

COMMENTARY

## DNA fingerprinting, dispute laid to rest.

*Eric S. Lander and Bruce Budowle,*

Two principals in the once-raging debate over forensic DNA typing conclude that the scientific issues have all been resolved.

The US public, usually indifferent to matters scientific, has suddenly become obsessed with DNA. Nightly newscasts refer to the polymerase chain reaction (PCR) and even the tabloids offer commentary on restriction fragment length polymorphisms (RFLPs). The new-found fascination with nucleic acids does not stem from recent breakthroughs in genetic screening for breast cancer susceptibility or progress in gene therapy -- developments which will indeed affect the lives of millions. Rather, it focuses on the murder case against the former US football star, O. J. Simpson.

The Los Angeles trial, starting in November and to be broadcast live by several major television networks, will probably feature the most detailed course in molecular genetics ever taught to the US people. This bold experiment in public education should, in principle, be a cause for rejoicing among scientists. The catch is that the syllabus is being prepared by attorneys whose primary roles are as adversaries; the likely result is confusion. Already, the news weeklies are preparing the ground with warnings that DNA fingerprinting remains "controversial", being plagued by major unresolved scientific issues.

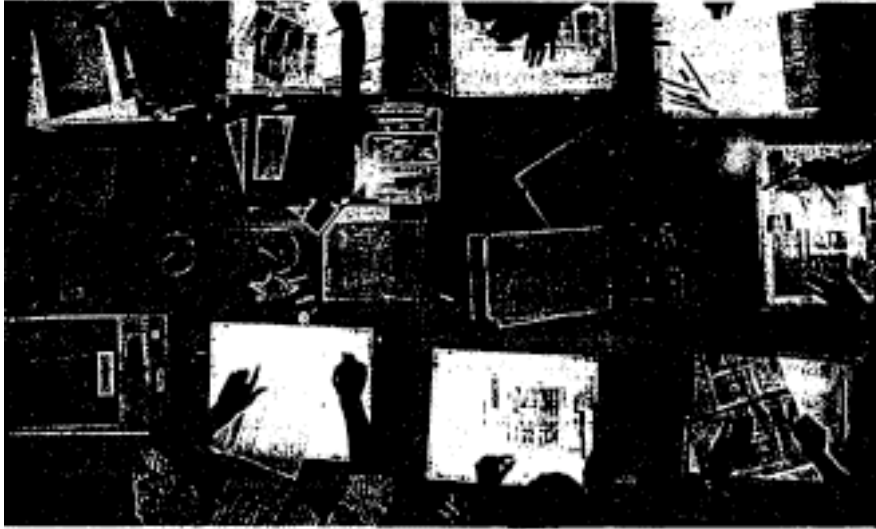
Forensic DNA typing certainly did provoke controversy soon after it was introduced into US courts in 1988. The technology itself represents perhaps the greatest advance in forensic science since the development of ordinary fingerprints in 1892, and is soundly rooted in molecular biology. The problem, however, stemmed from the manner of its introduction. Pioneered by biotech start-up companies with good intentions but no track record in forensic science, DNA typing was marred by several early cases involving poorly defined procedures and interpretation<sup>1</sup>. Standards were lacking for such crucial issues as: declaring a match between patterns; interpreting artefacts on gels; choosing probes; assembling databases; and computing genotype frequencies. There is broad agreement today that many of these early practices were unacceptable, and some indefensible. For its part, the US Federal Bureau of Investigation (FBI) moved much more deliberately in developing procedures, sought public comment and opted for conservative procedures.

As a result of these growing pains, forensic DNA typing was subjected to intense debate and scrutiny. When it first burst on the scene, the supporting scientific literature consisted of a mere handful of papers. By the middle of this year, there had been more than 400 scientific papers, 100 scientific conferences, 3 sets of guidelines from the Technical Working Group on DNA Analysis Methods (TWGDAM), 150 court decisions and, importantly, a 3-year study by a National Research Council (NRC) committee released in 1992 (ref. 2). In the light of this extraordinary scrutiny, it seems appropriate to ask whether there remains any important unresolved issue about DNA typing, or whether it is time to declare the great DNA fingerprinting controversy over.

