 8) is consistent with the dendrograms and shows the Portuguese as intermediate between Europeans and Africans.

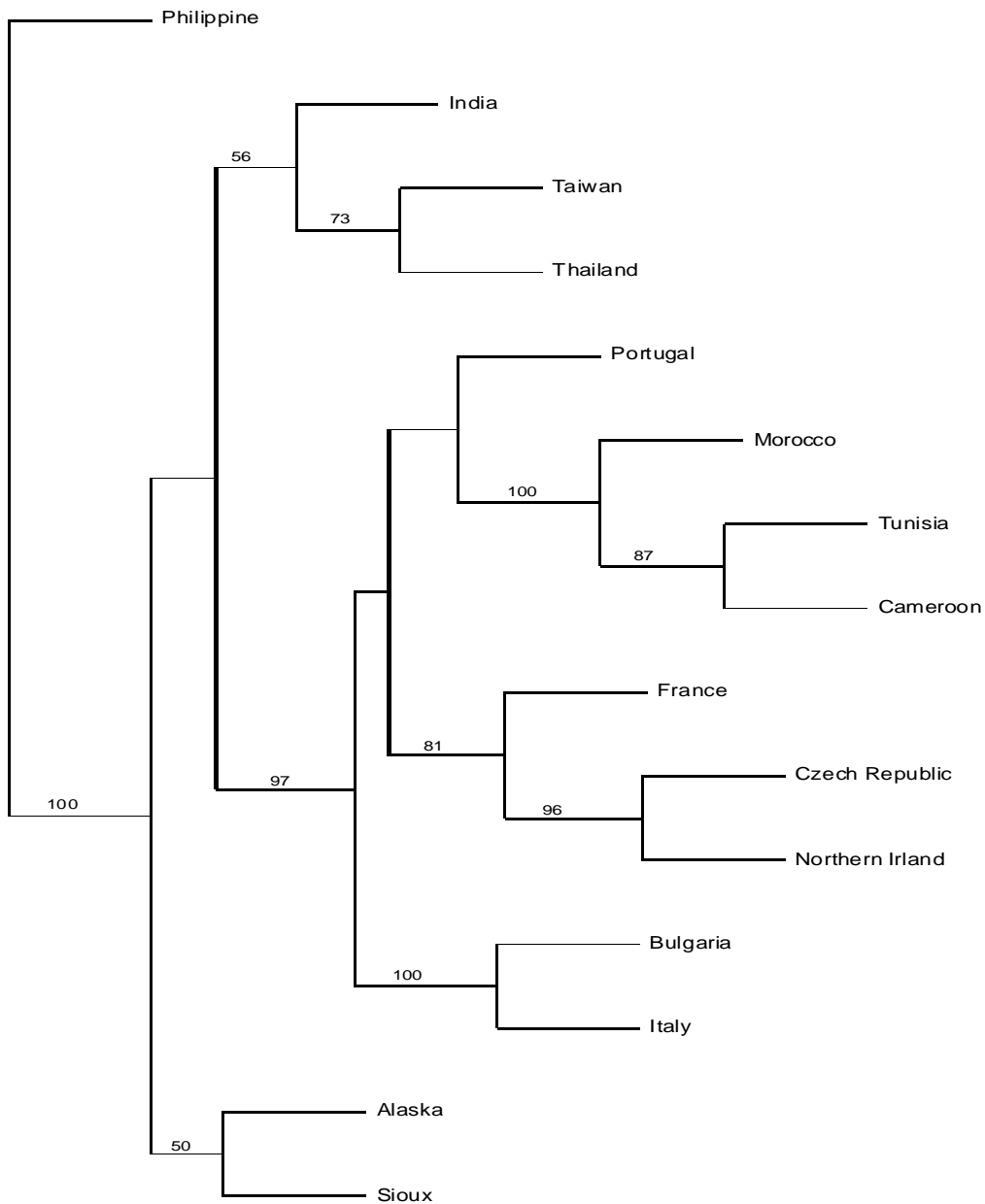


Figure 5. Neighbor-joining (NJ) dendrograms showing the comparative position of the Portuguese population with other populations typed with similar resolution. Standard genetic distances among populations were calculated using HLA-A, -B and -DRB1 allele frequencies. Numbers above branches are node support after the bootstrap technique implemented in PHYLIP package program BOOT. For references of the populations used see *data analysis*.

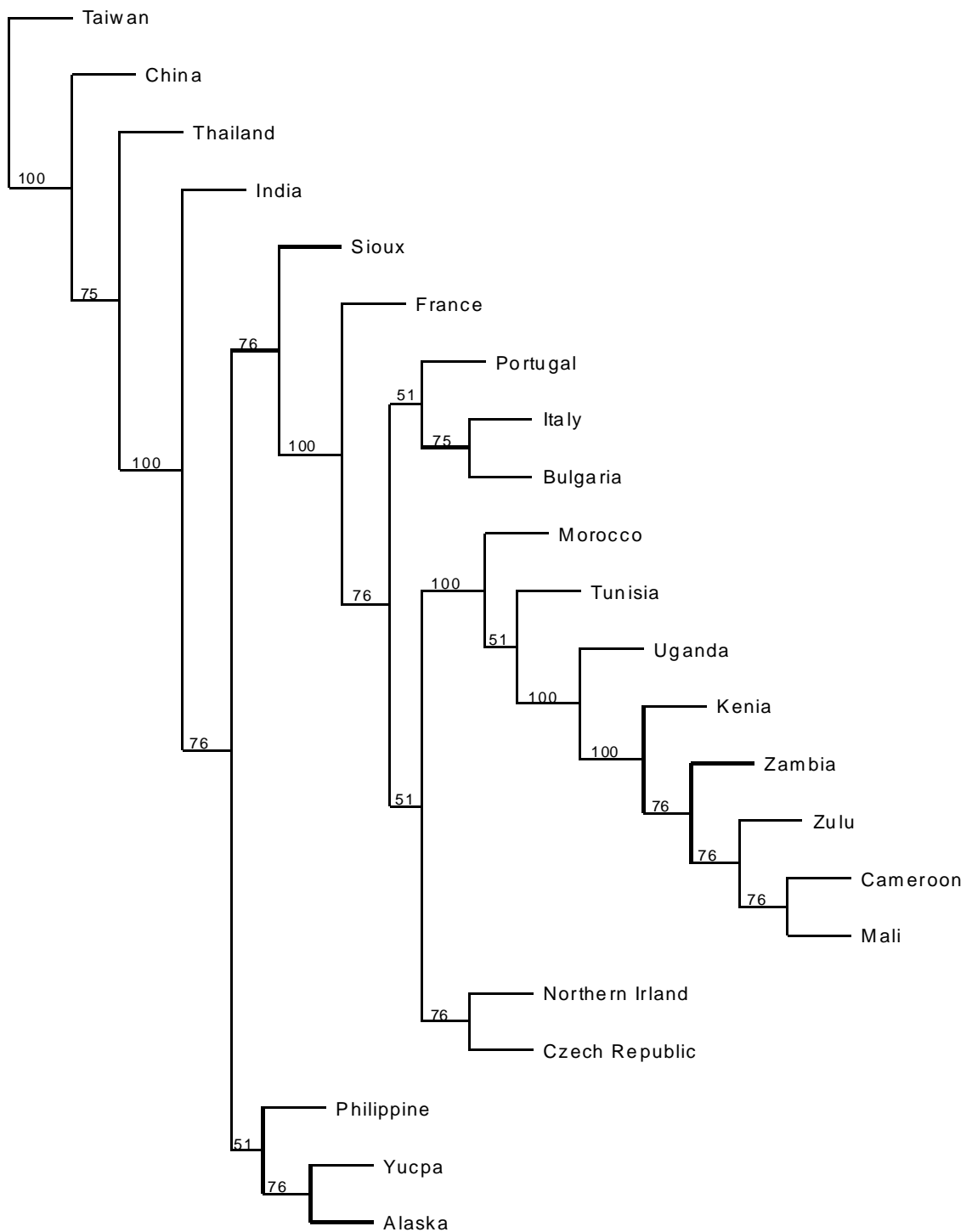


Figure 6. Neighbor-joining (NJ) dendrograms showing the comparative position of the Portuguese population with other populations typed with similar resolution. Standard genetic distances among populations were calculated using HLA-A and -B allele frequencies. Numbers above branches are node support after the bootstrap technique implemented in PHYLIP package program BOOT. For references of the populations used see *data analysis*.

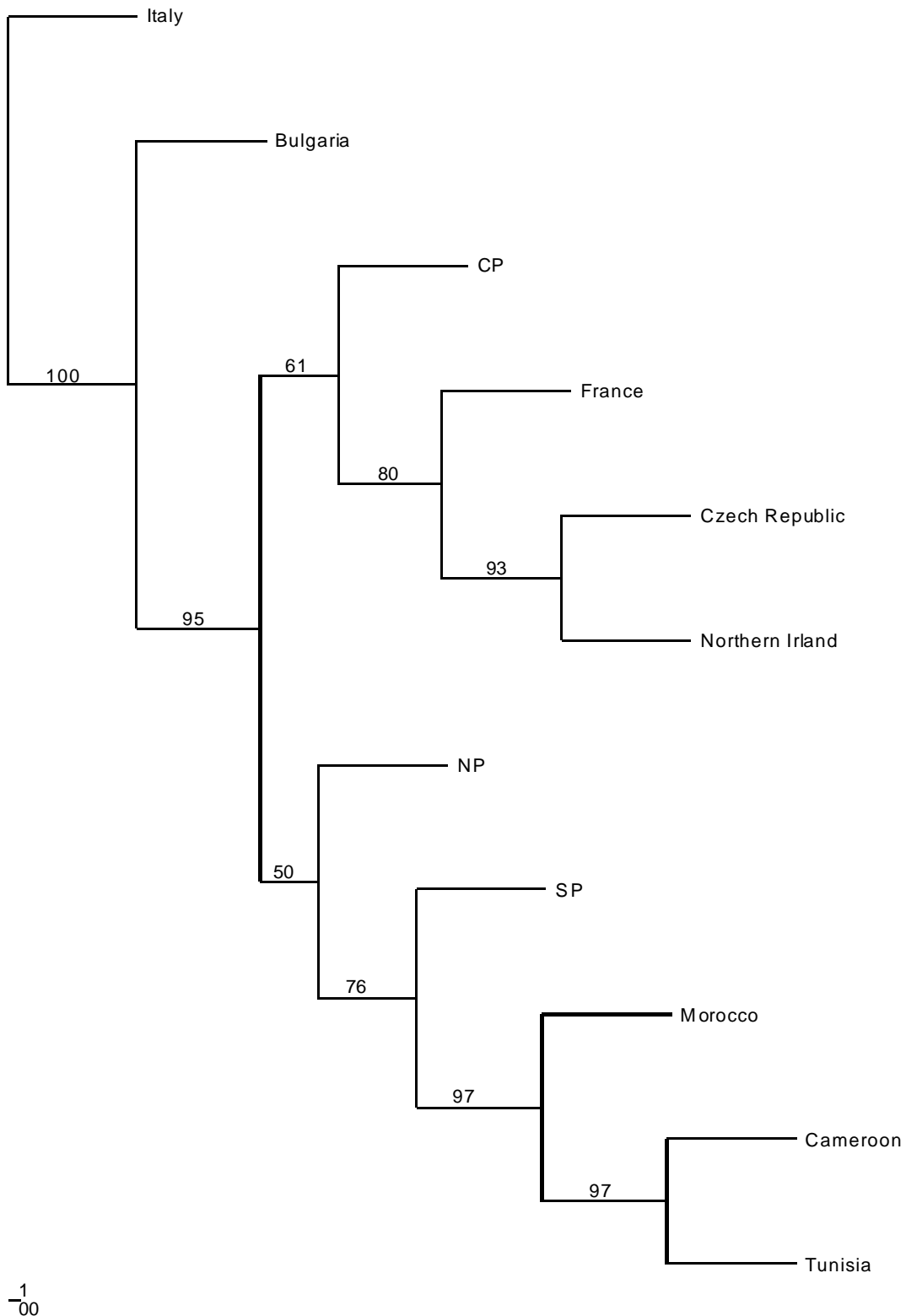


Figure 7. Neighbor-joining (NJ) dendrograms showing the comparative position of the Portuguese population with other populations typed with similar resolution. Standard genetic distances among populations were calculated using HLA-A, -B and -DRB1 allele frequencies on NP, CP and SP. Numbers above branches are node support after the bootstrap technique implemented in PHYLIP package program BOOT. For references of the populations used see *data analysis*.

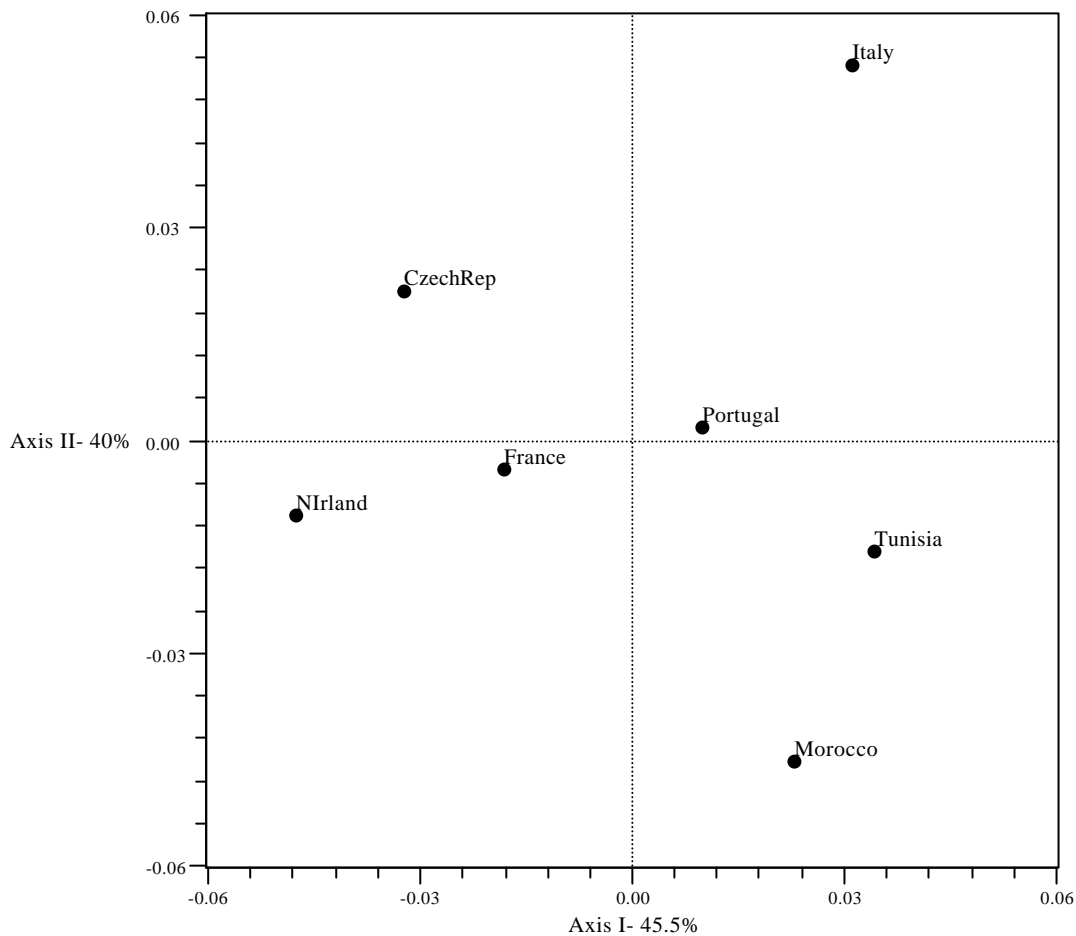


Figure 8. Principal coordinate analysis for 7 European and North African populations constructed with conventional F -statistics calculated from HLA-A, -B, and -DRB1 allele frequencies.

4.5. Discussion

Portugal has experienced the influence of various peoples along pre-historic and historic times. Arnaiz-Villena *et al.* (1999) and Gonzalez *et al.* (2003) postulated that around 6.000 years ago, Iberia could have been the receptacle for a substantial input of Berber migration when the Sahara experienced profound climate changes. Before Roman presence in Iberia (100 B.C.) several European and Mediterranean peoples are known to have settled in Iberia: Greeks, Phoenicians and Celts. After the fall of the Roman Empire, Germanics and Berbers from North Africa also settled in what is

known today as Portuguese territory (Arnaiz-Villena *et al.*, 1997). After the 15th century, a substantial amount of sub-Saharaners entered Portugal via the slave trade (Russell-Wood, 1998).

The present work describes for the first time HLA-A, -B and -DRB1 high resolution allele frequencies and estimated haplotypes in North, Centre and South Portuguese samples. This high resolution throughput shows that several HLA generic groups identified before (Spínola *et al.*, 2002) were in fact constituted by a variety of alleles (Table 9). This additional information is highly relevant for a detailed HLA structure characterization. Unfortunately, few world populations have been well characterized to HLA allele level, a situation that renders difficult the comparison of the available data sets.

The Portuguese position between Europe and North Africa, shown by dendrograms and Principal Coordinate Analysis (Figures 5-8), is also corroborated by particular allele intermediate frequencies (e.g. HLA-A*010101, A*030101, A*2301, B*0702, B*4901, B*4501, B*5002, DRB1*0101, DRB1*040501). Other allele frequencies show similar values to North African populations (e.g. HLA-A*3301, A*6601, B*440301, B*4901, DRB1*1102 and DRB1*110401) and Europeans (e.g. HLA-A*020101, A*110101, A*2601, A*2902, A*3101, A*3201, B*1402, B*1501, B*2705, B*3503, B*5802, DRB1*130101, DRB1*040101 and DRB1*130201). Several allele frequencies and haplotypes show no homogenous distribution along Portugal, probably reflecting different genetic influences. For example, haplotype A*010101-B*0801-DRB1*030101, considered of Celtic origin (Muro *et al.*, 2001), and A*020101-B*440301-DRB1*070101, North-Western European (Sanchez-Velasco *et al.*, 2003), is only present in the

Centre and North Portugal revealing a possible influence of Celts or Barbarian peoples. Consistently with the presence of these two haplotypes, B*0801 and A*020101 presents a North-South decreasing allele frequency in Portugal.

The Mediterranean haplotype A*3301-B*1402-DRB1*010201 (Gómez-Casado *et al.*, 2000) is exclusive to NP (4.3%), despite appearing in SP (1%) with B*140101, in agreement with the findings of Arnaiz-Villena and colleagues (1997) which didn't find the same haplotype in a population from the Centre Portugal. B*1402 shows a clearly North-South decreasing cline, but B*140101 is only present in SP (2%). A*020101-B*070201-DRB1*150101 (European-North African), A*010101-B*5801-DRB1*070101 (Basque-Berber), and A*020101-B*070201-DRB1*150101 (North African-Western European) were only found in CP. Otherwise, A*020101-B*510101-DRB1*130101 (1%) (Ibero-Berber), A*020101-B*180101-DRB1*1104 (1%) (Mediterranean), A*3001-B*1302-DRB1*070101 (1%) (Cretan and Middle East) and others autochthonous haplotypes (A*2301-B*4901-DRB1*130201: 4.3%, and A*020101-B*440201-DRB1*150101: 1%) were only found in NP. This data suggests that important ancient Mediterranean and North African markers, entered Portugal not only during pre-history but also in historic ages. The former were forced to migrate northward, due to the Arab expansion in 800 A.D via the Gibraltar Strait, and for this reason the population of NP shows distinctive features that could be the footprint of a more ancient genetic pool. This hypothesis is much more consistent when we confirm that SP do not present any well known haplotypes, except the Iberian Paleo-North African A*3002-B*180101-DRB1*030101(4%), absent in CP and NP, and A*2902-

B*440301-DRB1*070101, an ancient Western European haplotype (also present in North African), which was the only haplotype present in all the three NP, CP and SP sub-populations. These two haplotypes could be brought to SP by North African Berbers newcomers integrated with the Arab settlers. Apart from above, A*020101-B*180101-DRB1*030101 (SP 4% and CP 2%), identified before in Sardinians (Rendine *et al.*, 1998) and Iran (Farjadian *et al.*, 2004), is the only three *loci* SP haplotype common to CP. A*020101-B*3502-DRB1*110101, whose low resolution correspondent haplotype was found in Algerians (Arnaiz-Villena *et al.*, 1997), is present in SP at 1%.

Several alleles from SP and CP are absent in NP but all NP alleles are present in either one of the two other sub-populations (Tables 10-12). These differences in North-South allele frequencies reinforce a northward migration, because under hypothesis we expect in the North of Portugal a lost of some alleles and an increase in some frequencies because of founder effect. Surprisingly, in opposition to the above North-South clines, CP present several alleles with lower frequencies than NP and SP (e.g. HLA-A*2301, A*2601, B*350101, B*4901, B*5001, DRB1*0101, DRB1*0102, DRB1*0103, DRB1*040101 and DRB1*130201). This decreasing gradient from North and South to Centre of Portugal could reflect a bipolar distribution of ancient Portuguese (NP) and recent Berber movements (SP). In the dendrograms, CP systematically joins other European populations rather than NP and SP (or African populations), most probably due to the absence of common markers (Figure 7).

The input of sub-Saharan people in Portugal through the slave trade is difficult to detect. Most of sub-Saharan populations, especially West

Africans, do not present HLA Class I and Class II high resolution data, which renders difficult the identification of these markers in Portugal. Nevertheless two-*loci* haplotypes spread among sub-Saharan populations (Ellis *et al.*, 2000; Middleton *et al.*, 2003; Cao *et al.*, 2004) were found in Portugal (A*030101-B*4201, A*2301-B*0801, A*2301-B*440301, A*2301-B*5801, A*3001-B*5301 and A*3002-B*1402) attesting to a minute but visible genetic influence left by sub-Saharans.

Capítulo 5

5. HLA class I and II polymorphisms in Azores show different settlements in Oriental and Central islands.

Hélder Spínola, António Brehm, Bruno Bettencourt, Derek Middleton, Jácome Bruges-Armas.

Reprinted with minor modifications from Tissue Antigens 2005: 66: 217-230.

5.1. Abstract

HLA-A, -B, -Cw, -DRB1, -DQA1, and -DQB1 polymorphisms were examined in the Azorean population. The data was obtained at high-resolution level, using PCR-SSP, PCR-SSO, and SBT. The most frequent allele in each *locus* was: A*0201 (24.5%), B*510101 (9.8%), Cw*0401 (14.8%), DRB1*070101 (18.3%), DQA1*0201 (17.4%), and DQB1*0301 (19.4%). The predominant extended haplotype was A*0202-B*1503-Cw*0202-DRB1*090102-DQA1*0303-DQB1*0202 (1.9%), which was found to be absent in the Portuguese mainland.

The present study corroborates historical sources that say the Azores were populated not only by Portuguese but also by other Europeans, mostly Flemish people. Despite dendrogram analysis showing some remote Asian genetic affinities, the lack of specific alleles and haplotypes from those populations does not allow us to conclude for direct influence.

Haplotype and allele frequencies in Azores show no homogeneous distribution between Oriental and Central islands of this archipelago. The Oriental islands harbour several haplotypes already found in mainland Portugal and identified as Mediterranean and European. The Central group of islands on the contrary clearly shows an influence of central Europeans (most probably derived from a well documented Flemish settlement), with much less affinity to mainland Portugal.

5.2. Introduction

The highly polymorphic Human Leukocyte Antigen (HLA) System, located in the short arm of chromosome 6, consists of a closely linked set of genes important in transplantation, disease and anthropological studies (Anjos & Polychronakos, 2004). HLA allele frequencies and haplotypic patterns have great importance as markers to determine genetic relatedness and the degree of admixture between populations (Arnaiz-Villena *et al.*, 2002).

The Azores Islands (Portugal), located in the middle of the North Atlantic Ocean, were found to be uninhabited by Portuguese in the first half of the 15th century and were officially populated in 1439. The archipelago is constituted by nine islands subdivided into three groups (Oriental: Santa Maria and São Miguel; Central: Terceira, Graciosa, São Jorge, Pico and Faial; and Occidental: Flores and Corvo). During the 15th and 16th

centuries, the islands played a significant role in sea traffic between America and Europe (Monod, 1991) being important for commerce and slave trade between Africa, America and Europe. This central and strategic position between the three continents, brought to Azores people from different origins, mainly Portuguese but also Jews, Moorish prisoners, sub-Saharan slaves, Dutch (Flemish), French, Italians, English and Spaniards (Guill, 1993). Some historic registries also mention the selling of Caribbean Indians to the Azores. According to several authors, the first settlers of the Oriental and Central groups had different origins. The Oriental group was probably settled first, mostly by people from different regions of mainland Portugal (Bannerman & Mary, 1966; Morison, 1965) while the settlement of Central islands is known to be strongly influenced by Flemish people. By 1490 there were 2000 Flemings living on Central group islands (Rogers, 1979).

Few studies have been done in order to characterize the Azorean population genetic profile (Brehm *et al.*, 2003; Gonçalves *et al.*, 2005) but only one used HLA generic class I and class II typing in Terceira island (Bruges-Armas *et al.*, 1999). With the recent developments in DNA typing, HLA can be studied at a high resolution level, for example by sequencing. This high resolution throughput allows an additional amount of information with great relevance to a more specific HLA characterization, important in population's comparisons and disease studies (Piancatelli *et al.*, 2004; Ayed *et al.*, 2004). Unfortunately high resolution HLA allele typing data is still scarce, making comparisons difficult. Previous HLA generic typing indicates that the Azorean population most likely contains an admixture of high frequency Caucasoid, Mongoloid and, to a lesser degree, sub-Saharan HLA

genes (Bruges-Armas *et al.*, 1999). In the later study it was proposed that a Mongoloid population already existed in the Azores before the Portuguese discovery. Y-chromosome and mitochondrial DNA show major similarities to Portugal mainland and Europe and some genetic influence from sub-Saharan Africans (Fernandes & Brehm, 2003; Brehm *et al.*, 2003) but failed to detect any Asiatic contribution to the present day gene pool.

The main goal of the present work is to present a clear genetic profile of the Azorean based on high resolution typing of HLA-A, -B, -Cw, -DRB1, -DQA1, and -DQB1 *loci*. Allele and haplotype frequencies of this population and their relatedness to Portuguese mainland and other world populations will be evaluated. We expect that these comparisons may also contribute to clarify the origin of Azorean population and especially the Asian genetic influence previous reported (Bruges-Armas *et al.*, 1999).

5.3. Materials and methods

5.3.1. Population samples and HLA typing

The study population consisted of a total of 102 healthy unrelated individuals from the archipelago of Azores. Blood samples were collected after informed consent, from donors whose parents and grandfathers were born and living in the Azores (Az1). In order to analyse internal differences in the archipelago for HLA-A, HLA-B, and HLA-DRB1 *loci* this sample was further subdivided into Azores Oriental group of islands (AzOriental, n=43) and Azores Central group (AzCentral, n=59). The AzOriental comprise Santa Maria and São Miguel islands, and AzCentral is constituted by Terceira, Graciosa, São Jorge, Faial, and Pico islands. Genomic DNA was isolated from

whole blood containing EDTA using a phenol-chloroform procedure and frozen at -20°C until use. All subjects were typed for HLA-A, HLA-B, and HLA-DRB1 by sequence based typing (SBT) accordingly to Kurz *et al.* (1999) and Pozzi *et al.* (1999) for HLA-A and HLA-B *loci*, and amplifying and sequencing exon 2 of HLA-DRB1 *locus* with group specific primers used in SSOP typing (Williams *et al.*, 2004). Typing procedures and allele identification methodology were as previously outlined (Spínola *et al.*, 2005a).

5.3.2. Data analysis

In order to get a bigger sample for comparisons, we added data from Terceira island already published (Az2, n=129) (Armas *et al.*, 2004) and typed by SSP and SSOP for HLA-A, -B, -Cw, -DRB1, -DQA1, and -DQB1. These samples have been collected only with the information that ancestors were from Azores with no specification of the island's origin. Because of that we used Az1 and Az2 samples for a global Azores analysis.

Basic genetic parameters (allele and haplotype frequencies, gene diversity, and Hardy-Weinberg equilibrium) at the six *loci* were estimated with Arlequin v2.000 (Schneider *et al.*, 2000). The same population genetic software was employed to calculate Ewes-Watterson's, Slatkin's and Chakraborty's selective neutrality statistical tests in order to examine the presence of selective forces influencing allelic diversity at each *locus*. *Linkage disequilibrium* (D) and relative LD (D') between two alleles at two different *loci* and their level of significance (P) was calculated according to Weir (1996).

The Azores data was compared to mainland Portugal (Spínola *et al.*, 2005) and several populations which are available at the same typing resolution (Ellis *et al.*, 2000; Muro *et al.*, 2001; Layrisse *et al.*, 2001; Pimtanothai *et al.*, 2001; Tsuneto *et al.*, 2003; Ayed *et al.*, 2004; Middleton *et al.* 2004; Leffell *et al.*, 2004; Uyar *et al.*, 2004)¹. An analysis of molecular variance (AMOVA) was performed with these populations based on Euclidean distances between all pairs of haplotypes (Excoffier *et al.*, 1992). Comparative analysis of our data set with other populations available in the literature was achieved using the software included in the PHYLIP v.3.6 package (Felsenstein, 2004). First, SEQBOOT was used to perform a bootstrap analysis from gene frequency data. The program generates multiple data sets resampled from the original data. Distance matrices from each replicate data set were generated using GENDIST and used as input to NEIGHBOR to produce neighbor-joining trees. A single consensus bootstrapped tree was obtained with CONSENSE. The topology was visualized with TreeView (Page, 1996). Various dendrograms were made with different groups of *loci* depending on the level of HLA resolution typing available for the comparisons. Principal coordinate analysis using HLA-A, -B, and -DRB1 allele frequencies was carried out on the MultiVariate Statistical Package MVSP3 for Windows (Kovach Computing Services, Anglesey, Wales, UK)².

¹ www.allelefrequencies.net (20-01-2005)

² <http://www.kovcomp.com/mvsp> (20-01-2005)

5.4. Results

Tables 20 and 21 show the allele frequencies of HLA-A, HLA-B, HLA-Cw class I *loci* and HLA-DRB1, HLA-DQA1 and HLA-DQB1 class II *loci*, respectively, in the Azores population. Table 22 show allele frequencies of HLA-A, -B, and -DRB1 in Azores Oriental and Azores Central sub-populations. A total of 33 HLA-A, 53 HLA-B, 23 HLA-Cw, 40 HLA-DRB1, 14 HLA-DQA1 and 16 HLA-DQB1 alleles were found. The population is in Hardy-Weinberg equilibrium at each *loci* except HLA-B ($P < 0.02$). Overall, the heterozygosity was high (HLA-A: 0.90; HLA-B: 0.96; HLA-Cw: 0.92; HLA-DRB1: 0.94; HLA-DQA1 and HLA-DQB1: 0.89). Ewes-Watterson, Slatkin's and Chakraborty's tests of selective neutrality yielded non-significant results for HLA-A, -B, and -DRB1 suggesting lack of selection on these three *loci*.

An AMOVA between Azores and mainland Portugal, with HLA-A, -B, and -DRB1 *loci*, showed that only a residual genetic variation found can be attributed to differences among populations, the remaining is due to within population differences. An exact test of population differentiation showed significant results between Azores and mainland Portugal ($P = 0.0057$), but no differences within the Azores islands.

HLA-A n=231	Freq.	HLA-B n=231	Freq.	HLA-C n=129	Freq.
A*0101	0.102	B*0702	0.076	Cw*0102	0.012
A*0201	0.245	B*0705	0.002	Cw*0202	0.078
A*0202	0.022	B*0720	0.002	Cw*0303	0.054
A*0205	0.022	B*0801	0.035	Cw*0304	0.016
A*0206	0.002	B*1302	0.013	Cw*0401	0.148
A*0225	0.002	B*1401	0.015	Cw*0404	0.004
A*0301	0.106	B*1402	0.05	Cw*0501	0.062
A*0302	0.007	B*1501	0.028	Cw*0602	0.097
A*1101	0.065	B*1503	0.015	Cw*0701	0.124
A*1112	0.002	B*1516	0.002	Cw*0702	0.078
A*2301	0.041	B*1517	0.004	Cw*0704	0.019
A*2302	0.002	B*1518	0.002	Cw*0802	0.062
A*2402	0.11	B*1539	0.013	Cw*1202	0.008
A*2403	0.004	B*1801	0.056	Cw*1203	0.089
A*2417	0.002	B*2703	0.002	Cw*1401	0.012
A*2418	0.002	B*2705	0.026	Cw*1402	0.016
A*2501	0.013	B*2708	0.002	Cw*1502	0.054
A*2601	0.033	B*3501	0.074	Cw*1503	0.012
A*2602	0.002	B*3502	0.011	Cw*1505	0.004
A*2605	0.002	B*3503	0.03	Cw*1601	0.027
A*2901	0.009	B*3504	0.004	Cw*1602	0.008
A*2902	0.048	B*3508	0.002	Cw*1701	0.012
A*3001	0.007	B*3527	0.013	Cw*1703	0.004
A*3002	0.017	B*3533	0.002		
A*3101	0.011	B*3701	0.004		
A*3102	0.004	B*3801	0.046		
A*3201	0.035	B*3806	0.004		
A*3206	0.007	B*3901	0.02		
A*3301	0.015	B*3906	0.004		
A*3303	0.002	B*4001	0.017		
A*6601	0.007	B*4002	0.017		
A*6801	0.039	B*4005	0.002		
A*6802	0.013	B*4101	0.004		
		B*4202	0.004		
		B*4402	0.067		
		B*4403	0.054		
		B*4405	0.002		
		B*4501	0.009		
		B*4901	0.039		
		B*5001	0.037		
		B*5101	0.098		
		B*5108	0.007		
		B*5132	0.002		
		B*5201	0.007		
		B*5301	0.011		
		B*5401	0.002		
		B*5501	0.011		
		B*5601	0.004		
		B*5701	0.028		
		B*5703	0.004		
		B*5706	0.002		
		B*5801	0.013		
		B*6701	0.002		

Table 20. HLA-A, -B, -Cw, allele frequencies in the Azorean population.

HLA-DRB1 n=231	Freq.	HLA-DQA1 n=129	Freq.	HLA-DQB1 n=129	Freq.
DRB1*0101	0.076	DQA1*0101	0.112	DQB1*0201	0.058
DRB1*0102	0.039	DQA1*0102	0.109	DQB1*0202	0.173
DRB1*0103	0.015	DQA1*0103	0.097	DQB1*0203	0.035
DRB1*0106	0.002	DQA1*0104	0.016	DQB1*0301	0.194
DRB1*0301	0.084	DQA1*0105	0.008	DQB1*0302	0.101
DRB1*0302	0.002	DQA1*0201	0.174	DQB1*0303	0.05
DRB1*0304	0.015	DQA1*0202	0.004	DQB1*0304	0.012
DRB1*0401	0.03	DQA1*0301	0.089	DQB1*0402	0.047
DRB1*0402	0.03	DQA1*0302	0.019	DQB1*0501	0.112
DRB1*0403	0.017	DQA1*0303	0.058	DQB1*0502	0.019
DRB1*0404	0.022	DQA1*0401	0.05	DQB1*0503	0.016
DRB1*0405	0.013	DQA1*0501	0.128	DQB1*0601	0.004
DRB1*0407	0.013	DQA1*0505	0.132	DQB1*0602	0.054
DRB1*0408	0.002	DQA1*0601	0.004	DQB1*0603	0.101
DRB1*0701	0.183			DQB1*060401	0.012
DRB1*0801	0.037			DQB1*0609	0.012
DRB1*0803	0.007				
DRB1*0804	0.002				
DRB1*0806	0.002				
DRB1*0901	0.028				
DRB1*1001	0.009				
DRB1*1101	0.048				
DRB1*1102	0.02				
DRB1*1103	0.013				
DRB1*1104	0.033				
DRB1*1111	0.002				
DRB1*1201	0.024				
DRB1*1301	0.076				
DRB1*1302	0.024				
DRB1*1303	0.009				
DRB1*1305	0.004				
DRB1*1310	0.003				
DRB1*1322	0.003				
DRB1*1401	0.008				
DRB1*1404	0.002				
DRB1*1406	0.004				
DRB1*1501	0.073				
DRB1*1502	0.007				
DRB1*1601	0.017				
DRB1*1602	0.002				

Table 21. HLA-DRB1, -DQA1 and -DQB1 allele frequencies in the Azorean population.

HLA-A	Azores Oriental n= 43	Azores Central n= 59	HLA-B	Azores Oriental n= 43	Azores Central n= 59	HLA-DRB1	Azores Oriental n= 43	Azores Central n= 59
A*0101	0.11538	0.08036	B*0702	0.08978	0.08038	DRB1*0101	0.0641	0.08929
A*0201	0.25644	0.2589	B*0801	0.0641	0.01786	DRB1*0102	0.03846	0.05357
A*0202	0.01282	0.01786	B*1302	0.01282	0.02679	DRB1*0103	0.01282	0.03571
A*0205	0.01282	0.00893	B*1401	0.01282	0.01786	DRB1*0301	0.12821	0.08929
A*0301	0.0641	0.125	B*1402	0.05128	0.05357	DRB1*0401	0.03846	0.04464
A*1101	0.03846	0.05357	B*1501	0.01282	0.03571	DRB1*0402	0.02564	0.02679
A*1112	0.01282	0	B*1503	0	0.00893	DRB1*0403	0.02564	0.00893
A*2301	0.02564	0.07143	B*1516	0.01282	0	DRB1*0404	0.01282	0.01786
A*2402	0.10256	0.125	B*1801	0.07692	0.08036	DRB1*0405	0	0.01786
A*2403	0	0.00893	B*2705	0.02564	0.01786	DRB1*0407	0.02564	0
A*2501	0	0.01786	B*3501	0.08974	0.07143	DRB1*0701	0.20515	0.14288
A*2601	0.01282	0.04464	B*3502	0.01282	0.00893	DRB1*0801	0.02564	0.03571
A*2605	0.01282	0	B*3503	0.02564	0.01786	DRB1*0803	0	0.00893
A*2901	0.01282	0	B*3508	0	0.00893	DRB1*0806	0.01282	0
A*2902	0.0641	0.08036	B*3701	0.01282	0	DRB1*0901	0.01282	0.01786
A*3001	0	0.00893	B*3801	0.0641	0.05357	DRB1*1001	0.01282	0.00893
A*3002	0.05128	0.00893	B*3901	0	0.01786	DRB1*1101	0.0641	0.03571
A*3101	0.01282	0	B*3906	0.01282	0	DRB1*1102	0	0.00893
A*3102	0.01282	0	B*4001	0.01282	0.03571	DRB1*1103	0	0.00893
A*3201	0.05128	0.02679	B*4002	0.01282	0.03571	DRB1*1104	0.03846	0.03571
A*3206	0	0.00893	B*4402	0.0641	0.07143	DRB1*1111	0	0.00893
A*3301	0.05128	0.01786	B*4403	0.07692	0.07143	DRB1*1201	0.02564	0.03571
A*3303	0	0.00893	B*4405	0	0.00893	DRB1*1301	0.05128	0.0625
A*6801	0.03846	0.02679	B*4501	0	0.00893	DRB1*1302	0.01282	0.00893
A*6802	0.03846	0	B*4901	0.03846	0.03571	DRB1*1303	0.01282	0
			B*5001	0.02564	0.01786	DRB1*1305	0	0.00893
			B*5101	0.08974	0.09821	DRB1*1310	0	0.00893
			B*5108	0.01282	0.01786	DRB1*1322	0	0.00893
			B*5201	0.01282	0.00893	DRB1*1401	0.03846	0.05357
			B*5301	0.01282	0	DRB1*1501	0.08974	0.08036
			B*5501	0.02564	0.01786	DRB1*1502	0.01282	0.00893
			B*5601	0	0.00893	DRB1*1601	0.01282	0.01786
			B*5701	0.01282	0.03571	DRB1*1602	0	0.00893
			B*5801	0.02564	0.00893			

Table 22. HLA-A, -B, and -DRB1 allele frequencies in the Azores Oriental and Azores Central populations.

5.4.1. HLA-A *locus*

The most frequent of these alleles is by far HLA-A*0201 found in 24.5% of the samples, which is similar to Portugal mainland. Three alleles follow with frequencies ranging between 10-11% (A*0101, A*0301, and A*2402), slightly higher than in mainland Portugal (7-8%). Two alleles specific of the sub-Saharan region were found (*0202 and *0225). Surprisingly group A*24 (with four alleles in Azores) is not polymorphic in mainland Portugal, where it is represented only by B*2402.

5.4.2. HLA-B *locus*

HLA-B*5101 (9.8%) was the most frequent, an allele that appears with diverse values in Europe (3-23%). Other alleles show frequencies varying from 5-8% (B*0702, B*3501, B*4402, B*4403, B*1801, and B*1402). Allele B*3801 (4.6%), despite having similar frequencies with several Europeans populations, is rare in Portugal (0.3%). Otherwise, B*3502 (1.1%) has a much lower prevalence than in Portugal, where it is found at the world highest prevalence (5%).

5.4.3. HLA-Cw *locus*

Alleles Cw*0401 (15%) are the most common, with frequency higher than Europeans and similar to several sub-Saharan populations. The second most frequent allele, HLA-Cw*0701 (12.4%), has similar frequencies in sub-Saharan but lower among Caucasians. Alleles HLA-Cw*0602, Cw*1203, Cw*0202 and Cw*0702 appear in 8-10% of the samples with similar (Cw*0602 and Cw*0702) and lower (Cw*1203 and Cw*0202) frequencies than those found in European populations.

5.4.4. HLA-DRB1 *locus*

Allele DRB1*0701 (18.3%) is the most common showing higher frequency than in other Europeans with the exception of some Spanish populations (Middleton *et al.* 2004). Alleles DRB1*0301, DRB1*0101, DRB1*1301, and DRB1*1501 appear in 7-8% of the samples, with similar frequencies found in Portugal (DRB1*1301, DRB1*1501) and European (DRB1*0101, DRB1*0301) populations.

5.4.5. HLA-DQA1 *locus*

DQA1*0201 (17.4%) is the most common allele, presenting a frequency similar to Europeans (and even Chinese) but higher than sub-Saharan. Allele DQA1*0505 (13.2%) is absent in many populations (e.g. Spain and Cameroon), is found at similar frequencies in Tunisia, but has a much higher frequency in Italy (30%). DQA1*0501 (12.8%) in Azores has similar frequencies to some sub-Saharan populations (e.g. Cameroon, Congo, and Uganda) being always lower than in Europeans.

