Baldev R. Sharma,[1] Ph.D.; Monica Thompson,[1] M.S.; James R. Bolding,[1] M.S.; Yixi Zhong,[2] M.D.; Li Jin,[3] Ph.D.; and Ranajit Chakraborty,[2] Ph.D.

A Comparative Study of Genetic Variation at Five VNTR Loci in Three Ethnic Groups of Houston, Texas

REFERENCE: Sharma, B. R., Thompson, M., Bolding, J. R., Zhong, Y., Jin, L., and Chakraborty, R., "**A Comparative Study of Genetic Variation at Five VNTR Loci in Three Ethnic Groups of Houston, Texas,**" *Journal of Forensic Sciences*, JFSCA, Vol. 40, No. 6, November 1995, pp. 933–942.

ABSTRACT: Following the technique of Southern blot restriction fragment length polymorphisms (RFLP) analysis, we generated a database of DNA profiles at five Variable Number of Tandem Repeats loci (D1S7, D2S44, D4S139, D10S28, and D17S79) for 669 individuals of three major ethnic populations (Caucasians, Blacks, and Hispanics) of Houston, Texas. Analysis of fragment sizes at these loci within each sample, as well as their fixed-bin analyses, reveal that the assumptions of independence of allelic occurrences within and between loci are valid for this database. Fixed-bin allele frequency tables, therefore, are the best descriptors of this database for conservative forensic calculations. Finally, we demonstrate that this regional database from Houston, Texas, does not yield any meaningfully different forensic inference than the one obtained from the National database of the respective ethnic groups.

KEYWORDS: forensic science, DNA, population studies, genetic variation, RFLP data base, Hardy Weinberg expectation, linkage equilibrium

It is now well established that DNA typing provides a powerful tool for criminal investigations, as well as for civil litigations involving adjudication of relationships between individuals [*1,2*]. While the DNA technology has improved further than the Southern blot method [*3*] of restriction fragment length polymorphisms (RFLP) analysis of extracted DNA materials in forensic cases [*4–9*], most forensic applications of DNA typing in the USA still involve RFLP typing for a class of polymorphic loci, called the variable number of tandem repeats (VNTR) loci, because they exhibit large number of alleles, and consequently, high degree of inter-individual variability. These characteristics make the VNTR loci highly efficient in exonerating falsely accused individuals. The significance of a DNA match in legal cases is still a subject of debate [*10,11*]. Even though various statistical analyses have shown that a national database, such as the one generated by the Federal Bureau of Investigation [*2,12*], adequately resolves the legal implications of a DNA match found in forensic case-work [*13–17*], the existing concerns may be better addressed by developing regional databases of VNTR loci that are used by the US forensic laboratories.

The purpose of this research is to provide such a documentation. Specifically, we present the main features of polymorphisms at five VNTR loci (D1S7, D2S44, D4S139, D10S28, and D17S79) in three major ethnic populations of Houston, Texas. Analysis of actual fragment sizes at these loci within each sample, as well as their fixed-bin analyses [*2*], reveal that the assumptions of independence of allelic occurrences within and between loci are valid for this database. Fixed-bin allele frequency tables, therefore, are the best descriptors of this database for conservative forensic calculations. In addition, we show that the use of the multiplication rules (both within and between loci) is more strongly validated once the technical limitation of nondetectablity of aberrantly small sized alleles is accounted for in the interpretation of these databases. Finally, we demonstrate that the regional database from Houston, Texas, generated by following the protocol of Budowle and Baechtel [*18*] without any modification, does not yield any meaningfully different forensic inference than the one obtained from the national database of the FBI laboratory [*2,12*].

Materials and Methods

Populations Sampled and Laboratory Protocols

Whole blood samples were obtained from local health centers, blood banks, cadet volunteers at Houston Police Academy and volunteers from Houston Crime Laboratory personnel. Donors of both genders from three ethnic groups; Caucasians (n = 193), Blacks (n = 204), and Hispanics (n = 272) were included in the study. In addition, a small sample of Asians (n = 36) was also collected for analysis. Ethnic classifications of individuals were made by their self-recognized ethnic affiliation. All blood samples were collected during the time period of August 1990 through December 1992, and in order to maintain complete anonymity, all individual identification characteristics, except the ethnic classification records, were deleted from the datafiles. Although laboratory and statistical analyses of the Asian sample did not indicate anything contrary to the results observed for the other three ethnic groups, we have excluded this group from the discussion of statistical analysis, because of the insufficient sample size.

The DNA was extracted, purified, restricted with enzyme *Hae*III and analyzed for fragment size determinations (after electrophoresis and hybridization with the locus-specific probes) according to the protocol described by Budowle and Baechtel [*18*]. Approximately one microgram of DNA from each sample was loaded to the agarose gels during electrophoresis. The size standard markers

Received for publication 12 Sept. 1994; revised manuscript received 17 April 1995; accepted for publication 1 May 1995.

[1]Houston Police Department Crime Laboratory, Houston, TX.

[2]Center for Demographic and Population Genetics, University of Texas Health Science Center at Houston, Houston, TX.

[3]Genetics Department, Stanford University Medical School, Stanford, CA.

ranging in size from 526 to 22,621 base pairs were purchased from Gibco—BRL, Gaithersburg, Maryland. The probes were obtained from the following sources: Gibco—BRL (pH30 for the locus D4S139), Promega Corporation, Madison, Wisconsin (YNH24 for the D2S44 locus, and TBQ7 for the D10S28 locus), Cellmark Diagnostics, Germantown, Maryland (MS1 for the D1S7 locus), and Lifecodes Corporation, Stamford, Connecticut (V1 for the D17S79 locus). Fragment sizes, after electrophoresis and probe hybridizations, were measured with an IBM computer set up including a data translation board, a video camera, and software developed by the FBI [*19*].

Statistical Methods

Forensic applications of VNTR polymorphisms generally use binned classification of allele sizes, and hence, it is desirable to test the assumptions at binned levels of allelic definition [*10,13–15*]. We have, therefore, summarized the data for most of the analysis with fixed-bin classification of fragment sizes, following the fixed-bin boundaries as listed in [*2*]. Allele frequencies in the 31 fixed-bins were computed by the gene counting method [*20*], assuming that each single-band profile (for every locus) was truly homozygous. This procedure summarizes the locus-specific fragment size profiles into a multinomial distribution (analogous to genotype distribution) consisting of 31 possible "homozygous" (both fragments within the same bin, or single-banded patterns), and 465 (= 31 × 30/2) possible "heterozygous" (two fragments belonging to two different fixed-bins) categories.

Allelic independence, within as well as between loci, was tested at bin-level as well as with the actual, measured, fragment sizes. When the profiles were categorized by bins, three different test procedures were employed to test allelic independence within loci: a test based on total heterozygosity (Chi-square analysis), the likelihood ratio test [*21*], and an exact test [*22*]. Allelic independence within loci with the actual, measured, fragment sizes was tested by the intraclass correlation method [*13,15*]. Significance of the test statistics, for both the approaches, were judged by shuffling the observed fragment sizes across the individuals and the empirical levels of significance for each test were determined from 2000 replications of such shuffling (see [*15*] for details of this algorithm).

Allelic independence between each pair of loci was tested using the inter-class correlations of measured fragment sizes [*13,15*], for which, again, the empirical levels of significance were judged by shuffling alleles of both loci [*15*]. In addition, a large sample test based on the variance of the number of heterozygous loci (with binned classification of profiles) was employed to examine whether all five loci together conform to the independence hypothesis [*23*].

As mentioned, in all of these analyses we treated the single-band patterns as two copies of fragments of equal sizes. Empirical data now exists suggesting that for allele sizes detected through the Southern blot RFLP analysis, this may not be entirely correct, since the alleles that produce aberrantly small size fragments may remain undetected in such analysis. The occurrence of such alleles have been experimentally proven in many forensic databases [*1,2,24,25*]. Chakraborty et al. [*14*] have shown earlier that such possibilities may affect the interpretation of all of the above test results. Therefore, we used Gart and Nam's [*26*] statistical test, modified for the RFLP data as described by Chakraborty et al. [*25*], to examine the effect of "nondetectable" alleles on the allelic associations within and between loci.

