Correspondence

The AICEF/GITAD: Latin American Academy of Criminalistics and Forensic Studies

Sir:

The term "Iberoamerica" is often used in science and research to convey collaboration or association between Latin America and the two European countries Spain and Portugal, that predominately contributed to the language and culture of Latin America. Iberoamerica (from now on, IA) is therefore comprised of 23 countries with a total population over 500 million people. Two of these countries are Portuguese-speaking (Portugal and Brazil), and the other 21 countries are Spanish-speaking. Among the IA countries there are those with large populations, with federal political organizations (i.e., Brazil and Mexico), and others with smaller population sizes, such as the six countries that form Central America (i.e., Honduras, El Salvador, Guatemala, Nicaragua, Costa Rica, and Panama).

There are also important historical, socio-political anthropological and cultural differences among the IA countries. However, because of the common history since the XVI century due to the influence of Spain and Portugal, there are also common ties, even though the geographic distance is great, for example, between the Chilean Antarctic areas and the northern deserts of the Mexican New California. The fact is that the common language (Spanish and Portuguese are substantially similar and thus comprehendible) is a tool that facilitates the communication, understanding and cooperation among IA. This basis is the foundation for creation of the AICEF.

It is very difficult to describe, country by country, the organization of the forensic sciences. These differences make it difficult to establish a similar structure. The problem is exacerbated by peculiarities dependent on the political organization of some countries. In many countries around the world, forensic sciences and crime investigation have oversight from the State or Government, through the Ministeries (or equivalents) of the Interior (Internal Affairs, Home Office), and of Justice (Dept. of State). There are also National or Regional Institutes of Legal Medicine provided authority or mission from different Ministries or serve as a part of the Attorney's General Office. Also, Universities (usually public or State ones) sometimes play an important role by educating forensic professionals and/or by carrying out legal autopsies or toxicological analyses. Also, in some countries there exist mechanisms for private laboratories or companies to provide services, and obviously forensic experts that might legally play an important role in the forensic arena. The fact is as recent as 2000 the forensic structure in the different IA countries is notably different from one to another.

GITAD (Grupo Iberoamericano de Trabajo para el Análisis del DNA; Iberoamerican Working Group on DNA Analysis) was the original group of AICEF, and it was created to attempt to coordinate the efforts of all the forensic DNA typing laboratories of IA. The goal is to facilitate communication of technical knowledge and experiences and to help improve quality assurance and quality control programs.

The GITAD was founded in October 1998, during the Ninth International Symposium of Human Identification, held in Orlando, Florida. This first meeting was attended by representatives from 11 different IA countries (Argentina, Brazil, Chile, Colombia, Costa Rica, El Salvador, Mexico, Puerto Rico, Uruguay, Venezuela, and Spain). The help and support from Promega Corporation (Madison, WI) was welcomed and greatly appreciated.

A second GITAD meeting was held in Belo Horizonte, Brazil, in concert with the Second Latin American Symposium on Human Identification. At this meeting, the AICEF (Academia Iberoamericana de Criminalística y Estudios Forenses = Iberoamerican Academy of Criminalistics and Forensic Studies) was formally established. Different sections comprise the AICEF; these sections are Forensic and Legal Medicine, Toxicology, Crime Scene Investigation, Dactiloscopy, Ballistic & Graphology, Law, Anthropology, Odontology, Bioethics and the GITAD which is now the Forensic Genetic and Biology Section of AICEF.

By October 1999, all IA countries are represented in the AICEF/GITAD, regardless of the police/forensic structure of the country, the type of genetic techniques currently used or any other criteria. Official GITAD/AICEF members must be members of a Laboratory belonging to a public or Government institution, i.e., Ministry of Justice, State Police Departments, Federal Law EPolice, Attorney's Office, Institute of Legal Medicine.

A survey was conducted to determine the status of DNA laboratories in the area. A number of conclusions could be made regarding the need for collaboration and cooperation among all Latin American countries. For instance, most Latin American forensic laboratories are small in size and have few personnel (typically fewer than eight people). On a positive side, most laboratories contain highly qualified personnel who have Ph.D. university degrees and who run the laboratory. This is an optimistic situation. Although there have been limited international relationships, highly-educated personnel are in place—a prerequisite for high quality.

Another interesting observation is that most GITAD laboratories almost exclusively deal with criminal casework and only paternity analyses when required as part of a judicial investigation. Also to be noted are the differences in techniques used not only among the different countries, but also inside the same country. Some of the differences are such that the sharing of DNA profile data is not possible. This data incompatibility can be counterproductive, especially because investigative budgets are limited regarding DNA analysis.

