

HLA Allele and Haplotype Frequencies of the Portuguese Bone Marrow Donors Registry

Ricardo São João, Ana Luisa Papoila, Dário Ligeiro, and Helder Trindade

Abstract

Genes in human leucocyte antigens (HLA) System are important in the study of autoimmune diseases and responsible for the rejection of transplants of organs and tissues. HLA genes are part of the human major histocompatibility complex (MHC) which is characterized by the presence of several multigene families, extensive polymorphism at many loci and significant linkage disequilibrium between alleles at particular loci. We analysed HLA-A, -B, -DRB1 locus phenotypes through a sample of 1,021 subjects that were randomly selected among the volunteers recruited by the Portuguese Bone Marrow Donors Registry (Cedace) in order to evaluate allele, gene, haplotype and phenotype frequencies. Allelic frequencies in each of the studied locus were obtained by direct counting. Maximum-likelihood haplotype frequencies were estimated using an expectation-maximization (EM) algorithm [2]. Locus phenotype and gene relative frequencies were estimated according to Baur and Danilov [1]. Hardy-Weinberg equilibrium were tested. The data presented is a definition of HLA genetic repertoire of Cedace with relevance on the strategic management for the increase of a more diverse register with clinical utility.

R. S. João (✉)

Polytechnic Institute of Santarém-Business School, Complexo Andaluz Apt 295 2001-904 Santarém, Portugal

e-mail: ricardo.sjao@esg.ipsantarém.pt

A.L. Papoila

CEAUL and Faculdade de Ciências Médicas-Universidade Nova de Lisboa, Campo Mártires da

Pátua 130 1169-056 Lisboa, Portugal

e-mail: ana.papoila@fcm.unl.pt

D. Ligeiro · H. Trindade

Centro de Histocompatibilidade do Sul, Alameda das Linhas de Torres 117, 1769-001 Lisboa, Portugal

e-mail: dario@chsul.pt; helder.trindade@chsul.pt

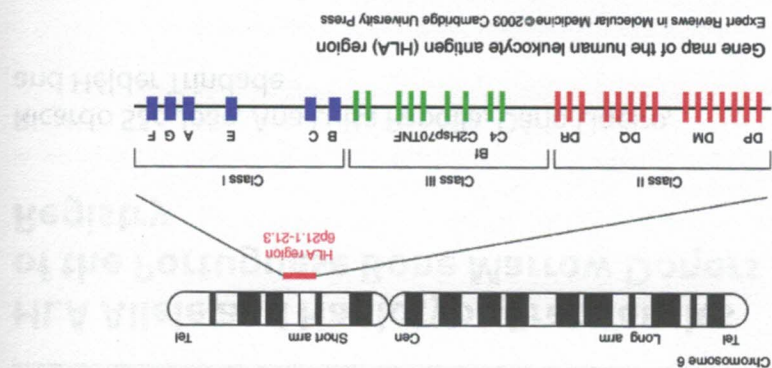


Fig. 1 Gene map of the human leukocyte antigen (HLA) region on chromosome 6p21.1 to p21.3, with class II, class III and class I genes located from the centromeric (Cen) to the telomeric (Tel) end. Figure from <http://www.expertreviews.org/> with authors', Narinder K. Mehra and Gurvinder Kaur, consent

1 Introduction

As a species, man has had his development supported in the capacity to generate human leucocyte antigens (HLA) diversity, as T cell restriction molecules. This evolution results in great antigen diversity that renders it virtually impossible to find two identical individuals, with the exception of twins. HLA antigens are also responsible for tissue compatibility and for that reason they are target for allogeneic immunological response, which means they are a biological barrier to cell, tissue and organ transplantation. In organ transplantation new immunosuppressor therapies allow transplant between donor-recipient pairs without full HLA identity. In haematopoietic stem cell transplantation, on the contrary, a high degree of HLA compatibility is necessary in order to achieve better patient survival. HLA identity is firstly sought by a low-resolution technique, looking for three main loci, HLA A, B and DRB1, and only after this first successful match, is another technical approach for allelic resolution run. In fact, due to intensive polymorphism of HLA genes, the selection of a non-related donor with the necessary degree of HLA gene compatibility to a patient is a difficult task, only possible at large databases of volunteers haematopoietic stem cell donors genotypes. A large database was created, the National Donor Registry, known as CEDACE (Centro Nacional de Dadores de Células de Medula Óssea, Estaminais ou de Sangue do Cordão) typed for more than 95% at HLA main loci A, B and DRB1.

