



Human leukocyte antigen–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 ^{a,*}, Bin Tu ^b, Ruyan Yang ^b, Carly Masaberg ^b, Jennifer Ng ^b, Carolyn Katovich Hurley ^b

^a Children's Hospital Oakland Research Institute, Oakland, California, USA

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

ARTICLE INFO

Article history:

Received 13 March 2010

Accepted 19 October 2010

Available online 23 October 2010

Keywords:

Human leukocyte antigen

Population study

European American

Italian American

Spanish American

ABSTRACT

High-resolution DNA sequencing was used to identify the human leukocyte antigen (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 and 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 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 as well as 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.

© 2011 American Society for Histocompatibility and Immunogenetics. Published by Elsevier Inc. All rights reserved.

1. 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 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, whereas the major wave of Italian immigration to the United States 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 US population speaks Spanish, whereas 0.4% of the US 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 we use these data to estimate multi-locus haplotypes and the degree of differentiation between these pop-

ulations and other US populations originating from other regions of Europe, as well as European populations.

2. Subjects and methods

2.1. 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 individuals, 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 United States. However, because these individuals are part of the larger US population, genetic contributions from European or non-European populations cannot be excluded simply on the basis of self-identification.

2.2. Identification of known HLA alleles

HLA alleles were isolated and characterized by DNA sequencing as previously described [2]. Sequence interpretation was based on

* Corresponding author.
E-mail address: sjmack@chori.org (S.J. Mack).

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 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.

2.3. 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 χ^2 testing when expected values were ≥ 5 . The χ^2 tests were investigated for overall common genotypes (those expected to be seen in at least five 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 less than 0.0056.

Two-, three-, and four-locus haplotype frequencies were estimated using the iterative expectation-maximization (EM) algorithm [14,15]. LD 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. An $A D'_{ij}$ 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'_{ij} [16], uses the products of the allele frequencies at each locus to weight the LD contribution of specific allele pairs; whereas the second, W_n [19], calculates a normalization of the χ^2 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 Table 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 with those for the following: 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 [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.

3. Results

3.1. Allele and genotype frequencies

Allele frequencies for the HLA–A, –C, –B, and –DRB1 loci are shown in Table S2. A total of 272 unique A–C–B–DRB1 phenotypes (based on four-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 Hardy–Weinberg

Download English Version:

<https://daneshyari.com/en/article/3351770>

Download Persian Version:

<https://daneshyari.com/article/3351770>

[Daneshyari.com](https://daneshyari.com)