Research paper

Production of high-fidelity electropherograms results in improved and consistent DNA interpretation: Standardizing the forensic validation process

Kelsey C. Peters[a], Harish Swaminathan[a], Jennifer Sheehan[a], Ken R. Duffy[b], Desmond S. Lun[c,d,e], Catherine M. Grgicak[a,c,f,*]

[a] Biomedical Forensic Sciences, Boston University School of Medicine, United States
[b] Hamilton Institute, Maynooth University, Ireland
[c] Center for Computational and Integrative Biology, Rutgers University, United States
[d] Department of Computer Science, Rutgers University, Camden, United States
[e] Department of Plant Biology and Pathology, Rutgers University, New Brunswick, USA
[f] Department of Chemistry, Rutgers University, Camden, United States

ARTICLE INFO

ABSTRACT

Samples containing low-copy numbers of DNA are routinely encountered in casework. The signal acquired from these sample types can be difficult to interpret as they do not always contain all of the genotypic information from each contributor, where the loss of genetic information is associated with sampling and detection effects. The present work focuses on developing a validation scheme to aid in mitigating the effects of the latter. We establish a scheme designed to simultaneously improve signal resolution and detection rates without costly large-scale experimental validation studies by applying a combined simulation and experimental based approach. Specifically, we parameterize an *in silico* DNA pipeline with experimental data acquired from the laboratory and use this to evaluate multifarious scenarios in a cost-effective manner. Metrics such as $signal_{1copy}$-to-noise resolution, false positive and false negative signal detection rates are used to select tenable laboratory parameters that result in high-fidelity signal in the single-copy regime. We demonstrate that the metrics acquired from simulation are consistent with experimental data obtained from two capillary electrophoresis platforms and various injection parameters. Once good resolution is obtained, analytical thresholds can be determined using detection error tradeoff analysis, if necessary. Decreasing the limit of detection of the forensic process to one copy of DNA is a powerful mechanism by which to increase the information content on minor components of a mixture, which is particularly important for probabilistic system inference. If the forensic pipeline is engineered such that high-fidelity electropherogram signal is obtained, then the likelihood ratio (LR) of a true contributor increases and the probability that the LR of a randomly chosen person is greater than one decreases. This is, potentially, the first step towards standardization of the analytical pipeline across operational laboratories.

1. Introduction

Samples containing low-copy numbers of DNA are routinely encountered in casework and are challenging to interpret because of the inherent complexity associated with determining contributing genotypes. Interpretation of DNA evidence is often carried out within the likelihood ratio (LR) framework, which assesses the weight of evidence by comparing the probability of observing the data under two different hypotheses. In the forensic context, the hypotheses compare the probability of observing the evidence, E, given that a specific person of interest was a contributor, H_1, versus the probability of evidence given that an unknown person contributed, H_2, and is expressed as

$$LR = \frac{\Pr(E|H_1)}{\Pr(E|H_2)} \qquad (1)$$

In recent years, a number of probabilistic genotyping systems have been developed [1–7] and perform this computation, where E is presumed to consist of information obtained through the amplification and electrophoresis of a set of forensically relevant short tandem repeats (STRs). This evidence contains information related to the length of the amplified fragments as well as the number of DNA target molecules amplified as measured by an electropherogram (EPG), wherein each peak in the EPG may originate from any combination of three sources: 1) true allele; 2) instrument noise; and 3) artefact. Previous work has demonstrated that the level of detail incorporated into probabilistic

models used for inference can substantially impact interpretation [8–10].

Core to any inference is, of course, the quality and information content of the EPG. Consequently, many strategies to improve EPG information content have been evaluated. They include the application of enhanced injection parameters [11], additional PCR cycles [11,12], upgraded amplification chemistries [13], purification of post-PCR product [14], and more sensitive instrumentation [15]. These strategies all result in an increase in signal with respect to the target number of DNA molecules, which we call α; this is the sensitivity of the process.

We utilize the recommendations set forth by the International Union of Pure and Applied Chemistry in distinguishing 'sensitivity' from the terms 'analytical threshold', 'limit of detection (LOD)' and 'signal-to-noise resolution'. Here, sensitivity is defined as

$$\alpha = \left(\frac{dS}{dT_c} \right) \tag{2}$$

where α is the slope of the tangent at nominal amplicon quantity T_c of an analytical curve, which is constant if the curve is linear. Consequently, a method may be extremely sensitive (i.e. large change in signal with quantity), but baseline noise may also be high leading to poor signal-to-noise resolution and a relatively large LOD. For discussions on sensitivity, LOD and signal-to-noise we refer the readers to [16–20].

In general, valuable EPG information can be garnered by lowering the LOD, or the number of molecules that can be detected. Along with improving sensitivity, another relatively simple way to accomplish this is to decrease the signal threshold (S_T) at which signal detection is demarcated [21]. The extreme version of this is to set S_T to the lowest possible level allowed by the analytical process and evaluate the signal in its entirety [22,23]. The S_T that is implemented into casework is often referred to as the analytical threshold (AT), and we reserve AT for that purpose.

There is an inherent relationship between α, S_T and the information content contained in the evidence E. Even in cases where no S_T is applied, the sensitivity of the laboratory process, which is dependent upon parameters such as PCR cycle number and injection conditions, determines the amount of information available for import into inference systems. Thus, development of a tenable DNA validation system that informs laboratories of α and S_T effects on the information content contained in E is warranted.

If no DNA target molecules are present at the start of amplification, then the signal is due to random noise or artefact. If one or a few target copies of a given allele are present then, depending on the laboratory conditions and sensitivity of the process, allele signal may be difficult to distinguish from noise. Alternatively, low-level signal may fall below some pre-defined S_T. In general, current practice relies upon the application of a S_T to the EPG signal, S, such that the probability of the signal observed due to noise surpasses that of S_T does not exceed a given γ [24],

$$\Pr(S > S_T | T_{c=0} = 0) \le \gamma. \tag{3}$$

Here, $T_c = 0$ is the number of DNA copies present at the start of amplification and γ can be any value defined by the laboratory. It may be based on the post-analysis interpretation scheme where a larger risk of detecting signal in the absence of target copies may be tolerated if a probabilistic system that models noise is utilized, while a small γ may be appropriate in cases where binary or manual interpretation techniques prevail.

With the advent of probabilistic genotyping, large signal thresholds that reduce the risk of noise detection to negligible levels may not be desirable since there is a tradeoff between the false detection of noise and allele dropout. There are two main reasons for allele dropout: 1) detection effects [21]; and 2) sampling effects [25–27]. The former occurs when DNA target molecules are successfully amplified but there

are too few amplicons produced to yield fluorescent signal that exceeds baseline levels. The latter arises when the DNA molecule is not available for amplification. This may occur when the DNA template is fragmented or when the DNA target molecule did not survive the pre-PCR steps and, as a result, is not present in the amplification tube. From a signal detection perspective, the risk of false non-detection of true signal, as defined in [24], is slightly modified for forensic DNA detection purposes such that we focus on defining the probability of signal not exceeding S_T when one copy of amplifiable DNA is present at the start of amplification,

$$\Pr(S < S_T | T_{c=0} = 1) \le \beta \tag{4}$$

Therefore, the S_T is the level at which the target substance is decidedly considered to be part, or not part, of E. Note that although Eq. (4) is related to the probability of allele dropout, it is not equivalent to it. Previous work has demonstrated that detection of a single DNA molecule is possible [28] as long as the sensitivity is modified to a degree that ensures allelic signal surpasses the pre-defined S_T [11].

Historically, forensic DNA detection has relied upon the implementation of an AT that minimizes γ [29]. As with the authors of [30], we distinguish allele drop-in from noise detection, focusing on the signal detection problem in an attempt to design a full laboratory procedure that will lead to good signal-to-noise resolution and an LOD of one copy of DNA. Despite its unexpected nature, allele drop-in is, therefore, categorized as allele signal rather than noise and an AT designed to detect most allele signal will, by definition, also detect allele drop-in. As such, allele drop-in and the propensity of the laboratory to observe drop-in ought to be considered at the interpretation stage. With the implementation of probabilistic genotyping programmes [1,5] there has been renewed interest in re-evaluating the practice of minimizing γ in favor of utilizing lower thresholds [10]. Despite this, we emphasize that evaluating the tradeoff between the false non-detection of alleles and the false detection of noise without first considering sensitivity and signal-to-noise resolution would inevitably lead to large overall signal detection error rates.

Recently a stochastic model of EPG generation encompassing all aspects of the entire forensic DNA laboratory process, from quantification to peak detection, was described [31], where it was shown that good signal to noise resolution was readily achieved by the application of slight modifications to the laboratory conditions under which the DNA samples are processed. We extend that work here by developing a validation strategy to determine an optimized AT for the forensic DNA process. This AT is established by careful evaluation of the false detection rates only after good noise to signal$_{1copy}$ (i.e., RFU signal obtained from 1 copy of amplifiable DNA) resolution is acquired. To determine the laboratory parameters that result in a sensitivity, α, that produces adequate resolution, we generate simulated capillary electrophoresis (CE) signal from $T_{c=0} = 1$ and $T_{c=0} = 0$ copies of DNA via the stochastic model described in [31]. For each laboratory condition, we compute the false positive and false negative detection rates for various S_T values for the large synthesized dataset to acquire a good approximation of the detection error rates. The PCR cycle number, injection parameters and S_T that minimize the false positive and negative detection rates while still maintaining a reasonable dynamic range (i.e., large mass range of DNA that can be analyzed before the detector is saturated) are chosen as the validated laboratory process. If appropriate, the AT may be conditioned on the false positive rate.