Finally, in order to examine the forensic implication of the differences between the present database and the National database generated by the FBI [*2,12*], we computed the predicted multilocus DNA profile frequencies for every individual included in this database, computed from the binned allele frequencies of the two (that is, HPD and FBI) databases.

Results

Fixed-Bin Allele Frequencies

Tables 1 through 5 present the fixed-bin allele frequencies (absolute allele counts as well as the relative frequencies, expressed as percent of the total) at each locus in each ethnic group. While the allele frequency distributions among the ethnic groups are different, and they reach statistical significance ($P < 0.001$) for each locus, the spread of the allele sizes for each locus across the ethnic groups are almost identical. This is consistent with the observations made in the Worldwide survey of VNTR polymorphisms [*27*]. The total number of allele counts (2n, n being the number of individuals sampled) for different loci for the same population (shown in the last row of Tables 1 through 5) vary because fragment size data are not available for every individual for each locus. While these allele frequencies are shown for each of the 31 fixed bins in each case, we recommend that for any forensic calculations bins with fewer than 5 allele counts should be merged (re-binning, [*12*]) to avoid any undue emphasis on the small frequency of a specific DNA profile.

TABLE 1—*Binned allele frequencies in three population groups from Houston, Texas at the D1S7 locus.*

Bin	Bin Boundary (bp)	Caucasians	Hispanics	Blacks
1	1– 639	0 (0.00)	0 (0.00)	0 (0.00)
2	640– 772	1 (0.30)	0 (0.00)	0 (0.00)
3	773– 871	0 (0.00)	1 (0.21)	0 (0.00)
4	872– 963	1 (0.30)	0 (0.00)	2 (0.55)
5	964– 1077	0 (0.00)	0 (0.00)	3 (0.83)
6	1078– 1196	0 (0.00)	3 (0.63)	1 (0.28)
7	1197– 1352	3 (0.89)	1 (0.21)	2 (0.55)
8	1353– 1507	2 (0.60)	7 (1.46)	4 (1.11)
9	1508– 1637	2 (0.60)	1 (0.21)	2 (0.55)
10	1638– 1788	2 (0.60)	7 (1.46)	3 (0.83)
11	1789– 1924	5 (1.49)	5 (1.04)	3 (0.83)
12	1925– 2088	2 (0.60)	9 (1.88)	3 (0.83)
13	2089– 2351	6 (1.79)	14 (2.92)	7 (1.93)
14	2352– 2522	8 (2.38)	8 (1.67)	6 (1.66)
15	2523– 2692	7 (2.08)	15 (3.13)	15 (4.14)
16	2693– 2862	7 (2.08)	10 (2.08)	13 (3.59)
17	2863– 3033	11 (3.27)	18 (3.75)	8 (2.21)
18	3034– 3329	19 (5.66)	27 (5.63)	24 (6.63)
19	3330– 3674	24 (7.14)	50 (10.42)	22 (6.08)
20	3675– 3979	18 (5.36)	29 (6.04)	19 (5.25)
21	3980– 4323	26 (7.74)	37 (7.71)	14 (3.87)
22	4324– 4821	18 (5.36)	40 (8.33)	23 (6.35)
23	4822– 5219	24 (7.14)	21 (4.38)	25 (6.91)
24	5220– 5685	19 (5.66)	24 (5.00)	27 (7.46)
25	5686– 6368	24 (7.14)	29 (6.04)	16 (4.42)
26	6369– 7241	26 (7.74)	34 (7.08)	19 (5.25)
27	7242– 8452	15 (4.46)	35 (7.29)	29 (8.01)
28	8453–10093	22 (6.55)	13 (2.71)	18 (4.97)
29	10094–11368	13 (3.87)	14 (2.92)	14 (3.87)
30	11369–12829	9 (2.68)	14 (2.92)	16 (4.42)
31	12830–25000	22 (6.55)	14 (2.92)	24 (6.63)
Total		336 (100.00)	480 (100.00)	362 (100.00)

TABLE 2—*Binned allele frequencies in three population groups from Houston, Texas at the D2S44 locus.*

Bin	Bin Boundary (bp)	Caucasians	Hispanics	Blacks
1	1– 639	0 (0.00)	1 (0.19)	0 (0.00)
2	640– 772	2 (0.65)	2 (0.39)	13 (3.67)
3	773– 871	0 (0.00)	7 (1.37)	5 (1.41)
4	872– 963	4 (1.31)	20 (3.91)	3 (0.85)
5	964– 1077	3 (0.98)	6 (1.17)	5 (1.41)
6	1078– 1196	10 (3.27)	8 (1.56)	28 (7.91)
7	1197– 1352	15 (4.90)	48 (9.38)	29 (8.19)
8	1353– 1507	13 (4.25)	75 (14.65)	42 (11.86)
9	1508– 1637	32 (10.46)	59 (11.52)	26 (7.35)
10	1638– 1788	25 (8.17)	34 (6.64)	40 (11.30)
11	1789– 1924	23 (7.52)	38 (7.42)	23 (6.50)
12	1925– 2088	21 (6.86)	22 (4.30)	18 (5.09)
13	2089– 2351	28 (9.15)	37 (7.23)	31 (8.76)
14	2352– 2522	7 (2.29)	16 (3.13)	8 (2.26)
15	2523– 2692	17 (5.56)	21 (4.10)	9 (2.54)
16	2693– 2862	15 (4.90)	28 (5.47)	11 (3.11)
17	2863– 3033	39 (12.75)	26 (5.08)	5 (1.41)
18	3034– 3329	20 (6.54)	28 (5.47)	10 (2.83)
19	3330– 3674	18 (5.88)	23 (4.49)	12 (3.39)
20	3675– 3979	5 (1.63)	4 (0.78)	11 (3.11)
21	3980– 4323	4 (1.31)	3 (0.59)	7 (1.98)
22	4324– 4821	0 (0.00)	3 (0.59)	7 (1.98)
23	4822– 5219	0 (0.00)	0 (0.00)	1 (0.28)
24	5220– 5685	0 (0.00)	1 (0.19)	7 (1.98)
25	5686– 6368	3 (0.98)	1 (0.19)	2 (0.57)
26	6369– 7241	1 (0.33)	1 (0.19)	1 (0.28)
27	7242– 8452	1 (0.33)	0 (0.00)	0 (0.00)
28	8453–10093	0 (0.00)	0 (0.00)	0 (0.00)
29	10094–11368	0 (0.00)	0 (0.00)	0 (0.00)
30	11369–12829	0 (0.00)	0 (0.00)	0 (0.00)
31	12830–25000	0 (0.00)	0 (0.00)	0 (0.00)
Total		306 (100.00)	512 (100.00)	354 (100.00)

TABLE 3—*Binned allele frequencies in three population groups from Houston, Texas at the D4S139 locus.*

