TECHNICAL NOTE

Christine S. Tomsey,[1] M.S.; Christopher J. Basten,[2] Ph.D.; Bruce Budowle,[3] Ph.D.; Beth Ann Giles,[4] B.S.; Scott Ermlick,[4] M.S.; and Sarah Gotwald,[4] M.S.

Use of Combined Frequencies for RFLP and PCR Based Loci in Determining Match Probability

REFERENCE: Tomsey CS, Basten CJ, Budowle B, Giles BA, Ermlick S, Gotwald S. Use of combined frequencies for RFLP and PCR based loci in determining match probability. J Forensic Sci 1999;44(2):385–388.

ABSTRACT: Statistical analysis was performed on a subset of the Pennsylvania State Police Caucasian, African American and Hispanic database for the purpose of determining Hardy-Weinberg equilibrium and associations across the RFLP loci D1S7, D2S44, D4S139, D5S110, D10S28 and D17S79 and the PCR-based loci HLA-DQA1, LDLR, GYPA, HBGG, D7S8 and Gc. Overall, the statistical results are consistent with a population in equilibrium both within and between loci. The assumption for independence is valid.

KEYWORDS: forensic science, DNA typing, restriction fragment length polymorphism, D1S7, D2S44, D4S139, D5S110, D10S28, D17S79, HLA-DQA1, LDLR, GYPA, Gc, D7S8, HBGG, population genetics, Hardy-Weinberg equilibrium, polymerase chain reaction, linkage equilibrium, independence

Restriction fragment length polymorphism (RFLP) technology in forensic DNA analysis has, for many years, been the analytical procedure of choice to aid in the characterization of physiological fluids. Due to the high degree of polymorphism at each genetic locus, coincidental matches between unrelated persons over several loci are vanishingly rare. Estimates for the frequencies of six loci profiles range from one in the billions to one in the trillions. However, at times, the forensic evidence is very limited in quantity or has been subjected to environmental insults that has rendered the sample degraded, thus precluding the typing of some or all of the genetic loci a laboratory may typically use for DNA RFLP analysis. When this occurs, the laboratory may turn to polymerase chain reaction (PCR)-based systems to further characterize the sample. A combination of the PCR and RFLP loci will further define the source of the evidence. The National Research Council has suggested that the assumption of independence between loci is reasonable for yielding meaningful estimates of the rarity of a profile (1). The assumption of independence between nine genetic markers, the protein markers PGM, EAP and ESD and the PCR-based loci LDLR, GYPA, HBGG, D7S8, Gc and HLA-DQA1 has already been shown by Budowle et al. (2). We demonstrate here, with a much larger set of loci, that the assumption of independence holds between 12 loci, six PCR-based and six VNTR loci. We have combined the data sets of six RFLP loci (D1S7, D2S44, D4S139, D5S110, D10S28, and D17S79) and the PCR based PM/DQA1 loci (LDLR, GYPA, HBGG, D7S8, Gc and HLA-DQA1) and demonstrated that there is little evidence for departures from Hardy-Weinberg equilibrium (HWE) and linkage expectations.

[1] DNA Laboratory manager, Pennsylvania State Police Bureau of Forensics and Criminal Identification, Forensics Division, Greensburg, PA 15601.

[2] Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203.

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

[4] Forensic science supervisor, laboratory manager, and forensic scientist, respectively, Pennsylvania State Police Greensburg Regional Laboratory, Greensburg, PA 15601.

Received 3 Feb. 1998; and in revised form 25 June 1998; accepted 29 June 1998.

Materials and Methods

Sample Preparation

Samples were collected in EDTA vacutainers and deposited on sterilized white cotton cloth, allowed to dry and stored at −80°C until analysis. The samples were obtained from unrelated individuals from Pennsylvania. The sample size that had both RFLP and PCR analysis completed consisted of 99 Caucasians, 100 African Americans and 102 Hispanics.

RFLP Method

RFLP analysis was performed using a slight modification of the FBI method (3). The loci D1S7, D2S44, D4S139, D5S110, D10S28 and D17S79 were typed. Sizing of the alleles was performed using the FBI computer sizing program (4) and binned according to the Lifecodes 23 kb ladder using a fixed bin analysis as outlined by Budowle et al. (5), and Budowle and Guisti (6). Table 1 lists the number of bins for each of the six VNTR loci in each of the racial groups, as well as the number of alleles for each of the PCR loci.

PCR Method

Extraction of the sample was achieved using the chelex method (7) and quantitation of the DNA was achieved using the Quantiblot Kit (Perkin Elmer) (8). Amplification was carried out using the Perkin Elmer 9600 Thermal Cycler following the manufacturer's instructions for the amplification conditions (9). PCR multiplexing was carried out using the Perkin Elmer Amplitype PM/DQA1 PCR

Amplification and Typing Kit (Perkin Elmer Corporation, Norwalk, CT).