5.4.6. HLA-DQB1 *locus*

DQB1*0301 (19.4%) and DQB1*0202 (17.3%) are the most common alleles in the Azores and have frequencies rather similar or slightly higher to other Europeans (Middleton *et al.* 2004). DQB1*0501, DQB1*0201 and DQB1*0603 alleles, have also values similar to other European populations (10-11%). Amazingly, DQB1*0203 (3.5%) is a very rare allele in other world populations.

5.4.7. Haplotype frequencies

The exact test of *linkage disequilibrium* between each of the 15 pairs of *loci* was statistically significant except for HLA-A, -DQB1. The most representative extended haplotypes and A-B-DRB1 haplotypes (Azores Central and Oriental) are listed in Tables 23 and 24 respectively. Table 25 shows the most frequent A-B-DRB1 haplotypes found in Azores Central and Oriental islands. Tables 26-33 shows the most common two *loci* haplotypes with statistical significance *linkage disequilibrium* found in Azores.

Only 2 of the 9 most frequent ($\geq 1.2\%$) extended haplotypes found in Azores, A*2902-B*4403-Cw*1601-DRB1*0701-DQA1*0201-DQB1*0202 (1.6%), considered of Western European origin, and A*0201-B*5101-Cw*1502-DRB1*0701-DQA1*0201-DQB1*0202 (1.2%), have the corresponding A-B-DRB1 partial haplotype in mainland Portugal with frequencies of 0.7% and 1.7% respectively. One of these extended haplotypes was not found in mainland Portugal: A*0301-B*0702-Cw*0702-DRB1*1501-DQA1*0102-DQB1*0602 considered of North African/Western European origin (Middleton *et al.* 2004).

Some three-*loci* haplotypes found in Azores were identified before as Pan European (A*0101-B*0801-DRB1*0301: 1.3%), Mediterranean (A*0201-B*1801-DRB1*1104: 1.3%), and North African/Western European (A*0301-B*0702-DRB1*1501: 1.3%) (Middleton *et al.* 2004). Six of the 11 most frequent three-*loci* haplotypes were common to mainland Portugal, including the most frequent A*0201-B*5101-DRB1*0701 (2% in Azores and 0.7% in mainland Portugal).

Haplotypes ≥ 0.008	Frequency
A*0101-B*0801-Cw*0202-DRB1*0102-DQA1*0101-DQB1*0501	0.008
A*0101-B*1402-Cw*0802-DRB1*0102-DQA1*0101-DQB1*0501	0.012
A*0101-B*1402-Cw*0802-DRB1*0407-DQA1*0501-DQB1*0203	0.008
A*0101-B*5701-Cw*0701-DRB1*0701-DQA1*0201-DQB1*0303	0.008
A*0201-B*0702-Cw*0702-DRB1*0101-DQA1*0101-DQB1*0501	0.016
A*0201-B*0702-Cw*0702-DRB1*1501-DQA1*0102-DQB1*0302	0.008
A*0201-B*1539-Cw*0303-DRB1*0401-DQA1*0301-DQB1*0302	0.008
A*0201-B*4402-Cw*0501-DRB1*1301-DQA1*0103-DQB1*0603	0.012
A*0201-B*4901-Cw*0701-DRB1*1102-DQA1*0505-DQB1*0301	0.008
A*0201-B*5001-Cw*0602-DRB1*0701-DQA1*0201-DQB1*0202	0.012
A*0201-B*5101-Cw*1401-DRB1*090102-DQA1*0302-DQB1*0303	0.008
A*0201-B*5101-Cw*1502-DRB1*0701-DQA1*0201-DQB1*0202	0.012
A*0201-B*5701-Cw*0602-DRB1*0701-DQA1*0201-DQB1*0303	0.008
A*0202-B*1503-Cw*0202-DRB1*090102-DQA1*0303-DQB1*0202	0.019
A*0205-B*4901-Cw*0701-DRB1*1102-DQA1*0505-DQB1*0301	0.008
A*0301-B*0702-Cw*0702-DRB1*1501-DQA1*0102-DQB1*0602	0.012
A*0301-B*2705-Cw*0202-DRB1*0404-DQA1*0301-DQB1*0302	0.008
A*0301-B*4001-Cw*0701-DRB1*0301-DQA1*0501-DQB1*0201	0.008
A*1101-B*1801-Cw*0501-DRB1*0301-DQA1*0501-DQB1*0201	0.008
A*1101-B*4402-Cw*0501-DRB1*0403-DQA1*0301-DQB1*0304	0.008
A*2402-B*3503-Cw*1203-DRB1*1101-DQA1*0505-DQB1*0301	0.008
A*2402-B*3901-Cw*1203-DRB1*1301-DQA1*0103-DQB1*0603	0.008
A*2402-B*5101-Cw*0202-DRB1*1101-DQA1*0505-DQB1*0301	0.008
A*2601-B*3801-Cw*1203-DRB1*1301-DQA1*0103-DQB1*0603	0.008
A*2902-B*4403-Cw*1601-DRB1*0701-DQA1*0201-DQB1*0202	0.016
A*3201-B*1401-Cw*0802-DRB1*0701-DQA1*0201-DQB1*0202	0.012
A*3201-B*1402-Cw*0802-DRB1*0801-DQA1*0401-DQB1*0402	0.008

Table 23. Most common HLA A-B-C-DRB1-DQA1-DQB1 extended haplotypes in the Azorean population as estimated by maximum likelihood.

HLA A-B-DRB1 Haplotypes ≥ 1	Frequency
A*0101-B*0801-DRB1*0301	0.013
A*0101-B*1402-DRB1*0102	0.013
A*0201-B*0702-DRB1*0101	0.015
A*0201-B*1801-DRB1*1104	0.013
A*0201-B*4402-DRB1*1301	0.013
A*0201-B*5101-DRB1*0701	0.02
A*0202-B*1503-DRB1*0901	0.011
A*0301-B*0702-DRB1*1501	0.013
A*0301-B*3501-DRB1*0101	0.011
A*2402-B*0702-DRB1*1501	0.015
A*2902-B*4403-DRB1*0701	0.024

Table 24. Most common HLA AB-DRB1 haplotypes in the Azorean population as estimated by maximum likelihood.

Haplotypes	Possible origin	Azores Oriental	Azores Central
A*0101-B*0801-DRB1*0301	Pan European	0.026	0
A*0101-B*0801-DRB1*0401	?	0.026	0
A*0101-B*1402-DRB1*0102	Mediterranean	0	0.018
A*0201-B*0702-DRB1*0101	?	0	0.018
A*0201-B*1302-DRB1*0701	Central European	0.013	0.018
A*0201-B*1801-DRB1*1104	Mediterranean	0.026	0
A*0201-B*4002-DRB1*1501	?	0	0.018
A*0201-B*4402-DRB1*0401	3.8% in Irish	0	0.027
A*0201-B*4402-DRB1*0701	North-Western European	0.026	0
A*0201-B*4403-DRB1*0701	North-Western European	0	0.018
A*0201-B*5101-DRB1*0701	?	0.026	0
A*0301-B*1801-DRB1*1104	?	0	0.018
A*0301-B*4403-DRB1*0701	?	0	0.027
A*1101-B*4402-DRB1*1201	?	0	0.018
A*1101-B*5101-DRB1*0101	?	0	0.018
A*2301-B*3501-DRB1*0101	?	0	0.018
A*2402-B*0702-DRB1*1501	?	0.038	0
A*2402-B*0801-DRB1*0301	?	0	0.018
A*2601-B*1401-DRB1*0701	?	0	0.018
A*2902-B*0702-DRB1*0103	?	0	0.018
A*2902-B*4001-DRB1*0301	?	0	0.018
A*2902-B*4403-DRB1*0701	Western European	0.038	0
A*2902-B*5701-DRB1*0701	?	0	0.018
A*3002-B*1801-DRB1*0301	Iberia/North African	0.026	0
A*3201-B*4901-DRB1*0701	?	0.026	0
A*3301-B*1402-DRB1*0102	Mediterranean	0.038	0.018

Table 25. Most common HLA A-B-DRB1 haplotypes in Azores Oriental and Central groups as estimated by maximum likelihood (only frequencies above 0.017 or haplotypes common to Oriental and Central Azores Islands are shown). Haplotypes possible origin is accordingly to Sanchez-Velasco *et al.* (2003), Arnaiz-Villena *et al.* (2001a), Gómez-Casado *et al.* (2000), and Middleton *et al.* (2004)¹.

¹ www.allelefrequencies.net (20-01-2005)

Haplotypes	Frequency	s.d.	D	D'	P
A*0101-B*0801	0.021	0.007	0.017	0.742	**
A*0101-B*1402	0.013	0.006	0.008	0.345	*
A*0101-B*5101	0.02	0.01	0.010	0.443	*
A*0101-B*5701	0.01	0.005	0.007	0.313	*
A*0201-B*4001	0.011	0.004	0.006	0.284	*
A*0201-B*4402	0.029	0.009	0.012	0.543	*
A*0201-B*4901	0.017	0.005	0.008	0.335	*
A*0201-B*5101	0.044	0.008	0.02	0.872	**
A*0202-B*1503	0.011	0.005	0.01	0.454	**
A*0301-B*0702	0.02	0.007	0.012	0.536	**
A*0301-B*3501	0.03	0.009	0.022	0.964	**
A*1101-B*3501	0.015	0.007	0.01	0.457	**
A*1101-B*4402	0.013	0.004	0.009	0.388	**
A*2402-B*0702	0.017	0.007	0.009	0.391	*
A*2402-B*1501	0.012	0.004	0.009	0.394	**
A*2402-B*3503	0.011	0.006	0.008	0.329	*
A*2402-B*3801	0.011	0.007	0.006	0.267	*
A*2601-B*3801	0.01	0.004	0.009	0.377	**
A*2902-B*4403	0.025	0.006	0.023	1	**
A*3301-B*1402	0.015	0.007	0.014	0.63	**

Table 26. Most common HLA-A, -B (n=231) two-*loci* haplotypes in the Azores population as estimated by maximum likelihood (only frequencies above 0.01 with statistical significance *linkage disequilibrium* are shown). *P<0.05, ** P<0.001.

Haplotypes	Frequency	s.d.	D	D'	P
A*0101-DRB1*0102	0.013	0.004	0.009	0.392	**
A*0101-DRB1*0301	0.019	0.008	0.011	0.475	*
A*0201-DRB1*0401	0.015	0.007	0.008	0.354	*
A*0201-DRB1*0402	0.016	0.006	0.009	0.383	*
A*0202-DRB1*0901	0.011	0.006	0.009	0.401	**
A*0301-DRB1*0101	0.02	0.008	0.012	0.5	*
A*0301-DRB1*0404	0.011	0.005	0.008	0.369	**
A*0301-DRB1*0801	0.013	0.006	0.009	0.373	*
A*0301-DRB1*1104	0.011	0.004	0.007	0.318	*
A*2301-DRB1*0701	0.015	0.009	0.008	0.346	*
A*2402-DRB1*1302	0.01	0.003	0.008	0.340	**
A*2402-DRB1*1501	0.017	0.007	0.009	0.391	*
A*2601-DRB1*1301	0.011	0.005	0.009	0.378	**
A*2902-DRB1*0701	0.032	0.01	0.023	1	**
A*3201-DRB1*0701	0.017	0.007	0.011	0.475	*

Table 27. Most common HLA-A, -DRB1 (n=231) two-*loci* haplotypes in the Azores population as estimated by maximum likelihood (only frequencies above 0.01 with statistical significance *linkage disequilibrium* are shown). *P<0.05, ** P<0.001.

Haplotypes	Frequency	s.d.	D	D'	P
B*0702-DRB1*0101	0.013	0.006	0.007	0.229	*
B*0702-DRB1*1501	0.034	0.007	0.029	0.949	**
B*0801-DRB1*0301	0.019	0.005	0.017	0.549	**
B*1302-DRB1*0701	0.011	0.005	0.008	0.28	**
B*1401-DRB1*0701	0.016	0.006	0.012	0.412	**
B*1402-DRB1*0102	0.028	0.005	0.026	0.848	**
B*1503-DRB1*0901	0.011	0.004	0.010	0.345	**
B*1801-DRB1*0301	0.016	0.008	0.011	0.37	**
B*1801-DRB1*1104	0.017	0.007	0.015	0.491	**
B*3501-DRB1*0101	0.026	0.009	0.021	0.684	**
B*3801-DRB1*1301	0.022	0.008	0.018	0.603	**
B*4402-DRB1*0401	0.011	0.007	0.009	0.287	**
B*4402-DRB1*1201	0.015	0.006	0.014	0.449	**
B*4403-DRB1*0701	0.04	0.008	0.030	1	**
B*5701-DRB1*0701	0.024	0.009	0.019	0.623	**

Table 28. Most common HLA-B, -DRB1 (n=231) two-*loci* haplotypes in the Azores population as estimated by maximum likelihood (only frequencies above 0.01 with statistical significance *linkage disequilibrium* are shown). *P<0.05, ** P<0.001.

Haplotypes	Frequency	s.d.	D	D'	P
A*0101-Cw*0602	0.035	0.007	0.025	0.781	**
A*0101-Cw*0701	0.028	0.014	0.015	0.469	*
A*0101-Cw*0802	0.019	0.009	0.013	0.406	*
A*0101-Cw*1402	0.011	0.004	0.009	0.281	**
A*0201-Cw*0401	0.04	0.017	0.003	0.009	**
A*0201-Cw*0702	0.042	0.012	0.023	0.719	*
A*0202-Cw*0202	0.019	0.008	0.017	0.531	**
A*0205-Cw*0701	0.015	0.005	0.012	0.375	**
A*0301-Cw*0401	0.047	0.012	0.032	1	**
A*1101-Cw*0401	0.022	0.006	0.012	0.375	*
A*1101-Cw*0501	0.016	0.008	0.011	0.344	*
A*1101-Cw*0602	0.019	0.008	0.013	0.406	*
A*2402-Cw*0202	0.023	0.008	0.014	0.438	*
A*2402-Cw*0303	0.03	0.009	0.025	0.781	**
A*2402-Cw*1203	0.026	0.014	0.016	0.5	*
A*2601-Cw*1203	0.017	0.01	0.014	0.438	**
A*2902-Cw*1601	0.019	0.006	0.018	0.563	**
A*3201-Cw*0802	0.019	0.007	0.017	0.531	**

Table 29. Most common HLA-A, -Cw (n=129) two-*loci* haplotypes in the Azores population as estimated by maximum likelihood (only frequencies above 0.01 with statistical significance *linkage disequilibrium* are shown). *P<0.05, ** P<0.001.

Haplotypes	Frequency	s.d.	D	D'	P
B*0702-Cw*0702	0.066	0.018	0.06	1	**
B*0801-Cw*0701	0.031	0.013	0.027	0.45	**
B*1402-Cw*0802	0.047	0.013	0.043	0.717	**
B*1501-Cw*0303	0.027	0.014	0.026	0.433	**
B*1539-Cw*0303	0.019	0.011	0.019	0.317	**
B*3501-Cw*0401	0.058	0.011	0.047	0.783	**
B*4402-Cw*0501	0.043	0.016	0.038	0.633	**
B*5001-Cw*0602	0.043	0.013	0.039	0.65	**
B*5101-Cw*1502	0.047	0.013	0.041	0.683	**

Table 30. Most common HLA-B, -Cw (n=129) two-*loci* haplotypes in the Azores population as estimated by maximum likelihood (only frequencies above 0.01 with statistical significance *linkage disequilibrium* are shown). *P<0.05, ** P<0.001.

Haplotypes	Frequency	s.d.	D	D'	P
DRB1*0101-DQB1*0501	0.054	0.013	0.046	0.39	**
DRB1*0102-DQB1*0501	0.031	0.015	0.027	0.229	**
DRB1*0301-DQB1*0201	0.035	0.013	0.03	0.254	**
DRB1*0301-DQB1*0203	0.016	0.006	0.013	0.11	**
DRB1*0304-DQB1*0201	0.016	0.006	0.015	0.127	**
DRB1*0401-DQB1*0302	0.012	0.005	0.009	0.076	*
DRB1*0402-DQB1*0302	0.031	0.012	0.028	0.237	**
DRB1*0404-DQB1*0302	0.023	0.011	0.021	0.178	**
DRB1*0405-DQB1*0302	0.016	0.007	0.014	0.119	**
DRB1*0701-DQB1*0202	0.15	0.036	0.118	1	**
DRB1*0701-DQB1*0303	0.028	0.009	0.019	0.161	**
DRB1*0801-DQB1*0402	0.043	0.007	0.041	0.347	**
DRB1*090102-DQB1*0202	0.02	0.008	0.016	0.136	**
DRB1*090102-DQB1*0303	0.011	0.006	0.009	0.076	**
DRB1*1101-DQB1*0301	0.05	0.01	0.041	0.347	**
DRB1*1102-DQB1*0301	0.027	0.01	0.023	0.195	**
DRB1*1103-DQB1*0301	0.016	0.011	0.013	0.11	**
DRB1*1104-DQB1*0301	0.027	0.01	0.021	0.178	**
DRB1*1201-DQB1*0301	0.019	0.01	0.015	0.127	**
DRB1*1301-DQB1*0603	0.089	0.015	0.081	0.686	**
DRB1*1401-DQB1*0503	0.012	0.004	0.012	0.102	**

Table 31. Most common HLA-DRB1, -DQB1 (n=129) two-*loci* haplotypes in the Azores population as estimated by maximum likelihood (only frequencies above 0.01 with statistical significance *linkage disequilibrium* are shown). *P<0.05, ** P<0.001.

Haplotypes	Frequency	s.d.	D	D'	P
DRB1*0101-DQA1*0101	0.054	0.016	0.046	0.331	**
DRB1*0301-DQA1*0501	0.054	0.018	0.044	0.317	**
DRB1*0304-DQA1*0501	0.027	0.012	0.025	0.18	**
DRB1*0401-DQA1*0301	0.012	0.008	0.009	0.065	*
DRB1*0402-DQA1*0301	0.031	0.01	0.029	0.209	**
DRB1*0403-DQA1*0301	0.016	0.008	0.014	0.101	**
DRB1*0404-DQA1*0301	0.019	0.012	0.017	0.122	**
DRB1*0405-DQA1*0303	0.016	0.006	0.015	0.108	**
DRB1*0701-DQA1*0201	0.171	0.023	0.139	1	**
DRB1*0801-DQA1*0401	0.043	0.015	0.041	0.295	**
DRB1*0901-DQA1*0302	0.016	0.006	0.015	0.108	**
DRB1*0901-DQA1*0303	0.023	0.011	0.022	0.158	**
DRB1*1101-DQA1*0505	0.047	0.011	0.040	0.288	**
DRB1*1102-DQA1*0505	0.027	0.015	0.024	0.173	**
DRB1*1103-DQA1*0505	0.012	0.005	0.01	0.072	**
DRB1*1104-DQA1*0505	0.016	0.012	0.011	0.079	*
DRB1*1201-DQA1*0505	0.016	0.007	0.012	0.086	**
DRB1*1301-DQA1*0103	0.089	0.015	0.082	0.59	**
DRB1*1302-DQA1*0102	0.023	0.009	0.021	0.151	**
DRB1*1401-DQA1*0104	0.012	0.008	0.012	0.086	**
DRB1*1501-DQA1*0102	0.05	0.015	0.042	0.302	**
DRB1*1601-DQA1*0102	0.012	0.01	0.01	0.072	**

Table 32. Most common HLA-DRB1, -DQA1 (n=129) two-*loci* haplotypes in the Azores population as estimated by maximum likelihood (only frequencies above 0.01 with statistical significance *linkage disequilibrium* are shown). *P<0.05, ** P<0.001.

Haplotypes	Frequency	s.d.	D	D'	P
DQA1*0501-DQB1*0201	0.058	0.014	0.051	0.481	**
DQA1*0201-DQB1*0202	0.136	0.026	0.106	1	**
DQA1*0303-DQB1*0202	0.023	0.008	0.013	0.123	*
DQA1*0501-DQB1*0203	0.035	0.011	0.03	0.283	**
DQA1*0505-DQB1*0301	0.132	0.017	0.106	1	**
DQA1*0301-DQB1*0302	0.074	0.016	0.065	0.613	**
DQA1*0303-DQB1*0302	0.019	0.009	0.014	0.132	*
DQA1*0201-DQB1*0303	0.031	0.011	0.022	0.208	**
DQA1*0301-DQB1*0304	0.012	0.007	0.011	0.104	**
DQA1*0401-DQB1*0402	0.047	0.016	0.044	0.415	**
DQA1*0101-DQB1*0501	0.097	0.025	0.084	0.792	**
DQA1*0102-DQB1*0502	0.016	0.01	0.013	0.123	**
DQA1*0104-DQB1*0503	0.016	0.004	0.015	0.142	**
DQA1*0102-DQB1*0602	0.047	0.01	0.041	0.387	**
DQA1*0103-DQB1*0603	0.093	0.022	0.083	0.783	**
DQA1*0102-DQB1*0604	0.012	0.006	0.01	0.094	**
DQA1*0102-DQB1*0609	0.012	0.006	0.01	0.094	**

Table 33. Most common HLA-DQA1, -DQB1 (n=129) two-*loci* haplotypes in the Azores population as estimated by maximum likelihood (only frequencies above 0.01 with statistical significance *linkage disequilibrium* are shown). *P<0.05, ** P<0.001.

The most frequent A-B-DRB1 haplotypes estimated from donors whose parents and grandfathers were born and living on the same island (Az1 samples, Table 25) do not show a homogeneous distribution in the Oriental and Central Azores islands. Only 2 out of 26 have a distribution common to both groups of islands: the Central European A*0201-B*1302-DRB1*0701, which is nevertheless absent from mainland Portugal, and the Mediterranean A*3301-B*1402-DRB1*0102 which exists in mainland Portugal.

The Western European haplotype A*2902-B*4403-DRB1*0701 was found only on the Oriental islands, at a frequency of 3.8%. The same happens with the pan-European A*0101-B*0801-DRB1*0301 (present at 2-6%), the Mediterranean A*0201-B*1801-DRB1*1104 (2-6%, roughly the same as in Portugal), the Iberian/Paleo North African A*3002-B*1801-DRB1*0301 (2.6%, but 1.7% in Portugal), the North-Western European A*0201-B*4402-DRB1*0701 (2.6%), and the North African/Western European A*0301-B*0702-DRB1*1501 (1-3%, Table 25). Other haplotypes already found in mainland Portugal in residual frequencies appear only in the Oriental group (A*0201-B*5101-DRB1*0701: 2.6%, and A*2402-B*0702-DRB1*1501: 3.8%). Three haplotypes are an exception to this situation: the north-western European A*0201-B*4403-DRB1*0701 haplotype (Azores Central: 1.8%; and Portugal mainland: 3.1%), the Mediterranean haplotype A*3301-B*1402-DRB1*0102 (Azores Central: 1.8%; Azores Oriental: 3.8%; and Portugal mainland: 2.1%), and the haplotype A*0201-B*4402-DRB1*1301 exclusive from the Central islands and found before in the South of Portugal at a similar frequency (1%).

The haplotypes A*0101-B*1402-DRB1*0102 (Mediterranean) and A*0201-B*0702-DRB1*0101 (0.3% in Northern Ireland) were present exclusively on the Central group of islands at a frequency of 1.8% each. The Central European haplotype A*0201-B*1302-DRB1*0701 is present in Azores (1.8%) but is absent in mainland Portugal. The two most frequent haplotypes in Central Azores (A*0301-B*4403-DRB1*0701, 2.8% and A*0201-B*4402-DRB1*0401, 2.7% each), were found in Northern Ireland at 0.5% and 3.8% respectively.

The most frequent haplotypes found in mainland Portugal have similar frequencies (A*2902-B*4403-DRB1*0701; and A*2301-B*4403-DRB1*0701), lower frequencies (A*0201-B*4403-DRB1*0701; A*3301-B*1402-DRB1*0102; A*3002-B*1801-DR*0301; and A*0201-B*0702-DRB1*1501), or are absent (A*0101-B*4403-DRB1*0701; A*0201-B*1801-DRB1*0301; A*2301-B*4901-DRB1*1302; A*6801-B*4501-DRB1*0405; and A*2402-B*5101-DRB1*0701) in the Azores.

5.4.8. Phylogenetic analyses

A phylogenetic tree constructed with HLA-A, HLA-B and HLA-DRB1 allele frequencies (Figure 9) clusters Azores to Asian and Native American populations. The same analysis without Asian and Ameridian populations, groups Azores to other Europeans rather than Portugal and North Africans. When considering the two Azorean sub-populations (Oriental and Central), dendrogram separate these two groups and cluster Central islands to Asiatic and Amerindian populations (Figure 10). When Oriental and Central Azores were analysed only with European and African populations, the former emerges more similar to Portugal and North Africans but the later clusters to

other Europeans (Figure 11). A principal coordinate analysis based on HLA-A, HLA-B, and HLA-DRB1 allele frequencies shows the Azorean in close proximity to Europeans (Figure 12).

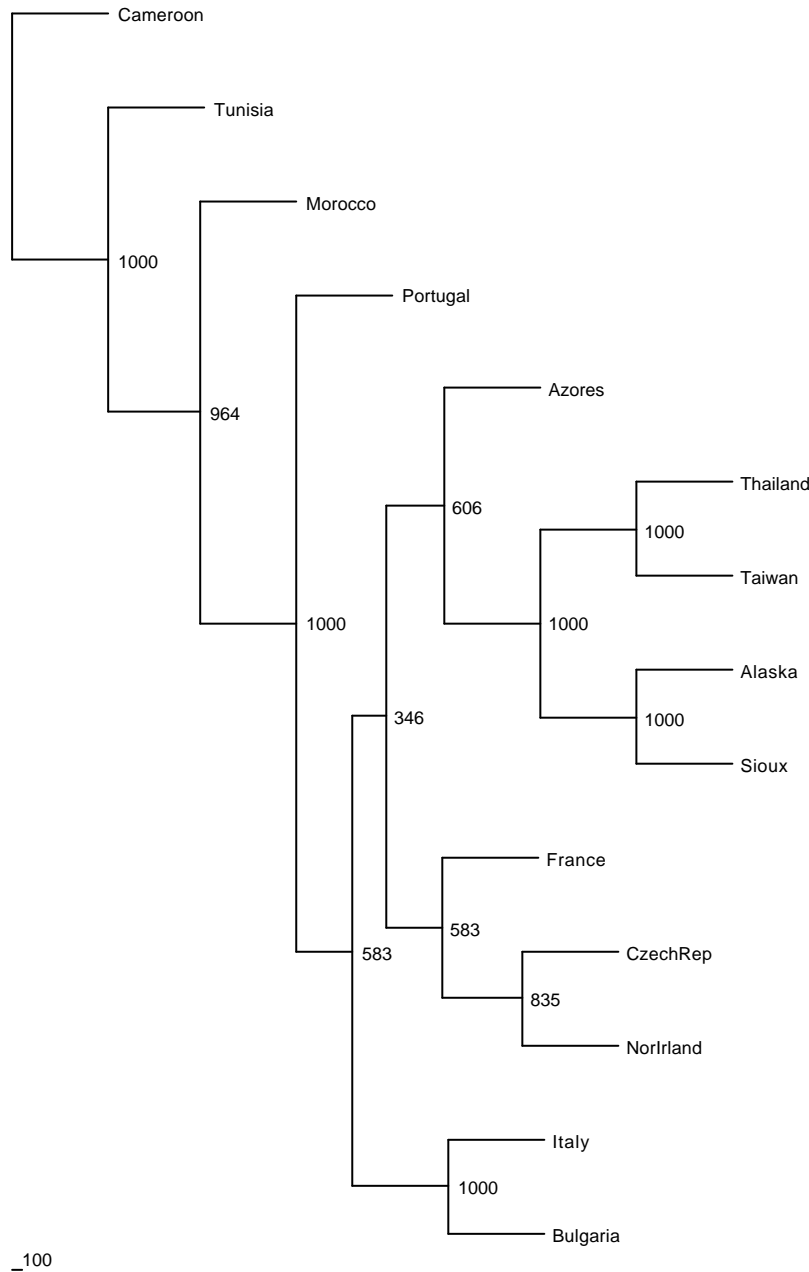
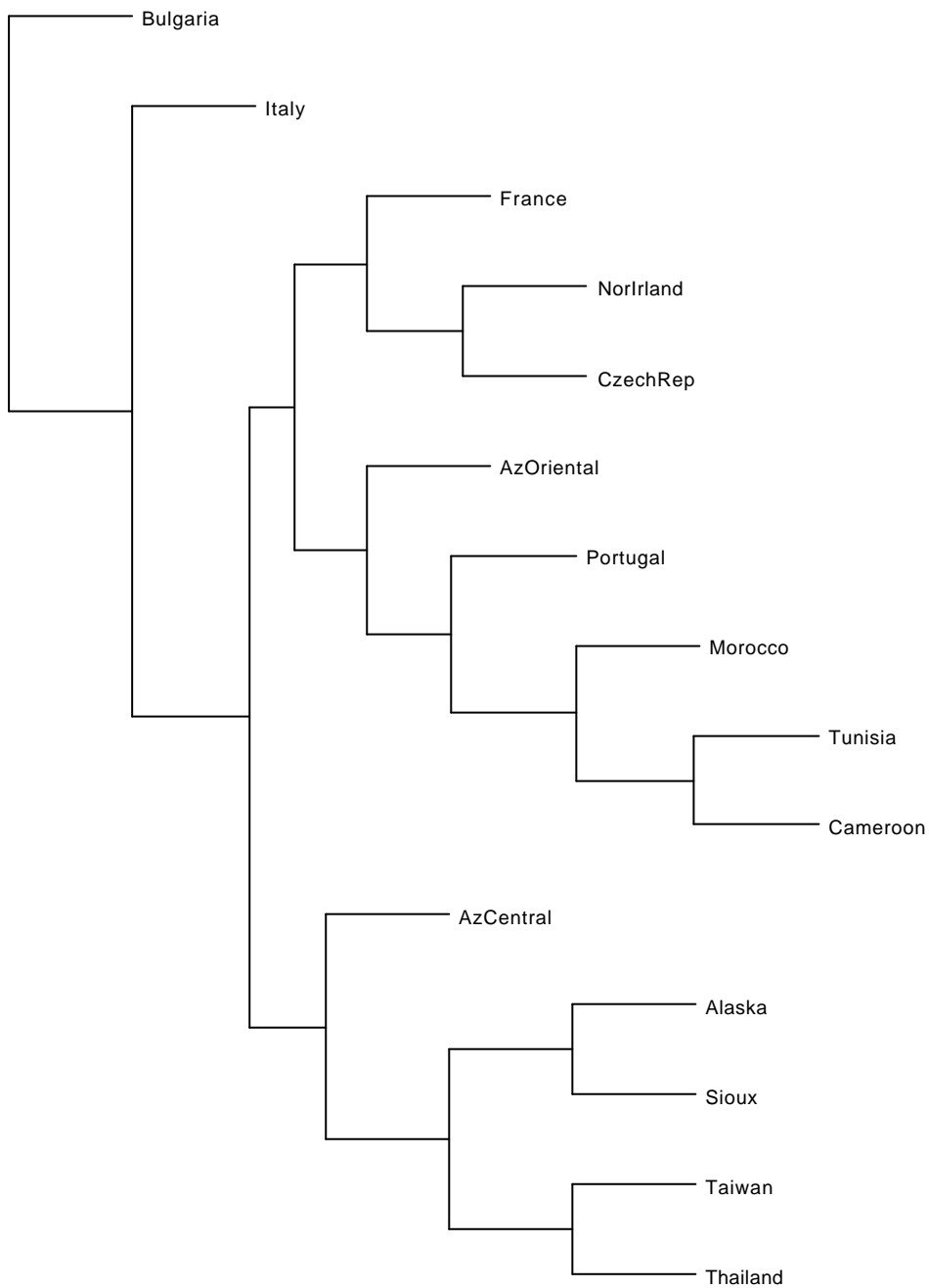


Figure 9. Neighbor-joining (NJ) dendrogram constructed with standard genetic distances calculated using HLA-A, -B and -DRB1 allele frequencies showing the comparative position of the Azores with other populations typed with similar resolution. Numbers above branches are node support after the bootstrap technique implemented in PHYLIP package program BOOT. For references of the populations used see *data analysis*.



_100

Figure 10. Neighbor-joining (NJ) dendrogram constructed with standard genetic distances calculated using HLA-A, -B and -DRB1 allele frequencies, considering the two sub-populations Azores Oriental (AzOriental) and Azores Central (AzCentral). Numbers above branches are node support after the bootstrap technique implemented in PHYLIP package program BOOT. For references of the populations used see *data analysis*.

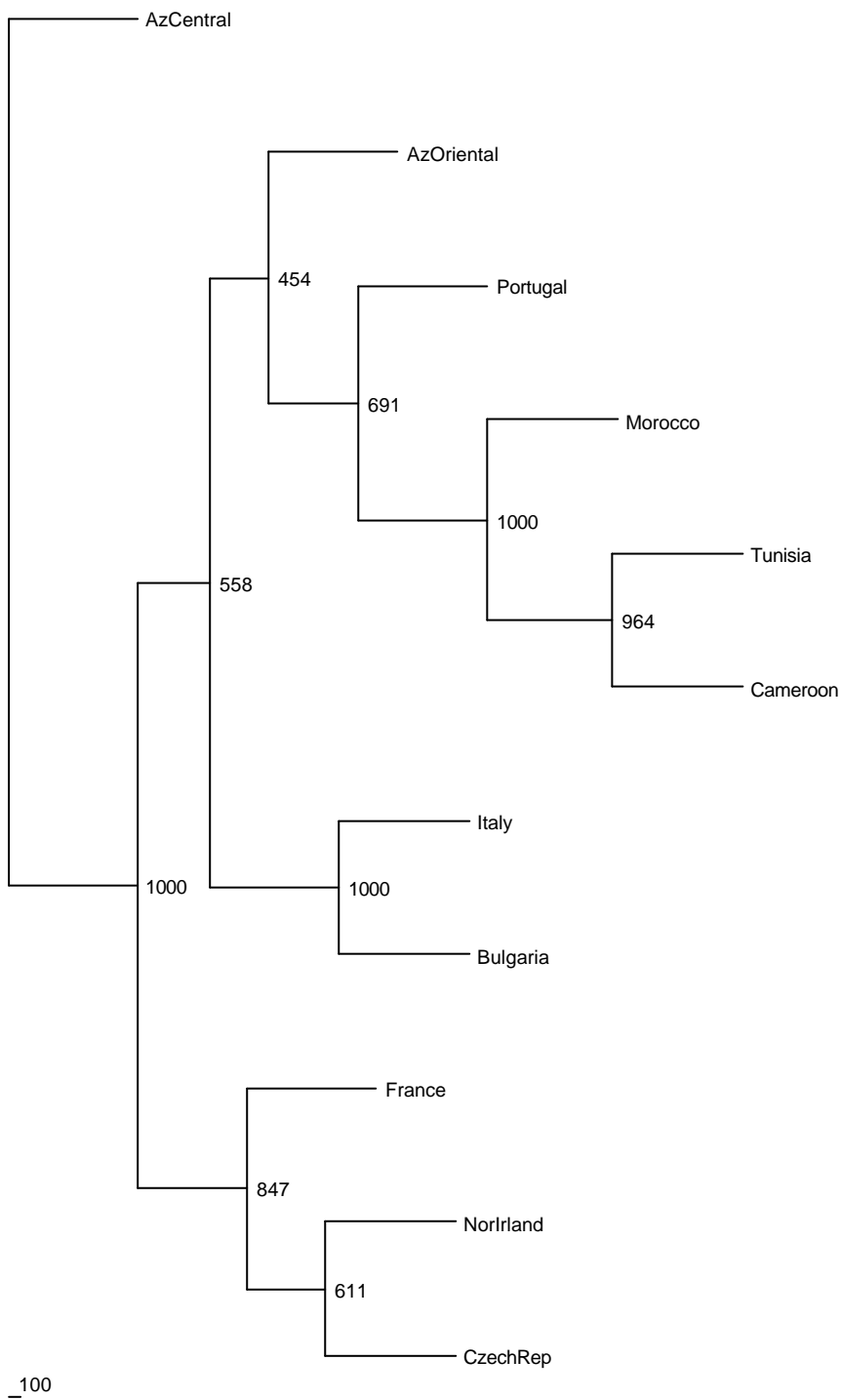


Figure 11. Neighbor-joining (NJ) dendrogram constructed with standard genetic distances calculated using HLA-A, -B and -DRB1 allele frequencies, considering the two sub-populations Azores Oriental (AzOriental) and Azores Central (AzCentral), analysed only with European and African populations. Numbers above branches are node support after the bootstrap technique implemented in PHYLIP package program BOOT. For references of the populations used see *data analysis*.

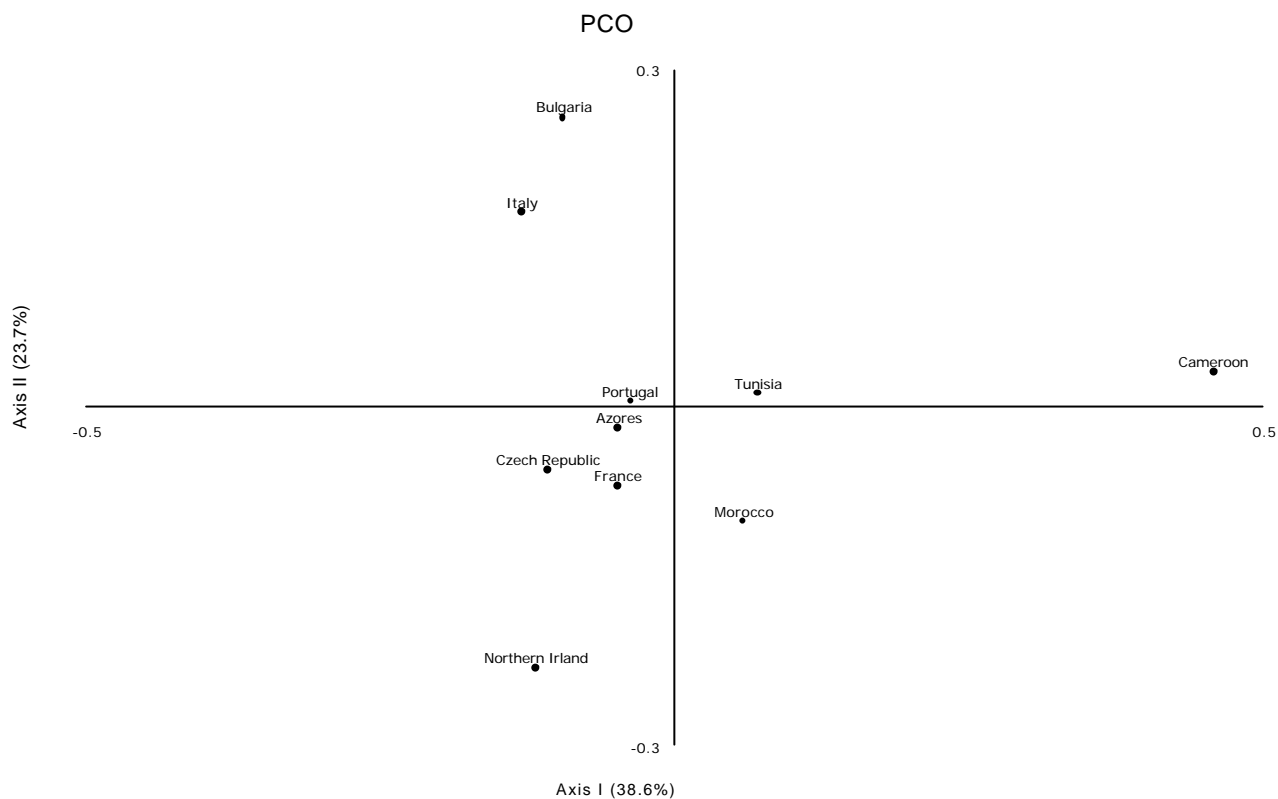


Figure 12. Principal coordinates analysis using HLA-A, -B, and -DRB1 allele frequencies. For references of the populations used see *data analysis*.

5.5. Discussion

During its settlement after 1432, the Azores experienced the influence of various peoples, especially Portuguese, Dutch (Flemish), Spaniards, Africans, Jewish, and Italians (Frutuoso, 1977; Guill, 1993; Mendonça, 1996). The important role played by Azores in sea traffic between America and Europe during the 15th and 16th century was one of the principal reasons for a settlement involving people from several origins.

The present study describe HLA-A, HLA-B, HLA-Cw, HLA-DRB1, HLA-DQA1, and HLA-DQB1 high resolution allele frequencies and estimated

haplotypes in Azores. This high resolution throughput shows that several HLA generic groups typed before by serology (Bruges-Armas *et al.*, 1999) were in fact composed of a variety of groups and alleles (Tables 20 and 21). This additional and much more complete information is highly relevant for a detailed HLA structure characterization. Unfortunately, many world populations do not have HLA allele and haplotype high resolution characterization, making difficult the comparison with the present Azores data. However, websites such as www.allelefrequencies.net have been of great help because it is bringing together in one source the data for HLA, and other immunogenetic *loci*, in world-wide populations.

A dendrogram based on HLA-A, -B, and -DRB1 allele frequencies clusters Azoreans to Asian and Amerindians (Figure 9), a result not consistent with those obtained when class I and class II allele frequencies are used separately (data not shown). In fact the data shows that the Azoreans are more similar to other Europeans than to mainland Portuguese. Otherwise, if we look at the haplotype content prevailing in Azores we cannot find a common stratus with the known haplotypes specific to Asiatic or Ameridian populations. Thus, our results are not in agreement with the previous report of high frequency Mongoloid haplotypes in Azores (Bruges-Armas *et al.*, 1999) and we could not find either the most common haplotypes of Amerindians (Layrisse *et al.*, 2001; Leffell *et al.*, 2004). The low resolution haplotypes previously found at high frequencies (2% each) and considered as of Mongoloid possible origin (Bruges-Armas *et al.*, 1999) were found at much lower frequencies (A*0201-B*5001-DRB1*0701: 0.9%; A*2402-B*4402-DRB1*1102: 0.2%; A*2902-B*4901-DRB1*0701: 0.2%). Two of these haplotypes were also present in Madeira island and

South Portugal (Spínola *et al.*, 2002) in higher frequencies than previously reported in Azores (Bruges-Armas *et al.*, 1999). The Azores similarity to Asians shown by the dendrogram analysis reflects similar allele content rather than a real genetic influence of those populations.