As co-authors, we can address these questions in an even-handed manner. B.B. was one of the principal architects of the FBI's DNA typing programme, whereas E. S. L. was an early and vigorous critic of the lack of scientific standards, and served on the NRC committee. In a world of soundbites, we are often pegged as, respectively, a "proponent" and an "opponent" of DNA typing. Such labels greatly oversimplify matters, but it is fair to say that we represent the range of scientific debate.

We recently discussed the current state of DNA typing, and could identify no remaining problem that should prevent the full use of DNA evidence in any court. What controversy existed seems to have been fully resolved by the NRC report, the TWGDAM guidelines and the extensive scientific literature. The DNA fingerprinting wars are over.

Our goal is to correct the lingering impression to the contrary. Our analysis below represents our unanimous opinions (apart from specific comments about the workings and intent of the NRC committee itself, which necessarily are based on E.S.L.'s recollection). We focus on the subject most often said to remain problematical: population genetics. Our main thesis is that the academic-debate that continues to swirl about population genetic issues is rooted in a misunderstanding of the NRC report and is, in any case, of no practical consequence to the courts. We also touch on how the legal and scientific community should cope with the continuing evolution of DNA typing technology. In particular, we question whether a steady succession of ad hoc committees, however distinguished, is a wise solution.



Comparing autoradiographs from DNA samples at Cellmark Diagnostics.

735

## Laboratory practices

The initial outcry over DNA typing standards concerned laboratory problems: poorly defined rules for declaring a match; experiments without controls; contaminated probes and samples; and sloppy interpretation of autoradiographs<sup>1</sup>.

Although there is no evidence that these technical failings resulted in any wrongful convictions, the lack of standards seemed to be a recipe for trouble. To address these problems, the NRC committee enunciated conservative standards for each laboratory step, based on, more than a decade of experience with human DNA analysis. TWGDAM also developed guidelines along similar lines. Today, there is no doubt about the correct laboratory protocols to ensure reliable DNA typing results. Since the NRC report, US courts have unanimously accepted the technical reliability of DNA evidence, both in principle and in practice.

The NRC committee also highlighted the importance of laboratory accreditation, rigorous quality assurance and quality control (QA/QC) programs, and external blind proficiency tests (tests administered by persons outside the testing lab itself). The importance of these practices has been universally acknowledged, and most forensic labs follow TWGDAM's voluntary quality-control guidelines.

## Population genetics

The controversy over population genetics began as a secondary issue. If DNA analysis reveals that two samples match at the loci tested, the final step is to estimate the frequency of the shared genotype in the general population, which indicates the probability that a randomly chosen person would carry this genotype. Such estimates depend on surveys of the appropriate population.

In some early cases, the rarity of genotype frequencies was greatly overstated owing to a technical error: the calculations were based on overoptimistic assumptions about the precision with which genotypes could be measured. One commercial lab, for example, reported the astronomical frequency of 1 in 738,000,000,000, based on a fourlocus match<sup>1</sup>. The NRC committee easily rectified these problems by requiring consistency between the measurement precision used for forensic analysis and for population genetic estimates (a practice that the FBI, in fact, had long followed).