The HLA system is located in the short arm of chromosome 6 (see Fig. 1). Within the HLA system, three constituent regions are distinguished. Near the centromere (Cen) of chromosome 6 is the class II region that contains the class II genes, while nearest the telomere (Tel) of the short arm of chromosome 6 is the class I region that contains the class I genes.

array is given by

If, for an m-allele

relationship between

frequencies remain

maturing population

and is known as the

in 1908, detected

and genotype frequencies

In population genetics

Hardy-Weinberg

3.1

Methodology

3

about the probability

haplotype frequencies

likelihood of finding

registries of prospective

donors; narrow donor

alternative donors are

the lack of donors. Because

genetics. The key limiting

important resource for

evaluate allele, gene

were randomly selected

HLA-A, -B, -DRB1 loci

while others are widely

may be rare in another

alleles differs among

than 1%). As is the case

lower than 10%) and

frequency greater than

alleles at particular

five polymorphisms at

genes are part, is

The human major

2

Data

HLA Allele and Haplotype

2 Data

The human major histocompatibility complex, of which the HLA class I and class II genes are part, is characterized by the presence of several multigene families, extensive polymorphism at many loci and significant linkage disequilibrium between alleles at particular loci. In most populations, a few alleles are frequent (gene frequency greater than 10%) but most occur at low frequency (gene frequency lower than 10%) and a number of the latter may be rare (gene frequency lower than 1%). As is the case for other genetic polymorphisms, the frequency of HLA alleles differs among populations. An allele that is common in one population may be rare in another. Some alleles are limited to particular ethnic populations, while others are widely shared among ethnically distinct populations. We analysed HLA-A, -B, -DRB1 locus phenotypes through a sample of 1,021 subjects that were randomly selected among the volunteers recruited by Cedace in order to evaluate allele, gene, haplotype and phenotype frequencies. These data represent an important resource for investigators in the fields of transplantation and population genetics. The key limiting factor in the use of bone marrow transplantation (BMT) is the lack of donors. Because only 25–30% of patients have an HLA-identical sibling, alternative donors are often required. Marrow can be procured from unrelated living donors; marrow donation is a simple, safe procedure. National and international registries of prospective volunteer donors are being expanded to increase the likelihood of finding an exact HLA match for any given recipient. The gene and haplotype frequencies of a registry can be used in advice clinicians and patients about the probability of finding an HLA match for BMT.

3 Methodology

3.1 Hardy-Weinberg Equilibrium

In population genetics, it is very important to study the relationship between allele and genotype frequencies. Godfrey Harold Hardy [4] and Wilhelm Weinberg [7], in 1908, detected, independently, a principle that describes the referred relationship and is known as the Hardy-Weinberg law (HWL). It says that, in a large random-mating population with no selection, mutation or migration, the genotype and allele frequencies remain stable from generation to generation and that there is a fixed relationship between allele and genotype frequencies.

If, for an *m*-allele autosomal locus with alleles A_1, A_2, \dots, A_m , the genotypic array is given by

$$\sum_i p_i^2 A_i A_i + \sum_{i < j} 2 p_i p_j A_i A_j,$$

where p_i is the allelic frequency of A_i , it is said that the population with these genotype frequencies, known as Hardy-Weinberg proportions (HWP), is in Hardy-Weinberg equilibrium at that locus.

