Published in final edited form as:

Tissue Antigens. 2009 January ; 73(1): 17–32. doi:10.1111/j.1399-0039.2008.01151.x.

HLA-A, -B, -C, -DRB1 Allele and Haplotype Frequencies Distinguish Eastern European Americans from the General European American Population

Steven J. Mack¹, Bin Tu², Ana Lazaro², Ruyan Yang², Alex K. Lancaster^{3,4}, Kai Cao², Jennifer Ng², and Carolyn Katovich Hurley²

¹Children's Hospital Oakland Research Institute, Oakland, CA USA

²CW Bill Young Marrow Donor Recruitment and Research Program, Departments of Pediatrics and Oncology, Georgetown University, Washington DC USA

³Department of Integrative Biology, University of California, Berkeley, CA USA

Abstract

Sequence based typing was used to identify HLA-A,B,C,DRB1 alleles from 558 consecutively recruited U.S. volunteers with Eastern European ancestry for an unrelated hematopoietic stem cell registry. Four of the 31 HLA-A alleles, 29 -C alleles, 59 -B alleles, and 42 -DRB1 alleles identified (A*0325, B*440204, Cw*0332, and *0732N) are novel. The HLA-A*02010101g allele was observed at a frequency of 0.28. Two-, three- and four-locus haplotypes were estimated using the expectation maximization algorithm. The highest-frequency extended haplotypes (A*010101g-Cw*070101g-B*0801g-DRB1*0301 and A*03010101g-Cw*0702-B*0702-DRB1*1501) were observed at frequencies of 0.04 and 0.03, respectively. Linkage disequilibrium values (D'_{ij}) of the constituent 2-locus haplotypes were highly significant for both extended haplotypes (p-values were less than 8×10^{-10}), but were consistently higher for the more frequent haplotype. Balancing selection was inferred to be acting on all four loci, with the strongest evidence of balancing selection observed for the HLA-C locus. Comparisons of the A-C-B haplotype and DRB1 frequencies in this population to those for African, European and western Asian populations revealed high degrees of identity with Czech, Polish, and Slovenian populations and significant differences from the general European American population.

Keywords

HLA; population study

Introduction

The United Nations Statistics Division (UNSD) defines Eastern Europe as including Belarus, Bulgaria, the Czech Republic, Hungary, Poland, the Republic of Moldova, Romania, the Russian Federation, Slovakia, and Ukraine. However, common understandings of the region also include Albania, Bosnia and Herzegovina, Croatia, the Republic of Macedonia, Montenegro, Serbia, and Slovenia (these last six nations having once constituted Yugoslavia), although the UNSD defines these nations as belonging to Southern Europe. In the 2000 US census, 61% of Americans (170 million individuals) identified themselves as

Corresponding author: Steven J. Mack, Children's Hospital Oakland Research Institute, 5700 Martin Luther King Jr Way, Oakland, CA USA 94609; Phone: 510-597-7145 Fax: 510-450-7910; sjmack@chori.org.

⁴Current address: Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, USA

being of European ancestry, and of these, 17.4 million Americans identified themselves as being of Czech, Czechoslovakian, Hungarian, Polish, Romanian, Russian, Slovakian, or Ukrainian ancestry, and 1.13 million Americans identified themselves as being of Albanian, Croatian, Serbian, Slovenian, or Yugoslavian ancestry. In total, Americans of Eastern European ancestry constitute 6.6% of the US population, and approximately 11% of the European American population (1).

The HLA loci are the most polymorphic loci in the human genome, with several hundreds of allelic products encoded at multiple loci (e.g., more than 1029 alleles encode over 870 different HLA B molecules). The HLA genes encode cell surface proteins that present foreign and self-derived peptides to T-lymphocyte antigen receptors for recognition. Sequence polymorphisms in the antigen binding domains of the HLA molecules determine the repertoire of peptides that can be presented and in turn influence an individual's immune response (2). HLA alleles and multi-locus haplotypes are differentially distributed in populations around the globe, and studies of allele and haplotype frequency diversity have been used to make inferences about population relationships and history, and the selective forces operating to maintain high levels of HLA diversity.

Here, we present HLA-A, B, C, and DRB1 alleles identified via DNA sequencing in a population of Eastern European Americans, and use these data to estimate multi-locus haplotypes and the degree of linkage disequilibrium between loci, and to examine selective forces influencing this population and the degree of differentiation between this population and other U.S. populations originating from other regions of Europe, as well as European populations.

Materials and Methods

Sample population

The study population included 558 individuals from the United States indicating Eastern European ancestry who were consecutively recruited as volunteer donors for a bone marrow donor registry from January 2003 through February 2005. Because of the recruitment setting, individuals are unlikely to be related and are likely to originate from different areas of the United States. All are self identified as being of Eastern European ancestry. However, because these individuals are part of the larger US population, genetic contributions from non-European populations cannot be excluded simply on the basis of self-identification.

Identification of known HLA alleles

Genomic DNA was prepared using the QIAamp 96 DNA blood kit (Qiagen Valencia, CA). Each individual was initially typed at intermediate resolution for HLA-A,-B,-C,-DRB1 by sequence specific probe based hybridization using the One Lambda LABType® SSO Kit (One Lambda, Canoga Park, California) following manufacturer's protocols. To identify the HLA-A,-B,-C alleles carried by each individual, PCR primers were used to amplify each locus as previously described (3). Applied Biosystems Big Dye terminator chemistry and sequencing primers were used to obtain the sequences of both strands of exons 2 and 3. DRB1 alleles were amplified and sequenced using the AlleleSEQR class II kit (Abbott Molecular Inc, Des Plaines, IL). Allele group specific amplification primers from the kit were used to isolate DRB1 alleles for sequencing. Additional in-house PCR and sequencing primers were added when needed to obtain resolution. Reactions products were identified with Applied Biosystems 3730xl DNA analyzers (PE Applied Biosystems, Foster City, CA) and sequence interpretation used Assign software (Conexio Genomics, Applecross, Western Australia) and was based on ImMunoGeneTics (IMGT)/HLA database release 2.7 (4).

Alleles identical in exons 2 and 3 (class I) or exon 2 (DRB1) were not resolved; expressed alleles in this category share the amino acid sequence of their antigen binding grooves. Unresolved alleles of this type that differ in four digit names ie, encode allelic products that vary in amino acid sequence outside of the antigen binding site, are indicated by the use of a “g” following the name of the lowest numbered allele in the group. For example, A*02010101g includes alleles A*02010101, A*0209, A*0243N, A*0266 as well as synonymous alleles A*02010102L and A*020108. A listing of these unresolved alleles can be found at <http://www.ebi.ac.uk/imgt/hla/ambig.html> (4) under database release 2.7. For those class I samples yielding alternative allele combinations, either allele specific sequencing primers or allele specific PCR amplification were used to link polymorphisms and to identify the specific allele combination (4). [In-house primer sequences used for all loci are available at www.dodmarrow.org.]

Characterization of new HLA alleles

Potentially new alleles were isolated and characterized as previously described (3,5). DNA sequencing of PCR products included primers annealing to both DNA strands for at least two independent PCR reactions with an ABI 3730 Automated DNA Sequencer (Applied Biosystems, Foster City, CA). Allele designations were assigned by the WHO Nomenclature Committee for Factors of the HLA System (6).

Statistical analysis

PyPop (Python for Population genetics, version 0.6.0 <http://www.pypop.org>) was used to carry out Hardy-Weinberg testing, Ewens-Watterson homozygosity analyses, and haplotype and linkage disequilibrium (LD) estimates (7,8). Allele frequencies were obtained by direct counting. Allele frequencies at each HLA locus were evaluated for deviations from Hardy-Weinberg equilibrium proportions using the exact test of Guo and Thompson (9), and by chi-square testing when expected values were ≥ 5 . Chi-square tests were investigated for overall common genotypes (those expected to be seen in at least 5 instances “lumped” genotypes (the set of all genotypes individually expected to be seen in fewer than 5 instances each), all heterozygotes, all homozygotes, as well as for individual common and heterozygote genotypes. These Hardy-Weinberg tests measure the degree to which observed genotype frequencies differ from those expected based on the allele frequencies for that population, assuming that the population is suitably large and experiences random mating (10).

The Ewens-Watterson test of homozygosity was applied to each locus (11) (12), using Slatkin’s Monte-Carlo implementation of the exact test (13,14). In this test, the observed homozygosity (F , the sum of the squares of the allele frequencies) is compared with the mean value of F expected for a population of the same size with the same number of alleles, undergoing neutral evolution. The normalized deviate of F (F_{nd} , the difference between the observed and expected values of F divided by the square root of the variance of the expected F was also calculated for each locus (15).

Two-, three-, and four-locus haplotype frequencies were estimated using the iterative expectation-maximization (EM) algorithm (16,17). Linkage disequilibrium (LD) between alleles at each pair of loci, and two overall (locus-pair-level) measures of linkage disequilibrium, normalized to values between zero and one, were calculated. The normalized allele-pair-level LD measure, D'_{ij} , is the disequilibrium coefficient (D) divided by the upper and lower bounds of D for the particular alleles at each locus (as described in (18,19,20), and ranges from +1 to -1. A D'_{ij} value of zero indicates linkage equilibrium, while a value of +1 indicates the complete association of a given pair of alleles in a single haplotype, and for the data reported here, a value of -1 indicates the complete absence of a haplotype

comprised by those alleles. [Note: The complete absence of a particular haplotype can only be inferred from a D'_{ij} value of -1 when none of the reported alleles has a frequency greater than 0.5.] The first of the locus-pair-level measures, D' (18), uses the products of the allele frequencies at each locus to weight the LD contribution of specific allele pairs; while the second, W_n (21), calculates a normalization of the chi-square statistic for deviations between observed and expected haplotype frequencies. The significance of the overall LD between any two loci was tested using the permutation distribution of the likelihood ratio test (17).

