

ORIGINAL ARTICLE

High-resolution analysis of the HLA-A, -B, -C and -DRB1 alleles and national and regional haplotype frequencies based on 120 926 volunteers from the Italian Bone Marrow Donor Registry

Nicoletta Sacchi¹  | Mauro Castagnetta¹  | Valeria Miotti² | Lucia Garbarino¹ | Annamaria Gallina¹

¹E.O. Galliera, Italian Bone Marrow Donor Registry, Genova, Italy

²Azienda Sanitaria Universitaria Integrata di Udine, Laboratory of Immunogenetica, Udine, Italy

Correspondence

Nicoletta Sacchi, E.O. Galliera, Italian Bone Marrow Donor Registry Genova, Italy.
Email: nicoletta.sacchi@galliera.it

HLA genes are highly polymorphic and structurally complex. They are located in the major histocompatibility complex (MHC) on chromosome 6, and the frequency of alleles and haplotypes varies widely among human populations.

In this paper, we calculated the allele and haplotype frequencies using the HLA data of more than 120 000 Italian unrelated bone marrow donors enrolled in the national registry (IBMDR) and typed them with a high-resolution (HR) method for the HLA-A, -B, -C and -DRB1 alleles. The allele frequency data were obtained by manual counting; haplotype frequencies were calculated using the expectation maximisation (EM) algorithm. The total numbers of observed alleles were 226 for HLA-A, 343 for HLA-B, 201 for HLA-C and 210 for HLA-DRB1, which account for 5.4%, 6.7%, 5.2% and 8.5%, respectively, of each locus allele (IPD-IMGT/HLA Database Release 3.32, April 2018). The three most frequent Italian haplotypes were *HLA-A*01:01~B*08:01~C*07:01~DRB1*03:01* (2.5%), *A*02:01~B*18:01~C*07:01~DRB1*11:04* (1.1%) and *A*30:01~B*13:02~C*06:02~DRB1*07:01* (1.1%). Moreover, for a relevant subset of the examined population (>100 000 individuals), the birthplace was available, and thus, we grouped the frequency data based on the corresponding Italian geographic areas, describing the HLA specificity of the Italian regional populations. The haplotype frequencies were also compared between national and regional data, and we observed remarkable differences in the regional haplotype frequencies, particularly in Sardinia. This study represents a valid tool to identify a more efficient haematopoietic stem cell unrelated donor recruitment and selection strategy, as well as for population genetic and HLA-disease association fields.

KEYWORDS

frequency, haplotype, regions, unrelated donors

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2019 The Authors. HLA: Immune Response Genetics published by John Wiley & Sons Ltd.

1 | INTRODUCTION

Allogenic haematopoietic stem cell transplantation (HSCT) is a widely used and effective therapy for haematopoietic malignant diseases and numerous other disorders. HR donor-recipient matching for the human leukocyte antigen (HLA) genes improves patient survival as the number of mismatches correlates with the risk of rejection and/or graft vs host disease (GVHD).¹⁻⁵

When a related HLA-identical donor is not available, an alternative donor of haematopoietic stem cells must be found. Although allo-HSCT from other sources, such as cord blood and related HLA-mismatched or haploidentical donors, can be an option,⁶⁻⁸ an unrelated donor who matches with the corresponding patient at least for HLA-A, -B, -C and -DRB1 at HR (HLA 8/8 matched) is still one of the best options.⁹ Consequently, there is the need to develop and maintain a useful and efficient unrelated donor registry. This can be done by increasing donor numbers,¹⁰ by recruiting young males,¹¹ or by increasing the possibility of identifying a fully matched donor using available HLA allele and haplotype frequencies¹²⁻¹⁴ to establish an efficient recruitment and selection strategy. In fact, the knowledge of population-specific HR HLA haplotype frequency distributions facilitates individual donor searches and provides the theoretical background for estimating the chance for a patient to find fully matched donors in the registry.

Unfortunately, due to the high level of HLA polymorphism in the Italian population,^{15,16} 50% of Italian patients who need HSCT are currently unable to find an HLA 8/8 matched donor.¹⁷ Consequently, the characterisation of HR HLA haplotype frequency distribution in the Italian population is helpful for applying a recruitment strategy that increases the diversity of the donor pool.

The Italian national Bone Marrow Donor Registry (IBMMDR) was started in 1989 with the objective of providing HLA-matched unrelated donors for Italian patients and worldwide patients at large.

IBMMDR manages the unrelated haematopoietic stem cell donor recruitment, maintenance and search country-wide. This registry is composed of a network of 17 active regional donor registries and 75 donor centres representing the Italian geographic regions. At the end of December 2017, IBMMDR listed 392 873 available donors, and approximately 25 000 new donors are registered every year. Since 2015, IBMMDR has performed HLA -A, -B, -C and -DRB1 typing at a HR level (two fields) in all newly recruited volunteers.

Additional HLA HR-typing data are obtained from patient-directed typing or due to prospective typing of partially typed young donors.

Due to these strategies, the IBMMDR has accumulated a large data set of HR HLA-typed individuals (more than 131 000 at the end of 2017).

Over the last 10 years, several papers have been published on HLA allele and haplotype distributions in the Italian population.^{16,18,19} These studies analysed unrelated donor HLA data and demonstrated that the genetic structure of the Italian population could be represented by volunteer haematopoietic stem cell donors enrolled in the IBMMDR. However, in those studies, the haplotype frequencies were estimated at the HLA low resolution level or were calculated from limited samples of two field-typed donors.

In this study, a sample of 120 926 randomly HR-typed donors was extracted from the IBMMDR database, and the frequencies of the HLA-A, -B, -C and -DRB1 alleles and haplotypes were estimated. The distribution of HLA allele and haplotype frequencies were reported nationally for Italy and then grouped into regional data to assess regional differences in the frequency data.

