Published in final edited form as:

Tissue Antigens. 2009 December ; 74(6): 520–533. doi:10.1111/j.1399-0039.2009.01381.x.

Molecular Variation at the HLA-A, B, C, DRB1, DQA1, and DQB1 Loci in Full Heritage American Indians in Arizona: Private Haplotypes and Their Evolution

Robert Williams,

NIH, PECRB/NIDDK, 1550 E. Indian School Road, Phoenix, Arizona, USA

Yao-Fong Chen,

Tzu Chi University, Human Development, Hualien, Taiwan

Robert Endres,

Blood Systems Laboratories, HLA Laboratory, Tempe, Arizona, USA

Derek Middleton,

Transplant Immunology, 3rd Floor, Duncan Building, Royal Liverpool and Broadgreen University Hospital Trust, Prescott Street, UK

Massimo Trucco, and

University of Pittsburgh School of Medicine, Dept of Pediatrics, Pittsburgh, Pennsylvania, USA

William Knowler

NIH/NIDDK, DECRS, Phoenix, Arizona, USA

Abstract

A sample of 492 full heritage, unrelated residents of the Gila River Indian Community (GRIC) of Arizona were characterized for their high resolution DNA alleles at the HLA-A, B, C, DRB1, DQA1, and DQB1 loci. Only 5 allelic categories are found at HLA-A, 10 at HLA-B, 8 at HLA-C and HLA-DR, and 4 at DQA1 and DQB1. There is little evidence for population structure at the 6 loci. Two “private” alleles, B*5102 and B*4005, that are found nearly exclusively in American Indian populations in the desert southwest and northern Mexico, are likely new mutations after the first inhabitation of the area, the evolution of which are reflected in the contemporary distribution of their respective haplotypes. DRB1*1402 has the highest reported frequency of any specificity at the DRB1 locus, 0.7461, and serves as a sensitive probe for locating related east Asian populations. The haplotypes in this population also exhibit a highly restricted distribution and strong genetic disequilibria, which has important implications for matching solid organ and bone marrow allografts. It is shown that, when one considers HLA-A-B-DRB1 homozygotes as allograft donors for all full heritage members of the GRIC, 50% of the community would find a non-mismatched organ within the homozygotes for the 6 most common haplotypes. This raises questions about transplantation policy and whether, in the presence of high frequency private alleles and a restricted number of haplotypes, the full heritage American Indian community of the desert southwest should act as its own pool of donors for its affected members.

Keywords

Pima; Private Alleles; Restricted Variation; Tohono O'odham; Transplantation

Introduction

Since 1965 the Pima and Tohono O'odham Indians (Pimans) of the Gila River Indian Community (GRIC), in the Sonoran Desert of Arizona, have participated in a long range study of Type 2 diabetes mellitus and its complications, obesity, arthritis, and cardiovascular disease by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) in Phoenix, Arizona (1,2). The HLA loci have played an important role in the many epidemiological studies that have been performed. It has been reported that HLA-A*02 is associated with Type 2 diabetes, an allele that also confers a mortality risk for cardiovascular disease (3,4). A meta-analysis revealed an association between the most common class II allele, HLA-DRB1*1402, and rheumatoid arthritis (5). In 1992 population descriptions for the serological variation at the class I and class II loci were published (6,7). Since that time many persons have been characterized for the high resolution DNA alleles at both class I and class II. The present report summarizes these molecular characterizations, discusses them in the context of the reported serological data, reveals the private allelic variation and haplotype evolution that are unique in Pimans, and explores the implications of this variation for solid organ and bone marrow transplantation in the traditional Piman cultural area of the desert southwest and northern Mexico.

Materials and Methods

Population

Typing for the DNA alleles at the HLA-A, B, and C loci was performed by YFC, RE, and DM; HLA-DRB1 typing was performed by YFC, RE, and MT; HLA-DQA1 and DQB1 typing was performed by MT; all tests were by standard methods (8-12). A total of 492 persons was typed by the 3 different laboratories and each person had at least 1 locus typed to the allele level, with allelic resolution at the level of the nucleotide and amino acid, included in the analysis. The sample was restricted to residents of the GRIC who are unrelated (no first degree relatives) with a self-reported Indian heritage of 8/8. There were 193 males with the average age at the time of sample collection of 43.8 years, and 299 females with an average age of 46.0.

Statistics

Allele frequencies were calculated by gene counting and by maximum likelihood while fitting an ABO-like model that included a blank allele (13). Haplotype frequencies and their genetic disequilibria were computed by a modified EM algorithm after the method of Long et al. (14) that included an algorithmic filter for false estimation of rare haplotypes. Tests for the significance of the disequilibria between loci, and the significance of the value of D for all individual 2-, 3-, and 4-allele haplotypes, were performed by a Monte Carlo simulation of 1000 samples using the estimated allele frequency values at each locus as the probability for drawing a random allele at each locus in the haplotype. Two random alleles were selected for each locus in the simulation for N phenotypes, where N represents the size of the observed sample. Each sample in the simulation is one in which there is both Hardy-Weinberg and gametic equilibrium. The EM algorithm was then applied to each of the 1000 samples after which allele and haplotype frequencies and likelihoods were calculated. Means and 95% confidence intervals were computed for the haplotype frequencies, disequilibria, and likelihoods. The empirical 95% confidence intervals represent the 25th and 975th observations of the ordered set of 1000 estimates for each variable. The variable D' , the proportion of maximum positive or negative disequilibria for 2 loci, was calculated by standard methods (13).

Wright's fixation index or inbreeding coefficient, F , was estimated for each locus (15).

This compares the observed total frequency of heterozygotes with that expected under Hardy-Weinberg equilibrium. The standard error for F was calculated by the jackknife, leave-one-out method (16). Each genotype was removed from the sample, 1 at a time, after which the allele frequencies and F were computed again. The standard error of the statistic was then calculated from the N leave-one-out samples. Tests for Hardy-Weinberg at a locus were performed by the method of Nam and Gart and the standard goodness-of-fit test (13,17). The null hypothesis of a Nam and Gart test is that the T statistic is 1.0 when the locus is in equilibrium and there is no evidence for blank alleles. Only genotype categories with expected values greater than 5, and a category that included all genotypes with smaller expectations, were included in the calculation of the goodness-of-fit chi square. The degrees of freedom is 1 less than the number of categories. Individual admixture estimates (IAE) were taken from Williams et al. (18). The standard error of the mean of IAE was estimated by the bootstrap procedure (16). The age of new mutations in the population is estimated from the method of Kimura and Ohta for neutral alleles in randomly mating populations of constant effective size of N (19).

Normalized Haplotype Matrix

For new mutations one expects that their distribution on class I, A-B-C haplotypes would be highly restricted. For instance, when a new mutation at HLA-B occurs, it would be linked with an allele at both HLA-A and HLA-C, and this haplotype would exclusively carry the new mutation until new combinations began to appear by genetic recombination between A and B, and B and C. The normalized haplotype frequency of the new mutant's haplotype would be 1.0, that is, its haplotype frequency divided by its allele frequency. The closer to 1.0 that this normalized value is, the newer the mutation and the less time for genetic recombination. (One qualification to this analysis is that there is no way to distinguish between the primary events, mutation and migration. For instance, if the haplotype were to enter the population by gene flow, rather than mutation, the distribution of normalized frequencies and the effect of genetic recombination would appear the same in the matrix. However, "private" genetic variation, restricted to a small group or area, is more likely to be the result of new mutations.) The matrix is created by putting the A alleles on the rows and the C alleles on the columns and filling each cell in the matrix with the normalized frequency of the respective haplotype for the new B allele.

Results

Genetic Admixture

Among the 492 persons with stated full heritage, 321 were full heritage Pima, 19 full heritage Tohono O'odham, and 145 full heritage people who were some combination of Pima and Tohono O'odham. Seven persons in the sample were a combination of Maricopa and either Pima or Tohono O'odham. Heritage was also measured in the sample by estimating an individual admixture estimates (IAE). Using a 2 parental model for genetic admixture with allele frequencies from full heritage Pima as the American Indian parental group and frequencies from European-Americans as the second parental group, individual estimates of European genetic admixture were computed for each member of the sample. The mean estimated European-American IAE in the sample was 0.024 with a 95% confidence interval of 0.017-0.030. Figure 1 presents estimated proportion of Indian heritage.

Allelic Variation

Allele frequencies, and their 95% confidence intervals, for the HLA-A, B, C, DRB1, DQA1, and DQB1 loci are found in Tables 1 and 2. Table 3 includes heterozygosity and goodness-

of-fit statistics for these 6 loci. Haplotype frequencies, disequilibria estimates, and their 95% confidence intervals are presented in Tables 4 (2-loci), 5 (3-loci), and 6 (4-loci).

Allelic variation at the HLA class I and II loci follows a simple, restricted pattern in Pima Indians; there are a small number of alleles and haplotypes with significant frequency with the remainder rarely occurring. Many of the rare alleles are probably the result of genetic admixture.

