Accountability in Research, Vol. 3, pp. 55–67
Reprints available directly from the publisher
Photocopying permitted by license only

The Use of DNA Typing in Forensic Science

Laurence D. Mueller

Department of Ecology and Evolutionary Biology, University of California, Irvine, California 92717

Controversy has surrounded the use of DNA typing in forensics especially with respect to providing estimates of the frequency of genetically similar individuals in human populations. To address these concerns the National Academy of Sciences has recently produced a report in which specific recommendations are made to alleviate some of this controversy. The reaction of the forensic community to this report is discussed along with many specific concerns about the interaction of the legal and scientific communities.

INTRODUCTION

In 1985 Alec Jeffreys described a class of genetic markers known as VNTR's (variable number of tandem repeat). These genes have small core units of DNA that are repeated in tandem several to hundreds of times. Examinations of many different people revealed a large array of different forms of these genes, or alleles, each with a different number of tandem repeat units. This observation lead Jeffreys to immediately suggest that these markers might be valuable for the identification of individuals since it appeared that different individuals tended to have different arrays of these VNTR's.

It was not long before these genetic markers were used in a forensic setting in England (the Pitchfork case which is the subject of the book, *The Blooding*, Wambaugh, 1989). An equally important benchmark in the development of DNA typing techniques in forensics came in 1989 with the case of People v. Castro in New York City. In this case the defense raised (largely through the efforts and insights of their expert Dr. Eric Lander) a large number of questions about the reliability of DNA typing techniques as practiced by Lifecodes Corporation (Lander, 1989). Since Castro the major areas of scientific contention have been with the manner in which statistical weight is assigned to DNA evidence.

The concerns of the scientific community have continued and grown since the Castro case leading to numerous challenges in the U.S. and Canadian court systems. While many trial courts have admitted DNA evidence, today there are a number of state supreme court (New Hampshire v. Vandebogart; Massachusetts v. Lanigan; Minnesota v. Schwartz) or appeal court decisions (New Mexico v. Anderson; California v. Barney, Howard) which have disallowed the use of DNA typing techniques based on the lack of scientific consensus on numerous issues crucial to the proper application of DNA typing techniques.

In recognition of the importance of DNA typing to forensics and the growing

scientific debate, the National Research Council (NRC) assembled a committee of scientists and legal scholars in 1990 to consider the current use of DNA typing techniques in forensics and offer recommendations for the solution of the most pressing problems. The report was released in April, 1992 (NRC, 1992) but it appears that the level of debate has only increased as a result of this report.

In this paper I will review some of the most important issues facing DNA typing in forensics and how the National Research Council has suggested solving them. In addition I consider some more general issues facing members of the scientific community when their opinions are solicited in criminal litigation.

Evaluating the Weight of DNA Evidence

The typical forensic case usually involves the comparison of genetic patterns (or genotypes) from two samples: evidence and suspect. The evidence may be semen from a rape victim, blood or other tissue samples found at the scene of a crime. The forensic laboratory must first determine if the genetic patterns from these two samples are sufficiently similar that they could have come from the same source. Although this is not a trivial task, I will not discuss the various procedures for declaring matches here (but see Thompson and Ford, 1991; Budowle et al., 1991).

Once the laboratory has declared that two genetic patterns match, it is necessary to determine the rarity of the matching characteristics in order to evaluate the weight that should be attached to this declaration. If 1 in 10 people would be expected to share the matching characteristics in a particular case, this evidence would deserve less weight than if the chances of a coincidental match were 1 in a million. In the course of developing statistical methodologies that would help to assess these probabilities it has been generally recognized that when there are two equally likely assumptions the one that is more conservative should be chosen. Here the conservative choice is the one less likely to underestimate the probability of observing this genetic pattern. Conservative choices minimize the likelihood of an estimation error that disadvantages an accused suspect.

There are two independent phenomena that could cause a match to be declared between an innocent person and the DNA found in an evidence sample (assume the evidence sample really contains DNA from the true criminal). The first possibility is that the suspect is coincidentally identical to the true criminal. The theories and techniques of population genetics should be useful in determining this probability. The second phenomenon is that the laboratory has incorrectly matched these two (different) samples due to some error. These types of errors are usually called false positives.

