Human Immunology 80 (2019) 955–965

Contents lists available at ScienceDirect

Human Immunology

journal homepage: www.elsevier.com/locate/humimm

Next generation sequencing and the classical HLA loci in full heritage Pima Indians of Arizona: Defining the core HLA variation for North American Paleo-Indians

Robert C. Williams[a,*], William C. Knowler[a], Alan R. Shuldiner[b], Nehal Gosalia[b], Cristopher Van Hout[b], Regeneron Genetics Center[b], Robert L. Hanson[a], Clifton Bogardus[a], Leslie J. Baier[a]

[a] *Phoenix Epidemiology and Clinical Research Branch, NIH, NIDDK, Phoenix, AZ 85014, United States*
[b] *Regeneron Genetics Center, Tarrytown, NY 10591, United States*

ABSTRACT

The Pima Indians of the Gila River Indian Community in Arizona have participated in a long-range study of type 2 diabetes mellitus since 1965 and have been the subject of HLA typing and population studies since the early days of serological assays. These data have been in numerous HLA workshops and conferences and have been the source of at least five novel alleles at the classical HLA loci. In recent time nearly the entire study group was subject to next generation sequencing by whole genome or exome technologies, which has allowed us to HLA type over 3000 full heritage persons with recently developed computer algorithms. We present here the results for the classical HLA Loci: *HLA-A, B, C, DRA, DRB1, DRB3, DRB4, DRB5, DPA1, DPB1, DQA1,* and *DQB1* to the third field of resolution for synonymous alleles and type the likely four field resolution alleles from the subset of whole genome sequences. Allele frequencies, and haplotype frequencies at up to five loci, are presented as well as measures of population structure and heterozygosity. We define a core set of HLA variation that approximates the distribution for the Paleo-Indians and impute nine-locus, 4-field haplotypes that are expected to be common in full heritage peoples.

1. Introduction

Since the very early years of HLA serological typing and population studies, the Pima Indians of the Gila River Indian Community in Arizona have been represented in HLA workshops, conferences, and population studies and have been the source of at least five new HLA allele designations [1–3]. The many full Pima heritage members of the community have been designated as the best contemporary American Indian architype of the Paleo-Indians [4]. This community is closely related—historically, culturally and biologically—with the Tohono O'odham Nation in southern Arizona and many people in the Gila River Indian Community share heritages from both tribes [5–7]. The Pima Indians have also participated in a long-range study of type 2 diabetes mellitus and obesity conducted by the Intramural Program of the National Institute of Diabetes and Digestive and Kidney Diseases [8,9]. As part of this study, DNA has been isolated and characterized by genome wide association studies and whole genome and/or exome sequencing to discover the genetic component for these disorders. Our intent in this paper is to present the results of HLA typing for the classical loci from the DNA sequences of full Pima heritage persons in the Gila River Indian Community, to characterize the alleles at the third field of resolution, and to impute the four field haplotype combinations that constitute the most common HLA variation in full heritage Pima Indians. The ultimate purpose of the extended HLA typing is to provide the information needed for large, powerful HLA-disease association studies in the Pima [10–13].

2. Materials and methods

2.1. Sample

All persons with full, stated American Indian heritage from either the Pima or the Tohono O'odham tribes, whose DNA had been characterized by either whole genome or exome typing were included in the analysis and will be referred to as "Pima".

2.2. DNA sequencing

Whole-genome sequencing (30 × average coverage) was performed by Illumina FastTrack Sequencing Services (San Diego, USA) [14]. The reads were aligned to the human genome reference GRCh37 using BWA software (version 0.7.15). Whole-exome sequencing (WES) was

[*] Corresponding author at: NIH/NIDDK, Phoenix Epidemiology and Clinical Research Branch, 445 North 5th Street, Suite 210, Phoenix, AZ 85004, United States. *E-mail address:* williamsr@mail.nih.gov (R.C. Williams).

https://doi.org/10.1016/j.humimm.2019.10.002
Received 16 April 2019; Received in revised form 10 October 2019; Accepted 11 October 2019
Available online 06 November 2019

performed by Regeneron Genetics Center (Tarrytown, NY). Targeted exonic regions were captured using a slightly modified version of the xGen probe library (Integrated DNA Technologies). Captured DNA was PCR amplified and sequenced with v4 chemistry using 75 bp paired-end reads on the Illumina HiSeq 2500. WES was performed such that >85% of targeted bases are covered at >20 × depth.

2.3. HLA typing

HLA typing was performed by versions 1.7 and 2.0 of the Explore software from Omixon Corporation [15]. The algorithm HLA Explore uses in this process is called the gene filtering algorithm. The sole goal is to determine if a read pair is mappable to any of the alleles contained in the IMGT/HLA database. It will select those read pairs where at least one of the reads can be aligned with no more than three mismatches and one soft-clip to at least one allele in the database and the orientation of the mapped reads is forward-reverse if both reads in the pair are aligned. Pairs that can be mapped only with indels are also discarded. The selected reads will be saved in the FASTQ file pair.

The algorithm used here is called the statistical genotyping algorithm (SG). It is an alignment-based algorithm that aligns every read pair to all alleles in the IMGT/HLA database and selects a best matching allele pair based on various measures. The first step of the algorithm is the short-read alignment. The software aligns all read pairs to all of the sequences in the IMGT/HLA database. Read pairs with no or very few mismatches (allowing for soft clips at read ends) will get an "alignment score" for each allele depending on the number of mismatches. A read pair will be assigned to the allele(s) with the best alignment score. Any read pair that cannot be aligned to an allele without indels, or contains too many mismatches, will be discarded. Reads that align equally well to different loci (cross-mapping reads) will also be discarded.

The second step of the algorithm is the best match selection. Best matching allele pairs (or best matches) are the top allele pairs that are considered the most likely correct typing results based on allele alignment statistics. Best matching allele pairs are selected in multiple steps:

1. A short list of allele candidates is created based on supporting reads. (Filtering out alleles that do not meet the minimum requirements, eg. parts of the key exons are not covered by any reads or supported by a low number of unique reads)
2. Allele pair candidates are generated using all possible combinations of the alleles selected in the previous steps.
3. Each allele pair candidate is compared to every other allele pair candidate. (Since the IMGT/HLA database contains a high number of partially defined alleles (eg. only the exon 2 and 3 regions are defined), we have developed a sophisticated alignment scoring system to balance out the differences between the partially defined and fully defined alleles.
4. The top allele pairs are selected based on the results of the pairwise comparison. (eg. a similar percentage of bases covered but better coverage depth will be preferred)

The resulting file will contain the list of best matching alleles and a graphical representation of how the read pairs from the sample align to the reference sequences.

2.4. Statistics

Data analysis was performed at the amino acid field of resolution, unless the 3-field allele at a locus was unique. When there were synonymous alleles present in the output, then these were grouped at the amino acid field of resolution, the P-group, and are indicated in the tables with the extension ":00" at the third field to maintain the uniformity of the nomenclature. A null allele refers to the deletion of *HLA-DRB3, DRB4,* and *DRB5* on selected chromosomes and will be designated as *NL:00:00*; a blank allele is the failure of the software to return a genotype for a DNA sequence. Allele frequencies were calculated either by gene counting or the maximum likelihood method in an ABO-like model with one blank allele [16]. Haplotypes and disequilibria were computed by the estimator-maximum (EM) algorithm [17].

The estimate "D" is a standard measure of genetic disequilibrium. It is defined as $D = F(a1\text{-}b1) - F(a1)*F(b1)$, where a1 and b1 are alleles at loci A and B, *F(a1-b1)* is the observed frequency of the haplotype estimated by the EM algorithm, and the product of the allele frequencies is the expected haplotype equilibrium frequency. "D" is a simple measure of how much the observed haplotype frequency differs from the expected. As can be seen from the formula, when the disequilibrium estimate has a negative value, the observed estimate is less than the expected.

Tests for HLA population structure were performed with the Nam-Gart procedure [18].

3. Results

3.1. Ordered raw allele counts from typing software

Supplementary Tables S1–S4 have the entire distribution of alleles typed by the Omixon Explore software for the classical loci included in this paper for full heritage Pimas: S1, *HLA-A, B,* and *C*; S2, *HLA-DRB1, DQB1,* an *DPB1*; S3, *HLA-DRA, DQA1,* and DPA1; S4, *HLA-DRB3, DRB4,* and *DRB5.* The alleles are ordered by decreasing frequency based on a total of 2 N alleles at each locus and include singletons. The largest number of DNA sequences was from exome sequencing. Therefore, the data were captured from the software at the 3-field of allele resolution. What follows are descriptions of the major alleles at the loci with a category labeled "Other" that combines specificities with frequencies <0.01.

3.2. HLA-A locus

The ordered HLA-A allele counts from the whole genome and exome typing are found in Supplementary Table S1.

Allele *HLA-A*02:01:01* has the highest allele frequency in the raw count. The typing software could not distinguish between *A*02:01:01* and *A*02:197:02* and assigned the latter allele to persons who were known to be A*02:01 in previous population studies; therefore, we assigned the previous allele designation *A*02:01:01*. There were also a small number of synonymous variants of the P-group for *A*02:01*. These alleles are also represented in the allele *A*02:01:00* in the population genetics tables. It has a frequency of 0.4779 (Table 1) and is nearly identical to the frequency of *A*02:01* reported for a much smaller sample in a previous study, 0.4742. A second allele in the *HLA-A*A2* antigen group, that types unambiguously, is *A*02:06:01* with an allele frequency of 0.0574.

