*Bruce Budowle,[1] Ph.D.; Keith L. Monson,[1] Ph.D.;
Alan M. Giusti,[1] B.S.; and Barry L. Brown,[1] Ph.D.*

The Assessment of Frequency Estimates of Hae III-Generated VNTR Profiles in Various Reference Databases

ABSTRACT: The likelihood of occurrence of 1964 Hae III-generated target DNA profiles was estimated using fixed bin frequencies from various regional and ethnic databases and the multiplication rule. The databases generally were from the following major categories: Black, Caucasian, Hispanic, Oriental, and American Indian. It was found that subdivision, either by ethnic group or by U.S. geographic region, within a major population group did not substantially affect forensic estimates of the likelihood of occurrence of a DNA profile. As expected, the greatest variation in estimates for within-group estimates was among American Indian databases. Because the greatest variation in statistical estimates occurs across-major population groups, in most cases, there will be no unfair bias applying general population database estimates. Therefore, based on empirical data, there is no demonstrable need for using alternate approaches, such as the ceiling approach, to derive statistical estimates. The current practice of using general population databases and the multiplication rule provides valid estimates of the likelihood of occurrence of a DNA profile.

KEYWORDS: pathology and biology, VNTR, allele frequency, population databases,

Although statistical inferences about the rarity of DNA profile matches should be a matter of the weight of the evidence, the admissibility of statistical inferences of DNA profile matches has confounded some courts. When the reliability of statistical estimates for DNA analyses has been challenged, the two main contentions that have been raised are that: 1) a reference population database should reflect the ethnic make-up of the suspect, and 2) potentially large differences in allele frequencies among subpopulation groups in the United States could result in significantly different estimates of the likelihood of occurrence of a DNA profile [1,2]. Because of the power of discrimination DNA typing affords, particularly restriction fragment length polymorphism (RFLP) analysis of variable number of tandem repeat (VNTR) loci, vigorous challenges in the legal arena were expected. However, the concerns for these issues are misplaced.

Before attempting to define the "proper" reference database to be used in DNA

analyses, the germane legal question must be stated for the general case. Under the United States Constitution, the suspect is guaranteed a right to a fair trial, and a basic element of this right is the presumption of innocence. Therefore, the legal question for the typical case is, "What is the likelihood that someone other than the suspect is the source of the evidentiary material?" The relative rarity of a DNA pattern in a suspect's ethnic subgroup, which might be of some academic interest, is not particularly relevant in the legal setting [3]. To use the specific ethnic background of the suspect (which may be impossible to define) would presuppose that he or she be the true perpetrator. However, if the true perpetrator were known a *priori*, there would be no need for statistical estimates. Furthermore, if a particular subgroup was chosen as the reference database, for the majority of cases this would insinuate that a member of one subgroup is a more likely source of the crime scene evidence. Since the ethnicity of those people who are potential perpetrators rarely, if ever, is known, statistical estimates must be based on some sort of general population database.

While the ethnic background of the suspect is not germane to selecting a reference database [3–7], the possible impact of different allele frequencies in subpopulations on statistical estimates has been a bit more elusive for a few courts. Some critics contend that reference databases based on subpopulations in the United States, rather than on general population groups, might produce large differences in the estimated likelihood of occurrence of DNA profiles [1]. They suggest it would be necessary to assess the frequencies of DNA profiles in a variety of ancestral ethnic groups before proceeding with assigning statistical estimates. The present debate is not "between alternative hypotheses of some substructure and no substructure" [8], but it is about whether substructure causes significant effects on statistical estimates. Indeed, it is universally accepted that substructure exists within major population groups. Rather, the important issue is whether general population group databases will yield probability estimates that would convey reliable and/or meaningful results in the forensic context. Because of this debate, alternate approaches, such as the "ceiling principle" [2] have been devised to address hypothetical issues arising from population substructure. Unfortunately, such approaches have not first addressed the germane legal question that needs to be answered, so that the issue of the use of general population databases would have been clarified. Furthermore, these approaches have not considered the available population data and their forensic significance [9–30]. The object of calculations for estimating DNA profile frequencies is to demonstrate the rarity of the profile. A profile is rare whether it has an estimated frequency of 1/5,000,000, 1/50,000,000, or 1/500,000,000. Obviously, the difference in the rarity of these estimates would have little consequence in a forensic context. Differences in statistical estimates are deemed "forensically significant" when the likelihood of occurrences of the DNA profile would be meaningfully different [31]. Comparisons of regional United States populations would be the most meaningful for determining forensic significance of statistical estimates for DNA profiles derived from evidence from crimes committed in the United States, because they would provide a valid reflection of the population of potential perpetrators. Additionally, VNTR population data from around the world can be evaluated for forensic significance. Such data should provide insight as to whether statistical estimates based on general population groups would produce forensically significant differences from other databases from regions of the United States, and from databases from around the world.

This paper makes use of VNTR population data generated by the forensic community using the restriction endonuclease Hae III to evaluate the forensic significance of subpopulations on statistical inferences drawn from general population databases. The conclusion of this study is that the current practice of using the multiplication rule and using general population databases to provide allele frequencies to estimate the likelihood of occurrence of a DNA profile in the general population would not yield a wrongful bias.

Materials and Methods

RFLP population data on several VNTR loci were kindly provided by the contributors displayed in Table 1. The data consisted of fragment lengths generated by digestion with the restriction endonuclease Hae III.

In this study, several conventions were followed to facilitate comparisons of populations. The conventions were generally those used by the FBI in casework analyses [12,32]. All fragment lengths in each population sample were sorted into 31 fixed bin categories as previously described [12]. The number of DNA fragments that fell into each bin was divided by the total number of alleles (that is, twice the number of individuals) in the sample population to determine the frequency of each bin. Additionally, frequencies were derived by a process termed "rebinning" whereby bins with fewer than five counts were merged with contiguous bins. After the bin tables were thus established, the frequency of an observed allele was estimated by determining in which bin(s) the fragment could reside, using a ±2.5% measurement error window [12]. Although measurement error can vary among laboratories, the ±2.5% measurement window was used to facilitate this study. If the measurement error window spanned a bin boundary, the frequency of the higher frequency bin was assigned to the allele [12]. The single-locus frequency of a two-band pattern was calculated using $2pq$, where p and q are the estimated binned allele frequencies for each VNTR band, while the frequency of occurrence of a single-band pattern was estimated using $2p$ [12]. The frequency of occurrence of a profile composed of multiple single-locus profiles was calculated as the product of the single-locus frequencies. Since measurement error for the FBI RFLP system can be greater than ±2.5% for fragments above approximately 10,000 base pairs (bp), any profile at a particular locus that contained an allele greater than 10,090 bp was assigned a locus frequency of 1.00. Additionally, since the size of fragments less than 640 bp can not be ascertained for the FBI RFLP system, any single-locus profile containing such a fragment was assigned a frequency of 1.00.

When using Hae III population data, target profiles of 1,964 individuals from the FBI's Caucasian ($N = 808$), Black ($N = 517$), and Hispanic ($N = 639$) databases were used. The FBI Black and Caucasian databases consist of individuals who reside in various regions of the United States, while the Hispanic database is subdivided into two groups, representing the southeastern and southwestern United States. The likelihood of occurrence of each profile, using the loci D1S7, D2S44, D4S139 and D10S28, was calculated in every available database using rebinned formats.

To compare databases, the inverse of the frequency estimated for the composite profile was plotted on a logarithmic (base 10) scale (that is, a scatter plot) for evaluation of forensic significance between/among various reference populations. Scatter plots were generated using population data either rebinned or sorted into the original 31 bins. With the latter approach, a minimum frequency of $1/n$ (where n = the total number of alleles at a locus for a particular database), or a minimum bin frequency of 0.001, was used to enable multiplication of a bin containing no observed alleles.

Because comparisons of rebinned databases of different sizes can exaggerate differences in estimates, an alternative approach involving random sampling of the larger database was performed to evaluate forensic significance when a large database is compared with a much smaller database. The number of individuals sampled from the larger database was defined by the number of samples in the smaller database at each locus. Five random samplings were performed, the 31 bin data resulting from each sampling were averaged, and then the averaged 31 bin tables were rebinned. Averaged rebinned data of the FBI Caucasian database were compared with the rebinned French and Israeli databases by scatter plot analysis. Additionally, the averaged rebinned data of the Black database were compared with the rebinned Haitian database, and the averaged rebinned

TABLE 1—*Reference databases and loci analyzed for Hae III-based data.[a]*

Laboratory	D1S7	D2S44	D4S139	D10S28
CA Caucasians[b]	212	215	217	215
FBI Caucasians[c]	595	792	594	429
FBI Israelis[c]	97	116	115	124
FL Caucasians[d]	239	240	215	204
GA Caucasians[e]	287	292	289	281
IL Caucasians[f]	309	300	300	306
MI Caucasians[g]	309	422	361	385
MN Caucasians[h]	251	255	242	242
NV Caucasians[i]	283	301	289	290
OR Caucasians[j]	272	273	272	273
VT Caucasians[k]	219	227	216	233
Montreal Caucasians[l]	658	457	611	749
Kingston Caucasians[m]	326	458	461	413
Vancouver Caucasians[m]	271	321	265	256
Toronto Caucasians[n]	228	229	219	228
French Caucasians[o]	156	128	203	116
Swiss Caucasians[p]	205	402	398	399
CA Blacks[b]	213	213	220	222
FBI Blacks[c]	359	475	448	288
FBI Haitians[c]	98	98	97	89
FL Blacks[d]	148	153	140	128
GA Blacks[e]	508	488	455	494
MI Blacks[g]	451	486	503	507
MI Atlanta Blacks[g]	423	486	440	473
MN Blacks[h]	213	213	211	210
SC Blacks[q]	230	245	241	245
AZ Hispanics[r]	212	216	207	208
CA Hispanics[b]	258	259	245	256
FBI SE Hispanics[c]	305	300	311	230
FBI SW Hispanics[c]	216	215	211	210
MI Hispanics[g]	87	91	87	91
AZ Amerindians[r]	185	185	153	189
MN Amerindians[h]	217	215	192	208
No. Ontario Amerindians[m]	195	215	182	208
Salishan Amerindians[m]	111	107	97	118
Saskatchewan Amerindians[m]	103	99	95	102
CA Chinese[b]	109	108	101	120
CA Japanese[b]	137	126	125	137
CA Korean[b]	100	99	93	100
Singapore Chinese[s]	150	200	202	212
Singapore Malaysians[s]	139	201	218	229
Asian Indians[m] (RCMP)	109	97	105	109
Asian Indians[s] (Singapore)	146	205	206	208

[a]The numbers in locus column represent the number of individuals typed.
[b]Orange County Sheriff's Coroner Department, Santa Ana, CA.
[c]FBI.
[d]Broward Sheriff's Office Crime Laboratory, Fort Lauderdale, FL.
[e]Georgia Bureau of Investigation, Decatur, GA.
[f]Illinois Department of State Police, Bureau of Forensic Sciences, Springfield, IL.
[g]Michigan State Police, East Lansing, MI—databases from Michigan and Atlanta, GA.
[h]Minnesota Bureau of Criminal Apprehension, St. Paul, MN.
[i]Washoe County Sheriff's Office Crime Laboratory, Reno, NV.
[j]Oregon State Police, Portland, OR.
[k]Vermont State Police, Waterbury, VT.
[l]Laboratoire de Police Scientifique, Montreal, Canada.
[m]Royal Canadian Mounted Police (RCMP) Central Forensic Laboratory, Ottawa, Ontario, Canada—Caucasian databases are from Kingston and Vancouver; American Indian databases are from Northern Ontario, British Columbia (Salishan), and Saskatchewan.
[n]Centre of Forensic Sciences, Toronto, Canada.
[o]Laboratoire de Génétique Moléculaire, Nantes, France.
[p]Institut für Rechtsmedizin, Bern, Switzerland.
[q]South Carolina State Police Law Enforcement Department (SLED), Columbia, SC.
[r]Arizona Department of Public Safety Crime Laboratory, Phoenix, AZ.
[s]Institute of Science and Forensic Medicine, Singapore.

TABLE 2—*Examples of operational constraints on frequency estimates of four locus profiles.*

	Specimen #F2390			
Locus	D1S7	D2S44	D4S139	D10S28
Base pair size	18144/18144	2731/1485	11035/7712	12545/2169
Fixed bin estimate[a]	1.000[b]	2(.041)(.124)	1.000[b]	1.000[b]

[a]Based on FBI Caucasian data, the frequency is approximately 0.010 (or 1/98).
[b]DNA profiles that contain an allele with size greater than 10,090 base pairs are assigned a frequency of 1.00 for the single locus profile.

southeastern Hispanic database was compared with the rebinned Michigan Hispanic database by scatter plot analysis.

