Forensic Science International: Genetics 13 (2014) 269–280

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Original Research

Interpreting forensic DNA profiling evidence without specifying the number of contributors

Duncan Taylor [a,b,*], Jo-Anne Bright [c], John Buckleton [c]

[a] Forensic Science South Australia, 21 Divett Place, Adelaide, SA 5000, Australia
[b] School of Biological Sciences, Flinders University, GPO Box 2100, Adelaide, SA, 5001, Australia
[c] ESR, PB92021, Auckland, New Zealand

ARTICLE INFO

Article history:
Received 31 March 2014
Received in revised form 11 August 2014
Accepted 31 August 2014

Keywords:
DNA profile interpretation
Mixtures
Number of contributors
MCMC
Continuous model
STRmix

ABSTRACT

DNA profile interpretation has benefitted from recent improvements that use semi-continuous or fully continuous methods to interpret information within an electropherogram. These methods are likelihood ratio based and currently require that a number of contributors be assigned prior to analysis. Often there is ambiguity in the choice of number of contributors, and an analyst is left with the task of determining what they believe to be the most probable number. The choice can be particularly important when the difference between two possible contributor numbers means the difference between excluding a person of interest as being a possible contributor, and producing a statistic that favours their inclusion. Presenting both options in a court of law places the decision with the court. We demonstrate here an MCMC method of correctly weighting analyses of DNA profile data spanning a range of contributors. We explore the theoretical behaviour of such a weight and demonstrate these theories using practical examples. We also highlight the issues with omitting this weight term from the *LR* calculation when considering different numbers of contributors in the one calculation.

1. Introduction

DNA profile interpretation has benefitted from recent improvements that use semi-continuous (e.g. LRmix, LikeLTD, LabRetriever) [1–5] or fully continuous (e.g. STRmix™, TrueAllele) [6–8] methods to interpret information within an electropherogram (EPG). These methods are likelihood ratio (*LR*) based and currently require that a number of contributors be assigned prior to analysis. Although it is possible in each of these systems to analyse the same profile under a number of different contributor options, the question still remains how to make use of the information. This is particularly true when the difference between two possible contributor numbers means the difference between excluding a person of interest (POI) as being a possible contributor, and producing a statistic that favours their inclusion. Presenting both options in a court of law places the decision with the court. If it is not possible for an expert to make this assignment it may be expecting a lot to ask the court to do so.

There are two recognised solutions to the problem, both of which have their foundation in the idea that the number of contributors is a nuisance variable and should be integrated out of the *LR* calculation. This is not a commonly held view. Many forensic biologists would consider the number of contributors to be something that should be determined from the data and hence are part of the output of the interpretation rather than a nuisance parameter that we will sum or integrate out. Budowle et al. state that "every effort should be made to provide the best estimate of the number of contributors" [9]. This, and many other statements of this type, gives voice to the commonly held, but incorrect, view that the number of contributors to a profile is knowable and is an important part of the output.

While other works have examined the issue of calculating *LR*s where the number of contributors is either unknown or can be bound in some sensible manner [10,11], the theory relies on having probabilities for the numerator and denominator of the *LR*, which simplifies the calculation. Commonly continuous systems, which utilise at least peak height information, work with probability densities. We explain here a Markov Chain Monte Carlo (MCMC) method for weighing different numbers of contributors against each other, and the practical consequences of including this information in the *LR*. In doing so we demonstrate that a single

* Corresponding author at: Forensic Science South Australia, 21 Divett Place, Adelaide, SA 5000, Australia. Tel.: +61 8 8226 7700; fax: +61 8 8226 7777.
E-mail address: Duncan.Taylor@sa.gov.au (D. Taylor).

http://dx.doi.org/10.1016/j.fsigen.2014.08.014
1872-4973/

exact number is not required, contrary to many currently held views.

1.1. Mathematics of the LR

The evidence obtained consists of a number of peaks (O^l_{ar} for allele '*a*' at locus '*l*' in replicate '*r*'), each with an associated height, molecular weight and DNA sequence. Whether this information is obtained from one or several replications of the same DNA sample the result is a series of observed peaks, which together are refer to as the observed data **O**.

We are interested in the probability of obtaining this observed evidence and in particular the probability of obtaining it given some competing propositions (or hypotheses), H_1 and H_2. We use Pr for probability and *p* for a probability density.

$$LR = \frac{Pr(\boldsymbol{O}|H_1)}{Pr(\boldsymbol{O}|H_2)}$$

In order to calculate these competing probabilities there are a number of nuisance parameters that we must consider, namely:

- The *j* genotype sets ($\boldsymbol{S}_j$) of contributors, each of which is made up of *n* single person genotypes for an *n* person mixture.
- The mass parameter (**M**), which is the term used for the grouping of parameters for template DNA amounts for each contributor (t_n), a degradation curve for each contributor (d_n), a replicate amplification efficiency for each replicated analysis (R_r), amplification efficiencies for each locus (A^l), and peak height variance constants for stutters and alleles.
- The number of contributors **N** to a DNA profile. Note that here **N** may signify one, or several possible numbers under consideration.

For many years the nuisance parameter that has been most concentrated is the genotype sets, which has been incorporated into the *LR* by:

$$\mathrm{LR} = \frac{\sum_j p(\boldsymbol{O}|\boldsymbol{S_j}, H_1)Pr(\boldsymbol{S_j}|H_1)}{\sum_{j'} p(\boldsymbol{O}|\boldsymbol{S}_{j'}, H_2)Pr(\boldsymbol{S}_{j'}|H_2)}$$

For many years simplifying assumptions have been made for $p(\boldsymbol{O}|\boldsymbol{S_j}, H_x)$, often to the point where these values where considered equal and simply removed from the equation, or designated as zero, resulting in the removal of the entire genotype set from the *LR* calculation. The likelihood, $p(\boldsymbol{O}|\boldsymbol{S_j}, H_x)$, is independent of the proposition because given a genotype set the likelihood will be the same regardless of whether a POI is being hypothesised as a contributor, i.e. $p(\boldsymbol{O}|\boldsymbol{S_j}, H_1) = p(\boldsymbol{O}|\boldsymbol{S_j})$. In order to assess $p(\boldsymbol{O}|\boldsymbol{S_j})$ it is helpful to introduce another set of nuisance parameters, the mass parameters. Following Taylor et al. [6] we denote these mass parameters **M**, and note $p(\boldsymbol{O}|\boldsymbol{S_j}) = \int_M p(\boldsymbol{O}|\boldsymbol{S_j}, \boldsymbol{M})\, p(\boldsymbol{M}) d\boldsymbol{M}$. We term $p(\mathbf{O}|\mathbf{S_j})$ as the 'weight' for genotype set *j* and give it the nomenclature w_j.

$$\mathrm{LR} = \frac{\sum_j w_j Pr(\boldsymbol{S_j}|H_1)}{\sum_{j'} w_{j'} Pr(\boldsymbol{S}_{j'}|H_2)} \quad (1)$$

Note the summations in the numerator and denominator of Eq. (1) may contain different numbers of non-zero elements, which we indicate by using summation indices of *j* and *j*′. We use this nomenclature throughout the remainder of the paper. To obtain weights we use the system of [6], whereby an MCMC process steps through different genotype sets from iteration to iteration. The weights obtained are residence times of each genotype set as the focus of the MCMC. In doing so we produce weights that sum to one, and are proportional to $p(\mathbf{O}|\mathbf{S_j})$. The value of $p(\mathbf{O}|\mathbf{S_j})$ is a density without its normalising constant, and we ask the reader to consider them as densities rather than probabilities as this assists between model comparisons discussed later in this paper. When considering the number of contributors, *N*, the weights must be considered as spanning models (consideration of different numbers of contributor). Note that Eq. (1) can still be thought of as having a term for number of contributors, however as the number is fixed for all components and there are no comparisons between models of different numbers of contributor then the *N* term is omitted. To obtain proportionality within and between models we note that the weights are products of a term for within model comparisons (which we will continue to use w_j) and a term for between model comparisons, which we term Z_n. Introducing Z_n into the *LR* gives:

$$\mathrm{LR} = \frac{\sum_n Z_n Pr(N_n|H_1) \sum_j w_j Pr(\boldsymbol{S_j}|N_n, H_1)}{\sum_{n'} Z_{n'} Pr(N_{n'}|H_2) \sum_{j'} w_{j'} Pr(\boldsymbol{S}_{j'}|N_{n'}, H_2)} \quad (2)$$

where N_n is a model specifying *n* contributors and we make explicit that there is no mathematical connection between the number of contributors under H_1, *n*, and under H_2, *n*′.

Details of the calculation and meaning of Z_n and the derivation of Eq. (2) are given in full in supplementary material 2 (note that the full derivation is mathematically dense, and the majority of this paper does not require an understanding to that depth).

