TITLE

Using Probabilistic Theory to Develop Interpretation Guidelines for Y-STR Profiles

AUTHORS

Duncan Taylor^{1,2}, Jo-Anne Bright³ and John Buckleton^{3,4}

1. Forensic Science South Australia, 21 Divett Place, Adelaide, SA 5000, Australia

2. School of Biological Sciences, Flinders University, GPO Box 2100 Adelaide SA, Australia
5001

3. ESR, Private Bag 92021, Auckland 1142, New Zealand

4. National Institute of Standards and Technology, 100 Bureau Drive, MS 8980 and 8314
Gaithersburg, MD 20899 United States of America

CORRESPONDING AUTHOR:

Duncan A. Taylor, PhD

Forensic Science South Australia

21 Divett Place

Adelaide SA

Australia 5000

Phone: +61-8 8226 7700

Fax: +61-8 8226 7777

Email: Duncan.Taylor@sa.gov.au

KEYWORDS

Y-STR; profile; interpretation; thresholds; probabilistic interpretation; guidelines.

26

27 **ABSTRACT**

28 Y-STR profiling makes up a small but important proportion of forensic DNA casework.
29 Often Y-STR profiles are used when autosomal profiling has failed to yield an informative
30 result. Consequently Y-STR profiles are often from the most challenging samples. In
31 addition to these points, Y-STR loci are linked, meaning that evaluation of haplotype
32 probabilities are either based on overly simplified counting methods or computationally
33 costly genetic models, neither of which extend well to the evaluation of mixed Y-STR data.
34 For all of these reasons Y-STR data analysis has not seen the same advances as autosomal
35 STR data. We present here a probabilistic model for the interpretation of Y-STR data. Due to
36 the fact that probabilistic systems for Y-STR data are still some way from reaching active
37 casework, we also describe how data can be analysed in a continuous way to generate
38 interpretational thresholds and guidelines.

39

40 **1.0 INTRODUCTION**

41

42 The last few years has seen substantial advances in the ability to interpret, analyse and
43 evaluate autosomal STR data [1-7]. Limited advances have been made for the interpretation
44 of Y chromosome STR profiles, although some investigation into stutter has appeared [8, 9].
45 Y chromosome work makes up a low proportion of casework and presents difficulties with
46 the evaluation of the data. The difficulties with the evaluation of Y-STR data can be broken
47 into two parts; firstly the lack of probabilistic modelling that has been applied to Y-STR data
48 and secondly the difficulties associated with the evaluation of the comparison of Y-STR
49 reference profiles/s to an evidence profile. We suspect that the difficulties in the latter are
50 largely responsible for the lack of work on the former. In this work we deal with the

probabilistic modelling of Y-STR data, for which much can be borrowed from already existing autosomal models.

This work is broken into two sections. The first of these is the deconvolution of a profile into its contributing haplotypes using probabilistic theory. We borrow from autosomal models that take into account factors that affect DNA profile behaviour such as template DNA amount, degradation, and locus specific amplification efficiencies. These models can be used to determine the probability of the observed Y-STR profile given potential contributing haplotypes. In doing this, the models allow a ‘weight’ to be given to each potential contributor haplotype set that acts as an indication of how well the proposed haplotypes describe the observed data. This then lends itself to development of interpretational guidelines.

As there is currently no tool available that can be used to analyse Y-STR data in the manner described within this paper the second section describes how the continuous thinking can be applied to assist in the creation of mixture interpretation guidelines and thresholds such as a stochastic threshold (ST).

1.1 Deconvolution of Y-STR data

1.1.1 Building an expected profile

The deconvolution of autosomal STR data relies on a number of models that are used to describe various properties of DNA profile behaviour. Ultimately, given some parameters we wish to describe the fluorescence observed in one or more electropherograms (epg). In this work we follow the autosomal model of [1] in that we consider the following ‘Mass’ parameters to describe observed fluorescence:

1. Template amount, t_n , for each of the n contributors. Collectively we refer to the vector of all template values as **T**.
2. Degradation, d_n , which models the decay with respect to molecular weight (m) in template for each of the contributors. We constrain degradation to be negative. Collectively we refer to the vector of all degradation values as **D**.
3. Amplification efficiency, A^l to allow for the observed amplification levels of each locus l . Amplification efficiencies are modelled by a normal distribution, $\log_{10}(A^l) \sim N(0, \sigma^2)$. Collectively we refer to the vector of all amplification efficiency values as **A**.
4. A replicate amplification multiplier R_r . This effectively scales all peaks up or down between replicates r . We constrain $\sum_r \log(R_r) = 0$. Collectively we refer to the vector of all replication amplification multiplier values as **R**.

Together we consider **D**, **A**, **R** and **T** as the mass parameters, **M**.

The total allelic product, T_{anr}^l , for an allele, a , at locus l , from contributor n in replicate r is modelled as

$$T_{anr}^l = A^l R_r t_n e^{d_n \times m_a^l} X_{an}^l \quad (1)$$

m_a^l is the molecular weight of allele a at locus l . Notice that this is the same formulation for total allelic product for autosomal STR data as described in [1], except that ‘dose’ (a term designated as X_{an}^l in [1], which took into account whether the individual was heterozygous or homozygous for allele a) can only takes values of 0 or 1 depending on whether contributor n contains allele a at locus l . The model we present, at this stage, only considers data for which the primer pair amplifies a single amplicon.

The total allelic product from an allele is split between back stutter, forward stutter and allelic peak. We model back stutter ratio, $\overleftarrow{\pi}_a^l$, and forward stutter ratio, $\overrightarrow{\pi}_a^l$, for allele a at locus l using models described in [10, 11]. The height of the allelic and stutter peaks formed from allele a are therefore:

$$E_{anr}^l = \frac{T_{anr}^l}{1 + \overleftarrow{\pi}_a^l + \overrightarrow{\pi}_a^l} \quad (2a)$$

$$E_{(a-1)r}^l = \overleftarrow{\pi}_a^l O_{ar}^l \quad (2b)$$

$$E_{(a+1)r}^l = \overrightarrow{\pi}_a^l O_{ar}^l \quad (2c)$$

Where $a-1$ signifies the back stutter product and $a+1$ signifies the forward stutter product. Note that we base the expected height of stutter products on the observed parent peak height rather than the expected parent peak height and so they do not possess a contributor term. Given the mass parameters the height of allelic peaks are expected to be independent within and between loci. Stutter peaks are dependent on their parent peak heights, but given this they are also expected to be independent within and between loci. This allows the deconvolution to occur in a locus by locus manner as described in [1], rather than having to consider the entire haplotype or haplotypic mixture as a whole entity. In this way then the deconvolution of Y-STR data becomes very similar to that of autosomal STR data, with the exception that, in our simplified model, a single allele is always expected at each locus from each contributor, rather than there being a possibility of one or two alleles being donated.