Second most common at HLA-A is *A*24:02:01*, which has a frequency of 0.3593 in Table 1. This includes 26 *A*24:353* alleles that could not be distinguished by the typing software from *A*24:02:01*. Allele *A*31:01:02* is the fourth common allele at this locus with allele frequency 0.0823. Alleles in the "Other" category are found in Supplementary Table S1. Together, the four major alleles at HLA-A represent over 97% of the allelic variation at the locus, which is in sharp contrast with other, non-American Indian groups, in which this locus presents many alleles with much smaller allele frequencies and much greater heterozygosity.

A test for population structure does not reject the null hypothesis of Hardy-Weinberg equilibrium. Observed heterozygosity and homozygosity fall within the 95% confidence intervals for the expected values, and the Nam-Gart statistic is not significantly different from 1.0 (Table 2).

Table 1
Class I P-Group Allele Frequencies in Full Heritage Pima Indians.

Allele	Allele Frequency	95% C.I.
HLA-A Locus, N = 3195		
02:01:00	0.4779	0.4657, 0.4902
02:06:01	0.0574	0.0517, 0.0631
24:02:01	0.3593	0.3475, 0.3711
31:01:02	0.0823	0.0756, 0.0891
Other	0.0230	0.0193, 0.0267
HLA-B Locus, N = 3202		
27:05:02	0.0867	0.0798, 0.0936
35:01:01	0.1686	0.1595, 0.1778
39:01:01	0.0412	0.0364, 0.0461
39:06:02	0.0795	0.0729, 0.0861
40:01:02	0.0214	0.0178, 0.0249
40:02:01	0.0764	0.0699, 0.0829
40:05:01	0.1676	0.1584, 0.1767
48:01:01	0.2007	0.1908, 0.2105
51:02:01	0.1321	0.1238, 0.1404
Other	0.0259	0.0220, 0.0298
HLA-C Locus, N = 3153		
02:02:02	0.0915	0.0844, 0.0986
03:04:01	0.2236	0.2133, 0.2339
04:01:01	0.1216	0.1136, 0.1297
07:02:01	0.1670	0.1578, 0.1762
08:01:01	0.2507	0.2400, 0.2614
08:03:01	0.0829	0.0761, 0.0897
Other	0.0626	0.0567, 0.0686

3.3. HLA-B locus

The major alleles that have been known to segregate at the HLA-B locus typed unambiguously at the third field of resolution in this sample (Table 1). Allele *B*48:01:01* represents >20% of the variation with a frequency of 0.2007, followed by *B*35:01:01*, 0.1686, B*40:05:01, 0.1676, and *B*51:02:01* with a frequency of 0.1321. These four alleles represent two thirds of the variation at HLA-B with a total of 0.6690, which once again demonstrates the restricted variation at class I loci in full heritage Pima. The *B*40* class of alleles has three amino acid variants while the *B*39* class has two. Alleles *B*48:01:01, B*51:02:01*, and *B*40:05:01* were partially or wholly defined from cells of the Pima tribe during numerous HLA workshops and conferences and further refined to alleles by gene sequencing and molecular analyses [19–23].

There is no evidence of population structure in the sample for HLA-B. Heterozygosity and homozygosity are in the expected ranges while the Nam-Gart statistic is not significantly different from 1.0 (Table 2). The remaining alleles, usually rare, are listed in Supplementary Table S1.

3.4. HLA-C locus

Consistent with the *HLA-A* and *HLA-B* loci, *HLA-C* also exhibits strong restriction of variation in the six defined alleles with large allele frequencies (Table 1). Four of the six alleles have a frequency >0.12, while two alleles, *C*08:01:01* and *C*03:04:01*, together represent nearly half of the allele frequency distribution at this locus, 0.4743. Allele *C*08:01:01* was also defined in a workshop and by gene sequencing from a white blood cell of the Pima Indians [24,25]. Alleles in the HLA-C Other category are found in Supplementary Table S1.

In contrast with *HLA-A* and *HLA-B*, there is evidence for population structure at *HLA-C*. The Nam-Gart statistic is significantly different from 1.0, though the lower confidence limit lies close to 1.0, 1.0438 (Table 2). This disturbance is also reflected in the observed and expected heterozygosity, 0.8069 and 0.8253, respectively, and homozygosity, 0.1931 and 0.1747. The deficit in heterozygotes could be caused by the failure of the software to detect alleles unique but undefined in the American Indian population. To estimate the frequency of so-called blank alleles, an ABO-like maximum likelihood model was fitted to the HLA-C locus distribution. It yielded a significant blank frequency of 0.0831 (95% C.I., 0.0784, 0.0878).

3.5. HLA-A, B, C haplotype frequencies and genetic disequilibria

Two-locus haplotype frequencies and their genetic disequilibria for all combinations of *HLA-A, B*, and *C* were calculated and are presented in Supplementary Tables S5–S10.

Table 3 presents the distribution of 3-locus *HLA-A, B*, and *C* haplotypes. For perspective, and to better understand the restricted variation of the class I loci in full heritage Pima, excluding alleles in the other category, we calculate the number of haplotype permutations and their theoretical uniform frequency; there is a total of 5 × 10 × 7 = 350 possible haplotypes with uniform frequency of 0.0029 (1/350). The most common haplotype is *A*02:01:00-B*51:02:01-C*08:01:01* with a frequency of 0.1086. The first 8 haplotypes together represent 55% of the total variation at these three loci. Only 22 haplotypes, 7% of the permutations, represent nearly 84% of the variation.

3.6. HLA-DRA and HLA-DRB1

The *HLA-DRA* and *HLA-DRB1* proteins form the functional heterodimer that is named *HLA-DRB1*, because *HLA-DRA* has traditionally been considered monomorphic and was ignored. However, in full heritage Pimas there is a simple, two-allele polymorphism, at the second and third fields of resolution, segregating at HLA-DRA with alleles *DRA*01:01:01*, frequency 0.1602, and *DRA*01:02:02* with frequency 0.8398 (Table 4). Observed heterozygosity and homozygosity

Table 2
Heterozygosity, Homozygosity, and Nam and Gart's Test for Population Structure at HLA Loci.

		Heterozygosity			Homozygosity			t Statistic for HLA H_0: t = 1.0	
Locus	N	Observed	95% C.I.	Expected	Observed	95% C.I.	Expected	t	95% C.I.
HLA-A	3195	0.6319	0.6152,0.6486	0.6319	0.3681	0.3514,0.3848	0.3681	0.9473	0.8779,1.0166
HLA-B	3202	0.8548	0.8426,0.8670	0.8633	0.1452	0.1330,0.1574	0.1367	1.0980	0.9941,1.2019
HLA-C	3275	0.8069	0.7931,0.8206	0.8253	0.1931	0.1794,0.2069	0.1747	1.1293[1]	1.0438,1.2148
HLA-DRA	3270	0.2722	0.2569,0.2874	0.2691	0.7278	0.7126,0.7431	0.7309	0.9887	0.9544,1.0230
HLA-DRB1	3239	0.4912	0.4740,0.5084	0.4936	0.5088	0.4916,0.5260	0.5064	1.0492	0.9581,1.1403
HLA-DRB3	3197	0.3253	0.3091,0.3415	0.3254	0.6747	0.6585,0.6909	0.6746	1.0004	0.9658,1.0351
HLA-DRB4	3160	0.1085	0.0977,0.1194	0.1298	0.8915	0.8806,0.9023	0.8702	1.1639[1]	1.1290,1.1988
HLA-DRB5	3242	0.1317	0.1201,0.1433	0.1457	0.8683	0.8567,0.8799	0.8543	1.0961[1]	1.0617,1.1306
HLA-DPA1	3269	0.0422	0.0353,0.0491	0.0426	0.9578	0.9509,0.9647	0.9574	1.0107	0.9622,1.0592
HLA-DPB1	3267	0.5852	0.5684,0.6021	0.5958	0.4148	0.3979,0.4316	0.4042	1.1432[1]	1.0839,1.2026
HLA-DQA1	1195	0.0854	0.0695,0.1012	0.1726	0.9146	0.8988,0.9305	0.8274	1.5055[1]	1.4488,1.5622
HLA-DQB1	2893	0.1977	0.1832,0.2122	0.2689	0.8023	0.7878,0.8168	0.7311	1.2646[1]	1.2282,1.3010

[1] $p < 0.05$.

Table 3
HLA-A, HLA-B, HLA-C Ordered 3-locus Haplotype Frequencies (>0.01) in Full Heritage Pimans, N = 3029.

	HLA-A	*HLA-B*	*HLA-C*	Frequency	D
1	02:01:00	51:02:01	08:01:01	0.1086	0.0336
2	24:02:01	40:05:01	03:04:01	0.0946	0.0305
3	02:01:00	48:01:01	08:03:01	0.0698	0.0262
4	24:02:01	48:01:01	08:01:01	0.0647	0.0243
5	02:01:00	27:05:02	02:02:02	0.0566	0.0127
6	24:02:01	35:01:01	04:01:01	0.0516	0.0073
7	24:02:01	39:06:02	07:02:01	0.0514	0.0179
8	02:01:00	35:01:01	04:01:01	0.0482	−0.0058
9	02:01:00	40:05:01	07:02:01	0.0324	0.0193
10	02:01:00	48:01:01	08:01:01	0.0311	−0.0356
11	02:01:00	39:06:02	07:02:01	0.0295	−0.0072
12	24:02:01	40:02:01	03:04:01	0.0287	0.0036
13	31:01:02	39:01:01	07:02:01	0.0257	0.0178
14	02:01:00	40:05:01	03:04:01	0.0219	−0.0224
15	02:01:00	40:02:01	03:04:01	0.0179	−0.0042
16	31:01:02	48:01:01	08:01:01	0.0168	0.0054
17	24:02:01	40:01:02	03:04:01	0.0165	0.0059
18	24:02:01	27:05:02	02:02:02	0.0164	−0.0121
19	02:06:01	48:01:01	08:01:01	0.0162	0.0071
20	02:06:01	35:01:01	04:01:01	0.0152	0.0060
21	31:01:02	27:05:02	02:02:02	0.0124	0.0043
22	02:01:00	35:01:01	03:04:01	0.0115	0.0113
	Total			0.8376	

Table 4
Class II P-Group Allele Frequencies in Full Heritage Pima Indians.