We did not estimate the frequencies of the DQB and DPB alleles even if they can be a matching criterion in the selection of an unrelated haematopoietic stem cell donor because these two loci are not routinely typed at recruitment. In Italy, the DQB and DPB typing is only performed if there is a request for a specific patient, and this fact could create bias in the frequency analysis of these two loci.

Furthermore, we investigated the prevalence of observed alleles by applying the criteria described by Mack et al.²⁰ to identify common (C) and well-documented (WD) alleles in our database and then compared our findings with those of the CWD European catalogue from EFI.²¹

2 | MATERIALS AND METHODS

2.1 | Sample population

Our analysis consisted of 120 926 individuals (dataset 1) who were included in the Italian Registry IBMMDR at the end of December 2017 and typed for HLA-A, -B, -C and -DRB1 at the HR level. To avoid bias, we included only the donors fully typed at the recruitment step in this study.

All donors included in the analysis fulfilled the eligibility criteria and were registered in the IBMMDR donor centres located in all 20 Italian regions. At the time of the analysis, all subjects were from 18 to 55 years old, and 77.8% were younger than 35 years. The sample was composed of 45.6% males and 54.4% females (Table 1). Upon recruitment, volunteers were asked to sign an informed consent form and to provide personal information. For a sample of 104 135 donors (data set 2), the birthplace and ethnic origin were

TABLE 1 Characteristics of the data sets

	Sample size	Gender Males (%)	Age				Median (25th-75th)
			18-25	26-35	36-45	46-55	
Data set 1	120 926	45.6	44 466 (36.8%)	49 582 (41.0%)	23 760 (19.6%)	3118 (2.6%)	28 (23-35)
Data set 2	104 135	45.9	37 459 (36.0%)	43 254 (41.5%)	20 761 (19.9%)	2661 (2.6%)	28 (24-35)
Data set 3	55 538	46.8	18 496 (33.3%)	22 441 (40.4%)	12 167 (21.9%)	2434 (4.4%)	29 (24-36)

available. In our study, the HLA data of these individuals were classified and grouped according to the Italian geographical region where they were born.

A sample of 55 538 individuals (data set 3) with no group “P” or group “G” alleles at any HLA-A, -B, -C and -DRB1 locus was used to calculate the number of observed alleles. This sample was also used to establish the catalogue of Italian national CWD alleles by applying the criteria described by Mack et al.²⁰ To put our findings into the context of relevant existing CWD catalogues, if necessary, allele designations were collapsed to two fields of resolution.

The reference datasets we used are the CWD 2.0.0 catalogue (ASHI CWD),²⁰ based on observations from several worldwide populations, and the EFI CWD catalogue,²¹ based on individuals of European origin.

2.2 | HLA typing and processing of HLA data

All donors included in this study were typed for HLA-A, B, C and -DRB1 loci at a HR level by ASHI or EFI-accredited tissue typing laboratories using HR molecular biology techniques (SBT, SSO, SSP and more recently NGS). The alleles are generally assigned in two field forms, and ambiguities within the relevant exons were resolved with different approaches.

Nevertheless, in the analysed population samples, some alleles were not distinguished by typing systems, and genomic regions were not defined outside the antigen recognition domain (ARD). Therefore, some HLA alleles with nucleotide sequences that encode the same protein sequence for the ARD (exons 2 and 3 for HLA class I and exon 2 for HLA class II alleles) were designated group “P” alleles; on the other hand, some alleles belonging to the same P group were unambiguously assigned (eg, *DRB1*14:01P* and *DRB1*14:54*). The same protocol was performed for some G group alleles (eg, *C*04:01:01G* and *C*04:01:01:01*).

To avoid bias, the same approach of Schmidt et al.¹² was applied: alleles with synonymous mutations inside or outside the relevant exons were merged to the corresponding two field alleles, and alleles that differed by nonsynonymous mutations outside the relevant exons were merged. The resulting alleles were characterised through the letter “g”

(for “group”) that was appended to the first possible allele.^{22,23}

2.3 | Statistical analysis

The HLA-A, -B, -C and -DRB1 allele frequencies were obtained by the direct counting method. The Arlequin software package (version 3.5.2.2, Excoffier & Lischer, 2010²⁴) was used to estimate maximum likelihood haplotype frequencies based on the expectation-maximisation (EM) algorithm and to assess the Hardy-Weinberg equilibrium (HWE) for each of the four loci. A total of 50 starting points and ≤ 1000 interactions were chosen as input parameters for the EM algorithm, and the threshold for stopping the algorithm was $\varepsilon = 10^{-7}$.

We did not assess HWE in data set 1 because the national sample has a very large size, and this contributed to the determination of a very small *P*-value and a small HWE deviation. A previous report showed²⁵ that deviations from HWE did not affect the accuracy of HF estimations with the EM algorithm in these cases; therefore, we decided to test HWE only at the regional level.

The input parameters for the Markov chain Monte Carlo test by Guo and Thompson²⁶ were 10^6 Markov chain steps and 10^5 dememorization steps. Values of *P* < 0.01 were regarded as significant.

2.4 | Evaluation of genetic variation in the Italian regions

Italy is subdivided into 20 geographical regions from North to South: Lombardia, Piemonte, Valle d'Aosta, Liguria, Veneto, Friuli Venezia Giulia, Province TN e BZ and Emilia Romagna (Northern Italy), Toscana, Marche, Umbria, Lazio, Abruzzo, Molise (Central Italy), Campania, Puglia, Basilicata, Calabria, Sicilia (Southern Italy), Sardinia.

