

Newton E. Morton

DNA in Court

CRC Research Group in Genetic
Epidemiology,
University of Southampton,
The Princess Anne Hospital,
Southampton, UK

Blood groups and isozymes have been used for many years as evidence on identity and paternity. With the introduction of HLA and other more polymorphic systems it became possible not only to exclude a suspect but to reach small probabilities of a random match. In that event the defence might question the laboratory technique or the sample on which matching probabilities were calculated but dispute among scientists was limited to statistical niceties, especially the debate between Bayesians who want to introduce the prior probability of a match and classicists who consider this too controversial to be of practical use.

Mild disagreement within the scientific fraternity was fanned into a civil war by DNA technology. Jeffreys applied Southern blots and his multilocus probes to exclude a defendant and identify the culprit in a large sample of potential suspects [1]. Rapid technical advances led to use of single-locus probes and

the polymerase chain reaction (PCR), with the prospect that fragment lengths will be replaced by sequences in the near future. Extremely small matching probabilities were obtained with a small number of probes, which if accepted would be conclusive. However, the very power of the methods created distrust, not only among defence attorneys and judges but also among some mathematicians and population geneticists. They could point to technical problems, disagreement among statisticians as to the best way of evaluating DNA evidence, and especially to subtleties in population structure. Suppose that the probability of a random match were infinitesimal in a plausible population: might there not be an even more relevant population in which the probability would be considerably larger? None of the hard critics was a human geneticist, but their attempts to evaluate evidence on population structure in man fuelled their concerns. Defenders of DNA identification counterattacked. The larger community of geneticists, except for the few with a special interest in population structure, read about the controversy in *Science* editorials that gave the flavour of strife between lawyers and expert witnesses, but no coherent account of the

This review and commentary represents an invited contribution to the EJHG. Its content reflects the views of a distinguished scholar and expert in the field whom the Editorial Board of the EJHG gratefully acknowledges for reviewing the current debate on Forensic Genetics.

issues that divided those witnesses. It is not surprising that some courts responded irrationally by rejecting DNA evidence or admitting it with little weight.

To end this controversy, the National Research Council established a committee chaired by a distinguished clinical geneticist (Victor McKusick) and including a mathematician (Eric Lander) and a molecular biologist active in DNA identification (Thomas Caskey). The latter was forced to resign over a questionable conflict of interest, and the published report has sown confusion in courts and disappointment in the two groups (forensic statisticians and population geneticists) who were not represented on the committee. Bruce Weir is especially concerned about errors in statistical theory and formulae, Ian Evett asks why matching criteria were not considered, and Neil Risch wonders what principles justify use of gene frequencies borrowed from irrelevant populations or simply fabricated. Several national groups, including the Federal Bureau of Investigation and the London Metropolitan Police, have scheduled conferences to consider the unresolved issues.

Any serious account of this controversy faces three obstacles. First, the DNA techniques themselves are evolving rapidly, with multilocus probes being now used only to a limited extent with DNA in prime condition, as in paternity trials, while PCR has recently been introduced and sequencing with its poorly understood error frequencies is not yet practical. Therefore current interest in single-locus RFLP fragment lengths has a short history and may not persist for long. Secondly, the statistical aspects are also evolving, with matching probabilities now seen as a component of the likelihood ratio that also subsumes matching criteria. There is diversity in these statistics in response to the continuous variation of fragment length estimates, which may

be smoothed, weighted, binned or windowed. The computer programs that produce matching probabilities and likelihood ratios have as much need for comparison and quality control as the DNA technology that generates the data, and this need has not been met. Thirdly, human biology has been preoccupied with population structures of long duration, thought to be typical of the pre-industrial condition. Therefore we have many studies of remnant populations, isolates, Amazonian villages, and Micronesian atolls, but fewer studies of national populations divided into communes or provinces. For industrial populations there is evidence that genetic substructure has been greatly reduced, but large samples of potential mates such as are generated by paternity trials have yet to be studied.