For juries it is of little significance what causes an innocent person to match, what matters is how often such matches might be expected. In the next two sections I review the methods that have been used for evaluating each of these phenomena (coincidentally matching genotypes and false positives) and discuss where controversies have arisen.

False Positives. Many supporters of the use of DNA typing techniques have argued from first principles that the frequency of false positives would be vanishingly rare. These arguments are premised on the (mistaken) belief that a false

positive would arise from multiple laboratory errors which would make different genetic patterns appear to be the same. In fact there are much simpler errors that can result in false positives. For instance, if DNA from a suspect is erroneously placed in two vials and one is labeled suspect and the other evidence then an analysis of these samples should give the appearance of a perfect match.

False positives has now been documented in a variety of laboratories. The private laboratory, Cellmark, has made two false positives in a comparison of 100 forensic like samples (Thompson and Ford, 1991). The FBI found a match in their data base between two different people who in fact had very different genotypes (Table 1; Sullivan, 1992). The FBI discovered this erroneous match while doing a computer search of their data bases for matching DNA profiles. These samples were obtained from the Texas College of Osteopathic Medicine. It was suggested by the director of this lab that this match may have been the result of a sampling handling error which inadvertently duplicated one of the samples.

Even top research laboratories make errors which can lead to false positives. In K. Kidd's laboratory at Yale University a study of 128 people from Amerindian populations resulted in one false positive (Table 2). These pairs of matching samples was inferred to be inadvertent duplicates due to a large number of additional loci at which these two samples matched.

These types of problems have lead the NRC committee on DNA typing to recommend that all forensic laboratories take part in regular blind proficiency testing administered by an outside agency. The committee recommends that the results of these proficiency tests be presented to the jury in a quantitative fashion (Lempert, 1991). It is worth pointing out that a laboratory that does 1000 proficiency tests (which could potentially result in 1000 false positives) and makes no false positives can not claim that its error rate is zero (or in practice much less than the genotype frequency reported in the case which may be one in millions (Saks and Koehler, 1991). What would be permissible is a statistical statement such as "the probability of a false positive in this laboratory is less than or equal to x, where $\alpha = (1 - x)^N$, N is the number of proficiency tests completed, and α is the chance that the true error rate is greater than x, usually assumed to be 0.05 by convention. Thus, for the laboratory that has completed 1000 tests without error $x = 0.003$.

For instance, as a July 1991 the FBI had completed 417 internal proficiency tests. Even if we assume that all of these could have produced false positives and that none were made, the best we can say is that with 95% confidence the FBI rate of false

TABLE 1. False positive in the FBI data base. The measured band sizes (in kilobases) for two people in the FBI data base are given below. These people show different profiles on retesting.

LOCUS I.D. #	D2S44	D17S79	D1S7	D4S139	D14S13
F4381	1.126	1.889	4.873	7.926	4.066
	1.033	1.889	3.528	5.372	3.179
F3988	1.115	1.876	4.819	7.661	4.005
	1.022	1.876	3.472	5.216	3.146

TABLE 2. The false positive in the Amerindian data base.

	INDIVIDUAL BAND SIZES (kilobases)	
I.D. # Tribe LOCUS	AI053 Karitiana	AI055 Karitiana
D2S44	10.53	10.54
	10.53	10.54
D17S79	3.98	3.96
	3.98	3.96
D14S1	3.85	3.85
	3.85	3.85
D14S13	8.23	8.20
	4.98	4.96
D18S27	4.61	4.60
	4.61	4.60
LILA5	7.56	7.52
	6.29	6.28
DXYS14	2.30	2.29
	2.16	2.15
	1.42	1.42

positives is less than or equal to 1 in 140. Similarly, Cellmark estimates their rate of false positives with a combination of the CALCD proficiency results and their own internal proficiency tests as 1 in 139 (through 1991).