In this study, we estimated the HLA Italian regional frequencies based on the donors' birthplace.

In particular, we extracted the birthplace from the Italian health insurance card (which contains the birthplace code) of the donors who self-reported Italian origin in the consent form at recruitment.

This sample of 104 135 donors (Dataset 2) was grouped into the 20 Italian regional populations with the goal of reflecting the genetic diversity of the populations in this country.

To estimate if the regional sample sizes (RSS) obtained were representative of the specific regional population, we applied the same approach as in other studies,¹⁸ which define the RSS as acceptable when the ratio between *sampled individuals/resident population* is $>5/10000$.

Table S1 reports the numbers of individuals of data set 2 grouped per region and this ratio index calculated on the resident population having the same age as our donors (18-55 years).

Like any other approach used to classify populations in geographic subgroups, our method is affected by a possible selection bias since it is not able to take into account possible gene flow due to population migrations, which, in Italy, usually occur from the South to the North.

Nevertheless, by comparing the obtained Italian HLA regional frequencies with those from previous studies based on living district information,^{18,19} we found comparable results.

For this reason, we suppose that these subsamples are sufficiently large to be representative of the Italian regional populations independently of demographic flows.

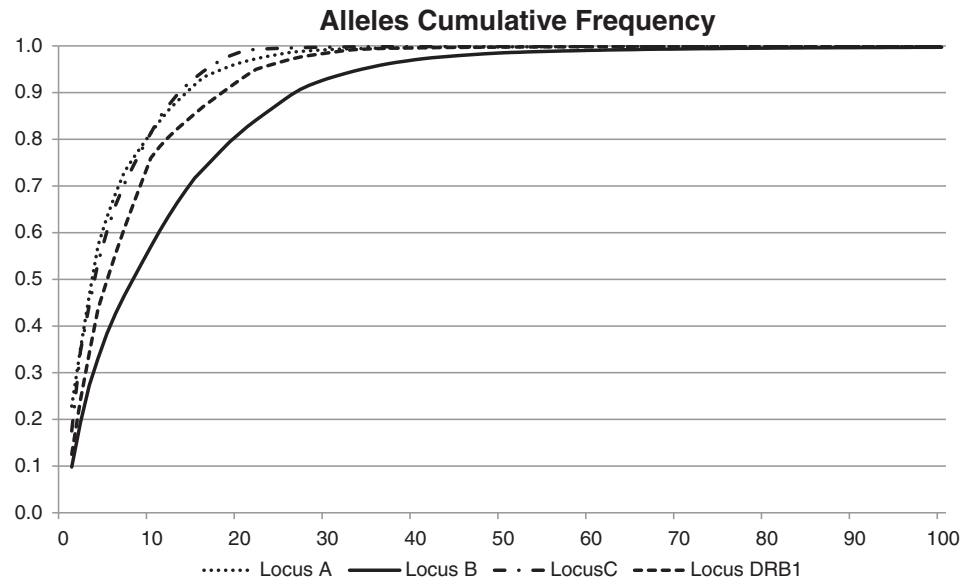
3 | RESULTS

3.1 | HLA allele frequencies

Allele frequencies were calculated from the genotypes of 120 926 individuals processed as previously described and are reported in Table 2 (first 20 frequency ranked alleles) and in Table S2. Cumulative allele frequencies for the four analysed HLA genes are displayed in Figure 1. The greatest allelic diversity in our data set occurs for HLA-B, and the lowest diversity occurs for HLA-C. For each locus, the three most frequent alleles we found were *A*02:01g* (22.8%), *A*24:02g* (12.3%) and *A*01:01g* (11.5%); *B*51:01g* (9.8%), *B*18:01g* (9.5%) and *B*35:01g* (8.0%); *C*04:01g* (17.5%), *C*07:01g* (17.1%) and *C*06:02g* (9.9%); *DRB1*07:01g* (12.5%), *DRB1*11:01g* (11.6%) and *DRB1*11:04g* (10.1%), as reported in Table 2. Hardy-Weinberg equilibrium (HWE) was tested at the regional level, and no significant difference from the expected equilibrium proportions was detected.

TABLE 2 The first 20 frequency-ranked HLA-A, -B, -C and -DRB1 alleles

Rank	HLA-A		HLA-B		HLA-C		HLA-DRB1	
	Allele	Freq. (%)	Allele	Freq. (%)	Allele	Freq. (%)	Allele	Freq. (%)
1	02:01g	22.820	51:01g	9.798	04:01g	17.526	07:01g	12.525
2	24:02g	12.286	18:01g	9.523	07:01g	17.108	11:01g	11.632
3	01:01g	11.528	35:01g	8.004	06:02g	9.921	11:04g	10.067
4	03:01g	10.628	08:01g	5.760	12:03g	9.028	03:01g	9.472
5	11:01g	5.973	07:02g	5.239	07:02g	6.582	01:01g	5.987
6	32:01g	5.217	44:02g	4.402	05:01g	5.773	15:01g	5.874
7	26:01g	4.585	44:03g	3.828	02:02g	4.600	14:01g	5.389
8	68:01g	2.939	49:01g	3.548	15:02g	3.845	13:01g	5.135
9	30:01g	2.697	38:01g	3.451	03:03g	3.604	16:01g	4.957
10	23:01g	2.661	13:02g	3.419	08:02g	3.413	13:02g	4.937
11	29:02g	2.556	35:03g	3.372	01:02g	3.381	01:02g	2.447
12	31:01g	2.336	15:01g	3.116	16:01g	2.533	08:01g	2.017
13	30:02g	1.873	14:02g	2.931	14:02g	2.315	10:01g	1.765
14	02:05g	1.839	35:02g	2.749	03:04g	1.932	04:03g	1.642
15	33:01g	1.745	57:01g	2.559	12:02g	1.637	04:01g	1.636
16	25:01g	1.645	58:01g	1.976	17:01g	1.488	11:03g	1.572
17	68:02g	0.802	50:01g	1.954	07:04g	1.253	15:02g	1.396
18	03:02g	0.792	55:01g	1.936	16:02g	1.114	04:02g	1.381
19	29:01g	0.758	39:01g	1.925	15:05g	0.915	04:05g	1.379
20	33:03g	0.623	52:01g	1.628	03:02g	0.689	12:01g	1.335

