

- Trends in Criminal DNA Databanking,” *Journal of Law, Medicine & Ethics* 28 (2000): 209-221.
36. T. Simoncelli, “Dangerous Excursions: The Case Against Expanding Forensic DNA Databases to Innocent Persons,” *Journal of Law, Medicine & Ethics* 34 (2006): 390-397.
37. Pub. Law 109-162, 119 Stat. 2960, codified at 42 U.S.C. §14132(a)(1) (2005). The amendment retains the prohibition on inclusion of samples voluntarily submitted solely for the purpose of elimination. *Id.*
38. L. Wacquant, “Deadly Symbiosis: When Ghetto and Prison Meet and Mesh” *Punishment and Society* 3, no. 1 (2000): 95-134. Reprinted in D. Garland, ed., *Mass Imprisonment: Social Causes and Consequences* (London: Sage, 2001): 82-120.
39. S. R. Gross, et al., “Exonerations in the United States, 1989 through 2003,” Unpublished manuscript, University of Michigan Law School (2004) available at <<http://www.law.umich.edu/newsandinfo/exonerations-in-us.pdf>> (last visited March 14, 2006).

What Does One in a Trillion Mean?

BY EDWARD UNGVARSKY

Observations Suggest “Not Much”

As part of the federally managed Combined DNA Index System (“CODIS”), every state in the country, in addition to the federal government, has a database of DNA profiles collected from convicted offenders.¹ The national database has over 3.6 million profiles, and has more than doubled in size since 2004.^{2,3} The federal and state databases will grow at even faster rates as Congress and state legislatures have vastly expanded the scope of authorized collection of DNA samples for inclusion of profiles in the databases.⁴

Cold hits (when an unknown DNA sample found at a crime scene is matched to a profile in one of these databases) serve as the foundation – and often the entirety – of the evidence against a suspect. Beyond this law enforcement purpose, these large DNA databases can also serve another valuable function: the basis for scientific inquiry into the accuracy of the theoretical assumptions that underlie forensic DNA profiling.

Away from the eyes of the media and the general public, a debate rages in courtrooms over these databases. On one side, defense attorneys and university scientists contend that the databases should be available for scientific research to allow testing of the assumptions underlying the statistics used in DNA profiling. On the other, prosecutors and scientists associated with crime laboratories defend the status quo and deny the value of empirically testing the appropriateness of the prevailing theoretical model of forensic DNA profiling. How this dispute is settled will ultimately affect the truth-seeking function of the criminal justice system.

CURRENT REPORTING OF FORENSIC DNA PROFILING RESULTS

When DNA evidence first became admissible in trials, one judge predicted that forensic DNA technology would be the “single greatest advance in the ‘search for truth’ ... since the advent of cross-examination.”⁵ DNA evidence is seen as highly persuasive and nearly infallible.⁶

However, a “match”* between an individual’s forensic DNA profile and a crime scene DNA sample does not necessarily mean that that individual left the evidence sample. DNA samples can be consistent with one another if they do in fact come from the same individual – but also if the match is mere coincidence.⁷ In fact, a match between the individual’s DNA profile and the evidence sample does not necessarily mean that that individual is the source of the evidence sample, “a statistical estimate of the significance of a [DNA] match is needed.”⁸ This statistical estimate is called the random match probability (“RMP”) – the chance that DNA from an unrelated person, selected at random from a particular population, would share the genetic profile with the evidence sample.

The RMP is determined following a population genetics model that rests upon several assumptions. First, the model

* “Consistency” is a better term than “match” to use for when similarities are observed from different sources of forensic DNA profiles; however, this article uses the more common, if sometimes misunderstood and misused, term.

assumes that the estimated allele frequencies are accurate. Second, the model assumes that the 13 loci used to develop the DNA profile are independent of each other. Forensic scientists multiply the allelic frequency estimates of each forensic locus against the others through the use of the product rule. Third, the model assumes that everyone is unrelated to everyone else such that there is no appreciable substructure in the population.⁹

Because forensic scientists recognize that the assumption that we are all unrelated is problematic, they have theorized a correction, known as theta, to the formula that generates the reported frequency. The modified product rule that is ultimately used to calculate the RMP is a theoretical construct, which, to date, has not been subject to rigorous empirical testing.