In general, it appears that the available data suggest that false positives may happen at rates of 1 in hundreds or perhaps 1 in thousands. If a jury truly understood these limitations then it should hardly matter to them if the probability of finding another person who would match the genetic profile of the evidence were 1 in a 10,000 or 1 in 10 billion, since it is so much more likely that a laboratory will incorrectly match two evidence samples. This perspective is worth keeping in mind since a major objection to the method of estimating genotype frequencies proposed by the NRC is that the method is "absurdly" conservative (Morton, 1992). In fact, what may be absurd are attempts to salvage controversial methods for estimating genotype frequencies of 1 in millions when acceptable methods that produce frequencies of 1 in thousands are available.

Genotype Frequencies. A DNA profile or genotype does not represent the whole human genome. Usually, this profile is just the genetic information from four (or fewer) loci out of hundreds of thousands in the human genome. Consequently, it is certainly possible that multiple people may have the same profile. This is especially so when the laboratories match criteria is considered, since these lump a variety of DNA bands into a common bin even though some of these may be from different alleles. The task for population geneticists, then, is to construct methods for estimating the likelihood of finding people at random in the reference population that might match the DNA profile or genotype in the evidence sample.

The typical method that had been used prior to the NRC report of April 1992 was

the product rule. This method assumed that bands or alleles at a single locus were independent (within a population) and that bands at different loci were also independent (again within a population). These two levels of independence translate into the population genetic assumptions known as Hardy-Weinberg and linkage equilibrium, respectively.

While there are numerous evolutionary forces that may invalidate these assumptions, e.g. small population size, inbreeding, and natural selection, there has been essentially one phenomena that has concerned population geneticists and that is population substructure. If, for instance, a population like Caucasians in the United States is not a randomly mating, genetically homogeneous population but is in fact composed of genetically differentiated subpopulations (like Swedes, Italians, etc.) then the assumptions of independence made in the product rule would be incorrect.

There are two distinct methodologies that could be used to address the problems of population substructure. The first method (Devlin, Risch and Roeder, 1991; Chakraborty and Daiger, 1991; Weir, 1992b,c; Risch and Devlin, 1992) suggests that a statistical analysis of the possibly heterogeneous data bases will be sufficient to detect the compromising effects of population substructure. The second method specifically rejects this approach for several reasons. First, these statistical methods generally lack power (Cohen, Lynch and Taylor, 1991; Green and Lander, 1991; Lewontin and Hartl, 1991; NRC, 1992) and available empirical data suggests the existence of population substructure is widespread (Balazs et al., 1992; Krane et al., 1992). Secondly, the people in the data base have not been sampled at random and thus there is no expectation that the mixture of subgroups within these samples is representative of any real population. In light of these concerns the second method proposes that population subgroups be sampled directly. This procedure provides data which can be used to directly assess the genetic differences between subgroups with greater statistical power. The subgroup samples can also serve as data base samples for the estimation of genotype frequencies.

Proponents of the first method have suggested that even if populations like Caucasians, blacks and Hispanics are composed of subgroups the potential magnitude of the errors that will be made by using the product rule are small and can be tolerated (Chakraborty and Kidd, 1991; Weir, 1992). However, those that support the second method are concerned about potentially large errors that can result from the product rule. For instance data collected by Kidd et al. (1991) shows a seven locus match between two related people from the Karitiana population despite Chakraborty and Kidd's (1991) claim that no such matches exist (Table 3). Of even greater significance is the six locus match in this data base between two unrelated people (Table 3). The product rule predicts the frequency of such patterns should be 1 in 96 million while the observed frequency is 1 in 37.

Some have suggested that arguments over the rarity of DNA profiles (or genotypes) are pointless because, under any reasonable assumptions, the profile will be rare enough to provide overwhelming evidence of guilt. It would be dangerous to accept this argument uncritically, however, because the manner in which DNA evidence is used varies so greatly from case to case. For instance, in a case where there is physical evidence other than DNA, such as eye witness evidence, even a probability of 1 in 100 might be considered rare enough to clinch the case. However,

TABLE 3. Matching genetic patterns between different people at six or seven VNTR loci.

