Forensic Science International: Genetics 4 (2009) 1–10

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Interpreting low template DNA profiles

David J. Balding [a,*], John Buckleton [b]

[a] Department of Epidemiology and Public Health, Imperial College, St Mary's Campus, Norfolk Place, London W2 1PG, UK
[b] ESR Private Bag 92021, Auckland, New Zealand

ARTICLE INFO

Article history:
Received 20 October 2008
Received in revised form 27 February 2009
Accepted 26 March 2009

Keywords:
Low copy number
LTDNA
Drop-out
Drop-in
Masking
Garside and Bates

ABSTRACT

We discuss the interpretation of DNA profiles obtained from low template DNA samples. The most important challenge to interpretation in this setting arises when either or both of "drop-out" and "drop-in" create discordances between the crime scene DNA profile and the DNA profile expected under the prosecution allegation. Stutter and unbalanced peak heights are also problematic, in addition to the effects of masking from the profile of a known contributor. We outline a framework for assessing such evidence, based on likelihood ratios that involve drop-out and drop-in probabilities, and apply it to two casework examples. Our framework extends previous work, including new approaches to modelling homozygote drop-out and uncertainty in allele calls for stutter, masking and near-threshold peaks. We show that some current approaches to interpretation, such as ignoring a discrepant locus or reporting a "Random Man Not Excluded" (RMNE) probability, can be systematically unfair to defendants, sometimes extremely so. We also show that the *LR* can depend strongly on the assumed value for the drop-out probability, and there is typically no approximation that is useful for all values. We illustrate that ignoring the possibility of drop-in is usually unfair to defendants, and argue that under circumstances in which the prosecution relies on drop-out, it may be unsatisfactory to ignore any possibility of drop-in.

1. Introduction

In 2003 James Garside and Richard Bates were convicted in London of the murder of Marilyn Garside, the estranged wife of James Garside. It was the prosecution's case that Garside had hired Bates to murder his wife. The initial conviction was successfully appealed in 2004, leading to a retrial in 2005 which was followed by an unsuccessful appeal in 2006 [1].

Mixed DNA profiles from the crime scene formed a key part of the evidence. The major component of these crime scene profiles (CSP) corresponded to the victim and was assumed to be from her (her full profile was available). In addition, up to eight minor-component alleles were identified. Bates was subsequently arrested and profiled, revealing that the 17 distinct alleles in his 10-locus profile (Table 1) included all 8 CSP minor-component alleles, as well as a further six that were masked by the victim. However three of Bates' alleles, one at locus D2 and two at D18, were not reported in any of the large number of electropherograms (epg) generated in the investigation, from different crime scene samples and different profiling runs under varying laboratory conditions.

The prosecution allegation that Bates was the origin of the CSP minor component implies that allelic drop-out must have occurred at D2 and D18. Drop-out arises when an allele that is carried by an individual contributing DNA to a sample is not reported in a DNA profile obtained from that sample. A related phenomenon, drop-in, occurs when trace amounts of DNA, for example from a crime scene environment or laboratory plasticware, generate one or more spurious alleles in the profile. It is rare for drop-out or drop-in to occur with good-quality samples not subject to degradation or inhibition, but they become more likely as the amount of DNA template is reduced or environmental exposure increases. In Bates, alleles not attributable to the defendant were evident in some crime scene profiles, but not in the most incriminating CSP described in Table 1. Under conditions in which drop-out is plausible, it is difficult to entirely rule out the possibility of drop-in and we allow for this possibility in our analyses below.

The prosecution in effect ignored locus D18 in computing its measure of evidential weight, which was equivalent to a "Random Man Not Excluded" (RMNE) probability, see [2] and Table 2. The widespread policy of ignoring loci showing no minor-component alleles is commonly thought to be neutral or conservative, a view sometimes supported by sayings such as "the absence of evidence is not evidence of absence." Below, we show that in the presence of masking it can be very unfair to defendants.

At D2, one allele corresponding to Bates was reported, but not his other allele. The RMNE calculation uses the "2*p* rule", under

* Corresponding author. Tel.: +44 20 7594 3309.
E-mail address: d.balding@imperial.ac.uk (D.J. Balding).

1872-4973/$ – see front matter © 2009 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.fsigen.2009.03.003

Table 1
Likelihood ratio (*LR*) for a CSP in R v Garside and Bates. The crime scene sample included ample DNA from the victim, Mrs Garside, and only the minor component not attributable to her is shown. Highlighted in bold are the three alleles of Bates that are neither masked by Mrs Garside nor appear in the CSP.

Locus	Mrs Garside	Bates	CSP: minor component	Likelihood ratio (*LR*) a = 0.5				RMNE
				D = 0.05		*D* = 0.5		
				C = 0	*C* = 0.05	*C* = 0	*C* = 0.05	
D3	16,16	13,16	13	59	59	29	28	18
VWA	15,17	16,16	16	4.3	4.2	4.6	4.6	2.0
D16	11,12	11,12	–	2.6	2.6	1.6	1.6	1
D2	20,20	**19**,22	22	2.8	2.8	9.6	9.3	11
D8	12,13	8,13	8	24	24	16	16	12
D21	30,32.2	30,31.2	31.2	10	10	6.2	6.1	4.5
D18	14,14	**12,15**	–	0.056	0.056	0.76	0.76	1
D19	12,14	12,15	15	4.4	4.4	3.2	3.2	2.6
THO1	9.3,9.3	7,7	7	5.4	5.4	5.5	5.4	2.3
FGA	23,25	21,21	21	6.7	6.7	5.9	5.8	2.3
Product				4.2M	4.0M	17M	14M	0.33M
a = 1				4.2M	3.9M	12M	9.7M	

which the allele frequency estimate is doubled for a single reported allele when drop-out is possible. This rule has previously been shown to be non-conservative [3] particularly when the drop-out probability is low. Here we show that the problem is more severe in the presence of masking, and that it can extend to the high-drop-out setting.

The three alleles of Bates that were discordant with the prosecution hypothesis were referred to in the judgment as "voids". The judge in the second trial described the choice facing the court in the following terms:

> "What are the consequences of the impossibility of assigning a statistical weight to the voids? The alternatives are to exclude the evidence entirely or to admit it subject to an appropriate warning to the jury of the limitations of the evidence, and particularly highlighting the fact that although what was found was consistent with Bates' DNA profile, the voids at D2 and D18 in particular may have contained an allele or alleles, the presence of which would have been wholly exculpatory. In arriving at the correct conclusion it is important to remember that scientific evidence frequently only provides a partial answer to a case, or to an issue in a case. However, the test of admissibility is not whether the answer is complete, but whether science can properly and fairly contribute to the matter in question. . .".

At the final appeal the court ruled:

> "We can see no reason why partial profile DNA evidence should not be admissible provided that the jury are made aware of its inherent limitations and are given a sufficient explanation to enable them to evaluate it."

Table 2
Summary of *LR* and RMNE formulae for single-locus crime scene and suspect profiles when there is a single contributor, assuming that common drop-out and drop-in probabilities, *D* and *C*, apply to all alleles. CSP = crime scene profile; RMNE = Random Man Not Excluded.