Because of the experience of the different members and because of the need to develop common guidelines, within a single year the AICEF/GITAD has made a number of decisions to meet its desired goal of communication and data sharing.

A set of six short tandem repeat loci have been chosen as common core set among the IA countries, in order to facilitate interchange of data and compatibility for future common databases or criminal collaboration. The GITAD six core loci are CSF1PO, TPOX, TH01, D7S820, D13S317, and D16S539. These loci were selected because: (a) they can be analyzed either by silver-staining

or fluorescent-based detection techniques. Currently, less than 20% of the labs in Latin America can use fluorescent detection methods; (b) they are well defined and reagents for analysis can be purchased from commercial companies to ensure compatibility and quality; (c) they have relatively high PDs and PEs; and (d) they are CODIS-compatible. GITAD recommends the use of these six STR loci in order to build databases that can be compatible among different countries. The use of these six STR loci does not preclude the use of additional loci. Each laboratory must use the number of loci necessary to achieve the desired PD or PE in a paternity case (i.e., in paternity, to reach a PI>1.000, it is usually necessary to use more than 6 loci).

Different working groups of the GITAD have been established. These are: (a) Quality assurance and quality control; (b) forensic statistics; (c) evidence collection and preservation; and (d) comparative legislation. The working groups are each developing common/similar guidelines for all countries, regardless of specific requirements due to national laws. The first document on "Recommendations for QA/QC Procedures in Forensic Genetic Laboratories" has been approved as of Sept. 2000, and it is available through our web site. Hard copies will also be distributed to all laboratory members and are available, free of charge, upon request to the GITAD President. By June 2001 all other working groups will release guidelines and recommendations.

An initial QC analysis was run in Autum 1999, including typing of four unknown dried bloodstains (spotted on cotton) for at least the six core STR loci (i.e., an open blind test). Results were received and processed, showing compatibility and reliability of the participating laboratories (data not shown). Pooling all data, up to 17 different loci were analyzed using both silver staining (85% of participating laboratories) and fluorescent-based techniques (15%).

A third AICEF/GITAD meeting was held in Montevideo, Uruguay (Feb. 16–18, 2000) and a number of major issues were addressed. These include: (a) potentially increasing the number of core STR loci for the IA database compatibility from the six first recommended to the same 13 loci in CODIS; (b) establishing minimum requirements regarding QA/QC procedures in Forensic DNA laboratories; (c) advocating minimum criteria for statistical calculations in final reports; and (d) including recommendations for legislators to ensure and facilitate international cooperation.

With such an active program, the AICEF is trying to ensure that the IA will have a prominent role in the 21st century in the forensic sciences. Although such efforts are laborious and time consuming, the AICEF/GITAD believes that the benefits of such endeavors are more than worthwhile.

Jose A. Lorente, M.D., Ph.D.
President, AICEF/GITAD
Dept. of Legal Medicine, University of Granada
Av. Madrid 11, 18012 Granada Spain

Misinterpretation of a Urinary 6-Monoacetyl Morphine Concentration

Sir:

This laboratory was recently involved in the investigation of a multiple-fatality vehicular homicide case. We feel that an opinion presented by an expert for the defense was seriously flawed. We thought that the opinion, and its underlying basis, might be of interest to others in the field.

The driver in the case was known to local police, and was suspected both of causing the accident, and of being "significantly impaired" by heroin at the time. The driver was injured in the crash, and received emergency medical care, including analgesia in the form of morphine. Hospital testing of a urine sample indicated the presence of opiates.

We were contacted by the County Prosecutor to determine if testing was available that could establish if the driver had indeed used heroin, and was impaired at the time of the accident (the Prosecutor recognized the potentially confounding presence of morphine).

We suggested that given the circumstances of the case, analysis of the urine sample for 6-monoacetyl morphine (6-MAM) might confirm use of heroin. We emphasized that detection of 6-MAM in a urine sample would not provide evidence of impairment at the time of the accident, merely confirmation that heroin had been used at some time prior to the collection of the sample. With this understanding, the prosecution requested that our laboratory perform the analysis; 0.267 mg/L 6-MAM in the urine sample was detected by GCMS using standard methods.

Prior to trial, we received a copy of an opinion provided to the defense by a reputable expert, which included the (unreferenced) statement "Literature reports indicate that 6-monoacetylmorphine is present in urine in 64% to 73% of all heroin users studied, averaging approximately 0.8 mg/L, and ranging up to 10 mg/L. Consequently, if 0.267 mg/L was accurately detected in (the subject's) urine, the detected concentration is relatively low compared to literature values, indicating a probable prolonged period of time between (the) last heroin use and collection of the urine sample."