Application of this theoretical model has served as the basis for the statistical estimates of DNA matches in courtrooms across the country; indeed, miniscule RMP estimates ranging from 1 in a trillion, to 1 in 26 quintillion are typical.¹⁰ Crime laboratory analysts then testify in court that such figures are reliable and accurate. With the exception of identical twins, no forensic scientist has yet observed a match at all 13 loci (which would be coincidental). FBI DNA analysts testify that, where match probabilities are smaller than 1 in 280 billion, they can conclude with a reasonable degree of scientific certainty that the individual whose DNA

profile matches the crime scene evidence is the source of the crime scene DNA.¹¹

For the mathematically trained, such figures can occasion laughter. As put by Stanford University mathematician Dr. Keith Devlin, “such a figure is total nonsense. Nothing in life ever comes remotely close to such a degree of accuracy. In most professions where numerical precision is important, including laboratory science, 1 in 10,000 is often difficult to achieve... As often happens when the computation of probabilities is concerned, such unsophisticated use of the product rule rapidly takes theory well beyond the bounds of reality.”¹²

Be that as it may, the above-described theoretical model governs the

admission of DNA match evidence in courtrooms across the United States. Courts have upheld its admission, crime laboratory analysts have testified to it, and jurors have relied upon the model to assess evidence in order to render verdicts for over twenty years.

Is application of the theory sound? And, if not, how and to what degree has it gone awry? Can the theoretical model be tested empirically – and, if so, how and by whom?



Keep *GeneWatch* Coming!

You can add an individual gift subscription for only \$20 when you order a year of *GeneWatch* for yourself

Call 617-868-0870 or visit www.gene-watch.org to subscribe.

EMPIRICAL OBSERVATIONS AND TESTING OF FORENSIC DNA'S THEORETICAL MODEL

When theoretical applications raise real world questions, the best place to turn for answers is often the real world. And, as it turns out, at least one crime laboratory has already conducted an empirical study that gives reason to believe that the model is flawed.

In 2001, the Arizona Department of Public Safety Crime Laboratory ("Arizona DPS") searched its convicted offender database and observed a 9 STR* locus match.¹³

The forensic science community treated this report of a 9-locus coincidental match between unrelated persons as unusual and noteworthy. Prior to that day, there had never been a report of a coincidental match of more than 6 loci, and even those infrequent occurrences themselves merited

great attention.¹⁴ Coincidental matches at 9 or more loci often result in RMP values that are smaller than 1 in 280 billion, the FBI's threshold figure for source attribution.¹⁵

As recently as the spring of 2005, well-known geneticists who frequently testify for the prosecution treated the 2001 Arizona DPS report as an outlier, and testified under oath that matches at 9, 10, or more loci were rare in the extreme. Scientists associated with crime laboratories testified that to observe a 9-or-10-locus match between two individuals would be exceedingly unusual, and that the only known 10-locus match was an example that involved an incestuous relationship.¹⁶

But science advances in response to increased study and can render such inadequately tested inductive conclusions erroneous. In the fall of 2005, Arizona DPS further examined its convicted offender database. Arizona DPS compared the DNA profiles of each of the 65,493 persons in its database against each other. From this comparison, Arizona DPS reported some remarkable findings: its database had 122 pairs of people who matched at 9 out of the 13 loci, 20 that matched at 10 loci, 1 that matched at 11 loci, and 1 that matched at 12 loci. The last two matches were confirmed to be between pairs of siblings.¹⁷ Recent studies emphasize the likelihood that siblings will have nearly perfect matches across the 13 STR loci, an issue of particular importance when suspects are charged based on a cold hit DNA profile.¹⁸

*STR markers, or short tandem repeat profiles, have become the standard genetic markers for human identification in genetic cases between two apparently unrelated individuals (one Caucasian, the other African American).