At HLA-A, 233 persons were typed and resolved at high resolution, which yielded 5 allele categories. Two alleles, A*0201 and A*2402, sum to 81.5% of the total allelic frequencies. The RARE allele category contains A*0101[2], A*0301[1], and A*680102[1] (Table 1). Observed heterozygosity (H) is 0.635 and is nearly identical to the expected value of 0.642 (Table 3). Wright's F is small, 0.011 and not statistically significant from 0.0, while the Nam-Gart T statistic for this locus includes 1.0 in its confidence interval. A goodness-of-fit chi square of 2.42, with 7 degrees of freedom, also does not reject the null hypothesis of Hardy-Weinberg equilibrium at the HLA-A locus. There is no evidence for blank alleles at the HLA-A locus: the maximum likelihood estimate included 0.0 (0.0, 0.0127); there were no double blank phenotypes.

There were 9 common high resolution alleles detected at the HLA-B locus for which 218 persons have been characterized (Table 1). Alleles B*2705, B*3501, B*3901, B*3906, B*4001, B*4002, B*4005, B*4801, and B*5102 have confidence intervals that exclude 0.0, and have a combined frequency of 0.9725. There were a total of 12 rare alleles detected at the locus: B*0702[2], B*0801[1], B*1401[1], B*1501[4], B*3903[1], B*3905[1], B*3910[1], and B*5101[1]. Observed heterozygosity was 0.821 as compared to an expected H of 0.861. Wright's F is 0.046 and not statistically different from 0.0 (Table 3). In contrast the hypothesis of Hardy-Weinberg equilibrium is rejected at HLA-B by the Nam-Gart T statistic of 1.614 and a confidence interval that excludes 1.0. The heterogeneity chi-square statistic is, however, not significant with an estimate of 13.98 and 14 degrees of freedom. When an ABO-like maximum likelihood model was fitted to the HLA-B locus there was a significant frequency, 0.0288 (0.0149, 0.0426), of a blank allele.

One hundred sixty-nine persons were typed for high resolution alleles at the HLA-C locus where 7 common specificities were detected that have a combined frequency of 0.9793. The most common allele is C*0801, 0.2988. Rare alleles include C*0102[2], C*0701[2], C*0802[1], and C*1502[2]. Observed heterozygosity is 0.799 with an expected H of 0.806 that lies within the 95% confidence interval of the observed value, and a Wright's F of 0.009 with confidence limits that include 0.0 (Table 3). Neither Nam-Gart's T nor the goodness-of-fit chi square statistic is statistically significant. There is also no evidence for blank alleles at HLA-C when the ABO-like, maximum likelihood, model is applied to the locus.

For the HLA-DRB1, class II, locus 321 persons were typed at high resolution (Table 2). Allele DRB1*1402 is most common at this locus, 0.7461, followed by DRB1*1602 with an allele frequency of 0.0919. Low resolution antigen and allele DRB1*04 is represented by 5 molecular variants in Pimans: DRB1*0403, DRB1*0407, DRB1*0410, and 2 alleles in the RARE category, DRB1*0401[1] and DRB1*0411[2]. There was also just 1 allele for DRB1*1301 in the sample. Observed heterozygosity is 0.411 with confidence limits that include the expected value, 0.428. Wright's F , 0.036, is not significantly different from 0.0, while Nam-Gart's T and the heterogeneity chi-square statistics support the presence of Hardy-Weinberg equilibrium at the locus. Fitting an ABO-like maximum likelihood model to the data did yield an estimate for the blank allele that is significantly different from 0.0, 0.0152 (0.0016, 0.0287). However, there were no double blank phenotypes.

The DQA1 and DQB1 loci share a pattern of variation in the 217 high resolution genotypes. Each locus has an allele with a frequency that approaches 0.9, DQA1*0501, 0.8894, and DQB1*0301, 0.8963, with 3 additional alleles at low frequencies (Table 2). Because of 1 highly frequent allele at each locus the observed heterozygosity is low, 0.221 for DQA1 and 0.207 for DQB1, while neither is different from expectations (Table 3). Wright's *F* is negative for each locus and of similar magnitude, -0.089 and -0.085, but falls within the 95% confidence interval created with the jackknife standard error of *F*. The Nam-Gart *T* and the heterogeneity Chi-square statistics support the presence of Hardy-Weinberg equilibrium at the 2 loci.

Haplotype Variation

A sample of 214 persons was typed for both the high resolution HLA-A and HLA-B loci from which their 2-locus haplotype frequencies were estimated (Table 4). The highest frequency haplotype is A*0201-B*5102, 0.1212, with a significant positive *D* of 0.0605. Second in magnitude is A*0201-B*4801, 0.1198, for which *D* is not significantly different from 0.0. With 5 alleles at HLA-A and 10 at HLA-B there are 50 2-locus haplotypes in the estimator maximum program for the haplotype frequencies. The 7 haplotypes in Table 4 alone represent more than 60% of the variation for the A-B combinations. For estimating A-C haplotype frequencies 165 persons were typed for high resolution alleles. Haplotypes A*0201-C*0801 and A*2402-C*0304 have the largest frequencies, 0.1731 and 0.1248, respectively (Table 4). While there are 40 A-C haplotypes in the model, the 9 A-C haplotypes in Table 4 represent 75.2% of the haplotypes for these loci. Variation for the 2-locus HLA-A and HLA-DRB1 haplotypes is represented by 5 entries in Table 4 that sum to 0.7368. As would be expected from their allele frequencies, haplotype A*0201-DRB1*1402 has the largest magnitude, 0.3438 followed by A*2402-DRB1*1402, 0.2261. None of the 5 disequilibrium estimates for the A-DRB1 haplotypes is significantly different from 0.0.

Eight haplotypes are found in Table 4 for HLA-B and HLA-C for which 164 persons were typed at high resolution. For 7 of the combinations the estimate for *D* is significantly different from 0.0 while it also represents more than 50% of the haplotype frequency. The highest is for B*4801-C*0801, 0.1761, with *D*=0.1080. Similarly haplotype B*3501-C*0401 is very frequent, 0.1616, with a *D*=0.1325. Together these 7 B-C haplotypes sum to 0.8050. Haplotypes for B-DRB1 in Table 4 are very common when the 2 alleles are frequent at each locus. Combinations B*4005-DRB1*1402, B*4801-DRB1*1402, and B*5102-DRB1*1402 have frequencies greater than 0.12. Loci HLA-C and HLA-DRB1 share the pattern of a few 2-locus haplotypes with very large frequencies (Table 4). For instance C*0801-DRB1*1402, 0.1902, C*0304-DRB1*1402, 0.1366, and C*0702-DRB1*1402, 0.1575, together represent 48.4% of the variation of this 2-locus combination.

The largest 2-locus haplotype frequencies in Table 4 are found in the class II loci combinations. Haplotypes DRB1*1402-DQA1*0501, 0.8180, DRB1*1402-DQB1*0301, 0.8179, and DQA1*0501-DQB1*0301, 0.8894, are more than 80% of the haplotype set while each also has a disequilibrium that is significantly different from 0.0.

Selected 3-locus haplotypes are found in Table 5. The class I loci haplotype A*0201-B*5102-C*0801 has the largest frequency, 0.1136, which alone represents more than 11% of the variation for this combination. When combined with class II loci, the alleles in this haplotype are part of other high frequency groups in Table 5: A*0201-B*5102-DRB1*1402, 0.1242, A*0201-C*0801-DRB1*1402, 0.1377, and B*5102-C*0801-DRB1*1402, 0.1180. For the 3-locus combinations, just the few, most frequent (*HF* > 0.05), haplotypes in the table are a significant proportion of each set: A-B-C with 400 possible haplotypes, the 8 in Table 5 are 55.5% of the total; A-B-DRB1 with 400 potential haplotypes, 4 haplotypes in the table, 36%; A-C-DRB1 with 320 combinations, 4 table haplotypes, 37%; and B-C-DRB1

with 640 haplotypes and 6 table haplotypes, 45%. The haplotype with the largest magnitude in Table 5 is for the class II loci, for which there are 128 possible haplotypes; combination DRB1*1402-DQA1*0501-DQB1*0301 alone has a frequency of 0.8180, with a significantly negative value of D , -0.0618.

Eight 4-locus haplotypes for HLA-A, B, C, and DRB1, with $HF > 0.04$, are found in Table 6 for 155 persons who were typed for high resolution alleles at all loci. The most frequent combination is A*0201-B*5102-C*0801-DRB1*1402, 0.1119, with $D=0.0591$, significantly differs from 0.0 and is 53% of the total haplotype set. In fact all of the most frequent 4-locus haplotypes have significantly positive disequilibrium values. Together the 8 haplotypes in Table 6 represent 48% of the 4-locus set with 3200 possible combinations.

