Available online at www.sciencedirect.com
SCIENCE DIRECT®
Forensic Science International 152 (2005) 115–119
Forensic Science International
www.elsevier.com/locate/forsciint

Relatedness and DNA: are we taking it seriously enough?

John Buckleton [a], Christopher M. Triggs [b,*]

[a] *ESR, The New Zealand Forensic Science Service, P.B. 92021, Auckland, New Zealand*
[b] *Department of Statistics, University of Auckland, P.B. 92019, Auckland, New Zealand*

Received 29 April 2004; received in revised form 29 July 2004; accepted 30 July 2004
Available online 29 September 2004

Abstract

In forensic DNA testimony most DNA laboratories report the match probability for an unrelated person from some relevant population. These laboratories typically make available the match probability for relatives when requested. This practice has served well for many years. However, as the discrimination power of our multiplexes has increased the estimated match probabilities for both related and unrelated people have become markedly smaller. Associated with this general reduction in match probabilities have been the observations that the relative balance between the match probabilities of the many unrelated people and the few relatives of a suspect has changed. We suggest that we should now report routinely the match probability for a sibling whenever the suspect has a non-excluded sibling.

Keywords: DNA interpretation; Relatedness; Match probability

1. Introduction

Typically in forensic DNA testimony a sample of biological material associated with a crime is compared to a sample from a suspect. The genotyping results from the scene and the suspect are compared. If the suspect can be excluded as a contributor to the material at the scene then that is usually the end of the matter. However, if he cannot be excluded it is customary to provide some assessment of the strength of the DNA evidence. Historically this has been undertaken by assessing the probability of this evidence if the suspect is, indeed, not the donor of the stain. If the suspect is not the donor of the stain then we need to explain the evidence. Possible explanations include:

- a laboratory error,
- a relative was the donor, or
- an unrelated person was the donor.

In this paper, we consider the relative impact of the latter two explanations. Evaluation of these options is not original. We, rather, call for a re-examination of reporting practice in view of the evolution of modern multiplexes to a larger number of loci.

The first DNA profiles were termed ‘‘multilocus profiles’’ [1–3]. These were based on minisatellite repeat sequences. After digestion of the DNA radiolabelled probes were applied at low stringency that bound to the fragments from a large number of loci. The resultant autoradiographs looked somewhat like a bar code and this was an analogy that was commonly used. At the time (and subsequently) it was not know how many loci were involved or which bands were allelic or linked [1,2,4,5].

The next stage of DNA evidence involved the same restriction fragment length technology (RFLP) but the probes were applied at high stringency and were designed

* Corresponding author.
E-mail addresses: john.buckleton@esr.cri.nz (J. Buckleton), triggs@stat.auckland.ac.nz (C.M. Triggs).

0379-0738/$ – see front matter © 2004 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.forsciint.2004.07.020

to visualise a single locus. These loci were still minisatellites. As expected, individuals typically showed only one or two alleles per locus. Laboratories originally implemented between three and nine loci for such tests.

The next evolution was the advent of PCR, coupled with the visualisation of STR loci. This is the method of choice today. Early multiplexes visualised three or four loci progressing through six loci and now many forensic laboratories implement multiplexes that visualise 9, 10, 13 or 15 STR loci and paternity testing laboratories may have more. These developments are expensive and involve high implementation costs, but are highly desirable from the point of view of discrimination.

At this point it is important to state that we should not get into a ''locus counting'' frame of mind. In court we have often experienced questions suggesting the more loci the better. Hence, why did we not use 10, 11, 12, or more loci? There are many factors that affect the utility of a multiplex and the number of loci is only one of them. Each locus may have differing polymorphisms and there may be other factors suggesting the inclusion of a particular locus in a multiplex system. These other factors may include such things as the availability of suitable primers, mutation rates, heterozygote balance, and performance under the compromise conditions of the multiplex.