This allows us to introduce the concept of dimensionality. There are more variables (dimensions) in the vector **M** for increased number of contributors. This means that the mass parameters are now dependent on contributor numbers so that $p(\mathbf{M}_n|N_n)$ is a probability density in multidimensional space. In MCMC parlance we could term the number of contributors different models and hence $N = 3$ could be one model and $N = 4$ another. There are a number of known ways to compare different models, some of which are within chain comparisons such as reversible jump MCMC [12]. The method used in this study is a between chain comparison by calculation of the marginal likelihoods.

1.2. The effect of propositions on the LR

1.2.1. The effect on genotype set weights

At this stage it is useful to discuss the propositions that are being considered. In Eq. (2) there are only two terms that depend on the propositions, $\Pr(\boldsymbol{S_j}|N_n, H_x)$ and $\Pr(N_n|H_x)$. The $\mathbf{S}_j$ term can be further decomposed into the genotypes of the known contributors (assumed to be present in the DNA sample by all parties, $\mathbf{S}_k$), the genotype sets of the POIs being postulated as contributors under one proposition but not all (typically in forensic contexts this will be a POI considered a contributor under H_1 but not in H_2, $\mathbf{S}_p$) and the genotypes of all the other, unknown contributors ($\mathbf{S}_U$) that must be present in the mixture to explain the total number of contributors to the profile that are not accounted for by known contributors or POIs. Using H_1 as an example, this gives:

$$Pr(\boldsymbol{S_j}|H_1) = Pr(\boldsymbol{S}_{U_j}|\boldsymbol{S_k}, \boldsymbol{S_p}, H_1) Pr(\boldsymbol{S_k}, \boldsymbol{S_p}|H_1) \text{ where } Pr(\boldsymbol{S_k}, \boldsymbol{S_p}|H_1) = 1$$

The proposition makes a difference here because out of all '*j*' possible combinations of *n* person genotypes that could describe **O**, many of them will not contain the genotype of the POI. In such a situation there are no combinations of genotypes of unknowns that can explain the data given $\mathbf{S_k}$ and $\mathbf{S_p}$, and so $\mathbf{S_U}$ is an empty set ($\boldsymbol{S}_{U_j} = \varnothing$) giving rise to $Pr(\boldsymbol{S}_{U_j}|\boldsymbol{S_k}, \boldsymbol{S_p}, H_1) = 0$. The remaining non-zero elements of $Pr(\boldsymbol{S}_{U_j}|\boldsymbol{S_k}, \boldsymbol{S_p}, H_1)$ are based on the rarity of the required genetic components (alleles) in whatever population is of interest.

Even though the same list of *j* genotype sets is considered under H_1 and H_2, with the same values of $p(\mathbf{O}|\mathbf{S_j})$, we often write the sum under H_1 as having *j* non-zero elements and under H_2 having *j*' non-zero elements. However because w_j are not dependent on

propositions, MCMC is able to be used to determine them based purely on the observed data, **O**.

1.2.2. The effect on contributor number weights

$\Pr(N_n|H_x)$ can be evaluated in one of two ways. If a proposition specifies exactly n contributors then:

$$Pr(N_{n*}|H_x) = \begin{cases} 1 & n* = n \\ 0 & \text{otherwise} \end{cases}$$

If there is ambiguity in the number of contributors that the propositions then a possible way forward is to consider all values for N_n equally likely:

$$Pr(N_{n*}|H_x) = \frac{1}{K} \quad \text{for all } n^*$$

where there are K different numbers of contributors being considered. However any values can be chosen for these probabilities if information exists to guide the choice.

1.2.3. The propositions themselves

Classically these propositions have always specified a number of contributors. However to treat uncertainty in the number of contributors, that number cannot be specified in the propositions. To allow uncertainty in the number of contributors it is necessary to consider propositions that do not specify an exact number of contributors, such as '*the POI is a contributor of DNA to this sample*' considered against '*the sample has originated from people unrelated to the POI*'.

Alternatively propositions can include contributor numbers as long as they encompass a range. In this instance the propositions become '*the POI and n to n' unrelated individuals are the sources of DNA*' considered against '*(n + 1) to (n' + 1) individuals, unrelated to the POI, are the sources of DNA*'. Both sets of hypotheses produce equivalent *LR*s.

1.3. Dealing with a range of contributors

We discuss the two options for dealing with a range of contributors within the *LR*. These are:

i. integrating the number of contributors out and
ii. assigning a number based on maximising $\text{pr}(\mathbf{O}|H_x)$ for each H_x.

which are explored in Sections 1.3.1 and 1.3.2 respectively.

1.3.1. Integrating out the number of contributors

This approach seeks to implement Eq. (2). We note that the total number of contributors to the profile and the genotypes of the unknown contributors to the profile are all nuisance variables. This is not a new concept and was discussed by Buckleton et al. [13] although given how obvious the equation is it is likely it was considered much earlier. There has been minimal uptake in the forensic community presumably due to a lack of software options that are able to implement it, and the difficulty in assigning the between model weights, Z_n.

Under this scenario the number of contributors is not specified under either propositions and rather a range is considered. Although informative prior probabilities for the number of contributors can be used it is more likely that an uninformative prior will be used so that the $\Pr(N_n|H_x)$ terms in Eq. (2) cancel each other out to give:

$$\text{LR} = \frac{\sum_n Z_n \sum_j w_j Pr(\boldsymbol{S_j}|N_n, H_1)}{\sum_{n'} Z_{n'} \sum_{j'} w_{j'} Pr(\boldsymbol{S_{j'}}|N_{n'}, H_2)}$$

1.3.2. Using the most probable number of contributors for each hypothesis (MPN)

This approach assigns the number of contributors (n) as that choice that produces the maximum posterior probability for the EPG, $\vee_n Z_n \sum_j w_j Pr(\boldsymbol{S_j}|N_n, H_1)$ and $\vee_{n'} Z_{n'} \sum_{j'} w_{j'} Pr(\boldsymbol{S_{j'}}|N_{n'}, H_2)$.

Using this approach the number of contributors for each competing hypotheses within the *LR* is assigned in such a way that it optimises the posterior probability of the EPG for that hypothesis. The number of contributors may be the same or different for each hypothesis, however a single number is chosen for each.

There are two competing forces that determine which choice of contributors is most favourable for a given EPG. Firstly there is a drive to minimise the number of unknown contributors under a hypothesis as each additional unknown contributor incurs an additional genotype probability in the calculation. This is true in the calculation of $\Pr(\boldsymbol{O}|H_x, \boldsymbol{M_n}, \boldsymbol{S_j}, N_n)$ and also in the calculation of Z_n. The number of imbalances or stochastic effects will drive an increase in the number of contributors as the profile will be described better with an additional contributor accounting for the imbalances.

The effect of each of these two competing components will dictate which choice of number of contributors produces the greatest posterior probability for the EPG. The *LR* calculated will ultimately be

$$\text{LR} = \frac{Z_n \sum_j w_j Pr(\boldsymbol{S_j}|N_n, H_1)}{Z_{n'} \sum_{j'} w_{j'} Pr(\boldsymbol{S_{j'}}|N_{n'}, H_2)}$$

where n and n' can be the same or different.

1.3.3. Choosing stratification or MPN

There are scenarios where a range may be more applicable than an MPN estimate for the choice in number of contributors, or vice versa. Note that the choice between stratification and MPN only affects the $\Pr(N_n|H_x)$ terms in Eq. (2) as outlined in Section 1.2.2.

Consider an intimate swab from a victim which has yielded a DNA profile with peaks at heights which could be reasonably explained by two contributors. The victim says she was raped by one man and has not had recent consensual sex. The main contributor to this profile corresponds with the victim, who is an assumed contributor. There are a number of minor alleles present all except one of which can be accounted for by the person of interest. Under the assumption of two contributors (and ignoring the possibility of drop-in for this scenario) the suspect would be excluded as a contributor of DNA to this profile. This is the position that defence may wish to take. The prosecution would take the stance that the profile has originated from three individuals and so would not exclude the suspect as a source of DNA. Under this scenario the MPN estimated values for number of contributors is arguably the better treatment of the profile.

Now consider a complex, low level DNA profile that has originated from an item where no-one can be assumed to have contributed DNA and there are several persons of interest for comparison. The profile can be described as two person profile, although there are indications that it may be from more than two, such as sub-threshold peaks, imbalances or drop-ins/drop-outs. Under this scenario, given the inability to reasonably assign a number of contributors to the profile, stratifying across a number of contributors is arguably the more appropriate choice.

1.4. The interaction of peak height variability and the number of contributors

There are four sources of ambiguity in assigning the number of contributors. These are:

1. Contributors sharing alleles, known as 'masking'
2. Artefactual peaks in allelic positions
3. Drop-out of alleles
4. Variability in peak heights

The masking of alleles has dominated considerations of the number of contributors [13–17], probably incorrectly. It is a common claim in court by defence experts the 'true' number of contributors to a profile could be different from the number used in statistical analyses. This of course misses the point that for evidence samples the 'true' number of contributors can never be known and is not required for *LR* calculations. It is also often forgotten that defence and prosecution have every right to nominate numbers of contributors in their own propositions, but have no jurisdiction over the other party's choice.

