Research paper

Probabilistic characterisation of baseline noise in STR profiles

Ullrich J. Mönich [a,*], Ken Duffy [d], Muriel Médard [a], Viveck Cadambe [c], Lauren E. Alfonse [b], Catherine Grgicak [b]

[a] Research Laboratory of Electronics, Massachusetts Institute of Technology, United States
[b] Biomedical Forensic Sciences, Boston University School of Medicine, United States
[c] Department of Electrical Engineering, Pennsylvania State University, United States
[d] Hamilton Institute, Maynooth University – National University of Ireland Maynooth, Ireland

ABSTRACT

There are three dominant contributing factors that distort short tandem repeat profile measurements, two of which, stutter and variations in the allelic peak heights, have been described extensively. Here we characterise the remaining component, baseline noise. A probabilistic characterisation of the non-allelic noise peaks is not only inherently useful for statistical inference but is also significant for establishing a detection threshold. We do this by analysing the data from 643 single person profiles for the Identifiler Plus kit and 303 for the PowerPlex 16 HS kit. This investigation reveals that although the dye colour is a significant factor, it is not sufficient to have a per-dye colour description of the noise. Furthermore, we show that at a per-locus basis, out of the Gaussian, log-normal, and gamma distribution classes, baseline noise is best described by log-normal distributions and provide a methodology for setting an analytical threshold based on that deduction. In the PowerPlex 16 HS kit, we observe evidence of significant stutter at two repeat units shorter than the allelic peak, which has implications for the definition of baseline noise and signal interpretation. In general, the DNA input mass has an influence on the noise distribution. Thus, it is advisable to study noise and, consequently, to infer quantities like the analytical threshold from data with a DNA input mass comparable to the DNA input mass of the samples to be analysed.

© 2015 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Short tandem repeat (STR) allele signal interpretation is a central tool in forensic analysis, as the number of repeated copies of a basic motif at given loci can be used to uniquely identify an individual. Currently, forensically relevant STR loci are processed via capillary electrophoresis. The relative intensity of the fluorescent signal is an indication of the number of amplicons in the sample. The transit time of the amplicons through the capillary is utilised to determine the size of the fragment, which is translated into the number of repeat units, i.e. allele. A DNA STR profile will likely contain the following elements: the desired DNA signal, disturbance due to the instrument, non-specific amplification, or other reasons, often termed noise, and artefacts, such as stutter, -A, and pull-up [1]. The interpretation of DNA profiles, especially from

low DNA input mass samples, is complicated, because of the distortion of the DNA signal by the latter elements.

The effects of noise in STR profiles are usually suppressed by applying an analytical threshold to the data [2–6]. All peaks smaller than the analytical threshold, which is also referred to as detection threshold, or minimum distinguishable signal threshold, are considered indistinguishable from noise and thus are ignored. Further, a second threshold, the stochastic threshold, has been used in casework. The stochastic threshold is defined as the value above which it is reasonable to assume that allelic dropout has not occurred within a single-source sample [7].

Recently, it was shown that the inclusion of peak height information results in some improvement over interpretation methods that use a drop-out probability for low-template samples [8]. Interpretation methods that utilise peak height or area information, rather than the binary information of whether a peak height or area exceeds a threshold or not, have garnered much attention, and a variety of fully-continuous interpretation platforms have been developed [9–16].

The peak area information is used in [9–12] to establish a quantitative method for the interpretation of STR mixtures. In

these methods the peak areas are usually modelled as Gaussian random variables and baseline noise is not handled explicitly. Sometimes, noise is modelled implicitly as small peaks with mean zero [10,12]. In [13], a continuous modelling of the peak height is used if the peak exceeds a user defined analytical threshold. A quantitative computer evaluation of DNA mixtures is proposed in [15,16]. In the latter work, peak heights are modelled as truncated Gaussian random variables and baseline noise is implicitly treated as small peaks with mean zero. Although these continuous methods use peak height information, an analytical threshold is typically still applied.

Taking the idea of the continuous methods further, it is consequent to remove the analytical threshold as well, and to consider the information in the peaks that were previously below the analytical threshold. The potential of this approach is illustrated in [17], where the authors present an algorithm to determine the number of contributors to a DNA mixture. They compare the maximum allele count method to their method, which uses no threshold, and show that the latter results in increased accuracy for samples with low DNA input mass. However, little is known about the properties of these peaks. In order to derive a meaningful signal model, it appears necessary to study the distribution of the baseline noise.

