SCIENCE

Published by the **American Association for the Advancement of Science (AAAS),** *Science* serves its readers as a forum for the presentation and discussion of important issues related to the advancement of science, including the presentation of minority or conflicting points of view, rather than by publishing only material on which a consensus has been reached. Accordingly, all articles published in *Science*—including editorials, news and comment, and book reviews—are signed and reflect the individual views of the authors and not official points of view adopted by the AAAS or the institutions with which the authors are affiliated.

Membership/Circulation
Director: Michael Spinella
Fulfillment: Marlene Zendell, *Manager*; Gwen Huddle, *Assistant Manager*; Mary Curry, *Member Service Supervisor*; Pat Butler, Helen Williams, Robert Smariga, *Member Service Representatives*
Promotions: Dee Valencia, *Manager*; Laurie Baker, Hilary Baar, *Assistants*
Research Manager: Kathleen Markey
Financial Analyst: Jacquelyn Roberts
Administrative Assistant: Nina Araujo de Kobes

Advertising and Finance
Associate Publisher: Beth Rosner
Advertising Sales Manager: Susan A. Meredith
Display Recruitment Sales Manager: Janis Crowley
Financial: Deborah Rivera-Wienhold, *Manager*; Julie Eastland, *Senior Analyst*; Andrew Joyce, *Junior Analyst*
Marketing Manager: Laurie Hallowell
Traffic Manager: Tina Turano
Traffic Manager (Display Recruitment): Daniel Moran
Line Recruitment: Michele Pearl, *Manager*; Millie Muñoz-Cumming, *Assistant*
Reprints Manager: Corrine Harris
Permissions Manager: Arlene Ennis
Advertising Assistants: Allison Pritchard, Kelly Nickerson, Debbie Cummings
Send materials to *Science* Advertising, 1333 H Street, NW, Washington, DC 20005, or FAX 202-682-0816.

SALES: Northeast/E. Canada: Fred Dieffenbach, Rt. 30, Dorset, VT 05251; 802-867-5581, FAX 802-867-4464 • **Mid-Atlantic:** Richard Teeling, 28 Kimberly Place, Wayne, NJ 07470; 201-904-9774, FAX 201-904-9701 • **Southeast:** Mark Anderson, 1915 Brickell Avenue, Ste. CC-1, Miami, FL 33129; 305-856-8567, FAX 305-856-1056 • **Midwest:** Donald Holbrook, 1110 North Harvey, Oak Park, IL 60302; 708-386-6921, FAX 708-386-6950 • **West Coast/W. Canada:** Neil Boylan, 2847 Fillmore, Ste. 3, San Francisco, CA 94123; 415-673-9265, FAX 415-673-9267 • **Germany/Switzerland/Austria:** Ric Bessford, World Media Services, Leopoldstrasse 52, 8000 Munich 40, Germany; +49-089-39-00-55, FAX +49-089-39-00-15 • **Japan and Far East:** Massy Yoshikawa, Orient Echo, Inc., 1101 Grand Maison Shimomiyabi-cho 2-18, Shinjuku-ku Toyko 162, Japan; +3 3235-5961, FAX +3 3235-5852 • **UK, Scandinavia, France, Italy, Belgium, The Netherlands:** Andrew Davies, 1 Newbridge View, Micklehurst Road, Mossley, Ashton-under-Lyne, OL5 9SE, Great Britain; +44-457-838-519, FAX +44-457-838-898 • **Other:** Contact *Science* Advertising: 202-326-6544, FAX 202-682-0816.

Information to Contributors appears on pages 36–38 of the 3 January 1992 issue. Editorial correspondence, including requests for permission to reprint and reprint orders, should be sent to 1333 H Street, NW, Washington, DC 20005. **Science Telephone:** 202-326-6500. London office: 071-435-4291. **Subscription/Member Benefits Questions:** 202-326-6417. **Other AAAS Programs:** 202-326-6400.

LETTERS

DNA Fingerprint Matches

I am writing to comment on two aspects of the report "On the probability of matching DNA fingerprints" by Neil J. Risch and B. Devlin (7 Feb., p. 717). Risch and Devlin searched several large databases to determine whether there were any samples with matching patterns across a number of gene loci. They found "the probability of a matching DNA profile between unrelated individuals to be vanishingly small. . . ."