We evaluate the validation protocol and confirm that it is applicable across two CE platforms resulting in signal detection capabilities that are at the single-copy level, regardless of instrument. We confirm that the optimized S_T values acquired from simulation coincide with those obtained from experimental samples and show the utility of this validation procedure by demonstrating that improving the information content through optimized detection strategies aids probabilistic interpretation. That is, we show that high-fidelity EPGs garnered from 1-, 2- and 3-person samples result in fewer LRs favoring the prosecution's

hypothesis when the defense's hypothesis is true and establish that there is an increase in the LR for true minor contributors to these samples. Further, we demonstrate that high-fidelity EPGs from two CE platforms produce consistent summary statistics.

2. Methods

2.1. Experimental data generation in the multi- and single- copy regimes

Ninety-five single source profiles of known genotype were obtained by extracting DNA from whole blood or proficiency test samples purchased from various manufacturers using phenol/chloroform purification and alcohol precipitation. The extracts were quantified using Quantifiler® Duo (Life Technologies, Carlsbad, CA) on the Applied Biosystems® 7500 (Applied Biosystems, Foster City, CA) using the manufacturer's recommended thermal cycling protocol and a validated, universal calibration curve [32]. The experimental data were used for two purposes: 1) to parameterize the model (Sec. 2.3); and 2) to evaluate the false negative and positive rates for a set of experimental data in the single-copy regime (Sec. 2.4). The data used to parameterize the model were not also used to evaluate it. Seven low template target amounts (0.25, 0.125, 0.063, 0.047, 0.031, 0.016, 0.008 ng) were amplified at 29 cycles using the AmpFlSTR® Identifiler® Plus kit (Applied Biosystems®, Foster City, CA). Fragment separation was performed on an ABI 3130 Genetic Analyzer (Applied Biosystems®, Foster City, CA) using a 3 kilovolt (kV) injection for 5, 10, and 20 s and on an ABI 3500 Genetic Analyzer (Applied Biosystems®, Foster City, CA) using a 1.2 kV injection for 5, 15, and 25 s. The EPGs were analyzed in GeneMapper® ID-X (Applied Biosystems®, Foster City, CA) using the Local Southern Method with an AT of 1 RFU. Pull-up, complex pull-up, and – A artefacts were filtered using CleanIt, a visual basic script that examines the intensity and position of the suspected artefact peak against the intensity and position of the peak of interest (http://sites.bu.edu/grgicak/software/). Pull-up was defined as a peak which appears in the same position (± 0.3 bp) as the peak-of-interest in another dye channel and has a peak height of ≤6% of the peak-of-interest. Complex pull-up was defined as a peak with a plateau-like shape located between two adjacent peaks-of-interest in a different dye channel and less than 6% of those peaks-of-interest; – A was defined as a peak one bp shorter (± 0.3 bp) than the peak of interest. There were no height restrictions for – A artefacts. Spikes or peaks of large RFU value that were present in multiple dye channels at the same bp, were manually filtered. The data were exported as .csv files and used for downstream analysis and interpretation.

2.2. Description of the model in RESOLVIt

The computational tool RESOLVIt carries out the simulation of a large number of artificial EPGs by *in silico* execution of a laboratory's amplification and electrophoresis protocols according to the SEEIt model specified in Ref. [31]. Each simulation begins with the random sampling of two different alleles at every locus. This effectively forces all contributors during simulation to be heterozygous thus avoiding complications that arise from assigning total RFU level (signal plus noise) to each allele in a homozygous sample. Since the objective of this work is to investigate the ability of a laboratory protocol to distinguish $signal_{1copy}$ from noise, dropout that stems from sampling effects is ignored and the number of copies of an allele that undergo amplification is forced to be 1; thus, $T_{c=0} = 1$ for every allele. The PCR process is simulated for the chosen number of cycles *via* a multitype Galton-Watson process, as detailed in [31]. That is, at each cycle, each amplicon of an allele is successfully copied with a probability corresponding to the amplification efficiency, and if copied it gives rise to stutter with a given stutter slippage probability, whereupon the stutter amplicons are then amplified at later rounds of PCR. After the last PCR cycle, we simulate the CE process, converting amplicons into

fluorescence and adding instrument noise.

Noise is a stochastic instrument artifact where spurious additional fluorescent data points appear in the EPG. If several noise points occur within a fluorescence bin that contains no allele or stutter signal, the combined effect can be such that peak detection software declares what we call a noise peak. For a given experimental data set created from known genotypes, the empirical probability of a noise peak occurrence, P_N, can be readily evaluated. With S_i being the fluorescent peak height in non-allele, non-stutter position $i \in \{1, ..., b\}$, as determined by the peak detection software, and x being the smallest peak threshold that can be set in the peak detection software, then P_N is the fraction of bins where a peak is declared due to noise:

$$P_N = \frac{1}{b} \sum_{i=1}^{b} \chi(S_i \geq x),\tag{5}$$

where χ is the indicator function taking the value 1 if the condition is true and 0 if it is false.

An empirical study employing GeneMapper IDx with $x = 1$ RFU for several large-scale data sets [33] has reported that the fraction, P_N, of non-allele, non-stutter bins that are labeled as peaks is ca. 15%. That paper also reports that given a noise peak is observed, the conditional distribution of the peak height in RFU is well described by a lognormally distributed random variable. The EPG simulation model we employ here, and in the model described in [31], describes the post-peak detection processed signal. Thus following those studies, the occurrence of each noise peak is modeled as an independent random variable that with probability 1-P_N is zero and with probability P_N it is lognormally distributed with mean μ_{noise} and variance σ_{noise}^2. As the likelihood of noise-peak occurrence, P_N, and the distribution of their peak heights when they do occur, as determined by μ_{noise} and variance σ_{noise}^2, depend on the peak declaration software used as well as that software's parameterization, they are user-defined quantities in the RESOLVIt tool.

The CE sensitivity parameter, α, is estimated from experimental data for every locus by summing the peak height corresponding to signal and stutter at a locus from experimental samples of known genotype and determining the linear function of the nominal amplicon number versus peak height [31]. This enables the calculation of the total fluorescence, corresponding to signal plus noise plus stutter, at every allele position based on the number of amplicons present after amplification.

After the completion of this step, we obtain information about the peak heights of 2X alleles at every locus, where X is the number of EPGs simulated. In order to understand the false positive detection rate, an equal number (2X) of noise alleles are randomly simulated at every locus and their heights are sampled from a lognormal distribution. The mean and standard deviation of the lognormal distribution are derived for every locus by modeling the mean and standard deviation of noise peaks from a large number of experimental known samples as a linear function of the nominal amplicon number.

Since the simulated peaks are of known origin (allele or noise), the false positive detection rate (proportion of noise peaks that fall above S_T) and the false negative detection rate (proportion of allele peaks that fall below S_T) are calculated for a range of S_T values. This information enables a forensic laboratory performing validation to investigate the impact on the two error rates with (increasing or decreasing) S_T. The S_T that meets the laboratories operational requirements is selected as the AT utilized in casework. The AT implemented in operations can be chosen such that it corresponds to the lowest false positive rate or the lowest false negative rate or both. Further, the computational system described herein allows for fast exploration of values for laboratory parameters, such as the number of PCR cycles and injection time that enhance the discernment of signal from noise. For reference, simulation of 10,000 profiles takes ∼25 min on a 1.9 GHz Intel® Xeon® processor with 4 GB RAM.

2.3. Parameterization of the model using laboratory data

The model is parameterized by utilizing a set of laboratory data generated with the methods described in Sec. *2.1*. Supplemental Table S1 details the sensitivity, α, determined for 5 (*Lab 1*), 10 (*Lab 2*) or 20 s (*Lab 3*) 3 kV ABI 3130 Genetic Analysis injection settings and for the 25 s (*Lab 4*) 1.2 kV injection settings for the ABI 3500 Genetic Analysis platform, as described in [31]. In brief, signal intensity was plotted against the nominal amplicon number to obtain the estimated rise in signal per amplicon. The nominal amplicon number at cycle, *c*, is approximated assuming a PCR efficiency, Eff, of 0.96 [31] and the following equation

$$N_c = N_0(1+\text{Eff})^{c-2} \tag{6}$$

where N_c is the number of amplicons produced at cycle *c*, N_0 is the target, or initial, amplicon number, and *c* is the cycle number.

Though stutter is not directly evaluated when determining signal-to-noise resolution, stutter production during PCR is expected to affect the number of allele amplicons produced. Slippage that occurs in early PCR cycles leads to lower numbers of full-allele amplicons. Thus we include stuttering in our model and utilize the stutter slippage probabilities shown in Supplemental Table S2. Specifically, stutter probabilities were estimated by comparison of the stutter ratios obtained through simulation and the ratios previously obtained [13]. The average stutter ratios obtained through simulation and the estimated stutter ratios provided by the manufacturer is found in Supplemental Table S2.

As reported in [33] we set the probability of non-zero noise peaks, P_N, to 0.15 (see Supplemental methods for an example) and the average and standard deviation of the non-zero noise peaks were estimated from examining 643 single-source samples amplified at targets ranging from 0.25 to 0.008 ng of DNA. Supplemental Table S3 reports the noise parameters used during simulation.

2.4. Detection error tradeoff analysis

Consistency between synthetic and experimental peak intensity had been previously confirmed [31]. Data from four distinct laboratory setups were used to create simulated emulations, *Lab 1-4*. S_T values, ranging from 1 to 50 RFU in 5 RFU increments, were used to compute the detection error tradeoff (DET) rates for simulated data obtained for simulations *Lab 1, 2,* and *3*, corresponding to the 3130 platform. For simulation *Lab 4,* corresponding to the more sensitive 3500 platform, S_T values ranging from 1 to 105 RFU in 5 RFU increments were used. False positive detection error rates were determined by counting the number of noise peaks with heights larger than S_T and dividing by the number of non-zero noise peaks. False negative rates were estimated by determining the proportion of signal obtained from one copy of DNA that fell below S_T.