A subtler but more challenging issue emerged in later cases, concerning the structure of human populations. The 'product rule', used by forensic labs to calculate genotype frequencies, assumed that the individual alleles comprising a genotype could be treated as statistically independent, and their frequencies multiplied<sup>2</sup>. However, some population geneticists asserted that the assumption of independence was appropriate for well-mixed populations (technically, those at Hardy-Weinberg equilibrium and linkage equilibrium), but was not necessarily valid for populations with substructure. According to this argument, the frequency of a common Japanese genotype might be underestimated because the product rule ignored the fact that common Japanese alleles tend to occur together in the US Asian population. Moreover, the frequency of genotypes arising from mixed ethnic ancestry, might be understated because the product rule was typically applied to separate racial databases (for the Caucasian, Black and Hispanic populations) and thus did not account for the presence of genotypes involving common alleles from different racial groups. The substructure argument became a *cause celebre*, pitting such luminaries as Lewontin and Hartl<sup>3</sup> against Chakraborty and Kidd<sup>4</sup>. Both sides conceded that substructure could matter in principle, but many doubted that its effect could be significant in practice (see ref. 5).

The NRC committee at first attempted to settle the issue on its merits. The members agreed that the product rule was probably near the mark, but were hard pressed to say just how close. The committee considered applying formulas from

theoretical population genetics based on the empirical measures of the degree of variation and admixture among and within populations. However, it concluded that there were, at the time, too few hard data about the loci used in forensic typing (most classical genetic surveys concerned protein polymorphisms, likely to be strongly influenced by natural selection) and about the precise structure of the US population. It would be too risky to base a recommendation on assumptions that might subsequently turn out to be faulty.

Thomas Caskey (Baylor College) eventually pointed the way out the quagmire when he asked, out of frustration, whether it was possible to ignore population substructure altogether. Taking up the notion, the NRC committee set out to fashion an extremely conservative rule having the virtue that it made virtually no assumptions.

## **The ceiling principle**

The solution turned out to be quite simple. Suppose that the US population is descended from a collection of populations  $P_1, P_2, \dots, P_n$ , each sufficiently old and well mixed to allow the product rule to be safely applied. Regardless of the population substructure, the multiplication rule requires only a slight modification to yield a strict upper bound on the frequency of any genotype  $G$ : for each allele in  $G$ , the allele frequency should be taken to be the maximum over the component subpopulations. In effect, the approach makes the worst-case assumption that the population may contain individuals who, for example, carry a common Caucasian allele at a locus on chromosome 2 and a common Black allele at a locus on chromosome 17. By assuming the worst, one is guaranteed to be conservative. Because it used the maximum frequency in any subpopulation, the method was dubbed the 'ceiling principle'<sup>2</sup>.

In practice, it is unnecessary to, survey every possible subpopulation. The committee concluded that the likely variation in allele frequencies could be reckoned by conducting modest surveys of 100 individuals from each of 10-15 representative subpopulations spanning the range of ethnic groups represented in the United States - such as English, Germans, Italians, Russians, Navahos, Puerto Ricans, Chinese, Japanese, Vietnamese and west Africans. Each allele frequency could then be taken to be the maximum over these subpopulations, although never less than 5%. (The latter provision was designed to deal with unexamined populations. If an allele was rare in the 10-15 subpopulations surveyed, genetic drift was not likely to have caused its frequency to drift much above 5% in other significant subpopulations.) Even in advance of detailed data about ethnic groups, the committee felt that same principle could be applied to the available racial databases (Caucasian, Black, Hispanic, Asian), although it recommended a 10% floor on allele frequencies to reflect the greater uncertainty about subpopulation variation: this slightly amended form was called the 'interim ceiling principle'. (The choices of 5% and 10% were based on the quantitative effect of genetic drift on the match odds - that is, on the reciprocal of the allele frequency - although none of this reasoning survived into the text of the final report.) The practical effect of these rules was to limit the contribution of any single locus to a factor of 50:1 odds based only on aggregate data for racial classifications and 200:1 odds based on more detailed ethnic surveys.

The ceiling principle was unabashedly conservative. It gave the benefit of every conceivable doubt to the defendant, so that it could withstand attacks from the most stubborn and creative attorneys. Some of the statistical power was sacrificed to neutralize all possible worries about population substructure.