Last summer I was trying a Federal Bureau of Investigation (FBI) case, *Minnesota* v. *Johnson* (*1*), and examined three FBI databases, C-3 (Caucasian), B-4 (black), and H-3 (Hispanic). During my examination, I discovered 25 apparent matches. Before my examination, the existence of these matches had been known by only a few individuals connected with the FBI. Bruce Budowle of the FBI subsequently testified in *Minnesota* v. *Johnson* that he was aware of these matches and that they had been discovered when the FBI examined its database with its computer matching program. The FBI was able to verify that most of these matches occurred because the Texas College of Osteopathic Medicine submitted more than one blood sample from the same individual. One false match was the result of a sample handling error.

The FBI also discovered three sets of matching samples from Florida. These samples were from the black and Hispanic databases. The FBI was not able to verify that the Florida matches were the result of duplicate submissions from the same individual or of submissions from identical twins. Budowle then asked Cellmark Diagnostics (Germantown, Maryland) to examine the matching samples. Its probes also yielded unclear results. The Florida matches were then deleted from the databases, even though there was no explanation for their occurrence.

The FBI again revised its databases in January 1992. The new databases are designated C-4, B-5, and H-4. Budowle testified (*2*) that all the matches have been edited out of these new databases and that this removal of matches is justified because it is not possible for two individuals to yield identical profiles when as many as seven probes are used. This may or may not be correct. My first point is this: Of what scientific value is a paper that seeks to draw any conclusion from the fact there are no matches in a database when the matches have been removed from the database before the analysis is done? The FBI's removal of matches from its databases before giving them to outside scientists guarantees that those scientists' conclusions will support the FBI's "self-fulfilling prophecy."

This is not an isolated practice. Budowle testified in *United States* v. *Yee* (*3*) that the FBI ran its match program over its South Carolina black database and found a large number of matches. The FBI's record-keeping was such that it could only speculate as to the cause of these matches. Again, the FBI removed them from its database.

The existence of individuals who match across a number of loci is not unprecedented. Kenneth Kidd's Amerindian (Karitiana) data (*4*) show a seven-probe match between two individuals, a four-probe match between another two individuals, and a number of three-probe matches. These matches occurred in a database of 54 donors from one Indian village. Despite this fact, which is well known to the FBI, the FBI chose simply to remove apparent matches from its databases. The apparent justification of this practice is that it eliminates the necessity of keeping records about the source of data. It is troubling to think that this approach has acceptance among scientists.

My second point relates to the match window used by Risch and Devlin. As described in their report (p. 718), they used "a bound of 2.4% of the mean of the two fragment sizes." The FBI uses a match window described as ±2.5% of the mean of the two fragment sizes. While these two descriptions sound the same, the match window used by the FBI in casework is more than twice as large as the window used by Risch and Devlin. A simple example illustrates this point. Assume that the FBI is examining two bands, a known band at 6000 base pairs and an unknown band at 6300 base pairs. The FBI would calculate the mean of these two bands to be 6150 and the match window to be ±2.5% of 6150, running from 5996.25 base pairs to 6303.75 base pairs. As a result the two bands would be within the FBI's match window and therefore would be said to "match." If one applies the formula of Risch and Devlin (p. 718), where x is 6000 and y is 6300, the result is 0.04878. This result is more than twice 0.024 and would not be a match according to Risch and Devlin. Yet a reader of the descriptions of the two match windows could easily believe they are the same.

It is my understanding that Risch and Devlin did not know that matches had been

removed by the FBI before they did their analysis or that they were using a significantly smaller match window than the FBI uses. I do not question their integrity. However, these factors place in doubt the conclusions they reached about the validity of FBI DNA fingerprinting data.

