Am. J. Hum. Genet. 47:515–523, 1990

HLA-DQα Allele and Genotype Frequencies in Various Human Populations, Determined by Using Enzymatic Amplification and Oligonucleotide Probes

Rhea Helmuth,* Nicola Fildes,† Edward Blake,‡ Michael C. Luce,§ J. Chimera,§ Roberta Madej,* C. Gorodezky,‖ Mark Stoneking,† Norma Schmill,# William Klitz,** Russell Higuchi,† and Henry A. Erlich†

*RDS-West; and †Department of Human Genetics, Cetus Corporation, Emeryville, CA; ‡Forensic Science Associates, Richmond, CA; §Department of Molecular Biology, Roche Biomedical Laboratories, Research Triangle Park, NC; ‖Instituto de Salubridad y Enfermadades Tropicales, Mexico City; #Alberta Hereditary Diseases Program, Alberta Children's Hospital, Calgary; and **Department of Integrative Biology, University of California, Berkeley

Summary

Allele and genotype frequencies at the HLA-DQα locus have been determined by the use of polymerase chain reaction (PCR) amplification and nonradioactive oligonucleotide probes. The probes define six alleles and 21 genotypes in a dot-blot format. A total of over 1,400 individuals from 11 populations has been typed by two different laboratories using this method. In contrast to some variable-number-of-tandem-repeat markers that have been used for identity determination, DQα genotype frequencies do not deviate significantly from Hardy-Weinberg equilibrium in all populations studied. The distribution of alleles varies significantly between most of these populations. In Caucasians, the allele frequencies range from 4.3% to 28.5%. In this population, the power of discrimination is .94, and, for paternity determination, the power of exclusion is .642. These population data will allow the use of the HLA-DQα marker in paternity determination, the analysis of individual identity in forensic samples, and anthropological studies.

Introduction

The identification of specific DNA polymorphisms has provided a wealth of genetic markers for mapping studies and for the analysis of populations. The development of the polymerase chain reaction (PCR) (Saiki et al. 1985, 1988*b*; Mullis and Faloona 1987) for the in vitro amplification of specific DNA sequences has made possible a number of rapid approaches for determining DNA polymorphisms. One of these approaches, analysis of amplified DNA by using non-radioactive-labeled oligonucleotide probes in a dot-blot format (Saiki et al. 1986, 1988*a*; Bugawan et al. 1988), represents a very simple and rapid method for determining allele and genotype frequencies in population surveys. Here we report the allele and genotype frequencies for the HLA-DQα locus, in U.S. Caucasian, U.S. black, Southeast Asian, Japanese, and various Hispanic populations, as well as those in more isolated populations (Indonesian, Papua New Guinean, Australian, Bedouin, and Nigerian), as determined by this method.

The DQα locus codes for the polymorphic α-chain of the HLA-DQ molecule, one of the class II α-β chain heterodimers encoded by the HLA-D or class II region on chromosome 6. These molecules are integral membrane proteins which bind and present antigen peptide fragments to the T cell receptor of $CD4^+$ T lymphocytes (reviewed in Trowsdale et al. 1985). Most of the polymorphism of HLA class II α- and β-chain loci is localized to the second exon encoding the NH_2-terminal outer domain (Trowsdale et al. 1985).

The detection of polymorphic alleles by oligonucleotide probes requires the prior determination of allelic sequence diversity within the population to be tested.

Received January 26, 1990; revision received April 23, 1990.

Address for correspondence and reprints: Henry Erlich, Department of Human Genetics, Cetus Corporation, 1400 Fifty-Third Street, Emeryville, CA 94608.

0002-9297/90/4703-0015$02.00

Table 1

Oligonucleotides Used in the Present Study

Name		Sequence (5′ to 3′)
PCR primers:		
GH26		GTGCTGCAGGTGTAAACTTGTACCAG
GH27		CACGGATCCGGTAGCAGCGGTAGAGTTG
	DQA Specificity	Sequence (5′ to 3′)
Dot-blot probes:		
GH75	1	CTCAGGCCACCGCCAGGCA
RH83[a]	1	TGAGTTCAGCAAATTTGGAG
GH68	2	TTCCACAGACTTAGATTTG
RH71[a]	2	TTCCACAGACTTAGATTTGAC
GH67	3	TTCCGCAGATTTAGAAGAT
GH66	4	TGTTTGCCTGTTCTCAGAC
GH88	1.1	CGTAGAACTCCTCATCTCC
GH76	All but 1.3	GTCTCCTTCCTCTCCAG
RH119[a]	All but 1.3	GTCTCCTTCCTCTCCAGG
GH89	1.2, 1.3, 4	GATGAGCAGTTCTACGTGG
GH77	1.3	CTGGAGAAGAAGGAGAC
RH117[a]	1.3	CCTGGAGAAGAAGGAGAC

[a] Alternative probe used in reverse dot-blot procedure.