FIGURE 1 Cumulative HLA-A, -B, -C, -DRB1 allele frequencies**TABLE 3** Number of common (C) and well-documented (WD) alleles (CWD) in 55 538 Italian individuals

Locus	C	WD	no CWD	CWD	Total number of HLA alleles	% CWD
A	30	38	84	68	152	44.7
B	50	72	126	122	248	49.2
C	24	33	77	57	134	42.5
DRB1	35	36	87	71	158	44.9
Total	139	179	374	318	692	46.0

TABLE 4 Comparison between the numbers of Italian common (C) and well-documented (WD) alleles (CWD), ASHI criteria vs. EFI catalogue

		Italian CWD according to ASHI	EFI CWD catalogue		
			C	WD	No CWD
Overall	C	139	132	7	0
	WD	179	35	75	69
	No CWD	374	5	125	244
	Total	692	172	207	313
Locus A	C	30	28	2	0
	WD	38	5	17	16
	No CWD	84	2	34	48
	Total	152	35	53	64
Locus B	C	50	46	4	0
	WD	72	15	32	25
	No CWD	126	1	42	83
	Total	248	62	78	108
Locus C	C	24	23	1	0
	WD	33	7	13	13
	No CWD	77	2	25	50
	Total	134	32	39	63
Locus DRB1	C	35	35	0	0
	WD	36	8	13	15
	No CWD	87	0	24	63
	Total	158	43	37	78

TABLE 5 The 50 most common haplotypes observed in the Italian population

Rank	HLA haplotype				Frequency
	A	B	C	DRB1	
1	01:01g	08:01g	07:01g	03:01g	0.025357
2	02:01g	18:01g	07:01g	11:04g	0.011435
3	30:01g	13:02g	06:02g	07:01g	0.010879
4	29:02g	44:03g	16:01g	07:01g	0.010829
5	03:01g	07:02g	07:02g	15:01g	0.010167
6	33:01g	14:02g	08:02g	01:02g	0.009459
7	24:02g	35:02g	04:01g	11:04g	0.009276
8	30:02g	18:01g	05:01g	03:01g	0.008925
9	03:01g	35:01g	04:01g	01:01g	0.007673
10	01:01g	57:01g	06:02g	07:01g	0.006078
11	11:01g	35:01g	04:01g	01:01g	0.005041
12	23:01g	44:03g	04:01g	07:01g	0.004771
13	02:01g	13:02g	06:02g	07:01g	0.004577
14	02:01g	35:01g	04:01g	14:01g	0.004511
15	02:01g	07:02g	07:02g	15:01g	0.004326
16	11:01g	35:01g	04:01g	14:01g	0.004191
17	24:02g	18:01g	12:03g	11:04g	0.004070
18	02:01g	18:01g	05:01g	03:01g	0.003987
19	02:05g	50:01g	06:02g	07:01g	0.003841
20	02:01g	08:01g	07:01g	03:01g	0.003554
21	24:02g	18:01g	07:01g	11:04g	0.003441
22	23:01g	49:01g	07:01g	11:01g	0.003120
23	26:01g	38:01g	12:03g	13:01g	0.003038
24	02:05g	58:01g	07:01g	16:01g	0.002856
25	24:02g	07:02g	07:02g	15:01g	0.002715
26	02:01g	51:01g	15:02g	11:01g	0.002711
27	02:01g	57:01g	06:02g	07:01g	0.002604
28	01:01g	52:01g	12:02g	15:02g	0.002543
29	01:01g	15:17g	07:01g	13:02g	0.002512
30	02:01g	44:02g	05:01g	04:01g	0.002458
31	02:01g	44:02g	05:01g	11:01g	0.002458
32	25:01g	18:01g	12:03g	15:01g	0.002451
33	01:01g	35:02g	04:01g	11:04g	0.002418
34	02:01g	51:01g	01:02g	11:01g	0.002297
35	11:01g	52:01g	12:02g	15:02g	0.002279
36	02:01g	18:01g	07:01g	11:01g	0.002149
37	02:01g	44:02g	05:01g	13:01g	0.002074
38	24:02g	13:02g	06:02g	07:01g	0.002074
39	01:01g	37:01g	06:02g	10:01g	0.002046
40	02:01g	39:01g	12:03g	16:01g	0.001893
41	24:02g	08:01g	07:01g	03:01g	0.001840

(Continues)

TABLE 5 (Continued)

Rank	HLA haplotype				Frequency
	A	B	C	DRB1	
42	11:01g	35:01g	04:01g	11:01g	0.001839
43	02:01g	38:01g	12:03g	13:01g	0.001800
44	68:01g	44:02g	07:04g	11:01g	0.001793
45	03:01g	07:02g	07:02g	11:01g	0.001788
46	02:01g	51:01g	14:02g	11:01g	0.001776
47	02:01g	51:01g	02:02g	11:01g	0.001755
48	02:01g	18:01g	12:03g	11:04g	0.001706
49	01:01g	08:01g	07:01g	11:01g	0.001701
50	02:01g	35:01g	04:01g	01:01g	0.001643

The number of observed alleles (Table S3) was then counted in the sample of 55 538 individuals (data set 3) with no “P” group or “G” group alleles at any locus, and they included 226 HLA-A, 343 HLA-B, 201 HLA-C and 210 HLA-DRB1 alleles, which accounted for 5.4%, 6.7%, 5.2% and 8.5%, respectively, of the alleles known at each locus (IPD-IMGT/HLA Database Release 3.32, April 2018; <http://www.ebi.ac.uk/ipd/imgt/hla/stats.html>).

