

An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA

Peter Gill^{a,*}, Jonathan Whitaker^a, Christine Flaxman^a, Nick Brown^a,
John Buckleton^b

^a*Forensic Science Service, Priory House, Gooch Street North, Birmingham B56QQ, UK*

^b*ESR, Private Bag 92021, Auckland, New Zealand*

Received 9 December 1999; received in revised form 12 February 2000; accepted 13 February 2000

Abstract

By increasing the PCR amplification regime to 34 cycles, we have demonstrated that it is possible routinely to analyse <100 pg DNA. The success rate was not improved (without impairing quality) by increasing cycle number further. Compared to amplification of 1 ng DNA at 28 cycles, it was shown that increased imbalance of heterozygotes occurred, along with an increase in the size (peak area) of stutters. The analysis of mixtures by peak area measurement becomes increasingly difficult as the sample size is reduced. Laboratory-based contamination cannot be completely avoided, even when analysis is carried out under stringent conditions of cleanliness. A set of guidelines that utilises duplication of results to interpret profiles originating from picogram levels of DNA is introduced. We demonstrate that the duplication guideline is robust by applying a statistical theory that models three key parameters — namely the incidence of allele drop-out, laboratory contamination and stutter. The advantage of the model is that the critical levels for each parameter can be calculated. This information may be used (for example) to determine levels of contamination that can be tolerated within the strategy employed. In addition we demonstrate that interpreting one banded loci, where allele dropout could have occurred, using $LR = 1/2f_a$ was conservative provided that the band was low in peak area. Furthermore, we demonstrate that an apparent mis-match between crime-stain and a suspect DNA profile does not necessarily result in an exclusion. The method used is complex, yet can be converted into an expert system. We envisage this to be the next step. © 2000 Elsevier Science Ireland Ltd. All rights reserved.

Keywords: STR; LCN; Bayes; Likelihood ratio

*Corresponding author. Tel.: +44-121-607-6871; fax: +44-121-622-2051.

E-mail address: dnapgill@compuserve.com (P. Gill)

1. Introduction

The sensitivity of PCR amplification can be improved by increasing the number of PCR cycles. Using the second generation multiplex (SGM) [17,18] DNA profiles have been obtained from single buccal cells simply by increasing the number of amplification cycles from 28 to 34 [10]. However, increased occurrence of stutters and artefacts were observed that reduced the quality of the DNA profile. The benefit of increased sensitivity derived from increasing the number of cycles of amplification has to be balanced against a reduction of profile quality. In addition there are increased risks of laboratory-based contamination to consider. Preparation and extraction of samples must be carried out under stringent conditions of cleanliness.

Nested PCR has been suggested as a method to improve sensitivity [19,20]. This method utilises two sets of primers in two separate PCR reactions. In the first, the STR and adjacent flanking regions are amplified. The primers used in the second round are designed to amplify a smaller product using an aliquot of the first round PCR as the DNA template. Nested PCR reduces the amount of non-specific or artefact products and can analyse contents of a single cell [16], but suffers from the disadvantage of necessitating transfer of PCR product into a separate tube.

To interpret low copy number (LCN) profiles we developed common-sense guidelines that are based on biological principles. An allele is reported only if duplicated in replicate analyses of a sample extract. However, it was useful to compare against a statistical model. Our aim was to discover whether the results derived from the two methods were reasonably concordant.

Accordingly, a Bayesian statistical model was developed that calculates likelihood ratios for a set of replicates by simultaneously taking account of spurious bands, allele dropout and stutter. Following Evett [6,7] we take an explicitly Bayesian approach. The model presented can easily be expanded to interpret mixtures, but the formulae are complex and are best utilised as part of an expert system. In accordance with our philosophy of interpretation, our aim is to discover if the guidelines based on biological principles are reasonable and also to discover any theoretical limitations to the implementation of these principles.

2. Materials and methods

A series of stock DNA dilutions ranging from 1 ng to 0.8 pg were prepared. The DNA was extracted from semen using the Chelex extraction method [21] and quantification was carried out as described by Walsh et al. [22]. DNA dilutions were prepared from the stock DNA using sterile distilled water (SDW). To enable direct comparisons to be made, the same DNA dilution series was used in all of the experiments described. However, it is difficult to accurately quantify low levels of DNA — hence we estimate ranges for quantities <1 ng. Accordingly, we have compared 100–25 pg with 12–0.8 pg.

2.1. Amplification protocols

2.1.1. SGM

SGM was prepared and used as described previously [17,18] except that two different regimes of PCR amplification were compared using either 28 or 34 amplification cycles.

2.1.2. AMPFISTR SGM plus

The AMPFISTR SGM plus™ reaction conditions [5] were: activation 95°C 11 min, 94°C 1 min; cycling 59°C 1 min, 72°C 1 min; extension 60°C 45 min; storage 25°C up to 18 h, 10°C after 18 h.

Each reaction vessel contained 20 µl sterilised distilled water/DNA, 19.1 µl PCR reaction mix, 10 µl ABD primer mix and 0.9 µl TaqGold (H0595). These volumes were contained in 0.2 ml bubble top tubes. No mineral oil was added to the tubes.

The number of amplification cycles was varied between 28 and 56 cycles.

The AMPFISTR SGM plus™ multiplex was also used in a modified nested primer regime as follows:

The first round was amplified with the AMPFISTR SGM plus™ multiplex for 28 cycles; an aliquot (1 µl) was removed and amplified a second time for 28 cycles with fresh AMPFISTR SGM plus™ reagents.

An alternative strategy used the second round amplification of each locus with singleplex primers, supplied by ABD, in separate reactions.

2.1.3. Gel electrophoresis

Electrophoresis of amplification products was performed using 6% acrylamide slab gels 36-cm well to read distance on a 377 ABD automated sequencer [12].

3. Results

3.1. Determination of the optimum number of PCR amplification cycles

To determine the optimum conditions, dilutions of DNA were varied between 28 and 56 cycles. Three parameters were measured: (a) number of alleles observed, (b) imbalance of heterozygote peaks, and (c) prevalence of stutter.

3.1.1. Number of alleles observed

At 28 cycles neither SGM nor AMPFISTR SGM plus™ produced alleles at DNA concentrations below 100 pg. At 34 cycles, full profiles could be obtained down to approximately 25–50 pg — the equivalent of four to ten cell nuclei. Below that level, allele drop-out occurred — it is unlikely that a full genetic complement is present below 25–50 pg.

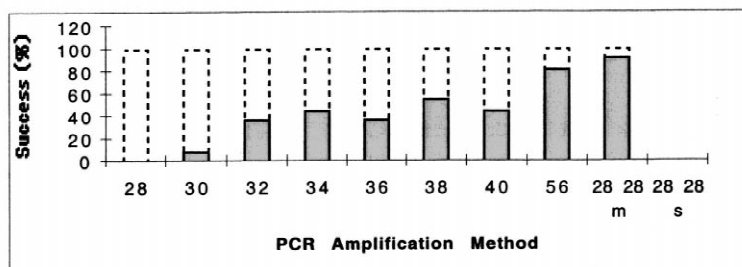
When <100 pg of DNA was analysed there was no advantage to using >34 cycles; the *Taq* enzyme becomes increasingly inefficient with increased cycles because of

degradation. There was some evidence to suggest that more alleles could be observed using nested singleplex primers (Fig. 1).

