HLA Haplotype Frequencies and Genetic Profiles of the Kuwaiti Population

Reem Ameen a  Salem H. Al Shemmari b  Steven G.E. Marsh c

a Medical Laboratory Department, Health Sciences Center, Kuwait University, Kuwait City, Kuwait; b Medicine Department, Health Sciences Center, Kuwait University, Kuwait City, Kuwait; c Anthony Nolan Research Institute and UCL Cancer Institute, Royal Free Campus, London, UK

Significance of the Study
- We describe here the HLA haplotype diversity in a Kuwaiti Arab population. It is the first attempt to describe the HLA genetic profile of Kuwaiti natives and investigate the Kuwaiti genetic profile by means of high-resolution analysis at the 5-locus level of A∼C∼B∼DRB1∼DQB1.

Keywords
Human leukocyte antigen · Kuwait · HLA haplotype frequency · Genetic profile

Abstract
Objective: The aim of this study was to assess the HLA haplotype frequencies and genetic profiles of the Kuwaiti population. Materials and Methods: Whole venous blood was obtained from 595 healthy, unrelated Kuwaiti volunteers. The study population was genotyped for HLA class I (HLA-A, HLA-B, and HLA-C) and class II (HLA-DRB1 and HLA-DQB1) loci using sequence-specific oligonucleotide (SSO) probe-based hybridization and high-resolution HLA genotyping. Haplotype frequencies were estimated using an implementation of the expectation maximization algorithm that resolves both phase and allelic ambiguity. The Kuwaiti population was compared with other populations from the US National Marrow Donor Program (NMDP), by running a principal component analysis (PCA) on the relevant haplotype frequencies. Results: The most common HLA class I alleles in Kuwait were HLA-A*02:01g, HLA-C*06:02g, and HLA-B*50:01g with frequencies of 16, 14, and 12%, respectively. The most common HLA class II alleles in Kuwait were HLA-DQB1*02:01g and HLA-DRB1*07:01 with frequencies of 29.7 and 16.5%, respectively. The most common Kuwaiti haplotype observed was HLA-A*02:01g~HLA-C*06:02g~HLA-B*50:01g~HLA-DRB1*07:01~HLA-DQB1*02:01g at a frequency of 2.3%. The PCA demonstrated close genetic proximity of the Kuwaiti population with Middle Eastern, Southeast Asian, and North African populations in the NMDP. Conclusion: Identifying the haplotype diversity in the Kuwaiti population will contribute to the selection of an HLA-match for HSCT, disease associations, pharmacogenomics, and knowledge of population HLA diversity.

Introduction
Hematopoietic stem-cell transplantation (HSCT) is a curative treatment for many life-threatening diseases, including blood cancers. The extent of human leukocyte antigen (HLA) matching between the donor and the re-
cipient largely influences the outcome of related and unrelated donor HSCT. In addition, the outcome of HSCT is significantly improved for HLA fully matched donor-patient pairs compared with mismatched cases, even if only a single HLA allele is different [1]. In general, an HLA-matched identical sibling is the ideal donor; however, only an average of 30% of patients can benefit from the related donor option [2]. When a matched sibling is not available, the second alternative is an unrelated donor. Due to the extreme diversity of HLA alleles and the resulting number of potential genotypic combinations, the chances of finding a matched donor are low when searching international registries as patients with certain ethnic and racial backgrounds are underrepresented.

The HLA region on human chromosome 6p21 is the most polymorphic region in the human genome [3–5]. In different ethnic groups, HLA polymorphisms and variations in the HLA allele and haplotype distributions are usually reflective of the origin, migration, and degree of admixture in these populations. In addition, linkage disequilibrium (LD) among different alleles and the predominant ancestral haplotypes vary widely across different human populations [6]. Knowledge of population HLA diversity can contribute to the selection of an HLA-match for HSCT, disease associations, and pharmacogenomics.

