Published in final edited form as:
Hum Immunol. 2011 February ; 72(2): 144–149. doi:10.1016/j.humimm.2010.10.017.

Human Leukocyte Antigens-A, -B, -C, -DRB1 allele and haplotype frequencies in Americans originating from Southern Europe: Contrasting patterns of population differentiation between Italian and Spanish Americans

Steven J. Mack[1], **Bin Tu**[2], **Ruyan Yang**[2], **Carly Masaberg**[2], **Jennifer Ng**[2], and **Carolyn Katovich Hurley**[2]

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

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

Abstract

High resolution DNA sequencing was used to identify the HLA-A, -B, -C, and -DRB1 alleles found in 552 individuals from the United States indicating Southern European (Italian or Spanish) heritage. A total of 46 HLA-A, 80 HLA-B, 32 HLA-C, and 50 DRB1 alleles were identified. Frequent alleles included A*02:01:01G (allele frequency = 0.26 in Italian Americans; 0.22 in Spanish Americans); B*07:02:01G (Italian Americans allele frequency = 0.11); B*44:03 (Spanish Americans allele frequency = 0.07); C*04:01:01G and C*07:01:01G (allele frequency = 0.13 and 0.16, respectively, in Italian Americans; 0.15 and 0.12, respectively, in Spanish Americans); and DRB1*07:01:01 (allele frequency = 0.12 in each population). The action of balancing selection was inferred at the HLA-B and -C loci in both populations. The A*01:01:01G-C*07:01:01G-B*08:01:01G-DRB1*03:01:01 haplotype was the most frequent A-C-B-DRB1 haplotype in Italian Americans (haplotype frequency = 0.049), and was the second most frequent haplotype in Spanish Americans (haplotype frequency = 0.021). A*29:02:01-C*16:01:01-B*44:03-DRB1*07:01:01 was the most frequent A-C-B-DRB1 haplotype in Spanish Americans (haplotype frequency = 0.023), and was observed at a frequency of 0.015 in Italian Americans. Pairwise F'_{st} values measuring the degree of differentiation between these Southern European-American populations and European and European-American populations suggest that Spanish Americans constitute a distinct subset of the European-American population, most similar to Mexican Americans, whereas Italian Americans cannot be distinguished from the larger European-American population.

Keywords

HLA; population study

Introduction

Beginning with their discovery by an Italian sailor with financial support from Spain in 1492, the Americas have had a long history with these two Mediterranean countries. The

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.

Spanish colonization of the southwest and western regions of what is now the United States, which started soon after Columbus' discovery of the New World, ended in the 1800s, while the major wave of Italian immigration to the U.S. occurred more recently, in the late 1800s. Today, 6.5% of the US population is of Italian (5.6%) or Spanish (0.9%) ancestry [1]. Conversely, Spanish and Italian are the second and tenth most common languages spoken in the United States; 9.3% of the U.S. population speaks Spanish, whereas 0.4% of the U.S. population speaks Italian [1]. Here, we present HLA-A, -B, -C, and -DRB1 alleles identified in a population of European-Americans self-described as having Italian or Spanish ancestry, and use these data to estimate multi-locus haplotypes and the degree of differentiation between these populations and other US populations originating from other regions of Europe, as well as European populations.

Subjects and Methods

Sample population

The study population included 552 individuals from the United States indicating Italian or Spanish ancestry who were consecutively recruited as volunteer donors for a bone marrow donor registry from August 2007 through March 2008. Of these 552, 273 self identified as having Italian ancestry and 279 as having Spanish ancestry. Because of the varied recruitment sites and the setting, individuals are unlikely to be related and are likely to originate from different areas of the U.S. However, because these individuals are part of the larger U.S. population, genetic contributions from European or non-European populations cannot be excluded simply on the basis of self-identification.

Identification of known HLA alleles

HLA alleles were isolated and characterized by DNA sequencing as previously described [2]. Sequence interpretation was based on ImMunoGeneTics (IMGT)/HLA database release 2.21 [3]. 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. 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 [2]. (In-house primer sequences used for all loci are available at http://www.dodmarrow.org.) The genotype data generated via this method can be found in Supplemental Table S1.

