Int J Legal Med (2001) 114 : 147–155

ORIGINAL ARTICLE

L. A. Foreman · I. W. Evett

Statistical analyses to support forensic interpretation for a new ten-locus STR profiling system

Received: 8 July 1999 / Accepted: 14 January 2000

Abstract A new ten-locus STR (short tandem repeat) profiling system was recently introduced into casework by the Forensic Science Service (FSS) and statistical analyses are described here based on data collected using this new system for the three major racial groups of the UK: Caucasian, Afro-Caribbean and Asian (of Indo-Pakistani descent). Allele distributions are compared and the FSS position with regard to routine significance testing of DNA frequency databases is discussed. An investigation of match probability calculations is carried out and the consequent analyses are shown to provide support for proposed changes in how the FSS reports DNA results when very small match probabilities are involved.

Key words Short tandem repeat (STR) loci · UK population data · Statistics · Significance testing · Allele distributions · Match probability · Bayes' theorem · Uniqueness

Introduction

The PCR-based STR profiling system used currently in forensic casework at the FSS and which is the basis for the UK National Criminal Intelligence DNA Database (Werrett 1997), the so-called SGM (second generation multiplex), consists of the amelogenin sex test plus the six loci (Sparkes et al. 1996) HUMTHO1 (THO1), HUMVWFA31/A (VWA), D18S51 (D18), D21S11 (D21), D8S1179 (D8) and HUMFIBRA (FGA). In June 1999, a new ten-locus profiling system, known as SGM-plus, was introduced to replace SGM. This types samples at all six SGM loci plus a further four STR loci, D16S539 (D16), D2S1338 (D2), D3S1358 (D3) and D19S433 (D19). The loci D2 and D3 are already used by the FSS as part of the TGM (third generation multiplex) profiling system (Watson et al. 2000) and in conjunction with SGM, this is the basis for profiling samples in paternity casework. The statistical analyses described here were carried out on data sets of SGM-plus profiles compiled by the FSS from individuals in three UK racial groups: Caucasian, Afro-Caribbean, Asian (of Indo-Pakistani descent).

When a match has been obtained between the DNA profile of a defendant, *s*, say, and that of a crime scene sample, it is now standard practice to report the weight of the DNA evidence in terms of a match probability. Given that *s* is not the source of the crime stain, this represents the probability that another individual in the relevant population would share the matching profile. Until now, much of the debate surrounding the use of DNA profiling in forensic science has been concerned with the method of calculating match probabilities and, more specifically, verification of the underlying allele independence assumptions made. For many years, the so-called product rule was adopted whereby component allele proportions are multiplied together within and across loci to give an estimate of the proportion for the complete multi-locus matching profile. It then became widely accepted that statistical significance testing of the underlying independence assumptions was a pre-requisite for implementation of any new profiling system.

More recently, concerns were raised (Balding and Nichols 1994) that use of the product rule could overstate the strength of the DNA evidence by ignoring within-locus correlation arising from the presence of substructuring in general populations. Consequently, the aforementioned authors derived a match probability formula to take account of this by means of a parameter that measures population differentiation/substructure (often denoted as θ or F_{ST}). Although many forensic organisations (including the FSS) currently use this formula as a basis for their calculations, statistical validation of new profiling systems is still focussed on carrying out a series of within- and between-locus significance tests.

The limitations of classical significance testing are well-documented e.g. with respect to DNA profiling (Evett and

L. A. Foreman (✉) · I. W. Evett
The Forensic Science Service, Metropolitan Laboratory,
109 Lambeth Road, London, SE1 7LP, UK
e-mail: LAForeman@compuserve.com; Fax: +44-171-2306432

Buckleton 1996) and medical applications (Matthews 1998). Results can be very sensitive to database size, outliers and the arbitrary nature of the significance level. Furthermore, it is well-known that the idealised conditions necessary for the allele independence assumptions to hold exactly are never satisfied in practice and that population substructure is the major factor responsible for departures from independence in forensic profiles. However, there is now ample evidence that these substructuring effects are minor at the STR loci used for forensic identification, i.e. F_{ST} estimates are small (e.g. Gill and Evett 1995; Foreman et al. 1998). Recognition of all these issues culminates in the paradoxical situation where authors describe the results of significance testing on DNA profiling data in order to justify the independence assumptions underlying the forensic calculations to be used in practice, only to ignore or "explain away" any significant results which are obtained as being of no practical consequence (e.g. Evett et al. 1996).

As part of the statistical validation of the current SGM system (Evett et al. 1997), independence tests were carried out on the three main FSS frequency databases, although the authors made clear their reservations about the relevance of such testing. However, with the introduction of the new SGM-plus system, it is our view that routine independence testing of DNA frequency databases should be abandoned; our reasons can be summarised as follows:

1. Testing within loci is irrelevant since within-locus independence is not assumed in the match probability calculations reported in FSS casework practice. The match probability formula due to Balding and Nichols (1994) is routinely adopted with an appropriate value of F_{ST}. Previous FSS guidelines recommended using values of 3% with Caucasian and Afro-Caribbean data, 5% with Asian data and 0% when considering alternative suspects belonging to a different racial group from the defendant. These "cautious" values were originally adopted following recommendations made by the US National Research Council (1996) at a time when PCR-based systems were relatively new. However, there is now an extensive body of literature describing analyses of STR data which support much lower F_{ST} values. In particular, analyses relevant to the three main FSS SGM frequency databases are described in Foreman et al. (1998), Foreman (1999) and Foreman and Lambert (2000) and these suggest a value of F_{ST} = 0.02 as being a reasonable "upper bound" for use with the Caucasian, Afro-Caribbean and Asian (Indo-Pakistani) populations of the UK. Until more comprehensive data become available at the new SGM-plus loci, it seems reasonable to extrapolate the results obtained for the SGM loci; hence, new FSS guidelines recommend using F_{ST} = 2% with all three frequency databases.