The fiction that matches at high numbers of loci are unexpected was demolished; prosecution-affiliated scientists now opine that such matches are "expected," while still hewing to the line that no report of a coincidental 13-locus match has been observed.¹⁹

Dr. Laurence Mueller of the University of California at Irvine has conducted preliminary modeling studies of those matches and observed that the number of observed 9 and 10 locus matches did not fit the expected rate based on RMP assumptions and required further exploration.²⁰

These studies and their results raise questions for scientists and criminal justice professionals alike. Are there problems with the theoretical model that lead to the exceedingly small RMP estimates that are presented in court for the 9-to-13-locus matches between suspects and evidence samples? When will we first observe coincidental matches at 13 loci? Will it be when a larger DNA profile database is searched?

We do not know the answers yet, but we do know that these questions cannot be answered by continued blind adherence to the theoretical model. These questions, and others, can only be answered by further empirical research.⁺ Some judges are ordering searches of their state convicted offender databases; the results from those searches remain to be seen.²¹

While state systems consider whether to study their databases and whether to publish the results of their studies, the FBI has steadfastly refused to examine its national CODIS database of 3.6 million profiles. Dr. Devlin observes that because the FBI's database is "easily large enough to yield statistically reliable results," an empirical study of the database is warranted.²²

Calling such a study "a matter of some urgency," Dr. Devlin expects that the national CODIS database "will contain not just one but several pairs that match on all 13 loci, contrary (and how!) to the prediction made by proponents of the currently much-touted RMP that you can expect a single match only when you have on the order of 15 quadrillion profiles."²³



+Empirical inquiry into the databases can also help determine an error rate for DNA profiling. For example, just as, across the 13 loci, there should be fewer full 13 loci matches than 12 loci matches, there should be fewer 25 out of 26 allele matches than 24 out of 26 allele matches. If the converse turned out to be true, that might suggest the presence of single-allele typing errors in entry of the profiles in the databases and could help establish an estimate for the error rate for DNA profiling.

Given the FBI's current protestations and those raised by its state-level CODIS partners, it seems highly unlikely that the FBI will conduct the analysis or, absent a court-order, even allow an outside scientist to examine the CODIS databases.²⁴ Perhaps at minimum the FBI will change its policy and preclude its analysts from opining that a particular individual is the source of crime scene DNA based on statistical estimates whose accuracy and reliability are unsupported by any empirical testing.

CONCLUSION

After over twenty years of nearly unqualified acceptance in the courtroom, forensic DNA has entered a new age. Coincidental matches at larger numbers of loci, once unheard of, are now "expected." A coincidental 13-locus match will be reported; it is only a matter of time. The twin myths that microscopic RMP estimates are statistically meaningful and that statements of source attribution are scientifically defensible are being shattered in the face of scientific inquiry.

Forensic DNA may still achieve its predicted status as the single greatest advance for the search for truth since cross-examination. The scientific method – observe, theorize, and test – has a long and successful pedigree. The time has come for the FBI and other crime laboratories to agree to empirical investigation of the accuracy of the theoretical model that sustains the courtroom use of forensic DNA. The crime laboratories should conduct this research themselves to validate their statistical estimates. And, if the crime laboratories remain uninterested in this work, it is hard to imagine what harm might occur in the event that university scientists are given appropriate access to the data for independent empirical analysis. Scientists should take advantage of statutory authorizations for release of this information for research purposes and approach the CODIS depositories directly.²⁵ With the expansion of the DNA profile databases, the time is ripe to explore the theoretical model upon which forensic DNA assumptions are based with real-world studies. ■■■

Edward Ungvarsky is Special Counsel to the Director at the Public Defender Service for the District of Columbia. He directs the Public Defender Service's Forensic Practice Group and has expertise in the litigation of DNA, eyewitness ID, false confession and other forensic evidence.