Table 1 Counts, phenotype, genotype and standard deviations (SD) frequencies of HLA-A,-B,-DRB1 loci allele groups

Allele	# of	Phenotype	Genotype	Std
		frequency	frequency	deviations
		(4.802)	(4.802)	
A*01	1.021	0.2126	0.1127	0.00343
A*02	2.730	0.5686	0.3432	0.00598
A*03	899	0.1872	0.0985	0.00320
A*11	627	0.1306	0.0676	0.00265
A*23	412	0.0858	0.0439	0.00214
A*24	1.006	0.2095	0.1109	0.00340
A*25	137	0.0285	0.0144	0.00122
A*26	346	0.0721	0.0367	0.00196
A*29	491	0.1023	0.0525	0.00234
A*30	334	0.0696	0.0354	0.00192
A*31	233	0.0485	0.0246	0.00160
A*32	359	0.0748	0.0381	0.00199
A*33	344	0.0716	0.0365	0.00195
A*34	43	0.0090	0.0045	0.00068
A*36	12	0.0025	0.0013	0.00036
A*43	1	0.0002	0.0001	0.00010
A*66	67	0.0140	0.0070	0.00085
A*68	479	0.0998	0.0512	0.00231
A*69	27	0.0056	0.0028	0.00054
A*74	24	0.0050	0.0025	0.00051
A*80	11	0.0023	0.0011	0.00035
B*07	606	0.1239	0.0640	0.00258
B*08	611	0.1250	0.0646	0.00259
B*13	139	0.0284	0.0143	0.00122
B*14	744	0.1522	0.0792	0.00287
B*15	500	0.1023	0.0525	0.00234
B*18	563	0.1151	0.0593	0.00249
B*27	287	0.0587	0.0298	0.00176
B*35	1.250	0.2556	0.1372	0.00378
B*37	115	0.0235	0.0118	0.00111
B*38	271	0.0554	0.0281	0.00171
B*39	143	0.0292	0.0147	0.00124
B*40	332	0.0679	0.0345	0.00190
B*41	112	0.0229	0.0115	0.00110
B*42	18	0.0037	0.0018	0.00044
B*44	1.439	0.2943	0.1599	0.00408
B*45	132	0.0270	0.0136	0.00119
B*46	1	0.0002	0.0001	0.00010
B*47	29	0.0059	0.0030	0.00056
B*48	5	0.0010	0.0005	0.00023

(continued)

even in closed and distant ethnic groups is significant to describe the degree of genetic heterogeneity of the Portuguese population. **DRB1 locus** has only 13 allele groups, all of them represented in the probed population. The most common are DRB1*13 (18,3%), DRB1*07 (15,9%), DRB1*04 (15,4%), DRB1*01 (12,7%) and DRB1*11 (12,5%). Less frequent are DRB1*09 (0,68%), DRB1*10 (1,4%), DRB1*12 (1,6%) and DRB1*16 (2,8%).

4.2 Multi-locus Analysis

HLA haplotypes are specific sets of HLA-A,-B,-DR locus alleles inherited together from a parent. Haplotypes are usually determined by genotyping a sufficient number of family members to establish a gametic assignment of the alleles detected. Because of the extraordinarily large number of possible HLA haplotypes, it is impractical to determine anything but the most common haplotype frequencies by doing family studies. Nevertheless, it is possible to estimate population haplotype frequencies by genotyping a sufficient number of unrelated individuals and using a computer algorithm, to estimate the allele associations that are consistent with the observed genotype data. For large populations that are in Hardy-Weinberg equilibrium, it is possible to estimate even relatively rare haplotypes (e.g. frequency <0.01%) with reasonable accuracy. In this study we used the Lancaster and Nelson [5] population genetics analysis package, PyPop (<http://alleles5.biol.berkeley.edu/pypop/>). This program implements an iterative expectation-maximization (EM) [2] algorithm on the genotyping data of a maximum of 1,021 randomly selected samples leading to the maximum-likelihood estimate of haplotype frequency for loci: A:B:DRB1. From the sample of 1,021 individuals it was reported 996 unique phenotypes, 3,381 genotypes and 2,082 haplotypes with an estimated frequency above 0.00001 and a log likelihood obtained via the EM algorithm $\ln(L_1) = -11296.3$. The exact test of Guo and Thompson [3] was performed for deviations of HWP. The *p*-value provided describes how probable the observed set of genotypes is, with respect to a large sample of other genotypic configurations (conditioned on the same allele frequencies and 2n). *p*-values lower than 0.05 can be interpreted as evidence that the sample does not fit HWP. Table 2 presents the HLA-A:B:DRB1 haplotypes with an estimated frequency greater than or equal to 0.5%. The well-known Caucasoid haplotype A*01-B*08-DRB1*03, due to hard disequilibrium linkage, comes out as the most frequent in the probed population. The five most frequent HLA HLA-A:B:DRB1 haplotypes are 01:08:03 (3,1%), 02:44:07 (2,3%), 02:44:04 (2,1%), 02:51:11 (1,9%) and 29:44:07 (1,6%) which are all typical haplotypes of European Caucasian populations. In fact from this analysis we can detect only 11 haplotypes with frequencies greater than or equal to 1%.