The observed likelihoods for 15 combinations of genetic disequilibria between loci for persons who were typed for HLA-A, B, C, DRB1, DQA1, or DQB1 were calculated. The observed likelihood falls within the range of the 95% confidence interval, generated from 1000 samples of size N in gametic equilibrium, for haplotypes at A-C, A-DRB1, B-DRB1, and C-DRB1. Loci HLA-B and HLA-C are in strong disequilibrium (observed likelihood = -696.9, 95% C.I. -1033.7, -970.8) and contribute to the statistically significant disequilibria of higher level combinations A-B-C, B-C-DRB1, and A-B-C-DRB1.

Normalized Haplotype Matrices and Haplotype Evolution

Table 7 presents the normalized haplotype frequency matrix for HLA-B*5102. It will be seen that only 3 cells of the matrix have non-zero values, that is, only 3 haplotypes were found in the sample. Haplotype A*0201-B*5102-C*0801 has a normalized value of 0.8712. This haplotype represents the primary, historical event in the population, either mutation or gene flow. The matrix is a window through which to view the appearance and subsequent evolution of the new B*5102 haplotype. Columns in Table 7 represent recombination between the A and B loci, while the rows reflect that between B and C. There has been at least 1 recombination event between A and B that replaced A*0201 with A*3101 on the B*5102-C*0801 chromosome segment; and at least 1 event between B and C that replaced C*0801 with C*0803 on the A*0201-B*5102 segment. There are no B*5102 haplotypes in the sample with A*0206 or A*2402, in spite of the fact that the latter allele has a very high frequency in the population.

The normalized haplotype matrix for HLA-B*4005 is presented in Table 8 and tells a very different story from that for B*5102. The original event appears to be older because of the larger departure from 1.0 for the normalized frequency of the parent haplotype, A*2402-B*4005-C*0304, 0.5562. The matrix further suggests that there was an ordered succession of events for this parental type, though the exact order cannot be known. One ordered set has 3 events: 1) there was a recombination between A and B that replaced A*2402 with A*0201 to produce A*0201-B*4005-C*0304; 2) a subsequent recombination replaced C*0304 with C*0702 from which arose A*0201-B*4005-C*0702, a haplotype with a normalized frequency of 0.3011; and 3) there was a recombination between A and B that replaced A*0201 with A*3101 that yielded A*3101-B*4005-C*0702. More complex sets of events can be formed because the exact order within a column of the matrix is not known. For instance, it might be that A*0206 replaced A*2402 on the parental haplotype before a further recombination replaced A*0206 with A*0201, which then underwent recombination at C. Nevertheless, the root haplotype is evident as well as the general pattern of events for its subsequent evolution in the population, even though their exact order cannot be determined from the data.

The estimated generations and years (20 years per generation) since the occurrence of HLA-B*5102 and B*4005 are presented in Table 9. The method of Kimura and Ohta requires first

the estimate of the scaled time unit (stu), which is 0.6047 for B*5102 and 0.6735 for B*4005. The time since the mutation occurred then depends on the effective population size, which is presented in increments of 100 persons.

Table 10 presents the expected distribution of homozygotes and heterozygotes for the 9 most common A-B-DRB1 haplotypes in the current study. For instance, the first, most common haplotype, A*0201-B*5102-DRB1*1402, would yield, in a sample size of 4418 persons, 68 homozygotes who could act as donors or recipients and 758 heterozygotes who share one copy of the haplotype and who would be potential recipients. This is about 18.7% of the full heritage population. The haplotype categories are independent of one another. For the second haplotype in Table 10, A*0201-B*4801-DRB1*1402, there are 37 expected homozygotes and 484 heterozygotes that exclude genotypes with haplotype 1. Table 11 has the expected distribution for the 36 heterozygotes formed from the 9 most common haplotypes in Table 10.

Discussion

Comparative Data in Other Populations

Williams and McAuley reported on the low resolution, class I, serological variation in this population in 1992 (6). At the HLA-A locus serological antigen HLA*A2 in that study is represented as high resolution molecular variants A*0201 or A*0206 in the present work while the second major antigen, HLA*A24 types as allele A*2402. Serological allele HLA*A31 is characterized at high resolution as A*3101. Of particular note is the highly restricted variation at this locus when compared with non-American-Indian populations. Only 4 alleles are polymorphic (an allele frequency ≥ 0.01), while 1 allele, A*0201, represents nearly half of the total frequency.

In a recent report by the National Marrow Donor Program (NMDP) on the high resolution allele and haplotype frequencies in 4 ethnic groups in the United States, European American, African American, Asian, and Hispanic, 27, 36, 37, and 45 alleles were defined at HLA-A as common, respectively, that is, with a frequency greater than 0.005 (20). This definition of the frequency of “common” alleles in the NMDP data base offers perspective on the variation in this report. It is one half of the frequency used in the conventional definition of an allele being common, or polymorphic, 0.01. Using the traditional measure, there are only 5 common alleles in the sample from the GRIC. Nearly a quarter, 0.226, of the population is expected be homozygous for the major allele, A*0201, which has a frequency of 0.4742. In the present work this high resolution allele is more common than in any of the major NMDP ethnic groups in which the highest frequency is in European Americans, 0.2960 (20). In 2 published studies of American Indian populations by the American Society for Histocompatibility and Immunogenetics (ASHI) minority workshops, the frequencies of A*0201 and A*0206 in the Lakota Sioux were 0.2966 and 0.1054, respectively (21). In contrast, in the Yup'ik Eskimo, allele A*0201 has a frequency of only 0.0238 while A*0206 is more common than in the present work, 0.1647 (22). Allele A*0201 could represent genetic admixture from European-Americans in the Yup'ik. For allele A*2402 its frequency in the Gila River sample, 0.3341, is higher than in any of the 4 NMDP ethnic groups, in which Asians have the highest value, 0.1824 (20); it is also higher than in the Lakota Sioux, 0.2623 (21); but it is lower than that observed in the Yup'ik Eskimo in which it has the largest frequency at HLA-A, 0.5814 (22).

The restriction of variation is also exhibited at the HLA-B locus in the GRIC sample, when compared to the NMDP data. Table 1 has 9 alleles common at high resolution whereas in African Americans there are 68, Asians 71, European Americans 44, and Hispanics 88 (20). High resolution typing at HLA-B, when compared with serological results, reveals a pattern

of new alleles, some of which were first defined in workshops and conferences in samples from the GRIC.

Serological antigen Bw48 in the earlier report (6) is characterized at high resolution as HLA-B*4801, the most common allele at HLA-B in the GRIC sample, 0.2179 (Table 1). Lymphocytes from the GRIC were important in both the serological characterization and molecular sequencing of this specificity that was also found in the Kaingang of South America (23,24). Allele B*4801 appears to be a genetic chimaera that was produced by an unknown, complex, sequence of molecular events. A part of allele B*3901 has been introduced into B*4001 to produce B*4801 which, in addition, differs from B*4001 at codons 97 and 245, and in its leader sequence (23). It is nearly absent in African and European Americans in the NMDP data base, with frequencies of 0.0005 and 0.0002, respectively, while Asians, Hispanics, and the Lakota Sioux have similar low frequencies, 0.0204, 0.0215, and 0.0221 (20,21). Once again, in Hispanics, the presence of this allele could be from the American Indian component of genetic admixture. In the Yup'ik Eskimo, B*4801 has an allele frequency of 0.1468, that is between the latter populations and the GRIC sample (22).

At the HLA-C locus, allele C*0801 was defined by the DNA sequence of a cell from the GRIC that was included in the earlier population report as typing for Cw8 (6). Its sequence helped clarify a confusion in the definition of the molecular variation at HLA-C when it replaced the previously defined specificity, Cw*1101 (25,26). When the high resolution typing is compared with the serology of the earlier report, there is, similar to HLA-A, a second level of polymorphism. For instance, the serological antigen, Cw8, has 2 underlying molecular alleles, C*0801 and C*0803, while for Cw3, C*0303 and C*0304 are present. Also mirroring the alleles A*0201 and A*0206 at HLA-A, in each case 1 allele is highly prevalent, C*0801 (0.2988) and C*0304 (0.1953), while the second has a more modest frequency, C*0803 (0.0621) and C*0303 (0.0237) (Table 1). While C*0801 has the largest allele frequency at the locus, it has a frequency of only 0.0180 in the Lakota Sioux and 0.0042 in the Yup'ik Eskimo (21,22). Asians in the NMDP population data base have the highest frequency of the C*0801 allele, 0.0799; it is 0.0283 in Hispanics, while it is found at only trace levels in European-Americans, 0.0001, and African-Americans, 0.0012 (20). Using the traditional frequency for the definition of a genetic polymorphism, 0.01, the 3 Native American samples have polymorphic levels of C*0803: GRIC, Lakota Sioux, 0.0128, and Yup'ik Eskimo, 0.0191 (21,22). This contrasts with the 4 NMDP populations where its largest allele frequency is in Asians, but not polymorphic, 0.0045 (20).