2. To obtain match probabilities for complete profiles, single-locus probabilities from Balding and Nichols' formula are multiplied across loci. Applying a significance testing approach to verification of the between-locus independence assumptions would involve carrying out a total of 1013 tests per SGM-plus database, covering all possible combinations of 2–10 loci. This is clearly unfeasible and, hence, between-locus testing of STR systems consisting of large numbers of loci is often restricted to 2-locus combinations (e.g. Hammond et al. 1994; Watson et al. 2000). Furthermore, testing on such a large scale is bound to yield a substantial number of significant results due to chance alone and Bonferroni-type adjustments to the *p*-values only serve to reduce the power of each multilocus test. However, since all the SGM-plus loci lie on distinct chromosomes in the genome, any departures from between-locus independence assumptions will be due to population substructure. Adopting larger values of F_{ST} than necessary within loci should more than compensate for any small dependencies which might exist across loci – these are likely to be of a smaller magnitude than those indicated within loci (Evett and Weir 1998).

3. With particular reference to the STR loci used in the SGM-plus system, results of within- and between-locus independence testing have already been reported: SGM loci (Evett et al. 1997), SGM loci plus D3 and D16 (Ga-

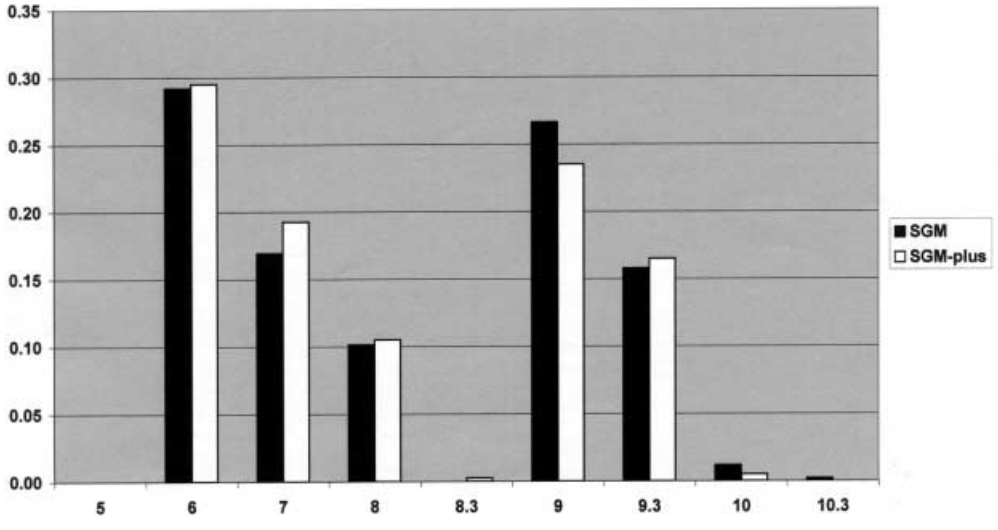

Fig. 1 Allele distributions at THO1 in the UK Asian population estimated from the FSS SGM (n = 257) and SGM-plus (n = 200) databases

rafano et al. 1998), SGM and TGM loci (Watson et al. 2000). These papers add to the existing body of literature which, to date, has found little evidence refuting the validity of multiplying match probabilities across loci.

However, despite our position with regard to the testing of independence assumptions, we recognise that there may still be merit in the use of within-locus significance tests as part of the quality control procedure when compiling databases; e.g. identification of anomalies during the typing process, detection of null alleles.

Materials and methods

Full details of the methods used in SGM-plus profiling are provided by Cotton et al. (2000) and rules for allelic designation are given in Gill et al. (1996).

The SGM-plus frequency databases analysed here were compiled from individuals belonging to the three main UK racial groups encountered in FSS casework (Cotton et al. 2000). A proportion of the SGM-plus samples are also present on the SGM frequency databases analysed in Evett et al. (1997), as indicated below:

1. 437 Caucasian samples from FSS staff (includes 28 SGM database samples)

2. 164 Afro-Caribbean samples from FSS staff and casework (includes 131 SGM database samples)

3. 200 Asian (Indo-Pakistani) samples from FSS and police staff, from patients at hospitals in Birmingham and Oxford and from immigration paternity testing (includes 130 SGM database samples)

The AmpF*l*STR© SGM Plus kit manual (Perkin-Elmer Corporation 1999, Foster City, Calif.) describes population statistics generated at each of the SGM-plus loci for two United States databases provided by Laboratory Corporation of America as follows:

1. 200 US Caucasian samples

2. 195 African-American samples

Inclusion of these data in our analyses facilitated within-racial group comparisons of allele distributions and match probabilities when considered alongside the FSS databases.

Allele distributions and discriminating power

Allele proportions estimated at the six SGM loci from the original FSS SGM frequency databases are tabulated in Evett et al. (1997). As would be expected, these agree closely with the corresponding estimates obtained using the SGM-plus databases; e.g. see Fig. 1 for a comparison of allele distributions at THO1 in the Asian population. Note that the nomenclature of alleles at locus D21 under the SGM-plus profiling system is different from that used with SGM; Gill et al. (1997) provide the conversion formula. Hence, Table 1 only gives allele proportions estimated for the four additional loci. For D2 and D3, close agreement is again obtained with separate estimates reported in Watson et al. (2000). As observed previously in analyses of SGM data (Foreman et al. 1998; Foreman and Lambert 2000), allele distributions exhibited at the new loci are very similar within racial groups (see Figs. 2, 3). The comparisons at D3 given in Fig. 2 include estimates