Bin	Bin Boundary (bp)	Caucasians	Hispanics	Blacks
1	1– 639	0 (0.00)	0 (0.00)	0 (0.00)
2	640– 772	0 (0.00)	0 (0.00)	0 (0.00)
3	773– 871	0 (0.00)	0 (0.00)	0 (0.00)
4	872– 963	0 (0.00)	0 (0.00)	0 (0.00)
5	964– 1077	0 (0.00)	0 (0.00)	0 (0.00)
6	1078– 1196	0 (0.00)	0 (0.00)	0 (0.00)
7	1197– 1352	0 (0.00)	0 (0.00)	0 (0.00)
8	1353– 1507	0 (0.00)	1 (0.19)	0 (0.00)
9	1508– 1637	0 (0.00)	0 (0.00)	0 (0.00)
10	1638– 1788	0 (0.00)	0 (0.00)	0 (0.00)
11	1789– 1924	0 (0.00)	0 (0.00)	0 (0.00)
12	1925– 2088	0 (0.00)	1 (0.19)	0 (0.00)
13	2089– 2351	0 (0.00)	1 (0.19)	17 (4.34)
14	2352– 2522	1 (0.29)	0 (0.00)	6 (1.53)
15	2523– 2692	3 (0.88)	1 (0.19)	3 (0.77)
16	2693– 2862	3 (0.88)	2 (0.37)	7 (1.79)
17	2863– 3033	3 (0.88)	2 (0.37)	5 (1.28)
18	3034– 3329	4 (1.17)	3 (0.56)	12 (3.06)
19	3330– 3674	9 (2.63)	4 (0.75)	19 (4.85)
20	3675– 3979	4 (1.17)	17 (3.17)	22 (5.61)
21	3980– 4323	6 (1.75)	13 (2.43)	19 (4.85)
22	4324– 4821	28 (8.19)	38 (7.09)	39 (9.95)
23	4822– 5219	23 (6.73)	23 (4.29)	20 (5.10)
24	5220– 5685	27 (7.90)	44 (8.21)	39 (9.95)
25	5686– 6368	48 (14.04)	89 (16.60)	34 (8.67)
26	6369– 7241	52 (15.21)	78 (14.55)	40 (10.20)
27	7242– 8452	42 (12.28)	106 (19.78)	42 (10.71)
28	8453–10093	30 (8.77)	53 (9.89)	36 (9.18)
29	10094–11368	15 (4.39)	24 (4.48)	12 (3.06)
30	11369–12829	11 (3.22)	13 (2.43)	9 (2.30)
31	12830–25000	33 (9.65)	23 (4.29)	11 (2.81)
Total		342 (100.00)	536 (100.00)	392 (100.00)

Independence of Fixed-bin Alleles Within Loci

Table 6 presents the summary of the three tests of allelic independence within loci, where each single-band profile (for each locus) was treated as a homozygote. These results alone might, erroneously, suggest an apparent significant departure from independence in this database at several loci, from each of the three tests performed. For example, the chi-square test, based on deficiency of total heterozygosity, as well as the exact test [22], both, show significant departure for the Black sample at the D2S44 locus. At the D1S7 locus, samples from all ethnic groups exhibit significant deficiency of heterozygosity, and in addition the likelihood ratio test and the exact test show significant departure from independence in the Caucasian sample at this locus. The sample from the Blacks shows significant departure from allelic independence at the D17S79 locus by all the three test procedures. In addition the chi-square test fails the independence test at this locus in Caucasians and Hispanics. The sample from the Blacks shows significant departure from the independence assumption at the D4S139 locus by the chi-square and exact test procedures, and finally, this sample also shows significant deficiency of heterozygosity (by chi-square test) at the D10S28 locus.

One would also reach the same conclusion, when Geisser and Johnson's quantile test [28] is performed on the same data base (data not presented). We contend that all of these results are the artifacts of the assumption that each single band profile is a true homozygote. This is shown in the re-analysis of the same data by allowing "nondetectability" of alleles. The summary results of this analysis is shown in Table 7, where we compute Gart and Nam's [26] score-statistic, T (equation 6 of [25]). When T is significantly larger than one, we might conclude that nondetectable alleles are present in the database [25], and hence, the assumption that each single-band profile is truly homozygous is fallacious. Estimates of the nondetectable allele frequency (r) and its standard error (computed using equations 11 and 12 of [25]; see also [26]) are presented in Table 7. The score-statistic T is significantly larger than one, wherever significant departures from the independence assumption were noted, by any of the three tests shown in Table 6.

Adjusting for the presence of nondetectable alleles, each of the three test procedures were repeated for the entire database (using the algorithm described in [25]), the resulting empirical significance levels are shown in Table 7. Once the adjustment for the presence of nondetectable alleles is made, we do not observe any significant departure from the allelic independence assumption, for any of the 15 locus-population combinations. The required null allele frequency (r) is also not larger than 5% for any locus-population combination, which is consistent with the empirical frequencies of "nondetectable" alleles seen in RFLP databases [17,24,25].

Intra- and Inter-class Correlations of Fragment Sizes

Test results for linear dependence of fragment sizes within and between loci, through the intraclass and interclass correlation of fragment sizes across the sampled individuals for each database, are shown in Table 8. The intraclass and interclass correlations of fragment sizes may be computed either by the analysis of variance approach [13], or by using the nonparametric method [29], but

TABLE 4—*Binned allele frequencies in three population groups from Houston, Texas at the D10S28 locus.*

Bin	Bin Boundary (bp)	Caucasians	Hispanics	Blacks
1	1– 639	1 (0.30)	0 (0.00)	0 (0.00)
2	640– 772	2 (0.60)	1 (0.20)	1 (0.27)
3	773– 871	1 (0.30)	1 (0.20)	1 (0.27)
4	872– 963	3 (0.89)	8 (1.59)	1 (0.27)
5	964– 1077	18 (5.36)	36 (7.17)	12 (3.24)
6	1078– 1196	12 (3.57)	40 (7.97)	17 (4.60)
7	1197– 1352	11 (3.27)	12 (2.39)	13 (3.51)
8	1353– 1507	30 (8.93)	44 (8.77)	26 (7.03)
9	1508– 1637	29 (8.63)	52 (10.36)	25 (6.76)
10	1638– 1788	21 (6.25)	52 (10.36)	21 (5.68)
11	1789– 1924	21 (6.25)	29 (5.78)	27 (7.30)
12	1925– 2088	27 (8.04)	25 (4.98)	23 (6.22)
13	2089– 2351	21 (6.25)	45 (8.96)	27 (7.30)
14	2352– 2522	9 (2.68)	12 (2.39)	15 (4.05)
15	2523– 2692	5 (1.49)	7 (1.39)	11 (2.97)
16	2693– 2862	20 (5.95)	18 (3.59)	17 (4.60)
17	2863– 3033	11 (3.27)	11 (2.19)	12 (3.24)
18	3034– 3329	11 (3.27)	11 (2.19)	14 (3.78)
19	3330– 3674	17 (5.06)	17 (3.39)	18 (4.87)
20	3675– 3979	16 (4.76)	18 (3.59)	17 (4.60)
21	3980– 4323	10 (2.98)	26 (5.18)	8 (2.16)
22	4324– 4821	27 (8.04)	29 (5.78)	19 (5.14)
23	4822– 5219	1 (0.30)	1 (0.20)	9 (2.43)
24	5220– 5685	4 (1.19)	3 (0.60)	6 (1.62)
25	5686– 6368	7 (2.08)	2 (0.40)	0 (0.00)
26	6369– 7241	1 (0.30)	2 (0.40)	10 (2.70)
27	7242– 8452	0 (0.00)	0 (0.00)	8 (2.16)
28	8453–10093	0 (0.00)	0 (0.00)	5 (1.35)
29	10094–11368	0 (0.00)	0 (0.00)	6 (1.62)
30	11369–12829	0 (0.00)	0 (0.00)	1 (0.27)
31	12830–25000	0 (0.00)	0 (0.00)	0 (0.00)
Total		336 (100.00)	502 (100.00)	370 (100.00)

TABLE 5—*Binned allele frequencies in three population groups from Houston, Texas at the D17S79 locus.*

