THE POPULATION GENETICS OF DNA PROFILING; WHY DO THE PROBLEMS PERSIST?

Simon Easteal
Human Genetics Group
The John Curtin School of Medical Research
The Australian National University

Introduction

The use of DNA profiling as evidence in a court case was seriously and successfully challenged for the first time in 1989 (Lander 1989). Many successful challenges have followed, and in the process the problems associated with producing and interpreting DNA profiles have become more widely understood and appreciated. The commercial and government laboratories involved with DNA profiling have taken seriously the need to deal with these problems. In the United States, standards to which acceptable evidence must conform have been established. Elsewhere, including Australia, efforts are being made to make the evaluation of evidence more straightforward. Despite this, there are still cases in which DNA profiling evidence is excluded from courtrooms and the issues involved continue to be hotly debated.

The first challenge to DNA profiling evidence, in the Castro case, focussed primarily on the laboratory procedures used in generating the profiles (Lander, 1989). The technical work had been done poorly and adequate experimental controls had not been performed. The result was that the data could not be interpreted and this was quite rightly pointed out in court. In 1990 in the New South Wales Supreme Court DNA profiling evidence was ruled inadmissible for the first time in Australia in the case of R v Tran. Once again, the main problem was technical inadequacies. In fact the inadequacies in this case were much worse than in the Castro case. In both these cases, however, the issue of how to determine the chance that an observed match between profiles is a coincidence was also raised. It is this issue that remains predominant in discussions about interpreting DNA profiles.

The problem is as follows: a DNA profile is a physical image (usually several dark bands on an X-ray film) that results from applying a complex set of laboratory procedures to DNA obtained from some biological material. The image represents a component of the genetic characteristics of the individual from which it was derived. If two profiles differ from each other then the material from which they were derived (be it blood, semen, hair, skin saliva, or any other part of the body) cannot have come from the same individual. If, however, two profiles match each other this
does not necessarily imply that they came from the same individual. We are all genetically unique, but DNA profiles only represent a small component of our genetic make up, and we are not all unique with respect to our DNA profiles. Two unrelated people may share the same DNA profile.

How do we know whether two identical DNA profiles (for example, one from a crime scene and one from a suspect) came from the same individual (the suspect) and that the crime scene profile did not in fact come from someone else who, by chance, had the same DNA profile? How do we know, in other words, that a match between two profiles is not simply a coincidence. The answer is that we can never know this with certainty. The best we can do is to estimate the probability of the match being a coincidence. This probability depends on how common the matched profile are. Common profiles are likely to be shared by different people and therefore to be found together as a match by coincidence.

In a sense DNA profiles are not different from many other forms of forensic evidence in this respect. Matching of hair or fabric or footprints can all occur by chance, and the probability that they will do so depends on how common they are. There are, however, some factors that are peculiar to estimating the frequency of DNA profiles. These arise for two reasons.

First, the frequencies of most DNA profiles are low enough that they cannot be estimated directly from population samples of a few hundreds or thousands of individuals (it is generally not practicable to collect larger population samples than this). Frequencies have to be estimated indirectly.

Second, in obtaining indirect estimates, we have to make certain assumptions about the population with which we are concerned and about the mating behaviour of the people in it. In most cases it is necessary to make assumptions that are unlikely to be valid.

**Estimating Population Frequencies of DNA Profiles**

The bands in a DNA profile are inherited according to the laws of Mendelian genetics. In a two-band profile from a particular region of DNA (or locus) each band represents a variant form of that DNA region (or allele), one band is inherited from each parent and only one band is passed on to any child. The overall DNA profile that is presented as evidence generally consist of several (usually four) individual one- or two-band profiles representing different regions of DNA.