Most DNA evidence is reported with a measure of the strength of the evidence. This measure is most often an estimated match probability or a likelihood ratio. These match probabilities or likelihood ratios are typically calculated for an ''unrelated'' person. This may be done using either the product rule or modifications that take account of subpopulation effects [6]. Sampling uncertainty may, or may not, be assessed for each case [7–12].

As more loci are added to multiplexes the various assumptions underlying the estimation of match probabilities become more difficult to test. Of note is that the number of assumptions of approximate linkage equilibrium within the population (when using the product rule) or subpopulation (when using Balding and Nichols' equations [6]) increases as a function of the number of loci. However, in general, as more loci have been added the discriminating power of the multiplexes is believed to have increased.

Occasionally the match probability for a relative may be requested by the defence and is duly reported. Sometimes close relatives are eliminated by genotyping. It is certainly not a novel suggestion that the effect of relatives should be assessed [6,10,13–19]. The formulae for match probabilities for siblings are quoted in Appendix A.

The question discussed here is whether or not the time has come to routinely report the match probability for a sibling as well as that for an unrelated person in all DNA casework. If decided upon, this change will involve some effort since the technology to implement such a suggestion for mixed samples exists but is not yet developed into user-friendly systems [20].

2. Results

It is necessary to consider by what criterion we may decide whether or not to include the estimated match probability for a sibling in addition to that for an unrelated person. Clearly, including an additional number makes an already complex evidential statement more so. There would need to be a good reason to include an additional number in a statement. It would seem reasonable to include this number if it was ''important'' for the decision making process.

How could we know whether or not the match probability for a sibling was important? David Balding has pointed the way clearly in his papers [14–16]. We take up his argument. As many multiplexes contain the amelogenin locus that allows sex determination we assume that the stain at the scene is from a known sex, say male. Hence, we confine ourselves to the consideration of male suspects. Consider a very simplistic view of the population of the US. Since we are restricting ourselves to males we model this population as containing the suspect, the single brother of this suspect, and 125 million unrelated males. Clearly, this model is simplistic but it allows us to make the necessary point most straightforwardly. Adding further relatives of the suspect, such as additional siblings, to the population simply strengthens the point.

Suppose that some non-genetic evidence has been produced or will be produced that suggests that the suspect is the donor of the stain DNA. The defence may also produce some evidence that suggests that he is not the donor. Such evidence affects the jurors' assessment of the probability that the suspect is the true donor of the material. Since this evidence is separate to the genetic evidence the probability based upon it is termed the ''prior probability'' even though it may not necessarily precede the genetic evidence. Typically this prior probability is unknown to the scientists. It is most likely to be a non-numerical subjective evaluation in the minds of the jurors.

There may also be non-genetic evidence for or against the other possible donors. These alternative donors are the single brother and the 125 million unrelated males in our simple population. There may be evidence that has eliminated some of these. They may have been genotyped or eliminated by other means. In more rare instances there may be evidence suggesting a specific alternative donor or donors. In all cases there will be information such as opportunity, ability, or access to the crime scene that may affect the probability that certain persons are donors. This information may be of such simple form as that small children or invalids are unlikely to have committed a criminal act involving physical strength.

It is important to consider any non-genetic evidence that points towards or away from the single brother of the suspect. For the purposes of this argument we assume that there is no evidence of this sort. If this assumption

is incorrect this alters the argument. If the non-genetic evidence points towards the brother being an alternative donor then the argument to include a match probability for him is stronger. If the non-genetic evidence points away from the brother being an alternative donor then the argument to include the match probability for him is weakened.

Consider the simplified situation where the crime scene stain is left either by the suspect, or by his brother, or by one of the 125 million unrelated males. In each case there is some probability, after the genetic and non-genetic evidence is given that the suspect is the true donor. This probability after the genetic and non-genetic evidence is called the posterior probability. This posterior probability becomes progressively higher as the match probabilities for the brother and unrelated males becomes lower. However, to calculate this posterior probability we require the prior probability that the suspect is the donor relative to the other potential donors. This is unlikely to be available. For the purposes of the following experiment we assume that the priors for the suspect and the brother are equal. We further assume that the suspect and the brother have the same prior as the average of the unrelated people.