Our experience suggests that the interpretation of small peaks in forward $(a+1)$ stutter positions, larger than expected peaks in back stutter $(a-1)$ positions, and peaks imbalances are larger sources of ambiguity. The drop-out of alleles and variability in peaks heights are manifestations of the same underlying phenomenon, that the peak height of an allele is not directly related to the template available in the extract.

In current practice the assignment of a number of contributors usually proceeds by assigning peaks as allelic or artefactual. There may be ambiguity in the assignment of peaks as artefactual and this arises most often when backstutter peaks are of a similar height to some unambiguously allelic peaks. This uncertainty is mentally "carried forward." A putative minimum number of contributors is then typically assigned as the maximum number of allelic peaks divided by 2. If this is not a whole number the result is rounded up. This putative number is then subjectively trialled against the EPG for potential fit to peaks heights and the ambiguously artefactual peaks.

Variation in peak heights has historically been treated using a threshold of acceptance or rejection for the ratio of two peaks from a heterozygote (*Hb*). Thresholds on *Hb* may be soft or hard. Consider, for example, two peaks of height 2000 and 1000RFU. The ratio of these is 2:1 and this ratio may be considered high if the sample is from a single donor. This would suggest that the inclusion of an additional contributor is warranted. This mental process is informally considering the likelihood of the observed data given a number of contributors. What is commonly lacking in this consideration is the effect of the genotype probabilities when adding a contributor. The prior probability of the genotype sets drops dramatically in some circumstances due to the multiplication of an additional profile frequency in the prior. In Example 1 the difference between likelihood and posterior probability (due to changes in prior probabilities) is demonstrated.

Empirical studies suggest that the variance in peak height is inversely proportional to the amount of template [18–21]. We model the system's tolerance to stochastic effects using a variance constant term. See [6] for a full explanation of the model, but to surmise the observed peak height (O) is compared to the expected peak height (E) using the model:

$$\log\left(\frac{O}{E}\right) \sim N\left(0, \frac{c^2}{E}\right)$$

where c^2 is the variance constant. This variance constant is not known with certainty and may vary between different samples and between models. We model the constant as having a gamma distributed prior. This choice appears reasonable from empirical data. However, following Balding [3] we allow the profile under consideration to influence the value for this constant.

Consider an apparently single source sample based on allele count. When treated as a single source profile the difference in peak heights within and between loci has to be treated as variance. If the variance is low (an intolerant variance) then a low probability will be produced for any imbalanced peaks. If the variance is high (a tolerant variance) then the system will produce moderate probabilities for imbalanced peaks.

We now consider the effect of adding a second contributor to a potentially single source profile. Certain combinations for this second contributor can improve the fit of the proposed genotypes to the profile. This will have the biggest effect for an intolerant variance and a lesser effect for a tolerant variance. We would therefore expect that the addition of a second contributor would improve the likelihood of the profile most when using an intolerant variance, or when large imbalance exists within the profile. This is also demonstrated in Example 1.

1.5. The effect of different EPGs will have on Z_n

1.5.1. If peak heights are high and balanced

The consideration of a number of contributors above the minimum required to reasonably describe the profile will not substantially improve the fit to the profile. In this instance the additional unknown contributor is likely to be considered as a very trace contributor in a probabilistic analysis.

This is contrary to many expectations that adding contributors to only one proposition's case is always detrimental to that case. For a mathematical demonstration of this concept we point the reader to Supplementary material 1. We provide a practical demonstration in Example 2. Note that this almost independence of the *LR* from the number of contributors is only true under the specific circumstances that the POI can account for one of the dominant contributors to the profile and that the addition of contributors provides no improvement to the description of the observed profile. When one or both of these is not the case the effects of changing numbers of contributors can be dramatic, as shown in Example 3 and Section 4.2.

1.5.2. If the peak heights are high and imbalanced

A choice of number of contributors that can account for the imbalances in some reasonable manner will describe the observed EPG much better than fewer contributors. In this instance the two effects of genotype probability and profile fit will be acting against one another and the weight will depend largely on the severity of the imbalance and rarity of the alleles. Example 3 explores this concept.

1.5.3. If peak heights are low

The scenario will be similar to the balanced profile considered above. The *LR* generated will depend on the balances within the profile, the intensity of the peaks, and rarity of alleles.

Example 1. The effect of the model on Z_n

Before we can assess a system's ability to generate weights that correspond to numbers of contributors we must ascertain how that system should be set up. Consider a strong, balanced single source profile (Fig. 1). We will interpret this as single source (correct) and as a two person mixture (incorrect).

Changing the shape of the variance constant prior can make the method more or less tolerant of stochastic effects. This has a direct effect on the values of Z_n, as imbalances that were acceptable given a more tolerant system, become vastly less probable.

The gamma prior distribution for the variance has two variables, shape and scale. Gamma distributions were chosen to produce a narrow distribution about a desired mode. Modes tested

were 0.1, 0.5 and 5. The narrowness of the distribution means the variance constant value will be very limited in how much the profile in question can affect it. This allows the effect of the variance and the number of contributors to be viewed with less confounding of effects.

For all calculations of *LR*s or Z_n values in this paper, allele frequencies were used from an Australian, self-declared Caucasian database [22].

Fig. 2 demonstrates that reducing the variance constant to very low levels leads to a marked favouring of the addition of a contributor to describe the profile. This is because at these levels the system is highly intolerant of imbalances, and so the slight differences between observed and expected peak heights are explained by the presence of a trace second contributor. Note that in Fig. 2 the distributions are only the likelihood of the observed data given the mass parameters within a model, they have not be multiplied by priors that include the allele frequencies. This can be seen by the distance between the modes of the distributions compared to the ratio of Z_n values being different by approximately the frequency of a full profile ($\sim 10^{12}$) when the variance constant is less than 0.5. This difference between distributions Z_n values is not seen when the variance constant is 5. The reason for the different behaviours can be explained by the ambiguity in the genotypes of the second contributor and the role that has to play in the calculation of Z_n.

Consider a profile originating from two individuals with a locus that specifies a mixture proportion and a variance constant that 'locks' that mixture proportion across the profile. For

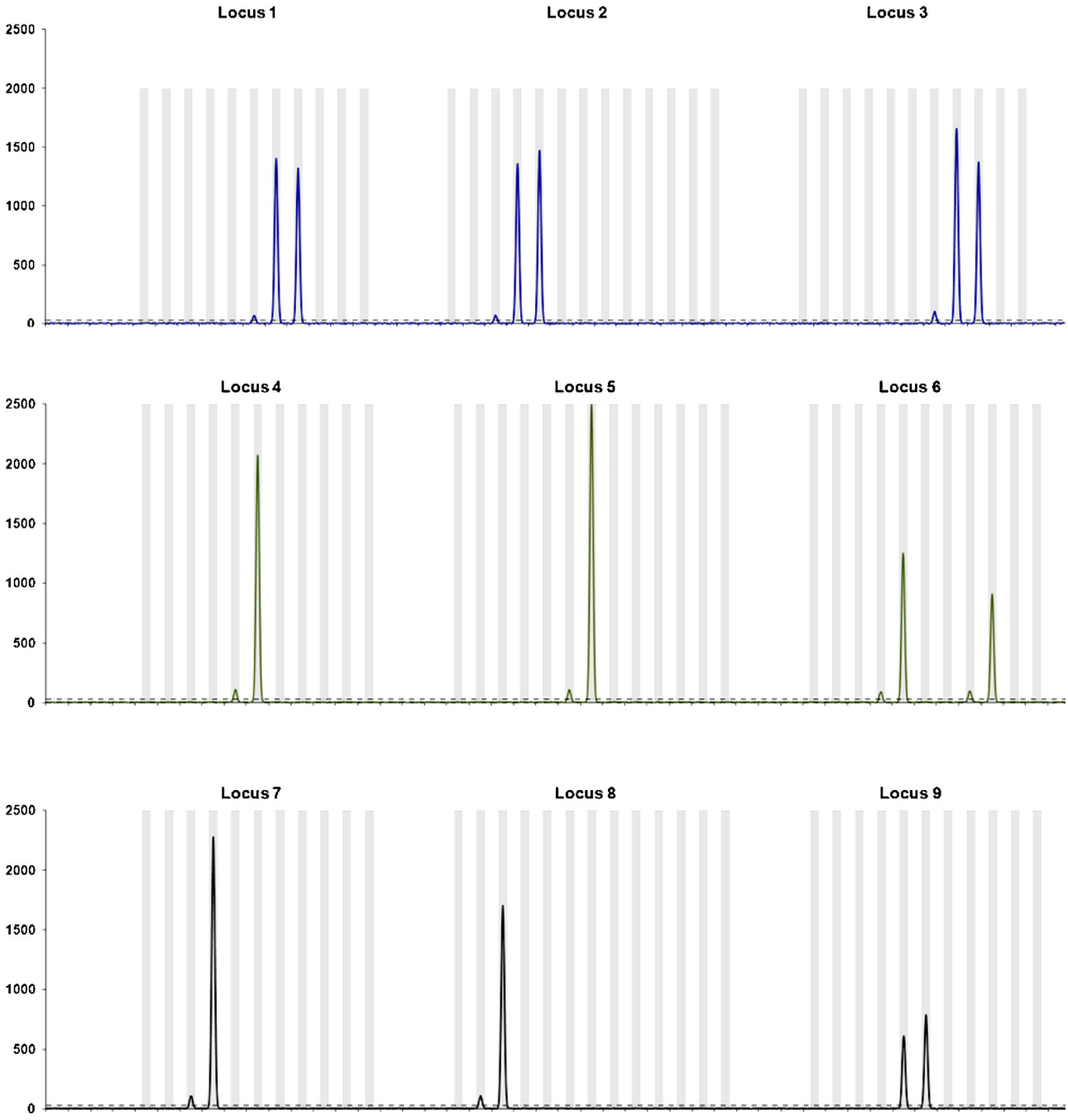

Fig. 1. Single source profile used for calculation.