Allele	Allele Frequency	95% C.I.
HLA-DRA, N = 3270		
01:01:01	0.1602	0.1514, 0.1691
01:02:02	0.8398	0.8309, 0.8486
HLA-DRB1, N = 3239		
04:03:01	0.0284	0.0244, 0.0324
04:07:01	0.0244	0.0206, 0.0281
04:10:01	0.0191	0.0158, 0.0225
08:02:01	0.0392	0.0345, 0.0439
14:02:01	0.6985	0.6873, 0.7097
14:06:01	0.0971	0.0899, 0.1043
16:02:01	0.0733	0.0670, 0.0797
Other	0.0199	0.0165, 0.0233
HLA-DRB3, N = 3197		
01:01:02	0.7954	0.7855, 0.8053
NL:00:00[1]	0.2046	0.1947, 0.2145
HLA-DRB4, N = 3160		
01:03:01	0.0698	0.0635, 0.0761
NL:00:00[1]	0.9302	0.9239, 0.9365
HLA-DRB5, N = 3242		
02:02:01	0.0791	0.0725, 0.0857
NL:00:00[1]	0.9209	0.9143, 0.9275
HLA-DPA1, N = 3269		
01:03:01	0.9783	0.9747, 0.9818
02:01:01	0.0190	0.0157, 0.0223
Other	0.0028	0.0015, 0.0040
HLA-DPB1, N = 3267		
02:01:02	0.0735	0.0671, 0.0798
04:01:01	0.3468	0.3353, 0.3583
04:02:01	0.5249	0.5128, 0.5371
Other	0.0548	0.0493, 0.0603
HLA-DQA1, N = 1195		
05:03:01	0.9046	0.8928, 0.9164
Other	0.0954	0.0836, 0.1072
HLA-DQB1, N = 2893		
03:01:01	0.8400	0.8305, 0.8494
Other	0.1600	0.1506, 0.1695

[1] NL:00:00 represents the null allele.

Table 5
2-locus Class II Haplotype Frequencies (>0.01) in Full Heritage Pimans.

N = 3239			
HLA-DRA	*HLA-DRB1*	Frequency	D
01:02:02	14:02:01	0.6985	0.1098
01:02:02	14:06:01	0.0971	0.0153
01:01:01	16:02:01	0.0726	0.0610
01:02:02	08:02:01	0.0392	0.0062
01:01:01	04:03:01	0.0284	0.0239
01:01:01	04:07:01	0.0244	0.0206
01:01:01	04:10:01	0.0190	0.0160
N = 3266			
HLA-DPA1	*HLA-DPB1*		
01:03:01	04:02:01	0.5250	0.0113
01:03:01	04:01:01	0.3468	0.0073
01:03:01	02:01:02	0.0735	0.0016
N = 1114			
HLA-DQA1	*HLA-DQB1*		
05:03:01	03:01:01	0.8588	0.0532
05:03:01	Other	0.0631	−0.0532
Other	03:01:01	0.0150	−0.0532

are within the expected values while the Nam-Gart statistic is not significantly different from 1.0, indicating a close approximation to Hardy-Weinberg equilibrium (Table 2).

In contrast, there are seven alleles with significant frequency segregating at *HLA-DRB1*. One allele represents almost 70% of the variation at the locus, *DRB1*14:02:01*, with allele frequency 0.6985, one of the highest allele frequencies reported in a human population for this locus. It too was first defined in American Indian peripheral blood cells, including the Pima, and was the subject of numerous HLA workshops and conferences, genetic sequencing, and molecular analysis [2]. The second most common allele is also from the *DRB1*14* antigen class, *DRB1*14:06:01*. Antigen class *DRB1*04* has three alleles segregating at polymorphic frequencies: *04:03:01, 04:07:01*, and *04:10:01*. Heterozygosity and homozygosity fall within the expected values and the Nam-Gart statistic is not significantly different from 1.0.

Two-locus haplotype frequencies and genetic disequilibria for *HLA-DRA* and *DRB1* are found in Table 5. Allele *DRA*01:02:02* is completely associated with *DRB1*14:02:01* and *DRB1*14:06:01*.

3.7. Haplotypes for classical HLA-A, B, C, DRA, and DRB1

In the historical development of HLA, the first serological loci that were defined were *HLA-A, B, C*, and then *DRB1*. These still are the nucleus for clinical applications such as organ and bone marrow transplantation. Two-locus haplotype frequency and disequilibria tables for *A-DRB1, B-DRB1*, and *C-DRB1* are found in Supplementary Tables S11–S16. Selected three- and four-locus tables are presented in Supplementary Tables S17-S23. Table 6 lists the five-locus haplotypes for the classical loci with frequencies ≥0.01. Because of the restricted variation at *HLA-DRA* and *HLA-DRB1*, and the close linkage between them and the class I loci, while the theoretical haplotype permutations for the five loci number 5600, only 24, 0.4% of the theoretical combinations, reflect >73% of the total variation. Haplotype *A*02:01:00-B*51:02:01-C*08:01:01-DRA*01:02:02-DRB1*14:02:01* alone constitutes >10% of the 5-locus haplotype frequency in these full heritage American Indians.

3.8. HLA-DRB3, HLA-DRB4, and HLA-DRB5

HLA-DRB3, *DRB4*, and *DRB5* pose significant problems for the estimation of allele and haplotype frequencies because of the presence of

Table 6
HLA-A, HLA-B, HLA-C, HLA-DRA, HLA-DRB1 Ordered 5-locus Haplotype Frequencies (≥0.01) in Full Heritage Pimans, N = 3003.

	A	*B*	*C*	*DRA*	*DRB1*	Frequency	D
1	02:01:00	51:02:01	08:01:01	01:02:02	14:02:01	0.1051	−0.0521
2	24:02:01	40:05:01	03:04:01	01:02:02	14:02:01	0.0754	−0.0405
3	02:01:00	48:01:01	08:03:01	01:02:02	14:02:01	0.0663	−0.0240
4	02:01:00	27:05:02	02:02:02	01:02:02	14:02:01	0.0582	−0.0395
5	24:02:01	48:01:01	08:01:01	01:02:02	14:02:01	0.0512	−0.0167
6	24:02:01	39:06:02	07:02:01	01:02:02	14:02:01	0.0493	−0.0198
7	02:01:00	40:05:01	07:02:01	01:02:02	14:02:01	0.0320	0.0001
8	24:02:01	35:01:01	04:01:01	01:02:02	14:06:01	0.0265	0.0002
9	02:01:00	35:01:01	04:01:01	01:02:02	14:02:01	0.0258	−0.0612
10	31:01:02	39:01:01	07:02:01	01:02:02	14:02:01	0.0254	0.0051
11	02:01:00	39:06:02	07:02:01	01:02:02	14:06:01	0.0223	0.0005
12	24:02:01	40:02:01	03:04:01	01:02:02	14:02:01	0.0200	−0.0200
13	24:02:01	40:05:01	03:04:01	01:01:01	04:03:01	0.0198	0.0063
14	24:02:01	35:01:01	04:01:01	01:01:01	16:02:01	0.0184	0.0043
15	24:02:01	40:01:02	03:04:01	01:02:02	14:02:01	0.0165	−0.0053
16	02:01:00	40:02:01	03:04:01	01:02:02	14:02:01	0.0156	−0.0251
17	31:01:02	48:01:01	08:01:01	01:01:01	16:02:01	0.0153	0.0072
18	02:01:00	48:01:01	08:01:01	01:02:02	14:06:01	0.0139	−0.0084
19	02:01:00	48:01:01	08:01:01	01:02:02	14:02:01	0.0136	−0.0973
20	24:02:01	48:01:01	08:01:01	01:02:02	14:06:01	0.0136	−0.0011
21	02:06:01	48:01:01	08:01:01	01:02:02	08:02:01	0.0134	0.0036
22	02:01:00	35:01:01	04:01:01	01:01:01	04:07:01	0.0128	0.0031
23	02:06:01	35:01:01	04:01:01	01:02:02	14:02:01	0.0118	−0.0039
24	02:01:00	40:05:01	03:04:01	01:02:02	14:02:01	0.0108	−0.0639
	Total					0.7332	

a null allele; the loci do not appear on all chromosomes in the class II region. We have therefore defined a null allele, *NL:00:00*, to represent this absence and have assigned the null in every instance that the program failed to assign a genotype. We then applied the standard analyses described above. The presence of the null allele complicates the testing of the loci for population structure; it is difficult to distinguish the null and a blank, that is, the failure of the software to resolve an allele that is present. *HLA-DRB3* exhibits the expected heterozygosity and has a Nam-Gart statistic not significantly different from 1.0. However, there is evidence for structure in *HLA-DRB4* and *DRB5* (Table 2), but its interpretation is uncertain.