Patrick J. Sullivan
Senior Attorney,
Assistant Hennepin County Public Defender,
317 2nd Avenue South, Suite 200,
Minneapolis, MN 55401–0809

REFERENCES

1. *Minnesota* v. *Johnson*, S.I.P. No. 89-072014 (Minnesota District Court, Fourth Judicial District, 1991).
2. *Minnesota* v. *Alt*, No. K4-90-2437 (Minnesota District Court, Third Judicial District, 1992).
3. *United States* v. *Yee*, 129 Federal Rules Decisions 629 (Northern District of Ohio, 1990).
4. K. Kidd, F. L. Black, M. Weiss, I. Balazs, K. K. Kidd, *Human Biol.* **63**, 775 (1991).

Response: The purpose of our analysis (*1*) of the probability of matching DNA fingerprints in the FBI and Lifecodes Corporation (Stamford, Connecticut) databases was to determine whether match probabilities were independent across loci and to calculate, under the assumption of independence, the probability of two random, unrelated individuals matching at multiple loci. Sullivan raises two issues regarding our analysis: (i) the exclusion of matching samples by the FBI from its database before our analysis and (ii) our criteria for defining band patterns as matching with the FBI database.

It is true that we were unaware that the FBI had excluded matching individuals from its database before we received it. We do not view the FBI's practice as unusual, however. Scientists performing large-scale studies always evaluate their databases for unusual observations: outliers, misrecordings, data-processing errors, and the like. From the standpoint of the FBI, it is understandable why it would remove duplications of the same person. Such duplications induce error in the estimates of general population bin frequencies, although the error is likely to be small. On the other hand, duplicate samples from the same individual can have an enormous impact on the analysis of between-locus independence.

The 25 samples that were deemed to be repeats of the same individuals by the FBI matched according to the FBI's criteria at all typed loci. There are three possible reasons why these matches occurred. Either samples were taken from (i) the same individuals (or equivalently monozygotic twins), or they were taken from (ii) related individuals (for example, siblings), or they were taken from (iii) unrelated individuals and match at all these loci merely by chance. For our analysis, it is essential that the sample consist entirely of unrelated individuals, so observations from (i) and (ii) must be deleted. Therefore, it is critical to determine whether the 25 matching samples are from (i), (ii), or (iii). In 22 cases, the FBI determined that the repeated samples actually did derive from the same individual (*2*). For the remaining three pairs, it was not possible to determine their origin. For these pairs, additional loci were tested, both by the FBI and by Cellmark Diagnostics (Germantown, Maryland). If these pairs belong to group (i), we would predict matching at all additional loci; however, if they belong to group (iii), we would predict discordance at some of these loci. These pairs matched at seven loci typed by the FBI and four additional loci typed by Cellmark, making it virtually certain that they came from group (i). Hence, it is not 25 *unrelated* matching individuals that were deleted from the FBI database, but 22 certain and 3 virtually certain duplicate samples. As we indicated above, there is little point in asking for the probability that two samples from a database match at a given number of loci when the database contains duplicates of the same individual, who obviously have to match at all loci.

We also evaluated the effect of not deleting these duplicate samples. We asked the FBI for the deleted sizings, added them to the database, and reanalyzed the data. The inclusion of the additional samples has essentially no impact on the single-locus match probabilities we reported [(*1*), table 1] because the databases are relatively large. Hence, assuming independence, the probability that two *unrelated* individuals match at three or more loci is still vanishingly small. However, the impact on the independence statistics was as one would predict. We detected large two-, three-, four-, and five-locus disequilibrium. This result demonstrates clearly the necessity of deleting duplicate samples.

Matching samples from the Karitiana tribe (*3*), mentioned by Sullivan, are irrelevant to forensic inference for general populations in the United States. The Karitiana tribe is an extremely inbred group founded by a few individuals (*4*). Human leukocyte antigen data show that the Karitiana tribe is an outlier even among the isolated inbred Amazonian tribes (*5*). We understand (*6*) that the Karitiana data now being circulated by defense lawyers are raw data files on seven loci requested by the defense attorneys in *United States* v. *Yee* (*7*). These data were provided before they could be fully evaluated by researchers and before publication. Two pairs of samples appear to match at seven variable number of tandem repeat (VNTR) loci. (One matches numerically; see below). One pair of samples matches at 30 loci and obviously came from the same individual; hence this match is an artifact that became obvious when the full data set was examined. The other pair of samples is from full siblings of an uncle and half-niece mating. The probability of family members matching at VNTR loci is always greater than the probability of a random match, a fact that has not been disputed; for Karitiana family members, the probability of matching is far greater than for members of the average American family.