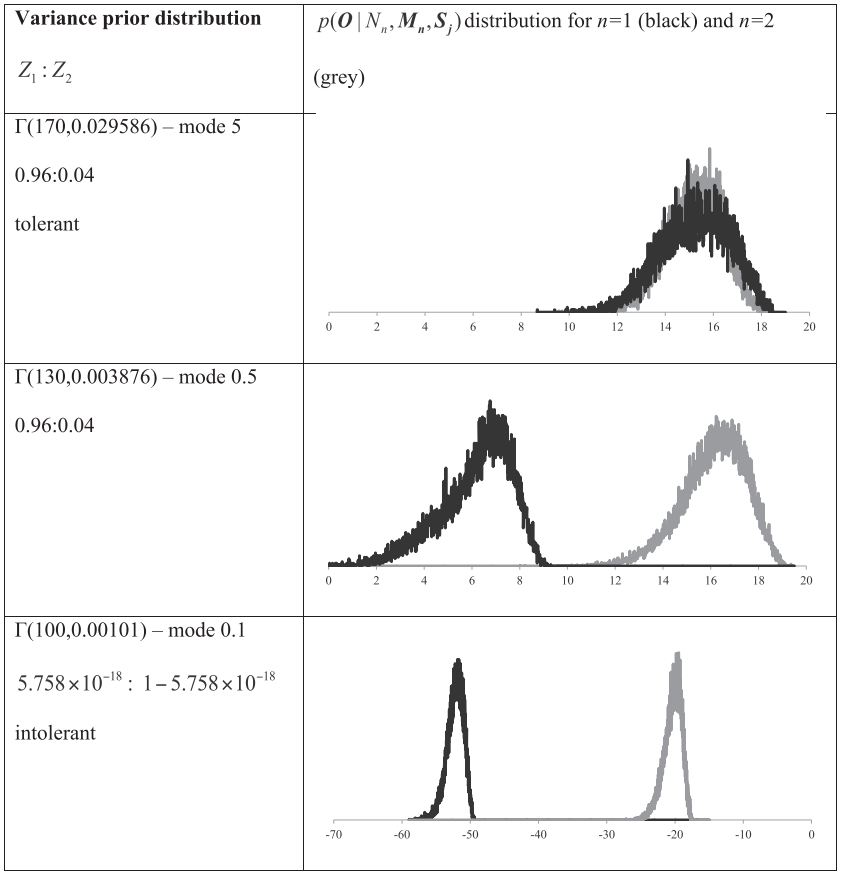

Fig. 2. Graphs showing distribution of $\log_{10}[p(O|N_n, \boldsymbol{M_n}, \boldsymbol{S_j})]$ on x-axis for profile seen in Fig. 1, when considered as originating from one (black) or two (grey) individuals. Note that these distributions are log likelihoods only and do not include priors i.e. for mass parameters, number of contrbutors or genotype set probabilities. Also note that these log likelihoods are produced by an MCMC process which steps through different genotype sets. The distributions produced have resulted from numerous genotype sets being the focus of the MCMC at various iterations during its run.

example picture a locus with two peaks [A,B] where [A] is 1000rfu and [B] is 250rfu. Imagine that the variance constant is small enough that only two genotypic explanations of the data are acceptable under N_2, and that is the genotypes are [A,A]&[A,B] with a mixture ratio of 0.6:0.4 or [A,A]&[B,B] with a mixture ratio of 0.8:0.2. If the profile was considered as originating from a single individual then the pairing of [A] and [B] will incur some penalty as the peaks are imbalanced i.e. $p(\boldsymbol{O}|N_2, \boldsymbol{M}_2, \boldsymbol{S_j}) > p(\boldsymbol{O}|N_1, \boldsymbol{M}_1, \boldsymbol{S_j})$. This will be offset by the difference in $\Pr(\boldsymbol{S_j}|N_n)$ within $\int_{M_n} \sum_j p(\boldsymbol{O}|N_n, \boldsymbol{M_n}, \boldsymbol{S_j})\, p(\boldsymbol{M_n}|N_n)\, p(\boldsymbol{S_j}|N_n)$ used to generate Z_n. Now consider that this is the only indication of a second contributor in the profile and we move on to the next locus, which is homozygous and has single peak [C] at 2000rfu. At this locus $p(\boldsymbol{O}|N_1, \boldsymbol{M}_1, \boldsymbol{S_j}) \approx p(\boldsymbol{O}|N_2, M_2, \boldsymbol{S_j})$ as the additional contributor does not increase the likelihood of the observed profile. The genotypes of the potential contributor under N_1 is [C,C] and under N_2 is [C,C]&[C,C]. There are no other acceptable genotype sets under N_2 as the first locus has dictated that the minor contributor is providing approximately 400 to 800rfu to [C] (based on the mixture proportion of 0.4–0.2). At this level dropout is sufficiently unlikely that it can be ignored. So in the calculation of Z_n $p(\boldsymbol{M}_1|N_1)p([\mathrm{C,C}]) \gg p(M_2|\boldsymbol{N}_2)p([\mathrm{C,C}]\,\&\,[\mathrm{C,C}])$ by approximately the frequency of a [C,C] genotype and the difference between $p(\boldsymbol{M_1}|N_1)$ and $p(\boldsymbol{M_2}|N_2)$ which is discussed later. This trend will continue across all loci where there is no improvement in fit from the addition of a contributor and the genotype sets are restricted by the mixture proportions.

If the penalties incurred from imbalances outweigh the rarity of an additional contributor's profile then $Z_2 > Z_1$ as seen in Fig. 2 at mode 0.1. As variance increases the imbalances are tolerated more and the penalty is lower, however if the variance is still small enough to restrict the genotype set then a point is reached where $\frac{p(\boldsymbol{O}|N_2, \boldsymbol{M_2}, \boldsymbol{S_{j'}})}{p(\boldsymbol{O}|N_1, \boldsymbol{M_1}, \boldsymbol{S_j})} < \frac{Pr(\boldsymbol{S_{j'}}|N_{n'})}{Pr(\boldsymbol{S_j}|N_n)}$ and this will result in $Z_2 < Z_1$ as seen in Fig. 2 at mode 0.5.

Now consider the same situation, but this time at locus 1 the peak height of peak [B] is 1010rfu (indicating some small level of imbalance that could be explained by standard stochastic effects). Imagine that at all loci the likelihood of the profile is not substantially increased by the addition of a second contributor, i.e. $p(\boldsymbol{O}|N_1, \boldsymbol{M_1}, \boldsymbol{S_j}) \approx p(\boldsymbol{O}|N_2, \boldsymbol{M_2}, \boldsymbol{S_{j'}})$ at all loci. Additionally the second contributor is deemed to be providing very little template to the observed profile. Using the second locus again, if we consider the priors, and particularly the genotype probability prior then $\sum_j p(\boldsymbol{M_1}|N_1)Pr(\boldsymbol{S_j}|N_1) = p(\boldsymbol{M_1}|N_1)Pr([C,C])$, as there is still only one genotype that can explain the observed profile. In the two person scenario $\sum_j p(\boldsymbol{M_2}|N_2)Pr(\boldsymbol{S_j}|N_2)$ is calculated by:

$$\begin{aligned} &= p(\boldsymbol{M_2}|N_2)\{Pr([C,C]\&[C,C]) + Pr([C,C]\&[C,Q]) + Pr([C,C]\&[Q,Q])\} \\ &\approx p(\boldsymbol{M_2}|N_2)Pr([C,C]) \end{aligned}$$

And hence the difference between Z_1 and Z_2 will be based on the difference between $p(\boldsymbol{M_1}|N_1)$ and $p(\boldsymbol{M_2}|N_2)$. Under these circumstances Z_1 will be mildly favoured over Z_2 as $p(\boldsymbol{M_2}|N_2)$ contains an

extra template term and an extra degradation term (or dimension), each with a prior. This result can be seen in Fig. 2 mode 5 where Z_2 is mildly less than Z_1, and demonstrates the theory that MCMC systems favour simplistic models.