There are three major alleles for the three loci: *DRB3*01:01:02*, *DRB4*01:03:01*, and *DRB5*02:02:01* (Table 4). We combined these with *HLA-DRA* and *HLA-DRB1* to estimate the 5-locus haplotypes found in Table 7. The alleles of each locus display a characteristic pattern of linkage with the HLA-DRB1 alleles. For *DRB1*14:02:01* and *DRB1*14:06:01*, it is with *DRB3*01:01:02* with a null allele at the two other loci; for *DRB1*16:02:01* it is *DRB5*02:02:01*; and for the *DRB1*04* antigen alleles, *DRB4*01:03:01*. The seven haplotypes in Table 7 combined make up nearly 97% of the variation at these class II loci.

3.9. HLA-DPA1 and HLA-DPB1

HLA-DPA1 and *HLA-DPB1* together form a class II protein heterodimer in the HLA class II region. *HLA-DPA1* is nearly monomorphic with allele *DPA1*01:03:01* having an allele frequency of 0.9783, while three major alleles segregate at HLA-DPB1: *DPB1*04:02:01*, *DPB1*04:01:01*, and *DPB1*02:01:02* (Table 4). *HLA-DPA1* has the expected distribution of heterozygosity and homozygosity with a Nam-Gart statistic not different from 1.0, while there is evidence for population structure in HLA-DPB1, though the differences between the observed and expected values of heterozygosity homozygosity are small and the lower confidence interval value for the t-statistic, 1.1, is close to 1.0 (Table 2).

Two-locus haplotypes for the loci are found in Table 5 with *DPA1*01:03:01-DPB1*04:02:01* representing the largest combination, 0.5250, followed by *DPA1*01:03:01-DPB1*04:01:01* with a frequency of 0.3468. Three-locus haplotypes were estimated for *HLA-A-DPA1-DPB1*, *HLA-B-DPA1-DPB1*, and *HLA-C-DPA1-DPB1* and are found in Supplementary Tables S24–S26. Four-locus haplotypes for *HLA-A, B*, *C*, and *DPB1* are presented in Supplementary Table S27.

3.10. HLA-DQA1 and HLA-DQB1

Among the classical loci typed, *HLA-DQA1* and *HLA-DQB1* posed the largest problems in allele resolution, sample size, variability, and population genetics.

For HLA-DQA1, only 1195 of the >3000 full heritage Pima were

Table 7
HLA-DRA, DRB1, DRB3, DRB4, DRB5 Ordered 5-locus Haplotype Frequencies (≥0.01) in Full Heritage Pimans, N = 3038.[1]

	DRA	DRB1	DRB3	DRB4	DRB5	Frequency	D
1	01:02:02	14:02:01	01:01:02	NL:00:00	NL:00:00	0.7017	NA
2	01:02:02	14:06:01	01:01:02	NL:00:00	NL:00:00	0.0988	−0.1451
3	01:01:01	16:02:01	NL:00:00	NL:00:00	02:02:01	0.0708	0.0141
4	01:02:02	08:02:01	NL:00:00	NL:00:00	NL:00:00	0.0402	−0.0218
5	01:01:01	04:03:01	NL:00:00	01:03:01	NL:00:00	0.0233	0.0048
6	01:01:01	04:07:01	NL:00:00	01:03:01	NL:00:00	0.0182	0.0038
7	01:01:01	04:10:01	NL:00:00	01:03:01	NL:00:00	0.0151	0.0032
	Total					0.9680	

[1] NL:00:00 represents the null allele.

resolved by the Omixon Explore software. The highest frequency allele in the raw output was *DQA1*05:03:01* (Supplementary Table S3). It is a member of the *DQA1*05:01P* protein group that also includes alleles *DQA1*05:03* and *DQA1*05:05* and their synonymous variants at the third field of resolution. Therefore, they were pooled with *DQA1*05:03:01*, with a pooled frequency of 0.9046, for the further analyses and estimates of haplotype frequencies.

While entries in Table 2 for *HLA-DQA1* suggest that there is significant population structure at this locus, with many fewer heterozygotes and more homozygotes than expected, this is a bit artificial because of the dominant allele and the many minor alleles that segregate at this locus. If one looks solely at the expected and observed frequencies of the 1030 *DQA1*05:03:01* homozygotes, which have an observed genotype frequency of 0.8619, and computes a goodness of fit chi-square statistic, it would contribute only 2.8 to the total chi-square. It is the presence of the Other, mostly rare, alleles that upset the population structure statistics.

HLA-DQB1 has an even more complex structure. In the raw data (Supplementary Table S2) there are 119 different *DQB1* alleles reported. However, only 8 have a raw frequency ≥0.01. The largest of these is *DQB1*03:01:01*, which is in the *DQB1*03:01P* group that also contains, in addition to the *DQB1*03:01* synonymous variants at the third field of resolution, *DQB1*03:22*, *DQB1*03:29*, and *DQB1*03:264*. The frequencies of these alleles were pooled into the *DQB1*03:01:01* category, with pooled frequency 0.8400 (Table 4), and used for the further analyses.

With this complex distribution of rare alleles at DQB1, many represented by only 1 or 2 alleles among the 2893 persons typed, it is to be expected that the goodness of fit tests would reject the hypothesis of Hardy-Weinberg equilibrium. Once again there is an excess of homozygotes and a much lower heterozygosity than expected. The Nam-Gart test rejects the null hypothesis with a 95% confidence interval that excludes 1.0.

The distribution for *DQA1-DQB1* haplotypes is found in Table 5. Haplotype *DQA1*05:03:01-DQB1*03:01:01* has the largest frequency, 0.8588. In Supplementary Tables S28–S30 are found 3-locus haplotypes with their frequencies and disequilibria for *HLA-A-DQA1-DQB1*, *HLA-B-DQA1-DQB1*, and *HLA-C-DQA1-DQB1*. Supplementary Table S31 presents the 4-locus combinations for *HLA-A, B, C*, and *DQB1*. Given the nearly monomorphic nature of *DQA1* and *DQB1*, the variation of the distributions is determined primarily by the class I loci. Supplementary Table S32 lists the major haplotypes for *HLA-DRA, DRB1, DQB1, and DPB1* with haplotype *DRA*01:02:02-DRB1*14:02:01-DQB1*03:01:01-DPB1*04:02:01* representing one third of the frequency variation, 0.3337, and a haplotype identical at the first three loci with *DPB1*04:01:01* the second most common combination with a frequency of 0.2538.

3.11. 5-Locus haplotypes for the primary classic polymorphic loci

Using the allele definitions and frequencies in Tables 1 and 4, selected 5-locus haplotype frequencies and disequilibria were computed. Table 8 presents the major polymorphic class II loci: *DRA*, *DRB1*, *DQB1*, *DPB1*, and *DRB3*. Two combinations make up over 60% of the variation at class II in full heritage Pima: *DRA*01:02:02-DRB1*14:02:01-DQB1*03:01:01-DPB1*04:02:01-DRB3*01:01:02*; the second haplotype differs from the first by substituting *DPB1*04:01:01* for the *04:02:01* allele (Table 8). To illustrate how the class II haplotypes combine with the class I loci, *DRA*, *DRB1*, *DQB1*, and *DPB1* were successively computed with *HLA-A*, *B*, and *C* and are found in Tables 9–11. Here the variation of these haplotypes is very much greater because of the increased polymorphism of the class I loci.

3.12. HLA alleles at 4-field resolution

Typing over 400 of the community for whole genome sequences gave us the intronic and intergenic variation that allowed us to impute the likely 4-field resolution alleles for each of the 3-field names. Table 12 lists the corresponding alleles that were found and the number of 4-field specificities that were resolved. In select cases, as for *DRB1*14:06:01*, there was no corresponding 4-field allele.

4. Discussion

4.1. Agreement and Kappa scores for old and new methods

As with any new typing method, resolving HLA alleles with a computer algorithm from whole genome and exome sequences requires a quality control assessment. Therefore, we compared the allelic genotypes of the new method with six loci that we reported earlier from different, DNA-based, methodologies (Table 13) [3]. At *HLA-A*, *B*, and *C*, there was a very high agreement between the first field of resolution, the antigen, each being from about 95–99%, with appropriately high Kappa scores. There was a lower agreement in the second field of amino acid differences because the whole genome and exome sequences contain more information than the earlier procedures. Still, the respective field 2 agreements of 90–97% are high, along with the accompanying Kappa scores. *HLA-DRB1* had a first field agreement of 98.9% but only an 91.2% agreement on the amino acid field of resolution. The additional information in the whole genome and exome sequences allowed a better differentiation of *DRB1*14:06* from *DRB1*14:02*. Locus *HLA-DQA1* had the fewest genotypes to compare but a high agreement of 97.1% for each field of resolution. Allele *DQA1*05:01* was reported as the predominant allele in the previous report [3], but *DQA1*05:03* in this one (Table 4). This is explained by allele *DQA1*05:03* being part of the *DQA1*05:01P* group with an identical amino sequence in the peptide binding groove of the protein. The information of the DNA sequences resolved the additional difference away from this structure. *HLA-DQB1*, while the antigen field agreement was relatively high, 96.3%, the amino acid field of resolution was only 89.6%. This follows from the difficulty we found in typing the locus and its extreme heterozygosity when typed from exomes.

In order to determine the agreement and Kappa scores for duplicate testing with the Omixon Explore program, at 3-field resolution, we retyped about 10% of the original *.fastq* files and then, using the same Omixon Explore typing database, db3.35.0.8, we copied their raw exome sequences to a new directory where we applied the complete typing pipeline again to create a duplicate set of results. From these we chose 500 random persons to compute the agreement and Kappa scores for the repeated genotypes at each locus. The results are presented in Table S33. Over 6000 genotypes and 12,000 alleles, there was not a single discordancy.