Toward this end, the sequence polymorphism at the HLA-DQα locus was determined by designing primers to conserved sequences, amplifying a 242 (or 239)-bp fragment containing the polymorphic sequences, and cloning and sequencing the amplified fragment (Scharf et al. 1986; Horn et al. 1988). More recently the amplified fragment has been sequenced directly without cloning (Gyllensten and Erlich 1988). The sequence analysis both of DQα genes from over 50 cell lines and of individual samples has revealed four major allelic types—DQA1, DQA2, DQA3, and DQA4. The DQA1 and DQA4 types can be further divided into the subtypes A1.1, A1.2, and A1.3 and A4.1, A4.2, and A4.3, for a total of eight sequence-defined alleles (Gyllensten and Erlich 1988; Horn et al. 1988).

The sequences of the oligonucleotide probes used for typing the allelic diversity at the DQα locus are shown in table 1. These probes have been labeled with HRP and hybridized to immobilized PCR-amplified DNA (dot-blot; Bugawan et al. 1988), or the oligonucleotides have been immobilized on a filter and hybridized to biotin-labeled PCR-amplified DNA (reverse dot-blot; Saiki et al. 1989). These probes do not distinguish the DQA4 subtypes (4.1, 4.2, and 4.3). The probe HE46 (Erlich and Bugawan 1989) hybridizes specifically to the DQA4.2 and DQA4.3 alleles but was not used in this population survey.

Material and Methods

Nomenclature

According to the most recent system of nomenclature adopted by the World Health Organization HLA Nomenclature Committee (Bodmer et al. 1990), the DQα locus is now known as DQA1, and the linked and homologous locus, previously known as DXα, is now termed DQA2. The allelic designations are DQA1*0101 (previously A1.1), *0102 (previously A1.2), *0103 (previously A1.3), DQA1*0201 (previously A2), *0301 (previously A3), DQA1*0401 (previously A4.2), DQA1*0501 (previously A4.1), and DQA1*0601 (previously A4.3). Since the new locus designation (DQA1) is the same as our previous allele designation, DQA1, in the present paper we have retained the older nomenclature to avoid confusion.

Populations/samples

Samples of unrelated individuals were obtained from sources as noted below, some as clean, extracted DNA

(Maniatis et al. 1982) and some as lymphocytes isolated by buffy coat and stored at −70°C in dimethylsulfoxide. The DNA was extracted from the lymphocyte samples and was made available for PCR by a quick lysis procedure described elsewhere (Higuchi 1989; Kellogg and Kwok 1990).

Cetus Samples

Black.—A group of 99 DNA samples collected for a sickle-cell screening program were obtained from Steve Embury (San Francisco General Hospital/University of California, San Francisco); 97 lymphocyte samples collected for a breast cancer screening program were obtained from Mary-Claire King (University of California, Berkeley); and 28 DNA samples were received from Ed Blake (forensic case studies). The total number of samples was 224.

Caucasian.—A group of 324 DNA samples were received from Ed Blake (forensic case studies); 80 DNA samples from the parents of the CEPH (Centre D'Etude du Polymorphisme Humaine) family collection were analyzed; five DNA samples were obtained from Allan Wilson (University of California, Berkeley), and four lymphocyte samples were collected from Cetus employees. The total number of samples was 413.

Japanese.—A total of 92 DNA samples, collected from medical students participating in a study on responsiveness to a hepatitis vaccine, were obtained from Takehiko Sasazuki (Kyushu University, Fukuoka, Japan).

Mestizo.—A total of 100 lymphocyte samples, collected in Mexico City and environs for this study, were obtained from Clara Gorodezky (Instituto de Salubridad y Enfermedades Tropicales, Mexico City).

Hispanic.—A total of 169 DNA samples, collected in Mexico by Norma Schmill through a Hereditary Disease Program, were obtained from David Hoar (Alberta Children's Hospital, Calgary).

Southeast Asian.—A total of 87 DNA samples (Chinese, Vietnamese, Korean, Thai, and Filipino) collected for an α-thalassemia screening program in San Francisco, were received from Steve Embury.

Isolated populations.—Genomic DNAs from Bedouins of Saudi Arabia (33 samples) and Nigerians (12 samples) were provided by J. Wainscoat. Native Australian DNAs (16 samples from Perth, Derby, Darwin, and Alice Springs) and Papua New Guinean DNAs (134 samples from localities described by Stoneking et al. [1990]) were purified from placental tissue. Indonesian DNAs (144 samples from the islands of Ternate, Hiri, Alor, Flores, Roti, and Timor) were purified from lymphocytes.

