



Significance of regional population HLA immunogenetic datasets in the efficacy of umbilical cord blood banks and marrow donor registries: a study of Cretan HLA genetic diversity

Helen Latsoudis¹, Emmanouil Stylianakis¹, Irene Mavroudi^{2,3}, Alexandros Kanterakis¹, Pavlos Pavlidis¹, Anthie Georgopoulou^{2,3}, Aristeia Batsali^{2,3}, Ioanna Gontika^{2,3}, Irene Fragiadaki^{2,3}, Maria Zamanakou⁴, Anastasios E Germenis⁴, Helen A. Papadaki^{2,3,*}

¹ Institute of Computer Sciences, Foundation for Research and Technology Hellas, Heraklion, Greece

² Public Cord Blood Bank of Crete, Department of Hematology, University Hospital of Heraklion, Heraklion, Greece

³ Haemopoiesis Research Laboratory, School of Medicine, University of Crete, Heraklion, Greece

⁴ CeMIA SA, Larissa, Greece

ARTICLE INFO

Article History:

Received 14 May 2021

Accepted 15 July 2021

Available online xxx

Key Words:

donor selection

HLA

public cord blood banks

regional genetic diversity

ABSTRACT

Background aims: The high genetic diversity of HLA across populations significantly confines the effectiveness of a donor or umbilical cord blood search for allogeneic hematopoietic stem cell transplantation (HSCT). This study aims to probe the HLA immunogenetic profile of the population of Crete, a Greek region with specific geographic and historical characteristics, and to investigate potential patterns in HLA distribution following comparison with the Deutsche Knochenmarkspenderdatei (DKMS) donor registry. It also aims to highlight the importance of regional public cord blood banks (PCBBs) in fulfilling HSCT needs, especially in countries with significant genetic diversity.

Methods: A cohort of 1835 samples representative of the Cretan population was typed for HLA class I (*HLA-A*, *HLA-B*, *HLA-C*) and class II (*HLA-DRB1*, *HLA-DQB1*, *HLA-DPB1*) loci by high-resolution second field next-generation sequencing. Data were compared with the respective HLA profiles of 12 DKMS populations ($n = 20\,032$). Advanced statistical and bioinformatics methods were employed to assess specific intra- and inter-population genetic indexes associated with the regional and geographic distribution of HLA alleles and haplotypes.

Results: A considerable HLA allelic and haplotypic diversity was identified among the Cretan samples and between the latter and the pooled DKMS cohort. Even though the HLA allele and haplotype frequency distribution was similar to regions of close geographic proximity to Crete, a clinal distribution pattern from the northern to southern regions was identified. Significant differences were also observed between Crete and the Greek population of DKMS.

Conclusions: This study provides an in-depth characterization of the HLA immunogenetic profile in Crete and reveals the importance of demographic history in HLA heterogeneity and donor selection. The novel HLA allele and haplotype frequency comparative data between the Cretan and other European populations signify the importance of regional PCBBs in prioritizing HLA diversity to efficiently promote the HSCT program at the national level and beyond.

© 2021 International Society for Cell & Gene Therapy. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Introduction

Allogeneic hematopoietic stem cell transplantation (HSCT) is an effective curative option for several malignant and non-malignant hematologic disorders, and the outcome of the treatment is highly

dependent on the HLA histocompatibility between the patient and donor. Stem cells derived from genotypically HLA-matched related donors represent the ideal grafts for this treatment approach. However, approximately 70% of patients in Western countries do not have a suitable related donor, and thus alternative sources are frequently considered, including HLA-matched unrelated donors (MUDs), umbilical cord blood units (CBUs) and haploidentical donors [1]. Currently, the criteria for an optimal graft are 10/10 HLA matching for *HLA-A*, *HLA-B*, *HLA-C*, *HLA-DRB1* and *HLA-DQB1* for MUDs and eight of eight HLA matching for

* Correspondence: Helen A. Papadaki, MD, PhD, Public Cord Blood Bank of Crete, Department of Hematology, University Hospital of Heraklion, P.O. Box 1352, Heraklion, Greece.

E-mail address: e.papadaki@uoc.gr (H.A. Papadaki).

HLA-A, *HLA-B*, *HLA-C* and *HLA-DRB1* for CBUs based on high-resolution HLA typing [1–3].

The application of next-generation sequencing (NGS) for increased sequence lengths in combination with advanced bioinformatics tools has emerged as the gold standard for both accurate high-resolution HLA typing and the generation of full-length HLA-inferred or phased haplotypes [4]. The identification of optimal MUDs or CBUs for patients in need has been improved recently due to a number of reasons, including the wide implementation of NGS-based high-resolution HLA typing across registries and public cord blood banks (PCBBs); the development of various software programs that can infer HLA haplotype frequencies in high resolution, which in turn can be used to predict the number of potentially matched donors; and the growing number of volunteer donors and CBUs worldwide [5–8]. However, because of inter-individual HLA allele disparities, a significant proportion of patients do not find a suitable MUD or CBU, and this is particularly evident in populations with high genetic diversity [1].

In addition, the determination of HLA allele and haplotype frequencies would be particularly important for the HSCT procedure in populations with significant ethnic and genetic diversity as well as in registries of donors and CBUs. Establishment of such databases would reveal novel/rare HLA alleles and haplotypes in representative populations, facilitate donor searches and accelerate the turnaround time of the procedure while at the same time helping the transplant center to rapidly decide whether to search for a MUD or CBU or proceed with a haploidentical donor. Access to this information would also contribute to the development of strategies by national registries and PCBBs aiming to increase not only the number of donors but also, most importantly, the HLA diversity through enrichment with appropriate donors from underrepresented population groups. Furthermore, although not yet routinely applied, HLA haplotype matching, in addition to HLA allele typing, has been associated with better transplantation outcomes [9,10].

Recent studies from the growing PCBB of the island of Crete, Greece (<https://share.wmda.info/display/WMDAREG/WO-2007>), have resulted in a number of novel HLA variants, indicating a diverse genetic background of the Cretan population. In the present study, the authors have performed high-resolution NGS typing of HLA class I (*HLA-A*, *HLA-B*, *HLA-C*) and class II (*HLA-DRB1*, *HLA-DQB1*, *HLA-DPB1*) genes in a representative sample of individuals of Cretan origin residing on the island and have mapped for the first time the immunogenetic profile of the Cretan population. The authors have also compared the HLA allele and haplotype frequencies of the Cretan population with representative European populations of the Deutsche Knochenmarkspenderdatei (DKMS) donor registry [11] using advanced bioinformatics and statistical tools. The novel HLA allele and haplotype frequency comparative data between the Cretan and other European populations highlight the significance of a regional PCBB in prioritizing the HLA diversity coverage of the population to efficiently promote the HSCT program at the national level and beyond.

Methods

Study cohorts

The original cohort consisted of 2000 individuals (292 CBUs and 1708 peripheral blood samples) representatively selected from the four prefectures of Crete according to the latest (2011) population census (total population, $n = 623\,000$), with self-assessed parentage and grand parentage at recruitment. All human blood samples were obtained under informed consent approved by the Institutional Review Board (the 7th Health Region of Crete) in accordance with the Declaration of Helsinki. Exclusion of 114 samples with non-Cretan origin ($n = 68$) and first-, second- and third-degree blood relations ($n = 46$) resulted in 1886 individuals (223 CBUs and 1663 peripheral

blood samples) available for sequencing analyses. Subsequent exclusion of 43 samples with ambiguities due to low sequencing quality control metrics—that is, low read counts (≤ 99 bp) and/or coverage depth (≤ 20 reads for HLA class I and II exon 2 or 3—and eight samples with novel HLA alleles resulted in a final cohort of 1835 samples available for comparative analyses. Specifically, data on allele and haplotype frequencies from the Cretan cohort were compared with a large dataset of 20 032 donors registered in the DKMS. In total, 12 minority countries with at least 1000 donors typed for HLA loci *A*, *B*, *C* and *DRB1* [11] were included: Austria (1698), Bosnia and Herzegovina (1028), Croatia (2057), France (1406), Greece (1894), Italy (1159), the Netherlands (1374), Portugal (1176), Romania (1234), Spain (1107), Turkey (4856) and the UK (1043). Among these, Turkey and Italy were also typed for *HLA-DQB1*, allowing a five-locus haplotype frequency comparison between these samples and the Cretan cohort. Only high-resolution HLA-typed DKMS datasets listed in Allele Frequency Net Database as of September 2019 [12] were used as input for corresponding comparative analyses. The study was conducted by the PCBB of Crete under the auspices of the 7th Health Region of Crete.

