J.J. Chen
J.A. Hollenbach
E.A. Trachtenberg
J.J. Just
M. Carrington
K.S. Rønningen
A. Begovich
M.-C. King
S. McWeeney
S.J. Mack
H.A. Erlich
G. Thomson

Hardy-Weinberg testing for HLA class II (DRB1, DQA1, DQB1, AND DPB1) loci in 26 human ethnic groups

Key words:
exact test; Hardy-Weinberg proportions; HLA; individual test; population genetics

Acknowledgments:
This work was supported in part by National Institutes of Health grant GM 35326 (J.C., J.H., S.M., G.T.), Novo Nordisk Fonden (K.S.R.), National Institutes of Health grant AI-29042 (A.B.), National Institutes of Health grant GMHD 25792 (M.C.K., J.J.J.), National Institutes of Health grant (S.M., E.H.), the National Cancer Institute of the National Institutes of Health, under contract number NO1-CO-56000 (M.C.). The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Received 18 June, revised,
accepted for publication 27 September 1999

Tissue Antigens . ISSN 0001-2815

Tissue Antigens 1999: **54**: 533–542
Printed in Denmark . All rights reserved

Abstract: Testing the fit of population data to Hardy-Weinberg proportions is crucial in the validation of many current approaches in population genetic studies. In this paper, we tested fit to Hardy-Weinberg proportions using exact approaches for both the overall and individual heterozygote genotype data of four HLA Class II loci: DRB1, DQA1, DQB1, and DPB1, from 26 human populations. Eighty of 99 overall tests fit the Hardy-Weinberg expectation (73% for DRB1, 89% for DQA1, 81% for DQB1 and 81% for DPB1). Deviations from Hardy-Weinberg proportions were both locus and group specific. Although we could not rule out other mechanisms at work, the individual test results indicated that the departure was possibly partly due to recent admixture. Evidence for selection and other sources of deviation are also discussed.

The genes of the human leukocyte antigen (HLA) region control many important functions involved in the immune response, and are known to influence susceptibility to over 50 diseases. A number of features of the highly polymorphic class II DR, DQ, and DP regions implicate selection as an important factor maintaining its variation (e.g. see (1) for review). The specific mechanism of the selection has not yet been identified, although models invoking heterozygote advantage (since individuals heterozygous for HLA present a wider range of foreign peptides to T cells) are strong candidates (1–4).

Many disease and population genetics studies assume Hardy-Weinberg (HW) equilibrium, or incorporate Hardy-Weinberg proportions (HWP) in the mathematical models. The HW law states that in a large random-mating population with no selection, mutation, or migration, allele frequencies predict genotypic frequencies at a locus, and both genotype and allele frequencies will remain constant in subsequent generations. Violating these assumptions may not always result in significant deviation from HWP. However, deviation from HWP will often imply violation of the above con-

Authors' affiliations:

J.J. Chen[1],
J.A. Hollenbach[2],
E.A. Trachtenberg[3],
J.J. Just[4],
M. Carrington[5],
K.S. Rønningen[6],
A. Begovich[7],
M.-C. King[8],
S. McWeeney[2],
S.J. Mack[3],
H.A. Erlich[7],
G. Thomson[2]

[1]School of Public Health, Saint Louis University, Saint Louis, Missouri, USA,

[2]Department of Integrative Biology, University of California at Berkeley, Berkeley, California, USA,

[3]Children's Hospital Oakland Research Institute, Oakland, California, USA,

[4]Department of Human Genetics, Millenium Pharmaceuticals, Cambridge, Massachusetts, USA,

[5]National Cancer Institute-Frederick Cancer Research and Development Center, Frederick, Maryland, USA,

[6]Section of Epidemiology, National Institute of Public Health, Oslo, Norway,

[7]Department of Human Genetics, Roche Molecular Systems, Alameda, California, USA,

[8]Departments of Genetics and Medicine, University of Washington, Seattle, Washington, USA

Correspondence to:
John J. Chen
School of Public Health
Saint Louis University
3663 Lindell Blvd.
St. Louis, MO 63108
USA
Tel: +1 314 977 8134
Fax: +1 314 977 8150
e-mail: chenjj@slu.edu

ditions and deserves further study. Checking HWP is important in validation of many current approaches in population studies. This is relevant not only to theoretical population genetic or evolutionary research, but also to many forensic and clinical studies. For example, a synergistic effect for the DR3 and DR4 haplotypes DRB1*0301-DQA1*0501-DQB1*0201 and DRB1*0401-DQA1*0301-DQB1*0302 which are strongly associated with type 1 diabetes in Caucasians (5), is demonstrated by deviation from HWP with excess of heterozygous versus homozygous patients.

