TECHNICAL NOTE

Linda B. Jankowski,[1] M.S.; Bruce Budowle,[2] Ph.D.; Nancy T. Swec,[1] B.S.; Joseph A. Pino,[1] B.S.; Sharon Freck-Tootell,[1] B.S.; Harry W. Corey,[1] B.S.; Robyn Schwartz,[1] B.S.; Edward J. LaRue,[1] B.A.; William L. Rochin,[1] B.S.; Charles J. Kearney,[1] B.S.; and Margaret L. Tarver,[1] J.D., M.S.

New Jersey Caucasian, African American, and Hispanic Population Data on the PCR-Based Loci HLA-DQA1, LDLR, GYPA, HBGG, D7S8, and Gc

REFERENCE: Jankowski LB, Budowle B, Swec NT, Pino JA, Freck-Tootell S, Corey HW, Schwartz R, LaRue EJ, Rochin WL, Kearney CJ, Tarver ML. New Jersey Caucasians, African American, and Hispanic population data on the PCR-based loci HLA-DQA1, LDLR, GYPA, HBGG, D7S8, and Gc. J Forensic Sci 1998; 43(5):1037–1040.

ABSTRACT: New Jersey Caucasian, African American, and Hispanic genotype and allele frequencies were determined for the six PCR-based loci, HLA-DQA1, LDLR, GYPA, HBGG, D7S8, and Gc. All but one locus (HLA-DQA1 for African Americans) meet Hardy-Weinberg expectations. However, observing one departure in 18 loci over the three New Jersey sample populations is not unexpected. There is little evidence for departures from independence between pairs of loci in the three populations studied. Thus, multiple locus profile frequencies can be determined using the product rule.

KEYWORDS: forensic science, DNA typing, population genetics, PCR, Hardy-Weinberg equilibrium, linkage equilibrium, HLA-DQA1, LDLR, GYPA, HBGG, D7S8, Gc

The predominant DNA PCR-based genetic loci used in forensic identity testing are HLA-DQA1 and the polymarkers, LDLR, GYPA, HBGG, D7S8, and Gc. We have previously shown in a New Jersey Caucasian population database that the six loci meet Hardy-Weinberg expectations and demonstrate gametic phase equilibrium (1). To date, there are no published data which describe the frequencies of the subtypes of the HLA-DQA1 4 allele (i.e., the 4.1 and the 4.2/4.3 delineation) in United States population groups (2–4). This study expands our previous report of New Jersey Caucasians to include the subtypes of the HLA-DQA1 4; in addition, New Jersey African American and Hispanic databases for HLA-DQA1 and polymarker loci have been established. The data presented here show that all but one locus meet Hardy-Weinberg expectations and that there is little evidence for association among the six loci studied. Thus, an estimate of the rarity of a multiple locus DNA profile can be determined from the frequency data presented here.

Materials and Methods

Sample Preparation

Whole blood samples were obtained by venipuncture in ethylenediametetraacetate (EDTA) vacutainer tubes or by fingerprick. For the Caucasian population, 164 samples from New Jersey Caucasians were collected from unrelated volunteers. The contributors to the 285 unrelated African Americans included 99 volunteers, 100 convicted sex offenders, and 86 known samples from casework. The Hispanic database of 128 unrelated individuals consisted of 58 volunteers, 41 convicted sex offenders, and 29 known samples from casework.

The DNA was isolated using chelex extraction (5,6). An estimate of the quantity of extracted DNA was determined by the slot-blot procedure previously described by Waye et al. (7) using the Perkin Elmer Quantiblot kit according to the protocol recommended by the manufacturer. Approximately 2 ng of DNA were used for subsequent amplification.

Typing

The HLA-DQA1 and polymarker loci were typed using the Amplitype PM + DQA1 PCR Amplification and Typing Kit (Per-

[1] DNA technical director, senior forensic scientists, and technical director, repectively, New Jersey State Police South Regional Laboratory, Hammonton, NJ 08037.