The second point raised, the issue of match criterion, is somewhat more complicated. We estimated the measurement error for the FBI methodology with fresh DNA to be approximately $0.00625L$ (*8*), where L is the fragment length; likewise, for Lifecodes methods, the estimate was $0.00575L$. Samples for both databases were fresh blood, and therefore almost all matching alleles should be within our match criteria, barring laboratory error. Alleles of different size would also fall within the match criteria.

Environmental factors sometimes cause a forensic sample to migrate slightly faster or slower than a fresh sample, a phenomenon called band shifting. Band shifting is visually obvious because measurement errors of proximal bands are correlated, preserving the overall banding pattern. Band shift is a problem in forensic samples, but it should be minimal for the fresh blood samples used by the FBI and Lifecodes to develop their databases. Hence, a 2.4% window, as we used, is more realistic than a 5% window.

The 5% window is a purely numerical criterion used by the FBI to allow for band shifting. The visual criteria it uses allow for correlated measurement errors and are therefore narrower. We can derive a crude rule for visual matching with the 5% window which also takes into account correlated error, at least for single loci (*9*). With this rule, we performed the same analyses on the FBI database as we reported earlier (*1*). Our conclusions are similar. For this database, we had found (*1*) P values less than 0.05 for 3 out of 40 tests of pairwise independence of loci; with the new rule, the number is 1 out of 40. With the old matching rule, one three-locus match was observed out of a total of 7.6 million pairwise comparisons; four three-locus matches were obtained with the new matching rule. There were again no four- or five-locus matches. Finally, assuming independence and using the old matching rule, we estimated the probability of a chance five-locus match of unrelated individuals in the black, Caucasian, Southeast Hispanic, and Southwest Hispanic populations to be 5.59×10^{-14}, 8.4×10^{-13}, 5.87×10^{-13}, and 1.32×10^{-12}, respectively; with the new rule, these estimates are 3.73×10^{-12}, 3.37×10^{-11}, 2.55×10^{-11}, and 5.78×10^{-11}, respectively. Even if we used a simple 5% numerical matching rule without any adjustment for band correlation, which does not mimic

visual match criteria, the probabilities of chance five-locus matches are similar: 4.08×10^{-11}, 4.18×10^{-10}, 3.55×10^{-10}, and 6.57×10^{-10}, respectively. In other words, the probability of a five-locus match in each population is still vanishingly small.

The use of DNA fingerprinting to exclude a suspect has not, to our knowledge, ever been challenged by defense lawyers. Because the probability that two unrelated individuals match at a set of five loci is so small, innocent suspects will virtually always be excluded as the source of the evidentiary material, and about 30% of suspects have this experience (*10*). We appreciate, however, the tremendous burden this fact places on defense lawyers when they attempt to create a line of defense to exonerate a client whose DNA fingerprint matches that of the evidentiary sample.

Neil Risch
B. Devlin
School of Medicine, Yale University, New Haven, CT 06510

REFERENCES AND NOTES

1. N. Risch and B. Devlin, *Science* **255**, 717 (1992).
2. See the accompanying letter by Bruce Bodowle.
3. J. R. Kidd, F. L. Black, M. Weiss, I. Balazs, K. K. Kidd, *Human Biol.* **63**, 775 (1991).
4. The Karitiana tribe is dominated by a father, mating with four females (six generations ago), and a son, mating with six females. The females were of unknown relationship to each other or to the males, except that one of the son's wives was also his stepdaughter. Many of the matings involve biological relationships closer than second cousins. In subsequent generations, there was limited mating outside of the family (a total of seven more individuals). The "population" or family now consists of about 130 individuals in one village.
5. F. L. Black, personal communication.
6. K. K. Kidd, personal communication.
7. *United States* v. *Yee*, 129 Federal Rules Decisions 629 (Northern District of Ohio, 1990).
8. B. Budowle *et al.*, *Am. J. Hum. Genet.* **48**, 841 (1991). See table 2.
9. In (*1*), we assumed two fragments of sizes x and y match if