There are two important issues with regard to HWP testing. First is testing the overall fit of a sample. Traditionally, a large sample goodness-of-fit χ^2 is performed. But, due to the highly polymorphic nature of these HLA class II genes, it is not uncommon for researchers to distinguish up to 20 or 30 alleles in a sample. As a result, in many studies the genotype counts within the sample are quite small. Therefore, the large sample goodness-of-fit χ^2 are not valid in these situations and simulation approaches are more appropriate than their asymptotic counterparts in testing for overall fit of HWP.

The second issue is testing at the individual heterozygote level. For a given sample, besides the overall fit, researchers are often interested in knowing whether individual heterozygotes significantly deviate from HWP. This is especially relevant when genetic data are available from more than one population. By comparing patterns across populations, consistent deviations, which may not be strong enough to be found in the overall tests of an individual sample, could be detected. This will allow researchers to generate future hypotheses on the mechanisms causing such patterns and facilitate further study of this genetic region.

Table 1

Populations studied, their locations, sample sizes and references

Population	Geographic location	Sample size	Reference
Mixtec Alta (Amerindian)	Oaxaca; Mexico	103	6
Mixe (Amerindian)	Oaxaca, Mexico	55	6
Zapotec (Amerindian)	Oaxaca, Mexico	90	6
Norwegian	Norway	180	7
Ethiopian	Ethiopia	40	8
Japanese	Japan	290	8
Venezuelan	Venezuela	27	8
Mexican American	USA	55	9
African-American	New York, New York, USA	241	10
Cayapa (Amerindian)	Ecuador	99	11
Cauca (Colombian African American)	Pacific, Colombia	20	12
Choco (Colombian African American)	North Pacific Coast,Colombia	20	12
Providencia (Colombian African American)	Caribbean Island, Colombia	30	12
Coreguaje (Amerindian)	Amazonas, Colombia	30	13
Embera (Amerindian)	Pacific, Colombia	20	13
Ijka (Amerindian)	Sierra Nevada, Colombia	30	13
Kogui (Amerindian)	Sierra Nevada, Colombia	31	13
Nukak (Amerindian)	Eastern Amazonas, Colombia	20	13
Sikuani (Amerindian)	Eastern Plains, Colombia	27	13
Tule (Amerindian)	NW Pacific Coast, Colombia	29	13
Waunana (Amerindian)	Pacific, Colombia	30	13
CEPH	France/Utah, USA	193	14
Maya (Amerindian)	Yucatan, Mexico	15	15
Ticuna (Amerindian)	Brazil	49	15
Yanomamo (Amerindian)	Venezuela	55	15
Pima (Amerindian)	Arizona, USA	17	16

In this study, we analyzed the genotypic distribution data of four HLA class II loci, DRB1, DQA1, DQB1, and DPB1, from 26 human populations. For each sample, Hardy-Weinberg proportions were examined by both the overall test for each locus and the individual heterozygote tests within each locus. Significant patterns were found and the implications of such patterns are discussed.

Material and methods

Samples

Genotype frequencies from 26 diverse human populations were examined. The ethnic group sampled, the sample sizes, and the geographic locations of these populations are listed in Table 1, together with their literature references (6–16).

Genomic DNA was amplified using the polymerase chain reaction (PCR) and typed using sequence-specific oligonucleotide probe methods (SSOP) for DRB1, DQA1, DQB1, and DPB1. For more details on the typing protocol in each study, see the individual references. For the following populations DPB1 typing was not included: Ethiopian, Japanese, Mexican American, Norwegian, and Venezuelan. In some cases, further allelic resolution was accomplished via direct sequencing from PCR product.

Statistical methods

HW proportions were tested with both the traditional χ^2 test and the exact test for overall deviation. The exact test (17) utilized a Markov Chain Monte Carlo approach to simulate samples and estimated the significance by comparing the probability of the observed genotype distribution with those of the samples simulated under the HW equilibrium distribution. Individual heterozygote deviations were tested by Monte Carlo simulation, using the disequilibrium coefficients (18). Conditioning on the allele frequencies, we generated 20,000 pseudo-samples by permuting the individual alleles and randomly forming diploid individuals. For a particular heterozygote, the observed test statistic (19) was compared with those of the pseudo-samples.

Results

Many HLA loci are known for their high level of polymorphism and this is mirrored in the samples studied (Table 2). For example, in African-Americans, the population with the highest levels of polymorphism, 37 alleles were detected for the DRB1 locus, 8 alleles for DQA1, 12 for DQB1, and 25 for DPB1. The Amerindian populations showed the lowest levels of polymorphism.