Although these obstacles cannot yet be removed, four basic points will be addressed:

- (1) problems in DNA typing;
- (2) evidence on population structure;
- (3) principles for evaluation of DNA evidence, and
- (4) methods that violate these principles.

DNA Typing

Many of the technical problems in forensic use of DNA are common to molecular biology. Samples must be correctly identified, typed blindly against a molecular standard, and reagents and protocol must be rigorously controlled. In practice this means that the forensic laboratory should submit to frequent assessment by a national agency, and only accredited laboratories should be accepted in court under close surveillance over identity of samples. The error frequencies reported in linkage studies have multiple causes, including misidentification of samples and clerical errors in recording results, that would be much less frequent if tests were blindly repli-

cated under conditions as stringent as obtain in an accredited forensic laboratory. The defence has every reason to question a reported match and to demand a repetition if that would serve the defendant, but there is general agreement that in responsible hands DNA technology is admissible and no more subject to error prejudicial to the defendant than other forensic evidence, despite the fact that some evidentiary samples are partially degraded, contaminated with misleading DNA, or of insufficient quantity for reliable typing with the technique used. A partially degraded or electrolytically different sample may migrate at a slightly different rate than fresh DNA, leading to false exclusion of the suspect. This can be controlled by monomorphic bands of similar size, which tend to show the same disturbance. PCR gives smaller fragments that are seldom affected and require small quantities of DNA. However, at very low concentration requiring many cycles of replication, there are artefacts including contamination that could lead to false exclusion, or rarely to a false match if a heterozygote were mistyped as a homozygote. This is controlled by restricting PCR to adequate amounts of DNA and to typing other loci. The ultimate protection for an innocent suspect is that artefact or error leading in exceptional circumstances to a false match at one locus will not prevent exclusion at other loci. Only if the quantity of DNA in the evidentiary sample is insufficient (and therefore the evidence of low weight) can DNA technology generate a problem of concern to an innocent suspect.

Population Structure

An expert witness who testifies that two DNA samples do not match must defend his molecular technology, but population genet-

ics is not at issue. On the contrary, evidence that two samples match requires an assessment of the probability that two random individuals from an appropriate population have indistinguishable genotypes at the loci tested.

Proponents of DNA profiling have argued that population structure is negligible within a sample of a particular racial group from a large region (for example, New York Caucasians or Blacks or California Hispanics). In support of this position they have calculated matching probabilities within the population of the suspect (which among valid calculations is the most favourable to the suspect) and with different populations. The two calculations have been in substantial agreement.

The standard way of describing population structure is by kinship, the probability that two alleles sampled in a particular way be identical by descent. The inbreeding of an individual equals the kinship between his parents. However, kinship is not limited to potential mates and may be estimated for two individuals of different generations, the same sex, or for populations that never intermarry. The importance of kinship is that it describes the relationship between gene and genotype frequencies, and so gives matching probabilities and likelihood ratios. Several isolates have been described in which kinship is greater than 0.1 because of few founders and/or preferential consanguineous marriage. However, none of these isolates occurs in an industrial society, where kinship by a variety of methods has been estimated to be about 0.001 [2], a value that is negligible for larger gene frequencies. Even such a small value must be diluted in many urban populations, although preferential consanguineous marriages persist in some circumstances and generate kinship that is not negligible [3]. Despite bias of population geneticists toward 'primitive' populations, there is information about kinship in forensic populations.

Classical evidence on population structure comes from migration, genealogy, surname concordance (is_{Hartl}ly), and bioassay of kinship from blood groups and isozymes. For the few hypervariable loci that have been bioassayed, differentiation among major racial groups is much less than for blood groups and isozymes, some of which are exposed to diversifying selection [4]. From myotonic dystrophy and the fragile X syndrome it may be inferred that hypervariable loci are subject to normalizing selection, opposed by high mutation rates [5]. This is the situation which Malécot [6] predicted would lead to low kinship, as observed. Within an ethnic group these forces are dominated by migration, and so no significant difference would be expected among estimates derived from DNA markers and other sources. Available data support this prediction.