Bin	Bin Boundary (bp)	Caucasians	Hispanics	Blacks
1	1– 639	2 (0.78)	1 (0.29)	1 (0.38)
2	640– 772	0 (0.00)	0 (0.00)	0 (0.00)
3	773– 871	1 (0.39)	1 (0.29)	0 (0.00)
4	872– 963	0 (0.00)	1 (0.29)	0 (0.00)
5	964– 1077	0 (0.00)	3 (0.86)	8 (3.05)
6	1078– 1196	8 (3.13)	8 (2.29)	10 (3.82)
7	1197– 1352	59 (23.05)	84 (24.00)	67 (25.57)
8	1353– 1507	60 (23.44)	54 (15.43)	57 (21.76)
9	1508– 1637	66 (25.78)	62 (17.71)	28 (10.69)
10	1638– 1788	38 (14.84)	40 (11.43)	26 (9.92)
11	1789– 1924	11 (4.30)	50 (14.29)	15 (5.73)
12	1925– 2088	8 (3.13)	41 (11.71)	25 (9.54)
13	2089– 2351	2 (0.78)	5 (1.43)	13 (4.96)
14	2352– 2522	1 (0.39)	0 (0.00)	1 (0.38)
15	2523– 2692	0 (0.00)	0 (0.00)	3 (1.15)
16	2693– 2862	0 (0.00)	0 (0.00)	3 (1.15)
17	2863– 3033	0 (0.00)	0 (0.00)	1 (0.38)
18	3034– 3329	0 (0.00)	0 (0.00)	2 (0.76)
19	3330– 3674	0 (0.00)	0 (0.00)	2 (0.76)
20	3675– 3979	0 (0.00)	0 (0.00)	0 (0.00)
21	3980– 4323	0 (0.00)	0 (0.00)	0 (0.00)
22	4324– 4821	0 (0.00)	0 (0.00)	0 (0.00)
23	4822– 5219	0 (0.00)	0 (0.00)	0 (0.00)
24	5220– 5685	0 (0.00)	0 (0.00)	0 (0.00)
25	5686– 6368	0 (0.00)	0 (0.00)	0 (0.00)
26	6369– 7241	0 (0.00)	0 (0.00)	0 (0.00)
27	7242– 8452	0 (0.00)	0 (0.00)	0 (0.00)
28	8453–10093	0 (0.00)	0 (0.00)	0 (0.00)
29	10094–11368	0 (0.00)	0 (0.00)	0 (0.00)
30	11369–12829	0 (0.00)	0 (0.00)	0 (0.00)
31	12830–25000	0 (0.00)	0 (0.00)	0 (0.00)
Total		256 (100.00)	350 (100.00)	262 (100.00)

the results are virtually identical with regard to their significant departures from independence [*15*]. Therefore, in Table 8, we present only the nonparametric correlation estimates, evaluated by the method of Chakraborty et al. [*15*]. Their empirical significance was judged by shuffling the observed fragment sizes across individuals. Except the intraclass correlation at D17S79 locus in the Caucasian sample (which was significantly larger than zero, at $P = 0.01$ level), no correlation coefficient was significant at 5% level, suggesting that even the presence of nondetectable alleles did not affect the correlation test of fragment size independence. While the occurrence of this single significance (in a total of 45 tests) may be attributed to chance alone, note that the null allele frequency estimate (r) for the D17S79 locus in the Caucasian sample is $4.0 \pm 3.6\%$ (Table 7), and this would be enough to explain the pseudo-dependence of fragment sizes within individuals at this locus.

Global Test of Gametic Disequilibrium Based on the Distribution of the Number of Heterozygous Loci

Table 9 shows the observed and expected proportions (in percent) of heterozygotes (with the fixed-bin definition of alleles) for each locus for the three ethnic groups. The expected proportions were calculated by two methods, first by ignoring the nondetectable alleles, whereby from the observed binned allele frequencies (Tables 1 through 5) the unbiased estimate of heterozygosity [*30*] were computed. These are the same estimates that were employed in the chi-square test of Hardy-Weinberg equilibrium (Table 6). Clearly, in all but two cases (D2S44 in Caucasians, and D4S139 in Hispanics) there are deficiencies of observed heterozygosities in relation to these expected heterozygosity values, and these are significant (at 5% level) in 9 cases (see Table 6). The second method of computing the expected heterozygosity was to invoke the null allele frequency estimates (shown in Table 7), so that the expectations of the binned homozygosity now included heterozygosity for null allele as well, in addition to the true homozygotes. The exact equation for this second expected heterozygosity, adjusted for the occurrence of nondetectable alleles, is given in [*25*]. Comparison of the observed and expected heterozygosities of Table 9 re-affirms our previous analysis (Tables 6 and 7) that all of the observed significant chi-square values are truly due to the presence of nondetectable alleles in the RFLP-based DNA typing method.

Table 10 shows the summary of the tests of independence of loci at the binned profile level. As expected, the observed variance (s_k^2) of the number of heterozygous loci (of the five loci tested) in individuals on whom data are available for all five loci is within the 95% confidence interval under the independence assumption, when the observed locus-specific binned heterozygosity values are used. However, when the locus-specific heterozygosity estimates were computed under the assumption that the alleles within each locus combine independently to form an individual's genotype, we find that of the three samples, in two (except the Caucasians) the observed value of s_k^2 is outside the 95% confidence limits. But, this happens only when we disregard the possibility of nondetectable alleles (shown in rows where the expected locus-specific heterozygosity estimates do not account for the presence of nondetectable alleles). However, such discrepancies did not occur when the locus-specific heterozygosity estimates were revised incorporating the null allele occurrences, suggesting that, like the case of

TABLE 6—*Test of Hardy-Weinberg Equilibrium by different test procedures without incorporating nondetectability of* HaeIII *restriction fragments.*

	Caucasians	Hispanics	Blacks
DIS7 (n)	168	240	181
H: Obs (Exp)	153 (159.15)	215 (226.76)	162 (171.92)
Chi-square (P)	4.52 (0.04)	11.07 (0.00)	11.41 (0.00)
−2*ln(L) (P)	280.79 (0.02)	284.98 (0.40)	278.91 (0.47)
Exact Test Prob.	0.02	0.24	0.23
D2S44 (n)	153	256	177
H: Obs (Exp)	143 (142.38)	231 (237.01)	158 (165.44)
Chi-square (P)	0.04 (0.89)	2.05 (0.15)	5.12 (0.03)
−2*ln(L) (P)	173.85 (0.56)	185.43 (0.93)	247.82 (0.06)
Exact Test Prob.	0.60	0.84	0.03
D4S139 (n)	171	268	196
H: Obs (Exp)	149 (154.89)	237 (236.98)	170 (181.97)
Chi-square (P)	2.38 (0.15)	0.00 (1.00)	11.00 (0.00)
−2*ln(L) (P)	126.18 (0.35)	124.44 (0.37)	186.30 (0.08)
Exact Test Prob.	0.39	0.44	0.02
D10S28 (n)	168	251	185
H: Obs (Exp)	158 (158.67)	234 (234.69)	170 (176.44)
Chi-square (P)	0.05 (0.87)	0.03 (0.89)	5.07 (0.03)
−2*ln(L) (P)	249.98 (0.14)	241.81 (0.14)	307.95 (0.22)
Exact Test Prob.	0.16	0.15	0.09
D17S79 (n)	128	175	131
H: Obs (Exp)	89 (102.74)	137 (147.28)	92 (111.56)
Chi-square (P)	9.31 (0.00)	4.53 (0.03)	23.11 (0.00)
−2*ln(L) (P)	45.55 (0.15)	53.27 (0.18)	96.72 (0.03)
Exact Test Prob.	0.05	0.10	0.00

within locus analysis, the presence of nondetectable alleles may induce a pseudo-dependence across loci as well.

Comparison with National (FBI) Database

The above analyses demonstrated that the DNA typing data on the five VNTR loci (D1S7, D2S44, D4S139, D10S28, and D17S79) in three ethnic samples from the Houston area conforms to allelic independence (at bin level, as well as at the level of measured fragment sizes) within as well as between loci. Nevertheless, the fixed-bin allele frequency data (Tables 1–5), when compared with the national database generated by the FBI laboratory, may reflect some statistical differences. An exact test of contingency chi-square analysis reveals that these allele frequencies are somewhat different from the published fixed-bin allele frequencies [*12*], and the difference occasionally reaches statistical significance (data not shown). We ascribe this to sample size differences of these two databases, since the national database is about 3-times larger (n = 2000, approximately, for the FBI database, in contrast of the total number of 669 individuals studied here.