Even in settings where an analytical threshold is applied, the information of the noise peak height distribution is of relevance because it can be used to specify the analytical threshold. The fact that the conservative nature of analytical thresholds has been criticised led [18] to analyse the baseline noise. Information loss may be kept low by applying an optimal analytical threshold that suppresses most of the noise while enhancing the detection of alleles [6]. However, a common validation platform or best practise have yet to be established for this purpose [19]. In order to establish a best practise, a detailed understanding of the behaviour of noise appears necessary.

Noise in STR profiles is commonly modelled via Gaussian distributions [18,17], though log-normal distributions have also been suggested [6]. In [17], a Gaussian noise model is used for determining the likelihood that a given number of individuals contributed to a mixed sample. Although utilisation of the Gaussian model to describe noise provides improved identification over previous techniques, the authors do not provide analysis independent of determining the most likely number of contributors to confirm its appropriateness. In this paper, we revisit this premise.

Using data sets created at a range of DNA input masses using both the Identifiler Plus kit and the PowerPlex 16 HS kit, we first ask in Section 3.2 if there is a mass dependent per-dye colour representation of the noise. This hypothesis is rejected, particularly at high input masses. That is, there is no single statistical distribution that can jointly describe the noise observed at distinct loci measured in the same dye colour. As a result, it is necessary to characterise noise on a mass dependent per-locus basis. In Section 3.4, using statistical analysis, we assess three classes of distributions, Gaussians, log-normals and gamma distributions, for their ability to describe measured noise data.

2. Methods

2.1. Data generation and representation

There are different ways to represent the information contained in an electropherogram. In an idealised signal, i.e., in a signal with no artefacts or noise, each peak corresponds to an allele that is present in the DNA sample. Hence, it is possible to specify each peak location by a tuple consisting of locus and allele name.

We choose a vector representation of the data, similarly to [20]. To obtain the vector representation from the electropherogram

peak list, we index all possible allele and stutter positions with an integer from 1 to I, where I denotes the total number of positions. Then we collect the signal peak heights from all the possible positions in a vector $\underline{y} = (y_1, y_2, \ldots, y_I)^{\mathrm{T}}$ and call this vector the measurement vector. If there is no peak detected at a particular position, a 0 is included in the vector $\underline{y}$ at the corresponding index.

The data that are used for the analysis in this paper consist of 643 distinct single person profiles from 60 individuals for the Identifiler Plus kit and 303 distinct single person profiles from 48 individuals for the PowerPlex 16 HS kit. For the IDplus kit, the 643 profiles were generated by targeting an input mass of 0.0078 ng (97 profiles), 0.0156 ng (94 profiles), 0.0313 ng (94 profiles), 0.047 ng (74 profiles), 0.0625 ng (90 profiles), 0.125 ng (97 profiles), and 0.25 ng (97 profiles). For the PowerPlex 16 HS kit, the 303 profiles were generated by targeting an input mass of 0.0078 ng (48 profiles), 0.0156 ng (47 profiles), 0.0313 ng (48 profiles), 0.0625 ng (48 profiles), 0.125 ng (48 profiles), and 0.25 ng (48 profiles).

It is to be noted that additional samples targeting 0.5 ng (45 profiles) and 1 ng (45 profiles) were amplified using both kits, but a substantial number of these resulted in at least one locus with "offscale" peaks, indicating that the detector was saturated. This resulted in only 29 and 14 0.5 ng Identifiler Plus and PowerPlex 16 HS usable profiles, respectively. In the 1 ng case, only 2 Identifiler Plus profiles showed no saturation. Given the number of saturated profiles, the statistical evaluations described below were not conducted on these "higher" template samples.

The data were generated with the GeneAmp PCR System 9700, the 3130 Genetic Analyzer, and the GeneMapper ID-X v1.1.1 software, where a constant RFU threshold of one was used. The injection time was 10 s and the injection voltage 3 kV. In the case of the Identifiler Plus amplification kit, the 29-cycle protocol was used and, in the case of the PowerPlex 16 HS kit, the 32-cycle protocol was chosen. The Identifiler Plus kit includes 15 tetranucleotide STR loci and the PowerPlex 16 HS kit 13 tetranucleotide and 2 pentanucleotide STR loci. All amplifications and sample preparations were performed according to the manufacturers' recommended protocols [21,22].