CSP	Profile of *s*	RMNE probability	*LR*	RMNE typically unfair?
AB	AB	$2pApB$	Eq. (3)	Yes (if C > 0)
A	AB	$2pA$	Eq. (2)	Yes
	AA		Eq. (2) but $\bar{D}_2$ replaces $D\bar{D}$ in numerator	No
Null	AB	1	Eq. (4)	No
	AA		Eq. (4) but D_2 replaces D^2 in numerator	Yes

It is welcome to hear from the judiciary that evidence interpretation does not have to be perfect to be admissible. However, the quoted statements are worrying in several respects. Talk of "voids" that may have contained an allele suggests an assumption that drop-out has occurred, whereas there may be no "missing" allele at these loci: the alleles of the true culprit could have been masked by the victim's alleles. In the presence of masking we cannot know if any alleles have dropped out, and if so which one(s). It is therefore dangerous to speak of "voids" without some form of qualification, such as "according to the prosecution case". Since such qualification is awkward to use frequently, we propose the term "discordant alleles" to indicate alleles that are not consistent with the prosecution case.

The Garside and Bates case is far from unique in grappling with difficulties associated with interpreting low template DNA (LTDNA) profiles with discordant alleles. The profiles shown in Table 3 arose in an ongoing US-based case (hereafter US case). The observation of more than two alleles at two loci in the CSP indicates the presence of DNA from at least two individuals, and the absence of reportable alleles at five loci indicates that substantial drop-out affects the profiles of all contributors. A known individual V is expected to be a contributor of DNA to the mixture and the issue at stake is whether or not *s* is another contributor. Although we cannot be sure that there are only two contributors, the low number of observed alleles suggests that this is the case [4] and we proceed under this assumption.

The prosecution's RMNE calculation ignored the eight loci at which no CSP alleles could be attributed to *s*, while the remaining seven loci were given equal weight, even though at four of these loci the alleles attributable to *s* could also have come from V. Neither of the two homozygote alleles in the profile of *s* appeared in the CSP, yet the two homozygous alleles of V did appear. We argue that failure to observe the homozygotes tends to exclude *s* as a potential contributor, whereas the RMNE calculation ignored both loci. Indeed, in an RMNE calculation evidence can never favour a defendant. We present below alternative calculations that lead to very different conclusions from those made by the prosecution in this case.

We agree with [5] that there does not yet exist a fully satisfactory analysis of DNA profiles obtained from LTDNA samples, and so approximations and simplifying assumptions are required. However we are concerned that some approximations and assumptions being used in courts have not been adequately justified. Relevant theory for the interpretation of LTDNA profiles based on the likelihood ratio (*LR*) has been published [5–7]. Software for computing *LR*s in complex scenarios has been described [8], but is not widely available for courtroom use. We believe that the *LR* forms the basis of the most

Table 3
Likelihood ratio (*LR*) calculation for the CSP in the two-person mixture crime sample (US case). The *LR* contrasts the hypothesis that v and *s* were the sources of the CSP, with the hypothesis that v and *i* were the sources, where *i* is an individual unrelated to both v and *s*. The "Multi-source RMNE" calculation only uses loci at which the CSP includes at least one allele in the profile of *s*. The "Single-source RMNE" is as described Table 2 using only the three CSP alleles that cannot be attributed to v.

Locus	Known contributor v	Suspected contributor *s*	CSP	*LR*: *D* = 0.5 a = 0.5, *C* = 0.05	Multi-source RMNE	Single-source RMNE
D8	13,16	12,13	12,13,16	4.5	4.0	3.4
D21	28,30	29,29	28	0.38	1	1
D7	8,10	9,10	–	1.1	1	1
CSF	8,10	10,11	–	1.1	1	1
D3	14,16	16,17	16	1.4	4.3	1
THO1	7,7	9.3,9.3	7	0.38	1	1
D13	11,13	8,12	–	1.1	1	1
D16	12,13	11,12	12,13	0.93	4.0	1
D2	19,24	17,25	24	0.84	1	1
D19	12,13	13,15	12,13	1.1	6.5	1
VWA	18,20	19,20	18,20	1.4	18	1
TPO	9,9	11,12	9,11	1.3	7.0	2.0
D18	13,15	12,17	–	1.1	1	1
D5	8,12	11,13	8,11,12	0.88	1.7	1.2
FGA	21,22	22,24	–	1.1	1	1
Product				2.0	96K	8.3

rational approach to evidence evaluation [9,10], and in particular the *LR* approach has very significant advantages for complex evidence interpretation problems, such as those involving discordant alleles [11–16]. In our experience (in several English-speaking countries), appropriate *LR*-based methods do not always underpin courtroom practice in such complex cases, when the approach is at its most advantageous.

In this paper we seek to synthesize previous work and advance the *LR*-based methodology for LTDNA profiles, to highlight its advantages and to illustrate its impact for the two cases outlined above. We hope thereby to encourage the adoption in courts of an *LR*-based approach, but we do not seek here to prescribe to forensic practitioners exactly how they should present evidence, nor which approximations should be used in court. The important issue of a null minor component in the presence of masking has not been explicitly addressed in the literature, even though methods for dealing with it are straightforward extensions of existing methods. Other novel aspects developed here involve drop-out at homozygote alleles, and uncertainty about whether a peak is allelic or stutter or masking. We also provide freely-available software that underpins the analyses of this paper and allows researchers to implement and to extend or alter our formulae.

2. Background to LTDNA interpretation

DNA profiling from small amounts of DNA sometimes uses additional PCR cycles to increase amplification and hence compensate for the low starting template, and this approach is sometimes labelled low copy number (LCN) profiling. Because of ambiguity as to whether or not LCN refers to a specific profiling technique, the broader term LTDNA is used when there is limited DNA template, whether or not additional PCR cycles or other enhancement techniques are employed. In addition to drop-out and drop-in, LTDNA profiles are sometimes characterised by low epg peak heights, exaggerated peak imbalance and large stutter artefacts (Fig. 1). Typically labs employ a "reporting threshold", often 25–50 RFU, below which any apparent peaks are not reported [12] and an "LTDNA threshold", often 150–300 RFU, above which drop-out is excluded, so that a single allele peak is assumed to correspond to a homozygote [5,16]. In what follows, we assume that all single peaks from an unknown source lie between these two thresholds. If replication of an LTDNA profiling run is possible, it provides an opportunity to reduce uncertainty arising from stochastic effects. We will return to this issue below, but for now we will take the viewpoint of a court that is presented with a single or consensus profile, and is not given details of individual replicate runs if they exist.

Computing an *LR* requires a probability model to describe the real-world processes underlying an LTDNA profile. However, LTDNA profiling is complex, and there are many plausible probability models, and hence many different *LR*s can be proposed, each with some justification. Inevitably, fully realistic models of all sources of variation are unachievable, and approximations must be made and assessed [15,16].

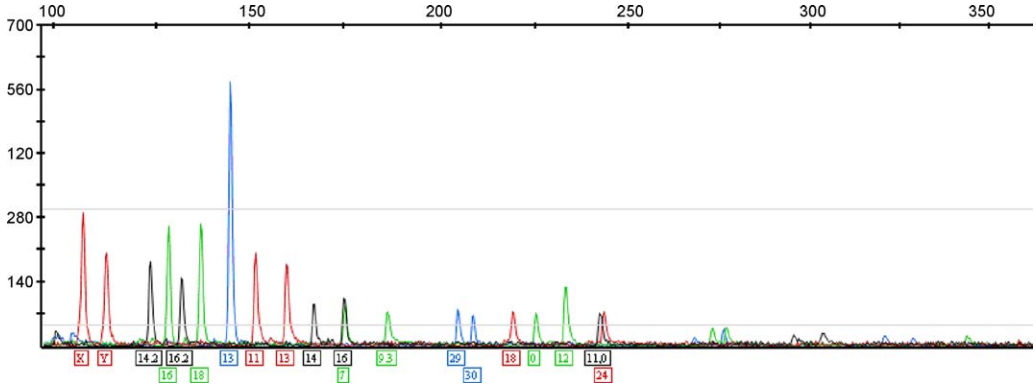

Fig. 1. An electropherogram (epg) showing several peaks close to the reporting threshold of 50 RFU (lower grey line). The highest peak, labelled 13, has no partner but since it is above the LTDNA threshold of 300 RFU (upper grey line) it is assumed to represent a 13,13 homozygote. In contrast, the genotypic designation 11,0 for a low peak allows the possibility of any other allele, including another 11, in the genotype of the contributor.