It should be noted that some of the 1964 profiles contain information from fewer loci than others in the scatter plots. Additionally, because of the conventions described, some operational constraints were placed on the data. Single-locus profiles that contain alleles whose sizes fall outside the 640–10,090 bp range are not considered when deriving a multiple locus frequency estimate. Therefore, as shown in Table 2, while data for four loci may be available for an individual sample, fewer loci might be involved in the final estimate; so values such as 1 in 98 should not be construed as reflecting the relative rarity of four locus profiles. Furthermore, scatter plots generally do not separate out 1, 2, 3, and 4 locus target profiles; to appreciate the range of estimates for these different classes of target profiles the reader should refer to the compendium VNTR Population Study: A Worldwide Study [33] and Chakraborty et al. [34]. Table 3 provides a breakdown of the target profiles and the number of loci per population group (after adjusting for operational constraints) that were used for estimating DNA profile frequencies.

In an effort to display the data in an alternative manner, ratios of the compared estimates were made. Ratios were determined by dividing the more common frequency by the less common frequency in the designated reference databases for each of the 1964 target profiles.

Finally, because of the large volume of data, only representative examples of all the binned data and cross-group scatter plot comparisons that were performed are provided in this report. The comparisons shown in this study were not selected because they are the ones that support the authors' contentions. The results are similar for all comparisons. More data are available in a separate compendium (VNTR Population Study: A Worldwide Study [33]).

Results and Discussion

The VNTR loci studied were highly polymorphic in all databases described in Table 1. In accordance with previous studies [7,22,31,32] the data strongly support that multiple

TABLE 3—*The number of loci carried by each target profile per population group after adjusting for operational constraints for estimating DNA profile frequencies.[a]*

	1964 Hae III Target Profiles			
Number of Loci	Blacks	Caucasians	Hispanics	Total
0	5	3	1	9
1	93	210	22	325
2	105	115	90	310
3	154	260	258	672
4	160	220	268	648

[a]Operational constraints are described in the Materials and Methods section.

TABLE 4—Ratios (percentages by number of loci contained within a target profile) of the maximum to minimum frequency estimates of 1964 target DNA profiles from rebinned data from various sample populations containing at least 200 individuals (Hae III-based data).

Reference Populations	No. of Groups	No. of Loci[b]	Ratio Interval[a]					
			1	>1–2	>2–5	>5–10	>10–100	>100
Caucasian[c]	15	0 (N = 9)	100[i]	0	0	0	0	0
		1 (N = 325)	0	44.6	49.5	5.2	0.6	0
		2 (N = 310)	0	7.7	69.4	16.8	6.1	0
		3 (N = 672)	0	0.6	63.1	27.2	9.1	0
		4 (N = 648)	0	0.2	39.5	40.6	19.3	0.5
		All	0.5	8.9	53.8	26.2	10.5	0.1
Black[d]	7	0 (N = 9)	100[i]	0	0	0	0	0
		1 (N = 325)	0	61.8	35.7	2.2	0.3	0
		2 (N = 310)	0	29.4	62.9	6.5	1.3	0
		3 (N = 672)	0	14.9	69.8	13.7	1.6	0
		4 (N = 648)	0	6.8	66.8	22.7	3.7	0
		All	0.5	22.2	61.8	13.6	2.0	0
Hispanic[e]	4	0 (N = 9)	100[i]	0	0	0	0	0
		1 (N = 325)	0	63.7	35.1	0.9	0.3	0
		2 (N = 310)	0	35.5	57.4	6.1	1.0	0
		3 (N = 672)	0	22.3	61.8	14.3	1.6	0
		4 (N = 648)	0	15.9	57.6	19.9	6.6	0
		All	0.5	29.0	55.0	12.6	3.0	0
Caucasian Black[f]	22	All	0.5[j]	0.2	10.8	16.8	62.9	9.0
Caucasian Black Hispanic[g]	26	All	0.5[j]	0.2	9.0	14.1	64.7	11.6
Caucasian FBI Black[h]	16	All	0.5[j]	0.5	18.9	24.9	50.9	4.3
Black FBI Caucasian[i]	8	All	0.5[j]	2.0	26.3	31.7	36.5	3.1

[a]Ratios were determined by dividing the largest frequency by the smallest frequency for each of the 1,964 target profiles observed across the designated set of reference databases. Each ratio interval represents the magnitude of the ratio for each target profile. The percentages in each interval are the portion of target profiles within each ratio interval.

[b]For the within group Caucasian, Black and Hispanic comparisons the ranges have been broken down into the number of loci each target profile carries. Because of space limitations, in the across group comparisons the various target profile locus categories have been combined and are designated "All."

[c]Caucasian reference populations are from the FBI; Orange County Sheriff's Coroner Department, Santa Ana, CA; Broward Sheriff's Office Crime Laboratory, Fort Lauderdale, FL; Georgia Bureau of Investigation, Decatur, GA; Illinois Department of State Police, Bureau of Forensic Sciences, Springfield, IL; Michigan State Police, East Lansing, MI; Minnesota Bureau of Criminal Apprehension, St. Paul, MN; Washoe County Sheriff's Office Crime Laboratory, Reno, NV; Oregon State Police, Portland, OR; Vermont State Police, Waterbury, VT; Laboratoire de Police Scientifique, Montreal, Canada; RCMP Central Forensic Laboratory, Ottawa, Ontario, Canada (provided two Caucasian population databases); Centre of Forensic Sciences, Toronto, Canada; and Institut für Rechtsmedizin, Bern, Switzerland.

[d]Black reference populations are from the FBI; Orange County Sheriff's Coroner Department, Santa Ana, CA; Georgia Bureau of Investigation, Decatur, GA; Michigan State Police, East Lansing, MI (provided two databases—Michigan and Atlanta, GA); Minnesota Bureau of Criminal Apprehension, St. Paul, MN; and South Carolina State Police, SLED, Columbia, SC.

[e]Hispanic reference populations are from the FBI (provided two databases—southeastern and southwestern Hispanics); Arizona Department of Public Safety Crime Laboratory, Phoenix, AZ; and Orange County Sheriff's Coroner Department, Santa Ana, CA.

[f]Reference populations listed in c and d.

[g]Reference populations listed in c, d, and e.

[h]Reference populations listed in c and the Black reference database from the FBI.

[i]Reference populations listed in d and the Caucasian reference database from the FBI.

[j]Because of operational constraints placed upon deriving estimates of DNA profile frequencies 9 target profiles contained zero loci and thus would have a frequency of 1.00 in all databases.

locus VNTR DNA profiles are rare events in any relevant database. It has long been accepted that the greatest contribution to human diversity (approximately 85% based on protein markers) is due to variation among individuals; only a small contribution is due to racial and ethnic differences [35]. The observation of the rarity of multiple locus highly polymorphic VNTR profiles in all relevant databases further supports the position that most variation is due to differences among individuals. Therefore, the effect that different reference population groups might have on the estimate of the likelihood of occurrence of a DNA profile was considered.

The concern for the forensic community is not the *statistical* significance but rather the *forensic* significance of estimates of the likelihood of occurrence of DNA profiles when using various DNA databases. Again, differences in statistical estimates are deemed "forensically significant" when the various likelihood of occurrences of the DNA profile would be meaningfully different [31]. To appreciate the forensic effects on the differences of target profile estimates the reader should also refer to volume IV of VNTR Population Data: A Worldwide Study [33].

The likelihood of occurrence of each of the 1964 target DNA profiles was estimated using the rebinned format for each of the reference databases listed in Table 1. The ratio of the maximum to minimum values of the estimates (max/min) obtained from each of the major-population groups represented in the 1964 target profiles is shown in Table 4. The approach of evaluating max/min estimates presents a worst-case scenario, particularly because the allele frequencies are subject to sampling variances. Although it would be more appropriate to consider the ratio of the maximum to the mean value for forensic significance, the max/min approach was utilized to show extreme examples. Rebinning of the 31 bin format population data intentionally places a more conservative frequency estimate on a DNA profile when using a small-sized reference database [12]. The effect would be to increase the range of max/min values when estimates within small and large databases are compared. The increased conservatism of estimates from small-sized databases compared with larger-sized databases is appropriate for forensic applications, but would be misleading for evaluating the potential variations due to population substructure. Therefore, not all databases, only those with at least 200 individuals, were considered for max/min estimate comparisons. Order-of-magnitude or greater max/min ratios of frequencies less common than 1/1,000,000 (or for that matter 1/100,000) would not alter the forensic significance of the rarity of a DNA profile. One should not construe that 1/1,000,000 (or 1/100,000) is a dividing line for what should be considered as an acceptable likelihood of occurrence for a DNA profile; the values simply provide a point where differences among less common occurrences would have no consequence on the rarity of the DNA profile.

In Table 4, the Caucasian, Black and Hispanic within-reference group comparison data were divided further by the number of loci contained within each target profile. Of the 1964 target DNA profiles evaluated in the 15 different Caucasian databases, 10.6% of the target profiles had max/min ratios that exceeded an order of magnitude (Table 4). However, only 36 of these profiles (or a total of 1.8% of all target profiles) had minimum frequencies that were more common than 1/1,000,000. The max/min ratios of these 36 profiles ranged from 10-to-70.7-fold. It should be noted, however, that the max/min ratios are inflated by the 1964 target profiles selected for analysis. The target profiles were derived from Caucasians, Blacks, and Hispanics, with more than 40% of the profiles being Caucasian samples. However, only four of the 36 profiles, with max/min ratios greater than an order of magnitude and frequencies more common than 1/1,000,000, were profiles from Caucasian individuals. The four Caucasian profiles had max/min ratios ranging from 10.5-to-22.5-fold, and the most common frequency was 1/229,000 (a rare event in itself). Thus, for the Caucasian databases used in this study there appear to be no forensically significant differences for DNA profile frequency estimates. Any of the

Caucasian databases could be used to derive an estimate of the likelihood of occurrence of a multiple single-locus DNA profile and the inference of the rarity of the DNA profile would not change.

The observation was expected that target profiles from Caucasians would have very few differences for DNA profile frequency estimates in Caucasian databases, while Black and Hispanic target profiles would be the major contributors, 24 and eight, respectively, of max/min ratios that are greater than an order of magnitude and frequencies more common than 1/1,000,000. Generally, for considering the significance of DNA frequency estimates among various reference population samples the experiment would be designed such that only Caucasian target profiles would be evaluated among Caucasian databases (or Black target profiles among Black databases, etc.). Due to the potentially greater genetic differences among major-population groups than for within groups, a DNA profile that would be relatively more common, for example, in the Black population could be more rare in the Caucasian sample populations.

Additionally, max/min ratios of approximately two orders of magnitude were not observed unless the most common frequency was less than 1/1,000,000. Only 24 DNA profiles (1.2%) had max/min ratios in the two orders of magnitude range. Consistent with the stated observation that Caucasian target profiles are unlikely to exhibit substantial differences among various reference Caucasian databases, only three out of the 24 DNA profiles that had max/min ratios approaching two orders of magnitude were from Caucasians, even though Caucasian DNA profiles comprise more than 40% of the target DNA profiles used for analyses.

The same trends hold for the max/min ratios for the Hispanic and Black databases (Table 4). Only 59 out of the 1964 target DNA profiles (3.0%) had max/min ratios exceeding an order of magnitude among Hispanic databases containing at least 200 individuals. Of these 59 target DNA profiles, only nine profiles had frequencies more common than 1/1,000,000. The max/min ratios for these nine profiles were 10.1-to 14.4-fold. Only one out of the 1964 target DNA profiles had a max/min ratio approaching two orders of magnitude, with its most common frequency being 1/375,000,000. Four Hispanic databases, each with at least 200 individuals, were evaluated, in contrast with 15 databases for Caucasians. The larger number of Caucasian databases may contribute to the higher percentage of Caucasian profiles with max/min ratios exceeding an order of magnitude compared with Hispanics (Table 4). However, the Hispanic databases came from Florida, Texas, Arizona, and California. Southeastern and southwestern Hispanics have different racial admixture [36,37] and therefore, if subpopulation differences exist between these two groups the max/min ratios should increase. Thus, as with Caucasians, there are very few differences for forensic applications among the Hispanic sample populations.

There were seven Black databases with at least 200 individuals available for analysis. Again, minimal max/min ratio differences were observed. Only 40 out of 1964 target DNA profiles (2.0%) had max/min ratios exceeding an order of magnitude. Of these 40 target DNA profiles only five profiles had frequencies more common than 1/1,000,000, with max/min ratios from 10.2-to-25.4-fold. None of the 1964 max/min ratios approached two orders of magnitude. Two of these five target DNA profiles were from Black individuals with max/min ratios of 10.5- and 11.5-fold.

Lewontin and Hartl [1] and the NRC Report [2] assert that differences between major-population groups cannot be used to provide a meaningful bound on the variation of DNA profile frequency estimates for forensic purposes, because the genetic diversity between subgroups within a major-population group is greater than the genetic variation between major-population groups. If this assertion were true, then the addition of one Black database to the 15 Caucasian databases should augment to a degree the max/min ratios (but certainly not any more than that observed just for the 15 Caucasian databases).

