likeLTD v6.1: an illustrative analysis, explanation of the model, results of validation tests and version history

David J. Balding, Christopher D. Steele
UCL Genetics Institute
Darwin Building, Gower Street
London WC1E 6BT
d.balding@ucl.ac.uk

June 5, 2016

Abstract

likeLTD ("likelihoods for Low Template DNA profiles") is an R package for computing likelihoods for DNA profiles. Version 6.0 included both a discrete model that uses allelic calls (present/uncertain/absent), with only minor changes from Version 5.5, and a new continuous model that uses the peak heights from an electropherogram. Both models can handle multiple profiled possible contributors and up to two unprofiled contributors, in addition to the queried contributor, as well as sporadic dropin. The continuous model explicitly accommodates stutter, double-stutter and over-stutter, which are typically called as uncertain or non-allelic when using the discrete model. The package also provides input files for example analyses (the "Laboratory case" described below).

This document describes the continuous model of likeLTD, including the modelling of peak heights, accounting for the effects of stutter and DNA degradation, as well as installation and running of the software. Much of the information in this guide has been published (Steele and Balding, 2016), which additionally includes a test of the linkage adjustment we propose, investigation of the behaviour of the continuous model in relation to multiple replicates, and comparison to theoretical predictions of the model. For corresponding information about the discrete model see the guide for Version 5.5 and Balding (2013). For background on forensic DNA profiling see Butler (2010), and for introductions to statistical methods for evaluating DNA profile evidence see Buckleton et al. (2004); Balding and Steele (2015).

We present some comparisons of results from running both continuous and discrete models on a range of single-contributor and mixed laboratory-generated DNA profiles. We also present results from the continuous model on a subset of those profiles subject to modifications, such as alteration of heights of individual peaks, or inclusion of extra peaks. All results reported here, unless otherwise stated, are from running Version 6.1 of likeLTD, with a standard allele frequency database of around 7000 UK Caucasians, $F_{ST} = 0.03$, a sampling adjustment adj = 1, and a detection threshold of 20 RFU for all loci.

Changes in v6.1

Version 6.1 achieves large improvements in computational efficiency by fine tuning some of the optimisation parameters of DEoptim and our convergence criteria. Furthermore, a crude estimation of the DNA contribution of unknown contributors through k-means clustering was added to the allele report, so that the allele report can give guidance on whether or not some unknown contributors can be modelled as dropin rather than a full contributor. If the estimated contribution of an unknown is $< 1/3$ that of Q, then we suggest that they can be treated as dropin.

Contents

1 Installation and example R script **3**
1.1 Input . . . 3
1.2 Allele report . . . 5
1.3 Arguments and optimisation . . . 5
1.4 Output report . . . 8
1.5 Computing time . . . 8

2 The likeLTD peak height model: overview **8**
2.1 Key features of likeLTD . . . 8
2.2 The contributors of DNA . . . 9
2.3 The parameters . . . 9
2.4 Dropin model . . . 10

3 The likeLTD model: further details **10**

4 Validation **11**
4.1 Model fit . . . 12
4.2 Altering the model for the example analysis . . . 12
4.3 Laboratory validation . . . 14
4.3.1 Single contributor . . . 15
4.3.2 Two contributors . . . 16
4.3.3 Three contributors . . . 17
4.3.4 Runtime . . . 18
4.3.5 Minor as dropin . . . 18
4.3.6 Summary . . . 19
4.4 Validation using artificial changes to input data . . . 19
4.4.1 Missing peak insertion . . . 19
4.4.2 Altering observed peaks . . . 21
4.4.3 Dropin peak insertion . . . 22
4.4.4 Summary . . . 24

5 Real case comparison: Meredith Kercher **24**
5.1 Case circumstances . . . 24
5.2 Results . . . 25
5.2.1 Raffaele Sollecito . . . 25
5.2.2 Amanda Knox . . . 26
5.3 Conclusions . . . 26
5.4 Validation summary . . . 26

6 Acknowledgements and version history **26**

A Allele report for Laboratory case **31**

B Output file for Laboratory case **38**

C Laboratory protocol **48**

1 Installation and example R script

Installing `likeLTD` (only needs doing once on any computer) and loading it (once per `R` session) are both very simple.

```
install.packages("likeLTD")
require(likeLTD)
```

The `install.packages` command may generate a request for you to choose a site from which to download the package. Choose any site near you.

The example analysis that comes with `likeLTD` is that of a laboratory-created mixture of three individuals, where each individual contributes approximately 250, 62 and 16pg of DNA. Reference profiles are available in input file `laboratory-reference.csv` for the 250pg contributor, who we treat as a known individual, and the 16pg contributor who we treat as the queried individual. The 62pg contributor is treated as unknown for all analyses (no reference profile is provided). The crime scene profile (CSP) consists of a single profiling run at the 17 loci of the NGM SElect^{TM} PCR amplification kit and is available in input file `laboratory-CSP.csv`. We wish to evaluate the evidence against the queried individual (Q) using a likelihood ratio (LR) of the form:

$$\text{LR} = \frac{\Pr(E|H_p)}{\Pr(E|H_d)} \tag{1}$$

where E is the DNA evidence (CSP and reference profiles), H_p is the prosecution hypothesis that assumes Q is a contributor to the CSP and H_d is the defence hypothesis that assumes an unknown individual, X, is a contributor to the CSP instead of Q. The hypotheses may specify a number of known contributors (K) and unknown contributors (U). LRs will be presented throughout as $\log_{10} LR$, which gives the weight of evidence (WoE) in bans.

The CSP shows a total of 38 peaks that are not attributable to either of the reference individuals, 15 of which appear to be allelic peaks (see Section 1.2 for criteria). No more than two of these estimated allelic peaks occurs at any one locus. This suggests a comparison of the following two hypotheses for the contributors of DNA to the sample:

$$\begin{aligned} H_p: &\quad \text{Q + K1 + U1} \\ H_d: &\quad \text{X + K1 + U1} \end{aligned}$$

where Q, X, K1 and U1 are all assumed unrelated to each other. In this example we know from the experimental design that H_p is true. If extra individuals are included in H_p beyond the true number of contributors, this can add to the computational cost but there will be little impact on the WoE because the amount of DNA from the additional contributors will be estimated to be small. Cowell et al. (2013) illustrate this with an example in which a log(LR) of 14.09 with three contributors barely changes as the number of contributors increases, reaching 14.04 with eight contributors.

1.1 Input

We now show how to calculate likelihoods under H_p and H_d using `likeLTD`. The first command below finds out where your system has stored the Laboratory case files, and saves that location in `datapath`. For your own analyses, you will need to create your own CSP and reference files, in the same format as `laboratory-CSP.csv` and `laboratory-reference.csv`. It is usually most convenient to create these files in a specific directory, and then set that to be the working directory for `R` using the command `setwd()` or using the `R` menu option (its location varies across operating systems). For example if your case files are in the directory `C:/Users/JoeBloggs/Cases/JoeBloggs1` then you enter the command `setwd("C:/Users/JoeBloggs/Cases/JoeBloggs1")`. In that case you can set `datapath = "."` in place of the first command below. A number of allele frequency database files are provided with `likeLTD`. To use your own database file instead

(must be in same format) set `databaseFile` to the filename, including path if not in the working directory. If you wish to choose a different individual to be Q, or to add or omit a profiled contributor then you must create a new reference file containing all relevant reference profiles with one tagged as "queried" and the others as "known" under the second column titled "known/queried". Homozygous loci for reference individuals should be input as two separate alleles e.g. "16,16" rather than "16".

```
datapath = file.path(system.file("extdata", package="likeLTD"),"laboratory")

# File paths and case name for allele report
admin = pack.admin.input.peaks(
            peaksFile = file.path(datapath, 'laboratory-CSP.csv'),
            refFile = file.path(datapath, 'laboratory-reference.csv'),
            caseName = "Laboratory",
            detectionThresh = 20
             )
```

The possible arguments for `pack.admin.input.peaks` are

`peaksFile`: Path to CSP file with peak heights. No default.

`refFile`: Path to file with reference profiles. No default.

`caseName`: Case name. Defaults to "dummy".

`databaseFile`: Path to database file. Defaults to `NULL`.

`kit`: Choice of database supplied with `likeLTD`. Can take the values "DNA17","SGMplus","Identifiler" and `NULL` (which is appropriate if `databaseFile` is set). Only used if `databaseFile=NULL`, at which point `kit` will default to "DNA17".

`linkageFile`: Path to file containing recombination rates between linked loci. Defaults to `NULL`, at which point the linkage file supplied with `likeLTD` will be used.

`detectionThresh`: Detection threshold used for analysing peaks. This is either a single value that is applied across all loci, or a named list giving the detection threshold at each locus. Defaults to a single value of 20 RFU.

`outputPath`: Output path for reports. Defaults to the current working directory.

In the script shown here, neither `databaseFile` nor `kit` has been specified, so `likeLTD` will use the NGM SElect™ database provided with the package, which is the `"DNA17"` database. If you wish to specify a different `detectionThresh` for each lane of the CSP the admin specification should be similar to:

```
# File paths and case name for allele report
admin = pack.admin.input.peaks(
            peaksFile = file.path(datapath, 'laboratory-CSP.csv'),
            refFile = file.path(datapath, 'laboratory-reference.csv'),
            caseName = "Laboratory",
            detectionThresh = list(D10S1248=20,vWA=20,D16S539=20,D2S1338=20, # blue
                                   D8S1179=30,D21S11=30,D18S51=30, # green
                                   D22S1045=40,D19S433=40,TH01=40,FGA=40, # black
                                   D2S441=50,D3S1358=50,D1S1656=50,D12S391=50,SE33=50) # red
            )
```

Designation	S	DS and OS
Non-allelic	$x < 0.05$	$x < 0.05$
Uncertain	$0.05 \leq x < 0.15$	$0.05 \leq x < 0.1$
Allelic	$x \geq 0.15$	$x \geq 0.1$

Table 1: Criteria for designating alleles in stutter (S), double-stutter (DS) or over-stutter (OS) positions as either non-allelic, uncertain or allelic when estimating `nUnknowns`. x indicates the ratio of the stutter position peak height to the parent peak height.

1.2 Allele report

```
# Next we generate an allele report
allele.report.peaks(admin)
```

The allele report is a .doc that will be created in the current working directory (set `outputPath` to specify a different directory). The report generated by the above command (`Laboratory-Allele-Report-1.doc`) is shown in Appendix A. It summarises the input data, highlights rare alleles, and suggests values for key parameters (and hence suitable hypotheses to compare), in particular specifying the number of unprofiled contributors required to explain the observed CSPs under H_p, and whether to model dropin or not. Peaks are called as non-allelic, uncertain or allelic according to the criteria given in Table 1. These calls are not used in computing the WoE. Note that the assumptions are based on modelling both over- and double-stutter, if these are not modelled manual evaluation of the number of unknown contributors should be performed. The peaks that are called as allelic are then used to suggest the number of unprofiled contributors and whether or not to model dropin. Here, the allele report indicates that one unknown contributor is sufficient under H_p to explain the observed alleles not attributable to Q or K1 or possible stutters of the two profiled contributors.

Version 6.1 adds an estimation of the DNA contribution of any unknown contributors based on k-means clustering on the heights of unattributable called-allelic peaks, where k is the minimum number of unknowns based on those peaks. If the DNA contribution is estimated at $< 1/3$ that of Q, we recommend that the unknown may be explainable as dropin. See Section 4.3.5 for a demonstration that explaining minor contributors as dropin has little-to-no effect on the resulting WoE against a non-minor Q.

1.3 Arguments and optimisation

Based on the allele report we specify the required hypotheses by setting the following arguments:

`nUnknowns`: The number of unknown contributors under the prosecution hypothesis (either 0, 1 or 2). `likeLTD` automatically adds an additional unknown contributor (X) under the defence hypothesis, who replaces Q from the prosecution hypothesis. Defaults to 0.

`doDropin`: Whether to model dropin or not (logical: `TRUE` or `FALSE`). Defaults to `FALSE`.

`ethnic`: The ethnic category of the queried contributor. The default database comes with "NDU1" (Caucasian),"NDU2" (African + Afro-Caribbean), "NDU3" (South Asian), "NDU4" (East Asian), "NDU6" (African) and "NDU7" (Afro Caribbean). If you use your own allele frequency database you will choose your own category labels (required even if there is only one category). Defaults to "NDU1".

`adj`: Sampling adjustment (scalar). Defaults to 1.

`fst`: F_{ST} adjustment (scalar) for distant relatedness (coancestry) of Q and X. Defaults to 0.03.

`relationship`: Assumed relationship between Q and X. Can take values between 0 and 7 (defaults to 0):

0 Unrelated.

1 Parent/offspring.
2 Siblings.
3 Uncle (or aunt)/nephew (or niece).
4 Half-uncle (or half-aunt)/half-nephew (or half-niece).
5 Cousins.
6 Grandparent/grandchild.
7 Half-siblings.

The direction of relationships does not alter the computation, so is unspecified e.g. if `relationship=1`, the relationship may be Q as parent and X as offspring, or Q as offspring and X as parent.

`combineRare`: Whether to combine rare alleles that have not been observed in the CSP or reference profiles (logical: `TRUE` or `FALSE`). Defaults to `TRUE`.

`rareThreshold`: Allele probability below which unobserved database alleles will be combined when `combineRare` is set to `TRUE`. Defaults to 1, meaning all unobserved database alleles will be combined.

`doDoubleStutter`: Whether to model double-stutter (stutter to two repeat units smaller than the parent peak) or not. Defaults to `TRUE`.

`doOverStutter`: Whether to model over-stutter (stutter to one repeat unit larger than the parent peak) or not. Defaults to `TRUE`.

```
# Enter arguments
args = list(
        nUnknowns = 1
        )

# Create hypotheses
hypP = do.call(prosecution.hypothesis.peaks, append(admin,args))
hypD = do.call(defence.hypothesis.peaks, append(admin,args))

# Get parameters for optimisation
paramsP = optimisation.params.peaks(hypP)
paramsD = optimisation.params.peaks(hypD)

# Run optimisation
results = evaluate.peaks(paramsP, paramsD)
```

Only values that you wish to be different from the default must be specified in `args`, which for the example shown is only `nUnknowns`. If instead we wished to model dropin, use a South Asian database and not model double- or over-stutter the `args` list would look like:

```
# Enter arguments
args = list(
    nUnknowns = 1,
    doDropin = TRUE,
    ethnic = "NDU3",
    doDoubleStutter = FALSE,
    doOverStutter = FALSE
    )
```

If `combineRare=TRUE` and `rareThreshold=1` the program combines all alleles in the database that were not observed in the CSP or reference profiles into a single allele labelled "-1", which is given the mean LUS and BP of the combined alleles, and the sum of their probabilities. During computation if a joint genotype allocation shares n "-1" alleles, these are assumed to be n distinct alleles e.g. peak heights of unobserved alleles do not stack. Fewer unobserved alleles are combined if `rareThreshold` < 1; setting `combineRare=FALSE` can greatly increase runtime when `nUnknowns=2`. Observed alleles that do not have either a LUS or BP value specified in the database used will have these values extrapolated. LUS values will be extrapolated from the allele closest in size that shares the same partial repeat (allele 15 will be extrapolated from allele 14, allele 15.1 will be extrapolated from allele 16.1), if no allele shares the same partial repeat as the allele then the LUS value will be extrapolated from the allele closest in size.