The potential for significant error in a quantitative inference derived from a urine value is well recognized in the forensic community, and needs no further comment. We were, however, interested in the basis for the "average value" of ~0.8 mg/L, and the source thereof. Our experience with 6-MAM suggested that the 0.267 mg/L was a relatively high number.

It appears that the source of information used by the expert for the defense was a recent edition of "Baselt" (1) which, under the section "Heroin," contained the following statement: "6-acetylmorphine is present in urine in 64% to 73% of all heroin users studied, averaging approximately 0.8 mg/L, and ranging up to 10 mg/L (Fehn & Megges, 1985; Derks et al., 1986)."

A review of the studies cited revealed that of the 47 urine samples collected from heroin users, and evaluated by Fehn and Megges (2), 6-MAM was detected in 24 cases. 6-MAM levels in 22 of those cases were less than 0.55 mg/L. One result of 8.0 mg/L and one of 10.0 mg/L were included in the data set, without comment. Interestingly, while the mean of the complete data set was 0.864 mg/L, the authors make no mention of this value in the article, (presumably recognizing its inherent lack of statistical reliability in their specific experiment). The 22 values with 6-MAM concentrations less than 0.550 mg/L comprise a markedly skewed, non-gaussian data set, for which the mean value (0.124 mg/L) is neither characteristic, nor of predictive value (SD = 0.156 mg/L). Because the study was not controlled for dose, or time post exposure, a "mean" value for the concentration of a metabolic intermediate such as 6-MAM is inherently meaningless. Indeed, in the absence of time and dose parametric constraints, the best description for the mean value is that it approaches 0 as a limit.

The key point, of course, is that the Fehn and Megges study was a methodology report, intended only to demonstrate the capability and reliability of detection of 6-MAM in urine samples. Because of the experimental design, the data do not provide a legitimate basis

for guiding comparative interpretations such as were provided by the defense expert. (Derks, et al. (3) is similarly oriented towards methodologic considerations.)

The defendant pled guilty to all charges prior to trial, so we did not have an opportunity to address these issues directly in court. Many in the forensic community are familiar with the problem with the mean value and range derived from the Fehn and Megges study. We did wish however, to note the potential for significant interpretive error regarding 6-MAM, when summary results are utilized without evaluation of the underlying literature, and/or reference to personal experience.

References

1. Baselt RC. Disposition of toxic drugs and chemicals in man, 5th ed., Chemical Toxicology Institute, Foster City, CA, 2000.
2. Fehn J, Megges G. Detection of O_6-monoacetylmorphine in urine samples by GC/MS as evidence for heroin use. J Anal Tox 1985;9:134–8.
3. Derks HJGM, Van Twillert K, Pereboom-DeFauw DPKH, Zomer G, Loeber G. Determination of the heroin metabolite 6-acetylmorphine by high-performance liquid chromatography using automated pre-column derivatization and fluorescence detection. J Chrom 1986;370:173–8.

Robert H. Powers, Ph.D.
Chief Toxicologist
Hamilton County Coroner's Office
3159 Eden Avenue
Cincinnati, Ohio 45219

Detection of Deviations from Genetic Equilibrium—A Commentary on Budowle B, Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM. Population Data on the Thirteen CODIS Core Short Tandem Repeat Loci in African Americans, US Caucasians, Hispanics, Bahamians, Jamaicans, and Trinidadians. J Forensic Sci 1999;44:1277–86

Sir:

The existence of the combined DNA databank, CODIS, is a very valuable contribution to the full use of DNA typing in forensic science in the United States of America. Thirteen short tandem repeat (STR) loci were chosen to be the core loci in CODIS (1). Budowle et al. (2) present allele frequency distributions for the 13 core loci in each of six ethnic groups resident in the United States. The population samples range in size from 80 individuals from Trinidad to 210 African Americans. We were disappointed that Budowle et al. suggest that this data supplies further validation of the forensic use of the product rule. This paper shows that testing for departures from Hardy-Weinberg equilibrium with datasets of this size is very unlikely to detect any such departures. The fact that a formal test does not detect a departure from Hardy-Weinberg equilibrium does not demonstrate that the population from which the sample was drawn is in equilibrium. Thus, it would be prudent if forensic scientists using data of this type assumed that population substructure exists and corrected for it. It would also be prudent to adjust estimates of genetic parameters calculated from the data for sampling uncertainty induced by small sample sizes.

The power of a statistical testing procedure is the probability that the procedure will detect a significantly difference or departure from assumptions when there is actually a difference to detect. High power means that the test should find any departures. In such an instance, failure to find departure is strong evidence that there is, indeed, no such departure. More formally we propose two complementary hypotheses

H_0: The locus is in Hardy-Weinberg Equilibrium (HWE)
H_1: The locus is not in HWE.