DET analysis was also performed on experimental data from 95 single-source samples of known genotype with target mass of 0.008 ng. In the case of experimental samples, we could not confirm the number of template copies that generated the signal. Thus, for experimental data the signal was categorized in two ways: 1) as signal falling in noise positions; and 2) as signal falling in allele positions. The range of S_T values was the same as for the corresponding simulated setting. The false positive rate was determined by counting the number of peaks in noise position that exceeded S_T divided by the total number of non-zero noise peaks in noise position. The false negative rate was determined by counting the number of peaks in allele position, based on the known genotype, that fell below S_T and dividing by the expected number of alleles. Only heterozygous loci were used to determine false positive and negative rates from experimental samples.

2.5. Impact of signal resolution and detection on the LR for a true contributor and on the probability that LR > 1 for a non-contributor

We examined both single source and mixture samples to investigate the effects of information content and detection errors on the LR, and on the proportion of LRs that exceeds one when one million random genotypes are tested in lieu of a known contributor. We utilized the previously published probabilistic model CEESIt [6] to determine these quantities. Briefly, CEESIt is built on a continuous mixture interpretation model that incorporates noise, stutter, stochastic PCR effects as well as random contributor levels. Let *E* denote the evidence, *i.e.* the sample's electropherogram. Let $H_1(g)$ denote the hypothesis that *E* arises from a contributor with genotype *g* in conjunction with a fixed number $k \geq 0$ of unknown contributors whose genotypes are selected randomly with given frequencies. Let H_2 denote the hypothesis that all $k + 1$ contributors have genotypes selected at random with those given frequencies. Define the likelihood ratio for a specific genotype *g* to be

$$LR(g) = \frac{\Pr(E|H_1(g))}{\Pr(E|H_2)} \tag{7}$$

With *s* denoting a suspect's genotype and *G* denoting a randomly selected genotype, CEESIt computes approximations to $LR(s)$ and to the distribution of the random variable $LR(G)$. It achieves this by computing an approximation of $\Pr(E|H_1(g))$, $\Pr(E|\hat{H}_1(g))$, through Monte Carlo sampling (see Ref. [6] for details). In addition, CEESIt generates a large number, *n,* of hypotheses $H_{2,1}, H_{2,2}, …, H_{2,n}$ where the genotypes of all $k + 1$ contributors are randomly chosen independently according to their frequencies in the population, excluding the genotype *s*. For each hypothesis $H_{2,i}$, it computes an approximation of $\Pr(E|H_{2,i})$, $\Pr(E|\hat{H}_{2,i})$, through Monte Carlo Sampling. It then approximates $\Pr(E|H_2)$ as

$$\Pr(E|\hat{H}_2) = \Pr(E|\hat{H}_1(s))\Pr(G = s) + (1 - \Pr(G = s))\frac{1}{n}\sum_{i=1}^{n} \Pr(E|\hat{H}_{2,i}). \tag{8}$$

CEESIt reports the approximate likelihood ratio for the suspect, LR(s) = $\Pr(E|\hat{H}_1(s))/\Pr(E|\hat{H}_2)$ and the Monte Carlo approximation to the distribution of $LR(G)$ obtained by $\Pr(E|\hat{H}_{2,i})/\Pr(E|\hat{H}_2)$ for $i = 1 … n$.

The single source samples were created as described in *Sec* 2.1, while a subset of the single source extracts were mixed to contain two or three contributors. The mixture ratios ranged from one part of the minor contributor to 19 parts of the other contributor(s) and DNA total target masses ranged from 0.008 from 0.5 ng. To test the impact of detection, we focus on samples where the minor contributor constituted between one and 20 cells' worth of DNA. A list of the mixtures and the mass of the minor contributors tested in CEESIt are available in Supplemental Table S4.

3. Results

3.1. Simulation as a means of developing highly resolved EPG signal

Synthesizing forensic EPGs with an *in silico* model of the entire forensic pipeline allows the evaluation of each signal component, such as noise and allele, in an efficient manner without the use of costly resources. This is desirable for laboratories interested in determining laboratory conditions that lead to high fidelity EPG signal, or EPG signal with minimal levels of dropout. One way to minimize dropout is to implement a post-PCR laboratory procedure designed to minimize detection error. If allele dropout and noise detection are carefully considered during validation, then unnecessary signal loss can be avoided. To do this effectively, good signal-to-noise resolution when $T_{c=0} = 1$ and $T_{c=0} = 0$ is required.

Previous studies have demonstrated that the noise component is different between colour-channels, loci and template mass, suggesting that locus-specific noise parameters ought to be determined for each kit
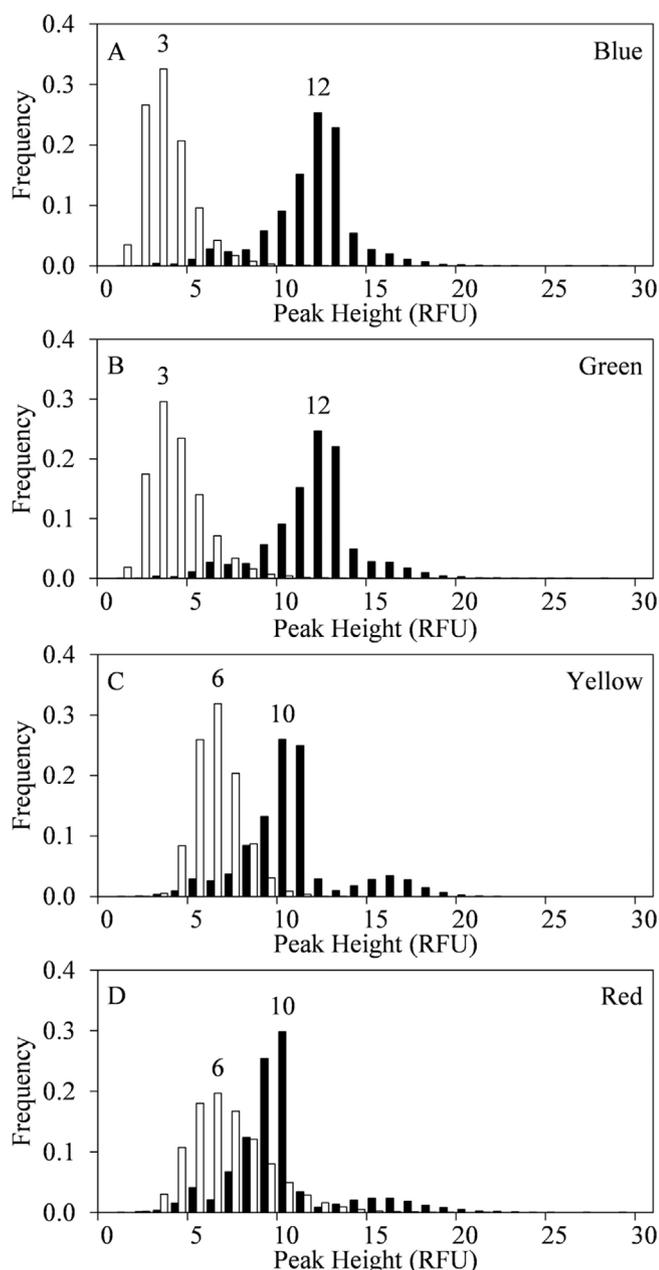
Fig. 1. Frequency of allele (black) and noise (white) simulated signal obtained when one copy of DNA is amplified at 29 cycles and separated using the sensitivities found under *Lab 1* (Table 1) for representative loci from each colour channel: A) D8S1179, B) D3S1358, C) D19S433 and D) D5S818 loci The modes of the peak height of noise and signal are provided. Most of the noise measurement peaks are zero; thus, the noise frequencies have been normalized and exclusively consider the non-zero noise measurement values.

and target mass [29,33]. Many sources of variation within the laboratory process exist and sources that introduce considerable changes in sensitivity would impact a simulation-based approach. Although previous studies have suggested that variation in signal due to the most common laboratory changes, such as capillary and kit lot modifications, have a negligible impact on the final signal compared to amplification effects [34], good practice dictates that the dataset used to parameterize RESOLVIt include data acquired over an extended time period, and chosen to allow ordinary factors that affect the result [35]. Substantive changes to the detector or light source require, at a minimum, performance checks to confirm that signal thresholds are still applicable. Within this work, we define noise as signal which is not easily characterized as either allele or artefact, and so it can encompass non-

reproducible uncharacterized amplification products as well as instrument noise. Fig. 1 displays the histograms of the peak heights for 20,000 simulations when $T_{c=0} = 0$ and 20,000 simulations when $T_{c=0} = 1$ using model parameterizations for *Lab 1* as summarized in Supplemental Tables S1–S3 for representative blue, green, yellow and red fluorescently tagged loci. Consistent with previously published experimental findings [29,33], we observe that the noise shifts right and widens as the color channel transitions from blue to red. In contrast, the peak height distributions for one target copy shifts left as the colour channels transition from blue to red. Notably, the peak height distribution obtained when one copy of DNA is amplified cannot be easily described by a simple distribution class. For example, the allele signal in Fig. 1C and D are of particular interest since we observe a complex distribution of peak heights that range from 1 to 28 RFU. Upon investigation, we find that the right-tails, which are pronounced in the yellow and red loci, are the result of additive noise contributing to the allele signal while the left tails are the result of missed amplifications in the first cycles of PCR. A high degree of overlap is observed between the noise and allele distributions when simulating the manufacturer recommended laboratory procedure (*i.e., Lab 1*) [36].