The Guo and Thompson exact test for HWP (see Table 3) reveals that the population submitted to haplotype estimation does fit the Hardy-Weinberg equilibrium. To be in HW equilibrium means that the sampled of individuals have random mating and does not suffer of evolutive pressures, which turns possible to apply the frequency data to a larger population.

Table 2 Sample output of HLA-A-B-DRB1 haplotype frequency estimation

Haplotype	# Copies	Frequency	SD
01:08:03	63.2	0.03097	0.0038
02:44:07	47.6	0.02332	0.0033
02:44:04	42.5	0.02079	0.0031
02:51:11	39.2	0.01918	0.0030
29:44:07	32.0	0.01569	0.0027
33:14:01	30.8	0.01510	0.0027
03:07:15	27.2	0.01331	0.0025
03:35:01	23.5	0.01150	0.0023
02:44:13	22.9	0.01124	0.0023
02:18:03	20.6	0.01007	0.0022
23:44:07	20.1	0.00982	0.0022
02:51:13	19.3	0.00946	0.0021
02:50:07	17.8	0.00871	0.0020
11:35:01	16.3	0.00800	0.0020
02:51:08	16.3	0.00797	0.0019
30:18:03	15.9	0.00778	0.0019
68:51:13	14.9	0.00730	0.0019
24:35:11	14.4	0.00706	0.0018
02:14:01	14.0	0.00687	0.0018
26:38:13	13.8	0.00678	0.0018
02:18:11	13.8	0.00673	0.0018
24:35:07	13.3	0.00653	0.0018
02:15:04	12.7	0.00622	0.0017
01:57:07	12.3	0.00602	0.0017
24:35:13	11.2	0.00548	0.0016
02:07:01	10.9	0.00534	0.0016
03:14:01	10.4	0.00511	0.0016
02:51:04	9.9	0.00485	0.0015
33:44:13	9.9	0.00483	0.0015

Table 3 Guo and Thompson exact test for Hardy-Weinberg proportions

	<i>p</i> -value	SD
HLA-A	0.6110	0.01086
HLA-B	0.6383	0.01185
HLA-DRB1	0.6557	0.008386

4.3 Most Common Phenotypes at PBMRD

In a sample of 20,000 individuals of Cedace it was detected 17,055 different HLA-A,-B-DRB1 phenotypes. The 2,945 that remains are repeated phenotypes at

Table 4 Output of HLA-A

HLA-A
A*01,29
A*01,02
A*01,03
A*01,11
A*02,33
A*01,33
A*02,03
A*01,02
A*01,23
A*01,68
A*02,29
A*03,29
A*01,02
A*01
A*01,02
A*02,03
A*02,29

5

Conclusion

different proportions, counts and relative influence of HLA-A*01-B*02:03:01 combinations. In fact the influence of the A*01-B*02:03:01 combinations is valuable to determine the haematological

to determine the HLA-A*01-B*02:03:01

HLA-A*01-B*02:03:01

Table 4 Output of HLA-A-B-DRB1 phenotype frequency in a random sample of 20,000 individuals