The Arabian Peninsula, situated at the nexus of Africa, Europe, and Asia, was an early human migration route out of Africa and part of the intercontinental trade route [7–10]. The State of Kuwait is located at the northwestern tip of the Arabian Gulf and is bordered by Iraq and Saudi Arabia. Kuwait was established by settlers from the Najd region of eastern Saudi Arabia in the 17th and 18th centuries [11]. The current population is about 3.5 million and is made up of 1.5 million Kuwaiti Arab nationals as well as expatriates, mainly Arabs and South Asians (Indians, Pakistanis, Filipinos, etc.). Kuwaiti Arab-national males represent 15.7% of the population and females represent 16.3% (a male-to-female ratio of 1:1). Thalassemia major and other hemoglobinopathies are prevalent in the Kuwaiti population, with 80% of the cases resulting from consanguineous marriages. This represents an important cause of morbidity and mortality in Kuwait [12–14].

To our knowledge, HLA diversity in the ethnic Kuwaiti population has not been previously studied. Most previous reports were related to the role of HLA, disease associations, and other populations [15–17]. This is a considerable disadvantage for unrelated HSCT locally and also for predicting the size of a national unrelated stem cell registry. The aim of this study was to assess the HLA haplotype frequencies and evaluate the genetic profiles of the Kuwaiti population.

**Materials and Methods**

**Study Population**

Whole venous blood was obtained from 595 healthy, unrelated Kuwaiti volunteers. The study population was evenly distributed geographically to obtain a representative sample of the 6 governorates of Kuwait. All individuals who participated in this study were Kuwaiti Arab citizens. The QIAamp DNA blood mini kit (Qiagen, Hilden, Germany) was used to extract DNA. Volunteers were selected independently of any patient need.

**HLA Typing**

Two-thirds of the study population (n = 395) were genotyped for HLA class I (HLA-A, HLA-B, and HLA-C) and class II (HLA-DRB1 and HLA-DQB1) loci by a sequence-specific oligonucleotide (SSO) probe-based hybridization technique, using the One Lambda LABtype sequence-specific oligonucleotide kit, which provided intermediate resolution. High-resolution HLA genotyping, up to the third-nomenclature field, was obtained for 200 individuals with DNA sequence-based typing (SBT), by using the Illumina MiSeq and targeting exon 2 and 3 for class I and exon 2 for class II.

**Statistical Analysis**

To minimize the loss of information from the use of allele codes, SSO primary data were directly interpreted into HLA assignments based on the IPD-IMGT/HLA database (v3.15.0) and then formatted as genotype-list strings [18]. Alleles are described with the first 2 fields of HLA allele nomenclature, representing protein level assignment. Alleles with nucleotide sequences that translated to amino acids identical at the antigen recognition site (ARS) were combined. They were indexed by the lowest-numbered allele followed by "g," since these sets of alleles are often not distinguished by current typing systems, and genomic regions are not defined outside the ARS for many of the alleles [3, 19].

Haplotype frequencies were estimated using an implementation of the expectation maximization (EM) algorithm that resolves both phase and allelic ambiguity [20]. It computes the most probable set of haplotypes that explain the unphased genotype input data via a maximum-likelihood estimation. Starting from arbitrary initial haplotype frequencies, it calculates genotype frequencies under the assumption of the Hardy-Weinberg equilibrium (expectation step). After normalizing, these genotype frequencies are used to estimate haplotype frequencies (maximization step). The expectation and maximization steps are repeated until a stop criterion with a predefined value is fulfilled [19].

The Kuwaiti population was compared with other populations from the US National Marrow Donor Program (NMDP) using principal component analysis (PCA) of the relevant haplotype frequencies [21]. The PCA was performed using MATLAB (MATLAB and Statistics Toolbox release 2015b, MathWorks, Inc., Natick, MA, USA). Comparisons were also performed using clustering analysis by means of CLUTO software [22]. This was done by clustering the top 100 haplotypes (rows) in the Kuwaiti population together with NMDP populations (columns) [21, 22] (Fig. 1), separating the total set of generated haplotypes into 9 prespecified clusters based on haplotype similarity across columns. Rows were clustered according to frequency similarities across populations, with bright-red entries corresponding to higher haplotype frequencies and the brightness of the entries decreasing as their value decreased.
Fig. 1. Clusters of the top 100 haplotypes in the Kuwaiti population with some of the NMDP populations. Darker color indicates higher haplotype frequency.
Results

HLA Allele Frequencies: Class I and Class II

We found the most common HLA class I alleles in Kuwait were HLA-A*02:01g, HLA-C*06:02g, and HLA-B*50:01g with frequencies of 16, 14, and 12%, respectively. The most common class II alleles in Kuwait were HLA-DQB1*02:01g and HLA-DRB1*07:01 with frequencies of 29.7 and 16.5%, respectively. The top 10 most common HLA class I and class II alleles are depicted in Tables 1 and 2.