To examine the forensic implications of such allele frequency differences, Fig. 1 shows the impact of fixed-bin allele frequency differences between the present database and that of the national database [*12*]. In these computations, we have taken all individuals from the database (including the ones for whom typings for some loci are missing), and computed their multilocus profile frequencies using the fixed bin allele frequencies of the present study and those of the national database (for which the fixed bin allele frequencies were taken from [*12*]). The profile frequency computations used the multiplication rules within as well as between loci, since for both databases these assumptions are now validated (from

TABLE 7—*Tests for Hardy-Weinberg Equilibrium of the fixed bin frequencies at five VNTR loci in four ethnic populations of Houston, Texas, incorporating the nondetectability of* HaeIII *restriction fragments.*

Locus	Statistic	Populations: Caucasians	Blacks	Hispanics
D1S7	Score (T)	1.74	2.96	2.51
	Prob. (T ≤ 1)	0.03*	0.00*	0.00*
	$r \pm$ se(r) in %	1.43 ± 1.31	3.63 ± 1.64	2.90 ± 1.48
	P-values:			
	for Chi-square	0.77	0.58	0.84
	for -2ℓnL	0.31	0.91	0.92
	for exact test	0.32	0.92	0.86
D2S44	Score (T)	0.83	1.77	1.42
	Prob. (T ≤ 1)	—	0.01*	0.08
	$r \pm$ se(r) in %	0.0	1.60 ± 1.33	0.88 ± 1.15
	P-values:			
	for Chi-square	0.77	0.71	0.82
	for -2ℓnL	0.83	0.43	0.99
	for exact test	0.78	0.42	0.87
D4S139	Score (T)	1.14	1.95	0.96
	Prob. (T ≤ 1)	0.33	0.01*	—
	$r \pm$ se(r) in %	0.41 ± 1.13	2.63 ± 1.33	0.0
	P-values:			
	for Chi-square	0.31	0.66	1.00
	for -2ℓnL	0.71	0.59	0.71
	for exact test	0.66	0.40	0.61
D10S28	Score (T)	1.04	1.71	0.89
	Prob. (T ≤ 1)	0.46	0.03*	—
	$r \pm$ se(r) in %	0.08 ± 0.82	1.31 ± 1.22	0.0
	P-values:			
	for Chi-square	1.00	0.79	1.00
	for -2ℓnL	0.48	0.66	0.45
	for exact test	0.50	0.56	0.48
D17S79	Score (T)	1.81	2.43	1.54
	Prob. (T ≤ 1)	0.00*	0.00*	0.02*
	$r \pm$ se(r) in %	4.04 ± 3.59	4.78 ± 3.11	2.47 ± 2.82
	P-values:			
	for Chi-square	0.33	0.08	0.63
	for -2ℓnL	0.89	0.80	0.75
	for exact test	0.54	0.44	0.56

TABLE 8—*Intraclass and interclass correlations of fragment sizes within and between VNTR loci in samples from three ethnic populations of Houston, Texas.*

Locus	D1S7	D2S44	D4S139	D10S28	D17S79
			Caucasians		
D1S7	−.012	−.040	.040	−.012	.031
D2S44		−.137	.024	.036	−.001
D4S139			.022	.045	.037
D10S28				−.049	.032
D17S79					.226[a]
			Blacks		
D1S7	−.025	.062	.023	.050	−.047
D2S44		.127	.027	.072	.004
D4S139			.085	−.055	.040
D10S28				.060	−.033
D17S79					−.042
			Hispanics		
D1S7	−.047	.006	.017	−.034	−.070
D2S44		−.024	.062	−.001	.048
D4S139			−.014	.001	−.043
D10S28				.022	−.064
D17S79					−.067

[a] $P < 0.05$, the only significant correlation.

TABLE 9—*Observed and expected heterozygosity[a] at five VNTR loci in three ethnic populations of Houston.*

Populations		Heterozygosity (in %) for the Locus				
		D1S7	D2S44	D4S139	D10S28	D17S79
Caucasians	obs.	91.1	93.5	87.1	94.1	69.5
	exp.[1]	94.7	93.1	90.6	94.5	80.3
	exp.[2]	91.8	92.8	89.6	94.0	73.7
Blacks	obs.	89.5	89.3	86.7	91.9	70.2
	exp.[1]	95.0	93.5	92.8	95.4	85.2
	exp.[2]	88.1	90.3	87.9	92.6	77.1
Hispanics	obs.	89.6	90.2	88.4	93.2	78.3
	exp.[1]	94.5	92.6	88.4	93.5	84.2
	exp.[2]	89.0	90.8	88.3	93.3	80.0

[a]Of the two methods of computing the expected heterozygosities, the first (exp.[1]) does not account for the existence of non-detectable alleles, while the second (exp.[2]) takes into account their presence. See text for details.

TABLE 10—*Test of independence of loci based on the variance (s_k^2) of the number of heterozygous (binned) loci in individuals.*

	Caucasians	Blacks	Hispanics
n	92	118	138
Observed s_k^2	0.452	0.549	0.564
Expected s_k^2 and its 95% Confidence Limits (CL)			
(1) Based on obs. het.:			
s_k^2	0.511	0.576	0.505
Lower 95% CL	0.352	0.418	0.370
Upper 95% CL	0.669	0.733	0.640
(2) Based on exp. het.[1]:			
s_k^2	0.411	0.347[a]	0.415[a]
Lower 95% CL	0.268	0.231	0.294
Upper 95% CL	0.555	0.463	0.535
(3) Based on exp. het.[2]:			
s_k^2	0.486	0.544	0.508
Lower 95% CL	0.328	0.389	0.371
Upper 95% CL	0.644	0.699	0.646

n = Number of individuals on whom data are available on all 5 loci.
[a]Observed value of s_k^2 is outside the confidence interval.
[1,2]Using the locus-specific expected heterozygosities are the ones shown in Table 9.

the present analyses for the Houston database, and from [*13–15,25*] for the FBI database). These profile frequencies are plotted against each other (in logarithmic scale, $-\log(P_{FBI})$ indicating the negative logarithm of a multilocus profile of an individual in the present database, based on fixed-bin allele frequencies of the national database, while $-\log(P_{HPD})$ is the same by using the fixed bin allele frequencies of Tables 1 through 5) for all individuals. Therefore, panel (*a*) of Fig. 1 shows the impact of such allele frequency differences for the 193 Caucasians, panel (*b*) for the 204 Blacks, and panel (*c*) for the 272 Hispanic individuals of the present study. In all cases, the estimates of the multilocus DNA profile frequencies reside around the 45-degree line, suggesting that even when some binned allele frequencies show statistical significance between databases, their impact on DNA profile frequency computations is noticeable only when in both populations this profile is extremely rare. For example, as seen in this diagram, there are some profiles where the two databases may exhibit profile frequency estimates that might be 100-fold different from each other (in particular in panel c, for the Hispanics). But, a closer examination may reflect such differences occur when the profile frequencies are of the order of 1 in a million, or rarer.