[2] Unit chief, Forensic Science Research and Training Center, FBI Academy, Quantico, VA 22135.

Received 9 June 1997; and in revised form 21 Nov. 1997; accepted 9 Feb. 1998.

kin-Elmer) according to the manufacturer's protocol. Approximately 2 ng of DNA were amplified in a Perkin-Elmer DNA Thermal Cycler 480.

Statistical Analysis

The allele frequencies were calculated from the observed genotypic frequencies using the gene count method. Unbiased estimates of the expected heterozygosity were calculated according to the method described by Edwards et al. (8). Possible divergence from Hardy-Weinberg equilibrium expectations (HWE) was determined by calculating the unbiased estimate of the expected homozygote/heterozygote frequencies (9–11), the likelihood ratio test (8,12,13) and the exact test (14), based on 1000 shuffling experiments. An interclass correlation criterion (15) was used for detecting disequilibrium between pairs of loci.

An RXC contingency table exact test was used to generate a G-statistic (2000 shuffling experiments) (16,17) to test for homogeneity between the New Jersey population samples and other U.S. population data. The program was kindly provided by R. Chakraborty (University of Texas School of Biomedical Sciences, Houston, TX).

Calculations to determine expected genotypic frequencies based on observed allele frequencies were performed using the product rule from the standard Hardy-Weinberg binomial expansion.

TABLE 1—*Observed allele frequencies in New Jersey Caucasian, African American, and Hispanic sample populations.*

Locus/Allele	Caucasians	African American	Hispanic
LDLR	**(n = 164)**	**(n = 285)**	**(n = 128)**
A	0.412	0.174	0.469
B	0.588	0.826	0.531
GYPA	**(n = 164)**	**(n = 285)**	**(n = 128)**
A	0.564	0.533	0.574
B	0.436	0.467	0.426
HBGG	**(n = 164)**	**(n = 284)**	**(n = 128)**
A	0.436	0.412	0.414
B	0.561	0.255	0.508
C	0.00305	0.333	0.781
D7S8	**(n = 164)**	**(n = 285)**	**(n = 128)**
A	0.628	0.649	0.523
B	0.372	0.351	0.477
Gc	**(n = 164)**	**(n = 285)**	**(n = 128)**
A	0.268	0.107	0.23
B	0.14	0.716	0.258
C	0.591	0.177	0.512
DQA1	**(n = 164)**	**(n = 285)**	**(n = 127)**
1.1	0.152	0.139	0.154
1.2	0.207	0.281	0.146
1.3	0.0701	0.0421	0.0591
2	0.152	0.119	0.122
3	0.146	0.0807	0.213
4.1	0.247	0.211	0.228
4.2/4.3	0.0244	0.128	0.0787

Results and Discussion

The distributions of observed allele frequencies for the HLA-DQA1 and polymarker loci are listed in Table 1. The observed number and the observed and expected genotype frequencies are shown in Table 2.

Of the 18 loci evaluated, only the HLA-DQA1 locus in African Americans does not meet Hardy-Weinberg expectations ($p = 0.002$, exact test) (Table 3). This is only one departure observed in a total of 18 loci analyzed. Moreover, the estimates of the HLA-DQA1 genotypes using the New Jersey African American data and those derived using U.S. African American data by Budowle et al. (2,18) would not be substantially different. Thus using the product rule under HWE assumptions provides a valid estimate for genotypic frequencies at this locus.

An interclass correlation test used to evaluate assumptions of independence demonstrated that there is little detectable evidence for association between the alleles at any of the pair-wise comparisons (Table 4). Out of a total of 45 pair-wise comparisons, significant departures were observed in only two pair-wise comparisons: D7S8 and HLA-DQA1 in African Americans ($p = 0.042$) and GYPA and HBGG in Hispanics ($p = 0.005$). The two examples of deviation from gametic phase equilibrium represent approximately 4.5% of the total number of pairwise comparisons; this is no more than would be expected by chance. Therefore the assumption of independence holds for the three New Jersey population groups.