The function `do.call` calls the function given in its first argument. `prosecution.hypothesis.peaks` and `defence.hypothesis.peaks` are both functions defined within `likeLTD`, which generate the necessary objects for H_p and H_d respectively.

The function `optimisation.params.peaks` sets the parameters needed for optimisation. These values can be altered if required but the default settings should be adequate for most analyses. One possible exception is the argument `maxDropin` of `optimisation.params.peaks` which has a default of 100. This is the maximum total contribution of dropin to peak heights summed over all alleles (in RFU). This will typically be adequate for the very low levels of dropin that typically arise. However we have shown (Steele and Balding, 2016) that minor contributors of DNA not of interest to the court can adequately be modelled as dropin (see Section 4.3.5), which may necessitate an increase in `maxDropin`.

Linked loci are those that are located close enough on the same chromosome that they are sometimes inherited as a block rather than as independent markers; if unaccounted for this tends to lead to an overstatement of the evidence against a Q who is closely related to an alternative contributor X. If the CSP contains linked loci, and Q and X are closely related (e.g. `args$relationship=2`) `likeLTD` will apply a correction factor to the LR of m_l/m_u where m_u is the match probability ignoring linkage and m_l is the match probability including linked loci, see Bright et al. (2013a) for full calculations of linked locus match probabilities. This correction can be turned off by specifying the argument `doLinkage=FALSE` to `optimisation.params.peaks`.

The `evaluate.peaks` function is defined within `likeLTD`, and is a wrapper function for the `DEoptim` function that performs optimisation. The `evaluate.peaks` function splits the convergence into a number of steps, with each subsequent step having more stringent convergence tolerance and an increased crossover rate (a parameter for `DEoptim`); the combination of these two behaviours means that the parameter space is searched extensively to start with, and gradually focuses to a more intensive local search towards the end. The program stops running new steps after convergence has been reached, which is defined as having a relative difference between the current step result and all of the last `nConverged` (defaults to four) steps results less than `tolerance` (defaults to 1e-6). Interim results after each step are available when the argument `interim` is set as `TRUE` (default), which writes the most recent results to `Interim.csv`, and saves the internal state of the `evaluate.peaks` function to `interim.RData`, in the current working directory. The file `interim.RData` can then be handed to `evaluate.from.interim.peaks` to restart a computation that has partially completed. The seed to be used for optimisation may be specified by handing the `seed.input` argument to `evaluate.peaks`; if this argument is not specified then `likeLTD` sets the seed to a numeric representation of the current date, time and process ID.

The object returned by `evaluate.peaks` is a list of seven elements: `Pros`, `Def`, `WoE`, `Lp`, `Ld`, `seed.used`, `seed.input` and `runtime`. `Pros` and `Def` correspond to the prosecution and defence results respectively, and have the same structure as the object returned by `DEoptim` (see `help(DEoptim)`). The final WoE (in bans) can be obtained through the command `results$WoE[length(results$WoE)]`. `Lp` and `Ld` give the prosecution and defence likelihoods at each step. `seed.used` gives the seed that was used by the optimisation, while `seed.input` is `NULL` if no seed was specified but gives the user defined seed otherwise, so should be the same as `seed.used`. `runtime` is a list of three elements: `elapsed`, `start` and `end`.

Hypothesis	Model	WoE	Runtime (mins)
Q/X + K1 + U1	Discrete	2.3	14
	Peak Height	8.2	23
Q/X + U1 + U2	Discrete	0.5	38
	Peak Height	7.8	200

Table 2: Runtime and WoE using both the peak height and discrete models for the Laboratory case assuming the hypotheses stated in the first column.

1.4 Output report

```
# Generate output report
output.report.peaks(hypP,hypD,results)
```

The results are given in the output file `Laboratory-Evaluation-Report-1.doc` (the numbering of the filename increments automatically, or a custom filename may be specified with `file="fileName.doc"`) which again summarises the input data, similar to the allele report, but also states the hypotheses compared and gives single-locus and overall LRs in favour of the prosecution hypothesis relative to the defence hypothesis, as well as overall WoE. The output file for the Laboratory case analysis is given in Appendix B.

The estimated DNA contributions are 152-172 RFU for Q/X, 982-993 RFU for K1 and 903-932 RFU for U1, indicating a minor contributor (Q/X) and two major contributors (K1 and U1). These values are not fully consistent with the target DNA input amounts of 16pg (Q/X), 250pg (K1) and 62pg (U1), however, a visual inspection of the CSP does indicate a minor contributor and two almost equal template major contributors. The discrepancy between intended DNA input and actual CSP contributions may be due to pipetting variability or some other source of error. See Section 4.1 for further discussion of the results obtained for this case.

1.5 Computing time

The peak height model has longer run times than the discrete model, but gives significantly higher WoEs for this case (Table 2). The run time for the peak height model scales with the number of observed peaks, the number of unknown contributors and the number of replicates in the CSP. Modelling over-stutter or double-stutter increases run time. The run times in Table 2 were obtained on a desktop computer with 15Gb of RAM, and an eight core Intel i7 processor (at 3.1GHz per core). Computing times may vary across machines. See Sections 4.3.4 and 5 for further information regarding run times.

2 The `likeLTD` peak height model: overview

2.1 Key features of `likeLTD`

Some key features of `likeLTD`:

- `likeLTD` uses peak height information directly; providing similar or greater statistical efficiency than the discrete model (which remains available and was the only model prior to v6.0). There is a substantial improvement in statistical efficiency relative to the discrete model for some CSPs.
- It combines information across all DNA profiling runs, thus avoiding the need for a "consensus" profile (Gill et al., 2000).
- DNA dose can decrease with fragment length due to degradation, based on the model of Tvedebrink et al. (2012).

- Stutter ratio has a linear relationship with longest uninterrupted sequence (LUS), as demonstrated by (Kelly et al., 2014), and this relationship is allowed to differ both across loci and across replicates.
- As a consequence of estimating the DNA contribution, a potential contributor can be considered in a hypothesis without implying that their DNA is present, because the contribution of DNA from that individual can be estimated at zero.
- Because the penalised likelihoods are maximised over the nuisance parameters, combining information over alleles, loci, replicates and individuals, there is little need for external calibration data. This is only required for a few hyperparameters – the parameters of the penalty functions. The underlying parameters are allowed flexibility to best fit the CSP data under each hypothesis, constrained by penalty functions that depend on these hyperparameters.

2.2 The contributors of DNA

Given the CSP and reference profiles, we seek to compare the likelihood of the CSP when a profiled individual Q is a contributor with the corresponding likelihood when Q is replaced by an unprofiled individual X. The ratio of those two likelihoods, each maximised over the nuisance parameters, is the likelihood ratio (LR). There can be up to two further unprofiled possible contributors of DNA, U1 and U2, and multiple profiled uncontested contributors (K1, K2, ...).

There can be several LRs of interest, considering X of different ethnicities and different relatedness with Q (the more genetically similar X is to Q, the smaller the LR). `likeLTD` allows X to be related to Q with the specification of one of eight possible relationships. In addition, we use an F_{ST} adjustment to allele fractions that allows for possible remote shared ancestry of Q with X. Within `likeLTD`, this adjustment only affects the alleles of Q and does not take into account any other profiled contributors. We assume U1 and U2 to be mutually unrelated, and they and the K are all assumed unrelated to X (when these individuals are included in both hypotheses, any relatedness to X will usually have little effect on the LR).

Because the relatedness coefficients and F_{ST} account for the positive correlations across loci due to shared ancestry of Q and X, it is reasonable to compute full-profile LRs by multiplication of single-locus LRs, which is standard practice in the assessment of DNA profile evidence (Buckleton et al., 2004). We thus focus below on the single-locus case.

2.3 The parameters

The "nuisance" parameters, which must be eliminated under each multi-locus likelihood before taking their ratio, are

- the DNA contributions of each hypothesised contributor in RFU.
- the parameters of the stutter model; mean gradient and multiplicative locus adjustment.
- the mean double- and over-stutter fraction, if modelled.
- one degradation parameter for each hypothesised contributor, and one degradation parameter for dropin peaks, if modelled.
- a multiplicative replicate adjustment; one for each replicate after the first, with the first as the "reference" replicate.
- a dropin dose (RFU), if modelled.
- the scale parameter for the gamma distribution, used to compute probabilities of observed peak heights given the expected peak height.

`likeLTD` maximises a (penalised) likelihood over these parameters using the `R DEoptim` function.

2.4 Dropin model

The dropin parameter in `likeLTD` is the expected total contribution of dropin to peak heights at a locus in one profiling run. Because dropin is ubiquitous for low-template profiles the default minimum dropin dose is 5 RFU and the default maximum is 100 RFU. Dropin of a given allele is assumed to occur in proportion to the frequency of that allele in the population, so if we have a given environmental DNA load in RFU, λ, then for each allele, i, in the population database we expect $p_i\lambda$ RFU dropin dose in each replicate. This dropin dose is subject to degradation at a separate rate to that of non-dropin doses, given by $(1+\delta)^{-f_i}$, where δ is the dropin specific degradation parameter and f_i is the mean adjusted fragment length for allele i in base pairs.

3 The `likeLTD` model: further details

Computations are performed separately under H_p and H_d. Let C denote the set of contributors under a given hypothesis. Suppose that the CSP replicates are indexed by the elements of a set R, and include loci in the set L, while I_l denotes the set of possible alleles at locus $l \in L$. Each element of G_l is an allocation of genotypes at locus l to each $c \in C$. The genotype of Q is constant over G_l, and similarly for other c with reference profile available, but the elements of G_l vary according to the genotypes allocated to unprofiled c. Population genotype probabilities are assumed given. In practice, allele probabilities are obtained from a database, possibly using a sampling adjustment, and genotype probabilities are derived as products of allele probabilities assuming Hardy Weinberg equilibrium, possibly with an F_{ST} adjustment (Balding and Steele, 2015).

Let χ_c denote the effective DNA mass at a heterozygote allele of $c \in C$ in the first replicate, expressed in RFU, a unit of peak height. To compute the expected contribution from c to the height of an epg peak at allele $i \in I_l$ for a given $g \in G_l$, we first adjust for the genotype of c specified by g, the replicate $r \in R$, and DNA degradation:

$$P_{l,r,g,c,i} = \frac{n_{g,c,i}\rho_r\chi_c}{(1+\delta_c)^{f_i}}, \tag{2}$$

where $n_{g,c,i} \in \{0,1,2\}$ indicates the number of i alleles in the genotype of c and ρ_r denotes a replicate adjustment ($\rho_1 = 1$), while δ_c is a parameter measuring the degradation of DNA from c and f_i is the mean adjusted length of allele i in base pairs. Each $P_{l,r,g,c,i}$ must next be adjusted for the fractions that stutter to allelic position $i-1$ (S), double-stutter to $i-2$ (D) or over-stutter to $i+1$ (O). Whereas D and O are global constants, because these are rare events and it would be difficult to parametrise the relationship, we propose a zero-intercept linear model for S:

$$S_{l,i} = \alpha_l u_i.$$

Here, α_l is the locus-specific coefficient of u_i, the longest uninterrupted sequence (LUS) of allele i (Brookes et al., 2012; Bright et al., 2013b; Kelly et al., 2014). To compute the expected peak height at allele i in replicate r for a given g, each $P_{l,r,g,c,i}$ is incremented with any stutter contribution from allele $i+1$, double stutter from $i+2$ and over-stutter from $i-1$, and summed over contributors c. Finally, a contribution from dropin is added. This gives the expected peak height as:

$$E_{l,r,g,i} = \frac{\lambda p_i}{(1+\delta)^{f_i}} + \sum_{c\in C}((1-S_{l,i}-D-O)P_{l,r,g,c,i} + S_{l,i+1}P_{l,r,g,c,i+1} + DP_{l,r,g,c,i+2} + OP_{l,r,g,c,i-1}). \tag{3}$$

where p_i is the population allele fraction and λ is a dropin parameter, in RFU. Note that dropin of an allele is assumed to occur in proportion to its population frequency, and is adjusted for degradation with a dropin-specific rate δ.

The peak height at allelic position i is then assumed to have a gamma distribution with expectation $E_{l,r,g,i}$ and variance $\sigma E_{l,r,g,i}$. The scale parameter σ is a global constant, so that values of l, r, g and i affect

peak-height variance only through the mean. In `likeLTD` we treat peak heights as discrete: observed values are recorded to the nearest integer RFU value, say j, and we compute the corresponding probability as the gamma probability mass between $j-0.5$ and $j+0.5$. The dropout probability is the gamma probability mass assigned to the interval $(0, t_l-0.5)$, where t_l is the detection threshold (the smallest recordable peak height).

In `likeLTD`, alleles that are not observed in any CSP replicate or any reference profile of an assumed contributor are combined into a single allelic class. When the unprofiled contributors are assigned > 1 allele in this class, these are assumed to be distinct: unprofiled contributors are assumed not to share any unobserved allele.

Parameter	Distribution	Mean	SD
$\mathrm{E}[\alpha_l]$	N	0.013	0.010
$\log_{10}(\alpha_l/\mathrm{E}[\alpha_l])$	N	0	0.300
D	Γ	0.02	0.019
O	Γ	0.02	0.019
δ	e	0.02	0.020
σ	e	100	0.010

Table 3: Penalties applied to the parameters of the peak-height model. Distributions: N=normal, Γ=gamma, e=exponential. The degradation parameters δ have the same penalty for each contributor and for dropin.

In order to encourage the optimisation algorithm to search in realistic regions of the parameter space, the penalty terms shown in Table 3 are imposed. Large values of δ and σ are penalised, while for both D and O a zero value is excluded but a broad range of positive values is supported. Two separate penalties on the α_l are intended to allow flexibility for its mean while limiting its variance over loci. Incorporation of these penalty terms into the likelihood function is analogous to imposing a prior distribution, but our approach is not Bayesian: elimination of nuisance parameters is achieved via maximisation and not integration, which is for example the approach adopted by `STRmix`, implemented using Markov chain Monte Carlo.