HLA-A	HLA-B	HLA-DRB1	Count	Frequency × 10 ⁴
A*01,29	B*08,44	DRB1*03,07	28	14,13
A*01,02	B*08,44	DRB1*03,04	17	8,58
A*02,29	B*44,51	DRB1*07,11	16	8,07
A*01,02	B*08,44	DRB1*03,07	15	7,57
A*01,03	B*08,51	DRB1*03,08	13	6,56
A*01,11	B*08,35	DRB1*01,03	12	6,06
A*02,33	B*14,44	DRB1*01,04	11	5,55
A*01,33	B*08,14	DRB1*01,03	10	5,05
A*02,03	B*35,44	DRB1*01,04	10	5,05
A*01,02	B*08,18	DRB1*03,11	10	5,05
A*01,23	B*08,44	DRB1*03,07	10	5,05
A*01,68	B*08,53	DRB1*03,13	10	5,05
A*02,29	B*44	DRB1*07,13	10	5,05
A*03,29	B*07,44	DRB1*07,15	10	5,05
A*01,02	B*08,50	DRB1*03,07	9	4,54
A*01	B*08	DRB1*03	9	4,54
A*01,02	B*44,57	DRB1*04,07	9	4,54
A*02,03	B*07,44	DRB1*04,15	9	4,54
A*02,29	B*44	DRB1*04,07	9	4,54

different proportions. Table 4 identifies the most common phenotypes, the absolute counts and relative frequency in the probed population. As expected, it is noted the influence of HLA-A,-B-DRB1 haplotypes frequencies on phenotype frequencies, the A*01-B*08-DRB1*03 appears on 63% of the 19 most frequent phenotypes. In fact the 100 most frequent ABDRB1 phenotypes at the south PBMDR, are direct combinations of the haplotypes with a frequency greater than 0,5% (28 haplotypes).

5 Conclusions

From the obtained results, the population under study revealed an anthropologic proximity with the European Caucasian* populations [1]. The characterization of HLA gene and haplotype frequencies of a Bone Marrow Donors Registry, is a valuable resource not only in the prediction of the probability of finding a matched haematopoietic stem cell donor, considering the receptor HLA phenotype, but also to determine donor recruitment goals and strategies. Furthermore, the extend of the populations at the registry represent an important source of information for investigators interested in population genetics and HLA-disease association studies.

References

1. Baur, M.P., Danilov, J.A.: Population Analysis of HLA-A,B,C,DR, and other genetic markers. In: Terasaki, P.I. (ed.) *Histocompatibility Testing*, pp. 955-957. PL UCLATissue Typing Lab, Los Angeles (1980)
2. Excoffier, L., Stakin, M.: Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol. Biol. Evol.* **12**, 921-927 (1995)
3. Guo, S., Thompson, E.: Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* **48**, 361-72 (1992)
4. Hardy, G.H.: Mendelian proportions in mixed populations. *Science* **28**, 49-50 (1908)
5. Lancaster, A., Nelson, M., Single, R., Meyer, D., Thomson, G.: PyPop a software framework for population genomics: analyzing large-scale multi-locus genotype data. In: Altman, R.B. et al. (eds.) *Pacific Symposium on Biocomputing*, vol. 8, pp. 514-525 (2003)
6. Levene, H.: On a matching problem arising in genetics. *Annals of Math. Stat.* **20**, 91-94 (1949)
7. Weinberg, W.: über den Nachweis der Vererbung beim Menschen. *Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg*, **64**, 368-383 (1908)

Abstract

Various techniques for time series, including initial lagged series, analysis in a dimension analysis (ICA) or extended time series, since the higher-order subspace obtain the principle necessarily independent some comparative methods of ordering influence the quality

F. Sebastião (✉)

Department of Mathematics

Leiria, and CM-UTAD

Leiria, Portugal

e-mail: fsebast@ipkcmf

I. Oliveira

Department of Mathematics

Apartado 1013, 5001-801

e-mail: ioliveir@utadp

J. Lita da Silva et al.

Markov Processes and

and Applied Statistics, DOI

© Springer-Verlag Berlin