REFERENCES

1. <http://www.fbi.gov/hq/lab/codis/clickmap.htm>.
2. Ibid.
3. See John Butler, Forensic DNA Typing and Prospects for Biometrics (May 12, 2004), available at http://www.cstl.nist.gov/biotech/strbase/pub_pres/BiometricIDMay2004.pdf (listing total number of convicted offender profiles present in national DNA database at 1,641,076 as of March 2004).
4. See Lisa Hurst, NIJ DNA Grantees Meeting (June 28, 2006), available at <http://www.dnaresource.info/presentations.html> (discussing expansion of databases at state and federal level to include juveniles, misdemeanants, arrestees, and federal detainees); DNA Fingerprint Act of 2005, Pub. L. No. 109-162, 119 Stat 2960 (2006) (dramatically expanding scope of federal database).
5. *People v. Wesley*, 533 N.Y.S.2d 643, 644 (Sup. Ct. 1988).
6. See Survey of D.C. Jurors conducted by the Public Defender Service in December 2003, questions 6, 20, available at <http://www.pdsdc.org/SpecialLitigation/SLDSysRes/Brady%20Poll%20Results,%20December%202003.pdf>. (conveying that on scale of 1 to 10, DNA evidence rated 9 for general persuasiveness and 9 for perceived general reliability, the highest scores of any type of evidence).
7. National Research Council, *The Evaluation Of Forensic DNA Evidence* 127 (1996) ("Suppose that a DNA sample from a crime scene and one from a suspect are compared, and the two profiles match at every locus tested. Either the suspect left the DNA or someone else did. We want to evaluate the probability of finding this profile in the 'someone else' case.").
8. John Butler, *Forensic DNA Typing* 270 (2d ed. 2005); see *People v. Barney*, 10 Cal. Rptr. 2d 731, 742 (Cal. Ct. App. 1992); *Porter v. United States*, 618 A.2d 629 (D.C. 1992); National Research Council DNA Technology in Forensic Science 9 (1992).
9. See generally National Research Council, *The Evaluation Of Forensic DNA Evidence*, supra note 7, at 122 (setting forth Recommendation 4.1 for calculation of profile frequency).
10. See *United States v. Jenkins*, 887 A.2d 1013 (D.C. 2005).
11. Bruce Budowle et al., Source Attribution of a Forensic DNA Profile, 2 *Forensic Science Communications* 3 (July 2000), available at <http://www.fbi.gov/hq/lab/fsc/backissu/july2000/source.htm>.
12. Keith Devlin, Damned Lies, *Devlin's Angle*, MAA online (Oct. 2006), available at http://www.maa.org/devlin/devlin_10_06.html.
13. Kathryn Troyer et al., "A Nine STR Locus Match Between Two Apparently Unrelated Individuals Using AmpflSTR Profile Plus and Cofiler," Promega 12th International Symposium (2001), available at <http://www.promega.com/geneticidproc/ussymp12proc/abstracts/troyer.pdf>.
14. See Department of Justice, National Institute of Justice, *The Future of Forensic DNA Testing* 25 n.13 (Nov. 2000) (reporting 10 6-locus DNA profile matches in New Zealand database of 10,907 records in which 8 of matches were brothers and 2 of matches were unrelated persons); Richard Willing, "Mismatch Calls DNA Tests into Question", *USA Today* (Feb. 8, 2000), at 3A (coincidental 6-locus match in United Kingdom database).
15. See, e.g. Brendan Shea, FBI Laboratory Dictation Report, *United States v. Berger*, No. 2004 FEL 003420 (July 12, 2004) (declaring source attribution based on 9-locus match).