4.2. Population structure

We made no attempt to control for relatedness in our sample of full heritage Pima and used every genotype of every person that was resolved by the software. Among the core classical *HLA-A, B, C*, and *DRB1* loci, only *HLA-C* showed evidence for structure. Even here, the lower confidence interval value for the t-statistic was 1.04, slightly larger than the expected value of 1.0. This situation illustrates a problem with a history that goes back to the serological days, blank alleles. When there was an antigen present on the surface of the cell, but no antibody in the tray to attach to it, then the antigen was not resolved and became a silent, recessive, “blank” allele. When typing with whole genomes and exomes, if there is not a sequence in the typing database that corresponds to the one found, then it too will not be reported and become a

Table 8
HLA-DRA, DRB1, DQB1, DPB1, DRB3 Ordered 5-locus Haplotype Frequencies (≥0.01) in Full Heritage Pimans, N = 2705.

	DRA	*DRB1*	*DQB1*	*DPB1*	*DRB3*	Frequency	D
1	01:02:02	14:02:01	03:01:01	04:02:01	01:01:02	0.3472	−0.2668
2	01:02:02	14:02:01	03:01:01	04:01:01	01:01:02	0.2652	−0.3600
3	01:02:02	14:02:01	03:01:01	02:01:02	01:01:02	0.0627	−0.1001
4	01:01:01	16:02:01	03:01:01	04:02:01	NL:00:00[1]	0.0425	0.0017
5	01:02:02	14:06:01	03:01:01	04:01:01	01:01:02	0.0362	−0.0495
6	01:02:02	14:06:01	03:01:01	04:02:01	01:01:02	0.0354	−0.0236
7	01:01:01	16:02:01	03:01:01	04:01:01	NL:00:00	0.0173	−0.0010
						0.8314	

[1] *NL:00:00* represents the null allele.

blank allele. One way to address this is to estimate the allele frequencies at the *HLA-C* locus with an ABO-like maximum likelihood model that has one recessive allele and multiple codominant ones and to assess the frequency of the blank. For *HLA-C*, the blank allele frequency was large, 0.0831 (95% C.I. 0.0784, 0.0878). This suggests that the efficiency of the typing at this locus was compromised either by new alleles not in the data base or by other technical issues of the software or the DNA sequencing in the *HLA-C* region. Actual population structure at *HLA-C* cannot also be excluded; but its close linkage with *HLA-B*, and the absence of structure for *HLA-B*, would argue against it.

Outside the core group of loci, there was also population structure at *HLA-DRB4, DRB5, DPB1, DQA1, and DQB1*. At *HLA-DRB4* and *DRB5* there is the complication of both null and blank alleles. A null allele occurs when a locus is absent from selected chromosomes in the *DRBX* region, whereas a blank is a failure to type an allele that is present. The problem is differentiating the null and blank allele when no genotype is reported for a locus. Testing these loci for structure is, therefore, problematic. *HLA-DPB1* and *DQB1* are among the most polymorphic of the class II loci and have many minor alleles segregating in full heritage Pima Indians. When there are one or two alleles with high frequency, and many with low frequencies, some existing in only three, two, or one copy (Table S2), then calculating goodness of fit statistics on the entire array becomes nearly impossible because for each allele, the expected and observed genotypes will not have a consistent pattern or number. Nearly all the rare alleles will be present only in heterozygotes. In addition, there may not be a uniform distribution of the rare alleles across the common variants, which makes grouping of the minor alleles in a “Other” category only a partial solution to the problem.

4.3. Variation in common American Indian haplotypes

The presence of over 400 whole genome sequences in this data base allowed us to type HLA alleles at the highest field of resolution by using the intergenic and intronic variation (Table 12). Combined with the haplotype frequencies calculated over many permutations of the loci, we impute what are the common haplotypes at the highest resolution of HLA variation in America’s first people (Table 14). Genetic drift and natural selection will vary the frequencies and occurrences of these combinations in each American Indian tribe that descends from the Paleo-Indian migration.

4.4. Core HLA variation in North American Indians

In all HLA population reports in the Americas, a fundamental confounder is non-native genetic admixture. In our center we have studied a population that has been relatively isolated in the Southwest United States since European contact and has little genetic admixture. We also have a control for admixture in the stated-proportion American Indian variable, in our ability to estimate individual non-Indian genetic admixture from vectors of allele frequencies, and our demonstration that the stated-statistic is a reliable measure of ethnicity [26–28]. We have included in this study only those persons who state that they are full heritage Pima. Therefore, the resulting vector of alleles at each of the major loci are core sets that represent the North American Paleo-Indian component for HLA and that are expected to be present in varying frequencies in all population studies that include some significant proportion of American Indians. Four examples illustrate the point.

A population report of 302 Lakota Sioux for the American Society for Histocompatibility and Immunogenetics Minority Workshops was performed in 2004 [29]. They are about 2000 miles removed from the Pima and inhabit a very different ecosystem. Yet in Table 4, page 82, in which the allele frequencies are in descending frequency order, the first four alleles at *HLA-A* are *A*02:01, A*24:02, A*31:01,* and *A*02:06*, the exact vector of common alleles in the Pima, the core variation at this locus. The remaining frequencies are likely genetic admixture, which was revealed in earlier studies in this population with a sensitive marker for gene flow, the gamma marker loci, Gm (data not shown). At

Table 9
HLA-A, DRA, DRB1, DQB1, DPB1 Ordered 5-locus Haplotype Frequencies in Full Heritage Pimans, N = 2815, with frequency ≥ 0.01.

	A	*DRA*	*DRB1*	*DQB1*	*DPB1*	Frequency	D
1	02:01:00	01:02:02	14:02:01	03:01:01	04:02:01	0.1715	−0.0657
2	02:01:00	01:02:02	14:02:01	03:01:01	04:01:01	0.1378	−0.1679
3	24:02:01	01:02:02	14:02:01	03:01:01	04:02:01	0.1069	−0.0515
4	24:02:01	01:02:02	14:02:01	03:01:01	04:01:01	0.0948	−0.0500
5	02:01:00	01:02:02	14:02:01	03:01:01	02:01:02	0.0391	−0.0495
6	31:01:02	01:02:02	14:02:01	03:01:01	04:02:01	0.0347	−0.0127
7	24:02:01	01:02:02	14:06:01	03:01:01	04:02:01	0.0287	−0.0055
8	02:01:00	01:02:02	14:06:01	03:01:01	04:01:01	0.0271	−0.0073
9	24:02:01	01:01:01	16:02:01	03:01:01	04:02:01	0.0236	−0.0151
10	02:06:01	01:02:02	14:02:01	03:01:01	04:02:01	0.0219	−0.0067
11	24:02:01	01:02:02	14:02:01	03:01:01	02:01:02	0.0204	−0.0146
12	24:02:01	01:01:01	16:02:01	03:01:01	04:01:01	0.0150	−0.0099
13	31:01:02	01:01:01	16:02:01	03:01:01	04:02:01	0.0147	−0.0014
						0.7363	

Table 10
HLA-B, DRA, DRB1, DQB1, DPB1 Ordered 5-locus Haplotype Frequencies in Full Heritage Pimans, N = 2812, with frequency ≥ 0.01.

	B	*DRA*	*DRB1*	*DQB1*	*DPB1*	Frequency	D
1	48:01:01	01:02:02	14:02:01	03:01:01	04:02:01	0.0912	−0.0609
2	40:05:01	01:02:02	14:02:01	03:01:01	04:02:01	0.0674	−0.0144
3	51:02:01	01:02:02	14:02:01	03:01:01	04:01:01	0.0664	−0.0713
4	27:05:02	01:02:02	14:02:01	03:01:01	04:02:01	0.0631	−0.0431
5	35:01:01	01:02:02	14:02:01	03:01:01	04:01:01	0.0462	0.0226
6	40:05:01	01:02:02	14:02:01	03:01:01	04:01:01	0.0413	−0.0140
7	40:02:01	01:02:02	14:02:01	03:01:01	04:01:01	0.0353	−0.0372
8	51:02:01	01:02:02	14:02:01	03:01:01	02:01:02	0.0306	−0.0328
9	39:01:01	01:02:02	14:02:01	03:01:01	04:02:01	0.0282	−0.0170
10	39:06:02	01:02:02	14:02:01	03:01:01	04:01:01	0.0274	−0.0480
11	39:06:02	01:02:02	14:02:01	03:01:01	04:02:01	0.0242	−0.0074
12	40:02:01	01:02:02	14:02:01	03:01:01	04:02:01	0.0214	−0.0090
13	48:01:01	01:01:01	16:02:01	03:01:01	04:02:01	0.0211	−0.0063
14	48:01:01	01:02:02	14:02:01	03:01:01	02:01:02	0.0182	−0.0191
15	39:06:02	01:02:02	14:06:01	03:01:01	04:01:01	0.0164	−0.0065
16	51:02:01	01:02:02	14:02:01	03:01:01	04:02:01	0.0163	−0.0164
17	35:01:01	01:01:01	16:02:01	03:01:01	04:02:01	0.0161	−0.0051
18	35:01:01	01:02:02	14:02:01	03:01:01	04:02:01	0.0156	0.0425
19	35:01:01	01:02:02	14:06:01	03:01:01	04:02:01	0.0149	−0.0088
20	48:01:01	01:02:02	14:06:01	03:01:01	04:02:01	0.0141	−0.0112
21	40:01:02	01:02:02	14:02:01	03:01:01	04:01:01	0.0137	−0.0160
22	35:01:01	01:01:01	16:02:01	03:01:01	04:01:01	0.0135	−0.0039
23	48:01:01	01:02:02	14:02:01	03:01:01	04:01:01	0.0134	−0.0237
						0.7158	

HLA-B the common alleles are *B*35:01, B*39:01, B*51:01, B*27:05, B*15:01, B*40:01, B*40:02,* and *B*48:01*, seven of which are part of the Pima core set, at least at the first field of resolution. Alleles *B*51:02* and *B*40:05* in the Pima are post-migration mutations, as we discussed in an earlier paper [3]. At *HLA-C* all six Pima core alleles are found at polymorphic levels (>0.01) with *C*07:02*, *C*03:04:01*, and *C*04:01:01* being among the most common. Among the Lakota Sioux at *HLA-DRB1*, alleles *DRB1*04:07*, *DRB1*14:02*, and *DRB1*16:02:01* have frequencies >0.10, all part of the Pima core set, while alleles *DRB1*04:03* and *DRB1*08:02* also segregate at polymorphic frequencies.