Roche Biomedical Laboratories (RBL) Samples

A total of 492 samples (172 black, 174 Caucasian, and 146 Hispanic) were obtained from a random sampling of blood specimens. These samples were collected, with no geographical preference, from individuals throughout the United States. Total cell DNA was extracted from leukocytes by using a modification of a procedure of Miller et al. (1988).

PCR Amplification

Amplification by PCR was performed on approximately 0.5 μg human genomic DNA by using 20 pmol each of primers GH26 and GH27 (Scharf et al. 1986). Two and one-half units of AmpliTaq™r*Taq* DNA Polymerase (Perkin Elmer Cetus Instruments [PECI]) (Saiki et al. 1988*b*) were added in a 100-μl vol of polymerase buffer containing 50 mM KCl, 10 mM Tris pH 8.4, 2.5 mM $MgCl_2$, and 0.01% gelatin. Each of the four deoxynucleotide triphosphates was present at 2 mM (8 mM total dNTP). The reactions were amplified for 30 cycles by using a PECI thermal cycler (denaturation was at 94°C for 30 s; annealing was at 55°C for 30 s; and extension was at 72°C for 30 s).

DQα Typing

A dot-blot procedure was done essentially according to a method described by Bugawan et al. (1988) by using the oligonucleotide probes listed in table 1 and hybridization and wash conditions that distinguish alleles that differ by as little as a single base. These probes were labeled by the conjugation of their 5′ ends with horseradish peroxidase (HRP). Hybridized probe was detected by an HRP-catalyzed reaction resulting in a colored precipitate that adheres to the dot-blot membrane.

Some samples were also typed by the "reverse dot-blot procedure" (Saiki et al. 1989) with biotinylated DQα primers and immobilized oligonucleotide probes (Erlich and Bugawan 1989; Saiki et al. 1989). The DQα genotypes obtained with both the dot-blot and the reverse dot-blot procedures have been identical for all samples tested.

Results

The allele frequencies for the 11 populations studied are shown in table 2. As shown, Caucasian, black, and

Table 2

Distribution (%) of DQα Alleles in Various Human Populations

HLA-DQα Allele	Population											
	Black (n = 448)	Black (RBL) (n = 344)	Caucasian (n = 826)	Caucasian (RBL) (n = 348)	Japanese (n = 184)	Mestizo (n = 200)	Hispanic (Mexican) (n = 338)	Hispanic (RBL) (n = 292)	Southeast Asian (n = 174)	Indonesian (n = 288)	Papua New Guinean (n = 268)	Australian (n = 32)	Bedouin (n = 66)	Nigerian (n = 24)
1.1	15	11.3	13.7	12.6	8.7	9	8	11.6	19	22.6	28.0	18.8	7.6	8.3
1.2	26.3	29.4	19.7	24.4	12	9	5.6	20.2	23.6	46.9	40.3	6.3	24.2	41.7
1.3	4.5	3.8	8.5	4.3	22.8	4	1.2	3.8	0	2.8	5.6	18.8	12.1	4.2
2	12.1	11.1	10.9	13.5	.5	8	5	7.5	6.9	2.4	.4	6.3	19.7	16.7
3	11.8	12.2	20.1	16.7	44.6	29.5	43.5	23.6	25.3	5.6	15.7	28.1	13.6	0
4	30.4	32.3	27.1	28.5	11.4	40.5	36.7	33.2	25.3	19.8	10.1	21.9	22.7	29.2

NOTE.—DQα typing was performed as described in Material and Methods. n = Number of alleles in sample. The similarity of the allele frequencies determined by Cetus and RBL for black, Caucasian, and Hispanic populations was tested by the G-test for heterogeneity (Sokal and Rohlf 1981). For the two black populations, the G-test yielded a value of 4.95, indicating a virtually identical distribution. The distribution of the two Caucasian populations was also very similar, with a G-test value of 11.9 (equivalent to χ^2 of 11.1 with 5 df, p = .05). The distributions in the RBL U.S. Hispanic and in the Cetus Mexican Hispanic populations were different, with a G-test value of 54.6 (equivalent to χ^2 of 20.5, or p = .001). The reasons for this difference are discussed in the text. The distribution of DQα alleles between pairwise combinations of the different ethnic populations was also evaluated by χ^2 analysis (see text) and was found to be significantly different for most pairs.