However, this is not what occurs. For the Caucasian max/min ratios, 10.6% of the target profiles exceeded an order of magnitude; but, with the addition of only one Black database (that is, the FBI Black database) to the 15 Caucasian sample population databases, 55.2% of the max/min ratios exceeded an order of magnitude (Table 4). Likewise, 2% of the Black max/min ratios exceeded an order of magnitude; but with the addition of one Caucasian database (that is, the FBI Caucasian database) to the seven Black sample population databases, 39.6% of the max/min ratios exceeded an order of magnitude (Table 4).

Likewise, the max/min ratios show an even greater increase when all reference databases, regardless of major-population group, are compared. 10.6%, 3.0%, 2.0% of the Caucasian, Black, and Hispanic max/min ratios, respectively, exceeded an order of magnitude (Table 4). However, when the 22 Caucasian and Black reference populations were considered together, 71.9% of the target profiles had max/min ratios greater than an order of magnitude. When the four Hispanic databases were included also, 76.3% of the max/min ratios exceeded an order of magnitude (Table 4).

These max/min ratio data show that the present approach of using general population group databases for deriving meaningful DNA frequency estimates is appropriate. However, the data do overstate the range of differences that would be encountered. The range of estimates of a DNA profile across all relevant databases is not the most informative criterion for evaluating the differences in target profile rarity when using various reference populations. Comparisons of the likelihood of occurrence of DNA profiles in various individual reference populations are more useful for evaluating the effect on DNA statistical estimates when using a database which may not precisely represent the demographics of the region where the crime was committed [32]. Therefore, the FBI Caucasian and Black databases, which represent composite databases, were compared with several regional and ethnic databases. Obviously, the differences in these estimate comparisons must be less than the max/min ratios discussed above.

Figures 1 and 2 compare the probability of occurrence of the 1964 target RFLP profiles (from one to four VNTR locus profiles, from Table 1), estimated by the fixed bin method using various reference populations. The x and y axes of each scatter plot describe the reference population used for each scatter plot comparison. Each point represents a probability comparison for a DNA profile of each of the 1964 individuals (defined as either Black, Caucasian, southeastern Hispanic, or southwestern Hispanic) in the FBI population databases. Thus, real DNA profiles were used to evaluate the effect of estimating the likelihood of occurrence of each DNA profile in different databases. The diagonal on each scatter plot indicates the theoretical line where both reference databases would produce the same estimate. Generally, when the data points fall close to the diagonal line, the databases would yield similar results.

Furthermore, points that fall closer to the origin tend to represent a single-locus profile, while those furthest from the origin tend to represent four single-locus profiles. Scatter plots that contain incomplete sample profiles can be informative in depicting forensic significance for situations encountered in casework: even though a four- or five-probe battery may be available to the forensic scientist for analyses, the quality and/or quantity of the DNA derived from the forensic specimen may preclude typing of all available loci. Figure 2 shows two representative examples of the rarity of one, two, three, and four locus target profiles on scatter plots.