The New Jersey African American data were statistically similar with U.S. African American data described by Budowle et al. (2,18) at all loci except HLA-DQA1 ($p = 0.007$). The frequencies of the HLA-DQA1 alleles 1.1, 1.2, 1.3, 2, 3, 4.1, 4.2/4.3 described by Budowle et al. are 0.112, 0.308, 0.051, 0.078, 0.148, 0.189, and 0.114, respectively. It is obvious that although statistically different, the estimates of HLA-DQA1 genotype frequencies would not be substantially different between the two African American sample populations. Furthermore, when a six-locus profile frequency is estimated, both African American databases would yield similar results (data not shown). Thus, a statistical difference at one locus has little forensic significance. For the Hispanic data, only the HBGG locus was significantly different from the Hispanic data described by Budowle et al. ($p < 10^{-3}$). These data are remarkably similar even though the two databases derive from geographically distinct regions (i.e., New Jersey vs. Texas). The New Jersey Caucasian population data were compared previously with other Caucasian data (1); therefore only the new HLA-DQA1 data were compared. The two Caucasian sample populations are similar at the HLA-DQA1 locus.

In conclusion, databases for three New Jersey population groups (Caucasian, African American, and Hispanic) have been established for six polymorphic PCR-based loci, HLA-DQA1, LDLR, GYPA, HBGG, D7S8, and Gc. Moreover, this study includes the subtyping data for the HLA-DQA1 4 allele. This evaluation demonstrates that for our New Jersey sample populations, the loci generally meet Hardy-Weinberg expectations, and there is little evidence for association across the loci. The population databases presented here can be used to determine an estimate of the rarity of a multiple locus profile for forensic identity purposes using the product rule.

TABLE 2—*Observed numbers, and observed (in parentheses) and expected [in brackets] genotype frequencies.*