Statistical analysis

PyPop (Python for Population genomics, version 0.7.0, available at http://www.pypop.org) was used to carry out Hardy-Weinberg testing, Ewens-Watterson homozygosity analyses, and haplotype and linkage disequilibrium (LD) estimates [4,5]. Allele frequencies were determined by direct counting. Allele Frequency data generated via this method can be found in Supplemental Table S2. Allele frequencies at each HLA locus were evaluated for deviations from Hardy-Weinberg equilibrium proportions using the exact test of Guo and Thompson [6], 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 heterozygous 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 [7].

The Ewens-Watterson test of homozygosity was applied to each locus [8,9], using Slatkin's Monte-Carlo implementation of the exact test [10,11]. 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 (2n) with the same number of alleles (k), 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 [12]. This normalization permits the direct comparison of homozygosity values for different loci in the same population, and for populations with different values of 2n and k. F_{nd} values significantly less than 0 are consistent with the action of balancing selection, resulting from allele frequency distributions that are significantly more "even", rather than skewed toward specific alleles. Because the null-hypothesis of the Ewens-Watterson test is neutral evolution ($F_{nd} = 0$), we used a paired sign test [13] to compare the signs of the F_{nd} values for each population and locus against the expectation of neutrality. To correct for the number of comparisons, the results of these tests were considered significant if the associated p-values were lower than 0.0056.

Two-, three-, and four-locus haplotype frequencies were estimated using the iterative expectation-maximization (EM) algorithm [14,15]. Linkage disequilibrium between alleles at each pair of loci, and two overall (locus-pair-level) measures of linkage disequilibrium, normalized to values between 0 and 1, 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 elsewhere [15–18]), and ranges from +1 to −1. A D'ij value of 0 indicates linkage equilibrium, whereas 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 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′ [16]uses the products of the allele frequencies at each locus to weight the LD contribution of specific allele pairs; whereas the second, Wn [19], 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 [15]. LD between any pair of loci was considered significant if the associated p-value was less than 0.00416 (corrected for multiple comparisons). Haplotype frequencies and associated allele-pair-level LD values (D and D'ij) are included in Supplemental Tables S3A, S3B, S4, and S5.

Comparison of populations was limited by the availability of allele-level data for HLA A-C-B haplotypes reported in the literature for other populations. Arlequin v3.11 [20] 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 [7,21], Mali [21], Rwanda [7], Senegal [7], South Africa [7], Uganda [21], Zambia [21], and Zimbabwe [7]; North African populations from Morocco [22–25] and Algiers [22]; European populations from Bulgaria [7] [26], Croatia [27], the Czech Republic [7,28], Finland [7], Georgia [7,29], Germany [30], Italy [31−34], Macedonia [35,36], Northern Ireland [7], Norway [37,38], Poland [39,40], Portugal [41], Russia[42], Slovenia [7,43], Spain [40,44–47], and Sweden [48]; Asian populations from India [49], and Israel [50]; two African American populations [2,51], five European American populations [52–54] [51,55], and a Mexican American population [56] by calculating pairwise F_{st} values (and associated p-values), and an exact test of population differentiation [57] for this entire set of populations using DRB1 allele frequencies and A-C-B haplotype frequencies. These tests evaluate population differences using different null-hypotheses; p-values for pairwise F_{st} values are generated under the assumption that there is

no difference in allele frequency between population samples, whereas the exact test assumes random mating (panmixia) between population samples. Given that HLA allele and haplotype data are characterized by large numbers of low-frequency alleles and haplotypes, it is possible that two population samples drawn randomly from the same population may more often appear significantly different under the F_{st} test than under the exact test. In accounting for small differences in population sample sizes, the F_{st} calculation [58] may result in small negative F_{st} values; these were treated as being equivalent to 0 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 [59]. Because all populations had not been genotyped at the same loci or for the same level of resolution, these comparisons were performed and the analysis focused on the amino acid sequences encoding the polymorphic antigen binding groove using the binning approach described by Solberg et al. [60]. After accounting for the number of comparisons, a given pair of population datasets was determined to differ significantly if the appropriate p-value associated was less than 0.00069.

Results

Allele and genotype frequencies

Allele frequencies for the HLA-A, -B, -C, and -DRB1 loci are shown in Table S2. A total of 272 unique A-C-B-DRB1 phenotypes (based on 4-digit allele names) were observed among the 273 Italian American individuals examined; 277 in the 279 Spanish American individuals. When the two populations were combined, significant deviations from expected Hardy-Weinberg equilibrium proportions were observed for HLA-B (p=0.00396) and DRB1 (p=0.03846) (data not shown). No overall deviations were observed at the HLA-A, -C, or -B loci for the individual populations although a minor but significant deviation from expected HWEP was observed at the DRB1 locus (p-value = 0.02683) for the Spanish American population. 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. Based this analysis, the two populations were analyzed separately.