2.1. Drop-out and drop-in probabilities

We will write D for the probability of drop-out of one allele of a heterozygote, and we assume that drop-out is independent across the two alleles of a heterozygote and also across loci. The value of D at a locus could depend on the peak height if a single peak is observed, and could increase with allele length, because degradation affects longer alleles more than shorter alleles. Here we follow previous authors [8,16] and assume that D is constant, but what follows is readily extended to allow D to vary over alleles.

Some theory and data are emerging to assist in the task of choosing a suitable estimate for D [8,27], but we agree with [5] that "judgment calls are still required by the reporting scientist". Assessments of D, whether quantitative or informal, can be based on the stochastic phenomena displayed in the epg over all loci, and informed by the results of laboratory trials that employ LTDNA profiling on samples of known origin and mixture proportions. It will usually be impossible to specify a precise value, and instead a range of plausible values should be considered. Fortunately, as noted by others [8,17], the *LR* often does not vary greatly over the range of plausible values for D. If a point estimate is required, because the whole-profile *LR* can involve high powers of D, and because the average of a power is larger than the power of the average, appropriate point estimates for D should typically be towards the high end of the plausible range.

We write D_2 for the homozygote drop-out probability. It is natural to seek to express D_2 in terms of D. One possibility is to assume that the allele on each homologous chromosome drops out independently with known probability D, so that $D_2 = D^2$ [2,8]. This is likely to overstate D_2 if both alleles can generate partial signals that combine to reach the reporting standard, whereas each individual signal would fail to reach this standard. For example, suppose that under the prevailing conditions a heterozygous allele would generate a peak with mean height 40 RFU and that $D = 0.7$. If the signal from a homozygote is a superposition of two signals, with expected height close to 80 RFU, the corresponding drop-out probability is likely to be much less than $D^2 = 0.49$. In 40 laboratory trials of 10 pg DNA samples, 9 instances of homozygote drop-out were observed, from which D_2 0.225. The value of D estimated from 160 heterozygote alleles at the same three loci was 0.66, so D^2 0.44 and hence D_2 $1/2D^2$ (Simon Cowen, personal communication). On the other hand, at very low template levels the epg peak for each allele may be closer to "all or nothing", in which case the independence assumption may be reasonable. Here, we assume that $D = aD^2$. The appropriate value of a should be chosen on the basis of laboratory trials. Typically we expect that $0 << a$ 1, and we use $a = 1/2$ and 1 in the numerical examples below.

2.2. Drop-in

Here we assume that at most one drop-in occurs per locus, with probability C, but this restriction is easily relaxed. We also exclude the possibility of a drop-out followed by drop-in of the same allele. Following [8], we treat drop-ins at different loci as mutually independent, and independent of any drop-outs. Multiple drop-ins in a profile may, depending on the values of C and D, be better interpreted as an additional unknown contributor. Typically we use C 0.05 below, reflecting that drop-in is rare [8], and such low values of C automatically penalise a prosecution hypothesis requiring multiple drop-in events.

2.3. Allele frequencies and shared ancestry

In the examples below, to make some allowance for sampling uncertainty we use $P_X = (n_X + 4)/(n + 4)$ if s is a XX homozygote, $P_X = (n_X + 2)/(n + 4)$ for each allele of a heterozygous s and $P_X = n_X/(n + 4)$ if allele X is not included in the profile of s, where n_X is the count of X in a database of size n; see [18] for a discussion and [19] for alternative approaches.

We do not here consider alternative sources of the CSP that are direct relatives of s, even though we regard this as an important possibility in practice. To allow approximately for the effects of remote shared ancestry between s and other possible sources of the CSP, we replace the allele proportion P, adjusted for sampling error as described above, with $(F_{ST} + (1 \quad F_{ST})P)/(1 + F_{ST})$ for each allele when s is heterozygote, and with $(2F_{ST} + (1 \quad F_{ST})P)/(1 + F_{ST})$ in the homozygote case [18,20]. For alleles not in the profile of s, the allele proportion P is replaced with $(1 \quad F_{ST})\,P/(1 + F_{ST})$. See [20] for a discussion of the appropriate value of F_{ST}. We recommend using a value of at least 0.01 in all cases, and use 0.02 in numerical examples below. In the numerical examples below, genotype proportions are obtained from allele proportions assuming Hardy-Weinberg Equilibrium.

3. LR for a single contributor

Explicit formulae can be complicated, but logically the computation of the *LR* when drop-out and drop-in are possible is straightforward [7]: the denominator involves a summation over every genotype for the unknown contributor, while the numerator consists only of the term in the denominator corresponding to the profile of the alleged contributor, s. To illustrate notation and concepts, consider first a single CSP that shows only an A allele, whereas the genotype of the alleged contributor s, obtained from a good-quality reference sample that is assumed to be profiled without error, is AB. This would normally constitute an exclusion of s, but if $D > 0$ then the prosecution hypothesis, H_s, that s is the source of the CSP, can be sustained. Writing "BB" for "has profile", and assuming that $\Pr(s \text{ BB} AB)$ does not depend on whether H_s or H_i is true, the *LR* can be written

$$LR = \frac{Pr(\text{CSP} \quad A|s \quad AB; H_s)}{Pr(\text{CSP} \quad A|s \quad AB; H_i)},$$

in which H_i is the alternative hypothesis that i, unrelated to s, is the source of the CSP. Large values of the *LR* imply strong evidence against s, provided that all his relatives can be excluded as alternative possible contributors to the CSP; otherwise the *LR*s for various relatives of s should also be provided.

Under H_s we have observed a non-drop-out allele (A), a drop-out (B), and no drop-in, so that the numerator is $\bar{D}D\bar{C}$, where we introduce the notation $\bar{X}$ for $1 \quad X$. In the denominator, if no drop-in has occurred then i can have genotype AA or AQ, where Q denotes any allele other than A, whereas if the A allele has dropped in then i can be QQ or QQ', where Q' is neither A nor Q. The *LR* is then:

$$LR = \frac{\bar{D}D\bar{C}}{(\bar{D}^2 P_{AA} + \bar{D}DP_{AQ})\bar{C} + (D^2 P_{QQ} + D^2 P_{QQ'})C P_A} \quad (1)$$

where P_{XY} and P_X denote the probabilities of genotype XY and allele X, given the alleles observed in the CSP. Note that the denominator of (1) does not depend on the profile of s except through the P_{XY}, for which dependence can arise through possible shared ancestry [18,20]. Eq. (1) is similar to the formula set out set out in Table 1 of [8], except that we distinguish homozygote drop-out as a special case, and also we have corrected an error in the first term of the denominator.