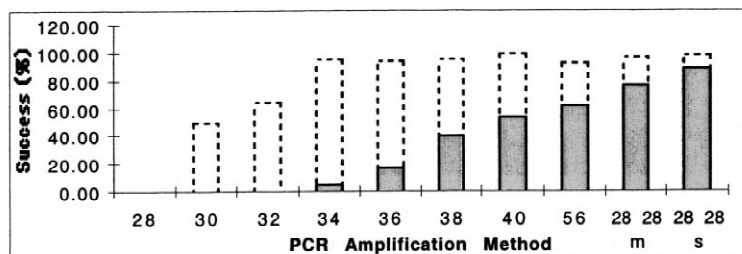
3.1.2. Stutters

The presence of stutters across all loci increased as the cycle number increased, and as the amount of template DNA increased (Fig. 1). When samples were over-amplified, this markedly increased the number of stutters observed. At 34 PCR cycles and 1 ng of DNA approximately 40% of alleles stuttered, whereas at <100 pg only 5% stuttered. Although the incidence of stutter decreased, the size relative to the associated allele

1ng



100pg–25pg



12pg–0.8pg

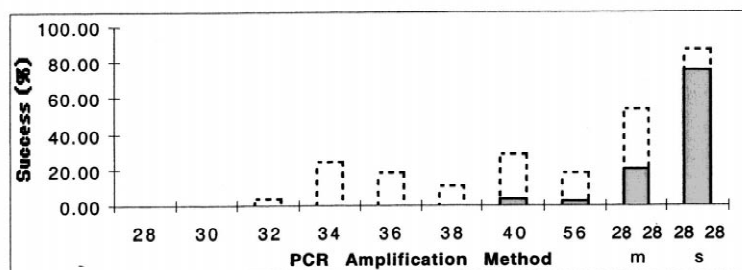


Fig. 1. Demonstration of stutter as a proportion of the success (measured as % of alleles observed across all loci). Success rates are denoted by the dashed histogram bars; the proportion of stutters by the solid histogram bars. Between 28 and 56 cycles are compared. 28 28 m=nested multiplex PCR. 28 28 s=nested multiplex followed by individual singleplex PCR.

increased (in preparation). Guidelines described by Gill et al. [13] cannot be used to interpret stutter.

3.1.3. Imbalance

An imbalanced heterozygote was defined as $(\phi a - \phi b)/\phi a \geq 0.2$, where ϕa is the area of the larger peak and ϕb is the area of the smaller peak, irrespective of molecular weight (Fig. 2). Again, there was a general increase in imbalance as the amount of DNA reduced, and as the number of amplification cycles increased.

When the amount of DNA was <100 pg, the best results were obtained using 34

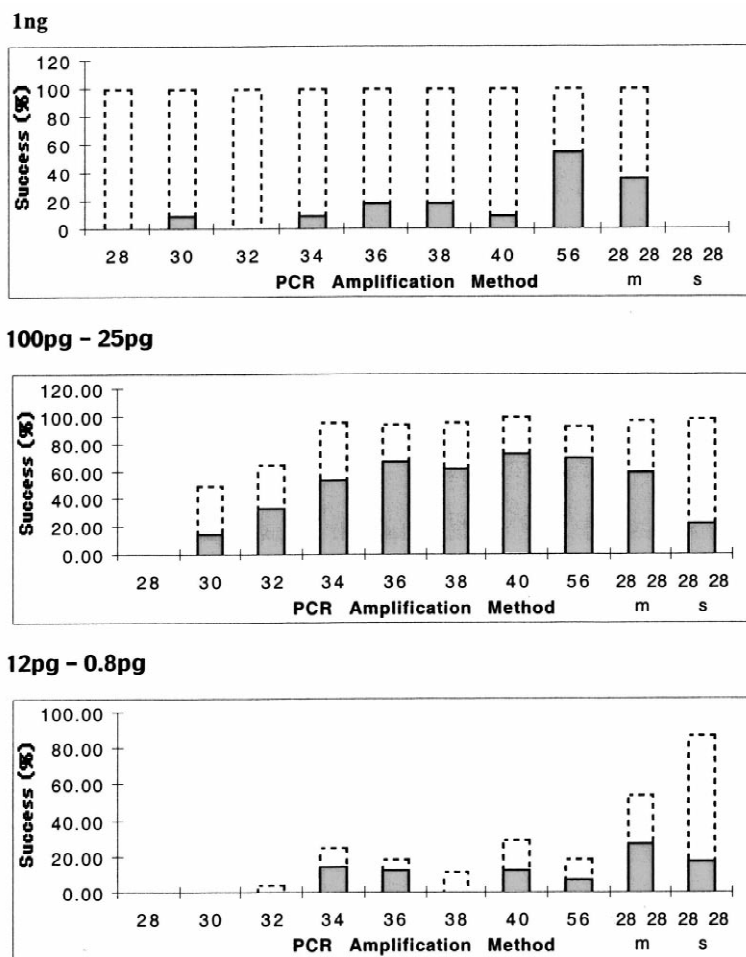


Fig. 2. Demonstration of imbalance as a proportion of the success (measured as % of alleles observed across all loci). Success rates are denoted by the dashed histogram bars; the proportion of imbalanced loci by the solid histogram bars. Between 28 and 56 cycles are compared. 28 28 m=nested multiplex PCR. 28 28 s=nested multiplex followed by individual singleplex PCR.

amplification cycles. The success rate was very high but the increased imbalance of heterozygotes will require a new set of guidelines [13] to interpret mixtures using peak area criteria [3]. Interpretation that does not rely upon peak area information is still possible using Evett et al. [8] rationale. There was no advantage to be gained by increasing the amplification regime above 34 cycles or by using a nested primer regime.

3.1.4. Allele drop-out

Because of stochastic variation any apparent homozygote is considered to be a potential heterozygote. The likelihood ratio can be estimated as $1/2f_a$, where f_a is the frequency of the allele. In Tables 2 and 3 apparent homozygotes are given an F' designation in order to denote the possibility of allele drop-out, where F' means that any allele may be present in the sample

4. Assessing the reliability of the low copy number (LCN) method

4.1. Spurious alleles

DNA extractions were carried out in facilities designed to minimise the chance of laboratory contamination. We use a purpose built laboratory that is fitted with a HEPA filter and maintained under positive air pressure. There is restricted access; all operators wear disposable gowns and masks. All plastic-ware used was guaranteed DNA-free by manufacturers, and was treated with UV prior to use. Nevertheless, negative controls still showed spurious alleles, albeit at low level. In the example shown in Table 1, 21 of 30 negative controls showed evidence of contamination. Typically, the allele peaks were close to the background level. Contaminants were more often associated with the low molecular weight loci, with five to six examples in Amelogenin, D19S433, D3S1358, D8S1179 and less than four in the higher molecular weight loci. Furthermore, 18 negatives showed just one spurious allele; two showed two alleles and one showed three alleles. There was just one example of two bands at a single locus in sample 23 at D2S1338. A total of five alleles (excluding amelogenin) could be attributable to the analyst.

A series of replicate samples from an individual were compared. Samples comprising approximately 3 pg DNA were amplified 50 times using 34 cycles of AMPFISTR SGM plus™ (Table 2). This sample was chosen because of the prevalence of several homozygous loci — a contaminant allele would transform the locus into a heterozygote. A total of six spurious alleles were observed even though all negative controls were ‘clean’; the random nature of the contamination was similar to that previously described.