Comparison of populations was limited by the availability of allele-level data for HLA A:C:B haplotypes. Arlequin v3.11 (22) was used to compare the HLA-A:C:B haplotypes and DRB1 genotypes in this population to those for Sub-Saharan African populations from Kenya (10, 23), Mali (23), Rwanda (10), Senegal (10), South Africa (10), Uganda (23), Zambia (23), and Zimbabwe (10); North African populations from Morocco (24,25,26,27) and Algiers (24); European populations from Bulgaria (10, 28), Croatia (29,30), the Czech Republic (10,31), Finland (10), Georgia (10,32), Germany (33), Italy (34), Macedonia (35,36), Northern Ireland (10), Norway (37,38), Poland (39,40), Portugal (41), Russia (42,43), Slovenia (10,44), Spain (45–48), and Sweden (49); Asian populations from India (50), Israel (51), and Turkey (52); two African American populations (53,3), and four European American populations (53–56), by calculating pairwise F_{st} values (and associated p-values) for this entire set of populations. Because the frequencies of the available A:C:B haplotypes for one of the Polish populations (39) summed to 0.78, a "blank" A:C:B haplotype with a frequency of 0.22 was considered for this population, and pairwise F_{st} values for this population were interpreted differently than for the rest (see below). In accounting for small differences in population sample sizes, the F_{st} calculation (57) may result in small negative F_{st} values; these were treated as being equivalent to zero and are reported as such. Pairwise standardized F_{st} values (F'_{st} values) were generated using Hedrick's method of dividing each value by the maximum F_{st} value (58). Because all populations had not been genotyped at the same loci or for the same level of resolution, three comparisons were performed and the analysis focused on the amino acid sequences encoding the polymorphic antigen binding groove. A given pair of population datasets was determined to differ significantly if the appropriate p-value associated was less than 0.05.

Results

Allele and genotype frequencies

Allele frequencies for the HLA-A, C, B and DRB1 loci are shown in Table 1. Five-hundred fifty unique HLA-A, C, B, DRB1 phenotypes were observed among the 558 individuals examined. No overall deviations from expected Hardy-Weinberg equilibrium proportions (HWE) were observed at the HLA-A, -C, or -B loci (p-values 0.4881, 0.2028, and 0.5792, respectively). However a minor but significant deviation from expected HWE was observed at the DRB1 locus (p-value = 0.0187). This deviation primarily resulted from an undercount of DRB1*0701:*1601 genotypes (0 observed, 5.23 expected, p-value = 0.0222). Given the lack of similar deviations at the class I loci, it seems unlikely that the unexpected genotypic ratios at the DRB1 locus stem from sampling error.

Allelic diversity was high at each locus; thirty-one HLA-A alleles, 29 HLA-C alleles, 59 HLA-B alleles, and 42 DRB1 alleles were observed in this population (with an allele defined as either a unique class I exon 2-3 sequence or a unique DRB1 exon 2 sequence). Novel exon 2-3 sequences were detected at each class I locus (see below). Of the detected four digit alleles that had been previously identified, all but two have been identified by Cano et al (59) as being "common or well-documented" (CWD) alleles. The two non-CWD alleles (A*3208 and Cw*0713) were each observed in single individuals.

Four HLA-A alleles (A*02010101g, *010101g, *03010101g, and *24020101g) were observed at frequencies greater than 0.1, and represented 65% of the allelic diversity observed at this locus. Most notably, the A*02010101g allele was observed at a very high frequency of 0.275. Three HLA-C alleles (Cw*070101g, *04010101g, and *0702) were observed with frequencies above 0.1, and represented 41% of the HLA-C diversity in this population. HLA-B*0702 is the only HLA-B allele observed with a frequency greater than 0.1 in this population (allele frequency = 0.104). Three DRB1 alleles (DRB1*0701, *1501, and *0301) with frequencies greater than 0.1 represented 34% of the allelic diversity observed at this locus.

Ewens-Watterson Homozygosity Test

The normalized deviate of the Ewens-Watterson homozygosity statistic (F_{nd}) was calculated based on the observed allele frequencies at each locus, and was used to infer the action of balancing or directional selection at each locus. The Ewens-Watterson model calculates a mean homozygosity value for a population of a given size, in which a given number of alleles has been detected, and that is experiencing neutral evolution. The observed homozygosity value for populations evolving under neutral conditions will be similar to the expected homozygosity value, and the resulting F_{nd} value will be close to 0. F_{nd} values significantly greater than 0 (resulting from frequency distributions that are significantly “skewed” in favor of one or a few high-frequency alleles) are consistent with either the action of directional selection on the locus in question or an extreme demographic effect (e.g., a population bottleneck), while F_{nd} values significantly lower than 0 are consistent with the action of balancing selection, resulting from allele frequency distributions that are significantly more “even”.

The results of the Ewens-Watterson Homozygosity test are shown in Table 2. Negative F_{nd} values were observed for all four loci, and a significantly low value ($F_{nd} = -1.1779$, p-value = 0.0308) was observed for the HLA-C locus. In addition, the application of a sign test to these F_{nd} values indicates an overall significant trend away from the null hypothesis of neutral evolution (p-value = 0.0455), suggesting that the allele frequency distributions at all four loci have been shaped by balancing selection.

New alleles

Three individuals carried novel alleles at the HLA-A, HLA-B, or HLA-C loci (Table 3), a fourth carried a newly reported non-expressed HLA-C allele. Novel allele, A*0325, has already been reported by us (60). It differs from A*03010101 by a single novel nucleotide change at polymorphic codon 36 altering TTC (phe) to TGC (cys). B*440204 differs from B*44020101 at polymorphic codon 138 altering ACC to ACT, a synonymous substitution which has not been previously observed. Cw*0332 differs from Cw*030401 at codon 90, changing GCC (ala) to GAC (asp), a common alternative in HLA-C alleles. Cw*0732N was previously reported in a Spanish individual (Dr. Jose L. Vicario, unpublished, IMGT/HLA database). The allele has an insertion (CGCAGATACCT) between codons 163.2 and 165.1 in exon 3 and a loss of 4 nucleotides previously found in that region. Based on shared alleles of the Spanish cell (CTM8689384) with cell NT00597, the non-expressed allele is most likely associated with A*03010101g, B*0702, DRB1*1501.

Haplotypes

Two-locus (A:B, C:B, and B:DRB1), three-locus (A:B:DRB1) and four-locus haplotypes were estimated for this population. Two-locus haplotypes are presented in Table 4; three- and four-locus haplotypes in Table 5. Because the outcome of the EM algorithm is unreliable for rare haplotypes ($n = 1$ or 2), only those haplotypes with at least three copies

are presented in the tables. In general, the five most frequent haplotypes were also common in an earlier study of European Americans (61).

Ninety six haplotypes were identified at least three times and accounted for almost 80% (h.f. =0.7936) of the A:B haplotypes (Table 4). The five most frequent A:B haplotypes were A*010101g:B*0801g (haplotype frequency [h.f.] = 0.0618), A*03010101g:B*0702 (h.f. = 0.0508), A*02010101g:B*15010101g (h.f. = 0.0337), A*02010101g:B*44020101g (h.f. = 0.0330), and A*02010101g:B*400101g (h.f. = 0.0286). Four of these haplotypes are also the four most frequent haplotypes in European Americans; the last haplotype listed above ranks sixth. Of the 55 C:B haplotypes accounting for 92% (h.f. =0.9189) of the total, the five most frequent haplotypes were identical to a previous study (61): C*0702:B*0702 (h.f. =0.1022), Cw*070101g:B*0801g (h.f. =0.0860), Cw*04010101g:B*350101g (h.f. =0.0573), Cw*0501g:B*44020101 (h.f. =0.0430), and Cw*0304: B*400101g (h.f. =0.04211). Ninety seven B:DRB1 haplotypes were observed in three or more individuals totaling to a haplotype frequency of 73% (h.f. =0.7279). The five most frequent B:DRB1 haplotypes were : B*0801g:DRB1*0301 (h.f. =0.0669), B*0702:DRB1*1501 (h.f. =0.05501), B*4403:DRB1*0701 (h.f. =0.0310), B*350101g:DRB1*0101 (h.f. =0.02833), and B*3502:DRB1*1104 (h.f. =0.0222). The last haplotype was ranked at 19 in the previous study of European Americans but the other four were also the four most frequent in that study.

Among the three locus haplotypes (Table 5), the five most common were: A*010101g:B*0801g:DRB1*0301 (h.f. =0.0454), A*03010101g:B*0702:DRB1*1501 (h.f. =0.0299), A*02010101g:B*1302:DRB1*0701 (h.f. =0.0151), A*02010101g:B*0702:DRB1*1501 (h.f. =0.0144), and A*02010101g:B*15010101g:DRB1*0401 (h.f. =0.0143). This included three of the top ranked haplotypes in the earlier study; the haplotype with DRB1*0701 ranked 19th and the haplotype with DRB1*0401 ranked sixth. Eighty three three locus haplotypes included 51% (h.f. =0.5100) of the total. Of the 65 haplotypes covering 43% of the total (h.f. =0.4309), the four locus haplotypes included three of the top ranked haplotypes in the earlier study: A*010101g:Cw*070101g:B*0801g:DRB1*0301 (h.f. =0.0435), A*03010101g:Cw*0702:B*0702:DRB1*1501 (h.f. =0.0295), and A*02010101g:Cw*0702:B*0702:DRB1*1501 (h.f. =0.0137). Haplotypes A*02010101g:Cw*0602:B*1302:DRB1*0701 (h.f. =0.0151) and A*02010101g:Cw*070101g:B*0801:DRB1*0301 (h.f. =0.0141) ranked 17th and 10th respectively in the earlier study. In summary, for the very common haplotypes, this Eastern European American population is similar to a European American population (61).

Global linkage disequilibrium (LD) estimates of the associations between HLA loci are shown in Table 6. All show significant LD. As already observed (62), C:B has the strongest association ($D' = 0.91160$; $W_n = 0.72619$). These genes lie adjacent to one another in the major histocompatibility complex (MHC) separated by a map distance of approximately 0.1 Mb (63). Separated by approximately 2.7 Mb, the A:DRB1 LD estimates have the lowest values ($D' = 0.41025$; $W_n = 0.31802$). This ranking of LD is the same as was observed observed for populations in the 13th International Histocompatibility Workshop Anthropology/Human Genetic Diversity Project, with LD generally decreasing in proportion to increasing map distance between loci (62). However, higher LD values were observed for A:B ($D' = 0.57324$; $W_n = 0.40388$) than A:C ($D' = 0.47558$; $W_n = 0.033870$). This was also observed in the 13th Workshop Anthropology project (62), and may be result from the low levels of polymorphism seen at HLA-C, relative to the HLA-A and -B loci; there are approximately half as many HLA-C alleles as HLA-A alleles, and one third as many HLA-C alleles as HLA-B alleles.