DNA extraction and HLA typing

Genomic DNA (gDNA) was extracted from peripheral blood or umbilical cord blood (UCB) lymphocytes using the iPrep PureLink gDNA blood kit (Thermo Fisher Scientific, Waltham, MA, USA). The complete gDNA sequence for HLA class I (*HLA-A*, *HLA-B*, *HLA-C*) and the gDNA region between exon 2 and 3'UTR for class II (*HLA-DRB1*, *HLA-DQB1*, *HLA-DPB1*) loci were genotyped by high-resolution NGS using the Ion Torrent S5 sequencing system (Thermo Fisher Scientific) at CeMIA SA laboratories (Larissa, Greece) as previously described [13–17] and presented in the supplementary methods. For alleles identical over exons 2 and 3 (for class I *HLA-A*, *HLA-B*, *HLA-C*) and exon 2 (for class II *HLA-DRB1*, *HLA-DQB1*, *HLA-DPB1*), definitions were based on release 3.39.0 of the ambiguous allele combinations file in the IPD-IMGT/HLA database [18], and allele “g” grouping was performed [11]. The HLA g-grouped alleles were used only for carrying out allele and haplotype comparative analyses between the Cretan and DKMS cohorts. All novel HLA alleles identified by NGS were verified by Sanger sequencing and were subsequently assigned a new name by the World Health Organization Nomenclature Committee for Factors of the HLA System [13–17].

Intra- and inter-population allele and haplotype frequency estimation

Although allele frequencies of the second field resolution-typed class I (*HLA-A*, *HLA-B*, *HLA-C*) and class II (*HLA-DRB1*, *HLA-DQB1*, *HLA-DPB1*) data for the Cretan cohort were determined by direct counting of alleles at each HLA locus, maximum likelihood HLA haplotype frequencies were estimated by the expectation–maximization algorithm implemented with Hapl-o-Mat v 1.1 software [19] using 1000 iterations (see supplementary methods).

Maximum likelihood haplotype frequencies as well as two- to five-point associations for the second field resolution-typed and g-grouped HLA class I (*HLA-A*, *HLA-B*, *HLA-C*) and class II (*HLA-DRB1* and *HLA-DQB1*) data were compared either (i) between the Cretan samples and the pooled DKMS cohort or (ii) between the Cretan and each of the 12 DKMS populations by employing a two-tailed Fisher's exact test using all haplotypes with estimated frequency (see supplementary methods) greater than $f_{\min} = 1/2n$ (n = sample size). The same test (significance, $\alpha \leq 0.05$), followed by the Holm–Bonferroni correction, was employed for the comparison of all g-grouped allele frequencies, which were determined by direct counting of alleles at each HLA locus between the Cretan samples and the pooled DKMS cohort or between the Cretan samples and each of the 12 DKMS minority populations.

Estimation of intra-population HLA genetic diversity measures

Hardy–Weinberg equilibrium (HWE) tests at HLA locus-by-locus adjusted for multiple testing (see supplementary methods) were performed by Arlequin 3.5.2.2 [20] and confirmed by PyPop 0.7.0 [21]. Allele frequencies for each HLA locus were also employed to test potential departures from selective neutrality in the direction of heterozygote advantage or disadvantage using the Ewens–Watterson homozygosity test of neutrality [22] with PyPop 0.7.0. All calculated differences between F_o (observed homozygosity) and F_e (expected homozygosity) were subsequently normalized (F_{ND}) and subjected to a two-tailed test that was also adjusted for multiple testing (see supplementary methods).

Linkage disequilibrium (LD) for pairwise non-random association between a pair of alleles at two different HLA loci [23] was initially calculated by custom scripts in R and subsequently corroborated by PyPop 0.7.0, using the haplotypes that were estimated by Hapl-o-Mat v 1.1. Since HLA loci are multiallelic, the authors employed the effect size statistic D' and the Cramer's V or W_n statistic to account for the different number of alleles at each locus in an attempt to express the overall LD strength or degree of deviation per HLA locus in the Cretan cohort [23–25]. Finally, the empirical significance values were corrected by the number of pair–locus comparisons performed on the data (see supplementary methods).

Estimation of inter-population HLA genetic diversity

The level of genetic diversity of each HLA locus was determined by calculating the genetic distance (D_1) between Crete and each of the 12 DKMS populations using Prevosti's metric [26]. Subsequently, any HLA genetic divergence or resemblance between the cohorts studied, represented by population–pairwise differences in HLA haplotypes, was graphically displayed in a hierarchical clustered heatmap (see supplementary methods) using Seaborn [27].

Principal component analysis

Principal component analysis (PCA) [28] was employed using Excel XLSTAT 2020 (Microsoft Corporation, Redmond, WA, USA) to visualize HLA genetic variation between the different populations studied since the dimension of the HLA haplotype frequency vectors made it hard to examine and visualize the respective data. Subsequently, scatter plots of the first two principal components (PC1 and PC2) of the dataset were produced based on the frequencies of the HLA-A, HLA-B, HLA-C and HLA-DRB1 alleles between Crete and each of the 12 populations of DKMS.

Linear regression analysis

Linear regression analysis was employed using custom scripts in R 4.0.3. to unravel possible patterns of correlation between the distribution of (i) the HLA allele frequencies and (ii) the HLA haplotype frequencies and geographic latitude of the populations studied (see supplementary methods). Specifically, comparisons were performed between Crete and each of the 12 DKMS populations using (i) the alleles with significantly different frequencies and (ii) the 10 most common HLA haplotypes obtained after comparing the 1835 Cretans and 20032 individuals encompassing the pooled DKMS cohort for both types of analyses.

Results

Sequencing quality control metrics

In the 1843 samples used for sequencing, the amplified exon/intron regions of each allele were nearly fully covered (detection

percentage) with an average coverage depth of >208 reads (see supplementary Table 1) per nucleotide position for both HLA class I (99.93%) and class II (99.72%) loci. The variance in coverage was attributed to 1.36% and 4.18% of samples that were not fully covered, with values ranging from 54% to 99% and from 50% to 99% for HLA class I and class II key exons (i.e., exons 2 and 3), respectively. Following quality control check, 2 200 579 and 2 127 547 total high-quality reads were produced, 37.3% and 62.7% of which were attributed to exons 2 and 3 of class I and 37.5% of which were attributed to exon 2 of class II, respectively (see supplementary Figure 1). The minimum average depths for exons 2 and 3 of the corresponding HLA class I and II loci are summarized in supplementary Table 1.

HLA allele frequencies in Crete

High-resolution sequencing resulted in the detection of eight novel alleles [13–17]. Among the typed alleles of the remaining 1835 Cretan samples (see supplementary Table 2), HLA-B exhibited the highest allelic variability ($n = 75$ alleles) compared with HLA-A ($n = 40$) and HLA-C ($n = 38$) class I loci, at 47 out of 75 alleles (62.67%), yet being observed with allele frequencies of <1% and contributing 10.76% of the cumulative locus allele frequency (Figure 1A). In accordance with previous studies [29,30], the top two common alleles for HLA-A, A*02:01 (19.37%) and A*24:02 (15.64%), and HLA-C, C*04:01 (22.89%) and C*07:01 (13.08%), accounted for 35.01% and 35.97%, respectively, of each cumulative locus allele frequency (Figure 1A). Similarly, the most common HLA-B*51:01 and HLA-B*35:01 accounted for 9.78% and 9.51% of HLA-B allelic diversity in our population (Figure 1A).

Predominant alleles were also observed in HLA class II loci, where DQB1*03:01 and DPB1*04:01 accounted for 30.08% and 33.5% of the cumulative allele frequency at HLA-DQB1 ($n = 18$ alleles) and HLA-DPB1 ($n = 41$ alleles), respectively (Figure 1B; also see supplementary Table 2). Interestingly, the alleles at HLA-DPB1 with frequencies <1% prevailed, and although they represented 73.17% of the total number of alleles (i.e., 30 of 41), they contributed only 5.23% of the cumulative locus allele frequency (Figure 1B). By contrast, the majority of HLA-DRB1 (i.e., 21 of 35) and HLA-DQB1 (i.e., 13 of 18) alleles were characterized by a frequency of >1%, and similar to HLA class I loci, 61.9% of DRB1 alleles (i.e., 13 of 21) were maintained at relatively even frequencies (range, 1–4%).