5. Defining the purpose of the ‘negative control’

There are two kinds of negative control. The first emulates the entire process, from extraction onwards (the extraction control), whereas the second is used to detect possible

Table 1

A compilation of spurious alleles found in 30 replicate negative controls (AMPFISTR SGM plus)^a

Sample	Amelo	D19	D3	D8	THO	VWA	D21	FGA	D16	D18	D2
1	–	–	–	–	–	–	–	–	–	–	–
2	–	–	15	–	–	–	–	–	–	–	–
3	–	–	–	–	–	–	–	–	–	–	–
4	–	–	17	–	–	–	–	–	–	–	–
5	–	–	–	–	–	–	–	–	–	–	–
6	–	–	–	–	–	–	–	–	–	–	–
7	–	14	–	–	–	–	–	–	–	–	–
8	X	–	–	13	–	–	–	–	–	–	–
9	–	–	14	–	–	–	–	–	–	–	–
10	X	–	–	–	–	–	–	–	–	–	–
11	X	–	–	–	–	16	–	–	–	–	–
12	–	–	–	–	–	–	–	–	–	–	–
13	–	13	–	–	–	–	–	–	–	–	–
14	–	–	–	–	–	–	–	–	–	–	–
15	–	–	16	–	–	–	–	–	–	–	–
16	–	–	–	15	–	–	–	–	–	–	–
17	X	15	–	–	–	–	–	–	–	–	–
18	X	14	–	14	–	–	–	–	–	–	–
19	–	–	–	–	–	–	28	–	–	–	–
20	–	–	–	–	–	–	–	–	–	13	–
21	–	–	–	–	–	–	33.2	–	–	–	–
22	–	–	–	–	–	–	–	–	–	–	–
23	–	–	–	10	–	–	25 27	–	–	–	–
24	–	–	–	–	–	–	–	–	–	–	–
25	–	–	15	–	–	–	–	–	–	–	–
26	–	–	–	–	–	–	–	–	–	–	–
27	X	–	–	10	–	–	–	–	–	–	–
28	–	15	–	–	–	–	–	–	–	–	–
29	–	–	15	–	–	16	–	–	–	–	–
30	–	15	–	–	–	–	–	–	–	–	–
+ ve	X Y	14 15	15 17	11 12	6 7	16 17	28 31.2	23 25	11 13	12 13	17 22
- ve	–	–	–	–	9.3	–	–	–	–	–	–

^a Amelo, amelogenin; THO, HUMTH01; D21, D21S11; D18, D18S51; D8, D8S1179; VWA, HUMVWA31/A; FGA, HUMFIBRA/FGA; D19, D19S433; D16, D16S539; D2, D2S1338; D3, D3S1179.

contamination introduced at the PCR stage (the amplification control) i.e. consists of a blank tube containing PCR reagents and distilled water. Both are used concurrently.

The negative control only detects gross contamination within a reagent (such as distilled water) that was used for extraction and PCR. It follows that if this occurs, then the same spurious alleles would be observed in every sample processed.

Within the context of LCN, we have shown that the negative control does not act as indicator of minor contamination within associated processed samples of the same batch. This is because the method is sensitive enough to detect a single molecule of DNA. By definition, one molecule cannot affect more than one tube hence the negative control cannot operate in the traditional sense. Conversely, casework samples could be affected by laboratory-based contaminants that do not appear in the negative control. Con-

Table 2
A partial compilation of results showing incidence of spurious alleles from 50 replicates where a single extract has been repeatedly amplified (AMPFISTR SGM plus)^a

Sample	Amelo	D19	D3	D8	THO	VWA	D21	FGA	D16	D18	D2
Control	X X	14,14	18,18	15,15	7 9.3	19,19	28 32.2	20, 23	9,12	12,16	17,23
1	–	14 F'	–	15 F'	–	–	28 32.2	20 F'	–	16 F'	–
2	X F'	–	18 F'	15 F'	–	19 F'	–	–	12 F'	–	–
3	X F'	–	–	15 F'	–	–	–	–	–	–	17 F'
4	X F'	14 F'	18 F'	–	–	–	–	–	9 12	–	–
5	X F'	–	18 F'	–	–	18 F'	–	–	–	–	–
6	X F'	14 F'	–	–	–	19 F'	28 32.2	20 F'	–	12 F'	–

^a See Table 1 for full locus designations. A spurious allele at the VWA locus is highlighted bold in the table.

taminants may be tube-specific and transfer could occur via minute dust particles, or plastic-ware.

Nevertheless, the negative control serves an important function as a ‘health check’ of the process, since the incidence of contaminant alleles in the negative is directly associated with the chance of observing a laboratory-based contaminant within casework samples of DNA. In the context of LCN, replication of extraction negative controls is recommended in order to determine if laboratory-based contaminants are reproducibly amplified. If not, we suggest that there is no a priori reason to suppose that any of the alleles observed in the negatives have affected the associated extracted samples.

In addition, it is useful to compare profiles against operator controls to guard against the possibility of gross contamination of a sample.

6. Importance of replication of results

Accepting that it was not possible completely to avoid laboratory contamination we used the 50 control samples as a model to test the reliability of a guideline that only allowed an allele to be reported if a duplicate result could be obtained from two or more replicate samples. Effectively, a consensus is reported. An assessment to evaluate the risk of designating a spurious allele can be made — consider the first two results in Table 2: Only D8S1179 allele 15 appears in the consensus and could be reported. If the first four replicates are analysed, five alleles are duplicated and appear in the consensus result.

The small, albeit low risk, that spurious alleles could be duplicated and reported in the consensus results was illustrated by pairwise comparisons of data. Just four occasions out of 1225 comparisons showed that the same spurious allele would be recorded in two different samples ($p = 4/1225 = 0.003$). Even if an allele was found in a crime sample that did not match the suspect this does not necessarily lead to an exclusion (demonstrated using the Bayesian model). Note that with real casework, it is not usually possible to carry out more than three separate tests of a DNA extract because of the limited sample size.

Table 3
Results from an actual case showing derivation of the consensus result^a

Sample	Amel	VWA	THO	D8	FG	D21	D18	D19	D3	D16	D2
R_1	XY	16,19	6,7	12,14	20,24	28,30	12,F	13,17	15,16	11,13	17,20
R_2	XY	16,19	6,F	12,14	20,24,25	28,30	12,F	13,17	15,16	11,13	17,20
Consensus	XY	16,19	6,F	12,14	20,24	28,30	12,F	13,17	15,16	11,13	17,20
Suspect	XY	16,19	6,7	12,14	20,24	28,30	12,12	13,17	15,16	11,13	17,20
Negative 1	X	14	—	14,15	—	—	—	15	15	—	—
Negative 2	X	14	—	—	—	—	—	14	16	5	—

^a See Table 1 for full locus designations.

7. An actual case

The following is an example from an actual case (Table 3). The consensus result is shown from the duplicate tests. Note that in the negative controls HUMVWFA31/A allele 14 was duplicated. However, this does not affect the consensus since it does not appear in the test samples (otherwise it could not have been designated). The likelihood ratio is calculated using $1/2f_a f_b$ for heterozygotes and $1/2f_a$ for homozygotes, where f_a is the frequency of allele a and f_b is the frequency of allele b .

8. A Bayesian model

We consider three separate phenomena as follows:

C: The event that an allele appears at a given locus in the crime sample as a result of contamination. Contamination in this sense is taken to mean additional genetic material added to one replicate during processing. Genetic material in the sample per se is treated as a component of a mixture. Strictly speaking, we should use C_L being the probability of a contaminant allele at a given locus — however, we currently approximate all loci as behaving equally. In addition we will use the term \bar{C} by which we mean that no contaminant allele is seen. Strictly we should allow for the possibility that a contaminant allele has appeared in a position of an existing allele (and therefore is not seen), however for simplicity we have not done this. For example if a locus shows ab alleles in the crime stain and the suspect is an ab genotype then we write $p(\bar{C})$ meaning that no contamination has occurred. Strictly we should allow for the possibility that contamination has occurred at position a or b and is not seen because it is masked by the presence of true allelic peaks at a and b . This would suggest a formulation $p(\bar{C}) + p(C)(f_a + f_b)$. Since, in most real cases C , f_a and f_b are small this should be a minor error.

D: The event that a given allele drops out, primarily because of stochastic variation of amplification of low copy number DNA. Initially we model the dropout of homozygotes (both alleles) and heterozygotes with the same probability estimate.