In an effort to diminish the degree of overlap between noise and signal$_{1copy}$, peak heights for 20,000 simulations when $T_{c=0} = 0$ and 20,000 simulations when $T_{c=0} = 1$ using model parameterizations for *Lab 2*, as summarized in Supplemental Tables S1, S2 and S3, were captured. Histograms for representative blue, green, yellow and red fluorescently tagged loci are displayed in Supplemental Fig. S1. Despite the complex signal spectra observed throughout the simulations, two informative outcomes emerge: 1) good signal to noise resolution can be obtained for all loci using common forensic DNA processes; and 2) a substantial level of signal overlap between noise and allele is caused by early stuttering events and PCR inefficiencies. To acquire a measure of the distance between our noise and single-copy signal distributions, we utilize the Bhattacharyya coefficient (BC) [37], defined as

$$BC = \sum_{i=1}^{b} \sqrt{N_i A_i} \tag{9}$$

where b is the number of RFU partitions, or bins, of size 1 RFU here, considered and N_i and A_i are the frequency of noise and allele, respectively. Non-overlapping distributions result in a BC of 0, while complete overlap results in BC's of 1. The BC results are summarized in Table 1.

DET curves for the four representative loci of simulations using sensitivities listed in Supplemental Table S1 are depicted in Fig. 2. DET plots are graphical representations of the error rates for binary classification systems. For the purposes of allele detection, we plot the false positive rate against the false negative rate at various S_T values in order

Table 1
BC values for each locus for data simulated to mimic one copy of DNA amplified for 29 cycles and injected for 5 s, 10 s or 20 s at 3 kV on the ABI3130 Genetic Analyzer or for 25 s at 1.2 kV on the 3500 Genetic Analyzer. Bin size = 1 RFU.

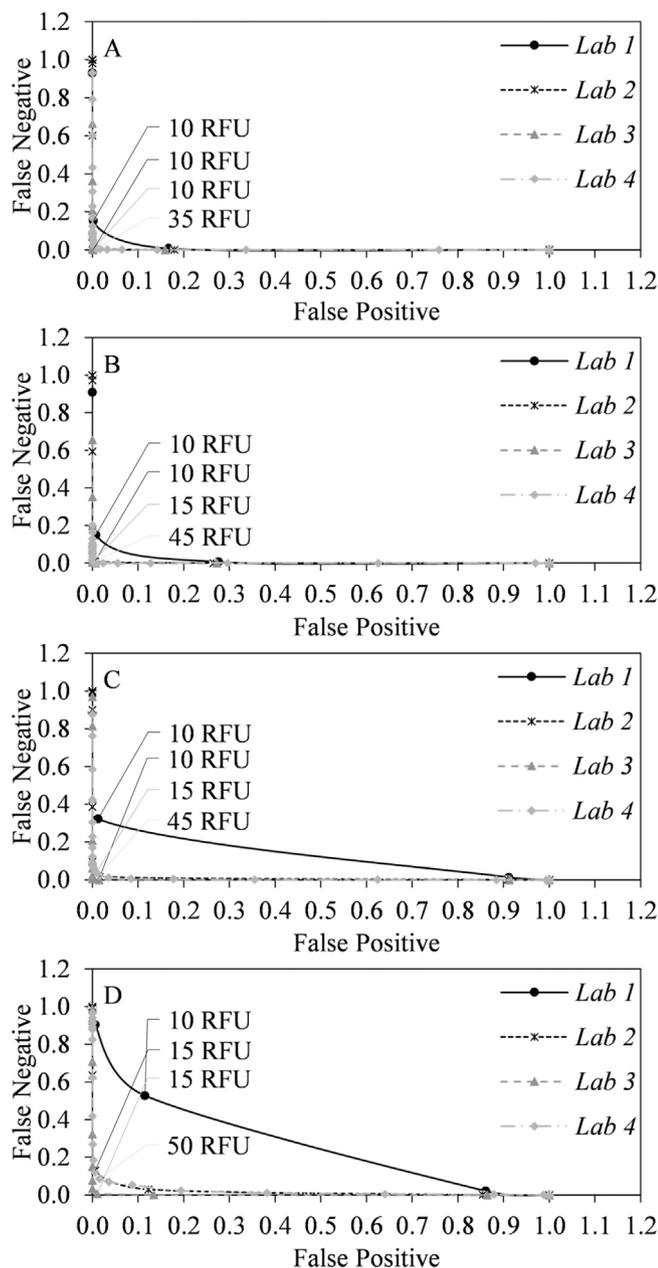| Locus	Dye	Lab 1	Lab 2	Lab 3	Lab 4
D8S1179	B	0.2233	0.0566	0.0039	0.0493
D21S11	B	0.5119	0.1710	0.0528	0.0723
D7S820	B	0.3053	0.0796	0.0135	0.0716
CSF1PO	B	0.4878	0.1662	0.0212	0.0111
D3S1358	G	0.2862	0.0686	0.0057	0.0298
TH01	G	0.2577	0.0690	0.0033	0.0341
D13S317	G	0.3975	0.1241	0.0208	0.0200
D16S539	G	0.2822	0.0671	0.0032	0.0308
D2S1338	G	0.3657	0.0977	0.0167	0.0276
D19S433	Y	0.5253	0.1305	0.0086	0.1282
vWA	Y	0.7020	0.2436	0.0390	0.1380
TPOX	Y	0.8240	0.2951	0.0540	0.0953
D18S51	Y	0.8672	0.3578	0.1041	0.1299
D5S818	R	0.7835	0.2904	0.0539	0.2867
FGA	R	0.8732	0.3965	0.0902	0.4412

Fig. 2. Detection Error Tradeoff (DET) analysis of signal obtained through simulation of one copy of DNA amplified at 29 cycles and separated for each condition summarized in Supplemental Tables S1–S3 for representative loci from each colour channel: A) D8S1179, B) D3S1358, C) D19S433 and D) D5S818 loci. The optimum signal threshold where the total error (sum of false positive and negative rates) is minimized is labeled.

Table 2

Analytical threshold in which error is minimized for data simulated to mimic one cell's worth of DNA amplified for 29 cycles and injected for 5, 10 or 20 s at 3 kV on the ABI 3130 Genetic Analyzer or for 25 s at 1.2 kV on the ABI 3500 Genetic Analyzer. Error is shown in parentheses. Supplemental Tables S5–S8 detail the false negative, false positive, and total error rates for each locus for Labs 1–4 simulated data.

Locus	Dye	*Lab 1*	*Lab 2*	*Lab 3*	*Lab 4*
D8S1179	B	10 (0.1553)	10 (0.0107)	10 (0.0023)	35 (0.0157)
D21S11	B	5 (0.3234)	10 (0.0532)	15 (0.0203)	25 (0.0237)
D7S820	B	5 (0.1023)	5 (0.0395)	10 (0.0032)	20 (0.0208)
CSF1PO	B	5 (0.2701)	10 (0.0662)	15 (0.0135)	15 (0.0025)
D3S1358	G	10 (0.1578)	10 (0.0158)	15 (0.0058)	45 (0.0081)
TH01	G	10 (0.0991)	10 (0.0166)	15 (0.0016)	45 (0.0095)
D13S317	G	10 (0.1728)	10 (0.0461)	15 (0.0068)	30 (0.0043)
D16S539	G	10 (0.2088)	10 (0.0145)	10 (0.0029)	35 (0.0076)
D2S1338	G	10 (0.2520)	10 (0.0258)	15 (0.0097)	30 (0.0047)
D19S433	Y	10 (0.3343)	10 (0.0335)	15 (0.0045)	45 (0.0395)
vWA	Y	10 (0.3861)	15 (0.0979)	15 (0.0093)	45 (0.0438)
TPOX	Y	10 (0.8980)	10 (0.1354)	15 (0.0126)	35 (0.0203)
D18S51	Y	5 (0.9061)	10 (0.1512)	15 (0.0235)	30 (0.0371)
D5S818	R	10 (0.6430)	15 (0.1368)	15 (0.0174)	50 (0.1019)
FGA	R	5 (0.9173)	10 (0.1856)	15 (0.0255)	45 (0.1652)

large signal thresholds.

An AT of 30 RFU has recently been suggested as a possible signal threshold with which to analyze EPGs [10], but these data suggest that an AT of 30 RFU coupled with these laboratory conditions (*i.e.*, 3130 Genetic analyzer; 10 s/3 kV injection; 29 cycles; Identifiler® Plus kit) would result in significant levels of dropout due to detection effects. Thus, these true signals would be excluded from interpretation resulting in lower information contents. Given that a number of probabilistic genotyping systems have been designed to interpret noise [6,23], the traditional practice of utilizing a large signal thresholds in an effort to reduce the risk of noise detection requires re-evaluation. For *Lab 2* simulations, the false detection rate was as high as 0.1201 for the FGA locus and the BC ranged from 0.0566 to 0.3965. This large false detection error may be considered too high for certain downstream interpretation systems. As such, obtaining fuller signal to noise resolution by injecting the sample for twice as long, or increasing the cycle number to 30 will likely result in an overall decrease in detection errors.