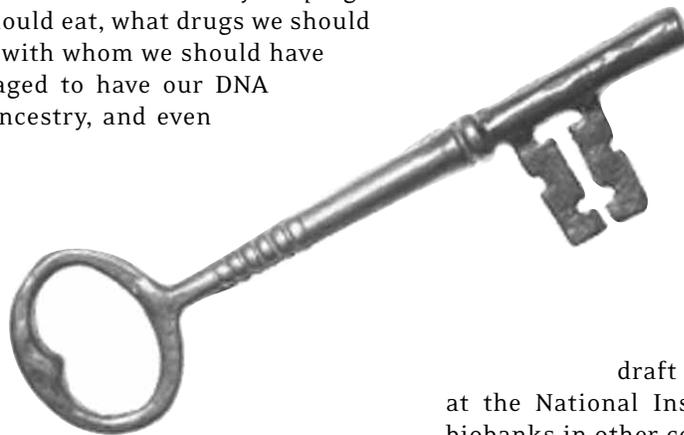
16. Transcript, *United States v. Jenkins*, F-320-00 (March 28, 2005), at 52-53 (testimony of Dr. Fred Bieber of Harvard Medical School); Transcript, *United States v. Jenkins*, F-320-00 (March 20, 2005), at 36-37 (testimony of Dr. Ranajit Chakraborty of the University of Cincinnati).
17. Arizona Department of Public Safety Crime Laboratory, 9+ Locus Match Summary Report (Oct. 2005).
18. See David R. Paoletti, et al., Assessing the Implications for Close Relatives in the Event of Similar but NonMatching DNA Profiles, 46 *Jurimetrics J.* 161 (2006); Bruce Weir, Matching and Partially-Matching DNA Profiles, 49 *J. Forensic Science Communications.* 1009 (2004).
19. See Bruce Budowle et al., Clarification of Statistical Issues Related to the Operation of CODIS, Pub. No. 0701 of the Laboratory Division of the Federal Bureau of Investigation (unpublished manuscript) [hereinafter "Clarification of Statistical Issues"], at 7; Affidavit of Frederick R. Bieber, Ph.D., *United States v. Berger*, No. 2004 FEL 003420 (Oct. 10, 2006), at 2, ¶ 6.
20. Declaration of Laurence D. Mueller, *United States v. Berger*, No. 2004 FEL 003420 (Oct. 6, 2006), at 2, ¶ 7 ("Based on models that I have run to date, the numbers of matches in the Arizona database cannot be accounted for simply by close relatives."); Declaration of Laurence D. Mueller, *United States v. Berger*, No. 2004 FEL 003420 (Oct. 22, 2006), at 1, 2, ¶ 4, 8 ("[T]he results I have seen from the much smaller Arizona database undermine the accuracy of current method of calculating frequency estimates.... While it is easy to find a level of relatives that accurately predicts the number of 9-locus matches in the Arizona database and it is easy to find a level that accurately predicts the number of 10-locus matches, it has been nearly impossible to find any combination of relatives or population substructure that will explain why both the number of 9-and-10 loci matches are anomalous.").
21. See, e.g., Order of the Honorable Vincent Gaughan, Circuit Court of Cook County, Illinois (July 11, 2006); Order of the Honorable Steven I. Platt, Circuit Court for Prince George's County, Maryland (Aug. 4, 2006).
22. Devlin, *supra* note 12.
23. *Ibid.*
24. See Clarification of Statistical Issues, *supra* note 19, at 2-7.
25. See, e.g., 42 U.S.C. § 14132(b)(3)(D) (2006) (disclosure permissible "if personally identifiable information is removed, for a population statistics database, for identification research and protocol development purposes, or for quality control purposes").

New Genetic Privacy Concerns

BY PATRICIA A. ROCHE AND GEORGE J. ANNAS

DNA Donors May Give Up More than They Realize

Everyone wants a piece of us, specifically a DNA sample. Private marketers assure us that our DNA will unlock useful secrets that can make our lives better by helping to determine what food we should eat, what drugs we should take, and even whether or with whom we should have children. We are encouraged to have our DNA decoded to discover our ancestry, and even mass marketed magazines, like *National Geographic*, encourage readers to submit their DNA samples for analysis to determine "where we come from." Federal and state officials discuss a national, or even international, DNA databank that can be used to hunt for criminals, and sometimes also to help exonerate criminal suspects. The promises of DNA testing seem endless to both individuals and society. But DNA testing has a dark side as well. As the source of genetic information, the DNA



molecule is also a separate entity that can be collected and stored for multiple, currently unknown (or at least unconsented to) uses. Since multiple copies can readily be made, once an individual's sample is obtained, no further contact with the person is required to procure additional material for testing. Should this be of concern to us, or do existing federal and state laws adequately protect our genetic privacy?

In May 2006, the Secretary's Advisory Committee on Genetics, Health, and Society (SACGHS) issued a draft report on a proposed national biobank at the National Institutes of Health (NIH).¹ Similar to biobanks in other countries, such as the U.K., Estonia, and Canada, this repository would contain tissue samples and health information from 500,000 to 1,000,000 individuals to be utilized for studying the gene-environment interactions underlying common, complex diseases. During the preparation of its final report, SACGHS asked for public