We perform two experiments to examine the power of Fisher's exact test to detect departures from Hardy-Weinberg equilibrium.

The first experiment considers the insensitivity of the exact test to perturbations in the observed data of Budowle et al. (2). Consider the genotype counts from 203 Caucasian individuals at the D3S1358 locus (3) given in Table 1. We see that 21 individuals were recorded as having the 15, 16 genotype at this locus.

Given an allelic array such as that in Table 1, Fisher's exact test for HWE enumerates all such arrays with the same allele counts and calculates the probability of observing that array given the marginal allele counts. The *p*-value for the exact test is given by the sum of such probabilities where the sum is over all arrays giving rise to a smaller probability than the observed data array. The total number of arrays will be very large, so in practice a complete enumeration is not possible. However, several Monte-Carlo techniques allow us to obtain an estimate of the *p*-value. A traditional permutation testing approach (4) "unlinks" the genotypes (so that the locus is in HWE), and constructs a "new" allelic array from the pool of individual alleles, thus preserving the counts of alleles. If the new permuted array gives a smaller probability than the observed array then a counter, *k*, is incremented. This procedure is repeated a fixed number, *N*, times and the *p*-value is estimated by $\hat{p} = k/N$. A Monte-Carlo Markov Chain (MCMC) test (5) samples a finite number of possible arrays by probabilistically perturbing the allelic array in such a way that the allelic counts are preserved and estimates the *p*-value from this. The former is the method used by Budowle et al. (2). To illustrate the lack of power of the exact test to detect deviations from HWE, additional copies of a single genotype were added to the array the rest of the array being left unchanged. The *p*-value of the exact test on these perturbed arrays was recorded.

In the second experiment, populations known to be disequilibrium were simulated and the behavior of the exact test on samples of genotypes drawn from them was examined. We model disequilibrium as a nonzero value of θ (sometimes known as Wright's F_{ST}). The simulation scheme is as follows: we generate a population of size 10,000 using the New Zealand Caucasian database frequencies for locus vWA. The initial population is in HWE because each allele of the genotype is chosen randomly and independently. It can be shown (4) that if several populations of size N_s are allowed to separate and breed (with replacement and

TABLE 1—*Observed genotype counts for* N = *203 Caucasians at the D3S1358 locus.*

Allele							
13	0						
14	0	2					
15	0	10	13				
16	0	11	21	11			
17	1	16	22	23	9		
18	0	16	21	16	5	4	
19	0	0	0	1	1	0	0
Allele	13	14	15	16	17	18	19
Allele count	1	57	100	93	86	67	2

nonselectively) for t generations that the genetic distance between them is given by

$$d = -\ln(1-\theta) = -t\ln\left(1-\frac{1}{2N_s}\right)$$

where $\theta = \overline{\theta}$. Therefore, choosing a subpopulation size of $N_s = 1000$, we allow our ten subpopulations to breed for

$$t = \frac{\ln(1-\theta)}{\ln\left(1-\frac{1}{2N_s}\right)}$$

generations and then recombine. The number of generations, t, for the different values of θ are

θ	t
0	0
0.01	20
0.03	61

We select a database of size n, where n is 80 or 200, from this "inbred" population. These are equivalent to the largest and smallest samples in Budowle et al. We then perform the exact test on the samples. The experiment was repeated 10,000 times for each combination of database size and θ, and the proportion of "significant" results recorded. When θ is zero this provides an estimate of the size of the test or Type I error rate, and when θ is not zero the proportion is an estimate of the power of the test.

If the exact test p-value is small, we would be unlikely to observe the genotype counts in the database if the null hypothesis that the locus was in HWE were true, and hence have some evidence for the alternative hypothesis H_1. Therefore, if data that is known not to be in HWE is simulated, the exact test is expected to return a small p-value more frequently. We expect to see a higher (correct) rejection rate (increased power) both as the departure from HWE becomes larger, and as the database size increases.

Experiment 1

Two genotypes were chosen as the source of perturbations in Table 1. The 15,16 genotype was chosen, as these are the two most common alleles at this locus. The observed count of this genotype is close to expectation under assumption H_O that the locus is in HWE. The second genotype was 17,18 involving two relatively common alleles. However the observed count of the 17,18 genotype had the largest standardized residual from expectations calculated under assumption H_O. The p-values of the exact test are plotted in Figs. 1*a* and 1*b*. The initial count of the 15,16 genotype was 21 and the p-value of the exact test is 0.082. To lower the p-value to the traditional 5% significance level an additional 13 copies of the 15,16 genotype must be added to the database.