We repeated the simulations, therefore, using the sensitivities listed under column *Lab 3* in Supplemental Table S1, emulating 29-cycle Identifiler® Plus amplification with a 20 s/3 kV injection on an ABI 3130 platform. Fig. 3 shows histograms of the signal obtained for four representative blue, green, yellow and red loci. Supplemental Table S7 summarizes the false positive and negative rates for all loci across all S_T values. As shown in Table 1, the BC value decreased from a maximum of 0.3965 for *Lab 2* to a maximum of 0.1041 for *Lab 3* simulations, indicating a higher degree of separation between noise and allele signal. Error rates ranging from 0.0016 to 0.0255 indicated the noise detection and allele dropout is greatly diminished by increasing injection time from 5 s to 10 s to 20 s. Fig. 4 plots log of total error, or sum of false negatives and false positives, against the log BC for each locus in all four simulations. *Lab 1* results in relatively high BC values and relatively high error rates. A correlation between BC and total error rate is observed as error increases as the degree of overlap between the noise and allele distributions increases (Fig. 4).

3.2. Improved 3500 genetic analyzer sensitivity does not necessarily improve signal resolution

Fragment analysis of STRs is accomplished using CE and there has recently been a transition from the 3130 to the 3500 Genetic Analyzer within forensic operational environments. Prior to implementation of a new platform, the instrument must be validated for forensic purposes [38]. At this stage, the dynamic range, sensitivities and other pertinent

to evaluate the separation of signal from noise and one DNA molecule. If perfect noise and signal resolution is achieved, such that BC = 0, then there exists an S_T that will yield false negative and false positive rates of zero. If small false detection rates are obtained at a given S_T, then this value ought to be considered as the operational AT of the laboratory since it affords strong protection against noise detection while minimizing allele dropout. Table 2 summarizes the AT that results in the lowest overall error rate for each locus, with the total error (sum of false positive and false negative rates) in parentheses. Supplemental Table S6 summarizes the false positive and negative rates obtained at various S_T values for *Lab 2* simulated data for all loci. As predicted from visual inspection of the histograms the red and yellow channels result in the largest error rates. Interestingly, we observe that the commonly employed AT of 50 RFU results in total error rates that are substantive; these are driven by the high allele dropout rates associated with such
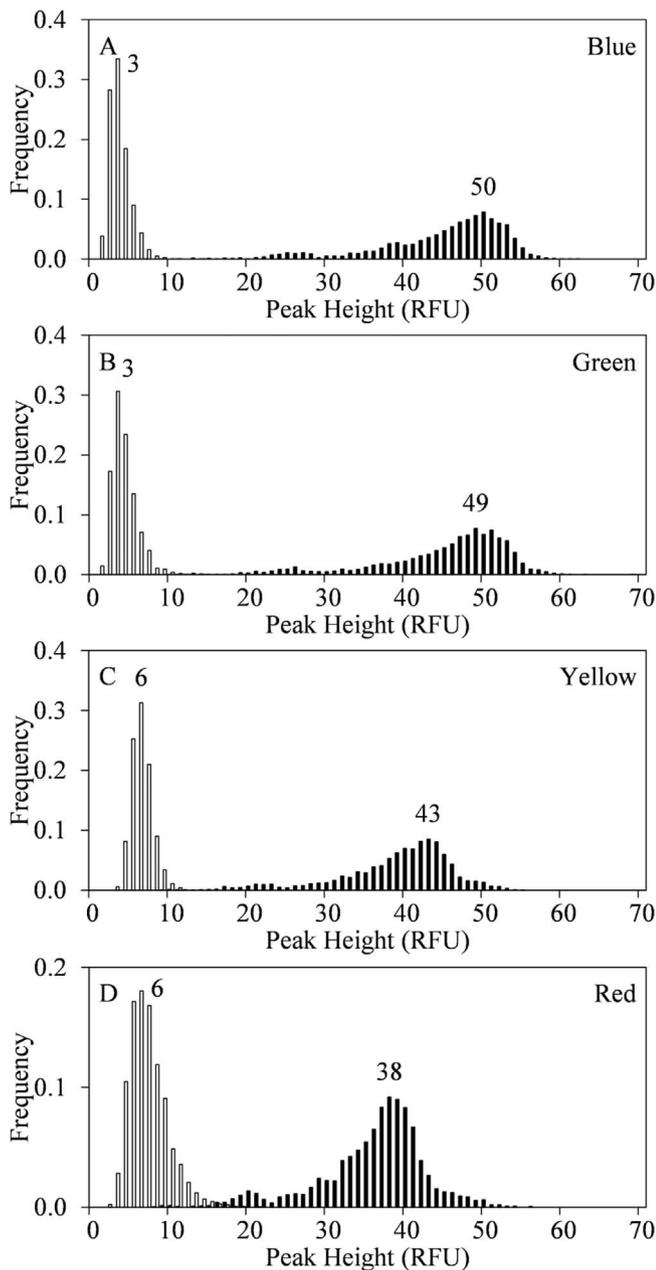
Fig. 3. Frequency of allele (black) and noise (white) simulated signal obtained when one copy of DNA is amplified at 29 cycles and separated using the sensitivities found under *Lab 3* (Supplemental Table S1) for representative loci from each colour channel: A) D8S1179, B) D3S1358, C) D19S433 and D) D5S818 loci. The noise frequencies have been normalized and exclusively consider the non-zero noise measurement values.
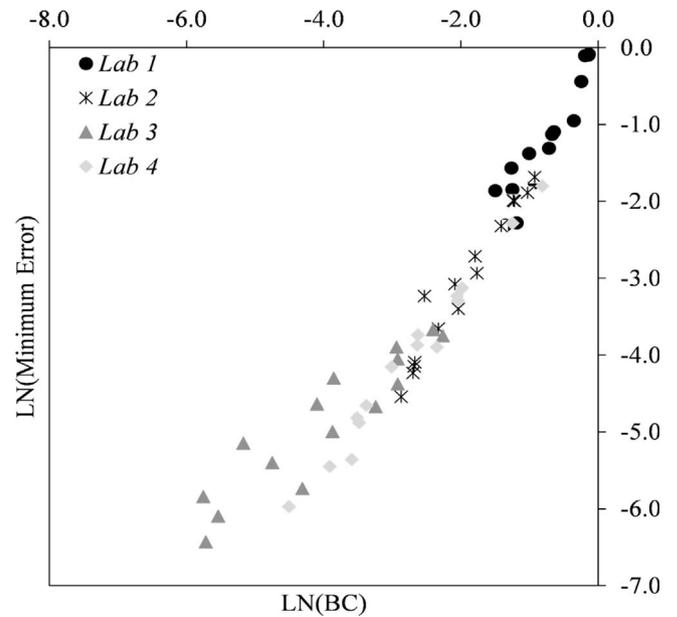

Fig. 4. The lognormal (base 10) BC value against the lognormal (base 10) total error (sum of false positive and false negatives rates) based upon signal simulated for each CE condition listed in Table 1.

majority of loci. Specifically, the largest total error rate (Table 2) for the 3500, *Lab 4*, simulations is 0.1652 (at $S_T = 45$). The S_T that resulted in the lowest total error rate ranged from 25 to 50 RFU, depending on the locus. An S_T of 55, however, was required if a false positive rate of ≤ 0.01 was desired (Supplemental Table S8). Improved resolution and lower total error rates are achievable with the 3500 platform *via* small laboratory modifications, such as increased injection voltages or, alternatively, the addition of one cycle to the amplification may be appropriate.

3.3. Simulated versus experimental data

In cases where signal thresholds and noise drop-in rates for casework are derived using simulation, demonstrable consistency between simulated and experimental data is a necessity. We provide in Supplemental Tables S9–S12 and Fig. 6 the DET results from experimental data generated using the 3130 and the 3500 platforms. As with the simulated data, the lowest error rate for the 3130 platform occurred when a 20 s injection was coupled with locus specific S_T's of 5 RFU to 20 RFU (Supplemental Table S11). The lowest signal error rates were obtained at S_T's between 15 and 50 RFU for the 3500 platforms (Supplemental Table S12). These signal thresholds are consistent with those derived through simulation. One notable difference, however, is the discrepancy between the total error rates between simulated and experimental data. In short, total overall error rates, regardless of S_T, for the experimental data are significantly larger than those obtained during simulation, as expected, as stochastic sampling events are known to occur during pre-PCR processing [11,26]. Unlike simulation, where we were able to set the copy number to $T_{c=0} = 1$, experimental data is produced by taking an aliquot of liquid ($V_{aliquot}$) from a total extract volume (V_{tot}) containing a total, T, of DNA molecules of an allele. Assuming that the sample is well mixed, a copy of each individual allele is randomly selected independently with probability $V_{aliquot}/V_{tot}$. This results in the total initial copies of the allele, $T_{c=0}$, that shall undergo amplification being captured by a binomial random variable

$$T_{c=0} \sim Binomial\left(T, \frac{V_{aliquot}}{V_{tot}}\right).$$

(10)