Given a set of values for mass parameters, $\mathbf{M}$, and a haplotype for each of the n contributors then an expected profile (which we call $\mathbf{E}$, and is made up of all individual E_{anr}^l terms) can be

built by summing the expected allelic peaks from the individual contributors and any back or forward stutters from alleles that differ by a single repeat unit.

$$E_{ar}^l = \overleftarrow{\pi}_{(a+1)r}^l O_{(a+1)r}^l + \overrightarrow{\pi}_{(a-1)r}^l O_{(a-1)r}^l + \sum_n E_{anr}^l$$

$$\mathbf{E} \equiv E_{1,1}^1 \dots E_{A,R}^L$$

1.1.2 Peak height variability

There exists a level of peak height variability within STR DNA profiles. We note that the expected peak heights (E) generated from the mass parameters may be different from the observed peak heights (O) in the epgs(s). In large part this is due to variation in the sampling of DNA molecules in a DNA extract for inclusion in PCR reaction [12]. As a base for modelling peak height variability we follow [1, 11]. We extend that theory here so that the difference in an observed peak from its expected height $\Pr(O_{ar}^l | E_{ar}^l)$ is modelled by equations 3a, 3b and 3c. In these equations the terms O_{ar}^l and E_{ar}^l on the left hand side of the equations refer to the peak being examined and the '+1' or '-1' on the right hand side of the equations signify peaks relative to the peak in question. Also note that 3a and 3b are the same but we list both to assist in the understanding of equation 3d.

$$\log\left(\frac{O_{ar}^l}{E_{ar}^l}\right) \sim N\left(0, \frac{c^2}{E_{ar}^l}\right) \quad \text{for purely allelic peaks} \quad (3a)$$

$$\log\left(\frac{O_{ar}^l}{E_{ar}^l}\right) \sim N\left(0, \frac{c^2}{E_{ar}^l}\right) \quad \text{for purely forward stutter peaks,} \quad (3b)$$

$$\log\left(\frac{O_{ar}^l}{E_{ar}^l}\right) \sim N\left(0, \frac{k^2}{O_{(a+1)r}^l}\right) \quad \text{for purely back stutter peaks and} \quad (3c)$$

$$\log\left(\frac{O_{ar}^l}{E_{ar}^l}\right) \sim N\left(0, \frac{\mathbb{S}_{ar}^l k^2}{O_{(a+1)r}^l} + \frac{(\mathbb{A}_{ar}^l + \mathbb{F}_{ar}^l) c^2}{E_{ar}^l}\right) \quad \text{for composite peaks} \quad (3d)$$

where $\mathbb{A}_{ar}^l$, $\mathbb{S}_{ar}^l$ and $\mathbb{F}_{ar}^l$ are allelic, back stutter proportion and forward stutter proportions of the peak, so that $\mathbb{A}_{ar}^l + \mathbb{S}_{ar}^l + \mathbb{F}_{ar}^l = 1$.

The values of k^2 and c^2 in autosomal work have prior gamma distributions and the point values for these terms is a parameter within the model, so that the analysis can adjust to account for profiles that are more or less variable than the ‘average variability’ (see [13] for work describing the nature of the k^2 and c^2 terms and factors that affect their prior distribution).

For autosomal STR analysis the values of k^2 and c^2 have separate prior distributions, $\Gamma(\alpha_1, \beta_1)$ and $\Gamma(\alpha_2, \beta_2)$ respectively, and are independent parameters within the model. The work in [13] describes how Markov Chain Monte Carlo (MCMC) can be used to provide the peak height variability framework described above, and profiles of known source to obtain the peak height variance constant prior distributions. This works in part because there are two allelic peaks expected at each heterozygous locus, and these peaks act as intra-locus calibrators of each other’s expected heights. In the case of haplotypic markers, such a calibration does not exist. The consequence within the MCMC of not having an intra-locus calibrator could be that a disproportionate amount of peak height variability is assigned to either allelic or stutter peaks (i.e. the model could explain all variability exists in allelic peak heights and that stutter peaks have no peak height variability). To counter this we restrict the evaluation of prior distributions of variance constants to be equal, $\Gamma(\alpha_1, \beta_1) = \Gamma(\alpha_2, \beta_2) = \Gamma(\alpha, \beta)$. In autosomal work this assumption is found to be approximately true depending on the resolution of the stutter model in use (data not shown).

As long as $\Gamma(\alpha, \beta)$ can be determined this now provides a means to assess the differences between expected and observed peak heights within Y-STR profiles.

1.1.3 Haplotypic weight

As mentioned in section 1.1.1, with given mass parameters the expected height of peaks within and between loci are independent. We follow the work of [1] by describing the weight (w_j^l) for a proposed genotype set j at locus l (S_j^l) as the probability of the observed data given the genotype set and integrated across the mass parameters:

$$w_j^l = \int \prod_a \prod_r \Pr(O_{ar}^l | S_j^l, \mathbf{M}) \Pr(\mathbf{M}) d\mathbf{M} \quad (4)$$

And that the weight for the entire profile haplotype is obtained by multiplication of the weights at each locus:

$$w_j = \prod_l w_j^l \quad (5)$$

which is an approximation for the integral that considers whole profile weights as integrals across the whole profile, rather than locus by locus. This approximation has been shown to be valid [1].

We use MCMC in the method described by [1] to evaluate equation 4. In brief, we change all mass parameters and the genotype at one randomly chosen locus within the MCMC at each iteration. The acceptance or rejection of the proposed values is decided by the Metropolis-Hastings algorithm and the weights for each genotype set become the proportion of iterations that contain them.

1.2 Determining the peak height variability of Y-STR data

We use Yfiler® Plus (Life Technologies, CA) as an example for determining peak height variability. The following settings were used:

- Saturation: 7000 rfu determined as per section 3.7 of [10].
- Analytical threshold: 30 rfu determined as per appendix 1 of [10].
- Back stutter ratio: Regressions of allele vs back stutter ratios were used in accordance with stutter ratio results obtained from Life Technologies Yfiler® Plus manual [14] chapter 5, Figures 11 to 17. The table of regression values used are given in Table 1.
- Forward stutter ratio: A profile-wide forward stutter ratio of 0.01 was used.