Table 1 Allele proportions estimated at the non-SGM loci from the three FSS frequency databases: (A) Caucasian (n = 437); (B) Afro-Caribbean (n = 164); (C) Asian, of Indo-Pakistani descent (n = 200)

Allele	Racial group		
	A	B	C
D16			
5	0.000	0.003	0.000
8	0.019	0.015	0.058
9	0.129	0.189	0.193
10	0.054	0.119	0.115
11	0.289	0.348	0.293
12	0.288	0.223	0.205
13	0.186	0.082	0.120
14	0.029	0.021	0.018
15	0.005	0.000	0.000
D2			
16	0.037	0.040	0.010
17	0.185	0.146	0.100
18	0.087	0.055	0.120
19	0.110	0.159	0.145
20	0.138	0.073	0.138
21	0.032	0.113	0.050
22	0.024	0.134	0.060
23	0.112	0.110	0.155
24	0.142	0.067	0.103
25	0.111	0.082	0.108
26	0.019	0.021	0.005
27	0.002	0.000	0.005
28	0.000	0.000	0.003
D3			
12	0.001	0.000	0.000
13	0.006	0.003	0.003
14	0.132	0.079	0.060
15	0.265	0.268	0.275
16	0.247	0.351	0.313
17	0.195	0.232	0.220
18	0.141	0.064	0.118
19	0.014	0.003	0.013
D19			
10	0.000	0.006	0.003
10.2	0.000	0.003	0.000
11	0.000	0.073	0.000
12	0.087	0.113	0.055
12.2	0.000	0.034	0.010
13	0.222	0.274	0.260
13.2	0.013	0.040	0.020
14	0.382	0.253	0.253
14.2	0.015	0.034	0.068
15	0.177	0.076	0.173
15.2	0.038	0.049	0.070
16	0.041	0.009	0.053
16.2	0.017	0.027	0.023
17	0.005	0.000	0.013
17.2	0.000	0.006	0.003
18	0.000	0.003	0.000
18.2	0.002	0.000	0.000
19.2	0.001	0.000	0.000

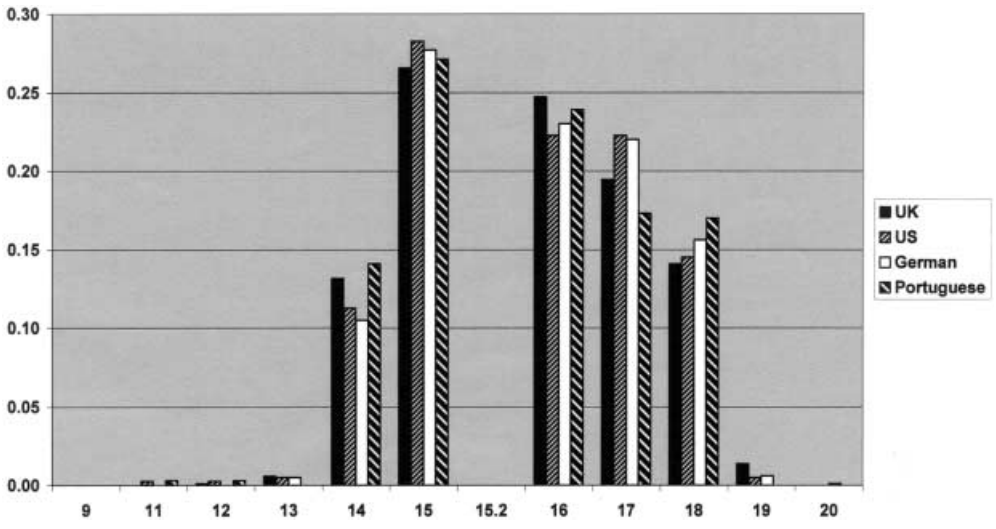

Fig. 2 Allele distributions at D3 exhibited among Caucasians in the UK (n = 437), US (n = 200), SW German (n = 499) and Portuguese (n = 153) populations

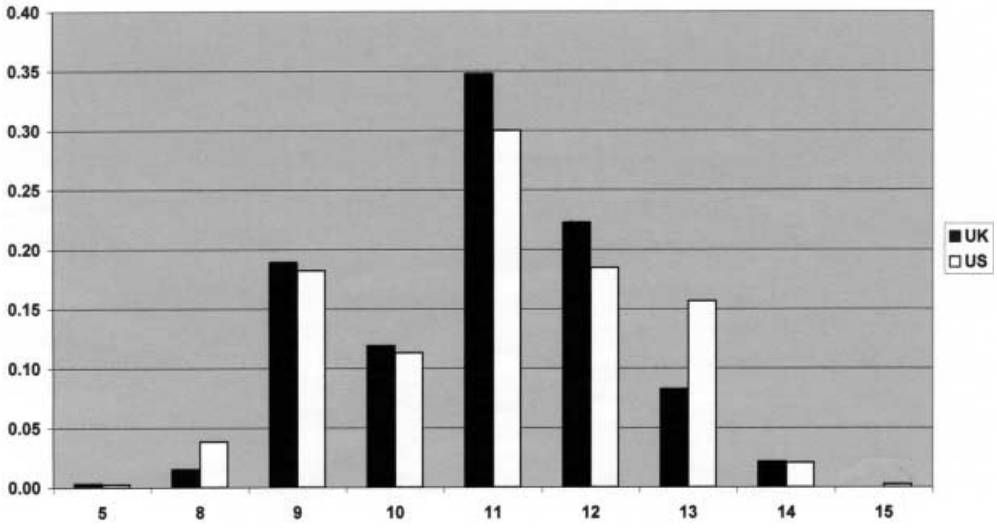

Fig. 3 Allele distributions at D16 for the UK Afro-Caribbean population (n = 164) and the African-American population (n = 195)

provided in Momhinweg et al. (1998) for a German and a Portuguese population.