For the Italian American population, 29 HLA-A alleles, 49 HLA-B, 25 HLA-C, and 40 DRB1 alleles were identified (with an allele defined as either a unique class I exon 2–3 sequence or a unique DRB1 exon 2 sequence). Spanish Americans showed a higher number of alleles: 42 HLA-A, 73 HLA-B, 29 HLA-C, and 48 HLA-DRB1. All alleles had been previously identified and were found in a list of common and well-documented (CWD) alleles [61] except for A*02:44, B*35:77, B*44:30, C*12:03:03, C*12:12; C*14:06, and DRB1*04:05:04. All of these alleles were found in single individuals in this study. The sequences of one HLA-A (A*25:01:01) and five HLA-C (C*12:03:03, 12:05, 12:12, 14:06, 15:05:01G) alleles had been observed in only single individuals reported to the IMGT/HLA database at the time of this study (i.e., not confirmed). Although three of the non-CWD or unconfirmed alleles were identified previously in individuals whose ethnicity is unknown (B*35:77, C*12:03:03, DRB1*04:05:04), the remaining alleles were identified in individuals with European, Spanish, or Hispanic backgrounds. All of the non-CWD and unconfirmed four digit HLA-C alleles had been previously observed in a study of the National Marrow Donor Program registry [62]. Comparisons of the HLA assignments of individuals in our study with those in the IMGT/HLA database have identified potential haplotypes that carry three "rare" alleles. Based on the assignments of three individuals carrying A*02:44 (two in IMGT/HLA), the allele is apparently associated with B*40:02:01G, DRB1*08:02:01. B*35:77 is found with A*02:01:01G, C*04:01:01G in the four individuals identified with this allele and with DRB1*08:01:01G in three of these individuals. B*44:30 is associated with A*29:02:01, DRB1*07:01:01 in three individuals characterized.

In Italian Americans, three HLA-A alleles (A*02:01:01G, *01:01:01G, *03:01:01G) were observed at frequencies greater than 0.1, and represented 53% of the allelic diversity observed at this locus. Most notably, the A*02:01:01G allele was observed at a very high frequency of 0.260. A single HLA-B allele, B*07:02:01G (0.108), and four HLA-C alleles (C*07:01:01G, *04:01:01G, *07:02:01G, *06:02:01G) were observed with frequencies above 0.1. These HLA-C alleles represented 50% of the HLA-C diversity in this population. Three DRB1 alleles (DRB1*07:01:01, *03:01:01, and*15:01:01) with frequencies greater than 0.1 represented 34% of the allelic diversity observed at this locus.

In Spanish Americans, two HLA-A alleles (A*02:01:01G, *24:02:01G) were observed at frequencies greater than 0.1, and represented 32% of the allelic diversity observed at this locus. Most notably, the A*02:01:01G allele was observed at a very high frequency of 0.215. No HLA-B alleles were present at frequencies over 0.10. Three HLA-C alleles (C*04:01:01G, *07:01:01G, *07:02:01G) were observed with frequencies above 0.1. These HLA-C alleles represented 38% of the HLA-C diversity in this population. One DRB1 allele (DRB1*07:01:01) was present with a frequency of greater than 0.1, and represented 12% of the allelic diversity observed at this locus.

Several alleles in the Spanish American population (HLA-A*02:04, *02:06:01G, *02:13, *02:17:01G, *02:22G, *68:03:01 and *68:05, HLA-C*03:05, HLA-B*15:15, *3504:01, *35:05, *35:12, *35:17, *35:43G, *39:02:02, *39:03 and *48:01:01, and DRB1*04:07:01, *04:11, *08:02:01, *08:07, *14:02, *14:06, and *16:02:01) were detected at low to intermediate frequencies, or were not detected in the Italian American population, and have been observed predominantly or exclusively in Native American populations [32,60]. In addition, HLA-A*23:05, *36:01 and *66:02, *C*18:01G, B*15:16, *53:01:01, *57:02, *57:03:01, *58:02, *78:01 and *81:01G, and DRB1*03:02:01 and *15:03, were also detected at low to intermediate frequencies in the Spanish American population, or were not detected in the Italian population; these alleles have been observed predominantly or exclusively in Sub-Saharan African populations [32,60]. Together, these Native American and African alleles account for approximately 12% of the HLA diversity in this population. Although some of these alleles were also detected in the Italian American population, they were observed at much lower frequencies. In addition, very few alleles were detected exclusively in the Italian American population, and of those that were, only A*02:44, and DRB1*12:02 have been observed predominantly in non-European populations (from East Asia and the Pacific). This pattern suggests that the Spanish American population is considerably admixed, in a manner characteristic of Hispanic populations.