Admixture

To estimate admixture from non-Eastern European and non-European populations, we compared pairwise F_{st} values for this Eastern European American population sample to those calculated for A:C:B haplotypes from Sub-Saharan African, European, and western Asian populations, and to F_{st} values calculated for DRB1 allele frequencies in Sub-Saharan African, European, North African, and western Asian populations. Preliminary analyses (data not shown) indicated that the Eastern European American population was uniformly closer to European and European American populations (and significantly different from all non-European populations). Therefore, only results of comparisons between the Eastern European American population and other European and European American populations are presented in Table 7.

At the DRB1 locus, the Eastern European American population did not differ significantly from eastern European populations from Poland (1 of 2 studies), the Czech Republic (2 of 2 studies), Slovenia (2 of 2 studies), and Russia (1 of 2 studies). However, the Eastern European American population was also not significantly different from European populations from Spain (1 of 5 studies), Norway (1 of 2 studies), Germany, and Portugal, or from general European American populations (2 of 3 studies).

Considering A:C:B haplotypes, the Eastern European American population was most similar to the Czech population and a general European American population ($F_{st} = 0.011$ and 0.0013 , respectively). The Eastern European American population did not differ significantly from the Czech population ($p\text{-value} = 0.17117 \pm 0.0286$) but differed significantly from the European American population ($p\text{-value} = 0.018 \pm 0.012$). A Polish population (39) with a "blank" A:C:B haplotype frequency of 0.22 was included in this haplotype analysis. The F_{st} values involving this Polish population and other European or European American populations were uniformly higher than those between all other European and European American populations (means = 0.6316 and 0.1015 , respectively) and this Polish population differed significantly from all other populations included in the analysis. However, the lowest pairwise F_{st} value involving this Polish population (0.5614) is observed for the comparison with the Eastern European American population.

Overall, the Eastern European American population was most similar to the Czech and European American populations when comparing A:C:B haplotypes, and was most similar (F_{st} equivalent to zero) to the Czech and Slovenian populations, as well as to German, European American and Spanish populations when considering DRB1 allele frequencies, suggesting that this group represents a distinct sub-sample of the European American population, with a considerable (but not exclusively) eastern European contribution.

Discussion

DNA sequencing was used to identify HLA alleles from a population of individuals from the United States who self identified as having Eastern European ancestry. The 2000 US census suggest that half of the Americans of Eastern European ancestry are at least partially of Polish descent, and a Polish contribution to the Eastern European American population is reflected our analyses, wherein comparison of the Eastern European American population to a variety of world populations demonstrated a strong similarity to Czech, Polish, Russian and Slovenian populations. Pairwise F_{st} values also showed similarities to European American populations, which might be expected since the latter would include individuals with eastern European ancestry. Similarities between the Eastern European American population and other European populations are indicated when considering DRB1 alleles, but not when considering A:C:B haplotypes; this reflects the subdivision of alleles that are common across Europe into multi-locus haplotypes that display regional restriction. Overall,

our analyses suggest that self-identified ethnicity, in spite of the variation in how race and ethnicity is perceived among individuals (64), is still a fairly accurate estimation of ancestry.

First settled by Europeans in the late 1500s, the United States is a nation of immigrants from predominantly Europe but with an increasing contribution from Latin America, Africa and Asia. In the U.S. census of 2000, almost 60% of individuals specified a single ancestry and 30% specified multiple ancestry (1). Understanding this complex and evolving population structure to better define subgroups of individuals who share a common genetic heritage will lead to a better understanding of disease risk and the outcomes of medical intervention (64,65). Studies employing genome-wide genetic markers including microsatellites, Alu markers, and single nucleotide polymorphisms have mapped human genetic diversity in world-wide populations for single markers and haplotypes (66,67). Within U.S. populations, the focus has been on broad population groups: European Americans, African Americans, Hispanic Americans, and Asian Americans. Based on our findings distinguishing Americans with Eastern European heritage from a more general European American population and to a certain extent from populations in other regions of Europe, these genome-wide studies should be extended to better define subpopulations of European Americans.

In the 2000 U.S. census, 7% of individuals reported their ancestry as American. It is likely that this “new” population arising from the admixture of many groups will continue to increase over time. Loss of ethnic identity will pose a significant challenge for registries supporting hematopoietic stem cell transplantation with an unrelated donor. Today more than one third of donors are identified through an international search and reside in a country different from that of the donor (68). Knowledge of the ancestry of an individual is critical in predicting the HLA alleles carried by a potential donor with only low resolution HLA typing results. This information is used by histocompatibility experts and by algorithms to predict the likelihood of HLA allele matching (69). Registries should begin to develop strategies that will allow them to predict population group membership to guide searches. Such strategies may require testing of other polymorphic markers such as microsatellites for comparison with population databases (64,65).

Acknowledgments

This research is supported by funding from the Office of Naval Research N00014-04-1-0795 (C.K.H., J.N.) and NIH grant GM35326 (G.T., A.L.). The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of the Navy, the Department of Defense, or the U.S. government. We would like to thank Olga Malinowski for technical assistance.

References

1. Brittingham A, de la Cruz GC. Ancestry: 2000. 2000 Brief. US Census Bureau. 2004
2. Parham P, Ohta T. Population biology of antigen presentation by MHC class I molecules. *Science*. 1996; 272:67–74. [PubMed: 8600539]
3. Tu B, Mack SJ, Lazaro A, Lancaster A, Thomson G, Cao K, Chen M, Ling G, Hartzman R, Ng J, Hurley CK. HLA-A, -B, -C, -DRB1 allele and haplotype frequencies in an African American population. *Tissue Antigens*. 2007; 69:73–85. [PubMed: 17212710]
4. Robinson J, Waller MJ, Parham P, de Groot N, Bontrop R, Kennedy LJ, Stoeckl P, Marsh SG. IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. *Nucleic Acids Res*. 2003; 31:311–314. [PubMed: 12520010]
5. Lazaro A, Cao K, Masaberg C, Steiner N, Xiao Y, Tu B, Turner V, Nickerson P, Stoll S, Schall C, Valdez R, Ng J, Hartzman R, Hurley CK. Twenty-three novel HLA-B alleles identified during intermediate-resolution testing. *Tissue Antigens*. 2006; 68:245–248. [PubMed: 16948646]
6. Marsh SG, Albert ED, Bodmer WF, Bontrop RE, Dupont B, Erlich HA, Geraghty DE, Hansen JA, Hurley CK, Mach B, Mayr WR, Parham P, Petersdorf EW, Sasazuki T, Schreuder GM, Strominger

- JL, Svejgaard A, Terasaki PI, Trowsdale J. Nomenclature for factors of the HLA system, 2004. *Tissue Antigens*. 2005; 65:301–369. [PubMed: 15787720]
7. Lancaster, AK.; Nelson, MP.; Single, RM.; Meyer, D.; Thomson, G. PyPop: a software framework for population genomics: analyzing large-scale multi-locus genotype data. In: Altman, RB.; Dunker, K.; Hunter, L.; Jung, T.; Klein, T., editors. Pacific Symposium on Biocomputing 8. Singapore: World Scientific; 2003. p. 514-525.
 8. Lancaster AK, Single RM, Solberg OD, Nelson MP, Thomson G. PyPop update - a software pipeline for large-scale multilocus population genomics. *Tissue Antigens*. 2007; 69:192–197. [PubMed: 17445199]
 9. Guo SW, Thompson EA. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*. 1992; 48:361–372. [PubMed: 1637966]
 10. Meyer, D.; Single, R.; Mack, SJ.; Lancaster, A.; Nelson, MP.; Erlich, HA.; Fernandez-Vina, M.; Thomson, G. 13th IHWS anthropology/human genetic diversity joint report. Chapter 4. Single locus polymorphism of classical HLA genes. In: Hansen, JA., editor. Immunobiology of the human MHC. Seattle: International Histocompatibility Working Group Press; 2006. p. 653-704.
 11. Ewens W. The sampling theory of selectively neutral alleles. *Theor Pop Biol*. 1972; 3:87–112. [PubMed: 4667078]
 12. Watterson G. The homozygosity test of neutrality. *Genetics*. 1978; 88:405–417. [PubMed: 17248803]
 13. Slatkin M. An exact test for neutrality based on the Ewens sampling distribution. *Genetical Research*. 1994; 64:71–74. [PubMed: 7958833]
 14. Slatkin M. A correction to the exact test based on the Ewens sampling distribution. *Genetical Research*. 1996; 68:259–260. [PubMed: 9062082]
 15. Salamon H, Klitz W, Easteal S, Gao X, Erlich HA, Fernandez-Vina M, Trachtenberg EA, McWeeney SK, Nelson MP, Thomson G. Evolution of HLA class II molecules: Allelic and amino acid site variability across populations. *Genetics*. 1999; 152:393–400. [PubMed: 10224269]
 16. Dempster A, Laird N, Rubin D. Maximum likelihood estimation from incomplete data using the EM algorithm. *J Royal Stat Soc*. 1977; 39:1–38.
 17. Excoffier L, Slatkin M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol*. 1995; 12:921–927. [PubMed: 7476138]
 18. Hedrick PW. Gametic disequilibrium measures: proceed with caution. *Genetics*. 1987; 117:331–334. [PubMed: 3666445]
 19. Lewontin RC. The interaction of selection and linkage. II. Optimum models. *Genetics*. 1964; 50:757–782. [PubMed: 14221879]
 20. Lewontin RC. On measures of gametic disequilibrium. *Genetics*. 1988; 120:849–852. [PubMed: 3224810]
 21. Cramer, H. Mathematical methods of statistics. Princeton, NJ: University Press; 1946.
 22. Excoffier L, Laval G, Schneider S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Bioinformatics Online*. 2005; 1:47–50.
 23. Cao K, Moormann AM, Lyke KE, Masaberg C, Sumba OP, Doumbo OK, Koech D, Lancaster A, Nelson M, Meyer D, Single R, Hartzman RJ, Plowe CV, Kazura J, Mann DL, Sztein MB, Thomson G, Fernandez-Vina MA. Differentiation between African populations is evidenced by the diversity of alleles and haplotypes of HLA class I loci. *Tissue Antigens*. 2004; 63:293–325. [PubMed: 15009803]
 24. Izaabel, H.; Djoulah, SGHJ.; Caillat-Zucman, S.; Beaurain, G.; Akhayat, O.; Bach, J-F.; Khalil, I.; Benhamamouch, S.; Degos, L.; Deschamps, I.; Hors, J.; Sanchez-Mazas, A. Populations: Moroccan from Souss, Morocco and Algerian from Oran, Algeria. In: Hansen, JA., editor. Immunobiology of the Human MHC. Seattle: International Histocompatibility Working Group Press; 2006. p. 591-592.
 25. Canossi, A.; Piancatelli, D.; Oumhani, K.; Aureli, A.; Di Rocco, M.; Del Beato, T.; Liberatore, G.; Casciani, CU.; El Aouad, R.; Adorno, D. Population: Chaouya from Morocco. In: Hansen, JA., editor. Immunobiology of the Human MHC. Seattle: International Histocompatibility Working Group Press; 2006. p. 592-594.