Using these assumptions we can see that as more loci are added the posterior probability that the suspect is the true donor approaches 1. In Table 1, we give the posterior probability for 10,000 simulated profiles for the 3 locus CTT triplex, the 9 locus Profiler plus set, the 13 locus CODIS set, the 15 autosomal locus Powerplex 16 set, or a profile of a set of 19 loci. Details of these calculations are given in Appendix A. From Table 1, we see that as the number of loci in a multiplex increases the posterior probability assigned to the suspect approaches one.

The assumption that the suspect has the same prior as his brother and the unrelated people is typically untenable. However, any such assumption allows us to make the point that the posterior probability that the suspect is the true donor increases as the match probability of his brother and the unrelated people drops. Further, this probability becomes very high unless the prior for the suspect is low relative to the other possible donors. These very high posterior probabilities or very low match probabilities have suggested to some commentators that there may be a threshold beyond which it is acceptable to state that the suspect is, indeed, the donor of the stain [17]. (see also [16,21,22]). However, most scientists have been reluctant to take this step on the reasonable basis that to do so:

1. Accepts the assumptions inherent in the model, such as equal priors.
2. Rounds the posterior probability to 1.0, so excluding the possibility that anyone other than the suspect could have left the crime scene stain.

In many commentators' view, including our own, these actions are deemed acceptable by the finders of fact but not by a forensic scientist.

We need to consider what may be done if we do not state that the suspect is, indeed, the donor of the stain. Most commentators attempt to equip the court with estimates upon which it can form a view of the strength of the evidence. As noted above this is typically done by quoting a match probability. The central thesis of this paper, then, is whether this match probability should be that for an unrelated person or for both the unrelated person and a brother. To do this we examine what contributes to the "remainder" of the posterior probability that is not assigned to the suspect. This remainder represents the doubt, if any. It must fall on either the sibling or on one or more unrelated males.

The posterior probability that the suspect is, indeed, the true donor is usually unknown since we do not know his prior probability. Nor do we need to know it for this exercise. What we need to do is consider the "remainder" of the posterior probability. What is contributing to this? If this is significantly arising from the single brother then we should report his estimated match probability.

If we make the assumption that the non-genetic evidence suggests that the brother is neither more, nor less likely to be the donor than the unrelated people then we can calculate what fraction of the remaining posterior probability has arisen from this brother.

This calculation was undertaken again for 10,000 profiles. The results are given in Table 2. Since brothers are more likely to share alleles than unrelated people their match probability is significantly higher. However, the difference varies. For common alleles the difference between the brother's match probability and the unrelated match probability will be smaller, for rare alleles it will be larger. Some of these simulated cases will involve many common alleles. For these cases it may transpire that the majority of the remaining posterior probability arises from the unrelated

Table 1
Posterior probabilities for the hypothesis that the suspect is the true donor of the stain from 10,000 simulated profiles for various multiplexes

	CTT triplex	Profiler	CODIS	Powerplex 16	19 Loci
Mininum	0.000000943	0.817554044	0.999927748	0.999998414	0.999999920
Mean	0.000030687	0.997389447	0.999995385	0.999999552	0.999999983
Maximum	0.008433091	0.999965490	0.999999534	0.999999958	0.999999998

In this simulation we have assumed that the suspect, his brother and the average of the unrelated people have equal prior probabilities.