The variation in breadth of the scatter plots is different for within-major population group comparisons (for example, FBI Caucasian vs. Swiss or Japanese vs. Chinese; Tables 5 to 10, Figs. 1 and 2) and between group comparisons (for example, FBI Black vs. FBI Caucasian or FBI Black vs. Japanese; Table 11, Figure 1). For Hae III Caucasian population sample databases, FBI Caucasian frequencies were compared with several of the United States Caucasian databases displayed in Table 1a (Oregon, California, Florida,
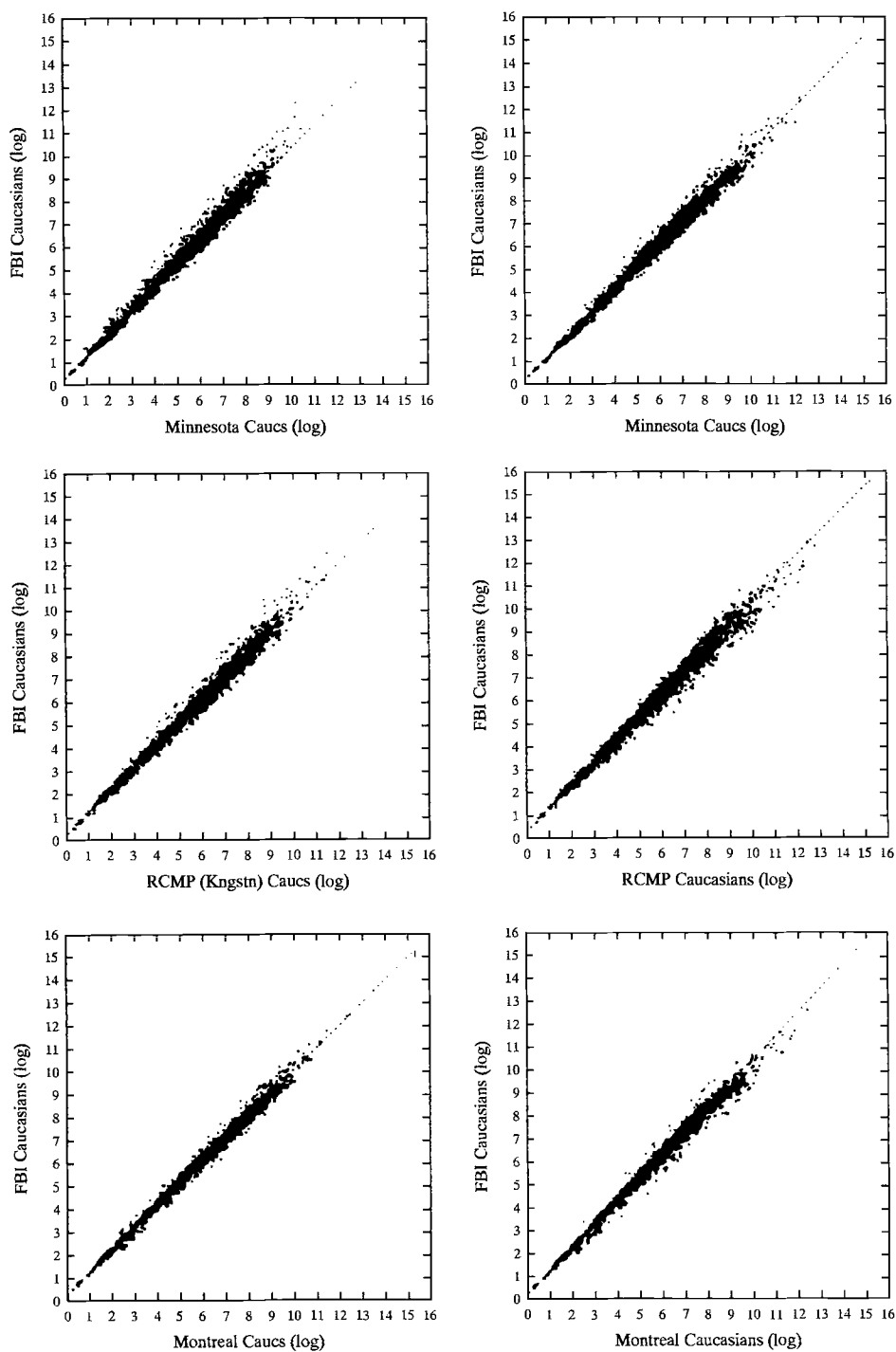
FIG. 1—*Examples of scatter plot comparisons of various reference population databases. The likelihood of occurrence of 1964 Hae III-generated target RFLP profiles (from one to four VNTR locus profiles) was estimated by the fixed bin method using various reference populations. The x and y axes of each scatter plot are labeled with each reference population used in a comparison. The first column of scatter plots displays comparisons using rebinned data; the second column of scatter plots displays comparisons using 31 bin data; and the third column of scatter plots displays comparisons using random sampling/rebinned data.*
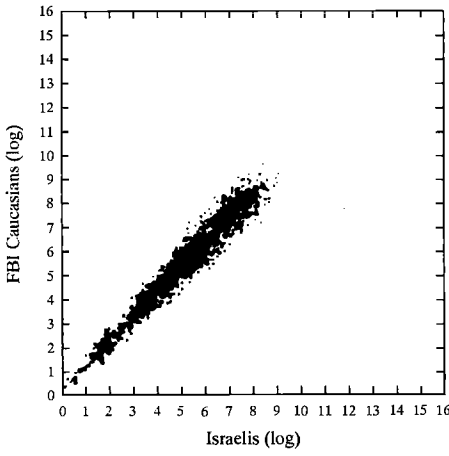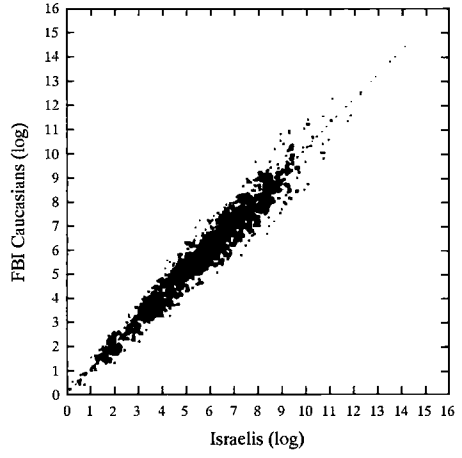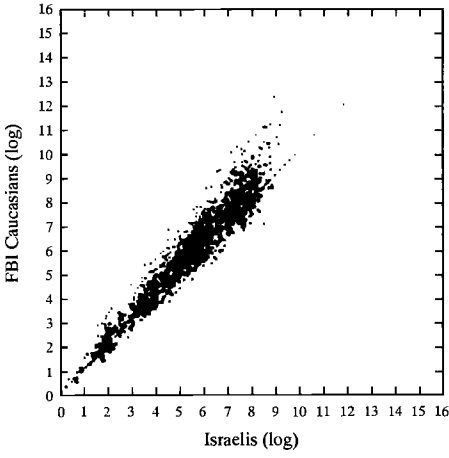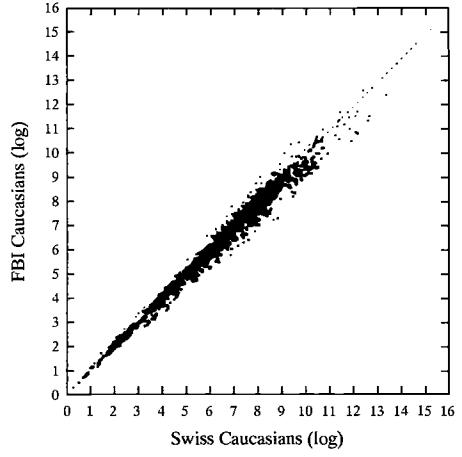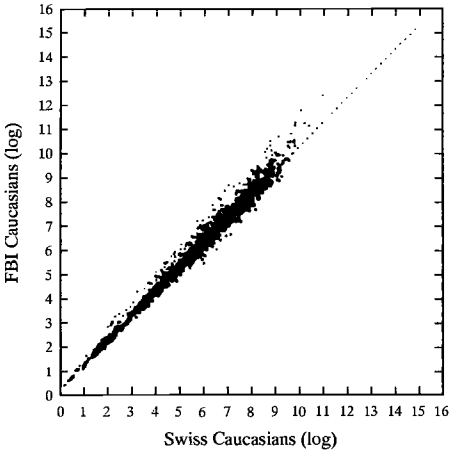
329
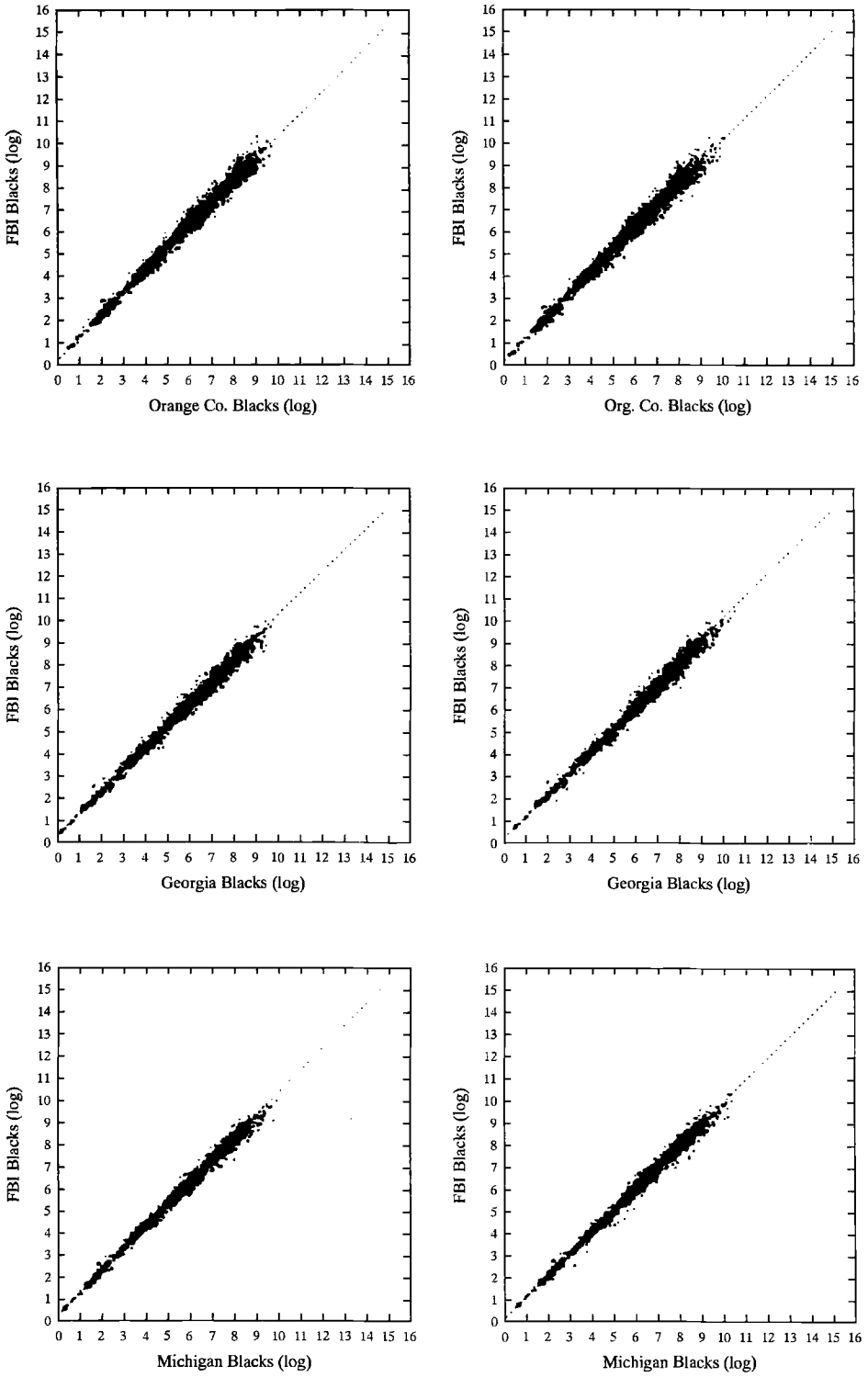
FIG. 1—*Continued.*
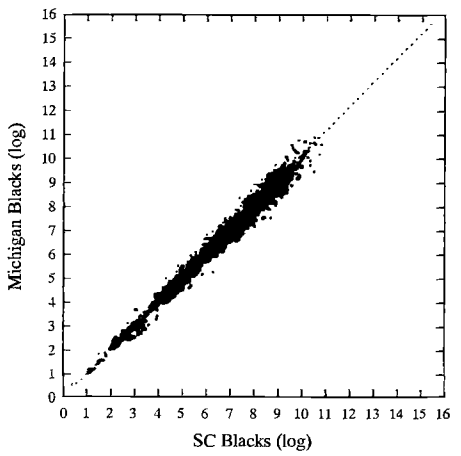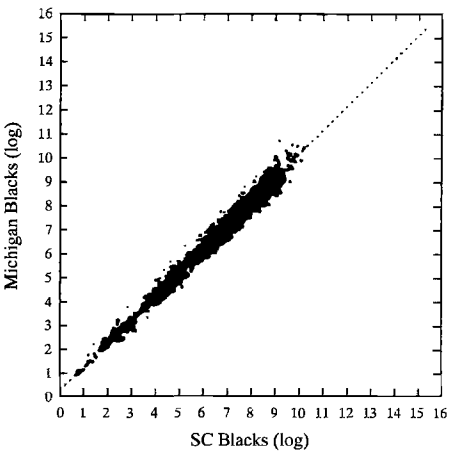
FIG. 1—*Continued.*
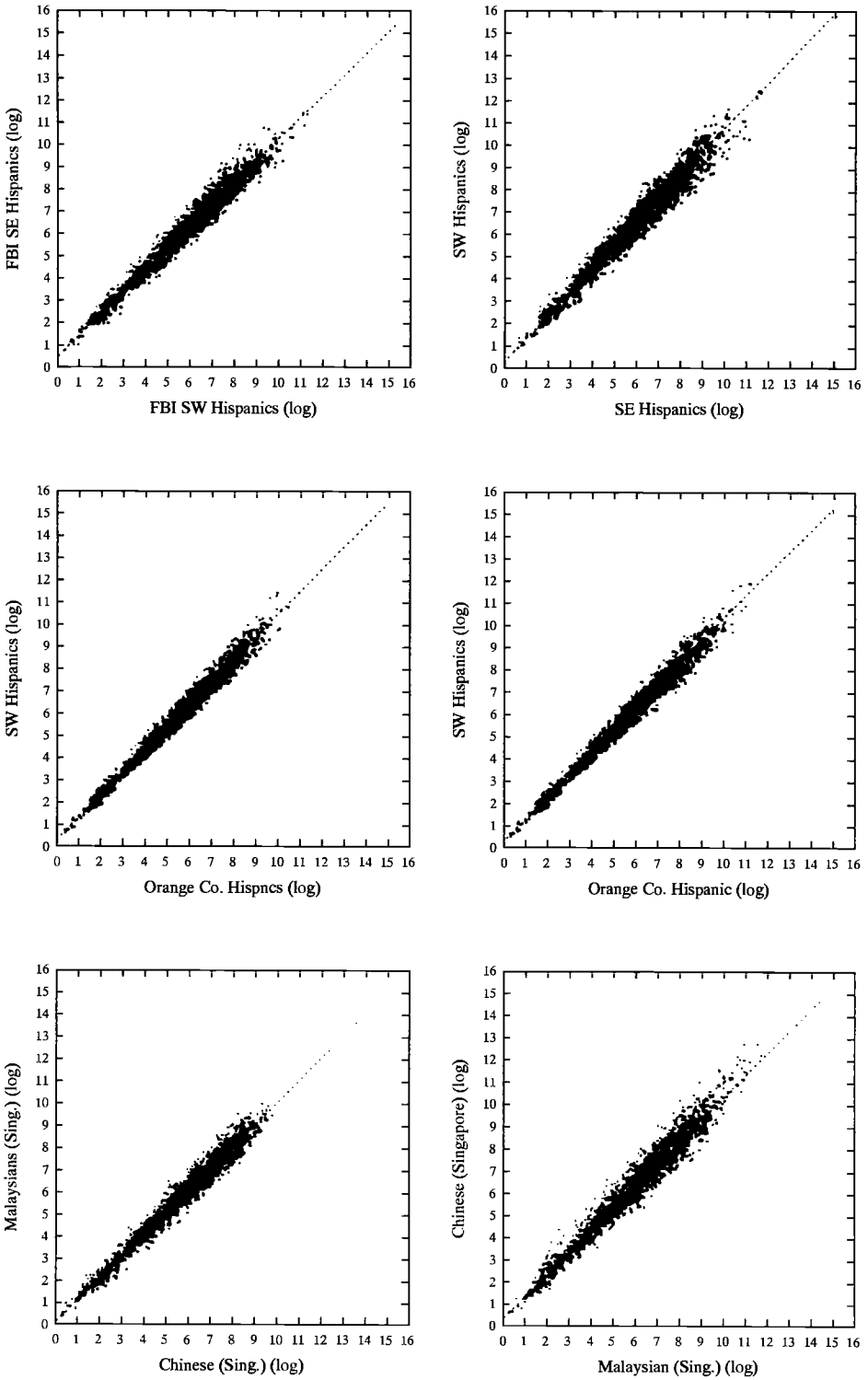
FIG. 1—*Continued.*
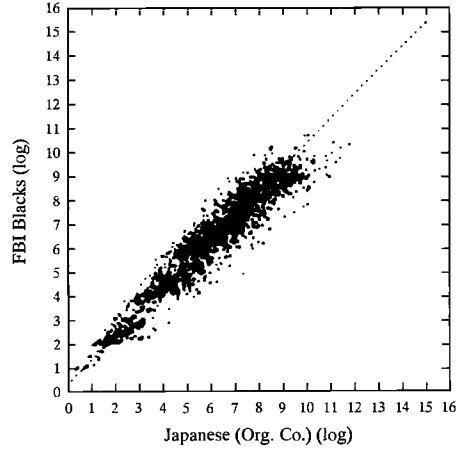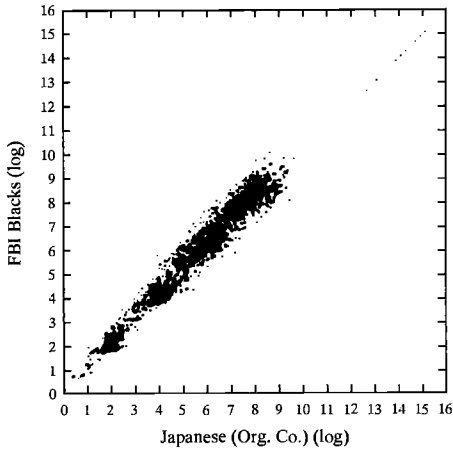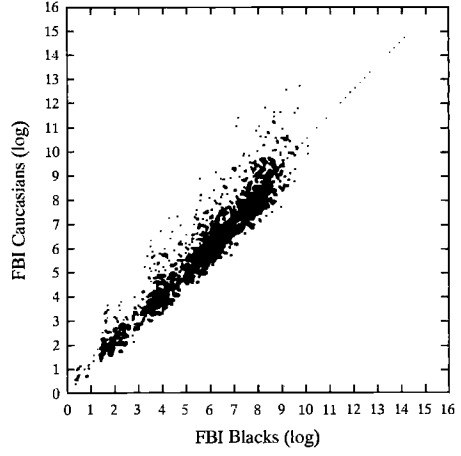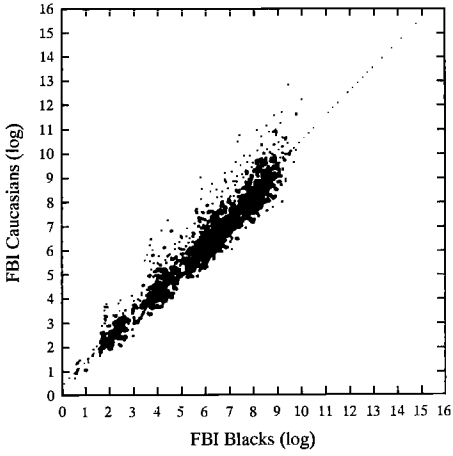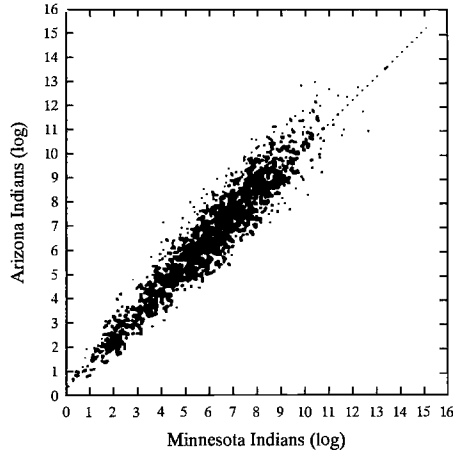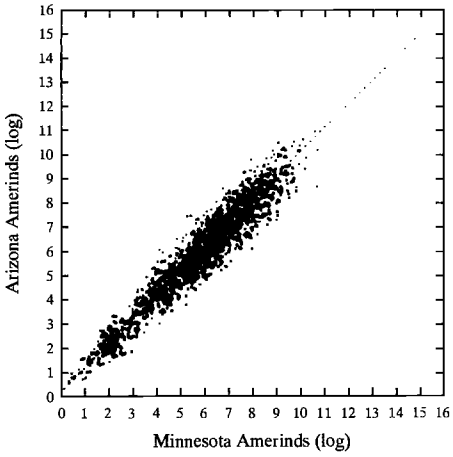
FIG. 1—*Continued.*
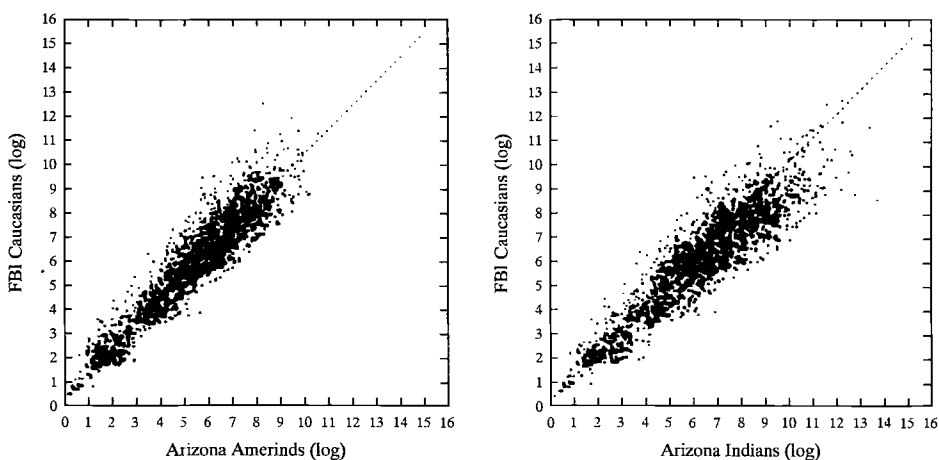
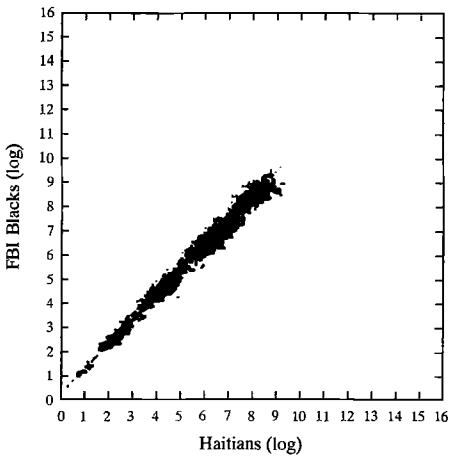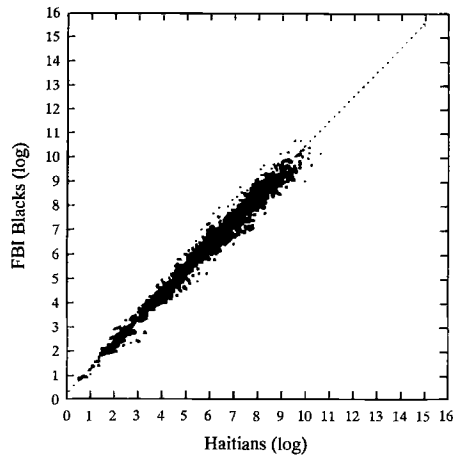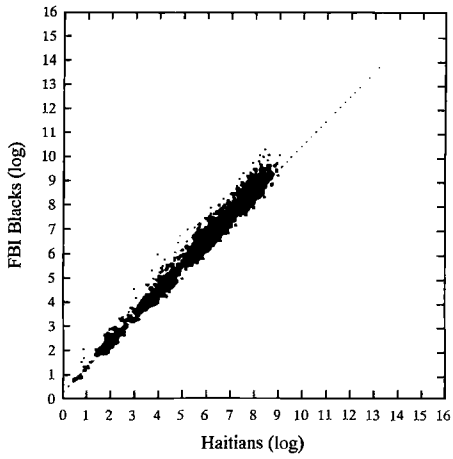FIG. 1—*Continued.*