Number of alleles sampled for each locus in 26 ethnic groups

	Number of alleles			
Population	DRB1	DQA1	DQB1	DPB1
Mixtec Alta	15	5	9	10
Mixe	9	4	4	5
Zapotec	16	6	9	12
Ethiopia	16	6	9	NA
Japan	30	8	15	NA
Mexican American	29	7	13	NA
Norwegian	16	7	10	NA
Venezuela	15	7	9	NA
African-American	37	8	12	25
Cayapa	14	3	4	6
Cauca	16	7	9	10
Choco	15	7	8	8
Providencia	19	7	9	15
Coreguaje	5	3	4	4
Embera	9	3	3	5
Ijka	10	6	6	6
Kogui	4	3	3	3
Nukak	5	3	3	3
Sikuani	8	4	5	6
Tule	10	4	4	4
Waunana	9	3	4	5
CEPH	28	7	13	17
Maya	9	4	5	5
Pima	5	3	2	4
Ticuna	9	3	5	9
Yanomamo	15	5	3	7

Table 2

Compared with large-sample test results (not shown), there are fewer populations and loci with significant deviations from HW expectations using the exact overall tests (Table 3). This indicates the less conservative nature of the large-sample χ^2 test when sample size is relatively small, compared with the number of genotypes in the sample. Out of 99 exact tests for different loci and populations, 80 tests showed fit to HWP (73% for DRB1, 89% for DQA1, 81% for DQB1 and 81% for DPB1). Even though the significance of the tests were grouped, for convenience, by ranges of *P*-values, it is important to note that less significant results, such as DRB1 in CEPH ($P<0.079$) could still be suggestive, especially when studied across samples.

Exact overall Hardy-Weinberg testing results *Table 3*

	P-value			
Population	DRB1	DQA1	DQB1	DPB1
Mixtec Alta	0.007893**	0.4133	0.06575	0.1647
Mixe	0.06312	0.01986*	0.1322	1.0000
Zapotec	0.04722*	0.03470*	0.05959	0.2073
Ethiopia	0.3121	0.2113	0.3179	NA
Japan	0.8037	0.7789	0.7834	NA
Mexican American	0.01468*	0.5490	0.02946*	NA
Norwegian	0.3060	0.4010	0.1298	NA
Venezuela	0.1715	0.07224	0.000439***	NA
African-American	0.5141	0.07932	0.02776*	0.0022**
Cayapa	0.02588*	0.3618	0.9881	0.3030
Cauca	0.8235	0.8016	0.4798	0.2880
Choco	0.3313	0.4185	0.2882	0.5257
Providencia	0.02866*	0.09465	0.1207	0.4015
Coreguaje	0.8890	0.3485	0.4421	0.3747
Embera	0.9515	0.6609	0.8947	0.4376
Ijka	0.9758	0.9802	0.9991	0.1701
Kogui	0.1806	0.5971	0.5958	0.00002***
Nukak	0.06184	0.00973**	0.00392**	0.1285
Sikuani	0.1339	0.1176	0.08064	0.1585
Tule	0.001704**	0.1311	0.1180	0.0870
Waunana	0.4043	0.3137	0.04134*	0.1433
CEPH	0.07863	0.7864	0.2161	0.7357
Maya	0.4381	0.3520	0.5511	0.8362
Pima	0.6518	1.0000	0.9697	0.0060**
Ticuna	0.004914**	0.5172	0.1032	0.0125*
Yanomamo	0.1155	0.3573	0.8011	0.8480

* $0.01<P<0.05$; ** $0.001<P<0.01$; *** $P<0.001$

It is of interest that the most polymorphic locus, DRB1, showed the greatest deviation from HWP, while the least deviation was for DQA1, which is consistently less polymorphic than the other loci. The DQB1 locus for the Venezuelan sample and the DPB1 locus for the Kogui sample showed the most significant deviations from HWP ($P<0.00044$ and $P<0.00002$, respectively). The Mixtec Alta, Tule and Ticuna samples for the DRB1 locus, the Nukak sample for the DQA1 and DQB1 loci, and the African-American and Pima samples for the DPB1 locus also showed significant effects at the $P<0.01$ level. Twelve of the 26 ethnic groups showed fit to HWP for all their loci considered. Five ethnic groups showed deviation at more than one locus: Zapotec at DRB1 ($P<0.048$) and DQA1 ($P<0.035$), Mexican American at DRB1 ($P<0.015$) and DQB1 ($P<0.030$), African-American at DQB1 ($P<0.028$) and DPB1 ($P<0.0022$), Nukak at DQA1 ($P<0.01$) and DQB1 ($P<0.004$), Ticuna at DRB1 ($P<0.005$) and DPB1 ($P<0.013$). No consistent patterns for deviations from HWP were seen in terms of ethnic group or year of study (which correlated with increasing number of alleles detected in a specific sample).