Hispanic populations were studied by two different labs (Cetus and RBL). The frequencies determined by the two labs for the Caucasian samples are very similar, as are the distributions of alleles for the two black populations (see Note to table 2). In contrast, the allele frequencies for the Cetus Hispanic samples (random samples from Mexico) are more similar to the known Mestizo samples (from Mexico City) than to the RBL Hispanic samples, (chosen, by Spanish surname, from U.S. samples) for which the allele distribution is more similar to that of the Caucasian or black population. The term "Hispanic," which denotes "defined by Spanish surname," can include Mexicans, Cubans, Puerto Ricans, and Spaniards, as well as other Latin American nationalities, and, therefore, does not really define a homogeneous racial group.

The allele distribution in the Japanese population differs dramatically from those of all the other groups, including the Southeast Asians. The Japanese frequency of the A2 allele is lower (0.5%, vs. 5%–13.5% in the other populations), as is the frequency of the A4 allele (11.4%, vs. 25%–40% in the other populations), and the frequency of the Japanese A1.3 allele is significantly higher (22.8%, vs. 0%–8.5% in the other populations).

Heterogeneity in DQα allele frequencies was assessed by the method of unplanned tests (Sokal and Rohlf 1981) for all pairwise comparisons of the 11 populations. Only 13 of the 55 pairwise comparisons were not statistically significant; 12 of these involve Nigerians, Australians, and/or Bedouins and presumably reflect the much smaller sample sizes of these populations. All of the other comparisons showed significant between-population heterogeneity in DQα allele frequencies.

The observed frequency of genotypes, as well as the expected frequencies based on the assumption of Hardy-Weinberg equilibrium for the large populations surveyed, are shown in table 3. No significant differences are observed. This observation is significant for the utility of the HLA-DQα marker in individual identification, because it indicates that genotype frequencies can be reliably estimated from allele frequency data. The expected genotype frequencies in five isolated populations of anthropologic interest (Indonesian, Papua New Guinean, Australian, Bedouin, and Nigerian; table 3) indicate, at the DQα locus, a genetic diversity (i.e., mean heterozygosity) that is quite high (.69 and .72, respectively). This is comparable to the range of diversity values observed in the less isolated populations.

The power of discrimination, PD (Fisher 1951) (PD = $1-\Sigma P_j^2$, where P_j is the frequency of each genotype), calculated from the genotype data in table 3 is

Table 3

Observed and Expected HLA-DQα Genotype Frequencies (%)