3.2 | Common and well-documented alleles

A total of 692 two field HLA-A, -B, -C and -DRB1 alleles were observed in the Italian population. According to the ASHI criteria,²⁰ we identified 318 (46%) common (C) and well-documented (WD) alleles (Table 3). Among these, 30, 50, 24 and 35 of the HLA-A, -B, -C and -DRB1 alleles, respectively, were common with frequencies greater than 0.001, and 38, 72, 33 and 36, respectively were well-documented alleles observed in at least five independent unrelated individuals or three times in a specific haplotype in unrelated individuals. Almost all individuals in our sample carry at least one CWD allele on each investigated HLA locus.

According to these results, the “no CWD” alleles in the Italian population represent 54% of the total. This high percentage is mainly due to the high number of alleles that appeared just once in data set 3 (from 52.4% at locus A to 68.8% at locus C).

The CWD Italian allele catalogue was also compared with the most recent EFI CWD²¹ catalogue (Table 4).

We found that 65% of the Italian catalogue is shared with the EFI catalogue, and notably, all our common alleles are inside the EFI CWD group. In this comparison, we also observed 130 “no CWD” Italian alleles that move to the C and, more frequently, to the WD allele category according to the EFI catalogue. This result is not surprising because the

EFI CWD catalogue has more WD alleles, reflecting its larger samples.

Vice versa, 69 Italian WD alleles (10%) are not represented in the EFI CWD catalogue, and the prevalence of “no CWD” alleles remained 45.2%, in confirmation of the previous results.

3.3 | Haplotype frequencies

The frequencies of HLA-A-B-C-DRB1 haplotypes were estimated from the sample of 120 926 IBMDR donors with the EM algorithm using the Arlequin programme. The algorithm calculated the frequencies of the 25 057 most frequent haplotypes according to a cumulative frequency of 99.99%. Table S4 lists the haplotypes with a frequency >0.0005.

The most frequent haplotypes (>1%) are *HLA-A*01:01~B*08:01~C*07:01~DRB1*03:01* (2.5%); *HLA-A*02:01~B*18:01~C*07:01~DRB1*11:04* (1.1%) and *HLA-A*30:01~B*13:02~C*06:02~DRB1*07:01* (1.1%). The list of the 50 most common haplotypes found in Italy is shown in Table 5. The frequencies of the 10 most common haplotypes sum up to 11%.

3.4 | Regional distribution of alleles

Haplotype frequencies were compared between national and regional data. The 10 most frequent haplotypes for each Italian region were tabulated and compared with the corresponding national data to assess regional differences in the frequency data (Table S5). The sum of squares (SS) was then calculated to evaluate which regions deviate from the national values (Figure 2), and we found no correlation between the SS value and the regional sample size ($P = 0.359$).

Our results showed that the regions with haplotype frequencies significantly deviating from the national

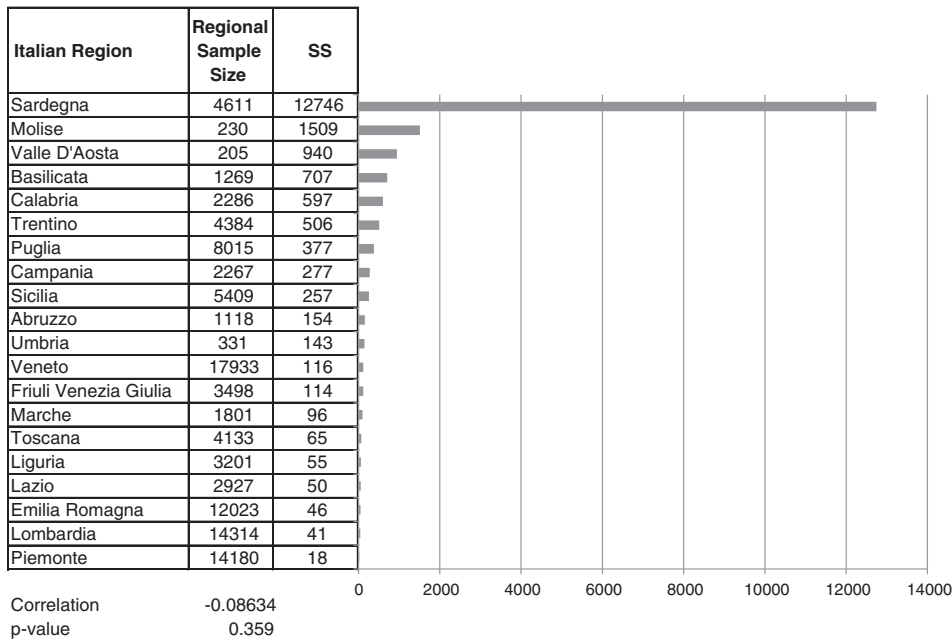


FIGURE 2 Sum of squares (SS) frequency differences between each region and the national data

frequencies are Sardinia (SS = 12 746), Molise (SS = 1509) and Valle D'Aosta (SS = 940).

The graphical representation of this comparison is shown in Figure 3 for the regions with SS > 500.

4 | DISCUSSION

In this study, we presented the HR allele and haplotype frequencies of the Italian population based on a data set of 120 926 unrelated stem cell donors of the Italian Bone Marrow Donor Registry. Compared to previous studies,^{16,18,19} our data included the largest number of HLA-A, -B, -C and -DRB1 alleles and haplotypes ever analysed in Italy.