In (1), LR ! 0 as D ! 0, corresponding to strong evidence for exclusion of s if drop-out is very unlikely. This is intuitive since, in this scenario, a case against s cannot be sustained unless drop-out has occurred. Several authors [3,11] have noted that the '2p rule' overstates the correct *LR* substantially for low values of D, but Fig. 2 shows that the effect can be more substantial than previously noted, especially in the presence of masking, discussed further below. It has

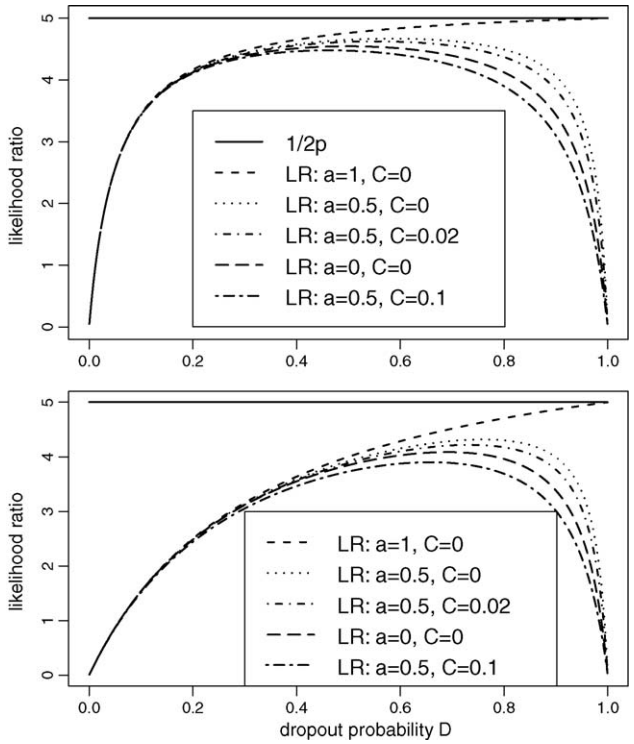

Fig. 2. Single-locus likelihood ratios (*LR*) when CSP = *A* and *s* = *AB*, without masking (top, Eq. (1)) and when there are two masking alleles (bottom, Eq. (2)). The locus is assumed to have ten equally frequent alleles (no F_{ST} or sampling adjustment), and a (a in legend box) is defined by $D_2 = aD^2$, where D_2 denotes the homozygote drop-out probability. The drop-in probability is denoted *C*.

not previously been noted that $LR \to 0$ as $D \to 1$, which occurs in (1) if $a < 1$. This arises because, if drop-out is very likely, observation of a single *A* allele suggests that the contributor to the CSP has genotype *AA*, so once again the evidence points away from *s* who is *AB*.

One important conclusion, here and in what follows, is that if drop-out is invoked to sustain the prosecution case then estimation of drop-out probabilities cannot be avoided. DNA-based prosecutions that rely on drop-out and do not explicitly estimate plausible ranges for the drop-out rate parameter are, in our view, defective.

4. LR for multiple contributors, one unknown

The two cases introduced above both involve mixed profiles with a known contributor, V, accepted by both prosecution and defence. Alleles from V can mask alleles from the unknown contributor in each case, either directly or via stutter peaks. We initially ignore the latter, but return to them below.

4.1. One minor-component allele

Let *M* denote the set of major component alleles from known contributors, while *s* = *AB*, the CSP minor component is *A*, and neither *A* nor *B* lies in *M*. The numerator is the same as in (1), but under H_i we must consider the possibility that alleles of *i* lie in *M* and hence are masked. Again using a derivation similar to that of [7], the *LR* is

$$LR = \frac{Pr(\mathrm{CSP} \mid s, AB, M, H_s)}{Pr(\mathrm{CSP} \mid s, AB, M, H_i)} = \frac{D\bar{D}\bar{C}}{\bar{D}_2 P_{AA} + D\bar{D}P_{AQ} + \bar{D}P_{AM}(\bar{C} + (D_2 P_{QQ} + D^2 P_{QQ'} + P_{MM} + DP_{MQ})C P_A} \quad (2)$$

where *Q* and *Q*′ denote distinct alleles that are both neither *A* nor in *M*. As expected, (2) reduces to (1) if *M* is null. Once again, $LR \to 0$ in (2) both as $D \to 0$ and, if $a < 1$, as $D \to 1$.

If the genotype of *s* is now *AA* while the CSP minor component is still *A*, the numerator of (2) becomes $\bar{D}_2\bar{C}$, while the denominator remains unchanged (the values of the P_{XY} change because of the changed effect of shared ancestry of *i* with *s*). A possible approximation is to assume that $D = C = 0$, so that

$$LR \approx \frac{1}{P_{AA} + P_{AM}}.$$

This would be acceptable for a good-quality profile, but in the LTDNA setting it is usually unfair to defendants. The RMNE approximation $LR = 1/2P_A$ is usually conservative in this setting (Fig. 3, top).

4.2. Two minor-component alleles

If *s* = *AB* and the minor component CSP = *AB*, it may seem natural to use the RMNE value $LR = 1/P_{AB}$. However, masking and the possibility of drop-in can make this value unfair to defendants. In fact,

$$LR = \frac{\bar{D}^2\bar{C}}{\bar{D}^2\bar{C}P_{AB} + (\bar{D}_2 P_{AA} + D\bar{D}P_{AQ} + \bar{D}P_{AM})(CP_B + (\bar{D}_2 P_{BB} + D\bar{D}P_{BQ} + \bar{D}P_{BM})CP_A}; \quad (3)$$

where *Q* denotes an allele that is not in {*A*,*B*} ∪ *M*. Only the first term of the denominator of (3) is included in an RMNE calculation. Fig. 3 (middle) shows a large effect of the additional terms when *D* is not small, but masking has little effect.

4.3. Null minor component

The effect of masking can be very important when the minor component CSP is null, especially if *D* is small. For *s* = *AA* the *LR* is

$$LR = \frac{D^2}{P_{MM} + DP_{MQ} + D_2 P_{QQ} + D^2 P_{QQ'}} \quad (4)$$

where *Q* and *Q*′ are any alleles not in *M*. If *s* = *AA*, then D_2 replaces D^2 in the numerator. In either case, (4) favours the defence hypothesis when there are masking alleles, strongly so unless *D* is large (Fig. 3, bottom). Further, even if *M* is also null so that the first two terms in the denominator of (4) are zero, $LR = 1$ unless $a = 1$. If $a < 1$, the observation of a null CSP supports the conclusion that *i* is heterozygous and so $LR > 1$. Conversely, if *s* ≠ *AA* then D_2 replaces D^2 in the numerator of (4) and $LR < 1$ when $a < 1$. In contrast, the RMNE value is always 1, irrespective of masking and the genotype of *s*.

5. Multiple unknown contributors

The principles that we have used to construct *LR*s, based on those set out in [7] and implemented in [8], extend readily to multiple unknown contributors. The denominator now involves summations over the possible genotypes for each unknown contributor. The result can be approximated by averaging over repeated simulations of their genotypes, which can reduce the computational burden at the cost of some approximation. However, for one or two unknown contributors, the exact value is fast to compute and so simulation-based approximation is not necessary. Our two example cases are assumed to involve only one unknown contributor, and to keep the exposition as simple as possible we limit attention to this case.