HLA-DQα Genotype	Population										
	Black (n = 224)	Black (RBL) (n = 172)	Caucasian (n = 413)	Caucasian (RBL) (n = 174)	Japanese (n = 92)	Mestizo (n = 100)	Hispanic (Mexican) (n = 169)	Hispanic (RBL) (n = 146)	Southeast Asian (n = 87)	Indonesian (n = 144)	Papua New Guinean (n = 134)
1.1, 1.1	3.6 (2.3)	.0 (1.3)	2.2 (1.9)	1.2 (1.6)	.0 (.8)	1.0 (.8)	.6 (.6)	2.7 (1.4)	5.8 (3.6)	3.5 (5.1)	5.2 (7.8)
1.1, 1.2	7.6 (7.9)	12.2 (6.6)	3.6 (5.4)	5.2 (6.2)	1.1 (2.1)	3.0 (1.6)	1.2 (.9)	4.8 (4.7)	10.3 (9.0)	22.9 (21.2)	23.1 (22.6)
1.1, 1.3	.9 (1.4)	.0 (.9)	2.9 (2.3)	1.7 (1.1)	6.5 (4.0)	1.0 (.7)	.0 (.2)	1.4 (.9)	.0 (.0)	1.4 (1.3)	3.7 (3.1)
1.1, 2	3.6 (3.6)	2.9 (2.5)	1.9 (3.0)	4.0 (3.4)	.0 (.1)	.0 (1.4)	.6 (.8)	1.4 (1.7)	1.2 (2.6)	.0 (1.1)	0 (.2)
1.1, 3	2.7 (3.5)	1.2 (2.8)	5.3 (5.5)	2.9 (4.2)	7.6 (7.8)	4.0 (5.3)	6.5 (7.0)	4.8 (5.5)	8.1 (9.6)	4.2 (2.5)	12.7 (8.8)
1.1, 4	8.0 (9.1)	6.4 (7.3)	9.2 (7.4)	9.2 (7.2)	2.2 (2.0)	8.0 (7.3)	6.5 (5.9)	5.5 (7.7)	6.9 (9.6)	9.7 (8.9)	6 (5.7)
1.2, 1.2	8.5 (6.9)	5.8 (8.6)	4.6 (3.9)	8.6 (6.0)	1.1 (1.4)	.0 (0.8)	.0 (.3)	4.8 (4.1)	8.1 (5.6)	21.5 (22.0)	19.4 (16.2)
1.2, 1.3	2.2 (2.4)	1.2 (2.2)	3.4 (3.4)	1.2 (2.1)	3.3 (5.5)	.0 (.7)	.0 (.1)	.7 (1.5)	.0 (.0)	4.2 (2.6)	2.2 (4.5)
1.2, 2	4.0 (6.4)	6.4 (6.5)	4.6 (4.3)	4.6 (6.6)	.0 (.1)	2.0 (1.4)	1.2 (.6)	4.8 (3.0)	3.5 (3.3)	3.5 (2.3)	.7 (.3)
1.2, 3	7.1 (6.2)	8.1 (7.2)	8.2 (7.9)	7.5 (8.2)	12.0 (10.7)	5.0 (5.3)	4.7 (4.9)	11.6 (9.5)	10.3 (11.9)	4.9 (5.3)	8.2 (12.7)
1.2, 4	14.7 (16.0)	19.2 (19.0)	10.4 (10.7)	13.2 (13.9)	5.4 (2.7)	8.0 (7.3)	4.1 (4.1)	8.9 (13.4)	6.9 (11.9)	15.3 (18.6)	7.5 (8.1)
1.3, 1.3	.0 (.2)	.6 (.1)	1.2 (.7)	.0 (.2)	4.4 (5.2)	.0 (.2)	.0 (.0)	.0 (.1)	.0 (.0)	.0 (0.1)	1.5 (0.3)
1.3, 2	2.2 (1.1)	1.2 (.8)	1.5 (1.9)	.6 (1.2)	1.1 (.2)	1.0 (.6)	.0 (.1)	.7 (.6)	.0 (.0)	.0 (0.1)	.0 (.0)
1.3, 3	1.3 (1.1)	.6 (.9)	1.7 (3.4)	.6 (1.4)	21.7 (20.3)	2.0 (2.4)	.6 (1.0)	.0 (1.8)	.0 (.0)	.0 (0.3)	2.2 (1.8)
1.3, 4	2.2 (2.7)	3.5 (2.5)	5.1 (4.6)	4.6 (2.5)	4.4 (5.2)	4.0 (3.2)	1.8 (.9)	4.8 (2.5)	.0 (.0)	.0 (1.1)	.0 (1.1)
2, 2	2.2 (1.5)	.6 (1.2)	2.2 (1.2)	2.3 (1.8)	.0 (.0)	1.0 (.6)	1.2 (.3)	.7 (.6)	1.2 (.5)	.0 (.1)	.0 (.0)
2, 3	1.3 (2.9)	5.2 (2.7)	4.8 (4.4)	4.6 (4.5)	.0 (.5)	5.0 (4.7)	1.8 (4.4)	1.4 (3.5)	3.5 (3.5)	.0 (.3)	.0 (.1)
2, 4	8.5 (7.4)	5.2 (7.2)	4.6 (5.9)	8.6 (7.7)	.0 (.1)	6.0 (6.5)	4.1 (3.7)	5.5 (5.0)	3.5 (3.5)	1.4 (1.0)	.0 (.1)
3, 3	.9 (1.4)	.0 (1.5)	4.4 (4.0)	3.5 (2.8)	19.6 (19.9)	9.0 (8.7)	22.5 (18.9)	7.5 (5.6)	5.8 (6.4)	.0 (.3)	3.0 (2.5)
3, 4	9.4 (7.2)	9.3 (7.9)	11.4 (10.9)	10.9 (9.5)	8.7 (10.2)	25.0 (23.9)	28.4 (31.9)	14.4 (15.7)	17.2 (12.8)	2.1 (2.2)	2.2 (3.2)
4, 4	8.9 (9.2)	10.4 (10.4)	6.8 (7.3)	5.2 (8.1)	1.1 (1.3)	15.0 (16.4)	14.2 (13.5)	13.7 (11.0)	8.1 (6.4)	5.6 (3.9)	2.2 (1.0)
PD	.92	.91	.94	.93	.88	.88	.83	.91	.92	.86	.87
h	.79	.77	.81	.80	.72	.73	.67	.78	.78	.69	.72

NOTE.—Expected genotype frequencies (shown in parentheses) were calculated, on the basis of Hardy-Weinberg equilibrium, from the allele frequencies in table 1. The power of discrimination values (PD) and allelic diversity values (h) are shown. n = Number of individuals in the sample. The allelic diversity ($[1 - \Sigma X_i^2][n/n-1]$; Nei and Royuchoudhury 1974) is an unbiased estimate of the expected heterozygosity (corrected for sample size) where X_i = allele frequencies and n = sample size (no. of alleles). The allelic diversities for the Australian, Bedouin, and Nigerian populations (genotype frequencies not shown because of small sample size) are .82, .84, and .73, respectively.

shown for each population in table 3. The PD values range from .83 to .94. For paternity determination, where the distribution of alleles—not genotypes—is the relevant parameter, the average power of exclusion for this DQα genotyping system was determined from the equation derived by Garber and Morris (1983). This equation gives the average power of exclusion of a genetic system of n codominant alleles, where $i = 1, \ldots n$ is the index for alleles of frequencies P_i, P_j, . . . Pn. On the basis of the allelic frequencies presented in table 1, the average power of exclusion for this genotyping system is 62.7%, 59.2%, and 58.8% for Caucasians, blacks, and Hispanics, respectively. These values are comparable to the combined power of exclusions for four relatively informative red-blood-cell antigens or serum protein genetic markers (Peck 1989).