HLA haplotype frequencies in Crete

The six-locus HLA haplotypes inferred using the expectation–maximization algorithm on the second field HLA typed data of the 1835 Cretan samples, together with the maximum likelihood estimated haplotype frequencies ($f_{min} \geq 1/2n$), are outlined in supplementary Table 3. Specifically, significant diversity was observed among the 1735 inferred six-locus haplotypes, with 63.3% estimated only once. Such variation resulted in very low frequencies of the top 10 haplotypes that ranged between 0.43% and 0.98%. It is noteworthy that both the first (A*33:01-C*08:02-B*14:02-DRB1*01:02-DQB1*05:01-DPB1*04:01) and second (A*24:02-C*04:01-B*35:02-DRB1*11:04-DQB1*03:01-DPB1*04:01) most common haplotypes are frequently observed in countries of the Mediterranean Sea basin, including Tunisia [12], South Italy and Sardinia [11,31,32], Spain [11,32,33] and Turkey [11]. The haplotype A*01:01-C*07:01-B*08:01-DRB1*03:01-DQB1*02:01-DPB1*01:01, reported as common in most European countries [11,29–35], was fifth in rank among the top 10 haplotypes of the Cretan cohort. The low frequency (i.e., 0.56%) of this ancestral haplotype as well as of all common haplotypes corroborates the suspected HLA diversity of the Cretan population and probably reflects the admixture of Mediterranean populations that might have occurred through the course of history in Crete.

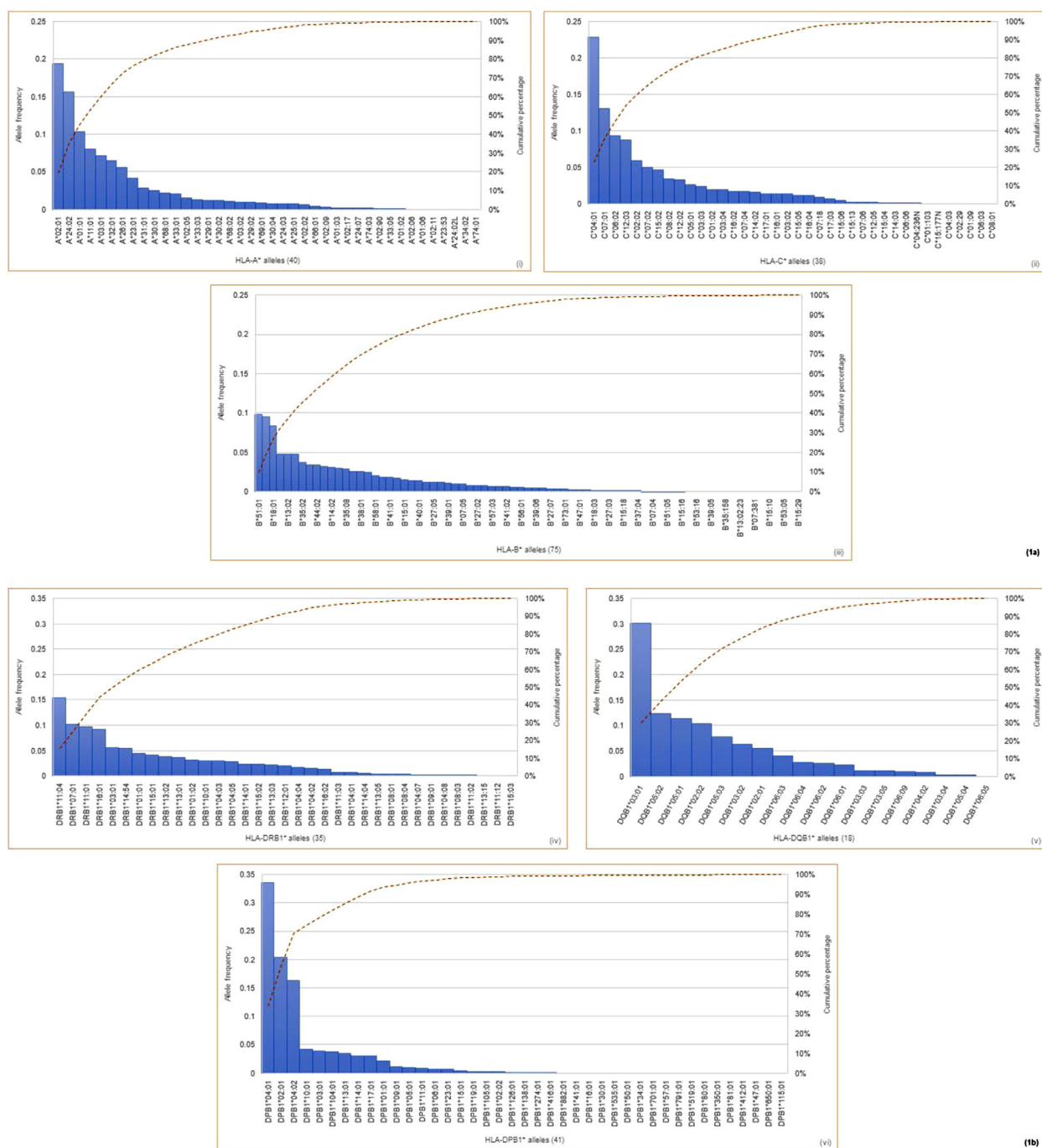


Fig. 1. Pareto plots of HLA Class I (A) and Class II (B) allele frequencies in the Cretan population. The distribution of data is plotted in descending order of allele frequency with a cumulative line on the secondary axis as a percentage of the total. The number in parentheses denotes the number of alleles at each HLA locus.

HLA genetic diversity in the Cretan population

HLA genetic diversity measures among the 1835 Cretans resulted in non-significant deviations from HWE expectations. Specifically, the estimated significance ($\alpha \leq 8.33\text{e-}03$) values were 0.71846 for *HLA-A*, 0.26119 for *HLA-B*, 0.49755 for *HLA-C*, 0.28310 for *HLA-DRB1*, 0.15420 for *HLA-DQB1* and 0.01178 for *HLA-DPB1*. The minor, but significant, departure at the *HLA-DPB1* locus, prior to correction, may be attributed to the increased number of low-frequency alleles that could have caused an overestimation of HWE deviation from its expected proportions. However, all deviations from expected heterozygosity had a reduced heterozygosity, with a mean difference of 0.00429.

To test whether such departures from HWE could be attributed to selection processes, the authors applied the Ewens–Watterson test of neutrality [21,22]. Subsequent analyses using the Cretan cohort ($2n = 3670$) suggested evidence of balancing selection for all HLA loci except *DPB1*. Specifically, the lower level of F_o compared with F_e under neutrality (see supplementary Table 4) implied increased genetic diversity relative to neutral expectations, probably due to frequency-dependent selection or heterozygote advantage [36]. Moreover, the loci with the higher diversity (*HLA-B*) and lower number of rare alleles (*HLA-DRB1* and *HLA-DQB1*) showed statistically significant deviation from neutrality, with $F_{ND} < 0$, exhibiting, therefore, a more pronounced balancing selection [34,36–38]. By contrast, the increased homozygosity compared with that expected under

neutrality with $F_{ND} > 0$ suggested that *HLA-DPB1* may be under directional selection, which, in agreement with previous studies [34,38], accords with the high frequency (33.51%) of *HLA-DPB1*04:01* (see supplementary Table 4).

Pairwise non-random association of estimated HLA haplotypes using the global LD measures D' and Cramer's V or Wn showed that despite the non-significant disequilibrium between *HLA-DPB1* and any of the class I genes, all remaining loci were in LD with each other (see supplementary Table 5; see supplementary Figure 2). Moreover, the strongest LD was calculated between physically close or neighboring HLA loci—that is, *DRB1-DQB1* allele pairs (~70 kb apart, GRCh37/hg19), followed by *C-B* pairs (~80 kb apart, GRCh37/hg19). However, elevated values of LD may not necessarily be attributed to only the physical distance between these loci but also to the action of balancing selection that maintains genetic diversity [25,38]. Undoubtedly, among the HLA genes, *HLA-B* and *HLA-DRB1* are considered the most polymorphic [18], contributing 26.702% and 10.021% of the 28 320 HLA alleles reported in the IPD-IMGT/HLA database (v.3.42) through October 2020. Such a hypothesis is supported by the observed LD between *HLA-A* and *HLA-C* loci, which, even though they are physically closer (i.e., ~1300 kb) than *HLA-A* and *HLA-B* (~1380 kb), exhibit looser (but still significant) linkage (see supplementary Table 5; see supplementary Figure 2). Significant long-distance LD was also observed between the physically distant *HLA-DRB1* and *HLA-B* (~1200 kb) or *HLA-C* (~1280 kb) loci. These results, which are in accordance with previous observations [30,33–35,38], indicate that in addition to physical distance, other parameters, including sequence diversity, recombination rate and selection, also determine the strength of LD between two HLA loci.

Comparative analyses with DKMS registry

Allele and haplotype comparisons were performed between the 1835 samples of the Cretan cohort and the DKMS registry using either the 20032 donors of the 12 pooled DKMS populations (see supplementary Tables 6A, 7A) or the corresponding number of donors from each individual population separately (see supplementary Tables 6B, 7B). For this purpose, typed alleles for *HLA-A*, *HLA-B*, *HLA-C* and *HLA-DRB1* loci were g-grouped and allele and haplotype frequencies calculated accordingly.