A second sample has its origin in the southern Dene people, the Navajo, which has traditionally been considered distinct in genetic history from the Paleo-Indians. In an unpublished HLA population study of 184 persons (data not shown), the *HLA-A* locus has, in order of frequency, *A*02:01, A*24:02, A*31:01,* and A**02:06*, which, together, make up >94% of the frequency variation at this locus. At *HLA-B* the most common allele is *B*27:05*, followed by *B*35:01, B*40:02, B*48:01, B*51:02, B*39:06*, and *B*39:01*, which are all part of the core set in Pima Indians and constitute >85% of the variation at this locus. In addition, alleles *B*15:01* and *B*15:07* have significant frequency in the Navajo whereas in the Pima *B*15:01* only has a frequency of 0.0020 (Table S1) and *B*15:07* is not found. Locus *HLA-C* in the Navajo is represented by the 6 core alleles found in the Pima in Table 1, with a combined frequency >86%. Allele *C*01:02* is also found at >10% frequency in the Navajo but only 0.33% in the Pima (Table S1). As in the Pima core set, the most common allele in the Navajo at the *HLA-DRB1* locus is *DRB1*14:02*, followed by *DRB1*08:02, DRB1*16:02, DRB1*04:07, DRB1*04:03, DRB1*14:06, and DRB1*14:10.* Together they comprise >80% of the allelic variation in Navajo. The Navajo also have a high allele frequency of *DRB1*14:01*, >12%, which was not found in the >3000 Pima Indians who were typed. The most common 4-locus haplotype in the Navajo is *A*02:01-B*35:01-C*04:01-DRB1*14:02* with a frequency >8%, all alleles of which are found in the Pima core set.

This core variation is also found in two recent publications outside North America. In 1101 Ecuadorians with mixed ancestry, and stratified

Table 11
HLA-C, DRA, DRB1, DQB1, DPB1 Ordered 5-locus Haplotype Frequencies in Full Heritage Pimans, N = 2780, with frequency ≥ 0.01.

	C	*DRA*	*DRB1*	*DQB1*	*DPB1*	Frequency	D
1	07:02:01	01:02:02	14:02:01	03:01:01	04:02:01	0.0886	−0.0540
2	03:04:01	01:02:02	14:02:01	03:01:01	04:01:01	0.0852	−0.0380
3	08:01:01	01:02:02	14:02:01	03:01:01	04:01:01	0.0719	−0.0713
4	02:02:02	01:02:02	14:02:01	03:01:01	04:02:01	0.0662	−0.0449
5	08:03:01	01:02:02	14:02:01	03:01:01	04:02:01	0.0655	−0.0552
6	03:04:01	01:02:02	14:02:01	03:01:01	04:02:01	0.0544	0.0023
7	08:01:01	01:02:02	14:02:01	03:01:01	02:01:02	0.0474	−0.0474
8	08:01:01	01:02:02	14:02:01	03:01:01	04:02:01	0.0431	−0.0224
9	07:02:01	01:02:02	14:02:01	03:01:01	04:01:01	0.0338	−0.0726
10	04:01:01	01:02:02	14:02:01	03:01:01	04:01:01	0.0294	0.0206
11	07:02:01	01:02:02	14:06:01	03:01:01	04:01:01	0.0178	−0.0019
12	08:01:01	01:01:01	16:02:01	03:01:01	04:02:01	0.0165	−0.0081
13	08:01:01	01:02:02	14:06:01	03:01:01	04:02:01	0.0155	0.0048
14	04:01:01	01:02:02	14:06:01	03:01:01	04:02:01	0.0140	−0.0090
15	04:01:01	01:01:01	16:02:01	03:01:01	04:02:01	0.0136	−0.0031
16	04:01:01	01:02:02	14:02:01	03:01:01	04:02:01	0.0115	0.0386
17	04:01:01	01:01:01	16:02:01	03:01:01	04:01:01	0.0114	−0.0022
	Total					0.6858	

Table 12
Four-Field Resolution Alleles from Whole Genome Sequences in Full Heritage Pima, N = 351, (number of alleles).

3-Field	4-Field Resolution from Whole Genome Sequences (Number alleles)
HLA-A	
02:01:01	02:01:01:01 (337)
02:06:01	02:06:01:01 (38)
24:02:01	24:02:01:01 (255)
31:01:02	31:01:02:01 (63)
HLA-B	
27:05:02	27:05:02:01 (68)
35:01:01	35:01:01:05 (101)
39:01:01	39:01:01:03 (30)
39:06:02	39:06:02:02 (62)
40:01:02	40:01:02:01 (1), 40:01:02:04 (10)
40:02:01	40:02:01:01 (44), 40:02:01:02 (1)
40:05:01	40:05:01:01 (119)
48:01:01	48:01:01:01 (146)
51:02:01	51:02:01:01 (91)
HLA-C	
02:02:02	02:02:02:01 (68)
03:03:01	03:03:01:01 (9)
03:04:01	03:04:01:02 (143), 03:04:01:01 (11)
04:01:01	04:01:01:11 (98)
07:02:01	07:02:01:01 (125), 07:02:01:03 (1)
08:01:01	08:01:01:01 (179)
08:03:01[1]	08:03:01 (60)
HLA-DRA	
01:01:01	01:01:01:01 (72), 01:01:01:02 (1), 01:01:01:03 (41)
01:02:02[1]	01:02:02 (552)
01:02:03[1]	01:02:03 (34)
HLA-DRB1	
04:03:01	04:03:01:01 (24), 04:03:01:02 (2)
04:07:01	04:07:01:02 (11)
04:10:01[1]	04:10:01 (3)
08:02:01	08:02:01:01 (33)
14:02:01	14:02:01:02 (441)
14:06:01[1]	14:06:01 (102)
16:02:01	16:02:01:03 (59)
HLA-DPA1	
01:03:01	01:03:01:05 (383), 01:03:01:02 (236), 01:03:01:01 (49) 01:03:01:03 (15), 01:03:01:04 (2)
02:01:01	02:01:01:02 (12), 02:01:01:01 (1)
HLA-DPB1	
02:01:02	02:01:02:05 (26), 02:01:02:01 (14), 02:01:02:32 (8), 02:01:02:04 (3)
04:01:01	04:01:01:01 (159), 04:01:01:04 (64)
04:02:01	04:02:01:02 (369)
HLA-DQA1	
03:01:01[1]	03:01:01 (25)
03:03:01	03:03:01:03 (13), 03:03:01:01 (2)
04:01:01	04:01:01:02 (40), 04:01:01:03 (3)
05:03:01	05:03:01:02 (321), 05:03:01:01 (232)
05:05:01	05:05:01:03 (60)
HLA-DQB1	
03:01:01	03:01:01:01 (612)
03:02:01	03:02:01:01 (26)
04:01:01	04:01:01:01 (12)
04:02:01	04:02:01:01 (43), 04:02:01:05 (1), 04:02:01:06 (1)
HLA-DRB3	
01:01:02	01:01:02:01 (548)
HLA-DRB4	
01:03:01	01:03:01:03 (32), 01:03:01:01 (11)
HLA-DRB5	
01:01:01[1]	01:01:01 (4)
02:02:01[1]	02:02:01 (59)

[1] No fourth field of resolution was reported by the typing software.

in three geographical areas, while 20 alleles were reported, the largest frequencies at the *HLA-A* locus were *A*02*, *A*24*, and *A*31*, which likely represent *A*02:01*, *A*02:06*, *A*24:02*, and *A*31:01*, the core vector of alleles at *HLA-A* in the Pima, with the balance of the reported alleles being admixture from European, African, and other groups [30]. At *HLA-B*, 46 alleles are reported but the most common represent the Pima core at this locus: *B*27, B*35, B*39, B*40(*40:01, *40:02), B*48, and B*51*. For *HLA-DRB1* the most frequent allele, as in the Pima, is *DRB1*14* with a frequency of 0.1735 along with significant frequencies for *DRB1*04, DRB1*08, and DRB1*16*, all of which mirrors the core alleles in Pima Indians.

As a second example, the NGS typing of 112 Mestizos from Oaxaca in Southeastern Mexico—a population that would roughly correspond to Mexican Americans in the United States with components of American Indian, European, and African genetic admixture—the core set of Pima alleles reflect the Native American fraction [31]. At *HLA-A*, among the most frequent alleles are found *A*02:01:01*, *A*24:02:01*, *A*02:06:01*, and *A*31:01:02*. At HLA-B and HLA-C, the authors report 46 and 28 alleles present, respectively, in this highly admixed group; the Pima core set is represented at each locus, except *C*08:03:01*, which absence could be explained by the combined action of genetic admixture and genetic drift. Among the 10 most frequent alleles at *HLA-DRB1* in the sample, six are represented in the Pima core set: **04:07:01, *08:02:01, *16:02:01, *14:06:01, *04:03:01*, and **14:02:01*.