Statistical Analysis

Since a subset of the entire Pennsylvania State RFLP database was typed for the PCR-based loci, some of the predefined bins for the RFLP loci contained fewer than five alleles. In order to do the statistical tests, it was necessary to merge adjacent bins as described previously by Budowle et al. (5).

Exact tests were used to determine the independence of alleles within and between loci (Maiste and Weir (10); Zaykin, Zhivotovsky, and Weir (11). Results of the analysis are shown in Tables 2 to 4. The upper triangle portion of each of the Tables 2 to 4 contain the exact test results for the analyses of all the genotypes at a locus. The lower triangle portion of the Tables 2 to 4 presents the results for the heterozygote genotypes only. The bottom two lines of each table describe the HWE test results for all genotypes (Full) and heterozygotes only (Het).

TABLE 1—*Number of loci in database.*

Number	Locus	Chromosome	# Alleles		
			Ca	AA	Hi
1	D2S44	2	21	22	21
2	D17S79	17	11	14	11
3	D10S28	10	25	25	24
4	D1S7	1	26	26	23
5	D4S139	4	16	19	16
6	D5S110	5	24	26	24
7	LDLR	19	2	2	2
8	GYPA	4	2	2	2
9	HBGG	11	3	3	3
10	D7S8	7	2	2	2
11	GC	4	3	3	3
12	DQA1	6	7	7	7

Results

Hardy-Weinberg Disequilibria

The Pennsylvania State Police database contains data on 12 loci in three population groups. This entails 36 independent tests of the HWE. For statistical tests at the 5% significance level, we would expect about 5% of the 36 tests to show significant departures from HWE. There was one significant result over the 36 tests with all genotypes (see Tables 2 to 4) at the D5S110 locus in the African American database ($p = 0.0216$). This result is about what would be expected if the three populations were in HWE at the 12 loci.

TABLE 2—*Exact tests for Caucasian database.*

Locus	1	2	3	4	5	6	7	8	9	10	11	12
1	. . .	0.0942	0.4425	0.1374	0.1872	0.1050	0.3505	0.4261	0.7340	0.3856	0.3486	0.7568
2	0.6802	. . .	0.4515	0.4025	0.2346	0.2092	0.4073	0.2516	0.1528	0.0290	0.1857	0.6707
3	0.6595	0.6485	. . .	0.2953	0.2658	0.2270	0.4428	0.3204	0.6255	0.2760	0.0517	0.8211
4	0.8206	0.8088	0.7998	. . .	0.1358	0.0944	0.3276	0.1184	0.3514	0.1785	0.0636	0.5450
5	0.8922	0.8457	0.8563	0.9445	. . .	0.0983	0.6206	0.2949	0.8559	0.3572	0.1218	0.4070
6	0.8678	0.8582	0.8494	0.9330	0.9604	. . .	0.3338	0.1227	0.6563	0.1310	0.1377	0.0996
7	0.2797	0.2536	0.5275	0.2402	0.1372	0.1685	. . .	0.2014	0.5980	0.2713	0.0860	0.2376
8	0.4144	0.4354	0.7504	0.4439	0.2937	0.3105	0.4028	. . .	0.2690	0.3349	0.4424	0.1811
9	0.7334	0.6632	0.7139	0.9110	0.5072	0.5882	0.5861	0.6990	. . .	0.2766	0.3693	0.9524
10	0.1277	0.1534	0.2073	0.0870	0.0563	0.0702	0.1543	0.2225	0.4644	. . .	0.2775	0.4379
11	0.3565	0.0897	0.9512	0.7127	0.0840	0.3211	0.0637	0.1194	0.2637	0.0262	. . .	0.1558
12	0.3059	0.3522	0.2361	0.4156	0.4863	0.4741	0.5751	0.7450	0.8912	0.4093	0.2781	. . .
Full	0.7502	0.0764	0.6735	0.0733	0.6548	0.1367	0.6796	1.0000	0.5442	0.3120	0.1039	0.4027
Hets	0.7381	0.6777	0.5891	0.8634	0.8858	0.8873	0.3610	0.5765	0.8229	0.1979	0.0637	0.2085