From the paternity cases analyzed (Caucasians and blacks), paternity index (PI) and random man not excluded (RMNE) were determined by the methods of Lee (1983) and Lee and Lebeck (1983), respectively. The mean PI was 2.56 (probability of paternity being 71.9%) for Caucasians and 2.28 (probability of paternity being 69.5%) for blacks. The mean RMNE for Caucasians was 40.2%, and that for blacks was 45.0%. An unusual example of the utility of DQα typing in paternity testing is the resolution of a case involving a deceased child, where the typing was based on the ability to PCR amplify informative sequences from a tissue section. In this case, DQα typing of the mother, two alleged fathers, and a kidney section from the deceased child excluded one potential father and included the other, on the basis of the shared A1.3 allele. This analysis yielded a 92% probability of paternity (J. Chimera and M. C. Luce, unpublished data).

Discussion

The HLA-DQα dot-blot typing system using PCR-amplified DNA and nonradioactive oligonucleotide probes is a powerful approach for analyzing the genetic structure of populations. It is a simple and rapid method for typing large numbers of samples, and only very small amounts (e.g., single hairs) of the individual samples are required (Higuchi et al. 1988). The definition of alleles by specific hybridization probes has some advantages over the definition by electrophoretic mobility, in that problems in identifying the precise size of DNA fragments in different lanes of the gel are eliminated. However, this method raises the issue of potential "blank" alleles—i.e., alleles which either fail to amplify with the primers or fail to type with the probes. In our survey of over 2,000 individuals and over 200 cell lines since this typing method was first reported (Saiki et al. 1986), we have observed no evidence of such putative blank alleles.

First, we have never observed a sample which amplified the DQα 242-bp fragment but failed to type with the oligonucleotide probes (i.e., a putative homozygous blank variant in the probe region). Second, we observe no excess of homozygotes over the expected frequency (table 2) and thus find no evidence of a putative heterozygote with one blank allele. Third, in family studies we have never observed a homozygote who failed to transmit his or her one DQα allele to *all* progeny. Fourth, phylogenetic analysis of DQα sequences from a variety of primate species indicates that the DNA sequences complementary to the primers and to the A1, A3, and A4 probes as well as to the subtyping probes have been conserved for 5–20 million years (Gyllensten and Erlich 1989). The human DQA2 allele appears to have arisen after speciation and represents a more recent allele derived from recombination between the ancestral A3 and A4 allelic types (Gyllensten and Erlich 1989). In fact, these DQα primers and probes can be used to type individual nonhuman primates (U. B. Gyllensten, unpublished data). It is possible that, as more ethnic groups are examined, very rare, new DQα allelic variants may be identified, but, if so, these are likely to be new subtypes of the existing allelic types (A1, A2, A3, and A4) and would not type as blanks.

Polymorphism in the HLA-D region previously has been characterized by using serologic reagents to type the subset of cells that expressed the class II antigens. A detailed discussion of the relationship between these sequence-defined alleles and the serologic specificities will be presented elsewhere (H. A. Erlich, unpublished data). In brief, however, the DQA1 alleles (1.1, 1.2, and 1.3) correspond to haplotypes encoding the DQw1 specificity, and the DQA1.1 subtype is associated with haplotypes encoding the DR1 specificity. The DQA2 allele corresponds to haplotypes encoding the DR7 specificity. The reported frequencies of these serologic specificities (Albert 1984) agree very closely with the frequency of the associated DNA-defined alleles reported here.

For the populations surveyed here, the observed frequencies of DQα genotypes do not deviate significantly from those expected on the assumption of Hardy-Weinberg equilibrium, although there is a slight excess of homozygotes in the RBL Hispanic and in the Cetus Southeast Asian groups (table 4). This observation is consistent with the known heterogeneity of these groups

Table 4

Observed versus Expected Frequency of Homozygous Genotypes

Marker and Population	% Observed	% Expected
DQα:		
Black (Cetus)	24	22
Black (RBL)	17	23
Caucasian (Cetus)	21	19
Caucasian (RBL)	21	20
Southeast Asian	29	22
Hispanic (Cetus)	38	34
Hispanic (RBL)	29	23
D2S44:		
Hispanic[a]	17	4
D17S79:		
Hispanic[a]	13	4

[a] Source: Lander (1989).

as defined. The discrepancy between the allele frequencies of the RBL Hispanic group and those of the Cetus Hispanic (Mexican) group also reflects the genetic heterogeneity of a population defined simply by Spanish surname (see Results). The dramatic difference in DQα allele frequencies between different populations (table 2) suggests that this will be a valuable marker for exploring anthropologic issues, such as the patterns of migration and the evolutionary history of human populations.