St: The event that a band has been caused by stuttering from an adjacent allelic peak.

9. Estimating $p(C)$, $p(D)$ and $p(St)$

We model these continuous events as binary for simplicity. This means that there is a loss of information. A continuous equivalent is envisaged but this cannot be implemented at this stage without further research. Therefore, under this simplified model an allele either drops out or it does not, an allele either stutters or it does not, contamination either happens or it does not. We accept the preliminary nature of this approach — the continuous method is likely to be markedly superior (although much more complex).

$p(C)$: The probability of laboratory induced contamination can be estimated from experimental observation. In the absence of sufficient data, and for the purposes of illustration, we make the preliminary simplified assumption that contamination is a

completely random event. Hence the probability of observing a given contaminant allele at a given locus is $p(C) \times f_a$, where f_a is the frequency of the allele in the population. A compilation of data comprising spurious bands from negative controls could be used to give a better estimate. It is probable that different laboratories will have different levels of $p(C)$.

$p(D)$: We consider the event that a given allele drops-out given that there is more than one allele present at the locus — $p(D)$ is calculated from experimental observation.

$p(St)$: A band is considered a possible stutter if it is four bases less than the size of an associated allele. When a nanogram of DNA is analyzed using 28 cycles of PCR, the size in peak area is usually less than 15% the size of the associated allele. For LCN, the relative size of the peak (area) is not as informative since a stutter peak may actually exceed the size of the associated allelic peak (unpublished data). We also consider that a stutter may occur in conjunction with drop-out of the associated allele.

10. A generalised statistical theory

Consider that we do $R_1, R_2 \dots R_n$ replicates that originate from separate PCR amplifications of a sample.

Hypotheses are defined as:

- H_1 : The DNA in the crime stain is the suspect's
- H_2 : The DNA in the crime stain is from someone else

Without loss of generality we can write:

$$LR = \frac{p(R_1 R_2 \dots | H_1)}{p(R_1 R_2 \dots | H_2)} \quad (1)$$

We assume that the events in one replicate are independent of those in the others. We accept that it is unlikely to be true, but useful nonetheless. First consider that it is the events of contamination, stutter and dropout that are assumed to be independent and not the profiles themselves. For example, we assume that a contaminant band in R_1 has no effect on the presence or absence of one in R_2 . However, the presence of a contaminant in one replicate is likely to increase the chance of one in another, especially if the contaminants originate from the scientist manipulating the samples — this is minimised by wearing protective clothing.

Future work will address this in some detail and will lead to improvement of the existing models, which were intended primarily to test the 'biologically' derived rules rather than replace them.

Nevertheless, it does seem plausible that any error caused by ignoring departures from independence is likely to benefit the defendant. The logic for this is that a laboratory contaminant has an increased chance of occurring in both duplicates. Therefore the probability of contamination occurring in R_2 conditional on it appearing in R_1 is higher

than $p(C)$. Our assumption of independence is likely to underestimate the numerator more than the denominator and hence is likely to be conservative:

$$p(R_1 \dots | M_j) = \prod_i p(R_i | M_j) \quad (2)$$

It is necessary to specify the ‘components’ of H_2 . By this we postulate that the ‘expert’ considers a set of possible ‘random man’ genotypes $M_1, M_2 \dots M_n$. These are by nature exclusive but not necessarily exhaustive. This step is actually unnecessary, as it is in theory possible to postulate all possible ‘random man’ genotypes and sum over all possibilities. It does however markedly simplify manual implementation if the list can be shortened by judgement.

Examples of the steps which detail examples of the criteria used in expert judgement are given by Clayton et al. [3] and Gill et al. [13]:

$$LR = \frac{p(R_1 R_2 \dots | H_1)}{\sum_i p(R_1 R_2 \dots | M_j) p(M_j)} \quad (3)$$

Hence:

$$LR = \frac{\prod_i p(R_i | H_1)}{\sum_j \prod_i p(R_i | M_j) p(M_j)} \quad (4)$$

It is convenient to analyse the components separately in tabular format as illustrated in Table 4.

To calculate $p(R_i | M_j)$, we assume that the events C , D , and St are independent of each other and of M_j .

11. The procedure to estimate a LR

We begin by considering spurious peaks and dropout only. Later we will consider stuttering. Suppose a crime stain was analysed three times (R_1 , R_2 and R_3) and three different phenotypes were observed at the D18S511 locus: The suspect (S) was 12,16; $R_1 = 12$; $R_2 = 16$; $R_3 = 12,16$. H_1 is the probability of the evidence if the profile is the

Table 4

A format to illustrate the mode of calculation of the LR^a

M_j	$P(M_j)$	R_1	R_2	Calculation
M_1	$P(M_1)$	$P(R_1 M_1)$	$P(R_2 M_1)$	Product of this row
M_2	$P(M_2)$	$P(R_1 M_2)$	$P(R_2 M_2)$	Product of this row
				Sum of this column

^a M_1 describes the list of potential genotypes. $p(M_j)$ is the probability of the genotype in the population (estimated from a population database). $R_1 \dots R_n$ describe the probabilities of the genotypes in the replicate samples, given M_j .

suspect's. H_2 is the probability of the evidence if the profile is from someone else. To calculate the LR , we use the format of Table 4 and proceed as follows:

Step 1: Assess the reasonable random man genotypes from the information in the replicates. List in column M_j

Step 2: Calculate $p(M_j)$ in the second column

Step 3: Calculate $p(R_i|M_j)$ in columns R_1 , R_2 and R_3

Step 4: Calculate the products of each row

Step 5: Sum the products

Step 6: The numerator is the product $p(R_i|M_j)$ corresponding to the genotype of the suspect. In the example (Table 5) this appears as part of the term at the right hand side of the second row corresponding to the genotype 12,16 but without the frequency terms

$$LR = \frac{1}{2f_{12}f_{16} \left[1 + \frac{f_{12} f_{16} p(C)^2}{p(D) p(\overline{D})^2 p(\overline{C})^2} \right]} \quad (5)$$

Provided that $p(C)$ is small (<0.3), the expression in the denominator

$$1 + \frac{f_{12} f_{16} p(C)^2}{p(D) p(\overline{D})^2 p(\overline{C})^2} \approx 1, \text{ hence:}$$

$$LR \approx \frac{1}{2f_{12} f_{16}} \quad (6)$$

12. Some common examples

We assess some examples that have been commonly encountered.

12.1. Apparently single banded profiles

When an apparent one-banded homozygote is encountered in a crime stain ($R_1 = a$) and the peak area is small, this may mean that allele dropout has occurred, i.e. the

Table 5

Calculation of the components of the likelihood ratio for an example where three replicates show evidence of spurious bands and allele drop-out

M_j	$P(M_j)$	R_1 (12)	R_2 (16)	R_3 (12,16)	Products
12 12	f_{12}^2	$p(\overline{D}) p(\overline{C})$	$p(D) p(\overline{C}) f_{16}$	$p(\overline{D}) p(\overline{C}) f_{16}$	$p(\overline{D})^2 p(\overline{C})^2 p(D) p(C)^2 f_{12}^2 f_{16}^2$
12 16	$2f_{12}f_{16}$	$p(D) p(C) p(\overline{D})$	$p(\overline{D}) p(\overline{C}) p(D)$	$p(\overline{D})^2 p(C)$	$2p(\overline{D})^4 p(\overline{C})^3 p(D)^2 f_{12}f_{16}$
16 16	f_{16}^2	$p(D) p(C) f_{12}$	$p(D) p(C)$	$p(D) p(C) f_{12}$	$p(D)^2 p(C)^2 p(D) p(C)^2 f_{12}^2 f_{16}^2$
					Denominator = sum of above

genotype may in fact be heterozygous. This is considered a possibility whenever the peak is close to background. At low peak areas, experimental observation confirms that the probability of allele dropout $p(D)$ is high. If the allele in the crime stain is type a and the suspect is type ab then it would seem reasonable to limit M_j to aa , ab or aF , where F stands for any allele other than type a or type b . We can approximate $LR \approx 1/2f_a$. The approximation can be tested using:

$$LR = \frac{1}{2f_a \left[1 + \frac{1 - 2p(D)}{2p(D)} f_a \right]} \quad (7)$$

We test if the scaling function $1/[1 + (1 - 2p(D))/(2p(D)) f_a] \geq 1.0$ in Fig. 3. Provided that $p(D) > 0.5$, which is always reasonable when the peak is close to the background, then the approximation is demonstrated to be conservative.