This considers the D3S1358 locus from the American Caucasian population in isolation. Budowle et al. (2) consider six ethnic groups and 13 loci within each group. To achieve an overall 5% significance level across the 78 exact tests the significance level of each individual test is reduced to 0.064% using the Bonferroni correction. If this locus in this population were viewed as just one amongst 78 possible tests to achieve the formal significance level the number of 15,16 genotypes in the array would have to be increased to 45.

Among the relatively common alleles at the D3S1358 locus the count of the 17,18 genotype, 5, is furthest from its HWE expected value. This genotype count is varied from 0 to 55 to assess its effect of the exact test p-value. We see from Fig. 1*b* that any genotype count between 5 and 40 will give a p-value greater than 5%, while any count between 1 and 50 will ensure that this locus population combination will not reject the tests combined across all 78 combinations. Thus, this observed allelic array, based on a sample of over 200 individuals, shows how insensitive the exact test is.

Another problem with the use of the Bonferroni correction is that the nominal significance level to detect a departure from HWE for any specific locus in a particular ethnic group depends on the total number of loci and ethnic groups considered. Use of the Bonferroni correction means that the more sources of data we have, the less sensitive our testing regime becomes. Several combined tests have been suggested, one at least by Fisher himself. If we hypothesize that every locus in all the ethnic groups considered is in HWE, then the p-value of the exact test is uniformly distributed on the interval 0,1. The sum of $-2\log(p\text{-value})$ has a chi-squared distribution with degrees of freedom equal to twice the number of tests carried out, (6). For the 13 loci in six populations the test statistic is

$$-2\sum_{\substack{loci,\\ populations}}\log(p) = 172.1$$

which gives rise to a p-value for the combined test of 0.18.

We propose a graphical test for departures from HWE. If every locus in every ethnic group is in HWE we may regard the 78 ob-

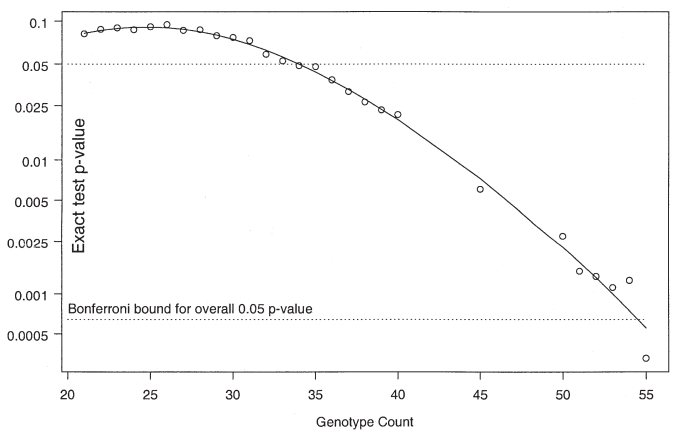

FIG. 1*a*—*Effect on the* p*-value of the exact test of adding additional 15,16 genotypes to the observed allelic array of Table 1.*

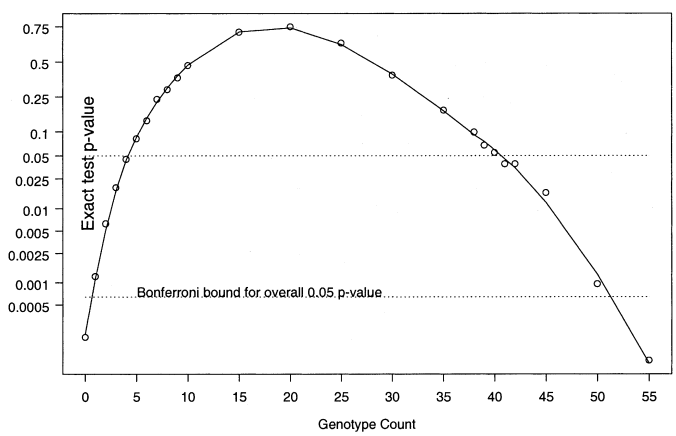

FIG. 1*b*—*Effect on the* p*-value of the exact test of adding additional 17,18 genotypes to the observed allelic array of Table 1.*