The haplotype and allele frequencies found in Azores reflect the influence of Portuguese and other Europeans in the settlement of this archipelago. Several haplotypes are common to mainland Portugal (e.g. A*0101-B*0801-DRB1*0301, A*0201-B*5101-DRB1*0701, A*0201-B*1801-DRB1*1104), reflecting the weight of the Portuguese in the Azorean settlement process. However, many haplotypes are absent from mainland Portugal (e.g. A*0301-B*0702-DRB1*1501, A*0101-B*1402-DRB1*0102, A*0201-B*0702-DRB1*0101, A*0201-B*4402-DRB1*1301), some of those with a clear European origin. Only 2 of the 9 most frequent (>1.2%) extended haplotypes (Table 23) found in Azores have their partial A-B-DRB1 haplotypes common to mainland Portugal. The haplotypes not common to mainland Portuguese could be brought to Azores by Flemish and others European peoples involved in settling these Atlantic islands. The same conclusions were reported on the *HFE* gene mutations frequency in the Central islands of the Azores archipelago. The frequencies identified on the C282Y or the H63D mutations in the group of individuals under investigation were similar to those previously reported in the groups which possibly populated the Azores (Couto *et al.*, 2003). None of the HLA three *loci* A-B-DRB1 haplotypes found in West Africa Guiné-Bissau population (unpublished data) was found in Azores in spite of the sub-Saharan input of slaves into the Portuguese Atlantic Islands. African specific alleles (e.g. A*0202, A*0225, B*1503, B*1516, B*2703, B*4202, and B*5703) (Cao *et al.*,

2004) were found but mostly at low frequencies. Another study performed in the Central Group of islands identified also HLA-B*2703 which was first described in African populations and later in other populations with known or possible African admixture (Couto *et al.*, 2004). This probably indicates a minor genetic influence left by the sub-Saharan slave trade which is also in agreement with recent Y-chromosome findings and mtDNA data (Brehm *et al.*, 2003; Gonçalves *et al.*, 2005).

5.5.1. Diversity between Azores islands

HLA-A, -B, and -DRB1 allele and haplotype frequencies calculated from Az1 samples show a clear heterogeneity between Azores Central and Oriental islands. A dendrogram based on allele frequencies separate these two groups and cluster Central islands to Asiatic and Amerindian populations (Figure 10). This similarity of Central islands to Asians should not reflect a real genetic influence of those populations as referred above. On the contrary, Oriental islands have more affinity to mainland Portugal in this analysis. A phylogenetic tree constructed only with European and Sub-Saharan populations (Figure 11) clarify the differences between these two groups of Azores islands. The Oriental group shows again its connection to Portugal and North Africans but the Central group emerges more similar to other Europeans. These results are in agreement with the different haplotype profiles found in the two groups of islands, since only 2 of the 26 most frequent tree *loci* haplotypes are common to both groups. The most frequent haplotypes found in one group are absent in the other (Table 25). Azores Oriental present several high frequency haplotypes common to mainland Portugal contrary to what happens in the Central group of islands

which presents only three of those haplotypes. Several most frequent haplotypes found in Central islands have unknown origin probably due to a lack of HLA characterization on several European populations, from where they could be introduced. Further HLA high resolution studies on Dutch, Belgium and other central European populations should clarify the origin of some of these haplotypes.

Both the dendrogram analysis and haplotype content clearly points to a different settlement of both groups of islands from the archipelago. This observation corroborates historical sources that say the Azores were populated not only by Portuguese, especially on Oriental group, but also by other Europeans, mostly Flemish people in Central islands.

Capítulo 6

6. HLA polymorphisms in Cabo Verde and Guiné-Bissau inferred from sequence-based typing

Hélder Spínola, Jácome Bruges-Armas, Derek Middleton, António Brehm

In press Human Immunology

6.1. Abstract

HLA-A, -B, and -DRB1 polymorphisms were examined in the Cabo Verde and Guiné-Bissau populations. The data was obtained at high-resolution level, using sequence based typing. The most frequent alleles in each *locus* was: A*020101 (16.7% in Guiné-Bissau and 13.5% in Cabo Verde), B*350101 (14.4% in Guiné-Bissau and 13.2% in Cabo Verde), DRB1*1304 (19.6% in Guiné-Bissau), and DRB1*110101 (10.1% in Cabo Verde). The predominant three *loci* haplotype in Guiné-Bissau was A*2301-B*1503-DRB1*110101 (4.6%) and in Cabo Verde was A*3002-B*350101-DRB1*1001 (2.8%), exclusive to Northwestern islands (5.6%) and absent in Guiné-Bissau.

The present study corroborates historical sources and other genetic studies that say Cabo Verde were populated not only by Africans but also by Europeans. Haplotypes and dendrogram analysis shows a Caucasian genetic influence in today's gene pool of Cabo Verdeans. Haplotypes and allele frequencies present a differential distribution between Southeastern and Northwestern Cabo Verde islands which could be the result of different genetic influences, founder effect or bottlenecks. Dendrograms and principal coordinates analysis show that Guineans are more similar to North Africans than other HLA studied sub-Saharan, probably due to ancient and recent genetic contacts with other peoples, namely East Africans Arabs and Berbers.

6.2. Introduction

One of the most polymorphic genetic systems in humans is the Human Leukocyte Antigen (HLA), which consist of a closely linked set of genes highly important in transplantation, anthropological and forensic fields (Riley & Olerup, 1992). The characterization of the HLA profile of a population, with high throughput techniques as Sequence Based Typing (SBT), is a helpful mechanism to trace genetic relationships between neighbouring populations. Previous 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; Arnaiz-Villena *et al.*, 1999; Clayton *et al.*, 1997).

Some sub-Saharan African populations have been studied from the HLA point of view (Ellis *et al.*, 2000; Pimtanonthai *et al.*, 2001; Cao *et al.*, 2004). However, high resolution HLA allele typing data is still scarce among sub-

Saharans, especially West African populations. Guiné-Bissau is one of these African Atlantic coast countries that were never HLA typed. The human settlement of the West Coast of Africa is the result of a continuous complex network of migrations, invasions and admixture of peoples from different origins that began around 40000 years before present (YBP) (Alimen, 1987). Before Sahara desertification took place (9000 YBP), several Neolithic cultures flourished in the area, bringing together people of sub-Saharan and North African origin (Dutour *et al.*, 1988). Later, around 4000 YBP, this region became the centre of migrations by different ethnic groups, namely Fula and East Africa Shuwa Arabs, through the Sahel Corridor used for travel between East and West coasts of Africa (Ellis *et al.*, 2000). The most recent historic events, such as admixture of Berber and native populations due to Muslim pressure in Northwestern Africa in the 9th century, also affected this region (Moreira, 1964).

This region was also at the origin of most slave trade in the 15th-16th centuries to Europe and America (Russell-Wood, 1998). The Cabo Verde archipelago was discovered inhabited around 1460 and settled by a few Europeans (less than 10%), while African slaves originating from the Guiné coast constituted most of the remaining population (Carreira, 1983).

Studies on mtDNA, STR and Y-Chromosome SNPs have been done for populations of Guiné-Bissau and Cabo Verde (Brehm *et al.*, 2002; Gonçalves *et al.*, 2002; Fernandes *et al.*, 2003; Gonçalves *et al.*, 2003; Rosa *et al.*, 2004). The population of Cabo Verde have been already HLA typed but in a low resolution level and in a small number of samples (Spínola *et al.*, 2002). The main goal of the present work is to present a clear genetic profile of Cabo Verdeans and Guineans based on high resolution typing of the HLA-A,

-B, and –DRB1 *loci* and evaluate their relatedness to each other and other African and European populations. HLA data comparison between Cabo Verde and Guiné-Bissau may also contribute to clarify the origin of the genetic heterogeneity already found but now using markers not sex-dependent and thus qualified for analysis of admixture which is not sex-biased (Brehm *et al.*, 2002 ;Gonçalves *et al.*, 2003; Rosa *et al.*, 2004).

6.3. Materials and methods

6.3.1. Population samples and HLA typing

The present study population consisted of a total of 189 healthy unrelated males randomly distributed in Cabo Verde Nothwestern islands (n=62), Cabo Verde Southeastern islands (n=62) and Guiné-Bissau (n=65). Blood samples were collected after informed consent, from donors whose parents and grandfathers were born and living in the same island (Cabo Verde) or belong to the same ethnic group (Guiné-Bissau). Guiné-Bissau samples were identified as belonging to seven different ethnic groups: Balanta (n=10), Papel (n=11), Mandinga (n=9), Felupe (n=5), Bijagós (n=10), Fula (n=10), and Mancanha (n=10). Genomic DNA was isolated from whole blood containing EDTA using a phenol-chloroform procedure and frozen at -20°C until use. All subjects were typed for HLA-A, HLA-B and HLA-DRB1 *loci*. Sequence based typing (SBT) of the HLA-A and –B *loci* was performed on exon 2 and exon 3 accordingly to Kurz *et al.* (1999) and Pozzi *et al.* (1999) with minor modifications. DNA fragments amplified by polymerase chain reaction (PCR) were purified and sequenced using ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kits (Applied

Biosystems, Foster City, CA, USA) in an ABI PRISM 310 GENETIC ANALYZER (Applied Biosystems), according to the manufacturer's instructions. HLA-DRB1 SBT was performed on exon 2, amplified and sequenced with group specific primers used in SSOP typing (Williams *et al.*, 2004). All samples were PCR amplified with each of the group specific pair of primers. The positive reactions were sequenced with the amplification primers according to the above description for HLA-A and -B. Sequencing was always performed in forward and reverse directions and processed using the Matchtools Allele Identification packet (Applied Biosystems) which assess the typing by an updated HLA-sequence library with alleles defined up to June 2001. In the present study we did not test for polymorphisms outside exons 2 and 3 in HLA-A and -B and outside exon 2 in HLA-DRB1. Therefore, a few groups of alleles could not be distinguished. A summary of the unresolved ambiguities and the reference used in each case in the text and for frequency analysis is given in Table 8. Likewise, some heterozygous samples could have more than one allele's combination. In these situations, for allele frequencies and haplotype calculations we used a maximum likelihood estimation assignment, choosing the combination of the most frequent single alleles.

6.3.2. Data analysis

Basic genetic parameters (allele and haplotype frequencies, gene diversity, and Hardy-Weinberg equilibrium) at the three *loci* were estimated with Arlequin v2.000 (Schneider *et al.*, 2000). In the present study haplotypes were inferred using a maximum likelihood approach. *Linkage disequilibrium* (D) and relative LD (D') between two alleles at two different

loci and their level of significance (P) was calculated according to Weir (1996). Ewes-Watterson's, Slatkin's and Chakraborty's selective neutrality statistical tests were applied to examine the presence of selective forces influencing allelic diversity at each *locus*.

The Cabo Verde and Guiné-Bissau data was compared to several populations which are available at the same typing resolution (Ellis *et al.*, 2000; Pimtanonthai *et al.*, 2001; Sanchez-Velasco *et al.*, 2003; Ayed *et al.*, 2004; Middleton *et al.* 2004)¹. An analysis of molecular variance (AMOVA) was performed with these populations based on Euclidean distances between all pairs of haplotypes (Excoffier *et al.*, 1992). The total genetic variation between Cabo Verde Northwestern and Cabo Verde Southeastern was estimated and the correspondent *Fst* value was used to evaluate if there was significant difference between them. Variance components were tested for significance by nonparametric randomisation tests using 10000 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 Cabo Verde and Guiné-Bissau data set with other populations available in the literature was achieved using the software included in the PHYLIP v.3.6 package (Felsenstein, 2004). First, SEQBOOT was used to perform a bootstrap analysis from gene frequency data. The program generates multiple data sets resampled 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 CONSENSE.

¹ www.allelefrequencies.net

The topology was visualized with TreeView (Page, 1996). Dendrograms were based on HLA-A, -B and -DRB1 or on HLA-A and -B, in order to include some relevant populations with no high resolution typing for HLA-DRB1 *loci*.

Principal coordinates analysis using HLA-A and HLA-B allele frequencies was carried out on the MultiVariate Statistical Package MVSP3 for Windows (Kovach Computing Services, Anglesey, Wales, UK)¹.

6.4. Results

Tables 33, 34, and 35 show the allele frequencies of HLA-A, HLA-B and HLA-DRB1 *loci*, respectively, in Cabo Verde (Cabo Verde total, CVT; Cabo Verde Northwestern, CVNW; Cabo Verde Southeastern, CVSE) and Guiné-Bissau (GB) populations. A total of 31 HLA-A, 45 HLA-B and 32 HLA-DRB1 alleles were found in Cabo Verde and 20 HLA-A, 31 HLA-B and 23 HLA-DRB1 alleles were found in Guiné-Bissau. The populations were in Hardy-Weinberg equilibrium at each *locus*, except HLA-DRB1 in Cabo Verde. The heterozygosity was high for all examined *loci* (Cabo Verde HLA-A: 0.94; HLA-B: 0.97; HLA-DRB1: 0.95; Guiné-Bissau HLA-A: 0.86; HLA-B: 0.94; HLA-DRB1: 0.89). Ewes-Watterson, Slatkin's and Chakraborty's tests of selective neutrality yielded significant results at HLA-A and HLA-DRB1 in Cabo Verde and HLA-DRB1 in Guiné-Bissau suggesting that selection was acting.

An AMOVA between the two sub-populations of Cabo Verde (Northwestern and Southeastern) shows that only 0.13% of the total genetic variation found can be attributed to differences among them, the

¹ <http://www.kovcomp.com/mvsp>

remaining is due to within sub-population differences. The exact test of population differentiation performed by Arlequin show significant results between Cabo Verde Northwestern and Cabo Verde Southeastern groups of islands.

HLA-A Alleles	Frequencies			
	Guiné-B	CV(T)	CV(NW)	CV(SE)
A*010101	0.046	0.049	0.048	0.048
A*0102	0.008	0.008	0	0.016
A*020101	0.167	0.135	0.149	0.121
A*0202	0.062	0.016	0.008	0.024
A*0205	0	0.004	0	0.008
A*0217	0	0.004	0	0.008
A*030101	0.062	0.089	0.113	0.065
A*030103	0	0.004	0.008	0
A*110101	0	0.024	0.024	0.024
A*2301	0.156	0.113	0.081	0.148
A*240201	0.015	0.069	0.048	0.089
A*240301	0	0.008	0.008	0.008
A*2601	0.031	0.040	0.048	0.032
A*2608	0	0.004	0.008	0
A*2902	0.015	0.032	0.016	0.048
A*3001	0.038	0.044	0.056	0.032
A*3002	0.031	0.093	0.121	0.065
A*310102	0	0.020	0.032	0.008
A*3201	0.038	0.044	0.048	0.040
A*3301	0.046	0.020	0.032	0.008
A*3303	0.085	0.032	0.024	0.040
A*3402	0.054	0.004	0	0.008
A*6601	0.015	0.016	0.024	0.008
A*680101	0.023	0.024	0.040	0.008
A*680102	0	0.004	0.008	0
A*6802	0.023	0.028	0.008	0.048
A*680301	0	0.004	0.008	0
A*6815	0	0.004	0	0.008
A*6901	0	0.032	0.040	0.024
A*7401	0.077	0.028	0	0.056
A*8001	0.008	0.004	0	0.008

Table 33. HLA-A allele frequencies in Guiné-Bissau and Cabo Verde populations estimated by maximum likelihood. CV(T)- Cabo Verde Total; CV(NW)- Cabo Verde Northwestern; CV(SE)- Cabo Verde Southeastern.

HLA-B Alleles	Frequencies			
	Guiné-B	CV(T)	CV(NW)	CV(SE)
B*070201	0.023	0.040	0.024	0.056
B*0705/06	0.023	0.012	0.008	0.016
B*0801	0.077	0.065	0.056	0.073
B*1302	0.023	0	0	0
B*1303	0	0.004	0	0.008
B*1402	0.015	0.065	0.032	0.097
B*150101	0	0.044	0.048	0.040
B*1503	0.108	0.056	0.048	0.065
B*1510	0.038	0.004	0.008	0
B*1515	0	0.004	0.008	0
B*1516	0	0.004	0	0.008
B*151701	0	0.004	0	0.008
B*1518	0.008	0.008	0.008	0.008
B*180101	0.038	0.036	0.040	0.032
B*2702	0	0.008	0.008	0.008
B*2703	0.015	0.004	0	0.008
B*2705	0	0.020	0.024	0.016
B*350101	0.144	0.132	0.174	0.089
B*3502	0	0.004	0	0.008
B*3503	0	0.008	0.008	0.008
B*3701	0.015	0.008	0.016	0
B*3801	0	0.004	0	0.008
B*3910	0.008	0	0	0
B*4001	0	0.004	0.008	0
B*4002	0.008	0.016	0.024	0.008
B*4006	0	0.004	0.008	0
B*4101	0	0.008	0	0.016
B*4102	0.008	0.004	0.008	0
B*4103	0.031	0	0	0
B*4201	0.015	0.024	0.032	0.016
B*4202	0.008	0.004	0	0.008
B*440201	0.008	0.004	0	0.008
B*440301	0.031	0.028	0.016	0.040
B*4405	0	0.004	0	0.008
B*4501	0.062	0.012	0.008	0.016
B*4701	0	0.008	0.008	0.008
B*4901	0.031	0.032	0.024	0.040
B*5001	0.015	0.012	0.008	0.016
B*510101	0.015	0.077	0.081	0.073
B*510102	0	0.004	0.008	0
B*5108	0	0.004	0	0.008
B*520101	0	0.004	0.008	0
B*520102	0	0.004	0.008	0
B*5301	0.10	0.097	0.089	0.107
B*5601	0.008	0	0	0
B*570101	0	0.016	0.032	0
B*5702	0.008	0	0	0
B*570301	0.008	0.028	0.032	0.024
B*5801	0.078	0.036	0.032	0.040
B*5802	0	0.032	0.056	0.008
B*7801	0.023	0	0	0
B*8201	0.008	0	0	0

Table 34. HLA-B allele frequencies in Guiné-Bissau and Cabo Verde populations estimated by maximum likelihood. CV(T)- Cabo Verde Total; CV(NW)- Cabo Verde Northwestern; CV(SE)- Cabo Verde Southeastern.

HLA-DRB1 Alleles	Frequencies			
	Guiné-B	CV(T)	CV-NW	CV-SE
DRB1*0101	0	0.048	0.081	0.016
DRB1*010201	0.023	0.052	0.016	0.089
DRB1*0103	0	0.004	0	0.008
DRB1*030101	0.046	0.081	0.056	0.105
DRB1*030201	0	0.048	0.073	0.024
DRB1*0303	0.008	0.008	0.008	0.008
DRB1*0307	0.008	0	0	0
DRB1*040101	0	0.02	0.024	0.016
DRB1*0402	0	0.008	0.016	0
DRB1*040301	0	0.008	0.016	0
DRB1*0404	0	0.008	0	0.016
DRB1*040501	0.077	0.040	0.040	0.040
DRB1*070101	0.062	0.069	0.065	0.073
DRB1*080101	0	0.008	0.008	0.008
DRB1*080302	0	0.004	0	0.008
DRB1*080401	0.008	0.008	0.008	0.008
DRB1*0806	0.008	0.008	0.016	0
DRB1*090102	0.077	0.060	0.032	0.089
DRB1*1001	0.10	0.060	0.081	0.048
DRB1*110101	0.108	0.101	0.106	0.089
DRB1*1102	0.077	0.028	0.016	0.040
DRB1*110401	0.008	0.016	0.008	0.024
DRB1*1127	0	0.008	0.016	0
DRB1*120101	0.008	0.016	0.024	0.008
DRB1*130101	0.031	0.069	0.073	0.065
DRB1*130201	0.077	0.088	0.089	0.081
DRB1*130301	0	0.012	0.016	0.008
DRB1*130302	0	0.004	0.008	0
DRB1*1304	0.196	0.048	0.024	0.073
DRB1*1317	0.008	0	0	0
DRB1*140101	0.015	0.012	0.024	0
DRB1*140701	0.008	0	0	0
DRB1*1412	0.008	0	0	0
DRB1*150101	0.008	0.040	0.032	0.048
DRB1*150201	0	0.012	0.024	0
DRB1*160201	0.031	0.004	0	0.008

Table 35. HLA-DRB1 allele frequencies in Guiné-Bissau and Cabo Verde populations estimated by maximum likelihood. CV(T)- Cabo Verde Total; CV(NW)- Cabo Verde Northwestern; CV(SE)- Cabo Verde Southeastern.

6.4.1. Guiné-Bissau allele frequencies

The most frequent HLA-A alleles found in Guiné-Bissau was A*020101 (16.7%) and A*2301 (15.6%). Two alleles follow with frequencies ranging between 7-9% (A*3303 and A*7401). HLA-A*3303 shows a higher frequency than other sub-Saharanans, except Mali (9.4%). HLA-A*7401 is a typical African allele that present similar frequencies in other sub-Saharan populations. Alleles A*6801 and A*6802 present the same frequency (2.3% each) in opposition to other sub-Saharan that have a higher prevalence of the second (Middleton *et al.* 2004)¹.

HLA-B *locus* present the allele B*350101 (14.4%) as the most frequent in Guiné-Bissau, a prevalence higher than in other sub-Saharanans. The next most frequent alleles found in Guiné were B*1503 (10.8%) and B*5301 (10%). HLA-B*1503 is a typical sub-Saharan allele and shows one of the highest frequencies found in sub-Saharan. HLA-B*5301 present similar frequencies in other sub-Saharanans (Middleton *et al.* 2004). HLA-B*4103 present a relatively high frequency (3.1%) as it's a world very rare allele and it was absent in sub-Saharan populations already studied (Middleton *et al.* 2004).

Allele DRB1*1304 (19.6%) was the most frequent at HLA-DRB1 *locus*, but is surprisingly absent in the Cameroon and residual in Morocco (0.5%) (Pimtanonthai *et al.*, 2001; Middleton *et al.* 2004). The next most frequent alleles were DRB1*110101 (10.8%) and DRB1*1001 (10%), with frequencies much higher than in Cameroon (4.8% and 1.1% respectively) (Pimtanonthai *et al.*, 2001).

¹ www.allelefrequencies.net

6.4.2. Cabo Verde allele frequencies

The most frequent HLA-A alleles found in the Cabo Verde was A*020101 (13.5%) and A*2301 (11.3%), typical frequencies in sub-Saharan populations (Middleton *et al.* 2004). Three alleles follow with frequencies ranging between 7-9% (A*240201, A*030101 and A*3002). HLA-A*240201 (6.9%) shows a higher frequency than other sub-Saharans and similar to Europeans (Middleton *et al.* 2004). HLA-A*030101 (8.9%) presents an intermediate frequency between sub-Saharan populations (3-8%) and Europeans (8-20%) (Middleton *et al.* 2004). HLA-A*3002 (9.3%) has a prevalence in the extended range found in sub-Saharan (3-23%). Alleles A*6801 and A*6802 present the same frequency (2.8% each) in opposition to other sub-Saharan that have a higher prevalence of A*6802 (Middleton *et al.* 2004).

Allele B*350101 (13.2%) is the most frequent at HLA-B *locus* in Cabo Verde, a prevalence higher than other sub-Saharans except Mali which is similar (Middleton *et al.* 2004). The next most frequent alleles found in Cabo Verde were B*5301 (9.7%) and B*510101 (7.7%), higher and similar to sub-Saharans respectively (Middleton *et al.* 2004).

The DRB1*110101 (10.1%), with a frequency similar to Guiné-Bissau, was the most frequent HLA-DRB1 allele in Cabo Verde, followed by DRB1*130201 (8.8%), similar to Guiné-Bissau, and DRB1*030101 (8.1%), similar to Cameroon (Pimtanothai *et al.*, 2001).

6.4.3. Haplotype frequencies

The exact test of *linkage disequilibrium* between the 3 pairs of *loci* was statistically significant in Guiné-Bissau and Cabo Verde. The most

representative (frequency >1%) 3-*loci* haplotypes are listed in Table 36. Tables 37-39 and 40-42 show the most common two *loci* haplotypes with statistical significance *linkage disequilibrium* found in Guiné-Bissau and Cabo Verde respectively.

Haplotypes	Frequencies			
	Guiné-B	CV(T)	CV(N)	CV(SE)
A*010101-B*350101-DRB1*030101	0	0.008	0	0.016
A*010101-B*3701-DRB1*070101	0	0.004	0.016	0
A*010101-B*440301-DRB1*070101	0	0.008	0	0.016
A*010101-B*5801-DRB1*1304	0.015	0	0	0
A*0102-B*4901-DRB1*040501	0	0.008	0	0.016
A*020101-B*0705/06-DRB1*1304	0.015	0	0	0
A*020101-B*1503-DRB1*110101	0	0.012	0	0.016
A*020101-B*2705-DRB1*0101	0	0	0.016	0
A*020101-B*350101-DRB1*110101	0	0	0.016	0
A*020101-B*350101-DRB1*1102	0.031	0	0	0
A*020101-B*350101-DRB1*130201	0.023	0	0	0
A*020101-B*440301-DRB1*160201	0.023	0	0	0
A*020101-B*4901-DRB1*090102	0	0.008	0	0.016
A*020101-B*510101-DRB1*130101	0	0.008	0	0.016
A*020101-B*570101-DRB1*070101	0	0.008	0.016	0
A*0202-B*2703-DRB1*1304	0.015	0	0	0
A*0202-B*350101-DRB1*1001	0.023	0	0	0
A*0202-B*350101-DRB1*1102	0	0.008	0	0.016
A*030101-B*0801-DRB1*1304	0	0.012	0.016	0
A*030101-B*350101-DRB1*040501	0.015	0	0	0
A*030101-B*4101-DRB1*070101	0	0.008	0	0.016
A*030101-B*5802-DRB1*030101	0	0.023	0.040	0
A*110101-B*350101-DRB1*0101	0	0.008	0.016	0
A*2301-B*070201-DRB1*150101	0	0.008	0	0.016
A*2301-B*1402-DRB1*0806	0	0.008	0.016	0
A*2301-B*1402-DRB1*110101	0	0.012	0	0.024
A*2301-B*1503-DRB1*110101	0.046	0.004	0	0
A*2301-B*4901-DRB1*130201	0	0.012	0.016	0
A*2301-B*5301-DRB1*090102	0.015	0.016	0	0.032
A*2301-B*5801-DRB1*090102	0	0.012	0.016	0.008
A*2301-B*5801-DRB1*110101	0.015	0	0	0
A*240201-B*070201-DRB1*150101	0	0.008	0	0.016
A*240201-B*2705-DRB1*040501	0	0.008	0	0.016
A*240201-B*510101-DRB1*1102	0	0.008	0	0.016
A*2601-B*0801-DRB1*1304	0.015	0	0	0
A*2601-B*5801-DRB1*1304	0	0.012	0	0.016
A*2902-B*1402-DRB1*010201	0	0.012	0	0.016
A*2902-B*4501-DRB1*110101	0.015	0	0	0
A*3001-B*510101-DRB1*110101	0	0.008	0.016	0
A*3001-B*570301-DRB1*070101	0	0.008	0.016	0
A*3002-B*0801-DRB1*040101	0	0.012	0.024	0
A*3002-B*0801-DRB1*1304	0.015	0	0	0
A*3002-B*180101-DRB1*030101	0	0.012	0	0.024
A*3002-B*180101-DRB1*030201	0	0.008	0.016	0
A*3002-B*350101-DRB1*1001	0	0.028	0.056	0
A*3201-B*070201-DRB1*120101	0	0.008	0.016	0
A*3301-B*1503-DRB1*1001	0	0.012	0.016	0
A*3301-B*5301-DRB1*090102	0.015	0	0	0
A*3303-B*070201-DRB1*1102	0.015	0	0	0
A*3303-B*1510-DRB1*1001	0.015	0	0	0
A*3402-B*180101-DRB1*040501	0.015	0	0	0
A*6601-B*4501-DRB1*1001	0.015	0	0	0
A*680101-B*350101-DRB1*040501	0	0	0.016	0
A*6901-B*150101-DRB1*130201	0	0.024	0.032	0.016
A*7401-B*0801-DRB1*010201	0	0.004	0	0
A*7401-B*350101-DRB1*1001	0.015	0	0	0
A*7401-B*4901-DRB1*1304	0.015	0	0	0
A*7401-B*5301-DRB1*1304	0.015	0.008	0	0.016

Table 36. Most common HLA-A, -B, -DRB1 three-*loci* haplotypes in the Cabo Verde and Guiné-Bissau populations as estimated by maximum likelihood. CV(T)- Cabo Verde Total; CV(NW)- Cabo Verde Northwestern; CV(SE)- Cabo Verde Southeastern.

Haplotypes	Frequency	s.d.	D	D'	P
A*010101-B*5801	0.015	0.01	0.012	0.28	*
A*020101-B*440301	0.023	0.01	0.018	0.425	*
A*020101-B*4501	0.03	0.01	0.02	0.468	*
A*0202-B*2703	0.015	0.011	0.014	0.343	**
A*0202-B*5801	0.023	0.016	0.018	0.433	*
A*030101-B*350101	0.031	0.015	0.022	0.518	*
A*2301-B*0705/06	0.015	0.012	0.012	0.278	*
A*2301-B*1503	0.059	0.018	0.042	1	**
A*2601-B*0801	0.015	0.01	0.012	0.295	*
A*2902-B*4501	0.015	0.013	0.014	0.343	**
A*3001-B*070201	0.015	0.007	0.015	0.344	**
A*3001-B*4201	0.015	0.012	0.015	0.351	**
A*3303-B*5001	0.015	0.01	0.014	0.335	**
A*3303-B*5301	0.023	0.017	0.015	0.346	*
A*3402-B*180101	0.015	0.007	0.013	0.316	**
A*3402-B*4103	0.015	0.01	0.014	0.325	**
A*6601-B*4501	0.015	0.004	0.014	0.343	**
A*680101-B*1402	0.015	0.014	0.015	0.357	**
A*7401-B*350101	0.031	0.021	0.02	0.467	*

Table 37. Most common HLA-A, -B two-*loci* haplotypes in the Guiné-Bissau population as estimated by maximum likelihood (only frequencies above 0.01 with statistical significance *linkage disequilibrium* are shown). *P<0.05, ** P<0.001.

Haplotypes	Frequency	s.d.	D	D'	P
A*020101-DRB1*1102	0.046	0.02	0.033	0.636	**
A*020101-DRB1*140101	0.015	0.008	0.013	0.246	*
A*020101-DRB1*160201	0.023	0.017	0.018	0.342	*
A*0202-DRB1*1001	0.023	0.013	0.017	0.322	*
A*0202-DRB1*1304	0.031	0.011	0.019	0.355	*
A*030101-DRB1*1304	0.031	0.018	0.019	0.355	*
A*2301-DRB1*110101	0.069	0.024	0.052	1	**
A*2601-DRB1*090102	0.015	0.018	0.013	0.248	*
A*2902-DRB1*110101	0.015	0.012	0.014	0.263	**
A*3201-DRB1*040501	0.015	0.008	0.012	0.238	*
A*3301-DRB1*090102	0.015	0.009	0.012	0.226	*
A*3301-DRB1*1001	0.023	0.009	0.018	0.353	**
A*3303-DRB1*070101	0.023	0.009	0.018	0.34	*
A*3402-DRB1*040501	0.023	0.013	0.019	0.361	**
A*6601-DRB1*1001	0.015	0.014	0.014	0.265	**
A*6802-DRB1*1304	0.015	0.008	0.011	0.208	*

Table 38. Most common HLA-A, -DRB1 two-*loci* haplotypes in the Guiné-Bissau population as estimated by maximum likelihood (only frequencies above 0.01 with statistical significance *linkage disequilibrium* are shown). *P<0.05, ** P<0.001.

Haplotypes	Frequency	s.d.	D	D'	P
B*070201-DRB1*070101	0.015	0.01	0.014	0.329	**
B*0705/06-DRB1*1304	0.023	0.016	0.019	0.437	**
B*0801-DRB1*1304	0.054	0.026	0.039	0.913	**
B*1503-DRB1*010201	0.015	0.008	0.013	0.304	*
B*1503-DRB1*110101	0.038	0.023	0.027	0.631	*
B*1503-DRB1*130101	0.015	0.009	0.012	0.283	*
B*1510-DRB1*1001	0.023	0.015	0.019	0.454	**
B*180101-DRB1*030101	0.015	0.012	0.014	0.321	**
B*2703-DRB1*1304	0.015	0.012	0.012	0.293	*
B*350101-DRB1*1001	0.044	0.032	0.03	0.695	*
B*350101-DRB1*1102	0.054	0.023	0.042	1	**
B*3701-DRB1*070101	0.015	0.006	0.014	0.34	**
B*4103-DRB1*040501	0.015	0.015	0.013	0.306	*
B*440301-DRB1*160201	0.023	0.021	0.022	0.521	**
B*4501-DRB1*110101	0.03	0.015	0.024	0.56	**
B*4901-DRB1*1304	0.023	0.012	0.017	0.4	*
B*5301-DRB1*090102	0.046	0.016	0.038	0.905	**

Table 39. Most common HLA-B, -DRB1 two-*loci* haplotypes in the Guiné-Bissau population as estimated by maximum likelihood (only frequencies above 0.01 with statistical significance *linkage disequilibrium* are shown). *P<0.05, ** P<0.001.

Haplotypes	Frequency	s.d.	D	D'	P
A*010101-B*0801	0.012	0.011	0.009	0.288	*
A*020101-B*4501	0.012	0.009	0.01	0.338	**
A*020101-B*510101	0.028	0.011	0.018	0.575	*
A*030101-B*5802	0.032	0.009	0.029	0.949	**
A*110101-B*350101	0.016	0.004	0.013	0.418	**
A*2301-B*1402	0.024	0.016	0.017	0.542	**
A*2301-B*4901	0.016	0.005	0.013	0.404	**
A*2301-B*5301	0.024	0.01	0.013	0.427	*
A*2301-B*5801	0.02	0.014	0.016	0.519	**
A*240201-B*510101	0.024	0.013	0.019	0.601	**
A*2601-B*5801	0.012	0.005	0.011	0.344	**
A*3001-B*4201	0.012	0.005	0.011	0.348	**
A*3001-B*570301	0.012	0.009	0.011	0.351	**
A*3002-B*0801	0.02	0.008	0.014	0.455	*
A*3002-B*180101	0.012	0.009	0.009	0.282	*
A*3002-B*350101	0.043	0.013	0.031	1	**
A*3301-B*1503	0.012	0.005	0.011	0.354	**
A*6601-B*5301	0.016	0.009	0.015	0.47	**
A*6802-B*5301	0.012	0.008	0.009	0.303	*
A*6901-B*150101	0.024	0.007	0.023	0.735	**

Table 40. Most common HLA-A, -B two-*loci* haplotypes in the Cabo Verde population as estimated by maximum likelihood (only frequencies above 0.01 with statistical significance *linkage disequilibrium* are shown). *P<0.05, ** P<0.001.

Haplotypes	Frequency	s.d.	D	D'	P
A*010101-DRB1*070101	0.012	0.008	0.009	0.284	*
A*020101-DRB1*040501	0.019	0.008	0.014	0.447	*
A*020101-DRB1*070101	0.024	0.015	0.015	0.484	*
A*020101-DRB1*130101	0.02	0.008	0.011	0.353	*
A*030101-DRB1*030101	0.032	0.013	0.025	0.816	**
A*2301-DRB1*090102	0.03	0.01	0.023	0.756	**
A*2301-DRB1*110101	0.026	0.01	0.015	0.49	*
A*240201-DRB1*110401	0.012	0.006	0.011	0.358	**
A*2601-DRB1*030201	0.012	0.008	0.01	0.314	**
A*2601-DRB1*1304	0.02	0.009	0.018	0.594	**
A*3001-DRB1*070101	0.016	0.007	0.013	0.412	**
A*3002-DRB1*1001	0.036	0.014	0.031	1	**
A*3002-DRB1*130101	0.016	0.007	0.01	0.316	*
A*3201-DRB1*010201	0.012	0.004	0.01	0.32	**
A*3201-DRB1*120101/06	0.012	0.007	0.011	0.371	**
A*3301-DRB1*1001	0.012	0.006	0.011	0.355	**
A*3303-DRB1*090102	0.014	0.007	0.012	0.405	**
A*6901-DRB1*130201	0.032	0.011	0.029	0.959	**
A*7401-DRB1*010201	0.012	0.009	0.011	0.347	**

Table 41. Most common HLA-A, -DRB1 two-*loci* haplotypes in the Cabo Verde population as estimated by maximum likelihood (only frequencies above 0.01 with statistical significance *linkage disequilibrium* are shown). *P<0.05, ** P<0.001.

Haplotypes	Frequency	s.d.	D	D'	P
B*070201-DRB1*150101	0.02	0.006	0.019	0.747	**
B*0801-DRB1*010201	0.012	0.007	0.009	0.351	*
B*0801-DRB1*030101	0.014	0.028	0.009	0.346	*
B*0801-DRB1*1304	0.016	0.013	0.013	0.524	**
B*1402-DRB1*010201	0.028	0.01	0.025	1	**
B*150101-DRB1*130201	0.028	0.009	0.024	0.98	**
B*1503-DRB1*1001	0.02	0.015	0.017	0.676	**
B*1503-DRB1*110101	0.024	0.009	0.018	0.74	**
B*180101-DRB1*130101	0.012	0.008	0.01	0.387	*
B*2705-DRB1*040501	0.011	0.004	0.011	0.43	**
B*350101-DRB1*0101	0.024	0.015	0.017	0.694	**
B*350101-DRB1*1001	0.023	0.008	0.015	0.588	*
B*350101-DRB1*110101	0.013	0.005	0.0001	0.006	*
B*350101-DRB1*110401	0.012	0.008	0.01	0.402	**
B*4201-DRB1*030201	0.02	0.005	0.019	0.765	**
B*440301-DRB1*070101	0.016	0.006	0.014	0.571	**
B*4901-DRB1*130201	0.016	0.011	0.013	0.536	**
B*510101-DRB1*110101	0.023	0.009	0.015	0.605	*
B*510101-DRB1*1102	0.012	0.01	0.01	0.4	**
B*5301-DRB1*090102	0.022	0.015	0.016	0.651	**
B*5301-DRB1*1304	0.014	0.007	0.01	0.388	*
B*570301-DRB1*070101	0.012	0.005	0.01	0.49	**
B*5801-DRB1*090102	0.018	0.008	0.016	0.644	**
B*5802-DRB1*030101	0.016	0.006	0.014	0.545	**

Table 42. Most common HLA-B, -DRB1 two-*loci* haplotypes in the Cabo Verde population as estimated by maximum likelihood (only frequencies above 0.01 with statistical significance *linkage disequilibrium* are shown). *P<0.05, ** P<0.001.

A*2301-B*1503-DRB1*110101 (4.6%) was the most frequent three-*loci* haplotype in Guiné-Bissau, also present in Cabo Verde but with a residual frequency (0.4%). Its partial two-*loci* haplotype A*2301-B*1503 (5.9% in Guiné-Bissau) has been found in several sub-Saharan populations (Cao *et al.*, 2004; Middleton *et al.* 2004) and in Cabo Verde, being exclusive to the Southeastern islands (4%). The second most frequent three-*loci* haplotype in Guiné-Bissau A*020101-B*350101-DRB1*1102 (3.1%) is absent in Cabo Verde. The partial A*020101-B*350101 haplotype (5.4% in Guiné-Bissau), also absent in Cabo Verde, has been found in Malians at 2.2% (Middleton *et al.*, 2004). Three other haplotypes (A*020101-B*350101-DRB1*130201; A*020101-B*440301-DRB1*160201; and A*0202-B*350101-DRB1*1001), were found in Guiné-Bissau each at a frequency of 2.3%, but absent in Cabo Verde.