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 results of the Ewens-Watterson Homozygosity Test are shown in Table 1. Negative F_{nd} values were observed for all four loci in both populations, and while significantly low values were observed for HLA-B and HLA-C for both populations and DRB1 for the Spanish American population, no F_{nd} values were significant after correction for multiple comparisons. However, the application of a sign test to all eight F_{nd} values reveals a significant ($p = 0.0047$) overall negative deviation from the expectation of neutral evolution ($F_{nd} = 0$), suggesting the action of balancing selection in shaping allelic diversity at all four loci in both populations.

Haplotypes

Two-locus (A-B, C-B, and B-DRB1), three-locus (A-B-DRB1) and four-locus haplotypes (at the 4 digit allele level) were estimated for these populations. Two-locus haplotypes are

presented in Tables S3(A) and S3(B); three- and four-locus haplotypes in Tables S4 and S5. Because the outcome of the EM algorithm is unreliable for rare haplotypes (n = 1 or 2), only those haplotypes with at least three copies should be considered reliable for making inferences. As shown in Table 2, global linkage disequilibrium (LD) estimates of the associations between HLA loci show significant LD as previously observed by others. In general, for Italian Americans, nine of the ten most frequent four-locus haplotypes are among the 15 most common European-American haplotypes [62]; whereas seven of the ten most frequent Spanish American four-locus haplotypes are found. For Italian Americans, 25 four-locus haplotypes were identified at least three times and accounted for almost 30% (haplotype frequency = 0.2963) of the A-C-B-DRB1 haplotypes (Table S5). For Spanish Americans, 21 four locus haplotypes were identified at least three times and accounted for almost 19% (haplotype frequency = 0.18613) of the A-C-B-DRB1 haplotypes (Table S5). In summary, for the very common haplotypes, these populations are similar to a European-American population, although the Spanish American population is less similar than the Italian population [62].

Differentiation of populations

The DRB1 allele frequencies and A-C-B haplotype frequencies of the Italian American and Spanish American populations were compared to available DRB1 allele and A-C-B haplotype frequencies in a set of African, European, Southwest Asian, African American, European-American and Mexican American populations by calculating pairwise F'_{st} values and the exact test of population differentiation. F'_{st} comparisons between the Italian American and Spanish American populations and African and southwest Asian populations indicated significant differentiation between these groups; therefore, only the results of these comparisons to 31 European and six Europe-derived populations are shown in Table 3.

The Italian American and Spanish American populations were shown to be significantly differentiated from each other via F'_{st} values calculated for DRB1 frequencies and A-C-B haplotype frequencies, as well as by the exact test of differentiation using DRB1 allele-frequencies. The exact test of differentiation did not reveal significant differentiation of the two populations through the comparison of A-C-B haplotype frequencies.

F'_{st} values calculated for DRB1 allele frequencies reveal Spanish Americans to be significantly differentiated from 23 European populations (including five Spanish populations); Spanish Americans were not significantly differentiated from populations from the Czech Republic [28], Germany (33), Slovenia (7), and Portugal (44), and two populations from Madrid, Spain [40]. Spanish Americans are also significantly differentiated from European-American populations, but could not be distinguished from Mexican Americans. These DRB1 results were largely consistent with those of the exact test for population differentiation; Spanish Americans were significantly differentiated from twenty-two European populations, but were not significantly differentiated from two populations from Madrid, Spain, two from the Czech Republic, two from Croatia, and one each from Germany and Slovenia. Spanish Americans were significantly differentiated from all European-American populations, but not from Mexican Americans.