	INDIVIDUAL BAND SIZES (kilobases)			
I.D. #	AI029	AI044	AI099	AI108
Tribe	Karitiana	Karitiana	Surui	Maya
LOCUS				
D2S44	10.68	10.69	10.60	10.82
	10.68	10.69	10.29	10.27
D17S79	3.79	3.93	3.57	3.56
	3.97	3.93	3.32	3.29
D14S1	3.82	3.82	4.57	4.55
	3.82	3.82	3.79	3.79
D14S13	10.65	10.64	12.79	12.78
	9.25	9.24	5.21	5.18
D18S27	4.59	4.59	4.75	4.75
	4.59	4.59	4.52	4.55
LILA5	12.58	12.70	7.33	7.32
	11.19	11.04	7.33	7.32
DXYS14	3.48	3.44		
	2.71	2.70		
	1.78	1.77		
	1.38	1.38		

The data are from Kidd et al. (1991), the matching criteria applied was the FBI's although the patterns would match under more stringent criteria. Individual AI029 matches AI044 and AI099 matches AI108.

if DNA is the only substantial evidence in a death penalty case then numbers like 1 in 10,000 may not be rare enough. Moreover, while most forensic laboratories attempt to assay four loci, cases appear with results from only three, two and even a single locus. With smaller numbers of loci, statistics are not always guaranteed to be vanishingly rare even when the assumptions of Hardy-Weinberg and linkage equilibrium are used in computing frequencies.

National Research Council Report on DNA Typing

The third chapter of the NRC report discusses many of the issues raised in this paper. In the NRC report the two opposing views of the status of population substructure are described. The report makes several important recommendations which I summarize below.

- The report correctly identifies the importance of quantifying laboratory false positives as error rates. The report calls for all forensic laboratories to take part in regular, blind proficiency tests that would be administered by an independent agency. The results of these tests should be quantified and reported to juries.
- The report has clearly sided with those population geneticists that have called for sampling of population subgroups. The major recommendation in this

report is to gather samples on 15–20 human subpopulations. These data bases would then serve as a basis for future calculations of genotype frequencies.

- Once the 15–20 data bases have been established the report describes a new procedure for calculating genotype frequencies (see also Lander, 1991). This method is called the ceiling principle. This technique would survey all 20 data bases for the frequency of each allele. Each allele would be assigned the highest frequency from these 20 data bases or a value of 0.05 whichever is greatest. The individual allele frequencies that result from this procedure would then be multiplied by each other and the appropriate constant.

The report recognizes that some time will be required to collect 15–20 new data bases. Consequently, they have provided suggestions for how frequencies might be calculated during this interim period. Of course laboratory error rates can always be reported. The interim recommendations, as I will call them, only concern the estimation of genotype frequencies.

- Laboratories should first check their existing data bases for people that match the genetic profile of the evidence sample. This technique, known as the counting method, while unsophisticated has the advantage that it does not require the assumptions of Hardy-Weinberg and linkage equilibrium and thus is robust to departures from these assumptions.
- Secondly, the laboratory should perform a modification of the ceiling principle. There are three major modifications to the ceiling principle that would be used during this interim period. (1) At **least** three data bases should be used rather than 15–20. (2) In each of these data bases the upper 95% confidence limit of the allele frequency is used rather than the observed frequency. (3) The lower limit for an individual allele frequency is set to 10% rather than 5%.

It must be emphasized that NRC report is not fully explicit on how to implement every detail of their recommendations. Perhaps correctly recognizing that these will require additional discussion by the scientific community. For instance, the report suggests that each data base used in the interim ceiling principle be checked by agreement to Hardy-Weinberg and linkage equilibrium. While the specific statistical test that might be used is not specified, presumably it might be one of the available tests used for this purpose previously (Devlin, Risch, Roeder, 1991; Weir, 1992b,c). However, the NRC does not describe what should be done if there is a departure from either of these levels of independence. One possible response could be to not use the offending data base. However, the practical effect of this is to reduce the number of data bases available for computing the ceiling statistic. This can only have the effect of producing a rarer frequency (and hence possibly prejudicial to the defendant) than if the data base had been kept. My own interpretation is to use all data bases whether they satisfy these tests or not. One rationale for this interpretation is that the result from the counting method will be robust to these departures anyway.