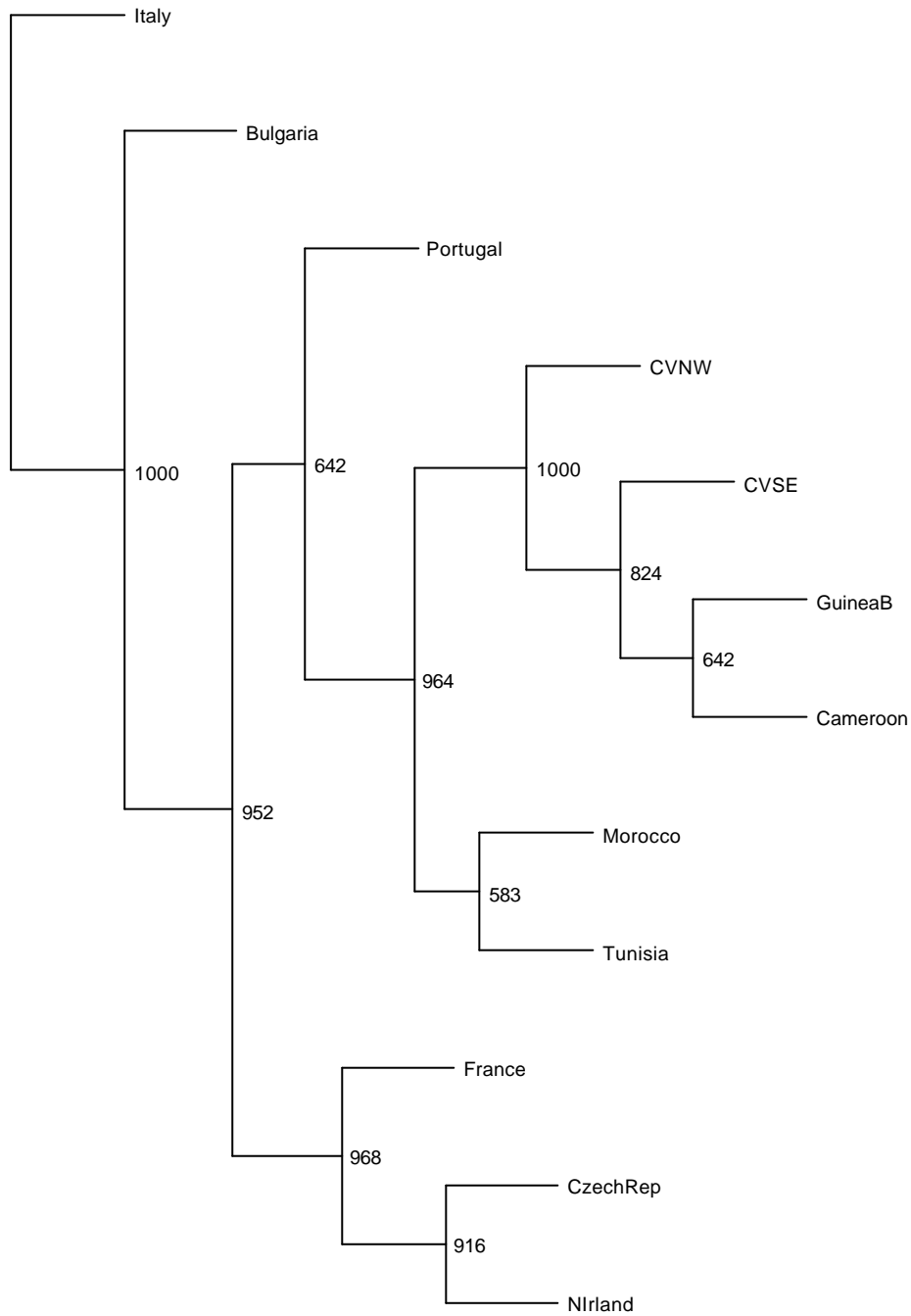
The most frequent three-*loci* haplotype in Cabo Verde was A*3002-B*350101-DRB1*1001 (2.8%), exclusive to Nothwestern islands (5.6%) and absent in Guiné-Bissau. The partial haplotype A*3002-B*350101 (7.3% in Cabo Verde Northwestern and 0.8% in Guiné-Bissau) has not been found in other sub-Saharans (Cao *et al.*, 2004; Middleton *et al.* 2004). The other two most frequent haplotypes in Cabo Verde, A*6901-B*150101-DRB1*130201 (2.4%) and A*030101-B*5802-DRB1*030101 (2.3%), were also absent in Guiné-Bissau. Partial haplotype A*6901-B*150101 was not found in other sub-Saharan and is more prevalent in Northwestern (3.2%) than Southeastern (1.6%) Cabo Verde islands. A*030101-B*5802 is exclusive to Northwestern islands (5.6%) and despite it was not being found in Guiné-Bissau is present in Kenyans, Zulus, and Cameroon's (Ellis *et al.*, 2000; Cao *et al.*, 2004).

From the haplotypes present at plus than 1% in the Cabo Verde islands only two (A*2301-B*5801-DRB1*090102, A*6901-B*150101-DRB1*130201) were present on both islands (Table 36). These two haplotypes are absent in Guiné-Bissau and more prevalent in Northwestern than Southeastern Cabo Verde islands. Only Southeastern islands present haplotypes common to the most frequent found in Guiné-Bissau (A*2301-B*5301-DRB1*090102 and A*7401-B*5301-DRB1*1304).

More than nine percent (9.3%) of all three-*loci* haplotypes found in Cabo Verde are common to Portugal (Spínola *et al.*, 2005) and only 2.2% are common to Guiné-Bissau. No haplotypes common to Guiné-Bissau and Portugal were found (data not shown).

6.4.4. Phylogenetic analyses

Phylogenetic trees constructed with HLA-A, -B, and -DRB1 (Figure 13) or just with class I (HLA-A and -B) allele frequencies (Figure 14), show a close relationship between Guiné-Bissau and Cabo Verde which appear clustered with other sub-Saharan populations. A principal coordinate analysis (Figure 15) is consistent with the dendrograms but shows also that Northwestern Cabo Verde islands are not so distance to North Africans and Europeans than Southeastern islands.



_100

Figure 13. Neighbor-joining (NJ) dendrograms showing the comparative position of Cabo Verde and Guiné-Bissau populations with other populations typed with similar resolution. Standard genetic distances among populations were calculated using HLA-A, -B and -DRB1 allele frequencies with Cabo Verde Northwestern (CVNW) and Cabo Verde Southeastern (CVSE). Numbers above branches are node support after the bootstrap technique implemented in PHYLIP package program BOOT. For references of the populations used see *data analysis*.

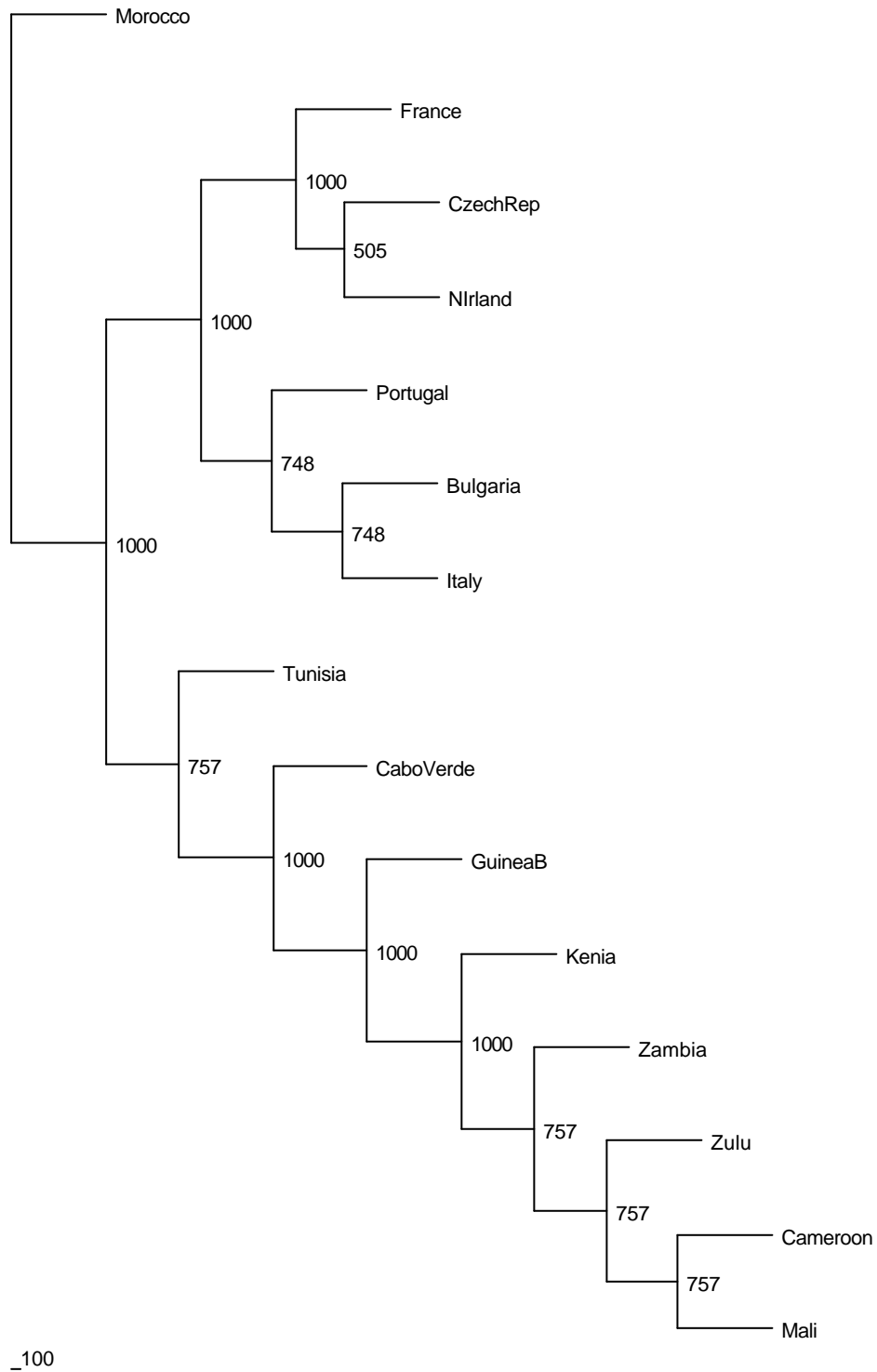


Figure 14. Neighbor-joining (NJ) dendrograms showing the comparative position of Cabo Verde and Guiné-Bissau populations with other populations typed with similar resolution. Standard genetic distances among populations were calculated using HLA-A and -B allele frequencies. Numbers above branches are node support after the bootstrap technique implemented in PHYLIP package program BOOT. For references of the populations used see *data analysis*.

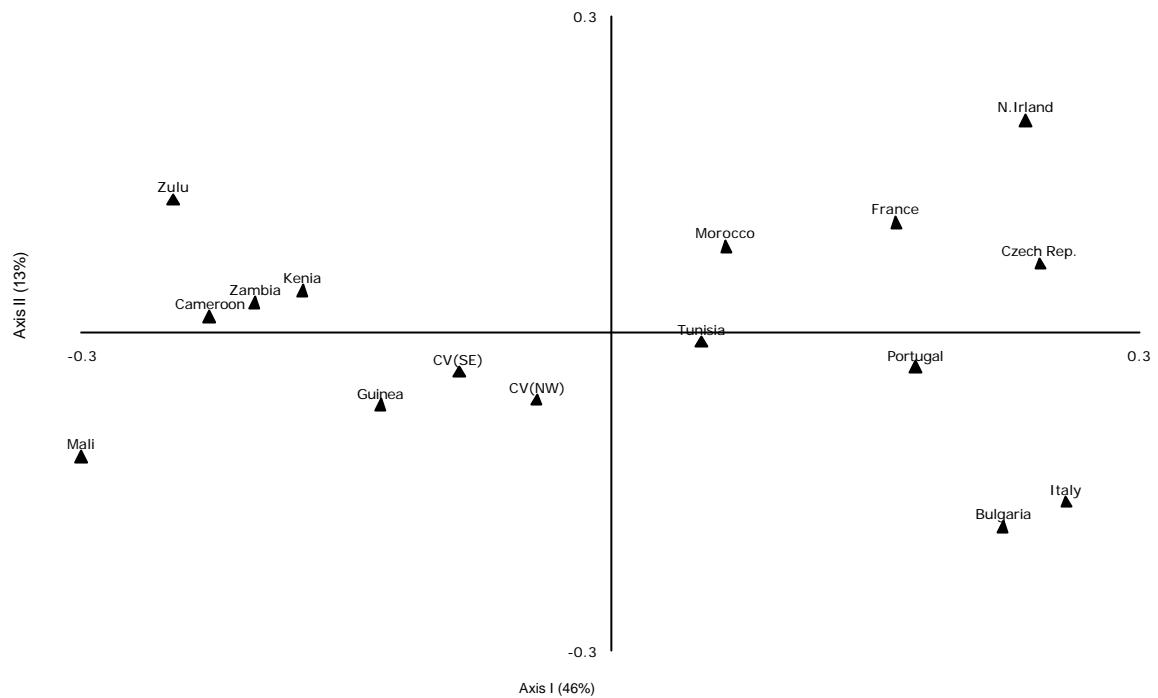


Figure 15. Principal coordinates analysis using HLA-A, and -B allele frequencies. CV(NW)- Cabo Verde Northwestern; CV(SE)- Cabo Verde Southeastern.

6.5. Discussion

Guiné-Bissau was strongly influenced by different historic and pre-historic events that bring together people from different origins in Africa. In different times during several thousands of years this area on the West Coast of Africa received people from North Africa, Sahara and East Africa. The coast of Guinea, known as Senegambia, was the origin of most slave trade in the 15th-16th centuries in which Portugal was involved. African people were captured from the coast of Guinea and taken to Cabo Verde islands, which served as a central outpost in the slave trade to Europe and America continents (Russell-Wood, 1998). Cabo Verde archipelago, after its

discovery by Portuguese in 15th century, was populated with slaves from the West Coast of Africa and some Caucasian, mostly from Iberia.

The present work describes for the first time HLA-A, -B, and –DRB1 high resolution allele and haplotype frequencies in Guiné-Bissau and Cabo Verde populations. This high resolution throughput shows that several HLA generic groups identified before in Cabo Verde (Spínola *et al.*, 2002) were in fact constituted by a variety of alleles (Tables 33-35). This additional information is highly relevant for a detailed HLA structure characterization. Unfortunately, few African populations have been well characterized to HLA allele level, a situation that renders difficult the comparison of the available data sets.

Guiné-Bissau and Cabo Verde position in dendrograms (Figure 13 and 14) reflects clearly their sub-Saharan origin. However, Principal Coordinates Analysis (Figure 15) shows that these two populations are not so distant to North Africans as other sub-Saharans. In Guiné-Bissau, this situation could result from ancient genetic contacts with other peoples, namely North Africans before Sahara desertification and East Africa Arabs through the Sahel Corridor (Ellis *et al.*, 2000). More recent influences such as admixture with Berbers due to Muslim pressure in Northwestern Africa in the 9th century (Moreira, 1964) are possible. A previous mtDNA study in Guiné-Bissau samples showed 6% of its haplogroups characteristic of North Africa, East Africa, Arabia, Middle East and even Europe (Rosa *et al.*, 2004).

The observed similarity between Guiné-Bissau and Cabo Verde is consistent with the well documented origin of the archipelago first settlers. Despite the prevalent sub-Saharan influence in Cabo Verde, we found many

more haplotypes common to Portugal (9.3%) than to Guiné-Bissau (2.2%) denoting a Caucasian contribution to the Cabo Verde settlement.

Principal Coordinate Analysis (Figure 15) shows that the Northwestern group of islands is more similar to Caucasians than Southeastern Cabo Verde islands, in agreement with previous studies based on STR's and Y-Chromosome typings (Fernandes *et al.*, 2003; Gonçalves *et al.*, 2003). In fact, these two groups have significant different HLA allele frequencies and each one shows a specific haplotype profile (Table 36). Southeastern islands have also more common haplotypes to Guiné-Bissau (3%) and Portugal (10%) than Northwestern islands do (1% common to Guiné-Bissau and 7% to Portugal). This reduced number of common haplotypes to Guiné-Bissau could mean that Cabo Verde sub-Saharan settlers do not belong to a single and homogeneous group. The lower number of common haplotypes to Guiné-Bissau and Portugal in Northwestern islands is most probably the result of a founder effect, as these islands were populated later from slaves that fled from the Southeastern group (Gonçalves *et al.*, 2003). Further studies on West African coast populations should clarify the origin of sub-Saharan settlers in Cabo Verde islands.

Capítulo 7

7. HLA genes in Madeira island (Portugal) inferred from sequence-based typing: footprints from different origins.

Hélder Spínola, Jácome Bruges-Armas, Marian Gantes Mora, Derek Middleton, António Brehm.

In press Molecular Immunology

7.1. Abstract

HLA-A, -B, and -DRB1 polymorphisms were examined in Madeira island populations. The data was obtained at high-resolution level, using sequence based typing. The most frequent alleles at each *loci* were: A*020101 (24.6%), B*5101 (9.7%), B*440201 (9.2%), and DRB1*070101 (15.7%). The predominant three *loci* haplotypes in Madeira were A*020101-B*510101-DRB1*130101 (2.7%) and A*010101-B*0801-DRB1*030101 (2.4%), previously found in North and Centre Portugal.

The present study corroborates historical sources and other genetic studies that say Madeira were populated not only by Europeans, mostly Portuguese, but

also sub-Saharan Africans due to slave trade. Comparison with other populations shows that Madeira experienced a stronger African influence due to slave trade, than Portugal mainland and even the Azores archipelago. Despite this African genetic input, haplotype and allele frequencies were predominantly from European origin, mostly common to mainland Portugal.

7.2. Introduction

The Human Leukocyte Antigen (HLA) is one of the most polymorphic genetic systems in humans and consists of a closely linked set of genes. HLA genes are highly important for medical purposes (namely transplantation, autoimmune diseases and allergies), anthropological and forensic fields (Riley and Olerup, 1992; Boehncke *et al.*, 1998; Gibert *et al.*, 2003).

Madeira island (Portugal), located in the North Atlantic Ocean, was uninhabited when the settlement process began by the Portuguese in the 15th century. Settlers were mainly Portuguese but, when the Cabo Verde islands were discovered (1462) and begun to be an outpost for sub-Saharan African slave trade, Madeira was one of the destinies for these African peoples. Most sub-Saharan slaves brought to Madeira were from Northwestern African coast (Carreira 1983; Russell-Wood 1998) and later were also introduced from Angola. In the middle of 16th century African slaves constituted already 10% of the Madeira population and eventually become completely integrated into the society.

Some studies have been done in order to characterize the Madeiran population genetic profile (Fernandes *et al.* 2001, Brehm *et al.* 2003) but only one used HLA generic class I and class II typing (Spínola *et al.* 2002). These

studies showed that, besides the greater influence of Caucasians in the Madeira settlement, sub-Saharanans left an important genetic imprint on the population.

With recent developments in DNA typing, HLA can be studied at a high resolution level, for example by sequencing. This high resolution throughput allows an additional amount of information with greater relevance to a more specific HLA characterization, important in population's comparisons and disease studies (Ayed *et al.* 2004). Previous HLA generic typing revealed that the Madeirean population has an admixture of Caucasoid and to a lesser degree, sub-Saharan alleles and haplotype frequencies (Spínola *et al.* 2002).

The main goal of this work is to present a clear genetic profile of Madeireans based on high resolution typing of HLA-A, -B, and -DRB1 *loci* and contribute to clarify the origin of Madeira population and especially its sub-Saharan component previously reported (Spínola *et al.* 2002).

7.3. Materials and methods

7.3.1. Population samples and HLA typing

The sample population consisted of 185 healthy unrelated males whose parents and grandfathers were born and living in Madeira island. Genomic DNA was isolated from blood using a phenol-chloroform procedure and all subjects were sequence based typed for HLA-A, HLA-B, accordingly to Kurz *et al.* (1999) and Pozzi *et al.* (1999), and HLA-DRB1 with specific primers used in SSOP typing (Williams *et al.* 2004), as previously described (Spínola *et al.* 2005a). We choose to study only one class II HLA *loci*, the HLA-DRB1, since this one is highly polymorphic in opposition to others class II (e.g. HLA-DQA1, HLA-DQB1) that are much more conserved.

7.3.2. Data analysis

Basic genetic parameters at the three *loci* were estimated with Arlequin v2.000 (Schneider *et al.* 2000). Comparative analysis of the Madeira island data set with other populations available in the literature was achieved using the software included in the PHYLIP v.3.6 package (Felsenstein, 2004).

7.4. Results

Table 43 shows the allele frequencies of HLA-A, HLA-B and HLA-DRB1 *loci* in the Madeira population. A total of 33 HLA-A, 42 HLA-B and 37 HLA-DRB1 alleles were found. The population was in Hardy-Weinberg equilibrium for each studied *locus*, except HLA-DRB1. The heterozygosity was high for all examined *loci* (HLA-A: 0.90; HLA-B: 0.95; HLA-DRB1: 0.92).

The exact test of population differentiation performed by Arlequin shows non significant results between the Madeira population and mainland Portugal (Spínola *et al.* 2005a).

7.4.1. Allele frequencies

The most frequent HLA-A alleles found in Madeira island were A*020101 (24.6%) and A*240201 (13%). Five HLA-A alleles considered of African-specific (A*0102, A*0202, A*3402, A*3601, and A*7403) (Cao *et al.* 2004) were found albeit in small frequencies. HLA-B *locus* present the alleles B*5101 (9.7%) and B*440201 (9.2%) as the most frequent in Madeira. Four HLA-B alleles considered of sub-Saharan origin (B*1503, B*1516, B*5703, and B*5802) (Cao *et al.* 2004) were found in Madeira. Alleles DRB1*070101(15.7%), DRB1*030101 (12.2%) and DRB1*130101 (12.7%) were the most frequent at HLA-DRB1 *locus*.

HLA-A	Frequency	HLA-B	Frequency	HLA-DRB1	Frequency
A*010101	0.059	B*070201	0.049	DRB1*0101	0.035
A*0102	0.008	B*0801	0.078	DRB1*010201	0.046
A*020101	0.246	B*1302	0.005	DRB1*0103	0.008
A*020104	0.003	B*1401	0.027	DRB1*030101	0.122
A*0202	0.008	B*1402	0.049	DRB1*040101	0.024
A*0205	0.022	B*150101	0.035	DRB1*0402	0.016
A*0208	0.005	B*1503	0.022	DRB1*040301	0.005
A*030101	0.070	B*1516	0.003	DRB1*0404	0.014
A*0302	0.003	B*1518	0.003	DRB1*040501	0.024
A*110101	0.059	B*180101	0.062	DRB1*0406	0.003
A*2301	0.062	B*2702	0.005	DRB1*040701	0.008
A*240201	0.130	B*2705	0.011	DRB1*0408	0.003
A*240301	0.011	B*2707	0.003	DRB1*070101	0.157
A*2405	0.003	B*350101	0.038	DRB1*080101	0.030
A*2426	0.003	B*3502	0.030	DRB1*080302	0.003
A*2501	0.005	B*3503	0.019	DRB1*080401	0.005
A*2502	0.003	B*3508	0.011	DRB1*0806	0.003
A*2601	0.035	B*3534	0.005	DRB1*090102	0.011
A*290101	0.003	B*3701	0.005	DRB1*1001	0.008
A*2902	0.062	B*3801	0.038	DRB1*110101	0.046
A*3001	0.003	B*3901	0.014	DRB1*1102	0.008
A*3002	0.014	B*3924	0.005	DRB1*1103	0.011
A*310102	0.022	B*4001	0.022	DRB1*110401	0.055
A*3108	0.003	B*4002	0.027	DRB1*1110	0.003
A*3201	0.049	B*4019	0.003	DRB1*1111	0.003
A*3301	0.030	B*4101	0.003	DRB1*120101	0.011
A*3402	0.003	B*4102	0.008	DRB1*130101	0.127
A*3601	0.003	B*440201	0.092	DRB1*130201	0.046
A*6601	0.011	B*440301	0.054	DRB1*130302	0.016
A*680101	0.032	B*440302	0.003	DRB1*1304	0.003
A*680102	0.019	B*4404	0.003	DRB1*1305	0.003
A*6802	0.008	B*4501	0.016	DRB1*1309	0.003
A*7403	0.003	B*4901	0.059	DRB1*1310	0.005
		B*5001	0.032	DRB1*140101	0.019
		B*5002	0.008	DRB1*150101	0.089
		B*510101	0.097	DRB1*150201	0.005
		B*520101	0.005	DRB1*160101	0.022
		B*5301	0.011		
		B*570101	0.016		
		B*570301	0.005		
		B*5801	0.016		
		B*5802	0.003		

Table 43. HLA-A, -B and -DRB1 allele frequencies in Madeira island population.

7.4.2. Haplotype frequencies

The most representative 3-*loci* and 2-*loci* haplotypes are listed in Tables 44-47. A*020101-B*510101-DRB1*130101 (2.7%) was the most frequent 3-*loci* haplotype in Madeira island, found before in North Portugal (1.1%). The second most frequent was A*010101-B*0801-DRB1*030101 (2.4%), considered Pan-European of Celtic origin (Arnaiz-Villena *et al.* 2001a), an haplotype also present in North (2.2%) and Centre (3%) Portugal (Spínola *et al.* 2005a). Most of the frequent 3-*loci* haplotypes found in Madeira island are common to mainland Portugal and none of these were found in Guiné-Bissau or Cabo Verde islands (Spínola *et al.* 2005b).

Haplotypes	Frequency	Possible origin
A*010101-B*0801-DRB1*030101	0.024 a)	European
A*020101-B*0801-DRB1*030101	0.023	?
A*020101-B*180101-DRB1*110401	0.016 a)	Mediterranean
A*020101-B*440201-DRB1*040101	0.011 b)	?
A*020101-B*440201-DRB1*130101	0.016 b)	?
A*020101-B*510101-DRB1*130101	0.027 a)	Ibero-Berber
A*030101-B*070201-DRB1*150101	0.014	North African/Western European
A*240201-B*1401-DRB1*070101	0.016	?
A*240201-B*3502-DRB1*030101	0.014 a)b)	?
A*2902-B*440301-DRB1*070101	0.019 a)	Western European
A*3301-B*1402-DRB1*010201	0.014 a)	Western European
a) Present in mainland Portugal; b) Present in Azores Islands.		

Table 44. Most common (≥ 0.01) HLA-A, -B, -DRB1 three-*loci* haplotypes in Madeira island population as estimated by maximum likelihood. Possible origin accordingly to Sanchez-Velasco *et al.* (2003).

Haplotypes	Frequency	s.d.	D	D'	P
A*020101-B*0801	0.019	0.007	0.00001	0.0003	*
A*020101-B*510101	0.069	0.012	0.045	1	**
A*030101-B*070201	0.022	0.008	0.018	0.406	**
A*110101-B*350101	0.019	0.008	0.017	0.372	**
A*2301-B*180101	0.012	0.007	0.008	0.183	*
A*240201-B*1401	0.016	0.007	0.013	0.284	**
A*240201-B*15010101	0.016	0.008	0.011	0.247	**
A*240201-B*3502	0.016	0.006	0.012	0.275	**
A*2601-B*3801	0.014	0.006	0.012	0.272	**
A*2902-B*1503	0.011	0.005	0.009	0.211	**
A*2902-B*440301	0.032	0.01	0.029	0.647	**
A*3201-B*4002	0.016	0.007	0.015	0.332	**
A*3301-B*1402	0.024	0.008	0.023	0.51	**

Table 45. Most common HLA-A, -B two-*loci* haplotypes in the Madeira island population as estimated by maximum likelihood (only frequencies above 0.01 with statistical significance *linkage disequilibrium* are shown). *P<0.05, ** P<0.001.

Haplotypes	Frequency	s.d.	D	D'	P
A*010101-DRB1*030101	0.018	0.008	0.011	0.535	*
A*010101-DRB1*130101	0.018	0.008	0.01	0.5	*
A*020101-DRB1*0101	0.019	0.008	0.01	0.502	*
A*020101-DRB1*110401	0.029	0.009	0.015	0.753	*
A*020101-DRB1*130101	0.052	0.015	0.021	1	*
A*240201-DRB1*070101	0.037	0.012	0.016	0.788	*
A*2902-DRB1*070101	0.021	0.008	0.011	0.558	*
A*3201-DRB1*130201	0.014	0.006	0.011	0.548	**
A*3201-DRB1*150101	0.014	0.007	0.007	0.354	*
A*3301-DRB1*010201	0.014	0.007	0.012	0.59	**

Table 46. Most common HLA-A, -DRB1 two-*loci* haplotypes in the Madeira island population as estimated by maximum likelihood (only frequencies above 0.01 with statistical significance *linkage disequilibrium* are shown). *P<0.05, ** P<0.001.

Haplotypes	Frequency	s.d.	D	D'	P
B*070201-DRB1*150101	0.019	0.008	0.015	0.229	**
B*0801-DRB1*030101	0.073	0.014	0.063	1	**
B*1401-DRB1*070101	0.023	0.008	0.019	0.3	**
B*1402-DRB1*010201	0.03	0.01	0.027	0.431	**
B*180101-DRB1*110401	0.019	0.007	0.015	0.24	**
B*180101-DRB1*150101	0.016	0.007	0.01	0.163	*
B*350101-DRB1*0101	0.011	0.006	0.009	0.149	**
B*3502-DRB1*030101	0.015	0.006	0.012	0.185	**
B*3801-DRB1*130101	0.022	0.008	0.017	0.265	**
B*4002-DRB1*130101	0.011	0.007	0.007	0.113	*
B*440201-DRB1*040101	0.011	0.006	0.009	0.136	**
B*440201-DRB1*130101	0.026	0.008	0.015	0.231	*
B*440301-DRB1*070101	0.027	0.01	0.018	0.29	**
B*4901-DRB1*040501	0.011	0.006	0.009	0.148	**
B*4901-DRB1*080101	0.011	0.006	0.009	0.142	**
B*4901-DRB1*130201	0.011	0.006	0.008	0.128	*
B*5001-DRB1*070101	0.019	0.007	0.0139	0.219	**
B*510101-DRB1*080101	0.011	0.006	0.008	0.124	*
B*510101-DRB1*130101	0.029	0.01	0.017	0.263	*
B*570101-DRB1*070101	0.011	0.005	0.008	0.131	**

Table 4.7. Most common HLA-B, -DRB1 two-*loci* haplotypes in the Madeira island population as estimated by maximum likelihood (only frequencies above 0.01 with statistical significance *linkage disequilibrium* are shown). *P<0.05, ** P<0.001.

7.5. Discussion

The present study describes for the first time HLA-A, HLA-B, and HLA-DRB1 high resolution allele and haplotype frequencies in Madeira island. This high resolution throughput shows that several HLA generic groups, typed before by a medium resolution technique (Spínola *et al.* 2002), were in fact composed of a high variety of alleles (Table 4.3). This additional and much more complete information is highly relevant for a detailed HLA structure characterization. Comparing our HLA data with other populations (Middleton *et al.*, 2004, Spínola *et al.* 2005a, Spínola *et al.* 2005b, Spínola *et al.* 2005c) reveals that Madeira population have an intermediate position

between European and African populations which should be expected under an hypothesis of a mixed origin of its settlers (Figures 16-18). This analysis shows also that Madeira experienced a stronger sub-Saharan influence, due to slave trade, than mainland Portugal and even the Azores islands. These results are corroborated by the finding in Madeira population of several sub-Saharan African specific alleles, albeit in low frequencies (e.g. B*1503, B*1516, B*5703, B*5802, A*0102, A*0202, A*3402, A*3601, and A*7403).

Looking at the haplotype content prevailing in Madeira island we do not find at the 3-*loci* level common characteristics with the already typed sub-Saharan populations. The haplotype structure clearly prints to a mostly European background, mostly of Portuguese origin but also reflecting other Caucasian affinities.

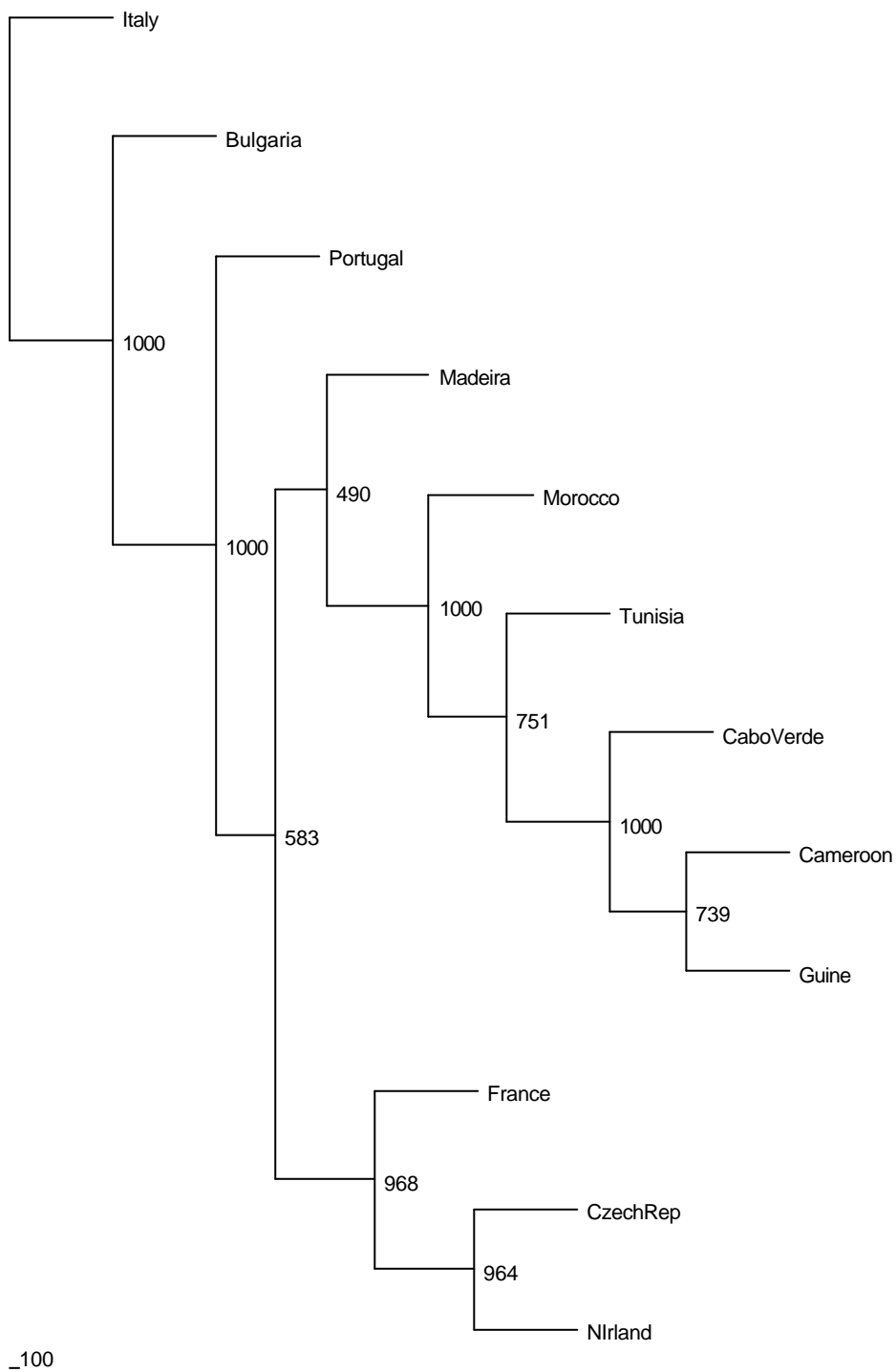


Figure 16. Neighbor-joining (NJ) dendrograms showing the comparative position of Madeira population with other populations typed with similar resolution. Standard genetic distances among populations were calculated using HLA-A, -B and -DRB1 allele frequencies. Numbers above branches are node support after the bootstrap technique implemented in PHYLIP package program BOOT. For references of the populations used see *data analysis*.

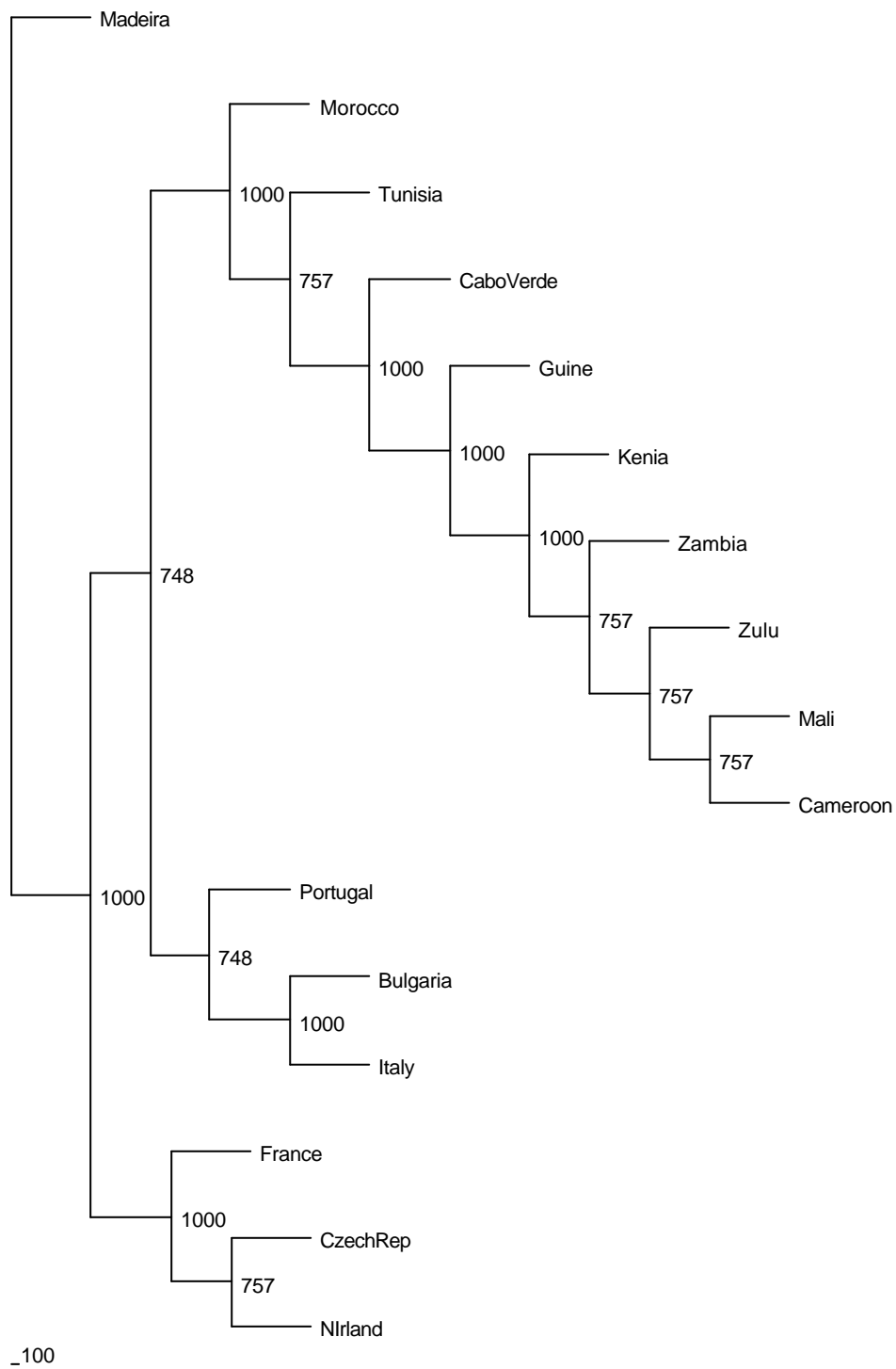


Figure 17. Neighbor-joining (NJ) dendrograms showing the comparative position of Madeira population with other populations typed with similar resolution. Standard genetic distances among populations were calculated using HLA-A and -B allele frequencies. Numbers above branches are node support after the bootstrap technique implemented in PHYLIP package program BOOT. For references of the populations used see *data analysis*.

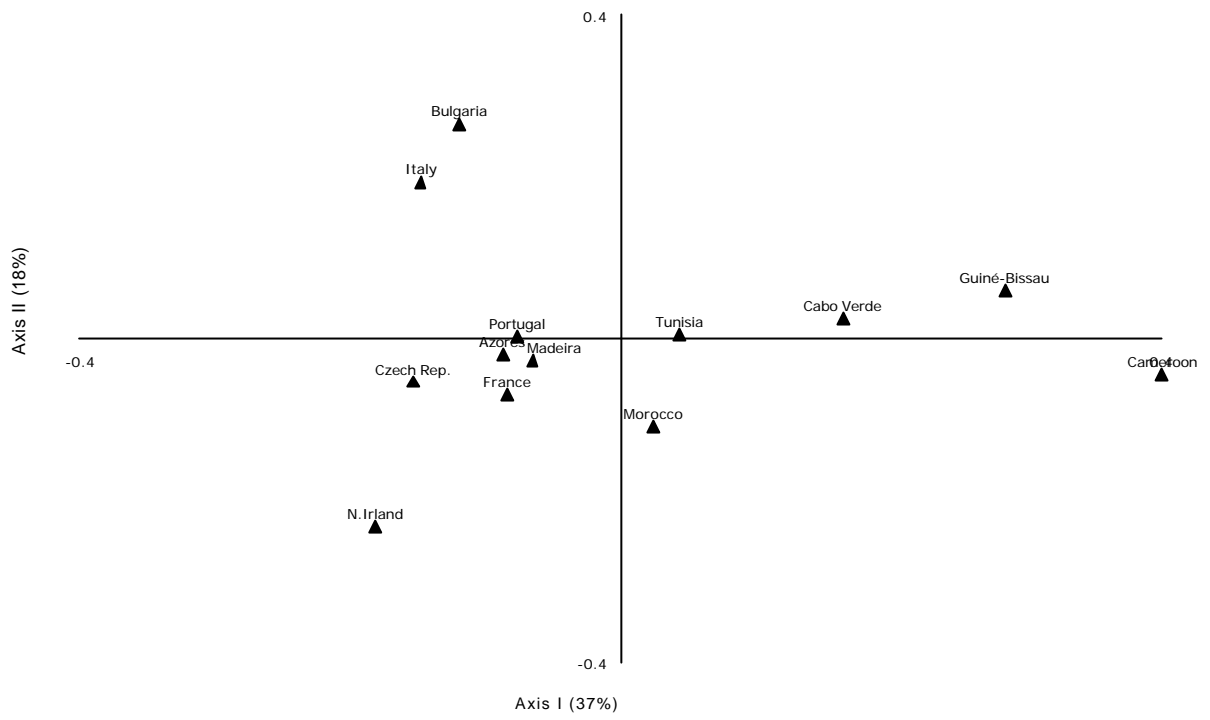


Figure 18. Principal coordinates analysis using HLA-A, -B, and -DRB1 allele frequencies.

Capítulo 8

8. Discrepancies on HLA typing by SSOP and SBT techniques: a case study.

Hélder Spínola, Jácome Bruges-Armas, António Brehm.

(Submitted)

8.1 Abstract

Six hundreds and twenty one samples from Portugal, Cabo Verde archipelago and Guiné-Bissau were SSOP and SBT typed for HLA-A, -B, and –DRB1 in order to characterize and compare results discrepancies between these two methods. Fifty seven alleles (4.6% of all 1242 chromosomes typed) identified by SSOP were not concordant with SBT results. Thirty five (2.8% of total chromosome typed) SSOP miss typing results were discrepancies inside the same allele group and 22 others (1.77% of total chromosome typed) were relative to non concordant results between different groups. SSOP allele miss typing is the result of interpretation difficulties due to less intense, absent or

dubious hybridization patterns. SSOP method procedures are highly exigent on human resources experience and high precision equipment.

8.2 Introduction

The Human Leukocyte Antigen (HLA) is one of the most polymorphic genetic systems in humans and consists of a closely linked set of genes. HLA genes are highly important for medical purposes (namely transplantation, autoimmune diseases and allergies), anthropological and forensic fields (Boehncke *et al.*, 1998; Riley & Olerup, 1992).