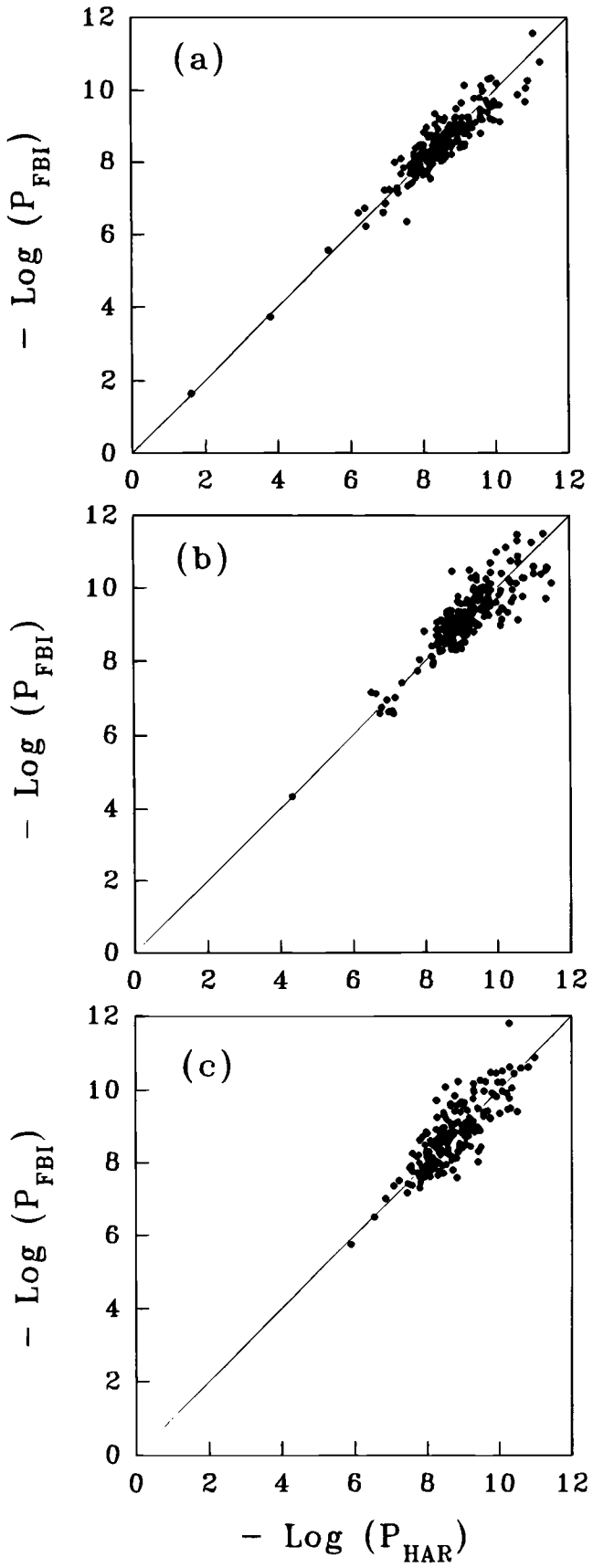

FIG. 1—*Scatter plots of frequencies of multilocus DNA profiles in the present database, computed from fixed bin allele frequency data of the present survey (horizontal axis, P_{HPD}) versus the same obtained from the fixed bin allele frequencies the FBI National database (vertical axis, P_{FBI}). Both frequencies are in logarithmic scale. Panels indicate profile frequency comparisons for* (a) *193 Caucasians,* (b) *204 Blacks, and* (c) *272 Hispanics.*

Discussion and Conclusions

In aggregate, the above results present a comprehensive analysis of a RFLP-based DNA typing database from the Houston area that may be used for forensic case analyses at a local level. We have shown that the assumptions of allelic independence within as well as between loci are appropriate when we invoke the presence of nondetectable alleles within the present database. One might argue that we have not demonstrated that such alleles are truly present in the same database. While this could have been done using a different restriction enzyme (for example, *Pvu*II, or *Pst*I, which yield fragment sizes larger than the *Hae*III enzyme used for digestion for the preparation of this database), we did not perform such experiments in our laboratory, because such work was done at least for two other databases within the US. In the Western California database, the *Hae*III null alleles at the D2S44 locus occur with a frequency of 1.7% in the Blacks [*24*]. Our estimate (1.6 ± 1.3%, see Table 7) compares reasonably well with this. *Hae*III null alleles at the D2S44 locus for the national data on Blacks occur with a frequency of 2.7%, where most of the Black individuals (for whom the experiment with the alternative enzyme, *Pvu*II was conducted) were from Texas [*25*]. They also report the occurrence of *Hae*III null alleles at the D17S79 locus with frequencies of 6.5% in Blacks, 3.0% in Hispanics of Texas, and 3.7% in Hispanics of Florida, which are comparable with the range of our statistical estimates (2.5–4.8%, see Table 7) for this locus.

Since the independence assumptions (within and across loci) appear reasonable, and nondetectable alleles must be invoked in such interpretations, one might question how valid would be the fixed-bin allele frequencies (Tables 1 through 5), because in their computations no adjustment for null allele occurrences were made. We contend that since for forensic use of allele frequency estimates it is desirable to be conservative, these unadjusted fixed-bin allele frequencies would suffice, because in the presence of null alleles such estimates are conservative [*14*]. Indeed, simple algebraic derivations would show that if we construct re-binned tables from the ones shown in Tables 1 through 5 with the convention that each (re)bin must have at least five allelic counts, the degree of conservativeness of these fixed-bin allele frequencies are so high (when null alleles are incorporated) that it is not even necessary to invoke the suggestion of the National Research Council report [*10*] to consider the upper 95% confidence bound of each individual allele frequency.

Although the tests of independence and comparisons of the present database with the national data [*12*] are done with the fixed-bin classification of alleles, we contend that the results would have been qualitatively the same if a floating bin concept of alleles is invoked. This conclusion is reached through two supplementary analyses.

First, even though at our laboratory we have not done a full-scale experiment to evaluate the degree of measuremental error, on every gel we used the K562 cell line as a positive sizing control. In total, we have 81 to 124 measurements on these fragment sizes (which vary from locus to locus), which gave average molecular sizes from 1187bp to 6497bp. The standard deviation of these multiple measurements (which involve both intra- and inter-gel comparisons) ranged from 8.35 to 55.11, respectively, which are 0.70% to 0.85% of the average sizes. Therefore, the fixed-bin window widths (ranging from 5.8% of the center for Bin 17 to 18.7% of the center for Bin 2) are conservative enough to account for measuremental errors for the protocol used in this laboratory.

Second, we examined whether or not the conservativeness of the fixed bin allele frequencies would be compromised if floating bins of width ±2.5% are used for the entire range of fragment size for each of these loci. The protocol of this numerical exercise was identical to the one described in [*31*]. In brief, we defined sliding floating bins of widths ±2.5% centered from the minimum to maximum, each time computing the number of fragments (f_1) in the database residing in the floating bin. For comparison, we also computed the fixed-bin frequencies, using the protocol currently practiced by the forensic community [*2*], where the fixed bin frequency is either taken from Tables 1 through 5 directly (when the ±2.5% sliding floating bin is within a specific fixed bin). In the case where the sliding ±2.5% floating bin overlapped two adjacent fixed bins, we took the larger of the respective fixed bin frequencies (f_2) for the specific database. The maximum of the ratio of f_1/f_2 is plotted against the centers of the sliding floating bins, which are shown in Fig. 2 (five panels for five loci) for the three ethnic samples. Since the ratio f_1/f_2 never exceeded one in any case, we conclude that the fixed-bin allele frequencies are always more conservative than the ±2.5% floating bins for the entire range of fragment sizes in this database. This is consistent with the analysis of the national database [*12,31*]. In addition to showing the conservative feature of the fixed-bin allele frequencies, this particular analysis further demonstrates that there is no need to add the fixed bin frequencies as suggested in the National Research Council report [*10*]. When a floating bin overlaps two adjacent fixed-bins, it is enough to consider the larger of the two fixed bin frequencies, as currently practiced [*2*], and this does not compromise the conservative features of the fixed-bin allele frequency computations.

An additional empirical test of independence of fragment sizes across loci was also done at the floating bin level for the present data sets by searching for matches between all possible pairs of profiles for each ethnic sample separately. A match is defined when the ±2.5% wide intervals of fragment sizes found in two individuals overlapped with each other. This generated locus-specific empirical match probabilities which were multiplied to generate the expected 2-, 3-, 4-, or 5-locus random match probabilities to compare with the observed ones in the database. Table 11 presents the summary of this analysis, which shows that the observed number of several multi-locus matches in this database occur in accordance with what is expected based on the allele frequencies alone. In other words, the use of empirical locus-specific matches for these calculations exhibits that the multiplication rule (across loci) is also valid with the floating bin concept of allelic definitions. We might note that this analysis supports the conclusion of Herrin [*32*] who performed a similar empirical search of multilocus match probabilities in another regional database of DNA typing.

Finally, one might argue that these analyses do not address the question of adequacy or representativeness of the individuals of the Houston area. With regard to the adequacy, we contend that the sample sizes (193 Caucasians, 204 Blacks, and 272 Hispanics) for the three of the four ethnic groups are more than adequate in contrast with the minimum sample size requirements prescribed [*33,34*]. These sample sizes are even larger than the ones suggested in the National Research Council report [*10*]. Although all conclusions described above hold for the Oriental sample as well, because of the small number (n = 36) of individuals in this sample, we recommend that until supplemented by more samples, the binned allele frequencies from this sample should be used with caution.