The committee was comfortable with such a lop-sided approach, because even these extreme assumptions did not undermine the practical use of DNA fingerprinting. A four-locus match performed by forensic labs could still provide odds of 6,250,000:1. If this were not enough, two additional loci could increase the odds to more than 15,000,000,000:1.

Finally, the ceiling principle was not intended to be exclusive. Expert witnesses were still free to provide their statistical "best estimate" of genotype frequencies based on the product rule. But if disagreement over such estimates arose, the ceiling principle provided an approach that all parties had to admit was biased to favour defendant. By all rights this seemingly solomonic solution should have ended the controversy over population genetics.

## **Hitting the ceiling**

Surprisingly, attacks came from an unexpected quarter. Some vocal theoretical population geneticists and statisticians concluded, that the committee had been too conservative. They argued that the effect of population substructure was slight and that it would best be treated by using formulas from theoretical population genetics. The ceiling principle was accused of being clumsy and scientifically flawed. Suddenly, a new controversy over population genetics seemed to emerge<sup>5-10</sup>.

The debate was based on a simple misunderstanding of the NRC Committee report but, with the committee disbanded, there was no easy way to address it. Moreover, the committee members had agreed to let the report speak for itself to avoid the emergence of conflicting gospels according to different members. In retrospect, this was probably an unwise decision because it has allowed a minor academic debate to snowball to the point that it threatens to undermine the use of DNA fingerprinting by suggesting that there is some problem with the use of population genetic estimates in court.

Six objections have been raised to the ceiling principle, which are worth briefly refuting:

(1) **The ceiling principle is premised on the flawed analysis of Lewontin and Hartl that there is significant population substructure<sup>8</sup>.** On the contrary, the committee was quite dubious that substructure had significant effects, but felt that the possibility needed to be taken seriously rather than dismissed based on theoretical or indirect arguments. The NRC report cites the Lewontin-Hartl article<sup>3</sup> only twice, both times balanced against a longer list of opposing articles.

(2) **The ceiling principle is scientifically flawed because it is not used in population genetics<sup>9,10</sup>. Moreover, the plan to sample 10-15 representative populations is statistically unsound<sup>8</sup>.** The choice of a statistical method necessarily depends on the dangers of overestimation versus underestimation. In forensics, there is strong agreement on the need to be conservative contrast, population geneticists do not need to be conservative in academic studies; they are content to err equally often on the high and low side, and thus ceiling approaches are unnecessary. However, ceiling approaches are common in a closely related genetic pursuit: the mapping of disease genes. To guard against falsely implicating or excluding a chromosomal contagion human geneticists often analyse their data under worst-case conditions, such as using an unrealistically low ceiling on the penetrance of a disease gene or an unrealistically high frequency for a marker allele<sup>11,12</sup>. Also, some authors have complained that surveys of 100 individuals do not allow accurate estimates of the frequency of rare alleles<sup>8</sup>. The purpose of suggested population studies, however was not to estimate low frequencies but rather to check that some alleles do not unexpectedly have extremely high frequencies (much more than 5-10%) in certain populations. For this purpose 100 individuals is quite adequate.

(3) **The ceiling principle makes ludicrous assumptions about the possible substructure of a population<sup>9,10</sup>.** Although the NRC report called for empirical studies of those groups that made significant contributions to the United States (such as English, Italians and Puerto Ricans), some commentators<sup>9,10</sup> were carried away by hyperbole - asserting that the ceiling principle assumes that "the culprit might be ... a Lapp for one allele and a Hottentot for the other". However, if Lapp and Hottentot are replaced by Italians and Puerto Ricans, the assumption is perfectly reasonable. Indeed, it is unreasonable to assume that such genotypes don't occur in the population.