12.2. Apparently one banded profiles — effect of one additional replicate

Taking the previous example, we now consider the advantages of replication, where an additional aliquot (R_2) of the same DNA extract is separately amplified. Suppose that

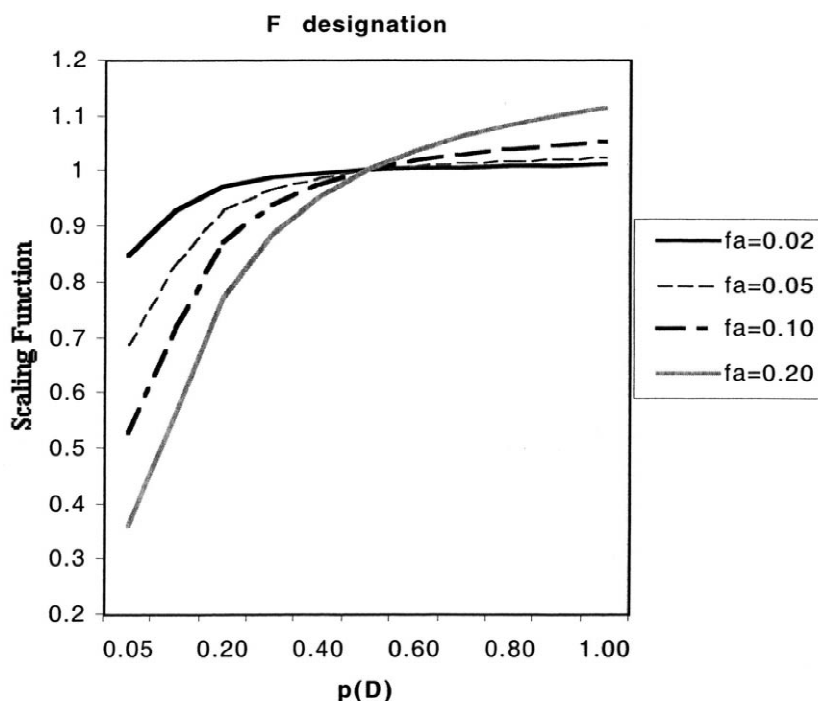


Fig. 3. Testing the robustness of the F designation. The F designation is conservative provided that the expression $1/[1 + (1 - 2p(D))/(2p(D)) f_a] = 1.0$ (from Eq. (7)). Allele frequencies (f_a) 0.02, 0.05, 0.10 and 0.20 are tested. Generally, the F designation is conservative provided $p(D) > 0.5$.

the second replicate yields a heterozygote ab that matches the suspect's profile (Suspect = ab ; $R_1 = a-$; $R_2 = ab$).

We take account of three possible explanations of the evidence — either $M_1 = aa$ homozygote or M_1 is a heterozygote and an allele has dropped out, else $M_2 = ab$ heterozygote. If the first explanation is true then the b allele must be a spurious band. In this example the LR would be reported as $LR = 1/2f_a$ because only the a allele was duplicated. The formula that describes this model is:

$$LR = \frac{1}{2f_a f_b \left[1 + \frac{f_a p(C)}{2p(D) \overline{p(D)} \overline{p(C)}} \right]}$$

This expression is always less than $1/2f_a f_b$, but the $1/2f_a$ evaluation is a conservative approximation whenever $1/f_b [1 + (f_a p(C))/(2p(D) \overline{p(D)} \overline{p(C)})] \geq 1.0$ and this is true for all reasonable estimates of $p(C)$ and $p(D)$.

If a locus appears homozygous with allele a , but allele drop-out could have occurred, so that the locus was really a heterozygote, then interpretation using $1/2f_a$ is reasonable provided that contamination is low and the allele peak area itself is small or close to the baseline.

12.3. Additional replicates increase the LR

Continuing with the previous example, we consider the effect of additional (n) replicates that have been analysed and demonstrated to all correspond to the genotype of the suspect (ab in this example). The suspect is ab ; $R_1 = a$; $R_2 \dots R_n = ab$ (i.e. a total of $n+1$ replicates were analysed). The general formula that describes $R_1 = a-$ and n = the number of replicates that are genotype ab is:

$$LR = \frac{1}{2f_a f_b \left[1 + \frac{f_a f_b^{n-1} p(C)^n}{2p(D) [\overline{p(D)} \overline{p(C)}]^n} \right]} \quad (9)$$

Provided that n is greater than or equal to 2, the guideline will allow the reporting of $LR = 1/2f_a f_b$ because both alleles are duplicated. The LR calculated from Eq. (9) will always be less than this but is nevertheless a very close approximation (Fig. 4a) for most intermediate values of $p(D)$. It is noted that the difference between $n=2$ and $n=3$ is minor in these simulations. Also that the actual values of $p(D)$ and $p(C)$ have very little effect on the final estimate (Fig. 4a,b), provided that the latter is less than 0.6 (which it should always be). If $n=1$ then the LR is conservative relative to $1/2f_a$.

13. An example where an allegedly contaminant band is observed in conjunction with allele dropout

The next example is more extreme than those previously discussed. Suppose that a replicate (R_1) matches the suspect at one allele (b), but has an additional allele (c) that is

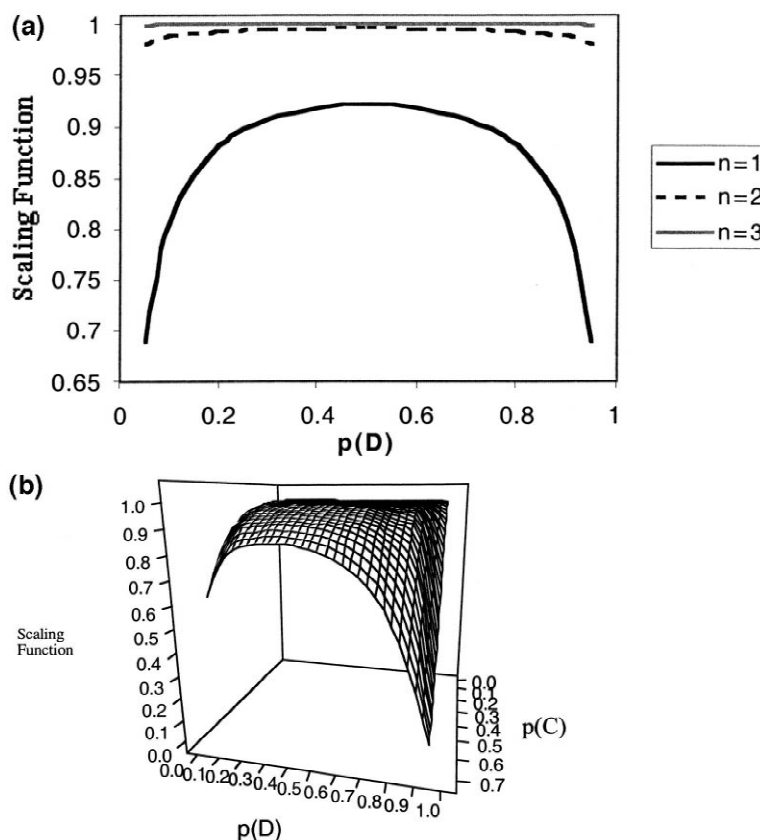


Fig. 4. (a) Evaluation of $1/[1 + (f_a f_b^{n-1} p(C)^n)/(2p(D)[p(D) p(C)]^n)]$ from Eq. (9) (called the scaling function). When n (the number of replicates where the genotype is ab) is greater than or equal to 2 and $R_1 = a$, then the $LR \approx 1/2f_a f_b$. When $n=1$, $1/2f_a$ would be used which is always conservative. In Fig. 2a, $p(C) = 0.3$; $f_a = f_b = 0.1$. (b) The three-dimensional model showing the relationships between $p(C)$ and $p(D)$, when $f_a = 0.1$. The scaling function ≈ 1.0 for moderate and low values of $p(C)$ and for all intermediate values of $p(D)$.