F'_{st} values calculated using A-C-B haplotypes distinguished Spanish Americans from all populations, including Mexican Americans, whereas the exact test for population differentiation revealed Spanish Americans to be significantly differentiated from the Northern Ireland population alone using A-C-B haplotype frequencies. This discrepancy between the result of the two tests likely reflects the broad spectrum of low-frequency A-C-B haplotypes (observed fewer than three times) in the populations tested; for example, 62% of Northern Ireland, 78% of Mexican American, 79% of Italian American, and 86% of Spanish American A-C-B haplotypes are seen fewer than three times (data not shown).

The overall degree of differentiation of Italian Americans from European and European-American populations was less extensive than that for Spanish Americans. F'st values calculated for DRB1 allele frequencies indicated that Italian Americans were not significantly differentiated from 20 of 30 European populations (five from Spain, two from Slovenia, two from the Czech Republic, two from Poland, two from Croatia, two from Norway, and populations from Portugal, Russia, Georgia, Sweden and Germany), although they were significantly differentiated from all three Italian populations (all p-values < 0.00001). Italian Americans could not be differentiated from European-Americans, although they were significantly differentiated from Mexican Americans. F'st values calculated for A-C-B haplotype frequencies revealed significant differentiation of Italian Americans from Georgia, Finland, Northern Ireland and Mexican American populations, but not from a Czech Republic, or two European-American populations. These results were consistent with those of the exact test for population differentiation, where Italian Americans were not significantly differentiated from 25 of 30 European populations, two of which were Italian populations [34], and were not significantly differentiated from European-American populations using DRB1 allele frequencies. As with Spanish Americans, the exact test revealed significant differentiation of Italian Americans from the Northern Ireland population via A-C-B haplotype frequencies.

Discussion

Immigrants from Europe comprise the major population of the United States. Data from the US Census demonstrate the diversity of countries represented in the United States, with some areas of Europe contributing substantially more immigrants that others [1]. The goal of our studies is to evaluate the similarity and differences in the HLA profiles of populations derived from various regions of Europe in order to better define subgroups of individuals who have a similar genetic background. These data will allow a better understanding of the frequencies of HLA alleles and haplotypes which, in turn, will strengthen the ability of HLA matching algorithms to predict allele matched donors for patients requiring a hematopoietic progenitor cell transplant. In addition, a clearer understanding of sub-population structure and differentiation within the greater European American population can facilitate the selection of control populations for case-control, disease-association and anthropology studies.

Our analyses suggest that Americans of self-defined Spanish ancestry constitute a subset of the European American population that is distinct from other European American populations, from non-Spanish European populations, and also from regionally isolated Spanish populations. In contrast, Americans of self-defined Italian ancestry appear similar to the general European American population, and to the population of Europe to a lesser extent. The Italian American population cannot be distinguished from the Eastern European American population, a subgroup we have previously shown to constitute a distinct subset of the European-American population [52]. Although the presence of a Spanish population in the US predates the major period of Italian immigration, the Spanish American group appears to be most similar to the population of Madrid, Spain, and to Mexican Americans, and demonstrates African and Native American admixture characteristic of other Hispanic populations. Given the large-size of the Spanish-speaking segment of the U.S. population, it seems possible that the self-identified Spanish ethnicity refers more to spoken language than to a comprehensive account of ancestral national identity. Similarly, while the self-identified Italian ethnicity may reflect some aspect of ancestry, Italian immigrants to the U.S. appear to have been assimilated into the general population. These contrasting patterns of population differentiation underscore the importance of understanding the history of linguistic and cultural assimilation in the U.S. for interpreting the significance of self-defined ethnicity.

Those in the U.S. who self-identify as being of Spanish or Italian ethnicity should consider the prospect that they enjoy an even richer family history.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research is supported by funding from the Office of Naval Research N00014-07-1-1108 and N00014-08-1-1078 (C.K.H., J.N.), and with the support of NIH grant U01AI067068 (S.J.M.). 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.

Abbreviations

HLA	human leukocyte antigen
a.f	allele frequency
h.f	haplotype frequency
U.S	United States of America
No	number

References

1. Brittingham, A.; de la Cruz, GC. 2000 Census Brief. US Census Bureau; 2004. Ancestry: 2000.
2. 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]
3. Robinson J, Waller MJ, Parham P, de Groot N, Bontrop R, Kennedy LJ, Stoehr 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]
4. 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.
5. 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]
6. Guo SW, Thompson EA. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 1992;48:361–372. [PubMed: 1637966]
7. 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.
8. Ewens W. The sampling theory of selectively neutral alleles. Theor Pop Biol 1972;3:87–112. [PubMed: 4667078]
9. Watterson G. The homozygosity test of neutrality. Genetics 1978;88:405–417. [PubMed: 17248803]
10. Slatkin M. An exact test for neutrality based on the Ewens sampling distribution. Genetical Research 1994;64:71–74. [PubMed: 7958833]
11. Slatkin M. A correction to the exact test based on the Ewens sampling distribution. Genetical Research 1996;68:259–260. [PubMed: 9062082]