Table 1: back stutter ratio regression results

Locus	In regression $\pi_a = \beta_0 + \beta_1 a$ value below are: β_0, β_1
DYS576	-0.0311, 0.0078
DYS389I	-0.0700, 0.0100
DYS635	-0.0922, 0.0078
DYS389II	-0.1600, 0.0100
DYS627	-0.0700, 0.0080
DYS460	-0.0525, 0.0125
DYS458	-0.0467, 0.0089
DYS19	-0.1300, 0.0150
YGATAH4	-0.0860, 0.0140
DYS448	-0.0275, 0.0025
DYS391	-0.0400, 0.0100
DYS456	-0.0900, 0.0133
DYS390	-0.2167, 0.0133
DYS438	-0.0200, 0.0057
DYS392	-0.0550, 0.0125
DYS518	-0.1220, 0.0080
DYS570	-0.0133, 0.0067
DYS437	-0.1000, 0.0100
DYS385	-0.0400, 0.0100
DYS449	-0.1300, 0.0100
DYS393	-0.0583, 0.0117
DYS439	-0.0500, 0.0100
DYS481	-0.0844, 0.0122
DYF387S1	-0.2280, 0.0090

DYS533	-0.0400, 0.0100
--------	-----------------

DNA from two male individuals was extracted twice using Chelex (Bio-Rad Laboratories, CA) and twice using DNA IQ (Promega Corporation, WI) and amplified at 500, 400, 100, 50, 20, 10, 5 and 1 pg in triplicate, giving a total of 96 samples. All samples containing saturated data or fewer than 10 data points were excluded from the dataset, giving total of 67 samples. This group of 67 samples were analysed using the method described in section 1 of [13]. The distribution for the peak height variance constant (not shown) has mean of approximately 9 and a mode of approximately 6. There is some arbitrariness as to choosing a point value from a distribution with which to develop thresholds, and an argument could be made to choose either the mode or the mean, or some other quantile. We have chosen the mean. This value can be ‘sanity checked’ by comparison with stochastic effects seen in observed data. This is carried out in a manner similar to that described in section 3.4 in [10], however instead of considering heterozygous imbalance we graph $\log\left(\frac{O_{ar}^l}{E_{ar}^l}\right)$ against E_{ar}^l for allelic peaks. The expected height for E_{ar}^l is obtained by the expected back stutter ratio and observed back stutter peak height:

$$E_{ar}^l = \frac{O_{(a-1)r}^l}{\frac{\pi_a^l}{\pi_a}} \quad (6)$$

Doing so for the 96 samples in the Yfiler Plus dataset produces the graph seen in Figure 1, where the dotted lines represent the 95% bounds on expected variability (see [10] for an explanation of the origin of these bounds) and the dashed line represents the approximate bound on the observed data based on an analytical threshold of 30 rfu (approximate because it uses an average value of $\pi_a^l = 0.09$, which is the average value for all peaks in the dataset).

Figure 1: Observed peak height variability showing 95% bounds (dotted lines) and bound on observed data (dashed line)

Because of the bound on observations it is difficult to assess the coverage of the expected 95% bounds on $\log\left(\frac{O_{ar}^l}{E_{ar}^l}\right)$. It would be expected that 2.5% of data falls above the upper bound dotted line, and should not be affected by the bound on observations. 1.5% of observations seen in Figure 1 fall above the upper bound 95% interval, which represents a reasonable alignment with theoretical expectations.

As described in [13] the distribution of locus amplification efficiency variances over a number of samples can be modelled during a component-wise MCMC using an exponential distribution (termed in [13] as a hyper-distribution). Doing so for the dataset described here

produced an exponential distribution with mean value 0.03. We then use this mean value as σ^2 in the locus amplification efficiency model by $\log_{10}(A') \sim N(0, 0.03)$.

2.0 Developing Y-STR guidelines from probabilistic assessment

Without access to probabilistic software that can apply a continuous interpretation method to Y-STR profiles, laboratories will require some rules or guidelines to assist with a ‘manual’ interpretation. Typically, such guidelines will include:

- An analytical threshold
- A saturation threshold
- A stochastic threshold (sometimes referred to as a dropout or homozygote threshold)
- Mixture ratio/proportion guidelines.

The analytical threshold and saturation threshold for Y-STR profiling systems can be determined in the same manner as for autosomal STR systems.

2.1 Stochastic threshold (ST)

It is worth considering whether the term ‘stochastic/dropout/homozygote threshold’ has meaning in the context of Y-STR profiles. Classically for autosomal STR profiles this denotes a height at which, if we see a single allelic peak, (and given the level of peak height variability for that laboratory/process) we would have a defined level of confidence that it does not possess a partner that had fallen below the analytical threshold. In the context of Y-STR data the same meaning cannot be ascribed as peaks within a locus do not have partners (even loci that have apparent multiple alleles are really amplifications of several different loci from the one primer pair). For single source Y-STR data a ST is more useful to distinguish an allele that has dropped out from a null allele. This by necessity is based on the heights of allelic peaks at other loci. In mixtures the consideration of dropouts (or distinguishing

dropout from null alleles) has consequences for haplotype interpretation. Andersen et al. [15] consider dropout probabilities and locus amplification efficiency in a model that includes the molecular weight of the locus in question in order to consider whether dropout or null allele is more likely. We use this same concept, but apply our models for peak height variability and locus amplification efficiency. We use the same probability of a null allele, $\Pr(N) = 0.0002$ as used in [15].

Therefore we consider two events that could cause an apparent dropout:

- 1) Underrepresentation of the allele during sampling the DNA extract for PCR
- 2) Poor amplification efficiency of the locus

We seek two values; firstly, the probability of dropout, of a peak with height E_{ar}^l (which we approximate as the average peak height of allelic peaks in the profile for single source profiles), and secondly, the point at which the probability that a peak has dropped out becomes more likely than the probability that the absence of an allele at that locus is due to the presence of a null allele. Let the haplotype with dropout be S_1 and the haplotype with a silent allele be S_2 . Formally the probability of the data given the haplotype is

$$\int \prod_M \prod_l \prod_a \prod_r \Pr(O_{ar}^l | S_j^l, \mathbf{M}) \Pr(\mathbf{M}) d\mathbf{M} \text{ which we can approximate with equations 4 and 5.}$$

Let us imagine that the locus with potential dropout is x then the probability of the data for each haplotype is:

$$\int \prod_M \prod_l \prod_a \prod_r \Pr(O_{ar}^l | S_1^l, \mathbf{M}) \Pr(\mathbf{M}) d\mathbf{M} \text{ for haplotype 1 and} \quad (7a)$$

$$\Pr(N) \times \int \prod_{M \neq x} \prod_l \prod_a \prod_r \Pr(O_{ar}^l | S_1^l, \mathbf{M}) \Pr(\mathbf{M}) d\mathbf{M} \text{ for haplotype 2.} \quad (7b)$$

We simplify Equations 7a and 7b by considering that the mass parameters template and degradation provide an expected height, E^* , before amplification efficiency has been

considered. By doing this we simplify the problem to consider just the locus where the potential dropout/null has occurred. Therefore to calculate the probability of dropout, $\Pr(O_{ar}^l < Z)$ we consider the combination of amplification efficiency and peak height variability that lead to potential dropout. This is achieved by:

$$\Pr(O_{ar}^l < Z) = \int_{A^l} \Pr(O_{ar}^l < Z | E^*, A^l) \Pr(A^l) dA^l \quad (8)$$

$$\text{where } E_{ar}^l = E^* \times A^l \text{ and } \Pr(O_{ar}^l < Z | E^*, A^l) = \int_{O=0}^Z \Pr(O) \Pr\left[\log\left(\frac{O}{E_{ar}^l}\right)\right] dO$$

In words, amplification efficiencies are considered that would change the expected height from E^* to E_{ar}^l and then the probability of dropout is considered for a peak at height E_{ar}^l , integrated across all values of amplification efficiency. Using these models the probability given by equation 8 is dominated by peak height variability at low expected heights and by amplification efficiencies at high peak heights. Figure 2 shows the results of applying equation 8 to peak heights E_{ar}^l from Z (30 rfu) to 1000 rfu and using $\sigma^2 = 0.03$. The model in equation 8 relies on having an expected peak height of E_{ar}^l , and this can be approximated from the average allelic peak height across a profile, or by application of a model that accounts for degradation and molecular weight. Also shown in Figure 2 is the probability of dropout (as seen in equation 8) without taking into account locus amplification efficiency.