In contradiction to this large body of evidence Lewontin and Hartl [7] have claimed strong differentiation among national populations in Europe. The gene frequencies they chose lie far outside most other estimates [8] corresponding to a kinship of 0.067. One of their samples, the AB0 group in Poles, has never been identified, and the others are extremely atypical. Less than 2% of the kinship in their samples is due to national differences to which they attribute it [4]. The Lewontin and Hartl paper is to population genetics what the Piltdown hoax was to paleontology.

Principles of DNA Evidence

Use of DNA in court is based on the ratio between the probability of the evidence if the suspect and the evidentiary sample are identical to the probability if they are different. This is called the likelihood ratio [9]. A ratio greater than 100 is usually considered strong evidence of identity, while a ratio less than

1/100 is strong evidence of exclusion [10]. In the interval, the evidence is suggestive but to reach a decision more loci should be tested [11]. However, if there is no other evidence against the suspect (for example if he were charged merely because of an apparent match in a group of n potential suspects) the likelihood ratio to assert identity should be more conservative: a reasonable choice is $100n$. Use of the likelihood ratio does not require these rules which are no more than a guide, and all the DNA evidence is given by the likelihood ratio whether these rules are accepted or not.

In general, evidence consists of a part relating to the suspect and a remainder associated with the evidentiary sample. The latter may for brevity be called the culprit, although in some cases this would be inappropriate. Since the evidence for the suspect is the same under all hypotheses about the culprit, the population and other characteristics of the suspect do not enter into the likelihood ratio. However, the population of the culprit is hypothetical, and there are various ways to handle this uncertainty. There may be testimony or circumstantial evidence favouring a particular population; the proportions found in the regional forensic database may be relevant; or courtesy may dictate that the culprit be assumed to belong to the same population as the suspect, since this is most favourable to the defence. These choices make surprisingly little difference to the likelihood ratio [12] but the defence naturally wants to exploit whatever disagreement there may be. Therefore calculation of the likelihood ratio under various hypotheses is a necessary part of the DNA evidence.

Besides claiming a favourable population for the culprit, the defence may argue that the culprit is related to the suspect in some way. This generates an infinite number of hypotheses. Relationship may be bilineal (monozygotic twins, sibs, double first cousins, and multiple remote relationship) or unilineal, the

suspect or culprit may or may not be inbred, and the relationship may be as close or closer than usually obtains between mates. Most of these possibilities have been considered in the forensic context [2], and population genetic theory provides for more complex relationships [13, 14].

Faced with this cornucopia of alternatives, the prosecution may argue that no suspicion falls on any relative, and therefore the culprit should be considered to be randomly drawn from the same large forensic population. The expert witness does not share the court's responsibility to weigh the evidence, but he must be prepared to determine the likelihood ratio under whatever hypotheses the court entertains, and this should generate much more computation than is currently expected. Although there is no limit to the variety of crime or the ingenuity of the defence, in most cases the argument reduces to the structure of forensic populations that can be described by kinship, and therefore to empirical studies of kinship in relevant populations. To define a 'relevant population' seven principles have been proposed [4]:

(1) There is no connection between a matching probability in one population and gene frequencies in an unrelated population.

(2) An upper bound to a probability is not a probability.

(3) Of the indefinitely large number of ways in which such an upper bound may be estimated, few conform to generally accepted theory.

(4) For every genotype-specific matching probability there is a confidence interval, a mean matching probability, a genotype-specific likelihood ratio, and a mean likelihood ratio under credible hypotheses about the population of the culprit and his relationship to the suspect. Consideration of these alternatives by the court protects adequately against excessive reliance on evidence of identity.