At the class II loci, allele DRB1*1402 has an important place and unique properties in the Gila River Indian sample. Its detailed serological definition and history were laid out in the earlier population description (7). Briefly, it was first identified in Arizona in cells from GRIC and in segregation studies in Yaqui Indian families from Guadalupe, a suburb of Phoenix, which were included in the 1984, 9th International HLA Workshop and Conference in Munich and Vienna (27). The pattern of reactivity was one with HLA anti-sera that were poly-specific for antigens DR3 and DR6 and strongly associated with DRw52 and DQw3; therefore Blood Systems HLA Laboratory gave the antigen the name DR3X6 (7). In this same workshop cells from South American Indians also shared the serological pattern including a homozygous cell 9W1701, that had the local name of AMALA, and that was subsequently DNA sequenced and given the official allele designation of DRB1*1402 (28). Subsequent to the 9th Workshop and Conference large samples from GRIC were typed for DRB1 and incorporated into population and disease association studies.

Allele DRB1*1402 segregates in American Indian populations but is either not found or found at only trace levels in European Americans, African Americans, and Asians in the

NMDP data base, 0.0000, 0.0006, and 0.0003, respectively, while its presence in Hispanics, 0.0238, arises from the American Indian component of the population (20). Its frequencies in the Lakota Sioux, 0.1510, and in the Yup'ik Eskimo, 0.2222, while high for a DRB1 allele, are much smaller than that in the GRIC sample, 0.7461 (Table 1) (21,22). In fact, the size of this DRB1 allele frequency may be unique for a human population. For perspective, it is 3.2 times as great as the largest DRB1 frequency in the Yup'ik Eskimo, DRB1*0401, 0.2321 (22); 4.0 times as great as the most common allele in the Lakota Sioux, DRB1*0407, 0.1881 (21); 5.2 times as great as the most frequent allele in European Americans, DRB1*1501, 0.1444 (20); 6.4 times as great as the largest allele frequency in African Americans, DRB1*1503, 0.1175 (20); and for Asians, with DRB1*0901 having the highest allele frequency, 0.1018, it is 7.3 times larger (20). The direct result of the high allele frequency of this DRB1 specificity is an even larger phenotype frequency and a high proportion of persons with 2 copies of the allele. In the GRIC DRB1 sample of 321 persons, 299, or 93.2%, have at least 1 copy of the DRB1*1402 allele. This is very close to its expected value, 93.6%, under Hardy-Weinberg equilibrium. More than half, 180, or 56.1%, are homozygotes, compared to the expected value of 55.7%.

The high frequency of DRB1*1402 is mirrored in the similarly high magnitudes of DQA1*0501 and DQB1*0301 with which it is in strong linkage disequilibrium (Tables 2 and 5). There are comparable data in the minority workshop reports for the 2-locus haplotype DRB1*1402-DQB1*0301. In the Yup'ik Eskimo its frequency is 0.2202, while in the Lakota Sioux it is 0.1071, which are much smaller than that for the GRIC, 0.8179 (21,22). In the NMDP samples the frequency of the haplotype is less than 0.0001 in European Americans, African Americans, and Asians, while its frequency in Hispanics, 0.0185, is probably the result of the American Indian component of Mexican Americans in the Hispanic database (20).

East Asian Affinities

Allele DRB1*1402 also acts as a sensitive marker for tracking affinities between American Indian populations and those of northeast Asia. Uinuk-ool et al. reported on the distribution of DRB1, DQA1, and DQB1 loci for indigenous populations in Siberia (29). DRB1*1402 is present in the Negidal (0.086), Ulchi (0.041), and Okhotsk Evenki (0.020) native groups who share the Tungusic-group language. The Ulchi and the "lower" Negidal populations live along the lower Amur River, while the Evenki people are one of the largest ethnic groups in northeast Asia comprising some 20,000 persons who are distributed west of the Sea of Okhotsk and in China and Mongolia. In terms of linkage groups, the common haplotype in the GRIC sample, DRB1*1402-DQA1*0501-DQB1*0301, is most likely found in the Amur River populations the Negidal (0.057) and the Ulchi (0.041). DRB1*1402 was not found in the 5 indigenous groups that were typed from the central Siberian region, the Mansi, Tuva, Todja, Tofalar, and Buryat. The allele is ubiquitous in the American Indian populations of North, Central, and South America (30). Therefore it acts as a sensitive thread that connects the people of northeast Asia and the North and South American continents. The authors suggest that the DRB1, DQA1, DQB1 data, when put in a neighbor-joining method, are consistent with the 3-migration model for introduction of humans to the Americas (29). A similar result was reported for the Gm locus for the GRIC (31). The combined HLA and Gm data, therefore, are consistent with the interpretation of 1 source population in northeast Asia and differentiation into 3 distinct genomic sets by genetic drift, either by Founder Effect in 3 distinct migrations from the same source, or by 1 migration followed by genetic drift in the PaleoIndian, Na Dene, and Aleut-Eskimo historical areas.

Private Alleles and Their Evolution

Neel introduced the idea of private alleles in 1973, taking as his guide the idea of “private” as first used by serologists to define rare antigenic variants in small human populations or groups of families. He expanded this idea to electrophoretic variants in South American Indian tribes (32). There were 2 categories of HLA*B5-related antigens in the earlier report, B5 and B51, which now are represented by 1 high resolution specificity, B*5102 (6). It was first defined in lymphocytes from Pima Indians and probably represents a new world mutation at the HLA-B locus (33). In the NMDP data the frequency of B*5102 is nearly 0.0: 6.0×10^{-5} in European-Americans, 0.0004 in African Americans, 0.0062 in Asians, and 0.0055 in Hispanics (20). The frequency in Asians and Hispanics could be due to an ethnic classification error in the NMDP data set or genetic admixture. It is not found in either of the ASHI minority reports for the Lakota Sioux and the Yup'ik Eskimo (21,22). Outside the American southwest and Mexico only 13 chromosomes have been reported to be carrying B*5102 (34-37).

Allele HLA-B*4005 was first defined in cells from the GRIC in the HLA International Workshop and Conference in 1987 using the SERAN serological data analysis programs and was given a provisional name of HLA-BN21 because it cross-reacted in the microlymphocytotoxicity test with epitopes of the HLA-B21 group (6,38). Its distribution and haplotype associations were first reported in 1992 (6). Gene sequencing, using lymphocytes from the GRIC, revealed a molecular, DNA, affinity with the HLA-B*40 alleles (39). Therefore the WHO Nomenclature Committee for factors of the HLA System gave it the designation of HLA-B*4005, the fifth DNA allele in the HLA-B*40 group. In the NMDP data base it has a frequency of 0.0 in European Americans, African Americans, and Asians and occurs at a frequency in Hispanics, 0.0055, that could be attributed to the American Indian component of this population (20). It was not reported as present in either of the 2 ASHI minority workshop reports for the Lakota Sioux and the Yup'ik Eskimo (21,22). Outside of the American southwest and Mexico only 5 chromosomes have been reported to be carrying B*4005 (36,37,40,41). It has a frequency of 0.1514 in the GRIC sample (Table 1).

While the present sample was drawn from the GRIC, it also represents the HLA distribution in the Tohono O'odham, to whom the Pima are closely related in language, culture, and history (42-44). The Tohono O'odham cultural and residential area, before the establishment of the current international boundary in the 19th century, encompassed large portions of northern Mexico, 100 miles south of the border and west to the Gulf of California and the Colorado River (43). In a recent study of 44 full heritage Tarahumara Indians of western Chihuahua State in Mexico, Garcia-Ortiz et al. (45) report high frequencies of B*5102 (0.1477) and B*4005 (0.1136). After contact the Spanish defined a large geographical area that they called *Pimeria* that was divided into the Upper, or northern Pima, the Gila River Pima and Tohono O'odham, and the Lower or southern Pima, which included the lowland and highland Mexican Pima (46). The western extent of the Tarahumara is in the Sierra Madre, next to the highland, Mexican Pima population who live in and around the town of Maycoba (47,48). The lowland Pima in Mexico, who live in the vicinity of Onavas, refer to the highland Pima as the ‘Tarahumara-like people’ because of the similarities of their lifestyle and culture (47). In the National Museum of Anthropology and History in Hermosillo, Mexico there is a large collection of Maycoba Pima material culture. It shows strong similarities with that of the Tarahumara (47). These 2 HLA alleles, B*5102 and B*4005, then, can be seen as markers, private alleles, that had their origin within the *Pimeria* language and culture area that constituted a large arc running from the Gila River through northern Mexico and then curving east across the State of Sonora, to the Sierra Madre, and into the western part of the State of Chihuahua.

The exact timing of the entry of the Pima-related populations to the southwestern United States and northern Mexico is not known. The first migration of humans to the area was probably in the range of 12,000-15,000 years ago (49,50). Archeologists debate whether the present population developed from an indigenous group of long continuity that is represented in later years by the Hohokam archeological assemblage, or whether the Hohokam represent a northern Mexican population that migrated to the Salt River Valley and displaced, or lived among, the earlier culture (51). These 2 private alleles are not found in the plains Indians represented by the Sioux or in the Eskimo (21,22). This suggests that the mutations to B*5102 and B*4005 occurred after the first inhabitation of the desert southwest.

