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

Evaluation of Independence Assumptions for PCR-Based and Protein-Based Genetic Markers in New Jersey Caucasians*

REFERENCE: Budowle B, Jankowski LB, Corey HW, Swec NT, Freck-Tootell S, Pino JA, Schwartz R, Kelley CA, Tarver ML. Evaluation of independence assumptions for PCR-based and protein-based genetic markers in New Jersey Caucasians. J Forensic Sci 1997;42(2):223–225.

ABSTRACT: Allele frequencies for six PCR-based loci and three protein-based (i.e., enzyme systems) loci were determined in a Caucasian sample population from New Jersey. The loci are LDLR, GYPA, HBGG, D7S8, Gc, HLA-DQA1, PGM_1, ESD, and EAP. All loci meet Hardy-Weinberg expectations. In addition, there is little evidence for association of alleles among the nine loci. The allelic frequency data generally are similar to another Caucasian population database.

KEYWORDS: forensic science, genetic markers, DNA, Caucasian, population genetics, polymarker, phosphoglucomutase 1, esterase D, erythrocyte acid phosphatase, Hardy-Weinberg equilibrium, linkage equilibrium, polymerase chain reaction, LDLR, GYPA, HBGG, D7S8, Gc

When DNA profiles from a known source and an evidentiary sample cannot be excluded as potentially originating from the same source, an estimate (or estimates) of the rarity of the evidentiary profile is provided (1). In addition to typing the evidence for DNA markers, at times, protein-based genetic markers have been analyzed. The most meaningful estimate of the rarity of the combined DNA and protein profiles would be derived by assuming independence and multiplying the individual locus frequencies together. However, one could argue that independence between or among the loci has not been demonstrated formally. There are no published data to date determining whether or not the loci in the polymarker system (LDLR, GYPA, HBGG, D7S8, Gc, and HLA-DQA1) and three common protein-based markers (PGM_1, EAP, and ESD) demonstrate gametic phase equilibrium (or linkage equilibrium) expectations. This paper presents allele frequency data for nine genetic markers in a Caucasian population sample from New Jersey. The data demonstrate that the allele frequencies for the loci meet the expectations of independence and can be useful for providing estimates of the frequency of a multiple locus profile.

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

[2]Principle forensic scientist, Senior forensic scientists and technical director, respectively. New Jersey State Police South Regional Laboratory, Hammonton, New Jersey 08037.

*This is publication number 96-09 of the Laboratory Division of the Federal Bureau of Investigation. Names of commercial manufacturers are provided for identification only, and inclusion does not imply endorsement by the Federal Bureau of Investigation.

Received 2 May 1996; accepted 1 July 1996.

Materials and Methods

Sample preparation: Whole blood samples, obtained in EDTA vacutainer tubes by venipuncture or by fingerprick, were collected from unrelated volunteers from the state of New Jersey. The DNA was extracted by the chelex method (2). The quantity of extracted DNA was estimated using the slot-blot procedure described by Waye, et al. (3) using the Quantiblot Kit (Perkin-Elmer) (4).

Multiplex PCR: The Polymarker loci were typed using the AmpliType® PM PCR Amplification and Typing Kit (Perkin Elmer Corporation, Norwalk, CT). The amplification conditions were those recommended by the manufacturer. Amplification was carried out in a Perkin-Elmer DNA Thermal Cycler 480. Typing PGM, ESD, and EAP was performed as described previously (5–7).

Statistical Analysis: The frequency of each allele for each genetic marker was calculated from the numbers of each genotype in the sample set (i.e., the gene count method). Unbiased estimates of expected heterozygosity were computed as described by Edwards et al. (8). Possible divergence from Hardy-Weinberg expectations (HWE) was tested 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) for two locus associations was used for detecting disequilibrium between the loci.

A R × C contingency table exact test was used to generate a G-statistic (1000 shuffling experiments) (16,17) to test for homogeneity for the allele frequency distributions between the New Jersey Caucasian population sample and other U.S. Caucasian data (18,19). The program was kindly provided by R. Chakraborty (University of Texas School of Biomedical Sciences, Houston, Texas).

Results and Discussion

The distributions of observed allelic frequencies for LDLR, GYPA, HBGG, D7S8, Gc, and HLA-DQA1 and PGM, ESD, and EAP are shown in Tables 1 and 2, respectively. There was no deviation from HWE for the nine loci based on the homozygosity test, likelihood ratio test, and the exact test (Tables 1 and 2). An interclass correlation test analysis demonstrated that there is little detectable evidence for correlation between the alleles at any of the pair-wise comparisons of loci. Out of a total of 36 pair-wise

TABLE 1—*Polymarker allele frequencies in a New Jersey Caucasian sample population of 164 individuals.*

Locus/ Allele	Frequency
LDLR A*	0.412
LDLR B*	0.588
GYPA A†	0.564
GYPA B†	0.436
HBGG A‡	0.436
HBGG B‡	0.561
HBGG C‡	0.003
D7S8 A§	0.628
D7S8 B§	0.372
Gc A‖	0.268
Gc B‖	0.140
Gc C‖	0.591
DQA1 1.1¶	0.152
DQA1 1.2¶	0.207
DQA1 1.3¶	0.070
DQA1 2¶	0.152
DQA1 3¶	0.146
DQA1 4¶	0.271

*LDLR—observed homozygosity = 49.4%; expected homozygosity (unbiased estimate) = 51.4%; HWE—homozygosity test (p = 0.603), likelihood ratio test (p = 0.626), exact test (p = 0.626).