After generation of the data, all peaks originating from pull-up or -A were manually removed. In order for peaks to be classified as pull-up, the peak in question had to be in the same position (± 0.3 bases) as the allelic peak in another colour channel and have a peak height of 5% or less of the allelic peak. Further, if a peak fell between two adjacent allelic peaks in a different dye colour and had a "plateau-like" shape, then the peak in question was classified as complex pull-up. A peak was determined to be -A if it was one base shorter (± 0.3 bases) than an allelic peak. There were no height restrictions for the complex pull-up and -A artefacts. For a definition and explanation of the previous terms we refer to [1].

The genotypes of all individuals had been determined beforehand from high DNA input mass samples.

2.2. Data analysis

We enumerate the profiles from 1 to P, where P denotes the total number of profiles that are available for a given kit, and denote the measurement vectors by $y^p = (y_1^P, y_2^P, \ldots, y_I^P)^{\mathrm{T}}$, $p = 1, \ldots, P$. The components y_i^p, $i = 1, \ldots, I$, of the measurement vector are called measurement values. Note that the measurement values that are given by the GeneMapper ID-X v1.1.1 software are non-negative integers.

Knowing the genotypes of the individuals, we can group the components $y_1^p, y_2^p, \ldots, y_I^p$ of each measurement vector into three categories: true peak, stutter, and noise. We call a component y_i^p of the measurement vector $\underline{y}^p$ true peak if at index i there is either a homozygote allele or a heterozygote allele. Stutter is an artefact
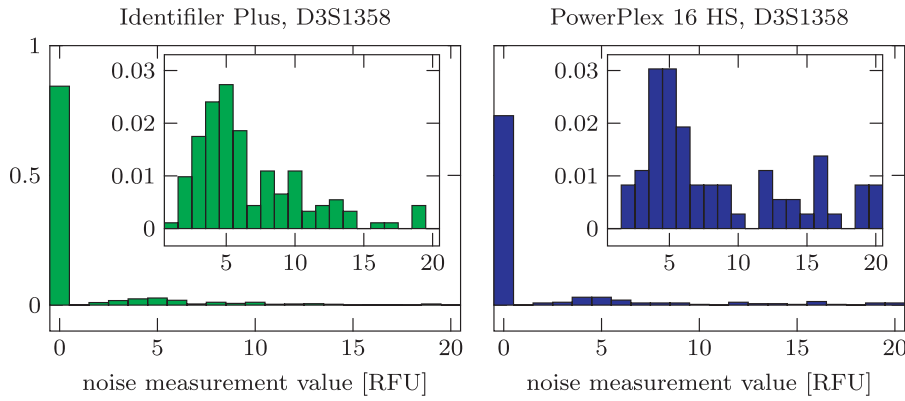
Fig. 1. Normalised histogram of the noise measurement values for locus D3S1358 and a DNA input mass of 0.25 ng. Left: Identifiler Plus kit. Right: PowerPlex 16 HS kit. Most of the noise measurement values are zero. The inset plot is a magnification without the first bin. Histogram colours (left: green, right: blue) reflect the fluorescent dye colour at that locus for that kit.

that is induced by the PCR. The amount of stutter is different for the Identifiler Plus kit and the PowerPlex 16 HS kit, in part because both kits use a different number of PCR cycles. Depending on where the stutter occurs, we distinguish among $N-2$, $N-1$, $N+1$, and $N+2$ stutter. $N-2$ stutter denotes a stutter fragment that is two repeat units shorter than the parent allele. Analogously, $N-1$ stutter is one repeat unit shorter, $N+1$ stutter one repeat unit longer, and $N+2$ stutter two repeat units longer than the parent allele. The most relevant stutter is $N-1$ stutter, because it is the strongest [23]; however, the PowerPlex 16 HS kit, which uses three additional PCR cycles, shows also stutter in the $N-2$ position, an issue we return to in Sections 3.2.1 and 3.4.4. Except in Sections 3.2.1 and 3.4.4, where we also consider $N-2$ stutter, we use the following convention, which is representative of what is done in practise: We call a component stutter if it is either in $N-1$ or in $N+1$ stutter location of a true peak. All remaining components are called noise. Note that the actual partitioning of the components of each measurement vector in the three categories depends on the genotype of the person whose DNA was used for generating the profile. Thus, for each profile p, we get in general a different set of noise components, the indices of which we denote by $\mathcal{I}^{n}(p) = \{i : y_i^p \text{ is a noise component}\}$. In Section 3.1, we study the frequency of noise measurement values being zero. In Section 3.2 we do a per-dye colour analysis to study if it is sufficient to have a per-dye colour description of the baseline noise. Since this is not the case, we continue in Section 3.4 with a per-locus analysis of the noise. The computations are done for all 15 STR loci. For visualisation, however, we restrict to CSF1PO, D3S1358, D5S818, and TPOX in the figures. The dye colours of these four loci are blue, green, red, and yellow for the Identifiler Plus kit, and green, blue, green, and yellow for the PowerPlex 16 HS kit.