Evidence of the Hohokam dates from about 300 B.C. (51,52). If one accepts Haury's (52) theory that the Hohokam were migrants from Northern Mexico, then, depending upon the effective population size, one can propose a time interval for the mutations. Under this hypothesis the 2 private mutations occurred in northern Mexico among Pima-related populations, more than 2300 years ago, and then came north with the Hohokam. A reasonable estimate of time-since-mutation is, from Table 9, between about 3600 and 8000 years ago with effective population sizes of 300 to 600. The highly restricted variation of contemporary full heritage American Indians suggests that these effective population sizes, which are usually smaller than the actual size of the group, would fit that of a pre-agricultural, hunting and gathering society, like the one that existed in the new world before the implementation of sedentary agriculture, the rise of the city state, and the resultant increase in population density. An alternative hypothesis is that the mutations to B*5102 and B*4005 occurred among the Pimeria, after their cultural and linguistic identity developed, and that the occurrence of these private variants spread through the area by gene flow. This would be consistent with the contemporary Pima and Tohono O'odham being descendants of the indigenous people of the original migration who then gave rise to the Hohokam. The many hundreds of years of cultural and material intercourse between what is now the southwestern United States and Mexico would support this view. Also consistent with this second hypothesis is the observation that the origin of European genes in the GRIC is primarily from admixed migrants from Mexico (53). In either case, the private variation demonstrated by the HLA-B locus in this geographical region lends a similarity of genetic variability to what was already known about the cultural and linguistic affinities of the extent people.

Implications for Transplantation and Matching

These results may have important implications for a major health problem in the Community, that is, end stage renal disease (ESRD) due to diabetes (54). In other populations ESRD is optimally treated by kidney transplantation, a treatment that can be problematic in small populations. Small human populations with private alleles segregating at the HLA loci pose particular problems for matching donor with recipient from national data bases. To investigate the situation it was estimated that there are 4418 living, full American Indian heritage persons between the ages of 15 and 65 years in the GRIC in which the current population study was performed. For solid organ transplantation, let us propose a strategy that persons who are homozygous for A-B-DRB1 haplotypes would act as living-related, living-unrelated, or cadaver donors for other homozygotes and for persons with 1 copy of the linkage group. For this pairing, in the host versus graft direction (HVG), there would be no HLA incompatibility. From Table 10 it will be seen that the 6 most common homozygotes could be potential donors for 2308 persons, over 50% of the entire full heritage population. Nearly 60% of the community would find a donor among the 9 most common homozygotes.

For bone marrow transplantation the incompatibility of homozygote donors with heterozygote recipients, in the graft versus host direction (GVH), would preclude this strategy. The homozygotes within each genotype category in Table 10 would, however, be candidates for donor-recipient pairs. In Table 11 are found the expected numbers of persons who would be identical for the 36 heterozygous genotypes for the 9 most common HLA-A, B, and DRB1 haplotypes. The persons within each genotype category would be potential candidates for donor-recipient stem cell or bone marrow donation as well as identically matched living-related, living-unrelated, and cadaver solid organs. For instance, the most common heterozygote in full heritage members of the GRIC, A*0201-B*5102-DRB1*1402 / A*0201-B*4801-DRB1*1402, has an expected frequency of 100 persons. Even the least frequent heterozygote from the 9 haplotypes would be carried by 11 persons.

The first, third, and eighth most common haplotype in Table 10 have an allele with a restricted, private, distribution in the GRIC, either B*5102 or B*4005. These alleles are not found in the NMDP data base to the extent that a likely match would be found for a full heritage American Indian from the desert southwest; also their distribution in this population makes it unlikely that a full heritage person would find a solid organ identical match in the United Network for Organ Sharing (UNOS) network.

This raises the policy issue about whether community-based donor and recipient pools should be formed for native groups. With high through-put DNA sequencing technology it is now possible to type thousands of persons with efficiency and low cost. While working with the local transplantation center and medical facilities in the desert southwest, when the need for a transplant in a resident American Indian became necessary, the community would act as the donor resource. Transplantation is an international exercise today. In this spirit the available donor-recipient pool could be extended to the Tohono O'odham and other Pima-related people in northern Mexico, who also share this private variation at the HLA loci.

In Table 10 the proportion of persons who could be matched in the community is an underestimate because only full heritage persons were considered. But there are many admixed members of the community who would nevertheless share 1 of the common haplotypes with the potential homozygous donor pool. Therefore such a strategy of identifying homozygotes as organ donors would have application far outside the group of full heritage persons. Also, many of the haplotypes in GRIC, without B*5102 and B*4005, are shared by all American Indian groups in the United States. There is the potential to expand the full heritage American Indian donor-recipient pool to include all tribes that share this variation.

Acknowledgments

We thank the members of the Gila River Indian Community for their cooperation and participation in this study, and we thank the staff of the Diabetes Epidemiology and Clinical Research Section, NIDDK, for conducting the examinations. This research was partially supported by grants BSF45-3 and BSF45-4 from Blood Systems Foundation, Blood Systems, Inc., Scottsdale, Arizona and by the Intramural Research Program of the NIDDK.

References

1. Knowler WC, Pettitt DJ, Saad MF, Bennett PH. Diabetes mellitus in the Pima Indians: Incidence, Risk Factors and Pathogenesis. *Diabetes Metab Rev.* 1990; 6:1–27. [PubMed: 2192853]
2. Knowler WC, Bennett PH, Hamman RF, Miller M. Diabetes incidence and prevalence in Pima Indians: a 19-fold greater incidence than in Rochester, Minn. *Am J Epidemiol.* 1978; 108:497–504. [PubMed: 736028]
3. Williams RC, Knowler WC, Butler WJ, et al. HLA-A2 and type 2 diabetes in Pima Indians: an association and decrease in allele frequency with age. *Diabetologia.* 1981; 21:460–463.

4. Williams RC, Hanson RL, Pettitt DJ, Sievers ML, Nelson RG, Knowler WC. HLA*A2 confers mortality risk for cardiovascular disease in Pimans. *Tissue Antigens*. 1996; 47:188–193. [PubMed: 8740767]
5. Williams RC, Jacobsson LTH, Knowler WC, et al. Meta-analysis reveals association between most common class II haplotype in full heritage Native Americans and rheumatoid arthritis. *Hum Immunol*. 1995; 42:90–94. [PubMed: 7751165]
6. Williams RC, McAuley JE. HLA class I variation controlled for genetic admixture in the Gila River Indian Community of Arizona: A model for the Paleo-Indians. *Hum Immunol*. 1992; 33:39–46. [PubMed: 1548148]
7. Williams RC, McAuley JE. HLA class II variation in the Gila River Indian Community of Arizona: Alleles, haplotypes, and a high frequency epitope at the HLA-DR locus. *Hum Immunol*. 1992; 33:29–38. [PubMed: 1372303]
8. Williams F, Meenagh A, Patterson C, Middleton D. Molecular diversity of the HLA-C gene identified in a Caucasian population. *Hum Immunol*. 2002; 63:602–613. [PubMed: 12072195]
9. Middleton, D.; Williams, F. PCR-sequence-specific oligonucleotide probe typing for HLA-A, -B, and -DR. In: Powis, SH.; Vaughan, RW., editors. *Methods in Molecular Biology*. New York: Humana Press; 2002. p. 67–112.
10. Thonnard J, Deldime F, Heusterspreute M, et al. HLA class II genotyping: two assay systems compared. *Clin Chem*. 1995; 41:553–556. [PubMed: 7720245]
11. Zetterquist H, Olerup O. Identification of the HLA-DRB1*04, -DRB1*07, and -DRB1*09 alleles by PCR amplification with sequence-specific primers (PCR-SSP) in 2 hours. *Hum Immunol*. 1992; 34:64–74. [PubMed: 1356957]
12. Ringquist, S.; Nichol, L.; Trucco, M. Transplantation genetics. In: Rimoin, DL.; Connor, M.; Pyeritz, RE.; Korf, B., editors. *Principles and Practice of Medical Genetics*. 5th. Philadelphia: Elsevier; 2007. p. 983–1010.
13. Weir, B. *Genetic Data Analysis*. Sunderland, Mass.: Sinauer Associates; 1990.
14. Long JC, Williams RC, Urbanek M. An EM algorithm and testing strategy for multiple locus haplotypes. *Am J Hum Genet*. 1995; 56:799–810. [PubMed: 7887436]
15. Wright, S. *Evolution of the Genetics of Populations, the Theory of Gene Frequencies*. Vol. 2. Chicago: The University of Chicago Press; 1969.
16. Efron, B.; Tibshirani, RJ. *An Introduction to the Bootstrap*. London: Chapman & Hall/CRC; 1998.
17. Nam J, Gart JJ. On two tests of fit for HLA data with no double blanks. *Am J Hum Genet*. 1987; 41:70–76. [PubMed: 3605098]
18. Williams RC, Long JC, Hanson RL, Sievers ML, Knowler WC. Individual estimates of European genetic admixture associated with lower body-mass index, plasma glucose, and prevalence of Type 2 diabetes in Pima Indians. *Am J Hum Genet*. 2000; 66:527–538. [PubMed: 10677313]
19. Kimura M, Ohta T. The age of a neutral mutant persisting in a finite population. *Genetics*. 1973; 75:199–212. [PubMed: 4762875]
20. 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]
21. Leffell MS, Fallin MD, Hildebrand WH, Cavett JW, Iglehart BA, Zachary AA. HLA alleles and haplotypes among the Lakota Sioux: Report of the ASHI Minority Workshops, Part III. *Hum Immunol*. 2004; 65:78–89. [PubMed: 14700599]
22. Leffell MS, Fallin MD, Erlich HA, et al. HLA antigens, alleles, and haplotypes among the Yup'ik Alaska natives: Report of the ASHI Minority Workshops, Part II. *Hum Immunol*. 2002; 63:614–625. [PubMed: 12072196]
23. Belich MP, Madrigal JA, Hildebrand WH, et al. Unusual HLA-B alleles in two tribes of Brazilian Indians. *Nature*. 1992; 357:326–329. [PubMed: 1317015]
24. Raffoux, C.; Williams, RC.; Gorodezky, C. Report of Antigen Society 7 (B12, B13, B21, B37, B40, B41, B48). In: Charron, D., editor. *HLA, Genetic Diversity of HLA. Functional and Medical Implications*. Vol. 1. Paris: EDK Publishers; 1997. p. 45–78.
25. Zemmour, J.; Gumperz, J.; Hildebrand, H., et al. HLA-Cw11 is a combination of Cw1 and a public epitope of HLA-Cw3 and HLA-B46 molecules. In: Tsuji, K.; Aizawa, M.; Sasazuki, T., editors. *HLA 1991*. Vol. 1. Oxford: Oxford University Press; 1991. p. 527–529.