For the remainder of the paper a variance constant prior gamma distribution of $\Gamma(1.62, 3.98)$ was used for alleles and $\Gamma(2.57, 3.57)$ for stutter peaks. It is known that stutter and allele peaks have different peak variance values [20]. These values were optimised to Profiler Plus control data (analysis not shown).

Example 2. Considering different dimensions in H_1 and H_2

We consider the results from Example 1 but this time consider it as originating from one, two, three or four individuals. The known source was used as the POI and was then compared with the analysis at each stage using propositions:

H_1: POI + $(n-1)$ unknowns

H_2: n unknowns

where n can be one, two or three and n' can be one, two, three or four and n is independent of n'. Table 1 shows $\sum_j w_j Pr(\boldsymbol{S_j}|N_n, H_1)$, $\sum_j w_j Pr(\boldsymbol{S_j}|N_n, H_2)$ and Z_n for the four contributor scenarios.

Note a mild decrease in likelihoods as the number of contributors increases, even without the Z_n terms. This demonstrates the ability of the continuous MCMC method to overcome one of the problems associated with purely probabilistic systems that work by maximum likelihood estimation, which is that an increase in the number of contributors always increases likelihood (see Section 5.1 of [23], who demonstrates this phenomenon). Table 2 shows the $\log_{10}$(LR) considering different combinations of individuals in H_1 and H_2 when it is assumed that $Z_n = Z_{n+1}$, i.e Z_n from Table 1 is not included in the *LR* calculation.

Table 3 is a repeat of Table 2, but this time including Z_n in the *LR* calculation.

Table 1
Probabilities densities of the observed profile in Fig. 1 given a varying number of contributors and the Z_n values associated with those numbers of contributors.

n	$\sum_j w_j Pr(\boldsymbol{S_j}\|N_n, H_1)$	$\sum_{j'} w_{j'} Pr(\boldsymbol{S_{j'}}\|N_n, H_2)$	Z_n
1	6.76×10^{23}	1.87×10^{13}	0.8087
2	1.74×10^{22}	4.8×10^{11}	0.0818
3	3.23×10^{20}	8.23×10^{9}	0.0227
4	7.28×10^{19}	2.01×10^{9}	0.0868

Table 2
$\log_{10}$(LR) considering differing number of contributors under H_1 and H_2.

		Contributors under H_2			
		1	**2**	**3**	**4**
Contributors under H_1	**1**	10.6	11.6	12.6	12.8
	2	9.2	10.3	11.3	11.5
	3	8.0	9.1	10.1	10.3
	4	7.7	8.8	9.8	10.0

Table 3
$\log_{10}$(LR) considering differing number of contributors under H_1 and H_2 and including Z_n.

		Contributors under H_2			
		1	**2**	**3**	**4**
Contributors under H_1	**1**	10.6	12.6	14.2	13.8
	2	8.2	10.3	11.8	11.4
	3	6.5	8.5	10.1	9.7
	4	6.7	8.8	10.4	10.0

In this instance the change in *LR*s caused by including Z_n is slight. This is because the addition of contributors in the model beyond one, does not significantly improve the description of the observed peak heights. Additionally there is ambiguity in the genotypes of the additional (unnecessary) contributors, such that there are many possibilities, including complete dropout, that they can take. The sum across these genotypic probabilities is high and therefore has a small impact on Z_n. The results seen in Tables 2 and 3 demonstrate the theory shown in Section 1.5 that under the tested circumstances the *LR* should remain reasonably constant regardless of the addition of contributors under both or either one of the propositions. It should also be noted that due to the slight favouring of simpler (lower contributor) models, there is still no advantage in artificially increasing the number of contributors to one or both of the hypotheses as this will tend to drive the *LR*'s support away from the proposition with the greater number of contributors.

If the number of contributors increases under both propositions (moving down the diagonal of Tables 2 and 3), the *LR* decreases, but not due to any loss of resolution in the 'major' contributor's genotype. Note that the *LR* for the two, three and four person scenarios are approximately one half, one third and one quarter the *LR* of the single source equivalent (note the $\log_{10}(LR)$ if you are comparing $10^{10.6}$, $10^{10.3}$, $10^{10.1}$ and 10^{10}). This is because the propositions being considered do not nominate a specific contributor position for the POI (see [24] for a full explanation of the concept).

Example 3. Imbalanced peaks

We use again the profile given in Fig. 1, but introduce imbalance at locus 1. The height of the second allele at locus 1 is artificially adjusted in the range 40 to 1322rfu (its original height). The peak heights assigned to peak 2 in locus 1 were 40, 250, 500, 750, 1000 and 1322rfu, which produced heterozygote balance *Hb* values that fill the range between zero and one. The profile was again analysed as either originating from one or two contributors. Fig. 3 shows the ratios of the Z_2/Z_1 across the range of *Hb* and demonstrates the effects of the relationship between $p(\boldsymbol{O}|N_n, \boldsymbol{M_n}, \boldsymbol{S_j})p(\boldsymbol{M_n}|N_n)$ and $\Pr(\boldsymbol{S_j}|N_n)$ within $\sum_j p(\boldsymbol{O}|N_n, \boldsymbol{M_n}, \boldsymbol{S_j})\, p(\boldsymbol{M_n}|N_n) Pr(\boldsymbol{S_j}|N_n)$ as described in Example 1. At this point we omit the possibility of drop-in from the calculation as we are interested in showing only the effects of the imbalance on the values of Z_n:

There are a number of observations that can be made from Fig. 3:

- $Hb < 0.2$—the severity of the imbalance means that two contributors explains the profile sufficiently better to outweigh the additional profile probability. At the most extreme *Hb* the height of the less intense peak means that at other loci, when considered as a two person mixture, the genotypes the second contributor can take include complete dropout. This further adds support to the two contributor model.
- $0.2 < Hb < 0.4$—as *Hb* increase the imbalance penalty decreases, and the genotype sets for the second contributor are constricted by the mixture proportion, determined by the imbalance at locus 1. The additional contributor's genotype frequency outweighs the penalty from the imbalance at locus 1. This relationship reaches the pinnacle at $Hb = 0.35$.
- $0.4 < Hb < 0.8$—the penalty from the imbalance is still steadily decreasing, however the difference between the two peaks means that the optimal contribution of the additional contributor is becoming less. Consequently, moving along the *x*-axis from *Hb* of 3.5–5.5 the genotype set weights are being spread over more possible genotypes.

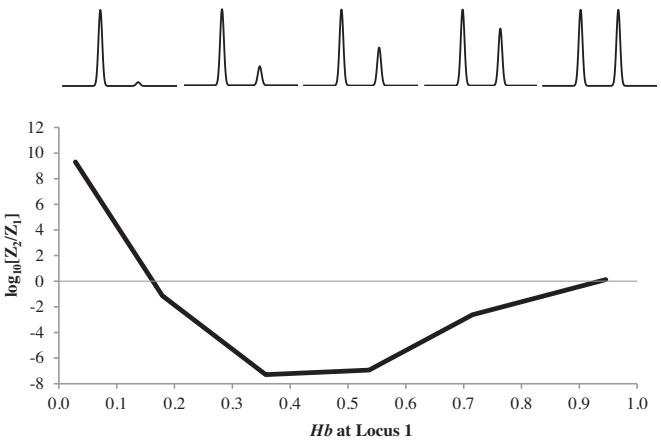

Fig. 3. Improvement in model description obtained by an additional contributor as a function of Hb, with Hb shown diagrammatically above.

- $0.8 < Hb$—after this point the penalty from Hb is relatively minor. The mixture proportion of the second contributor is optimised to close to 0% and so there is complete ambiguity in their genotype. The result is that $Z_1 \approx Z_2$, with Z_1 only being slightly higher due to the presence of an additional template and degradation prior under N_2.

We can consider how this might affect an *LR* calculated where the comparison is to a POI who is homozygous at the first locus i.e. the presence of the second peak at locus 1 needs to be described as either artefactual (e.g. drop-in, which we are not considering) or a second contributor for the POI to produce an $LR > 0$. The classical treatment of this problem would be for the two competing scenarios to nominate their number of contributors and then for the *LR* to be calculated giving equal weight to the numbers of contributors. Providing equal weight to either scenario could substantially disadvantage the case where an additional contributor provides a better, or worse, explanation for the observed profile. We show the effect of this in Fig. 4 by comparing the *LR*s produced by assuming $Z_1 = Z_2$ (solid line) i.e. Z_n effectively terms are not included in the *LR* calculation, and by using the informed Z_n values (dashed line), from the results seen in Fig. 3.

We also include in Fig. 3, a comparison to systems that do not take peak height into account. Under such a system each genotype set weight is composed of fixed probabilities of dropout (which we set to 0.05) and drop-in (which, to be consistent with other systems being displayed, we set to 0). As peak height is not taken into account then by definition changing the height of one of the peaks should have no effect on the *LR*, hence the horizontal line seen in Fig. 4.