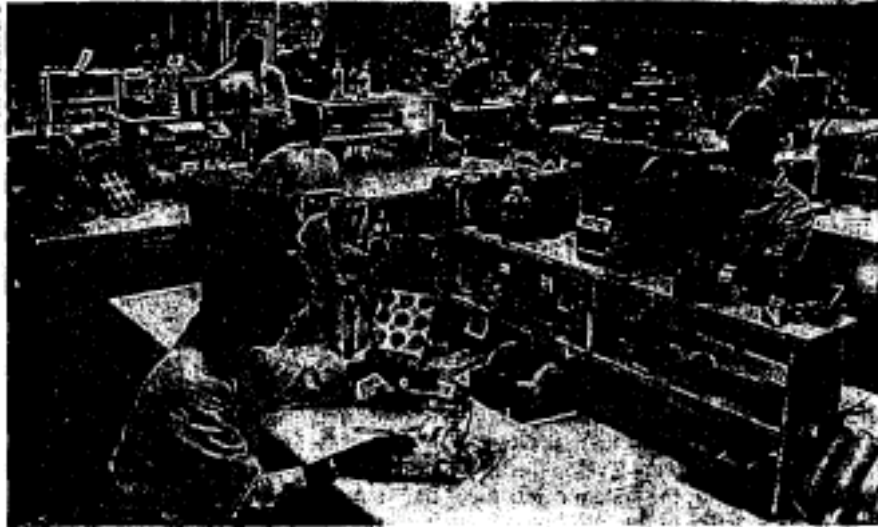
(4) **The ceiling principle is so conservative that it hampers the courtroom application of DNA fingerprinting.** Despite fears that an ultraconservative standard would clip the wings of the DNA fingerprinting, published analyses by several groups<sup>13,14</sup> agree that the effect is modest: Whereas the product rule typically gives four-locus genotype frequencies of about  $10^{-8}$ - $10^{-9}$ , the ceiling principle pares them back to about  $10^{-6}$ - $10^{-7}$ . That extreme assumptions have so little effect only underscores the power of DNA typing and the wisdom of taking a conservative approach.

(5) **The ceiling principle is not actually guaranteed to be conservative.** This concern appears in only a single paper<sup>15</sup>, in which the authors point out that the conservativeness of the ceiling principle depends on the component populations  $P_1 \dots P_0$  being themselves well-mixed. The authors take great pains to construct counterexamples in the event that the populations are themselves substructured. In fact, the NRC committee considered this point to be self-evident (although failed to state it clearly enough), as can be seen from the trivial observation that the ceiling principle has no effect on genotype frequencies when one combines two identical, but substructured populations. In applying the ceiling principle in practice, the committee was confident that any residual substructure in the component subpopulations could safely be ignored in the context of such a conservative scheme.

(6) **The NRC report is causing DNA fingerprinting cases to be thrown out of court<sup>8</sup>.** To the contrary, most courts have used the NRC report as strong evidence that, notwithstanding disagreement over the best solution, there is at least one approach that is indisputably conservative. Of the few cases in which DNA evidence has ever been rejected on population-genetic grounds, virtually all involved evidence predating the NRC report. These courts have cited the report solely for its acknowledgement that a controversy existed and was a reason for constituting the committee.

The NRC report, to be sure, has important flaws. The ceiling principle was not an elegant solution, but simply a practical way to sidestep a contentious and unproductive debate. The report had more than its share of miswordings, ambiguities and errors, many of which have been corrected by a vigilant commentator<sup>16</sup>. A few poorly worded sentences have been seized upon by lawyers trying to undermine the straightforward calculation of ceiling frequencies (although such arguments have not succeeded). Most important, the report failed to state clearly enough that the ceiling principle was intended as an ultra-conservative calculation, which did not bar experts from providing their own 'best estimates' based on the product rule. This failure was responsible for the major misunderstanding of the report. Ironically, it would have been easy to correct.

## A law-enforcement perspective



Opting for conservative procedures - FBI serology laboratory In Washington, DC.