(5) An acceptable bound must not violate statistical or genetic principles or known values of gene frequency or kinship.

(6) Gene frequencies should be estimated in large samples to minimize sampling error.

(7) In the absence of evidence to the contrary, the suspect and culprit should be assumed to be randomly drawn from a forensic population. Contrary evidence may be accommodated by the affinal or other model of population genetics and by an appropriate estimate of kinship, without altering gene frequencies in the reference population.

Testimony that violates any of these principles should not be admissible in court.

Invalid Approaches

Nichols and Balding [15] tried to incorporate kinship into matching probabilities. They chose a high value of kinship (0.05) that lies far above the range for industrial populations, and the approximation they give does not correspond to any population structure. Even the claim that it provides an upper bound is not strictly true [4], and it violates principles 2, 3, 4, 5, and 7. However, Nichols and Balding, who are not human geneticists, are to be praised for attempting to use kinship in forensics.

The approach of Lander [16], adopted with exaggeration by the NRC Committee, cannot be justified. It violates all seven principles by taking gene frequencies (if greater than 0.1) from an alien and irrelevant sample (for example, Lapps or Cambodians) and as 0.1 otherwise. To defend this argument in court it is necessary to claim a coherent scientific theory behind this arbitrary rule and its acceptance by most human geneticists. Neither claim can be sustained.

The proposal of Lewontin and Hartl [7] looks for a population most favourable to the

defence, to which the suspect need not belong. Thus the culprit might be assumed to be a Basque for one locus and a Maltese for another. Contrarily, they also propose the subpopulation to which the suspect may belong, rejecting the larger forensic population of which it is a part. However, in general there is no reason to claim that the culprit comes from exactly the same ancestry as a suspect unless the latter is guilty. The population of the suspect can be known only through his testimony, appearance, or limited civil records, and would be modified by a Fifth Amendment plea. Therefore, only a general population should be specified, with infrequent instances of a crime in an isolated community requiring separate treatment. In any case, kinship provides both the theory and the set of relevant estimates.

Discussion

Early DNA profiling was done by molecular biologists who were naive in statistics and population genetics, and therefore vulnerable to a number of criticisms. Matching probabilities are clearly inferior to likelihood ratios unless classification error is negligible. As the distinguished statistician Chernoff [17] has remarked 'The match/binning approach, as described, doesn't make much inferential sense'. The assumption of no kinship between suspect and culprit must always be defended, and in some circumstances is significantly wrong. An expert witness who is unprepared to handle these complexities does the court a disservice. The strong criticism to which DNA profiling has been exposed is understandable, to some extent justified, and has stimulated improvements in evaluation of evidence.

We are now in a phase of synthesis. Which of the elegant likelihood ratio procedures is

best, or are they equivalent? What models and parameters of population structure are most appropriate to forensic samples? How will DNA profiling evolve with advances in molecular biology, including PCR, digitize [18] and sequencing? We may expect to see increased use of PCR, possibly including sperm typing. Digitizing methods that recognize nonoverlapping DNA sequences eliminate errors implicit in fragment size but introduce other errors through haplotype inference. The ultimate advance will be to sequencing, which (if without error) would allow return to matching probabilities. Much remains to be done to develop and control the quality of DNA profiling, including statistical and genetic aspects, which forensic laboratories have not yet assimilated. However, the pace of progress has already made the recommendations of the NRC obsolete. They do not rest on a well-supported theory and are rejected by statisticians like Berry, Evett, Chernoff, and Devlin and by population geneticists like Lange, Risch, Weir, Chakraborty, and Kidd.