12. 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]
13. Conover, W. Practical Nonparametric Statistics. New York: John Wiley and Sons; 1980.
14. Dempster A, Laird N, Rubin D. Maximum likelihood estimation from incomplete data using the EM algorithm. J Royal Stat Soc 1977;39:1–38.
15. Excoffier L, Slatkin M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. Mol Biol Evol 1995;12:921–927. [PubMed: 7476138]
16. Hedrick PW. Gametic disequilibrium measures. proceed with caution. Genetics 1987;117:331–334. [PubMed: 3666445]
17. Lewontin RC. The interaction of selection and linkage. II. Optimum models. Genetics 1964;50:757–782. [PubMed: 14221879]
18. Lewontin RC. On measures of gametic disequilibrium. Genetics 1988;120:849–852. [PubMed: 3224810]
19. Cramer, H. Mathematical methods of statistics. Princeton, NJ: University Press; 1946.
20. 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.
21. 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]
22. 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.
23. 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.
24. 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.
25. 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.
26. 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]
27. Martinovic I, Bakran M, Chaventre A, Janicijevic B, Jovanovic V, Smolej-Narancic N, Kastelan A, Grubic 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]
28. Cerna M, Fernandez-Vina M, Ivaskova 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]
29. Sanchez-Velasco P, Leyva-Cobian 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]
30. 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]

31. Lampis R, Morelli L, Congia M, Macis MD, 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. Hum Mol Genet 2000;9:2959–2965. [PubMed: 11115839]
32. Middleton, D.; Menchaca, L.; Rood, H.; Komerofsky, R. New allele frequency database; Tissue Antigens. 2003. p. 403-407.http://www.allelefrequencies.net
33. Adorno, D.; Piancatelli, D.; Canossi, A.; Aureli, A.; Papola, F. Caucasian Italian normal. In: Gjertson, DW.; Terasaki, P., editors. HLA 1998. American Society for Histocompatibility and Immunogenetics; 1998.
34. Ferrara, GB.; Delfino, L.; Longo, A.; Morabito, A.; Pera, C. Caucasian Italian normal. In: Gjertson, DW.; Terasaki, P., editors. HLA 1998. American Society for Histocompatibility and Immunogenetics; 1998.
35. Arnaiz-Villena A, Dimitroski K, Pacho A, Moscoso J, Gomez-Casado E, Silvera-Redondo C, Varela P, Blagoevska M, Zdravkovska V, Martinez-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. Ronningen 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-Koros 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. In: Charron, D., editor. Report of the anthropology component; HLA: Proceedings of the Twelfth International Histocompatibility Workshop and Conference; Sevres, EDK; 1997.
41. Spinola 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. 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]
43. Petrone A, Battelino T, Krzisnik C, Bugawan T, Erlich H, Di Mario U, 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]
44. Sanchez-Velasco P, Gomez-Casado E, Martinez-Laso J, Moscoso J, Zamora J, Lowy E, Silvera C, Cemborain A, Leyva-Cobian F, Arnaiz-Villena A. HLA alleles in isolated populations from North Spain: origin of the Basques and the ancient Iberians. Tissue Antigens 2003;61:384–392. [PubMed: 12753657]
45. Sanchez-Velasco P, Escribano dD, Paz-Miguel JE, Ocejo-Vinyals G, Leyva-Cobian F. HLA-DR, DQ nucleotide sequence polymorphisms in the Pasiegos (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]
46. Vidal S, Morante MP, Moga E, Mosquera AM, Querol S, Garcia J, Rodriguez-Sanchez J. Molecular analysis of HLA-DRB1 polymorphism in north-east Spain. Eur J Immunogenet 2002;29:75–77. [PubMed: 11841495]
47. Muro M, Marin L, Torio A, Moya-Quiles MR, Minguela A, Rosique-Roman J, Sanchis MJ, Garcia-Calatayud MC, Garcia-Alonso AM, Alvarez-Lopez MR. HLA polymorphism in the