FIG. 1—*Continued.*

Georgia, Minnesota, and Vermont) to determine whether or not there are any substantial differences for US regional databases, as well as with a French (Montreal) and an English Canadian database (Kingston), one Swiss database, and an Israeli database (Table 5, Fig. 1). There do not appear to be substantial differences among statistical estimates using different Caucasian databases. The breadth of the scatter plots is narrow. Using the rebinned data formats, 95.7%–99.9% of the 1964 target DNA profile frequency estimates from the FBI Caucasian sample population compared with regional/ethnic Caucasian databases were within an order of magnitude (excluding the French and Israeli samples). For those scatter plot Caucasian population comparisons with databases of more similar size (that is, FBI Caucasian vs. either English Canadian or French Canadian or Swiss), 98.8%–99.9% of the frequency estimate comparisons, with rebinned formatted data, were within an order of magnitude. Only one target DNA profile of the larger database comparisons, FBI Caucasian vs. English Canadians, had an estimate that exceeded an order of magnitude and a frequency more common than 1/1,000,000; however, it was not a Caucasian target profile.

Due to an intentionally designed characteristic of the fixed bin method, cross-group comparisons may not fall on the diagonal even if the population databases are similar. Bins with fewer than five counts are merged with contiguous bins; therefore, smaller databases generally will have fewer bins and larger databases will have more bins. The intent of merging bins is to eliminate the possibility of artificially small allele probabilities when smaller databases are used [12,32]. Another effect of merging bins for small databases is that bin frequencies that are truly small (as observed in the larger databases) will be overestimated. Thus, when comparing databases with scatter plots, the data points may cluster above or below the diagonal depending on which database has the greater number of samples (for example, FBI Caucasian vs. Israelis, Table 1 and Fig. 1).

Therefore, when comparing statistical estimates with different sized databases using the fixed bin method, it may be necessary to evaluate the breadth of the data point cluster by scatter plots of probability estimates calculated using an alternate method. If some population data are sorted into the original 31 bins rather than rebinned categories, the additional conservative bias placed on smaller-sized databases by rebinning for use in forensic casework analyses for frequency estimations then would not confound the interpretation of the data with differences that might be ascribable to subgroups. Alterna-
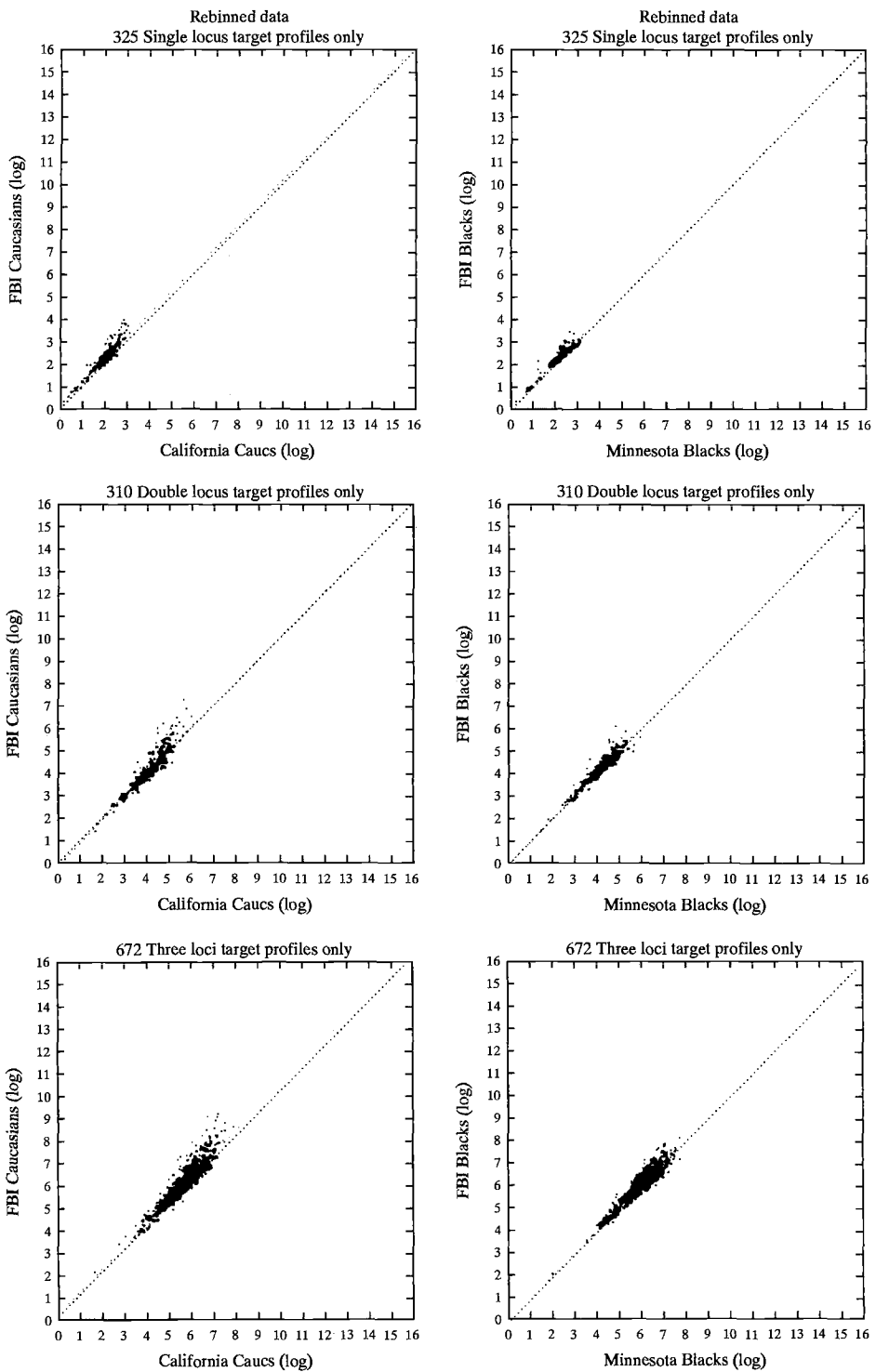
FIG. 2—*Examples of scatter plot comparisons for target profiles consisting of 1, 2, 3, or loci. The likelihood of occurrence of 1964 Hae III-generated target RFLP profiles was estimated by the fixed bin method. The x and y axes of each scatter plot are labeled with each reference population used in a comparison. The first column of scatter plots displays comparisons using rebinned data of FBI Caucasians v. California Caucasians and the second column of scatter plots displays comparisons using rebinned data of the FBI Blacks v. Minnesota Blacks. The first row of scatter plots displays target profiles containing only one locus, while the second, third and fourth rows of scatter plots display target profiles containing only two, only three, and only four loci, respectively.*

336
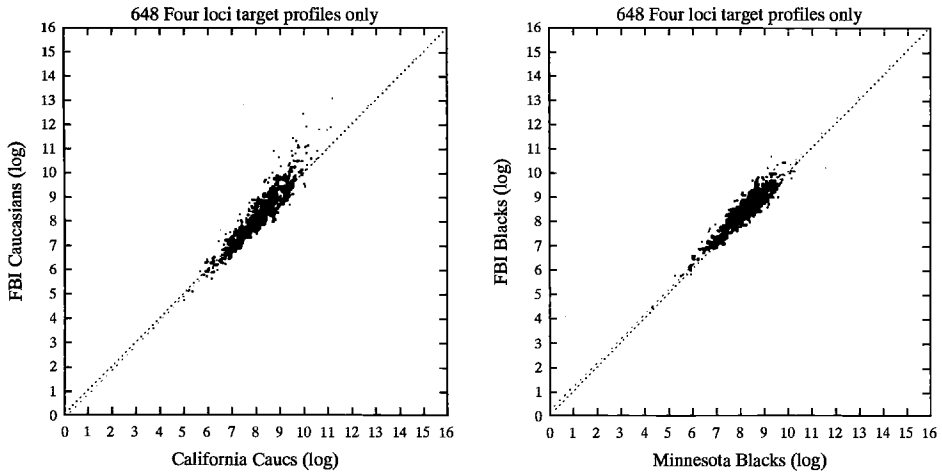
FIG. 2—*Continued.*

tively, the random sampling approach of the larger database can be used to evaluate forensic significance (Fig. 1).

When comparing within major group databases of different size, the sorting of data into 31 bin categories can reduce the variation in statistical estimates between the two sample populations, compared with using rebinned data (Table 5). Compared with the FBI Caucasian sample population, California, Florida, Georgia, Minnesota, Oregon, Vermont, French, and Israeli Caucasian databases have considerably smaller sample sizes. Therefore, to determine whether any substantial differences in estimates would be encountered using these various databases, either 31 bin data or the random sampling of the larger database should be evaluated. As shown in Table 5, 99.6%–99.9% of the regional U.S. comparisons are within an order of magnitude when 31 bin data are used. Of these regional U.S. 31 bin Caucasian data comparisons, only one target profile, that had one order of magnitude difference in the estimate and had a frequency more common than 1/1,000,000, was from a Caucasian individual (FBI Caucasian vs. Minnesota with an 11.4 fold difference). If within-major group databases are more equivalent in size, the 31 bin format and rebinned data generally yield a more similar range of estimates (if anything, the 31 bin formatted data should increase the range of differences slightly, particularly for rare events, and these differences would not alter the fact that a DNA profile was rare) (see FBI Caucasian vs. English Canadians; FBI Caucasians vs. Swiss; FBI Caucasians vs. French Canadians; Table 5). Because deviations based on ratios are going to show a large variance due to sampling, the very few observed differences are extreme examples. Therefore, there generally are no real differences for forensic purposes among these databases.