In Tables 4–7, the individual heterozygote HWP testing results from each of the four loci are summarized. Significant *P*-values were categorized by their levels ($P<0.001$; $0.001< P<0.01$; $0.01< P<0.05$) for convenience of discussion and whether the heterozygotes were in deficit or in excess. In order to make biologically meaningful comparisons, only heterozygotes which have a difference larger than two between the observed and expected counts were studied.

Some specific heterozygotes showed up as significantly different from HWP in more than one ethnic group. For the DQA1 locus, hetero-

DRB1 locus individual Hardy-Weinberg testing results

Population	Heterozygotes[#]	Heterozygote excess	Heterozygote deficit
Mixtec Alta	105	(0407, 1406){10}[5.8]<0.04>*	(0802, 1406){1}[4.3]<0.04>*, (0802, 1602){4}[8.1]<0.04>*
Zapotec	120	(0404, 0408){3}[0.5]<0.004>**, (1402, 1602){5}[1.5]<0.005>**	(0407, 0802){1}[4.4]<0.03>*
Ethiopia	120	(0102, 0301){3}[0.7]<0.02>*, (1301, 1302){3}[0.7]<0.02>*	
Japan	435	(0101, 0405){9}[4.5]<0.03>*, (0803, 1302){6}[3.0]<0.03>*, (0901, 1405){5}[2.5]<0.05>*	(0101, 0901){1}[5.1]<0.02>*
Mexican American	406	(0102, 0701){3}[0.2]<0.0003>***, (0301, 1101){3}[0.5]<0.007>**	
Norwegian	120	(1101, 1301){5}[2.0]<0.04>*, (0301, 1501){13}[8.5]<0.04>*	(0401, 1301){2}[6.1]<0.03>*, (0301, 0701){3}[6.2]<0.04>*, (0301, 1101){0}[2.3]<0.03>*
Venezuela	105	(0400, 1101){3}[0.4]<0.002>**	
African-American	666	(0301, 0702){4}[1.7]<0.05>*, (1001, 1101){3}[0.7]<0.03>*, (0803, 1302){3}[0.9]<0.04>*	
Cayapa	91	(0407, 1402){17}[10.8]<0.02>*, (0407, 0411){4}[1.9]<0.05>*	(0802, 1402){0}[3.3]<0.02>*, (0407, 08043){0}[2.5]<0.04>*
Kogui	6	(0403, 0802){3}[0.8]<0.03>*	
CEPH	378	(0701, 1302){5}[1.6]<0.02>*, (0401, 1101){5}[2.2]<0.04>*	(1101, 1501){0}[3.4]<0.02>*
Ticuna	36	(0403, 0807){4}[0.9]<0.003>**, (0411, 0802){3}[0.9]<0.02>*, (0901, 1305){3}[0.2]<0.0006>***	(0901, 1602){0}[2.0]<0.05>*
Yanomamo	105	(0403, 0407){5}[2.3]<0.05>*, (0701, 1602){4}[1.3]<0.009>**	

(Allele1, Allele2){Observed}[Expected]<*P*-value>
[#]: number of heterozygote classes possible in each sample
* $0.01<P<0.05$; ** $0.001<P<0.01$; *** $P<0.001$

Table 4

zygote 0301/0501 was significantly in deficit in both the Zapotec and Nukak samples. Several heterozygotes were found to be in significant excess for the DQA1 locus, i.e., 0401/0501 in the Zapotec and the Norwegian samples, 0101/0501 in the Ethiopia and the Providencia samples, and 0201/0501 in the African-American and the Yanomamo samples. For the DQB1 locus, 0301/0303 was significantly in deficit in the Zapotec, Japanese and the Venezuelan samples, whereas 0301/0501 was significantly in deficit in the Mexican-American, the Norwegian and the Venezuelan samples. Other significant deficit heterozygotes in the DQB1 locus include: 0301/0302 in the Mexican-American and Nukak samples, 0201/0402 in the Cauca and the CEPH samples. Also for the DQB1 locus, heterozygote 0201/0602 was found significant in excess in the Ethiopia, the Norwegian, and the Providencia samples, and heterozygote 0302/0402 in the Mixe and the Ticuna samples. For DPB1, heterozygote 0402/1401 was significantly in excess in the Ijka, Kogui, and Tule samples. 0402/1701 was in ex-