Even as academics debated fine points, forensic scientists got on with their business. The FBI and TWGDAM found the ceiling principle to be unnecessarily conservative, but nonetheless promptly adopted precise guidelines for implementing the ceiling principle, correctly clarifying a few ambiguous statements in the NRC report, such as which population databases to include and whether to sum adjacent bins in a frequency distribution. Forensic labs adopted a two-tiered approach, in which experts are prepared to quote both their best estimate and the conservative ceiling bound. As new population-genetic issues arise (such as how to modify or replace the ceiling principle to accommodate the less polymorphic PCR-based systems), the community is preparing to develop further guidelines. Overall, the system meets the spirit and the letter of the NRC report.

Conservative calculations have had no noticeable impact on the use of, DNA evidence. In the vast majority of cases, a jury needs to know only that a particular DNA pattern is very rare to weigh it in the context of a case: the distinction between frequencies of  $10^{-4}$ ,  $10^{-6}$ , and  $10^{-8}$  is irrelevant in the case of suspects identified by other means.

The FBI has also rapidly carried out population surveys, as recommended by the NRC committee. FBI scientists have studied more than 25 distinct subpopulations, as well as 50 separate samples from the US population<sup>17-21</sup>. The effort has yielded a remarkable database for examining allele frequency variation among ethnic groups. Reassuringly, the observed variation is modest for the loci used in forensic analysis and random matches are quite rare, supporting the notion that the FBI's implementation of the product rule is a reasonable best estimate. Nonetheless, the FBI has taken the scientifically sound position that it remains wiser to study new loci empirically than to assume that significant variation can never occur.

Most important, the admissibility of such DNA evidence prepared in accordance with the NRC recommendations is firmly established in virtually all US jurisdictions. In a few, the appellate courts have yet to rule formally, but there is little doubt that they will find such evidence acceptable.

### Modest proposals

Although the system ain't broke, there is no shortage of proposals about how to fix it. Some academic commentators advocate a return to the product rule; others propose an approach based on the kinship statistic  $F_{ST}$ ; and still others recommend an approach involving likelihood ratios that combine gel electrophoresis artefacts and population-genetic considerations into a single statistic<sup>9,10,22-24</sup>. Some seek to determine genotype frequencies exactly, while others prefer conservative estimates. The NRC - at the urging of the National Institute of Justice, representing the academic wing of forensic scientists - has concluded that the best solution is to constitute another *ad hoc* committee on DNA fingerprinting, composed primarily of statisticians and population geneticists.

It is easy to forget that this new debate is purely academic. The most extreme positions range over a mere two orders of magnitude: whether the population frequency of a typical four-locus genotype should be stated, for example, as  $10^{-5}$  or  $10^{-7}$ . The distinction is irrelevant for courtroom use.

Rehashing issues may be a harmless pastime in the academic world, but not so in a legal system that lives by the dictum *stare decisis* (let the decision stand). From the standpoint of law enforcement, it is better to have a settled, if slightly imperfect, rule than ceaselessly to quest after perfection. Already the NRC's intention to re-examine forensic DNA typing has been seized upon by some lawyers as evidence that there remain fundamental problems.

*Ad hoc* committees typically imagine that they will be able to accomplish their mandate with speed and finality. The original NRC study was anticipated to take one year, but required three. The idea of a second NRC panel was first floated in June 1993 with the optimistic projection that it could report by the end of that year. In fact, the committee has only just begun meeting and will probably not issue a report before late 1995. Even then, any recommendations will take 3 years to ripple through the legal system - guaranteeing that finality will not be achieved on these issues before early 1999. Despite the committee's best efforts, any new report will probably offer new opportunities for misunderstanding that will become apparent only after the panel is disbanded. And, if the new report endorses a different standard, some attorneys are sure to argue, rightly or wrongly, that differences between the reports demonstrate a lack of scientific consensus. These observations are not meant to dissuade the new NRC committee from its mission, but rather to point out the challenge facing any ad hoc group.