HLA Haplotype Frequencies

The haplotype A*02:01g~C*06:02g~B*50:01g~DRB1*07:01~DQB1*02:01g was the most common in our study population, with a frequency of 2.3%. This haplotype is the 159th and 65th most common haplotype in the US European Caucasian (i.e., individuals of European ancestry) (EURCAU) and Middle Eastern/North African Coast/Mediterranean (MENAFC) populations, respectively [21]. The second most common haplotype was A*26:01g~C*07:02g~B*08:01g~DRB1*03:01~DQB1*02:01g, at a frequency of 1.6%. This is the 1923rd and 66th most common haplotype in the US EURCAU and MENAFC populations, respectively. The third most common Kuwaiti haplotype was A*02:05g~C*06:02g~B*50:01g~DRB1*07:01~DQB1*02:01g, at a frequency of 1.5%. This haplotype is the 34th and 10th most common haplotype in the US EURCAU and MENAFC populations, respectively. The top 10 most common haplotypes are shown in Table 3.

Regular LD and asymmetric LD (ALD) are described. The definition of LD is the nonrandom association of alleles at 2 loci. The LD can assess the degree to which pairs of alleles are likely to be observed on the same haplotype and has important implications in analyzing immunogenetic data for population and disease-association studies [23]. The ALD measures are symmetric and equivalent to the correlation coefficient r when both loci are biallelic. When the numbers of alleles at the 2 loci differ, the ALD measures capture this asymmetry and provide additional detail about the LD structure [24].

Population Genetics Analysis

The result of the PCA showed that the first 7 principal components (PCs) account for all the variance in the Kuwait haplotype frequencies with the first 3 PCs accounting for >95% of the variances (69.4, 20.9, and 5.8% variance for the first 3 PCs) (Fig. 2). The first and second PCs demonstrate the close genetic proximity of the Kuwaiti population with the NMDP Middle Eastern (MIDEAS), Southeast Asian (SCSEA1), and combined MENAFC populations.

Table 1. Top 10 HLA class I allele frequencies in the Kuwaiti population

| HLA-A | Frequency, % | HLA-B | Frequency, % | HLA-C | Frequency, % |
|-------|--------------|-------|--------------|-------|--------------|
| 1     | A*02:01g     | 15.89 | B*50:01g     | 12.02 | C*06:02g     | 13.96 |
| 2     | A*24:02g     | 9.37  | B*51:01g     | 10.49 | C*04:01g     | 13.20 |
| 3     | A*01:01g     | 7.73  | B*08:01g     | 7.23  | C*07:02g     | 12.89 |
| 4     | A*11:01g     | 6.48  | B*52:01g     | 4.20  | C*15:02g     | 9.20  |
| 5     | A*68:01g     | 5.79  | B*41:01      | 4.03  | C*07:01g     | 7.29  |
| 6     | A*02:05g     | 4.78  | B*35:01g     | 3.74  | C*17:01g     | 5.38  |
| 7     | A*26:01g     | 4.44  | B*07:02g     | 3.70  | C*12:03g     | 5.13  |
| 8     | A*03:01g     | 4.28  | B*18:01g     | 3.61  | C*12:02      | 4.45  |
| 9     | A*31:01g     | 4.21  | B*40:06      | 3.61  | C*14:02g     | 3.78  |
| 10    | A*33:03g     | 3.82  | B*58:01g     | 3.19  | C*08:02g     | 3.19  |