The French and Israeli (as well as the Swiss and Canadian, previously) databases should not be considered as relevant databases for the population of potential perpetrators in the United States. However, under the highly unlikely assumption of no gene flow among subgroups in the United States, these population groups can be used to gain insight of the effects of subgroups on forensic DNA statistical estimates. Because of the much smaller size of the French Caucasian database, the random sampling approach was used. When sample size was taken into consideration in this way, 99.8% of the comparisons with the FBI Caucasian database were within an order of magnitude. There were only two target DNA profiles that had frequency estimate differences exceeding an order of magnitude and were more common than 1/1,000,000; one of the two target

TABLE 5—Distribution (percentages) of ratios of frequency estimates calculated in various pairs of Caucasian reference population (Hue III-based data).[a]

Reference Population Comparison	Bin Fmt[c]	Ratio Interval[b]					
		1	>1-2	>2-5	>5-10	>10-100	>100
FBI[d] v. California[e]	R	0.8	69.9	21.3	5.7	2.2	0.1
	31	1.0	77.3	19.5	1.8	0.4	0
FBI[d] v. Florida[f]	R	0.9	73.8	20.4	3.1	1.9	0
	31	0.9	78.6	18.6	1.7	0.3	0
FBI[d] v. Georgia[g]	R	1.3	77.9	19.5	1.3	0.1	0
	31	1.2	79.7	18.2	0.8	0.1	0
FBI[d] v. Minnesota[h]	R	0.7	73.1	21.1	3.6	1.6	0
	31	0.8	78.1	19.3	1.6	0.2	0
FBI[d] v. Oregon[i]	R	0.7	74.1	21.1	2.9	1.2	0.1
	31	0.7	76.8	20.6	1.7	0.2	0
FBI[d] v. Vermont[j]	R	0.8	70.7	19.5	4.7	4.3	0
	31	0.8	75.8	19.8	3.4	0.2	0
FBI[d] v. English Canadian[k]	R	1.1	74.3	21.4	2.2	1.0	0
	31	0.6	74.3	21.2	2.6	1.3	0
FBI[d] v. French Canadian[l]	R	0.9	86.3	12.2	0.6	0.1	0
	31	0.9	81.0	15.1	2.2	0.8	0
FBI[d] v. Swiss[m]	R	1.7	77.3	16.6	3.1	1.2	0
	31	1.5	77.5	16.6	3.2	1.2	0
FBI[d] v. French[n]	R	1.2	52.8	30.2	9.1	6.2	0.5
	31	1.1	58.1	31.1	7.7	2.0	0
	RS/R	0.7	67.6	28.3	3.3	0.2	0
FBI[d] v. Israeli[o]	R	0.5	38.6	33.2	13.0	13.4	1.3
	31	0.5	48.3	35.4	10.7	5.1	0
	RS/R	0.5	53.1	34.4	9.0	3.1	0

[a]The databases for this table were selected from a larger collection of databases contained within the compendium VNTR Population Study: A Worldwide Study (1993). The databases were chosen to display the effects of databases from various geographic US regions and several international databases compared with the FBI Caucasian database. Although there were three English Canadian databases, one was chosen arbitrarily for this table.

[b]Ratios were determined by dividing the more common frequency by the less common frequency in the designated reference databases for each of the 1,964 target profiles. Each ratio interval represents the magnitude of the ratio for each target profile. The percentages are the portion of target profiles falling within each ratio interval.

[c]Bin Formats (Fmt) are: R = rebinned data; 31 = 31 bin data; and RS/R = rebinned random sampling data of the larger-sized reference population compared with rebinned data of the smaller-sized reference population.

[d]Reference population from the FBI.

[e]Reference population from Orange County Sheriff's Coroner Department, Santa Ana, CA.

[f]Reference population from Broward Sheriff's Office Crime Laboratory, Fort Lauderdale, FL.

[g]Reference population from Georgia Bureau of Investigation, Decatur, GA.

[h]Reference population from Minnesota Bureau of Criminal Apprehension, St. Paul, MN.

[i]Reference population from Oregon State Police, Portland, OR.

[j]Reference population from Vermont State Police, Waterbury, VT.

[k]Reference population from RCMP Central Forensic Laboratory, Ottawa, Ontario, Canada.

[l]Reference population from Laboratoire de Police Scientifique, Montreal, Canada.

[m]Reference population from Institut für Rechtsmedizin, Bern, Switzerland.

[n]Reference population from Laboratoire de Génétique Moléculaire, Nantes, France.

[o]Reference population from the FBI.

TABLE 6—Distribution (percentages) of ratios of frequency estimates calculated in various pairs of Black reference populations (Hae III-based data).[a]

Reference Population Comparison	Bin Fmt[c]	Ratio Interval[b]					
		1	>1–2	>2–5	>5–10	>10–100	>100
FBI[d] v. California[e]	R	0.5	75.9	21.4	2.2	0.1	0
	31	0.6	76.0	21.3	2.0	0.2	0
FBI[d] v. Florida[f]	R	0.9	70.4	24.6	3.6	0.5	0
	31	1.0	70.5	24.6	3.5	0.5	0
FBI[d] v. Georgia[g]	R	0.9	83.9	14.5	0.7	0.1	0
	31	1.1	83.0	15.1	0.7	0.1	0
FBI[d] v. Minnesota[h]	R	0.6	78.1	19.7	1.5	0.1	0
	31	0.6	77.6	18.8	2.3	0.6	0
FBI[d] v. Michigan[i]	R	1.3	87.8	10.5	0.3	0.1	0
	31	1.3	87.4	10.4	0.9	0.1	0
Michigan[i] v. South Carolina[j]	R	0.8	70.1	27.1	1.9	0.1	0
	31	0.9	68.4	28.0	2.4	0.4	0
FBI[d] v. Haitian[k]	R	0.7	66.8	24.7	5.6	2.2	0
	31	1.0	74.5	20.8	2.9	0.8	0
	RS/R	0.9	72.4	24.7	1.9	0.2	0

[a]The databases for this table were selected from a larger collection of databases contained within the compendium VNTR Population Study: A Worldwide Study (1993). The databases were chosen to display the effects of databases from various geographic US regions compared with the FBI Black database. The Michigan vs. South Carolina comparison was included to display the effects of using a northern or a southern US database.

[b]Ratios were determined by dividing the more common frequency by the less common frequency in the designated reference databases for each of the 1,964 target profiles. Each ratio interval represents the magnitude of the ratio for each target profile. The percentages are the portion of target profiles falling within each ratio interval.

[c]Bin Formats (Fmt) are: R = rebinned data; 31 = 31 bin data; and RS/R = rebinned random sampling data of the larger-sized reference population compared with rebinned data of the smaller-sized reference population.

[d]Reference population from the FBI.

[e]Reference population from Orange County Sheriff's Coroner Department, Santa Ana, CA.

[f]Reference population from Broward Sheriff's Office Crime Laboratory, Fort Lauderdale, FL.

[g]Reference population from Georgia Bureau of Investigation, Decatur, GA.

[h]Reference population from Minnesota Bureau of Criminal Apprehension, St. Paul, MN.

[i]Reference population from Michigan State Police, East Lansing, MI.

[j]Reference population from South Carolina State Police, SLED, Columbia, SC.

[k]Reference population from the FBI.

TABLE 7—Distribution (percentages) of ratios of frequency estimates calculated in various pairs of Hispanic reference populations (Hae III-based data).[a]

Reference Population Comparison	Bin Fmt[c]	Ratio Interval[b]					
		1	>1–2	>2–5	>5–10	>10–100	>100
FBI Southeast[d] v. FBI Southwest[e]	R	0.8	58.5	33.9	5.5	1.4	0
	31	0.6	53.3	35.8	7.8	2.6	0
FBI Southwest[e] v. California[f]	R	0.8	69.0	27.3	2.8	0.1	0
	31	0.7	67.6	28.6	2.8	0.4	0
FBI Southeast[d] v. California[f]	R	0.7	62.7	31.0	4.5	1.1	0
	31	0.8	58.1	31.6	7.0	2.5	0
FBI Southwest[e] v. Michigan[g]	R	0.5	46.6	33.0	12.0	7.7	0.2
	31	0.9	60.4	31.1	5.5	2.1	0
FBI Southeast[d] v. Michigan[g]	R	0.6	42.7	34.0	12.8	9.7	0.3
	31	0.6	51.8	34.2	9.1	4.4	0
	RS/R	0.6	60.1	31.0	6.3	2.1	0

[a]The databases for this table were selected from a larger collection of databases contained within the compendium VNTR Population Study: A Worldwide Study (1993). The databases were chosen to display the effects of databases from various geographic US regions compared with the FBI Hispanic databases.
[b]Ratios were determined by dividing the more common frequency by the less common frequency in the designated reference databases for each of the 1,964 target profiles. Each ratio interval represents the magnitude of the ratio for each target profile. The percentages are the portion of target profiles falling within each ratio interval.
[c]Bin Formats (Fmt) are: R = rebinned data; 31 = 31 bin data; and RS/R = rebinned random sampling data of the larger-sized reference population compared with rebinned data of the smaller-sized reference population.
[d]Reference population from the FBI.
[e]Reference population from the FBI.
[f]Reference population from Orange County Sheriff's Coroner Department, Santa Ana, CA.
[g]Reference population from Michigan State Police, East Lansing, MI.

TABLE 8—Distribution (percentages) of ratios of frequency estimates calculated in various pairs of Oriental and Asian Indian reference populations (Hae III-based data).[a]

Reference Population Comparison[b]	Bin Fmt[d]	Ratio Interval[c]					
		1	>1-2	>2-5	>5-10	>10-100	>100
Chinese[e] v. Japanese[e]	R	0.8	59.7	31.4	6.6	1.5	0
	31	0.8	54.4	33.1	8.7	3.1	0
Chinese[e] v. Malaysian[g]	R	1.0	60.4	31.7	5.7	1.2	0
	31	1.0	49.3	33.4	10.9	5.4	0
Asian Indian[h] v. Asian Indian[i]	R	0.9	61.8	29.3	6.7	1.3	0
	31	0.8	63.7	29.3	5.3	0.9	0

[a]The databases for this table were selected from a larger collection of databases contained within the compendium VNTR Population Study: A Worldwide Study (1993). Only a few database comparisons were chosen arbitrarily to demonstrate effects.

[b]Note that more Oriental databases are available in VNTR Population Data: A Worldwide Study (1993).

[c]Ratios were determined by dividing the more common frequency by the less common frequency in the designated reference databases for each of the 1,964 target profiles. Each ratio interval represents the magnitude of the ratio for each target profile. The percentages are the portion of target profiles falling within each ratio interval.

[d]Bin Formats (Fmt) are: R = rebinned data; 31 = 31 bin data.

[e]Reference population from Orange County Sheriff's Coroner Department, Santa Ana, CA.

[f]Reference population from Orange County Sheriff's Coroner Department, Santa Ana, CA.

[g]Reference population from Institute of Science and Forensic Medicine, Singapore.

[h]Reference population from Institute of Science and Forensic Medicine, Singapore.

[i]Reference population from RCMP Central Forensic Laboratory, Ottawa, Ontario, Canada.

TABLE 9—Distribution (percentages) of ratios of frequency estimates calculated in various pairs of Native American Indian reference populations (Hae III-based data).[a]

Reference Population Comparison	Bin Fmt[c]	Ratio Interval[b]					
		1	>1–2	>2–5	>5–10	>10–100	>100
Arizona[d] v. Saskatchewan[e]	R	0.6	33.5	33.1	17.4	14.7	0.8
	31	0.6	31.5	33.7	16.2	16.5	1.5
Arizona[d] v. Salishan[f]	R	0.5	29.4	29.5	16.2	22.0	2.3
	31	0.7	26.5	28.6	16.6	24.0	3.6
Arizona[d] v. Northern Ontario[g]	R	1.0	33.2	31.9	15.5	17.0	1.3
	31	0.7	25.9	28.2	15.0	23.4	6.8
Arizona[d] v. Minnesota[h]	R	0.5	38.1	36.1	14.7	10.4	0.2
	31	0.6	30.2	32.0	17.1	18.4	1.6

[a]The databases for this table were selected from a larger collection of databases contained within the compendium VNTR Population Study: A Worldwide Study (1993). Only a few databases were chosen arbitrarily to demonstrate effects compared with Arizona Indians. Note that more American Indian databases are available in VNTR Population Data: A Worldwide Study (1993).

[b]Ratios were determined by dividing the more common frequency by the less common frequency in the designated reference databases for each of the 1,964 target profiles. Each ratio interval represents the magnitude of the ratio for each target profile. The percentages are the portion of target profiles falling within each ratio interval.

[c]Bin Formats (Fmt) are: R = rebinned data; 31 = 31 bin data.

[d]Reference population from Arizona Department of Public Safety Crime Laboratory, Phoenix, AZ.

[e]Reference population from RCMP Central Forensic Laboratory, Ottawa, Ontario, Canada.

[f]Reference population from RCMP Central Forensic Laboratory, Ottawa, Ontario, Canada.

[g]Reference population from RCMP Central Forensic Laboratory, Ottawa, Ontario, Canada.

[h]Reference population from Minnesota Bureau of Criminal Apprehension, St. Paul, MN.