The probability of obtaining a match between two distinct and unrelated individuals (PM) provides a measure of the discriminating power of the profiling system. Probabilities are calculated at each locus based on the number of matches observed when each profile in the database is compared with every other profile. These are then multiplied to obtain a figure for complete profiles. Table 2 gives PM values calculated from the SGM-plus data. Figures obtained at the SGM loci are virtually indistinguishable from those reported in Table 2 of Evett et al. (1997) and comparison of both tables reveals the increased discriminating power attainable under the SGM-plus profiling system; i.e. PM values are of the order 1 in 100 million for SGM as compared with 1 in 10,000 billion for SGM-plus.

It should be clear that the magnitude of match probabilities calculated from SGM-plus profiles will be very small and are typically of the order of one in thousands of billions for unrelated people. Figures such as these are un-

Table 2 Probability of a match between two unrelated people (PM) estimated at each locus from database between-person comparisons

	D16	D2	D3	VWA	D18	D21	D8	D19	FGA	THO1	Combined
Caucasian	0.086	0.027	0.073	0.061	0.028	0.050	0.061	0.088	0.030	0.083	1.90×10^{-13}
Afro-Caribbean	0.078	0.024	0.112	0.047	0.023	0.041	0.066	0.043	0.027	0.098	6.91×10^{-14}
Asian	0.067	0.024	0.101	0.065	0.034	0.041	0.043	0.050	0.032	0.087	8.87×10^{-14}
US Caucasian	0.103	0.024	0.078	0.065	0.030	0.045	0.067	0.078	0.036	0.094	2.99×10^{-13}
African-American	0.066	0.021	0.102	0.058	0.028	0.033	0.075	0.039	0.035	0.102	7.91×10^{-14}

likely to convey much to a jury if presented in isolation without providing assistance in their interpretation. In the remainder of the paper, we focus on determining match probabilities for various categories of alternative suspect which might be considered and how these can be combined with other, non-DNA evidence in the case to aid the jury in its task of assessing whether or not the defendant is the source of the crime stain. The analysis carried out here for the SGM-plus data follows closely the approach to reporting DNA evidence which is detailed in Balding (1999).

Match probability calculations

When considering possible sources of the crime stain DNA other than the defendant, *s*, we can calculate match probabilities for a variety of specified alternatives corresponding to individuals exhibiting different degrees of relatedness to *s*. Currently, figures reported in the main body of FSS statements refer to people unrelated to the defendant, although mention is made of the fact that blood relatives have a greater chance of matching and that separate match probabilities can be evaluated if necessary. Probabilities for unrelated individuals when there is a full SGM profile match are typically of the order of 1 in tens of millions. However, we shall see that adopting a similar approach with the SGM-plus system yields values many orders of magnitude smaller than this.

Adopting the notation used in Balding (1999), let *P* denote the population containing all possible sources of the crime stain excluding *s*. We assume that *P* can be split into a number of disjoint categories within which the match probability takes a fixed value. For illustration, we consider six categories corresponding to individuals exhibiting different degrees of relatedness to the defendant, *s*:

1. Sibling
2. Parent/child
3. Half-sibling or uncle/nephew
4. First cousin
5. Unrelated (subpopulation)
6. Unrelated (population)

A value of F_{ST} = 2% is used with categories 1–5 and F_{ST} = 0% with category 6. The category for unrelated individuals has been divided into members of the defendant's "subpopulation" (category 5) plus the remainder of the population (category 6). Match probabilities in each category are calculated using the formula given in Balding and Nichols (1994).

Theoretically, we can determine the most common SGM-plus profile in each population from the database allele proportions; in each case, this happens to be a completely heterozygous profile made up of the two most common alleles at each locus. Match probabilities obtained for the commonest profile exhibited in each of the three UK racial groups are given in Table 3. Two sets of values are shown; one set corresponds to the case where allele proportions are estimated from the raw database counts while the second set is based on allele counts adjusted to take account of sampling error using the size-bias correction (Balding and Nichols 1994). There is very close agreement between values obtained using both the raw and adjusted allele proportions; this is as one might expect since the sampling adjustment has greatest effect when applied to rare alleles. Furthermore, it can be seen that figures within each category are very similar across racial groups; i.e. the crucial factor determining the order of magnitude of figures (match probabilities, PMs, etc.) is the number of loci profiled. Thus, our dependence on databases can be seen to be diminishing since calculations are essentially invariant to racial group. To emphasise this point further, the most common profile was identified for two further UK populations, Arabic and Oriental, using SGM data analysed in Foreman and Lambert (2000). The match probability for unrelated individuals is of the order 1 in a million using SGM data from any of the five populations considered: Caucasian, Afro-Caribbean, Asian (Indo-Pakistani), Arabic, Oriental. Hence, even in cases where there is good information that the source of the crime stain originates from a population for which no data are available, the general results discussed here should be seen to apply. We note further that SGM-plus figures for the US Caucasian and African-American data corresponded closely to those obtained from the FSS Caucasian and Afro-Caribbean databases, respectively.

In order to gain some idea of the range of match probabilities we might expect to see with the SGM-plus system, simulated data sets of 10,000 profiles were generated from the database allele proportions, independently within and across loci. For each profile, match probabilities were calculated using adjusted allele proportions. Table 3 shows the range of values exhibited for each category of relatedness in the three UK racial groups considered. Maximum match probabilities arising from the simulated profiles can be seen to be of the same order of magnitude as those obtained for the commonest profiles. Comparing median (50% quantile) values across racial groups, we again see close agreement and these suggest that "typical" match probabilities which might be expected from applying current calculation methods to SGM-plus profiles range from approximately 1 in 30,000 for siblings to 1 in 50 trillion for unrelated individuals (a trillion is defined as 1 million million).