Murcia population (Spain): in the cradle of the archaeologic Iberians. Hum Immunol 2001;62:910–921. [PubMed: 11543893]

48. 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]
49. 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.
50. 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.
51. 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]
52. Mack SJ, Tu B, Lazaro A, Yang R, Lancaster AK, Cao K, Ng J, Hurley CK. HLA-A, -B, -C, and -DRB1 allele and haplotype frequencies distinguish Eastern European Americans from the general European American population. Tissue Antigens 2009;73:17–32. [PubMed: 19000140]
53. 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]
54. Klitz W, Maiers M, Spellman S, Baxter-Lowe LA, Schmeckpeper B, Williams TM, Fernandez-Vina M. 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]
55. 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]
56. Klitz W, Gragert L, Maiers M, Tu B, Lazaro A, Yang R, Xu Q, Masaberg C, Ng J, Hurley CK. Four-locus high-resolution HLA typing in a sample of Mexican Americans. Tissue Antigens 2009;74:508–513. [PubMed: 19845916]
57. Raymond M, Rousset F. An exact test for population differentiation. Evolution 2010;49:1280–1283.
58. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution 2010;38:1358–1370.
59. Hedrick PW. A standardized genetic differentiation measure. Evolution 2005;59:1633–1638. [PubMed: 16329237]
60. Solberg OD, Mack SJ, Lancaster AK, Single RM, Tsai Y, Sanchez-Mazas A, Thomson G. Balancing selection and heterogeneity across the classical human leukocyte antigen loci: a meta-analytic review of 497 population studies. Hum Immunol 2008;69:443–464. [PubMed: 18638659]
61. 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 histocompatiblity and immunogenetics. Hum Immunol 2007;68:392–417. [PubMed: 17462507]
62. 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]

Table 1

Ewens-Watterson Homozygosity Test

Population	2n	Locus	k	F_{obs}	F_{nd}	p-value
Italian American	546	A	28	0.1223	−0.3314	0.449
		B	48	0.0493	−1.1795	0.040
		C	24	0.0882	−1.2602	0.016
		DRB1	40	0.0671	−0.8851	0.138
Spanish American	558	A	41	0.0856	−0.1674	0.530
		B	70	0.0327	−1.2702	0.031
		C	29	0.0768	−1.2303	0.023
		DRB1	46	0.0507	−1.2117	0.031

Table 2

Global measures of linkage disequilibrium at the HLA-A, -B, -C, and -DRB1 loci in Italian and Spanish Americans.

	Italian American			Spanish American		
Locus Pair	**D′**	**Wn**	**p-value**	**D′**	**Wn**	**p-value**
A–B	0.64	0.48	< 0.00001	0.71	0.47	< 0.00001
A–C	0.56	0.45	< 0.00001	0.56	0.37	< 0.00001
A-DRB1	0.50	0.43	< 0.00001	0.61	0.39	< 0.00001
C-B	0.94	0.75	< 0.00001	0.90	0.75	< 0.00001
B-DRB1	0.70	0.51	< 0.00001	0.80	0.50	< 0.00001
C-DRB1	0.60	0.43	< 0.00001	0.64	0.37	< 0.00001

Table 3

Pairwise measures of population differentiation between Italian and Spanish Americans and other European and European-American populations.