The probability assigned to allelic position i, whether or not there is an observed above-threshold peak, is computed as a gamma probability mass as described above. Denoting this probability $a(l, r, g, i, \sigma)$, the penalised likelihood is computed by multiplying over alleles and replicates, summing over genotype allocations each multiplied by the product of genotype probabilities for the unprofiled contributors, and then multiplying over loci including the penalty term:

$$\prod_{l \in L} \pi_l \sum_{g \in G_l} \Big[\prod_{c \in C} Pr(\mathcal{G}_{g,c}) \Big] \prod_{r \in R} \prod_{i \in I_l} a(l, r, g, i, \sigma) \tag{4}$$

where $\mathcal{G}_{g,c}$ denotes the genotype allocated to c in g, while π_l is the combined penalty on the likelihood at locus l given the values for α_l, D, O, σ and the δ. (4) is then maximised over these parameters. `likeLTD` uses a genetic algorithm `Deoptim` that simulates mutation, recombination and selection on parameter vectors to search for the vector that maximises the penalised likelihood (Mullen et al., 2011). Maximisation is performed separately under H_p and H_d and the LR is the ratio of the maximised values.

4 Validation

To validate the peak height model we have carefully designed a series of tests to verify that the model adheres to expected behaviours under a number of conditions.

Using the Laboratory case (see Section 1) we have verified that the optimised model adheres to expected behaviour (see Section 4.1). Still looking at the Laboratory case, we have then altered the model assumptions used to run the case and ensured that any resulting change in the WoE or lack thereof is consistent with the altered model assumptions (see Section 4.2).

Next, we have generated a large number of laboratory CSPs ranging from one to three contributors, and have compared the WoE under both peak height and discrete models for each contributor in each CSP (see Section 4.3), and expect a greater WoE using the peak height model for unequal-contribution CSPs and similar WoEs for equal-contribution CSPs. Subsequently, we have altered a single peak at a time in a single of the laboratory-generated CSPs and evaluated the WoE using the peak height model (see Section 4.4), and expect introducing peaks congruent with H_p to increase the WoE, removing peaks congruent with H_p to decrease the WoE and introducing peaks incongruent with H_p to decrease the WoE.

Lastly we used the peak height model to evaluate the epg obtained from the bra clasp, item 165B, from the Meredith Kercher case, using `likeLTD`, `STRmix` and `EuroForMix`; all continuous models. We expect the three models to give similar answers to each other (see Section 5).

4.1 Model fit

For the Laboratory case (Appendixes A and B), `likeLTD` returns a WoE of 8.2 bans (Table 4, column 1) indicating extremely strong support for H_p, despite the low DNA contribution of approximately 16pg for Q and the complex nature of the CSP. The strongest support for H_p is seen at D21 and D12; both loci where the alleles of Q are not masked by allelic peaks of the major contributors. Conversely, D22 and D19 support H_d; at D22 Q is homozygous and masked by a major allele, so `likeLTD` explains the over-stutter at 17 as allelic for X under H_d (with a correspondingly lower $\widehat{\theta}$ under H_d), while at D19 Q has dropped out an allele while the corresponding 15 allele is observed unmasked which `likeLTD` finds to be more likely explained as X being heterozygous for 15 and a non-15 allele that is masked by one of the major contributors.

We assess the fit under H_d of the optimised `likeLTD` model for this case by observing the fraction of observed peak heights that lie within the central 50% and 95% intervals of their fitted gamma distributions, given both the most likely joint genotype allocation and the fitted parameters. The fit of the optimised parameters to the observed data can be investigated using the `peaks.results.plot` function included with `likeLTD`. This function plots boxplots for each hypothesised peak assuming the most likely joint genotype allocation, with boxes displaying the central 50% (inter-quartile range) of the gamma distribution, whiskers displaying the 95% equal-tailed probability interval, and red bars indicating the observed peak heights.

For the Laboratory case under H_d the proportions of observed peaks within the 50% and 95% probability intervals were 0.51 and 0.94, both close to their respective expected value (Figure 1).

4.2 Altering the model for the example analysis

Here we alter the assumptions of the model used to evaluate the WoE in the Laboratory case, modelling all combinations of double- and over-stutter with dropin, and removing the locus dependency of the stutter gradient. We expect that removing modelling assumptions that have no explanatory power for the given CSP to return an unaltered WoE, while removing modelling assumptions that are necessary to fully explain the CSP will results in an altered WoE.

Modelling dropin does not change the WoE for the laboratory case (Table 4), as dropin is not necessary to explain the CSP when double- and over-stutter are both modelled, as evidenced by the dropin estimates of 5 and 5 RFU under H_p and H_d respectively, equal to the minimum dropin value of 5.0. Similarly, removing double-stutter from the model does not change the WoE as there are no peaks in the CSP that can only be explained through double-stutter. Conversely, removing over-stutter from the model reduces the WoE, particularly because the 17 peak at D22 can no longer be explained by over-stutter (D22 WoE decreases from -0.5 bans with SDO to -0.8 and -0.7 bans with SD and S respectively), so must be assumed to be allelic by the program. D22 is subject to over-stutter more commonly than any other locus in the NGM Select^{TM} kit due to being the only locus with repeat units that are three base pairs long, rather than the standard four base pairs. In the peak height model the stutter ratio is assumed linear with the longest uninterrupted sequence (LUS) of the allele, with the gradient of the linear relationship allowed to differ between loci. When the stutter gradient is instead assumed to not vary between loci ($\alpha_l = E[\alpha_l]$) the WoE increases to 8.4 bans. This change in WoE is driven by the defence likelihood at D2S1338; at this locus Q is 17,22 but the most likely genotype for X is 17,18 meaning that the truly allelic peak at 22 is estimated

Model	SDO	SDO+dropin	SO+dropin	SD+dropin	S+dropin	$\alpha_l = E[\alpha_l]$
Parameters						
Dropin	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE
DS	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE
OS	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE
WoE						
D10S1248	0.6	0.6	0.6	0.6	0.6	0.6
vWA	1.0	1.0	1.1	1.1	1.1	1.1
D16S539	0.5	0.5	0.5	0.5	0.5	0.5
D2S1338	0.5	0.5	0.5	0.4	0.4	0.6
D8S1179	1.1	1.1	1.1	0.9	0.9	1.1
D21S11	1.7	1.7	1.7	1.7	1.7	1.7
D18S51	1.0	1.0	1.0	1.0	1.0	1.1
D22S1045	-0.5	-0.5	-0.5	-0.8	-0.7	-0.5
D19S433	-2.8	-2.8	-2.8	-2.8	-2.8	-2.8
TH01	0.5	0.5	0.5	0.5	0.5	0.5
FGA	0.6	0.6	0.6	0.5	0.5	0.5
D2S441	1.2	1.2	1.2	1.2	1.2	1.2
D3S1358	0.7	0.7	0.7	0.7	0.7	0.7
D1S1656	0.7	0.7	0.7	0.7	0.7	0.7
D12S391	1.4	1.3	1.4	1.3	1.3	1.3
SE33	0.1	0.1	0.1	0.1	0.1	0.1
Overall	8.2	8.2	8.2	7.6	7.7	8.4

Table 4: Locus and overall WoE for the Laboratory case provided with `likeLTD`, under different modelling assumptions. Columns four to six alter whether double or over stutter are being modelled while in column seven the stutter gradient is constant over loci (see Section 3).

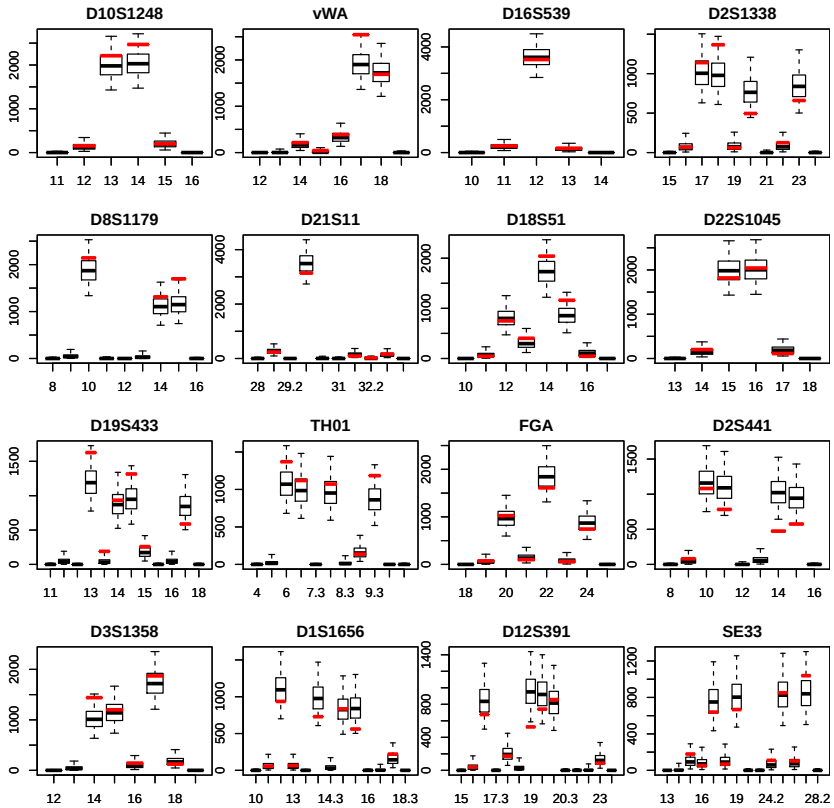


Figure 1: Boxes show the central 50% (inter-quartile range) of the gamma distribution for each hypothesised peak, whiskers represent the 95% equal-tailed probability interval and red bars show observed peak heights. RFU is displayed on the y-axes while allele labels corresponding to boxplots are displayed on the x-axes.

to be stutter from one of the majors under H_d (K1=18,23), requiring a large stutter gradient which is not possible when the stutter gradient cannot vary by locus. This means that the defence hypothesis has a higher likelihood at D2S1338 when the stutter gradient is allowed to vary by locus, leading to a lower locus LR with a locus variant gradient (0.46) than with a fixed gradient (0.61).

We have demonstrated here that modelling dropin and removing the modelling of double-stutter does not change the WoE for the Laboratory case, as these phenomena are not required to explain this particular CSP. Conversely removing the modelling of over-stutter or locus-dependent stutter gradients has an effect on the WoE as these phenomena are important in explaining the CSP under either H_p or H_d. This fits the expected behaviour of explanatory modelling assumptions altering the WoE and non-explanatory modelling assumptions having no effect on the WoE.

4.3 Laboratory validation

Here we compare the results of the peak height and discrete models on a set of 72 one to three contributor CSPs that were laboratory generated. We expect the two models to provide similar results for many cases, but the peak height model is expected to return a higher WoE in favour of a true hypothesis when the peak heights are informative, such as when Q contributes much less DNA to the CSP than one or more other contributors.

Single-, two- and three-contributor CSPs were generated in the laboratory (see Appendix C) from the DNA of 36 donors. Single-contributor CSPs were created at DNA contributions of 4, 16, 62 and 250pg,

with nine CSPs at each level. Two-contributor CSPs were created at both 16:250pg (12 CSPs) and 31:31pg (12 CSPs) DNA contribution ratios. Three-contributor CSPs were created at both 16:62:250pg (six CSPs) and 31:31:31pg (six CSPs) DNA contribution ratios. The WoE for each resulting CSP was evaluated using both discrete and continuous models of `likeLTD`. For multi-contributor CSPs, each contributor was queried in turn, leading to 36, 48 and 36 evaluations for the single-, two- and three-contributor CSPs respectively.

Here, the WoE will presented as an information gain ratio (IGR) which is WoE/$\log_{10}$IMP, where IMP is the inverse match probability, the theoretical maximum LR for a given Q. This allows for intuitive comparison of the WoE across different queried individuals.

4.3.1 Single contributor

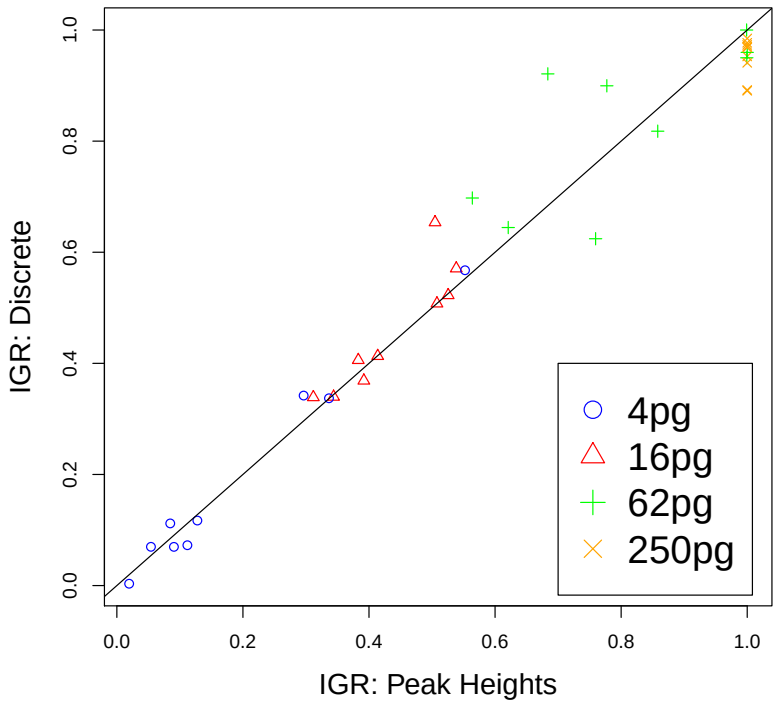


Figure 2: Information gain ratio (WoE/IMP) for 36 single-contributor CSPs using both the peak height (x-axis) and discrete (y-axis) models. Legend indicates the approximate DNA mass used to generate the CSPs.