One of the questions which is often asked in court is "how can you give such a number (match probability) when your database contains only a few hundred individuals?". In fact, the size of the database only determines the precision with which we can estimate the population proportions of the component alleles which make up the profile – these are typically in the range of 5% to 20% for the STR loci used in forensic identification. The small match probability comes from combining the individual allele proportions using the established methods described earlier; i.e. adopting Balding and Nichols' match probability formula at each individual locus and making an extremely generous allowance (F_{ST}) for possible structuring effects in the population. The match probabilities for each of the loci are then multiplied together. This process of

Table 3 Match probabilities obtained in each UK racial group for alternative suspects exhibiting various degrees of relatedness to the defendant, *s*: *1* sibling, *2* parent/child, *3* half-sibling or uncle/nephew, *4* first cousin, *5* unrelated (subpopulation), *6* unrelated (population)

	Category of relatedness to *s*					
	(1)	(2)	(3)	(4)	(5)	(6)
Caucasian						
Commonest profile						
Raw	1.1×10^{-4}	7.4×10^{-7}	3.5×10^{-8}	4.8×10^{-9}	3.8×10^{-10}	1.8×10^{-10}
Adjusted	1.2×10^{-4}	7.7×10^{-7}	3.7×10^{-8}	5.2×10^{-9}	4.2×10^{-10}	2.0×10^{-10}
Simulations						
Maximum	8.5×10^{-5}	3.5×10^{-7}	1.2×10^{-8}	1.4×10^{-9}	8.9×10^{-11}	3.1×10^{-11}
(5%, 50% 95%) quantiles	(1.9, 3.1, 5.0) $\times 10^{-5}$	(0.26, 1.7, 7.8) $\times 10^{-8}$	(0.41, 3.5, 22) $\times 10^{-10}$	(0.17, 2.2, 19) $\times 10^{-11}$	(0.073, 3.1, 64) $\times 10^{-13}$	(0.017, 2.6, 120) $\times 10^{-14}$
Afro-Caribbean						
Commonest profile						
Raw	1.1×10^{-4}	6.7×10^{-7}	3.1×10^{-8}	4.1×10^{-9}	3.1×10^{-10}	1.4×10^{-10}
Adjusted	1.2×10^{-4}	7.7×10^{-7}	3.7×10^{-8}	5.1×10^{-9}	4.0×10^{-10}	1.9×10^{-10}
Simulations						
Maximum	7.7×10^{-5}	2.5×10^{-7}	9.4×10^{-9}	1.2×10^{-9}	7.8×10^{-11}	2.4×10^{-11}
(5%, 50% 95%) quantiles	(1.8, 3.0, 4.8) $\times 10^{-5}$	(0.19, 1.4, 6.4) $\times 10^{-8}$	(0.29, 2.7, 17) $\times 10^{-10}$	(0.11, 1.5, 14) $\times 10^{-11}$	(0.038, 1.8, 40) $\times 10^{-13}$	(0.0083, 1.4, 67) $\times 10^{-14}$
Asian						
Commonest profile						
Raw	9.3×10^{-5}	4.7×10^{-7}	2.0×10^{-8}	2.5×10^{-9}	1.7×10^{-10}	7.7×10^{-11}
Adjusted	9.8×10^{-5}	5.3×10^{-7}	2.3×10^{-8}	3.0×10^{-9}	2.2×10^{-10}	1.0×10^{-10}
Simulations						
Maximum	7.7×10^{-5}	2.9×10^{-7}	1.0×10^{-8}	1.1×10^{-9}	5.6×10^{-11}	2.1×10^{-11}
(5%, 50% 95%) quantiles	(1.8, 2.9, 4.4) $\times 10^{-5}$	(0.26, 1.4, 6.0) $\times 10^{-8}$	(0.40, 2.9, 15) $\times 10^{-10}$	(0.17, 1.8, 13) $\times 10^{-11}$	(0.087, 2.5, 42) $\times 10^{-13}$	(0.026, 2.3, 80) $\times 10^{-14}$

multiplying across loci is widely accepted and it appears to be universal practice in the DNA profiling field. However, it necessarily invokes independence assumptions, as we discussed earlier. Over the years, various experiments have been carried out which have been based on millions of between-person comparisons (e.g. Lambert et al. 1995; Risch and Devlin 1992). These show that the independence assumptions are sufficiently reliable to infer probabilities that are of the order of 1 in tens of millions, as is typical with SGM profiles. However, Table 3 shows the tiny probabilities that could be calculated for 10-locus SGM-plus matches (of the order 1 in trillions for unrelated people). These invoke independence assumptions to a scale of robustness which we could not begin to investigate by statistical experiment; this is not to say that the figures are "wrong", rather they are without meaning. The Starr report (www.zdnet.com/yil/starrreport/7grounds.htm) on the recent legal proceedings involving the US President provides a good (or, should we say, bad) example of many of the issues which surround the proper representation of very powerful DNA evidence. After giving a conclusive opinion as to the source of the semen stain (the President), the report goes on to make the further (superfluous) statement:

The chance that the semen is not the President's is one in 7.87 trillion.

Apart from committing the Prosecutor's Fallacy (Balding and Donnelly 1994), we believe that quoting a match probability of the order 1 in a trillion in this way, to such a high degree of precision, has no scientific justification.