DQA1 locus individual Hardy-Weinberg testing results

Population	Heterozygotes[#]	Heterozygote excess	Heterozygote deficit
Mixtec Alta	10		(0401, 0501){7}[11.7]<0.04>*
Mixe	6	(0301, 0401){14}[8.4]<0.006>**	
Zapotec	15	(0401, 0501){17}[11.7]<0.05>*	(0301, 0501){13}[20.7]<0.01>**, (0301, 0401){10}[14.9]<0.04>*
Ethiopia	15	(0101, 0501){6}[2.2]<0.006>**, (0102, 0103){3}[0.8]<0.006>**	
Mexican American	21		(0101, 0501){3}[6.2]<0.04>*
Norwegian	21	(0401, 0501){5}[2.7]<0.05>*	(0101, 0300){5}[9.5]<0.04>*
African-American	28	(0201, 0301){9}[4.6]<0.02>*, (0201, 0501){14}[8.3]<0.03>*	(0101, 0201){2}[5.9]<0.03>*, (0102, 0201){10}[15.2]<0.05>*, (0102, 0501){23}[31.3]<0.03>*,
Choco	21	(0101, 0301){3}[0.8]<0.02>*	
Providencia	21	(0101, 0401){4}[1.1]<0.0007>***, (0101, 0501){5}[2.9]<0.05>*	
Nukak	3		(0301, 0501){3}[7.2]<0.02>*
Tule	6	(0301, 0501){16}[11.1]<0.03>*	
Yanomamo	10	(0201, 0501){5}[2.9]<0.05>*	(0201, 0301){0}[2.3]<0.02>*

(Allele1, Allele2){Observed}[Expected]<*P*-value>
[#] number of heterozygote classes possible in each sample
* 0.01<*P*<0.05; ** 0.001<*P*<0.01; *** *P*<0.001

Table 5

cess in the African-American and the Providencia samples. 1401/3501 in the Ijka and the Kogui samples. 0401/1401 in the CEPH and the Ticuna samples. On the other hand, heterozygote 0401/0402 was significantly in deficit in both the Mixtec Alta and the African-American samples. Note that specific subsets of alleles appear to be involved in the individual significant deviations from HWP (both in excess and deficit), namely DRB1*04 (38%), DQA1*05 (57%), DQB1*03 (58%) and DPB1*04 (64%).

Although the overall tests were significant, no significant individual heterozygotes were identified in the following loci and samples: the DRB1 locus for the Providencia and the Tule samples; the DQB1 loci for the Waunana sample. Conversely, a number of significant individual heterozygote effects may be seen, even when the overall test shows a reasonable fit to HWP. This reflects the effects of a large number of alleles and genotype frequencies, most of which are in close fit to HWP.

Discussion

From the overall tests, the vast majority of the ethnic groups showed a good fit to HWP in most of their loci. This agrees with results of Rivas et al. (20) who found fit to HWP in 90 out of 102 populations typed for DQA1 variation. Even though the test results shown here might be limited in statistical power for some of the samples with relatively small sizes, some interesting patterns were noteworthy. The Ethiopian and the Colombian African samples showed good fit to HWP, however the U.S. African-American sample deviated for the DQB1 and DPB1 loci. In the Japanese sample, which was the only Asian population studied, a good fit was found for all four loci. A good fit to HWP was also found in the samples of European populations. Conversely, in the Amerindian groups, which had the lowest levels of HLA variation, a quite striking devi-

DQB1 locus individual Hardy-Weinberg testing results

Population	Heterozygotes[#]	Heterozygote excess	Heterozygote deficit
Mixtec Alta	36	(0301, 0302){29}[21.5]<0.03>*	(0301, 0402){7}[13.5]<0.009>**, (0302, 0503){2}[4.4]<0.04>*
Mixe	6	(0302, 0402){13}[8.7]<0.03>*	
Zapotec	36	(0302, 0303){8}[2.5]<0.0001>***, (0402, 0606){4}[1.1]<0.007>**	(0301, 0303){0}[2.5]<0.04>*
Ethiopia	36	(0201, 0602){4}[1.5]<0.02>*	
Japan	105		(0301, 0303){6}[12.6]<0.02>*
Mexican-American	78		(0301, 0302){1}[5.1]<0.02>*, (0301, 0501){1}[3.8]<0.03>*
Norwegian	45	(0201, 0602){21}[13.7]<0.008>**, (0301, 0603){8}[4]<0.04>*, (0201, 0402){5}[2.9]<0.05>*	(0301, 0501){2}[5.3]<0.05>*
Venezuela	36	(0501, 0601){4}[0.9]<0.004>**	(0301, 0303){0}[2.3]<0.05>*, (0301, 0501){0}[2.7]<0.02>*
African-American	66	(0501, 0502){4}[1.7]<0.04>*, (0501, 0604){4}[1.7]<0.04>*, (0201, 0301){23.0}[16.4]<0.04>*	(0501, 0605){0}[3.1]<0.004>**, (0302, 0602){1}[4.9]<0.02>*, (0201, 0602){14}[23.2]<0.008>**
Cauca	36		(0201, 0402){0}[2.0]<0.007>**
Providencia	36	(0402, 0501){3}[0.9]<0.02>*, (0201, 0602){4}[1.6]<0.03>*, (0301, 0602){4}[1.6]<0.03>*	
Nakuk	3		(0301, 0302){3}[7.2]<0.01>**
CEPH	78	(0201, 0303){6}[3.0]<0.02>*, (0201, 0603){6}[2.8]<0.02>*, (0201, 0604){4}[1.7]<0.03>*, (0502, 0602){3}[0.9]<0.05>*	(0201, 0301){8}[13.1]<0.03>*, (0301, 0602){7}[11.7]<0.05>*, (0201, 0402){0}[2.4]<0.05>*, (0303, 0602){0}[2.7]<0.02>*
Ticuna	10	(0302, 0402){17}[10.6]<0.007>**	