The ranking and HR haplotype frequencies indicate that differences between all studies are generally small, but our data have significantly expanded the numbers of alleles observed at the HLA-A, -B, -C and -DRB1 loci.

We used the ASHI criteria²⁰ to identify CWD alleles in the Italian population, and we compared the results with the classification of CWD alleles published in Reference.²¹

The percentage of “no CWD” HLA alleles (~50%) vs “common” and “well documented” alleles in the Italian population can be suggestive of some evolutionary pressure toward allele diversification or be the result of several migratory waves, and it could in part justify the high polymorphism of our population with the consequent major difficulty of Italian patients finding an 8/8 matched unrelated donor.

Previous studies^{15,16,18} have already described the genetic heterogeneity of the Italian population as a result of the history of ancient peoples who settled in the country. Genetic drift, the cause of differentiation among the different

ethnic groups and migrations among them over the centuries, probably produced the observed genetic pattern in modern Italy. In addition, other factors contributed to the variable genetic composition of Italy, such as the effect of selection following exposure to pathogens that may have favoured or disfavoured carriers of specific HLA genes.

The haplotype *A*01:01~B*08:01~C*07:01~DRB1*03:01* is the most frequent nationwide and is the most common in 12 out of 20 regions; nevertheless, as described in the results, some HLA four-locus haplotypes exhibit specific regional characteristics.

In Valle d'Aosta, the abovementioned national most frequent haplotype goes down to the 4th ranking position, while in the same region, the most frequent haplotype is *A*02:01~B*35:01~C*04:01~DRB1*14:01*, which interestingly is only the 50th at the national level; furthermore, the third most frequent haplotype in the Valle d'Aosta *A*02:01~B*44:02~C*07:04~DRB1*11:01* is only at the 137th position nationwide.

In Molise, among the first 10 haplotypes is a relatively rare haplotype, *A*02:01~B*14:02~C*08:02~DRB1*11:01*, which in the national list is ranked 2027th (0.000076%).

Sardinia calls for a peculiar consideration: ranking the Sardinian regional data from the most frequent to the less frequent haplotype, the data clearly show that 8 out of 10 Sardinian more frequent haplotypes are virtually absent in the national ranking (frequency <0.01).

This phenomenon is associated with the insular nature of Sardinia, a typical cause of geographic barriers to gene flow and local selective pressure. The most represented haplotype in Sardinia is *A*30:02~B*18:01~C*05:01~DRB1*03:01*, which is significantly associated with celiac disease²⁷ and multiple sclerosis.²⁸

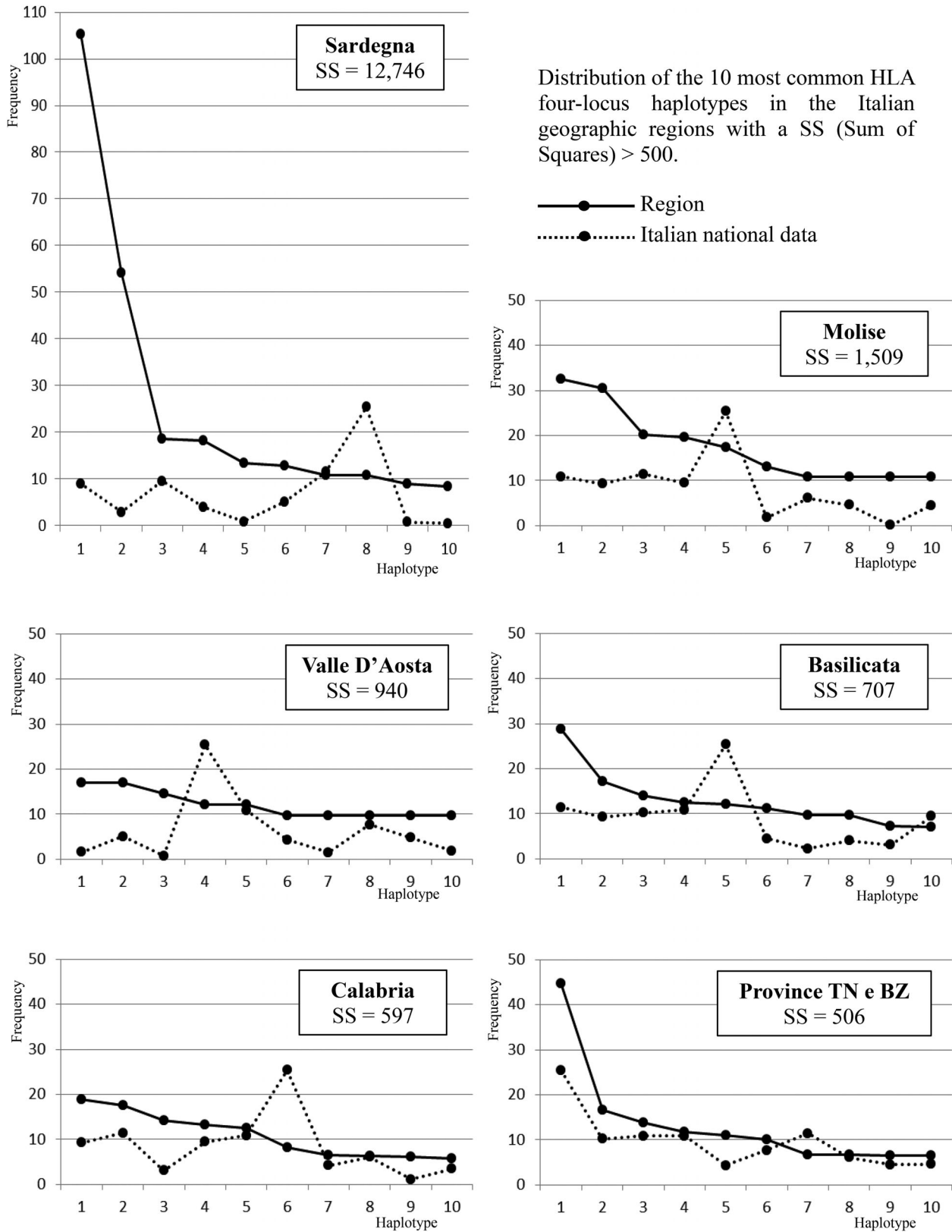


FIGURE 3 Comparison between national and regional haplotype frequencies (per thousand)