Recent court cases on the admissibility of DNA typing evidence have focused on the interim recommendations since these are the procedures that many laboratories have begun using. This opens up the possibility that if courts rule the interim procedures are acceptable then there would be no incentive to complete the study of

additional subpopulations and implement the final recommendations of the NRC. This is especially troubling since the manner in which various laboratories have chosen to interpret these interim recommendations is highly variable. There are three major deficiencies in the manner in which some forensic laboratories have applied these recommendations.

Data Bases Used: The NRC has recommended that at least three data bases be used when the interim ceiling principle is used. The words "at least" clearly implies that more may be used if they are available. The ultimate recommendation is that ceiling principle computations be based on 15–20 data bases. In the spirit of this recommendation it would seem appropriate to use more data bases than three if they exist. Several labs have not used data bases that are available to them, claiming that these are not major ethnic populations (Budowle and Monson, 1993).

Counting Method: The NRC report states the forensic labs should check their data base for people which match the genetic pattern of the evidence sample. Weir (1993) describes how this type of information can be used to provide frequency estimates. Many laboratories have ignored this recommendation. It should be emphasized that this recommendation is part of the interim methods to be used until the 15–20 data bases are available for the ceiling principle. Since there is no reason to consider the interim ceiling principle applied to only three data bases conservative, calculation of the counting method is a needed check.

Switching Bins: Several laboratories which had used the FBI's fixed bin system with the product rule have switched to a floating bin when computing the ceiling principle. Acceptable statistics can be produced with both fixed and floating bins. However, as discussed in the NRC report the size of a floating must be closely tied to the reproducibility of measurements within and between gels. The size of the floating bins that have been used by some labs are only about half the appropriate size thereby producing allele frequency estimates that are too small.

Critiques of the NRC Report. Many of the scientists that had supported the use of the product rule for forensic calculations have been highly critical of the NRC report (Budowle and Monson, 1993; Chakraborty et al., 1993; Devlin, Risch and Roeder, 1993; Weir, 1992a, 1993). Many of these critiques have chosen to focus on the ceiling principle, saying little about the major recommendation of studying 15–20 populations. Weir (1993), has called the ceiling principle, "The major population genetic recommendation of the report. . .".

The major arguments made against studying an additional 15–20 subpopulations have been made previously and were acknowledged in the NRC report (Lander, 1993). These arguments (see Devlin, Risch and Roeder, 1993) can be summarized as follows: existing data usually show no statistically significant departures from the independence and the possible magnitude of errors caused by population substructure are likely too small in any case. On this particular point I can see no hope for resolution of the disagreement among population geneticists until data on population subgroups becomes available.

Other arguments challenging the soundness of the NRC report are less substantive in nature. For instance Weir (1993) refers to the following passage in the NRC report (pg 76), "If the pattern occurred in 1 of 100 samples, the estimated frequency would be 1%, with an upper confidence limit of 4.7%. (The upper bound is the

traditional 95% confidence limit, whose use implies that the true value has only a 5% chance of exceeding the upper bound.)" Weir then goes on to characterize this result as incorrect and state that the correct limit can be found by applying an approximation based on normal theory. In fact the number cited in the NRC report is the confidence limit that follows from applying the binomial distribution which will be more accurate for small frequencies than the normal approximation.

Weir (1993) has suggested that the arbitrary cutoff of 10% produces a discontinuity in allele frequencies. The error made in this particular example is that Weir compares the allele frequencies to 10% prior to placing the confidence limits on these frequencies. This problem does not occur if the confidence intervals are placed on the frequencies first and then the values are compared to 10%.

The NRC formula for approximate 95% confidence interval has been called incorrect based on the interpretation of the words "upper 95% confidence interval" (Weir, 1993). In fact there are several possible interpretations of these words including a two tailed interval with a total of 5% probability mass in both tails or a one tailed interval with 5% probability mass in the upper tail. The bottom line is the NRC report gives an exact formula which can be interpreted unambiguously.