not found in the suspect under the prosecution hypothesis (H_1). Furthermore, we assume that there is no trace of allele a . We assess the condition where the suspect is ab ; R_1 is bc ; R_2 is ab by consideration of the genotypes (M_j) ab , ac , bc and bb :

$$LR = \frac{1}{2f_a f_b \left[2 + \frac{f_b p(D) p(C)}{p(D) p(C)} + \frac{f_b p(C)}{2p(D) p(D) p(C)} \right]} \quad (10)$$

The reporting guideline would only allow reporting of the duplicated b allele hence the reported likelihood ratio would be $LR = 1/2f_b$. There was very little effect contributed by $p(D)$ since the scaling function was always greater than 1.0 even when $p(C)$ was

moderately high (Fig. 5a,b), demonstrating the conservative nature of the reporting guideline.

14. Example where an allegedly contaminant band is observed

Suppose that the suspect is ab ; $R_1 = abc$ (where c is a supposedly a contaminant allele under H_1) and $R_2 = ab$. We limit the possible (M_j) genotypes to ab , ac or bc and we evaluate against the guideline $LR \approx 1/2f_a f_b$ (Fig. 6a,b). Evaluated against $p(C) = 0.3$, the approximation is reasonable provided that f_a is less than 0.10 and $p(D)$ is less than 0.50:

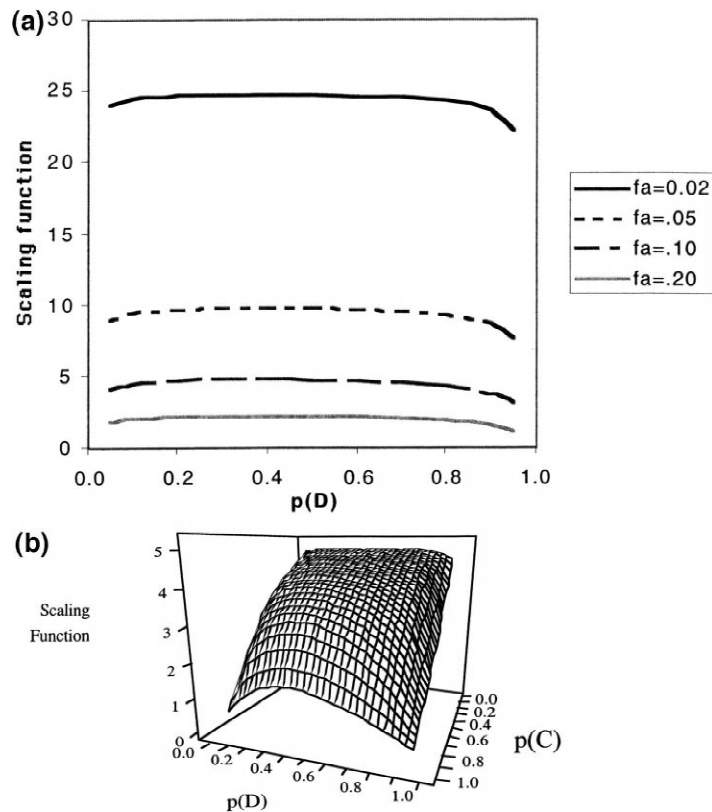


Fig. 5. (a) Provided that the scaling function $1/f_a[2 + (f_b p(D) p(C))/(p(\bar{D}) p(\bar{C}) + (f_b p(C))/(2p(D) p(D) p(C)))] \geq 1.0$ (from Eq. (10)) then $1/2f_b$ is conservative. Allele frequencies are $f_a = f_b = 0.02, 0.05, 0.1, 0.2$, respectively, and $p(C) = 0.3$. (b) Three-dimensional graph to show the combined effect of $p(D)$ and $p(C)$, when $f_a = 0.10$. The critical point where the scaling function > 1 is reached when $p(C) < 0.9$ and $p(D) < 0.9$.

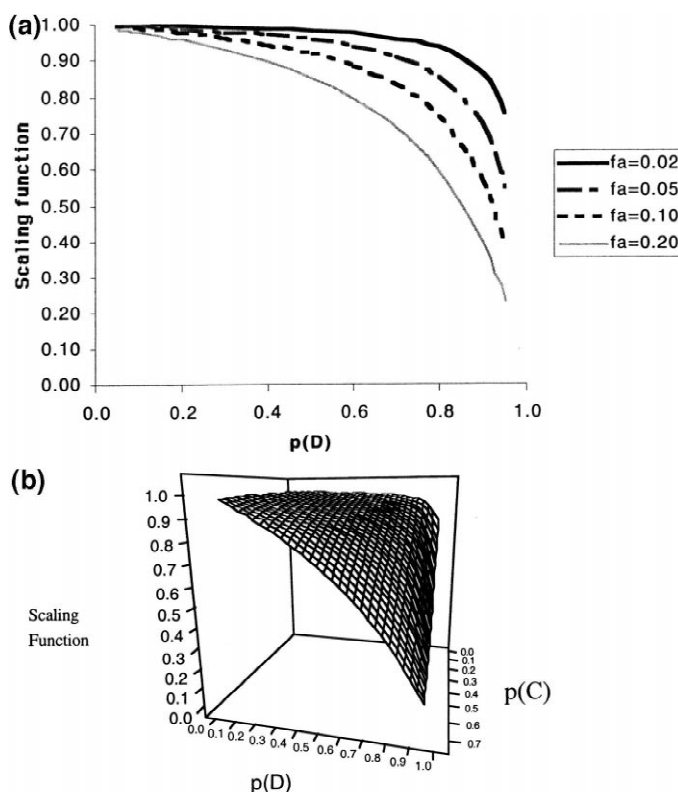


Fig. 6. (a) Evaluation of scaling function $1/[1 + p(D) p(C)(f_a + f_b)/(p(D) p(C))]$ (Eq. (11)) for $f_a = f_b = 0.02, 0.05, 0.10$ and 0.20 , respectively; $p(C) = 0.30$. (b) When $f_a = 0.1$, the $LR \approx 1$ when $p(D) < 0.5$ and $p(C) < 0.3$.

$$\frac{1}{2f_a f_b \left[1 + p(D) p(C) \left(\frac{f_a + f_b}{p(D) p(C)} \right) \right]} \quad (11)$$

15. Evaluation of stutters

A stutter is defined as an allelic product that is 1 repeat unit less than an associated allele. Stutters are well characterised for conventional PCR systems that use 28 PCR cycles. With the LCN technique, their presence may be dependent on multiple variables such as concentration of DNA, level of PCR inhibition and the sample type. It has been observed with LCN that the stutter peaks can exceed 15% of the peak area of the allelic peak (the guideline used for 28-cycle PCR [13]). In fact, stutters can exceed the area of the allelic peak. Because $p(D)$ and $p(C)$ are easier parameters to assess, a system that is relatively independent of stutter is desirable. We now consider the scenario where the