Figure 2: Probability of dropout (solid line) for expected heights E_{ar}^l when $Z=30$ rfu. The dashed line is $\Pr(N)$ (the silent allele probability) and the dotted line represents the probability of dropout when amplification efficiency is not taken into account.

From the results seen in Figure 2 an approximate value that could be used for distinguishing dropout from a null allele could be approximately 350 rfu. The probability of dropout line in Figure 2 (using equation 8) is that which would be used to consider the probability of dropout if there were multiple allelic peaks present in a profile. For example a 1:1 mixture where only a single peak is present could be treated similarly to an autosomal profile when deciding whether the risk that not all information has amplified above the analytical threshold (AT) is negligible. A probability cutoff value could be chosen, such as 0.001, and the corresponding height used as a threshold (from the dotted line in Figure 2 this would be approximately 200 rfu).

2.2 Interpreting a major component

We will use the results of the probabilistic estimation of peak height variability from section 1.2 to address mixture proportion guidelines.

316

317 Consider first a two person Y-STR profile, with L loci. The individuals have contributed
 318 DNA in unequal amounts. We define the proportion of contributor 1 as m_1 and of contributor
 319 2 as m_2 . These are unknown. To develop guidelines, which allow us to assign the apparent
 320 major component to the major contributor with some level of confidence we need to consider:

- 321 • Varied mixture proportions
- 322 • Varied profile intensities
- 323 • What probability do we accept for the incorrect assignment of a major contributor.

324 There are $L-d$ loci showing two peaks. d loci show one peak and that is either because both
 325 peaks are the same allele, or one allele has dropped, or one allele is silent.

326 Consider initially the loci showing two peaks. If we call the larger peak at each two allele
 327 locus E_M and the smaller E_m , then we could imagine that all $L-d$ of the larger peaks are
 328 attributable to the major and the remainder to the minor. Term this arrangement S_1 . However
 329 there are many other potential arrangements. Consider initially the swap of two peaks at one
 330 specific locus such that the smaller peak is assigned to the major and the larger to the minor.
 331 Term this arrangement S_2 .

332 The probability of any particular assignment, j , is estimated as

333
$$\int \prod_M \prod_l \prod_a \prod_r \Pr(O_{ar}^l | S_j^l, \mathbf{M}) \Pr(\mathbf{M}) d\mathbf{M} .$$
 We require

334
$$\frac{\int \prod_M \prod_l \prod_a \prod_r \Pr(O_{ar}^l | S_1^l, \mathbf{M}) \Pr(\mathbf{M}) d\mathbf{M}}{\int \prod_M \prod_l \prod_a \prod_r \Pr(O_{ar}^l | S_2^l, \mathbf{M}) \Pr(\mathbf{M}) d\mathbf{M}} > \alpha$$

335 Where α is a likelihood ratio threshold we wish to apply. The formulation above avoids
 336 assuming a known ratio, but is complex to apply. We again make a simplifying assumption
 337 that the profile ratio is known and available to be applied to the methodology below. In

practise this is likely to be average ratio observed across the profile. There will be some increased uncertainty associated with this practise.

We start with a mixture that possesses contributors in proportions 2/3 and 1/3 (or 2:1 in ratio). Two peaks are observed at O_1 and O_2 where $O_1 \sim 2 \times O_2$. There is a possibility that O_1 has been donated by the major contributor and O_2 by the minor. There is also possibility that O_1 has been donated by the minor contributor and O_2 by the major, but peak height variability has caused the apparent flip in relative peak heights. Possibility 1 is going to be more likely than possibility two, as it is a better fit with the mixture proportions, but the question for interpretation is whether the alternative is so unlikely that it can be discounted. This will depend on the absolute peak heights of O_1 and O_2 .

If the profile is adhering to the mixture proportions then the expected heights for the major (E_1) and minor (E_2) are:

$$E_1 = \frac{2}{3}(O_1 + O_2) \text{ and } E_2 = \frac{1}{3}(O_1 + O_2)$$

Therefore the probability of the observed data if O_1 is from the major and O_2 is from the minor (denoted as haplotype set 1, S_1) is:

$$p(\mathbf{O} | \mathbf{E}, S_1) = \Pr \left[\log \left(\frac{O_1}{E_1} \right) \right] \times \Pr \left[\log \left(\frac{O_2}{E_2} \right) \right] = \Pr \left[\log \left(\frac{O_1}{\frac{2}{3}(O_1 + O_2)} \right) \right] \times \Pr \left[\log \left(\frac{O_2}{\frac{1}{3}(O_1 + O_2)} \right) \right]$$

where the probabilities are enumerated using equation 3a.

And the alternative where O_2 is from the major and O_1 is from the minor (denoted as haplotype set 2, S_2) is:

$$p(\mathbf{O} | \mathbf{E}, S_2) = \Pr \left[\log \left(\frac{O_2}{\frac{2}{3}(O_1 + O_2)} \right) \right] \times \Pr \left[\log \left(\frac{O_1}{\frac{1}{3}(O_1 + O_2)} \right) \right]$$

Whether or not we are willing to interpret O_1 as the major then is determined by the peak heights and the confidence we desire in the assignment. Let us say our criterion was:

$\frac{p(\mathbf{O} | \mathbf{E}, S_1)}{p(\mathbf{O} | \mathbf{E}, S_2)} > \alpha$. This criterion requires the evidence to be α times more likely under S_1 than

under S_2 . This should not be confused with stating that S_1 is α times more likely than S_2 .

Figure 3 shows the value of $\frac{p(\mathbf{O} | \mathbf{E}, S_1)}{p(\mathbf{O} | \mathbf{E}, S_2)}$ over a range of heights for O_1 and O_2 , where we

consider a 2:1 mixture and assume $O_1 = 2O_2$, a 3:1 mixture and assume $O_1 = 3O_2$, a 4:1

mixture and assume $O_1 = 4O_2$ and a 5:1 mixture and assume $O_1 = 5O_2$.

Figure 3: Likelihood ratio considering the probability of the observed peaks if the higher peak is from the major contributor as opposed to the smaller peak being from the major contributor.