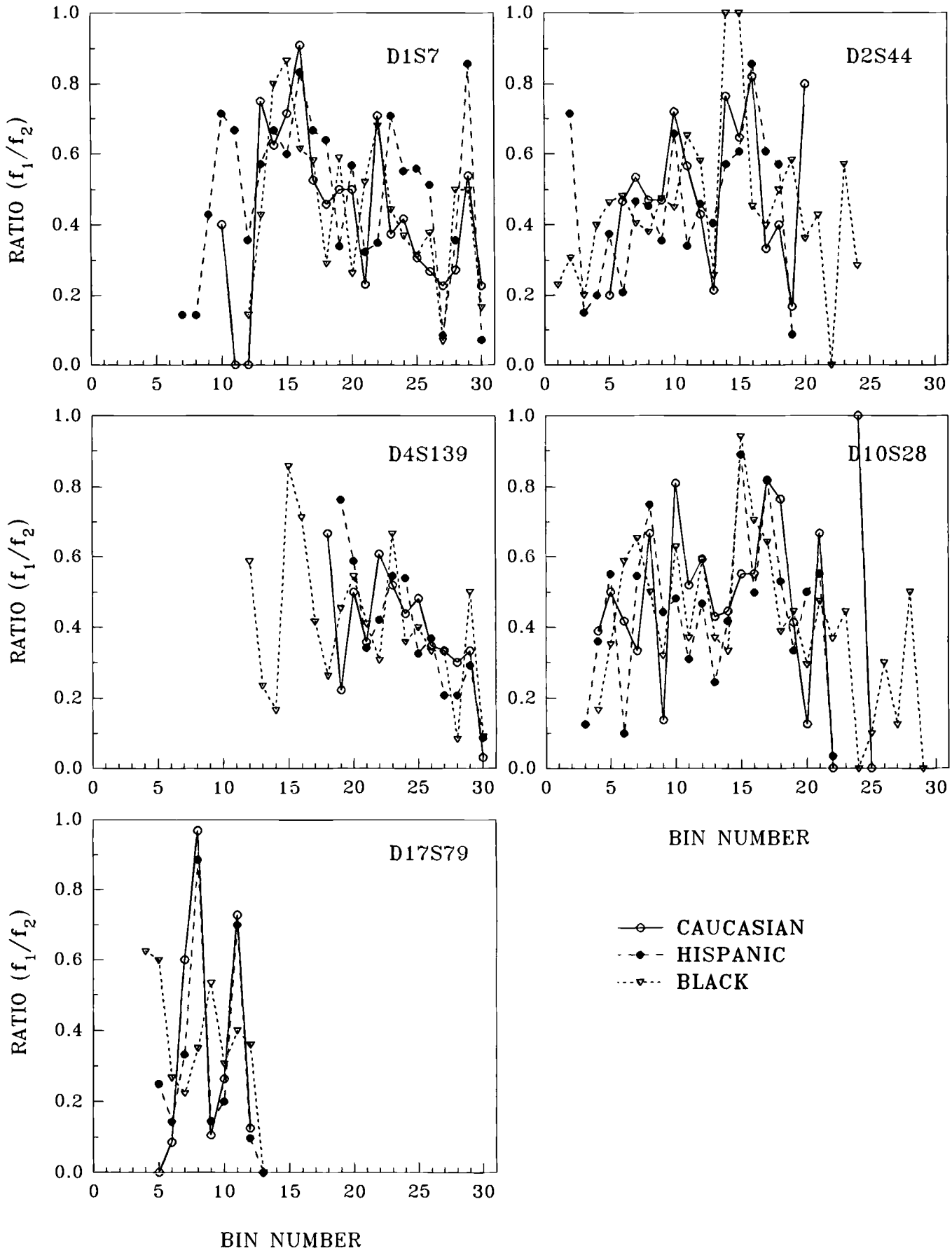

FIG. 2—*Ratio of ±2.5% floating bin (f_1) and fixed bin (f_2) frequencies of DNA fragment size in the present database for any DNA fragment size whose ±2.5% floating bin window crosses the boundaries of any of the 31 bins. Each panel represents the binned data for the locus indicated on the panel.*

TABLE 11—*Observed and expected numbers of multi-locus matches in the samples from three ethnic groups of Houston, Texas.*

Matches at[a]	Number of comparisons	Number of matches observed	Number of matches expected	Probability[b] of (obs. ≥ exp.)
		Caucasians		
2 loci	90,248	22	22.545	0.546
3 loci	68,761	0	0.219	1.0
4 loci	26,905	0	0.001	1.0
		Blacks		
2 loci	113,134	11	9.763	0.346
3 loci	92,037	1	0.069	0.067
4 loci	38,188	0	<0.001	1.0
		Hispanics		
2 loci	224,675	43	53.422	0.923
3 loci	173,578	1	0.581	0.441
4 loci	65,319	0	0.003	1.0

[a]Data on all combinations of multiple loci are pooled. No 5-loci match was observed for any sample.

[b]These probabilities are computed based on the assumption that the number of matches follows a Poisson distribution.

The representativeness of any sample for a cosmopolitan population such as the Houston residents is a difficult issue to address, because even if the random sample was drawn based on random numbers generated from any registry (or household-list), that would not remain a random sample at a different time point because of high mobility of the society in such a large city. Since population movement does not occur on the basis of DNA types at loci which have no physiological or functional attributes, we contend that the randomness of the present data is not compromised. Furthermore, it has been demonstrated that data collected through such convenient sources (such as blood donors, police cadets, Health Centers) are comparable with the ones obtained from structured statistical surveys, both in terms of phenotype frequencies at genetic markers [*35*], as well as in terms of genetic diversity within and between populations [*36*].

In summary, this comprehensive analysis demonstrates that a RFLP database of DNA typing for the major ethnic groups in Houston area is now available, and this satisfies the assumptions of independence of DNA fragment sizes within and across loci. Therefore, for forensic applications this database should meet the requirements of statistical calculations for judging the significance of a DNA match found in a forensic case, or for calculation of the odds of paternity when an accused male is not excluded in a parentage testing by these loci. The fixed-bin method of allele frequency computations is shown to be conservative, and the variance among multiple measurements of the same fragment size, nor the presence of nondetectable RFLP alleles compromise the quality of the data. The comparability with the national database suggests that if other refined definition of ethnicity is needed, for a cosmopolitan population for a city as large as Houston, comparable data at the National level may be enough to derive conservative population-specific estimates of specific DNA profiles.

Acknowledgments

The statistical analysis of this work was supported by Grants GM 41399 and GM 45861 from the National Institutes of Health, and grant 92-IJ-CX-K024 from the National Institute of Justice.

References

[*1*] Jeffreys, A. J., Turner, M., and Debenham, P., "The Efficiency of Multilocus DNA Fingerprint Probes for Individualization and Establishment of Family Relationships, Determined from Extensive Casework," *American Journal of Human Genetics*, Vol. 48, 1991a, pp. 824–840.

[*2*] Budowle, B., Giusti, A. M., Waye, J. S., Baechtel, F. S., Fourney, R. M., Adams, D. E., Presley, L. A., Deadman, H. A., and Monson, K. L., "Fixed-Bin Analysis for Statistical Evaluation of Continuous Distributions of Allelic Data from VNTR Loci for Use in Forensic Comparisons," *American Journal of Human Genetics*, Vol. 48, 1991a, pp. 841–855.

[*3*] Southern, E. M., "Detection of Specific Sequences Among DNA Fragments Separated by Gel Electrophoresis," *Journal of Molecular Biology*, Vol. 98, 1975, pp. 503–517.

[*4*] Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H., "Specific Enzymatic Amplification of DNA In Vitro: The Polymerase Chain Reaction," *Cold Spring Harbor Symposium on Quantitative Biology*, Vol. 51, 1986, pp. 263–273.

[*5*] Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A., "Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase," *Science*, Vol. 239, 1988, pp. 487–491.

[*6*] Jeffreys, A. J., Wilson, V., Neumann, R., and Keyte, J., "Amplification of Human Minisatellites by the Polymerase Chain Reaction: Towards DNA Fingerprinting of Single Cells," *Nucleic Acids Research*, Vol. 16, 1988, pp. 10953–10971.