IGR increases as the DNA mass increases, for both the peak height and discrete models (Figure 2). IGR is approximately equal between the two models for the majority of CSPs. At 16pg there is one exception to this equality, in which the discrete model returns a larger WoE than the peak height model, while greater variability is seen at 62pg. At 250pg the peak height model outperforms the discrete model for many CSPs because a minority of stutter peaks have been called as allelic, while many more have been called as uncertain. On reviewing the underlying CSPs, we found that in general the discrete model outperforms the peak height model when there is high variability in the observed CSP peak heights because the variance of the peak height model is constrained through a penalty on σ while the discrete model ignores peak height. For instance, some of the CSPs included loci where Q was heterozygous, but a single large peak was observed, while the other allele had dropped out, which in reality requires a high variance but may instead be well

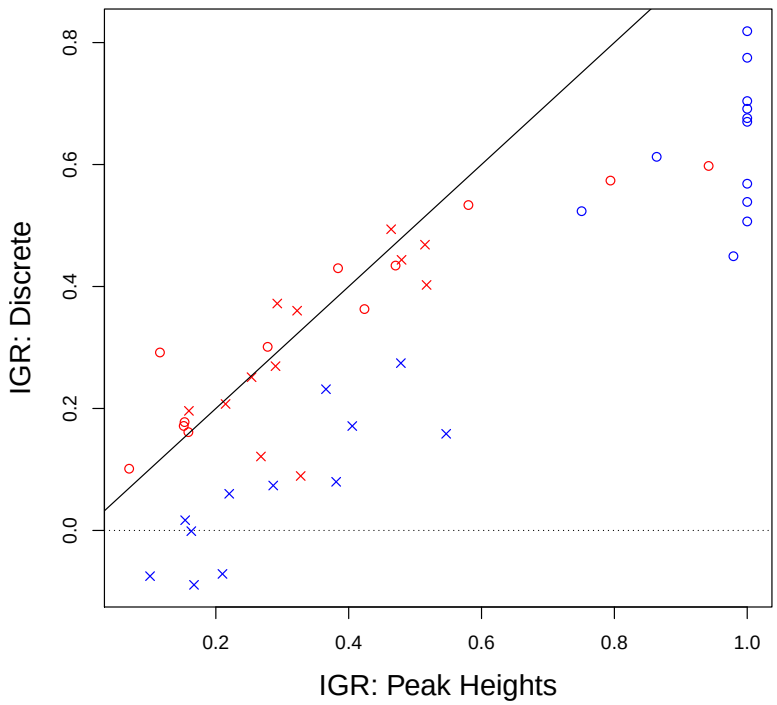


Figure 3: Information gain ratio (WoE/$\log_{10}$IMP) for 12 two-equal-contributor CSPs (red) and 12 two-contributor major/minor CSPs (blue) using both the peak height (x-axis) and discrete (y-axis) models. Both contributors to each CSP were queried, with circles and crosses indicating the first and second contributor respectively.

explained as a homozygote under H_d. Contrastingly, the peak height model outperforms the discrete model when an allele has been misassigned as allelic for the discrete CSP.

4.3.2 Two contributors

The IGR is approximately equal using the peak height and discrete models when the equal-contribution CSPs are queried (Figure 3, red). Two of the equal-contribution cases in Figure 3 localise with the major/minor cases. Visual inspection of the CSPs indicated that there was in fact a large discrepancy in contributions despite the intention to create equal contributions, perhaps due to pipetting error. One CSP performs noticeably better with the discrete model than with the peak height model; once again visual inspection revealed an unusually high variation in peak heights causing the peak height model to be conservative because very high variability is penalised in the model.

All of the major/minor CSPs return an IGR that is larger with the peak height model than with the discrete model (Figure 3, blue). Two of the major-queried evaluations have an IGR < 0.9; each of these CSPs have been confirmed by manual inspection to have peak heights closer to equal contributions than suggested by the specified DNA contributions of 16pg and 250pg. Note that when the minor is queried, four CSPs support H_d (IGR<0) using the discrete model, but support H_p using the peak height model; we know that H_p is true in all of these cases. Similarly, when querying the major contributor, the discrete model IGR ranges from 0.4 to 0.8, while the peak height model is able to obtain close to full information (IGR=1.0) for the majority of CSPs, reflecting the fact that the peak height model is able to exploit more information in

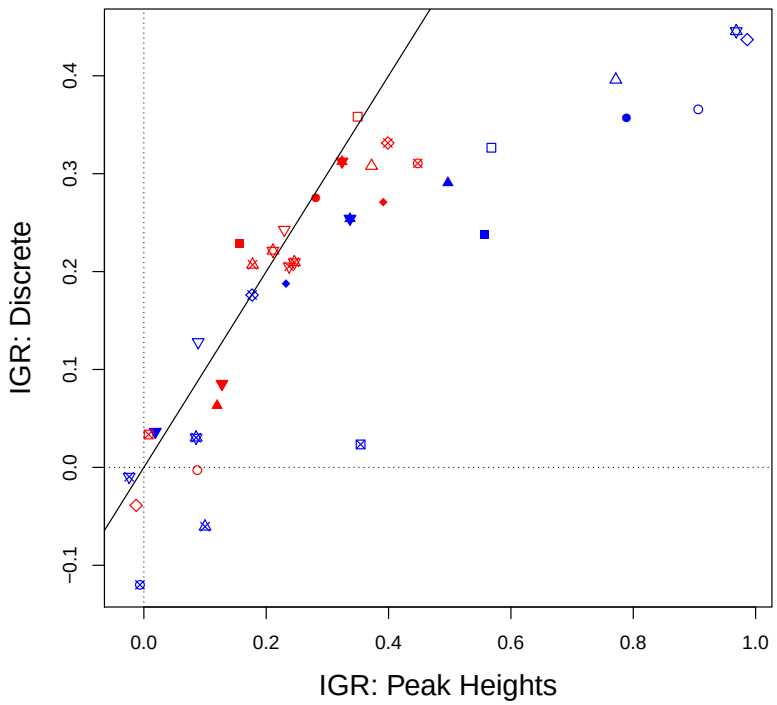


Figure 4: Information gain ratio (WoE/IMP) for 6 three-contributor equal-contribution CSPs (red) and 6 three-unequal-contributor CSPs (blue) using both the peak height (x-axis) and discrete (y-axis) models. The six cases of each condition are represented by square, circle, up-triangle, down-triangle, diamond and star symbols. Empty, filled and crossed symbols indicate that the first, second and third contributor were queried.

the CSP than the discrete model.

4.3.3 Three contributors

Of the six unequal-contribution CSPs evaluated (Figure 4, blue) one was a CSP for which whole-locus dropout was observed at 13 of the 16 used loci (downwards triangle), for which the peak height model is slightly more conservative than the discrete model for all evaluations perhaps due to insufficient information in the few observed peaks to estimate parameters of the model. Ignoring these three evaluations, all five 250pg-queried evaluations return a greater IGR with the peak height model than with the discrete model, all five do so for the 62pg-queried evaluations, and four of the five do so for the 16pg-queried evaluations with approximate equality in the 5th case. One 16pg-queried evaluation supports H_p using the peak height model, but supports H_d using the discrete model while in two 16pg-evaluations both the peak height and discrete models support H_d, despite H_p being true.

When equal-contributions CSPs are queried (Figure 4, red), the peak height and discrete models return approximately equal IGRs for all evaluations. Although peak heights can potentially distinguish single from multiple copies of an allele among the contributors (e.g. heterozygote from homozygote), in practice these results indicate that the variability in peak heights means that there is in fact little usable information in the equal-contributor scenario. There is one evaluation for which the peak height model supports H_p while the discrete model supports H_d, and one evaluation for which both the peak height and

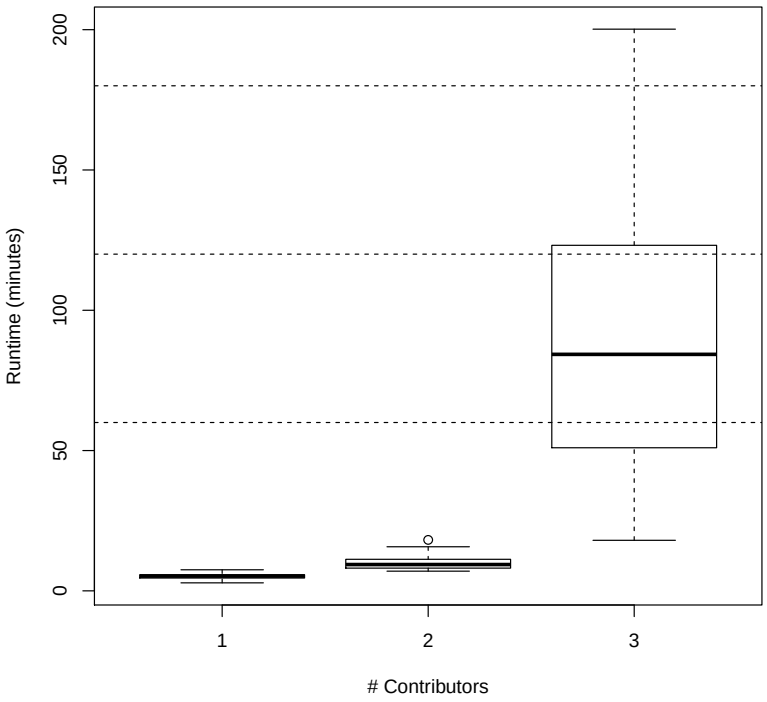


Figure 5: Runtime for the laboratory validation evaluations. Horizontal dashed lines indicate whole hours. The single- and two-contributor hypotheses included dropin, while the three-contributor hypotheses did not.

discrete models support H_d, despite H_p being true.

4.3.4 Runtime

The runtime for the peak height model over all laboratory validation evaluations ranges from 3 to 200 minutes, increasing with the number of contributors to the CSP (Figure 5).

4.3.5 Minor as dropin

In some scenarios it may be possible to explain peaks from a minor contributor as having originated from dropin rather than from an extra contributor, as suggested by the allele report (see Section 1.2). This not only reduces the computation complexity of the WoE calculation, but also eliminates the sometimes difficult decision of whether to treat low-level epg peaks as dropin or as an extra unknown contributor. The minor-as-dropin procedure is analogous to increasing the detection threshold to eliminate low-level peaks that are not of interest, as is sometimes performed currently for complex epgs. Additionally, this procedure reduces the difficulty of assigning the number of contributors to a mixture, as conceptually any number of minor contributors may be explained by dropin. This also allows CSPs with more than two unknown contributors to be evaluated by `likeLTD`, given that Q is not the minor contributor.

We re-evaluate the unequal-contributions two- and three-contributor CSPs with each contributor other than the minor as Q, demonstrating that low-level non-Q contributors can be explained as dropin with little-to-no effect on WoE against Q compared to treating them as unknown contributors (Figure 6).

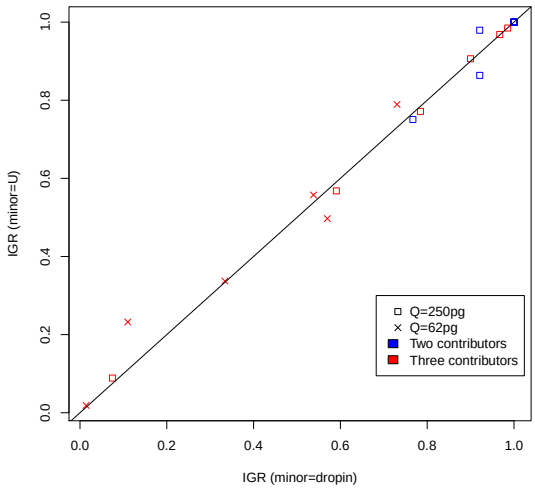


Figure 6: Information gain ratio (IGR) for 12 two- and/or 6 three-contributor CSPs (blue and red respectively) treating the minor contributor as dropin (x-axis) and as an additional contributor (y-axis).

4.3.6 Summary

The results presented here demonstrate that for a large number of laboratory CSPs the peak height model behaves as expected when compared to the discrete model; equal contribution mixtures return similar IGRs with both models, while the peak height model is able to utilise extra information in unequal contribution CSPs to return greater IGRs than the discrete model in favour of a true hypothesis. This is particularly useful for minor contributors, as highlighted by a number of cases where the discrete model supports H_d but the peak height model correctly supports H_p.

4.4 Validation using artificial changes to input data

Here we select one of the laboratory generated CSPs which we alter one peak at a time to verify that the resulting change in WoE is as expected. We expect that introducing dropped out alleles of Q will increase the WoE against Q, dropping out an allele of Q will decrease the WoE against him, introducing a dropin allele will decrease the WoE against Q, and that changes in peak heights that require greater variance of peak heights to explain the CSP under H_p should decrease the WoE and vice versa.

The single contributor CSP from donor 26 (16pg DNA) was used to investigate the behaviour of the peak height model when altering the CSP, as it had a mixture of locus dropouts (both heterozygote and homozygote), single dropouts (heterozygote) and non-dropouts (both heterozygote and homozygote). See Table 5 for a summary of the changes made to the CSP throughout this section.

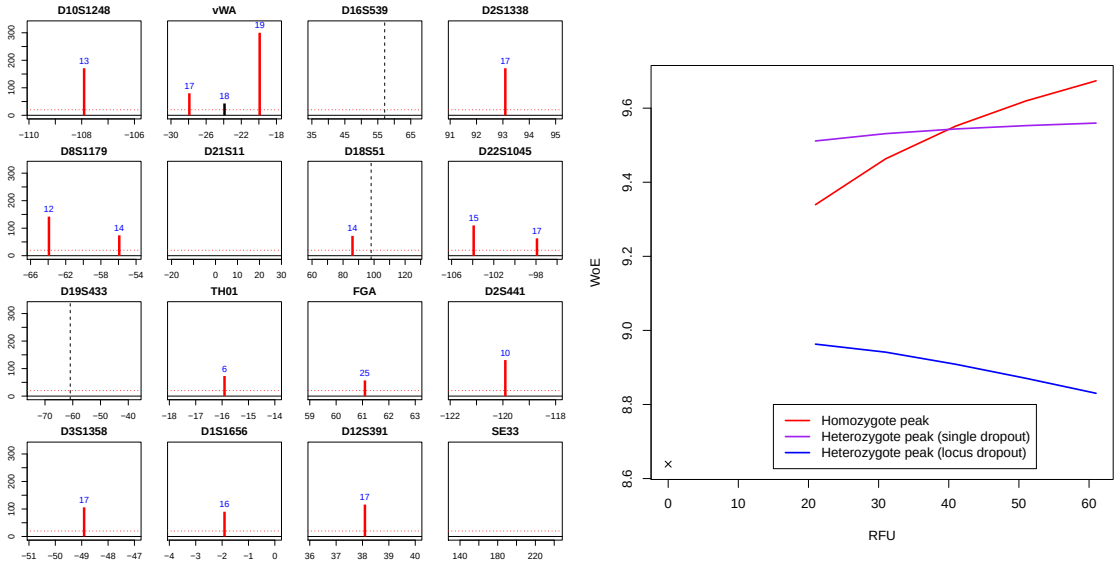
4.4.1 Missing peak insertion

A peak at the position of a single allele of Q which had dropped out was inserted into the CSP with varying peak height. This was done at three separate loci with:

1. No observed peaks, Q is homozygous (D16): homozygous locus dropout.
2. No observed peaks, Q is heterozygous (D19): heterozygous locus dropout.
3. One observed peak, Q is heterozygous (D18): heterozygous single dropout.