HLA allele frequencies

Allele g-grouping resulted in 38, 72, 33, 34 and 30 allele groups for *HLA-A*, *HLA-B*, *HLA-C*, *HLA-DRB1* and *HLA-DQB1* loci, respectively. Interestingly, 15 alleles, all at HLA class I loci (Table 1), were detected in the Cretan cohort but not in any of the 20032 DKMS donors. Among these, *C*15:177N* was also absent from the global catalog of common, intermediate and well-documented HLA alleles (3.0.0), whereas *B*58:65* was reported once in a sample of South or Central America/Hispanic/Latino ancestry [39].

Subsequent comparative analyses between the Cretan cohort and the pooled DKMS registry revealed several significant differences in the distribution of common allele frequencies following Bonferroni correction (Figure 2; also see supplementary Table 6A). The strongest associations were detected for HLA class II *DRB1* alleles (i.e., 58.82%) (Figure 2D), whereas *HLA-C* and *HLA-A* loci followed thereafter, encompassing 12 (i.e., 42.86%) (Figure 2B) and 15 (i.e., 41.67%) (Figure 2A) significant differentially distributed alleles, respectively. Even though *HLA-B* exhibited the highest allelic diversity among the class I and II loci, only 31.25% of its alleles had significantly different frequencies between the cohorts studied (Figure 2C).

Briefly, among the 15 *HLA-A* alleles, the strongest associations between Crete and the pooled DKMS registry were noted for *A*03:01* ($P = 2.616 \times 10^{-17}$) and the most commonly observed alleles in the Cretan cohort, *A*24:02* ($P = 3.084 \times 10^{-14}$) and *A*02:01* ($P = 5.664 \times 10^{-14}$). Specifically, although *A*24:02* was found more

Table 1

Frequencies and absolute counts of the 15 known *HLA-A*, *HLA-B* and *HLA-C* alleles that were identified in the 1835 samples from Crete but not detected in the combined cohort of 20032 individuals registered in DKMS.

HLA allele	Allele counts	Allele frequency
<i>A*02:90</i>	4	1.089×10^{-03}
<i>A*23:53</i>	1	2.724×10^{-04}
<i>C*01:09</i>	1	2.724×10^{-04}
<i>C*01:103</i>	2	5.449×10^{-04}
<i>C*04:236N</i>	3	8.174×10^{-04}
<i>C*06:03</i>	1	2.724×10^{-04}
<i>C*15:177N</i>	2	5.449×10^{-04}
<i>B*07:18</i>	1	2.724×10^{-04}
<i>B*08:37</i>	1	2.724×10^{-04}
<i>B*15:37</i>	2	5.449×10^{-04}
<i>B*35:158</i>	1	2.724×10^{-04}
<i>B*51:24</i>	1	2.724×10^{-04}
<i>B*53:05</i>	1	2.724×10^{-04}
<i>B*53:16</i>	1	2.724×10^{-04}
<i>B*58:65</i>	1	2.724×10^{-04}

frequently in populations from the Mediterranean basin (i.e., Cretans, Greeks and Turks), *A*02:01* was the least frequent in Crete and Turkey (see supplementary Table 6B), showing, however, a direct correlation with the decrease in latitude (see supplementary Figure 3A). It is noteworthy that although the difference in distribution of this allele between Cretans and the Greek DKMS minority was highly significant ($P = 8.311 \times 10^{-12}$), it did not reach significance following comparison with the Turks or Italians (see supplementary Table 6B). Even though the *HLA-A* locus showed a strong sign of latitudinal cline shaped by the distribution of HLA alleles, only four remained significant following Bonferroni correction (see supplementary Figure 3A). Interestingly, *HLA-A*03:01* was among these four, characterized by strong direct correlation with latitude, with decreasing frequency from north to south.

Regarding the remaining class I loci, only one of the top two frequent alleles was classified among the 12 and 20 significant differentially distributed alleles for *HLA-C* (Figure 2B) and *HLA-B* (Figure 2C), respectively. Specifically, the predominant allele *C*04:01* was listed as rank 1 in order of significance, as it was more frequent in Crete following comparison not only with the pooled cohort ($P = 2.011 \times 10^{-32}$) (see supplementary Table 6A) but also with each of the 12 populations (see supplementary Table 6B), showing a strong inverse correlation to decreased latitude (see supplementary Figure 3B). The reason the top *HLA-B* frequent allele, *B*35:01*, was listed lower in terms of significance (i.e., rank 9) compared with the pooled DKMS registry ($P = 1.450 \times 10^{-10}$) (see supplementary Table 6A) is probably attributable to the non-significant distribution of the allele between Cretans and Greek, Italian and Turkish populations. By contrast, the strongest associations reported for *B*08:01* ($P = 6.503 \times 10^{-33}$) and *B*07:02* ($P = 8.072 \times 10^{-30}$) alleles were attributed to their low frequencies in Crete (see supplementary Table 6B), with both showing a direct correlation with latitude, with decreasing frequencies from north to south (see supplementary Figure 3C). Both *HLA-C* and *HLA-B* were characterized by the strongest latitudinal cline among the class I and II loci studied, with 50% and 60% of differentially distributed alleles, respectively, showing either inverse or direct correlation with latitude.

The most frequent member within the *DRB1*11* family of alleles was the one recorded with the strongest association between Cretans and either the pooled cohort (*DRB1*11:04*, $P = 1.232 \times 10^{-53}$) or most of the individual populations (see supplementary Table 6B)—but not the Greeks—of DKMS. Moreover, the authors' results corroborate previous studies [29] regarding the absence of specific members of the *DRB1*03*, *DRB1*04*, *DRB1*11* and *DRB1*13–*16* groups of alleles in Cretans. However, subsequent comparisons with the DKMS registry

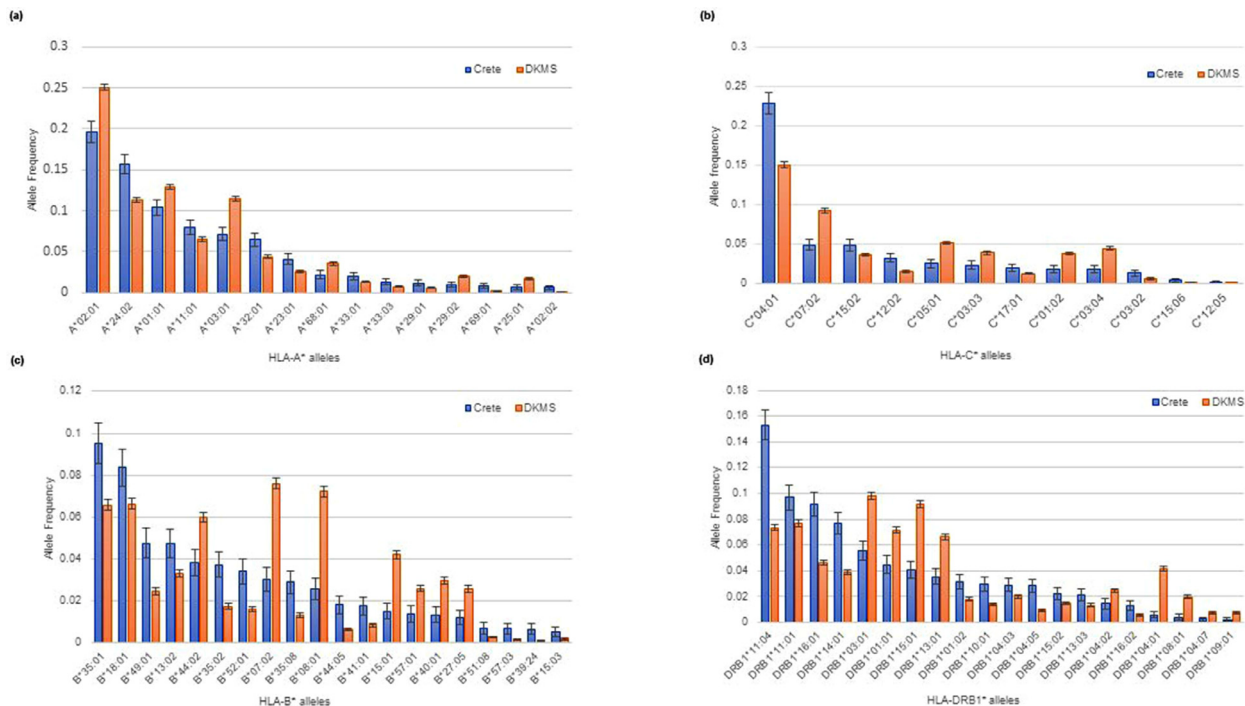


Fig. 2. Frequencies of the statistically significant *HLA-A*, *HLA-C*, *HLA-B* and *HLA-DRB1* alleles between Crete and the 12 populations of DKMS. (A) The 15 *HLA-A* alleles with statistically significant probability after Bonferroni correction for 36 multiple comparisons at $\alpha < 0.05$. (B) The 12 *HLA-C* alleles with statistically significant probability after Bonferroni correction for 28 multiple comparisons at $\alpha < 0.05$. (C) The 20 *HLA-B* alleles with statistically significant probability after Bonferroni correction for 64 multiple comparisons at $\alpha < 0.05$. (D) The 20 *HLA-DRB1* alleles with statistically significant probability after Bonferroni correction for 34 multiple comparisons at $\alpha < 0.05$. All comparisons were performed using the two-tailed Fisher's exact test followed by Bonferroni correction according to the number of alleles analyzed. Error bars represent the 95% confidence intervals that were used to describe the strength of association between the data.

cannot substantiate the specificity of these alleles in Greeks but not in Cretans, as has been previously suggested, since these specific allele members were also absent in the Greek—as well as in most of the Balkan, Iberian and Italian—populations of DKMS.