$$|x - y| / [0.5(x + y)] \leq b = 0.024 \qquad (1)$$

so that the fragments must be within approximately 4 measurement standard deviations of each other. To crudely mimic the visual evaluation of a match between a fresh and forensic sample, with the database of fresh samples, we redefine b = 0.05. Let x_1,x_2 be the ordered fragment size estimates at a locus for one individual and y_1,y_2 be these sizes at the same locus for another individual. If the pairs of fragments match, condition (1) with b = 0.05 must be met for each pair x_1,y_1 *and* x_2,y_2. To account for correlated measurement error, we impose an additional constraint. The pairs x_1,x_2 *and* y_1,y_2 are expected to covary, but x,y pairs run on different gels are not. In addition, the expected covariance of the x and y values are similar (call it ρ_x) under the hypothesis of a match, as are the measurement error variances of the individual matching fragments (call them σ_1^2 and σ_2^2). Defining $\delta_1 = (x_2 - x_1)$ and $\delta_2 = (y_2 - y_1)$, the expected value of $\delta_2 - \delta_1 = 0$ for a match with variance $\tau^2 = 2\sigma_1^2 + 2\sigma_2^2 -$

$4\sigma_1\sigma_2\rho_x$. As a second condition on matching we then used $|\delta_2 - \delta_1| < 4\tau$.

10. D. Bigbee, Federal Bureau of Investigation, personal communication.

Response: Since 1988, the FBI has analyzed blood samples from individuals of different population groups to establish a DNA database that is used to provide estimates of DNA profile frequencies for these groups. It is the intent of the FBI in the development of all its databases to use only samples from unrelated individuals. When duplicate samples or samples from identical twins are unintentionally included, it is proper to remove one of the matching profiles.

The FBI receives samples for its databases from different sources: Baylor University, Texas College of Osteopathic Medicine, Miami Red Cross, California Department of Justice, and elsewhere. However, the names of the individual donors are deleted before submission to the FBI to maintain the anonymity and privacy of the sample source. Once received by the FBI laboratory, the samples are given individual identification numbers. The quality of these records from this point on is not at issue; they are complete and correct. Accidental duplications that have occurred are not a consequence of the record-keeping of the FBI laboratory.

To ensure that no duplicate samples made it through the sample collection process, the FBI searched its database of samples from more than 2000 individuals with a computer matching program and found 25 matching sample sets. The FBI attempted to account for these duplicates by contacting the contributing laboratories. The Texas College of Osteopathic Medicine confirmed that 22 of the 25 matched pairs came from the same individuals. The Miami Red Cross could not confirm that the remaining three matched pairs came from the same individuals.

To address the possibility that the three matched pairs were in fact duplicates, the samples were typed by the FBI and also by Cellmark Diagnostics with additional loci. In this regard, Sullivan states that I "asked Cellmark Diagnostics . . . to examine the matching samples. Its probes also yielded unclear results. The Florida matches were then deleted from the databases, even though there was no explanation for their occurrence." This statement misrepresents evidence presented in *Minnesota* v. *Johnson* (*1*), in which Sullivan was the defense attorney.

The effort of the FBI laboratory to determine whether or not these matching samples from Florida were indeed duplicates was summarized in an affidavit submitted in *Minnesota* v. *Johnson*. It states (*1*),

additional probings with different probes were performed on these samples at the FBI. In addition, these three duplicate pairs were sent to Cellmark Diagnostics so that additional probings using different probes and a different restriction enzyme could be performed; the profiles from Cellmark Diagnostics with respect to the duplicate pairs were consistent using their cocktail approach.

The three pairs in question exhibited variable number of tandem repeats (VNTR) profiles consistent with the hypothesis that each pair was derived from the same individual on the basis of at least seven VNTR loci typed by the FBI and an additional multilocus cocktail typed by Cellmark Diagnostics (*1*, p. 90). The FBI concludes that these three sample pairs are from the same individuals or from identical twins. This process was completed 6 to 8 months before the FBI sent its database to Risch and Devlin.