Table 6
Derivation of Eq. (12)^a

M_j	$P(M_j)$	$R_1 = abc$	$R_2 = ac$	Products
ab	$2f_a f_b$	$p(\overline{D})^2 p(\overline{C}) f_c p(\overline{St})^2$	$p(\overline{D}) p(D) p(\overline{C}) f_c p(\overline{St})^2$	$2f_a f_b f_c^2 p(\overline{D})^3 p(D) p(\overline{C})^2 p(\overline{St})^4$
ac	$2f_a f_c$	$p(D)^2 p(St) p(\overline{C}) p(\overline{C})$ $+ p(St) p(\overline{C}) f_b$	$p(D)^2 p(\overline{C}) p(\overline{St})^2$	$2f_a f_c p(D)^4 p(\overline{C}) p(\overline{St})^3$ $[p(St) p(\overline{C}) + p(St) p(\overline{C}) f_b]$
bc	$2f_b f_c$	$p(D)^2 p(\overline{C}) f_a p(St)$	$p(\overline{D}) p(D) p(\overline{C}) f_a p(\overline{St})^2$	$2f_a^2 f_b f_c p(D)^3 p(D) p(\overline{C})^2 p(\overline{St})^3$

^a Note that $p(St)$ only appears once when $M_j = bc$ and $R_1 = abc$ because b must be in part or wholly allelic.

suspect is ac , $R_1 = abc$, where b is in a stutter position and $R_2 = ac$. We limit the possible M_j genotypes to ab , ac and bc (recall that this limitation is for convenience and simplicity only and that the approach can be applied to a complete list of any possible genotypes) (Table 6):

$$LR = \frac{1}{2f_a f_c \left[1 + \frac{p(\overline{C})^2 p(D) f_b (f_a + p(\overline{St}) f_c)}{p(\overline{C}) p(\overline{D}) \{p(St) p(\overline{C}) + p(\overline{St}) p(\overline{C}) f_b\}} \right]} \quad (12)$$

Fig. 7 demonstrates that the LR is always less than $1/2f_a f_c$ (since the scaling function is always less than 1) but gives a reasonable approximation provided that $p(St)$ is high (greater than 0.5 and $p(C) = 0.3$). Although the LR increases as $p(D)$ decreases, the effect is small provided $p(St)$ is high.

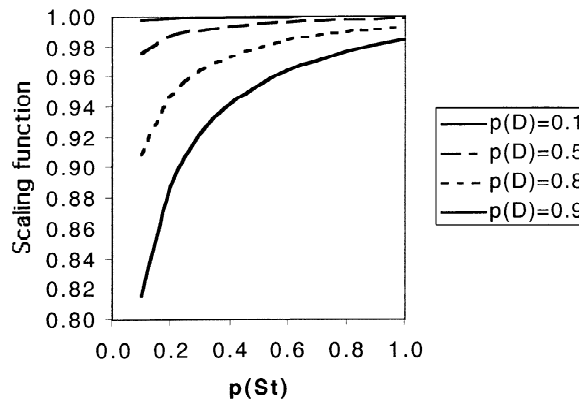


Fig. 7. Plots of scaling function $1/[1 + (p(\overline{C})^2 p(D) f_b (f_a + p(\overline{St}) f_c) / (p(\overline{C}) p(\overline{D}) \{p(St) p(\overline{C}) + p(St) p(\overline{C}) f_b\})]$ (Eq. (12)) vs. $p(St)$ for several levels of $p(D)$ ranging from 0.1 to 0.9. $p(\overline{C}) = 0.3$ and $f_a = f_b = f_c = 0.1$.

16. Extension of the dropout definition and evaluation of an actual example

Referring to an analysis of 50 PCR replicates of a sample (Table 5) that had four homozygous STRs and six heterozygous loci were analysed, we now expand the definition of $p(D)$ as follows:

- $p(D_{Ho})$ — the probability of dropout given that the locus is homozygous
- $p(D_{He})$ — the probability that a given allele drops out given that the locus is heterozygous
- $p(D_{He2})$ — the probability that both alleles drop out given that the locus is heterozygous

In the following calculations, the $p(D)$ parameters were either assigned the actual observed values in Table 7 or if not available, e.g. (pD_{He}) for locus D8S51179, the mean across available loci was used instead. By observation, $p(C)$ did not exceed 0.2 for any locus and this value was adopted throughout. Allele frequencies were used from a white Caucasian database.

From two of the replicates, we consider an extreme observation in Table 8 for the locus D3S1358. The suspect is 18,18; $R_1 = 15,18$ and $R_2 = 15,18$, i.e. an apparent mismatch/exclusion. Conventional analysis may indicate the results to be either inconclusive or an exclusion since a spurious allele is duplicated in the replicates. Using Eq. (13), limiting our considerations of M_j genotypes to 15,15; 15,18 and 18,18; the $LR = 0.068$. The evidence supports exclusion, but importantly, the LR is greater than zero:

$$LR = \frac{1}{f_{18}^2 \left[2 + \frac{2\{p(\overline{D}_{He})^2 p(\overline{C})\}^2}{\{p(\overline{D}_{Ho}) p(\overline{C})\}^2 f_{15} f_{18}} \right]} \quad (13)$$

Examination of the LR values calculated for individual loci revealed that all were either conservative or very close to the estimates derived by calculating $LR = 2f_a$ for homozygotes or $LR = 2f_a f_b$ for heterozygotes (Table 8). When complete locus dropout is observed in a replicate this has very little effect, i.e. the scaling function ≈ 1.0 . The combined LR across all loci = 68 000 (using a white Caucasian database [11]) and this serves to demonstrate the important principle that apparent allele mismatches caused by contamination do not necessarily lead to exclusions.

17. Extension to include subpopulation effects

In theory the extension to include subpopulation effects is straightforward but complex. Such an extension should be relatively easy in an expert system. This is desirable both because it is more consistent with the approaches employed in other areas of DNA interpretation but more importantly because it follows logically from the correct

Table 7
Analysis of $p(D)$ parameters from analysis of 50 replicates

$p(D)$	Amelo	D19S433	D3S1358	D8S51179	HUMTH01	HUMVWFA31/A	D2S13381	HUMFIBRA (FGA)	D16S539	D18S51	D2S1338	Mean
$p(D_{Ho})$	0.40	0.60	0.50	0.56	–	0.62	–	–	–	–	–	0.54
$p(D_{He})$	–	–	–	–	0.32	–	0.36	0.28	0.44	0.32	0.20	0.32
$p(D_{He2})$	–	–	–	–	0.64	–	0.52	0.68	0.46	0.56	0.70	0.59

Table 8
Analysis of LR s from an example case

Sample	Amelo	D19S433	D3S1358	D8S51179	HUMTH01	HUMVWFA31/A	D2S13381	HUMFIBRA (FGA)	D16S539	D18S51	D2S1338
Suspect	XX	14 14	18 18	15 15	7 9.3	19 19	28 32.2	20 23	9 12	12 16	17 23
R_1	XF	–	15 18	15 F	7 F	–	–	–	9 12	12 16	–
R_2	–	–	15 18	15 F	–	19 F	–	–	–	12 16	23 F
Consensus	XF	–	15 18	15 F	–	–	–	–	–	12 16	–
LR (calculated)	–	–	0.068	18.1	2.9	10.9	–	–	12.2	25.9	4.6
LR ($1/2f_a$) or ($1/2f_a f_b$)	–	–	–	4.9	2.8	5.9	–	–	13.5	25.9	4.46

consideration of the conditional nature of the probability of a genotype of possible offenders *given* the suspect's genotype [9]. All that is required is to replace $p(M_j)$ with the conditional probability $p(M_j|M_s)$. The methodology is standard in mixture interpretation [1,2,4,9,15].