Table 2
The percentage of simulated cases against the fraction of the remaining posterior probability that has arisen from the brother

Fraction of the remaining posterior probability assigned to the single brother	CTT triplex (%)	Profiler plus (%)	CODIS (%)	Powerplex 16 (%)	19 Loci (%)
0.0–0.1	100	31	0	0	0
0.1–0.2	0	17	0	0	0
0.2–0.3	0	11	0	0	0
0.3–0.4	0	8	0	0	0
0.4–0.5	0	6	0	0	0
0.5–0.6	0	6	1	0	0
0.6–0.7	0	5	1	0	0
0.7–0.8	0	5	3	0	0
0.8–0.9	0	5	7	0	0
0.9–1.0	0	7	86	100	100

people. For example we see that for 31% of simulated Profiler Plus profiles 0.0–0.1 of the remaining posterior probability has arisen from the brother and hence 0.9–1.0 has arisen from the unrelated people. For 7% of simulated Profiler Plus profiles 0.9–1.0 of the remaining posterior probability has arisen from the brother, while the fraction arising from unrelated people drops to 0.0–0.1.

From Table 2 we see that the posterior probability assigned to the brother increases as the number of loci in a multiplex increases. Thus, for most multiplexes currently in use a significant or even dominating fraction of the remaining posterior probability arises from the single brother.

We therefore conclude that it is time that the match probabilities for a sibling are reported in all casework involving many loci where the suspect has a non-excluded sibling.

Acknowledgements

We thank SallyAnn Harbison, Matthew Phipps and two unnamed referees for comments that greatly improved this paper.

Appendix A

The simulation has been undertaken using allele probabilities from a published source that had data for 19 loci [23]. 10,000 profiles were selected randomly based on their probabilities. These profiles were either the 3 locus CTT triplex, the 9 locus Profiler plus set, the 13 locus CODIS set, the 15 autosomal locus Powerplex 16 set, or a profile of all 19 loci. The match probabilities were calculated for "unrelated" people using Balding and Nichols' equations [6] (also published as equations 4.10 of NRC II [9], 4.20 of Evett and Weir, or 3.4 of Buckleton, Triggs and Walsh [20].)

$$P_i = \begin{cases} \dfrac{[3\theta + (1-\theta)\,p_{i1}][2\theta + (1-\theta)\,p_{i1}]}{(1+\theta)(1+2\theta)}, & A_{i1} = A_{i2} \\ \dfrac{2[\theta + (1-\theta)\,p_{i1}][\theta + (1-\theta)\,p_{i2}]}{(1+\theta)(1+2\theta)}, & A_{i1} \neq A_{i2} \end{cases}$$

Match probabilities for siblings [6,10,20] were calculated using

$$P_i = \begin{cases} \dfrac{1}{4}\left(1 + \dfrac{2(2\theta + (1-\theta)\,p_a)}{1+\theta} + \dfrac{(2\theta + (1-\theta)\,p_a)(3\theta + (1-\theta)\,p_a)}{(1+\theta)(1+2\theta)}\right), & A_{i1} = A_{i2} \\ \dfrac{1}{4}\left(1 + \dfrac{2\theta + (1-\theta)(p_a + p_b)}{(1+\theta)} + \dfrac{2(\theta + (1-\theta)\,p_a)(\theta + (1-\theta)\,p_b)}{(1+\theta)(1+2\theta)}\right), & A_{i1} \neq A_{i2} \end{cases}$$

In both cases $P = \prod_i P_i$

A θ value of 0.01 was used.

To calculate the posterior probability that the suspect is the donor of the stain we assess the formula discussed extensively previously [10,14,15,20]

$$\text{Posterior probability} = \frac{1}{1 + \text{match probability for the sibling} + 125{,}000{,}000 \times \text{match probability for unrelated}}$$

Assuming a population of one sibling and 125 million unrelated persons and that the prior for the sibling is equal to the average for the unrelated people then the fraction of the posterior attributable to the sibling is

$$\frac{\text{match probability for the sibling}}{\text{match probability for the sibling} + 125{,}000{,}000 \times \text{match probability for unrelated}}$$

References

[1] A.J. Jeffreys, V. Wilson, S.L. Thein, Individual-specific 'fingerprints' of human DNA, Nature 316 (1985) 76–79.