Locus	G_Q	CSP	Observation	Alteration
D16	13,13	$\emptyset$	Dropout of homozygous 13 allele	Reintroduction of 13 allele
				Introduction of 11 or 15 dropin peak
D18	14,17	14	Dropout of heterozygous 17 allele	Reintroduction of 17 allele
				Introduction of 8 or 12 dropin peak
D22	15,17	15,17	Fully observed heterozygote	Alteration of peak height at allele 17
				Introduction of 16 or 19 dropin peak
D19	13,14	$\emptyset$	Full heterozygous dropout	Reintroduction 13 allele
				Introduction of 15 or 18 dropin peak
TH01	6,6	6	Observed homozygote allele	Alteration of peak height at 6
				Introduction of 8.3 or 9.3 dropin peak
FGA	23,25	25	Dropout of heterozygous 23 allele	Alteration of peak height at 25
				Introduction of 21 or 22.1 dropin peak

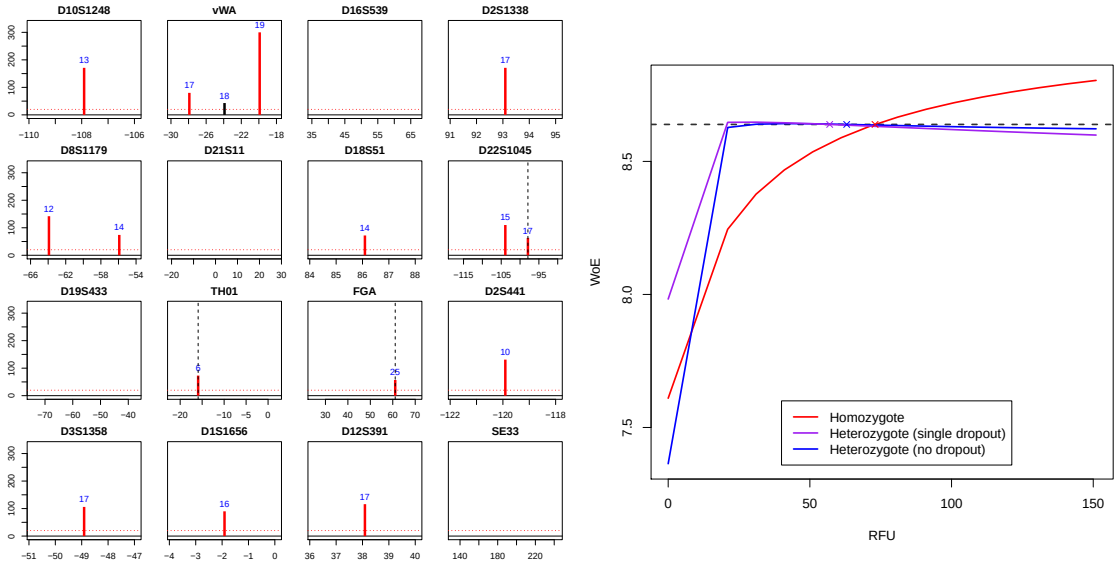
Table 5: Alterations applied to a single-contributor 16pg CSP at six loci. G_Q indicates the genotype of Q, the true contributor. $\emptyset$ under CSP indicates no observed peaks above the detection threshold at that locus. Observation gives the true effect seen at the locus. Alteration gives the two changes that were made at each locus. Reintroductions of dropped-out alleles ranged from 0 to 61 RFU, introductions of dropin peaks ranged from 0 to 61 RFU and alterations of observed peaks ranged from 0 to 151 RFU.



(a) CSP

(b) Results

Figure 7: (a) The single-contributor CSP for which peak heights were altered. Vertical dashed lines indicate the position of dropped-out alleles that were inserted. (b) WoE for a single CSP when a dropped out allele is artificially inserted at differing RFUs.



(a) CSP

(b) Results

Figure 8: (a) The single-contributor CSP for which peak heights were altered. Vertical dashed lines indicate the position of altered peaks. (b) WoE for a single CSP when the peak heights of an observed peak is artificially altered, from 0 RFU to 151 RFU. Crosses and the dashed horizontal line indicate the WoE and RFU when no peak is altered.

Inserting a homozygous dropout peak of Q increases the WoE, which is further increased as the RFU of the peak increases (Figure 7, red).

Inserting a heterozygous dropout peak of Q for which the corresponding allele was observed increases the WoE (Figure 7, purple) by more than when a homozygous allele was inserted, but the WoE increases less with increasing RFU of the inserted peak, so above 40 RFU the WoE is less with the inserted heterozygous peak than with the previously inserted homozygous peak. This is intuitive, as a small heterozygous peak is more likely than a small homozygous peak, leading to a greater WoE for the heterozygous peak at small RFUs. Similarly, a large heterozygous peak is less likely than a large homozygous peak, leading to a greater WoE for the homozygous peak at large RFUs.

Inserting a heterozygous dropout peak of Q for which the corresponding allele also dropped out increases the WoE initially (Figure 7, purple), but as the RFU of the peak is increased the WoE decreases. This is because the remaining dropout at this locus becomes less likely as the height of the artificial peak is increased; the variability in peak heights required to explain this observation increases with the increasing RFU of the introduced peak.

4.4.2 Altering observed peaks

A single observed peak in the CSP was given an altered RFU, from below the detection threshold (shown as 0 RFU here, analogous to dropout) to 150 RFU. This was performed for peaks at three separate loci with:

1. One observed peak, Q is homozygous (TH01): homozygous peak.
2. One observed peak, Q is heterozygous (FGA): heterozygous peak with dropout.
3. Two observed peaks, Q is heterozygous (D22): heterozygous peak.

When the peak height of a homozygous peak of Q is altered, the WoE has a strong positive relationship with the RFU of the peak (Figure 8, red), while removing the peak entirely decreases the WoE substantially.

When the peak height of a heterozygous peak of Q for which the corresponding allele dropped out is altered, the WoE has a weak negative relationship with the RFU of the peak (Figure 8, purple), with a large decrease in WoE when the peak is removed entirely.

When the peak height of a heterozygous peak of Q for which the corresponding allele was also observed is altered, the WoE decreases slightly as the RFU of the peak deviates from that observed in the unaltered CSP (Figure 8, blue). Once again, removing the peak entirely decreases the WoE substantially.

Dropout of a heterozygote peak of Q for which the corresponding allele was observed is less likely than dropout of a heterozygous allele for which the corresponding allele has also dropped out (Figure 8, RFU=0, blue and purple), which make intuitive sense. However, dropout of a homozygous peak of Q is more likely than dropout of a heterozygote allele for which the corresponding allele has also dropped out (Figure 8, RFU=0, red and blue); this is counter-intuitive but results from the penalty on `scale` that `likeLTD` imposes, meaning the variance introduced under H_p by pairing a dropout peak with a non-dropout peak, which can be explained as a homozygous allele under H_d, is penalised greater than the dropout of a homozygous peak.

4.4.3 Dropin peak insertion

A single peak was inserted into the CSP at the six previously altered loci, with the newly inserted peak being at a non-Q allele, and so the inserted peak simulates a dropin event. At each of the six loci both the highest frequency non-Q allele and lowest frequency allele in the DNA17 NDU1 database (Caucasian) were inserted separately. Inserted alleles, and their associated population probabilities (without sampling or F_{ST} adjustment) are given in Table 6.

Locus	Common		Rare	
	Allele	Probability	Allele	Probability
D16S539	11	0.317	15	0.001
D18S51	12	0.149	8	0.000
D22S1045	16	0.369	19	0.001
D19S433	15	0.179	18	0.000
TH01	9.3	0.334	8.3	0.001
FGA	21	0.179	22.1	0.000

Table 6: Dropin alleles that were inserted into the donor 26 16pg DNA CSP. Common alleles were chosen as the highest frequency allele in the DNA17 NDU1 database not-shared with Q. Rare alleles were chosen as the lowest frequency allele in the database.

As expected, at all loci introducing a dropin peak decreases the WoE from the non-dropin WoE of 8.6 bans to between 7.0 and 8.5 bans (Figure 9). For all conditions the WoE is further reduced as the peak height of the dropin peak increases from 21 RFU to 61 RFU. The reduction in WoE varies substantially between dropin peaks at different loci, ranging from 0.05 bans at D22 with a 21 RFU dropin of a common allele to 1.6 bans at D19 with a 21 RFU dropin of a rare allele.

At D22 (red) both of the alleles of Q are observed in the CSP, plus the third introduced dropin peak. The WoEs with introduction of a common (solid line) or rare (dashed line) allele diverge as the RFU of the introduced peak increases, because the dropin peak must be assigned as a dropin by `likeLTD` under H_p, which is plausible for a common allele, but implausible for a rare allele, and becomes increasingly implausible as the RFU of the dropin peak increases.

At TH01 (yellow), FGA (orange) and D18 (blue), a single allelic peak (homozygous, heterozygous and heterozygous respectively) was observed in the CSP, plus the introduced dropin peak. At these loci, H_d

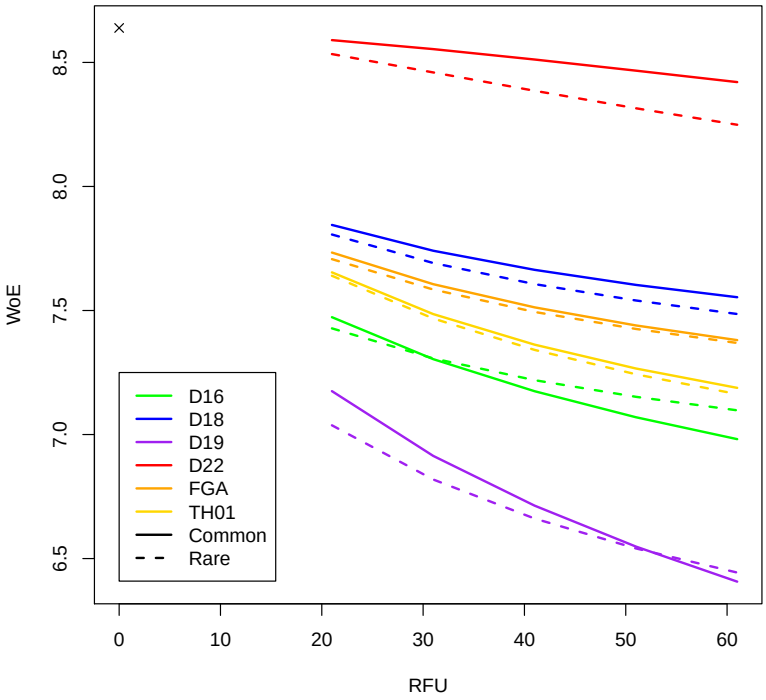


Figure 9: Weight-of-evidence for a single-contributor 16pg DNA CSP when a single rare or common dropin peak is inserted at one of six loci. See Table 6 for inserted alleles and their associated population probabilities.

explains the CSP as a heterozygous genotype composed of the observed true-allelic peak and the introduced dropin peak. Compared to the H_p explanation of a dropout and a dropin, this H_d explanation fits better when the dropin peak is rare than when it is common, leading to the WoE for the rare dropin being lower than that for the common dropin.

At D16 (green) and D19 (purple) no peaks were observed in the original CSP, so the CSPs here consist of just the introduced dropin peak. When the dropin peak is common in the population, under H_d `likeLTD` explains the observed peak as heterozygous at low RFU, but switches to explaining it as homozygous at high RFU. Conversely, when the dropin peak is very rare a homozygote is *a priori* unlikely, as under Hardy-Weinberg assumptions the probability of a homozygote is p_Z^2, which is 6.1e-7 and 1.5e-7 for the rare dropin allele at D16 and D19 respectively. For a common dropin the H_d explanation of a common homozygote allelic peak has an increasingly better likelihood compared to the H_p explanation of a common dropin as the RFU increases, leading to the reduction in WoE seen. However, for a rare dropin, the H_d explanation of a rare heterozygote peak does not increase its likelihood as much when the RFU increases, while the H_p explanation also performs less well as the RFU increases, so there is less discrepancy between the H_p and H_d explanations, leading to the lower drop in WoE seen in Figure 9.

4.4.4 Summary

We have demonstrated here that the peak height model behaves as expected when a CSP is altered in a number of ways; introducing a dropped out allele increases the WoE against Q, dropping out an allele decreases the WoE against Q and introducing a dropin peak decreased the WoE against Q. Increasing the RFU of homozygous Q alleles consistently increases the WoE against Q, as homozygous alleles are expected to be large. Increasing the RFU of heterozygous Q alleles has less of an effect on the WoE, but when the corresponding allele has dropped out the WoE decreases as a large observed peak with a dropout peak requires a large peak height variability to explain the CSP under H_p; these are often explained as homozygote peaks under H_d as would be expected for a large single peak. These are sensible and expected behaviours of the peak height model in response to altering the RFU of an allele contributed by Q. The WoE against Q decreases as the RFU of a dropin peak is increased, as would be expected, and the severity of the decrease in WoE when the dropin peak is introduced depends on the other observed peaks at the locus; if a dropin peak can only sensibly be explained as dropin it has little effect on the weight of evidence against Q, whereas if the dropin peak can be sensibly explained as an allele of X then the WoE against Q is significantly reduced. All of these behaviours are as expected, and make sense intuitively.

5 Real case comparison: Meredith Kercher

A CSP from a real-world crime was evaluated with three continuous models. This comparison benchmarks the models against each other in a real-world scenario. Assuming that all of the models are valid, the results obtained with each should be similar.

5.1 Case circumstances

In November 2007, Meredith Kercher was murdered in her flat in Perugia, Italy. While Rudy Guede was tried and convicted for the crime in under a year with little controversy, the accusation that Raffaele Sollecito and Amanda Knox were involved in the murder was much more controversial. The two were found guilty in December 2009, acquitted in October 2011, found guilty again in January 2014, and finally ruled innocent by Italy's Supreme Court of Cassation in March 2015. One of the key, and controversial, pieces of evidence in the case against Knox and Sollecito was Meredith Kercher's bra clasp, item 165B, found on the floor of the room Meredith was murdered in, over a month after the murder occurred. Here, the WoE of the epg arising from the bra clasp for both Knox and Sollecito to be a contributor will be evaluated using the `likeLTD`, `STRmix` and `EuroForMix` peak height models. The hypotheses compared are of the form:

$$H_p^S\text{: Q (Raffaele Sollecito) + K1 (Meredith Kercher) + U1,}$$

Q	Sollecito			Knox		
Program	likeLTD	STRmix	EuroForMix	likeLTD	STRmix	EuroForMix
D8	0.7	0.2	0.8	-0.2	1.0	-0.2
D21	0.5	0.8	0.8	-0.1	0.1	0.1
D7	0.4	0.5	0.5	-0.2	-0.4	-0.1
CSF	0.7	0.3	0.6	-0.1	0.0	0.0
D3	0.8	0.8	1.0	-0.4	0.4	-0.2
TH01	1.1	0.8	1.2	-0.3	-0.6	-0.3
D13	0.7	0.7	0.8	0.0	-0.3	-0.1
D16	0.8	0.9	0.7	0.0	0.0	0.1
D2	2.4	1.6	2.1	0.1	0.5	0.0
D19	1.3	1.4	1.6	-0.9	-1.4	-1.4
vWA	1.5	1.9	1.8	-0.2	-0.7	-0.4
TPOX	0.9	0.7	0.7	-0.1	0.1	-0.1
D18	1.4	1.2	1.4	0.2	0.3	0.3
D5	-0.4	0.0	-0.5	0.0	0.3	0.1
FGA	-1.2	0.0	-0.5	0.0	0.1	0.0
Overall	11.5	11.8	13.0	-2.3	-0.7	-2.1

Table 7: Locus and overall weight of evidence (WoE) for the epg generated from item 165B (bra clasp) in the Kercher case. WoE was evaluated against Raffaele Sollecito or Amanda Knox with three continuous models; `likeLTD`, `STRmix` and `EuroForMix`. In all evaluations Meredith Kercher was assumed to be a contributor, with another unknown individual and Q/X. The IMP for Sollecito is 18.5 bans. A detection threshold of 50 RFU was used in all evaluations.

$$H_d^S\text{: X + K1 (Meredith Kercher) + U1,}$$

and:

$$H_p^K\text{: Q (Amanda Knox) + K1 (Meredith Kercher) + U1,}$$
$$H_d^K\text{: X + K1 (Meredith Kercher) + U1.}$$

5.2 Results

5.2.1 Raffaele Sollecito

When Raffaele Sollecito is queried all three programs return a WoE≥11.5 bans (Table 7) with `likeLTD` and `STRmix` having similar WoEs ($\Delta = 0.3$ bans) but `EuroForMix` having a WoE > 1 ban larger (Δ=1.5 and 1.2 bans for `likeLTD` and `EuroForMix` respectively). The three programs have largely good correlation between locus WoEs, with two exceptions:

D5: `likeLTD` and `EuroForMix` have similar WoEs supporting H_d, `STRmix` supports neither hypothesis. Sollecito is homozygous and masked by a heterozygous peak of Kercher.

