Human Immunology 73 (2012) 399-404

Contents lists available at SciVerse ScienceDirect

journal homepage: www.elsevier.com/locate/humimm

Estimation of human leukocyte antigen class I and class II high-resolution allele and haplotype frequencies in the Italian population and comparison with other European populations

Sabina Rendine [a,b,*], Norma Maria Ferrero [a], Nicoletta Sacchi [c], Cristina Costa [c], Simona Pollichieni [c], Antonio Amoroso [a,b]

[a] A.O.U.S.G. Battista, Transplant Immunology Service, Torino, Italy
[b] University of Turin, Department of Genetics, Biology and Biochemistry, Torino, Italy
[c] Italian Bone Marrow Donor Registry (IBMDR) E.O. Ospedali, Galliera, Genoa, Italy

ARTICLE INFO

Article history:
Received 2 August 2011
Accepted 24 January 2012
Available online 1 February 2012

Keywords:
Alleles
Haplotypes
High-resolution typing
HLA
Linkage disequilibrium

ABSTRACT

The high-resolution (HR) allele and haplotype frequencies of class I and II human leukocyte antigen (HLA) system were determined in the Italian population from a sample of donors recruited in the Italian Bone Marrow Donor Registry (IBMDR). This study analyzed the HLA-A, -B, -C, -DRB1, and -DQB1 loci. Two different samples were used: donors HR typed at least for one allele, usually when selected for donor–recipient matching (respectively: 3596, 7591, 4715, 57345, and 8196), to make a list of the observed alleles and determine the relative frequencies of the alleles in each class of the corresponding antigen; donors HR randomly typed for both the alleles (respectively: 975, 1643, 1569, 22114, and 2087) to estimate the allele and haplotype frequencies, and two loci linkage disequilibrium. The number of alleles showing a frequency >1% on the total number of observed alleles are 18/75 HLA-A, 28/142 -B, 17/57 -C, 23/154 -DRB1, and 13/31 -DQB1. In each locus they account for more than 88% of the total cumulative frequencies. The most frequent alleles are A*02: 01, B*35: 01, C*04:01, DRB1*07:01, DQB1*03:01. The most frequent five-locus haplotype in the 338 donors randomly typed is A*01: 01-C*07:01-B*08: 01-DRB1*03:01-DQB1*02:01. The genetic comparison of the Italian population with 16 European populations shows a south–north gradient.

1. Introduction

The human leukocyte antigen (HLA) genetic system plays a crucial role in unrelated stem-cell transplantation; HLA matching between donors and recipients affects the success of hematopoietic stem-cell transplantation by minimizing graft rejections and graft-versus-host disease [1]. Reports on strong adverse effects of HLA mismatch at high-resolution levels on transplantation outcome re-emphasize that the selection of optimal unrelated donor should be based on allele matching [2–4]. Furthermore, matching at the haplotype level has a better likelihood of matching at other loci within the HLA region than for donors merely matched at the individual allelic level [5].

Moreover, the extreme allelic polymorphism located at several loci of this system tends to make the distribution of HLA alleles unique in each ethnic group. Hence, alleles and haplotypes determination at high-resolution level in different populations is very important in transplantation medicine for the best HLA-matching donor searching.

Several studies have been performed to characterize the Italian population [6–10]. These studies analyzed serologic or molecular data at low-resolution levels of HLA-A, -B, and -DR loci, and demonstrated that the genetic structure of the Italian population can be represented by volunteer hematopoietic stem cell donors enrolled in the Italian Bone Marrow Registry (IBMDR), which is the primary source of unrelated marrow donor searches in Italy. Notwithstanding this, a high-resolution level alleles and haplotypes distribution report for HLA class I and class II in the Italian population is still not available. In this study, samples of randomly HR typed donors were extracted from IBMDR Registry, and frequencies of HLA-A, -B, -C, -DRB1, and -DQB1 alleles and haplotypes were estimated with the aim of predicting the most common alleles and two, three, four, and five loci haplotypes in the Italian population.

2. Subjects and methods

2.1. Subjects

IBMDR is the third European Registry (sixth in the world) making available HLA data from volunteer hematopoietic stem cells donors on international networks. Its activity started in 1989, and

* Corresponding author.
E-mail address: sabina.rendine@unito.it (S. Rendine).

0198-8859/12/$36.00 - see front matter © 2012 American Society for Histocompatibility and Immunogenetics. Published by Elsevier Inc. All rights reserved.
doi:10.1016/j.humimm.2012.01.005

Table 1
Most common estimated allele frequencies (EAF; >1%) for the loci HLA-A, -B, and -C

HLA-A	*N* = 975	HLA-B	*N* = 1643	HLA-C	*N* = 1569
Allele	**EAF**	**Allele**	**EAF**	**Allele**	**EAF**
02:01	24.05	35:01	8.74	04:01	16.48
24:02	12.15	51:01	8.03	07:01	14.02
01:01	10.15	18:01	5.80	12:03	11.22
03:01	10.15	14:02	5.68	06:02	8.29
11:01	4.56	49:01	4.28	07:02	6.25
26:01	4.05	08:01	4.07	05:01	5.70
32:01	3.69	07:02	4.01	15:02	5.29
30:01	3.38	35:03	3.83	02:02	4.17
68:01	3.33	15:01	3.48	14:02	4.05
31:01	2.62	35:02	3.30	08:02	3.79
02:05	2.26	44:02	3.30	16:01	3.73
23:01	2.15	78:01	3.06	03:03	3.47
29:02	2.05	15:17	3.03	01:02	3.28
33:01	2.00	38:01	2.88	16:02	2.01
30:02	1.95	44:03	2.73	12:02	1.66
25:01	1.69	57:01	2.44	03:04	1.50
03:02	1.49	35:08	2.23	07:04	1.21
24:03	1.18	50:01	2.23		
		13:02	2.11		
		40:02	1.84		
		40:01	1.72		
		39:01	1.63		
		27:05	1.55		
		58:01	1.55		
		55:01	1.31		
		14:01	1.22		
		52:01	1.22		
		37:01	1.19		
Total	92.90	Total	88.46	Total	96.12