(Allele1, Allele2){Observed}[Expected]<*P*-value>
[#] number of heterozygote classes possible in each sample
* 0.01<*P*<0.05; ** 0.001<*P*<0.01; *** *P*<0.001

Table 6

ation from HWP was detected in some samples. Seven out of 9 highly significant overall tests ($P<0.01$) were found in the Amerindian populations. Despite the close linkage and strong linkage disequilibrium between the HLA class II loci, especially the DR-DQ complex, it is clear that departure from HWP is locus specific. Also, the deviation from HWP is group specific when comparing neighboring populations. HLA class II loci variation is known to show population structure, with some alleles differing in frequency between ethnic and geographic groups, while others are unique to particular groups (21).

Studies of different ethnic groups have often led to the identification of previously unknown alleles, or uncommon alleles. The DRB1*0807 allele was found in the Ticuna sample and the deviation from HWP in this locus was possibly due to the excess of heterozygotes involving *0807 (16). Our individual analysis using the exact approach showed that out of eight DRB1*0807 heterozygotes

DPB1 locus individual Hardy-Weinberg testing results

Population	Heterozygotes[#]	Heterozygote excess	Heterozygote deficit
Mixtec Alta	45		(0401, 0402){1}[3.0]<0.03>*
African-American	300	(0301, 1101){4}[0.9]<0.005>**, (0402, 1701){12}[5.4]<0.003>**, (1701, 2701){3}[0.6]<0.01>**	(0101, 1801){7}[13.1]<0.03>*, (1701, 1801){0}[3.6]<0.02>*, (0401, 0402){1}[5.4]<0.005>**
Cayapa	15		(0201, 1401){1}[3.5]<0.04>*
Providencia	105	(0101, 0201){3}[0.9]<0.03>*, (0401, 1601){3}[0.5]<0.002>**, (0402, 1701){3}[1.0]<0.05>*	
Ijka	15	(1401,3501){3}[0.8]*, (0402, 1401){13}[9.6]<0.05>*, (1401, 3501){3}[0.8]<0.002>**	
Kogui	3	(1401, 3501){16}[7.7]<0.0001>***, (0402, 1401){12}[7.7]<0.02>*	(0402, 3501){1}[4.7]<0.003>**
Tule	6	(0402, 1401){14}[9.6]<0.02>*	
CEPH	136	(0401, 1401){5}[2.0]<0.0006>***, (0401, 0601){5}[2.8]<0.03>*, (0401, 1501){4}[2]<0.02>*	
Pima	6	(0401, 2301){5}[1.3]<0.0006>***	(0402, 2301){0}[2.6]<0.003>**
Ticuna	36	(0401, 1401){7}[2.8]<0.007>**, (0301, 2901){3}[0.4]<0.002>**	(0402, 2901){0}[2.1]<0.03>*

(Allele1, Allele2){Observed}[Expected]<*P*-value>
[#] number of heterozygote classes possible in each sample
* 0.01<*P*<0.05; ** 0.001<*P*<0.01; *** *P*<0.001

Table 7

observed, only one heterozygote, 0403/0807, showed a significant excess, whereas all others showed a deficit. The heterozygote that showed highly significant deviation was instead 0901/1305. DRB1*1305 is a common African allele. All three individuals with allele *1305 in the Ticuna data were with *0901, a common allele in the population. Therefore, it is quite possible, as suggested by Mack and Erlich (16), that the deviation reflects African admixture.