From Figure 3 the thresholds on the height of the largest peak that are suggested for the assignment of the largest peak as coming from the major component for 2, 3, 4 or 5 to 1 mixtures (where both peaks have been observed in approximately that configuration) are 1022 rfu, 455 rfu, 304 rfu and 234 rfu respectively using a threshold of $\alpha = 1000$.

We can also consider the situation where only a single peak has been observed at a locus and we wish to know when it can be assigned to the major with confidence. Again a formal treatment would see the probability of the observed data considered in light of the two competing scenarios:

S_1 : The major contributor is the source of the observed peak,

S_2 : The minor is the source of the observed peak with the major having dropped out,

and integrated across all mass parameters so that assumptions regarding mixture proportions, DNA amount, degradation and amplification efficiency are not required. Again we make some simplifying assumptions in order to simply determine a threshold for this data. If the single peak is height O and the mix proportions are m_1 and m_2 then we evaluate the described scenario by:

$$p(\mathbf{O} | \mathbf{E}, S_1) = \frac{1}{2} \Pr \left[\log \left(\frac{O_1}{O_1} \right) \right] \Pr(O_{ar}^l < Z | E = \frac{m_2}{m_1} O_1) + \frac{1}{2} \Pr \left[\log \left(\frac{O_1}{O_1} \right) \right] \text{ and}$$

$$p(\mathbf{O} | \mathbf{E}, S_2) = \Pr(O_{ar}^l < Z | E = \frac{m_1}{m_2} O_1) \Pr \left[\log \left(\frac{O_1}{O_1} \right) \right]$$

Note that under $p(\mathbf{O} | \mathbf{E}, S_1)$ we consider that either the minor has dropped out (left part of the equation) or that it is masked under the major (right part of the equation), each with equal prior probability. Notice here that we simplify to the point that we are not considering the impact that amplification efficiency may be having on the locus, however this is likely to lead to a conservative threshold (i.e. requiring higher fluorescence in order to interpret) and so is

justifiable. Using the above formulations the points at which $\frac{p(O|E, S_1)}{p(O|E, S_2)} > 1000$ is obtained for 2:1, 3:1 or 5:1 are 93, 60 or 40 rfu respectively. The threshold of interpretation can be plotted against the mixture ratios in order to obtain a relationship between the two properties that allows thresholds to be considered for a range of ratio. This can be seen in Figure 4, where a power trend line has been fitted to the data.

Figure 4: Relationships between the mixture ratio for two person mixtures and the threshold required for the higher peak to be assigned to a major component

A similar approach can be used to consider more complex mixtures just as long as all genotypic sets are considered that include, dropout or masking when determining the likelihood ratio. Without deriving in the same detail as above, we provide some scenarios in Table 2 as examples for 3 person profiles. Note that the first row considers the same imbalances as the two person scenario above and so the thresholds are the same.

Table 2: DNA profile examples and potential thresholds for assigning the larger peak to a major component of a 3 person mixture

DNA Profile description	Scenarios	Value of x when
-------------------------	-----------	-------------------

		$\frac{p(O E, S_1)}{p(O E, S_2)} > 1000$
3 person profile with 3 peaks $O_A = x$ rfu $O_B = y$ rfu $O_C = y$ rfu $x > y$	S_I : Major = [A], minors = [B] and [C] S_2 : Major = [B] or [C], minors = [A] and [C] or [B]	$x:y = 2:1 - 1022$ rfu $x:y = 3:1 - 455$ rfu $x:y = 4:1 - 304$ rfu $x:y = 5:1 - 234$ rfu
3 person profile with 2 peaks $O_A = x$ rfu $O_B = y$ rfu $x > y$	S_I : Major = [A], minors = [A] and/or [B] and/or [Q] S_2 : Major = [B] or [Q], minors = [A] and [Q] and/or [B]	$x:y = 2:1 - 4500$ rfu $x:y = 3:1 - 738$ rfu $x:y = 4:1 - 389$ rfu $x:y = 5:1 - 274$ rfu
3 person profile with 1 peak $O_A = x$ rfu	S_I : Major = [A], minors = [A] and/or [Q] S_2 : Major = [Q], minors = [A] and possibly [Q]	$x:y = 1:1 - 453$ rfu $x:y = 2:1 - 286$ rfu $x:y = 3:1 - 227$ rfu $x:y = 4:1 - 199$ rfu $x:y = 5:1 - 182$ rfu

413

414 Although the same system of threshold determination could be generated for profiles of any
415 complexity, we suggest that the higher order mixtures (for example greater than three males,
416 or profiles complicated in some other way such as differential degradation or inhibition) may
417 be out of the scope of manual interpretation and better suited to analysis in probabilistic
418 systems.

419

420 2.3 interpreting a minor component

We consider here the situation where a single peak has been observed at a locus in a profile where the two contributors are in uneven proportions. Interpreting a minor component can be treated in the same way as a major component, although the added complexity that the minor peak could be masked under the major must be taken into account. There are two scenarios that we consider:

- 1) S_1 : The peak at height O is made up of two individuals donating an A allele and its observed height is equal roughly to what is expected. The numerator of the LR may be approximated by the density of the normal distribution $N\left(0, \frac{c^2}{O}\right)$ using the mode of the distribution for the variance
- 2) S_2 : The peak at height O is only from the major contributor and the other peak, expected to have occurred as height $O \frac{m_2}{m_1}$ has dropped out. This probability can be determined by equation 8.

Given an observed peak height O for allele A , we must consider the point where the ratio of likelihoods $\frac{p(O | E, S_1)}{p(O | E, S_2)}$ exceeds some value (α). To be consistent with previous interpretations we will use a value of $\alpha = 1000$. Define the proportion that the major contributor is donating to the mixture as m_1 and the proportion from the minor contributor is m_2 (where for a two person mixture $m_1 = 1 - m_2$). In this simplified treatment of the data we are not going to consider the effects that amplification efficiency may have on the interpretation and so are losing some power.

Figure 5: The probability of a single observed peak given the scenario that the minor is masked under the major peak vs that the minor peak has dropped out for various mixture proportions.

From Figure 5, and using an *LR* threshold of 1000 (arbitrarily chosen) it would be acceptable to designate the minor peak in a mixture as masked by the major for 2, 3, 4 and 10 to 1 mixtures when the height of the single observed peak was 270, 520, 1000 and 2340 rf units respectively. Again, the threshold of peak height required for interpretation of a masked minor component can be plotted against the mixture ratio as in Figure 6.

451

452

453 *Figure 6: Relationships between the mixture ratio for two person mixtures and the threshold*
 454 *required for assigning a minor allele as masked by the major*

455

456 2.4 Variation in mixture proportions across a profile

457 The thresholds given above are developed using simplifying assumptions. It is likely that
 458 there will be variability in the mixture proportions across a profile and the question may be
 459 asked as to whether the mixture proportion at any one locus deviates sufficiently from a
 460 profile average that it casts doubt on the interpretation, particularly for higher order mixtures.