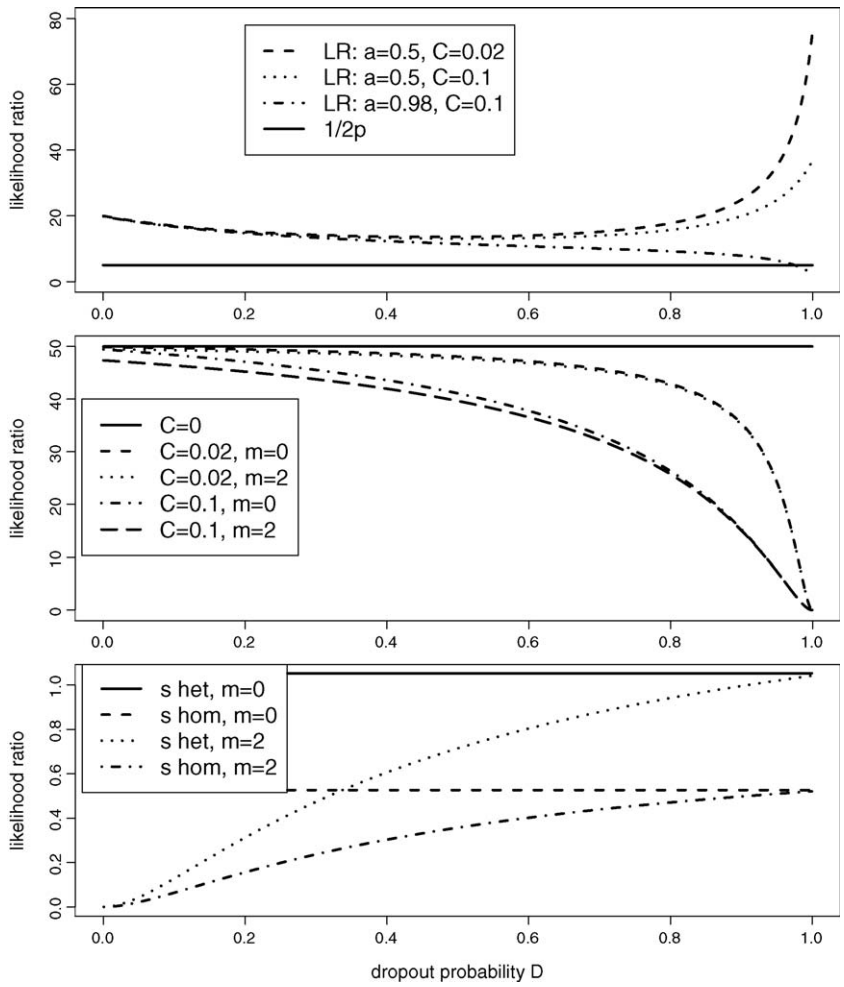

Fig. 3. Top: Single-locus *LR*s when CSP ■■*A* and *s* ■■*AA* and there are two masking alleles. Middle: Single-locus *LR*s when CSP ■■*AB* and *s* ■■*AB*, without masking ($m = 0$) and when there are two masking alleles ($m = 2$). Bottom: Single-locus *LR*s for a null CSP, with and without masking, in each case with a = 0.5 and $C = 0.02$. Other details as for Fig. 2. When $m = 0$ the *LR* is constant, 1.053 when *s* is heterozygous, and 0.546 when *s* is homozygous.

6. Uncertain allele calls

6.1. Stutter

Stutter peaks are usually located one repeat unit shorter than the allelic peak, but stutters at minus two or plus one repeat units are sometimes observed. Interpretation guidelines often specify a threshold height, as a proportion of the allelic peak height, above which a peak in a stutter position is called as an allele, and otherwise is regarded as stutter [12]. Inevitably, such thresholds do not always correctly distinguish allelic from stutter peaks, as noted by [11] at Recommendation 6. Calling a peak as stutter tends to favour *s* if he has the allele, but if not the apparent stutter could mask a minor-component allele that, if reported, would tend to exonerate *s*. Conversely, if a possible stutter is called as allelic, this favours *s* if he has no allele at that position, but otherwise is unfair to *s* because the allele is counted as evidence against him with no allowance for the possibility that it could be just a stutter artefact.

Stutter is most problematic when there are multiple contributors to the CSP, but it can also cause problems for single-contributor LTDNA profiles. For example, a low peak at allele $X \quad 1$ and a higher peak at *X* could be consistent with an *XX* contributor and substantial stutter, or an $(X \quad 1)X$ contributor with unbalanced peaks, or an *XY* contributor with both stutter and drop-out of the *Y*. Gill et al. [5] proposed upper and lower bounds on the *LR* obtained by considering the possible assignments of a peak in a stutter position. Below we propose weighted averages of *LR* numerator and denominator, similar to [21].

6.2. LTDNA masking

In the US case the known contributor's DNA is also at low template and subject to drop-out, and so masking is uncertain: the appearance in the epg of a peak in a masking position could partly reflect an allele of the unknown contributor. The simplest way to deal with uncertain masking, usually erring in favour of *s*, is to include in the masking set *M* every above-threshold peak that either corresponds to an allele of a known contributor or is in a stutter position to such an allele. Below we develop a more satisfactory quantitative analysis of uncertain masking, by assigning probabilities that an allele is masking.

6.3. Near-threshold peaks

An allele of *s* that is not reported in the CSP tends to exonerate, and this effect can be large when drop-out is unlikely. Intuitively, the effect is tempered if there is an apparent peak in the epg at the appropriate position, but which fails to achieve the reporting threshold. An ideal analysis would take into account all available information in a continuous way, without reference to thresholds. In practice it is extremely difficult to take into account all the information contained in a complex epg, and no algorithm for this

currently seems to be available although there have been promising developments [22,23].

In principle, a full *LR* analysis of a sub-threshold peak requires probabilities of observing the peak if the unknown contributor does, and does not, have an allele at this position. Intuitively, if there is a peak at *A* in the CSP epg that is well above the reporting threshold, while no sub-threshold peaks are obvious, then the locus should be exculpatory if *s* is *AB*. If, however, the *A* peak is only just above threshold, while a peak at *B* is clearly visible just below the threshold, then the locus might informally be regarded as inculpatory, even though one allele is not reportable. Our analysis at (1) does not discriminate between these two possibilities and makes no direct use of the heights of reported peaks or any apparent sub-threshold peaks.

6.4. Probabilistic allele designations

Towards a more satisfactory analysis of uncertain stutter, masking and near-threshold allele calls, we propose that, rather than just make a single judgment "call" of an epg to designate peaks as stutter, masking or allelic, an expert assesses probabilities for alternative possible designations. In computing the *LR*, both numerator and denominator must be averaged over the possible calls, weighted according to their respective probabilities (NB this is not equivalent to a weighted average of *LR*s). Most laboratories would have information available to help assess such probabilities. Typically the baseline noise has been quantified in general, in order to establish the drop-out threshold, and further information could come from assessing the baseline for the epg at hand.

For example, if *s* is *AB* and the CSP peak at allele *A* is above the reporting threshold, while that at *B* is slightly sub-threshold, an expert may assess probabilities of 1 and 0.6 that the peaks at *A* and *B* are allelic. The numerator of the *LR* is then 0.4 times the numerator of (2) plus 0.6 times the numerator of (3). Similarly, the denominator is 0.4 times the denominator of (2) plus 0.6 times that of (3). Fig. 4 illustrates the effect of averaging over two equi-probable allele calls.

7. Replication

If replicate epgs are available for the same crime sample, a common approach is to decide on a consensus profile and then to compute an *LR* for it, without explicitly modelling the details of the replication. Another possibility is to use the results of replicate epgs to inform the assessment of allele calling probabilities, described above.