Sullivan also states that the removal of duplicates "is not an isolated practice. Budowle testified in *United States* v. *Yee* [2] that the FBI ran its match program over its South Carolina black database and found a large number of matches." He faults the FBI's record-keeping again and omits the fact that the Medical University of South Carolina (MUSC) has acknowledged that it accidentally forwarded a large number of duplicate African-American samples to the FBI. Because of the anonymity afforded the sample donors, MUSC could not confirm which of the samples were duplicates. This event motivated the FBI to develop a computer program to search the database for samples matching across all VNTR loci. The program first was tested on other database samples; it detected no matching samples, and none was expected. Thereafter, the South Carolina African-American database was searched, matched profiles across all loci were identified, and the duplicates were removed. By comparing the South Carolina African-American database with and without the duplicate samples, the FBI ensured that the two databases were statistically similar at each locus analyzed (*3*). Even so, a subsequent administrative decision was made to remove all of the South Carolina African-American samples from the African-American database.

Finally, Sullivan makes a puzzling reference to the Karitiana population study (*4*). Any reference to the Karitiana should be accompanied by the caveat that the Karitiana are an isolated, inbred kinship living in the Amazon basin of western Brazil. The members of the kinship are much more closely related than family members found in populations in the United States. There is no relevance of data about matching probabilities derived from the Karitiana to that of unrelated individuals in the United States.

Bruce Budowle
Forensic Science Research and Training Center, Laboratory Division, FBI Academy, Quantico, VA 22135

REFERENCES AND NOTES

1. *Minnesota* v. *Johnson*, S.I.P. No. 89-072014 (Minnesota District Court, Fourth Judicial District, 1991).
2. *United States* v. *Yee*, 129 Federal Rules Decisions 629 (Northern District of Ohio, 1990).
3. We used a chi-square test to determine homogeneity and found *P* values from 0.991 to 1.00 for D1S7, D2S44, D14S13, D16S85, and D17S79. As there was not statistical difference between the database with duplicates and with duplicates removed, there is no anticipated difference in the final estimate of the likelihood of occurrence of a DNA profile with either database.
4. J. R. Kidd, F. L. Black, M. Weiss, I. Balazs, K. K. Kidd, *Human Biol.* **63**, 775 (1991).

Corrections and Clarifications

In the photograph on page 1142 accompanying the article "Chemical prospecting: Hope for vanishing ecosystems?" by Leslie Roberts (Research News, 22 May, p. 1142), Gerald Bills of Merck was on the right and Rodrigo Gamez of INBio was on the left.

In the News & Comment article by Richard Stone "Peer review catches congressional flak" (15 May, p. 959), Senator Robert C. Byrd (D–WV) and Representative William Natcher (D–KY) were incorrectly identified as having ordered a review of National Science Foundation (NSF) grants. The NSF grants were targeted by Byrd's Senate Appropriations Committee, as reported in Joseph Palca's article "Congress sends a message" (News & Comment, 29 May, p. 1274).

In Richard Stone's News article "Hard times in the promised land" (8 May, p. 728), the Russian city Novosibirsk, in Siberia, was incorrectly stated to be in Ukraine.

In the acknowledgements (note 25, p. 221) of the report "Centriole duplication in lysates of *Spisula solidissima* oocytes" by R. E. Palazzo *et al.* (10 Apr., p. 219), the American Cancer Society's grant to R.E.P. (JFRA 162121) was incorrectly stated to be from the American Chemical Society.

In Robert Pool's article "Bringing the computer revolution down to a personal level" ("Computing in Science" special section, 3 Apr., p. 55), the **Axiom** symbolic math program should have been listed as being available from the Numerical Algorithms Group in Downers Grove, Illinois.

In the news briefing "The world's most prolific scientists" (17 Jan., p. 283), Arnold L. Rheingold, a crystallographer at the University of Delaware, was inadvertently left off the list of the top 20 most prolific scientists. The Institute for Scientific Information has corrected its list and states that Rheingold published 391 articles and other scientific communications between 1981 and the end of 1990. He should have ranked 13th on the list.