Statistical analysis must evolve with molecular techniques. 'Band-shifting' may require covariance adjustment by a monomorphic standard. There is urgent need for as much quality control over forensic inference as for laboratory competence. Inevitably visual matching and bin matching of fragment lengths will be replaced by a continuous metric. There is an infinity of methods and parameters for density estimation, which are redundant and may be misleading. A simple alternative is to use raw frequencies in which the evidentiary sample has been included so that the likelihood ratio must be finite, and to compare with the mean likelihood ratio that is insensitive to sampling errors: the larger the number of probes the less difference there will be between the mean and genotype-specific estimates. Recent population admixture reduces disequilibrium between unlinked loci

only by 1/2 per generation, but in practice VNTR alleles at different loci have been found to be independent [19] or so weakly and inconsistently dependent as to be manageable by pairwise analysis [2]. Although all these and other developments remain for the

near future, they lie within the competence of professionals who know enough statistics and genetics to differ only in details. In a knowledgeable court DNA profiling is no longer exposed to risk of illogical presentation, blind acceptance or arbitrary rejection.

References

- 1 Wernett DJ, Lygo JE, Sutton JG: The introduction of DNA analysis into Home Office forensic science laboratories in England and Wales; in Ballantyne J, Sensabaugh G, Witkowsky J (eds): *DNA Technology and Forensic Science*. Cold Spring Harbor Laboratory Press, 1989.
- 2 Morton NE: Genetic structure of forensic populations. *Proc Natl Acad Sci USA* 1992;89:2556-2560.
- 3 Morton NE: Kinship and inbreeding in populations of Middle Eastern origin and controls; in Crawford MH, Mielke JH (eds): *Current Developments in Anthropological Genetics: Ecology and Population Structure*. New York, Plenum Press, 1981, vol 2, pp 449-466.
- 4 Morton NE, Collins A, Balazs I: Bioassay of kinship for hypervariable loci in Blacks and Caucasians. *Proc Natl Acad Sci USA*, in press.
- 5 Morton NE, Macpherson JN: Population genetics of the fragile-X syndrome - Multiallelic model for the FMR1 locus. *Proc Natl Acad Sci USA* 1992;89:4215-4217.
- 6 Malécot G: *Les mathématiques de l'hérédité*. Paris, Masson, 1948.
- 7 Lewontin RC, Hartl DL: Population genetics in forensic DNA typing. *Science* 1991;254:1745-1750.
- 8 Ckarakorty R, Kidd KK: The utility of DNA typing in forensic work. *Science* 1991;254:1735-1739.
- 9 Berry DA: Inferences using DNA profiling in forensic identification and paternity cases. *Stat Sci* 1991;6:175-205.
- 10 Devlin B, Risch N, Roeder K: Forensic inference from DNA fingerprints. *J Am Stat Assoc* 1992;87:337-350.
- 11 Wald A: *Sequential Analysis*. New York, Wiley, 1947.
- 12 Berry DA, Evett IW, Pinchin R: Statistical inference in crime investigations using deoxyribonucleic acid profiling. *Appl Stat* 1992;41:499-531.
- 13 Cotterman CW: A calculus for statistico-genetics; PhD Thesis, Ohio State University; in Ballonoff P (ed): *Genetics and Social Structure*. Stroudsburg, Dowden, Hutchinson & Ross, 1940, pp 157-272.
- 14 Jacquard A: *Genetics of human populations*, transl. Yermanos DM. San Francisco, Freeman, Cooper, 1978.
- 15 Nichols RA, Balding DJ: Effects of population structure of DNA fingerprint analysis in forensic science. *Heredity* 1991;66:297-302.
- 16 Lander ES: Research on DNA typing catching up with courtroom application. *Am J Hum Genet* 1991;48:819-823.
- 17 Chernoff H: Comment. *Stat Sci* 1992;6:192-196.
- 18 Jeffreys AJ, MacLeod A, Tamaki K, Neil DL, Monckton DG: Minisatellite repeat coding as a digital approach to DNA typing. *Nature* 1991;354:204-209.
- 19 Weir BS: Independence of VNTR alleles defined as fixed bins. *Genetics* 1992;130:873-887.