TABLE 10—Distribution (percentages) of ratios of frequency estimates in various reference within group population comparisons, based on number of loci in each target profile (Hae III-based data).[a]

Reference Populations	Bin Fmt[c]	No. of Loci[d]	Ratio Interval[b]					
			1	>1–2	>2–5	>5–10	>10–100	>100
FBI Caucasian v. CA Caucasian	R	0	100[e]	0	0	0	0	0
	31	0	100[e]	0	0	0	0	0
	R	1	0	91.1	6.5	2.5	0	0
	31	1	0	95.7	4.3	0	0	0
	R	2	1.0	74.2	17.1	5.5	2.3	0
	31	2	1.0	82.3	15.2	1.3	0.3	0
	R	3	0.4	69.2	23.4	4.9	2.1	0
	31	3	0.7	76.6	21.1	1.2	0.3	0
	R	4	0.3	58.8	28.9	8.3	3.4	0
	31	4	0.5	67.6	27.6	3.6	0.8	0.3
FBI Black v. MN Black	R	0	100[e]	0	0	0	0	0
	31	0	100[f]	0	0	0	0	0
	R	1	0	94.2	5.2	0.6	0	0
	31	1	0	94.2	4.0	1.5	0.3	0
	R	2	0	86.8	11.6	1.3	0.3	0
	31	2	0	85.5	11.6	2.6	0.3	0
	R	3	0.4	75.7	22.8	1.0	0	0
	31	3	0.3	74.7	21.9	2.5	0.6	0
	R	4	0	69.4	27.9	2.5	0.2	0
	31	4	0	69.8	26.9	2.5	0.9	0
SE Hispanic v. SW Hispanic	R	0	100[e]	0	0	0	0	0
	31	0	100[f]	0	0	0	0	0
	R	1	0.3	80.9	17.8	0.9	0	0
	31	1	0.3	75.7	21.5	2.5	0	0
	R	2	0.6	66.5	30.0	2.9	0	0
	31	2	0	57.7	35.8	5.2	1.3	0
	R	3	0	54.9	37.9	6.6	0.6	0
	31	3	0	51.2	37.8	9.2	1.8	0
	R	4	0.5	48.0	40.0	8.0	3.6	0
	31	4	0.2	42.7	41.4	10.3	5.4	0

Comparison	Fmt	No. loci						
Chinese v. Japanese	R	0	100^e	0	0	0	0	0
	31	0	100^e	0	0	0	0	0
	R	1	0.9	88.9	10.2	0	0	0
	31	1	1.2	82.5	16.0	0.3	0	0
	R	2	0	70.0	26.8	2.6	0.7	0
	31	2	0	64.8	29.4	4.8	1.0	0
	R	3	0.3	52.7	39.3	6.7	1.0	0
	31	3	0.3	49.7	37.9	9.1	3.0	0
	R	4	0.2	48.3	36.6	11.9	3.1	0
	31	4	0	40.9	38.9	14.5	5.7	0
AZ Indian v. N. Ont. Indian	R	0	100^e	0	0	0	0	0
	31	0	100^e	0	0	0	0	0
	R	1	3.1	55.7	35.7	5.5	0	0
	31	1	0.6	45.8	40.3	12.3	0.9	0
	R	2	0	40.3	35.5	13.9	10.3	0
	31	2	0	30.0	36.1	12.6	19.4	1.9
	R	3	0	27.7	32.4	18.2	20.1	1.6
	31	3	0.2	21.7	25.1	17.9	27.7	7.4
	R	4	0.2	24.8	28.2	18.8	25.8	2.2
	31	4	0.3	18.5	21.9	14.8	32.4	12.0

[a]The reference population comparisons were selected from Tables 4–9. Because of space limitations, only one example per table is displayed.
[b]Ratios were determined by dividing the largest frequency by the smallest frequency for each of the 1,964 target profiles observed across the designated set of reference databases. Each ratio interval represents the magnitude of the ratio for each target profile. The percentages in each interval are the portion of target profiles within each ratio interval.
[c]Bin Formats (Fmt) are: R = rebinned data; 31 = 31 bin data.
[d]The column has been broken down into the number of loci each target profile carries. There were 9 zero locus target profiles, 325 one locus target profiles, 310 two locus target profiles, 672 three locus profiles, and 648 four locus target profiles.
[e]Because of operational constraints placed upon deriving estimates of DNA profile frequencies 9 target profiles contained zero loci and thus would have a frequency of 1.00 in all databases.

TABLE 11—Distribution (percentages) of ratios of frequency estimates calculated in various pairs of cross-major group reference populations (Hae III-based data).[a]

Reference Population Comparison	Bin Fmt[c]	Ratio Interval[b]					
		1	>1-2	>2-5	>5-10	>10-100	>100
FBI Caucasian[d] v. FBI Black[e]	R	0.6	39.5	37.8	13.4	8.0	0.7
	31	0.6	37.9	37.0	13.6	9.4	1.5
FBI Caucasian[d] v. Chinese[f]	R	0.6	36.9	36.2	14.0	11.8	0.5
	31	0.5	33.4	31.9	13.0	18.4	2.9
FBI Caucasian[d] v. Japanese[g]	R	0.5	34.5	32.6	14.6	15.9	1.9
	31	0.5	34.4	33.0	15.8	15.3	1.1
FBI Black[e] v. Japanese[g]	R	0.5	37.0	35.5	15.7	11.3	0
	31	0.6	37.6	34.6	14.2	12.0	1.0
FBI Caucasian[d] v Asian Indian[h]	R	0.5	41.3	34.6	11.7	11.2	0.7
	31	0.5	43.8	36.2	12.5	6.9	0.1
FBI Caucasian[d] v. Arizona Amerindian[i]	R	0.5	30.5	31.4	17.8	17.8	2.0
	31	0.5	26.3	22.9	15.2	27.8	7.4
FBI Caucasian[d] v. Minnesota Amerindian[j]	R	0.6	38.3	36.3	13.2	10.8	0.7
	31	0.6	36.9	35.7	13.8	12.3	0.7

[a]The databases for this table were selected from a larger collection of databases contained within the compendium VNTR Population Study: A Worldwide Study (1993). Only a few databases were chosen arbitrarily to demonstrate effects across databases from different major population groups compared with FBI databases. Note that more databases are available in VNTR Population Data: A Worldwide Study (1993).

[b]Ratios were determined by dividing the more common frequency by the less common frequency in the designated reference databases for each of the 1,964 target profiles. Each ratio interval represents the magnitude of the ratio for each target profile. The percentages are the portion of target profiles falling within each ratio interval.

[c]Bin Formats (Fmt) are: R = rebinned data; 31 = 31 bin data.

[d]Reference population from the FBI.

[e]Reference population from the FBI.

[f]Reference population from Institute of Science and Forensic Medicine, Singapore.

[g]Reference population from Orange County Sheriff's Coroner Department, Santa Ana, CA.

[h]Reference population from Institute of Science and Forensic Medicine, Singapore.

[i]Reference population from Arizona Department of Public Safety Crime Laboratory, Phoenix, AZ.

[j]Reference population from Minnesota Bureau of Criminal Apprehension, St. Paul, MN.

profiles was from a Caucasian and the difference was 10.9-fold. The slightly wider breadth of the scatter plot in the FBI Caucasian vs. Israeli comparison was anticipated due to the genetic differences between the two groups; however, the estimates still did not produce substantial differences. 94.9% of the comparisons between the FBI Caucasian and Israeli estimates were less than an order of magnitude different from each other, using 31 bin data, and 96.9% of these comparisons were within an order of magnitude when using a randomly sampled FBI Caucasian database (Table 5). With the random sampling approach, 16 target DNA profiles exceeded an order of magnitude difference, while having a frequency more common than 1/1,000,000. Only four of these target profiles were from Caucasians, differences ranging from 10.9 to 45.1-fold, with the most common frequency being 1/163,000. Because a degree of gene flow among groups in the United States can be anticipated, these very few differences will be even more diminished when considering the United States population of potential perpetrators.

The same trends hold for the rest of the Hae III data comparisons of reference populations within a major population category. The FBI Black database yields estimates similar to those for regional Black databases (Table 6); intra-Hispanic (Table 7), intra-Oriental (Table 8), intra-Asian Indian (Table 8), and to a degree even intra-Amerindian (Table 9) database comparisons rarely produce substantial differences in the estimated likelihood of occurrence of a particular DNA profile.

The data from comparisons of regional U.S. Black population samples are very telling of the absence of effects of population substructure on the estimate of the likelihood of occurrence of a DNA profile in the general population (Table 6). Previous studies [38,39] have shown that there are differences in the degree of Caucasian admixture for Southern versus Northern or Western Blacks. Although Caucasian admixture has been estimated to be as little as 4–10% in the South and as much as 30% in the North, there were no apparent differences in the likelihood of occurrence of the 1964 target DNA profiles among regional Black population samples (Table 6). 99.5%–99.9% of the estimates of the target DNA profiles from the FBI Black database compared with regional U.S. Black sample populations using a rebinned format were within an order of magnitude. Out of all the regional Black sample population comparisons only three target profiles had differences exceeding an order of magnitude and frequencies more common than 1/1,000,000; none of these target profiles were from Black individuals. The Michigan Black vs. South Carolina Black data comparisons yielded very similar results, that is, no Black target DNA profiles with frequencies more common than 1/1,000,000 had estimate differences exceeding an order of magnitude (Table 6). These observations were expected since the frequencies of VNTR alleles are more rare in general in Black populations studied than in other major population groups [9,29,31,40], and the differences among allele frequencies in different population samples are diminished when the set of alleles comprising a DNA profile is used [29].

The Haitian database also showed very few differences in estimates compared with the FBI Black database. With the random sampling approach, 99.8% of the FBI Black vs. Haitian comparisons were within an order of magnitude. There was only one target DNA profile in the FBI Black vs. Haitian comparisons that exceeded an order of magnitude and had a frequency more common than 1/1,000,000; it was a Black target profile (two loci) with a 12.0-fold difference and the more common frequency was 1/9150.

Even though southeastern and southwestern Hispanic populations have different racial admixture [36,37], there were very few differences for frequency estimates of the 1964 target profiles among the Hispanic sample population comparisons (Table 7). The increase in the ratios with Michigan Hispanics is due to the smaller sample size in that database (Table 1). To illustrate, the random sampling approach was applied to the larger FBI southeastern Hispanic database and compared with rebinned data from Michigan Hispanics. Using this approach, 2.1% of the target DNA profiles had differences in

estimates greater than one order of magnitude between FBI southeastern Hispanics and Michigan Hispanics (Table 6). However, only 0.7% of the total target profiles (that is, 14 profiles) had frequency differences greater than one order of magnitude (ranging from 10.2–20.4-fold) where estimates were more common than 1/1,000,000.

Despite the genetic differences between Chinese, Japanese, and Malaysians, there were few differences in profile frequency estimates among these groups (Table 8). Using rebinned data, no more than 1.5% of the target profiles had estimates differing more than one order of magnitude. To obtain a further assessment of frequency estimates in Oriental reference databases, the frequency of 577 target DNA profiles from Chinese ($N = 120$), Japanese ($N = 138$), Korean ($N = 100$), and Vietnamese ($N = 219$) individuals were assessed in several Oriental databases. In a comparison of Japanese vs. Korean reference databases, none of the 577 Oriental target profiles differed by an order of magnitude unless the estimates were less common than 1/1,000,000. Similarly, for Korean vs. Vietnamese, large differences were not found. Only one target profile, from a Japanese individual, demonstrated more than one order of magnitude difference in the estimate (a 12.5-fold difference), and the more common frequency was 1/563,000. For a Chinese vs. Japanese comparison, there were only six target profiles with differences exceeding one order of magnitude and having a frequency more common than 1/1,000,000. The differences ranged from 10.4–16.7-fold. Of the six target profiles, three were from Chinese individuals and none were from Japanese target profiles; the most common frequency among these three Chinese target profiles was 1/268,000. The three remaining target profiles were Vietnamese with the most common frequency being 1/83,000. Additionally, the comparison between the Asian Indian reference populations from Canada and Singapore demonstrate few differences (Table 8).

Despite the genetic differences known to exist among American Indians [41–45], the data comparisons are still quite similar (Table 9). With the rebinned data format only, 10.6%–24.3% of the estimates from the various American Indian database comparisons exceeded one order of magnitude. The increased differences between 31 bin format comparisons, relative to rebinned format comparisons, may reflect the greater genetic differences among American Indian subgroups than among Caucasian, Black, Hispanic, or Oriental subgroups.

Because of limitations of space it was not possible to subdivide all the data into one, two, three, and four locus target profile frequency estimate comparisons. Therefore, Table 10 displays one such example from each of Tables 5–9. Based on the data in Table 10, the number of target loci that contain three loci (672) and four loci (648) (Table 3) and the scatter plots (figures 1 and 2), it is unlikely to observe substantial differences in DNA profile frequency estimates.

Comparisons across-major groups consistently yield wider breadth scatter plots than within-group comparisons (Table 11). Although the databases were not tested specifically for genetic differences or similarities, it would be expected that reference data bases from genetically more similar groups would yield estimates that are more similar than those drawn from different groups. However, when databases are not genetically similar, such as would be anticipated with across-major group comparisons, a 31 bin format generally increases the range of estimates between the two sample populations (Table 11). In contrast, the smoothing effect of rebinning on allele frequencies decreases DNA profile frequency estimate differences across major population group databases.

Lewontin and Hartl [1] and the NRC report [2] assert that genetic diversity between subgroups within a race is greater than between races, while others have refuted this position [3,31,40,43,46–51]. The scatter plots of within-group comparisons tend to cluster far more than between-group comparisons. Differences greater than two orders of magnitude in DNA profile frequency estimates from different U.S. databases are unlikely events and usually occur when frequencies are less common than 1/1,000,000. Such

differences in this frequency estimate range do not alter the implication that the DNA profile is rare.