Locus/type	Caucasians	African Americans	Hispanics
DQA1	**(n = 164)**	**(n = 285)**	**(n = 127)**
1.1,1.1	4 (0.0244) [0.0231]	11 (0.0386) [0.0193]	5 (0.0394) [0.0237]
1.2,1.2	9 (0.0549) [0.0428]	27 (0.0947) [0.0790]	4 (0.0315) [0.013]
1.3,1.3	0 (0.0) [0.00491]	2 (0.00702) [0.00177]	0 (0.0) [0.00349]
2,2	6 (0.0366) [0.0231]	5 (0.0175) [0.0142]	3 (0.0236) [0.0149]
3,3	4 (0.0244) [0.0213]	2 (0.00702) [0.00651]	4 (0.0315) [0.0454]
4.1,4.1	8 (0.0488) [0.0610]	13 (0.0456) [0.0445]	7 (0.0551) [0.0520]
	0 (0.0) [0.000595]	1 (0.00351) [0.0164]	1 (0.00787) [0.00619]
4.2/4.3,4.2/4.3			
1.1,1.2	10 (0.0610) [0.0629]	9 (0.0316) [0.0781]	4 (0.0315) [0.0450]
1.1,1.3	6 (0.0366) [0.0213]	9 (0.0316) [0.0117]	2 (0.0157) [0.0182]
1.1,2	5 (0.0305) [0.0462]	8 (0.0281) [0.0331]	3 (0.0236) [0.0376]
1.1,3	8 (0.0488) [0.0444]	3 (0.0105) [0.0224]	8 (0.0630) [0.0656]
1.1,4.1	13 (0.0793) [0.0751]	17 (0.0596) [0.0587]	10 (0.0787) [0.0702]
1.1,4.2/4.3	0 (0.0) [0.00742]	11 (0.0386) [0.0356]	2 (0.0157) [0.0242]
1.2,1.3	6 (0.0366) [0.0290]	3 (0.0105) [0.0237]	3 (0.0236) [0.0173]
1.2,2	11 (0.067) [0.0629]	25 (0.0877) [0.0669]	5 (0.0394) [0.0356]
1.2,3	6 (0.0366) [0.0604]	9 (0.0316) [0.0454]	10 (0.0787) [0.0622]
1.2,4.1	16 (0.0976) [0.102]	36 (0.126) [0.119]	4 (0.0315) [0.0666]
1.2,4.2/4.3	1 (0.00610) [0.0101]	24 (0.0842) [0.0719]	3 (0.0236) [0.0230]
1.3,2	3 (0.0183) [0.0213]	2 (0.00702) [0.0100]	2 (0.0157) [0.0144]
1.3,3	1 (0.00610) [0.0205]	1 (0.00351) [0.00679]	4 (0.0315) [0.0252]
1.3,4.1	6 (0.0366) [0.0346]	2 (0.00702) [0.0178]	3 (0.0236) [0.0269]
1.3,4.2/4.3	1 (0.00610) [0.00342]	3 (0.0105) [0.0108]	1 (0.00787) [0.00930]
2,3	7 (0.0427) [0.0444]	6 (0.0211) [0.0192]	5 (0.0394) [0.0520]
2,4.1	11 (0.0671) [0.0751]	9 (0.0316) [0.0502]	7 (0.0551) [0.0556]
2,4.2/4.3	1 (0.00610) [0.00742]	8 (0.0281) [0.0305]	3 (0.0236) [0.0192]
3,4.1	16 (0.0976) [0.0721]	14 (0.0491) [0.0341]	15 (0.118) [0.0971]
3,4.2/4.3	2 (0.0122) [0.00712]	9 (0.0316) [0.0207]	4 [0.0315] [0.0335]
4.1,4.2/4.3	3 (0.0183) [0.0121]	16 (0.0561) [0.0540]	5 (0.0394) [0.0359]
LDLR	**(n = 16.4)**	**(n = 285)**	**(n = 128)**
AA	26 (0.159) [0.170]	9 (0.0316) [0.0303]	31 (0.242) [0.220]
BB	55 (0.335) [0.346]	195 (0.684) [0.682]	39 (0.305) [0.282]
AB	83 (0.506) [0.485]	81 (0.284) [0.287]	58 (0.453) [0.498]
GYPA	**(n = 16.4)**	**(n = 285)**	**(n = 128)**
AA	46 (0.280) [0.318]	80 (0.281) [0.284]	45 (0.352) [0.329]
BB	25 (0.152) [0.190]	61 (0.214) [0.218]	26 (0.203) [0.181]
AB	93 (0.567) [0.492]	144 (0.505) [0.498]	57 (0.445) [0.489]
HBGG	**(n = 164)**	**(n = 284)**	**(n = 128)**
AA	34 (0.207) [0.190]	49 (0.173) [0.170]	21 (0.164) [0.171]
BB	55 (0.335) [0.315]	23 (0.0810) [0.0650]	31 (0.242) [0.258]
CC	0 (0.0) [0.00000930]	35 (0.123) [0.111]	0 (0.0) [0.00610]
AB	74 (0.451) [0.489]	58 (0.204) [0.210]	56 (0.438) [0.421]
AC	1 (0.00610) [0.00266]	78 (0.275) [0.274]	8 (0.0625) [0.0647]
BC	0 (0.0) [0.00342]	41 (0.144) [0.170]	12 (0.0938) [0.0793]
D7S8	**(n = 164)**	**(n = 285)**	**(n = 128)**
AA	60 (0.366) [0.394]	122 (0.428) [0.421]	38 (0.297) [0.274]
BB	18 (0.110) [0.138]	37 (0.130) [0.123]	32 (0.250) [0.228]
AB	86 (0.524) [0.467]	126 (0.442) [0.456]	58 (0.453) [0.499]
Gc	**(n = 164)**	**(n = 285)**	**(n = 128)**
AA	10 (0.0610) [0.0718]	6 (0.0211) [0.0114]	11 (0.0859) [0.0529]
BB	4 (0.0244) [0.0196]	149 (0.523) [0.513]	9 (0.0703) [0.0666]
CC	54 (0.329) [0.349]	13 (0.0456) [0.0313]	37 (0.289) [0.262]
AB	10 (0.0610) [0.0750]	42 (0.147) [0.153]	14 (0.109) [0.119]
AC	58 (0.354) [0.317]	7 (0.0246) [0.0379]	23 (0.180) [0.236]
BC	28 (0.171) [0.165]	68 (0.239) [0.253]	34 (0.266) [0.264]