3. Results

3.1. Frequency of noise measurement value zero

For each kit, we gather the noise components of all available measurement vectors in a list $\mathcal{N} = (n_1, n_2, \ldots, n_{|\mathcal{N}|}) = ((y_i^p)_{i \in \mathcal{I}^n(p)})_{p=1}^P$, the elements of which we call noise measurement values.

In Fig. 1 the normalised histogram of the noise measurement values is plotted for locus D3S1358 and both kits. It can be seen that most of the probability mass is concentrated in zero. The empirical frequency with which noise measurements are zero, which we denote by p_z, is given in Table 1 for the different loci and for data with a DNA input mass of 0.25 ng. This table is representative of all DNA input masses, as the frequency p_z does not significantly change with DNA input mass.

Let $\mathcal{I}^{n+}(p) = \{i \in \mathcal{I}^n(p) : y_i^p > 0\}$ denote the set of indices of the noise components with non-zero measurement value. Further, we call the elements of the list $\mathcal{N}^+ = (n_1^+, n_2^+, \ldots, n_{|\mathcal{N}^+|}^+) = ((y_i^p)_{i \in \mathcal{I}^{n+}(p)})_{p=1}^P$ non-zero noise measurement values. In the rest of the paper, we exclusively consider the non-zero noise measurement values $\mathcal{N}^+$.

Depending on the desired goal, for our analysis we restrict further the collection $\mathcal{N}^+$ to noise values coming from a specific dye colour, locus or DNA input mass. With a slight abuse of notation, we use the same symbol $\mathcal{N}^+$ for the restrictions, because the kind of restriction is clear from the context. In Table 1, for example, we consider in each row only the non-zero noise measurement values coming from samples with a DNA input mass of 0.25 ng and from the given locus. $|\mathcal{N}^+|$ denotes the number of elements in the collection $\mathcal{N}^+$.

3.2. Baseline noise distribution and dye colour

It is well-known that the statistical properties of the baseline noise are different for the different dye colours ([24], p. 42). In a first step, we also confirm this property for our data by performing the G-test of independence. This tests the hypothesis that two or more data sets are statistically consistent with being independent observations from a single source ([25], p. 68). The test indicates that the distribution of the non-zero noise measurement values is

Table 1
Empirical frequency p_z of the noise measurement values being zero and the number of non-zero noise measurement values $|\mathcal{N}^+|$ for the Identifiler Plus kit and the PowerPlex 16 HS kit for data with a DNA input mass of 0.25 ng. The high values of p_z show that most of the noise measurement values are zero.

Locus	IDplus		PowerPlex	
	p_z	$|\mathcal{N}^+|$	p_z	$|\mathcal{N}^+|$
CSF1PO	0.86	119	0.78	61
D13S317	0.81	94	0.66	71
D16S539	0.87	109	0.73	64
D18S51	0.89	328	0.79	203
D19S433	0.88	174	–	–
D21S11	0.85	382	0.79	272
D2S1338	0.84	169	–	–
D3S1358	0.84	143	0.73	98
D5S818	0.83	135	0.71	80
D7S820	0.86	135	0.74	56
D8S1179	0.81	173	0.80	84
FGA	0.88	469	0.83	254
Penta D	–	–	0.88	59
Penta E	–	–	0.86	111
TH01	0.91	110	0.84	47
TPOX	0.89	60	0.88	34
vWA	0.88	159	0.74	115

Table 2
List of loci for a given dye colour for the Identifiler Plus and the PowerPlex 16 HS kit.