Ordering the sum of the SS data for each region, from the most different to the most similar to the national distribution, Sardinia is the first region, followed by Molise, Valle D'Aosta,

Basilicata, Calabria and so on. Apparently, the Italian southern regions are more “different” than the central and northern regions, confirming the south-north genetic gradient observed

in previous studies.^{18,19} This phenomenon can hardly be associated with a single cause; although the Sardinian population suffers from past regional isolation, in other southern regions, there might be other causes explaining this phenomenon.

These data showed that the Italian donor population can be split into at least four areas, with Sardinia representing an area per se and the remaining regions being broadly grouped in north/centre/south, as previously stated.¹⁸

Because donor-recipient matching in haematopoietic stem cell transplantation should, at least, consider the HLA-A, -B, -C and -DRB1 genes at HR,² these data are highly relevant for both individual donor searches and strategic donor registry planning.

The allele and haplotype frequencies obtained in this study are useful for the following purposes: (a) to determine which alleles are suitable to be defined by HR techniques because of the higher heterogeneity; (b) to assign the most likely types in the donor population, recruited in the past and still typed with low- and medium-resolution methods; (c) to categorise a patient's haplotype at the beginning of the unrelated donor search as either common or uncommon in the national donor population, giving an important predictive estimation of the probability of finding a matched unrelated donor.

Moreover, the determination of HR HLA diversity of donors in IBMDR, particularly at the regional level, is a valid tool to develop better recruitment and unrelated donor search strategies to satisfy requests for haematopoietic stem cell donation and will consequently be relevant for resource allocation.

In addition, these data will contribute to HLA and disease association studies as well as forensic and anthropology fields.

ACKNOWLEDGMENTS

The authors want to thank the IBMDR Regional Registries directed by:



A. Zoli—A.O. Ospedali Riuniti Ancona, Immunologia Clinica, Laboratorio Tipizzazione Tissutale, Ancona; F. Papola—A.O. S. Salvatore, Servizio di Immunematologia, Laboratorio Tipizzazione Tissutale, L'Aquila; T. Mininni—A.O. Policlinico Consorziato, Pat. Clinica II, Laboratorio Tipizzazione Tissutale, Bari; A. Bontadini—Ospedale S. Orsola, Laboratorio Immunogenetica SIMT, Bologna; C. Carcassi—P.O. R. Binaghi, Centro Regionale Trapianti, (CRT), Cagliari; G. Rombolà—A.O. Careggi, Piastra dei Servizi Immunogenetica Biologia dei Trapianti, Firenze; L. Garbarino—E.O. Ospedali Galliera, Lab. di Istocompatibilit , Genova; M. Cardillo—Ospedale Maggiore Policlinico, Immunologia dei Trapianti, Milano; G. Giannella—Ospedale Madonna delle Grazie, S.I.T., Matera; L. Del Vecchio—II Universita' degli Studi di Napoli, SIMT, Napoli; B. Piazza—A.R. N.A.S. Civico

C.R.T. Sicilia, Palermo; M. Marchesi—Policlinico Montelucente, CRRT, Perugia; G. Romeo—A.O. Bianchi Melacrino Morelli, Laboratorio Tipizzazione Tissutale, Reggio Calabria; P. Grammatico—A.O. S. Camillo Forlanini, Laboratorio Genetica Medica, Roma; A. Amoroso—A.O.U. S. Giovanni Battista, Servizio Immunologia dei Trapianti Torino; V. Miotti—A.O. S. Maria Misericordia, Medicina Trasfusionale, Laboratorio Tipizzazione Tissutale, Udine; A.M. Gandini—Ospedale Civile Maggiore, SIT, Verona.

CONFLICT OF INTEREST

The authors have declared no conflicting interests.

ORCID

Nicoletta Sacchi  <https://orcid.org/0000-0002-7164-401X>
Mauro Castagnetta  <https://orcid.org/0000-0002-8052-9515>