TABLE 3—*Exact tests for African American database.*

Locus	1	2	3	4	5	6	7	8	9	10	11	12
1	. . .	0.0881	0.5360	0.0511	0.0162	0.0039	0.9860	0.9249	0.3244	0.4680	0.8151	0.0930
2	0.9805	. . .	0.2127	0.0379	0.0096	0.0036	0.7632	0.7871	0.4849	0.8684	0.1936	0.0568
3	0.6158	0.6884	. . .	0.7202	0.3244	0.2428	0.6225	0.8214	0.3129	0.3253	0.4976	0.7485
4	0.9619	0.9602	0.4204	. . .	0.0320	0.0065	0.8767	0.8924	0.1903	0.9092	0.2060	0.2670
5	0.9900	0.9918	0.7989	0.9856	. . .	0.0014	0.8311	0.2199	0.2733	0.3966	0.3482	0.1797
6	0.9979	0.9971	0.8473	0.9961	0.9997	. . .	0.1643	0.3112	0.0105	0.1393	0.0171	0.0401
7	0.4797	0.3585	0.9905	0.6903	0.3303	0.2667	. . .	0.1752	0.9688	0.3553	0.4709	0.9797
8	0.9995	0.5181	0.1122	0.5930	0.6729	0.9576	0.9288	. . .	0.4432	0.9665	0.2960	0.5211
9	0.8799	0.7373	0.4362	0.7864	0.9366	0.9554	0.6465	0.6815	. . .	0.8416	0.6246	0.0087
10	0.4736	0.3366	0.6138	0.6813	0.3333	0.3350	0.9200	0.8953	0.5719	. . .	0.9350	0.8168
11	0.1386	0.1358	0.7966	0.2161	0.0652	0.0358	0.7303	0.7431	0.3330	0.6569	. . .	0.6558
12	0.8498	0.8995	0.2579	0.7548	0.9234	0.9625	0.7454	0.7244	0.8967	0.6684	0.3784	. . .
Full	0.9711	0.3577	0.0709	0.7938	0.4295	0.0216	0.2963	0.4207	0.7723	0.6729	0.9316	0.9261
Hets	0.9342	0.9304	0.0843	0.8581	0.9780	0.9962	0.9650	0.8784	0.3625	0.8069	0.3979	0.5906

TABLE 4—*Exact tests for Hispanic database.*

Locus	1	2	3	4	5	6	7	8	9	10	11	12
1	. . .	0.0547	0.0501	0.0493	0.4186	0.0795	0.0413	0.1256	0.2951	0.1374	0.2622	0.1272
2	0.9791	. . .	0.1438	0.1322	0.3164	0.2266	0.3144	0.5243	0.5873	0.4221	0.3663	0.3196
3	0.9410	0.8909	. . .	0.2900	0.7889	0.5071	0.1439	0.6137	0.3906	0.0950	0.3358	0.3868
4	0.9456	0.8877	0.7124	. . .	0.7980	0.4832	0.4085	0.7121	0.3558	0.1859	0.5603	0.5206
5	0.7198	0.6376	0.3285	0.3358	. . .	0.8013	0.8077	0.7441	0.7237	0.3556	0.7824	0.7447
6	0.9198	0.8345	0.6266	0.6223	0.2523	. . .	0.5869	0.6229	0.1287	0.1921	0.5247	0.6054
7	0.1212	0.1903	0.3945	0.3841	0.6279	0.4156	. . .	0.8823	0.9148	0.2235	0.9321	0.3599
8	0.1904	0.2479	0.3969	0.4187	0.6926	0.4747	0.4777	. . .	0.5351	0.3620	0.7273	0.1768
9	0.5189	0.1134	0.8956	0.8951	0.9896	0.8106	0.2945	0.3225	. . .	0.3800	0.5335	0.0472
10	0.0108	0.0266	0.0602	0.0580	0.1829	0.0871	0.1111	0.1422	0.0471	. . .	0.0566	0.0102
11	0.9045	0.3858	0.6509	0.6725	0.4182	0.5578	0.3861	0.4237	0.2721	0.0867	. . .	0.1858
12	0.8299	0.7576	0.4541	0.5108	0.2398	0.4509	0.4637	0.5519	0.3843	0.1340	0.8522	. . .
Full	0.1856	0.8345	0.4610	0.7047	0.9826	0.6034	0.8472	1.0000	0.8918	0.1384	0.9601	0.1384
Hets	0.9679	0.8966	0.5684	0.6836	0.1922	0.5327	0.4846	0.5601	0.2664	0.0708	0.4097	0.4515

If we test the heterozygotes only, then there is no evidence for departures from HWE. Only the D5S110 locus in the African American population demonstrates a possible departure from HWE. However, when null alleles are taken into account (as demonstrated by Chakraborty et al. (12–14)) no departures were observed.