The NRC recommends that samples from population subgroups be composed of 100 people. Critics (Devlin, Risch and Roeder, 1993) have stated that these samples are too small. While 100 individuals is certainly too few to conduct tests of Hardy-Weinberg or especially linkage equilibrium and have good power, it is not too few to contrast allele frequencies. Thus, if samples of 100 individuals were available from Japanese, Korean, Chinese and Vietnamese populations one could determine if there were significant differences in allele frequencies between these populations. If not then a data base which was a mixture of these populations, called "Oriental" for instance, might be justified. The exact number of people included in these samples is a relatively minor issue and can not be viewed as a fatal flaw in this report. In fact the NRC report has called for the establishment of an oversight committee which would made detailed recommendations about the particular populations that would be included in these studies. Presumably, this committee could also entertain suggestions to alter and increase the sample sizes used in these studies. It is also worth noting that many data bases used today by forensic laboratories are not much larger than 100 people (e.g. Center for Forensic Science, Toronto, Canada: 97, Black data base; Cellmark: 168, Hispanic data base; Washoe County, Nevada: 139, Black data base).

Another criticism has been the arbitrary lower limits used in the ceiling principle (Budowle and Monson, 1993; Devlin, Risch and Roeder, 1993; Weir, 1993). These limits are seen as being without a scientific basis and thus unacceptable. In statistics it is customary to set certain limits on the levels of type I and type II error. The exact level that these are set to is to some extent arbitrary and represents the scientist's own evaluation of risks he is willing to take when making conclusions based on these statistical analyses. In a similar vein the limits set by the NRC can be viewed as an evaluation of the potential risks of errors if the true allele frequencies are much greater than the lower limit chosen. These types of arbitrary limits are used by the FBI, for instance, when they decide to merge fixed bins with fewer than 5 observations. This limit could have been either 4 or 6 and achieved the same result.

Scientific Scrutiny of Forensic Research

Many of the principles that are utilized in forensic DNA typing have been developed in other research laboratories. Nevertheless, there are a number of important issues that each laboratory must investigate by collecting its own in-house data. Some but not all of the topics which are subject to some degree of original empirical research include,

- establishing the match criteria,
- estimating the variation in repeat measurement on the same and different gels for establishing bin sizes,
- collection of data base samples. In some instances local forensic laboratories may be in the best position to collect samples from local ethnic groups that are unusually concentrated. For instance in Southern California there are relatively large populations from several Southeast Asian Countries. As a result the Orange County Crime laboratory has been able to create data bases for Chinese, Japanese, Korean and Vietnamese populations. Another example would be areas with local Indian populations (Lewontin, 1993).

This type of information is crucial to establishing laboratory protocols and eventually evaluating evidence samples. It is absolutely essential that this type of information be freely available to experts that defense attorneys may consult. A goal of the outside scientists who reviews these data is not only to comment on the specific case but also to discuss the general acceptance of the techniques and procedures utilized by the forensic laboratory. The procedures which foster and facilitate general acceptance are publication and discussion. Consequently, it is equally important that experts who have reviewed scientific evidence from a forensic laboratory be able to discuss their opinions and results of their analyses, if need be in scientific publications.

Obstacles to this process have been encountered by myself and others. These obstacles include: refusal of the forensic laboratory to turn over supporting scientific evidence since they are considered "trade secrets," protective orders on the material reviewed by scientists which limit their ability to discuss the information they have reviewed outside of the court and claims that the material is the intellectual property of the forensic laboratory scientists and thus can not be used in publications until the forensic laboratory scientists have published their own analyses.

The scientific research which has been developed to support the results of forensic testing should be considered no less important than the evidence used to support conclusions in a scientific paper. As such, it should be subject to the same level of scrutiny. If the laboratory personnel feel they need to publish these results (which I think is a good idea) it must be done prior to accepting case work (NRC, Chapter 3).