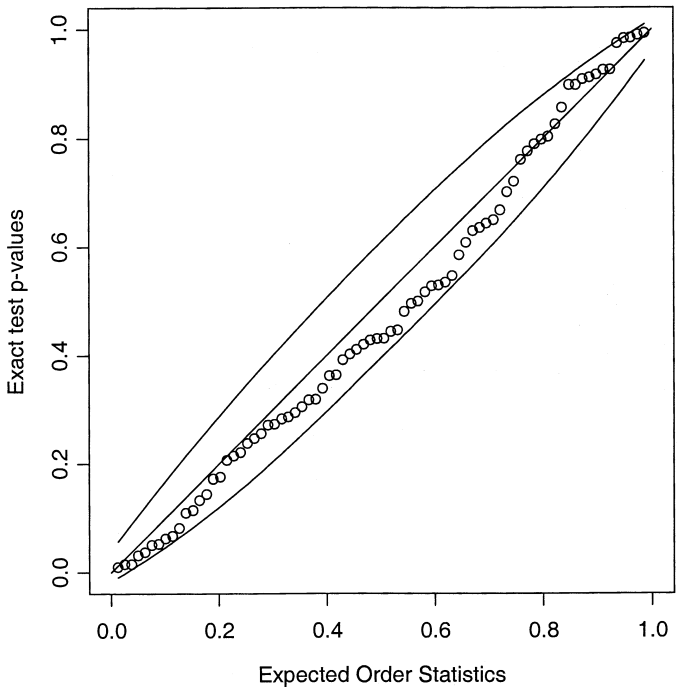

FIG. 2—*Results of a graphical test for departures from Hardy-Weinberg equilibrium for 13 CODIS STR loci combined across six ethnic groups.*

served *p*-values as a random sample of size 78 from the uniform distribution on 0,1, and plot their ordered values against the expected values of an ordered sample from the uniform distribution, a QQplot. In the spirit of Monte Carlo testing, a large number of random samples, which are known to come from a uniform distribution, may be generated. These samples can be used to construct a 95% envelope for the observed plot, Fig. 2 (7). If the observed *p*-values all lie within the envelope we may conclude that, as a whole, the datasets do not provide evidence of departures from HWE at the 5% significance level.

Experiment 2

The lack of power of the exact test can also be demonstrated by simulating populations with population substructure that are known not to be in HWE. Very few human populations have been observed with inbreeding coefficients, θ, in excess of 0.03. The results of the simulations are shown in Table 2. We see that at a nominal significance level of 5% even with a database of 200 individuals the exact test is so insensitive that it will reject a hypothesis of HWE at a locus just over 10% of the time.

We have shown that the exact test will not detect departures from HWE due either to high or low counts at a particular genotype, when tested on individual loci, or with loci combined across several ethnic groups or in samples drawn from populations that show extreme levels of inbreeding. We contend that Budowle et al. (2) cannot claim that their data "provide little evidence of departures from HWE," or that "based on these observations, the data do not support any significant departure from independence between pairs of loci in any sample population." We would go further and claim that there is no reason to test for departures from HWE with databases of this size.

TABLE 2—*Empirical power of Fisher's exact test for HWE. The standard error of the estimated powers is at most 0.5%.*

	Sample Size	
θ	80	200
0	4.75%	4.87%
0.01	5.80%	5.63%
0.03	8.13%	10.65

Of particular concern to us is the statement that "The application of the product rule is valid for estimating the rarity of a multiple loci profile for these tests." We contend that formal testing procedures, when applied to databases of at most a few hundred individuals do not have sufficient power to show that the underlying population is not in HWE, or is substructured or is admixed. In this note we only consider the performance of the exact test at single loci. Law et al. (8) show that the exact test has low power to detect linkage disequilibrium due to population substructure or admixture across several loci. The history of human populations of forensic importance shows that they have been subject to admixture and are substructured. In view of the limited power of the tests we feel that a more balanced conclusion would be: "These tests cannot differentiate between the model of independence and the model of mild departure, therefore it is in the interests of balanced testimony to concede that mild departure may exist."

We therefore suggest that multiple locus profile probabilities be calculated from the NRC formulae 4.10a and 4.10b (9), using conservative values of the inbreeding coefficient, θ. We also suggest that if allele frequencies are going to be estimated using databases drawn from samples of 200 or fewer individuals that the resulting estimates are adjusted for the induced sampling variation (Curran JM, Buckleton JS, Triggs CM, Weir BS. J Am Stat Assn 2000; submitted).