HLA typing has experienced a rapid evolution since DNA techniques appeared twenty years ago, especially due to the introduction of the Polymerase Chain Reaction (PCR) method (Mullis e Faloona, 1987). Nowadays, SSP (Sequence Specific Priming) (Bunce *et al.*, 1995), SSOP (Sequence Specific Oligonucleotide Probes) (Middleton, 2000) and SBT (Sequence Based Typing) (Kurz *et al.*, 1999) are the most generalized HLA typing techniques allowing a high resolution throughput. The accuracy of these typing techniques is of much importance since organ and tissue transplantation rejection, particularly bone marrow, depends on it. HLA matching for transplantation if not correctly assigned could be responsible for graft failures (Cicciarelli, 2004). Several comparative studies between serology and DNA techniques have been done in the past years showing big discrepancies that lead to the conclusion that DNA techniques are much more reliable than serology (Mytilineos *et al.*, 1990; Mickelson *et al.*, 1993; Donadi *et al.*, 2000; Ferraz *et al.*, 2002; Borelli *et al.*, 2004). However, DNA typing techniques are not free of errors since they depend on sensible procedures with specific temperatures, good DNA quality, adequate reagents and also a correct results interpretation. Despite the

importance of the reliability of different HLA DNA typing methods, few comparative studies have been performed between these techniques (Smith *et al.*, 1997).

In the present study we compared SSOP and SBT typing results in Portugal, Cabo Verde and Guiné-Bissau populations in order to identify and characterize discrepancies between them.

8.3 Materials and methods

The sample population consisted of 621 healthy unrelated males from different origins (Portugal mainland n=145, Madeira island n=185, Azores archipelago n=102, Cabo Verde archipelago n=124, and Guiné-Bissau n=65). Genomic DNA was isolated from blood using a phenol-chloroform procedure and all subjects were high resolution sequence based typed for HLA-A, HLA-B, accordingly to Kurz *et al.* (1999) and Pozzi *et al.* (1999), and HLA-DRB1 with specific primers used in SSOP typing (Williams *et al.*, 2004), as previously described (Spínola *et al.*, 2005a). The same samples were also DNA-typed for HLA Class I (A and B) and Class II (DRB1) by a medium resolution PCR-SSOP method as previously described (Middleton, 2000). Results obtained in both methods were compared for discrepancies.

8.4 Results and discussion

The allele typing results of 621 samples detected some discrepant results between SSOP and SBT results (Tables 48 and 49). Fifty seven alleles identified by SSOP were not concordant with SBT results, which mean 4.6% of all 1242 chromosomes typed. These 57 miss typing results had a different distribution along the 3 *loci* studied, with 16 cases for HLA-A, 39 for HLA-B, and only 2 for

HLA-DRB1. Thirty five (2.8% of total chromosome typed) SSOP missed typing results were discrepancies inside the same allele group and 22 (1.77% of total chromosome typed) were in different groups. Seven of all non concordant results revealed false homozygous samples, all of them in HLA-B *locus* and showing, by SBT, another allele belonging to a different group of the already identified by SSOP. The incongruence found between SSOP and SBT in this study is the result of the high sensibility of the former (SSOP) technique. These 57 (4.6%) cases of miss HLA allele typing by SSOP, assuming that SBT characterization is the one that is correct, is the result of interpretation difficulties due to less intense, absent or dubious hybridization patterns. These problematic hybridization patterns could be explained with poor quality and quantity DNA amplification and the occurrence of some variations on the high sensitive specific temperatures that probe hybridization needs on SSOP method. The washing membrane process is also another sensitive step that, if not completely achieved, could left behind some "background" that makes difficult the correct interpretation of the hybridization pattern. Otherwise, SSOP is only covering certain nucleotide bases whereas SBT covers the all nucleotide sequence and, in addition, the computer program to analyse de SSOP results would not take account of as many alleles as the SBT analysis program.

SSOP method procedures are highly exigent on human resources experience and high precision equipment. For laboratory and researchers that are give first steps on HLA typing field, SSOP will not be a recommended technique since it will need a high time and equipment investment until achieve a satisfactory quality level. On other way, SBT relay on most automated steps that are very easy to implement and are not sensible to technique and interpretation errors.

HLA discrepancies between SSOP and SBT typing		
SSOP typing	SBT typing	Number of discrepancies
A*23	A*2901	2
A*030102	A*030101	8
A*3203	A*0101	1
A*0226	A*0201	5
B*3520/28	B*3502	2
B*15	B*4002	1
B*40	B*5001	1
B*3520	B*3503	1
-	B*3508	2
B*5107	B*5101	10
B*0804/05	B*0801	3
B*3520/28	B*3508	1
-	B*1401	1
B*5107	B*5201	1
-	B*5101	2
B*40	B*4403	1
B*40	B*4402	1
B*2708	B*2705/13	1
-	B*1402	1
B*50	B*49	1
B*23	B*0301	1
B*35	B*5301	1
B*40	B*5001	2
B*3520	B*3501	2
B*3520	B*3534	2
-	B*3502	1
DRB1*07	DRB1*1301	1
DRB1*11	DRB1*1301	1
	Total	57

Table 48. Discrepancies between SSOP and SBT HLA-A, -B, and -DRB1 typing in 1242 chromosomes (621 samples) from Portugal, Cabo Verde and Guiné-Bissau.

	HLA-A		HLA-B		HLA-DRB1		Total	
	N°	%	N°	%	N°	%	N°	%
No discrepancy	1226	98.7%	1203	96.9%	1240	99.84%	1186	95.5%
Discrepancy	16	1.29%	39	3.14%	2	0.16%	57	4.6%
Discrepancy in the same allele group	13	1.05%	22	1.77%	0	0	35	2.8%
Discrepancy in different allele groups	3	0.24%	17	1.37%	2	0.16%	22	1.77%
False homozygous	0	0	7	—	0	0	7	—

Table 49. Comparisons between SBT and SSOP HLA-A, -B, and -DRB1 typing in 1242 chromosomes (621 samples) from Portugal, Cabo Verde and Guiné-Bissau.

Capítulo 9

9. Resultados globais

9.1. Parâmetros gerais

Um total de 54 alelos do gene HLA-A (15,5% de todos os alelos do *locus*), 78 do gene HLA-B (12,5% de todos os alelos do *locus*) e 55 do gene HLA-DRB1 (14,1% de todos os alelos do *locus*) foram encontrados nas populações em estudo. Os números mais baixos de alelos foram encontrados na Guiné-Bissau e os mais elevados nos Açores, para cada um dos três *loci* estudados. Todas as populações estavam em equilíbrio Hardy-Weinberg para cada um dos *loci* estudados, excepto o HLA-DRB1 na Madeira e em Cabo Verde e o HLA-B nos Açores. A heterozigotia nas populações estudadas era elevada em todos os *loci* variando de 0,86 a 0,94 no HLA-A, de 0,94 a 0,97 no HLA-B e de 0,89 a 0,95 no HLA-DRB1. Os valores mais elevados de heterozigotia em todos os *loci* estudados foram encontrados na população de Cabo Verde e os mais baixos na da Guiné-Bissau (Tabela 51).

As tabelas 4-6, 9-12, 20-22, 33-35 e 43 apresentam as frequências alélicas dos genes HLA-A, HLA-B e HLA-DRB1 nas cinco populações

estudadas (Portugal continental, Madeira, Açores, Cabo Verde e Guiné-Bissau).

Populações	Heterozigotia			Nº alelos			Alelos >80%			E-W			L.des.			Eq. H-W		
	A	B	DRB1	A	B	DRB1	A	B	DRB1	A	B	DRB1	A-B	A-DRB1	B-DRB1	A	B	DRB1
PT	0,89	0,94	0,93	27	43	36	10	15	14	ns	ns	ns	s	ns	s	s	s	s
Az	0,90	0,96	0,94	33	53	40	10	19	15	ns	ns	ns	s	s	s	s	ns	s
M	0,90	0,95	0,92	33	42	37	10	17	12	ns	s	ns	s	s	s	s	s	ns
CV	0,94	0,97	0,95	31	45	32	13	16	13	s	ns	s	s	s	s	s	s	ns
GB	0,86	0,94	0,89	20	31	23	11	14	9	ns	ns	s	s	s	s	s	s	s

Tabela 51. Quadro resumo de parâmetros gerais. PT: Portugal continental; Az: Açores; M: Madeira; CV: Cabo Verde; GB: Guiné Bissau; ns: resultado não significativo; s: resultado significativo; Alelos >80%: N° alelos que compreendem >80% das frequências observadas; E-W: Teste Neutralidade (Ewens-Watterson); L.des.: *Linkage desequilibrium*; Eq. H-W: Equilíbrio Hardy-Weinberg.

A caracterização alélica em alta resolução por SBT detectou alguns erros nos resultados anteriormente obtidos em média resolução por SSOP. Estes casos correspondem a 4,6% (57 casos) dos 1242 cromossomas caracterizados em média resolução por SSOP (tabelas 48 e 49). Estes 57 casos de caracterização alélica errada distribuem-se de uma forma muito díspar entre os três *loci* estudados, com 16 casos no HLA-A, 39 no HLA-B e apenas 2 no HLA-DRB1. Trinta e cinco casos (2,8% de todos os cromossomas caracterizados) foram erros de caracterização dentro do mesmo grupo de alelos e 22 (1,77% de todos os cromossomas caracterizados) foram erros entre grupos diferentes. Com a caracterização dos alelos por SBT foram também identificados 7 casos de falsos

homozigóticos nos dados obtidos por SSOP, todos eles referentes ao *locus* HLA-B. A técnica SBT revelou que os alelos não detectados nestes 7 falsos homozigóticos pertenciam a grupos diferentes dos respectivos alelos já identificados por SSOP.

Os testes de neutralidade selectiva Ewens-Watterson, Slatkin's e Chakraborty's revelaram resultados significativos nos *loci* HLA-A, em Cabo Verde, HLA-B, na Madeira, e HLA-DRB1, em Cabo Verde e Guiné-Bissau, sugerindo que nesses casos a selecção natural estaria a actuar.

Uma AMOVA entre as cinco populações estudadas mostra que apenas 0,37% da variação genética total encontrada pode ser atribuída a diferenças entre elas, sendo a restante devido a diferenças intra-populacionais. O teste de diferenciação realizado pelo programa Arlequin demonstrou a existência de diferenças significativas entre todas as populações excepto entre a Madeira e os Açores e entre a Madeira e Portugal Continental (M/Az: $P=0,89$; M/P: $P=0,65$).

9.2. HLA-A

O alelo HLA-A*0201 foi o mais frequente deste *locus* nas cinco populações estudadas, variando de 13,5% (Cabo Verde) a 26,1% (Portugal continental). Os alelos que se seguem com maiores frequências não coincidem entre as populações Portuguesas (Continente, Madeira e Açores) e as Africanas (Cabo Verde e Guiné-Bissau). Os alelos A*0101, A*0301, A*1101 e A*2402 possuem frequências de 6 a 13% nas três populações Portuguesas. Nas populações Africanas o alelo A*2301 (Guiné-Bissau 15,6%; Cabo Verde 11,3%), os alelos A*3303 e A*7401 (Guiné-Bissau 7-9%) e os alelos A*2402, A*0301 e A*3002 (Cabo Verde 7-9%) são os que se seguem ao mais

predominante. O alelo A*2301 surge em Portugal continental e na Madeira com uma frequência de 6% o que se pode explicar por uma influência Norte Africana ou sub-Sahariana. Por outro lado, o alelo A*2402 tem uma prevalência de 6,9% em Cabo Verde o que resulta provavelmente de influência Caucasiana.

O grupo mais diverso nas populações Portuguesas estudadas foi o A*02 com 4 a 5 alelos. Em Cabo Verde, apesar do grupo A*02 apresentar 4 alelos, o A*68 é o mais polimórfico (5 alelos). Na Guiné-Bissau não foi encontrado nenhum grupo com mais de 2 alelos. Alguns dos grupos polimórficos mostram frequências mais equilibradas entre alguns dos alelos presentes; é o caso do grupo A*68 nas cinco populações estudadas, do grupo A*30 na Guiné-Bissau, Cabo Verde e Portugal continental e do grupo A*33 na Guiné-Bissau e Cabo Verde.

9.3. HLA-B

O alelo HLA-B*5101 foi o mais frequente deste *locus* nas populações Portuguesas estudadas, variando de 9 a 12%. Em Cabo Verde, apesar de não ser o mais frequente, este alelo tem uma prevalência relevante (7,7%), o que pode ser o resultado de uma influência Caucasiana. Em Portugal continental o B*4403 tem a mesma prevalência que o B*5101 (12%). Em Cabo Verde e Guiné-Bissau o alelo mais encontrado foi o B*3501 (13,2% e 14,4% respectivamente). Alguns dos alelos que se seguem com maiores frequências coincidem entre algumas das populações estudadas, nomeadamente o B*180101 em Portugal continental (8,5%) e na Madeira (6,2%), o B*0801 na Guiné-Bissau (7.7%), Cabo Verde (6,5%) e na Madeira (7.8%), o B*4402 na Madeira (9,2%) e nos Açores (6.7%), o B*3501 em Portugal continental

(8,2%) e nos Açores (7,4%) e o B*5301 na Guiné-Bissau (10%) e em Cabo Verde (9,7%). Outros alelos apresentam frequências nesta ordem de grandeza em apenas uma das populações estudadas no presente trabalho, designadamente o B*0702 nos Açores (7,6%), o B*1503 (10,8%), o B*5801 (7,8%), o B*4501 (6,2%) na Guiné-Bissau e o B*1402 (6,5%) em Cabo Verde, este último poderá resultar também de influência Europeia.

O grupo mais diverso nas populações Portuguesas estudadas foi o B*35 com 4 a 7 alelos. O grupo B*15 é o mais polimórfico na Guiné-Bissau (3 alelos) e em Cabo Verde (7 alelos), apresentando em Portugal continental o mesmo número encontrado no B*35 (4 alelos) e uma diversidade relevante nos Açores (6 alelos). Alguns dos grupos polimórficos mostram frequências mais equilibradas nalguns dos alelos presentes; é o caso dos grupos B*35, B*40 e B*44 nas três populações Portuguesas, do grupo B*15 nos Açores, Madeira, Guiné-Bissau e Cabo Verde, do grupo B*07 na Guiné-Bissau e em Cabo Verde, do grupo B*14 na Madeira e do grupo B*57 em Cabo Verde.

9.4. HLA-DRB1

O alelo HLA-DRB1*0701 foi o mais frequente deste *locus* nas três populações Portuguesas (15-18%). Na Guiné-Bissau o alelo predominante foi o DRB1*1304 (19,6%) e em Cabo Verde o DRB1*110101 (10,1%). Alguns dos alelos que se seguem com maiores frequências coincidem entre algumas das populações estudadas, nomeadamente o DRB1*0301 e o DRB1*1301 em Portugal continental (13,3% e 7,9% respectivamente), nos Açores (8,4% e 7,6%), na Madeira (12,2% e 12,7%) e em Cabo Verde (8,1% e 6,9%), o DRB1*1501 em Portugal continental (7,6%), nos Açores (7,3%) e na Madeira (8,9%) e os alelos DRB1*090102, DRB1*110101 e

DRB1*130201, este último considerado protector contra a malária (Hill *et al.*, 1991), na Guiné-Bissau (7,7%, 10,8% e 7,7%) e em Cabo Verde (6%, 10,1% e 8,8%). Por outro lado, os alelos DRB1*040501 e DRB1*1102 (7,7% cada) apresentam frequências nesta ordem de grandeza apenas na Guiné-Bissau.

O grupo mais polimórfico nas populações Portuguesas estudadas foi o DRB1*04 com 7 a 8 alelos. O grupo DRB1*13 é o mais polimórfico na Guiné-Bissau (4 alelos) e, em conjunto com o DRB1*04, em Cabo Verde (5 alelos cada).

Alguns grupos polimórficos mostram frequências mais equilibradas entre alguns dos alelos que os compõem; é o caso dos grupos DRB1*11 e DRB1*13 em todas as populações do presente estudo, dos grupos DRB*01 e DRB1*04 em todas as populações excepto na Guiné-Bissau e do DRB1*15 em Cabo Verde.

9.5. Frequências haplotípicas

Todos os pares de *loci* estudados (A-B, A-DRB1 e B-DRB1) estavam em *linkage desequilibrium* estatisticamente significativo nas cinco populações, excepto o A-DRB1 em Portugal continental. As tabelas 13-19; 23-33, 36-42 e 44-47 apresentam os haplotipos mais frequentes nas populações estudadas.

Cada uma destas populações possui um haplotipo mais frequente diferente. Portugal continental apresenta em maior frequência o haplotipo A*020101-B*440301-DRB1*070101 (3,1%), os Açores o A*0201-B*5101-DRB1*0701 (2%) e a Madeira o A*020101-B*510101-DRB1*130101 (2,7%), três haplotipos com alguns alelos em comum. Os haplotipos mais frequentes encontrados na Guiné-Bissau (A*2301-B*1503-DRB1*110101 4,6%) e em

Cabo Verde (A*3002-B*350101-DRB1*1001 2,8%) são constituídos por alelos que não são comuns entre si. O haplotipo mais prevalente na Madeira surge também em Portugal continental (1,4%) e no Barlavento Cabo-Verdiano (1,6%).

Considerando apenas os haplotipos mais frequentes (>1%), não foi encontrado nenhum comum às cinco populações estudadas. As 3 populações Portuguesas (Continente, Madeira e Açores) apresentam em comum dois destes haplotipos mais comuns: A*0101-B*0801-DRB1*0301 e A*2902-B*4403-DRB1*0701, dois haplotipos considerados de origem Europeia. Portugal continental apresenta mais haplotipos frequentes comuns à Madeira (5 haplotipos) do que aos Açores (3 haplotipos). Portugal continental possui ainda 3 haplotipos frequentes em comum com o Sotavento Cabo-Verdiano mas apenas um com as ilhas Cabo-Verdianas do Barlavento. Entre a Madeira e os Açores 3 destes haplotipos mais comuns são coincidentes. Entre os Açores e o Sotavento Cabo-Verdiano, assim como entre a Madeira e este último grupo de ilhas, apenas um haplotipo com frequência superior a 1% é comum a ambas as populações. Entre as populações Africanas estudadas apenas um haplotipo mais frequente é comum entre o Barlavento e o Sotavento Cabo-Verdiano e dois entre a Guiné-Bissau e as ilhas do Sotavento.

Apesar de na Guiné-Bissau ter sido encontrada uma diversidade alélica inferior, o número de haplotipos frequentes é superior ao encontrado nas outras populações estudadas.

9.6. Análises filogenéticas

As árvores filogenéticas construídas a partir das frequências alélicas dos genes HLA-A, HLA-B e HLA-DRB1 (figuras 5-7, 9-14, 16 e 17) colocam

Portugal continental entre a Europa e o Norte de África e reflectem uma maior proximidade do Norte e Sul de Portugal às populações Norte Africanas. Os Açores associam-se, nos dendrogramas, às populações Asiáticas e Ameríndias. Excluindo estas últimas e dividindo os Açores em grupo Oriental e Central, a análise filogenética mostra comportamentos diferentes nestas duas sub-populações. O grupo Oriental revela uma maior proximidade a Portugal continental e o grupo Central a outras populações Europeias. A Madeira tem um comportamento semelhante a Portugal continental posicionando-se entre as populações Europeias e Africanas. As análises filogenéticas da população Cabo-Verdiana mostram uma grande proximidade à Guiné -Bissau, reflectindo a sua origem Africana. Os dendrogramas revelam ainda que Cabo Verde se posiciona entre as populações sub-Saharianas e as Norte Africanas, mais evidente nas ilhas do Barlavento.

A comparação das populações através da análise das coordenadas principais (PCO) (figuras 8, 12, 15 e 18) é em geral consistente com os resultados das análises filogenéticas. Os Açores, nesta análise, afastam-se definitivamente das populações Asiáticas e Ameríndias.

Capítulo 10

10. Discussão geral

10.1. Distribuição dos alelos HLA em Portugal e Cabo Verde.
Influências do comércio de escravos.

O povoamento da Madeira, logo após a sua descoberta, ocorreu maioritariamente com indivíduos provenientes de Portugal. Em 1670 a Madeira possuía já um quinto da sua população actual. Para além da contribuição Ibérica e Norte Europeia, outros povos contribuíram para o seu povoamento: Mouros (escravos Caucasianos provenientes da Mauritânia), escravos sub-Saharianos introduzidos directamente a partir de Cabo Verde (a certa altura constituíam 10% da população), Judeus, escravos da Índia e alguns nativos das ilhas Canárias (Guanches) (Pereira, 1989).

É possível identificar as marcas genéticas deixadas por alguns destes povos. Dois métodos estão disponíveis para estudar composição genética de uma população e eventualmente discernir a sua estrutura actual: variação das frequências alélicas e distribuição haplotípica. A população Madeirense apresenta alelos raros de origem Africana (HLA-A*34, HLA-A*36 e HLA-A*74) que a distingue de outras populações Portuguesas. Por outro lado,

3,5% dos seus haplotipos são de origem Africana. Informação obtida a partir do DNA mitocondrial mostra também a presença na Madeira de haplogrupos tipicamente Africanos (L e M1), 20,4% do total, comparado com apenas 6,5% nos Açores (Brehm *et al.*, 2003). O mesmo se aplica à distribuição de microsátélites STR (Fernandes *et al.*, 2001). É mais difícil identificar a influência Norte Africana devida ao comércio de escravos. As três sub-populações de Portugal continental mostram diferenças marcantes entre si, que não são tão visíveis a partir da comparação isolada das frequências alélicas. Por exemplo, apesar dos alelos A*02 e B*44 serem comuns às cinco populações estudadas, eles combinam com diferentes alelos DRB1 formando haplotipos distintos entre as populações. A tabela 8 apresenta vários exemplos destes casos em que os haplotipos apesar de frequentes são únicos de cada população (A*02-B*44-DRB1*02, A*02-B*44-DRB1*04, A*02-B*44-DRB1*07). A heterogeneidade das três populações de Portugal continental revela que, ao contrário do que foi encontrado previamente (Arnaiz-Villena *et al.*, 1997), existe uma forte base genética em Portugal continental (e também na Madeira e Cabo Verde) proveniente do Mediterrâneo e Norte de África, tal como é sugerido pelas frequências elevadas dos haplotipos A*33-B*14 e A*30-B*18.

A presença Europeia em Cabo Verde, embora nunca mais de 10% da população, é perceptível ao nível do sistema HLA. Cabo Verde possui alelos encontrados em frequências relativamente elevadas nas populações Portuguesas, ao contrário do que sucede nos Mandenka ou na população dos Camarões (tabelas 5 e 6, HLA-A*01, HLA-B*14, HLA-B*40 e HLA-B*51). O mesmo sucede relativamente aos haplotipos Ibero-Mediterrânicos mais frequentes (A*29-B*44-DRB1*07 e A*01-B*08-DRB1*03) (Arnaiz-

Villena *et al.*, 1997) ou mesmo aos marcadores Ibéricos do Neolítico (A*30-B*18) (Martinez-Laso *et al.*, 1995). Só poderemos ter a certeza de que estes haplotipos são originais da Península Ibérica, e não foram trazidos com os próprios escravos, depois de uma investigação completa sobre as populações da costa Oeste Africana ter sido executada. Em conjunto, os haplotipos de Cabo Verde sem origem sub-Sahariana compreendem 15% do total. A presença destes marcadores mostram que a população actual Caboverdiana é atípica quando comparada com outras populações Africanas, reflectindo o quanto a sua origem se deveu à mistura de um substrato Africano com influência genética Europeia.

10.2. Caracterização dos genes HLA na população Portuguesa por sequenciação: na encruzilhada entre a Europa e África.

Portugal experimentou a influência de vários povos ao longo do período pré-histórico e histórico. Arnaiz-Villena *et al.* (1999) e González *et al.* (2003) sugeriram que há cerca de 6.000 anos atrás a Península Ibérica foi o receptáculo de uma migração substancial de povos Berberes proveniente do Sahara, fugindo às profundas alterações climáticas que começaram a ocorrer nessa região. Antes da presença Romana na Ibéria (100 a.C.), vários povos Europeus e Mediterrânicos povoaram esta península: Gregos, Fenícios e Celtas. Após a queda do Império Romano, Germânicos e Berberes do Norte de África também povoaram o território que hoje é conhecido por Portugal (Arnaiz-Villena *et al.*, 1997). Depois do século XV, um número substancial de sub-Saharianos entraram em Portugal através do comércio de escravos (Russell-Wood, 1998).

O presente trabalho descreve pela primeira vez em alta resolução as frequências alélicas dos genes HLA-A, HLA-B e HLA-DRB1 e estima os haplotipos no Norte, Centro e Sul de Portugal continental. Esta análise por alta resolução mostra que vários grupos identificados anteriormente (Spínola *et al.*, 2002) nos genes HLA estudados são na realidade constituídos por vários alelos (tabela 10). Esta informação adicional é altamente relevante para uma caracterização detalhada da estrutura dos genes HLA. Infelizmente, poucas populações mundiais têm sido convenientemente caracterizadas até ao nível alélico, situação que torna difícil uma análise comparativa da informação obtida neste trabalho.

A posição de Portugal entre a Europa e o Norte de África, revelada pelas árvores filogenéticas e análise das coordenadas principais (figuras 5-8) é corroborada pela existência de frequências intermédias de determinados alelos (e.g. HLA-A*010101, A*030101, A*2301, B*0702, B*4901, B*4501, B*5002, DRB1*0101, DRB1*040501). Outras frequências alélicas mostram valores similares às populações Norte Africanas (e.g. HLA-A*3301, A*6601, B*440301, B*4901, DRB1*1102 e DRB1*110401) e Europeias (e.g. HLA-A*020101, A*110101, A*2601, A*2902, A*3101, A*3201, B*1402, B*1501, B*2705, B*3503, B*5802, DRB1*130101, DRB1*040101 e DRB1*130201). Várias frequências alélicas e haplotípicas revelam uma distribuição heterogénea ao longo do território Português, reflectindo provavelmente diferentes influências genéticas. Por exemplo, o haplotipo A*010101-B*0801-DRB1*030101, considerado de origem Celta (Muro *et al.*, 2001), e o A*020101-B*440301-DRB1*070101, do Norte e Oeste Europeu (Sanchez-Velasco *et al.*, 2003), está apenas presente no Centro e Norte de Portugal revelando uma possível influência Celta e dos povos Bárbaros. De

forma coerente com a presença destes dois haplotipos, o HLA-B*0801 e o HLA-A*020101 possuem um gradiente decrescente das frequências alélicas de Norte para Sul em Portugal (Figuras 19 e 20).

O haplotipo Mediterrânico A*3301-B*1402-DRB1*010201 (Gómez-Casado *et al.*, 2000) aparece exclusivamente no Norte de Portugal (4.3%), apesar de aparecer no Sul (1%) associado ao B*140101, em concordância com Arnaiz-Villena e colegas (1997) que não encontraram este haplotipo numa população do Centro de Portugal. O alelo B*1402 apresenta um claro gradiente decrescente entre o Norte e o Sul, mas o B*140101 está apenas presente no Sul (2%) (Figuras 21 e 22).

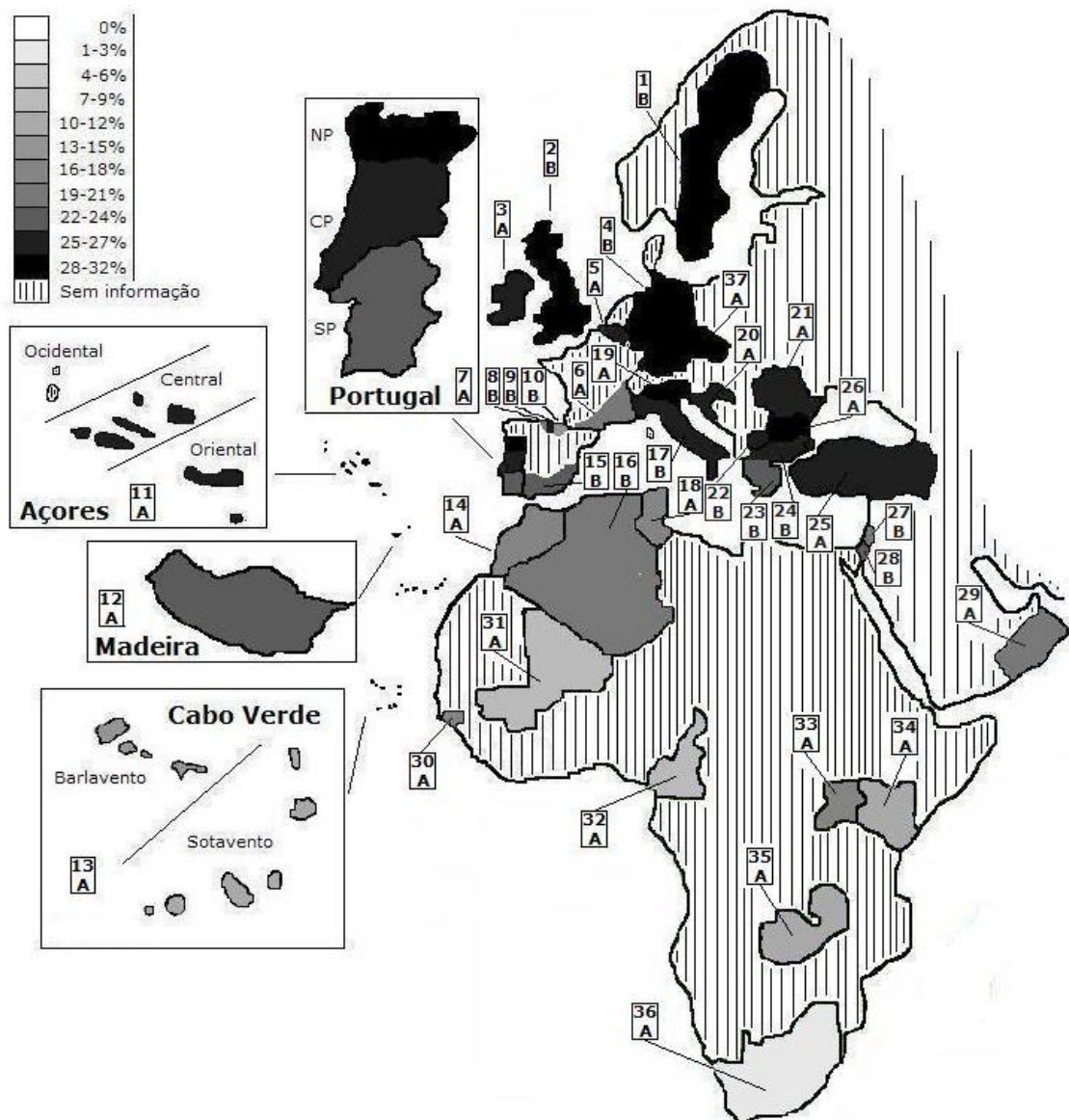


Figura 19- Distribuição geográfica das frequências do alelo A*0201 (A) ou do grupo A*02 (B) no continente Europeu e Africano. 1- Suécia (A*02: 28,8%); 2- Escócia (A*02: 31%) e País de Gales (A*02: 30,2%); 3- Irlanda do Norte (A*0201: 26,5%); 4- Alemanha (A*02: 28,2%); 5- Bélgica (A*0201: 26,6%); 6- França Sudeste (A*0201: 20,4%); 7- Portugal continental (A*0201: NP 29%, CP 26%, SP 23,5%); 8- Espanha Norte Cabuerniga (A*02: 20,5%); 9- Espanha Norte Catalanian (A*02: 27,1%); 10- Espanha Bascos Arratia (A*02: 16%); 11- Açores (A*0201: grupo Central 25,9%, grupo Oriental 25,6%); 12- Madeira (A*0201: 24,9%); 13- Cabo Verde (A*0201: Barlavento 14,9%, Sotavento 12,1%); 14- Marrocos (A*0201: 17,6%); 15- Espanha Murcia (A*02: 23%); 16- Argélia (A*02: 21,5%); 17- Itália (A*02: 25,4%); 18- Tunísia (A*0201: 17%); 19- Itália Norte (A*0201: 28,9%); 20- Croácia (A*0201: 26,3%); 21- Roménia (A*0201: 25%); 22- Macedónia (A*02: 25,6%); 23- Grécia (A*02: 22,15%); 24- Grécia Norte (A*0201: 27%); 25- Turquia (A*0201: 25,7%); 26- Bulgária (A*0201: 30%); 27- Israel (A*02: 24,7%); 28- Palestina (A*02: 20,3%); 29- Oman (A*0201: 21,6%); 30- Guiné-Bissau (A*0201: 16,7%); 31- Mali (A*0201: 8,3%); 32- Camarões (A*0201: 7,1%); 33- Uganda (A*0201: 18,5%); 34- Kenia (A*0201: 11,5%); 35- Zâmbia (A*0201: 12,8%); 36- Zulu África do Sul (A*0201: 2,5%); 37- República Checa (A*0201: 30,8%). (Ver referências em material e métodos).

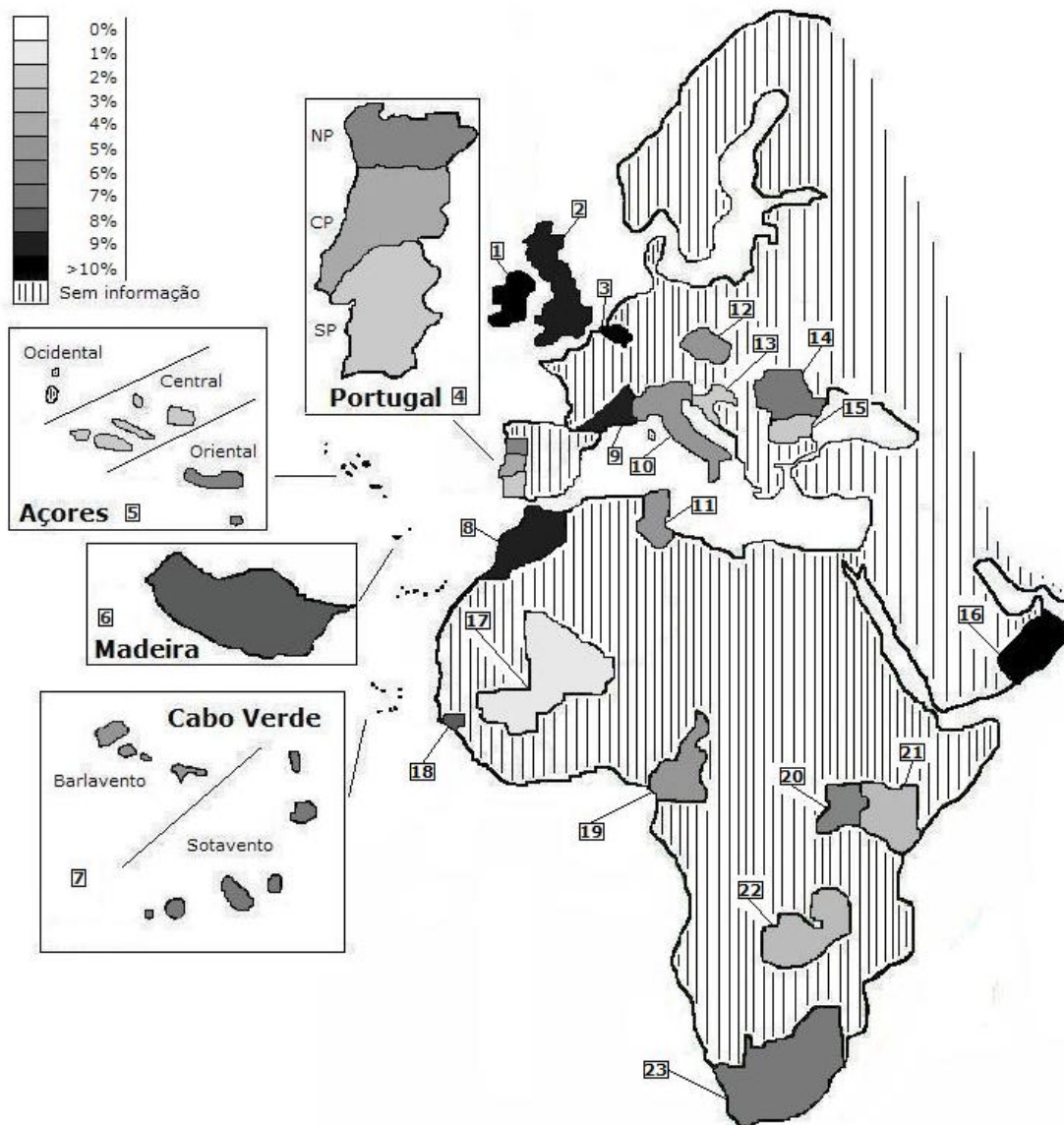


Figura 20- Distribuição geográfica das frequências do alelo B*0801 no continente Europeu e Africano. 1- Irlanda do Norte (15,7%); 2- Escócia (8,8%); 3- Bélgica (12,6%); 4- Portugal continental (NP 6,5%, CP 4%, SP 2,04%); 5- Açores (grupo Oriental 6,4%, grupo Central 1,8%); 6- Madeira (7,8%); 7- Cabo Verde (Barlavento 5,6%, Sotavento 7,3%); 8- Marrocos (8,9%); 9- França Sudeste (9,5%); 10- Italia Bergamo (5,3%); 11- Tunísia (4,6%); 12- República Checa (5,2%); 13- Croácia (1,7%); 14- Roménia (7,3%); 15- Bulgária (1,8%); 16- Oman (11%); 17- Mali (0,7%); 18- Guiné-Bissau (7,7%); 19- Camarões (5,4%); 20- Uganda (6,6%); 21- Kenia (3%); 22- Zâmbia (3,4%); 23- Zulu África do Sul (7,5%). (Ver referências em material e métodos).

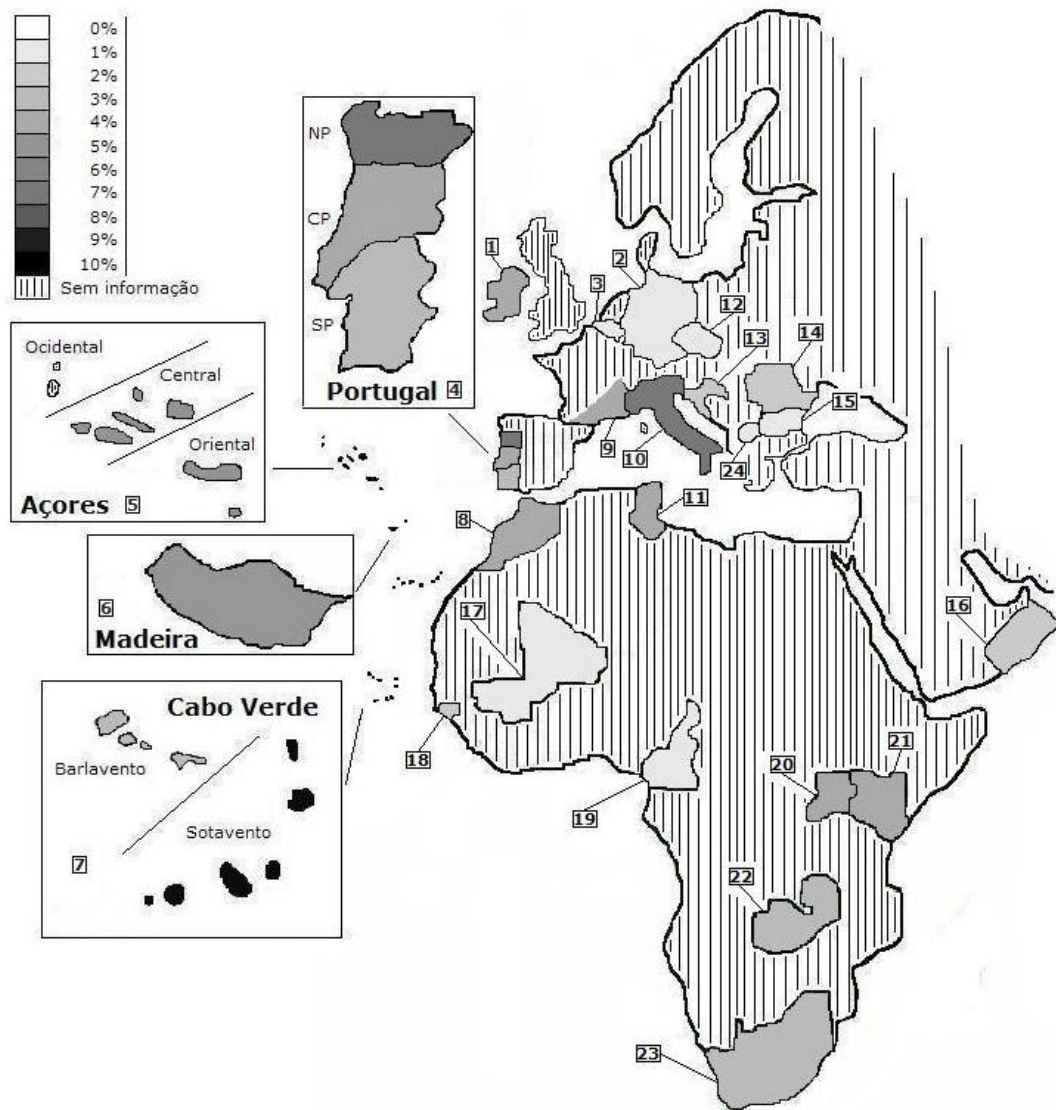


Figura 21- Distribuição geográfica das frequências do alelo B*1402 no continente Europeu e Africano. 1- Irlanda do Norte (3,6%); 2- Alemanha (1,4%); 3- Bélgica (1%); 4- Portugal continental (NP 6,5%, CP 4%, SP 3,06%); 5- Açores (grupo Oriental 5,1%, grupo Central 5,4%); 6- Madeira (4,9%); 7- Cabo Verde (Barlavento 3,2%, Sotavento 9,7%); 8- Marrocos (3,6%); 9- França Sudeste (3,5%); 10- Itália (6,7%); 11- Tunísia (4,08%); 12- República Checa (1,4%); 13- Croácia (2%); 14- Roménia (1,7%); 15- Bulgária (0,9%); 16- Oman (2,1%); 17- Mali (1,1%); 18- Guiné-Bissau (1,5%); 19- Camarões (0,5%); 20- Uganda (3,7%); 21- Kenia (4,3%); 22- Zâmbia (3,4%); 23- Zulu África do Sul (2,5%); 24- Macedónia (0,7%). (Ver referências em material e métodos).