Estimation of the population frequency of all the bands of a particular DNA profile involves combining the frequencies of the individual bands under certain assumptions about the way in which genes behave in populations. The populations must be in Hardy-Weinberg and linkage (or more appropriately
gametic-phase) equilibrium (see Easteal et al. 1991 and references therein for a description of these, and for a more extensive discussion of their relevance to the problem of estimating the population frequencies of DNA profiles). Populations exhibiting these properties only occur under certain very specific circumstances. Most importantly they only apply in populations that do not consist of a mixture of people who originate from separate populations. Almost all modern human populations consist of people of different racial and ethnic backgrounds. In countries where most people are non-indigenous, such as Australia, this is particularly the case. Modern human populations are almost never in strict Hardy-Weinberg or gametic-phase equilibrium. This is the point that is frequently made by expert witnesses. The assumptions that underlie estimates of the population frequencies of DNA profiles are unlikely ever to be met. So one answer to the question "Why do the problems persist" is that certain population geneticists quite rightly assert that the assumptions underlying the interpretation of DNA profiling evidence are generally invalid. They argue that the violation of these assumptions may have an appreciable effect on the interpretation of results. The available data, however, suggest that this is not the case. The assumptions are violated but this does not appear to make much difference (Evatt 1992; Evatt et al. 1993; Risch & Devlin 1992; Weir 1992). The argument, however, is still made that, while this may generally be true, it need not necessarily be the case in any particular instance (for example, Lewontin & Hartl, 1991). The answer to this argument is provided by the National (USA) Research Council Committee on DNA Technology in Forensic Science (1992). They have advocated a procedure that essentially removes the need to make any population genetic assumptions.

The Conservative Solution

The approach the Committee has recommended is that allele frequencies at the loci used in DNA profiling be estimated in samples from the major ethnic groups that comprise the United States population. The highest frequency in any of the samples is then used to estimate of the population allele frequency, or, if the frequency in none of the samples is higher than 5 per cent, the population frequency is taken as 5 per cent. This approach is referred to as the ceiling principle. The Committee also recommended that until appropriate populations samples have been analysed, a higher ceiling of 10 per cent should be adopted. This approach is conservative in the sense that it provides an over-estimate of the actual population frequency and, thereby, an over-estimate of the probability of coincidental matching. In other words, it favours the defence. The approach has been criticised on the grounds that the choice of ceiling is arbitrary making the whole procedure
unscientific (Devlin et al. 1993; Aldhouse 1993). In response to this, however, Lander (1993) has pointed out that the conservative nature of the estimates obtained from this approach mean that the argument can no longer be made that the probability of coincidental matching has been underestimated. In this sense interpretation of evidence using the Committee’s approach should be generally acceptable.

Old Problems from New Technology

As discussed above, the assumptions of Hardy-Weinberg and gametic-phase equilibrium are rarely, if ever, strictly valid in modern human populations. For the DNA loci routinely used in DNA profiling, allele frequencies do not vary substantially across populations, and the deviation from these assumptions is thus not of sufficient magnitude to cause any appreciable problems with interpretation. The lack of allelic variation across populations is probably attributable, in large part, to high mutation rates and low average allele frequencies. Even in populations that have been isolated from each other for long periods of time the same alleles are continually being regenerated by mutation preventing allele frequency divergence. DNA profiling evidence based on a completely different technical approach is starting to appear. The technique at the heart of this new approach is the Polymerase Chain Reaction (PCR). There are numerous advantages to PCR-based methods. They are cheaper, faster, easier to interpret, and they can be applied to much smaller and older material (they have recently been used to analyse DNA obtained from a 120 Ma old insect; they have not, however, yet been used to re-create dinosaurs).

These new methods, involve the analysis of DNA loci that are less variable and have lower mutation rates than the loci routinely analysed up to now. It is likely that allele frequencies at these new loci will vary across populations to a much greater extent. The effect of such variation on the estimation of coincidental matching probabilities will not be so easy to ignore. For this reason it is that conservative standards be introduced. The extensive time and money that will be expended in disputing the appropriateness of interpretation if such standards are not introduced is entirely predictable and can be avoided. There could be no satisfactory, lasting outcome to such disputation. The information necessary to obtain a statistically correct interpretation of the data would never be available. The issue needs to be put beyond the reach of population geneticists and statisticians, as it has been in the United States by the Committee on DNA Technology in Forensic Science (1992). Failure to do this is in the interests of no-one but those who might, as a consequence, avoid prosecution for crimes committed, and
those who stand to gain financially by providing expert testimony.

References