26. Zemmour J, Gumperz JE, Ward FE, Marsh SGE, Williams RC, Parham P. The molecular basis for reactivity of anti-Cw1 and anti-Cw3 alloantisera with HLA-B46 haplotypes. *Tissue Antigens*. 1992; 39:249–257. [PubMed: 1384166]
27. Schreuder, GMT.; Kennedy, LJ.; Gebuhrer, L., et al. Antigen report: HLA-DRw6 and its subgroups HLA-DRw13 and HLA-DRw14. In: Albert, ED.; Baur, MP.; Mayr, WR., editors. *Histocompatibility Testing 1984*. Berlin: Springer Verlag; 1984. p. 192-197.
28. Gorski J. First domain sequence of the HLA-DRB1 chain from two HLA-DRw14 homozygous typing cells: TEM(Dw9) and AMALA(Dw16). *Hum Immunol*. 1989; 24:145–149. [PubMed: 2466818]
29. Uinuk-ool TS, Takezaki N, Sukernik RI, Nagl S, Klein J. Origin and affinities of indigenous Siberian populations as revealed by HLA class II gene frequencies. *Hum Genet*. 2002; 110:209–226. [PubMed: 11935333]
30. Imanishi, T.; Akaza, T.; Kimura, A.; Tokunaga, K.; Gojobori, T. Allele and haplotype frequencies for HLA and complement loci in various ethnic groups. In: Tsuji, K.; Aizawa, M.; Sasazuki, T., editors. *HLA 1991; Proceedings of the Eleventh International Histocompatibility Workshop and Conference Vol. 1*; Oxford: Oxford University Press; 1992. p. 1065-1220.
31. Williams RC, Steinberg AG, Gershowitz H, et al. Gm allotypes in Native Americans: Evidence for three distinct migrations across the Bering Land Bridge. *Am J Phys Anthropol*. 1985; 66:1–19. [PubMed: 3976868]
32. Neel JV. “Private” genetic variants and the frequency of mutation among south American Indians. *Proc Nat Acad Sci*. 1973; 70:3311–3315. [PubMed: 4519625]
33. Kawaguchi G, Hildebrand WH, Hiraiwa M, et al. Two subtypes of HLA-B51 differing by substitution at position 171 of the $\alpha 2$ helix. *Immunogenetics*. 1992; 37:57–63. [PubMed: 1428012]
34. Chen S, Ren X, Liu Y, Hu Q, Hong W, Xu A. Human leukocyte antigen class I polymorphism in Miao, Bouyei, and Shui ethnic minorities of Guizhou China. *Hum Immunol*. 2007; 68:928–933. [PubMed: 18082574]
35. Hong W, Fu Y, Chen S, Wang F, Ren X, Xu A. Distributions of HLA class I alleles and haplotypes in northern Han Chinese. *Tissue Antigens*. 2005; 66:297–304. [PubMed: 16185325]
36. Chen S, Hong W, Shao H, et al. Allelic distribution of HLA class I genes in the Tibetan ethnic population. *Int J Immunogenet*. 2006; 33:439–445. [PubMed: 17117954]
37. Hong W, Chen S, Shao H, Fu Y, Hu Z, Xu A. HLA class I polymorphism in Mongolian and Hui ethnic groups from northern China. *Hum Immunol*. 2007; 68:439–448. [PubMed: 17462512]
38. Williams, RC.; Chen, SN.; Gill, DK., et al. Antigen Society #6 Report (B21, B49, Bw50, BN21, B12, B44, B45). In: Dupont, B., editor. *Immunobiology of HLA, Volume 1, Histocompatibility Testing 1987*. New York: Springer Verlag; 1987. p. 133-147.
39. Hildebrand WH, Madrigal JA, Belich MP, et al. Serologic cross-reactivities poorly reflect allelic relationships in the HLA-B12 and HLA-B21 groups. *J Immunol*. 1992; 149:3563–3568. [PubMed: 1385528]
40. Rani R, Marcos C, Lazaro AM, Zhang Y, Stastny P. Molecular diversity of HLA-A, -B, and -C alleles in a north Indian population as determined by PCR-SSOP. *Int J Immunogenet*. 2007; 34:201–208. [PubMed: 17504510]
41. Sanchez-Velasco P, Karadsheh NS, Garcia-Martin A, Ruiz de Alegria C, Leyva-Cobian F. Molecular analysis of HLA allelic frequencies and haplotypes in Jordanians and comparison with other related populations. *Hum Immunol*. 2001; 62:901–909. [PubMed: 11543892]
42. Fontana, BL. Pima and Papago: Introduction. In: Ortiz, A., editor. *Handbook of North American Indians*. Vol. 10. Washington, D.C.: Smithsonian Institution; 1983. p. 125-136.
43. Fontana, BL. History of the Papago. In: Ortiz, A., editor. *Handbook of North American Indians*. Vol. 10. Washington, D.C.: Smithsonian Institution; 1983. p. 137-148.
44. Ezell, PH. History of the Pima. In: Ortiz, A., editor. *Handbook of North American Indians*. Vol. 10. Washington, D.C.: Smithsonian Institution; 1983. p. 149-160.
45. Garcia-Ortiz JE, Sandoval-Ramirez L, Rangel-Villalobos H, et al. High-resolution molecular characterization of the HLA class I and class II in the Tarahumara Amerindian population. *Tissue Antigens*. 2006; 68:135–146. [PubMed: 16866883]

46. Di Peso, CC. Prehistory: O'otam. In: Ortiz, A., editor. Handbook of North American Indians. Vol. 9. Washington, D.C.: Smithsonian Institution; 1979. p. 91-99.
47. Dunnigan, T. Lower Pima. In: Ortiz, A., editor. Handbook of North American Indians. Vol. 10. Washington, D.C.: Smithsonian Institution; 1983. p. 217-299.
48. Pennington, CW. Tarahumara. In: Ortiz, A., editor. Handbook of North American Indians. Vol. 10. Washington, D.C.: Smithsonian Institution; 1983. p. 276-289.
49. Woodbury, RB. Prehistory: Introduction. In: Ortiz, A., editor. Handbook of North American Indians. Vol. 9. Washington, D.C.: Smithsonian Institution; 1979. p. 22-30.
50. Irwin-Williams, C. Post-Pleistocene Archeology, 7000-2000 B.C.. In: Ortiz, A., editor. Handbook of North American Indians. Vol. 9. Washington, D.C.: Smithsonian Institution; 1979. p. 31-42.
51. Gumerman, GJ.; Haury, EW. Prehistory: Hohokam. In: Ortiz, A., editor. Handbook of North American Indians. Vol. 9. Washington, D.C.: Smithsonian Institution; 1979. p. 75-90.
52. Haury, EW. The Hohokam, Desert Farmers and Craftsmen: Excavations at Snaketown, 1964-1965. Tucson: The University of Arizona Press; 1976.
53. Williams RC, Knowler WC, Pettitt D, et al. The magnitude and origin of European-American admixture in the Gila River Indian Community of Arizona: A union of genetics and demography. *Am J Hum Genet.* 1992; 51:101-110. [PubMed: 1609790]
54. Pavkov ME, Knowler WC, Bennett PH, Looker HC, Krakoff J, Nelson RG. Increasing incidence of proteinuria and declining incidence of end-stage renal disease in diabetic Pima Indians. *Kidney Int.* 2006; 70:1840-1846. [PubMed: 17003816]

Figure 1. Individual Admixture Estimates in GRIC Sample

The distribution of individual admixture estimates (IAE) is shown in the 492 residents of the Gila River Indian Community who were typed for high resolution HLA alleles. A 2 sample model was employed with European-American and Full Heritage Pima as the parental populations. Mean European admixture in the sample is 0.024 (0.017, 0.030).