In this study, $\frac{V_{aliquot}}{V_{tot}}$ was ca. $\frac{7.8\mu L}{100\mu L}$, or 0.078 and the total number of DNA

analytical metrics, including the minimum signal thresholds are evaluated [39]. Validation studies of the new capillary platform have shown that, in general, it is more sensitive than its 3130 counterpart [15,39]. To evaluate if this increase in sensitivity impacts the signal-to-noise resolution, we simulated signal from noise and single DNA molecules using the parameters, summarized in column *Lab 4* of Table 1, which were inferred from the noise and signal obtained from samples injected for 25 s and 1.2 kV on this laboratory's 3500 platform. Fig. 5 demonstrates the signal obtained when $T_{c=0} = 0$ and $T_{c=0} = 1$. Like the 3130 data, the 3500 noise becomes larger and more varied as the colour channels transition from blue to red. Further, we again observe that single copy DNA signal has both a right- and left- tail. Upon evaluation, the corresponding BC's (Table 1) signify that *Lab 4* conditions result in marginal improvements to resolution over *Lab 2* conditions for the
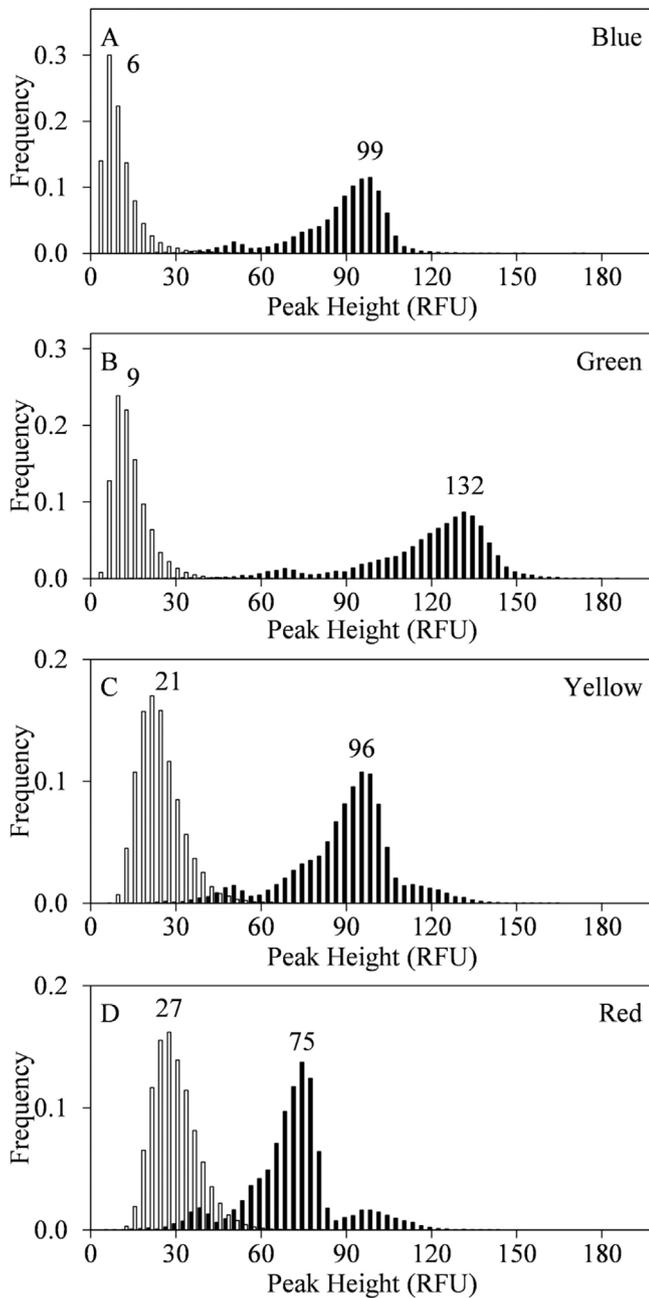
Fig. 5. Frequency of allele (black) and noise (white) simulated signal obtained when one copy of DNA is amplified at 29 cycles and separated using the sensitivities found under *Lab 4* (Supplemental Table S1) for representative loci from each colour channel: A) D8S1179, B) D3S1358, C) D19S433 and D) D5S818 loci. The modes of peak height of noise and signal are provided. The noise frequencies have been normalized and exclusively consider the non-zero noise measurement values.

molecules, T, was $\left\lceil \dfrac{0.001\frac{ng}{\mu L}}{0.0063\frac{ng}{cell}} \cdot 99\mu L \right\rceil$, or 16 DNA molecules. Thus, the probability of obtaining $T_{c=0}=0$ is

$$\Pr(\text{Binomial}(16, 0.078) = 0) = 0.27 \tag{11}$$

This represents the expected probability of dropout due to sampling effects and is in line with the proportion of times alleles were not detected when $S_T = 10$ RFU or 45 RFU for the 3130 and 3500 platforms, respectively.
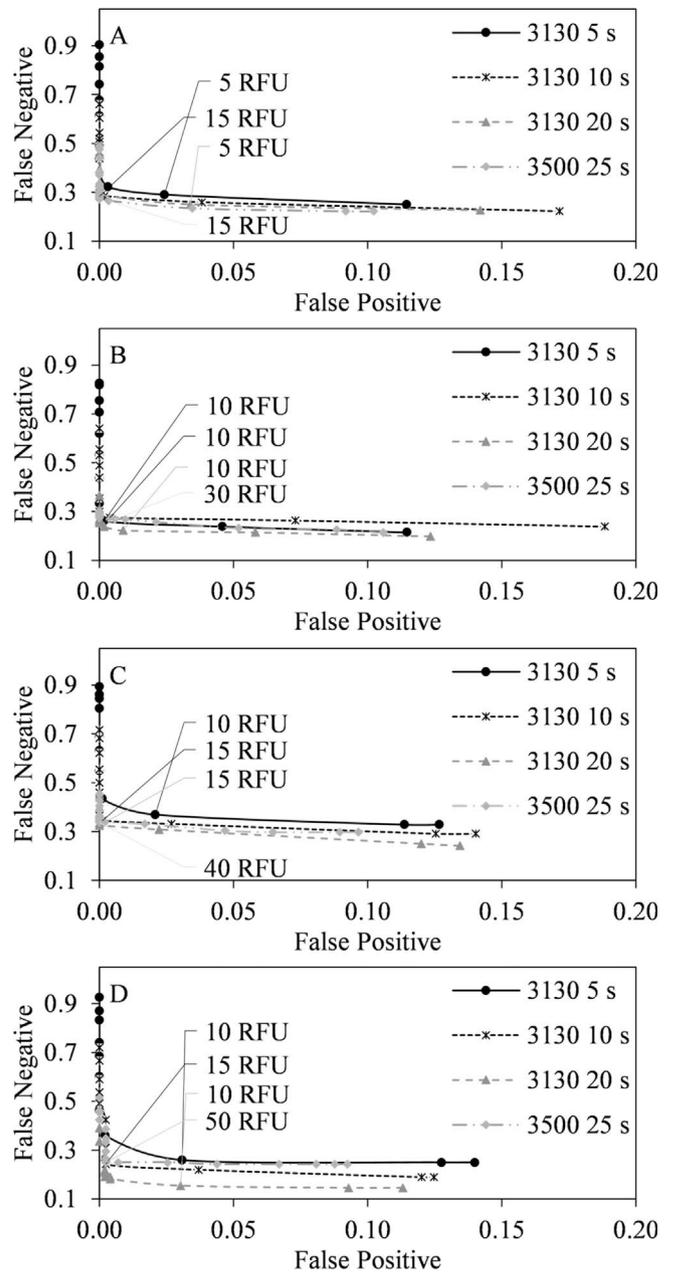

Fig. 6. Detection Error Tradeoff (DET) analysis of signal obtained when 95 samples of known genotype with a target mass of 0.0078 ng were amplified at 29 cycles and separated using a 5, 10, or 20 s 3 kV injection on an ABI 3130 Genetic Analyzer and a 25 s 1.2 kV on an ABI 3500 Genetic Analyzer for representative loci from each colour channel: A) D8S1179, B) D3S1358, C) D19S433 and D) D5S818 loci. The optimum signal threshold where the total error (sum of false positive and negative rates) is minimized is labeled.

3.4. Impact of information content on low-template probabilistic interpretation

We evaluated the impact of information content on interpretation by exploring the effects of signal thresholds and injection time on the LR of a contributor and on the proportion of LRs greater than 1 for the randomly sampled non-contributors, which we denote

$$\Pr(LR(\hat{G}) > 1) = \Pr(G = s)\chi(\Pr(E|\hat{H}_1)(s) > \Pr(E|\hat{H}_2))$$
$$+ \frac{1}{n}(1 - \Pr(G = s))\sum_{i=1}^{n}\chi(\Pr(E|\hat{H}_{2,i}) > \Pr(E|\hat{H}_2)) \tag{12}$$

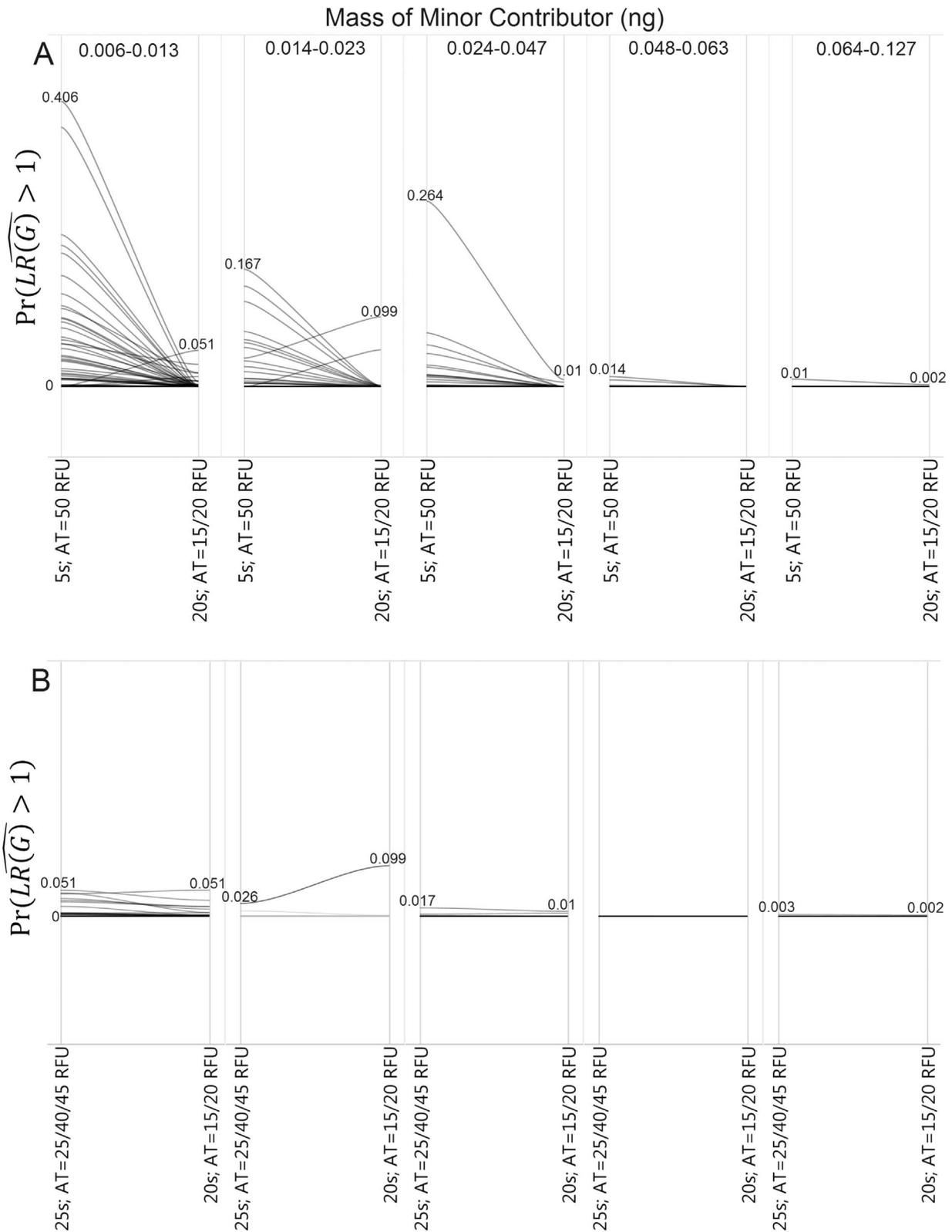where χ denotes the indicator function, which takes the value 1 if the