Associations Across Loci

Exact tests were performed between pairs of loci, considering all the genotypes at a locus. Over the three databases, 17 of the 196 tests have an exact probability less than 5%. We would expect about 10. Seven of those are due to a single locus D5S110 in the African American database, which also displays Hardy-Weinberg disequilibria at the 5% level. If the tests are restricted to heterozygotes, only six of the 196 tests are significant at the 5% level. This suggests that most of the disequilibria are due to the presence of null alleles at the D5S110 locus in the African American database. In addition, Karlin Interclass Correlation tests (15,16) were performed between all loci. When pairwise comparisons are made, only three of 66 comparisons in the Caucasians (D2S44-Gc $p = 0.017$; D10S28-HBGG $p = 0.023$; LDLR-GYPA $p = 0.039$), three of 66 in the African Americans (D1S7-HBGG $p = 0.032$; LDLR-GYPA $p = 0.011$; GYPA-HBGG $p = 0.028$), and two of 66 in the Hispanics (D4S139-LDLR $p = 0.047$; D4S139-Gc $p = 0.048$) demonstrate departures from expectations. This is what would be expected to occur by chance. These results argue strongly against any systematic associations across loci.

Discussion

Three loci, D4S139, GYPA and Gc, reside on chromosome 4. If there were any disequilibria between loci, then these loci would be the best candidates to display it. Only in the Karlin correlation test was there a borderline example of departure between Gc and D4S139 ($p = 0.048$) in the Hispanic population. Thus the overall data support that there is little association between loci in the three databases.

Overall, the statistical results are consistent with a population in equilibrium both within and between loci. The assumption for independence for the Pennsylvania State Police Caucasian, African American and Hispanic databases for the genetic loci D1S7, D2S44, D4S139, D5S1190, D10S28, D17S79, LDLR, GYPA, HBGG, D7S8, Gc and HLA-DQA1 is valid.

Acknowledgments

The authors thank Walter Lorenz and Thomas Meyers from the Allegheny County Crime Laboratory for their contribution of typing a portion of the PCR samples in the Caucasian and African American databases.

References

1. National Research Council. The Evaluation of Forensic DNA Evidence, Washington, DC: National Academy Press. 1996.
2. Budowle B, Jankowski L, 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;42:223–5.
3. Budowle B, Baechtel FS. Modifications to improve the effectiveness of RFLP typing. Appl Theor Electrophoresis 1990;1:181–7.
4. Budowle B, Monson KL. A system for semi-automated analysis of DNA autoradiograms. Proceedings of the International Symposium on Forensic Aspects of DNA Analysis; 19–23 June; FBI, FSRTC Quantico, VA, 1989;127–32.
5. Budowle B, Guisti AM, Waye JS, Baechtel FS, Forney RM, Adams DE, et al. Fixed bin analysis for statistical evaluation of continuous distribution of allele data from VNTR loci for use in forensic comparisons. Am J Hum Genet 1991;48:841–55.
6. Budowle B, Guisti AM. Fixed bin frequency distribution for the VNTR locus D5S110 in general United States reference databases. J Forensic Sci 1995;40:236–8.
7. Walsh PS, Metzger DA, Higuchi R. Chelex 100 as a medium for simple extractions of DNA for PCR based typing from forensic material. Biotechniques 1991;10:506–13.
8. Walsh PS, Valero J, Reynolds R. A rapid chemiluminescent method for quantitation of human DNA. Nucleic Acids Res 1992;20: 5061–5.
9. Amplitype User Guide, Version 2, Perkin Elmer Corporation 1993.
10. Maiste PJ, Weir BS. A comparison of tests for independence in the FBI RFLP database. In: Weir BS, editor. Human identification: the use of DNA markers. Boston, MA: Kluwer Academic Publishers, 1995;125–38.
11. Zaykin DL, Zhivotovsky, Weir BS. Exact tests for associations between alleles at arbitrary numbers of loci. In: Weir BS, editor. Human identification: the use of DNA markers. Boston, MA: Kluwer Academic Publishers, 1995;169–78.
12. Chakraborty R, Jin L. Heterozygote deficiency, population sub-

structure and their implication in DNA fingerprinting. Hum Genet 1992;88:267–72.
13. Chakraborty R, DE Andrade RM, Daiger SP, Budowle B. Apparent heterozygote deficiencies observed in DNA typing data and their implication in forensic applications. Ann Hum Genet 1992;56: 45–57.
14. Chakraborty R, Zhong Y, Jin L, Budowle B. Nondetectability of restriction fragment sizes within and between loci in RFLP typing of DNA. Am J Hum Genet 1994;55:391–401.
15. Karlin, S, Ladunga I, Blaisdell BE. Heterogeneity of genomes: measures and values. Proc Nat Acad Sci USA 1991;12837–44.
16. Karlin S, Cameron EC, Williams PT. Siblings and parent-offspring correlation estimates with variable family size. Proc Nat Acad Sci USA 1981;78:2664–8.

Additional information and reprint requests:
Christine S. Tomsey
Pennsylvania State Police
Bureau of Forensics and Criminal Identification
Forensic Division
Greensburg, PA 15601