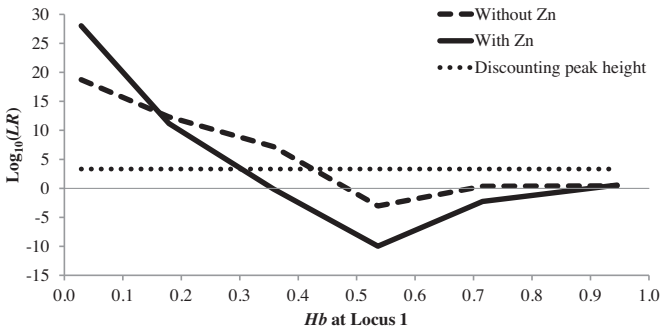

Fig. 4. LR with and without Z_n and also discounting peak height information.

- When $Hb < 0.2$, the profile strongly supports N_2 and if this is not taken into account the prosecution case is unfairly disadvantaged. Note that the *LR* is actually greater than the inverse of the profile frequency as in this example we have forced H_2 to consider only one contributor, which is a very poor description of the observed profile, and this low genotype set weight is used in the *LR* under H_2.
- In the range $0.2 < Hb < 0.8$ the additional contributor model is not supported by the data and omitting Z_n terms disadvantages the defence. A point worth noting is $0.4 < Hb < 0.5$ where the classic treatment favours inclusion of the POI slowly trending down to $LR = 1$, (note that by random chance variation the MCMC analysis has resulted in the dotted line dipping below the line of equality at $Hb = 0.54$, but the general trend expected would be for the *LR* to slowly to decrease to one as Hb increases to one) and including the Z_n values results in an *LR* favouring exclusion of the POI, in some cases by many orders of magnitude.
- When $Hb > 0.8$, then the profile can again be reasonably considered single source, but this time not matching the POI (at locus one). The additional contributor in H_1 adds no further fit to the observed profile. As $Z_1 \approx Z_2$ and there is complete redundancy in the additional contributor's genotype then the rarity of the single unknown's genotype in H_1 approximately equals the rarity of the single unknown's genotype in H_2 and the *LR* is driven towards one.

The difference in *LR* between the dotted line and solid line in Fig. 4 therefore represents the information being lost when assuming $Z_1 = Z_2$ in the *LR* calculation. Using informative weightings for number of contributors rather than uninformative can make a significant difference to the *LR*.

Most existing treatments of trans-contributor problems use uninformative weights for contributor number, and so are not appropriately assessing the two scenarios against each other. The alternative is to require a human interpretation and some system of conventional thresholds to ensure that the choice of number of contributors under both propositions can reasonably describe the observed EPG(s).

2. Validation of model

2.1. Variability in Z_n

The values for Z_n are determined using the system described from a MCMC system and so are subject to run to run variation. We investigate the variability in Z_n by running each of the analyses used the generation of Fig. 3 five times. The average number of post-burn-in iterations was 5.94×10^5 when considered a single source profile and 2.18×10^6 when considered a two person mixture. We carried out the same analyses as lead to Fig. 3 again in Fig. 5 but with all values for $\log_{10}(Z_2/Z_1)$ displayed, and the trend line showing the average $\log_{10}(Z_2/Z_1)$ value.

Fig. 5 shows that running the MCMC in the way described produced estimates for Z_n values that span approximately three orders of magnitude at the most variable Hb values. Note that the remainder of Section 3.1 delves into the components that make up Z_n in order to identify the source of the variation seen in Fig. 5, and hence an understanding of the contents of Supplementary material 2 is required.

Let the maximum likelihood value of posterior sample of the MCMC $p(\boldsymbol{O}|N_n, \boldsymbol{M_n}, \boldsymbol{S_j})\, p(\boldsymbol{M_n}|N_n) = \vee_{M_n,n}\, p(\boldsymbol{O})$, which selects the set of mass parameters that best describes the observed profile under each number of contributors. Ideally this value would be close to the maximum possible theoretical mode of the problem, however as the mode can be very 'pointy' [25] it is not always reached. Therefore the value of $\vee_{\boldsymbol{M_n},n}\, p(\boldsymbol{O})$ varies from run to run by approximately one to three orders of magnitude. This value is

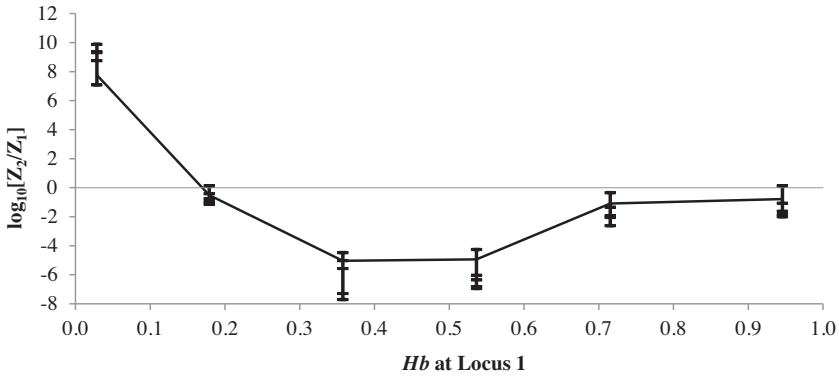

Fig. 5. Fig. 3 reproduced 10 times to show reproducibility. The line intercepts the average value for each Hb bracket.

also directly proportional to the final Z_n values as seen in (data not shown). These results suggest that long runtimes may be necessary in some instance for MCMC to get into the hyper-dimensional sample space close to the theoretical maximum mode. Also, due to the additional dimensions when considering a higher number of contributors, reaching the theoretical mode is likely to take more iterations as the model becomes more complex.

This suggests that much extended runtimes may still be necessary in some instances despite our best efforts. This significantly affects the practicality of the method if it is to be used in a forensic casework setting where significant time pressures commonly exist.

2.2. Test on complex data

The experiments shown in Example 3 indicate that a single extreme imbalance, or a number of mild imbalances, can drive the support for a higher number of contributors ($Z_n < Z_{n+1}$). However in Example 3 the imbalance needed to be very large in order for a higher number of contributors to be strongly favoured and in reality this level of imbalance would give cause to a scientist to choose that higher number unambiguously anyway via human interpretation. We have therefore demonstrated the theory but shown limited practical advantage. We turn now to a more complex scenario. Fig. 6 shows a complex profile originating from three individuals in proportions 0.17:0.42:0.42. This profile has been constructed so that peak masking and dropout means that by peak counting alone the mixture could be described by two contributors with some imbalances, high stutter and mixture proportion inconsistencies across the profile.

The MCMC analysis was run for 2×10^6 iterations under N_2 and 5.4×10^6 iterations under N_3. The likelihood ratios calculated by comparison to the known contributors are given in Table 4 when using propositions:

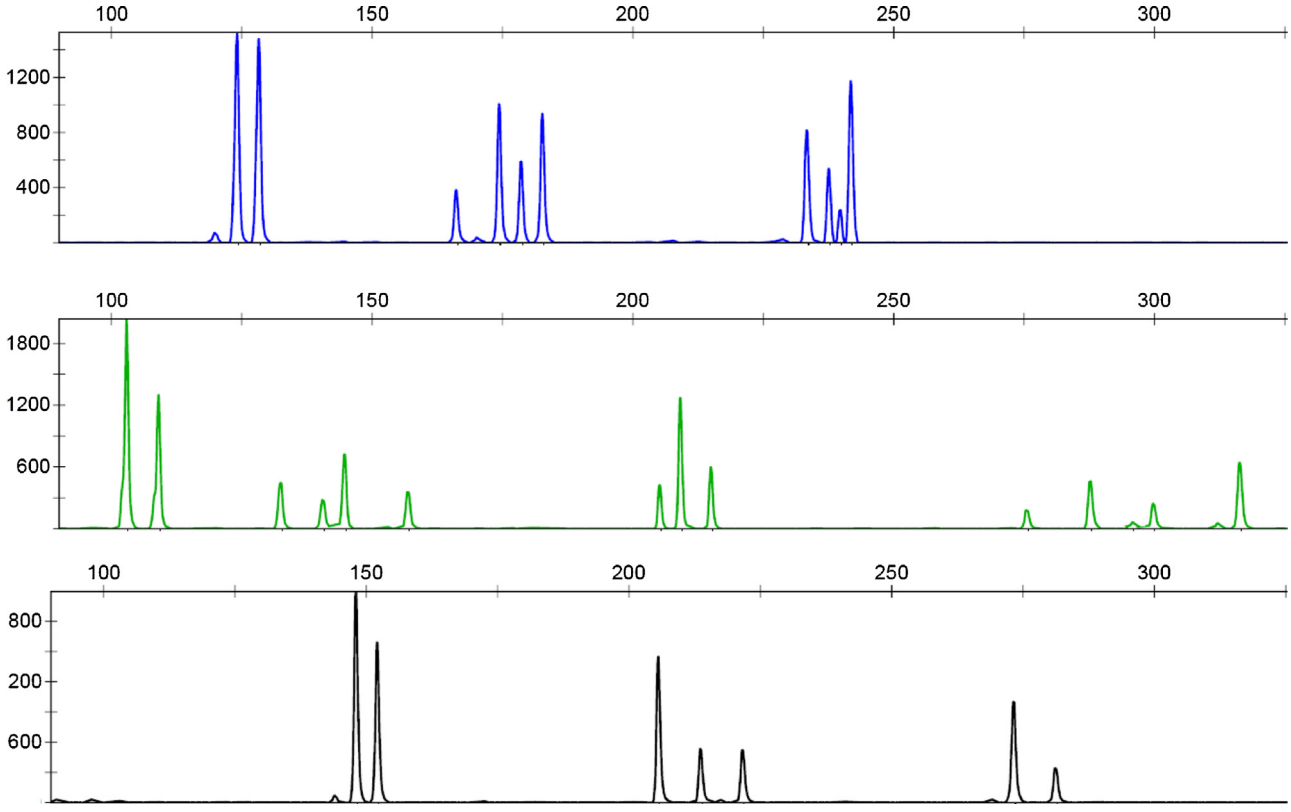

Fig. 6. Complex three person mixture masquerading as a two person mixture.