For these reasons, we believe that case-specific match probabilities should not be calculated as a matter of principle. Instead, we advocate use of the general figures given in Table 4 when reporting results for full SGM-plus profiles. The analyses described in this section provide a strong justification for adopting this approach, not least because the match probabilities which would arise from doing the calculation (see values for simulated profiles in Table 3) would typically be several orders of magnitude smaller than

Table 4 General match probability values recommended for use when reporting full SGM-plus profile matches

Relationship with *s*		Match probability
(1)	Sibling	1 in 10,000
(2)	Parent/child	1 in 1 million
(3)	Half-sibling or uncle/nephew	1 in 10 million
(4)	First cousin	1 in 100 million
(5) & (6)	Unrelated	1 in a billion

Table 5 Values of Pr(C|E) and the lower bound on Pr(U|E) for various alternative suspect populations, *P*, defined according to the number of individuals, N_i (i = 1,2,…,6), belonging to each of the following categories of relatedness to *s*: *1* Sibling, *2* parent/child, *3* half-sibling or uncle/nephew, *4* first cousin, *5* unrelated (subpopulation), *6* unrelated (population)

	Population category							
	(1)	(2)	(3)	(4)	(5)	(6)	Pr(C|E)	Pr(U|E)
Population numbers, N_i								
(a)	5	0	20	100	10,000,000	0	0.9895	0.9789
(b)	5	5	20	100	100,000	13,000,000	0.9865	0.9726
(c)	0	1	20	100	100,000	13,000,000	0.9871	0.9738
(d)	1	1	15	50	100,000	13,000,000	0.9870	0.9736
(e)	5	1	20	50	10,000	4,000,000	0.9954	0.9908
(f)	5	1	20	100	16,000,000	0	0.9837	0.9669
(g)	0	0	10	20	13,000,000	0	0.9872	0.9740
(h)	5	1	20	100	1,000	500,000	0.9989	0.9978
(i)	0	1	20	100	1,000	500,000	0.9995	0.9990

those given in Table 4 (which have been determined from the theoretically most common SGM-plus profile).

Posterior probabilities and uniqueness

Balding (1999) discussed two different approaches to reporting and explaining DNA evidence in court. The first approach involves presenting (the lower bound on) a probability of uniqueness which represents the chance that no member of *P* shares the crime scene profile. The second approach, which is our preferred method, implements Bayes' theorem to address the more pertinent issue of whether or not the defendant is the source of the crime stain by considering the posterior probability of this event. Both probabilities contain the same information although expressed in a slightly different way.

Let *C* denote the event that the defendant *s* is the source of the crime stain and let *E* denote all the evidence presented in the case. Then Balding and Donnelly (1995) show that the posterior odds in favour of *C* can be expressed as

$$O(C|E) = \frac{1}{\sum_{x \in P} R_s(x) w_s(x)} \quad (1)$$

where $R_s(x)$ denotes the match probability for alternative suspect x and $w_s(x)$ represents the weight of the non-DNA evidence against *x* relative to its weight against *s*. For the sake of illustration, we focus here on considering the initial non-DNA evidence available to the jury in relation to the geography of the crime and the defendant's whereabouts, say, and how this interacts with the DNA evidence presented by the scientist; i.e. in this case, *E* represents the DNA evidence plus prior information relating to the population of all possible sources of the crime stain, $P \cup \{s\}$. Other non-DNA evidence can be considered separately by the jury. If the jury considers that the defendant is as likely to have left the crime stain as any member of the specified population *P*, then the prior weights, $w_s(x)$, are all equal to 1. Combining this with the evidence that the defendant's DNA matches that of the crime stain gives

1. Posterior probability of *C*, $\Pr(C|E) = 1/(1+R)$
2. Probability of uniqueness, $\Pr(U|E) > 1–2R$ (Balding)

where *U* denotes the event that the matching profile is unique in $P \cup \{s\}$ and

$$R = \sum_{x \in P} R_s(x). \quad (2)$$

In particular, if we choose to decompose *P* into the six categories of relatedness defined earlier, then *R* simply reduces to a sum of products, $N_i \times P_i$, i = 1,2,....,6, where N_i and P_i denote the population size and match probability for category *i*. Values for the P_i are provided by the scientist (in accordance with the 1997 Appeal Court ruling in R. v. Doheny and G. Adams); we recommend the match probabilities given in Table 4. Appropriate figures (or a range of illustrative figures) for the N_i could be agreed upon by prosecution and defence counsel prior to the trial, based on information specific to the case and using general population figures available from the last census or more up-to-date figures provided by the Government Statistical Service (GSS).

We now give an illustration of the probability calculations described above and show how these vary with different specifications for the population sizes, N_i. For the P_i, we use the match probabilities in Table 4.

Table 5 gives posterior probabilities and lower-bound probabilities of uniqueness for a series of alternative suspect populations, *P*. Case (a) corresponds to the population *P* considered for illustration in Balding (1999). In cases (b)–(g), numbers, N_i, used in categories (5) & (6) are based on up-to-date population statistics (May 1999) summarised on the GSS website, Britain Update (www.statistics.gov.uk/news/brup1/.htm). For example, there are approximately 16 million white males between the ages of 16 and retirement age (59/64 years) in the whole of the UK, 13 million in England alone, 4 million of whom are resident in London and the South East. Cases (h) & (i) use population numbers for categories (5) & (6) estimated from the 1991 census for the Asian (Indo-Pakistani) population of the UK; from a total of approximately 1.5 million, about 500,000 will be male between the ages of 16 and 64 years (extrapolating from the GSS figures).

In cases (a)–(g), where N_5 or N_6 is large (≥10 million), the figures corresponding to individuals unrelated to *s* make the greatest contribution to the sum *R*. This can be seen by

comparing values of Pr(C|E) in cases (b)–(d), where only the population numbers for blood relatives are varied – P*r*(C|E) is the same (98.7%) in each case. Similarly, Pr(C|E) is 98.7% in case (g). The values obtained for Pr(C|E) in cases (e) & (f) illustrate the effect of considering larger and smaller population sizes for the unrelated category; i.e. the defendant is more likely to have left the crime stain when the alternative suspect population is small. This point is further emphasised by comparison of the figures obtained in cases (h) & (i), where N_6 is at least an order of magnitude smaller than considered in the earlier cases. The sum *R* in this case is dominated equally by figures for both siblings and people unrelated to *s*. As noted in Balding (1999), it can be seen that the contribution to *R* (and, hence, to Pr(C|E) and Pr(U|E)) from blood relatives of *s* other than siblings is negligible in all cases.