[*7*] Boerwinkle, E., Xiong, W., Fourest, E., and Chang, L., "Rapid Typing of Tandemly Repeated Hypervariable Loci by the Polymerase Chain Reaction: Application to the Apolipoprotein B 3′ Hypervariable Region," *Proceedings of the National Academy of Sciences, USA*, Vol. 86, 1989, pp. 212–216.

[*8*] Budowle, B., Chakraborty, R., Giusti, A. M., Eisenberg, A. J., and Allen, R. C., "Analysis of the VNTR Locus D1S80 by the PCR Followed by High-Resolution PAGE," *American Journal of Human Genetics*, Vol. 48, 1991b, pp. 137–144.

[*9*] Sajantilla, A., Budowle, B., Ström, M., Johnsson, V., Lukka, M., Peltonen, L., and Ehmholm, C., "PCR Amplification of Alleles at the D1S80 Locus: Comparison of a Finnish and a North American Caucasian Population Sample, and Forensic Casework Evaluation," *American Journal of Human Genetics*, Vol. 50, 1992, pp. 816–825.

[*10*] National Research Council, *DNA Technology in Forensic Science*, National Academy Press, Washington, DC, 1992.

[*11*] Ayala, F. J. and Black, B., "Science and the Courts," *American Scientist*, Vol. 81, 1993, pp. 230–239.

[*12*] Budowle, B., Monson, K. L., Anoe, K., Baechtel, F. S., Bergman, D., Buel, E., Campbell, P., Clement, M. E., Coery, H. W., Davis, L. A., Dixon, A., Fish, P., Giusti, A. M., Grant, T. L., Gronert, T. M., Hoover, D. M., Jankowski, L., Kilgore, A. J., Kimoto, W., Landrrum, W. H., Leone, H., Longwell, C. B., MacLaren, D. C., Medlin, L. E., Narveson, S. D., Pierson, M. L., Pollock, J. M., Raquel, R. J., Reznicek, J. M., Rogers, G. S., Smerick, J. E., and Thompson, R. M., "A Preliminary Report on Binned General Population Data on Six VNTR Loci in Caucasians, Blacks and Hispanics from the United States," *Crime Laboratory Digest*, Vol. 18, 1991c, pp. 9–26.

[*13*] Weir, B. S., "Independence of VNTR Alleles Defined As Fixed Bins," *Genetics*, Vol. 130, 1992, pp. 873–887.

[*14*] Chakraborty, R., de Andrade, M., Daiger, S. P., and Budowle, B., "Apparent Heterozygote Deficiencies Observed in DNA Typing Data and Their Implications in Forensic Applications," *Annals of Human Genetics*, Vol. 56, 1992, pp. 45–57.

[*15*] Chakarborty, R., Srinivasan, M. R., and de Andrade, M., "Intraclass and Interclass Correlations of Allele Sizes Within and Between Loci in DNA Typing Data," *Genetics*, Vol. 123, 1993a, pp. 411–419.

[*16*] Devlin, B., Rische, N., and Roeder, K., "No Excess of Homozygosity at Loci Used for DNA Fingerprinting," *Science*, Vol. 249, 1990, pp. 1416–1420.

[*17*] Devlin, B. and Risch, N., "A Note on Hardy-Weinberg Equilibrium of VNTR Data by Using Federal Bureau of Investigation's Fixed-Bin Method," *American Journal of Human Genetics*, Vol. 51, 1992, pp. 549–553.

[*18*] Budowle, B. and Baechtel, F. S., "Modifications to Improve the Effectiveness of Restriction Fragment Length Polymorphism Typing," *Applied and Theoretical Electrophoresis*, Vol. 1, 1990, pp. 181–187.

[*19*] Monson, K. L. and Budowle, B., "A System for Semi-Automated Analysis of DNA Autoradiograms," In *Proceedings of an Interna-*

tional Symposium on the Forensic Aspects of DNA Analysis, US Government Printing Office, Washington, DC, 1990, pp. 127–132.

[*20*] Li, C. C., *First Course in Population Genetics,* Boxwood, Pacific Grive, CA, 1976.

[*21*] Weir, B. S., *Genetic Data Analysis,* Sinauer, Sunderland, MA, 1991.

[*22*] Guo, S. W. and Thompson, E. A., "Performing the Exact Test of Hardy-Weinberg Proportion for Multiple Alleles, *Biometrics,* Vol. 48, 1992, pp. 361–372.

[*23*] Brown, A. D. H., Feldman, M. W., and Nevo, E., "Multilocus Structure of Natural Populations of *Hordeum spontaneum,*" *Genetics,* Vol. 96, 1980, pp. 523–536.

[*24*] Steinberger, E. M., Thompson, L. D., and Hartmann, J. M., "On the Use of Excess Homozygosity for Subpopulation Detection," *American Journal of Human Genetics,* Vol. 52, 1993, pp. 1275–1277.

[*25*] Chakraborty, R., Zhong, Y., Jin, L., and Budowle, B., "Nondetectability of Restriction Fragments and Independence of DNA Fragment Sizes Within and Between Loci in RFLP Typing of DNA," *American Journal of Human Genetics,* Vol. 55, 1994, pp. 391–401.

[*26*] Gart, J. J. and Nam, J., "A Score Test for the Possible Presence of Recessive Alleles in Generalized ABO-like Genetic Systems," *Biometrics,* Vol. 40, 1984, pp. 887–894.

[*27*] Laboratory Division Federal Bureau of Investigation, *VNTR Population Data: A Worldwide Study,* Vol. I-V. Forensic Science Research and Training Center, FBI Academy, Quantico, VA, 1993.

[*28*] Geisser, S. and Johnson, W., "Testing Hardy-Weinberg Equilibrium on Allelic Data from VNTR Loci," *American Journal of Human Genetics,* Vol. 51, 1992, pp. 1084–1088.

[*29*] Karlin, S., Cameron, E. C., and Williams, P. T., "Sibling and Parent-Offspring Correlation Estimation with Variable Family Size," *Proceedings of National Academy of Sciences, USA,* Vol. 78, 1981, pp. 2664–2668.

[*30*] Nei, M., "Estimation of Average Heterozygosity and Genetic Distance from a Small Number of Individuals," *Genetics,* Vol. 89, 1978, pp. 583–590.

[*31*] Chakraborty, R., Jin, L., Zhong, Y., Srinivasan, M. R., and Budowle, B., "On Allele Frequency Computation from DNA Typing Data," *International Journal of Legal Medicine,* Vol. 106, 1993b, pp. 103–106.

[*32*] Herrin, G., "Probability of Matching RFLP Patterns from Unrelated Individuals," *American Journal of Human Genetics,* Vol. 52, 1993, pp. 491–497.

[*33*] Chakraborty, R., "Sample Size Requirements for Addressing the Population Genetic Issues of Forensic Use of DNA Typing," *Human Biology,* Vol. 64, 1992, pp. 141–159.

[*34*] Evett, I. W. and Gill, P., "A Discussion of the Robustness of Methods for Assessing the Evidential Value of DNA Single Locus Profiles in Crime Investigations," *Electrophoresis,* Vol. 12, 1991, pp. 226–230.

[*35*] Cohen, B. H., Bias, W. B., Hamill, P. V. V., Drizd, T. A., and Diamond, E. L., "Selected Genetic Markers of Blood and Secretions for Youths 12–17 Years of Age," *U.S. Department of Health, Education and Welfare Publication No. (PHS) 80-1664,* National Center of Health Statistics, Hyattsville, MD, 1980.

[*36*] Chakraborty, R., "Analysis of Genetic Structure of Populations: Meaning, Methods, and Implications," *Human Population Genetics,* P. P. Majumder, Ed., Plenum Press, New York, 1993, pp. 189–206.

Address requests for reprints or additional information to
Ranajit Chakraborty, Ph.D.
Human Genetics Center
The University of Texas Houston Health Science Center
P.O. Box 20334
Houston, TX 77225