Table 2. Top 10 HLA class II allele frequencies in the Kuwaiti population

| HLA-DRB1 | Frequency, % | HLA-DQB1 | Frequency, % |
|----------|--------------|----------|--------------|
| 1        | DRB1*07:01   | 16.4     | DQB1*02:01g  | 29.69 |
| 2        | DRB1*03:01   | 13.24    | DQB1*03:01g  | 12.58 |
| 3        | DRB1*13:01g  | 5.29     | DQB1*05:02g  | 8.62  |
| 4        | DRB1*13:02   | 4.72     | DQB1*05:01g  | 8.57  |
| 5        | DRB1*16:02g  | 4.71     | DQB1*03:02g  | 8.33  |
| 6        | DRB1*15:01   | 4.62     | DQB1*06:02g  | 7.12  |
| 7        | DRB1*11:01g  | 4.04     | DQB1*06:03g  | 5.23  |
| 8        | DRB1*11:04   | 3.85     | DQB1*06:01g  | 4.76  |
| 9        | DRB1*15:02   | 2.99     | DQB1*06:04g  | 3.76  |
| 10       | DRB1*10:01   | 2.94     | DQB1*04:02g  | 1.98  |
The clustering of the top 100 haplotypes (rows) in the Kuwaiti population together with NMDP populations [21, 22] is shown in Figure 1. The total set of generated haplotypes is separated into 8 clusters based on haplotype similarity across columns. Rows are clustered by similarity in frequency across populations. The clustering analysis showed trends similar to the PCA, specifically the clustering of the Kuwaiti population with the NMDP MIDEAS, MENAFC, and SCSEAI populations (cluster 5). Cluster 4 shows some overlap between the Kuwaiti and US populations of European descent, albeit at lower frequencies (indicated by the lighter color in the heat map).

### Discussion

This study describes HLA haplotype diversity in a Kuwaiti Arab population. To our knowledge, this is the first attempt to describe the HLA genetic profile of Kuwaiti natives and investigate Kuwaiti genetic profiles using high-resolution (i.e., the first 2 nomenclature fields) HLA haplotype frequencies at 5 loci (A\sim C\sim B\sim DRB1\sim DQB1). We are aware that there is limited haplotype frequency information available from surrounding Arab populations at the same level of resolution studied here. However, keeping this resolution discrepancy in mind, we ob-

---

**Table 3.** The 10 most common haplotype frequencies in the Kuwaiti population

| Haplotype | Frequency, % |
|-----------|--------------|
| A*02:01g  C*06:02g  B*50:01g  DRB1*07:01  DQB1*02:01g | 2.26 |
| A*26:01g  C*07:02g  B*08:01g  DRB1*03:01  DQB1*02:01g | 1.58 |
| A*02:05g  C*06:02g  B*50:01g  DRB1*07:01  DQB1*02:01g | 1.5 |
| A*33:03g  C*03:02g  B*58:01g  DRB1*03:01  DQB1*02:01g | 1.15 |
| A*24:02g  C*07:02g  B*08:01g  DRB1*03:01  DQB1*02:01g | 1.1 |
| A*33:01  C*08:02g  B*14:02  DRB1*01:02  DQB1*05:01g | 1.09 |
| A*31:01g  C*15:02g  B*51:01g  DRB1*13:01g  DQB1*06:03g | 0.98 |
| A*11:01g  C*15:02g  B*40:06  DRB1*16:02  DQB1*05:02g | 0.92 |
| A*11:01g  C*12:02  B*52:01g  DRB1*15:02  DQB1*06:01g | 0.92 |
| A*23:01g  C*06:02g  B*50:01g  DRB1*07:01  DQB1*02:01g | 0.88 |
served that the most common Kuwaiti haplotypes were closely related to neighboring populations in Saudi Arabia [25, 26], Tunisia [27], and Jordan [28]. In particular, the third most common Kuwaiti haplotype, A*02:05g~C*06:02g~B*50:01g~DRB1*07:01~DQB1*02:01g, is common in NMDP Middle Eastern populations and also one of the most common in eastern Mediterranean populations [25, 29–31]. Hajeer et al. [26] demonstrated, at the 4-locus level, some overlap between the most common Saudi and Kuwaiti haplotypes, i.e., A*24:02:01:01~B*08:01:01~C*07:02:01:01~DRB1*03:01:01:01 and A*23:01:01~B*50:01:01~C*06:02:01:01~DRB1*07:01:01:01. Additionally, we found some overlap with the HLA haplotype frequencies observed in Arab populations in allelefrequencies.net, albeit at a much lower resolution.