Fig. 7. Parallel plots of $\Pr(\hat{LR(G)}>1)$ for experimental 1-, 2- and 3-person samples a) injected for 5 s on the 3130 platform coupled with an AT of 50 RFU and injected for 20 s on the 3130 platform coupled with an optimized AT of 15 and 20 for the blue/green and yellow/red channels, respectively and b) injected for 20 s on the 3130 platform coupled with ATs of 15 (blue/green) and 20 (yellow/red) versus the same samples injected for 25 s on the 3500 platform coupled with ATs 25 (blue), 40 (green/yellow) and 45 (red). All samples were amplified with the Identifiler* Plus set of loci using 29 PCR cycles.

condition in braces is true and zero otherwise. If the information content is substantial, and of high quality, then one assumes that the LR for a known contributor will be large and the probability that the likelihood ratio for a non-contributor is greater than one will be small. To evaluate the impact of signal content on these summary statistics we plot $\Pr(\hat{LR(G)}>1)$ when using an AT of 50 RFU and data generated from a 5 s injection of 1-, 2- and 3-person samples on a 3130 Genetic Analyzer against $\Pr(\hat{LR(G)}>1)$ when the same samples were analyzed
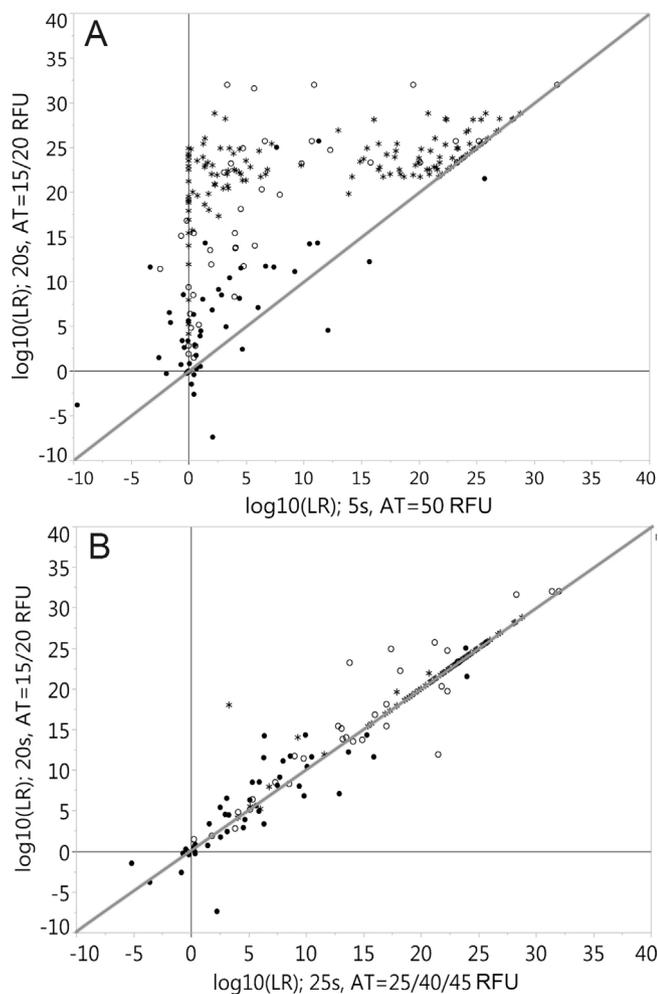
Fig. 8. Scatter plot of LRs for the true minor contributor of experimental (*) 1-, (○) 2- and (●) 3-person samples A) injected for 5 s on the 3130 platform coupled with an AT of 50 RFU versus those injected for 20 s on the 3130 platform coupled with an optimized AT of 15 and 20 for the blue/green and yellow/red channels, respectively and B) samples injected for 25 s on the 3500 platform with ATs of 25/40/45 versus the optimized 3130 parameters. All samples were amplified with the Identifiler® Plus set of loci using 29 PCR cycles. The x = y line is also shown.

instrument. Figs. 7 b and 8 b demonstrate that the summary statistics are similar between platforms, indicating that laboratory parameters that are chosen based upon an evaluation of signal error detection rates result in consistent interpretation across platforms.

4. Discussion

For any analytical bio-assay, the ultimate performance benchmarks are accuracy and reliability within the established dynamic range. One way to assess the limits of DNA testing is by evaluating the probability of dropout. Allele dropout rates are influenced by four factors: 1) the absence of the DNA target during amplification; 2) too few DNA fragment molecules produced during amplification to surpass signal detection levels; 3) CE settings that result in insufficient sensitivities; and 4) high detection thresholds. To alleviate instances of dropout occurring from the first factor, improved front end processing techniques are required. In contrast, allele dropout due to detection effects may be controlled by applying optimized post-PCR processes such that sufficient numbers of amplicons are produced and injected. If signal from a single copy of DNA is sufficiently large as to surpass noise levels, then optimized signal detection thresholds may be applied in a manner that will mitigate both the false detection of noise and allele dropout. In addition, processes that favor examination of the entire signal may be deemed appropriate.

In this study, the detection of approximately one copy of DNA was examined. Previous studies have demonstrated that baseline is affected by the target quantity of DNA [29,33]. As a consequence, the ability to resolve signal from a single copy of DNA from noise in the presence of a major contributor would also be of value. The validation scheme presented herein is a generalized method and may easily be applied to test such scenarios. As an example, Supplemental Table S13 details the false negative, false positive, and total error rates for simulated noise data that correspond to the amplification of 0.25 ng of DNA coupled with the allele signal corresponding to one copy of DNA. The noise simulated assuming a larger (i.e., 0.25 ng) target mass corresponds to an increase in total error rates and corresponds well with experimental data. Supplemental Table S14 depicts the false negative, false positive, and total error rates for noise garnered from samples amplified at a target mass of 0.25 ng and allele signal garnered from samples amplified at a target mass of 0.008 ng.

To apply signal detection thresholds effectively a cost-effective method, based on simulation, that allows the laboratory to evaluate many scenarios without incorporating significant enhancement procedures must be established. The data presented herein suggests that this validation scheme would consist of simulating data under numerous laboratory scenarios when DNA is absent ($T_{c=0} = 0$) and is present at the single copy ($T_{c=0} = 1$) level. The simulated data are evaluated to confirm that good signal-to-noise resolution is obtained. Once optimal laboratory conditions are chosen, DET analysis can be utilized as a means to determine the S_T that results in minimal total detection errors. If necessary, the detection threshold, or AT, can be conditioned on the noise detection rate. This approach is an effective means to establish good post-PCR laboratory practices, regardless of platform. It is also an effective way to compare injection conditions between platforms.

If optimized laboratory conditions are implemented, the limit of detection for the forensic DNA process is one target molecule of DNA and the dropout and noise detection rates can be well characterized for downstream probabilistic interpretation. We show that decreasing the LOD of the forensic process to one copy is a powerful mechanism by which to increase the information content introduced into probabilistic systems in a consistent manner. This work introduces a validation scheme that can be used to simultaneously improve signal resolution and detection rates for both true and noise signal without costly large-scale studies by applying a simulation and experimental based approach to validation.

using an AT of 15 RFU (Blue and Green channels) or 20 RFU (Yellow and Red channels) and a 20 s injection (Fig. 7a). We note that the 5 s injection time is the manufacturer's recommendation [36]. When coupled with an AT of 50 RFU the detection error rates for the recommended laboratory settings were one due to high levels of drop-out when in the single-copy regime. In contrast, the 20 s injection time coupled with an AT of 15/20 RFU corresponded to detection error rates that did not exceed 0.07 (Supplemental Tables S5 and S7). As anticipated, maximizing information content by utilizing EPGs generated with laboratory conditions that result in good allele-to-noise resolution and reasonable signal thresholds significantly reduces the probability that a non-contributor results in a LR > 1. Fig. 8a also demonstrates that less EPG data utilized during interpretation results in low LRs for the true minor contributors.