461 Consider a locus, which has mixture proportions of m_1 and m_2 and a total observed height
 462 (the sum of the peaks) of O_T rfu. Peaks are expected to be present at $E_1 = O_T m_1$ and
 463 $E_2 = O_T m_2$ rfu. We know from equation 3a that $O_T m_1$ will be observed at height O_1 for some
 464 proportion of the time as given by equation:

465 $O_1 = O_T m_1 \times 10^x$

where x is determined as the value of $\log\left(\frac{O_1}{E_1}\right)$ for which a normal distribution $N\left(0, \frac{c^2}{E_1}\right)$ is at the desired quantile. If we consider both peaks together (and make the simplifying assumption that the imbalance probability is apportioned evenly between the two peaks) then we are interested in the point at which both values of $\log\left(\frac{O}{E}\right)$, considered together, would reach the desired quantile. To demonstrate this concept we have considered mixtures with ratios of 1:1, 2:1 and 10:1. Figure 7 shows the 95% intervals on the mixture proportion variation that would be expected for a 1:1 mixture.

Figure 7: Variation expected in 1:1 mixture ratio for a range of peak intensities. Dashed lines show the 95% interval bounds.

Typically, interpretation thresholds that deal with allowable variation in the relative contribution of contributors are based on mixture proportions. Threshold such as $|D| \leq 0.2$ when average peak height is greater than 300 rfu have been suggested for autosomal profiling

systems [16]. That is, the mixture proportion of any individual locus being interpreted must fall within ± 0.2 of the profile-wide average value. We demonstrate the performance of a mixture proportion (Mx) threshold of ± 0.2 on data that has mixture ratios of 1:1, 2:1 and 10:1 (or mixture proportions of 0.5, 0.34, 0.09) in Figure 8.

Figure 8: Variation expected in 0.5 (left), 0.34 (middle) and 0.09 (right) mixture proportions, m , for a range of peak intensities. Data is shown only for the minor contributor mixture proportion. Dashed lines show the 95% interval bounds. Solid lines show a value of ± 0.2 from the known mixture proportion.

Note that in Figure 8 for the 10:1 profile there is no lower bound on the acceptable mixture proportion of the minor contributor. Also from all three graphs in Figure 8 it appears that the mixture proportion threshold of ± 0.2 covers most data that is above a typically ‘stochastic range’. Another method that could be used would be to determine at which peak intensity point the mixture proportion reached 0.5, which for the 2:1 and 10:1 ratios would be 580 rfu and 164 rfu respectively. Whilst this second method may allow more interpretation due to its increased flexibility it is also more difficult to apply.

2.5 determining peak height variability without continuous analysis software

If a laboratory does not have access to computing systems that can determine the peak height variability in the manner described by section 1.2 then an approximate value will need to be determined using empirical observations. This can be done by comparing the observed peak heights with the expected peak heights (determined from the stutter peak heights and the stutter ratio for that allele and locus combination) and adjusting the value of c^2 until quantile bounds have their expected coverage. This is basically retrofitting the variance constant to empirical observations as seen in Figure 1. This can be done by:

- a) Determine the expected height for the allelic peak at locus l in profile i , E_i^l , by

$$E_{ar}^l = \frac{O_{(a-1)r}^l}{\pi_a}$$

- b) Graph $\log\left(\frac{O_{ar}^l}{E_{ar}^l}\right)$ against E_{ar}^l for allelic peaks

- c) Graph the 95% lower bound by $-Z\sqrt{\frac{2c^2}{E}}$, where Z determines the quantile

- d) Adjust c^2 until only 2.5% of the data falls above the upper 95% bound

Doing the above for the dataset described in section 1.2 gives a value of $c^2 = 5.4$, which is in the mode of the variance constant distribution. This is lower than the value of $c^2 = 9$, which is the mean of the distribution that was used for thresholds. This difference is likely to decrease as the dataset used to derive it increases in size, at which point the manually derived value is likely to converge to the continuously determined value.

The second aspect required for interpretation is an inter-locus variance, σ^2 , for the amplification efficiency model:

$$\log_{10}(A^l) \sim N(0, \sigma^2) \quad (9)$$

Determination of a value for σ^2 can be carried out by:

- a) Determining the average peak height for each of the I profiles, $\overline{E}_i$
- b) Calculate the amplification efficiency factor for each locus that would be required for it to equal $\overline{E}_i$. For example if the observed height of the allele at locus l in profile i is O_i^l , then the amplification efficiency is calculated by $A_i^l = \frac{O_i^l}{\overline{E}_i}$
- c) The likelihood of the dataset, L , is calculated by $L = \prod_i \prod_l \Pr(A_i^l)$, where $\Pr(A_i^l)$ is modelled by equation 9
- d) The value of σ^2 is varied to maximise L .

Carrying out this procedure for the dataset described in 1.2 yields an amplification efficiency variance value of $\sigma^2 = 0.04$, which is close to the value given in section 1.2 of $\sigma^2 = 0.03$. The increased value for σ^2 when using the method described above is likely due to the fact that in the manual method above there is no degradation term, therefore any differences in peak heights will need to be accounted for by amplification efficiency. The continuous model described in section 1.2 does have a model for degradation and hence the lower amplification efficiency variance. The approximation of no degradation using the manual method is likely to have a small impact on the value of σ^2 as long as the profiles being considered do not exhibit marked degradation.

3.0 Assessing the interpretation models

In order to assess the performance of the suggested continuous and binary interpretation models we generated mixed DNA profiles from two, three and four male individuals in dilutions as specified in Table 3. Mixed DNA samples were amplified in triplicate as per manufacturer's instructions to obtain a total of 90 Yfiler Plus profiles.

Table 3: Mixtures constructed for analysis

Tubes	mixture proportions for contributor				Total DNA added to PCR (pg)
	One	Two	Three	Four	
1-4	0.50	0.50			500,200,50,20
5-8	0.33	0.67			
9-12	0.17	0.83			
13-15	0.09	0.91			
16-18	0.02	0.98			
19-21	0.33	0.33	0.33		500,200,50
22-24	0.50	0.33	0.17		
25-27	0.25	0.25	0.25		
28-30	0.40	0.30	0.20	0.10	

3.1 Probabilistic models

The method described in [1] and section 1 of this paper described the biological and statistic models used to analyse Y-STR data in a fully continuous manner. Section 2 uses these continuous models in order to develop interpretational thresholds than can be applied in a manual fashion, or with the assistance of an Excel spreadsheet in a relatively straightforward manner. Before applying the thresholds to mixture data we must first show that the continuous models on which they are based are indeed functioning correctly. We apply the continuous methodology using a modified version of STRmix™ (<http://strmix.esr.cri.nz/>) [1] (which we refer to here as STRmixY, and is currently in a developmental state) to carry out this function. It is impractical to display all results for all mixtures analysed in this study and so in Table 4 we show the results for the 1:1 mixture deconvolution, displaying all haplotype combinations and weights generated by the continuous method. The mixture, when generated, was in reality slightly divergent from the target 1:1 and the continuous interpretation picked this up, hence the weights in table 4 are not split evenly between contributor 1 and 2. A copy of the profile is given in Figure 9 and this divergence can be seen. In Table 4 the genotype sets associated with the known contributors are bolded. It can

be seen in all but one instance the correct genotype set was given the highest weight. In the one instance that the highest weight was not given to the known genotypes set (in DYS448), the weights correctly reflect the results obtained in the epg (Figure 9).