Our PCA and clustering analyses showed that the Kuwaiti population has genetic proximity with the NMDP MIDEAS, MENAFC, and SCSEAI populations due to some shared haplotypes: A*11:01g~C*12:02~B*52:01g~DRB1*15:02~DQB1*06:01g, A*01:01g~C*07:01g~B*15:17~DRB1*13:02~DQB1*06:04g, A*24:02g~C*07:02g~B*08:01g~DRB1*03:01~DQB1*02:01g, A*11:01g~C*12:02~B*52:01g~DRB1*15:02~DQB1*06:01g, A*01:01g~C*07:01g~B*15:17~DRB1*13:02~DQB1*06:04g as well as A*02:05g~C*06:02g~B*50:01g~DRB1*07:01~DQB1*02:01g (common throughout the eastern Mediterranean region). While the haplotype overlap with the NMDP Middle Eastern population was expected, the genetic proximity with the SCSEAI population can possibly be explained by a recent genetic flow from the 37% South Asian immigrant population in Kuwait [10, 32]. The proximity to the AAFA (African-American) population is better explained as shared alleles. The Kuwaiti population shares some alleles with the AAFA population (A*02:01g, A*33:03g, A*68:01g, B*08:01g, B*07:02g, C*07:02g, C*06:02g, DRB1*03:01, DRB1*07:01, DRB1*15:01, DQB1*02:01g, and 05:01g). Moreover, one of the top ranked AAFA and Kuwaiti haplotypes is A*30:01g~C*17:01g~B*42:01~DRB1*03:02~DQB1*04:02g.

The results of this study are important for supporting research investigating HLA linkage in diseases with etiologies, susceptibilities, or frequencies that are related to the major histocompatibility complex, and also to facilitate the search for HLA-matched unrelated donors for patients in need of HSCT. Currently, patients of Arab or Kuwaiti descent who require HSCT face challenges in the identification of matched unrelated donors because of the lower representation of donors of Middle Eastern heritage in international bone marrow registries and cord blood banks. Donors of predominantly European descent are unable to fully meet the needs of the Kuwaiti patient population. This study provides information about a relatively underrepresented population and can potentially provide much-needed insight into their genomics.

Conclusion

Identifying the haplotype diversity in the Kuwaiti population will contribute to the selection of HLA-matches for HSCT, disease associations, and pharmacogenomics as well as more knowledge of HLA diversity in the population.

Acknowledgements

We acknowledge Michael Halagan and Abeer Madbouly for their help in reviewing the manuscript and providing insights into the implementation of the EM algorithm.

Statement of Ethics

Only those who signed the informed consent form were recruited. The Ethics Committee of the Health Sciences Center approved the design of the study.

Funding Source

This research was supported by a grant from the Kuwait Foundation for Advancement of Science (KFAS) (grant No: 2012-130-204), Kuwait.

References

1. Copelan EA. Hematopoietic stem-cell transplantation. N Engl J Med. 2006 Apr;354(17):1813–26.
2. Rosenmayr A, Pointner-Prager M, Mitter- schiffthaler A, Bozic L, Pelzmann B, Tüchner H, et al. What are a patient’s current chances of finding a matched unrelated donor? Twenty years’ central search experience in a small country. Bone Marrow Transplant. 2012 Feb;47(2):172–80.
3. Marsh SG, Albert ED, Bodmer WF, Bontrop RE, Dupont B, Erlich HA, et al. Nomenclature for factors of the HLA system, 2010. Tissue Antigens. 2010 Apr;75(4):291–455.
4. Mungall AJ, Palmer SA, Sims SK, Edwards GA, Ashurst JL, Wilming L, et al. The DNA sequence and analysis of human chromosome 6. Nature. 2003 Oct;425(6960):805–11.
5. Robinson J, Halliwell JA, Hayhurst JD, Fliceck P, Parham P, Marsh SG. The IPD and IMGT/HLA database: allele variant databases. Nucleic Acids Res. 2015 Jan;43(Database issue):D423–31.
HLA Haplotype Frequencies in Kuwait

Med Princ Pract 2020;29:39–45
DOI: 10.1159/000499593