For a full SGM-plus profile, the sex of the individual will always be available from the amelogenin result and this information can be used in determining *P*. In most cases, considering the entire UK or English population as possible sources of the crime stain will be highly unrealistic, even when this is reduced by considering information regarding the sex and age range of alternative suspects, as was illustrated in the analyses described above. Therefore, we believe the populations *P* specified in Table 5 represent the largest which might be considered relevant in UK casework and, thus, the resulting probabilities, Pr(C|E) and Pr(U|E), can be thought of as a lower bound on those values which might be expected to arise from the use of more realistic population sizes. Furthermore, the match probabilities used to represent the weight of DNA evidence are among the largest that would be obtained from adopting the calculation methods currently implemented with SGM.

We recall from Balding (1999) that the lower bound on P*r*(U|E) only applies when the non-DNA evidence considered in *E* does not favour the defendant *s*. In contrast, the posterior probability Pr(C|E) can be calculated based on any population *P* considered relevant in a case, where the specification of *P* might include case-specific information which supports the defence position. Furthermore, using posterior probabilities to assist the jury in assessing the DNA evidence can be simply applied to just one population category if necessary; e.g. interpreting the match probability of 1 in a billion for unrelated people in the context of a specific case.

Discussion

With the introduction of new STR profiling systems consisting of an ever-increasing number of loci, there is a need to review how DNA evidence is reported and presented in court. Evett et al. (2000) discuss some of the issues involved and propose a new reporting policy in such cases. Results from the statistical analyses described here provide a justification for adopting this policy where full SGM-plus profiles are concerned.

When a DNA match has been observed between defendant and crime stain, the weight of the DNA evidence is either presented as a population proportion for the matching profile using the product rule (this remains the standard approach in the US; e.g. National Research Council 1996) or as a match probability using the established formula due to Balding and Nichols (1994). Both methods of calculation rely on between-locus independence assumptions which cannot be verified by experiment for systems involving large numbers of loci. We have outlined the main reasons why implementing a significance testing approach is unhelpful in this regard and, hence, our position on routine independence testing of frequency databases; i.e. that this practice should be abandoned.

A detailed investigation of match probability calculations was carried out which shows that applying current methods to full SGM-plus profiles would typically result in values of the order 1 in trillions for unrelated people. As already stated, it would be difficult to provide any sound statistical support for probabilities of such a small magnitude and, therefore, a more robust approach which avoids calculation of case-specific figures is appropriate. Instead, we recommend reporting the match probabilities given in Table 4 in all cases. These should be sufficient to convey the impression that full SGM-plus profiles are extremely rare; moreover, the probabilities are not many orders of magnitude different from the sort of numbers UK courts are used to dealing with under the SGM system. Despite the concerns raised in relation to invoking the between-locus independence assumptions in general, it is our opinion, based on the analyses and discussion put forward in this paper and elsewhere (e.g. Balding 1999), that the probabilities in Table 4 provide a fair and reasonable assessment of the weight of DNA evidence for each category and are not unfavourable to the defendant *s*.

Giving the same DNA statement, whatever the racial group of the source of the crime stain, has a simplistic appeal and the SGM-plus analyses fully support this; i.e. the recommended match probabilities in Table 4 are not based on any particular database and, thus, the composition of the frequency databases and their size would cease to be the central issues when reporting DNA evidence. Furthermore, the scientist would not be required to determine which F_{ST} values and databases were relevant in each individual case.

Given that the scientist has reported a DNA match probability of the order 1 in a billion for people unrelated to the defendant, say, it is likely that it would be expected that the jury be provided with some assistance in its interpretation. Reporting a (lower-bound) probability of uniqueness (Balding 1999) might appear to follow the recommendations of the 1997 Appeal Court ruling in R. v. Doheny and G. Adams more closely than presenting a posterior probability, assuming that a population of alternative suspects has been suggested by the defence. However, the uniqueness probability lower bound only applies when there is no non-DNA evidence in favour of the defendant and does not address the issue of most relevance to the court.

Alternatively, the scientist can use Bayes' theorem to illustrate what impact the DNA evidence might have on the jury's assessment of whether or not the defendant is

the source of the crime stain; there is considerable support for this being the most coherent approach among forensic practitioners (e.g. Kaye 1993). Non-DNA evidence in a case and general population information can be used to establish an appropriate population, *P*, of alternative suspects who are considered as likely as the defendant to have left the crime stain. This defines the prior probability; i.e. Pr(C) = 1/N, where N denotes the size of *P*. Even for the most unrealistic of suspect populations *P* considered in our illustrative analyses, where N is 16 million, the posterior probability Pr(C|E)>98%. This probability can be updated further using Bayes' theorem given additional non-DNA evidence (including any which favours the defence) not already considered in the specification of *P*.

In this paper, we have given particular attention to how the statistical analyses address issues relating to the reporting of full SGM-plus profiles in simple cases; i.e. when the defendant matches a crime stain. Although the technical improvements which come with the SGM-plus profiling system should yield full profiles from a greater proportion of samples than SGM, experience with SGM suggests that approximately 20% of cases will result in a partial profile. The approach described above for full profiles can easily be modified to deal with partial profiles if considered necessary. However, it should be stressed that analyses based on the commonest SGM-plus profile may not be appropriate in more complex cases; e.g. mixtures, parentage analyses, missing persons. This is because, in general, the weight of DNA evidence in such cases is reported in terms of a likelihood ratio, which requires consideration of both a numerator and denominator.