Table 4: Weights for 500pg 1:1 mixture PCR 1. Contributor 1 and 2 percentages are the posterior mean mixture proportions from STRmixY. Bolded values indicate the correct genotypic assignments.

Locus	Contributor 1 Genotype (59%)	Contributor 2 Genotype (41%)	Weight
DYS576	[16] [17]	[17] [16]	0.914 0.086
DYS389I	[12]	[12]	1.000
DYS635	[21]	[21]	1.000
DYS389II	[29] [28]	[28] [29]	0.906 0.094
DYS627	[19]	[19]	1.000
DYS460	[11] [10]	[10] [11]	0.926 0.074
DYS458	[15]	[15]	1.000
DYS19	[14]	[14]	1.000
YGATAH4	[10] [11]	[11] [10]	0.923 0.077
DYS448	[19] [20]	[20] [19]	0.753 0.247
DYS391	[10]	[10]	1.000
DYS456	[15] [14]	[14] [15]	0.904 0.096
DYS390	[23] [22]	[22] [23]	0.916 0.084
DYS438	[10]	[10]	1.000
DYS392	[11]	[11]	1.000
DYS518	[40] [39]	[39] [40]	0.634 0.366
DYS570	[18] [20]	[20] [18]	0.926 0.074
DYS437	[16]	[16]	1.000
DYS449	[26] [28]	[28] [26]	0.855 0.145
DYS393	[13]	[13]	1.000
DYS439	[11]	[11]	1.000

DYS481	[27]	[25]	0.922
	[25]	[27]	0.078
DYF387S1	[38]	[37]	0.917
	[37]	[38]	0.083
DYS533	[11]	[11]	1.000

570

571

572

573

574

Figure 9: epg for 500pg 1:1 mixture PCR 1

Figure 10 shows the target mixture proportions of the analysed profiles and the mean of the posterior distribution for mixture proportions obtained from deconvolution by STRmixY.

Ability of STRmixY to predict mixture proportions

Figure 10: Target mixture proportions compared with posterior mean mixture proportions from deconvolution

From these analyses the deconvolution of mixed Y-STR data can proceed in the same fashion as for autosomal data. There are other aspects of the continuous deconvolution of these samples, such as the assessment of locus amplification efficiencies, allelic and stutter peak height variability, degradation levels and a number of analysis diagnostics that we do not show here for reasons of brevity, but give further assurances that the continuous methods and underlying models are describing the observed data well. From the results of the probabilistic

modelling we are confident to proceed on to testing the manual interpretation thresholds based on these models.

3.2 Threshold models

We now subject the two person profiles detailed in Table 3 to the threshold based interpretation rules generated in section 2. We generate the average mixture ratio, $\overline{M}_r$, by

1. Averaging value of the log of highest observed peak divided by the lowest observed peak at each locus to provide the apparent mixture ratio at a locus, $\hat{M}_r^l$, and

2. Taking 10 to the power of this value $\overline{M}_r = 10^{\frac{1}{L} \sum \log(\hat{M}_r^l)}$. This can then be converted to an average mixture proportion by, $\overline{M}_x = \frac{\overline{M}_r}{1 + \overline{M}_r}$.

We then apply rules rigidly using Excel to all two person mixtures. For the major component at a locus to be interpretable:

If there are two observed peaks:

1. The observed mixture ratio must be within ± 0.2 of $\overline{M}_x$ (as per section 2.4) (in terms

of $\hat{M}_r^l$, this is $\frac{\overline{M}_x - 0.2}{1.2 - \overline{M}_x} \leq \hat{M}_r^l \leq \frac{\overline{M}_x + 0.2}{0.8 - \overline{M}_x}$) and

2. The peak height of the major peak must be greater than or equal to $2221.2 \times \overline{M}_r^{-1.223}$ (as per Figure 4 left pane).

If there is one peak:

3. Then it must be greater than or equal to $181.03 \times \overline{M}_r^{-1.055}$ (as per Figure 4 right pane).

Even given these three rules there is a requirement for an additional rule that prevents interpretation of a major (or minor) component in a mixture when the observed peak heights

are too close. We create this rule with consideration of the created and observed ratios for the two person mixtures at different concentrations, as seen in Figure 11.

Figure 11: Target and apparent mixture ratios for two person mixtures at different total DNA amounts. Profile averages are given as black bars and individual locus values are given as hollow circles

Figure 11 shows that there is overlap between individual locus mixture ratios at all levels of input DNA and overlap of profile averages when 50 pg or less is added to the PCR. As expected the observed ratio for the more disparately prepared mixtures is much less than the target ratios. This is because the expected peak heights of the minor components are below the analytical threshold, and so all observed values will be those that are stochastically higher

than expected. Given the results of Figure 11 we implement one additional interpretation threshold for interpreting major components:

4. For a locus with two peaks the observed mixture ratio at that locus must be greater than 2:1, i.e. $\hat{M}_r^l \geq 2$.

Note that rule 4 is somewhat more arbitrarily chosen than the other interpretational thresholds. The modelling of data to develop this threshold is more complex because it involves the observed average mixture ratio, the mixture ratio at the locus being examined and the height of the peak involved. Multiple regression could be performed to develop this rule, however we feel that it is likely to become overly cumbersome to apply in practise and so we simply choose a value that should perform conservatively.

To interpret a minor component, thresholds 1, 2 and 4 apply, and the modified rule 3 becomes:

5. If there is one peak present then it must be greater than or equal to $237.27 \times \overline{M}_r$ (as per Figure 6).

The results of applying these rules to the two person profiles on all non-duplicated loci (i.e. all loci excluding DYS385) are shown in Table 6. The maximum number of loci at which a component could be interpreted is therefore 24. The number of loci that can potentially be interpreted will depend on the ratio (i.e. only single peak loci can be interpreted from 1:1 mixtures) and the number of loci where information is detected.

Table 6: The number of loci at which the major and minor contributor can be assigned for the two person mixtures outlined in Table 3 (results of 3 PCRs combined). The asterisked results are explained within the text.