26. Piancatelli, D.; Canossi, A.; Oumhani, K.; Aureli, A.; Di Rocco, M.; Del Beato, T.; Liberatore, G.; Casciani, CU.; El Aouad, R.; Adorno, D. Population: Metalsa from Morocco. In: Hansen, JA., editor. *Immunobiology of the Human MHC*. Seattle: International Histocompatibility Working Group Press; 2006. p. 594-595.
27. Arnaiz-Villena, A.; Moscoso, J.; Martinez-Laso, J.; Zamora, J.; Lowy, E.; del Moral, P.; Gomez-Casado, E. Moroccan from El Jadida, Morocco. In: Hansen, JA., editor. *Immunobiology of the Human MHC*. Seattle: International Histocompatibility Working Group Press; 2006. p. 595-596.
28. Ivanova M, Rozemuller E, Tyufekchiev N, Michailova A, Tilanus M, Naumova E. HLA polymorphism in Bulgarians defined by high-resolution typing methods in comparison with other populations. *Tissue Antigens*. 2002; 60:496–504. [PubMed: 12542743]
29. Martinovi I, Bakran M, Chaventré A, Jani ijevi B, Jovanovi V, Smolej-Naranci N, Kastelan A, Grubi Z, Zunec R, Roberts DF, Rudan P. Application of HLA class II polymorphism analysis to the study of the population structure of the Island of Krk, Croatia. *Hum Biol*. 1997; 69:819–829. [PubMed: 9353977]
30. Martinovi I, Bakran M, Chaventré A, Jani ijevi B, Jovanovi V, Smolej-Naranci N, Kastelan A, Grubi Z, Zunec R, Roberts DF, Rudan P. Application of HLA class II polymorphism analysis to the study of the population structure of the Island of Krk, Croatia. *Hum Biol*. 1997; 69:819–829. [PubMed: 9353977]
31. Cerná M, Fernandez-Viña M, Ivásková E, Stastny P. Comparison of HLA class II alleles in Gypsy and Czech populations by DNA typing with oligonucleotide probes. *Tissue Antigens*. 1992; 39:111–116. [PubMed: 1598683]
32. Sánchez-Velasco P, Leyva-Cobián F. The HLA class I and class II allele frequencies studied at the DNA level in the Svanetian population (Upper Caucasus) and their relationships to Western European populations. *Tissue Antigens*. 2001; 58:223–233. [PubMed: 11782273]
33. Reil A, Bein G, Machulla HK, Sternberg B, Seyfarth M. High-resolution DNA typing in immunoglobulin A deficiency confirms a positive association with DRB1*0301, DQB1*02 haplotypes. *Tissue Antigens*. 1997; 50:501–506. [PubMed: 9389325]
34. Lampis R, Morell L, Congia M, Doloretta Macis M, Mulargia A, Loddo M, De Virgiliis S, Marrosu MG, Todd JA, Cucca F. The inter-regional distribution of HLA class II haplotypes indicates the suitability of the Sardinian population for case-control association studies in complex diseases *Human Molecular Genetics*. 2000; 9:2959–2965.
35. Arnaiz-Villena A, Dimitroski K, Pacho A, Moscoso J, Gómez-Casado E, Silvera-Redondo C, Varela P, Blagoevska M, Zdravkovska V, Martínez-Laso J. HLA genes in Macedonians and the sub-Saharan origin of the Greeks. *Tissue Antigens*. 2001; 57:118–127. [PubMed: 11260506]
36. Petlichkovski A, Efinska-Mladenovska O, Trajkov D, Arsov T, Strezova A, Spiroski M. High-resolution typing of HLA-DRB1 locus in the Macedonian population. *Tissue Antigens*. 2004; 64:486–491. [PubMed: 15361127]
37. Rønningen KS, Spurkland A, Markussen G, Iwe T, Vartdal F, Thorsby E. Distribution of HLA class II alleles among Norwegian Caucasians. *Hum Immunol*. 1990; 29:275–281. [PubMed: 2286539]
38. Lin L, Jin L, Kimura A, Carrington M, Mignot E. DQ microsatellite association studies in three ethnic groups. *Tissue Antigens*. 1997; 50:507–520. [PubMed: 9389326]
39. Nowak J, Mika-Witkowska R, Polak M, Zajko M, Rogatko-Koro? M, Graczyk-Pol E, Lange A. Allele and extended haplotype polymorphism of HLA-A, -C, -B, -DRB1 and -DQB1 loci in Polish population and genetic affinities to other populations. *Tissue Antigens*. 2008; 71:193–205. [PubMed: 18179647]
40. Bodmer, J.; Cambon-Thomsen, A.; Hors, J.; Piazza, A.; Sanchez-Mazas, A. Report of the Anthropology Component. In: Charron, D., editor. *HLA: Proceedings of the Twelfth International Histocompatibility Workshop and Conference*; 1997. p. 269-274.
41. Spínola H, Middleton D, Brehm A. HLA genes in Portugal inferred from sequence-based typing: in the crossroad between Europe and Africa. *Tissue Antigens*. 2005; 66:26–36. [PubMed: 15982254]
42. Arnaiz-Villena A, Martinez-Laso J, Moscoso J, Livshits G, Zamora J, Gomez-Casado E, Silvera-Redondo C, Melvin K, Crawford MH. HLA genes in the Chuvashian population from European

- Russia: admixture of Central European and Mediterranean populations. *Hum Biol.* 2003; 75:375–392. [PubMed: 14527201]
43. Kapustin S, Lyshchov A, Alexandrova J, Imyanitov E, Blinov M. HLA class II molecular polymorphisms in healthy Slavic individuals from North-Western Russia. *Tissue Antigens.* 1999; 54:517–520. [PubMed: 10599891]
 44. Petrone A, Battelino T, Krzysnik C, Bugawan TL, Erlich HA, Di Mario A, Pozzilli P, Buzzetti R. Similar incidence of type 1 diabetes in two ethnically different populations (Italy and Slovenia) is sustained by similar HLA susceptible/protective haplotype frequencies. *Tissue Antigens.* 2002; 60:244–253. [PubMed: 12445307]
 45. Sánchez-Velasco P, Gomez-Casado E, Martinez-Laso J, Moscoso J, Zamora J, Lowy E, Silvera C, Cemborain A, Leyva-Cobián F, Arnaiz-Villena A. HLA alleles in isolated populations from North Spain: Origin of the Basques and the ancient Iberians. *Tissue Antigens.* 2003; 61 3804-392.
 46. Sánchez-Velasco P, Escribano de Diego J, Paz-Miguel JE, Ocejo-Vinyals G, Leyva-Cobián F. HLA-DR, DQ nucleotide sequence polymorphisms in the Paisesos (Pas valleys, Northern Spain) and comparison of the allelic and haplotypic frequencies with those of other European populations. *Tissue Antigens.* 1999; 53:65–73. [PubMed: 10082432]
 47. Muroa M, Marína L, Torío A, Moya-Quilesa MR, Minguelaa A, Rosique-Romana J, Sanchisa MJ, García-Calatayuda MC, García-Alonsoa AM, Álvarez-López MR. HLA polymorphism in the murcia population (Spain): in the cradle of the archaeological Iberians. *Human Immunology.* 2001; 62:910–921. [PubMed: 11543893]
 48. Vidal S, Morante MP, Moga E, Mosquera AM, Querol S, Garcia J, Rodriguez-Sanchez JI. Molecular analysis of HLA-DRB1 polymorphism in north-east Spain. *Eur J Immunogenet.* 2002; 29:75–77. [PubMed: 11841495]
 49. Allen M, Sandberg-Wollheim M, Sjogren K, Erlich HA, Petterson U, Gyllensten U. Association of susceptibility to multiple sclerosis in Sweden with HLA class II DRB1 and DQB1 alleles. *Hum Immunol.* 1994; 39:41–48. [PubMed: 8181961]
 50. Mack, SJ.; Crawford, MH.; Saha, N.; Jani, AJ.; Geyer, LN.; Reddy, M.; Erlich, H. Populations: Golla from Andhra Pradesh, and north Indians from New Delhi, India. In: Hansen, JA., editor. *Immunobiology of the Human MHC.* Seattle: International Histocompatibility Working Group Press; 2006. p. 605-607.
 51. Gazit, E.; Kalt, R.; Roewenthal, R. Population: Ashkenazi and non-Ashkenazi Jews from Israel. In: Hansen, JA., editor. *Immunobiology of the Human MHC.* Seattle: International Histocompatibility Working Group Press; 2006. p. 607-608.
 52. Saruhan-Direskeneli, G.; Uyar, FA. Population: Turkish from Marmara, Turkey. In: Hansen, JA., editor. *Immunobiology of the Human MHC.* Seattle: International Histocompatibility Working Group Press; 2006. p. 609-610.
 53. Cao K, Hollenbach J, Shi X, Shi W, Chopek M, Fernandez-Vina MA. Analysis of the frequencies of HLA-A, B, and C alleles and haplotypes in the five major ethnic groups of the United States reveals high levels of diversity in these loci and contrasting distribution patterns in these populations. *Hum Immunol.* 2001; 62:1009–1030. [PubMed: 11543903]
 54. Bugawan TL, Klitz W, Blair A, Erlich HA. High-resolution HLA class I typing in the CEPH families: analysis of linkage disequilibrium among HLA loci. *Tissue Antigens.* 2000; 56:392–404. [PubMed: 11144287]
 55. Klitz W, Maiers M, Spellman S, Baxter-Lowe LA, Schmeckpeper B, Williams TM, et al. New HLA haplotype frequency reference standards: high-resolution and large sample typing of HLA DR-DQ haplotypes in a sample of European Americans. *Tissue Antigens.* 2003; 62:296–307. [PubMed: 12974796]
 56. Rossman MD, Thompson B, Frederick M, Maliarik M, Iannuzzi MC, Rybicki BA, Pandey JP, Newman LS, Magira E, Beznik-Cizman B, Monos D. ACCESS Group. HLA-DRB1*1101: a significant risk factor for sarcoidosis in blacks and whites. *Am J Hum Genet.* 2003; 73:720–735. [PubMed: 14508706]
 57. Weir BS, Cockerham CC. 1984 Estimating F-statistics for the analysis of population structure. *Evolution.* 38:1358–1370.