	DRB1						A-C-B haplotypes					
	Italian American			Spanish American			Italian American			Spanish American		
	F'st	Fst p-value	Exact Test p-value	F'st	Fst p-value	Exact Test p-value	F'st	Fst p-value	Exact Test p-value	F'st	Fst p-value	Exact Test p-value
Slovenia [7]	0.000	0.487	0.373	0.063	0.0091	0.0015						
Czech Republic [7]	0.000	0.622	0.799	0.060	< 0.00001*	0.0026	0.018	0.073	0.613	0.079	< 0.00001*	0.577
Georgia [7]							0.145	< 0.00001*	0.015	0.124	< 0.00001*	0.491
United States-East European [52]	0.000	0.892	0.98	0.047	< 0.00001*	< 0.00001*	0.000	0.46	0.056	0.050	< 0.00001*	0.001
Finland [7]	0.155	< 0.00001*	0.001	0.227	< 0.00001*	< 0.00001*	0.119	< 0.00001*	0.068	0.191	< 0.00001*	0.004
Northern Ireland [7]	0.063	< 0.00001*	< 0.00001*	0.154	< 0.00001*	< 0.00001*	0.068	< 0.00001*	< 0.00001*	0.185	< 0.00001*	< 0.00001*
Poland [39]	0.005	0.244	0.005	0.066	< 0.00001*	< 0.00001*						
Poland [40]	0.041	0.0288	< 0.00001*	0.088	< 0.00001*	< 0.00001*						
Croatia [27][a]	0.171	0.0091	0.014	0.176	< 0.00001*	0.0087						
Macedonia [35]	0.137	< 0.00001*	< 0.00001*	0.199	< 0.00001*	< 0.00001*						
Georgia [29]	0.060	0.01	< 0.00001*	0.089	< 0.00001*	< 0.00001*						
Croatia [27][b]	0.126	0.028	0.081	0.183	< 0.00001*	0.0354						
Bulgaria [26]	0.184	< 0.00001*	0.004	0.219	< 0.00001*	< 0.00001*						
Czech Republic [28]	0.000	0.784	0.905	0.037	0.055	0.0102						
Macedonia[36]	0.143	< 0.00001*	< 0.00001*	0.180	< 0.00001*	< 0.00001*						
Norway [37]	0.036	0.028	0.001	0.133	< 0.00001*	< 0.00001*						
Norway [38]	0.014	0.217	0.255	0.088	< 0.00001*	0.00066*						
Portugal [41]	0.000	0.577	0.387	0.024	0.037	0.0004*						
Russia [42]	0.000	0.613	0.015	0.048	< 0.00001*	< 0.00001*						
Slovenia [43]	0.000	0.748	0.771	0.037	0.028	0.0001*						
Sweden [48]	0.043	0.037	0.013	0.140	< 0.00001*	< 0.00001*						
Germany [30]	0.000	0.982	0.986	0.039	0.037	0.0111						

Hum Immunol. Author manuscript; available in PMC 2011 August 1.

	DRB1						A-C-B haplotypes					
	Italian American			Spanish American			Italian American			Spanish American		
	F'st	Fst p-value	Exact Test p-value	F'st	Fst p-value	Exact Test p-value	F'st	Fst p-value	Exact Test p-value	F'st	Fst p-value	Exact Test p-value
United States [51][c]							0.000	0.505	0.758	0.077	< 0.00001*	0.013
United States [53]	0.000	0.541	0.592	0.083	< 0.00001*	0.0001*						
United States [54]	0.011	0.0991	0.004	0.099	< 0.00001*	< 0.00001*						
United States [55]	0.000	0.739	0.892	0.049	< 0.00001*	< 0.00001*						
Italy [31]	0.299	< 0.00001*	< 0.00001*	0.344	< 0.00001*	< 0.00001*						
Italy [32,33]	0.061	< 0.00001*	0.001	0.104	< 0.00001*	< 0.00001*						
Italy [32,34]	0.059	< 0.00001*	0.204	0.132	< 0.00001*	< 0.00001*						
Spain [44]	0.326	< 0.00001*	< 0.00001*	0.442	< 0.00001*	< 0.00001*						
Spain [45][d]	0.012	0.163	0.02	0.066	< 0.00001*	< 0.00001*						
Spain [45][e]	0.279	< 0.00001*	< 0.00001*	0.371	< 0.00001*	< 0.00001*						
Spain[47]	0.037	0.037	0.032	0.055	< 0.00001*	< 0.00001*						
Spain [46]	0.015	0.019	0.144	0.040	< 0.00001*	< 0.00001*						
Spain [40][f]	0.001	0.397	0.059	0.025	0.082	0.046						
Spain [40][g]	0.000	0.928	0.954	0.026	0.091	0.233						
United States-Mexican American [56]	0.107	< 0.00001*	< 0.00001*	0.005	0.19	0.8717	0.087	< 0.00001*	0.61	0.018	< 0.00001*	0.319
United States–Italian American[h]				0.036	< 0.00001*	< 0.00001*				0.043	< 0.00001*	0.466

* p-values < 0.00069 (adjusted for multiple comparisons) indicate that populations differ significantly.

[a] This is the Dubasnica population

[b] This is the Vrbnik population

[c] This is the Caucasian population.

[d] This is the Cantabrian population.

[e] This is the Pasiego population.

[f] This is population number 104.

[g] This is population number 82.

[h] This is the Italian American population reported here