Expert Witnesses

A tradition in our legal system is to provide opposing counsel the opportunity to cross examine all witnesses. Experts are no exception, and the major focus of many

cross examinations (by both prosecution and defense lawyers) is to attack the credibility of the expert. Since I have only served as a defense expert my first hand knowledge has been limited to interactions with prosecutors although similar problems almost certainly occur for prosecution witnesses. A favorite angle these cross examinations take is to explore the amount of compensation an expert receives for his time and testimony (Roberts, 1992a). The implication of this line of cross examination is that the particular opinion held by the expert is one of convenience since it permits him to collect consulting fees and that without this sort of remunerations neither this expert nor anyone else would hold these opinions. These arguments are unconvincing when it is possible to show that large numbers of other scientists share these concerns. Access to experts is especially important to the defense bar since they will not usually have access to the personnel of forensic laboratories nor are the personnel in forensic laboratories likely to be critical of the techniques used by their own lab. While I and other population geneticists have consulted on cases *pro bono* it is doubtful whether there will be sufficient numbers of scientists willing to do *pro bono* work to provide all the expertise needed by the legal profession. I also doubt that employees of forensic laboratories would be willing to testify if they were required to do this on their "own" time. Thus, there does not appear to be any simple alternative to providing experts with compensation for their time. Consequently, the persistent questions about income are likely to continue since attorney's need to have the option to expose all possible sources of bias no matter how improbable.

There are other practices which I have seen used that appear to have the express purpose of discouraging experts from testifying and which I feel can only have a damaging effect on the relationship between the legal and scientific communities. I outline some of these below.

Junk Mail: Some prosecutors have engaged in letter writing campaigns long after their particular cases were completed (Roberts, 1992b). These letters might be directed to journal editors or administrative personnel at an expert's home institution. The letters may generally contain a rehashing of the prosecutor's dissatisfaction with the scientific testimony of the expert in question. The clear intent of such letters is to make any expert think twice about whether testifying is worth having your colleagues submitted to these verbal barrages. While I recognize that these types of practices are limited to a small number of lawyers, my colleagues do not. Consequently, I think the legal profession has to take a long hard look at these sorts of practices because their effect may be to alienate a large group of scientists from interacting with the legal professionals.

Other scientists has received calls from prosecutors asking that they retract publications (Roberts, 1991). While the exact intentions of these calls is difficult to interpret, they caused concern on the part of the scientist involved.

Critical Reviews: In several cases anonymous reviews of an expert's scientific papers have been turned over in court cases. These reviews are then utilized to attack the credibility of the author by demonstrating that other scientists disagree with him. These reviews have an appearance of objectivity since they were ostensibly made outside of the legal proceedings. However, there have been cases in which negative, anonymous reviews used to attack an expert in court were in fact written by an expert testifying for the opposing side. I think there is an obvious conflict of

interest when a scientist reviews an article written by a litigation opponent, particularly when the review may itself become evidence in the litigation. Ethical scientists should refuse to review papers of litigation opponents unless they are willing to sign the reviews and disclose the conflict to the journal editor.

CONCLUSIONS

DNA typing can be a very powerful tool in forensics. The results of DNA tests will be useful even if a laboratory were to follow all of the NRC's interim recommendations. Such results will only be more powerful when the study of additional subpopulations are completed. The current attacks on the scientific credibility of the NRC report are only likely to prolong the controversy.

ACKNOWLEDGEMENT

I thank W. Thompson and D. Krane for useful comments on the manuscript, and NIH grant AG09970 for research support.

REFERENCES

Balazs, Neuweiler, I.J., Gunn, P., Kidd, J., Kidd, K.K., Kuhl, J., and Mingjun, L. Human population genetic studies using hypervariable loci. I. Analysis of Assamese, Australian, Cambodian, Caucasian, Chinese and Melanesian populations. *Genetics* 131:191–198, 1992.

Budowle, B., Guisti, 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. *Amer. J. Hum. Genet.* 48:841–855, 1991.

Budowle, B., and Monson, K.L. Perspectives on the fixed bin method and the floor approach/ceiling method. In *Proceedings of the 1992 Symposium on Human Identification*, in press, 1993.

Cohen, J.E., Lynch, M., and Taylor, C.E. Forensic DNA tests and Hardy-Weinberg equilibrium: technical comment. *Science* 253:1037–1038, 1991.