## A sounder approach

The real solution is to recognize that forensic DNA typing has become a mature field and requires a more systematic approach. The NRC report anticipated, that rapid evolution of technology would pose a steady succession of questions requiring attention. Its most important recommendation was the establishment of a permanent national committee on forensic DNA typing (NCFDT) to address issues as they arose. If such a committee had been appointed in 1992, it could have made short work of the population genetics issues, by clarifying, changing or discarding the original NRC recommendations.

It is encouraging that this NRC recommendation has recently been adopted. The newly enacted DNA Identification Act of 1994 mandates the FBI to establish a DNA advisory board to recommend standards for laboratory procedures, quality assurance and proficiency testing. The act requires open meetings and broad representation, including molecular and population geneticists not affiliated with forensic laboratories, with board nominations to come from professional organizations including the National Academy of Sciences. Ideally, the panel will provide a forum for weighing important issues, including new laboratory techniques, population genetics and proficiency testing.

Most of all, the public needs to understand that the DNA fingerprinting controversy has been resolved. There is no scientific reason to doubt the accuracy of forensic DNA typing results, provided that the testing laboratory and the specific tests are on a par with currently practiced standards in the field. The scientific debates served a salutary purpose: standards were professionalized and research stimulated. But now it is time to move on.

*Eric S. Lander is in the Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, Massachusetts 02142, USA and the Department of Biology, MIT, Cambridge, Massachusetts 02139, USA. Bruce Budowle is in the Forensic Science Research and Training Center, FBI Laboratory, FBI Academy, Quantico, Virginia 22135, USA.*

1. Lander, E. S. *Nature* **339**, 501-505 (1989).
2. National Research Council. *DNA Technology in Forensic Science* (National Academy Press, 1992).
3. Lewontin, R. C. & Hartl, D. L. *Science* **254**, 1745-1750 (1991).
4. Chakraborty, R. & Kidd, K. K. *Science* **254**, 1735-1739 (1991).
5. *Science* **268**, 201-203 (1994).
6. Weir, B. S. *Proc. natn. Acad. Sci. U.S.A.* **89**, 11654-11659 (1992).
7. Aldhouse, P. *Science* **269**, 755-756 (1993).
8. Devlin, B., Risch, N. & Roeder, K. *Science* **259**, 748-749 (1993).
9. Morton, N. E., Collins, A. & Salust, I. *Proc. natn. Acad. Sci. U.S.A.* **90**, 1892-1896 (1993).
10. Morton, N. E. *Genetics* (in the press).
11. Schellenberger, G. D. et al. *Science* **258**, 668 (1992).
12. Terwilliger, J. D. & Ott, J. *Handbook of Human Genetic Linkage* (Johns Hopkins Univ. Press, 1994).
13. Duncan, G. T., Nopping, K. & Tracy, M. *Genetics* **88**, 51-57 (1993).
14. Krane, D. E. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 10583-10587 (1992).
15. Szymowitz, J. R. & Cohen, J. E. *Am. J. Hum. Genet.* **53**, 314-323 (1993).
16. Weir, B. S. *Am. J. Hum. Genet.* **52**, 437-440 (1993).
17. Federal Bureau of Investigation. *YSTR Population Data: A Worldwide Study* (1993).
18. Budowle, B., Monson, K. L., Ghazi, A. M. & Brown, B. L. *J. Forensic Sci.* **39**, 319-352 (1994).
19. Budowle, B., Monson, K. L., Ghazi, A. M. & Brown, B. L. *J. Forensic Sci.* **39**, 988-1008 (1994).
20. Budowle, B. et al. *J. Forensic Sci.* (in the press).
21. Huang, N. E. & Budowle, B. *J. Forensic Sci.* (in the press).
22. Morton, N. E. *Proc. natn. Acad. Sci. U.S.A.* **89**, 2556-2560 (1992).
23. Evelt, I. E., Sorraige, J. & Pinchin, R. *Am. J. Hum. Genet.* **52**, 498-506 (1993).
24. Balding, D. J. & Nichols, R. A. *Forensic Sci. Int.* **64**, 125-140 (1994).