When employed for individual discrimination and exclusion, the use of an allele-specific typing system, as described here for the DQα locus, has several advantages over variable-number-of-tandem-repeat (VNTR) systems defined by RFLP. An allele-specific system identifies discrete traits which are transmitted in a clear Mendelian fashion. In the case of the DQα locus, the distinctness and permanence of the allelic variants are emphasized by their maintenance for millions of years (Gyllensten and Erlich 1989). Comparison of the observed genotype frequencies to Hardy-Weinberg expected genotype frequencies can help validate the typing method—and, if there is homozygote excess, reveal population substructure (Wahlund 1928). The close fit to Hardy–Weinberg expectation for the DQα genotype frequencies reported here (tables 3 and 4) affirms the typing methodology and genetic model and further suggests that neither great departures from panmixis nor population substructuring exists for most of the populations as defined.

The assumption of Hardy-Weinberg equilibrium allows the estimation of genotype frequencies from observed allele frequencies and makes possible high levels of individual discrimination from relatively limited population data. For many VNTR systems the discrimination of alleles is limited by the resolution of the gel electrophoresis system. Consequently, distortions from Hardy-Weinberg equilibrium are difficult to interpret. The great excess of homozygotes found in VNTR population samples (see table 4; Lander 1989; Mueller, in press) cannot be clearly assigned to a single cause. Both the inability to fully discriminate classes because of limitations of the typing methodology and potential subpopulation structure may contribute to a homozygote excess. The high mutation rate present in some VNTR systems, which may be as great as 5% (Jeffreys et al. 1988), also creates the expectation of homozygote excess—but, in this case, because of subpopulation structuring.

Although the power of discrimination for the DQα marker is less than that for some VNTR systems (Balazs et al. 1989), this PCR-based typing system is a simple and rapid method capable of analyzing minute amounts of sample as well as significantly degraded DNA templates. As more PCR-based markers become available, a panel of such tests should be able to provide valuable information for individual inclusions, in addition to the exclusionary value provided by the DQα test. Finally, the use of HLA-DQα genotyping for individual identification has been significantly facilitated by the recent development of a reverse dot-blot format which utilizes immobilized oligonucleotide probes and a biotin-labeled PCR-amplification product (Saiki et al. 1989). This nonradioactive method, which allows the analysis of a sample in a single hybridization reaction with a typing strip to which all of the probes have been fixed, is now available commercially.

The HLA-DQα typing system described here (either dot-blot or reverse dot-blot or both) has been used, as of December 1989, in 106 forensic cases involving the analysis of over 1,000 evidence samples, including hair, blood stains, semen stains, and bone fragments. A detailed discussion of its use in forensics casework will be reported elsewhere (E. Blake, R. Higuchi, and H. A. Erlich, unpublished data).

Acknowledgments

We are grateful to David Hoar, Steve Embury, Mary-Claire King, Allan Wilson, Takehiko Sasazuki, Ann Bowcock, and the CEPH for providing samples. We thank Glenn Horn, Ann Begovich, and Ulf Gyllensten for helpful discussions. In addition, we thank Donald Cain and Kimberly Smith for excel-

lent technical assistance; George Maha for calculating RMNE and PI values; Corey Levenson, Lauri Goda, and Dragan Spasic for oligonucleotide preparation; and Kathy Levenson for careful preparation of the manuscript. The work of W.K. was supported by NIH grant GM35326.

References

Albert ED, Baur MP, Mayr WR (eds) (1984) Histocompatibility testing. Springer, New York

Balazs I, Baird M, Clyne M, Meade E (1989) Human population genetic studies of five hypervariable DNA loci. Am J Hum Genet 44:182–190

Bodmer JG, Marsh SGE, Parham P, Erlich HA, Albert E, Bodner WF, Dupont B, et al (1990) Nomenclature for factors of the HLA system, 1990. Tissue Antigens 35:1–8

Bugawan TL, Saiki RK, Levenson CH, Watson RM, Erlich HA (1988) The use of non-radioactive oligonucleotide probes to analyze enzymatically amplified DNA for prenatal diagnosis and forensic typing. Bio/Technology 6: 943–947