REFERENCES

1. Kunze-Schumacher H, Blasczyk R, Bade-Doeding C. Soluble HLA technology as a strategy to evaluate the impact of HLA mismatches. *J Immunol Res*. 2014;2014:1-8.
2. Lee SJ, Klein J, Haagenson M, et al. High-resolution donor-recipient HLA matching contributes to the success of unrelated donor marrow transplantation. *Blood*. 2007;110(13):4576-4583.
3. Verneris MR, Lee SJ, Ahn KW, et al. HLA mismatch is associated with worse outcomes after unrelated donor reduced-intensity conditioning hematopoietic cell transplantation: an analysis from the Center for International Blood and Marrow Transplant Research. *Biol Blood Marrow Transplant*. 2015;21(10):1783-1789.
4. Petersdorf E. Optimal HLA matching in hematopoietic cell transplantation. *Curr Opin Immunol*. 2008;20(5):588-593.
5. Furst D, Muller C, Vucinic V, et al. High-resolution HLA matching in hematopoietic stem cell transplantation: a retrospective collaborative analysis. *Blood*. 2013;122(18):3220-3229.
6. Passweg JR, Baldomero H, Bader P, et al. Use of haploidentical stem cell transplantation continues to increase: the 2015 European Society for Blood and Marrow Transplant activity survey report. *Bone Marrow Transplant*. 2017;52:811-817.
7. D'Souza A, Lee S, Zhu X, Pasquini M. Current use and trends in hematopoietic cell transplantation in the United States. *Biol Blood Marrow Transplant*. 2017;23:1417-1421.
8. Ustun C, Giannotti F, Zhang MJ, et al. Outcomes of UCB transplantation are comparable in FLT3+ AML: results of CIBMTR, EUROCORD and EBMT collaborative analysis. *Leukemia*. 2017 Jun;31(6):1408-1414.
9. Piemontese S, Ciceri F, Labopin M, et al. A comparison between allogeneic stem cell transplantation from unmanipulated haploidentical and unrelated donors in acute leukemia. *J Hematol Oncol*. 2017;10:24.
10. Kollman C, Abella E, Baitty RL, et al. Assessment of optimal size and composition of the U.S. National Registry of hematopoietic stem cell donors. *Transplantation*. 2004;78(1):89-95.

11. Schmidt AH, Biesinger L, Baier D, Harf P, Rutt C. Aging of registered stem cell donors: implications for donor recruitment. *Bone Marrow Transplant*. 2008;41(7):605-612.
12. Schmidt AH, Baier D, Solloch UV, et al. Estimation of high-resolution HLA-A, -B, -C, -DRB1 allele and haplotype frequencies based on 8862 German stem cell donors and implications for strategic donor registry planning. *Hum Immunol*. 2009;70(11):895-902.
13. Gragert L, Madbouly A, Freeman J, Maiers M. Six-locus high resolution HLA haplotype frequencies derived from mixed-resolution DNA typing for the entire US donor registry. *Hum Immunol*. 2013;74:1313-1320.
14. Zhou XY, Zhu FM, Li JP, et al. High-resolution analyses of human leukocyte antigens allele and haplotype frequencies based on 169,995 volunteers from the China Bone Marrow Donor Registry Program. *PLoS One*. 2015;10:e0139485. <https://doi.org/10.1371/journal.pone.0139485>.
15. Piazza A, Cappello N, Olivetti E, Rendine S. A genetic history of Italy. *Ann Hum Genet*. 1988;52:203-213.
16. Rendine S, Ferrero N, Sacchi N, Costa C, Pollichieni S, Amoroso A. 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. *Hum Immunol*. 2012;73:399-404.
17. Picardi A, Arcese W, Pollichieni S, et al. The Rome Transplant Network model compared to the Italian Bone Marrow Donor Registry activity for unrelated donor search process and transplant efficiency for hematologic malignancy. *Transfusion*. 2017;57(7):1734-1743. <https://doi.org/10.1111/trf.14131>.
18. Rendine S, Borelli I, Barbanti M, Sacchi N, Roggero S, Curtioni ES. HLA polymorphisms in Italian bone marrow donors: a regional analysis. *Tissue Antigens*. 1998;52:135-146.
19. Amoroso A, Ferrero NM, Rendine S, Sacchi N. Le caratteristiche HLA della popolazione Italiana: Analisi di 370.000 volontari iscritti all' IBMDR. *Analysis*. 2010;23:1-2.
20. Mack SJ, Cano P, Hollenbach JA, et al. Common and well-documented HLA alleles: 2012 update to the CWD catalogue. *Tissue Antigens*. 2013;81:194-203. <https://doi.org/10.1111/tan.12093>.
21. Sanchez-Mazas A, Nunes JM, Middleton D, et al. Common and well-documented HLA alleles over all of Europe and within European sub-regions: a catalogue from the European Federation for Immunogenetics. *HLA*. 2017;89(2):104-113. <https://doi.org/10.1111/tan.12956>.
22. Waller MJ, Robinson J, Fail SC, Marsh SGE. Exon identities and ambiguous typing combinations—July 2008. http://www.ebi.ac.uk/imgt/hla/pdf/ambiguity_v2220.pdf. Accessed 20 May, 2009.
23. Robinson J, Halliwell JA, Hayhurst JH, Flicek P, Parham P, Marsh SGE. The IPD and IPD-IMGT/HLA database: allele variant databases. *Nucleic Acids Res*. 2015;43:D423-D431.
24. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour*. 2010;10:564-567.
25. Fallin D, Schork NJ. Accuracy of haplotype frequency estimation for biallelic loci, via the expectation-maximization algorithm for unphased diploid genotype data. *Am J Hum Genet*. 2000;67(4):947-959.
26. Guo SW, Thompson EA. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*. 1992;48:361-372.
27. Congia M, Frau F, Lampis R, et al. A high frequency of the A30, B18, DR3, DRw52, DQw2 extended haplotype in Sardinian celiac disease patients: further evidence that disease susceptibility is conferred by DQ A1*0501, B1*0201. *Tissue Antigens*. 1992;39(2):78-83.
28. Bitti PP, Murgia BS, Ticca A, et al. Association between the ancestral haplotype HLA A30B18DR3 and multiple sclerosis in central Sardinia. *Genet Epidemiol*. 2001;20(2):271-283.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Sacchi N, Castagnetta M, Miotti V, Garbarino L, Gallina A. High-resolution analysis of the HLA-A, -B, -C and -DRB1 alleles and national and regional haplotype frequencies based on 120 926 volunteers from the Italian Bone Marrow Donor Registry. *HLA*. 2019;94:285–295. <https://doi.org/10.1111/tan.13613>