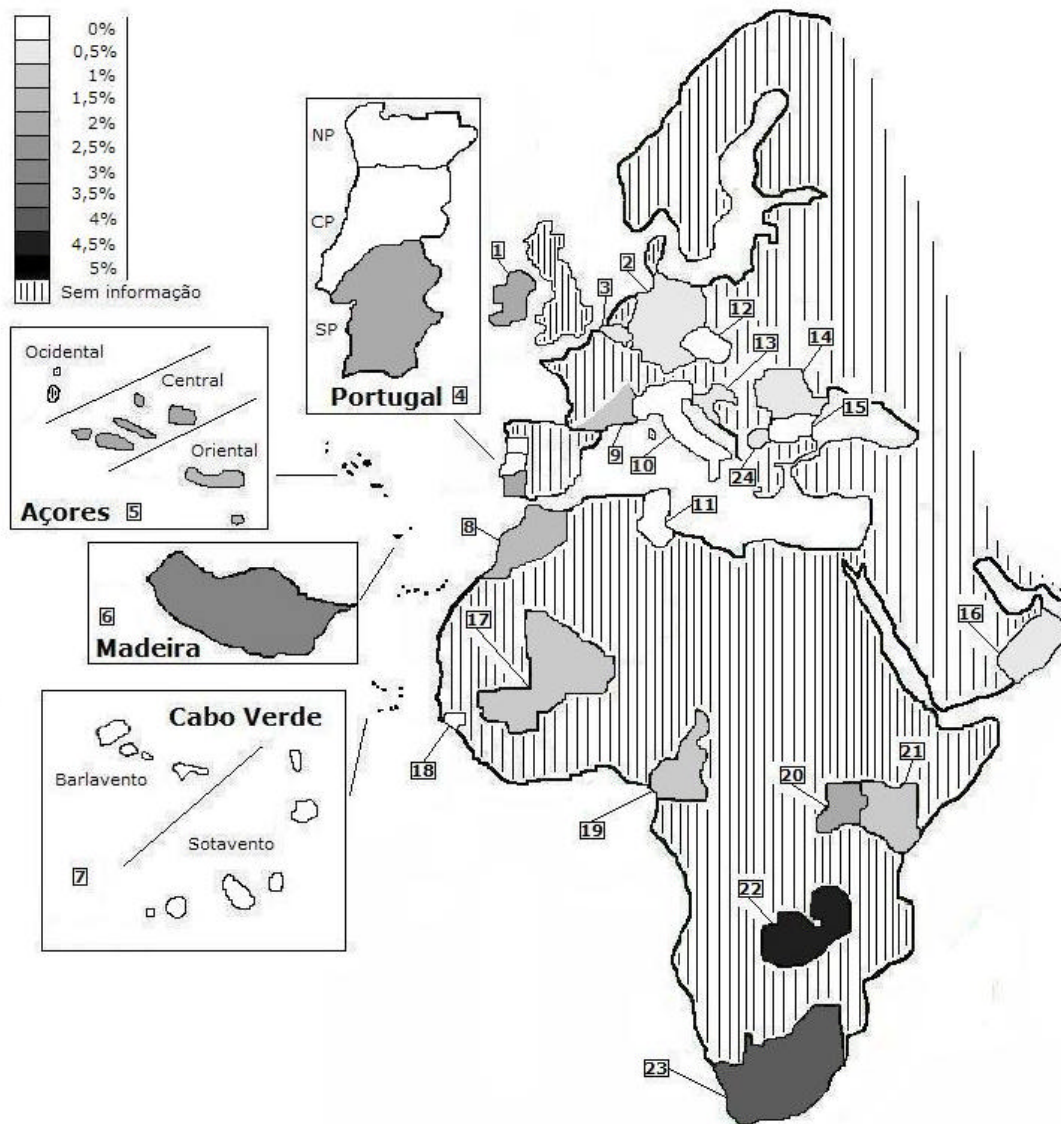


Figura 22- Distribuição geográfica das frequências do alelo B*1401 no continente Europeu e Africano. 1- Irlanda do Norte (2,2%); 2- Alemanha (0,3%); 3- Bélgica (0,5%); 4- Portugal continental (NP 0%, CP 0%, SP 2,04%); 5- Açores (grupo Oriental 1,3%, grupo Central 1,8%); 6- Madeira (2,7%); 7- Cabo Verde (Barlavento 0%, Sotavento 0%); 8- Marrocos (1,4%); 9- França Sudeste (1,2%); 10- Itália (0%); 11- Tunísia (0%); 12- República Checa (0%); 13- Croácia (0,3%); 14- Roménia (0,6%); 15- Bulgária (0%); 16- Oman (0,4%); 17- Mali (1,1%); 18- Guiné-Bissau (0%); 19- Camarões (1,1%); 20- Uganda (1,6%); 21- Kenia (1,1%); 22- Zâmbia (4,6%); 23- Zulu África do Sul (4%); 24- Macedónia (0,2%). (Ver referências em material e métodos).

Os haplotipos A*020101-B*070201-DRB1*150101 (Europeu e Norte Africano), A*010101-B*5801-DRB1*070101 (Basco e Berbere) e A*020101-B*070201-DRB1*150101 (Norte Africano e do Oeste Europeu) foram apenas encontrados no Centro de Portugal. Por outro lado, os haplotipos A*020101-B*510101-DRB1*130101 (1%) (Ibero-Berbere), A*020101-B*180101-DRB1*1104 (1%) (Mediterrânico), A*3001-B*1302-DRB1*070101 (1%) (Creta e Médio Oriente) e outros haplotipos autóctones (A*2301-B*4901-DRB1*130201: 4.3%, e A*020101-B*440201-DRB1*150101: 1%) foram apenas encontrados no Norte de Portugal. Esta informação sugere que importantes marcadores genéticos do Mediterrâneo e Norte de África deverão ter entrado em Portugal não apenas em períodos mais recentes mas também em tempos pré-históricos. Esta influência genética mais antiga terá sido forçada a migrar para Norte devido à expansão Árabe, através do Estreito de Gibraltar, em 800 d.C. Desta forma, a população do Norte de Portugal mostra características que podem resultar de um conjunto genético mais antigo. Esta hipótese torna-se muito mais consistente quando se constata que o Sul de Portugal não apresenta nenhuns haplotipos bem conhecidos, excepto o Ibérico e paleo-Norte Africano A*3002-B*180101-DRB1*030101 (4%), inexistentes no Centro e Norte de Portugal, e o A*2902-B*440301-DRB1*070101, um antigo haplotipo do Oeste Europeu (também presente no Norte de África), o qual constitui o único haplotipo presente nas três sub-populações de Portugal continental (Norte, Centro e Sul). Estes dois haplotipos podem ter sido trazidos para o Sul de Portugal pelos novos Berberes do Norte de África que chegaram com os Árabes. Para além destes, o A*020101-B*180101-DRB1*030101 (4% no Sul e 2% no Centro), identificado anteriormente na

Sardenha (Rendine *et al.*, 1998) e Irão (Farjadian *et al.*, 2004), é o único haplotipo do Sul comum ao Centro de Portugal. O A*020101-B*3502-DRB1*110101, cujo correspondente haplotipo em baixa resolução foi encontrado também na Algéria (Arnaiz-Villena *et al.*, 1997), está presente no Sul de Portugal com uma frequência de 1%.

Vários alelos do Sul e Centro de Portugal não estão presentes no Norte mas todos os alelos do Norte estão presentes em pelo menos uma das outras duas sub-populações. Estas diferenças nas frequências alélicas Norte-Sul reforça a ideia de uma migração em direcção ao Norte pois sob esta hipótese espera-se no Norte a perda de alguns alelos e o aumento de algumas frequências devido ao efeito fundador. Surpreendentemente, em oposição ao gradiente acima descrito, o Centro de Portugal apresenta vários alelos com frequências mais baixas do que as encontradas no Sul e no Norte (e.g. HLA-A*2301, A*2601, B*350101, B*4901, B*5001, DRB1*0101, DRB1*0102, DRB1*0103, DRB1*040101 e DRB1*130201) (Figuras 23-32). Este gradiente decrescente do Norte e do Sul para o Centro de Portugal pode reflectir uma distribuição bipolar dos Portugueses mais antigos (no Norte) e movimentos Berberes mais recentes (no Sul). Nos dendrogramas, o Centro de Portugal junta-se sistematicamente com as populações Europeias ao contrário do Norte e Sul de Portugal, provavelmente devido à inexistência de marcadores comuns (figura 7).

A influência sub-Sahariana em Portugal devido ao comércio de escravos é difícil de detectar. A maior parte das populações sub-Saharianas, especialmente as da costa Ocidental Africana, não apresentam caracterização alélica em alta resolução dos genes classe I e II do sistema HLA, o que torna difícil a identificação destes marcadores em Portugal. No

entanto, vários haplotipos dois-*loci* frequentes em populações sub-Saharanas (Ellis *et al.*, 2000; Middleton *et al.*, 2003; Cao *et al.*, 2004) foram encontrados em Portugal (A*030101-B*4201, A*2301-B*0801, A*2301-B*440301, A*2301-B*5801, A*3001-B*5301 e A*3002-B*1402) demonstrando uma pequena mas visível influência genética deixada pelos sub-Saharanos.

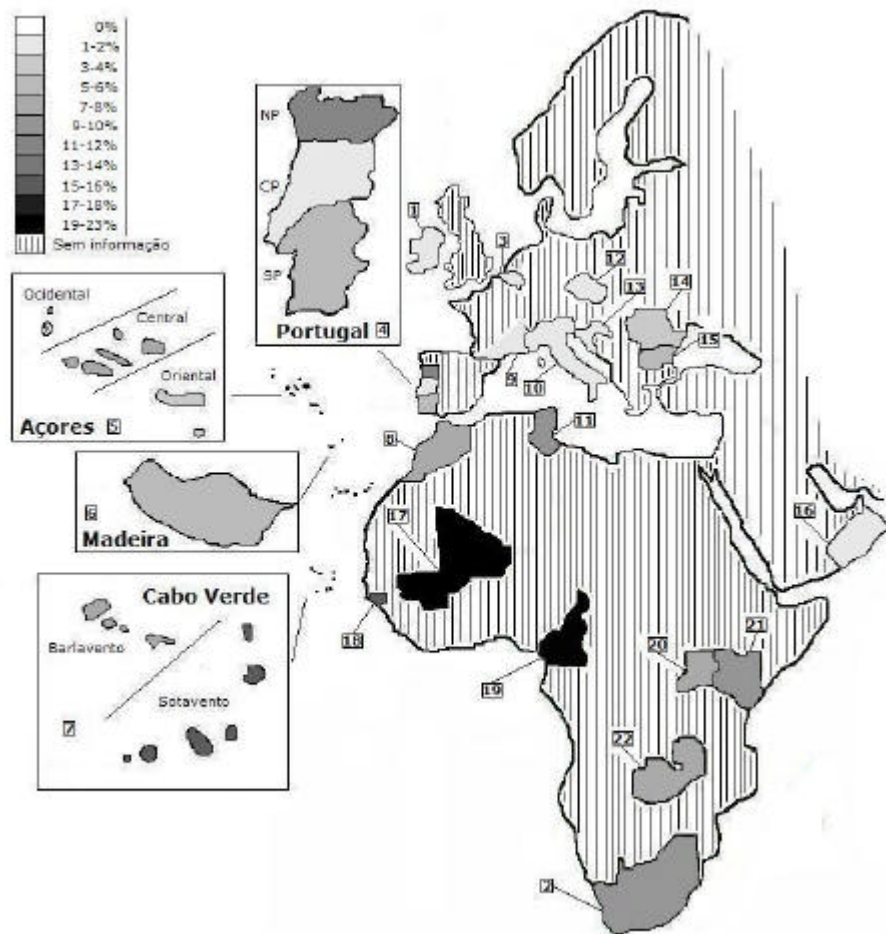


Figura 23- Distribuição geográfica das frequências do alelo A*2301 no continente Europeu e Africano. 1- Irlanda do Norte (1,4%); 2- Zulu África do Sul (10%); 3- Bélgica (1,6%); 4- Portugal continental (NP 10,9%, CP 1%, SP 6,1%); 5- Açores (grupo Oriental 2,6%, grupo Central 7,14%); 6- Madeira (6,2%); 7- Cabo Verde (Barlavento 8,1%, Sotavento 14,8%); 8- Marrocos (8,2%); 9- França Sudeste (1,2%); 10- Itália Bergamo (1,1%); 11- Tunísia (9,5%); 12- República Checa (2,19%); 13- Croácia (2,3%); 14- Roménia (2,5%); 15- Bulgária (5,5%); 16- Oman (1,7%); 17- Mali (23,3%); 18- Guiné-Bissau (15,6%); 19- Camarões (18,7%); 20- Uganda (6,8%); 21- Kenia (8,9%); 22- Zâmbia (8,1%). (Ver referências em material e métodos).

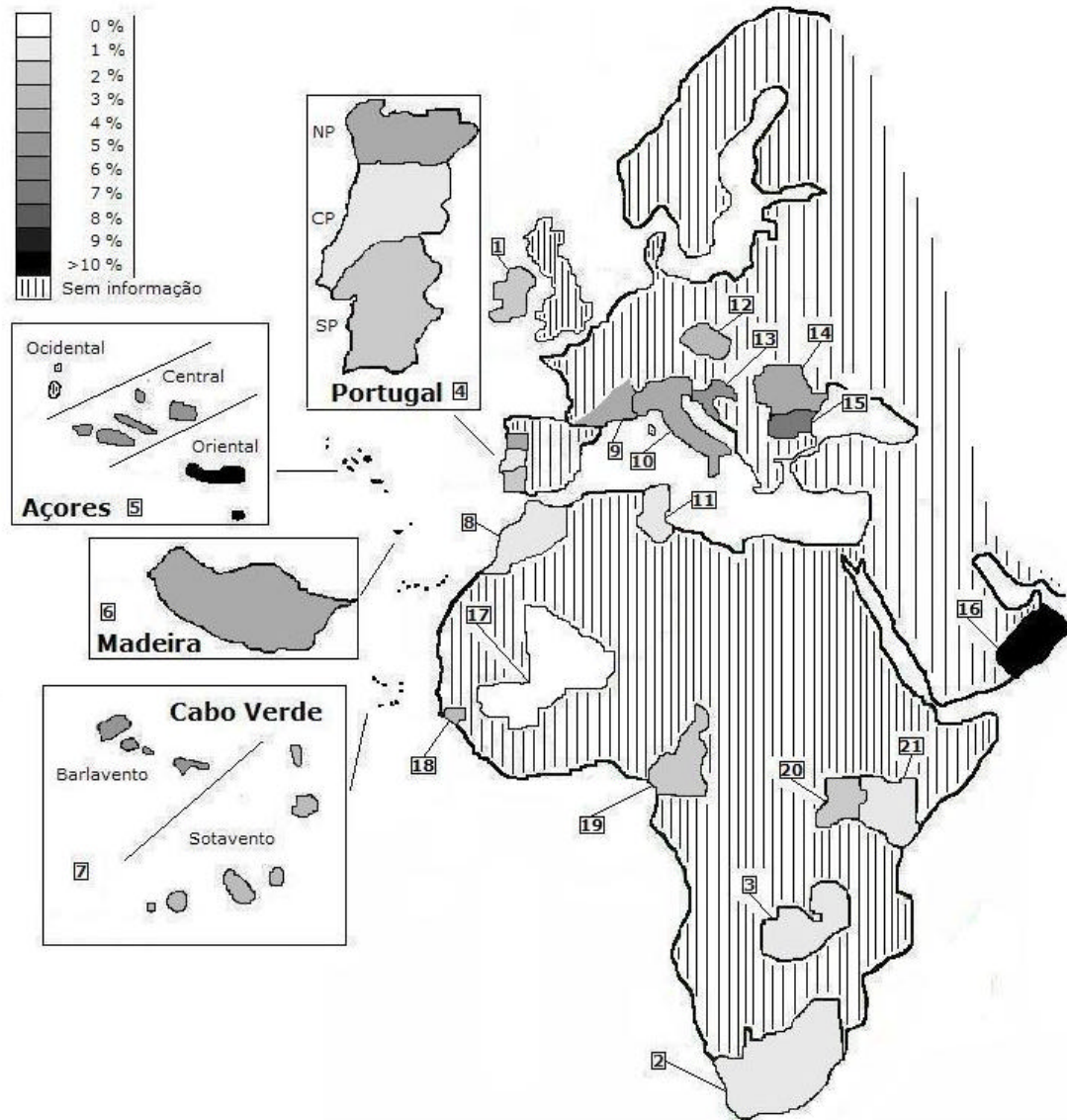


Figura 24- Distribuição geográfica das frequências do alelo A*2601 no continente Europeu e Africano. 1- Irlanda do Norte (1,8%); 2- Zulu África do Sul (1%); 3- Zâmbia (1,2%); 4- Portugal continental (NP 4,3%, CP 1%, SP 2%); 5- Açores (grupo Oriental 12,8%, grupo Central 4,5%); 6- Madeira (3,5%); 7- Cabo Verde (Barlavento 4,8%, Sotavento 3,2%); 8- Marrocos (0,7%); 9- França Sudeste (4,3%); 10- Itália Norte (3,8%); 11- Tunísia (1%); 12- República Checa (3,1%); 13- Croácia (5,3%); 14- Roménia (3,8%); 15- Bulgária (7,3%); 16- Oman (10,2%); 17- Mali (0%); 18- Guiné-Bissau (3,1%); 19- Camarões (1,7%); 20- Uganda (1,5%); 21- Kenia (0,6%). (Ver referências em material e métodos).

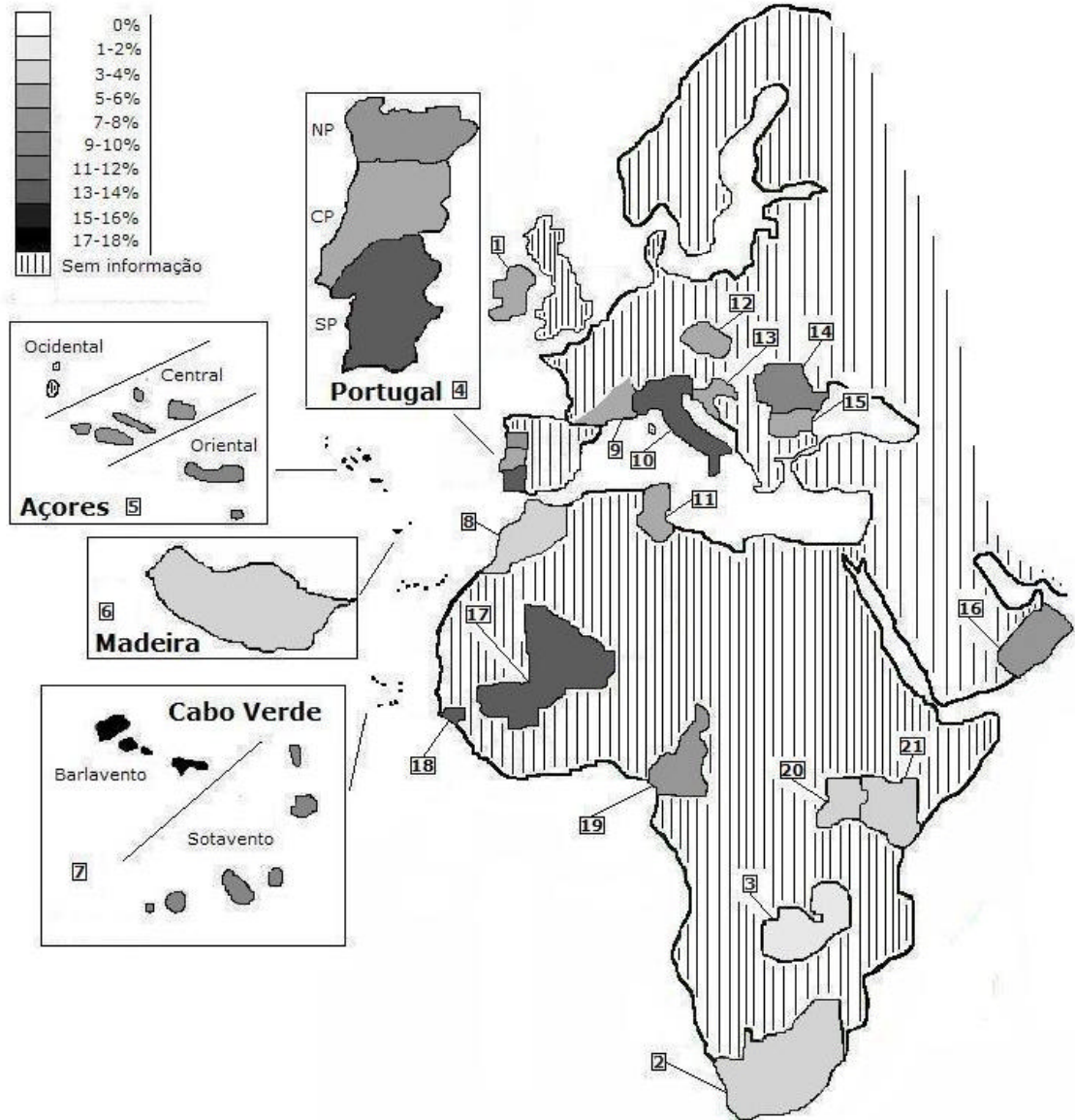


Figura 25- Distribuição geográfica das frequências do alelo B*3501 no continente Europeu e Africano. 1- Irlanda do Norte (4,6%); 2- Zulu África do Sul (4%); 3- Zâmbia (2,3%); 4- Portugal continental (NP 6,5%, CP 5%, SP 13,3%); 5- Açores (grupo Oriental 9%, grupo Central 7,1%); 6- Madeira (3,8%); 7- Cabo Verde (Barlavento 17,4%, Sotavento 8,9%); 8- Marrocos (4,3%); 9- França Sudeste (5%); 10- Itália Norte (13,3%); 11- Tunísia (4,9%); 12- República Checa (6,1%); 13- Croácia (6,4%); 14- Roménia (10,1%); 15- Bulgária (4,6%); 16- Oman (7,6%); 17- Mali (12,9%); 18- Guiné-Bissau (14,4%); 19- Camarões (7,1%); 20- Uganda (3,7%); 21- Kenia (3,4%). (Ver referências em material e métodos).

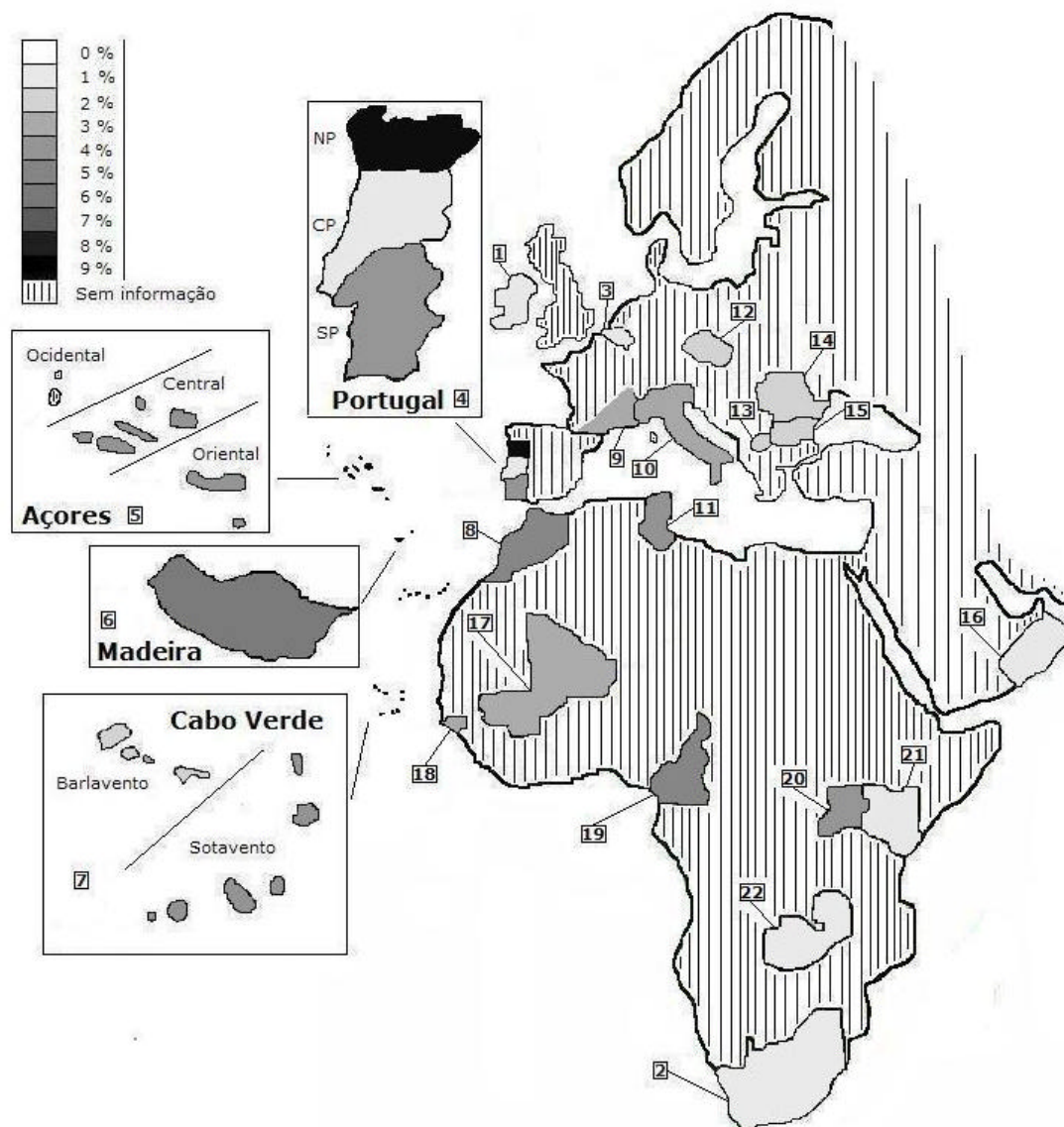


Figura 26- Distribuição geográfica das frequências do alelo B*4901 no continente Europeu e Africano. 1- Irlanda do Norte (1%); 2- Zulu África do Sul (1%); 3- Bélgica (1%); 4- Portugal continental (NP 8,7%, CP 1%, SP 4,1%); 5- Açores (grupo Oriental 3,8%, grupo Central 3,6%); 6- Madeira (5,9%); 7- Cabo Verde (Barlavento 2,4%, Sotavento 4%); 8- Marrocos (5,1%); 9- França Sudeste (2,7%); 10- Itália Norte (3,3%); 11- Tunísia (3,6%); 12- República Checa (1,9%); 13- Macedónia (1,9%); 14- Roménia (1,6%); 15- Bulgária (1,8%); 16- Oman (0,4%); 17- Mali (2,5%); 18- Guiné-Bissau (3,1%); 19- Camarões (5,4%); 20- Uganda (3,7%); 21- Kenia (0,4%); 22- Zâmbia (1,1%). (Ver referências em material e métodos).

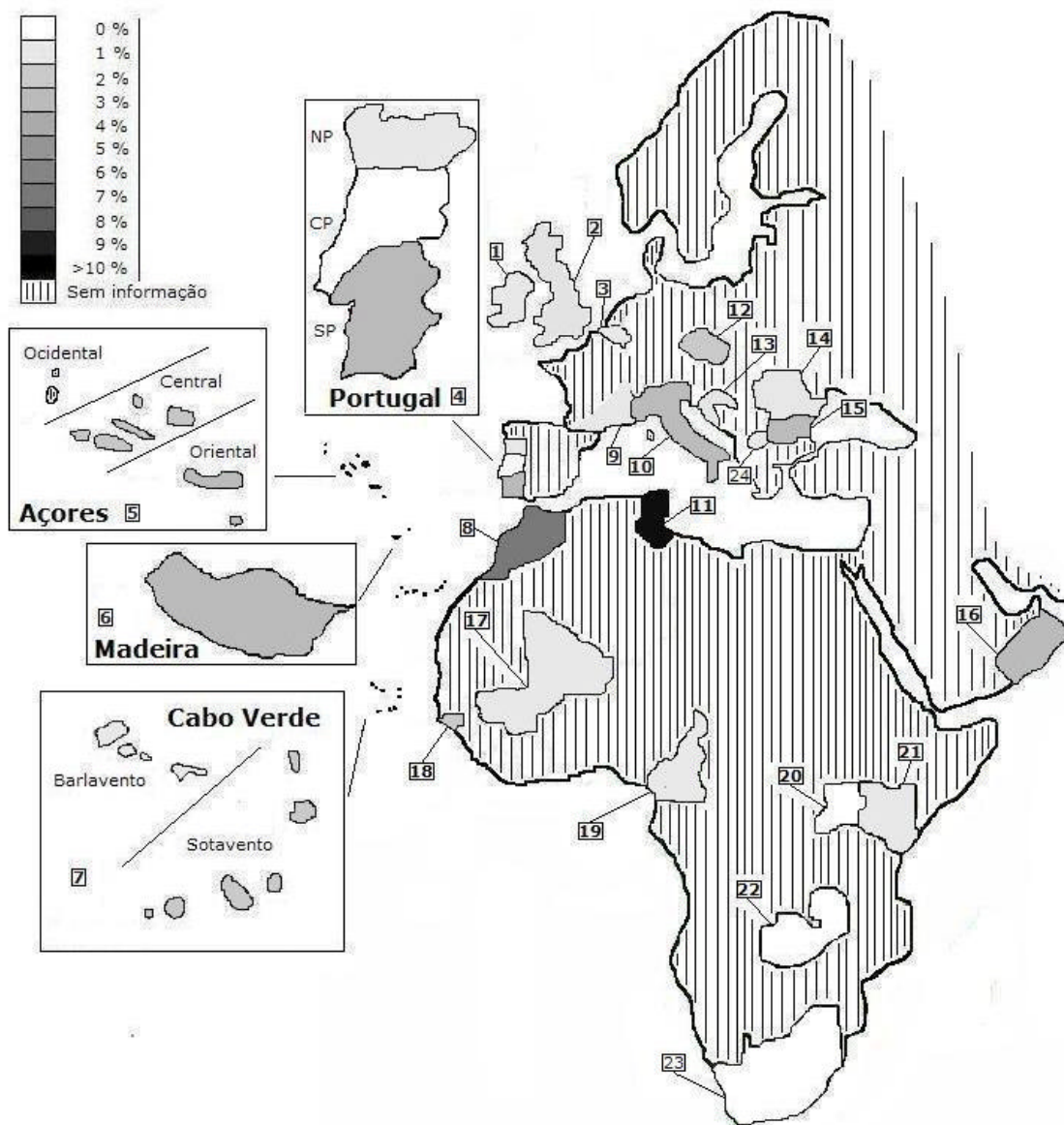


Figura 27- Distribuição geográfica das frequências do alelo B*5001 no continente Europeu e Africano. 1- Irlanda do Norte (1%); 2- País de Gales (0,8%); 3- Bélgica (1%); 4- Portugal continental (NP 1,1%, CP 0%, SP 3,1%); 5- Açores (grupo Oriental 2,6%, grupo Central 1,8%); 6- Madeira (3,2%); 7- Cabo Verde (Barlavento 0,8%, Sotavento 1,6%); 8- Marrocos (7,3%); 9- França Sudeste (1,2%); 10- Itália Norte (3,3%); 11- Tunísia (11,8%); 12- República Checa (2,4%); 13- Croácia (1,3%); 14- Roménia (0,7%); 15- Bulgária (2,7%); 16- Oman (3%); 17- Mali (1,1%); 18- Guiné-Bissau (1,5%); 19- Camarões (0,5%); 20- Uganda (0%); 21- Kenia (0,2%); 22- Zâmbia (0%); 23- Zulu África do Sul (0%); 24- Macedónia (0,7%). (Ver referências em material e métodos).

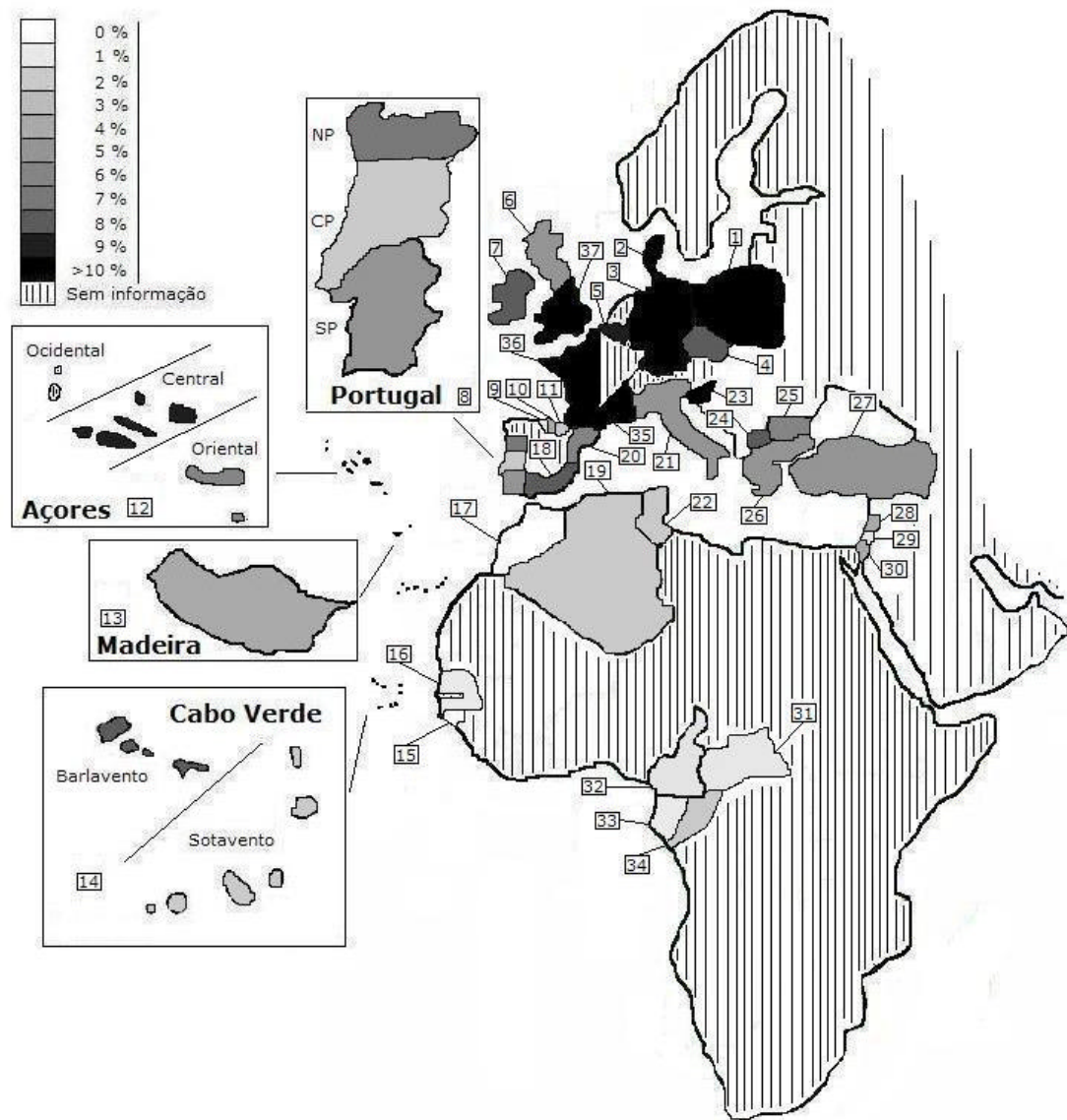


Figura 28- Distribuição geográfica das frequências do alelo DRB1*0101 no continente Europeu e Africano. 1- Polónia (9,6%); 2- Dinamarca (13%); 3- Alemanha (10,6%); 4- República Checa (7,6%); 5- Bélgica (8,6%); 6- Escócia (5,3%); 7- Irlanda do Norte (7,9%); 8- Portugal continental (NP 6,5%, CP 2%, SP 5,1%); 9- Norte de Espanha Cabuernigo (0,4%); 10- Norte de Espanha Cantabrian (4,2%); 11- Espanha Bascos (2%); 12- Açores (grupo Oriental 6,4%, grupo Central 8,9%); 13- Madeira (3,5%); 14- Cabo Verde (Barlavento 8,1%, Sotavento 1,6%); 15- Guiné-Bissau (0%); 16- Senegal (1,1%); 17- Marrocos (0%); 18- Sul de Espanha Múrcia (7,5%); 19- Algéria (1,5%); 20- Espanha Barcelona (5,7%); 21- Itália Norte (4,6%); 22- Tunísia (2%); 23- Eslovénia (10%); 24- Macedónia (8,2%); 25- Bulgária (5,5%); 26- Grécia (4,5%); 27- Turquia (4,6%); 28- Líbano (4,3); 29- Palestina (0,9%); 30- Israel (3,8%); 31- República Centro Africana Aka Pigmeus (0,6%); 32- Camarões (0,4%); 33- Gabão (0,9%); 34- Congo (2,2%); 35- França Sudeste (9,6%); 36- França Oeste (10,5%); 37- Inglaterra (10,2%). (Ver referências em material e métodos).

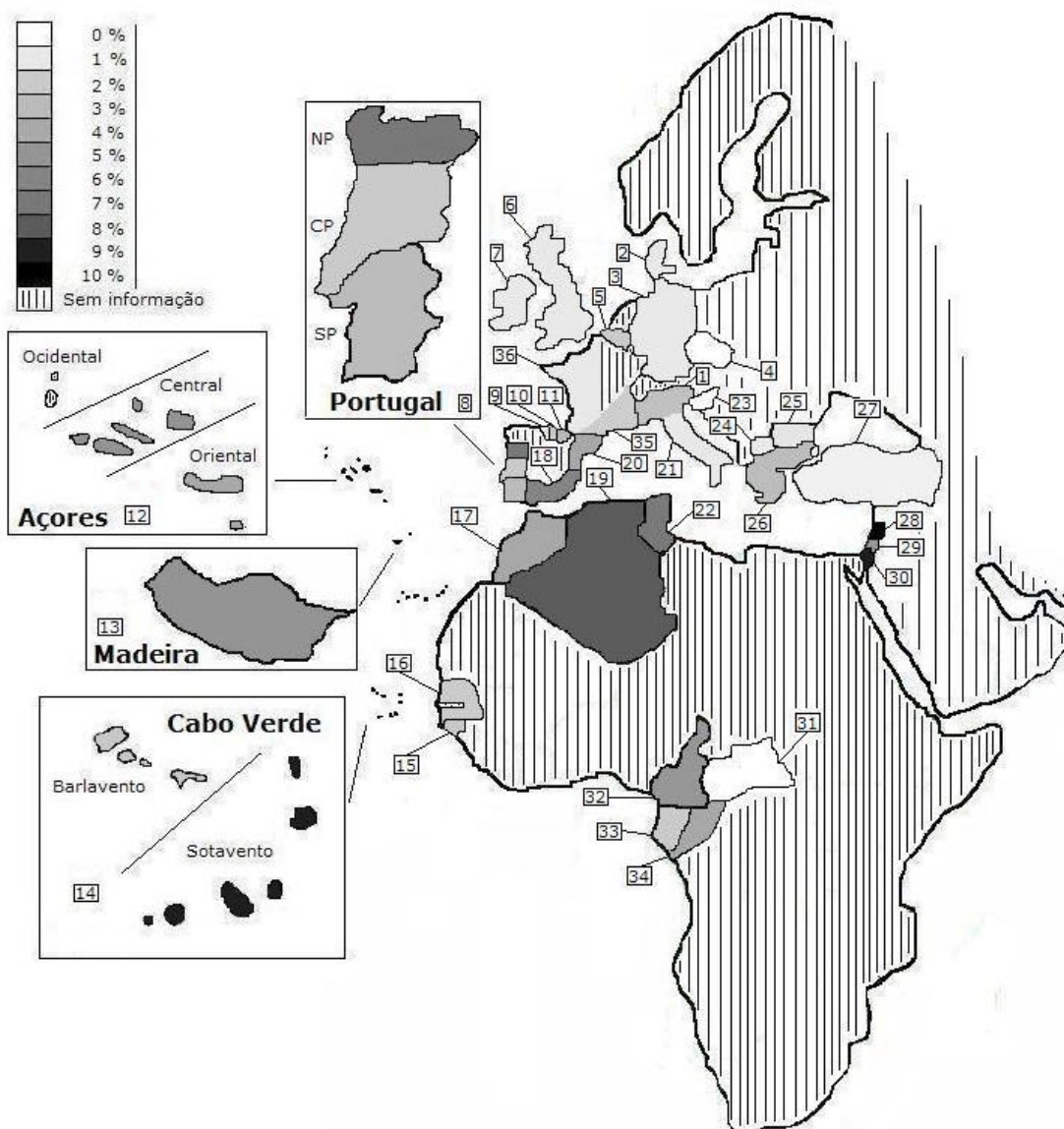


Figura 29- Distribuição geográfica das frequências do alelo DRB1*0102 no continente Europeu e Africano. 1- Itália Norte (2,6%); 2- Dinamarca (0,9%); 3- Alemanha (1,4%); 4- República Checa (0%); 5- Bélgica (2%); 6- Inglaterra (0,6%); 7- Irlanda do Norte (0,8%); 8- Portugal continental (NP 6,5%, CP 2%, SP 3,1%); 9- Norte de Espanha Cabuernigo (0,5%); 10- Norte de Espanha Cantabrian (2,4%); 11- Espanha Bascos (2,7%); 12- Açores (grupo Oriental 3,8%, grupo Central 5,4%); 13- Madeira (4,6%); 14- Cabo Verde (Barlavento 1,6%, Sotavento 8,9%); 15- Guiné-Bissau (2,3%); 16- Senegal (2,2%); 17- Marrocos (4%); 18- Sul de Espanha Múrcia (6%); 19- Argélia (8%); 20- Espanha Barcelona (3,7%); 21- Itália Centro (1,3%); 22- Tunísia (6,5%); 23- Eslovénia (0%); 24- Macedónia (1%); 25- Bulgária (0,9%); 26- Grécia (2,5%); 27- Turquia (1,2%); 28- Líbano (9,7); 29- Palestina (4,5%); 30- Israel (8,5%); 31- República Centro Africana Aka Pigmeus (0%); 32- Camarões (4,8%); 33- Gabão (1,5%); 34- Congo (4,4%); 35- França Sudeste (1,6%); 36- França Oeste (0,5%). (Ver referências em material e métodos).