Table 1
Class I Allele Frequencies. The RARE category is the sum of all alleles at a locus with frequency < 0.01. (N = Number of Persons)

	Allele	Frequency	95% C.I.
HLA-A	0201	0.4742	.4289, .5196
N = 233	0206	0.0687	.0457, .0916
	2402	0.3412	.2982, .3842
	3101	0.1073	.0792, .1354
	RARE	0.0086	.0002, .0170
HLA-B	2705	0.0619	.0393, .0846
N = 218	3501	0.1812	.1450, .2173
	3901	0.0505	.0299, .0710
	3906	0.0734	.0489, .0979
	4001	0.0298	.0139, .0458
	4002	0.0780	.0528, .1032
	4005	0.1514	.1177, .1850
	4801	0.2179	.1791, .2566
	5102	0.1284	.0970, .1598
	RARE	0.0275	.0122, .0429
HLA-C	0202	0.0592	.0340, .0843
N = 169	0303	0.0237	.0075, .0399
	0304	0.1953	.1530, .2375
	0401	0.1627	.1234, .2021
	0702	0.1775	.1368, .2183
	0801	0.2988	.2500, .3476
	0803	0.0621	.0364, .0879
	RARE	0.0207	.0055, .0359

Table 2**Class II Allele Frequencies**

The RARE category is the sum of all alleles at a locus with frequency < 0.01. (N = Number of Persons)

	Allele	Frequency	95% C.I.
HLA-DRB1	1402	0.7461	.7124, .7798
N = 321	1406	0.0467	.0304, .0631
	1602	0.0919	.0696, .1142
	0802	0.0623	.0436, .0810
	0403	0.0202	.0094, .0311
	0407	0.0171	.0071, .0272
	0410	0.0093	.0019, .0168
	RARE	0.0062	.0001, .0123
HLA-DQA1	0501	0.8894	.8599, .9189
N=217	0401	0.0599	.0376, .0822
	0301	0.0484	.0282, .0686
	0102	0.0023	.0000, .0068
HLA-DQB1	0301	0.8963	.8676, .9250
N=217	0302	0.0392	.0209, .0574
	0402	0.0622	.0395, .0849
	0602	0.0023	.0000, .0068

Table 3
Heterozygosity and Goodness-of-Fit Statistics for HLA Loci

Locus	N	Observed <i>H</i> (95% C.I.)	Expected <i>H</i>	Wright's <i>F</i> (95% C.I.)	Nam-Gart <i>T</i> statistic (95% C.I.)	Heterogeneity Chi-square (d.f., p)
HLA-A	233	0.635 (0.573, 0.697)	0.642	0.011 (-0.077, 0.099)	0.992 (0.735, 1.248)	2.42 (7, >0.05)
HLA-B	218	0.821 (0.770, 0.872)	0.861	0.046 (-0.014, 0.106)	1.614 (1.199, 1.996)	13.98 (14, >0.05)
HLA-C	169	0.799 (0.738, 0.859)	0.806	0.009 (-0.060, 0.079)	0.937 (0.538, 1.335)	11.07 (11, >0.05)
HLA-DRB1	321	0.411 (0.357, 0.465)	0.428	0.039 (-0.037, 0.115)	1.121 (0.832, 1.410)	1.80 (6, >0.05)
HLA-DQA1	217	0.221 (0.116, 0.276)	0.203	-0.089 (-0.114, -0.065)	0.876 (0.645, 1.106)	2.51 (3, >0.05)
HLA-DQB1	217	0.207 (0.153, 0.261)	0.191	-0.085 (-0.109, -0.060)	0.884 (0.654, 1.115)	2.10 (3, >0.05)

Table 4
Selected Haplotype (HF) Frequencies and Genetic Disequilibrium for 2 Loci (#H = number of haplotypes in estimate, *= $p < 0.001$)

#H	Locus 1 Allele	Locus 2 Allele	HF	Cum. HF	D	D'	95% Confidence Interval for D_{GE}
428	A	B					
	0201	5102	0.1212		0.0605*	0.8917	-0.0237, 0.0239
	0201	4801	0.1198		0.0184	0.1621	-0.0279, 0.0278
	2402	3501	0.0946		0.0312*	0.2627	-0.0258, 0.0267
	2402	4005	0.0789		0.0260*	0.2626	-0.0220, 0.0245
	0201	3501	0.0759		-0.0101	0.1174	-0.0282, 0.0283
	2402	3906	0.0578		0.0325*	0.6886	-0.0189, 0.0178
	0201	4005	0.0564	0.6046	-0.0153	0.2134	-0.0258, 0.0227
330	A	C					
	0201	0801	0.1731		0.0303	0.2046	-0.0359, 0.0374
	2402	0304	0.1248		0.0575*	0.4332	-0.0280, 0.0312
	2402	0401	0.0748		0.0198	0.1824	-0.0290, 0.0262
	0201	0401	0.0736		-0.0067	0.0834	-0.0271, 0.0311
	2402	0702	0.0663		0.0072	0.0617	-0.0272, 0.0287
	0201	0803	0.0636		0.0324*	1.0	-0.0195, 0.0195
	0201	0702	0.0620		-0.0243	0.2816	-0.0285, 0.0294
	2402	0801	0.0602		-0.0377*	0.3852	-0.0340, 0.0347
	0201	0304	0.0539	0.7523	-0.0443*	0.4512	-0.0322, 0.0316
328	A	DRB1					
	0201	1402	0.3438		0.0084	0.0510	-0.0401, 0.0375
	2402	1402	0.2261		0.0032	0.0292	-0.0337, 0.0396
	3101	1402	0.0640		-0.0117	0.1546	-0.0231, 0.0229
	0201	1602	0.0521		-0.0073	0.1228	-0.0238, 0.0249
	2402	1406	0.0508	0.7368	0.0204	0.3339	-0.0230, 0.0211
328	B	C					
	4801	0801	0.1761		0.1080*	0.6857	-0.0324, 0.0316
	3501	0401	0.1616		0.1325*	0.9998	-0.0246, 0.0232

#H	Locus 1 Allele	Locus 2 Allele	HF	Cum. HF	D	D'	95% Confidence Interval for D_{GF}
	5102	0801	0.1227		0.0822*	0.8779	-0.0270, 0.0269
	4005	0304	0.1001		0.0699*	0.5720	-0.0242, 0.0225
	3906	0702	0.0762		0.0623*	1.0	-0.0151, 0.0173
	2705	0202	0.0610		0.0573*	1.0	-0.0055, 0.0101
	4002	0304	0.0549		0.0398*	0.8986	-0.0163, 0.0170
	4005	0702	0.0524	0.8050	0.0245	0.4208	-0.0242, 0.0252
322	B	DRB1					
	4801	1402	0.1315		-0.0191	0.1268	-0.0302, 0.0350
	4005	1402	0.1243		0.0175	0.3388	-0.0295, 0.0300
	5102	1402	0.1216		0.0274*	0.6010	-0.0259, 0.0266
	2705	1402	0.0621		0.0182*	0.8560	-0.0192, 0.0171
	3906	1402	0.0617		0.0136	0.5841	-0.0194, 0.0175
	4002	1402	0.0565	0.5577	0.0042	0.1660	-0.0211, 0.0198
326	C	DRB1					
	0801	1402	0.1902		-0.0138	0.0676	-0.0328, 0.0368
	0702	1402	0.1575		0.0400*	0.6972	-0.0293, 0.0285
	0304	1402	0.1366		0.0027	0.0413	-0.0285, 0.0299
	0202	1402	0.0644		0.0211*	0.9983	-0.0195, 0.0158
	0803	1402	0.0543		0.0172*	0.9494	-0.0195, 0.0153
	0401	1602	0.0512	0.6542	0.0313*	0.3136	-0.0195, 0.0206
412	DRB1	DQA1					
	1402	0501	0.8180		0.0792*	0.9702	-0.0092, 0.0186
	1602	0501	0.0607	0.8787	0.0060	0.9934	-0.0111, 0.0066
412	DRB1	DQB1					
	1402	0301	0.8179		0.0772*	0.9691	-0.0093, 0.0184
	1602	0301	0.0607	0.8786	0.0059	1.0	-0.0100, 0.0064
434	DQA1	DQB1					
	0501	0301	0.8894		0.0922*	0.9997	-0.0091, 0.0132
	0401	0402	0.0530	0.9424	0.0493*	0.8776	-0.0051, 0.0090