Erlich HA, Bugawan T (1989) HLA class II gene polymorphism: DNA typing, evolution, and relationship to disease susceptibility. In: H Erlich (ed) PCR technology: principles and applications for DNA amplification. Stockton, New York, pp 193–208

Fisher RA (1951) Heredity 5:95–102

Garber RA, Morris JW (1983) General equations for the average power of exclusion for genetic system of n codominant alleles in one-parent and no-parent cases of disputed parentage. In: Walker RH (ed) Inclusion probabilities in parentage testing. American Association of Blood Banks, Arlington, VA, pp 277–280

Gyllensten UB, Erlich HA (1988) Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA–DQα locus. Proc Natl Acad Sci USA 85:7652–7656

——— (1989) Ancient roots for polymorphism at the HLA-DQα locus in primates. Proc Natl Acad Sci USA 86: 9986–9990

Higuchi R (1989) Simple and rapid preparation of samples for PCR. In: Erlich HA (ed) PCR technology: principles and applications for DNA amplification. Stockton, New York, pp 31–38

Higuchi R, von Beroldingen CH, Sensabaugh GF, Erlich HA (1988) DNA typing from single hairs. Nature 332:543–546

Horn GT, Bugawan TL, Long C, Erlich HA (1988) Allelic sequence variation of the HLA-DQ loci: relationship to serology and insulin-dependent diabetes susceptibility. Proc Natl Acad Sci USA 85:6012–6016

Jeffreys AJ, Royle NJ, Wilson V, Wong Z (1988) Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. Nature 332:278–281

Kellogg DE, Kwok S (1990) Detection of human immunodeficiency virus. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ (eds) PCR protocols: a guide to methods and applications. Academic Press, San Diego, pp 337–347

Lander E (1989) DNA fingerprinting on trial. Nature 339: 501–505

Lee CL (1983) Paternity case analysis: estimation of paternity index (X/Y). In: Walker RH (ed) Inclusion probabilities in parentage testing. American Association of Blood Banks, Arlington, VA, pp 507–524

Lee CL, Lebeck LK (1983) Paternity case analysis: derivation of random men not excluded values: a practical solution. In: Walker RH (ed) Inclusion probabilities in parentage testing. American Association of Blood Banks, Arlington, VA, pp 525–534

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Miller SA, Dykes KK, Polesky HF (1988) A simple salting-out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16:1215

Mueller L. Population genetics of hypervariable human DNA In: Farley M, Harrington J (eds) Genetics in forensics. Lewis, Chelsea, MI (in press)

Mullis K, Faloona F (1987) Specific synthesis of DNA in vitro via a polymerase catalysed chain reaction. Methods Enzymol 155:335–350

Nei M, Roychoudhury AK (1974) Sampling variances of heterozygosity and genetic distance. Genetics 76:379–390

Peck S (1989) Parentage assessment and the technology available. J. Int Fed Clin Chem 1:150–157

Saiki R, Scharf S, Faloona F, Mullis K, Horn G, Erlich HA, Arnheim N (1985) Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230:1350–1354

Saiki R, Walsh PS, Levenson CH, Erlich HA (1989) Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. Proc Natl Acad Sci USA 86:6230–6234

Saiki RK, Bugawan TL, Horn GT, Mullis KB, Erlich HA (1986) Analysis of enzymatically amplified β-globin and HLA-DQα DNA with allele-specific oligonucleotide probes. Nature 324:163–166

Saiki RK, Chang CA, Levenson CH, Warren TC, Boehm CD, Kazazian HH Jr, Erlich HA (1988*a*) Diagnosis of sickle cell anemia and β-thalassemia with enzymatically amplified DNA and non-radioactive allele-specific oligonucleotide probes. N Engl J Med 319:537–541

Saiki RK, Gelfand DH, Stoffel S, Scharf S, Higuchi RH, Horn GT, Mullis KB, et al (1988*b*) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487–491

Scharf SJ, Horn GT, Erlich HA (1986) Direct cloning and sequence analysis of enzymatically amplified genomic sequences. Science 223:1076–1078

Sokal RR, Rohlf FJ (1981) Biometry: the principles and prac-

tice of statistics in biological research, 2d ed. WH Freeman, New York

Stoneking M, Jorde LB, Bhatia K, Wilson AC (1990) Geographic variation in human mitochondrial DNA from Papua New Guinea. Genetics 124:717–733

Trowsdale J, Young JAT, Kelly AP, Austin PJ, Carson S, Meunier H, So A, et al (1985) Structure, sequence and polymorphism in the HLA-D region. Immunol Rev 85:5–43

Wahlund S (1928) Zusammensetzung von Populationen und Korrelationserscheinungen von Standpunkt der Vererbungslehre aus Betrachtet Hereditas 11:65–106