NOTE: Expected genotype frequencies were determined from the observed allele frequencies; homozygotes were calculated using the formula p^2; heterozygotes were calculated using the formula $2pq$.

TABLE 3—*Test for Hardy-Weinberg equilibrium.*

	Caucasian	African American	Hispanic
HLA-DQA1			
Observed Homozygosity	18.9%	21.4%	18.9%
Exp. Homozygosity*	17.5%	18.0%	16.3%
Homozygosity Test†	0.633	0.132	0.436
Likelihood Ratio Test†	0.805	0.004	0.984
Exact Test†	0.825	0.002	0.958
LDLR			
Observed Homozygosity	49.4%	71.6%	54.7%
Exp. Homozygosity*	51.4%	71.2%	50.0%
Homozygosity Test†	0.604	0.901	0.289
Likelihood Ratio Test†	0.615	1.000	0.360
Exact Test†	0.615	0.847	0.280
GYPA			
Observed Homozygosity	43.3%	49.5%	55.5%
Exp. Homozygosity*	50.7%	50.1%	50.9%
Homozygosity Test†	0.059	0.823	0.302
Likelihood Ratio Test†	0.061	0.815	0.349
Exact Test†	0.061	0.911	0.349
HBGG			
Observed Homozygosity	54.3%	37.7%	40.6%
Exp. Homozygosity*	50.3%	34.4%	43.3%
Homozygosity Test†	0.313	0.252	0.538
Likelihood Ratio Test†	0.262	0.463	0.656
Exact Test†	0.262	0.428	0.890
D7S8			
Observed Homozygosity	47.6%	55.8%	54.7%
Exp. Homozygosity*	53.1%	55.4%	49.9%
Homozygosity Test†	0.152	0.630	0.280
Likelihood Ratio Test†	0.148	0.684	0.363
Exact Test†	0.148	0.603	0.277
Gc			
Observed Homozygosity	51.5%	58.9%	44.5%
Exp. Homozygosity*	44.0%	55.4%	37.9%
Homozygosity Test†	0.517	0.234	0.122
Likelihood Ratio Test†	0.603	0.158	0.226
Exact Test†	0.606	0.107	0.174

* Unbiased estimate.
† Probability values.

TABLE 4—*Two locus interclass correlation test for HLA-DQA1 and PM loci for unrelated New Jersey Caucasians, African Americans and Hispanics.*

Loci	Caucasian	African American	Hispanic
LDLR/GYPA	0.385	0.737	0.283
LDLR/HBGG	0.085	0.319	0.103
LDLR/D7S8	1.000	1.000	0.454
LDLR/Gc	0.620	0.574	0.557
LDLR/HLA-DQA1	0.401	0.733	0.492
GYPA/HBGG	0.466	0.795	0.005*
GYPA/D7S8	1.000	0.246	0.419
GYPA/Gc	0.600	0.498	0.914
GYPA/HLA-DQA1	0.738	0.585	0.640
HBGG/D7S8	0.646	1.000	0.681
HBGG/Gc	0.681	0.239	0.419
HBGG/HLA-DQA1	0.968	0.331	0.398
D7S8/Gc	0.758	0.800	0.546
D7S8/HLA-DQA1	0.176	0.042*	0.280
Gc/HLA-DQA1	0.350	0.256	0.268

* Deviation at $p = 0.05$ level.