58. Hedrick PW. A Standardized Genetic Differentiation Measure. *Evolution*. 2005; 59:1633–1638. [PubMed: 16329237]
59. Cano P, Klitz W, Mack SJ, Maiers M, Marsh SG, Noreen H, Reed EF, Senitzer D, Setterholm M, Smith A, Fernandez-Vina M. Common and well-documented HLA alleles: report of the Ad-Hoc committee of the american society for histocompatibility and immunogenetics. *Hum Immunol*. 2007; 68:392–417. [PubMed: 17462507]
60. Lazaro AM, Xiao Y, Cao K, Masaberg C, Nichol L, Ng J, Hurley CK, Posch PE. Thirty-two novel HLA-A alleles identified during intermediate resolution testing. *Tissue Antigens*. 2008; 71:165–168. [PubMed: 18069934]
61. Maiers M, Gragert L, Klitz W. High-resolution HLA alleles and haplotypes in the United States population. *Hum Immunol*. 2007; 68:779–788. [PubMed: 17869653]
62. Single, R.; Meyer, D.; Mack, SJ.; Lancaster, A.; Nelson, MP.; Erlich, H.; Fernandez-Viña, MA.; Thomson, G. 13th International Histocompatibility Workshop anthropology / Human genetic diversity joint report. Haplotype frequencies and linkage disequilibrium among classical HLA genes. In: Hansen, JA., editor. *Immunobiology of the Human MHC*. Seattle: International Histocompatibility Working Group Press; 2006. p. 705-746.
63. Horton R, Gibson R, Coggill P, Miretti M, Allcock RJ, Almeida J, Forbes S, Gilbert JG, Halls K, Harrow JL, Hart E, Howe K, Jackson DK, Palmer S, Roberts AN, Sims S, Stewart CA, Traherne JA, Trevanion S, Wilming L, Rogers J, de Jong PJ, Elliott JF, Sawcer S, Todd JA, Trowsdale J, Beck S. Variation analysis and gene annotation of eight MHC haplotypes: The MHC Haplotype Project. *Immunogenetics*. 2008; 60:1–18. [PubMed: 18193213]
64. Bamshad M, Wooding S, Salisbury BA, Stephens JC. Deconstructing the relationship between genetics and race. *Nature Reviews Genetics*. 2004; 5:598–609.
65. Rosenberg NA, Li LM, Ward R, Pritchard JK. Informativeness of genetic markers for inference of ancestry. *Am J Hum Genet*. 2003; 73:1402–1422. [PubMed: 14631557]
66. Conrad DF, Jakobsson M, Coop G, Wen X, Wall JD, Rosenberg NA, Pritchard JK. A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nat Genet*. 2006; 38:1251–1260. [PubMed: 17057719]
67. Guthery SL, Salisbury BA, Pungliya MS, Stephens JC, Bamshad M. The structure of common genetic variation in United States populations. *Am J Hum Genet*. 2007; 81:1221–1231. [PubMed: 17999361]
68. Hurley CK, Fernandez VM, Setterholm M. Maximizing optimal hematopoietic stem cell donor selection from registries of unrelated adult volunteers. *Tissue Antigens*. 2003; 61:415–424. [PubMed: 12823765]
69. Hurley CK, Wagner JE, Setterholm MI, Confer DL. Advances in HLA: Practical implications for selecting adult donors and cord blood units. *Biol Blood Marrow Transplant*. 2006; 12:28–33. [PubMed: 16399581]

Table 1

HLA allele^a frequencies in 558 random Americans with Eastern European ancestry^b

HLA-A Allele	Frequency	n	HLA-C allele	Frequency	n	HLA-B Allele	Frequency	n	DRB1 Alleles	Frequency	n
010101g	0.14247	159	0102	0.03674	41	0702	0.10394	116	0101	0.09319	104
02010101g	0.27509	307	0202/0210	0.05824	65 ^d	0705g	0.00538	6	0102	0.01703	19
0205	0.01613	18	0302	0.00358	4	0714	0.00090	1	0103	0.00896	10
0206	0.00358	4	0303	0.05466	61	0801g	0.09050	101	0301	0.10215	114
0217	0.00090	1	030401/030403	0.06093	68	1301	0.00090	1	0302	0.00090	1
03010101g	0.12903	144	030402	0.00090	1	1302	0.03047	34	0401	0.06810	76
0302	0.00627	7	0332 (V)	0.00090	1	1401	0.00358	4	0402	0.01344	15
0325 (V) ^c	0.00090	1	04010101g	0.12993	145	1402	0.02957	33	0403	0.00896	10
1101	0.06362	71	0403	0.00269	3	15010101g	0.05556	62	0404	0.04032	45
2301g	0.01792	20	0501g	0.05556	62	1503	0.00358	4	0405	0.00627	7
24020101g	0.10215	114	0602	0.09767	109	1506	0.00090	1	0406	0.00090	1
240301g	0.00090	1	070101g	0.161239	180	1510	0.00179	2	0407	0.00986	11
2501	0.02688	30	0702	0.11559	129	1517	0.00717	8	0408	0.00179	2
2601	0.04122	46	070401g	0.01971	22	1518	0.00627	7	0410	0.00090	1
29010101g	0.00448	5	0713	0.00090	1	1534	0.00090	1	0426	0.00090	1
2902	0.02151	24	0732N (V)	0.00090	1	180101g	0.05466	61	0701	0.12455	139
3001	0.01075	12	0801	0.00179	2	1803	0.00179	2	0801	0.02688	30
3002	0.00538	6	0802	0.03136	35	2702	0.00986	11	0802	0.00090	1
3004	0.00269	3	0803	0.00090	1	270502g	0.03853	43	0803	0.00269	3
3101	0.02419	27	1202	0.00806	9	2707	0.00090	1	0804	0.00090	1
3201	0.02867	32	1203	0.06989	78	350101g	0.06541	73	0901	0.01165	13
3208	0.00090	1	1402	0.01613	18	3502	0.02599	29	1001	0.01075	12
3301	0.01344	15	1502	0.01882	21	3503	0.02330	26	1101	0.05376	60
3303	0.00358	4	1505	0.00538	6	3508	0.00806	9	1102	0.00269	3
3402	0.00090	1	1601	0.02778	31	3512	0.00090	1	1103	0.00448	5
3601	0.00090	1	1602	0.00448	5	3701	0.01523	17	1104	0.05824	65

HLA-A Allele	Frequency	n	HLA-C allele	Frequency	n	HLA-B Allele	Frequency	n	DRB1 Alleles	Frequency	n
6601	0.00627	7	1604	0.00538	6	3801	0.02599	29	120101g	0.01344	15
680101	0.01254	14	1701g	0.00986	11	3901	0.00986	11	1202	0.00179	2
680102g	0.02867	32				3906	0.00538	6	1301	0.06631	74
6802	0.00717	8	Total	1.00000	1116	400101g	0.04391	49	1302	0.04480	50
7401g	0.00090	1				400201g	0.01971	22	1303	0.01882	21
						4006	0.00179	2	1305	0.00448	5
Total	1.00000	1116				4010	0.00090	1	1401	0.01434	16
						4101	0.00538	6	1403	0.00090	1
						4102	0.00448	5	1404	0.00179	2
						4201	0.00090	1	1406	0.00179	2
						44020101g	0.06362	71	1501	0.11022	123
						4403	0.04122	46	1502	0.00896	10
						4405	0.00627	7	1503	0.00179	2
						440204 (V)	0.00090	1	1601	0.03763	42
						4501g	0.00538	6	1602	0.00179	2
						4601	0.00090	1			
						4701	0.00448	5	Total	1.00000	1116
						4801g	0.00269	3			
						4901	0.02957	33			
						5001	0.01254	14			
						5002	0.00090	1			
						510101g	0.05287	59			
						5108	0.00179	2			
						5201	0.00806	9			
						5301	0.00269	3			
						5501	0.01613	18			
						5601	0.00717	8			
						5701	0.02778	31			
						5703	0.00090	1			

\$watermark-text

\$watermark-text

\$watermark-text

HLA-A Allele	Frequency	n	HLA-C allele	Frequency	n	HLA-B Allele	Frequency	n	DRB1 Alleles	Frequency	n
						5801	0.00806	9			
						5802	0.00090	1			
						7801	0.00090	1			
						Total	1.00000	1116			

^a Alleles which were identical in exons 2 and 3 were not distinguished. These alleles are indicated by the addition of the a “g.” to the name of the lowest numbered allele in that group.

^b Alleles were assumed to be homozygous if the typing was consistent with a single allele in both intermediate and high resolution testing.

^c (V) indicates a novel allele defined in this study.

^d Includes one cell with a Cw*0210 allele.

Table 2

Ewens-Watterson homozygosity test of neutrality

Locus	Observed F	Expected F	Normalized deviate of F (Fnd)	p
HLA-A	0.1335	0.1449	-0.2169	0.5090
HLA-B	0.0484	0.0704	-1.0892	0.0644
HLA-C	0.0879	0.1551	-1.1779	0.0308 *
HLA-DRB1	0.0698	0.1045	-1.0000	0.0843

* Significant at the 5% level.