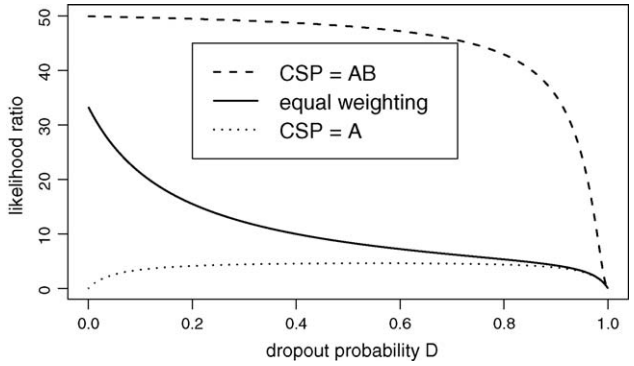

Fig. 4. Single locus *LR*s as a function of *D* for a single-contributor profile for which the alleged contributor has genotype *AB*. Dashed curve: CSP is called as *AB*; dotted curve: CSP is called as *A*; solid curve: *LR* obtained by averaging both numerator and denominator over the two possible calls corresponding to certainty about allele *A* but only a 0.5 probability that the unknown contributor has allele *B*. We use $a = 0.5$ and $C = 0.02$; other details as for Fig. 2.

If the replicates can be assumed to be independent, it is straightforward to extend our *LR* formulae to simultaneously analyse all the replicates [6,7]. The denominator is still a sum over possible genotypes for the unknown contributor, *i*. Every term in this sum includes a probability for that genotype, times the product of probabilities for the drop-outs and drop-ins required in all replicates in order to reconcile this genotype with the observed CSP alleles. The numerator consists of the term in the denominator for which *i* has the same genotype as *s*.

We applied this approach to the example in Table 2 of [8], which has two replicate epgs for a CSP that we assume has two known and one unknown contributor, all subject to drop-out. Following [8], we assumed a common *D* for all contributors. We also modified our formulae to allow up to two drop-ins per locus, as this was required under H_s at locus D18 (we allowed for the possibility at every locus). Our result has features in common with that shown in Fig. 1 of [8]: the *LR* is maximised at very low values of *D* (around $D = 0.005$) and declines rapidly with increasing *D*. However, our computed *LR* is several orders of magnitude smaller: our maximum value is around 10^7 versus 10^{12} in [8]. The source of that discrepancy is currently under investigation. In contrast, the *Locomation* software described in [8] gave similar results to those reported below for the Bates case (James Curran, personal communication).

8. Results

8.1. Garside and Bates

The most incriminating of the many CSPs presented in evidence is summarised in Table 1. The prosecution in effect presented an RMNE computation (see Table 2) applied to the CSP alleles not masked by the known contributor. In general, this approach is favourable to defendants if drop-out is not required under the prosecution case, but it does not use any value for *D* and we have noted above that it can be unfair to defendants in the presence of discordant alleles. The reported value was 0.61 million (M). We perform a similar calculation (Table 1), but because we have used different F_{ST} and sampling adjustments, our corresponding value is 0.33M.

Fig. 5 shows single-locus and overall *LR*s as functions of *D*, using allele frequencies reported in [24]. Because of the three discordant alleles, the overall *LR* is zero (exclusion of Bates) if $D = 0$. However, from Fig. 5 we see that the *LR* rapidly increases with *D*. Using our *LR* calculations, the RMNE value is in fact conservative for $0.02 < D < 0.98$. Given the small number of alleles in the minor component CSP and the stochastic nature of the results of repeat profiling runs, it would not have been difficult to argue that *D* lies in this range. But in the absence of any such argument, the claim that the *LR* was conservative was not adequately justified.

The RMNE value of 1 for locus D18 is not conservative for any value of *D*, and Table 1 illustrates that the non-conservative error can be small or large depending on the value of *D*. This conclusion conflicts with [12], who asserted in the context of the Bates case that drop-out of both alleles "must be regarded as neutral", but did not consider the effect of masking alleles that were present in this case. The prosecution in Bates defended its assumption by arguing that drop-out is more likely at locus D18 than at other loci because of its high molecular weight, but provided no quantitative backing for this argument. We have here assumed the same drop-out and drop-in probabilities at each locus.

At locus D2 the "2*p* rule" gives 11 (Table 1), which is non-conservative compared with an *LR* of 2.8 ($D = 0.05$) or 9.3 ($D = 0.5$). In fact, a peak was observed at the 19 allele, but did not reach a height such that it could be confidently distinguished from stutter artefacts generated by Mrs Garside's two 20 alleles. Treating the 19

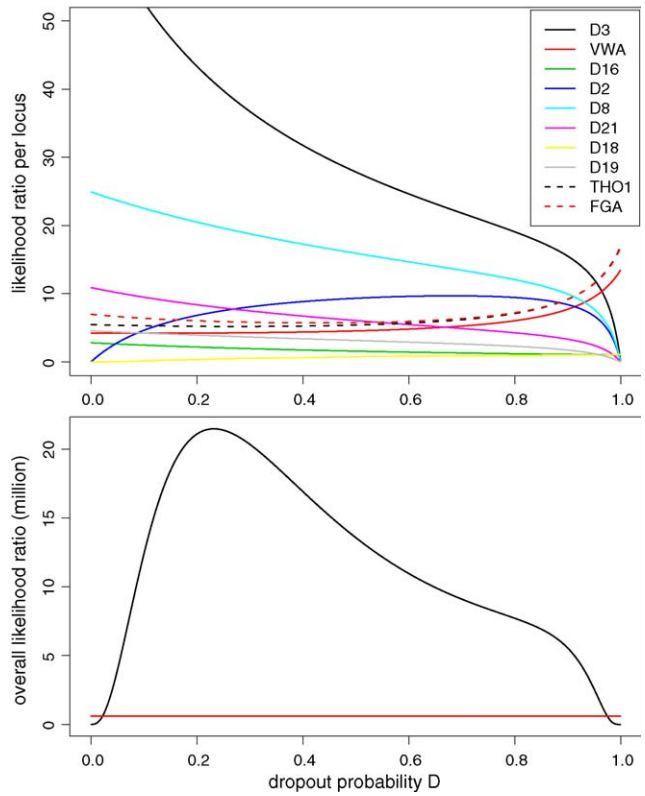

Fig. 5. Top. Single-locus *LR*s as a function of single-allele drop-out probability *D*, for the CSP of Table 1, using a = 0.5 and $C = 0.05$ and $F_{ST} = 0.02$. Bottom. Black curve: whole profile *LR* (=product of single-locus *LR*s). Red line: RMNE value reported in court.

allele as stutter was favourable to Bates. We recalculated the *LR* at D2 allowing equal probabilities for the peak at allele 19 to be (i) allelic and (ii) stutter. With $D = 0.5$ and $C = 0.05$, the *LR* increases from 9.3 to 13, and the profile *LR* increases to 19M.

Because all three of Bates' homozygous alleles appeared in the CSP, the lower probabilities of homozygote drop-out when a = 1/2 corresponds to slightly stronger evidence against him than when a = 1 (Table 1, final row).

The crime scene samples were analysed numerous times under various conditions and in none of these replicates were the 12 or 15 allele reported at D18. Repeat epgs were usually not strict replicates because of varying conditions and samples, but for illustration let us assume that there were two replicates both giving the same, null, result (i.e. the CSP includes only the 14 allele of v). Assuming independence of replicates, in place of (4) the *LR* is

$$LR = \frac{D^4}{P_{14;14} + D^2 P_{14;Q} + D^4 P_{QQ'} + D_2^2 P_{QQ}}.$$

This would represent much stronger evidence in favour of *s* than if just a single replicate was considered (and similarly the inculpatory loci would also become more inculpatory).