18. Discussion

Increasing the number of amplification cycles can expand the utility of DNA profiling in forensic casework. DNA profiles can be derived from minute traces of evidence, such as fingerprints. The purpose of this investigation was to examine the potential of this technique under controlled laboratory conditions. It was desirable to keep the number of PCR cycles to a minimum, otherwise increased incidence of artefacts could compromise the quality of the result. At the same time the sensitivity of the technique had to be sufficient to detect DNA from <100 pg of DNA. At these very low levels of DNA concentration, stochastic variation will preferentially amplify different alleles and allele drop out can occur. It was preferable to carry out duplicate analysis rather than to concentrate samples, because this did not usually increase the overall concentration of DNA above the stochastic threshold. Accordingly, it was demonstrated that under the optimal 34 cycles regime, there was increased heterozygote imbalance. The incidence of stutter decreased with smaller DNA quantities but their peak areas were correspondingly greater than described by Gill et al. [14] and this means that the guidelines described cannot be used. In addition, with increased PCR cycle number there was increased incidence of laboratory-based contamination that was unavoidable. Negative controls do not act as a reliable indicator of low-level contamination and cannot be used as such. To report DNA profiles we use replication of PCR amplification results and their collation to produce a consensus that is derived from a demonstration of reproducibility. By comparing results against a novel statistical model, we have demonstrated that this approach is reasonable and generally very conservative provided that levels of contamination were low. In addition, we demonstrated that interpreting one banded loci, where allele dropout could have occurred, using $LR = 1/2f_a$ was conservative provided that the band was low in peak area. Importantly, we demonstrate that a mis-match between crime-stain profile and a suspect profile does not always result in an exclusion. The use of 34 cycles appeared to be optimal. We are now carrying out an additional studies on the 34 cycle AMPF/STR SGM plus™ system in order to assess levels of stuttering, drop-out and contamination for each locus with different levels of template DNA in order to refine the guidelines for reporting purposes

19. Summary of the current guidelines

1. Duplication of every allele is demonstrated before reporting
2. If the negative controls show duplicated alleles that correspond to those in the samples, they are not reported — where possible, samples are reanalysed. If alleles

are found in the negatives, which do not correspond to those in samples, then this is inconsequential

3. If there is one allele in a sample which does not match the suspect's profile then further work is indicated. Alternatively, standard mixture analysis may be indicated. Different samples, e.g. alternative stain material from the case should be processed. This may not be possible however. Under this circumstance the profile may still be reported but the strength of the evidence is reduced
4. Ultimately, these guidelines will be superseded by expert systems utilising the Bayesian principles described in this paper

References

- [1] D.J. Balding, R.A. Nichols, Forensic relatedness, database selection and single bands, *Forensic Sci. Int.* 64 (1994) 125–140.
- [2] D.J. Balding, R.A. Nichols, A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity, *Genetica* 96 (1995) 3–12.
- [3] T.M. Clayton, J.P. Whitaker, R. Sparkes, P. Gill, Analysis and interpretation of mixed forensic stains using DNA STR profiling, *Forensic Sci. Int.* 91 (1998) 55–70.
- [4] J.M. Curran, C.M. Triggs, J. Buckleton, B.S. Weir, Interpreting DNA mixtures in structured populations, *J. Forensic Sci.* 44 (1999) 987–995.
- [5] E.A. Cotton, R. Allsop, J. Guest, R. Frazier, P. Koumi, I.P. Callow, A. Seager, R. Sparkes, Validation of the AMPF/STR[®] SGM Plus[™] system for use in forensic casework (2000) in preparation.
- [6] I.W. Evett, What is the probability that this blood came from that person: A meaningful question?, *J. Forensic Sci. Soc.* 23 (1983) 35–59.
- [7] I.W. Evett, Establishing the evidential value of a small quantity of material found at a crime scene, *J. Forensic Sci. Soc.* 33 (1993) 83–86.
- [8] I.W. Evett, G. Buffery, G. Willott, D. Stoney, A guide to interpreting single locus profiles of DNA mixtures in forensic cases, *J. Forensic Sci. Soc.* 31 (1991) 41–47.
- [9] I.W. Evett, B.S. Weir, *Interpreting DNA Evidence*, Sinauer Associates, Sunderland, MA, 1998.
- [10] I. Findlay, R. Frazier, A. Taylor, A. Urquart, Single cell DNA fingerprinting for forensic applications, *Nature* 389 (1997) 555–556.
- [11] L.A. Foreman, I.W. Evett, Statistical analyses to support forensic interpretation for a new ten-locus STR profiling system. *Int. J. Leg. Med.* (2000) in press.
- [12] R. Frazier, E. Millican, S. Watson, N. Oldroyd, R. Sparkes, K. Taylor, S. Panchal, L. Bark, C. Kimpton, P. Gill, Validation of the Applied Biosystems Prism 377 automated sequencer for forensic short tandem repeat analysis, *Electrophoresis* 1 (1996) 1.
- [13] P. Gill, R. Sparkes, C. Kimpton, Development of guidelines to designate alleles using a STR multiplex system, *Forensic Sci. Int.* 89 (1997) 185–197.
- [14] P. Gill, R. Sparkes, R. Pinchin, T.M. Clayton, J.P. Whitaker, J. Buckleton, Interpreting simple STR mixtures using allele peak area, *Forensic Sci. Int.* 91 (1998) 41–53.
- [15] S.A. Harbison, J.S. Buckleton, Applications and extensions of subpopulation theory: a caseworkers guide, *Sci. Just.* 38 (1998) 249–254.
- [16] M.C. Snabes, S.S. Chong, S.B. Subramanian, K. Kristjansson, D. Disepio, M.R. Hughes M R, Preimplantation single-cell analysis of multiple genetic loci by whole genome amplification, *Proc. Natl. Acad. Sci. USA* 91 (1994) 6181–6185.
- [17] R. Sparkes, C. Kimpton, S. Watson, N. Oldroyd, T. Clayton, L. Barnett, J. Arnold, C. Thompson, R. Hale, J. Chapman, A. Urquart, P. Gill, The validation of a 7-locus multiplex STR test for use in forensic casework. I. Mixtures, ageing, degradation and species studies, *Int. J. Leg. Med.* 109 (1996) 186–194.
- [18] R. Sparkes, C. Kimpton, S. Gilbard, P. Carne, J. Anderson, N. Oldroyd N, A. Urquart, P. Gill, The validation of a 7 locus multiplex STR test for use in forensic casework. II. Artefacts, casework studies and success rates, *Int. J. Leg. Med.* 109 (1996) 195–204.

- [19] C.M. Strom, S. Rechitky, Use of nested PCR to identify charred human remains and minute amounts of blood, *J. Forensic Sci.* 43 (1998) 696–700.
- [20] C.M. Strom, S. Rechitsky, G. Wolf, Y. Verlinsky, Reliability of polymerase chain reaction (PCR) analysis of single cells for preimplantation genetic analysis, *J. Assist. Reprod. Genet.* 11 (1994) 55–62.
- [21] P.S. Walsh, D.A. Metzgar, R. Higuchi R, Chelex™ 100 as a medium for the simple extraction of DNA for PCR based typing from forensic material, *BioTechniques* 1 (1991) 91–98.
- [22] P.S. Walsh, J. Varanio, R. Reynolds, A rapid chemiluminescent method for quantification of human DNA, *Nucleic Acids Res.* 20 (1992) 5061–5065.