\$watermark-text

\$watermark-text

\$watermark-text

Table 3

Cells and their novel HLA alleles

Cell	HLA-A	HLA-B	HLA-C	HLA-DR	GenBank Accession No. ^a
NT00682	A*03010101, *0325^b	B*0702, *4701	Cw*0602, 0702	DRB1*0701, *1301	DQ987874
NT00681	A*02010101g, *24020101g	B*440204 , *130201	Cw*050101g, *0602	DRB1*0101, *0701	DQ987875
NT00678	A*02010101g, 03010101g	B*15010101g, *510101g	Cw* 0332 , 150201	DRB1*0401, *1302	DQ984199
NT00597	A*03010101g, *1101	B*070201, *400201g	Cw* 0732N , 020202	DRB1*1301, *1501	DQ372911-DQ372913

^a Accession number of novel allele
^b Allele previously reported (33).
^c Novel allele is in bold type. The name have been officially assigned by the WHO Nomenclature Committee.

Table 4

HLA two-locus haplotype frequencies and linkage disequilibrium values

Locus Pair	Haplotype	Frequency	Observed (n)	D'ij	p-value
A-B	010101g:0801g	0.06179	69.0	0.63008	1.16582E-59
	03010101g:0702	0.05082	56.7	0.4132	2.60221E-34
	02010101g:15010101g	0.03365	37.5	0.45598	2.00543E-09
	02010101g:44020101g	0.03301	36.8	0.34645	1.17214E-06
	02010101g:400101g	0.02857	31.9	0.51828	1.71095E-09
	02010101g:180101g	0.02048	22.9	0.13735	n.s.
	02010101g:0702	0.02033	22.7	-0.28894	0.042847471
	03010101g:350101g	0.02005	22.4	0.20372	2.89971E-06
	02010101g:1302	0.01667	18.6	0.37525	0.000309397
	02010101g:510101g	0.01662	18.5	0.05423	n.s.
	02010101g:270502g	0.01616	18.0	0.19896	0.030776795
	2902:4403	0.01612	18.0	0.73862	1.16216E-69
	1101:350101g	0.01550	17.3	0.19075	3.42285E-10
	2501:180101g	0.01462	16.3	0.51749	6.55176E-33
	02010101g:5701	0.01455	16.2	0.34321	0.001654855
	02010101g:0801g	0.01439	16.1	-0.42218	0.00613189
	3301:1402	0.01163	13.0	0.86095	2.00292E-82
	24020101g:0702	0.01138	12.7	0.00831	n.s.
	010101g:5701	0.01065	11.9	0.28102	9.90095E-05
	02010101g:350101g	0.01027	11.5	-0.42929	0.01942872
	24020101g:3502	0.01010	11.3	0.31917	2.42787E-07
	010101g:3502	0.00989	11.0	0.27757	0.000202656
	2601:3801	0.00927	10.3	0.32898	4.81405E-18
	2301g:4403	0.00896	10.0	0.4785	2.12777E-25
	680102g:44020101g	0.00856	9.6	0.25154	2.39008E-08
	24020101g:350101g	0.00823	9.2	0.02633	n.s.
	24020101g:180101g	0.00807	9.0	0.05064	n.s.
	02010101g:4901	0.00751	8.4	-0.07727	n.s.
	24020101g:15010101g	0.00719	8.0	0.03037	n.s.
	010101g:3701	0.00715	8.0	0.38117	0.000102148
	02010101g:3503	0.00712	7.9	0.04206	n.s.
	010101g:4901	0.00708	7.9	0.11296	n.s.
	03010101g:510101g	0.00695	7.8	0.00278	n.s.
	3101:400101g	0.00675	7.5	0.24586	1.58941E-09
	03010101g:3503	0.00663	7.4	0.1787	0.016599135
	03010101g:1402	0.00625	7.0	0.09452	n.s.
	24020101g:510101g	0.00623	7.0	0.01745	n.s.

Locus Pair	Haplotype	Frequency	Observed (n)	D'ij	p-value
	1101:510101g	0.00620	6.9	0.05732	n.s.
	2601:510101g	0.00567	6.3	0.08942	0.0087538
	24020101g:4403	0.00538	6.0	0.0315	n.s.
	6802:1402	0.00538	6.0	0.74238	1.47839E-33
	2301g:4901	0.00538	6.0	0.27867	5.83564E-13
	03010101g:3901	0.00538	6.0	0.47811	3.46862E-05
	0205:5001	0.00538	6.0	0.4192	6.39536E-35
	010101g:4403	0.00535	6.0	-0.08977	n.s.
	1101:44020101g	0.00531	5.9	0.02253	n.s.
	24020101g:3801	0.00488	5.4	0.09519	n.s.
	1101:0702	0.00487	5.4	-0.26364	n.s.
	3201:0702	0.00481	5.4	0.07119	n.s.
	02010101g:2702	0.00471	5.3	0.28001	n.s.
	1101:3503	0.00458	5.1	0.14182	0.00499784
	010101g:1302	0.00456	5.1	0.00835	n.s.
	1101:15010101g	0.00453	5.1	0.01914	n.s.
	02010101g:5001	0.00428	4.8	0.09137	n.s.
	24020101g:400201g	0.00419	4.7	0.12305	n.s.
	3201:400201g	0.00419	4.7	0.18926	1.80609E-07
	02010101g:5201	0.00406	4.5	0.31511	n.s.
	02010101g:3801	0.00376	4.2	-0.47459	n.s.
	680102g:510101g	0.00373	4.2	0.08167	0.047248712
	24020101g:270502g	0.00372	4.2	-0.0545	n.s.
	2501:44020101g	0.00369	4.1	0.07973	n.s.
	03010101g:15010101g	0.00365	4.1	-0.49044	n.s.
	0205:4901	0.00358	4.0	0.19852	1.14771E-06
	3001:1302	0.00358	4.0	0.31238	8.37536E-10
	010101g:1517	0.00358	4.0	0.41693	0.00368987
	24020101g:3508	0.00358	4.0	0.38124	0.000662789
	2601:4901	0.00358	4.0	0.08343	0.018950042
	03010101g:3701	0.00358	4.0	0.122	n.s.
	010101g:0702	0.00356	4.0	-0.75992	0.000424585
	03010101g:3502	0.00348	3.9	0.0056	n.s.
	680102g:270502g	0.00345	3.8	0.08504	0.014760515
	3101:510101g	0.00345	3.8	0.0947	0.034990683
	24020101g:4405	0.00343	3.8	0.49569	9.60466E-05
	1101:5501	0.00342	3.8	0.15822	0.009420807
	2601:44020101g	0.00334	3.7	0.01949	n.s.
	02010101g:5501	0.00333	3.7	-0.25014	n.s.

Locus Pair	Haplotype	Frequency	Observed (n)	D'ij	p-value
	1101:400101g	0.00330	3.7	0.01241	n.s.
	24020101g:3503	0.00317	3.5	0.03779	n.s.
	1101:270502g	0.00308	3.4	0.0175	n.s.
	680102g:350101g	0.00308	3.4	0.04484	n.s.
	010101g:44020101g	0.00307	3.4	-0.65677	0.020694838
	24020101g:44020101g	0.00306	3.4	-0.52313	n.s.
	6801:350101g	0.00298	3.3	0.18454	0.008625701
	2601:270502g	0.00297	3.3	0.03742	n.s.
	03010101g:3801	0.00294	3.3	-0.12207	n.s.
	02010101g:400201g	0.00293	3.3	-0.45991	n.s.
	03010101g:0801g	0.00287	3.2	-0.7543	0.002217044
	3201:15010101g	0.00280	3.1	0.04475	n.s.
	010101g:270502g	0.00274	3.1	-0.50177	n.s.
	0302:4403	0.00269	3.0	0.40401	2.32238E-07
	2601:5601	0.00269	3.0	0.34813	1.87831E-06
	3002:180101g	0.00269	3.0	0.47109	1.49544E-06
	02010101g:0705g	0.00269	3.0	0.31026	n.s.
	6601:4102	0.00269	3.0	0.59748	9.84576E-64
	010101g:5801	0.00267	3.0	0.21953	n.s.
	010101g:180101g	0.00267	3.0	-0.65757	0.031313371
	Total	0.79356	885.6		
C-B					
	0702:0702	0.10215	114	0.98051	4.3092E-209
	070101g:0801g	0.08602	96	0.94104	5.7888E-114
	04010101g:350101g	0.05732	64	0.85786	1.04758E-85
	0501g:44020101g	0.04301	48	0.75908	6.1171E-125
	0304:400101g	0.04211	47	0.95632	4.1105E-159
	0602:1302	0.03047	34	1	1.96738E-72
	0303:15010101g	0.03043	34	0.53069	3.72594E-69
	0802:1402	0.02957	33	1	2.1101E-230
	070101g:4901	0.02867	32	0.96379	7.67956E-38
	1203:3801	0.02599	29	1	3.66438E-88
	1203:180101g	0.02599	29	0.43599	2.2296E-37
	0602:5701	0.02509	28	0.89263	5.58313E-53
	070101g:180101g	0.0233	26	0.31662	5.91871E-09
	1601:4403	0.02324	25.9	0.82971	4.7159E-113
	0102:270502g	0.0215	24	0.56869	1.10709E-76
	04010101g:3502	0.02055	22.9	0.75977	7.50689E-27
	0202:400201g	0.01699	19	0.8535	1.94262E-59