References

1. Budowle B, Moretti TR, Niezgoda SJ, Brown BL. CODIS and PCR based short tandem repeat loci: law enforcement tools. In: Second European Symposium on Human Identification. Madison, WI: Promega Corporation, 1998;73–88.
2. Budowle B, Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM. Population data on the thirteen CODIS core short tandem repeat loci in African Americans, US Caucasians, Hispanics, Bahamanians, Jamaicans and Trinidadians. J Forensic Sci 1999;44:1277–86.
3. www.fbi.gov/programs/lab/fsc/backissu/july1999/dnaloci.txt
4. Weir BS. Genetic data analysis II. Sunderland, MA: Sinauer, 1996; 194.
5. Guo SW, Thompson EA. Performing the exact tests of Hardy-Weinberg proportion for multiple alleles. Biometrics 1992;48:361–72.
6. Cox DR, Hinkley DV. Theoretical statistics. London: Chapman and Hall, 1974; 80.
7. Venables WN, Ripley BD. Modern applied statistics with splus, 3rd ed. New York: Springer, 1999; 72.
8. Law BY-L, Buckleton JS, Triggs CM, Weir BS. Effects of population structure and admixture on exact tests for association between loci. Genetics. In press.
9. National Research Council. The evaluation of forensic DNA evidence. Washington DC: National Academy Press, 1996;114–5.

John Buckleton, Ph.D., ESR,
Private Bag 92021
Auckland, New Zealand

Christopher M. Triggs, Ph.D.
Department of Statistics
University of Auckland
Private Bag 92019
Auckland, New Zealand

James M. Curran, Ph.D.
Department of Statistics
University of Waikato
Private Bag 3105
Hamilton, New Zealand

Authors' Response

The letter by Buckleton, et al. (1) criticizes the study by Budowle et al. (2) by suggesting that the sample sizes used do not have adequate power to detect departures from HWE. Additionally, Buckleton, et al. (1) claim that "the use of simple corrections to single-locus tests is shown to lead to misleading conclusions." Their criticisms are misleading and raise no new issues; also, their experiments are technically flawed.

The correspondence by Buckleton, et al. (1) creates the impression that the issues they perceive saw their genesis only with the publication by Budowle, et al. (2). Buckleton, et al. (1) failed to cite or at least acknowledge the large body of scientific literature (too numerous to list here) that uses the same or similar tests and arrives at the same conclusions as Budowle, et al. (2). In the *Journal of Forensic Sciences* alone, there have been at least 30 studies using the same or similar approaches as Budowle, et al. (2), in just the past year. Use of such tests has been advocated for forensic analyses since the early 1990s (for a few examples see 3–6).

Regardless of their contentions and experiments (to be addressed below), the concluding paragraph in Buckleton, et al.'s (1) paper suggests that the remedy for their concerns is to estimate probabilities essentially by following the recommendations of the NRC II Report (7). Apparently, Buckleton, et al. (1) are unaware of the routine methods for estimating DNA profile frequencies (at least in the United States) that have been practiced for several years (for an explanation of the methods used see 8–11). Strict adherence to Hardy-Weinberg expectations (HWE) is not assumed and correction for sampling variation is taken into account either by the "10-fold" rule or by analytical computations of the upper confidence limit of the point estimate (12).

Since Buckleton, et al. (1) appear to advocate an approach for estimating profile frequencies similar to one that has been widely accepted in the forensic community, the issues they raise surrounding HWE are trivial and academic. The realization by Buckleton, et al. (1) of certain limitations of HWE testing is far from novel, and there are technical problems in both of their experiments due to improper description of their experiments and due to hidden alterations of nuisance parameters of the resulting data from their experiments.

At least since 1970 (see Ward and Sing (13)), there have been many studies suggesting that reasonable levels of departures from HWE are practically impossible to detect (with high power) with data from a single population, unless the sample sizes are prohibitively large. In fact, some of the references cited by Buckleton, et al. (1) address the subject as well. Thus, there would seem to be no need for the authors to conduct experiments to demonstrate what is already known. It is also interesting that again Buckleton, et al. (2) did not cite several critical studies on this topic. For example, we refer them to Chakraborty and Zhong (14).

Buckleton, et al. (1) do not correctly present the findings of Budowle, et al. (2). Contrary to the assertion by Buckleton, et al. (1) "that samples of this size are unlikely to provide any evidence of departures from equilibrium" Budowle, et al. (2) show that ***in spite of the limited sample size*** (sic) significant departures from HWE were found. The Budowle, et al. (2) paper attributed this finding to multiple tests. Thus, analyses on the sample sizes studied by Budowle, et al. (2) can indeed (at least sometimes) detect departures from HWE.

The Buckleton, et al. (1) study contains several flaws. The hypothesis of HWE is a relationship of expected genotype frequencies with allele frequencies. In the statistical literature, this is described as a hypothesis testing problem with allele frequencies appearing as nuisance parameters (15,16). Fisher's exact test, or a variant of the test, is formulated so that the perturbed genotype frequencies are constructed keeping the allele frequencies unaltered. Buckleton, et al.'s (1) description of their first experiment ("additional copies of a single genotype were added to the array, the rest of the array being left unchanged") does not appear to keep the allele frequencies unchanged. Thus, the results in Figs. 1*a* and 1*b* do not appear to be independent of the allele frequencies.