If, hypothetically, the 12 allele had been reported from one replicate, we would have

$$LR = \frac{D^3 \bar{D} \bar{C}}{(P_{14;14} + D^2 P_{14;Q} + D_2^2 P_{QQ} + D^4 P_{QQ'})\bar{C} P_{12} + (D_2 \bar{D}_2 P_{12;12} + D\bar{D} P_{12;14} + D^3 \bar{D} P_{12;Q})\bar{C}};$$

where Q and Q' are neither 12 nor 14. If C and D are both small, we have approximately

$$LR \approx \frac{D^3}{(P_{14;14} + D^2 P_{14;Q})C P_{12} + D_2 P_{12;12} + D P_{12;14}}.$$

For example, if $C = 0.02$, $D = 0.05$, and a = 1/2, and the locus has 10 alleles with genotype proportions equal to 0.01 for each homozygote and 0.02 for each heterozygote, then the *LR* calculated from the above two formulae are, respectively, 0.118 and 0.121. Thus, for these parameter values, the evidence still points away from *s* and towards *i*, despite one observation of a 12 allele.

8.2. US case

The prosecution proposed an RMNE probability of 1 in 96K for Caucasians (Table 3, "multi-source RMNE" column), based on the seven loci (D8, D3, D16, D19, VWA, TPO, D5) at which the CSP included an allele that could be attributed to the suspected contributor, *s*. Unlike the single-source RMNE used in Bates, the probability computed was for the profile of a random man to be included among all the CSP alleles at each locus, without distinguishing known from unknown contributors. For zero or one observed allele, the computation is the same in each approach. For two observed alleles, *A* and *B*, the multi-source RMNE value of $P_{AA} + P_{AB} + P_{BB}$ exceeds the value P_{AB} for the single-source RMNE. This approach may seem generous to defendants but it makes no allowance for drop-out when two or more alleles are observed, nor for drop-in. In particular, although v is an acknowledged contributor to the CSP and *s* is an alleged contributor, because the calculation makes no allowance for drop-out, neither v nor even *s* is included in the 1 in 96K proportion of the population evaluated by the multi-source RMNE. Thus the RMNE value bears little relevance for evidential weight, and it could be highly misleading to a court.

In fact, at four of the loci included in the RMNE computation, all the alleles in the CSP can be attributed to the known contributor v, and so these loci provide little if any evidence for *s* being a contributor to the CSP. Removing these four loci brings the RMNE value from 1 in 96K to 1 in 47. More appropriate is to recognise that v is a contributor to the CSP and use the single-source RMNE calculation that is based only on the (three) alleles in the CSP that could have come from *s* but not from v, giving an RMNE probability of about 1 in 8 (Table 3, final column). However, an *LR* calculation shows that even this much reduced value still overstates evidential strength, because of the five loci at which the evidence favours the proposition that *s* is not a contributor to the CSP: no RMNE can accommodate this possibility. In fact (Table 3 and Fig. 6), under our assumptions the LTDNA evidence only very weakly supports the prosecution case, with $LR < 3$ when a = 1/2. We used the California Caucasian allele frequencies reported by [25], with a sampling adjustment and $F_{ST} = 0.02$. We also assumed the same value of *D* for both sources of the CSP.

The five loci at which the CSP showed no alleles are all slightly incriminating under our analysis ($LR \approx 1.1$ when a = 1/2), because *s* is heterozygous at each of these loci, whereas an alternative contributor might be homozygous and hence less susceptible to drop-out. Our novel treatment of homozygote drop-out is potentially important here because *s* is homozygous at two loci yet neither of these alleles appears in the CSP. Fig. 6 (bottom) illustrates the sensitivity of the *LR* to choice of a, but even with a = 1 (most favourable to prosecution) we have $LR < 7$ for all *D*.

Our *LR* analysis treats the alleles of v that were observed in the CSP as masking any alleles of a possible alternative contributor. It might be argued that this is generous to the defendant, since the five CSP alleles shared by *s* and v do not count as evidence for the prosecution yet, because of possible drop-out of the alleles of v, some of these could in fact represent alleles from the unknown contributor. Indeed, only three of eight alleles shared by *s* and v did not appear in the CSP, which is a lower drop-out rate than for the heterozygote alleles of v not shared by *s* (10 drop-outs from 18

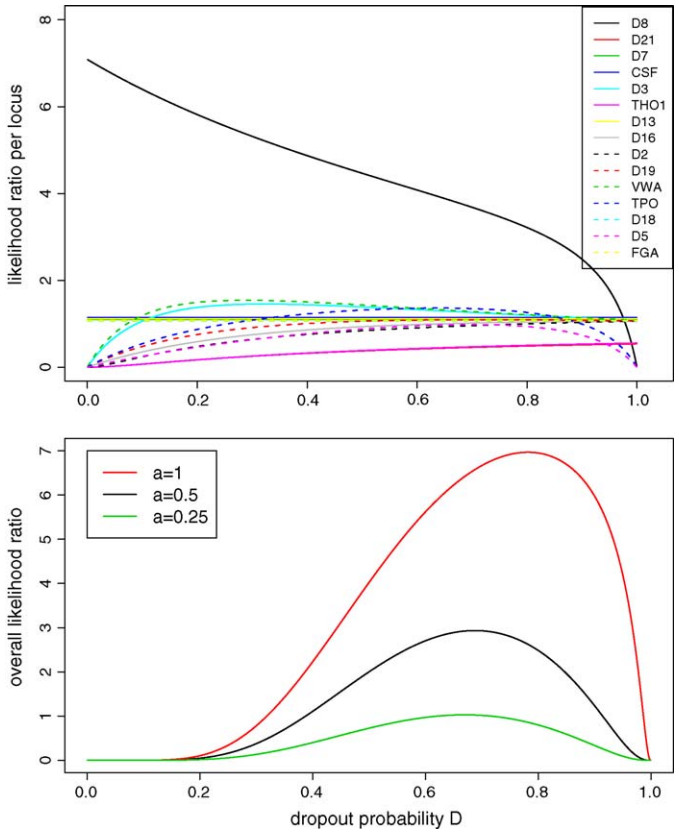

Fig. 6. Top. Single locus *LR*s as a function of *D* for the CSP of Table 3, using a = 0.5 and $C = 0.05$ and $F_{ST} = 0.02$. CSP alleles that are attributable to v are assumed to be masking. Bottom. Black curve: whole profile *LR* (=product of single-locus *LR*s). Red and green curves show the effect of varying the homozygote drop-out parameter a (see text for discussion).

alleles). Conversely, however, the nine CSP alleles attributed to v that are not shared by *s* might represent a contribution from a different contributor and so could be regarded as providing evidence for the defence. In accord with the principle that CSP allele calls should not refer to the profile of *s*, we recalculated the *LR* treating all the alleles of v that appear in the CSP as masking with probability 0.5 if v is heterozygous, and with probability 1 if v is homozygous. This has the effect of weakening the prosecution case: the overall *LR* of 2.0 in Table 3 is reduced to 1.2. Thus, attributing CSP alleles to v whenever possible, and ignoring the possibility that they could be from the unknown contributor, is here favourable to the prosecution.

Although our analysis has limitations as discussed above, we believe that it suffices to establish that the LTDNA evidence at best only very weakly supports the prosecution case. Further, an uncritical application of RMNE can very significantly misrepresent the evidence.