We are then ready to define the likely core variation in North American Paleo-Indians (Table 15). The first field of HLA nomenclature will be most certain, while the second, amino acid field, is less so. Since the first people entered North America, and even more since European contact >500 years ago, the evolutionary mechanisms of mutation, natural selection, gene flow, and genetic drift have worked to create many highly diverse populations at the HLA loci throughout North, Central, and South America. There is evidence that mutation has created new alleles, particularly at *HLA-B* [3,9]. We have recently reported the large differences in the allele frequencies at *HLA-DRB1* between the Arizona Pima, represented in this report, and their closely related Mexican Pima, the Maycoba, which might reflect the action of genetic drift and/or natural selection [32]. It is therefore impossible to know with exactness what the first HLA distributions were. Also, we do not claim that this core set is complete and exclusive. As more Native American tribes are studied, the core set may very well expand. But the echoes of the first peoples are present in their genomes and we feel that the closest approximation of first peoples can currently be found by consulting the HLA distribution in the Pima Indians of Southern Arizona.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

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 supported in part by the Intramural Research Program of the NIH, The National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK).

Regeneron Genetics Center Banner Author List and Contribution Statements

All authors/contributors are listed in alphabetical order. **RGC Management and Leadership Team:** Goncalo Abecasis, Ph.D., Aris Baras, M.D., Michael Cantor, M.D., Giovanni Coppola, M.D., Aris Economides, Ph.D., John D. Overton, Ph.D., Jeffrey G. Reid, Ph.D., Alan R. Shuldiner, M.D. **Sequencing and Lab Operations**: Christina Beechert, Caitlin Forsythe, M.S., Erin D. Fuller, Zhenhua Gu, M.S.,

Table 13
Comparison of 2009 HLA Oligonucleotide Typing (*Williams et al. Tissue Antigens, 2009*) with Whole Genome and Exome Typing by Omixon Explore.

	Resolution					
	Amino Acid Field (2)			Antigen Field (1)		
HLA Locus	Alleles	Agreement %	Kappa (95% C.I.)	Alleles	Agreement %	Kappa (95% C.I.)
A	464	93.0	0.85 (0.81, 0.90)	476	96.8	0.94 (0.91, 0.97)
B	434	96.7	0.94 (0.90, 0.97)	434	98.6	0.97 (0.95, 0.99)
C	280	96.4	0.93 (0.89, 0.97)	292	94.9	0.90 (0.85, 0.95)
DRB1[1]	560	91.2	0.83 (0.78, 0.87)	560	98.9	0.99 (0.96, 1.00)
DQA1	104	97.1	0.94 (0.88, 1.00)	104	97.1	0.94 (0.88, 1.00)
DQB1	298	89.6	0.79 (0.72, 0.86)	298	96.3	0.93 (0.88, 0.97)

[1] The reduction in the correlation was primarily the better differentiation of DRB1*14:06 from DRB1*14:02 in the whole genome and exome typing.

Table 14
Imputed Nine Locus HLA Haplotypes Common in Full Heritage American Indians.

A	*B*	*C*	*DRA*	*DRB1*	*DPA1*	*DPB1*	*DQA1*	*DQB1*
02:01:01:01	51:02:01:01	08:01:01:01	01:02:02	14:02:01:02	01:03:01:05	04:01:01:01	05:03:01:02	03:01:01:01
24:02:01:01	40:05:01:01	03:04:01:02	01:02:02	14:02:01:02	01:03:01:05	04:01:01:01	05:03:01:02	03:01:01:01
02:01:01:01	48:01:01:01	08:03:01	01:02:02	14:02:01:02	01:03:01:05	04:02:01:02	05:03:01:02	03:01:01:01
02:01:01:01	27:05:02:01	02:02:02:01	01:02:02	14:02:01:02	01:03:01:05	04:02:01:02	05:03:01:02	03:01:01:01
24:02:01:01	48:01:01:01	08:01:01:01	01:02:02	14:02:01:02	01:03:01:05	04:02:01:02	05:03:01:02	03:01:01:01
24:02:01:01	39:06:02:02	07:02:01:01	01:02:02	14:02:01:02	01:03:01:05	04:02:01:02	05:03:01:02	03:01:01:01
24:02:01:01	35:01:01:05	04:01:01:11	01:02:02	14:06:01	01:03:01:05	04:02:01:02	05:03:01:02	03:01:01:01
02:01:01:01	35:01:01:05	04:01:01:11	01:02:02	14:02:01:02	01:03:01:05	04:01:01:01	05:03:01:02	03:01:01:01
24:02:01:01	35:01:01:05	04:01:01:11	01:01:01:01	16:02:01:03	01:03:01:05	04:02:01:02	05:03:01:02	03:01:01:01
31:01:02:01	39:01:01:03	07:02:01:01	01:02:02	14:02:01:02	01:03:01:05	04:02:01:02	05:03:01:02	03:01:01:01
02:01:01:01	39:06:02:02	07:02:01:01	01:02:02	14:06:01	01:03:01:05	04:01:01:01	05:03:01:02	03:01:01:01
24:02:01:01	40:02:01:01	03:04:01:02	01:02:02	14:02:01:02	01:03:01:05	04:02:01:02	05:03:01:02	03:01:01:01
24:02:01:01	40:05:01:01	03:04:01:02	01:01:01:01	04:03:01:01	01:03:01:05	04:02:01:02	05:03:01:02	03:01:01:01
24:02:01:01	35:01:01:05	04:01:01:11	01:01:01:01	16:02:01:03	01:03:01:05	04:02:01:02	05:03:01:02	03:01:01:01
24:02:01:01	40:01:02:04	03:04:01:02	01:02:02	14:02:01:02	01:03:01:05	04:01:01:01	05:03:01:02	03:01:01:01
02:01:01:01	40:02:01:01	03:04:01:02	01:02:02	14:02:01:02	01:03:01:05	04:01:01:01	05:03:01:02	03:01:01:01
31:01:02:01	48:01:01:01	08:01:01:01	01:01:01:01	16:02:01:03	01:03:01:05	04:02:01:02	05:03:01:02	03:01:01:01
02:01:01:01	48:01:01:01	08:01:01:01	01:02:02	14:06:01	01:03:01:05	04:01:01:01	05:03:01:02	03:01:01:01
02:01:00:01	48:01:01:01	08:01:01:01	01:02:02	14:02:01:02	01:03:01:05	04:01:01:01	05:03:01:02	03:01:01:01
24:02:01:01	48:01:01:01	08:01:01:01	01:02:02	14:06:01	01:03:01:05	04:02:01:02	05:03:01:02	03:01:01:01
02:06:01:01	48:01:01:01	08:01:01:01	01:02:02	08:02:01:01	01:03:01:05	04:02:01:02	05:03:01:02	03:01:01:01
02:01:01:01	35:01:01:05	04:01:01:11	01:01:01:01	04:07:01:02	01:03:01:05	04:01:01:01	05:03:01:02	03:01:01:01

Table 15
Core HLA Variation in North American Paleo-Indians Imputed from the Arizona Pima.

HLA-A		*HLA-B*		*HLA-C*		*HLA-DRB1*	
Field 1	Field 2	Field 1	Field 2	Field 1	Field 2	Field 1	Field 2
*02	*02:01, *02:06	*27	*27:05	*02	*02:02	*04	*04:03, *04:07
*24	*24:02	*35	*35:01	*03	*03:04	*08	*08:02
*31	*31:01	*39	*39:01, *39:06	*04	*04:01	*14	*14:02
		*40	*40:01, *40:02	*07	*07:02	*16	*16:02
		*48	*48:01	*08	*08:01, *08:03		
		*51	*51:01				

Michael Lattari, Alexander Lopez, M.S., John D. Overton, Ph.D., Thomas D. Schleicher, M.S., Maria Sotiropoulos Padilla, M.S., Karina Toledo, Louis Widom, Sarah E. Wolf, M.S., Manasi Pradhan, M.S., Kia Manoochehri, Ricardo H. Ulloa. **Genome Informatics**: Xiaodong Bai, Ph.D., Suganthi Balasubramanian, Ph.D., Leland Barnard, Ph.D., Andrew Blumenfeld, Yating Chai, Ph.D., Gisu Eom, Lukas Habegger, Ph.D., Young Hahn, Alicia Hawes, B.S., Shareef Khalid, Jeffrey G. Reid, Ph.D., Evan K. Maxwell, Ph.D., John Penn, M.S., Jeffrey C. Staples, Ph.D., Ashish Yadav, M.S. **Planning, Strategy, and Operations:** Paloma M. Guzzardo, Ph.D., Marcus B. Jones, Ph.D., Lyndon J. Mitnaul, Ph.D.

Appendix A. Supplementary data

Supplementary tables for haplotype frequencies and genetic disequilibria (D) for the classical loci for the full Pima heritage sample. Supplementary data to this article can be found online at https://doi.org/10.1016/j.humimm.2019.10.002.

References

[1] R.C. Williams, J.E. McAuley, HLA class I variation controlled for genetic admixture in the Gila River Indian Community of Arizona: a model for the Paleo-Indians, Hum. Immunol. 33 (1992) 39–46.

[2] R.C. Williams, J.E. McAuley, 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. 33 (1992) 29–38.

[3] R.C. Williams, Y.F. Chen, R.O. Endres, D. Middleton, M. Trucco, J.D. Williams,

W.C. Knowler, 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, Tissue Antigens 74 (2009) 520–533.