Table 5
Selected Haplotype (HF) Frequencies and Genetic Disequilibria for 3 HLA Loci (#H = number of haplotypes in estimate, *= $p < 0.001$)

#H	Locus1 Allele	Locus2 Allele	Locus3 Allele	HF	Cum.HF	D	95% Confidence Interval for D_{GE}
322	A	B	C				
	0201	5102	0801	0.1136		0.0311*	-0.0166, 0.0168
	2402	4005	0304	0.0846		0.0350*	-0.0131, 0.0162
	2402	3501	0401	0.0779		0.0150*	-0.0132, 0.0140
	0201	3501	0401	0.0704		-0.0046	-0.0143, 0.0142
	0201	4801	0801	0.0686		-0.0268*	-0.0218, 0.0212
	2402	4801	0801	0.0564		0.0138	-0.0201, 0.0197
	0201	4801	0803	0.0518		0.0187*	-0.0093, 0.0087
	2402	3906	0702	0.0516	0.5749	0.0210*	-0.0076, 0.0106
314	A	B	DRB1				
	0201	5102	1402	0.1242		0.0145	-0.0171, 0.0162
	0201	4801	1402	0.0914		0.0149	-0.0218, 0.0199
	2402	4005	1402	0.0849		0.0166	-0.0187, 0.0171
	2402	3906	1402	0.0503	0.3508	0.0134*	-0.0105, 0.0102
318	A	C	DRB1				
	0201	0801	1402	0.1377		0.0154	-0.0250, 0.0217
	2402	0304	1402	0.1187		0.0274*	-0.0190, 0.0178
	2402	0702	1402	0.0568		0.0027	-0.0183, 0.0177
	0201	0803	1402	0.0566	0.3698	0.0080	-0.0090, 0.0104
316	B	C	DRB1				
	5102	0801	1402	0.1180		0.0215*	-0.0135, 0.0139
	4801	0801	1402	0.0794		-0.0301*	-0.0200, 0.0173
	4005	0304	1402	0.0704		-0.0014	-0.0140, 0.0128
	2705	0202	1402	0.0665		0.0188*	-0.0053, 0.0043
	3906	0702	1402	0.0617		0.0079*	-0.0076, 0.0068
	4005	0702	1402	0.0502	0.4462	0.0060	-0.0121, 0.0129
412	DRB1	DQA1	DQB1				

#H	Locus1 Allele	Locus2 Allele	Locus3 Allele	HF	Cum.HF	D	95% Confidence Interval for D _{GE}
	1402	0501	0301	0.8180		-0.0618*	-0.0071, 0.0027
	1602	0501	0301	0.0607	0.8787	-0.0047*	-0.0017, 0.0045

Table 6
Selected Haplotype (HF) Frequencies and Genetic Disequilibria for 4 HLA Loci (310 Haplotypes, $*=p < 0.001$)

A	B	C	DRB1	HF	Cum.HF	D	95%Confidence Interval for D _{GE}
0201	5102	0801	1402	0.1119		0.0591*	-0.0110, 0.0235
2402	4005	0304	1402	0.0832		0.0424*	-0.0088, 0.0128
2402	4801	0801	1402	0.0518		0.0271*	-0.0107, 0.0251
2402	3906	0702	1402	0.0509		0.0278*	-0.0049, 0.0065
0201	4801	0803	1402	0.0480		0.0233*	-0.0070, 0.0071
0201	2705	0202	1402	0.0442		0.0281*	-0.0036, 0.0043
0201	4005	0702	1402	0.0434		0.0206*	-0.0092, 0.0156
2402	3501	0401	1406	0.0432	0.4766	0.0272*	-0.0029, 0.0041

Table 7
A normalized haplotype matrix for all 3 locus HLA-A and C combinations with HLA-B*5102 present at the B locus

HLA-A	HLA-B*5102							
	HLA-C							
		0202	0303	0304	0401	0702	0801	0803
0201	0	0	0	0	0	0	0.8712	0.0790
0206	0	0	0	0	0	0	0	0
2402	0	0	0	0	0	0	0	0
3101	0	0	0	0	0	0	0.0498	0

Table 8
A normalized haplotype matrix for all 3 locus HLA-A and C combinations with HLA-B*4005 present at the B locus

HLA-A	HLA-B*4005							
	HLA-C							
		0202	0303	0304	0401	0702	0801	0803
0201	0	0	0	0.0730	0	0.3011	0	0
0206	0	0	0	0.0243	0	0	0	0
2402	0	0	0	0.5562	0	0	0	0
3101	0	0	0	0	0	0.0454	0	0

Table 9
Generations and years since mutation for HLA-B*5102 and B*4005 with 20 years per generation and scaled time units (stu) by the method of Kimura and Ohta

Effective Population Size	HLA-B*5102 (stu = .6047)		HLA-B*4005 (stu = 0.6735)	
	Generations	Years	Generations	Years
100	60.5	1209.4	67.4	1347.0
200	120.9	2418.8	134.7	2694.0
300	181.4	3628.2	202.1	4041.0
400	241.9	4837.6	269.4	5388.0
500	302.4	6047.0	336.8	6735.0
600	362.8	7256.4	404.1	8082.0
700	423.3	8465.8	471.5	9429.0
800	483.8	9675.2	538.8	10776.0
900	544.2	10884.6	606.2	12123.0
1000	604.7	12094.0	673.5	13470.0

Table 10
The cumulative proportion of no mismatched recipients when using HLA-A, B, DRB1 haplotype homozygotes as bone marrow donors in a GRIC full heritage population of size 4418, ages 15 to 65, for the 9 most common haplotypes (HF = haplotype frequency; #Hom. = expected number of homozygotes; #Het. = expected number of heterozygotes)

Haplotype	HF	#Hom.	#Het.	Total	Proportion
1. A*0201-B*5102-DRB1*1402	0.1242	68	758	826	0.187
2. A*0201-B*4801-DRB1*1402	0.0914	37	484	1347	0.305
3. A*2402-B*4005-DRB1*1402	0.0849	32	386	1765	0.400
4. A*2402-B*3906-DRB1*1402	0.0503	11	206	1983	0.449
5. A*0201-B*4002-DRB1*1402	0.0443	9	164	2156	0.488
6. A*0201-B*3501-DRB1*1406	0.0432	8	144	2308	0.522
7. A*0201-B*2705-DRB1*1402	0.0420	8	124	2440	0.552
8. A*0201-B*4005-DRB1*1402	0.0372	6	98	2544	0.576
9. A*2402-B*4801-DRB1*1402	0.0341	5	79	2629	0.595

Table 11

Expected Number of Identical Full Heritage Heterozygotes for the 9 Most Common HLA-A, B, and DRB1 Haplotypes. N = 4418 Persons

Haplotype1	Haplotype2	Expected Number
0201-5102-1402	0201-4801-1402	100
0201-5102-1402	2402-4005-1402	93
0201-5102-1402	2402-3906-1402	55
0201-5102-1402	0201-4002-1402	49
0201-5102-1402	0201-3501-1406	47
0201-5102-1402	0201-2705-1402	46
0201-5102-1402	0201-4005-1402	41
0201-5102-1402	2402-4801-1402	37
0201-4801-1402	2402-4005-1402	69
0201-4801-1402	2402-3906-1402	41
0201-4801-1402	0201-4002-1402	36
0201-4801-1402	0201-3501-1406	35
0201-4801-1402	0201-2705-1402	34
0201-4801-1402	0201-4005-1402	30
0201-4801-1402	2402-4801-1402	28
2402-4005-1402	2402-3906-1402	38
2402-4005-1402	0201-4002-1402	33
2402-4005-1402	0201-3501-1406	32
2402-4005-1402	0201-2705-1402	32
2402-4005-1402	0201-4005-1402	28
2402-4005-1402	2402-4801-1402	26
2402-3906-1402	0201-4002-1402	20
2402-3906-1402	0201-3501-1406	19
2402-3906-1402	0201-2705-1402	19
2402-3906-1402	0201-4005-1402	17
2402-3906-1402	2402-4801-1402	15
0201-4002-1402	0201-3501-1406	17
0201-4002-1402	0201-2705-1402	16
0201-4002-1402	0201-4005-1402	15
0201-4002-1402	2402-4801-1402	13
0201-3501-1406	0201-2705-1402	16
0201-3501-1406	0201-4005-1402	14
0201-3501-1406	2402-4801-1402	13
0201-2705-1402	0201-4005-1402	14
0201-2705-1402	2402-4801-1402	13
0201-4005-1402	2402-4801-1402	11