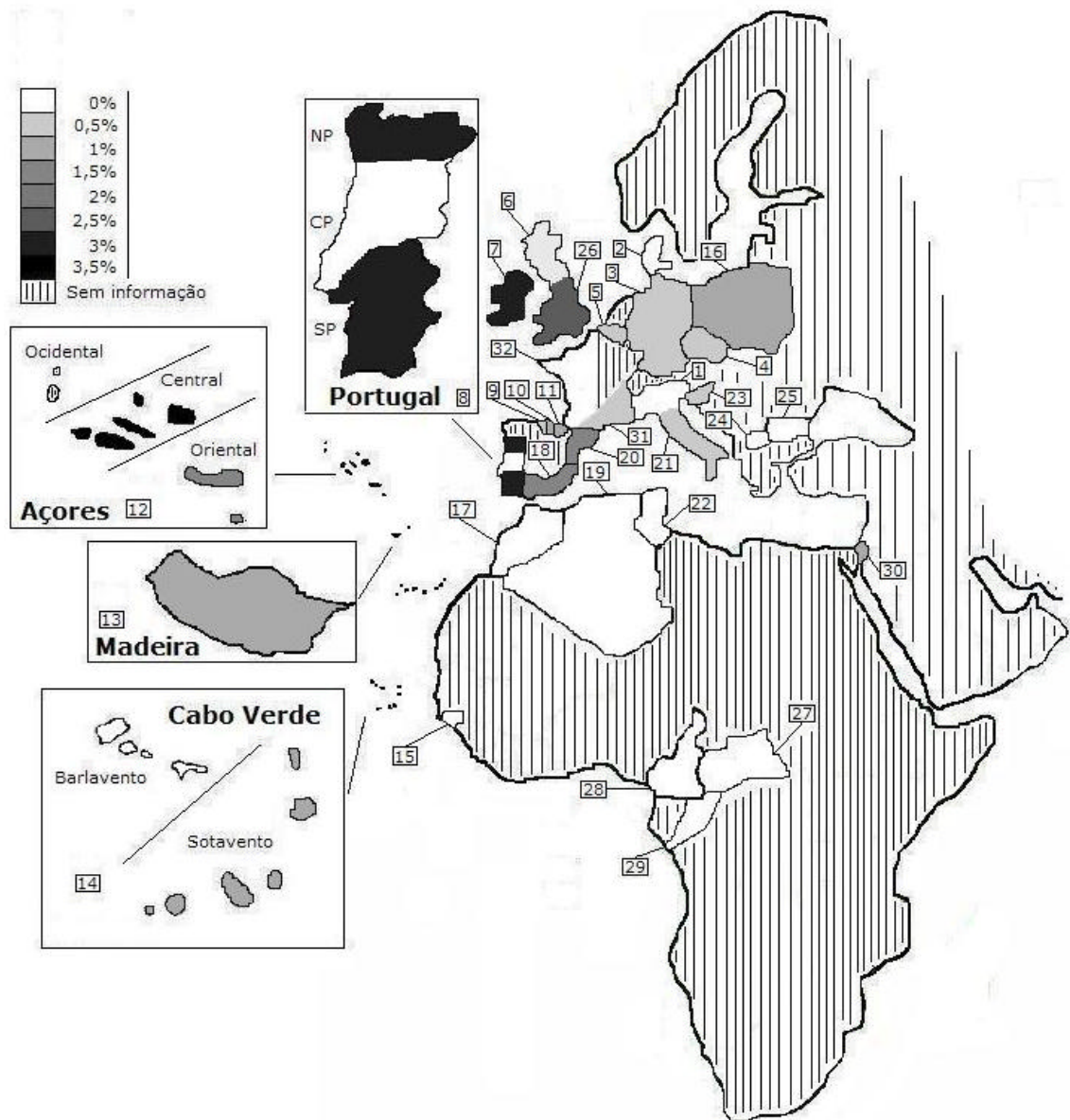


Figura 30- Distribuição geográfica das frequências do alelo DRB1*0103 no continente Europeu e Africano. 1- Itália Norte (0%); 2- Dinamarca (0%); 3- Alemanha (0,3%); 4- República Checa (0,5%); 5- Bélgica (0,5%); 6- Escócia (0,1%); 7- Irlanda do Norte (3,3%); 8- Portugal continental (NP 3,3%, CP 0%, SP 3,1%); 9- Norte de Espanha Cabuernigo (0,5%); 10- Norte de Espanha Cantabrian (1,2%); 11- Espanha Bascos (0,7%); 12- Açores (grupo Oriental 1,3%, grupo Central 3,6%); 13- Madeira (0,8%); 14- Cabo Verde (Barlavento 0%, Sotavento 0,8%); 15- Guiné-Bissau (0%); 16- Polónia (1%); 17- Marrocos (0%); 18- Sul de Espanha Múrcia (1,6%); 19- Algéria (0%); 20- Espanha Barcelona (1,4%); 21- Itália Centro (0,3%); 22- Tunísia (0%); 23- Eslovénia (0,5%); 24- Macedónia (0%); 25- Bulgária (0%); 26- Inglaterra (2,5%); 27- República Centro Africana Aka Pigmeus (0%); 28- Camarões (0%); 29- Congo (0%); 30- Israel (1,1%); 31- França Sudeste (0,4%); 32- França Oeste (0%). (Ver referências em material e métodos).

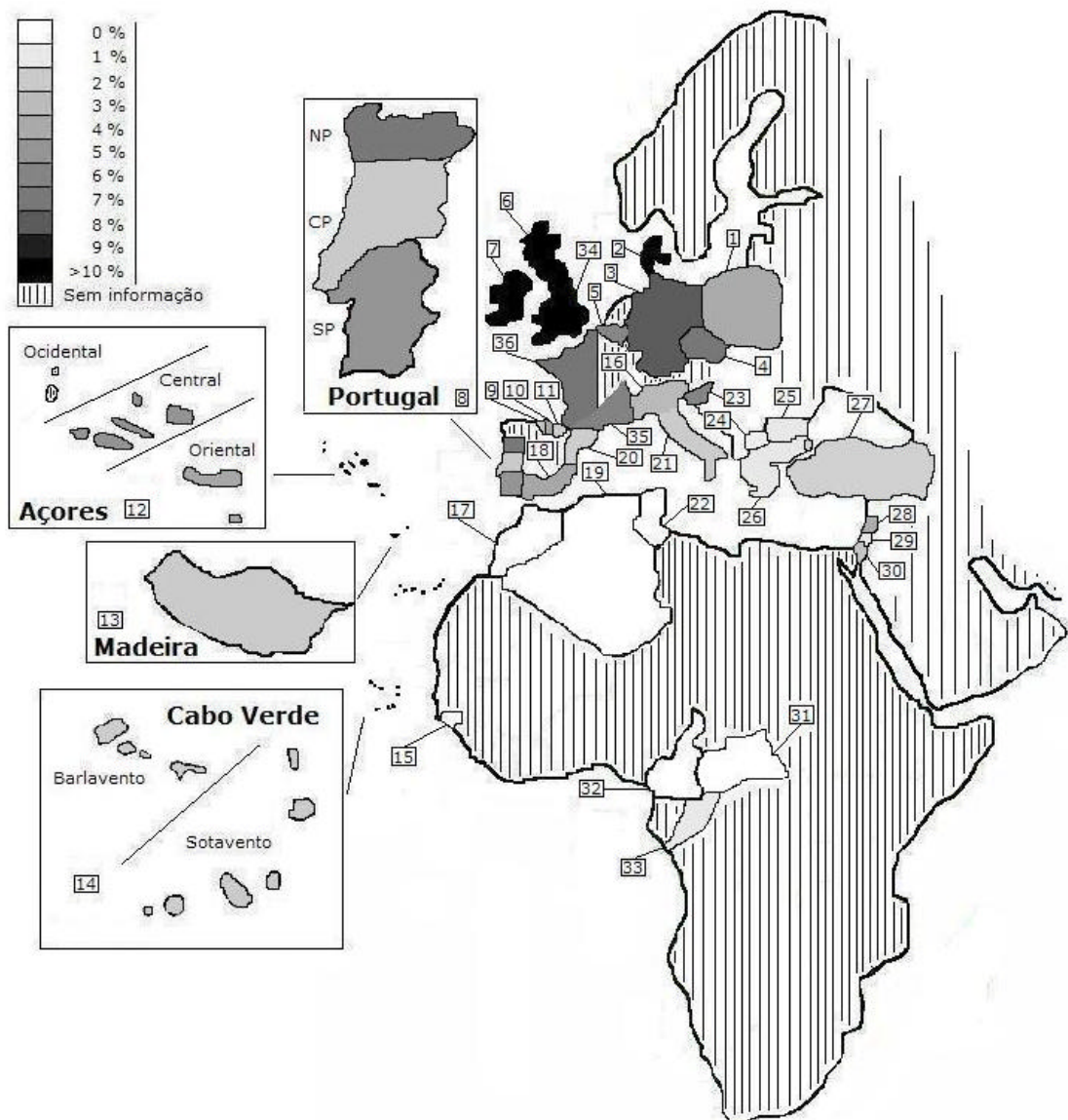


Figura 31- Distribuição geográfica das frequências do alelo DRB1*0401 no continente Europeu e Africano. 1- Polónia (3,5%); 2- Dinamarca (17,6%); 3- Alemanha (7,5%); 4- República Checa (6,7%); 5- Bélgica (6,1%); 6- Escócia (10,3%); 7- Irlanda do Norte (11,1%); 8- Portugal continental (NP 6,5%, CP 2%, SP 5,1%); 9- Norte de Espanha Cabuernigo (2,1%); 10- Norte de Espanha Cantabrian (3,6%); 11- Espanha Bascos (2%); 12- Açores (grupo Oriental 3,8%, grupo Central 4,5%); 13- Madeira (2,4%); 14- Cabo Verde (Barlavento 2,4%, Sotavento 1,6%); 15- Guiné-Bissau (0%); 16- Itália Norte (3%); 17- Marrocos (0%); 18- Sul de Espanha Múrcia (3,6%); 19- Algéria (0%); 20- Espanha Barcelona (2%); 21- Itália Centro (2%); 22- Tunísia (0%); 23- Eslovénia (6%); 24- Macedónia (1,3%); 25- Bulgária (0,9%); 26- Grécia (0,8%); 27- Turquia (2%); 28- Líbano (3,8); 29- Palestina (1,2%); 30- Israel (1,9%); 31- República Centro Africana Aka Pigmeus (0%); 32- Camarões (0%); 33- Congo (1,1%); 34- Inglaterra (12,4%) 35- França Sudeste (5,9%); 36- França Oeste (6,5%). (Ver referências em material e métodos).

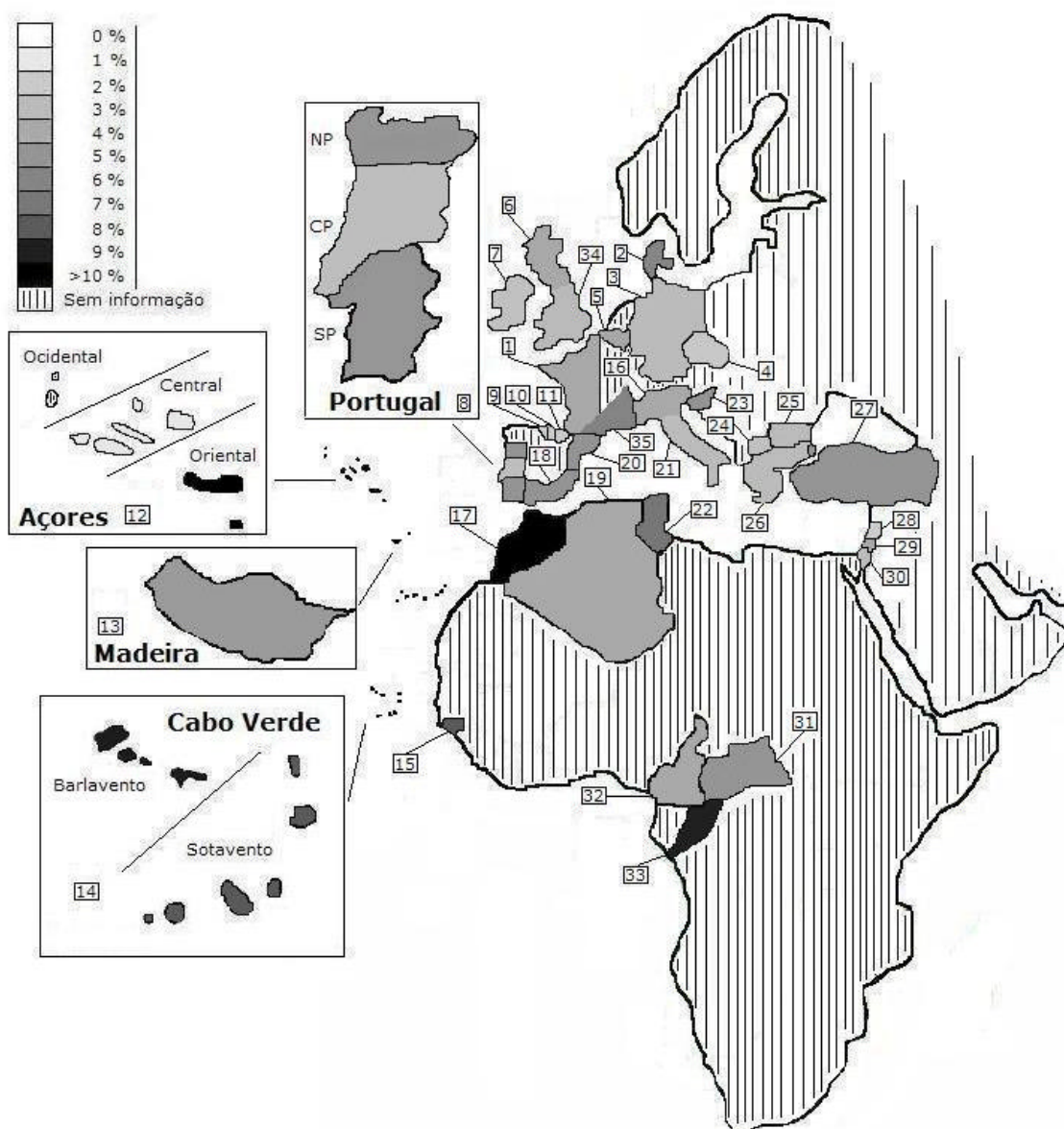


Figura 32- Distribuição geográfica das frequências do alelo DRB1*1302 no continente Europeu e Africano. 1- França Oeste (3,5%); 2- Dinamarca (5,6%); 3- Alemanha (2,6%); 4- República Checa (2,4%); 5- Bélgica (3,5%); 6- Escócia (4,3%); 7- Irlanda do Norte (3,3%); 8- Portugal continental (NP 5,4%, CP 3%, SP 5,1%); 9- Norte de Espanha Cabuernigo (1,6%); 10- Norte de Espanha Cantabrian (1,8%); 11- Espanha Bascos (2%); 12- Açores (grupo Oriental 12,8%, grupo Central 0,9%); 13- Madeira (4,6%); 14- Cabo Verde (Barlavento 8,9%, Sotavento 8,1%); 15- Guiné-Bissau (7,7%); 16- Itália Norte (3,6%); 17- Marrocos (11,1%); 18- Sul de Espanha Múrcia (5,2%); 19- Algéria (4%); 20- Espanha Barcelona (4,6%); 21- Itália Centro (2,9%); 22- Tunísia (6,5%); 23- Eslovénia (5%); 24- Macedónia (3,2%); 25- Bulgária (2,7%); 26- Grécia (3,3%); 27- Turquia (5%); 28- Líbano (1,6); 29- Palestina (3,9%); 30- Israel (2,5%); 31- República Centro Africana Aka Pigmeus (5,4%); 32- Camarões (4,3%); 33- Congo (9,4%); 34- Inglaterra (2,5%) 35- França Sudeste (5,9%). (Ver referências em material e métodos).

10.3. Polimorfismos dos genes HLA mostram diferenças no povoamento dos grupos Central e Oriental do arquipélago dos Açores.

O povoamento dos Açores, iniciado em 1432, teve o contributo de vários povos, especialmente Portugueses, Flamengos, Espanhóis, Africanos, Judeus e Italianos (Frutuoso, 1977; Guill, 1993; Mendonça, 1996). O papel central dos Açores no comércio marítimo entre a América e a Europa durante os séculos XV e XVI foi uma das principais razões para um povoamento envolvendo povos de diferentes origens.

O presente estudo descreve as frequências alélicas e haplotípicas dos genes HLA-A, HLA-B, HLA-Cw, HLA-DRB1, HLA-DQA1, e HLA-DQB1 nos Açores. Esta análise em alta resolução mostra que vários grupos HLA anteriormente caracterizados por métodos serológicos (Bruges-Armas *et al.*, 1999) são na verdade constituídos por uma variedade de grupos e alelos (tabelas 21 e 22). Esta informação adicional e completa é altamente relevante para uma caracterização da estrutura do sistema HLA.

A árvore filogenética construída com as frequências alélicas dos genes HLA-A, HLA-B, e HLA-DRB1 associa os Açores aos Asiáticos e Ameríndios (figura 9), um resultado que não está de acordo com o obtido quando apenas os genes classe I ou classe II são utilizados separadamente na construção do dendrograma. Na verdade, os dados obtidos mostram que os Açorianos são mais próximos de outros Europeus do que dos Portugueses do continente. Por outro lado, se considerarmos as frequências haplotípicas dos Açores não encontramos um estrato comum com os haplotipos específicos das populações Asiáticas e Ameríndias. Assim, os nossos resultados não estão de acordo com o trabalho anterior que revela

haplotipos Mongóis nos Açores (Bruges-Armas *et al.*, 1999) e também não nos foi possível encontrar os haplotipos mais comuns dos Ameríndios (Layrisse *et al.*, 2001; Leffell *et al.*, 2004). Os haplotipos em baixa resolução encontrados anteriormente em elevadas frequências (2% cada) e considerados de possível origem Mongol (Bruges-Armas *et al.*, 1999) foram encontrados com frequências muito mais baixas (A*0201-B*5001-DRB1*0701: 0.9%; A*2402-B*4402-DRB1*1102: 0.2%; A*2902-B*4901-DRB1*0701: 0.2%). Dois destes haplotipos foram também encontrados na Ilha da Madeira e no Sul de Portugal (Spínola *et al.*, 2002) com frequências mais elevadas do que as encontradas previamente nos Açores (Bruges-Armas *et al.*, 1999). As semelhanças entre os Açorianos e os Asiáticos, reveladas pela análise das árvores filogenéticas, reflectem mais uma composição alélica similar do que uma influência genética dessas populações.

As frequências alélicas e haplotípicas encontradas nos Açores reflectem a influência dos Portugueses e de outros Europeus no povoamento deste arquipélago. Vários haplotipos são comuns a Portugal continental (e.g. A*0101-B*0801-DRB1*0301, A*0201-B*5101-DRB1*0701, A*0201-B*1801-DRB1*1104), reflectindo a importância dos Portugueses no povoamento dos Açores. No entanto, vários haplotipos encontrados nos Açores não estão presentes em Portugal continental (e.g. A*0301-B*0702-DRB1*1501, A*0101-B*1402-DRB1*0102, A*0201-B*0702-DRB1*0101, A*0201-B*4402-DRB1*1301), alguns deles com uma origem claramente Europeia. Apenas 2 dos 9 haplotipos extensos mais frequentes (>1.2%) (tabela 24) encontrados nos Açores possuem o seu haplotipo parcial A-B-DRB1 comum a Portugal continental. Os haplotipos que não são comuns a

Portugal continental podem ter sido levados para os Açores pelos Flamengos e outros Europeus envolvidos no povoamento destas ilhas Atlânticas. A mesma conclusão foi reportada no estudo das frequências das mutações do gene *HFE* no grupo Central do arquipélago dos Açores. As frequências identificadas nas mutações C282Y ou H63D no grupo de indivíduos investigados eram semelhantes às previamente reportadas nos grupos que provavelmente povoaram os Açores (Couto *et al.*, 2003). Nenhum dos haplotipos A-B-DRB1 existentes na população da Guiné-Bissau foi encontrado nos Açores apesar do envolvimento destas ilhas Atlânticas no comércio de escravos. Alelos específicos das populações Africanas (e.g. A*0202, A*0225, B*1503, B*1516, B*2703, B*4202 e B*5703) (Cao *et al.*, 2004) foram encontrados mas a maioria em baixas frequências. Outro estudo desenvolvido nas ilhas do grupo Central encontrou também o alelo HLA-B*2703, o qual foi pela primeira vez descrito em populações Africanas e mais tarde noutras populações com mistura Africana conhecida (Couto *et al.*, 2004). Estes dados indicam provavelmente uma pequena influência genética deixada pelos escravos sub-Saharianos o que está de acordo com os dados obtidos recentemente através de estudos no cromossoma Y e DNA mitocondrial (Brehm *et al.*, 2003; Gonçalves *et al.*, 2005).

10.3.1. Diversidade entre as ilhas Açorianas

Os alelos HLA-A, -B e -DRB1 e as frequências haplotípicas calculadas com as amostras Az1 mostram uma clara heterogeneidade entre as ilhas do grupo Central e as ilhas do grupo Oriental do arquipélago dos Açores. Um dendrograma construído com as frequências alélicas separa estes dois grupos e junta as ilhas Centrais aos Asiáticos e Ameríndios (figura 10). Esta

similaridade entre as ilhas Centrais e os Asiáticos não deverá reflectir uma verdadeira influência genética entre estas populações como já foi referido anteriormente. Por outro lado, as ilhas Orientais do arquipélago dos Açores revelam uma maior afinidade nesta análise a Portugal continental. Uma árvore filogenética construída apenas com populações Europeias e Africanas (figura 11) clarifica as diferenças entre estes dois grupos de ilhas Açorianas. O grupo Oriental mostra novamente a sua ligação a Portugal continental e Norte de África mas o grupo Central apresenta-se mais próximo a outros Europeus. Estes resultados estão de acordo com o diferente perfil de haplotipos encontrado em cada um dos grupos de ilhas uma vez que apenas dois dos 26 haplotipos três-*loci* mais frequentes são comuns aos dois grupos. Os haplotipos mais frequentes encontrados num dos grupos está ausente no outro (tabela 26). As ilhas do grupo Oriental apresentam vários haplotipos frequentes que são comuns a Portugal continental, ao contrário do que acontece no grupo Central que possui apenas três desses haplotipos. É desconhecida a origem de vários haplotipos frequentes do grupo Central provavelmente devido a inexistência de uma caracterização adequada de várias populações Europeias, a partir das quais podem ter sido introduzidos. Estudos adicionais em alta resolução nos genes HLA em Holandeses, Belgas e em outras populações da Europa central poderão clarificar a origem de alguns destes haplotipos.

Tanto as árvores filogenéticas como a composição haplotípica apontam claramente para um povoamento com origens diferentes nos dois grupos de ilhas do arquipélago dos Açores. Estes dados corroboram fontes históricas que referem que os Açores foram povoados não apenas por Portugueses,

com maior presença no grupo Oriental, mas também por outros Europeus, principalmente Flamengos nas ilhas do grupo Central.

10.4. Caracterização do sistema HLA em Cabo Verde e Guiné-Bissau por sequenciação nucleotídica

A Guiné-Bissau foi fortemente influenciada por diferentes eventos históricos e pré-históricos que reuniram povos provenientes de diferentes origens do continente Africano. Em momentos diferentes, durante vários milhares de anos, esta zona da costa Oeste Africana recebeu povos do Norte de África, do Sahara e do Este Africano. A costa da Guiné foi a origem da maior parte dos escravos traficados por Portugal nos séculos XV e XVI. Estes povos Africanos eram capturados na costa da Guiné e levados para o arquipélago de Cabo Verde, o qual servia de entreposto no tráfico de escravos para os continentes Europeu e Americano (Russell-Wood, 1998). O arquipélago de Cabo Verde, depois da sua descoberta no século XV pelos Portugueses, foi povoado por escravos provenientes da costa Oeste Africana e alguns Caucasianos, principalmente provenientes da Península Ibérica.

Este estudo descreve pela primeira vez, em alta resolução, os alelos dos *loci* HLA-A, -B e -DRB1 e as frequências haplotípicas nas populações da Guiné-Bissau e Cabo Verde. Esta análise em alta resolução mostra que vários grupos genéricos dos *loci* HLA identificados anteriormente em Cabo Verde (Spínola *et al.*, 2002) são na realidade constituídos por uma variedade de alelos (tabelas 34-36). Esta informação adicional é altamente relevante para a caracterização detalhada da estrutura do sistema HLA.

A posição da Guiné-Bissau e Cabo Verde nos dendrogramas (figuras 13 e 14) reflecte claramente a sua origem sub-Sahariana. Contudo, a análise das

coordenadas principais (figura 15) mostra que estas duas populações não estão tão distantes dos Norte Africanos como outros sub-Saharianos. Na Guiné-Bissau esta situação pode resultar de antigos contactos com outros povos, nomeadamente Norte Africanos antes da desertificação do Sahara e Árabes do Este Africano através do corredor do Sahel (Ellis *et al.*, 2000). Influências mais recentes tais como o contacto com os Berberes devido à pressão Muçulmana no Noroeste Africano durante o século IX (Moreira, 1964) podem ter ocorrido. Um estudo anterior ao nível do DNA mitocondrial da população da Guiné-Bissau mostrou que 6% dos seus haplogrupos são característicos do Norte de África, Este de África, Arábia, Médio Oriente e mesmo Europa (Rosa *et al.*, 2004).

A similaridade observada entre a Guiné-Bissau e Cabo Verde é consistente com a origem bem documentada dos primeiros povoadores do arquipélago. Apesar da influência sub-Sahariana prevalente em Cabo Verde, encontramos muitos mais haplotipos comuns a Portugal (9.3%) do que à Guiné-Bissau (2.2%) denotando contribuição Caucasiana no povoamento de Cabo Verde.

A análise das coordenadas principais (figura 15) mostra que a população das ilhas do Barlavento Cabo-Verdiano são mais semelhantes aos Caucasianos do que a população das ilhas do Sotavento, em concordância com estudos anteriores em STR's e cromossoma Y (Fernandes *et al.*, 2003; Gonçalves *et al.*, 2003). De facto, estes dois grupos de ilhas são significativamente diferentes ao nível das frequências alélicas nos genes HLA estudados e cada um deles revela um perfil haplotípico específico (tabela 37). As ilhas do Sotavento possuem também mais haplotipos comuns à Guiné-Bissau (3%) e Portugal (10%) do que as ilhas do

Barlavento (1% comuns com a Guiné-Bissau e 7% com Portugal). Este número reduzido de haplotipos comuns à Guiné -Bissau pode significar que os povoadores sub-Saharianos do arquipélago de Cabo Verde não pertenciam apenas a um único grupo homogéneo. O mais baixo número de haplotipos comuns à Guiné -Bissau e a Portugal nas ilhas do Barlavento é provavelmente resultado de um efeito fundador, uma vez que estas ilhas terão sido povoadas mais tarde por escravos que fugiram das ilhas do Sotavento (Gonçalves *et al.*, 2003). Novos estudos nas populações da costa Oeste Africana poderão clarificar a origem sub-Sahariana dos povoadores do arquipélago de Cabo Verde.

10.5. Os genes HLA na ilha da Madeira (Portugal) caracterizados por sequenciação nucleotídica: marcas de diferentes origens.

O povoamento da ilha da Madeira começou em 1420 com os Portugueses e teve também a influência de povos sub-Saharianos trazidos como escravos. O Norte de África foi a primeira origem dos escravos presentes na Madeira, mas rapidamente a região sub-Sahariana da costa Oeste Africana tornou-se a origem prevalente dos povos cativos que trabalharam na agricultura e noutras actividades. Devido ao forte desenvolvimento das actividades comerciais na ilha da Madeira, para além dos Portugueses, outros povos Europeus também contribuíram para o povoamento da ilha, nomeadamente Espanhóis, Italianos, Franceses e Ingleses.

Este trabalho descreve pela primeira vez em alta resolução os alelos dos genes HLA-A, HLA-B e HLA-DRB1 e as respectivas frequências haplotípicas na ilha da Madeira. Esta caracterização em alta resolução mostra que vários

grupos de alelos caracterizados anteriormente em media resolução (Spínola *et al.*, 2002), são na verdade constituídos por uma grande variedade de alelos (tabela 44). Esta informação adicional e mais completa é altamente relevante para uma caracterização detalhada da estrutura HLA.

As árvores filogenéticas construídas com as frequências alélicas dos genes HLA-A, -B, e -DRB1 colocam a Madeira entre as populações Europeias e as Africanas, situação que suporta a hipótese de uma origem mista dos seus povoadores (figuras 16 e 17). Por outro lado, a análise das coordenadas principais (figura 18) mostra que a Madeira sofreu uma mais forte influência sub-Sahariana, devido ao comércio de escravos, do que Portugal continental e mesmo que os Açores. Estes resultados são corroborados pelos vários alelos específicos de populações sub-Saharianas encontrados na Madeira, apesar de aparecerem apenas em frequências muito baixas (e.g. B*1503, B*1516, B*5703, B*5802, A*0102, A*0202, A*3402, A*3601 e A*7403).

Analisando os haplotipos prevalentes na ilha da Madeira não foi possível encontrar ao nível dos 3 *loci*, características comuns com as populações sub-Saharianas já caracterizadas ao nível dos genes HLA. A estrutura haplotípica na Madeira mostra ser Europeia, a maior parte de origem Portuguesa (e.g. A*010101-B*0801-DRB1*030101, A*020101-B*180101-DRB1*110401, A*020101-B*510101-DRB1*130101) juntamente com alguma influência de outras populações Caucásicas (e.g. A*020101-B*0801-DRB1*030101, A*020101-B*440201-DRB1*040101, A*030101-B*070201-DRB1*150101). A influência genética directa das populações do Norte de África é difícil de discriminar uma vez que os Portugueses do continente podem ter carregado consigo essas mesmas características.

Dois dos haplotipos mais frequentes na Madeira, ausentes em Portugal continental, são comuns ao arquipélago dos Açores (A*020101-B*440201-DRB1*040101 2.7% no grupo Central dos Açores, e A*020101-B*440201-DRB1*130101 1.3% no arquipélago dos Açores). Estes resultados podem significar que os dois arquipélagos Portugueses tiveram povoadores Caucásios com origem comum, para além dos Portugueses do continente. Por outro lado, os Madeirenses estiveram envolvidos no povoamento dos Açores e estes dois haplotipos podem ter sido levados por esta via.

Os haplotipos encontrados na ilha da Madeira reflectem principalmente a influência dos Portugueses e de outros Europeus envolvidos no povoamento. Os haplotipos que não são comuns a Portugal continental podem ter sido levados para a Madeira por outros Europeus envolvidos no povoamento, nomeadamente Franceses, Italianos e Ingleses. Contudo, para além da influência Caucásica, as frequências alélicas reflectem uma influência genética sub-Sahariana claramente visível nos dendrogramas, na análise das coordenadas principais e na presença de alguns alelos e haplotipos 2-*loci* específicos das populações sub-Saharianas.

Não tendo sido encontrados haplotipos 3-*loci* específicos das populações sub-Saharianas, nomeadamente da Guiné-Bissau, pode significar que a influência genética Africana presente na Madeira provem de outras populações que ainda não foram estudadas ao nível do sistema HLA. As populações Africanas são mais distintas entre si do que as populações de outros continentes, de acordo com a teoria "Out of Africa" e vários estudos (Ellis *et al.*, 2000; Cao *et al.*, 2004). Futuros estudos ao nível dos HLA noutras populações sub-Saharianas poderão clarificar a origem da influência genética Africana na ilha da Madeira.

10.6. Discrepâncias na caracterização do sistema HLA por SSOP e SBT: um caso de estudo.

Os resultados da caracterização alélica de 621 amostras de Portugal, Cabo Verde e Guiné-Bissau detectaram alguns resultados discrepantes entre a técnica SSOP e SBT (tabelas 49 e 50). Cinquenta e sete alelos identificados por SSOP não eram concordantes com os resultados obtidos por SBT, o que significa 4,6% dos 1242 cromossomas caracterizados.

As incongruências detectadas entre a técnica SSOP e SBT neste trabalho resultam da grande sensibilidade da primeira. Estes 57 casos de falhas na técnica SSOP, assumindo que a caracterização por SBT é a que está correcta, resultam das dificuldades de interpretação decorrentes da existência de padrões de hibridação menos intensos, ausentes ou dúbios. Estas hibridações deficientes podem ser explicadas pela existência de problemas ao nível da qualidade e quantidade do DNA utilizado na amplificação por PCR e pela ocorrência de variações nas temperaturas específicas, altamente sensíveis, necessárias à hibridação das sondas no método SSOP. O processo de lavagem das membranas é outro dos passos sensíveis que, caso não seja eficaz, pode deixar algum "ruído de fundo" que torna difícil a interpretação correcta do padrão de hibridação. Por outro lado, o método SSOP apenas cobre algumas bases nucleotídicas enquanto que o SBT faz uma leitura de toda a sequência e, dessa forma, o programa de análise dos resultados SSOP não tomará em consideração tantos alelos quanto o programa utilizado na técnica SBT.

Os procedimentos do método SSOP são altamente exigentes no que diz respeito à experiência dos recursos humanos e da elevada precisão dos equipamentos. Para os laboratórios e investigadores que estão a dar os

primeiros passos no campo da caracterização alélica dos *loci* HLA, a técnica SSOP não será a mais recomendada uma vez que exigirá um grande investimento em tempo e equipamentos até atingir níveis de qualidade satisfatória. Por outro lado, o SBT baseia-se principalmente em passos automatizados fáceis de implementar e pouco sensíveis a erros técnicos e de interpretação.

10.7. Discussão global

A caracterização alélica dos genes A, B e DRB1 do sistema HLA em três populações Portuguesas (Continente, Madeira e Açores) e duas Africanas (Guiné-Bissau e Cabo Verde) permitiu estimar vários parâmetros genéticos importantes para a sua caracterização e comparação.

As diferenças encontradas entre a caracterização por SSOP e por sequenciação directa resultam na maior parte dos casos das dificuldades de interpretação dos padrões menos intensos e ambíguos resultantes da sensibilidade da primeira técnica.

Os testes de neutralidade aplicados nas cinco populações revelam, nalguns casos, a actuação da selecção natural sobre diferentes genes. Estes resultados estão de acordo com os encontrados noutros estudos (Ellis *et al.*, 2000; Piancatelli *et al.*, 2004; Cao *et al.*, 2004) e são coerentes com os mecanismos que se consideram envolvidos na manutenção da grande variabilidade genética que caracteriza o sistema HLA. A importância dos genes do sistema HLA na protecção contra doenças infecciosas (Doherty & Zinkernagel, 1975) e na interacção materno fetal (Hedrick & Thomson, 1988), entre outros, podem explicar a presença desta pressão selectiva.

A ausência de equilíbrio de Hardy-Weinberg é consistente com a ocorrência de selecção natural e de outros factores, nomeadamente fluxo genético e migração. Nas populações insulares estudadas, os *loci* HLA-DRB1 (Cabo Verde e Madeira) e HLA-B (Açores) não estavam em equilíbrio de Hardy-Weinberg, situação concordante com outros estudos que revelam ser frequente esta situação em populações isoladas, nomeadamente devido a um maior efeito da pressão selectiva (Chen *et al.*, 1999).

A diversidade alélica encontrada em cada *locus* é elevada em todas as populações estudadas, situação de acordo com a vantagem dos heterozigóticos pela sua maior capacidade no reconhecimento de um número superior de agentes patogénicos (Doherty & Zinkernagel, 1975).

A menor diversidade alélica na Guiné-Bissau, considerando os valores de heterozigotia e a distribuição alélica observados nos três *loci*, é consistente com os resultados obtidos em outras populações sub-Saharianas (Cao *et al.*, 2004) mas pode ser também resultante da dimensão da amostra, até porque a diversidade haplotípica é maior nesta população. O número elevado de alelos encontrado em Cabo Verde, comparativamente à Guiné - Bissau (tabelas 34-36), reflecte o facto do seu povoamento ter tido o contributo de várias populações Africanas e Europeias. A diversidade de contributos no povoamento dos Açores pode também justificar o maior número de alelos encontrados comparativamente a Portugal continental, embora neste caso a maior dimensão da amostra também possa ter tido influência com a detecção de vários alelos raros.

Os *loci* HLA-A, -B e -DRB1 apresentam, nas cinco populações estudadas, alguns casos de alelos relativamente frequentes pertencentes ao mesmo grupo (tabelas 5-7, 10, 21, 22, 34-36 e 44). A selecção de alelos

pertencentes ao mesmo grupo deverá ser menos intensa do que a que ocorre entre alelos de grupos diferentes, maximizando a amplitude do repertório de elementos patogénicos identificáveis pelo organismo (Gao *et al.*, 1997; Jakobsen *et al.*, 1998). Desta forma estes alelos mais frequentes pertencentes ao mesmo grupo podem ter origem em misturas entre populações ou ser resultado do efeito fundador, não sendo portanto o efeito directo da selecção natural sobre estas populações.

Os valores de *linkage disequilibrium* (D) encontrados entre os alelos que compõem os haplotipos mais frequentes (tabelas 14-20, 24-34, 37-43 e 45-48) podem resultar de vários factores. Mais do que a distância física entre os *loci*, o grau de recombinação, a selecção natural, a migração e a deriva genética são factores fundamentais na definição do *linkage disequilibrium* entre dois ou mais alelos de diferentes *loci* (Hedrick & Thomson, 1988; Sanchez-Mazas *et al.*, 2000). Os haplotipos mais frequentes, com valores de *linkage disequilibrium* mais elevados e estatisticamente significativos, constituem assim elementos importantes para traçar as influências genéticas entre as populações.

A análise dos haplotipos mais frequentes e das frequências alélicas revelam a existência de afinidades entre as populações em estudo. Portugal continental possui uma posição intermédia entre o Norte de África e a Europa, revelando as influências deixadas por várias migrações com origem nos continentes Africano e Europeu. As migrações para a Península Ibérica através do estreito de Gibraltar, forçadas pela desertificação do Sahara há cerca de 6 mil anos, poderão ser a origem de algumas das características alélicas e haplotípicas encontradas no Norte de Portugal. Estas influências, em conjunto com outras de origem Europeia, parecem constituir uma base

genética mais antiga que foi forçada a concentrar-se no Norte de Portugal devido à pressão da presença Muçulmana que se iniciou no século VIII e só terminou 700 anos depois. Por sua vez, acompanhando esta invasão Árabe vieram os Berberes que deverão ser os responsáveis pela influência Norte Africana detectada no Sul de Portugal.

Esta influência Africana em Portugal, que decorre de migrações ocorridas em vários momentos ao longo de milhares de anos, explica a proximidade às populações Marroquina e Tunisina (figuras 5 e 7).

A Madeira, povoada a partir de 1420, apresenta uma proximidade às populações Africanas semelhante à revelada por Portugal continental. Apesar do forte contributo Europeu no seu povoamento, a presença de escravos trazidos da costa Ocidental Africana deixou um contributo importante no perfil genético da população Madeirense, perfeitamente detectada ao nível das frequências alélicas.

O arquipélago dos Açores, povoado essencialmente por Europeus a partir da primeira metade do século XV, revela interessantes diferenças entre o grupo Oriental e o grupo Central. Coincidente com os registos históricos do seu povoamento, o grupo Oriental revela uma maior proximidade à população de Portugal continental enquanto que o grupo Central se aproxima mais de outras populações Europeias. No povoamento das ilhas do grupo Central estiveram envolvidos largas centenas de Europeus, essencialmente Flamengos. Apesar das análises filogenéticas que associam os Açores às populações Asiáticas, não foram encontrados alelos específicos ou haplotipos que corroborem esses resultados pelo que não é possível concluir sobre a existência de influências directas.

O arquipélago de Cabo Verde, que serviu de entreposto no comércio de escravos da costa Ocidental Africana controlado por Europeus, possui uma grande proximidade com a Guiné-Bissau. Revela, no entanto, a presença de alguma influência Caucasiana, essencialmente nas ilhas do Sotavento Cabo-Verdiano, que deverá resultar da presença dos Europeus envolvidos no comércio de escravos.

A Guiné-Bissau assemelha-se mais às populações Norte Africanas do que qualquer outra população sub-Sahariana estudada ao nível do sistema HLA, o que deve resultar de contactos com populações Berberes e Árabes do Este Africano.

10.8. Perspectivas futuras

Os resultados dos estudos de caracterização do sistema HLA noutras populações Africanas (e.g. Angola, São Tomé e Príncipe e Moçambique), que neste momento estamos a desenvolver no Laboratório de Genética Humana da Universidade da Madeira, poderão clarificar a origem de algumas frequências alélicas e haplotípicas encontradas em Portugal continental, na Madeira e nos Açores que podem ter sido introduzidas pelos escravos. A caracterização alélica e haplotípica dos principais genes do sistema HLA em mais populações Africanas constituirá ainda um contributo fundamental para a compreensão da sua própria diversidade genética e das relações existentes entre si.

A caracterização dos genes A, B e DRB1 do sistema HLA nas populações Portuguesas e Africanas, desenvolvida no âmbito desta tese de doutoramento, permite estimar a probabilidade de encontrar em cada uma das populações um dador compatível com um determinado paciente que

necessite de um transplante. Tendo em conta que muitos dos alelos dos diferentes *loci* estão em *linkage*, a frequência de um determinado fenotipo não deve ser estimada a partir das frequências alélicas. A probabilidade de encontrar o fenotipo pretendido é igual à soma das probabilidades dos pares de haplotipos que podem constituir esse mesmo fenotipo. A probabilidade de ocorrência de cada um desses pares de haplotipos é igual a duas vezes o produto da frequência de cada um desses dois haplotipos, quando são diferentes entre si, ou, simplesmente, o quadrado da frequência do haplotipo quando os dois são iguais (Schipper *et al.*, 1996).