Conclusions

Subdivision, either by ethnic group or by U.S. geographic region, within a major population group does not substantially affect forensic estimates of the likelihood of occurrence of a DNA profile. Because the binning process defines a statistical class for all observed and unobserved alleles in a population, all populations share the same alleles. The smoothing effects of fixed binning will reduce differences among ethnic groups [12,20]. Estimated frequencies among regional groups and several subgroups of a major population category are similar. Estimates of the likelihood of occurrence of a DNA profile using major population group databases (for example, Caucasian, Black, and Hispanic) provide a greater range of frequencies than would estimates from sub-groups of a major population category. Furthermore, Chakraborty et al. [34] recently calculated the confidence intervals for DNA profile estimates using FBI general population databases. The range of the confidence intervals was narrower than the range of estimates observed in the across-major group scatter plots. The most appropriate approach, therefore, is to estimate the likelihood of occurrence of a particular DNA profile in each major group. Additionally, the significance of the magnitude of the very few differences that were observed wanes when it is taken into consideration that the binning procedure used yields conservative estimates. On average, each allele frequency is over-estimated at least two-fold [52]. Furthermore, these very few differences also are exaggerated due to sampling fluctuations for each database and measurement biases that exist in the methods used by the various laboratories that have generated the population data. Therefore, based on empirical data, there is no demonstrable need for employing alternate approaches, such as the ceiling approach, to derive statistical estimates. VNTR frequency data from major population groups provide valid estimates of DNA profile frequencies without significant consequences for forensic inferences.

This paper has demonstrated in an informal way that estimated profile frequencies differ more between, than within major population groups. Not all possible comparisons have been presented, but the numerical results demonstrate the falsity of assertions that human populations differ more between, than within, major population groups. A formal demonstration of the effects of various levels of categorization (that is, individual, ethnic, and racial) is possible, and such studies are underway.

References

[1] Lewontin, R. C. and Hartl, D. L., "Population Genetics in Forensic DNA Typing," *Science*, Vol. 254, 1991, pp. 1745–1750.
[2] National Research Council, "DNA Typing: Statistical Bases for Interpretation," *DNA Technology in Forensic Science*, Chapter 3, Washington, D.C., National Academy Press, 1992, pp. 74–96.

[3] Budowle, B., Monson, K. L., and Wooley, J. R., "Reliability of Statistical Estimates in Forensic DNA Typing," *DNA on Trial: Issues in the Use of DNA Identification in Courts*, P. R. Billings, Ed., Cold Spring Harbor Laboratory Press, New York, 1992, pp. 79–90.

[4] Budowle, B. and Stafford, J., "Response to Expert Report by D. L. Hartl Submitted in the Case of United States Versus Yee," *Crime Laboratory Digest*, Vol. 18, 1991, pp. 101–108.

[5] Budowle, B. and Stafford, J., "Response to Population Genetic Problems in the Forensic Use of DNA Profiles by R. C. Lewontin Submitted in the Case of United States Versus Yee," *Crime Laboratory Digest*, Vol. 18, 1991, pp. 109–112.

[6] Evett, I. W. and Weir, B. S., "Flawed Reasoning in Court," *Chance*, Vol. 4, 1992, pp. 19–21.

[7] Weir, B. S. and Evett, I. W., "Whose DNA?", *American Journal of Human Genetics*, Vol. 50, 1992, p. 869.

[8] Brookfield, J., "Brookfield Replies," *Nature*, Vol. 358, 1992, p. 483.

[9] Balazs, I., Baird, M., Clyne, M., and Meade, E., "Human Population Genetic Studies of Five Hypervariable DNA Loci," *American Journal of Human Genetics*, Vol. 44, 1989, pp. 182–190.

[10] Balazs, I., Neuweiler, J., Gunn, P., Kidd, J., Kidd, K. K., Kuhl, J., Mingjun, L., "Human Population Genetic Studies Using Hypervariable Loci. I. Analysis of Assamese, Australian, Cambodian, Caucasian, Chinese, and Melanesian Populations," *Genetics*, Vol. 131, 1993, (in press).

[11] Brinkmann, B., Rand, S., and Wiegand, P., "Population and Family Data of RFLP's Using Selected Single and Multi-Locus Systems," *International Journal of Legal Medicine*, Vol. 104, 1991, pp. 81–86.

[12] 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, 1991, pp. 841–855.

[13] Buffery, C., Burridge, F., Greenhalgh, M., Jones, S., and Willott, G., "Allele Frequency Distributions of Four Variable Number Tandem Repeat (VNTR) Loci in the London Area," *Forensic Science International*, Vol. 52, 1991, pp. 53–64.

[14] Chakraborty, R. and Daiger, S. P., "Polymorphisms at VNTR Loci Suggest Homogeneity of the White Population of Utah," *Human Biology*, Vol. 63, 1991, pp. 571–587.

[15] Chakraborty, R., DeAndrade, 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.

[16] Chakraborty, R. and Jin, L., "Heterozygote Deficiency, Population Substructure and Their Implications in DNA Fingerprinting," *Human Genetics*, Vol. 88, 1992, pp. 267–272.

[17] Chakraborty, R., Srinivasan, M. R., Jin, L., and DeAndrade, M., "Effects of Population Sub-division and Allele Frequency Differences on Interpretation of DNA Typing for Human Identification," *Proceedings of the 1992 International Symposium on Human Identification*, Promega Corporation, Madison, Wisconsin, 1992, pp. 205–222.

[18] Chow, S. T., Tan, W. F., Yap, K. H., and Ng, T. L., "The Development of DNA Profiling Database in an Hae III Based RFLP System for Chinese, Malays, and Indians in Singapore," *Journal of Forensic Sciences*, Vol. 38, No. 4, July 1993, pp. 874–884.

[19] Deka, R., Chakraborty, R., and Ferrell, R. E., "A Population Genetic Study of Six VNTR Loci in Three Ethnically Defined Populations," *Genomics*, Vol. 11, 1991, pp. 83–92.

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

[21] Devlin, B., Risch, N., and Roeder, K., "No Excess Homozygosity at DNA Fingerprint Loci," *Science*, Vol. 249, 1990, pp. 1416–1420.

[22] Devlin, B., Risch, N., and Roeder, K., "Forensic Inference from DNA Fingerprints," *Journal of the American Statistics Association*, Vol. 87, 1992, pp. 337–350.

[23] 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.

[24] Gasparini, P., Trabetti, E., Savoia, A., Margio, M., and Pignatti, P. F., "Frequency Distribution of the Alleles of Several Variable Number of Tandem Repeat DNA Polymorphisms in the Italian Population," *Human Heredity*, Vol. 40, 1990, pp. 61–68.

[25] Gill, P., Woodroffe, S., Lygo, J. E., and Millican, E. S., "Population Genetics of Four Hypervariable Loci," *International Journal of Legal Medicine*, Vol. 104, 1991, pp. 221–227.

[26] Risch, N. and Devlin, B., "On the Probability of Matching DNA Fingerprints," *Science*, Vol. 255, 1992, pp. 717–720.

[27] Smith, J. C., Anwar, R., Riley, J., Jenner, D., and Markham, A. F., "Highly Polymorphic

Minisatellite Sequences: Allele Frequencies and Mutation Rates for Five Locus-Specific Probes in a Caucasian Population," *Journal of the Forensic Science Society*, Vol. 30, 1990, pp. 19–32.

[28] Van Eede, P. H., Henke, L., Fimmers, R., Henke, J., and de Lange, G. G., "Size Calculation of Restriction Enzyme Hae III-Generated Fragments Detected by Probe YNH24 by Comparison of Data from Two Laboratories: The Generation of Fragment-Size Frequencies," *Forensic Science International*, Vol. 49, 1991, pp. 21–31.

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

[30] Yokoi, T., Nata, M., Odaira, T., Sagisaka, K., "Hypervariable Polymorphic VNTR Loci for Parentage Testing and Individual Identification," *Japanese Journal of Human Genetics*, Vol. 35, 1990, pp. 179–188.

[31] Chakraborty, R. and Kidd, K. K., "The Utility of DNA Typing in Forensic Work," *Science*, Vol. 254, 1991, pp. 1735–1739.

[32] Monson, K. L. and Budowle, B., "VNTR Population Frequency Estimation for Forensics: Effect of Reference Population and Calculation Method," *Journal of Forensic Sciences*, Vol. 38, 1993, pp. 1037–1050.

[33] *VNTR Population Data: A Worldwide Study*, Volumes I–IV, Federal Bureau of Investigation, Washington, D.C., 1993.

[34] Chakraborty, R., Srinivasan, M. R., and Daiger, S. P., "Evaluation of Standard Error and Confidence Interval of Estimated Multilocus Genotype Probabilities and Their Implications in DNA Forensics," *American Journal of Human Genetics*, Vol. 52, 1993, pp. 60–70.

[35] Lewontin, R. C., "The Apportionment of Human Diversity," *Evolutionary Biology*, Vol. 6, 1972, pp. 381–398.

[36] Cerda-Flores, R. M., Kshatriya, G. K., Barton, S. A., Leal-Garza, C. H., Garza-Chapa, R., Schull, W. J., and Chakraborty, R., "Genetic Structure of the Populations Migrating from San Luis Potosi and Zacatecas to Nuevo Leon in Mexico," *Human Biology*, Vol. 63, 1991, pp. 309–327.

[37] Cerda-Flores, R. M., Kshatriya, G. K., Bertin, T. K., Hewett-Emmett, D., Hanis, C. L., and Chakraborty, R., "Genetic Diversity and Estimation of Genetic Admixture Among Mexican-Americans of Starr County, Texas," *Annals of Human Biology*, Vol. 19, 1992, pp. 347–360.

[38] Reed, T. E., "Caucasian Genes in American Negroes," *Science*, Vol. 165, 1969, pp. 762–768.

[39] Chakraborty, R., Kamboh, M. I., Nwankwo, M., and Ferrell, R. E., "Caucasian Genes in American Blacks," *American Journal of Human Genetics*, Vol. 50, 1992, pp. 145–155.

[40] Devlin, B. and Risch, N., "Ethnic Differentiation at VNTR Loci, with Special Reference to Forensic Applications," *American Journal of Human Genetics*, Vol. 51, 1992, pp. 534–548.

[41] Neel, J. V., " 'Private' Genetic Variants and the Frequency of Mutation Among South American Indians," *Proceedings of the National Academy of Sciences USA*, Vol. 70, 1973, pp. 3311–3315.

[42] Neel, J. V. and Ward, R. H., "Village and Tribal Genetic Differences Among American Indians, and the Possible Implications for Human Evolution," *Proceedings of the National Academy of Sciences USA*, Vol. 65, 1970, pp. 323–330.

[43] Smouse, P. E., Spielman, R. S., and Park, M. H., "Multiple-Locus Allocation of Individuals to Groups as a Function of the Genetic Variation Within and Differences Among Human Populations," *American Naturalist*, Vol. 119, 1982, pp. 445–463.

[44] Kidd, J. R., Black, F. L., Weiss, K. M., Balazs, I., and Kidd, K. K., "Studies of Three Amerindian Populations Using Nuclear DNA Polymorphisms," *Human Biology*, Vol. 63, 1991, pp. 775–794.

[45] Chakraborty, R. and Weiss, K. M., "Genetic Variation of the Mitochondrial Genome in American Indians Is at Mutation-Drift Equilibrium," *American Journal of Physical Anthropology*, Vol. 86, 1991, pp. 497–506.

[46] Latter, B. H. D., "Genetic Differences Within and Between Populations of the Major Human Subgroups," *American Naturalist*, Vol. 116, 1980, pp. 220–237.

[47] Nei, M. and Roychoudhury, A. K., "Genetic Relationship and Evolution of Human Races," *Evolutionary Biology*, Vol. 14, 1982, pp. 1–59.

[48] Morton, N., "Genetic Structure of Forensic Populations," *Proceedings of the National Academy of Sciences USA*, Vol. 89, 1992, pp. 2556–2560.

[49] Morton, N. E., Collins, A., and Balazs, I., "Bioassays of Kinship for Hypervariable Loci in Blacks and Caucasians," *Proceedings of the National Academy of Sciences USA*, Vol. 90, 1993, pp. 1892–1896.

[50] Mitton, J. B., "Genetic Differentiation of Races of Man As Judged by Single-Locus and Multilocus Analyses," *American Naturalist*, Vol. 111, 1977, pp. 203–212.

[*51*] Mitton, J. B., "Measurement of Differentiation: Reply to Lewontin, Powell, and Taylor," *American Naturalist*, Vol. 11, 1978, pp. 1142–1144.
[52] 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*, 1993, (in press).

Address requests for reprints or additional information to
Bruce Budowle, Ph.D.
FSRTC, FBI Academy
Quantico, VA 22135