Locus Pair	Haplotype	Frequency	Observed (n)	D'ij	p-value
	0202:270502g	0.01613	18	0.38265	7.82735E-25
	04010101g:4403	0.01592	17.8	0.29452	1.29805E-07
	0303:5501	0.01523	17	0.94109	6.68502E-63
	1402:510101g	0.01521	17	0.94013	5.8022E-65
	04010101g:3503	0.01521	17	0.60085	1.03816E-15
	0602:3701	0.01434	16	0.93481	3.65869E-32
	070401g:44020101g	0.01434	16	0.70893	1.53908E-38
	0304:15010101g	0.01432	16	0.20965	2.58447E-11
	1502:510101g	0.01341	15	0.69646	2.39111E-42
	0602:5001	0.01075	12	0.84168	5.81097E-22
	0202:510101g	0.00899	10	0.11876	0.000163772
	0202:2702	0.00806	9	0.80694	2.91884E-27
	1202:5201	0.00806	9	1	1.1E-244
	070101g:1517	0.00717	8	1	8.327E-11
	04010101g:15010101g	0.00632	7.1	-0.12461	n.s.
	04010101g:3508	0.00627	7	0.74459	6.47678E-09
	0202:4405	0.00627	7	1	1.37115E-26
	0102:5601	0.00627	7	0.87023	1.13364E-36
	1203:3503	0.00626	7	0.21361	5.80702E-05
	1203:3901	0.00538	6	0.5113	5.07126E-10
	1505:0705g	0.00538	6	1	1.1E-244
	0501g:180101g	0.00538	6	0.04532	n.s.
	070401g:1518	0.00448	5	0.70854	3.89536E-40
	0602:3502	0.00448	5	0.08283	n.s.
	0602:4701	0.00448	5	1	9.63885E-12
	1701g:4102	0.00448	5	1	9.8096E-112
	070101g:5801	0.00448	5	0.47065	0.00118046
	0102:510101g	0.00448	5	0.07294	0.043991193
	0702:3901	0.00448	5	0.38326	0.000410173
	0602:4501g	0.00448	5	0.81529	1.15469E-09
	0303:350101g	0.00361	4	0.00063	n.s.
	0702:3906	0.00358	4	0.6231	2.30398E-05
	1701g:4101	0.00358	4	0.66335	6.08874E-60
	1604:44020101g	0.00358	4	0.64436	9.49095E-10
	0302:5801	0.00358	4	1	2.1555E-109
	070101g:5701	0.00269	3	-0.39597	n.s.
	0102:15010101g	0.00269	3	0.01865	n.s.
	0501g:510101g	0.00269	3	-0.08475	n.s.
	Total	0.91891	1025.7		

Locus Pair	Haplotype	Frequency	Observed (n)	D'ij	p-value
B-DRB1					
	0801g:0301	0.0669	74.7	0.70951	7.2272E-109
	0702:1501	0.05501	61.4	0.4709	2.47214E-52
	4403:0701	0.03101	34.6	0.71718	1.30617E-39
	350101g:0101	0.02833	31.6	0.37478	5.02831E-25
	3502:1104	0.0222	24.8	0.84511	9.34717E-77
	1302:0701	0.02208	24.6	0.68553	5.15892E-27
	15010101g:0401	0.01759	19.6	0.26668	1.31806E-15
	44020101g:0401	0.01604	17.9	0.2013	1.23338E-10
	5701:0701	0.01515	16.9	0.48088	6.06376E-13
	180101g:1501	0.01413	15.8	0.16675	0.000141231
	270502g:0101	0.01379	15.4	0.29182	1.14487E-09
	1402:0102	0.01344	15	0.78306	1.38818E-86
	180101g:1104	0.01329	14.8	0.19628	2.29935E-10
	400101g:0404	0.01253	14	0.27919	4.60633E-19
	510101g:1301	0.01215	13.6	0.1752	2.11494E-07
	15010101g:1301	0.01042	11.6	0.12992	7.81136E-05
	510101g:0101	0.0096	10.7	0.09754	0.016327566
	400101g:1302	0.00887	9.9	0.16452	5.38013E-08
	44020101g:0101	0.00813	9.1	0.04015	n.s.
	5001:0701	0.00806	9	0.59205	3.41744E-09
	44020101g:1601	0.0079	8.8	0.15695	6.12482E-05
	0702:1301	0.00722	8.1	0.00552	n.s.
	44020101g:1104	0.00722	8.1	0.06528	0.036039816
	3801:0402	0.00717	8	0.52088	1.69081E-35
	5201:1502	0.00717	8	0.88788	4.6651E-174
	400201g:1101	0.00681	7.6	0.30846	8.76433E-10
	0702:0401	0.00673	7.5	-0.04866	n.s.
	0702:0701	0.00668	7.5	-0.48427	0.037683092
	44020101g:0701	0.00648	7.2	-0.17063	n.s.
	350101g:1501	0.0063	7	-0.12668	n.s.
	0702:0301	0.00628	7	-0.40878	n.s.
	2702:1601	0.00627	7	0.62214	1.00044E-25
	4901:1302	0.00627	7	0.17517	2.39947E-06
	270502g:0404	0.00575	6.4	0.11345	0.000214345
	0801g:1501	0.00563	6.3	-0.43566	n.s.
	1517:1302	0.00538	6	0.73827	3.79482E-22
	400101g:0801	0.00536	6	0.1625	2.55113E-05
	0801g:0401	0.00528	5.9	-0.14319	n.s.

Locus Pair	Haplotype	Frequency	Observed (n)	D'ij	p-value
	510101g:1101	0.00519	5.8	0.04691	n.s.
	0702:0101	0.00508	5.7	-0.4758	n.s.
	4901:1501	0.00495	5.5	0.06438	n.s.
	510101g:0404	0.00487	5.4	0.07161	0.037946254
	400101g:0101	0.00475	5.3	0.01641	n.s.
	15010101g:0101	0.00472	5.3	-0.08872	n.s.
	5501:1301	0.00469	5.2	0.2407	0.00011185
	1402:0701	0.00448	5	0.0308	n.s.
	350101g:0103	0.00448	5	0.465	2.35967E-08
	4901:1101	0.00445	5	0.10238	0.012260941
	44020101g:120101g	0.00444	5	0.31092	6.02564E-06
	1402:1303	0.00438	4.9	0.20954	2.80067E-08
	180101g:1101	0.00438	4.9	0.02832	n.s.
	270502g:0801	0.0041	4.6	0.11837	0.001027175
	3503:1101	0.00408	4.6	0.12841	0.005444055
	15010101g:1501	0.004	4.5	-0.34666	n.s.
	3801:1301	0.00395	4.4	0.09178	n.s.
	400201g:1301	0.00394	4.4	0.14299	0.011026634
	270502g:0301	0.00388	4.3	-0.01371	n.s.
	4403:1101	0.00385	4.3	0.04179	n.s.
	44020101g:1301	0.00375	4.2	-0.09787	n.s.
	1302:1501	0.0037	4.1	0.01278	n.s.
	180101g:0701	0.00364	4.1	-0.46466	n.s.
	0702:0801	0.00362	4	0.03432	n.s.
	15010101g:0701	0.00362	4	-0.47724	n.s.
	15010101g:1101	0.00358	4	0.01176	n.s.
	5701:1301	0.00358	4	0.06718	n.s.
	5601:0101	0.00358	4	0.44862	7.11435E-05
	4102:1303	0.00358	4	0.79616	5.51865E-38
	3701:1101	0.00358	4	0.19159	0.000839126
	44020101g:1302	0.00358	4	0.0183	n.s.
	3906:0801	0.00357	4	0.65528	3.54525E-22
	0801g:0701	0.00357	4	-0.68328	0.006609304
	3901:1601	0.00357	4	0.33706	1.3379E-08
	350101g:1301	0.00356	4	-0.17882	n.s.
	1402:1302	0.00354	4	0.07846	0.034630944
	3701:1001	0.00352	3.9	0.3173	6.36729E-19
	180101g:1601	0.00352	3.9	0.04117	n.s.
	3503:1401	0.00349	3.9	0.22521	4.22484E-09

Locus Pair	Haplotype	Frequency	Observed (n)	D'ij	p-value
	180101g:0401	0.00347	3.9	-0.06794	n.s.
	180101g:0301	0.00337	3.8	-0.39588	n.s.
	0702:0404	0.00328	3.7	-0.21706	n.s.
	3801:0101	0.00317	3.5	0.03157	n.s.
	400101g:0401	0.00281	3.1	-0.06139	n.s.
	5701:1501	0.00277	3.1	-0.0961	n.s.
	350101g:0407	0.00275	3.1	0.22884	0.003938248
	510101g:1104	0.00269	3	-0.12699	n.s.
	0702:1401	0.00269	3	0.09325	n.s.
	510101g:0403	0.00269	3	0.26093	0.000451137
	350101g:0403	0.00269	3	0.25101	0.002579612
	5701:0801	0.00269	3	0.07429	0.014681015
	4405:1601	0.00269	3	0.40623	4.97935E-08
	0705g:1001	0.00269	3	0.49457	2.26745E-31
	5801:1302	0.00269	3	0.30206	2.65521E-05
	5001:0301	0.00269	3	0.12489	n.s.
	4501g:0401	0.00269	3	0.46346	2.54411E-05
	350101g:0402	0.00268	3	0.14346	0.034477034
	3801:1101	0.00265	3	0.0511	n.s.
	180101g:1302	0.00265	3	0.00481	n.s.
	Total	0.72793	812.8		

Table 5

HLA three- and four-locus haplotype frequencies

Loci	Haplotype	Frequency	Observed (n)
A-B-DRB1	010101g:0801g:0301	0.04544	50.7
	03010101g:0702:1501	0.02985	33.3
	02010101g:1302:0701	0.01510	16.9
	02010101g:0702:1501	0.01440	16.1
	02010101g:15010101g:0401	0.01425	15.9
	2902:4403:0701	0.01344	15.0
	03010101g:350101g:0101	0.01325	14.8
	02010101g:0801g:0301	0.01211	13.5
	02010101g:5701:0701	0.00986	11.0
	24020101g:3502:1104	0.00976	10.9
	02010101g:44020101g:0401	0.00896	10.0
	03010101g:0702:1301	0.00828	9.2
	010101g:3502:1104	0.00824	9.2
	02010101g:180101g:1104	0.00813	9.1
	3301:1402:0102	0.00806	9.0
	2501:180101g:1501	0.00803	9.0
	1101:350101g:0101	0.00759	8.5
	2301g:4403:0701	0.00717	8.0
	3101:400101g:0404	0.00704	7.9
	24020101g:0702:1501	0.00701	7.8
	02010101g:270502g:0101	0.00627	7.0
	2601:3801:0402	0.00627	7.0
	02010101g:44020101g:1601	0.00627	7.0
	02010101g:44020101g:0101	0.00627	7.0
	02010101g:2702:1601	0.00538	6.0
	0205:5001:0701	0.00538	6.0
	02010101g:400101g:1302	0.00530	5.9
	02010101g:400101g:0801	0.00501	5.6
	010101g:5701:0701	0.00473	5.3
	010101g:0801g:1501	0.00451	5.0
	02010101g:400101g:0404	0.00448	5.0
	010101g:4403:0701	0.00448	5.0
	1101:44020101g:0401	0.00448	5.0
	02010101g:5201:1502	0.00448	5.0
	02010101g:510101g:1301	0.00448	5.0
	03010101g:510101g:0101	0.00445	5.0
	02010101g:44020101g:0701	0.00442	4.9