Finally, subsequent comparisons between Crete and Italy or Turkey using high-resolution second field *HLA-DQB1* data revealed eight and nine significantly different frequencies, respectively, among the 16 common typed alleles (Figure 3). Briefly, the strongest association was observed for *DQB1*05:02*, which was almost twice as common in Crete as in either of the other two populations studied (see supplementary Table 6B).

HLA haplotype frequencies

Maximum likelihood estimation of the four-locus haplotype frequencies resulted in a total of 1248 haplotypes, with *A*33:01-*

*C*08:02-B*14:02-DRB1*01:02* and *A*24:02-C*04:01-B*35:02-DRB1*11:04* being the two most common (Figure 4; also see supplementary Table 7A), as was observed in the 1735 inferred six-locus haplotype data. Subsequent comparative analyses between Crete and each of the 12 populations of DKMS confirmed that these two haplotypes were among the top 10 most common in Mediterranean countries (see supplementary Table 7B). The strongest association between Crete and the pooled DKMS registry, however, is reported for the most common European haplotype, *A*01:01-C*07:01-B*08:01-DRB1*03:01* ($P = 4.452 \times 10^{-21}$), which could be attributed to a strong direct correlation with latitude, with decreasing frequency from north to south (see supplementary Figure 3E).

Moreover, 60% of the 10 most frequent haplotypes were significantly different between Crete and DKMS (see supplementary Table 7A), and 16% (cumulative frequency) of the 50 most common Cretan

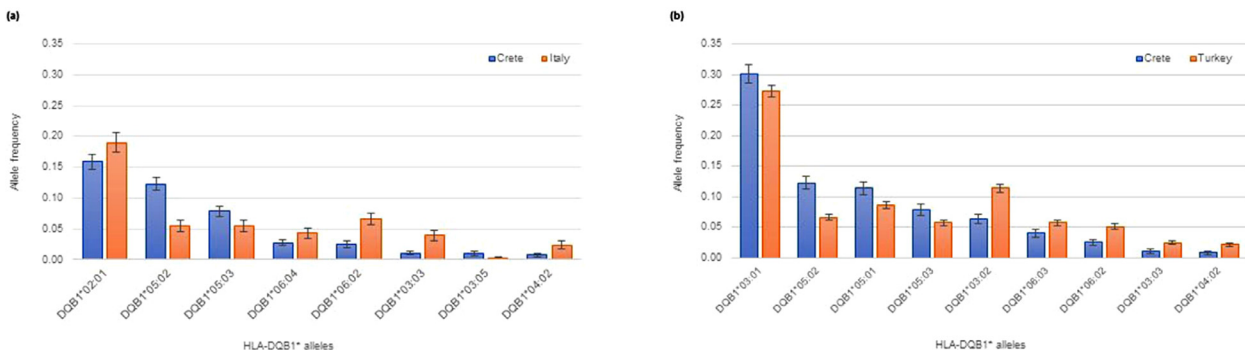


Fig. 3. Frequencies of the statistically significant *HLA-DQB1* alleles between Crete and DKMS populations for Italy or Turkey. The 16 multiple comparisons at $\alpha < 0.05$ between Crete and Italy (A) or between Crete and Turkey (B) revealed eight and nine *HLA-DQB1* alleles, respectively, with statistically significant probability after Bonferroni correction. All comparisons were performed using the two-tailed Fisher's exact test followed by Bonferroni correction according to the number of alleles analyzed. Error bars represent the 95% confidence intervals that were used to describe the strength of association between the data.

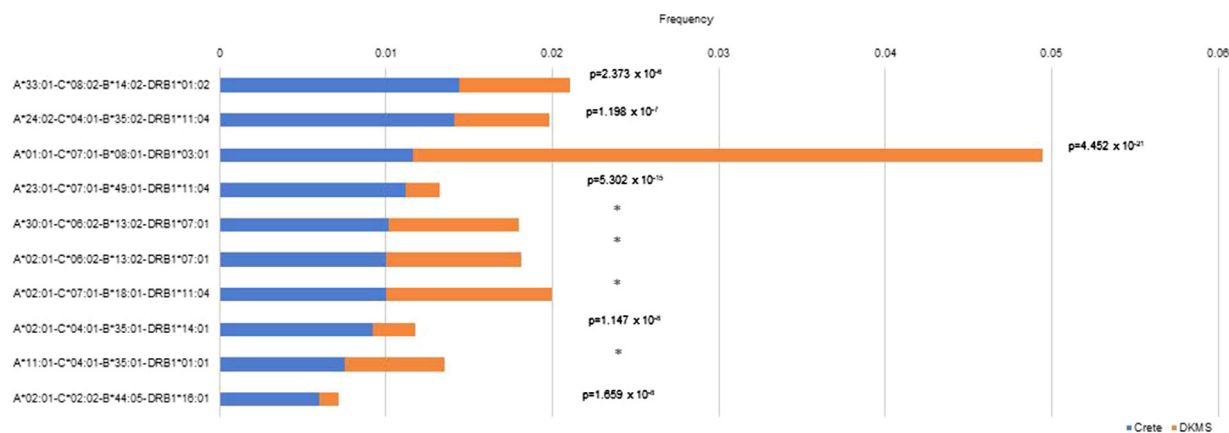


Fig. 4. Frequencies of the 10 most common HLA haplotypes in Crete compared with their counterparts from the 12 populations encompassing the combined DKMS cohort. All comparisons were performed using the two-tailed Fisher's exact test with $\alpha < 0.05$, and significantly different values are shown. Asterisks denote non-significantly different frequencies of the corresponding haplotypes between Crete and the pooled DKMS sample.

haplotypes (e.g., A*01:01-C*06:02-B*37:01-DRB1*04:04, A*26:01-C*07:01-B*51:01-DRB1*16:01, A*23:01-C*07:01-B*49:01-DRB1*16:01, A*11:01-C*04:01-B*51:01-DRB1*11:04) were either distinctive to Crete or predominantly found in countries of the Mediterranean basin, such as Spain, Italy and Turkey (see supplementary Table 7B).

Among the 1248 estimated Cretan haplotypes, 531 (i.e., 42.55%) were different compared with the pooled DKMS cohort (Table 2, Figure 5A), with 66.67% (i.e., 354) being observed only once. Similarly, the difference in the distribution of haplotypes that were observed only once among the 717 common haplotypes was also highly significant ($P < 0.00001$), being 37.52% (i.e., 269) and 16.32% (i.e., 117) for Crete and DKMS, respectively. It is interesting that the extent of variance in the distribution of common haplotypes (range of common haplotypes to Crete: 13.8–21.0%) (Table 2) was pronounced in countries of close genetic distance to Crete according to Prevosti's metric D1 calculations (Figure 5B,C), with the highest variance (13.8% common haplotypes) being attributed to the closest genetically Turkey (Prevosti's $D_1 = 0.675$). Even though 16% and 18.3% of the estimated haplotypes were common between the Cretan and more distant British (Prevosti's $D_1 = 0.838$) and Dutch (Prevosti's $D_1 = 0.824$) cohorts, they were characterized by the highest variation since approximately 20.9% and 21.9% of these, respectively, were differentially distributed compared with Crete (Table 2). Scaled genetic distances visualized by PCA (Figure 6) using high-resolution HLA-A, HLA-B, HLA-C and HLA-DRB1 allele frequencies

revealed that, in agreement with previous results [11], all DKMS populations scale in accordance with their geographic position. Basically, populations in the southeastern Mediterranean (i.e., Greece, Italy and Turkey) are clearly separated from populations in Northern and Western Europe (i.e., UK, Austria, the Netherlands and France) by the principal component with the highest significance. Moreover, the Mediterranean countries seem to scale into two clusters: the western Mediterranean (or Iberian), with Portugal being close to Spain, and the eastern Mediterranean, with Greece and Italy being close to Turkey. In accordance with the lowest genetic distance (Prevosti's $D_1 = 0.675$), Crete scales next to the eastern Mediterranean cluster, close to Turkey.