H_1: POI + $(n-1)$ unknowns

H_2: n unknowns

where n was two or three.

The values for Z_n are $Z_3 = 0.94$ and $Z_2 = 0.06$ indicating support for the three person scenario. To some analysts this may not be as much support for N_3 as expected, there are two reasons for this. Firstly the addition of the third contributor improves the fit to the observed data and consequently they do not have complete redundancy in the genotypes which they can take. This restricted genotype set prior probability therefore has an effect on Z_3:Z_2. Secondly we allow the peak variance constant to be optimised by the profile under N_2 model. Therefore the poor fit to the observed data has increased the variance constant and consequently become more tolerant of imbalances. In effect this analysis is telling us that the profile could be a poor fitting two person mixture with a high peak height variability or a better fitting three person profile with a lower peak height variability. Again the Z_n values appropriately weigh these scenarios against each other in the *LR* calculation.

We now calculate the *LR*s for known contributors using MPN for number of contributors:

H_1: The POI and $(n-1)$ unknowns are the sources of DNA

H_2: n' unknowns are the sources of DNA

(where n can be either 1 or 2 and n' can be either 2 or 3)and stratification across the range of contributors (2–3):

H_1: The POI and 1 or 2 unknowns are the sources of DNA

H_2: 2 or 3 unknowns are the sources of DNAand display the results in Table 5.

Profiles that masquerade as lower order mixtures can strongly favour their correct number of contributors as there will be multiple loci with imbalances or deviations from consistent mixture proportions across a profile. These combined effects can strongly drive Z_n towards the higher order scenario as seen in this example.

The profile seen in Fig. 6 may not be immediately recognised as a three person mixture, particularly if the DNA profiling system is known to have high peak height variability. However, incorrectly treating this profile as two person mixture excludes the known contributors as seen in Table 4. When analysed under the correct number of contributor all known contributors give much larger *LR*s and none are excluded, as would be expected. However the scientist cannot (and certainly should never) base their decision on number of contributors purely so that comparison of the POIs in a case yield an $LR > 1$. The analyst is left with a choice as to whether the stochastic events within the profile provide enough evidence to analyse it as a three person mixture, potentially overestimate the number of contributors required and falsely include a non-contributor. Or the analyst could surmise that the stochastic events do not warrant a third contributor, and so analyse it as a two person mixture and potentially falsely lower the *LR* or exclude true contributors. In some forensic laboratories, current practise would be for the analyst to analyse the profile as originating from three people under H_1 and two under H_2. The calculation of the *LR* would then be carried out without any weight terms for the number of contributors. Alternatively some laboratories would analyse the profile as a two person mixture under both propositions and, then analyse the profile as a three person mixture under both propositions and report both *LR*s. Neither of these two options appropriately weights the evidence. It is this instance that the Z_n values appropriately deal with the ambiguity in number of contributors.

It can be seen in Table 5 that either the MPN or stratification methods for dealing with ambiguous number of contributors gives an appropriate statistical weighting to the evidence for each known contributor, and is the same as the values given in Table 4 at the significance level show. The combination of Z_3/Z_2 and w_j means that N_3 is chosen using the MPN method for both H_1 and H_2, hence the values being the same as in Table 4. In the stratification method $p(\boldsymbol{O}|N_3, \boldsymbol{M}_3, \boldsymbol{S}_j)$ dominates the *LR* and so again at the significance level show the values in Table 5 are the same as in Table 4.

Table 4
Results from analysis of constructed three person mixture when analysed as $n = 2$ and $n = 3$.

	LR under N_2	LR under N_3
POI 1	0	1.03×10^6
POI 2	0	3.0×10^5
POI 3	0	2.6×10^8

Table 5
Results from analysis of constructed three person mixture when considered as a two to three person mixture.

	MPN	Stratified
POI 1	1.03×10^6	1.03×10^6
POI 2	3.0×10^5	3.0×10^5
POI 3	2.6×10^8	2.6×10^8

3. Conclusion

The advent of more sophisticated techniques of analysing DNA profiles has led to informative statistical weighting being obtained from a greater number of DNA profiles. Currently most systems of DNA profile interpretation require that a number of contributors be set prior to analysis. To do this the analyst must rely on their knowledge of DNA profile behaviour and make assessments on whether certain stochastic events are acceptable given a posited number or contributors. The acceptability of stochastic events requires rules and thresholds (even if values are not specified exactly), which is thorn in the side of advocates of continuous DNA interpretation systems that are designed to remove all such thresholds.

We attempt here to extend the model of [6] so that a posterior probability of a number of contributors to a model can be obtained as part of the MCMC process being used to analyse the profile. Obtaining these relative weights for profile dimensionality, Z_n, allows the calculation of an *LR* where either different numbers of contributors are posited under the two propositions, or a range of contributors is desired using weights informed by the data itself. This is an advance to the most common current method of dealing with trans-contributor analyses, which is either to:

(1) increase the number of contributors under both propositions or
(2) calculate the *LR* with a different numbers of contributors under the different propositions, but not taking into account the relative fits of these models to the observed profiles.

We have shown that both of these options can bias the result by many orders of magnitude. The first is biased when the increase in number of contributors is not required under one of the propositions (typically H_2) and the second can be biased when a number of unlikely stochastic events are required to explain the observed EPG(s) under a specific value of n.

Conceptually this work provides examples that show DNA profile evidence can be analysed assuming different numbers of contributors, and the results of these separate analyses can be combined by stratification for each proposition, using a model

weight, Z_n. The generation of Z_n will depend on how well each model describes the observed data, but also encompasses the additional profile frequencies required by the addition of contributors. The balance of these two factors ultimately dictates when one model is favoured over the other.

DNA profile analyses present some difficult challenges for MCMC methods, which other applications do not, namely:

- Time pressures for active casework results, typically results are desired with only minutes of runtime.
- Consistency of results between runs, especially challenging given the point above as the typical solution to this is to run the MCMC analysis for more iterations.
- A huge number of values that categorical parameters (genotype sets per locus) can take.
- The process must be able to work on a range of profile complexities using standard settings that do not require tuning from profile to profile.
- The process cannot be a 'black-box' as biology analysts with a wide variety of training (which usually does not include mathematics or statistics) must be able to implement the mathematics and explain the results in court.

Given these pressures we trial here calculating the Z_n values by calculation of the relative likelihood of one model over the other given the data using the method of Weinberg [25,26] which performs in an intuitively sensible manner given the tests we have run. Given all criteria mentioned above the method has worked well but we recognise the following issues:

(1) The process can require a number of MCMC iterations that may be impractical in forensic laboratories if using a range of contributors were to become a common event, rather than an exception. There is evidence of substantial variation in the results displayed in Fig. 5, although further investigation is already underway is aimed at reducing this variability further without requiring longer MCMC runs.
(2) The mathematics is approaching the complexity of a black-box. However, the conceptual introduction of a weight for the number of contributors is intuitively sensible and mathematics may not need to be fully understood to present these results in court. We have tried to structure the paper with this in mind, allowing readers to digest the majority of the paper without the need to refer to the complex mathematics in Supplementary material 2. It is thought by the authors that a conceptual level of understanding of what Z_n represents would be sufficient for almost any court challenge.

Another issue, although not relevant directly to the mathematics described in this paper, that will need to be addressed is one of communicating DNA results to court. Courts in the authors' countries have been used to scientists providing opinions on a number of contributors to a profile. Using a range of contributors indicates something which DNA statisticians have realised for some time but not had the means to act on, which is that the number of contributors to a DNA profile is not known and can never be known for most forensic casework. Further, the number of contributors does not need to be known to assess DNA evidence as long as the ambiguity in this number can be accounted for appropriately in the statistical model.

Acknowledgements

This work was supported in part by grant 2011-DN-BX-K541 from the US National Institute of Justice. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice. We would like to thank Ian Evett (Principal Forensic Services), Martin Weinberg (Department of Astronomy, University of Massachusetts) and Darfiana Nur (School of Computer Science, Engineering and Mathematics, Flinders University) for helpful discussions and contributions to this work. Finally we would like to thank two anonymous reviewers, whose comments improved this work.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.fsigen.2014.08.014.