[2] A.J. Jeffreys, V. Wilson, S.L. Thein, D.J. Weatherall, B.A.J. Ponder, DNA 'fingerprints' and segregation analysis of multiple markers in human pedigrees, Am. J. Hum. Genet. 39 (1986) 11–24.

[3] P. Gill, A. Jeffreys, D. Werrett, Forensic application of DNA 'fingerprints', Nature 318 (1985) 577–579.

[4] A. Jeffreys, V. Wilson, S.L. Thein, Hypervariable 'minisatellite' regions in human DNA, Nature 314 (1985) 67–73.

[5] A.J. Jeffreys, J.F.Y. Brookfield, R. Semeonoff, Positive identification of an immigration test-case using human DNA fingerprints, Nature 317 (1985) 818–819.

[6] D.J. Balding, R.A. Nichols, DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands, Forensic Sci. Int. 64 (1994) 125–140.

[7] D.J. Balding, Estimating products in forensic identification using DNA profiles, J. Am. Stat. Assoc. 90 (1995) 839–844.

[8] R. Chakraborty, M.R. Srinivasan, S.F. Daiger, Evaluation of standard errors and confidence intervals of estimated multilocus genotype probabilities and their implications in DNA, Am. J. Hum. Genet. 52 (1993) 60–70.

[9] NRC II, National Research Council Committee on DNA Forensic Science, The Evaluation of Forensic DNA Evidence, National Academy Press, Washington, DC, 1996.

[10] I.W. Evett, B.S. Weir, Interpreting DNA Evidence–Statistical Genetics for Forensic Scientists, Sinauer Associates, Inc., Sunderland, MA, 1998.

[11] J.M. Curran, J.S. Buckleton, C.M. Triggs, B.S. Weir, Assessing uncertainty in DNA evidence caused by sampling effects, Sci. Justice 42 (2002) 29–37.

[12] B.S. Weir, J. Curran, J. Buckleton, C. Triggs, I.S. Painter, Assessing sampling error in DNA evidence, 2000, personal communication.

[13] I.W. Evett, Evaluating DNA profiles in a case where the defence is "it was my brother", J. Forensic Sci. Soc. 32 (1992) 5–14.

[14] D.J. Balding, P. Donnelly, Inference in forensic identification, J. R. Stat. Soc. Ser. A 158 (1995) 21–53.

[15] D.J. Balding, P. Donnelly, Inferring identity from DNA profile evidence, Proc. Natl. Acad. Sci. USA 92 (1995) 11741–11745.

[16] D.J. Balding, When can a DNA profile be regarded as unique? Sci. Justice 39 (1999) 257–260.

[17] B. Budowle, R. Chakraborty, G. Carmody, K.L. Monson, Source attribution of a forensic DNA profile, Forensic Sci. Commun. 2 (2000), http://www.fbi.gov/programs/lab/fsc/backissu/july2000/source.html.

[18] P. Donnelly, Nonindependence of matches at different loci in DNA profiles: quantifying the effect of close relatives on the match probability, Heredity 75 (1995) 26–34.

[19] N. Fukshansky, W. Bar, Biostatistics for mixed stains: the case of tested relatives of a non-tested suspect, Int. J. Leg. Med. 114 (2000) 78–82.

[20] J.S. Buckleton, C.M. Triggs, S.J. Walsh, Forensic DNA Evidence Interpretation for Caseworkers and Lawyers, CRC Press, Boca Raton, FL, 2004.

[21] B. Budowle, R. Chakraborty, G. Carmody, K.L. Monson, Reply to weir, Forensic Sci. Commun. 3 (2001), http://www.fbi.gov/programs/lab/fsc.

[22] B.S. Weir, DNA match and profile probabilities: comment on Budowle et al. (2000) and Fung and Hu (2000), Forensic Sci. Commun. 3 (2001), http://www.fbi.gov/programs/lab/fsc (with replies).

[23] P. Koziol, M. Ciesielka, R. Madro, A. Krajka, Genetic data on 19 STR in south-east Poland, Forensic Sci. Int. 139 (2004) 89–92.