At the DPB1 locus, the DPB1*3501 allele differs from the *1401 allele by only one amino acid replacement at position 65. The DPB1*3501 allele was found in three Colombian Amerindian samples (frequencies ranging from 0.03 to 0.27), and has been found at extremely low frequencies in African and Portuguese populations. The high frequency of DPB1*1401 in most South American Amerindian groups supports the hypothesis that the *3501 allele arose from the *1401 in the Colombian Amerindian populations, and represents an example of convergent evolution and of independent origins of the DPB1*3501 allele in these Columbian Amerindian populations. It is of interest, also, that many significant deviations from HWP from individual heterozygotes involved the DPB1*3501 and *1401 alleles in the Colombian Amerindian samples.

It is important to note that in many cases the genotypes showing significant deviation contain alleles suspected of being the result of admixture or recent origin within the group. It is possible that admixture has effectively caused the deviation from HWP, suggesting its very recent history to the group, perhaps within one generation. In the study of Rivas et al. (20) on DQA1, the 12 population samples showing deviation from HWP were also of substantial mixed origin.

One interesting case from our study is the CEPH data, which is composed mainly of Caucasians sampled from Utah and France. But the sample also contains five Venezuelan individuals. This provides us with a situation where we can check for the effect of admix-

ture on deviation from HWP. As the French and Utah data share many haplotypes (22), the major potential admixture comes from the Venezuelans. We reanalyzed the CEPH data, excluding the five Venezuelan individuals. The overall tests were less significant. While DQA1, DQB1, and DPB1 were still non-significant, the marginally significant *P*-value of DRB1 (0.08) was 0.1 after removal of the Venezuelan individuals. At the individual heterozygote level, fewer significant deviations was detected (n=9 vs. 14).

Certainly, other explanations for the deviation from HWP, including nonrandom mating patterns with regard to admixed individuals within the ethnic group, must also be considered, as significant deviations from HWP are also seen in populations which showed very little evidence of any admixture, e.g. the Mixe (6). These deviations could be due to the impact of selection pressures from infectious diseases on these genes, especially when taking into consideration the difficulty of detecting inbreeding effects in human populations with relatively small samples. Indeed, it is possible that selection is also responsible for the deviations observed in many of the samples. Further, the fact that the DQB1 allele *0301 was involved in significant individual heterozygote deficiencies in 8 samples could well implicate selection, and this observation warrants study in more populations. The African-American sample is comprised of children born to HIV-infected mothers. Most of these children were themselves long-term HIV-positive survivors (23). The allele associated with AIDS long-term survival (DPB1*0101) was significantly in deficit in heterozygote 0101/1801, and the allele found associated with fast AIDS progression in study of this sample, i.e. DRB1*0301 was found to be in excess in heterozygote 0301/0702. In light of a recent paper by Carrington et al. (24), it would be very interesting and useful to see whether an excess of specific heterozygotes, or all heterozygotes, exists in such a sample. This area is under current investigation.

Testing for fit to HWP can also serve a useful role in assessing the initial quality of typing results. Should deviations from expectations occur, it is possible to look again at the raw data and the methods employed to generate them, and if deemed necessary repeat typing or use alternative methods to confirm results. This may be particularly relevant in non-Caucasian populations, where new alleles which may not be directly detected by the typing system are more likely to occur.

We have to caution here that the individual heterozygote testing approach is most useful if one has an a priori hypothesis about one or a few heterozygotes in the sample, regarding their excess or deficit from HWP. By applying the method to all heterozygotes, we face the problem of multiple tests, and it is very likely that some of the significant heterozygotes were due to type I error. Therefore the resulting *P*-values need to be judged with care. Since in this paper we are mainly interested in detecting patterns and generating future hypotheses, instead of classical hypothesis testing, we did not formally address the issue of adjusting the significance level here for multiple tests. In this analysis, we concentrated on testing only individual heterozygotes. The deviation of individual homozygote can also be derived, but was not discussed here. It is well-known that HLA is a highly linked region. Alleles at the loci studied tend to form common haplotypes. Extending the analysis to the linked loci as a whole, and testing its common haplotypes might provide us with more insight on this region.

In summary, based on the results of this study, the majority of the ethnic groups and loci studied show fit to HWP. The departure from HPW seems to be both group and locus specific. Although possible type I error, mistyping, and other natural forces, such as selection, may be responsible as well in causing such deviation, recent admixture seems to play a role in the violation of HWP in these data sets.