The current results confirm previous studies [29] showing that Cretans display genetic variations compared with other Greeks (Figure 5B). This suggestion is substantiated by the observation that although the European A*01:01-C*07:01-B*08:01-DRB1*03:01 was the most frequent haplotype in the Greek sample of DKMS, the top Cretan haplotype, A*33:01-C*08:02-B*14:02-DRB1*01:02, together with another four of the 10 most common Cretan haplotypes, was low ranking when Crete was compared with the Greek DKMS population. This could be associated with the insular nature of Crete or the composition of the Greek minority population of DKMS. Specifically, the closer genetic distance of Greece to Turkey (GD = 0.656) or Croatia (GD = 0.656) and Romania (GD = 0.686) combined with the higher frequency of common

Table 2

Percent distribution of the common and different HLA haplotypes between the 1835 Cretan samples and either the pooled DKMS cohort or each of the 12 DKMS populations studied.

Population (n)	HLA haplotypes	Unique HLA haplotypes (%)	Common HLA haplotypes (%) ^a	$\alpha < 0.05$ (%)
CRE (1835)	1248	531 (42.5) ^b	717 (57.5) ^b	252 (35.1) ^b
AUS (1698)	1304	500 (38.3)	222 (17.0)	37 (16.7)
BOS (1028)	797	222 (27.9)	184 (23.1)	33 (17.9)
CRO (2057)	1341	450 (33.6)	291 (21.7)	49 (16.8)
FRA (1406)	1207	509 (42.2)	209 (17.3)	30 (14.4)
GRE (1894)	1546	621 (40.2)	324 (21.0)	30 (9.3)
ITA (1159)	1064	442 (41.5)	217 (20.4)	22 (10.1)
NET (1374)	998	347 (34.8)	183 (18.3)	40 (21.9)
POR (1176)	1049	428 (40.8)	196 (18.7)	30 (15.3)
ROM (1234)	1063	355 (33.4)	242 (22.8)	35 (14.5)
SPA (1107)	1015	453 (44.6)	177 (17.4)	30 (16.9)
TUR (4856)	3255	1848 (56.8)	449 (13.8)	83 (18.5)
UK (1043)	806	290 (36.0)	129 (16.0)	27 (20.9)

AUS, Austria; BOS, Bosnia; CRE, Crete; CRO, Croatia; GRE, Greece; FRA, France; ITA, Italy; NET, the Netherlands; POR, Portugal; ROM, Romania; SPA, Spain; TUR, Turkey.

^a Statistically significant HLA haplotypes were estimated using Fisher's exact test ($\alpha < 0.05$).

^b Comparison with the pooled DKMS registry composed of 20 032 individuals.

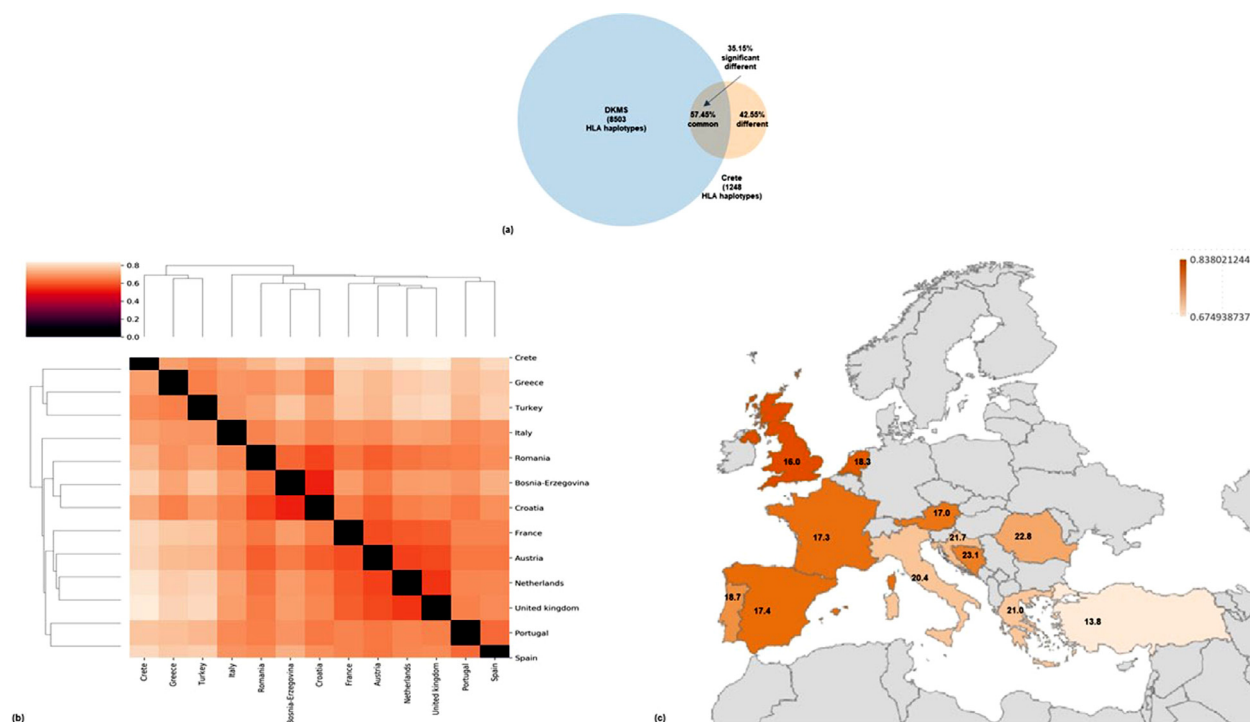


Fig. 5. (A) Distribution (%) of the common and different HLA haplotypes between the 1835 Cretan samples and the combined DKMS cohort of 20 032 individuals from the 12 populations studied. The statistically significantly different HLA haplotypes among the “common haplotypes” group were estimated using Fisher’s exact test ($\alpha < 0.05$). (B) Heatmap and hierarchical clustering of genetic distances in HLA among Crete and the 12 populations of DKMS. The genetic distance clustering was based on the calculated Prevosti’s metric using HLA haplotypic frequencies between Crete and each population separately. The intensity of the color illustrates the genetic relationship from higher (darker orange) to lower (lighter orange). (C) Geographic distribution (%) of common HLA haplotypes between Crete and the 12 populations of DKMS. Analyses were based on the comparison of 1248 Cretan HLA haplotypes with their common counterparts from each DKMS population and visualized in accordance with the genetic distance in HLA between Crete and each population studied using Prevosti’s metric calculations. The intensity of the color illustrates the genetic distance in HLA from lower (lighter orange) to higher (darker orange).

haplotypes between these cohorts (see supplementary Table 7B) suggests that the Greek DKMS population most likely consists of individuals from the northern rather than southern parts of Greece.

Discussion

The present study provides an in-depth characterization of the HLA immunogenetic profile in Crete, the largest and most populous

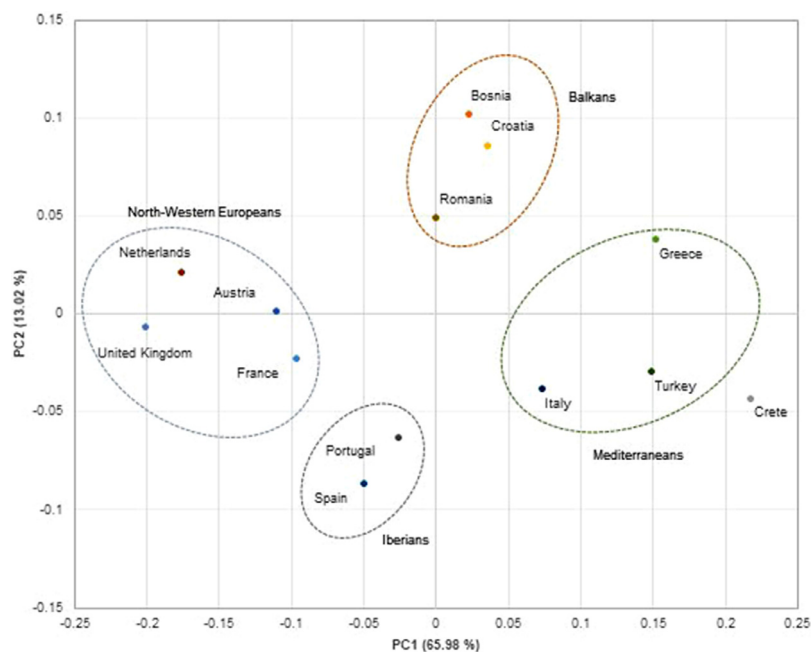


Fig. 6. Scatter plot of the first two principal components (PC1 and PC2) of the dataset. Multidimensional scaling was based on the frequencies of *HLA-A*, *HLA-B*, *HLA-C* and *HLA-DRB1* alleles obtained for Crete and each of the 12 populations of DKMS. The number in parentheses denotes the variance explained by the principal components (axes) following projection of the alleles by linear combination. Mediterraneans are clearly separated from Northern and Western Europeans by the principal component with the highest significance.