Chakraborty, R., and Daiger, S.P. Polymorphism at VNTR loci suggest homogeneity of the white population of Utah. *Human Biology* 63:571–587, 1991.

Chakraborty, R., and Kidd, K.K. The utility of DNA typing in forensic work. *Science* 254:1735–1739, 1991.

Chakraborty, R., Srinivasan, M.R., Jin, L., and de Andrade, M. Effects of population subdivision and allele frequency differences on interpretation of DNA typing data for human identification. In *Proceedings of the 1992 Symposium on Human Identification*, in press, 1993.

Devlin, B., Risch, N., and Roeder, K. No excess of homozygosity at loci used for DNA fingerprinting. *Science* 249:1416– 1420, 1990.

Devlin, B., Risch, N., and Roeder, K. Statistical evaluation of DNA fingerprinting: a critique of the NRC's report. *Science* 259:748, 1993.

Green, P., and Lander, E.S. Forensic DNA tests and Hardy-Weinberg equilibrium: technical comment. *Science* 253:1038–1039, 1991.

Jeffreys, A.J., Wilson, V., and Thein, S.L. Hypervariable "minisatelite" regions in human DNA. *Nature* 314:67–73, 1985.

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* 63:775–794.

Krane, D., Allen, R.W., Sawyer, S.A., Petrov, D.A., and Hartl, D.L. Genetic differences at four DNA typing loci in Finnish, Italian, and mixed Caucasian populations. *Proc. Natl. Acad. Sci. USA* 89:10583–10587, 1992.

Lander, E.S. DNA fingerprinting on trial. *Nature* 339:501–505, 1989.

Lander, E.S. Research on DNA typing catching up with court room application. *Am. J. Hum. Genet.* 48:819–823, 1991.

Lander, E.S. Letter to the editor. *Science* 260:1221, 1993.

Lempert, R. Some caveats concerning DNA as criminal identification evidence: with thanks to the Reverend Bayes. *Cardozo Law Review* 13:303–341, 1991.

Lewontin, R.C. Letter to the editor. *Am. J. Hum. Genet.* 52:205, 1993.

Lewontin, R.C., and Hartl, D.L. Population genetics in forensic DNA typing. *Science* 254:1745–1750, 1991.

Morton, N.E. Genetic structure of forensic populations. *Proc. Natl. Acad. Sci. USA* 89:2256–2260, 1992.

National Research Council, *DNA Technology in Forensic Science*. National Academy Press, Washington, D.C., 1992.

Risch, N., and Devlin, B. On the probability of matching DNA fingerprints. *Science* 255:717–720, 1992.

Roberts, L. Fight erupts over DNA fingerprinting. *Science* 254:1721, 1991.

Roberts, L. Hired guns or true believers? *Science* 257:735, 1992a.

Roberts, L. Prosecutor v. scientists: a cat-and-mouse relationship. *Science* 257:733, 1992b.

Saks, M.J., and Koehler, J.J. What DNA "fingerprinting" can teach the law about the rest of forensic science. *Cardozo Law Review* 13:361–372, 1991.

Sullivan, P.J. Letter to the editor. *Science* 256:1743–1744, 1992.

Thompson, W., and Ford, S. The meaning of a match: sources of ambiguity in the interpretation of DNA prints. In, *Forensic DNA Technology*, M.A. Farley and J.J. Harrington (eds), Lewis, Chelsea, MI, 1991.

Wambaugh, J. *The Blooding*. William Morrow, New York, 1989.

Weir, B.S. Population genetics in the forensic DNA debate. *Proc. Natl. Acad. Sci.* 89:11654–11659, 1992a.

Weir, B.S. Independence of VNTR alleles defined as fixed bins. *Genetics* 130:873–887, 1992b.

Weir, B.S. Independence of VNTR alleles defined as floating bins. *Am. J. Hum. Genet.* 51:992–997, 1992c.

Weir, B.S. Forensic population genetics and the NRC. *Am. J. Hum. Genet.* 52:437–440, 1993.