A caracterização alélica dos genes HLA-A, -B e -DRB1 na população Portuguesa, Cabo-Verdiana e Guineense constitui ainda a base para a concretização de vários estudos de grande importância médica. A título de exemplo destacam-se os estudos sobre a associação entre estes genes e o desenvolvimento de alergias (Boehncke *et al.*, 1998), de espondilartrites (Calin, 2002), de artrite reumatóide (Gibert *et al.*, 2003), de diabetes tipo1 (Larsen & Alper, 2004) e de abortos espontâneos recorrentes (Hedrick & Thomson, 1988; Alberts & Ober, 1993). Neste momento está já em desenvolvimento o estudo das Espondilartrites nos Arquipélagos da Macaronésia com o objectivo de estabelecer a sua associação com os alelos HLA, particularmente o HLA-B27, nestas populações. Estes estudos relativos à associação entre doenças auto-imunes e os genes HLA nas populações Portuguesa, Cabo-Verdiana e Guineense só são possíveis tendo por base a caracterização alélica e haplotípica já desenvolvida e apresentada nesta tese. Esta informação constitui uma referência imprescindível para a definição dos alelos dos diferentes *loci* HLA que influenciam a predisposição para o desenvolvimento de determinadas doenças.

Os dados aqui apresentados permitem ainda conhecer a prevalência, nas populações estudadas, dos alelos e haplotipos que se sabem possuir uma associação forte a determinadas doenças auto-imunes. Desta forma é possível ter uma ideia da maior ou menor susceptibilidade genética da população a essas mesmas doenças e desenvolver medidas preventivas, nomeadamente a redução de factores de risco quando aplicável.

Com a experiência técnico-científica obtida através dos trabalhos de caracterização do sistema HLA nas populações em estudo, o Laboratório de Genética Humana da Universidade da Madeira fica capacitado para auxiliar o Sistema Regional de Saúde, ou outros, no diagnóstico e identificação de indivíduos com elevado risco de desenvolver estas doenças auto-imunes. Da mesma forma, criaram-se ainda condições técnico-científicas para que a Região Autónoma da Madeira venha a desenvolver um registo regional de dadores de medula óssea que, em sintonia com o registo nacional e internacional, possa contribuir para aumentar a probabilidade de encontrar dadores compatíveis necessários para o tratamento de Leucemias e outras doenças hematológicas.

Os dados obtidos e as capacidades desenvolvidas permitem ainda associar os marcadores HLA aos procedimentos do Laboratório de Genética Humana da Universidade da Madeira no desenvolvimento de testes de paternidade. Pelo facto da taxa de mutação nos *loci* HLA ser inferior à que ocorre nos STR, a combinação destes dois sistemas na resolução de disputas de paternidade oferece resultados mais fiáveis, conferindo maior informação e diminuição das probabilidades de falsas exclusões devido a mutações (Grubic *et al.*, 2004).

11. Referências bibliográficas

- Alberts SC & Ober C (1993). Genetic variability in the major histocompatibility complex: a review of non-pathogen-mediated selective mechanisms. *Yearbook of Physical Anthropology* 36: 71-89.
- Alimen H. (1987). Evolution du climat et des civilisations depuis 40000 ans du nord au sud du Sahara occidental. (Premières conceptions confrontées aux données récentes). *Bull. L'Assoc. Franç. L'Etude Quaternaire* 4: 215-227.
- Al Jabri AA. (2002). HLA and in vitro susceptibility to HIV infection. *Molecular Immunology* 38: 959-967.
- Amiel JL. (1967). Study of the leukocyte phenotypes in Hodgkin's disease. In: Curtoni ES, Mattiuz PL, Tosi RM, (eds.). *Histocompatibility testing*. Munksgaard. Copenhagen. Pp. 79-81.
- Anjos S & Polychronakos C. (2004). Mechanisms of genetic susceptibility to type I diabetes: beyond HLA. *Molecular Genetics and Metabolism* 81: 187-195.
- Armas J, Couto R, Santos M, Bettencourt B. (2004). HLA-A, -B, -Cw, -DQA1, -DQB1 and -DRB1 alleles in a population from the Azores. *Human Immunology* 65: 862-864.
- 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, Martínez-Lazo J, Gómez-Casado E, Díaz-Campo N, Santos 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, 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, Dimitroski K, Pacho A, Moscoso J, Gómez-Casado E, Silvera-Redondo C, Varela P, Blagoevska M, Zdravkovska V, Martínez-Laso J. (2001a). HLA genes in Macedonians and the sub-Saharan origin of Greeks. *Tissue Antigens* 57: 118-127.
- Arnaiz-Villena A, Karin M, Bendikuzė N, Gomez-Casado E, Moscoso J, Silvera C, Oguz FS, 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.

- Arnaiz-Villena A, Gomez-Casado E, Martinez-Laso J. (2002). Population genetic relationships between Mediterranean populations determined by HLA allele distribution and a historic perspective. *Tissue Antigens* 60: 111-121.
- Ayed K, Ayed-Jendoubi S, Sfar I, Labonne MP, Gebuhrer L. (2004). HLA class-I and HLA class-II phenotypic, gene and haplotypic frequencies in Tunisians by using molecular typing data. *Tissue Antigens* 64: 520-532.
- Bach FH & Hirschhorn K. (1964). Lymphocyte interaction: a potential histocompatibility test in vitro. *Science* 142: 813-814.
- Bain B, Vaz MR, Lowenstein L. (1964). The development of large immature mononuclear cells in mixed lymphocyte culture. *Blood* 23: 108-116.
- Bannerman DA & Mary W. (1966). Birds of the Atlantic islands. A history of the birds of the Azores. (Vol. 3). Oliver & Boyd. Edinburgh.
- Bender K, Bissbort S, Hiller C, Mayerová A, Wienker TF. (1983). On regional mapping of human chromosome 6. Review and own findings. *Acta Anthropogenetica* (New Delhi) 7: 85-105.
- Benjamin R & Parham P. (1990). Guilt by association: HLA-B27 and ankylosing spondylitis. *Immunol Today*. 11: 137-142.
- Barcellos LF, Oksenberg JR, Begovich AB, Martin ER, Schmidt S, Vittinghoff E, Goodin DS, Pelletier D, Lincoln RR, Bucher P, Swerdlin A, Pericak-Vance MA, Haines JL, Hauser SL; Multiple Sclerosis Genetics Group. (2003). HLA-DR2 dose effect on susceptibility to multiple sclerosis and influence on disease course. *Am. J. Hum. Genet.* 72: 710–716.
- 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
- Berah M & Dausset J. (1963). The use of the inverted microscope in leuco-agglutination. *Vox Sang.* 8: 371-375.
- Bingulac-Popovic J, Figueroa F, Sato A, Talbot WS, Johnson SL, Gates M, Postlethwait JH, Klein J. (1997) Mapping of mhc class I and class II regions to different linkage groups in the zebrafish, *Danio rerio*. *Immunogenetics* 46: 129-134.
- Birinci A, Bilgici A, Kuru O, Durupinar B. (2005). HLA-B27 polymorphism in Turkish patients with ankylosing spondylitis. *Rheumatol Int.* Jul 20: (Epub ahead of print).
- Bjorkman PJ, Saper MA, Samraoui B, Bennet WS, Strominger JL, Wiley DC. (1987). Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* 329: 506-512.

- Blanco-Gelaz MA, López-Vázquez A, García-Fernández S, Martínez-Borra J, González S, López-Larrea C. (2001). Genetic variability, molecular evolution, and geographic diversity of HLA-B27. *Human Immunology* 62: 1042-1050.
- Bodmer JG, Marsh SGE, Albert E, Bodmer WF, Dupont B, Erlich HA, Mach B, Mayr WR, Parham P, Sasasuki T, Schreuder GMT, Strominger JL, Svejgaard A, Terasaki PI. (1991). Nomenclature for factors of the HLA system, 1990. *Tissue Antigens* 37: 97–104.
- Bodmer W. (1972). Evolutionary significance of the HLA system. *Nature* 237: 139–145.
- Boehncke WH, Loeliger C, Kuehnl P, Kalbacher H, Böhm 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.
- Borelli SD, Souza D, Freitas KFF, Botini FF, Melo FC, Bedendo J. (2004). Comparative study between serological and PCR-SSP methods to determine class II HLA molecules. *Acta Scientiarum. Health Sciences Maringá* 26: 337.
- Braun J, Bollow M, Remlinger G, Eggens U, Rudwaleit M, Distler A, Sieper J. (1998). Prevalence of spondylarthropathies in HLA-B27 positive and negative blood donors. *Arthritis Rheum* 41: 58-67.
- Brehm A, Pereira L, Bandelt HJ, Prata MJ, Amorim A. (2002). Mitochondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of Atlantic slave trade. *Ann. Hum. Genet.* 66: 49-60.
- Brehm A, Pereira L, Kivisild T, Amorim A. (2003). Mitochondrial portraits of the Madeira and Açores archipelagos witness different genetic pools of its settlers. *Hum Genet* 114: 77-86.
- Brown JH, Jardetsky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, Wiley DC. (1993). Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* 364: 33-39.
- 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 Azores archipelago: possible presence of Mongoloid genes. *Tissue Antigens* 54: 349-359.
- Bunce M, Barnardo M, Browning M, Morris P, Welsh K. (1995) Phototyping: comprehensive DNA typing for HLA-A, B, C, DRB1, DRB3, DRB4, DRB5 & DQB1 by PCR with 144 primer mixes utilizing sequence-specific primers (PCR-SSP). *Tissue Antigens* 46: 355-367.
- Calin A. (2002). Définir l'évolution de la spondylarthrite ankylosante: d'où venons-nous, où en sommes-nous, où allons-nous ? *Rev Rhum* 69: 183-187.

- Cao K, Hollenbach J, Shi X, Shi W, Chopek M, Fernández-Viña MA. (2001). Analysis of the frequencies of HLA-A, B, and C alleles and haplotypes in the five major ethnic groups of the United States reveals high levels of diversity in these *loci* and contrasting distribution patterns in these populations. *Human Immunology* 62: 1009-1030.
- 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-Viña 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.
- Carreira A. (1983). Migrações nas ilhas de Cabo Verde. (2nd ed). Instituto Caboverdeano do Livro. Lisboa.
- Chakraborty R. (1990). Mitochondrial DNA polymorphism reveals hidden heterogeneity within some Asian populations. *Am. J. Hum. Genet.* 47: 87- 94.
- Chabas D, Taheri S, Renier C, Mignot E. (2003). The genetics of narcolepsy. *Annu. Rev. Genomics Hum. Genet.* 4: 459–483.
- Chen JJ, Hollenbach JA, Trachtenberg EA, Just JJ, Carrington M, Ronningen KS, Begovich A, King MC, McWeeney S, Mack SJ, Erlich HA, Thomson G. (1999). Hardy-Weinberg testing for HLA class II (DRB1, DQA1, DQB1, and DPB1) *loci* in 26 human ethnic groups. *Tissue Antigens* 54: 533-542.
- 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.
- Cicciarelli JC. (2004). HLA typing immunogenetics and transplantation. *Current Opinion in Organ Transplantation* 9: 1-7.
- Clayton J, Lonjou C, Whittle D. (1997). Allele and haplotype frequencies for HLA *loci* in various ethnic groups. *In: Charron D. (ed.). Genetic diversity of HLA. Functional and medical implications.* EDK. Paris. Pp. 665-820.
- Cope AP & Sonderstrup G. (1998). Evaluating candidate autoantigens in rheumatoid arthritis. *Springer Semin Immunopathol.* 20: 23–39.
- Couto AR, Peixoto MJ, Garrett F, Laranjeira F, Cipriano T, Armas JB. (2003). Linkage disequilibrium between S65C *HFE* mutation and HLA A29-B44 haplotype in Terceira island, Azores. *Hum Immunol.* 64: 625-628.
- Couto AR, Gonzalez T, Spínola H, Santos MR, Gantes Mora M, Brehm A, Peixoto MJ, Garrett F, Bruges Armas J. (2004). HLA-B*2703 in Azorean Caucasoids. *Genes Immun* 5: 233.

- Daniel WW. (1987). *Biostatistics: a foundation for analysis in the health sciences*. John Wiley & Sons. London.
- 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.
- Dausset J. (1958). Iso-leuco-anticorps. *Acta Haematologica* 20: 156-166.
- Dempster A, Laird N, Rubin D. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B* 39: 1-38.
- Doherty PC & Zinkernagel RM. (1975). A biological role for the major histocompatibility antigens. *Lancet* 1: 1406-1409.
- Donadi EA, Cássia M, Santos P, Nepom GT. (2000). HLA-DR and HLA-DQ typing: A comparative study using serology and restriction fragment length polymorphism (RFLP) analysis. *Medicina Ribeirão Preto* 33: 27.
- Dupont B, (ed.). (1988). *Immunobiology of HLA. Histocompatibility testing 1987*. Springer-Verlag. New York.
- Dutour O, Vernet R, Aumassip G. (1988). Le peuplement pré historique du Sahara. In: Aumassip G. et al. (eds): *Milieux, hommes et techniques du Sahara pré historique*. Problèmes actuels: L'Harmattan. Paris. Pp. 39-52.
- Ejele OA & Nwauche CA. (2004). Determination of paternity disputes in the niger delta region of Nigeria. *Niger Postgrad Med J.* 11: 187-189.
- Elder JT, Nair RP, Guo SW, Henseler T, Christophers E, Voorhees JJ. (1994). The genetics of psoriasis. *Arch Dermatol* 130: 216-224.
- Ellis JM, Mack SJ, Leke RFG, Quakyi I, Johnson AH, Hurley CK. (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.
- Ewens WJ. (1972). The sampling theory of selectively neutral alleles. *Theor. Pop. Biol.* 3: 87-112.
- 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 & Slatkin M. (1995). Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol. Biol. Evol.* 12: 921-927.
- Farjadian S, Naruse T, Kawata H, Ghaderi A, Bahram S, Inoko H. (2004). Molecular analysis of HLA allele frequencies and haplotypes in Baloch of

- Iran compared with related populations of Pakistan. *Tissue Antigens* 64: 581-587.
- Felsenstein J. (2004). PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington. Seattle.
- Fernandes A & Brehm A. (2003). Y-chromosome STR haplotypes in the Azores archipelago (Portugal). *Forensic Sci. Int.* 135: 239-242.
- Fernandes AT, Brehm A, Alves C, Gusmão L, Amorim A. (2001). Genetic profile of the Madeira archipelago population using the new Power Plex2 16 System Kit. *Forensic Sci. Int.* 3228: 1-3.
- Fernandes AT, Velosa R, Jesus J, Carracedo A, Brehm A. (2003). Genetic differentiation of the Cabo Verde archipelago population analysed by STR polymorphisms. *Ann. Hum. Genet.* 67: 340-347.
- Ferraz AS, Saber LS, Voltarelli JC, Mytilineos J, Opelz G, Donadi EA. (2002). Comparative study of HLA-DR typing by serology and sequence-specific primer analysis in a genetically highly diverse population of kidney transplant recipients. *Transplantation Proceedings* 34: 463-465.
- Frutuoso G. (1977). Saudades da terra (written in 1580-1591). Livro IV, vol I. Edição Instituto Cultural de Ponta Delgada. Ponta Delgada.
- Gao X, Lester S, Boettcher B, McCluskey J. (1997). Diversity of HLA genes in populations of Australia and the Pacific. In: Charron D (ed.). *HLA genetic diversity of HLA. Functional and medical implication*. Vol.I. Pp 298-306.
- Garrigan D & Hedrick PW. (2003). Perspective: Detecting adaptive molecular polymorphism: lessons from the MHC. *Evolution* 57: 1707–1722.
- Gatti E & Pierre P. (2003). Understanding the cell biology of antigen presentation: the dendritic cell contribution. *Current Opinion in Cell Biology* 15: 468–473.
- Gibert M, Balandraud N, Touinssi M, Mercier P, Roudier J, Revirion D. (2003). Functional categorization of HLA-DRB1 alleles in rheumatoid arthritis: the protective effect. *Human Immunology* 64, 930–935.
- Godinho VM. (1965). Os descobrimentos e a economia mundial. (Vol. 1-2). Arcádia. 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.

- 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 Science International* 129: 78-80.
- Gonçalves R, Rosa A, Freitas A, Fernandes A, Kivisild T, Villems R, Brehm A. (2003). Y-Chromosome lineages in Cabo Verde islands witness the diverse geographic origin of its first male settlers. *Hum. Gent.* 113: 467-472.
- Gonçalves R, Freitas A, Branco M, Rosa A, Fernandes AT, Zhivotovsky LA, Underhill PA, Kivisild T, Brehm A. (2005). Y-chromosome lineages from Portugal, Madeira and Azores record elements of Sephardim and Berber ancestry. *Annals Human Genetics* 69: 443-454.
- González A, Brehm A, Pérez J, Maca-Meyer N, Flores C, Cabrera V. (2003). Mitochondrial DNA affinities at the Atlantic fringe of Europe. *American Journal of Physical Anthropology* 120: 391-404.
- Gonzalez S, Garcia-Fernandez S, Martinez-Borra J, Blanco-Gelaz MA, Rodrigo L, del Río JS, López-Vazquez A, Torre-Alonso JC, López-Larrea C. (2002). High variability of HLA-B27 alleles in ankylosing spondylitis and related spondyloarthropathies in the population of Northern Spain. *Human Immunology* 63: 673–676.
- Gorer PA. (1937). The genetic and antigenic basis of tumour transplantation. *Journal of Pathology & Bacteriology* 44: 691-697.
- Grubic Z, Stingl K, Martinez N, Palfi B, Brkljacic-Kerhin V, Kastelan A. (2004). STR and HLA analysis in paternity testing. *International Congress Series* 1261: 535– 537.
- Guill JH. (1993). A History of the Azores islands. Golden Shield International. Tulare. Pp 124-127.
- Hardy GH. (1908). Mendelian proportions in a mixed population. *Science* 28: 49-50.
- Healey D, Ozegbe P, Arden S, Chandler P, Hutton J, Cooke A. (1995). In vivo activity and in vitro specificity of CD4+ Th1 and Th2 cells derived from the spleens of diabetic NOD mice. *J Clin Invest.* 95: 2979-85.
- Hedrick PW & Thomson G. (1983). Evidence for balancing selection at HLA. *Genetics* 104: 449-456.
- Hedrick PW & Thomson G. (1988). Maternal-fetal interactions and the maintenance of HLA polymorphism. *Genetics* 119: 205-212.
- Hedrick PW, Whittam TS, Parham P. (1991). Heterozygosity at individual amino acid sites: Extremely high levels for HLA-A and HLA-B genes. *Proc Natl Acad Sci USA* 88: 5897-5901.

- Hedrick PW. (1999). Balancing selection and MHC. *Genetica* 104: 207-14.
- Hill AV, Allsopp CE, Kwiatkowski D, Anstey NM, Twumasi P, Rowe PA, Bennett S, Brewster D, McMichael AJ, Greenwood BM. (1991). Common West African HLA antigens are associated with protection from severe malaria. *Nature* 352: 595-600.
- Hillert J & Olerup O. (1993). HLA and MS. *Neurology* 43: 2426-2427.
- Howell WM, Evans PR, Spellerberg MB, Wilson PJ, Smith JL. (1989). A comparison of serological, cellular and DNA-RFLP methods for HLA matching in the selection of related bone marrow donors. *Bone Marrow Transplant.* 4: 63-68.
- Hughes AL & Nei M. (1988). Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 335: 167-170.
- Huttley GA, Smith MW, Carrington M, O'Brien SJ. (1999). A scan for linkage disequilibrium across the human genome. *Genetics* 152: 1711-1722.
- Imanishi T, Akaza T, Kimura A, Tokunaga K, Gojobori T. (1992). Allele and haplotype frequencies for HLA and complement loci in various ethnic groups. *In: Tsuji K, Aizawa M, Sasazuki T (eds.). HLA 1991. (Vol. 1). Oxford University Press. Oxford. Pp. 1065-220.*
- 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.
- Jaini R, Kaur G, Mehra NK. (2002). Heterogeneity of HLA-DRB1*04 and its associated haplotypes in the North Indian population. *Human Immunology* 63: 24-29.
- Jakobsen IB, Gao X, Eastaugh S, Chelvanayagam G. (1998). Correlating sequence variation with HLA-A allelic families: implications for T cell receptor binding specificities. *Immunology and Cell Biology* 76: 135-142.
- Kasahara M, Hayashi M, Tanaka K, Inoko H, Sugaya K, Ikemura T, Ishibashi T. (1996). Chromosomal localization of the proteasome Z subunit gene reveals an ancient chromosomal duplication involving the major histocompatibility complex. *Proc Natl Acad Sci USA* 93: 9096-9101.
- Kasahara M, Nakaya J, Satta Y, Takahata N. (1997). Chromosomal duplication and the emergence of the adaptive immune system. *Trends in Genetics* 13: 90-92.
- Katz JD, Benoist C, Mathis D. (1995). T helper cell subsets in insulin-dependent diabetes. *Science* 268: 1185-1188.
- Khan MA. (2000). HLA-B27 polymorphism and association with disease. *J Rheumatol* 27: 1110-1114.

- Kippax PJ, Sheldon S, Dyer PA. (1985). A simple and rapid method for the detection of lymphocytotoxic antibodies using cell panels frozen on Terasaki plates. *Tissue Antigens* 26: 318-322.
- Klein J & Sato A. (2000). The HLA system. First of two parts. *New England Journal of Medicine* 343: 702-709.
- 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.
- Larsen CE & Alper CA. (2004). The genetics of HLA-associated disease. *Current Opinion in Immunology* 16: 660–667.
- Layrisse Z, Guedez Y, Domínguez E, Paz N, Montagnani S, Matos M, Herrerra F, Ogando V, Balbas O, Rodrigues-Larralde A. (2001). Extended HLA haplotypes in a Carib Amerindian population: The Yucpa of the Perija Range. *Hum. Immunol.* 62: 992-1000.
- Le Bouteiller P. (1997). HLA-G: on the track of immunological functions. *Eur J Immunogenet* 24: 397-408.
- Leffell MS, Fallin MD, Hildebrand WH, Cavett JW, Iglehart BA, Zachary AA. (2004). HLA alleles and haplotypes among the Lakota Sioux: Report of the ASHI minority workshops, part III. *Hum. Immunol.* 65: 78-89.
- Li N & Stephens M. (2003). Modelling *linkage disequilibrium* and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* 165: 2213-2233.
- Luo M, Embree J, Ramdahin S, Ndinya-Achola J, Njenga S, Bwayo JB, Pan S, Mao X, Cheang M, Stuart T, Brunham RC, Plummer FA. (2002). HLA-A and HLA-B in Kenya, Africa: Allele frequencies and identification of HLA-B*1567 and HLA-B*4426. *Tissue Antigens* 59: 370-380.
- Lyon E & Frank EL. (2001). Hereditary hemochromatosis since discovery of the *HFE* gene. *Clinical Chemistry* 47: 1147–1156.
- Mack SJ, Bugawan TL, Moonsamy PV, Erlich JA, Trachtenberg EA, Paik YK, Begovich AB, Saha N, Beck HP, Stoneking M, Erlich HA. (2000). Evolution of Pacific/Asian populations inferred from HLA class II allele frequency distributions. *Tissue Antigens* 55: 383-400.
- Malissen M, Malissen B, Jordan BR. (1982). Exon/intron organization and complete nucleotide sequence of an HLA gene. *Proc. Natl. Acad. Sci. U.S.A.* 79: 893–897.
- Marker-Hermann E & Hoehler T. (1998). Pathogenesis of human leukocyte antigen B27-positive arthritis: information from clinical materials. *Rheum Dis Clin North Am* 24: 865-881.

- Marsh SGE, Albert ED, Bodmer WF, Bontrop RE, Dupont B, Erlich HA, Geraghty DE, Hansen JA, Mach B, Mayr WR, Parham P, Petersdorf EW, Sasazuki T, Schreuder GMT, Strominger JL, Svejgaard A, Terasaki PI. (2002). Nomenclature for factors of the HLA system, 2002. *Tissue Antigens* 60: 407–464.
- Marsh SGE, Albert ED, Bodmer WF, Bontrop RE, Dupont B, Erlich HA, Geraghty DE, Hansen JA, Hurley CK, Mach B, Mayr WR, Parham P, Petersdorf EW, Sasazuki T, Schreuder GMT, Strominger JL, Svejgaard A, Terasaki PI, Trowsdale J. (2005). Nomenclature for factors of the HLA system, 2004. *Tissue Antigens* 65: 301–369.
- 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.
- McFarlin DE & McFarland HF. (1982a). Multiple sclerosis (first of two parts). *N. Engl. J. Med.* 307: 1183–1188.
- McFarlin DE & McFarland HF. (1982b). Multiple sclerosis (second of two parts). *N. Engl. J. Med.* 307: 1246–1251.
- McGinnis MD, Conrad MP, Bouwens AGM, Tilanus MGJ, Kronick MN. (1995) Automated solid-phase sequencing of DRB region genes using T7 sequencing chemistry and dye-labeled primers. *Tissue Antigens* 46: 173-179.
- Mendonça L. (1996). História dos Açores. Visão geral (séc.XV-XIX). Edição Centro de Apoio Tecnológico à Educação. Ponta Delgada.
- MHC sequencing consortium. (1999). Complete sequence and gene map of a human major histocompatibility complex. *Nature* 401:921-923.
- Mickelson E, Smith A, McKinney S, Anderson G, Hansen JA. (1993). A comparative study of HLA-DRB1 typing by standard serology and hybridization of non-radioactive sequence-specific oligonucleotide probes to PCR-amplified DNA. *Tissue Antigens* 41:86-93.
- Middleton D, Manavalan JS, Fernandez-Vina MA, (eds.). (2004). *Human Immunology* 65: 855-1272.
- Middleton D. (2000). PCR-SSOP class I and class II (DRB1). In: Land GA, Strothman RM, Hahn AB. (eds.). *ASHI Laboratory Manual*, 4th edition. American Society for Histocompatibility and Immunogenetics. Lenexa.

- Middleton D, Williams F, Hamill MA, 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.
- Middleton D & Williams F. (1997). A history of DNA typing for HLA. In: *HLA 1997*. Terasaki and Gjertson (Eds.). UCLA Tyssue Typing Laboratory. Los Angeles. California. Pp 61.
- Middleton D, Menchaca E, Rood H, Komerofsky H. (2003). New allele frequency database: <http://www.allelefreqencies.net>. *Tissue Antigens* 61: 403-407.
- Miller MM, Goto R, Bernot A, Zoorob R, Auffray C, Bumstead N, Briles WE. (1994). Two Mhc class I and two Mhc class II genes map to the chicken Rfp-Y system outside the B complex. *Proc Natl Acad Sci USA* 91:4397-4401.
- Mitsunaga S, Tokunaga K, Kashiwase K, Akaza T, Tadokoro K, Juji T. (1998). A nested PCR-RFLP method for high-resolution typing of HLA-A alleles. *Eur J Immunogenet.* 25: 15-27.
- Mizuki N, Kimura M. (1996). Gene structure of the human MHC region. *Nippon Rinsho* 54:1705-17.
- Modiano D, Luoni G, Petrarca V, Sodiomon Sirima B, De Luca M, Simpore J, Coluzzi M, Bodmer JG, 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 PH, de Knijff P, Schipper RF, Kazemi M, Lambooij SLE, van der Meer R, Naipal A, Pesonen N, van der Slik A, Gliphart MJ. (1998). Selective co-evolution of the D6STNFA microsatellite region with HLA class I and II *loci*. *Tissue Antigens* 52: 213-219.
- Monod T. (1991). Spectacles de la Terre. The Azores. Selection du Reader's Digest. Paris.
- Moore BS, Ploeger NA, DeGoey SR. (1997). HLA antibody screening: comparison of a solid phase enzyme-linked immunoassay with antiglobulin-augmented lymphocytotoxicity. *Transplantation* 64: 1617-1620.
- Moreira JM. (1964). Os Fulas da Guiné Portuguesa na panorâmica geral do mundo Fula. *Boletim Cultural a Guiné Portuguesa* 19: 417-432.
- Morison SE. (1965). Portuguese voyages to America in the fifteenth century. Octagon Books. New York.
- Mullis KB & Faloona F. (1987). Specific synthesis of DNA *in vitro* via a polymerase catalysed chain reaction. *Meth Enzymol* 155: 335-350.

- Muro M, Marín L, Torío A, Moya-Quiles M, Minguela A, Rosique-Roman J, Sanchis M, Garcia-Calatayud M, Garcia-Alonso A, Ivarz-López M. (2001). HLA polymorphism in the Murcia population (Spain): in the cradle of the archeological Iberians. *Hum. Immunology* 62: 910-921.
- Mytilineos J, Scherer S, Opelz G. (1990). Comparison of RFLP-DR beta and serological HLA-DR typing in 1500 individuals. *Transplantation* 50: 870-873.
- Nei M & Gojobori T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* 3: 418-426.
- Nei M. (1987). Molecular evolutionary genetics. Columbia University Press. New York. Pp180.
- Olerup O & Zetterquist H. (1992). HLA-DR typing by PCR amplification with sequence-specific primers (PCR-SSP) in 2 hours: an alternative to serological DR typing in clinical practice including donor-recipient matching in cadaveric transplantation. *Tissue Antigens* 39: 225-235.
- Olerup O & Zetterquist H. (1993). DR "low-resolution" PCR-SSP typing – a correction and an up-date. *Tissue Antigens* 41: 55-56.
- Ota M, Seki T, Fukushima H, Tsuji K, Inoko H. (1992). HLA-DRB1 genotyping by modified PCR-RFLP method combined with group-specific primers. *Tissue Antigens* 39:187-202.
- Page RDM. (1996). TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* 12: 357-358.
- Parham P, Lomen CE, Lawlor DA, Ways JP, Holmes N, Coppin HL, Salter RD, Wan AM, Ennis PD. (1988). Nature of polymorphism in HLA-A, -B, and -C molecules. *Proc Natl Acad Sci USA* 85: 4005-4009.
- Pereira ECN. (1989). Ilhas de Zarco (4th edition, Vol. II). Câmara Municipal Funchal. Funchal.
- Pereira L, Prata MJ, Amorim A. (2000). Diversity of mtDNA lineages in Portugal: not a genetic edge of European variation. *Ann. Hum. Genet.* 64, 491-506.
- Piancatelli D, Canossi A, Aureli A, Oumhani K, Del Beato T, Di Rocco M, Liberatore G, Tessitore A, Witter K, El Aouad R, Adorno D. (2004). Human leukocyte antigen-A, -B, and -Cw polymorphism in a Berber population from North Morocco using sequence-based typing. *Tissue Antigens* 63: 158-172.
- Piazza A & Lonjou C. (1997). HLA in Europe and other Mediterranean countries. In: Charron D. (ed.). *Genetic diversity of HLA. Functional and medical implications.* (Vol. I). EDK. Paris. Pp. 374-386.

- Pimtanonthai N, Hurley CK, Leke R, Klitz W, Johnson AH. (2001). HLA-DR and –DQ polymorphism in Cameroon. *Tissue Antigens* 58: 1-8.
- Pinto F, González AM, 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.
- Pozzi S, Longo A, Ferrara GB. (1999). HLA-B locus sequence-based typing. *Tissue Antigens* 53: 275-281.
- Ramos M & Lopez de Castro JA. (2002). HLA-B27 and the pathogenesis of spondyloarthritis. *Tissue Antigens* 60: 191-205.
- Rando JC, Pinto F, González AM, Hernández M, Larruga JM, Cabrera VM, Bandelt HJ. (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, Curtoni ES. (1998). HLA polymorphisms in Italian bone marrow donors: a regional analysis. *Tissue Antigens* 52: 135-146.
- Renquin J, Sanchez-Mazas A, Halle L, Rivalland S, Jaeger G, Mbayo K, Bianchi F, Kaplan C. (2001). HLA class II polymorphism in Aka Pygmies and Bantu Congolese and a reassessment of HLA-DRB1 African diversity. *Tissue Antigens* 58: 211-222.
- Reveille JD, Ball EJ, Khan MA. (2001). HLA-B27 and genetic predisposing factors in spondyloarthropathies. *Curr Opin Rheumatol.* 13: 265-272.
- Riley E & Olerup O. (1992). HLA polymorphism and evolution. *Immunol Today* 13: 333-5.
- Rogers FM. (1979). Atlantic islanders of the Azores and Madeira. Christopher House. North Quincy.
- Rohlf FJ. (1997). NTSYS-PC, Numerical Taxonomy and Multivariate Analysis System. V. 2.0. Exeter Software, Applied Biostatistics, Inc, Setauket.
- Ronningen KS, 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.
- 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. Gen.* 68: 340-352.
- Russell-Wood AJ. (1998). The Portuguese empire, 1415-1808. A world on the move. Johns Hopkins Univ. Press. Baltimore, London.

- Saiki R, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N (1985). Enzymatic amplification of b-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230: 1350-1354.
- Sambrook J, Fritsch EF, Maniatis T. (1989). *Molecular cloning- a laboratory manual* (2nd edition, Vol. I, II, III). EUA: Cold Spring Harbor Laboratory Press.
- Sanchez-Mazas A, Djoulah S, Busson M, Le Monnier de Gouville I, Poirier JC, Dehay C, Charron D, Excoffier L, Schneider S, Langaney A, Dausset J, Hors J. (2000). A *linkage disequilibrium* map of the MHC region based on the analysis of 14 *loci* haplotypes in 50 French families. *Eur J Hum Genet* 8: 33-41.
- 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 JE, Ocejo-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.
- Sanchez-Velasco P, Gomez-Casado E, Martinez-Lazo J, Moscoso J, Zamora J, Lowy E, Silvera C, Cemborain A, Leyva-Cobián F, Arnaiz-Villena A. (2003). HLA alleles in isolated populations from North Spain: origin of the Basques and ancient Iberians. *Tissue Antigens* 61: 384-392.
- Santos P, Fonseca AM, Martinho A, Ferreira RV, Simões O, Mateus MR, Gomes J, Silva JM, 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 RF, D'Amato J, Rood JJV, Oudshoorn M. (1996). The probability of finding a haplotypically identical unrelated bone marrow donor. *In: Anonymous genetic diversity of HLA: functional and medical implications*. 2. Sèvres: EDK.
- Schipper RF, D'Amato J, de Lange P, Schreuder GM, van Rood JJ, Oudshoorn M. (1998). Validation of haplotype frequency estimation methods. *Hum Immunol.* 59: 518-523.
- Schmidt D, Goronzy JJ, Weyand CM. (1996). CD4+, CD7-, CD28- T cells are expanded in rheumatoid arthritis and are characterized by autoreactivity. *J Clin Invest.* 97: 2027-2037.
- Schneider S, Kueffer J M, Roessli D, Excoffier L. (2000). Arlequin v2.000: a software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva. Geneva, Switzerland.

- Scozzari R, Cruciani F, Pangrazio A, Santolamazza P, Vona G, Moral P, Latini V, Varesi L, Memmi MM, 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.
- Shankarkumar U, Prasanavar D, Ghosh K, Mohanty D. (2003). HLA A*02 allele frequencies and B haplotype associations in Western Indians. *Human Immunology* 64: 562-566.
- Shevach EM, Paul WE, Green I. (1972). Histocompatibility linked immune response gene function in guinea pigs. Specific inhibition of antigen induced lymphocyte proliferation by alloantisera. *J. Exp. Med.* 136: 1213-1229.
- Singal DP, Blajchman MA. (1973). Histocompatibility (HL-A) antigens, lymphocytotoxic antibodies and tissue antibodies in patients with diabetes mellitus. *Diabetes* 22: 429-432.
- Slatkin M. (1994). An exact test for neutrality based on the Ewens sampling distribution. *Genet. Res.* 64: 71-74.
- Slatkin M. (1996). A correction to the exact test based on the Ewens sampling distribution. *Genet. Res.* 68: 259-260.
- Smith AG, Matsubara K, Mickelson E, Marashi A, Regen L, Guthrie LA, Hansen JA. (1997). A comparative study of HLA-DRB typing by transcription-mediated amplification with the hybridization protection assay (TMA/HPA) versus PCR/SSOP. *Hum Immunol* 55: 74-84.
- Sollid LM, Markussen G, Ek J, Gjerde H, Vartdal F, Thorsby E (1989). Evidence for a primary association of celiac disease to a particular HLA-DQ alpha/beta heterodimer. *Journal of Experimental Medicine* 169: 345-350.
- Sospedra M & Martin R. (2005). Immunology of multiple sclerosis. *Annual Review of Immunology* 23: 683-747.
- Spencer HG & Marks RW. (1988). The maintenance of single locus polymorphism: numerical studies of a viability selection model. *Genetics* 120: 605-613.
- 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, Middleton D, Brehm A. (2005a). 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. (2005b). HLA polymorphisms in Cabo Verde and Guiné-Bissau inferred from sequence-based typing. *Human Immunology (in press)*.
- Spínola H, Brehm A, Bettencourt B, Middleton D, Bruges-Armas J. (2005c). HLA class I and II polymorphisms in Azores show different settlements in Oriental and Central islands. *Tissue Antigens* 66: 217-230.
- Stephens M, Smith NJ, Donnelly P. (2001). A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* 68: 978-989.
- Thorsby E. (1997). Invited anniversary review: HLA Associated diseases. *Human Immunology* 53: 1-11.
- Thorsby E & Lie BA. (2005). HLA associated genetic predisposition to autoimmune diseases: genes involved and possible mechanisms. *Transplant Immunology* 14: 175 – 182.
- Tiercy JM, 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.
- Tiercy JM. (2002). Molecular basis of HLA polymorphism: implications in clinical transplantation. *Transplant Immunology* 9: 173-180.
- Tosi R, Tanigaki N, Centis D, Ferrara GB, Pressman D. (1978). Immunological dissection of human Ia molecules. *J. Exp. Med.* 148: 1592-1611.
- Tsuneto LT, Probst CM, Hutz MH, Salzano FM, Rodriguez-Delfin LA, Zago MA, Hill K, Hurtado AM, Ribeiro-dos-Santos AKC, Petzl-Erler ML. (2003). HLA class II diversity in seven Amerindian populations. Clues about the origins of the Aché. *Tissue Antigens* 62: 512-526.
- Uko GP, Asuquo MA, Dawkins RL, Udim ER, Banjoko SO, Jaja O. (1997). HLA class I phenotype and gene frequencies in parts of southern Nigeria. *Disease Markers* 13: 123-130.
- Uyar FA, Dorak MT, Saruhan-Direskeneli G. (2004). Human leukocyte antigen -A, -B and -C alleles and human leukocyte antigen haplotypes in Turkey: relationship to other populations. *Tissue Antigens* 64: 180-187.
- Valdes AM, McWeeney SK, Meyer D, Nelson MP, Thomson G. (1999). Locus and population specific evolution in HLA class II genes. *Annals of Human Genetics* 63:2 7-43.
- Van de Wal Y, Kooy YM, Drijfhout JW, Amons R, Koning F. (1996). Peptide binding characteristics of the coeliac disease-associated DQ (alpha1*0501, beta1*0201) molecule. *Immunogenetics* 44: 246-53.

- Van Rood JJ & van Leeuwen A. (1963). Leukocyte grouping. A method and its application. *Journal of Clinical Investigation* 42: 1382-1390.
- Vartdal F, Johansen BH, Friede T, Thorpe CJ, Stevanovic S, Eriksen JE, Sletten K, Thorsby E, Rammensee HG, Sollid LM. (1996). The peptide binding motif of the disease associated HLA-DQ (alpha 1* 0501, beta 1* 0201) molecule. *Eur J Immunol* 26: 2764-2772.
- Vigilant L. (1999). An evaluation of techniques for the extraction and amplification of DNA from naturally shed hairs. *Biol. Chem.* 380: 1329 – 1331.
- Walsh SP, Metzger DA, Higuchi R. (1991). Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* 10: 506-513.
- Watterson G. (1978). The homozygosity test of neutrality. *Genetics* 88: 405-417.
- Weir BS. (1996). Genetic data analysis II. Sunderland: Sinauer Associates. Massachusetts.
- Weyand CM, Hicok KC, Conn D, Goronzy JJ. (1992). The influence of HLA-DRB1 genes on disease severity in rheumatoid arthritis. *Ann. Intern. Med.* 117: 801-806.
- Weyand CM & Goronzy JJ. (2000). Association of MHC and rheumatoid arthritis: HLA polymorphisms in phenotypic variants of rheumatoid arthritis. *Arthritis Res.* 2: 212–216.
- WHO- Nomenclature Committee. (1968). Nomenclature for factors of the HL-A system. *Bull. Wld. Hlth. Org.* 39: 483–486.
- Williams F, Meenagh A, Maxwell AP, Middleton D. (1999). Allele resolution of HLA-A using oligonucleotide probes in a two stage typing strategy. *Tissue Antigens* 54: 59-68.
- Williams F, Meenagh A, Single R, McNally M, Kelly P, Nelson MP, Meyer D, Lancaster A, Thomson G, Middleton D. (2004). High resolution HLA-DRB1 identification of a Caucasian population. *Human Immunology* 65: 66-77.
- Xie T, Rowen L, Aguado B, Ahearn ME, Madan A, Qin S, Campbell RD, Hood L. (2003). Analysis of the dense-major histocompatibility complex class III region and its comparison to mouse. *Genome Res.* 13:2621–2636.
- Yuhki N, Beck T, Stephens RM, Nishigaki Y, Newmann K, O'Brien SJ. (2003). Comparative genome organization of human, murine, and feline MHC class II region. *Genome Res.* 13: 1169–1179.

Yunis EJ & Amos DB. (1971). Three closely linked genetic systems relevant to transplantation. *Proc. Natl. Acad. Sci. USA* 68: 3031-3035.

Zachary AA, Klingman L, Thorne N, Smerglia AR, Teresi GA. (1995). Variations of the lymphocytotoxicity test. An evaluation of sensitivity and specificity. *Transplantation* 60: 498-503.

12. Portais na Internet

Adams SD, Krausa P, McGinnis M, Simonis TB, Stein J, Marincola FM. Practicality of high throughput HLA sequence-based typing. http://www.ashi-hla.org/publicationfiles/ASHI_Quarterly/25_2_2001/highthrusbt3.htm (data de acesso 27-09-2005).

Flanner M. Human leukocyte antigens and histocompatibility testing. <http://www.pharmacy.wisc.edu/courses/718-430/2001presentation/Flanner.pdf> (data de acesso 26-09-2005).

<http://www.allelefreqencies.net> (data de acesso 15-03-2005)

<http://www.genome.org> (data de acesso 10-10-2005)

<http://www.hla.ucla.edu/labServices/labServicesABT.htm> (data de acesso 26-09-2005).

<http://www.kovcomp.com/mvsp> (data de acesso 25-04-2003)