References

1. Budowle B, Jankowski LB, Corey HW, Swec NT, Freck-Tootell S, Pino J, et al. Evaluation of independence assumptions for PCR-based and protein-based genetic markers in New Jersey Caucasians. J Forensic Sci 1997;42(2):223–5.
2. Budowle B, Lindsay JA, DeCou JA, Koons BW, Guisti AM, Comey CT. Validation and population studies of the loci LDLR, GYPA, HBGG, D7S8, and Gc (PM loci), and HLA-DQα using a multiplex amplification and typing procedure. J Forensic Sci 1995; 40(1):45–54.
3. Helmuth R, Fildes N, Blake E, Luce MC, Chimera J, Madej R, et al. HLA DQ alpha allele and genotype frequencies in various human populations, determined by using enzymatic amplification and oligonucleotide probes. Am J Hum Genet 1990;47:515–23.
4. Scholl S, Budowle B, Radecki K, Salvo M. Navajo, Pueblo, and Sioux population data on the loci HLA-DQA1, LDLR, GYPA, HBGG, D7S8, Gc, and D1S80. J Forensic Sci 1996;4:47–51.
5. Comey CT, Koons BW, Presley KW, Smerick JB, Sobieralski CA, Stanley DM, et al. DNA extraction strategies for amplified fragment length polymorphism analysis. J Forensic Sci 1994;39:1254–69.
6. Jung JM, Comey CT, Baer DB, Budowle B. Extraction strategy for obtaining DNA from bloodstains for PCR amplification and typing of the HLA DQα gene. Int J Leg Med 1991;104:145–8.
7. Waye JS, Presley L, Budowle B, Shutler GG, Fourney RM. A simple method for quantifying human genomic DNA in forensic specimen extracts. Bio Techniques 1991;7:852–5.
8. Edwards A, Hammond H, Jin L, Caskey CT, Chakraborty R. Genetic variation at five trimeric and tetrameric repeat loci in four human population groups. Genomics 1992;12:241–53.
9. Chakraborty R, Smouse PE, Neel JV. Population amalgamation and genetic variation: observations on artificially agglomerated tribal populations of Central and South America. Am J Hum Genet 1988; 43:709–25.
10. Nei M, Roychoudhury AK. Sampling variances of heterozygosity and genetic distance. Genetics 1974;76:379–90.
11. Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 1978;89:583–90.
12. Chakraborty R, Fornage M, Guegue R, Boerwinkle E. Population genetics of hypervariable loci: analysis of PCR based VNTR polymorphism within a population. In: Burke T, Dolf G, Jeffreys AJ, Wolff R, editors. DNA fingerprinting: Approaches and applications. Berlin: Birkhauser Verlag, 1991;127–43.
13. Weir BS. Independence of VNTR alleles defined by fixed bins. Genetics 1992;130:873–87.
14. Guo SW, Thompson EA. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 1992;48: 361–72.
15. Karlin S, Cameron EC, Williams PT. Sibling and parent-offspring correlation estimation with variable family size. Proc Natl Acad Sci USA 1981;78:2664–8.
16. Roff DA, Bentzen P. The statistical analysis of mitochondrial DNA polymorphisms: X^2 and the problem of small samples. Molec Biol Evolution 1989;6:539–45.
17. Lewontin RC, Felsenstein J. The robustness of homogeneity tests in 2 X N tables. Biometrics 1965;21:19–33.
18. Budowle B, Koons BW, Moretti TR. Subtyping of the HLA-DQA1 locus and independence testing with PM and STR/VNTR loci. J Forensic Sci (in press).

Additional information and reprint requests:
Linda B. Jankowski, M.S.
New Jersey State Police
South Regional Laboratory
1101 S. White Horse Pike
Hammonton, NJ 08037