The error in the construction of the second experiment is more egregious. For the coancestry measure of genetic distance, the consequent time of separation of the subpopulations is based on an evolutionary model. However, Buckleton, et al. (1) do not allow for mutation in their experiment. The current voluminous literature on evolution of repeat polymorphisms indicate that the short tandem repeat loci do not fit a model where no mutation is allowed. Furthermore, in different replications, there are obvious changes in allele frequencies within subpopulations, as well as in the combined sample. Thus, the power computations shown in Table 2 are again affected by allele frequency changes.

Although we generally agree with Buckleton, et al. (1) about the process for computing DNA profile frequencies, their specific recommendations regarding the NRC II statistical approaches are misleading, if not partially erroneous. As explicitly stated in the NRC II Report, formulae 4.10a and 4.10b are intended to be applied when the sources of the matched profiles are assumed to be from the same reference population; thus, use of the formulae for a different population cannot be supported by any population genetic argument.

In conclusion, Buckleton, et al. (1) fail to take cognizance of the scientific literature, and their studies are technically flawed. They revisit an old and familiar issue lending nothing new to the topic. Their final recommendations to rectify a perceived ***problem*** already are well-practiced and thus are superfluous.

References

1. Buckleton J, Triggs CM, Curran JM. Detection of deviations from genetic equilibrium—a comment on population data on the thirteen CODIS core loci, Budowle, et al. (1999). J Forensic Sci 2001;46(1) 198–200.
2. Budowle B, Moretti TR, Baumstark AL, Defenbaugh DA Keys KM. Population data on the thirteen CODIS core short tandem repeat loci in African Americans, US Caucasians, Hispanics, Bahamians, Jamaicans, and Trinidadians. J Forensic Sci 1999;44:1277–86.
3. Chakraborty R, Jin L Heterozygote deficiency, population substructure and their implications in DNA fingerprinting. Hum Genet 1992; 88:267–72.
4. Chakraborty R, Jin L, Zhong Y, Srinivasan MR, Budowle B. On allele frequency computation from DNA typing data. Int J Leg Med 1993; 106:103–6.
5. Devlin B, Risch N. A note on Hardy-Weinberg equilibrium of VNTR data using the FBI's fixed bin method. Amer J Hum Genet 1992; 51:549–53.
6. Weir BS. Independence of VNTR alleles defined by fixed bins. Genetics 1992;130:873–87.
7. National Research Council II Report. The Evaluation of Forensic Evidence. National Academy Press, Washington, 1996.
8. Budowle B, Chakraborty R, Carmody G, Monson KL. Source attribution of a forensic DNA profile. Forens Sci Comm 2000;2(3) Available: *http://www.fbi.gov/programs/lab/fsc/current*
9. Chakraborty R, Stivers DN, Su B, Zhong Y, Budowle B. The utility of STR loci beyond human identification: implications for the development of new DNA typing systems. Electrophoresis 1999;20:1682–96.
10. DNA Advisory Board. Statistical and population genetics issues affecting the evaluation of the frequency of occurrence of DNA profiles calculated from pertinent population database(s) (approved February 23, 2000). Forens Sci Comm 2000;2(3) Available: *http://www.fbi.gov/programs/lab/fsc/current*

11. Smith JAL, Budowle B. Source identification of body fluid stains using DNA profiling. In: Proceedings from the Second European Symposium on Human Identification, 1998. Madison, WI: Promega Corporation, 1998, 89–90.
12. Chakraborty R, Srinivasan MR, Daiger SP. Evaluation of standard error and confidence interval of estimated multilocus genotype frequencies, and their implication in DNA forensics. Am J Hum Genet 1993; 52:60–70.
13. Ward RH, Sing CF. A consideration of the power of the χ^2 test to detect inbreeding effects in natural populations. Amer Nat 1970;104:355–65.
14. Chakraborty R, Zhong Y. Statistical power of an exact test for Hardy-Weinberg proportions of genotype data at a multilallelic locus. Hum Hered 1994;44:1–9.
15. Mitra SK. On the limiting power of the frequency chi-square test. Ann Math Stat 1958;29:1221–33.
16. Haber M. Detection of inbreeding effects by the χ^2 test on genotypic and phenotypic frequencies. Am J Hum Genet 1980;32:754–60.

Bruce Budowle, Ph.D.
FBI
935 Pennsylvania Ave. NW
Washington, DC 20535.
Ranajit Chakraborty, Ph.D.
University of Texas—Houston Health Science Center
P.O. Box 20334
Houston, TX 77225