Greek island and the fifth largest island in the Mediterranean basin, with an intent to optimize the UCB collection strategy of the regional PCBB. The island of Crete was chosen as an elegant example to unravel the HLA genetic diversity in populations with specific geographic and historical characteristics and to highlight the important role that a regional PCBB may play in fulfilling the HSCT needs of such populations.

Without doubt, the introduction of NGS [4] to the standard HLA typing methodology of population samples from different geographic locations [30–35] has not only unraveled the complexity of this genetic region [22,36,37] but has also revealed the importance of demographic history in HLA heterogeneity and UCB biobanking [5,6]. Under this prism, mapping the regional HLA haplotype spectrum may improve donor availability, particularly in countries of relatively small population size and significant ethnic and genetic diversity.

By applying a high-resolution second field NGS methodology to HLA class I (*HLA-A*, *HLA-B*, *HLA-C*) and class II (*HLA-DRB1*, *HLA-DQB1*, *HLA-DPB1*) loci, the authors identified a broad HLA diversity in Crete. The variation was evident in the distinctive alleles (see supplementary Table 2) and haplotypes (see supplementary Table 3), with 63.6% of the 1735 inferred six-locus HLA haplotypes being estimated only once and the top 10 haplotypes being observed at very low frequencies (range, 0.43–0.98%). Such a diversity was further corroborated by the identification of eight novel HLA alleles [13–17] and the differential distribution of HLA allele and haplotype frequencies that was evident after comparing the Cretan cohort with the 12 minority populations of the German DKMS registry. For instance, a substantial proportion of HLA alleles, ranging between 31.25% (*HLA-B*) and 58.82% (*HLA-DRB1*), were significantly different when Crete was compared with the pooled DKMS cohort (Figure 2A–D).

The allele frequency distribution for most HLA loci in the Cretan population was similar to regions of close geographic proximity to Crete (e.g., Italy, Turkey) and varied following a geographic pattern of differentiation (see supplementary Figure 3A–D) from northern (e.g., UK, the Netherlands) to southern regions, corroborating previous studies [40,41]. Moreover, the close correlation between genetic and geographic distances, illustrated by PCA associated with a southeast to northwest cline (Figure 6), may reflect the possibility of gene flow among neighboring populations. The authors' results support previous studies [41,42] showing that effective European maps may naturally arise as “an efficient two-dimensional summary of genetic variation in Europeans” reflecting either migration or “isolation by distance” events [42,43]. It is intriguing, however, that even though genetic distances visualized by PCA confirmed that all DKMS populations scale in accordance with their geographic position [11,40,43], the extent of variance in the distribution of common haplotypes (Figure 5B,C) was pronounced in neighboring countries of close genetic distance to Crete (e.g., Turkey, Italy).

A further north-to-south gradient in the distribution of HLA alleles and haplotypes was suggested when Cretans were compared with the Greek population of DKMS. The observed differences in distribution of the top common allele and haplotype frequencies can be attributed to the insular nature of Crete, which might have resulted in barriers to gene flow and local selective pressure, as previously described for Mediterranean islands [31]. Such a north-to-south gradient phenomenon has also been suggested for other countries, such as Italy, which displays significant genetic differences between the northern, central and southern regions. The possibility that members of the Greek minority of DKMS do not necessarily represent a random sample of the original population is also highly plausible, as previously reported [11]. Indeed, the fact that the authors' results suggest a closer genetic distance of Greece to Turkey and Balkan countries (i.e., Romania and Croatia) increases the possibility that the Greek population of DKMS consists of immigrants from mostly the central and northern, rather than southern, regions of Greece.

However, patterns of genetic differentiation between distinct populations may also be caused by a series of parameters that include, among others, admixture, selection processes and founder effects [5,6,29,31,36,37]. For instance, the similar distribution of the top common haplotype, *A*33:01-C*08:02-B*14:02-DRB1*01:02*, between Cretans and DKMS populations from regions of the Mediterranean basin (Turkey, Italy, Spain) is supported not only by genetic data from Tunisia [12], South Italy [11], Sardinia [31,32], Spain [11,32,33] and Turkey [11] but also by historical data of possible admixture events that point to the Phoenicians as the original ancestors of this HLA haplotype [29,32]. The increased number of descriptive haplotypes reported for Crete and Turkey, Italy or Spain (Table 2) further supports the possibility of gene flow events between populations across the Mediterranean route. It is interesting that the uniqueness of the distinctive haplotypes *A*01:01-C*07:01-B*57:03-DRB1*03:01*, *A*24:02-C*07:04-B*35:01-DRB1*12:01*, *A*32:01-C*07:01-B*49:01-DRB1*10:01* and *A*69:01-C*02:02-B*51:01-DRB1*11:01* was corroborated not only when compared with the 12 DKMS populations but also following comparison with every cohort reported in the Allele Frequency Net Database [12].

Moreover, selection processes associated with biological processes such as heterozygote advantage, frequency-dependent selection and population-specific immune responses resulting in increased genetic diversity relative to neutral expectations, as previously suggested [36,37], could have affected the level of HLA allele frequency differentiation between the populations studied. Perhaps the strongest balancing selection observed for *HLA-DRB1* (see supplementary Table 4) could justify the strongest associations detected for *DRB1* (with regard to the remaining HLA loci) when the distribution of common allele frequencies was compared between the Cretan population and the pooled DKMS cohort (Figure 2D; also see supplementary Table 6A).

Conclusions

The authors' comprehensive NGS-based high-resolution comparative analyses of the HLA immunogenetic profiles between Cretans and the 12 minority populations of DKMS show for the first time that although the ancestral and common European haplotypes are present at low frequency, unique alleles exist, and geographically distinctive haplotypes prevail in the Cretan population. The authors' data substantiate previously published predictors of HLA genetic variation in Europe [43]; they also highlight the importance of PCBBs for efficient donor recruitment, especially in regions with distinctive geographic characteristics and complex historical events. Strategies in regional PCBBs for collection of CBUs that take into consideration the HLA allele and haplotype frequencies associated with specific inter-population characteristics (e.g., genetic and geographic distances, gene flow) and intra-population natural selection or evolutionary processes (e.g., genetic drift, balancing and positive selection) will increase the efficiency of donor search. This is particularly important for countries like Greece, where only 22% of patients in need of HSCT and without a matched sibling donor manage to find an appropriate MUD or matched CBU from the National Registry (personal communication with Hellenic Transplant Organization). Furthermore, building representative population reference datasets will significantly improve variant interpretation globally while at the same time accelerating the discovery of HLA alleles associated with rare phenotypes, particularly in populations characterized by unique geographic, historical and/or cultural properties. The accurate knowledge of HLA allelic and haplotypic distribution among populations of different ethnicity or among different population clusters of the same ethnicity but diverse genetic background may also elucidate the impact of environmental–genetic interactions in the context of disease pathogenicity and adaptation.

Declaration of Competing Interest

The authors have no commercial, proprietary or financial interest in the products or companies described in this article.

Funding

This work was funded by the Public Investment Fund of the Region of Crete (decision no. 135857/18–06–2018).

Author Contributions

Conception and design of the study: HAP, IM, HL, AK, PP, AEG and ES. Acquisition of data: HAP, IM, ES, AG, AB, IG, IF, MZ and HL. Analysis and interpretation of data: ES, HL, AK, PP, AEG and HAP. Drafting or revising the manuscript: HL, HAP, IM, PP, ES, AK and AEG. All authors have approved the final article.

Acknowledgments

The authors thank the heads, staff and doctors of the Health Region of Crete for their contribution to collecting the peripheral blood samples from the population of Crete.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:[10.1016/j.jcyt.2021.07.010](https://doi.org/10.1016/j.jcyt.2021.07.010).