at the end of 2011 the donor's file counted approximately 390.000 donors, all typed with serologic or molecular techniques, usually at low resolution level (LR).

The samples used in this study included individuals recruited as volunteer donors for the IBMDR Registry and typed at high resolution level (high resolution [HR], four digits) at least for one of the two alleles.

In all: 3596 donors were analyzed for HLA-A, 7591 for HLA-B, 4715 for HLA-C, 57345 for HLA-DRB1, and 8196 for HLA-DQB1.

2.2. HLA typing

At the beginning, all volunteer donors were serologically typed for HLA-A and -B when they joined the IBMDR Registry, for HLA-DR usually when selected for a patient (patient-directed typing), and, only if matched, also for one or more of the loci C, DP, and DQ, depending on the transplantation center protocol. Afterward, -DR typing also was necessary for Registry admission.

Over time, the HLA DNA-based testing was implemented first at low resolution, then was increased to intermediate and high resolution. Subsequently, molecular testing of a donor was principally obtained after selection and retyping from a transplant center who believed he might be a potential match for their patient or, alternatively, if selected for allele frequency studies (random typing). Therefore, some volunteer donor HLA assignments were obtained by serology at the broad or split-level, and others by DNA-based testing at low, intermediate, or allele resolution (SBT-, SSO-, or SSP-based).

2.3. Allele and haplotype frequencies

The list of the HR alleles identified in the Italian population, and the relative frequency of each of them in the class of the corresponding LR allele (equivalences in "Nomenclature for Factors of the HLA system 2010" [11]) were obtained taking in account all the alleles determined at least at four digits, but only the first four digits were considered. In the case of typings with one allele determined at HR and one at LR (e.g., for typing ambiguities at the allelic group level) only the HR allele was considered. The data are shown in Supplementary Tables S1–S5.

Allele and haplotype frequencies (two, three, four, and five loci) were estimated by the expectation maximization (EM) algorithm [12] in the subsets of donors randomly typed at least at four digits for both the alleles of the considered locus: 975 donors for HLA-A, 1643 for -B, 1569 for -C, 22,114 for -DRB1, 2087 for -DQB1, 625 for A-C, 711 for C-B, 956 for B-DRB1, 1447 for DRB1-DQB1, 553 for A-C-B, 496 for A-C-B-DRB1, and 338 for A-C-B-DRB1-DQB1.

All of the previous analyses were performed using the statistical software SAS, version 9.2 of the SAS system for Windows [13].

The goodness of fit for Hardy–Weinberg equilibrium was tested by the G2 likelihood ratio statistic [14], defined as $G2 = 2\,\Sigma_i\,[o_i\,ln(oi/e_i)]$, where o_i and e_i are the observed and the expected values, respectively. This statistic is asymptotically distributed as a χ^2, and is less sensitive to low expected values than the Pearson χ^2 statistic.

2.4. Genetic comparison

The Italian population was compared with other 16 European, or with European origin, populations [15–18] typed at HR level for the most common alleles (58 in all) of the loci HLA-A, -B, -DRB1. Missing data (observed in seven populations and in a maximum of seven variables) were substituted with the allele mean frequencies. Other neighboring populations were excluded for the high number of missing data.

A principal component analysis was performed to condense the total variance into a lower number of variables, and a dendrogram, based on Nei distance, was reconstructed using the maximum likelihood method. The robustness of the dendrogram was assessed by the bootstrap method: the frequency of each splitting into 100 trees, obtained by resampling the original data, was calculated and indicated near each node in the displayed tree. A cluster was considered robust if its frequency exceeded 50%. This last analysis was performed by using the package PHYLIP [19].

Table 2
Most common estimated allele frequencies (EAF; >1%) for the locus HLA-DQB1 and -DRB1

HLA-DQB1	*N* = 2087	HLA-DRB1	*N* = 22,114
Allele	**EAF**	**Allele**	**EAF**
03:01	30.33	07:01	12.48
05:01	11.79	11:01	12.01
02:01	7.79	11:04	9.99
02:02	7.71	03:01	9.28
05:02	7.00	01:01	6.44
05:03	6.88	15:01	5.58
03:02	5.92	13:01	5.35
06:03	5.70	16:01	4.88
03:03	3.86	13:02	4.83
06:02	3.45	14:54	2.93
06:04	3.31	01:02	2.49
04:02	3.14	14:01	2.28
06:01	1.32	08:01	2.13
		10:01	1.93
Total	98.20	04:01	1.66
		04:03	1.59
		11:03	1.53
		04:02	1.43
		15:02	1.36
		13:03	1.29
		04:05	1.27
		12:01	1.23
		04:04	1.19
		Total	95.15