Loci	Haplotype	Frequency	Observed (n)
	02010101g:0702:0801	0.00437	4.9
	1101:0702:1501	0.00411	4.6
	02010101g:510101g:0404	0.00407	4.5
	03010101g:0801g:0301	0.00390	4.4
	02010101g:350101g:0101	0.00378	4.2
	010101g:1302:0701	0.00362	4.0
	02010101g:4901:1101	0.00358	4.0
	02010101g:4901:1302	0.00358	4.0
	03010101g:0702:0101	0.00358	4.0
	02010101g:44020101g:120101g	0.00358	4.0
	03010101g:0702:0401	0.00358	4.0
	02010101g:15010101g:1101	0.00358	4.0
	6802:1402:1303	0.00358	4.0
	24020101g:4403:1101	0.00358	4.0
	2601:510101g:1301	0.00358	4.0
	3101:510101g:1301	0.00358	4.0
	02010101g:5701:1301	0.00358	4.0
	03010101g:350101g:1501	0.00356	4.0
	1101:15010101g:1301	0.00344	3.8
	03010101g:0702:0701	0.00327	3.7
	1101:350101g:0103	0.00312	3.5
	24020101g:350101g:0101	0.00312	3.5
	02010101g:15010101g:1501	0.00295	3.3
	03010101g:0702:0301	0.00289	3.2
	02010101g:1302:1501	0.00282	3.1
	010101g:0801g:0701	0.00278	3.1
	010101g:180101g:1104	0.00276	3.1
	02010101g:44020101g:1104	0.00275	3.1
	02010101g:400101g:0101	0.00269	3.0
	02010101g:0801g:0401	0.00269	3.0
	010101g:270502g:0301	0.00269	3.0
	680102g:44020101g:1302	0.00269	3.0
	2601:270502g:0101	0.00269	3.0
	680102g:350101g:0407	0.00269	3.0
	24020101g:3801:1301	0.00269	3.0
	03010101g:1402:0102	0.00269	3.0
	02010101g:15010101g:0901	0.00269	3.0
	02010101g:350101g:1501	0.00269	3.0
	02010101g:180101g:1101	0.00269	3.0

Loci	Haplotype	Frequency	Observed (n)
	24020101g:350101g:0404	0.00269	3.0
	03010101g:3503:0801	0.00269	3.0
	6601:4102:1303	0.00269	3.0
	24020101g:180101g:0301	0.00269	3.0
	02010101g:510101g:0901	0.00269	3.0
	680102g:270502g:1301	0.00265	3.0
	24020101g:1302:0701	0.00265	3.0
	Total	0.50999	569.3
A-C-B-DRB1	010101g:070101g:0801g:0301	0.04352	48.6
	03010101g:0702:0702:1501	0.02949	32.9
	02010101g:0602:1302:0701	0.01511	16.9
	02010101g:070101g:0801g:0301	0.01410	15.7
	02010101g:0702:0702:1501	0.01373	15.3
	2902:1601:4403:0701	0.01286	14.4
	03010101g:04010101g:350101g:0101	0.01251	14.0
	02010101g:070101g:180101g:1104	0.00986	11.0
	02010101g:0501g:44020101g:0401	0.00941	10.5
	24020101g:04010101g:3502:1104	0.00896	10.0
	02010101g:0304:15010101g:0401	0.00894	10.0
	1101:04010101g:350101g:0101	0.00886	9.9
	03010101g:0702:0702:1301	0.00829	9.3
	3301:0802:1402:0102	0.00806	9.0
	02010101g:0602:5701:0701	0.00805	9.0
	2501:1203:180101g:1501	0.00802	8.9
	3101:0304:400101g:0404	0.00717	8.0
	2301g:04010101g:4403:0701	0.00716	8.0
	24020101g:0702:0702:1501	0.00699	7.8
	2601:1203:3801:0402	0.00627	7.0
	010101g:04010101g:3502:1104	0.00550	6.1
	02010101g:0304:400101g:0801	0.00538	6.0
	02010101g:0202:510101g:1301	0.00538	6.0
	0205:0602:5001:0701	0.00538	6.0
	02010101g:0304:400101g:1302	0.00528	5.9
	02010101g:0501g:44020101g:0701	0.00506	5.6
	010101g:070101g:0801g:1501	0.00459	5.1
	02010101g:0202:2702:1601	0.00448	5.0
	02010101g:0304:400101g:0404	0.00448	5.0
	02010101g:0702:0702:0801	0.00448	5.0
	03010101g:0702:0702:0101	0.00448	5.0

Loci	Haplotype	Frequency	Observed (n)
	02010101g:1202:5201:1502	0.00448	5.0
	02010101g:0501g:44020101g:0101	0.00448	5.0
	02010101g:070101g:4901:1101	0.00448	5.0
	02010101g:04010101g:350101g:1501	0.00448	5.0
	1101:0702:0702:1501	0.00448	5.0
	02010101g:0304:400101g:0101	0.00403	4.5
	1101:0303:15010101g:1301	0.00358	4.0
	010101g:0602:1302:0701	0.00358	4.0
	02010101g:0202:270502g:0101	0.00358	4.0
	010101g:0602:3502:1104	0.00358	4.0
	6802:0802:1402:1303	0.00358	4.0
	010101g:0602:5701:0701	0.00354	4.0
	02010101g:0303:15010101g:1301	0.00343	3.8
	03010101g:0702:0702:0701	0.00315	3.5
	1101:0501g:44020101g:0401	0.00314	3.5
	02010101g:0602:1302:1501	0.00282	3.1
	010101g:070101g:0801g:0701	0.00278	3.1
	02010101g:070101g:0801g:0401	0.00269	3.0
	010101g:0102:270502g:0301	0.00269	3.0
	680102g:070401g:44020101g:1302	0.00269	3.0
	010101g:1601:4403:0701	0.00269	3.0
	03010101g:04010101g:3503:0701	0.00269	3.0
	24020101g:0602:1302:0701	0.00269	3.0
	03010101g:0802:1402:0102	0.00269	3.0
	24020101g:0202:4405:0101	0.00269	3.0
	03010101g:0303:15010101g:0401	0.00269	3.0
	02010101g:0303:15010101g:1101	0.00269	3.0
	3002:0501g:180101g:0301	0.00269	3.0
	03010101g:0702:0702:0401	0.00269	3.0
	24020101g:070101g:180101g:1101	0.00269	3.0
	2601:070401g:44020101g:1601	0.00269	3.0
	24020101g:0501g:180101g:0301	0.00269	3.0
	680102g:0102:270502g:1301	0.00269	3.0
	6601:1701g:4102:1303	0.00269	3.0
	Total	0.43091	480.8

Table 6Pairwise global linkage disequilibrium estimates^a

Locus Pair	D'	Wn	p-value
A:C	0.47558	0.33870	<0.0001
A:B	0.57324	0.40388	<0.0001
A:DRB1	0.41025	0.31802	<0.0001
C:B	0.91160	0.72619	<0.0001
C:DRB1	0.52904	0.36512	<0.0001
B:DRB1	0.63355	0.46386	<0.0001

^aFor D', values above 0 show a positive association between loci; values above 0.5 indicate a very strong association. Wn is an alternative measurement of linkage disequilibrium which is interpreted in a similar fashion to D'.

Table 7

Pairwise F_{st} values for HLA loci comparing Eastern European Americans to other European and European American populations

Population (reference)	A:C:B Haplotypes	p-value ^a	DRB1 Alleles	p-value ^a
Poland (39)	0.56140 ^b	0.00000 *	0.00722	0.18359
Poland (40)			0.03986	0.00586 *
Czech Republic (10)	0.011018	0.17117	0.00000	0.61230
Czech Republic (31)			0.00000	0.82324
Slovenia (10)			0.00000	0.66895
Slovenia (44)			0.00000	0.59473
Russia (43)			0.00159	0.37402
Russia (42)			0.08583	0.00391 *
Bulgaria (28)			0.18488	0.00098 *
Croatia (30)			0.16865	0.01855 *
Croatia (29)			0.14136	0.03906 *
Georgia (10)	0.125765	0.00000 *		
Georgia (32)			0.10337	0.00293 *
Macedonia (35)			0.15017	0.00000 *
Macedonia (36)			0.14089	0.00000 *
Finland (10)	0.089037	0.00000 *	0.13423	0.00000 *
Norway (37)			0.04690	0.00391 *
Norway (38)			0.02035	0.166113
Sweden (49)			0.04840	0.00586 *
Germany (33)			0.00000	0.94922
N. Ireland (10)	0.086255	0.00000 *	0.07945	0.00000 *
Spain (45) ^c			0.38327	0.00000 *
Spain (46)			0.00000	0.44824
Spain (45) ^d			0.31151	0.00000 *
Spain (47)			0.04005	0.01855 *
Spain (48)			0.03452	0.00000 *
Portugal (41)			0.00694	0.26172
Italy (34)			0.35147	0.00000 *
USA (54)			0.01595	0.16699
USA (55)			0.02054	0.00195 *
USA (56) ^e			0.00000	0.71875
USA (53)	0.012799	0.01802 *		

* p<0.05, F_{st} values are significantly different

^a p values for unstandardized F_{st} values

^bBecause the A:C:B haplotypes for this population included a "blank" haplotype with a frequency of 0.22, the F_{st} value for this population is interpreted differently than others. See the Results section for details.

^cThis is the Cabuernigo population described by Sanchez-Velasco et al. (45).

^dThis is the Pasiego population described by Sanchez-Velasco et al. (45).

^eThis is the White population described by Rossman et al. (56).