FGA: `likeLTD` and `EuroForMix` support H_d, `STRmix` supports neither hypothesis, the `likeLTD` and `EuroForMix` WoEs are now considerably different. Sollecito is heterozygous and both alleles are masked by alleles of Kercher.

The runtime for `likeLTD` was between 16 and 17 minutes, while `EuroForMix` and `STRmix` took less than a minute to run.

5.2.2 Amanda Knox

When Amanda Knox is queried, all three programs support H_d (Table 7), with `likeLTD` and `EuroForMix` having similar WoEs (≈-2 bans), while `STRmix` has a noticeably larger WoE (-0.7 bans). There are some notable locus differences between the programs:

D8: `EuroForMix` and `likeLTD` support H_d, `STRmix` has the strongest support for H_p of any program and any locus when querying Knox. Knox has one observed allele in a stutter position of a Kercher allele, and a dropout allele in the double-stutter position of the same Kercher allele.

D3: `EuroForMix` and `likeLTD` support H_d, `STRmix` supports H_p. Knox has one allele masked by Kercher, and another allele that has dropped out.

D13: `EuroForMix` and `STRmix` support H_d, `likeLTD` supports neither hypothesis. One allele of Knox is masked by Kercher, while the other has dropped out.

The runtime for `likeLTD` was between 25 and 30 minutes, while `EuroForMix` and `STRmix` once again required less than a minute for computation.

5.3 Conclusions

The three models return similar results for all evaluations, all providing extremely strong support for Sollecito contributing to the sample, and all supporting H_d for the Knox evaluation, between limited support with `STRmix` to moderately strong support with `likeLTD` and `EuroForMix`. The largest difference between the models is 1.6 bans for the Knox evaluation, with differences likely due to divergent modelling choices between the programs.

5.4 Validation summary

The `likeLTD` peak height model has been demonstrated to behave as expected for numerous laboratory CSPs, when artificially altering the observed peaks of a single CSP, and in relation to results from other peak height models (`STRmix` and `EuroForMix`). The peak height model provides support for H_p in 16/18 minor contributor evaluations of laboratory-generated CSPs (16pg), and also does so in 41/42 equal contribution low-template evaluations (31pg), demonstrating high sensitivity.

6 Acknowledgements and version history

The underlying mathematical model and its implementation in the `likeLTD` `R` code were developed by DJB. Input into the model came from John Buckleton, as described in Balding and Buckleton (2009). A number of other academics and forensic scientists have given feedback and encouragement, among them Norah Rudin and Kirk Lohmueller in California, Torben Tvedebrink and Niels Morling in Denmark, Peter Gill (Norway), Hinda Haned (Netherlands), and Roberto Puch-Solis (UK).

Since Version 4.0, DJB has been helped to develop the `R` code by Adrian Timpson, and more recently Christopher Steele has developed and coded the continuous model and implemented the tests described in this document.

The early work in developing Version 5.0 was done by Adrian Timpson, the bulk of the recoding was done by Mayeul d'Avezac of the Research Software Development team in UCL Information Services Division, and some final enhancements were implemented by Christopher Steele.

There has been no external funding for this project, although DJB has benefited from fees paid to UCL Consultants Ltd for expert witness work. His employer University College London, and in particular the UCL Genetics Institute, have supported the project by continuing to pay him a salary during the many months of work time that he has devoted to it.

- **Version 1**
  - Release 1-0, 19/1/10. The initial code had separate files LR1unk.R and LR2unk.R for 1 and 2 unprofiled contributors. Each included functions LRnumer() and LRdenom()
  - Release 1-1, 23/1/10. Restructured code for LR1unk.R to make it more similar to LR2unk.R
  - Release 1-2, 26/3/10. Fixed small bug reported by Kirk Lohmueller, affecting the assignment of allfracs in 3 places
  - Release 1-3, 24/5/10. Changed way dropin is modelled.
- **Version 2**
  - Release 2-0, 21/6/10. Merged previous LR1unk.R and LR2unk.R into a single file LTDNALR.R with the functions LRnumer() from those files renamed as LRnumer1() and LRnumer2(), respectively, and similarly for LRdenom().
  - Release 2-1. The change introduced in V2.1 has since been undone in V3.0, by introduction of a better way to deal with rare alleles
- **Version 3**
  - Release 3-0, 12/10/11. The previous functions LRnumer1(), LRnumer2(), LRdenom1() and LRdenom2() were all replaced by a single function `likeLTD`. There is now a distinct dropout rate for each replicate (DO). The dropout rate for other individuals is determined as a function of DO and the amount of DNA from that individual relative to the amount contributed by the reference individual (Q or U). We now strip out alleles with zero database frequency. If an allele of Q or CSP is not found in rownames(acbp) this allele is inserted into acbp with count 1. This has speeded up computations so that it now becomes feasible to allow three unprofiled contributors to the crime scene profile when Qcont=F, otherwise two unprofileds + Q. The model for dropout is now improved: the previous kdrop function has gone, and both dropout and dropin calculations are included in a new function Calclik(). Stutter alleles, or other apparent artefacts, can be entered as uncertain alleles allowing the possibility that they could be allelic.
  - Release 3-1, 4/1/12. Previously the dropin parameter DI was the non-dropout rate for a hypothetical extra individual, but this is now modified so that the dropin rate for each replicate is DI times the non-dropout rate (1-DO) for that replicate. As before, if DI=0 then all CSP alleles must come from one of the specified contributors. We now allow any of the profiled possible contributors to be unaffected by dropout, including Q. This option should only be used if the individual's alleles are observed in the CSP in every replicate at every locus; otherwise an error is generated. Alleles of profiled possible contributors not subject to dropout are converted to uncertain and removed from the CSP in the preprocessing step and (except for Q) don't play any further role in `likeLTD`. There has been some rearrangement of the code so that more work is done in a preprocessing function that is called only once, rather than being repeated in every call to the main function. Some changes have been made to the way parameters are named and passed; function calls to previous versions of `likeLTD` will not work without modification.
- **Version 4**
  - Release 4-0, 19/3/12. The main innovation is to allow dropout rates to increase with fragment length. Thus, fragment lengths for each allele in the profiling system being employed must be supplied (in base-pairs, bp, centred so that 0 represents an average length). These are passed to `likeLTD` in column 2 of matrix afbp, which replaces vector allfracs in Version 3.1; column 1 is the previous allfracs, and specifies population allele fractions. The program uses the model of Tvedebrink et al. (2012) and essentially the "dose" of DNA contributed by an individual at an allele is adjusted by a geometric function of fragment length (increased for below-average fragment lengths, and decreased for above-average). The rate of the geometric distribution is a parameter deg (for degradation), which is a vector with one entry per contributor subject to dropout.

- Release 4-1, 8/5/12. Improvement to computation of number of simulations used when denNu=3 and also starting values for nupa and depa. Release of test document giving results from performance tests of `likeLTD`.
- Release 4-2, 26/6/12. These are mainly minor changes to improve the output and program clarity documentation. The test results document distributed with this code is also updated to include new test results. The most important change is an improved assignment of the simulation size for the likelihood approximation invoked for three unprofiled contributors (i.e. denNu=3). For one or two unknown contributors there should be no changes to results from Version 4.1. BB is now passed as a parameter rather than being assigned as a constant.
- Release 4-3, 10/8/12. Mostly just a few minor changes to documentation but there is one important bug fix that affected the likelihood calculations when DI > 0; thus any V4.2 runs that modelled dropin (Drin = TRUE in the wrapper) should be rerun with V4.3. Further improvements to output and to value for nsim.
- Release 4-4, 2/11/12. Two changes:
  * A new block of code can provide much faster computation when Nunp=2 or 3 and DI=0. The speed-up is greatest when the CSPs determine many alleles in the genotypes of the unprofiled contributors. The new code uses combinatorial functions that require the `R gtools` library; `library(gtools)` is now included in the Wrapper, but the package must first be installed using `install.packages("gtools")`. The result of the computation is unchanged from the original code that uses "for" loops. Both codes are kept, and the initial likelihood calculation is done once using each code in order to set flags indicating which is quickest; the faster code is then used for all subsequent calculations at that locus (there are separate flags for the calculations under H_p and H_d). Because of this improvement, the previous code that performed a simulation-based approximation to the likelihood when Nunp=3 has been removed, and so `nsim` has been removed from the list of parameters passed to `likeLTD`.
  * Locus adjustment terms are now included in the dropout model, as in Tvedebrink et al. (2009). However, rather than estimate the locus effects on dropout from external data, they are estimated from the input data for the profile being analysed. Because this may be relatively little information, a strong prior is imposed on the locus adjustments: gamma with both parameters equal so that the mean is 1. The default value of this parameter (lap) is 50, giving a prior standard deviation for the locus adjustments of 0.14, the same as the SD of the estimates of Tvedebrink et al. (2009).

  Also the inverse of the exact match probability is output for comparison with the LR for the observed CSP: this is the the standard match probability that would apply if the CSP showed exactly the reference profile of Q, and it is assumed that there is only one contributor. The LR for any other CSP should not exceed the inverse of the match probability.
- Release 4-5, 2/11/12. The power parameter β has been fixed in previous versions at -4.35 (Tvedebrink et al., 2009). In this version it is updated in the simulated annealing, separately under H_p and H_d, subject to a Gaussian prior/penalty with mean -4.35 and SD 0.38, the values obtained by Tvedebrink et al. (2009). This is a relatively minor and sensible change, and we have checked that it has little impact. However all the test results reported in this document are for V4-4 and not V4-5.

- **Version 5**
  - Release 5-0. This is a complete re-writing of the basic code, which is now established as an official `R` package on CRAN. The simulated annealing algorithm used in previous versions for parameter optimisation is replaced with a differential evolution algorithm for optimisation. The underlying likelihood model remains the same as version 4.5, however, significant speed improvements have been gained through re-factoring of R code (e.g. converting for loops into vector/matrix operations), re-writing computationally intensive steps in C, and implementing parallel computation of

the C code. Steps that have been implemented in C code include the computation of genotype combinations for unknown contributors, computing allele doses for each genotype combination, dose adjustments for relatedness, heterozygocity, dropout and power. Uploading the package to CRAN comes with improved documentation, version control and ease of access.

- Release 5-1. This update improved the calculation of the LR when close-relatedness is taken into account.
- Release 5-2. This update adds the function `get.likely.genotypes` that returns the most probable genotypes for each locus, and the most probable whole-profile genotype. There is an option to return marginal genotype probabilities for each contributor subject to dropout, or joint probabilities for all contributors subject to dropout.
- Release 5-3. This update improves the generation of both allele and output reports. These are now output as .doc files instead of .pdf files, and will now scale with the number of loci and the number of replicates correctly. The change to .doc files was motivated by client requests, and .pdf files can still be easily obtained by opening the .doc file in MS Word and saving as a pdf. There are additional improvements to the checks for unusual alleles (which will now recognize typos and alleles not present in the database), and to the suggestion of appropriate hypotheses to test.
- Release 5-4. This update improves the optimisation procedure, replacing the simple convergence threshold with a geometric progression of convergence. This includes a geometric progression of the `DEoptim::DEoptim.control CR` variable, which controls the crossover rate of the optimisation algorithm. The combination of these two means that the parameter space is more thoroughly searched in the initial stages, leading to improved optimisation. L_p and L_d are now optimised together (within each step), allowing for estimation of the progress of optimisation (and an associated progress bar). Interim results after each step are now available. These changes are incorporated in the new optimisation function, `evaluate`. Small changes to the outputs are included, namely altered default file names (including the case name in the file name) and including which database file is used in the information section.
- Release 5-5. This update allows database alleles that are unobserved in both the CSP and reference profiles to be combined into a single "rare" allele, greatly improving the speed of computation. Three databases are now provided with `likeLTD`, for NGMSelect, SGM+ and Identifiler. The new default database is that for NGMSelect. A correction for linkage has been added, that will be utilised when Q and X are assumed to be siblings. The function `evaluate.from.interim` allows for a partial computation to be restarted from a generated interim result. The full posterior probability for genotypes can be returned, allowing for sensitivity testing of the LR to choices of alleles when a reference profile is only partially known.

- **Version 6**
  - Release 6-0. This major update introduces a new peak height model into `likeLTD`, which can utilise the full peak heights information available in a CSP, incorporating stutter, over-stutter, double-stutter, dropin, degradation, multiple replicates and multiple contributors. The peak height model can be run in a similar fashion to the discrete model, but with `.peaks` appended to each function e.g. `evaluate` becomes `evaluate.peaks`. The adjustment to the LR for linked loci has been extended to include uncle (or aunt)/nephew (or niece), half-uncle (or half-aunt)/half-nephew (or half niece), cousins, grandparent/grandchild and half siblings relationships. With this comes a new way of specifying relatedness, through an index of what relationship you wish to assume Q and X have, rather than the previous relatedness coefficients. This is applied to the discrete model as well as the peak height model. A seed to be set before running maximisation can now be handed to `evaluate` and `evaluate.peaks`, if unspecified an integer representation of the current date, time and process ID will be used. The seed used is now printed in the output report for both models.

– Release 6-1. This release includes a substantial speed up of the program. Runtime for the Laboratory case has been reduced from approximately 363 and 2749 minutes for the 1U and 2U hypothesis pairs to 23 and 200 minutes. This was achieved by altering optimisation parameters: `nConverged` 5->4, `iterMax` 75->25, `searchPopFactor` 4->1. The function to determine the number of steps to run after the first was also altered to $\lceil (c+1)(r+1)\log_8 max(\sigma_p^2, \sigma_d^2) \rceil$ where c and r are the number of hypothesised contributors and the number of replicates in the CSP. An estimate of the contributions of unknown contributors was added to the allele report; this uses k-means clustering on the peak heights of unattributable alleles to assign each allele to an unknown contributor. If the estimated contribution of an unknown contributor is $< 1/3$ the estimated DNA contribution of Q then the allele report suggests that it may be possible to explain that unknown contributor as dropin rather than an extra unknown.