†GYPA—observed homozygosity = 43.3%; expected homozygosity (unbiased estimate) = 50.7%; HWE—homozygosity test (p = 0.059), likelihood ratio test (p = 0.056), exact test (p = 0.056).

‡HBGG—observed homozygosity = 54.3%; expected homozygosity (unbiased estimate) = 50.3%; HWE—homozygosity test (p = 0.313), likelihood ratio test (p = 0.277), exact test (p = 0.277).

§D7S8—observed homozygosity = 47.6%; expected homozygosity (unbiased estimate) = 53.1%; HWE—homozygosity test (p = 0.152), likelihood ratio test (p = 0.132), exact test (p = 0.132).

‖Gc—observed homozygosity = 41.5%; expected homozygosity (unbiased estimate) = 44.0%; HWE—homozygosity test (p = 0.517), likelihood ratio test (p = 0.626), exact test (p = 0.635).

¶DQA1—observed homozygosity = 20.7%; expected homozygosity (unbiased estimate) = 18.7%; HWE—homozygosity test (p = 0.503), likelihood ratio test (p = 0.691), exact test (p = 0.739).

TABLE 2—*ESD, PGM_1, and EAP allele frequencies in a New Jersey Caucasian sample population 155, 152, and 150 individuals, respectively.*

Locus/ Allele	Frequency
ESD 1*	0.897
ESD 2*	0.094
ESD 5*	0.010
PGM 1+ †	0.572
PGM 1− †	0.164
PGM 2+ †	0.217
PGM 2− †	0.046
EAP A‡	0.377
EAP B‡	0.580
EAP C‡	0.043

*ESD—observed homozygosity = 81.9%; expected homozygosity (unbiased estimate) = 81.2%; HWE—homozygosity test (p = 0.826), likelihood ratio test (p = 0.697), exact test (p = 0.697).

†PGM—observed homozygosity = 34.2%; expected homozygosity (unbiased estimate) = 40.2%; HWE—homozygosity test (p = 0.132), likelihood ratio test (p = 0.530), exact test (p = 0.670).

‡EAP—observed homozygosity = 54.0%; expected homozygosity (unbiased estimate) = 47.8%; HWE—homozygosity test (p = 0.131), likelihood ratio test (p = 0.256), exact test (p = 0.318).

TABLE 3—*Test of homogeneity between New Jersey Caucasians and published United States Caucasian data.**

Locus	p Value
LDLR	0.3410 ± 0.0150
GYPA	0.6170 ± 0.0154
HBGG	0.4790 ± 0.0158
D7S8	0.7480 ± 0.0137
Gc	0.5420 ± 0.0158
DQA1	0.0320 ± 0.0056
ESD	0.0750 ± 0.0083
PGM_1	0.0110 ± 0.0033
EAP	0.5950 ± 0.0155

*Data derived from Budowle, et al. (18,19).

comparisons, only one comparison between Gc and ESD departed significantly from the expectation of independence (p = 0.029). This number of departures is no more than would be expected by chance. Therefore, the expectation of independence within a locus and between loci holds for the New Jersey Caucasian sample population.

The New Jersey Caucasian allele frequency data were compared with other United States Caucasian data (18,19) (Table 3). Seven out of the nine loci were statistically similar. Only HLA-DQA1 and PGM_1 were statistically different between the two populations (p = 0.032 and p = 0.011, respectively). Despite these differences, an estimate of a multiple locus frequency would not be expected to be substantially different if our New Jersey Caucasian sample population were used instead of another United States Caucasian population. For example, the PGM_1 and HLA-DQA1 allele frequencies between the two Caucasian population samples were statistically different; but no allele frequencies differed by as much as two-fold.

The vast majority of genetic markers used in forensic analyses meet expectations of independence (e.g., see 18–33). Therefore, it would be reasonable in most situations to assume independence to derive a multiple locus profile frequency estimate for the PCR-based and protein-based loci. However, there have been no published data demonstrating that some of the forensically common PCR-based and protein-based genetic markers meet gametic phase equilibrium expectations. This study demonstrates that for our New Jersey Caucasian sample population the assumption of independence for the loci LDLR, GYPA, HBGG, D7S8, Gc, HLA-DQA1, PGM_1, EAP, and ESD is valid.