References

[1] P. Gill, J.P. Whitaker, C. Flaxman, N. Brown, J.S. Buckleton, An investigation of the rigor of interpretation rules for STR's derived from less that 100 pg of DNA, Forensic Sci. Int. 112 (2000) 17–40.
[2] D.J. Balding, J. Buckleton, Interpreting low template DNA profiles, Forensic Sci. Int.: Genet. 4 (2009) 1–10.
[3] D. Balding, Evaluation of mixed-source, low-template dna profiles in forensic science, Proceedings of the National Academy of Sciences of USA. 110 (2013) 12241–12246.
[4] H. Haned, Forensim: an open-source initiative for the evaluation of statistical methods in forensic genetics, Forensic Sci. Int.: Genet. 5 (2011) 265–268.
[5] K. Lohmueller, N. Rudin, Calculating the weight of evidence in low-template forensic DNA casework, J. Forensic Sci. 58 (2013) 234–259.
[6] D. Taylor, J.-A. Bright, J. Buckleton, The interpretation of single source and mixed DNA profiles, Forensic Sci. Int.: Genet. 7 (2013) 516–528.
[7] M.W. Perlin, A. Sinelnikov, An information gap in DNA evidence interpretation, PLoS One 4 (2009) e8327.
[8] M.W. Perlin, M.M. Legler, C.E. Spencer, J.L. Smith, W.P. Allan, J.L. Belrose, et al., Validating TrueAllele® DNA mixture interpretation, J. Forensic Sci. 56 (2011) 1430–1447.
[9] B. Budowle, A.J. Onorato, T.F. Callaghan, A.D. Manna, A.M. Gross, R.A. Guerrieri, et al., Mixture interpretation: defining the relevant features for guidelines for the assessment of mixed dna profiles in forensic casework, J. Forensic Sci. 54 (2009) 810–821.
[10] C. Brenner, R. Fimmers, M.P. Baur, Likelihood ratios for mixed stains when the number of donors cannot be agreed, Int. J. Legal Med. 109 (1996) 218–219.
[11] S.L. Lauritzen, J. Mortera, Bounding the number of contributors to mixed DNA stains, Forensic Sci. Int. 130 (2002) 125–126.
[12] P.J. Green, Reversible Jump Markov Chin Monte Carlo Computation and Bayesian Model determination, Biometrika 82 (1995) 711–732.
[13] J.S. Buckleton, J.M. Curran, P. Gill, Towards understanding the effect of uncertainty in the number of contributors to DNA stains, FSI Genet. 1 (2007) 20–28.
[14] H. Haned, L. Pene, J.R. Lobry, A.B. Dufour, D. Pontier, Estimating the number of contributors to forensic DNA mixtures: does maximum likelihood perform better than maximum allele count? J. Forensic Sci. 56 (2011) 23–28.
[15] H. Haned, L. Pène, F. Sauvage, D. Pontier, The predictive value of the maximum likelihood estimator of the number of contributors to a DNA mixture, Forensic Sci. Int.: Genet. 5 (2011) 281–284.
[16] D.R. Paoletti, T.E. Doom, C.M. Krane, M.L. Raymer, D.E. Krane, Empirical analysis of the STR profiles resulting from conceptual mixtures, J. Forensic Sci. 50 (2005) 1361–1366.
[17] T. Tvedebrink, Overdispersion in allelic counts and θ-correction in forensic genetics, Forensic Sci. Int.: Genet. Suppl. Ser. 2 (2009) 455–457.
[18] J.-A. Bright, E. Huizing, L. Melia, J. Buckleton, Determination of the variables affecting mixed MiniFiler(TM) DNA profiles, Forensic Sci. Int.: Genet. 5 (2011) 381–385.
[19] J.-A. Bright, K. McManus, S. Harbison, P. Gill, J. Buckleton, A comparison of stochastic variation in mixed and unmixed casework and synthetic samples, Forensic Sci. Int.: Genet. 6 (2012) 180–184.
[20] J.-A. Bright, D. Taylor, J.M. Curran, J.S. Buckleton, Developing allelic and stutter peak height models for a continuous method of DNA interpretation, Forensic Sci. Int.: Genet. 7 (2013) 296–304.
[21] J.-A. Bright, J. Turkington, J. Buckleton, Examination of the variability in mixed DNA profile parameters for the Identifiler(TM) multiplex, Forensic Sci. Int.: Genet. 4 (2009) 111–114.
[22] S.J. Walsh, J.S. Buckleton, Autosomal microsatellite allele frequencies for a nationwide dataset from the Australian Caucasian sub-population, Forensic Sci. Int. 168 (2007) e47–e50.
[23] R.G. Cowell, T. Graversen, S.L. Lauritzen. Analysis of forensic DNA mixtures with artefacts, areXiv:13024404v2 [statME], 2013.

[24] D.A. Taylor, J.-A. Bright, J.S. Buckleton, The 'factor of two' issue in mixed DNA profiles, J. Theor. Biol. (2014), http://dx.doi.org/10.1016/j.jtbi.2014.08.021.
[25] M.D. Weinberg, I. Yoon, N. Katz, A remarkably simple and accurate method for computing the Bayes Factor from a Markov chain Monte Carlo Simulation of the posterior distribution in high dimension, areXiv:13013156v1 [astro-phIM], 14 January, 2013.
[26] M.D. Weinberg, Computing the Bayes factor from a Markov Chain Monte Carlo Simulation of the posterior distribution, Bayesian Anal. 7 (2012) 737–770.

Glossary

a: allele

c^2: variance constant used in the modelling of deviations of elements in $\mathbf{O}$ from elements in $\mathbf{E}$, $\log\left(\frac{O^l_{ar}}{E^l_{ar}}\right) \sim N\left(0, \frac{c^2}{E^l_{ar}}\right)$

$\bar{d}$: the distance chosen to encompass subset Ω_s around the mode of the entire posterior MCMC sample Ω

$d^{(y)}$: the distance of mass parameter values in posterior sample y from those in iteration y^*

D_n: the dimensionality of model n

$\mathbf{E}$: vector of expected peak intensities

H_x: proposition x

K: the number of different models being examine i.e. the range of potential contributor numbers

l: locus

LR: the likelihood ratio

$\mathbf{M}$: mass parameters; template amount for each contributor, degradation for each contributor, amplification efficiency for each locus, replicate amplification strength per replicate, stutter peak height variance and allele peak height variance

$\mathbf{M_n}$: mass parameters in model n

$\boldsymbol{M}_n^{(y*)}$: the set of mass parameter values that lead to the highest observed posterior likelihood in the MCMC sample Ω for model n

$M_{i,n}^{(y)}$: the value for mass parameter i at MCMC iteration y for model n (note the n can be dropped when talking about values all within a model)

n: number of contributors

N_n: model n

$\mathbf{N}$: vector of all models under consideration

$\mathbf{O}$: vector of observed peak intensities

P: the average value for a full profile genotype probability

r: replicate

R_r: replicate amplification efficiency for replicate r

$\mathbf{S_j}$: genotype set j

${}^{R}S$: a genotype set that is present in non-zero weights 100% of the time for a specific contributor position

${}^{U}S$: an unresolved genotype set, one where there are more than one possibility for a specific contributor position

$\boldsymbol{S_{U_j}}$: unknown, untyped contributor genotype set

$\mathbf{S_P}$: contributor genotype(s) known under H_1 but not H_2

$\mathbf{S_K}$: contributor genotypes known under both H_1 and H_2

$t_n^{(y)}$: the value for the template amount for contributor n in posterior sample y

V: the hyperrectangle volume

w_j: the weight for genotype set j

w_{jn}: the weight for genotype set j for model n

y^*: the iteration that lead to the highest observed posterior likelihood in the MCMC sample Ω

Z_n: a scalar for the weights in model n

$\Gamma(a, b)$: gamma distribution with shape of 'a' and a scale of 'b'

Ω: the entire posterior sample produced from the MCMC

$|\Omega|$: the number of individual samples in Ω

Ω_s: a subset of the posterior sample

$|\Omega_s|$: the number of individual samples in Ω_s

$\boldsymbol{\sigma}_q$: the vector of hyperrectangle edge lengths in iteration q of the algorithm outlined in supplementary material 2 section "Identifying Ω_s"

σ_{qp}: the hyperrectangle edge length in iteration q for mass parameter p

$\vee_{\boldsymbol{M_n},n}\, p(\boldsymbol{O})$: the maximum likelihood of the posterior MCMC sample Ω considering mass parameters $\boldsymbol{M}_n$ and model N_n

$\vee_y\, t_n^{(y)}$: the maximum observed template parameter value for contributor n in the y iterations of the posterior MCMC sample Ω

$\wedge_y\, t_n^{(y)}$: the minimum observed template parameter value for contributor n in the y iterations of the posterior MCMC sample Ω