	Dye colour	Loci
IDplus	Blue	CSF1PO, D7S820, D8S1179, D21S11
	Green	D2S1338, D3S1358, D13S317, D16S539, TH01
	Yellow	D18S51, D19S433, TPOX, vWA
	Red	D5S818, FGA, AMEL
PowerPlex	Blue	D3S1358, TH01, D21S11, D18S51, Penta E
	Green	D5S818, D13S317, D7S820, D16S539, CSF1PO, Penta D
	Yellow	AMEL, vWA, D8S1179, TPOX, and FGA

not independent of the dye colour. We do not present the detailed results here, because the test confirms what has been known before ([24], p. 42). A description of how the test is performed follows below.

However, it is less clear whether the statistical properties also depend on the locus itself, rather than on the dye colour. If there was no dependence on the locus except for the dye colour, it would be sufficient to model the baseline noise on a per-dye colour basis, which in turn would result in a simplified noise model.

In this section, we analyse the dependence of the noise distribution on the locus, performing the G-test of independence. The test is done separately for each kit, DNA input mass, and dye colour. The dye colours and the corresponding loci for the Identifiler Plus kit and the PowerPlex 16 HS kit are given in Table 2.

In order to create the contingency table for the test, we first fix the kit, DNA input mass, and dye colour. Each row of the contingency table contains the data for one locus. If L denotes the number of loci for the given dye colour, then the table has L rows. The data are binned so that the observed frequency in each bin is larger than or equal to 5, simultaneously for all loci. The exact binning procedure is described in Appendix A. If the binning results in K bins, the contingency table has K columns.

After the contingency table is created, we perform the G-test of independence, which is described in Appendix B. Additionally, we interpret the tests for the different DNA input masses as a compound test and perform the Holm–Bonferroni correction [26].

The Holm–Bonferroni correction is a method to control the family-wise error rate when several statistical tests are being performed by adjusting the alpha level, or equivalently, the p-value of each of the individual hypotheses. Details about the Holm–Bonferroni test can be found in Appendix C. For our commentary, we select a significance level of 5% so that if the p-value is smaller than 5%, the hypothesis that the noise distribution is independent of the locus is rejected. If the p-value is larger than 5%, the hypothesis is not rejected.

In Tables 3 and 4, we show the results of the independence test for the Identifiler Plus and PowerPlex 16 HS kit, respectively. For the Identifiler Plus kit, we observe that, for most of the dye colours and DNA input masses, the hypothesis is not rejected, indicating that the distribution of the non-zero noise measurement values is independent of the locus for these dye colours. However, for the blue dye the hypothesis is rejected for the three largest DNA input masses 0.0625 ng, 0.125 ng, and 0.25 ng, meaning that the distribution of the non-zero noise measurement values is not independent of the locus. Interestingly, for the PowerPlex 16 HS kit, the same three DNA input masses are rejected for the blue dye. Further, it can be seen that, for the PowerPlex 16 HS kit and the DNA input masses 0.125 ng, and 0.25 ng, the hypothesis is rejected for all dye colours. This gives us a strong indication that for this kit and the higher DNA input masses, the distribution of the non-zero noise measurement values depends on the locus.

Since this dependency is more pronounced for the higher DNA input masses, it is reasonable to assume that the allelic peaks contribute to the noise. One possible way of contribution is in the form of $N-2$ stutter.

3.2.1. Influence of $N-2$ stutter

For the PowerPlex 16 HS kit, some loci result in some unusual high noise peaks, as shown in Fig. 6 at locus D3S1358. A closer analysis shows that some of these peaks are in $N-2$ or $N+2$ stutter position, suggesting they are part of the PCR stutter pattern.

If we distinguish between all noise components that are in $N-2$ stutter position and those that are not, we observe the following difference between both groups. For the data from the PowerPlex 16 HS kit and a DNA input mass of 0.25 ng, for 7 of the

Table 3
Identifiler Plus kit: G-test of independence for loci. The column p contains the p-value (in %) of the G-test, the column p_{HB} the p-value (in %) after Holm–Bonferroni correction, and the column Rej. the rejection decision according to the Holm–Bonferroni correction using a significance level of 5%. For the blue dye colour and high DNA input masses, the hypothesis of independence is rejected.