Fig. 7b plots $Pr(\hat{LR}(G)>1)$ when the same amplified work products were injected into the 3130 and 3500 Genetic Analyzers for 20 and 25 s. The LRs obtained for the known first minor contributor of the mixtures when H_1 is true are plotted in Fig. 8b. The ATs were set to 15/20 for the 3130 instrument and to 25(Blue)/40(Green, Yellow)/45(Red) for the 3500 instrument. These ATs were chosen based on the DET analysis described above and result in similar error rates for both platforms. Further, the injection parameters were chosen because they afforded similar, and good, signal-to-noise resolution for each

5. Conclusion

We modified a previously developed full mechanistic model of the forensic laboratory process for purposes of forensic validation. Production of a full DNA pipeline with the objective of recovering all signal results in positive interpretation effects. Specifically, we show that if the forensic pipeline is engineered such that high-fidelity EPG signal is obtained then the LR for a true contributor is, in general, larger and the probability that the LR determined for a non-contributor is greater than one is reduced, and therefore positively affected. We also show this validation scheme provides a means to ensure conformity in interpretation, regardless of platform. One of the primary features of the system proposed in the present paper is its flexibility to test multifarious scenarios in a matter of minutes. As a result, any concern about imprecision in an input model parameters, such as those defining the frequency of occurrence of noise peaks, can be readily alleviated by performing sensitivity analysis for a range of values describing that uncertainty. Thus this, potentially, is the first step towards standardization of the post-PCR validation process across laboratories.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fsigen.2017.09.005.

References

[1] M.W. Perlin, M.M. Legler, C.E. Spencer, J.L. Smith, W.P. Allan, J.L. Belrose, Validating TrueAllele® DNA mixture interpretation, J. Forensic Sci. 56 (2011).

[2] K. Inman, N. Rudin, K. Cheng, C. Robinson, A. Kirschner, L. Inman-Semerau, et al., Lab Retriever: a software tool for calculating likelihood ratios incorporating a probability of drop-out for forensic DNA profiles, BMC Bioinf. 16 (2015) 298.

[3] C.D. Steele, M. Greenhalgh, D.J. Balding, Evaluation of low-template DNA profiles using peak heights, Stat. Appl. Genet. Mol. Biol. 15 (2016) 431–445.

[4] R. Puch-Solis, L. Rodgers, A. Mazumbder, S. Pope, I. Evett, J. Curran, et al., Evaluating forensic DNA profiles using peak heights, allowing for multiple donors, allelic dropout and stutters, Forensic Sci. Int. Genet. 7 (2013) 555–563.

[5] J.-A. Bright, D. Taylor, C. McGovern, S. Cooper, L. Russell, D. Abarno, et al., Developmental validation of STRmix™, expert software for the interpretation of forensic DNA profiles, Forensic Sci. Int. Genet. 23 (2016) 226–239.

[6] H. Swaminathan, A. Garg, C.M. Grgicak, M. Medard, D.S. Lun, CEESIt: a computational tool for the interpretation of STR mixtures, Forensic Sci. Int. Genet. 22 (2016) 149–160.

[7] Ø. Bleka, G. Storvik, P. Gill, EuroForMix: an open source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts, Forensic Sci. Int. Genet. 21 (2016) 35–44.

[8] C.D. Marsden, N. Rudin, K. Inman, K.E. Lohmueller, An assessment of the information content of likelihood ratios derived from complex mixtures, Forensic Sci. Int. Genet. 22 (2016) 64–72.

[9] D. Taylor, J. Buckleton, Do low template DNA profiles have useful quantitative data, Forensic Sci. Int. Genet. 16 (2015) 13–16.

[10] D. Taylor, J. Buckleton, J.-A. Bright, Does the use of probabilistic genotyping change the way we should view sub-threshold data, Aust. J. Forensic Sci. 49 (2015) 78–92.

[11] R. Hedell, C. Dufva, R. Ansell, P. Mostad, J. Hedman, Enhanced low-template DNA analysis conditions and investigation of allele dropout patterns, Forensic Sci. Int. Genet. 14 (2015) 61–75.

[12] T. Caragine, R. Mikulasovich, J. Tamariz, E. Bajda, J. Sebestyen, H. Baum, et al., Validation of testing and interpretation protocols for low template DNA samples using AmpFlSTR identifier, Croat. Med. J. 50 (2009) 250–267.

[13] D.Y. Wang, C.-W. Chang, R.E. Lagacé, L.M. Calandro, L.K. Hennessy, Developmental validation of the AmpFℓSTR® Identifiler® Plus PCR Amplification Kit: an established multiplex assay with improved performance, J. Forensic Sci. 57 (2012) 453–465.

[14] P.J. Smith, J. Ballantyne, Simplified low-copy-number DNA analysis by post-PCR purification, J. Forensic Sci. 52 (2007) 820–829.

[15] J.-A. Bright, S. Neville, J.M. Curran, J.S. Buckleton, Variability of mixed DNA profiles separated on a 3130 and 3500 capillary electrophoresis instrument, Aust. J. Forensic Sci. 46 (2014) 304–312.

[16] IUPAC, Compendium of Chemical Terminology, 2.3.3 ed., Blackwell Scientific Publications, Oxford, 1997.

[17] D.A. Skoog, F.J. Holler, T.A. Nieman, Principles of Instrumental Analysis, 5th ed., Harcourt Brace & Company, Orlando, 1998.

[18] J. Miller, J. Miller, Statistics for Analytical Chemistry, 3rd ed., Ellis Horwood Limited, Chichester, 1993.

[19] L.A. Currie, Detection and quantification limits: origins and historical overview, Anal. Chim. Acta 391 (1999) 127–134.

[20] J.D. Winefordner, G.L. Long, Limits of detection: a closer look at the IUPAC definition, Anal. Chem. 55 (1983) 712A–724A.

[21] C.A. Rakay, J. Bregu, C.M. Grgicak, Maximizing allele detection: effects of analytical threshold and DNA levels on rates of allele and locus drop-out, Forensic Sci. Int. Genet. 6 (2012) 723–728.

[22] D. Taylor, D. Powers, Teaching artificial intelligence to read electropherograms, Forensic Sci. Int. Genet. 25 (2016) 10–18.

[23] M.W. Perlin, A. Sinelnikov, An information gap in DNA evidence interpretation, PLoS One 4 (2009).

[24] J. Inczedy, T. Lengyel, A.M. Ure, Compendium of Analytical Nomenclature, 3rd ed., Blackwell Science, Ltd., Oxford, UK, 1997.

[25] J. Weusten, J. Herbergs, A stochastic model of the processes in PCR based amplification of STR DNA in forensic applications, Forensic Sci. Int. Genet. 6 (2012) 17–25.

[26] K.E. Lohmueller, N. Rudin, K. Inman, Analysis of allelic drop-out using the Identifiler® and PowerPlex® 16 forensic STR typing systems, Forensic Sci. Int. Genet. 12C (2014) 1–11.

[27] P. Gill, J. Curran, K. Elliot, A graphical simulation model of the entire DNA process associated with the analysis of short tandem repeat loci, Nucleic Acids Res. 33 (2005).

[28] I. Findlay, A. Taylor, P. Quirke, R. Frazier, A. Urquhart, DNA fingerprinting from single cells, Nature 389 (1997) 555–556.

[29] J. Bregu, D. Conklin, E. Coronado, M. Terrill, R.W. Cotton, C.M. Grgicak, Analytical thresholds and sensitivity: establishing RFU thresholds for forensic DNA analysis, J. Forensic Sci. 58 (2012) 120–129.

[30] O. Hansson, P. Gill, Characterisation of artefacts and drop-in events using STR-validator and single-cell analysis, Forensic Sci. Int. Genet. 30 (2017) 57–65.

[31] K.R. Duffy, N. Gurram, K.C. Peters, G. Wellner, C.M. Grgicak, Exploring STR signal in the single- and multicopy number regimes: deductions from an in silico model of the entire DNA laboratory process, Electrophoresis 38 (2017) 855–868.

[32] C.M. Grgicak, Z.M. Urban, R.W. Cotton, Investigation of reproducibility and error associated with qPCR methods using Quantifiler® Duo DNA Quantification Kit, J. Forensic Sci. 55 (2010) 1331–1339.

[33] U.J. Mönich, K. Duffy, M. Médard, V. Cadambe, L.E. Alfonse, C. Grgicak, Probabilistic characterisation of baseline noise in STR profiles, Forensic Sci. Int. Genet. 19 (2015) 107–122.

[34] K.E. Rowan, G.A. Wellner, C.M. Grgicak, Exploring the impacts of ordinary laboratory alterations during forensic DNA processing on peak height variation, thresholds, and probability of dropout, J. Forensic Sci. 61 (2016) 177–185.

[35] EURACHEM/CITAC Guide CT4, Quantifying Uncertainty in Analytical Measurement, 2nd ed., (2000).

[36] Life Technologies Inc, AmpFlSTR® Identifiler® Plus PCR Amplification Kit User Guide, Life Technologies, Foster City, CA, 2014.

[37] S.H. Cha, Comprehensive survey on distance/similarity measures between probability density functions, Int. J. Math. Models Appl. Methods Appl. Sci. 1 (2007) 200–307.

[38] SWGDAM, Validation Guidelines for DNA Analysis Methods, (2012).

[39] E.L.R. Butts, M.C. Kline, D.L. Duewer, C.R. Hill, J.M. Butler, P.M. Vallone, NIST validation studies on the 3500 genetic analyzer, Forensic Sci. Int. Genet. Suppl. Ser. 3 (2011) e184–e185.