References

Balding, D. J. 2013. Evaluation of mixed-source, low-template dna profiles in forensic science. *Proc. Natn. Acad. Sci. USA*.

Balding, D. J. and Buckleton, J. 2009. Interpreting low template DNA profiles. *Forensic Sci. Int.-Gen.*, 4:1–10.

Balding, D. J. and Steele, C. D. 2015. *Weight-of-evidence for Forensic DNA Profiles*. John Wiley & Sons.

Bright, J.-A., Curran, J. M., and Buckleton, J. S. 2013a. Relatedness calculations for linked loci incorporating subpopulation effects. *Forensic Sci. Int.-Gen.*, 7(3):380–383.

Bright, J.-A., Taylor, D., Curran, J. M., and Buckleton, J. S. 2013b. Developing allelic and stutter peak height models for a continuous method of dna interpretation. *Forensic Sci. Int.-Gen.*, 7(2):296–304.

Brookes, C., Bright, J.-A., Harbison, S., and Buckleton, J. 2012. Characterising stutter in forensic str multiplexes. *Forensic Sci. Int.-Gen.*, 6(1):58–63.

Buckleton, J., Triggs, C., and Walsh, S. 2004. *DNA Evidence*. CRC Press.

Butler, J. 2010. *Fundamentals of Forensic DNA Typing*. Elsevier Academic Press.

Cowell, R. G., Graversen, T., Lauritzen, S., and Mortera, J. 2013. Analysis of DNA Mixtures with Artefacts. *ArXiv e-prints*.

Gill, P., Whitaker, J., Flaxman, C., Brown, N., and Buckleton, J. 2000. An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA. *Forensic Sci. Int.*, 112:17–40.

Kelly, H., Bright, J.-A., Buckleton, J. S., and Curran, J. M. 2014. Identifying and modelling the drivers of stutter in forensic dna profiles. *Aust. J. Forensic Sci.*, 46(2):194–203.

Mullen, K. M., Ardia, D., Gil, D. L., Windover, D., Cline, J., et al. 2011. Deoptim: An r package for global optimization by differential evolution. *J. Stat. Softw.*, 40(6):1–26.

Steele, C. and Balding, D. 2016. Evaluation of low-template dna profiles using peak heights. *Stat. Appl. Genet. Mol. Biol.* In press.

Tvedebrink, T., Eriksen, P., Mogensen, H., and Morling, N. 2009. Estimating the probability of allelic drop-out of STR alleles in forensic genetics. *Forensic Sci. Int.-Gen.*, 3:222–226.

Tvedebrink, T., Eriksen, P., Mogensen, H., and Morling, N. 2012. Statistical model for degraded DNA samples and adjusted probabilities for allelic drop-out. *Forensic Sci. Int.-Gen.*, 6:97–101.

A Allele report for Laboratory case

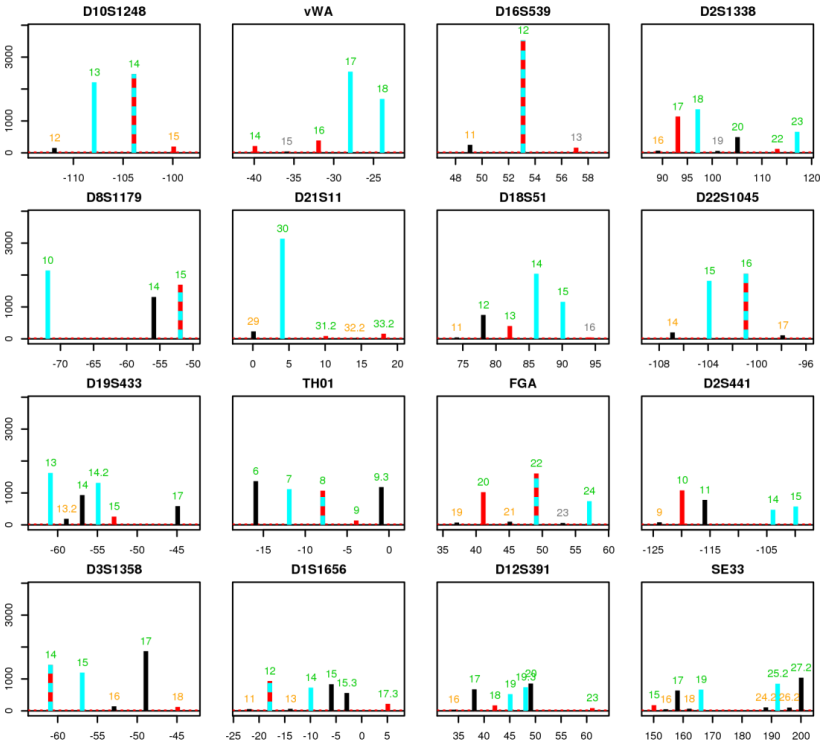
Laboratory-Allele-Report

Laboratory

Label	Reference profile
Q	Suspect
K1	Victim

Crime scene profiles (CSP)

Replicate: 3cont_3.u.1.t



The peak heights in RFU (y-axis) and mean adjusted allele length in base pairs (x-axis), with peaks at alleles in the profile of Q coloured in red, peaks at alleles of other assumed contributors shown with other colours, while black peaks are not attributable to Q or any other assumed contributor. Allele labels are coloured according to their possible allelic status (this is intended as a guide and is not assumed by the software): green=allelic, orange=uncertain, grey=non-allelic.

Reference profiles

All peaks in the provided profiles

Profile	D10S1248	vWA	D16S539	D2S1338	D8S179	D21S11	D18S51	D22S1045	D19S433	TH01	FGA	D2S441	D3S1358	D1S1656	D12S391	SE33
Suspect	*14,15*	*14,16*	*12,13*	*17,22*	*15,15*	*31.2,33.2*	*13,16*	*16,16*	*12,15*	*8,9*	*20,22*	*10,10*	*14,18*	*12,17.3*	*18,23*	*15,24*
Victim	*13,14*	*17,18*	*12,12*	*18,23*	*10,15*	*30,30*	*14,15*	*15,16*	*13,14.2*	*7,8*	*22,24*	*14,15*	*14,15*	*12,14*	*19,19.3*	*19,25.2*
Other	*12*	*15*	*11*	*16,19,20*	*14*	*29,32.2*	*11,12*	*14,17*	*13.2,14,17*	*6,9.3*	*19,21,23*	*11,9*	*16,17*	*11,13,15,15.3*	*16,17,20*	*16,17,18,24.2,26.2,27.2*

After removal of peaks that are possibly non-allelic (intended as a guide only - not assumed by software)

Profile	D10S1248	vWA	D16S539	D2S1338	D8S179	D21S11	D18S51	D22S1045	D19S433	TH01	FGA	D2S441	D3S1358	D1S1656	D12S391	SE33
Suspect	*14,15*	*14,16*	*12,13*	*17,22*	*15,15*	*31.2,33.2*	*13,16*	*16,16*	*12,15*	*8,9*	*20,22*	*10,10*	*14,18*	*12,17.3*	*18,23*	*15,24*
Victim	*13,14*	*17,18*	*12,12*	*18,23*	*10,15*	*30,30*	*14,15*	*15,16*	*13,14.2*	*7,8*	*22,24*	*14,15*	*14,15*	*12,14*	*19,19.3*	*19,25.2*
Other				*20*	*14*		*12*		*14,17*	*6,9.3*		*11*	*17*	*15,15.3*	*17,20*	*17,27.2*

Unobserved, *unreplicated* and **replicated** peaks in provided reference profiles and those unattributable to any reference profile.

Summary

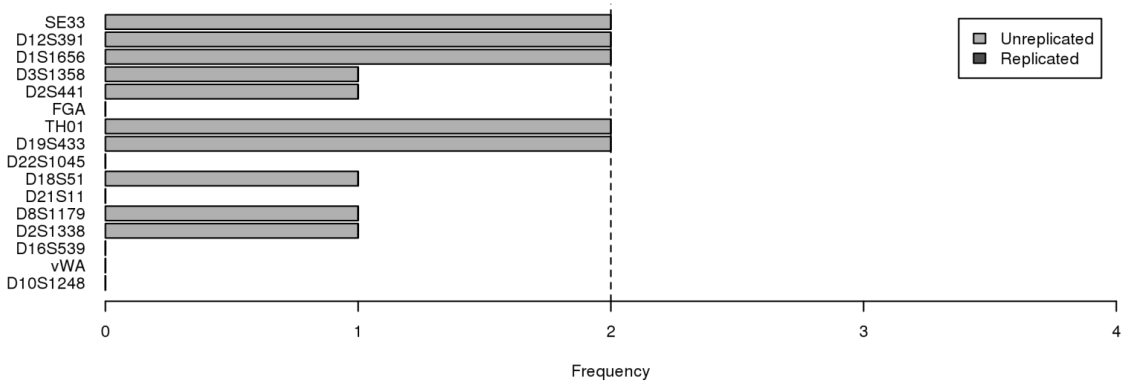
Reference profile	Suspect	Victim
Replicate: 3cont_3.u.1.t	0.9375	1
Overall	0.9375	1

Approximate representation (observed/total) for each reference profile per replicate and overall.

Reference profile	Suspect	Victim
Replicate: 3cont_3.u.1.t	807.0625	1587.375
Overall	807.0625	1587.375

Mean RFU for each reference profile per replicate and overall. This may be an over-estimate of the average DNA contribution of an assumed contributor due to sharing of alleles with other contributors. These values are for information purposes only, they are not used by the software.

Unattributable alleles



Number of unreplicated (light grey) and replicated (dark grey) unattributable alleles per locus, for the likely-allelic peaks (green allele labels shown in the CSP plots).

Alleles that are rare in at least one database

Locus	File	Profile	Allele	Issue	NDU1	NDU2	NDU3	NDU4	NDU6	NDU7
D2S1338	CSP	3cont_3.u.1.t	16	Rare allele	113	48	1	3	23	25
D18S51	CSP	3cont_3.u.1.t	11	Rare allele	35	4	3	1	1	3
D19S433	CSP	3cont_3.u.1.t	17	Rare allele	14	1	4	0	0	1
D2S441	CSP	3cont_3.u.1.t	9	Rare allele	6	1	5	0	0	1
D1S1656	CSP	3cont_3.u.1.t	15.3	Rare allele	169	13	8	0	5	8

Locus	File	Profile	Allele	Issue	NDU1	NDU2	NDU3	NDU4	NDU6	NDU7
D12S391	CSP	3cont_3.u.1.t	16	Rare allele	72	58	3	1	24	34
D12S391	CSP	3cont_3.u.1.t	19.3	Rare allele	23	0	1	0	0	0
D12S391	Reference	Victim	19.3	Rare allele	23	0	1	0	0	0
SE33	Reference	Suspect	24	Rare allele	0	4	0	1	2	2

Suggested parameter values

nU	doDropin	Recommendation
1	No	Recommended

If an nU value >2 is indicated, an approximate result can be obtained using nU=2 and doDropin=Yes. Please check the allele designations shown in the CSP plots that were used to generate these hypotheses; if you disagree with the suggested designations the recommendations here may need to be altered.

Minor as dropin

Mean RFU Q	Mean RFU U1	# as dropin
656	926	0

Mean peak height for Q, clustered mean peak heights for unknowns using k-means clustering with 1 clusters, and the number of unknowns that may be explainable as dropin (mean Q peak height/mean U peak height > 3).

System information

Type	Details
Date report generated:	Fri Apr 29 10:55:28 2016
Package	likeLTD
Title	Tools to Evaluate DNA Profile Evidence
Description	Tools to determine DNA profile Weight of Evidence. For further information see the likeLTD guide provided, or Balding, D.J. (2013) <DOI:10.1073/pnas.1219739110>.
Depends	R (≥ 2.10), DEoptim, ggplot2, gtools, rtf
Suggests	svUnit, scales
Imports	gdata, tools, tcltk
Version	6.1.0
Date	2016-02-21
Author	David Balding, Adrian Timpson, Christopher Steele, Mayeul d'Avezac, James Hetherington.
Maintainer	Christopher Steele <c.steele.11@ucl.ac.uk>
License	GPL-3
URL	https://sites.google.com/site/baldingstatisticalgenetics/
Packaged	2016-04-29 09:11:22 UTC; csteele
Built	R 3.1.3; x86_64-unknown-linux-gnu; 2016-04-29 09:11:24 UTC; unix
sysname	Linux
release	2.6.32-573.7.1.el6.x86_64
version	#1 SMP Tue Sep 22 22:00:00 UTC 2015
nodename	ugi-151057.ugi.ucl.ac.uk
machine	x86_64
login	csteele
user	csteele
effective_user	csteele

B Output file for Laboratory case

Laboratory-Evaluation-Report
Laboratory

Prosecution hypothesis: Suspect (Q) + Victim (K1) + U1

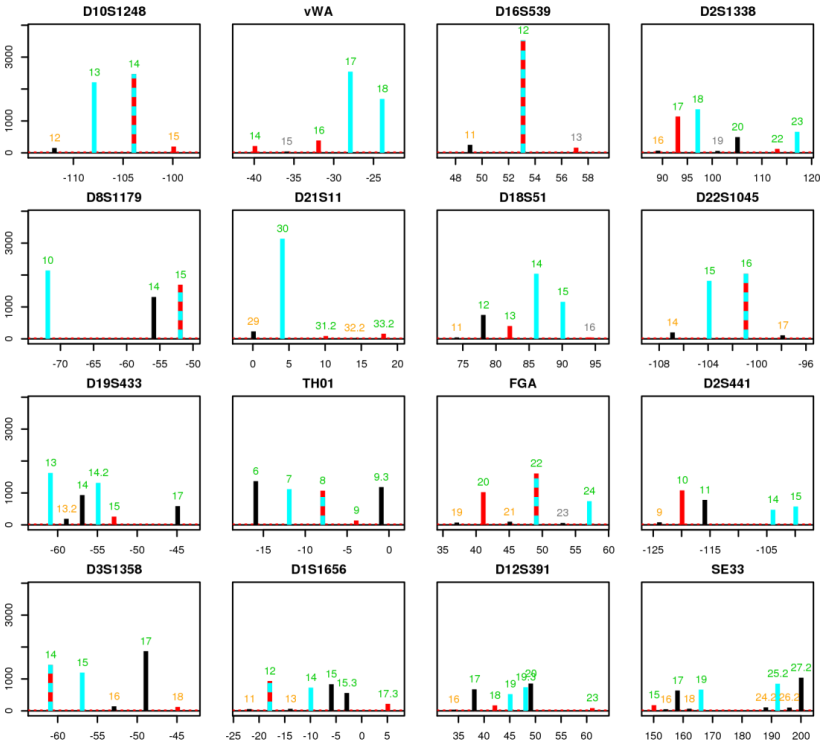
Defence hypothesis: Unknown (X) + Victim (K1) + U1

Overall Likelihood

calculation	estimate
Prosecution.log10	-302.2
Defence.log10	-310.4
Ratio.log10	8.2
Ratio	1.62e+08

Crime scene profiles (CSP)

Replicate: 3cont_3.u.1.t



The peak heights in RFU (y-axis) and mean adjusted allele length in base pairs (x-axis), with peaks at alleles in the profile of Q coloured in red, peaks at alleles of other assumed contributors shown with other colours, while black peaks are not attributable to Q or any other assumed contributor. Allele labels are coloured according to their possible allelic status (this is intended as a guide and is not assumed by the software): green=allelic, orange=uncertain, grey=non-allelic.