References

- [1] Carreras E, Dufour C, Mohty M, Kröger N. The EBMT Handbook: Hematopoietic stem cell transplantation and cellular therapies. 7th ed. Cham (CH): Springer; 2013.
- [2] Eapen M, Wang T, Veys PA, Boelens JJ, St Martin A, Spellman S, et al. Allele-level HLA matching for umbilical cord blood transplantation for non-malignant diseases in children: a retrospective analysis. *Lancet Haematol* 2017;4:e325–33.
- [3] Shaw BE, Mayor NP, Szydlo RM, Bultitude WP, Anthias C, Kirkland K, et al. Recipient/donor HLA and CMV matching in recipients of T-cell-depleted unrelated donor haematopoietic cell transplants. *Bone Marrow Transplant* 2017;52:717–25.
- [4] Cereb N, Kim HR, Ryu J, Yang SY. Advances in DNA sequencing technologies for high resolution HLA typing. *Hum Immunol* 2015;76:923–7.
- [5] Tiercy J-M. How to select the best available related or unrelated donor of hematopoietic stem cells? *Haematologica* 2016;101:680–7.
- [6] Gragert L, Eapen M, Williams E, Freeman J, Spellman S, Baitty R, et al. HLA match likelihoods for hematopoietic stem-cell grafts in the US registry. *N Engl J Med* 2014;371:339–48.
- [7] Dehn J, Setterholm M, Buck K, Kempenich J, Beduhn B, Gragert L, et al. HapLogic: A Predictive Human Leukocyte Antigen-Matching Algorithm to Enhance Rapid Identification of the Optimal Unrelated Hematopoietic Stem Cell Sources for Transplantation. *Biol Blood Marrow Transplant* 2016;22:2038–46.
- [8] Dubois V, Detrait M, Sobh M, Morisset S, Labussière H, Giannoli C, et al. Using Easy-Match® to anticipate the identification of an HLA identical unrelated donor: A validated efficient time and cost saving method. *Hum Immunol* 2016;77:1008–15.
- [9] Jöris MM, Lankester AC, von dem Borne PA, Kuball J, Bierings M, Cornelissen JJ, et al. The impact of frequent HLA haplotypes in high linkage disequilibrium on donor search and clinical outcome after unrelated haematopoietic SCT. *Bone Marrow Transplant* 2013;48:483–90.
- [10] Petersdorf EW, Malkki M, Horowitz MM, Spellman SR, Haagenson MD, Wang T. Mapping MHC haplotype effects in unrelated donor hematopoietic cell transplantation. *Blood* 2013;121:1896–905.
- [11] Pingel J, Solloch UV, Hofmann JA, Lange V, Ehninger G, Schmidt AH. High-resolution HLA haplotype frequencies of stem cell donors in Germany with foreign parentage: how can they be used to improve unrelated donor searches? *Hum Immunol* 2013;74:330–40.
- [12] Gonzalez-Galarza FF, McCabe A, Santos EJMD, Jones J, Takeshita L, Ortega-Rivera ND, et al. Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. *Nucleic Acids Res* 2020;48:D783–8.
- [13] Fragiadakis I, Latsoudis H, Zamanakou M, Kanterakis A, Papadaki HA. Two novel HLA-A alleles, HLA-A *03:339 and -A*24:17:01:02, detected in inhabitants from the island of Crete. *HLA* 2021;97(4):353–6.
- [14] Gontika I, Latsoudis H, Zamanakou M, Stylianakis E, Papadaki HA. Detection of the novel HLA-B *51:232:02 variant in an inhabitant from the island of Crete. *HLA* 2021;97(4):364–6.
- [15] Batsali A, Latsoudis H, Vatsiou S, Stylianakis E, Papadaki HA. Two novel HLA-C alleles, HLA-C *15:228 and -C*04:434, detected in inhabitants from the island of Crete. *HLA* 2021;97(3):243–5.
- [16] Mavroudi I, Latsoudis H, Zamanakou M, Kanterakis A, Papadaki HA. Two novel HLA-DRB1 alleles detected in inhabitants from the island of Crete. *HLA* 2021;97(2):163–6.
- [17] Georgopoulou A, Latsoudis H, Vatsiou S, Stylianakis E, Papadaki HA. Detection of the novel HLA-DQB1 *03:439 variant in an inhabitant from the island of Crete. *HLA* 2021;97(1):88–90.
- [18] Robinson J, Barker DJ, Georgiou X, Cooper MA, Flicek P, Marsh SGE. IPD-IMGT/HLA Database. *Nucleic Acids Res* 2020;48:D948–55.
- [19] Schäfer C, Schmidt AH, Sauter J. Hapl-o-Mat: open-source software for HLA haplotype frequency estimation from ambiguous and heterogeneous data. *BMC Bioinformatics* 2017;18:284.
- [20] 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;10:564–7.
- [21] Lancaster AK, Single RM, Solberg OD, Nelson MP, Thomson G. PyPop update—a software pipeline for large-scale multilocus population genomics. *Tissue Antigens* 2007;1(69 Suppl):192–7.
- [22] Hedrick PW, Thomson G. Evidence for balancing selection at HLA. *Genetics* 1983;104:449–56.
- [23] Lewontin RC. The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. *Genetics* 1964;49:49–67.
- [24] Thomson G, Single RM. Conditional asymmetric linkage disequilibrium (ALD): extending the diallelic r2 measure. *Genetics* 2014;198:321–31.
- [25] Hedrick PW. Gametic disequilibrium measures: proceed with caution. *Genetics* 1987;117:331–41.
- [26] Eberhard HP, Madbouly AS, Gourraud PA, Balère ML, Feldmann U, Gragert L, et al. Comparative validation of computer programs for haplotype frequency estimation from donor registry data. *Tissue Antigens* 2013;82(2):93–105.
- [27] Waskom ML. Seaborn: statistical data visualization. *Journal of Open Source Software* 2021;6(60):3021.
- [28] Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. *Philos Trans A Math Phys Eng Sci* 2016;374:20150202.
- [29] Arnaiz-Villena A, Iliakis P, Gonzalez-Hevilla M, Longas J, Gomez-Casado E, Sfyradaki K, et al. The origin of Cretan populations as determined by characterization of HLA alleles. *Tissue Antigens* 1999;53(3):213–26.
- [30] Tokić S, Žizkova V, Štefanić M, Glavaš-Obrovac L, Marczai S, Samardžija M, et al. HLA-A, -B, -C, -DRB1, -DQA1, and -DQB1 allele and haplotype frequencies defined by next generation sequencing in a population of East Croatia blood donors. *Scientific Reports* 2020;10(1):5513.
- [31] 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;94:285–95.
- [32] Martinez-Laso J, Ramirez-Puga A, Rivas-García E, Fernández-Tagarro E, Auyanet-Saavedra I, Guerra-Rodríguez R, et al. North African-Mediterranean HLA genetic contribution in a population of the kidney transplant waiting list patients of Canary origin (Gran Canaria). *HLA* 2018;92(1):12–23.
- [33] Enrich E, Campos E, Martorell L, Herrero MJ, Vidal F, Querol S, et al. HLA-A, -B, -C, -DRB1, and -DQB1 allele and haplotype frequencies: an analysis of umbilical cord blood units at the Barcelona Cord Blood Bank. *Hladnikia* 2019;94:347–59.
- [34] Montero-Martín G, Mallempati KC, Gangavarapu S, Sánchez-Gordo F, Herrero-Mata MJ, Balas A, et al. High-resolution characterization of allelic and haplotypic HLA frequency distribution in a Spanish population using high-throughput next-generation sequencing. *Hum Immunol* 2019;80:429–36.
- [35] Sanchez-Mazas A, Djoulah S, Busson M, Le Monnier de Gouville I, Poirier JC, Dehay C, et al. A linkage disequilibrium map of the MHC region based on the analysis of 14 loci haplotypes in 50 French families. *Eur J Hum Genet* 2000;8:33–41.
- [36] Brandt DYC, César J, Goudet J, Meyer D. The Effect of Balancing Selection on Population Differentiation: A Study with HLA. *Genes*. G3. 2018;8:2805–15.
- [37] Alcaide M. On the relative roles of selection and genetic drift in shaping MHC variation. *Molecular Ecology* 2010;19(18):3842–4.
- [38] Solberg OD, Mack SJ, Lancaster AK, Single RM, Tsai Y, Sanchez-Mazas A, et al. Balancing selection and heterogeneity across the classical human leukocyte antigen loci: a meta-analytic review of 497 population studies. *Hum Immunol* 2008;69:443–64.
- [39] Hurley CK, Kempenich J, Wadsworth K, Sauter J, Hofmann JA, Schefzyk D, et al. Common, intermediate and well-documented HLA alleles in world populations: CIWD version 3.0.0. *HLA* 2020;95(6):516–31.
- [40] Sanchez-Mazas A, Buhler S, Nunes JM. A New HLA Map of Europe: Regional Genetic Variation and Its Implication for Peopling History, Disease-Association Studies and Tissue Transplantation. *Human Heredity* 2013;76:162–77.
- [41] Reich D, Price AL, Patterson N. Principal component analysis of genetic data. *Nature genetics* 2008;40:491–2.
- [42] Novembre J, Stephens M. Interpreting principal component analyses of spatial population genetic variation. *Nat Genet* 2008;40:646–9.
- [43] Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, et al. Genes mirror geography within Europe. *Nature* 2008;456:98–101.