9. Software

One of us (DJB) has written R code (R is freely available with documentation at www.r-project.org) that computes single-locus *LR*s for a CSP with one unknown contributor, allowing for drop-out of one or both alleles, and drop-in of up to two alleles. The drop-out and drop-in probabilities can vary over loci but must currently be constant within a locus (it would not be difficult to alter the code to allow allele-specific probabilities). One version of the code allows for uncertain calls of stutter, masking, and near-threshold peaks, while another allows for replicate calls.

The software has been used to generate all the figures in this paper. It is available on request and has been checked extensively but comes with no guarantee of accuracy. It should be regarded as research-level software: we do not currently have the resources to test if fully rigorously, nor to make it widely accessible.

10. Conclusion

We have built on the contributions of previous authors to develop an approach to analysis of LTDNA profiles. In this approach, drop-out and drop-in probabilities are parameters that need to be assigned or estimated, as in [8] and its precedents [2,6,7]. An alternative approach would be to build into the *LR* calculation an automatic estimation of some of these parameters. Even better would be to model the entire epg, taking all peak heights into account without reference to thresholds. Although these approaches may have advantages in principle, they would be difficult to implement.

We allow uncertain allele calls for stutter, masking and near-threshold peaks by averaging over the possible calls when computing the numerator and denominator of the *LR*. Here we have limited attention to cases involving only one unknown contributor, but the principles extend to multiple unknown contributors [8]. We believe that the approach developed here overcomes many of the challenges presented by LTDNA evidence, and can be explained and defended in court.

We deplore uncritically ignoring loci that are discordant with the prosecution case, and recommend that drop-out and drop-in probabilities be assessed, and their implications be considered. We have also shown through argument and example that the Random Man Not Excluded (RMNE) approach to evidence evaluation has serious flaws in the LTDNA setting, and can potentially lead to serious misrepresentation of the evidence; we recommend that it not be used. Recently [26] have described an RMNE calculation allowing for up to a given number of drop-out alleles. They give no numerical comparison with an *LR* approach, and they acknowledge that their approach does not make full use of the evidence. In fact, for multi-source CSPs when multiple drop-out is allowed, the approach loses a large amount of evidential information. It also remains subject to the many criticisms of RMNE described above: evidence can never favour a defendant, and it is difficult to interpret for multi-contributor samples because no use is made of the suspect's profile.

Acknowledgments

Thanks to James Curran for assistance with comparisons of our software with LoComatioN, Peter Gill for supplying allele frequencies from [24], and Simon Cowen for making available unpublished data on homozygote drop-out. We also thank Jo-Anne Bright and SallyAnn Harbison for comments that improved this paper. Much of the paper was written while DJB was a scientific visitor to the Institut de Biologie de Lille, France.

References

[1] R v Garside, Bates, EWCA Crim 1395, 2006, Royal Courts of Justice, Strand, London.
[2] J.S. Buckleton, C.M. Triggs, S.J. Walsh, DNA Evidence, CRC Press, Boca Raton, Florida, 2004.
[3] J. Buckleton, C.M. Triggs, Is the 2*p* rule always conservative? Forensic Science International 159 (2006) 206–209.
[4] J.S. Buckleton, J.M. Curran, P. Gill, Towards understanding the effect of uncertainty in the number of contributors to DNA stains, Forensic Sci. Int. Genet. 1 (1) (2007) 20–28.
[5] P. Gill, et al., Interpretation of complex DNA profiles using empirical models and a method to measure their robustness, Forensic Sci. Int. Genet. 2 (2) (2008) 91–103.
[6] P. Gill, et al., An investigation of the rigor of interpretation rules for STR's derived from less that 100 pg of DNA, Forensic Science International 112 (1) (2000) 17–40.
[7] J.M. Curran, P. Gill, M.R. Bill, Interpretation of repeat measurement DNA evidence allowing for multiple contributors and population substructure, Forensic Science International 160 (2005) 47–55.

[8] P. Gill, A. Kirkham, J. Curran, LoComatioN: A software tool for the analysis of low copy number DNA profiles, Forensic Science International 166 (2–3) (2007) 128–138.
[9] B. Robertson, G.A. Vignaux, Interpreting Evidence—Evaluating Forensic Science in the Courtroom, John Wiley & Sons, Chichester, 1995.
[10] I.W. Evett, B.S. Weir, Interpreting DNA Evidence—Statistical Genetics for Forensic Scientists, Sinauer Associates, Inc, Sunderland, 1998.
[11] P. Gill, et al., DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures, Forensic Science International 160 (2006) 90–101.
[12] P. Gill, et al., National recommendations of the Technical UK DNA working group on mixture interpretation for the NDNAD and for court going purposes, Forensic Sci. Int. Genet. 2 (1) (2008) 76–82.
[13] N. Morling, et al., Interpretation of DNA mixtures—European consensus on principles, Forensic Sci. Int. Genet. 1 (3–4) (2007) 291–292.
[14] P. Stringer, et al., Interpretation of DNA mixtures—Australian and New Zealand consensus on principles, Forensic Sci. Int. Genet. 3 (2) (2009) 144–145.
[15] T. Tvedebrink, et al., Amplification of DNA mixtures—missing data approach, Forensic Science International: Genetics Supplement Series 1 (1) (2008) 664–666.
[16] P. Gill, R. Puch-Solis, J. Curran, The low-template DNA (stochastic) threshold—its determination relative to risk analysis for national DNA databases, Forensic Sci. Int. Genet. 3 (2) (2009) 104–111.
[17] J. Buckleton, C. Triggs, Dealing with allelic dropout when reporting the evidential value in DNA relatedness analysis, Forensic Science International 160 (2006) 134–139.
[18] D.J. Balding, R.A. Nichols, DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands, Forensic Science International 64 (1994) 125–140.
[19] J.M. Curran, et al., Assessing uncertainty in DNA evidence caused by sampling effects, Science and Justice 42 (1) (2002) 29–37.
[20] D.J. Balding, Weight-of-Evidence for Forensic DNA Profiles, John Wiley and Sons, Chichester, 2005.
[21] P. Gill, B. Sparkes, J.S. Buckleton, Interpretation of simple mixtures when artefacts such as stutters are present—with special reference to multiplex STRs used by the Forensic Science Service, Forensic Science International 95 (3) (1998) 213–224.
[22] I.W. Evett, P.D. Gill, J.A. Lambert, Taking account of peak areas when interpreting mixed DNA profiles, Journal of Forensic Sciences 43 (1) (1998) 62–69.
[23] M.W. Perlin, B. Szabady, Linear mixture analysis: a mathematical approach to resolving mixed DNA samples, Journal of Forensic Sciences 46 (6) (2001) 1372–1377.
[24] I.W. Evett, et al., Statistical analysis of data for three British ethnic groups from a new STR multiplex, International Journal of Legal Medicine 110 (1997) 5–9.
[25] B. Budowle, et al., Population data on the thirteen CODIS core short tandem repeat loci in African Americans, US Caucasians, Hispanics, Bahamanians, Jamaicans and Trinidadians, Journal of Forensic Science 44 (1999) 1277–1286.
[26] F. Van_Nieuwerburgh, et al., Impact of allelic drop-out on evidential value of forensic DNA profiles using RMNE, Bioinformatics 25 (2) (2009) 225–229.
[27] T. Tvederbrink, P.S. Eriksen, H.S. Mogensen, N. Morling, Estimating the probability of allelic drop-out of STR alleles in forensic genetics, Forensic Sci. Int. Genet. (2009), doi:10.1016/j.fsigen.2009.02.002.