[4] R.C. Williams, A.G. Steinberg, H. Gershowitz, P.H. Bennett, W.C. Knowler, D.J. Pettitt, W. Butler, R. Baird, L. Dowda-Rea, T.A. Burch, H.G. Morse, C.G. Smith, Gm allotypes in Native Americans: Evidence for three distinct migrations across the Bering Land Bridge, Am J Phys Anthropol 66 (1985) 1–19.

[5] B.L. Fontana, Pima and Papago: Introduction, in: A. Ortiz (Ed.), Handbook of North American Indians, Vol. 10 Smithsonian Institution, Washington, D.C., 1983, pp. 125–136.

[6] B.L. Fontana, History of the Papago, in: A. Ortiz (Ed.), Handbook of North American Indians, Vol. 10 Smithsonian Institution, Washington, D.C., 1983, pp. 137–148.

[7] P.H. Ezell, History of the Pima, in: A. Ortiz (Ed.), Handbook of North American Indians, Vol. 10 Smithsonian Institution, Washington, D.C., 1983, pp. 149–160.

[8] W.C. Knowler, D.J. Pettitt, M.F. Saad, P.H. Bennett, Diabetes mellitus in the Pima Indians: incidence, risk factors and pathogenesis, Diabetes Metab. Rev. 6 (1990) 1–27.

[9] W.C. Knowler, P.H. Bennett, R.F. Hamman, M. Miller, Diabetes incidence and prevalence in Pima Indians: a 19-fold greater incidence than in Rochester, Minn, Am. J. Epidemiol. 108 (1978) 497–504.

[10] R.C. Williams, W.C. Knowler, W.J. Butler, D.J. Pettitt, J.R. Lisse, P.H. Bennett, D.L. Mann, A.H. Johnson, P.I. Terasaki, HLA-A2 and type 2 diabetes in Pima Indians: an association and decrease in allele frequency with age, Diabetelogia 21 (1981) 460–463.

[11] R.C. Williams, L.T.H. Jacobsson, W.C. Knowler, A. del Puente, D. Kostyu, J.E. McAuley, P.H. Bennett, D.J. Pettitt, Meta-analysis reveals association between most common class II haplotype in full heritage Native Americans and rheumatoid arthritis, Hum. Immunol. 42 (1995) 90–94.

[12] R.C. Williams, R.L. Hanson, D.J. Pettitt, M.L. Sievers, R.G. Nelson, W.C. Knowler, HLA*A2 confers mortality risk for cardiovascular disease in Pimans, Tissue Antigens 47 (1996) 188–193.

[13] R.C. Williams, Y.L. Muller, R.L. Hanson, W.C. Knowler, C.C. Mason, L. Bian, V. Ossowski, K. Wiedrich, Y.F. Chen, S. Marcovina, J. Hahnke, R.G. Nelson, L.J. Baier, C. Bogardus, HLA-DRB1 reduces the risk of type 2 diabetes mellitus by increased insulin secretion, Diabetologia 54 (2011) 1684–1692.

[14] F.E. Dewey, M.F. Murray, J.D. Overton, L. Habegger, J.B. Leader, S.N. Fetterolf, C. O'Dushlaine, C.V. Van Hout, J. Staples, C. Gonzaga-Jauregui, R. Metpally, S.A. Pendergrass, M.A. Giovanni, H.L. Kirchner, S. Balasubramanian, N.S. Abul-Husn, D.N. Hartzel, D.R. Lavage, K.A. Kost, J.S. Packer, A.E. Lopez, J. Penn, S. Mukherjee, N. Gosalia, M. Kanagaraj, A.H. Li, L.J. Mitnaul, L.J. Adams, T.N. Person, K. Praveen, A. Marcketta, M.S. Lebo, C.A. Austin-Tse, H.M. Mason-Suares, S. Bruse, S. Mellis, R. Phillips, N. Stahl, A. Murphy, A. Economides, K.A. Skelding, C.D. Still, J.R. Elmore, I.B. Borecki, G.D. Yancopoulos, F.D. Davis, W.A. Faucett, O. Gottesman, M.D. Ritchie, A.R. Shuldiner, J.G. Reid, D.H. Ledbetter, A. Baras, D.J. Carey, Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study, Science 354 (6319) (2015), https://doi.org/10.1126/science.aaf6814.

[15] E. Major, K. Rigó, T. Hague, A. Bérces, S. Juhos, HLA typing from 1000 genomes whole genome and whole exome data, PLoS One 8 (11) (2013) e78410.

[16] B. Weir, Genetic Data Analysis, Sinaurer Associates Mass, Sunderland, 1990.

[17] J.C. Long, R.C. Williams, M. Urbanek, An EM algorithm and testing strategy for multiple locus haplotypes, Am. J. Hum. Genet. 56 (1995) 799–810.

[18] J. Nam, J.J. Gart, On two tests of fit for HLA data with no double blanks, Am. J. Hum. Genet. 41 (1987) 70–76.

[19] M.P. Belich, J.A. Madrigal, W.H. Hildebrand, J. Zemmour, R.C. Williams, R. Luz, M.L. Petzl-Erler, P. Parham, Unusual HLA-B alleles in two tribes of Brazilian Indians, Nature 357 (1992) 326–329.

[20] C. Raffoux, R.C. Williams, C. Gorodezky, Report of Antigen Society 7 (B12, B13, B21, B37, B40, B41, B48), in: D. Charron (Ed.), HLA, Genetic Diversity of HLA. Functional and Medical Implications, Vol. 1 EDK Publishers, Paris, 1997, pp. 45–78.

[21] G. Kawaguchi, W.H. Hildebrand, M. Hiraiwa, S. Karaki, T. Nagao, N. Akiyama, H. Uchida, K. Kashiwase, T. Akaza, R.C. Williams, T. Juji, P. Parham, M. Takiguchi, Two subtypes of HLA-B51 differing by substitution at position 171 of the α_2 helix, Immunogenetics 37 (1992) 57–63.

[22] R.C. Williams, S.N. Chen, D.K. Gill, J.T. Lane, J.E. McAuley, R. Strothman, K.K. Mittal, Antigen Society #6 Report (B21, B49, Bw50, BN21, B12, B44, B45), in: B. Dupont (Ed.), Immunobiology of HLA, Histocompatibility Testing, Volume 1 Springer Verglag, New York, 1987, pp. 133–147.

[23] W.H. Hildebrand, J.A. Madrigal, M.P. Belich, J. Zemmour, F.E. Ward, R.C. Williams, P. Parham, Serologic cross-reactivities poorly reflect allelic relationships in the HLA-B12 and HLA-B21 groups, J. Immunol. 149 (1992) 3563–3568.

[24] J. Zemmour, J. Gumperz, W.H. Hildebrand, S.G.E. Marsh, F.E. Ward, R.C. Williams, P. Parham, HLA-Cw11 is a combination of Cw1 and a public epitope of HLA-Cw3 and HLA-B46 molecules, in: K. Tsuji, M. Aizawa, T. Sasazuki (Eds.), HLA 1991, Vol. 1 Oxford University Press, Oxford, 1991, pp. 527–529.

[25] J. Zemmour, J. Gumperz, W.H. Hildebrand, F.E. Ward, S.G.E. Marsh, R.C. Williams, P. Parham, The molecular basis for reactivity of anti-Cw1 and anti-Cw3 alloantisera with HLA-B46 haplotypes, Tissue Antigens 39 (1992) 249–257.

[26] R.C. Williams, A.G. Steinberg, W.C. Knowler, D.J. Pettitt, Gm3;5,13,14 and stated-admixture: independent estimates of admixture in American Indians, Am. J. Hum. Genet. 39 (1986) 409–413.

[27] R.C. Williams, W.C. Knowler, D.J. Pettitt, J.C. Long, D.A. Rokala, H.F. Polesky, R.A. Hackenberg, A.G. Steinberg, P.H. Bennett, The magnitude and origin of European admixture in the Gila River Indian Community of Arizona: a union of genetics and demography, Am. J. Hum. Genet. 51 (1992) 101–110.

[28] R.C. Williams, J.C. Long, R.L. Hanson, M.L. Sievers, W.C. Knowler, 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. 66 (2000) 527–538.

[29] M.S. Leffell, M.D. Fallin, W.H. Hildebrand, J.W. Cavett, B.A. Inglehart, A.A. Zachary, HLA alleles and haplotypes among the Lakota Sioux: Report of the ASHI Minority Workshops, Part III, Hum. Immunol. 65 (2004) 78–89.

[30] J.M. Galarza, R. Barquera, A.M.T. Álvarez, D.I. Hernández Zaragoza, G.P. Sevilla, A. Tamayo, M. Pérez, D. Dávila, L. Birnberg, V.A. Alonzo, J. Krause, M. Grijalva, Genetic diversity of the HLA system in human populations from the Sierra (Andean), Oriente (Amazonian) and Costa (Coastal) regions of Ecuador, Hum. Immunol. 79 (2018) 639–650.

[31] B.A. González-Quezada, L.E. Creary, A.J. Munguia-Saldaña, H. Flores-Aguilar, M.A. Fernández-Viña, C. Gorodezky, Exploring the ancestry and admixture of Mexican Oaxaca Mestizos from Southeast Mexico using next-generation sequencing of 11 HLA loci, Hum. Immunol. 80 (2019) 157–162.

[32] Wen-Chi Hsueh, P.H. Bennett, J. Esparza-Romero, R. Urquidez-Romero, M.E. Valencia, E. Ravussin, R.C. Williams, W.C. Knowler, L.J. Baier, L.O. Schulz, R.L. Hanson, Analysis of type 2 diabetes and obesity genetic variants in Mexican Pima Indians: marked allelic differentiation from U.S. Pimas at HLA, Ann. Hum. Genet. (2018) 1–13, https://doi.org/10.1111/ahg.12252.