References

1. Edwards SV, Hedrick PW. Evolution and ecology of MHC molecules: from genomics to sexual selection. *TREE* 1998: **13**: 305–11.
2. Takahata N, Nei M. Allelic genealogy under overdominant and frequency dependent selection and polymorphism of major histocompatibility complex loci. *Genetics* 1990: **124**: 967–78.
3. Takahata N, Satta Y, Klein J. Polymorphism and balancing selection at major histocompatibility complex loci. *Genetics* 1992: **130**: 925–38.
4. Parham P, Ohta T. Population biology of antigen presentation by MHC class I molecules. *Science* 1996: **272**: 67–74.
5. Noble J, Valdes AM, Cook M et al. The role of HLA class II genes in insulin dependent diabetes mellitus (IDDM): molecular analysis if 180 Caucasian multiplex families. *Am J Hum Genet* 1996: **59**: 1134–48.
6. Hollenbach JA. Genetic character and evolution of the human leukocyte antigen complex in Mesoamerican natives. Ph.D. Dissertation, UC-Berkeley, Berkeley, CA, USA, 1998.
7. Rønningen KS, Spurkland A, Iwe T, Vartdal F, Thorsby E. Distribution of HLA-DRB1, -DQA1 and -DQB1 alleles and DQA1-DQB1 genotypes among Norwegian patients with insulin-dependent diabetes mellitus. *Tissue Antigens* 1991: **37**: 105–11.

8. Rønningen KS, Spurkland A, Tait BD. HLA class II associations in insulin-dependent diabetes mellitus among Blacks, Caucasoids, and Japanese. In: Tsuji K, Aizawa M, Sasazuki T, eds. *HLA 1991*. Vol. 1. New York: Oxford University Press, 1991: 713–22.
9. Erlich H, Rotter JI, Chang JD et al. HLA class II alleles and susceptibility to insulin-dependent diabetes mellitus in Mexican American families. *Nat Genet* 1993: **3**: 358–64.
10. Just, JJ, King MC, Thomson G, Klitz W. African-American HLA class II allele and haplotype diversity. *Tissue Antigens* 1997: **49**: 547–55.
11. Titus-Trachtenberg EA, Rickards O, De Stefano GF, Erlich HA. Analysis of HLA class II haplotypes in the Cayapa Indians of Ecuador: A novel DRB1 allele reveals evidence for convergent evolution and balancing selection at position 86. *Am J Hum Genet* 1994: **55**: 160–7.
12. Trachtenberg EA, Keyeux G, Bernal JE, Rhodas MC, Erlich HA. Results of expedition Humana I. Analysis of HLA class II (DRB1-DQA1-DQB1-DPB1) alleles and DR-DQ haplotypes in nine Amerindian populations from Colombia. *Tissue Antigens* 1996: **48**: 174–81.
13. Trachtenberg EA, Keyeux G, Bernal JE, Noble JA, Erlich HA. Results of expedition Humana II. Analysis of HLA class II alleles in three African American populations from Colombia using PCR/SSOP: identification of a novel DQB1*02(*0203) allele. *Tissue Antigens* 1996: **48**: 192–8.
14. Carrington M, Marti D, Wade J et al. Microsatellite markers in complex disease: Mapping disease-associated regions within the human major histocompatibility complex. In: Goldstein DB, Schlötterer C, eds.. *Microsatellites: evolution and applications*. New York: Oxford University Press, 1999: 225–37.
15. Mack S. Molecular evolution of mitochondrial control region sequences and class II HLA loci in Native American populations. Ph.D. Dissertation, UC-Berkeley, Berkeley, CA, USA, 1996.
16. Mack S, Erlich HA. HLA class II polymorphism in the Ticuna of Brazil: evolutionary implications of the DRB1*0807 allele. *Tissue Antigens* 1998: **51**: 41–50.
17. Guo SW, Thompson EA. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 1992: **48**: 361–72.
18. Hernández JL, Weir BS. A disequlibrium coefficient approach to Hardy-Weinberg testing. *Biometrics* 1989: **45**: 53–70.
19. Chen JJ, Thomson G. The variance for the disequilibrium coefficient in the individual Hardy-Weinberg test. *Biometrics* 2000: in press.
20. Rivas F, Zhong Y, Olivares N, Cerda-Flores RM, Chakraborty R. Worldwide genetic diversity at the HLA-DQA1 locus. *Am J Hum Biol* 1997: **7**: 735–49.
21. Charron DJ, ed. *HLA 1997: Genetic diversity of HLA functional and medical implications*. Vol. 1. Paris: EDK, 1997.
22. Carrington M, Stephens JC, Klitz W, Begovich AB, Erlich HA, Mann D. Major histocompatibility complex Class II haplotypes and linkage disequilibrium values observed in the CEPH families. *Hum Immunol* 1994: **41**: 234–40.
23. Just J, Louie L, Abrahams E et al. Genetic risk factors for perinatally-acquired HIV-1 infection. *Paediatr Perinat Epidemiol* 1992: **6**: 215–24.
24. Carrington M, Nelson GW, Martin MP et al. HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. *Science* 1999: **283**: 1748–52.