Reference profiles

All peaks in the provided profiles

Profile	D10S1248	vWA	D16S539	D2S1338	D8S179	D21S11	D18S51	D22S1045	D19S433	TH01	FGA	D2S441	D3S1358	D1S1656	D12S391	SE33
Victim	*13,14*	*17,18*	*12,12*	*18,23*	*10,15*	*30,30*	*14,15*	*15,16*	*13,14.2*	*7,8*	*22,24*	*14,15*	*14,15*	*12,14*	*19,19.3*	*19,25.2*
Suspect	*14,15*	*14,16*	*12,13*	*17,22*	*15,15*	*31.2,33.2*	*13,16*	*16,16*	*12,15*	*8,9*	*20,22*	*10,10*	*14,18*	*12,17.3*	*18,23*	*15,24*
Other	*12*	*15*	*11*	*16,19,20*	*14*	*29,32.2*	*11,12*	*14,17*	*13.2,14,17*	*6,9.3*	*19,21,23*	*11,9*	*16,17*	*11,13,15,15.3*	*16,17,20*	*16,17,18,24.2,26.2,27.2*

After removal of peaks that are possibly non-allelic (intended as a guide only - not assumed by software)

Profile	D10S1248	vWA	D16S539	D2S1338	D8S179	D21S11	D18S51	D22S1045	D19S433	TH01	FGA	D2S441	D3S1358	D1S1656	D12S391	SE33
Victim	*13,14*	*17,18*	*12,12*	*18,23*	*10,15*	*30,30*	*14,15*	*15,16*	*13,14.2*	*7,8*	*22,24*	*14,15*	*14,15*	*12,14*	*19,19.3*	*19,25.2*
Suspect	*14,15*	*14,16*	*12,13*	*17,22*	*15,15*	*31.2,33.2*	*13,16*	*16,16*	*12,15*	*8,9*	*20,22*	*10,10*	*14,18*	*12,17.3*	*18,23*	*15,24*
Other				*20*	*14*		*12*		*14,17*	*6,9.3*		*11*	*17*	*15,15.3*	*17,20*	*17,27.2*

Unobserved, *unreplicated* and **replicated** peaks in provided reference profiles and those unattributable to any reference profile.

Summary

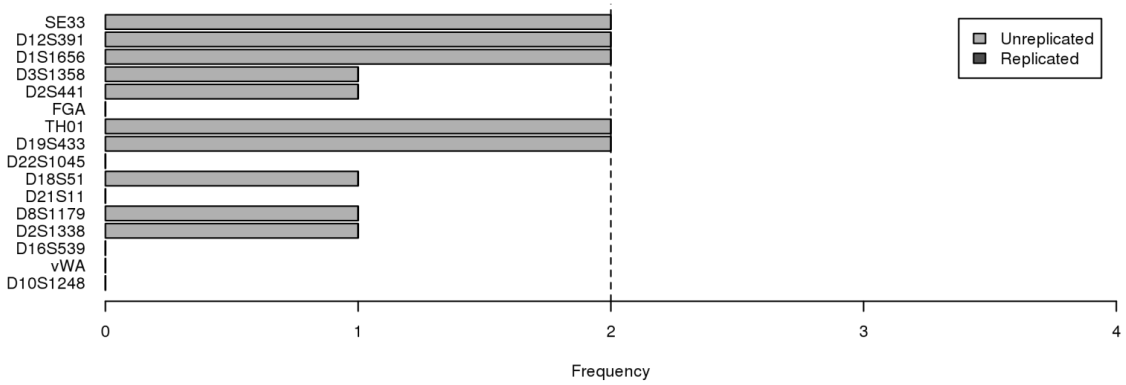
Reference profile	Victim	Suspect
Replicate: 3cont_3.u.1.t	1	0.9375
Overall	1	0.9375

Approximate representation (observed/total) for each reference profile per replicate and overall.

Reference profile	Victim	Suspect
Replicate: 3cont_3.u.1.t	1587.375	807.0625
Overall	1587.375	807.0625

Mean RFU for each reference profile per replicate and overall. This may be an over-estimate of the average DNA contribution of an assumed contributor due to sharing of alleles with other contributors. These values are for information purposes only, they are not used by the software.

Unattributable alleles



Number of unreplicated (light grey) and replicated (dark grey) unattributable alleles per locus, for the likely-allelic peaks (green allele labels shown in the CSP plots).

Alleles that are rare in at least one database

Locus	File	Profile	Allele	Issue	NDU1	NDU2	NDU3	NDU4	NDU6	NDU7
D2S1338	CSP	3cont_3.u.1.t	16	Rare allele	113	48	1	3	23	25
D18S51	CSP	3cont_3.u.1.t	11	Rare allele	35	4	3	1	1	3
D19S433	CSP	3cont_3.u.1.t	17	Rare allele	14	1	4	0	0	1
D2S441	CSP	3cont_3.u.1.t	9	Rare allele	6	1	5	0	0	1
D1S1656	CSP	3cont_3.u.1.t	15.3	Rare allele	169	13	8	0	5	8

Locus	File	Profile	Allele	Issue	NDU1	NDU2	NDU3	NDU4	NDU6	NDU7
D12S391	CSP	3cont_3.u.1.t	16	Rare allele	72	58	3	1	24	34
D12S391	CSP	3cont_3.u.1.t	19.3	Rare allele	23	0	1	0	0	0
D12S391	Reference	Victim	19.3	Rare allele	23	0	1	0	0	0
SE33	Reference	Suspect	24	Rare allele	0	4	0	1	2	2

Likelihoods at each locus

Likelihood	D10S1248	vWA	D16S539	D2S1338	D8S1179	D21S11	D18S51	D22S1045	D19S433	TH01	FGA	D2S441	D3S1358	D1S1656	D12S391	SE33
Prosecution.log10	-13.74	-17.18	-10.84	-22.53	-14.25	-15.83	-19.26	-14.69	-27.63	-17.78	-18.18	-19.04	-17.68	-22.77	-22.57	-28.19
Defence.log10	-14.39	-18.22	-11.39	-22.99	-15.30	-17.50	-20.27	-14.20	-24.79	-18.25	-18.74	-20.28	-18.39	-23.47	-23.92	-28.27
Ratio.log10	0.64	1.04	0.55	0.46	1.06	1.68	1.01	-0.49	-2.84	0.47	0.56	1.23	0.70	0.70	1.35	0.08
Ratio	4.40	10.95	3.54	2.87	11.45	47.35	10.35	0.32	0.00	2.98	3.63	17.08	5.03	4.98	22.56	1.20

Theoretical maximum LR = Inverse Match Probability (IMP)

calculation	estimate
likelihood ratio	9.35e+21
Log10 likelihood ratio	22.0

DNA contribution (RFU) and degradation estimates

Prosecution	U1	Victim	Suspect
Replicate: 3cont_3.u.1.t	931.92	981.81	152.21
Degradation	0.00042	0.00028	0.00259
Defence	**U1**	**X**	**Victim**
Replicate: 3cont_3.u.1.t	171.91	902.55	993.2
Degradation	0.00251	0.00035	0.00041

Dropin parameter estimates

hypothesis	dropin
Prosecution	-
Defence	-

User defined parameters

Parameter	User input
nUnknowns	1
ethnic	NDU1
adj	1
fst	0.03
relatedness1	0
relatedness2	0
relationship	0
doDropin	No
doDoubleStutter	Yes
doOverStutter	Yes
detectionThresh	20

Input files

File	Used
CSP	laboratory-CSP.csv
Reference	laboratory-reference.csv
Database	DNA17.txt (Default)

Seed used

Seed	Origin
1461927077	Randomly generated

Optimised parameters

Prosecution parameters

parameter	estimate	lower bound	upper bound
degradation1	-3.378	-20.000	-1.000
degradation2	-3.547	-20.000	-1.000
degradation3	-2.586	-20.000	-1.000
DNAcont1	931.922	0.000	5000.000
DNAcont2	981.807	0.000	5000.000
DNAcont3	152.207	0.000	5000.000
scale	53.923	0.000	1000.000
gradientS	0.005	0.000	0.010
gradientAdjust1	1.033	0.200	5.000
gradientAdjust2	1.348	0.200	5.000
gradientAdjust3	1.077	0.200	5.000
gradientAdjust4	1.164	0.200	5.000
gradientAdjust5	0.559	0.200	5.000
gradientAdjust6	1.235	0.200	5.000
gradientAdjust7	1.214	0.200	5.000
gradientAdjust8	1.298	0.200	5.000
gradientAdjust9	0.729	0.200	5.000
gradientAdjust10	0.798	0.200	5.000
gradientAdjust11	1.049	0.200	5.000
gradientAdjust12	0.900	0.200	5.000
gradientAdjust13	0.823	0.200	5.000
gradientAdjust14	0.963	0.200	5.000
gradientAdjust15	0.964	0.200	5.000
gradientAdjust16	1.197	0.200	5.000
meanD	0.001	0.000	0.100
meanO	0.003	0.000	0.100

Defence parameters

parameter	estimate	lower bound	upper bound
degradation1	-2.600	-20.000	-1.000
degradation2	-3.459	-20.000	-1.000
degradation3	-3.386	-20.000	-1.000
DNAcont1	171.905	0.000	5000.000
DNAcont2	902.548	0.000	5000.000
DNAcont3	993.200	0.000	5000.000
scale	49.208	0.000	1000.000
gradientS	0.005	0.000	0.010
gradientAdjust1	1.016	0.200	5.000
gradientAdjust2	1.369	0.200	5.000
gradientAdjust3	0.985	0.200	5.000
gradientAdjust4	1.371	0.200	5.000
gradientAdjust5	0.536	0.200	5.000
gradientAdjust6	1.231	0.200	5.000
gradientAdjust7	1.358	0.200	5.000
gradientAdjust8	1.264	0.200	5.000
gradientAdjust9	0.748	0.200	5.000
gradientAdjust10	0.776	0.200	5.000
gradientAdjust11	1.001	0.200	5.000
gradientAdjust12	0.909	0.200	5.000
gradientAdjust13	0.822	0.200	5.000
gradientAdjust14	0.947	0.200	5.000
gradientAdjust15	0.986	0.200	5.000
gradientAdjust16	1.136	0.200	5.000
meanD	0.001	0.000	0.100
meanO	0.002	0.000	0.100

Runtime

Parameter	Time
elapsed	22.931 mins
start	2016-04-29 10:32:30
end	2016-04-29 10:55:26

System information

Type	Details
Date report generated:	Fri Apr 29 10:55:30 2016
Package	likeLTD
Title	Tools to Evaluate DNA Profile Evidence
Description	Tools to determine DNA profile Weight of Evidence. For further information see the likeLTD guide provided, or Balding, D.J. (2013) <DOI:10.1073/pnas.1219739110>.
Depends	R (≥ 2.10), DEoptim, ggplot2, gtools, rtf
Suggests	svUnit, scales
Imports	gdata, tools, tcltk
Version	6.1.0
Date	2016-02-21
Author	David Balding, Adrian Timpson, Christopher Steele, Mayeul d'Avezac, James Hetherington.
Maintainer	Christopher Steele <c.steele.11@ucl.ac.uk>
License	GPL-3
URL	https://sites.google.com/site/baldingstatisticalgenetics/
Packaged	2016-04-29 09:11:22 UTC; csteele
Built	R 3.1.3; x86_64-unknown-linux-gnu; 2016-04-29 09:11:24 UTC; unix
sysname	Linux
release	2.6.32-573.7.1.el6.x86_64
version	#1 SMP Tue Sep 22 22:00:00 UTC 2015
nodename	ugi-151057.ugi.ucl.ac.uk
machine	x86_64
login	csteele
user	csteele
effective_user	csteele

C Laboratory protocol

To generate mixtures for validation purposes cheek swab samples were collected from 36 donors. DNA was extracted using a PrepFiler Express BTA™ Forensic DNA Extraction Kit and the Life Technologies Automate Express™ Instrument as per the manufacturer's recommendations.

Single-contributor and multi-contributor samples were created from the 36 DNA samples as shown in Table 8. These created samples were amplified using the AmpFℓSTR® NGMSelect® PCR kit as per the manufacturer's recommendations on a Veriti® 96-Well Fast Thermal Cycler for 30 cycles. The amplified PCR products were size separated by capillary electrophoresis using an ABI 3130 Sequencer, with 1 µL of the PCR products, 10 second injections and 3kV voltage. The results were analysed using GeneMapper® ID v3.2 with a detection threshold of 20 RFU, and no stutter threshold, so that both non-allelic and allelic peaks were recorded.

Peak height CSPs were converted to discrete CSPs using the same protocol as is used to which which peaks are called as allelic for the allele report (Table 1). Designations defaulted to the lowest confidence of calling a peak if a peak had multiple possible designations e.g. if we have a CSP with peaks 13,14,15 and peak heights 800,35,600, the 14 allele would be called as non-allelic if believed to be an OS of the 13 allele ($x = 0.044$), but uncertain if believed to be a S of the 15 allele ($x = 0.058$). In this situation the allelic call defaults to non-allelic due to the non-allelic call from the 13 parent peak.

# Cont	Single replicate			Multiple replicates		
	# Samples	Condition	# Reps	# Samples	Condition	# Reps
1	9	250	x1			
	9	62	x1			
	9	16	x1			
	9	4	x1			
2	12	Maj:Min (250:16)	x1	4	Maj:Min/2	x2
				4	Maj:Min/3	x3
				4	Maj:Min/4	x4
	12	Equal (31:31)	x1	4	Equal/2	x2
				4	Equal/3	x3
				4	Equal/4	x4
3	6	Unequal (250:62:16)	x1	2	Unequal/2	x2
				2	Unequal/3	x3
				2	Unequal/4	x4
	6	Equal (31:31:31)	x1	2	Equal/2	x2
				2	Equal/3	x3
				2	Equal/4	x4

Table 8: Laboratory protocol for generation of single-contributor and multiple-contributor CSPs from 36 donated DNA samples.