References

1. Budowle B, Monson KL, Wooley JR. The reliability of statistical estimates in forensic DNA typing. In: DNA Identification (Billings PR, editor.) Cold Spring Harbor Press, New York, 1992;79–90.
2. Walsh PS, Metzger DA, Higuchi R. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. Biotechniques 1991;10:506–13.
3. Waye JS, Presley L, Budowle B, Shutler GG, Fourney RM. A simple method for quantifying human genomic DNA in forensic specimen extracts. Biotechniques 1991;7:852–5.
4. Walsh PS, Valaro J, Reynolds R. A rapid chemiluminescent method for quantitation of human DNA. Nucleic Acids Res 1992; 20:5061–5.
5. Budowle B, Murch RS, Davidson LC, Gambel AM, Kearney JJ. Subtyping phosphoglucomutase-1 in semen stains and bloodstains: A report on the method. J Forensic Sci 1986;31(4):1341–8.
6. Wraxall BGD, Stolorow MD. The simultaneous separation of the enxymes glyoxalase I, esterase D, and phosphoglucomutase. J Forensic Sci 1986;31:1439–49.

7. Dykes D, Poleskey H, Miller S. Frequency of the EsD 5 allele in three ethnic groups in Minnesota. Hum Genet 1982;62:162–3.
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, Birkhauser Verlag, Berlin, 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 Nat Acad Sci USA 1981;78:2664–8.
16. Roff DA, Bentzen P. The statistical analysis of mitochondrial DNA polymorphisms:χ^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 × N tables. Biometrics 1965;21:19–33.
18. Budowle B, Sundaram S, Wenk R. Population data on the forensic genetic markers: Phosphoglucomutase-1, esterase D, erythrocyte acid phosphatase and glyoxalase I. Forensic Sci Int 1985;28:77–81.
19. Budowle B, Lindsey JA, DeCou JA, Koons BW, Giusti 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.
20. Sajantila A, Budowle B, Strom M, Johnsson V, Lukka M, Peltonen L, et al. PCR amplification of alleles at the D1S80 locus: Comparison of a Finnish and a North American Caucasian population sample, and forensic case-work evaluation. Am J Hum Genet 1992;50(4):816–25.
21. Kloosterman AD, Budowle B, Riley EL. Population data of the HLA DQα locus in Dutch Caucasians. Comparison with seven other population studies. Int J Leg Med 1993;105(4):233–8.
22. Hochmeister MN, Budowle B, Borer UV, Dirnhofer R. Swiss population data on the loci HLA-DQα, LDLR, GYPA, HBGG, D7S8, Gc, and D1S80. Forensic Sci Int 1994;67(3):175–84.
23. Hochmeister MN, Jung MM, Budowle B, Borer UV, Dirnhofer R. Swiss population data on three tetrameric short tandem repeat loci—VWA, HUMTHO1, and F13A1—derived using multiplex PCR and laser fluorescence detection. Int J Leg Med 1994;107:34–6.
24. Lorente JA, Lorente M, Budowle B, Wilson MR, Villanueva E. Analysis of HUMTHO1 allele frequencies in the Spanish population. J Forensic Sci 1994;39(5):1270–4.
25. Lorente JA, Lorente M, Budowle B, Wilson MR, Villanueva E. Analysis of the short tandem repeat (STR) HUMVWA in the Spanish population. Forensic Sci Int 1994;65(3):169–75.
26. Budowle B, Baechtel FS, Smerick JB, Presley KW, Giusti AM, Parsons G, et al. D1S80 population data in African Americans, Caucasians, Southeastern Hispanics, Southwestern Hispanics, and Orientals. J Forensic Sci 1995;40(1):38–44.
27. Budowle B, Giusti AM. Fixed bin frequency distributions for the VNTR locus D5S110 in general United States reference databases. J Forensic Sci 1995;40:236–8.
28. Huang NE, Budowle B. Chinese population data on the PCR-based loci HLA-DQα, LDLR, GYPA, HBGG, D7S8, and Gc. Hum Heredity 1995;45:34–40.
29. Huang NE, Budowle B. Fixed bin population data for the VNTR loci D1S7, D2S44, D4S139, D5S110, and D17S79 in Chinese from Taiwan. J Forensic Sci 1995;40:287–90.
30. Huang NE, Schumm J, Budowle B. Chinese population data on three tetrameric short tandem repeat loci—HUMTHO1, TPOX, and CSF1PO—derived using multiplex PCR and manual typing. Forensic Sci Int 1995;71:131–6.
31. Martin P, Alonso A, Budowle B, Albarran C, Garcia O, Sancho M. Spanish population data on seven tetrameric short tandem repeat loci. Int J Leg Med 1995;108(3):145–9.
32. Monson KL, Moisan JP, Pascal O, McSween M, Aubert D, Giusti A et al. Description and analysis of allele distribution for four VNTR markers in French and French Canadian populations. Hum Heredity 1995;45:135–43.
33. 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;41:47–51.

Additional information and reprint requests:
Bruce Budowle, Ph.D.
Forensic Science Research and Training Center
F.B.I. Academy
Quantico, VA 22135