Ratio	DNA (pg)	major loci interpreted	minor loci interpreted	loci with alleles detected
1:1	500	37*	37*	72
2:1		38	38	72
5:1		70	70	72
10:1		72	63	72
50:1		72	16	72
1:1	200	36	36	72
2:1		36	36	72
5:1		63	49	72
10:1		70	24	72
50:1		72	20	72
1:1	50	32	3	72
2:1		33	4	72
5:1		36	0	72
10:1		36	1	71
50:1		28	2	72
1:1	20	9	0	62
2:1		3	0	54
5:1		13	0	52

652

653 For the results in Table 6 all interpreted alleles for both major and minor contributors were
654 assigned correctly. The sample asterisked in table 6, possessed a genotypic assignment for
655 major and minor at a locus where the alleles of major and minor contributors were different
656 from each other. For the locus in question the observed average mixture ratio was 1.5:1 and
657 the locus at which the interpretation was made possessed two alleles; 10 at 891 rfu and 11 at
658 1976 rfu (locus DYS460 from Figure 9). Applying the thresholds 1 to 5 interpreted a major
659 contributor at this locus as having an 11 allele. All other loci interpreted for this profile
660 contained only a single peak (and so both contributors could be interpreted as having this
661 allele). While the above interpretation is correct, in that one of the contributors has this allele,
662 we highlight it as the target mixture proportion for this profile was 1:1 and so we expected
663 that no component would have been interpreted at this locus. It is worth noting that the
664 highest weight given by the continuous interpretation method to this sample (table 4) for any

of the loci with two distinct alleles was 0.93, at locus DYS460 and DYS570. In total 756 assignments were made of a major contributor and 399 assignments of a minor contributor out of the (54×24) 1296 potential assignments.

In a broader sense the results of the manual interpretation give reassurance that the continuous models used to describe DNA profile behaviour can be manipulated to generate binary thresholds for use in manual profile interpretations. It is also worth noting that the dataset to which these thresholds have been applied is not the training set used to generate the thresholds in the first place. This has been done so as to evaluate their use as they will be applied in casework.

4.0 Conclusion

Y-STR data can be crucial to forensic investigations, particularly when autosomal STR information has failed to yield an informative result. The relatively low proportion of cases that utilise Y-STRs, combined with difficulties in evaluating an *LR*, due to their mode of inheritance, has meant that the advances seen in autosomal STR analysis have not been mirrored for Y-STR data. This leaves many laboratories with limited power to develop interpretation guidelines and calculate evidential weights for Y-STR data, particularly for mixed profiles. In this work we develop methods that overcome some of these limitations by suggesting ways that laboratories can develop interpretational guidelines, based on the same probabilistic theory that has advanced autosomal interpretation. We demonstrate the generation of interpretation guidelines based on dropout, mixture proportion, peak balances and major/minor component configuration and applied them to laboratory generated mixtures. The reason we demonstrate methods for generating interpretational guidelines at all is due to the realisation that a publically available probabilistic system for Y-STR data

interpretation is unlikely in the short term. In the interim some methods for manual interpretation are still required.

We also applied the probabilistic methods directly (without the use of thresholds) to the same data. As has been demonstrated in a number of works (for example see [1]) the application of thresholds (which are typically designed to be conservative) is wasteful of data and leads to only a fraction of the generated mixed DNA profiles able to be used for calculation of an *LR*.

We note that we have not considered the effects of multiple PCR replicates on the ability to manually interpret Y-STR mixtures in this work. The use of multiple replicates will increase the confidence of the analyst in their interpretation and hence could be used to reduce the interpretational requirements (i.e. lower peak heights for interpreting a major component).

CONFLICTS OF INTEREST:

The Authors of this work are the technical developers of the commercially available software product STRmix™. The developers in no way profit financially from commercial STRmix™ activities.

ACKNOWLEDGEMENTS:

This work was supported in part by grant 2014-DN-BX-K028 from the US National Institute of Justice. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice or Commerce. Certain commercial equipment, instruments, or materials (or suppliers, or software) are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor

does it imply that the materials or equipment identified are necessarily the best available for the purpose.

REFERENCES:

- [1] Taylor D, Bright J-A, Buckleton J. The interpretation of single source and mixed DNA profiles. *Forensic Science International: Genetics*. 2013;7:516-28.
- [2] Haned H. Forensim: An open-source initiative for the evaluation of statistical methods in forensic genetics. *Forensic Science International: Genetics*. 2011;5:265-8.
- [3] Lohmueller K, Rudin N. Calculating the weight of evidence in low-template forensic DNA casework. *Journal of Forensic Sciences*. 2013;58:234-59.
- [4] Steele CD, Balding DJ. Statistical Evaluation of Forensic DNA Profile Evidence. *Annual Review of Statistics and Its Application*. 2014;1:361-84.
- [5] Puch-Solis R, Rodgers L, Mazumder A, Pope S, Evett I, Curran J, et al. Evaluating forensic DNA profiles using peak heights, allowing for multiple donors, allelic dropout and stutters. *Forensic Science International: Genetics*. 2013;7:555-63.
- [6] Balding DJ, Buckleton J. Interpreting low template DNA profiles. *Forensic Science International: Genetics*. 2009;4:1-10.
- [7] Perlin MW, Legler MM, Spencer CE, Smith JL, Allan WP, Belrose JL, et al. Validating TrueAllele® DNA mixture interpretation. *Journal of Forensic Sciences*. 2011;56:1430-47.
- [8] Bright J-A, Curran J, Buckleton J. Modelling PowerPlex® Y stutter and artefacts. *Forensic Science International: Genetics*. 2014;11:126-36.
- [9] Andersen MM, Olofsson J, Mogensen HS, Eriksen PS, Morling N. Estimating stutter rates for Y-STR alleles. *Forensic Science International: Genetics Supplement Series*. 2011;3:e192-e3.
- [10] Taylor D, Bright J-A, McGoven C, Hefford C, Kalafut T, Buckleton J. Validating multiplexes for use in conjunction with modern interpretation strategies. *Forensic Science International: Genetics*. 2015;<http://dx.doi.org/10.1016/j.fsigen.2015.09.011>.
- [11] Bright J-A, Taylor D, Curran JM, Buckleton JS. Developing allelic and stutter peak height models for a continuous method of DNA interpretation. *Forensic Science International: Genetics*. 2013;7:296-304.
- [12] Gill P, Curran J, Elliot K. A graphical simulation model of the entire DNA process associated with the analysis of short tandem repeat loci. *Nucleic Acids Research*. 2005;33:632-43.

- [13] Taylor D, Buckleton J, Bright J-A. Factors affecting peak height variability for short tandem repeat data. (under review). 2015.
- [14] Butler JM. Chapter 7 - Low-Level DNA and Complex Mixtures. In: Butler JM, editor. Advanced Topics in Forensic DNA Typing: Interpretation. San Diego: Academic Press; 2015. p. 159-82.
- [15] Andersen MM, Mogensen HS, Eriksen PS, Olofsson JK, Asplund M, Morling N. Estimating Y-STR allelic drop-out rates and adjusting for interlocus balances. Forensic Science International: Genetics. 2013;7:327-36.
- [16] Bright J-A, Turkington J, Buckleton J. Examination of the variability in mixed DNA profile parameters for the Identifiler(TM) multiplex. Forensic Science International: Genetics. 2010;4:111-4.