Likelihood ratios resulting from parentage and missing person analyses are typically many orders of magnitude smaller than those obtained in simple "suspect matching crime stain" cases; hence they are unlikely to give rise to the same concerns that have been addressed in this paper. In contrast, although the approach discussed here for full SGM-plus profiles would still apply in straightforward mixtures cases where the likelihood ratio reduces to the inverse of one or more match probabilities, it may not be appropriate in general. The increased sensitivity of the SGM-plus profiling technique means that we might expect to see a greater proportion of cases involving mixtures in practice; the issues involved in their interpretation will be addressed in a future paper.

Acknowledgements The authors would like to thank Jim Lambert, Mike Barber and Peter Gill for their helpful comments on an earlier draft of the paper.

References

Balding DJ (1999) When can a DNA profile be regarded as unique? Sci Justice 39: 257–260

Balding DJ, Donnelly P (1994) The prosecutor's fallacy and DNA evidence. Crim Law Rev October: 711–721

Balding DJ, Donnelly P (1995) Inference in forensic identification (with discussion). J R Statist Soc A 158: 21–53

Balding DJ, Nichols RA (1994) DNA profile match probability calculations: how to allow for population stratification, relatedness, database selection and single bands. Forensic Sci Int 64:125–140

Cotton EA, Allsop RF, Guest JL, Callow IP, Seager A, Sparkes RL (2000) Validation of the AmpF*l*STR SGM Plus system for use in forensic casework. Forensic Sci Int 112: 151–161

Evett IW, Buckleton JS (1996) Statistical analysis of STR data. In: Carracedo A, Brinkmann B, Bär W (eds) Advances in haemogenetics 6. Springer, Berlin Heidelberg New York, pp 79–86

Evett IW, Weir BS (1998) Interpreting DNA evidence: statistical genetics for forensic scientists. Sinauer, Sunderland Mass.

Evett IW, Gill PD, Scranage JK, Weir BS (1996) Establishing the robustness of short-tandem-repeat statistics for forensic applications. Am J Hum Genet 58: 398–407

Evett IW, Gill PD, Lambert JA, Oldroyd N, Frazier R, Watson S, Panchal S, Connolly A, Kimpton C (1997) Statistical analysis of data for three British racial groups from a new STR multiplex. Int J Legal Med 110: 5–9

Evett IW, Foreman LA, Jackson G, Lambert JA (2000) DNA profiling: A discussion of issues relating to the reporting of very small match probabilities. Crim Law Rev May, pp 341–355

Foreman LA (1999) Analyses to investigate appropriate measures of differentiation between European Caucasian populations using STR data. FSS Research Report FSS-RR-804

Foreman LA, Lambert JA (2000) Genetic differentiation within and between four UK ethnic groups. Forensic Sci Int 114: 7–20

Foreman LA, Lambert JA, Evett IW (1998) Regional genetic variation in Caucasians. Forensic Sci Int 95: 27–37

Garafano L, Pizzamiglio M, Vecchio C, Lago G, Floris T, D'Errico G, Brembilla G, Romano A, Budowle B (1998) Italian population data on thirteen short tandem repeat loci: HUMTHO1, D21S11, D18S51, HUMVWFA31, HUMFIBRA, D8S1179, HUMTPOX, HUMCSF1PO, D16S539, D7S820, D13S317, D5S818, D3S1358. Forensic Sci Int 97: 53–60

Gill P, Evett IW (1995) Population genetics of short tandem repeat (STR) loci. Genetica 96: 69–87

Gill P, Urquhart A, Millican ES, Oldroyd N, Watson S, Sparkes R, Kimpton CP (1996) A new method of STR interpretation using inferential logic –- development of a criminal intelligence database. Int J Legal Med 109: 14–22

Gill P, d'Aloja E, Andersen J, Dupuy B, Jangblad M, Johnsson V, Kloosterman AD, Kratzer A, Lareu MV, Meldegaard M, Phillips C, Pfitzinger H, Rand S, Sabatier M, Scheithauer R, Schmitter H, Schneider P, Vide MC (1997) Report of the European DNA Profiling group (EDNAP): an investigation of the complex STR loci D21S11 and HUMFIBRA (FGA). Forensic Sci Int 86: 25–33

Hammond HA, Jin L, Zhong Y, Caskey CT, Chakraborty R (1994) Evaluation of 13 short tandem repeat loci for use in personal identification applications. Am J Hum Genet 55: 175–189

Kaye DH (1993) DNA evidence: probability, population genetics and the courts. Harvard J Law Technol 7: 101–172

Lambert JA, Scranage JK, Evett IW (1995) Large scale database experiments to assess the significance of matching DNA profiles. Int J Legal Med 108: 8–13

Matthews R (1998) The great health hoax. In: The Sunday Review (The Sunday Telegraph). September 13 1998: 1–2

Momhinweg E, Luckenbach C, Fimmers R, Ritter H (1998) D3S1358: Sequence analysis and gene frequency in a German population. Forensic Sci Int 95: 173–178

National Research Council (1996) The evaluation of forensic DNA evidence. National Academy Press, Washington DC

Risch NJ, Devlin B (1992) On the probability of matching DNA fingerprints. Science 255: 717–720

Sparkes R, Kimpton C, Gilbard S, Carne P, Anderson J, Oldroyd N, Thomas D, Urquhart A, Gill P (1996) The validation of a 7-locus multiplex STR test for use in forensic casework. Int J Legal Med 109: 195–204

Watson S, Allsop R, Foreman L, Kelsey Z, Gill P (2000) Sequenced allelic ladders and population genetics of a new STR multiplex system. FSS Research Report FSS-RR-811

Werrett DJ (1997) The National DNA database. Forensic Sci Int 88: 33–42