DNA input (ng)	Blue			Green			Yellow			Red		
	p	p_{HB}	Rej.	p	p_{HB}	Rej.	p	p_{HB}	Rej.	p	p_{HB}	Rej.
0.0078	5.8	17.4	No	42.6	100.0	No	0.3	2.4	Yes	35.4	100.0	No
0.0156	62.8	62.8	No	29.5	100.0	No	51.1	100.0	No	88.7	100.0	No
0.0313	4.3	17.3	No	43.4	86.9	No	4.9	29.1	No	80.7	100.0	No
0.047	13.5	27.1	No	1.6	9.8	No	15.7	62.9	No	18.4	92.1	No
0.0625	0.2	1.2	Yes	70.3	70.3	No	60.4	60.4	No	96.5	96.5	No
0.125	0.4	2.1	Yes	40.6	100.0	No	58.3	100.0	No	14.2	99.5	No
0.25	0.8	3.9	Yes	0.6	4.4	Yes	5.0	24.9	No	14.6	87.8	No

Table 4
PowerPlex 16 HS kit: G-test of independence for loci. The column p contains the p-value (in %) of the G-test, the column p_{HB} the p-value (in %) after Holm–Bonferroni correction, and the column Rej. the rejection decision according to the Holm–Bonferroni correction using a significance level of 5%. For high DNA input masses, the hypothesis of independence is rejected for all dye colours.

DNA input (ng)	Blue			Green			Yellow		
	p	p_{HB}	Rej.	p	p_{HB}	Rej.	p	p_{HB}	Rej.
0.0078	18.8	37.6	No	17.2	51.6	No	7.7	7.7	No
0.0156	12.9	38.6	No	0.0	0.0	Yes	6.9	13.8	No
0.0313	56.5	56.5	No	70.6	70.6	No	0.3	1.3	Yes
0.0625	0.1	0.2	Yes	61.6	100.0	No	6.0	18.0	No
0.125	0.0	0.0	Yes	0.0	0.0	Yes	0.1	0.3	Yes
0.25	0.0	0.0	Yes	0.0	0.0	Yes	0.0	0.0	Yes

Table 5
Identifiler Plus kit with $N-2$ stutter removed: G-test of independence for loci. The column p contains the p-value (in %) of the G-test, the column p_{HB} the p-value (in %) after Holm–Bonferroni correction, and the column Rej. the rejection decision according to the Holm–Bonferroni correction with significance level 5%. Excluding $N-2$ stutter reduces the number of rejections.

DNA input (ng)	Blue			Green			Yellow			Red		
	p	p_{HB}	Rej.	p	p_{HB}	Rej.	p	p_{HB}	Rej.	p	p_{HB}	Rej.
0.0078	17.8	71.0	No	54.6	100.0	No	0.5	3.2	Yes	30.3	100.0	No
0.0156	44.5	44.5	No	47.2	100.0	No	69.1	69.1	No	79.5	100.0	No
0.0313	11.9	59.6	No	81.3	100.0	No	2.4	12.1	No	67.9	100.0	No
0.047	31.7	63.4	No	1.7	11.7	No	42.2	100.0	No	4.2	29.1	No
0.0625	0.2	1.3	Yes	85.8	100.0	No	45.5	100.0	No	94.5	94.5	No
0.125	5.3	32.1	No	98.7	98.7	No	47.9	95.8	No	46.2	100.0	No
0.25	31.7	95.1	No	50.9	100.0	No	1.5	9.1	No	43.9	100.0	No

15 loci, the maximum observed noise measurement value in $N-2$ stutter location is larger than 50 RFUs. In contrast, the maximum observed value of the noise peaks that are not in $N-2$ stutter location is larger than 50 RFUs only for 2 of the 15 loci.

It is a matter of convention, whether to treat peaks in $N-2$ stutter position as noise peaks or not. We chose to categorise them not as stutter, but as noise peaks in the previous section, because this seems to be closer to what is done in practise. However, from a modelling perspective, it might be beneficial to treat peaks in $N-2$ stutter position as artefacts in order to separate the effects of stutter more rigorously.

To study the impact of $N-2$ stutter, we remove all noise peaks in $N-2$ stutter location and redo the test of independence. The results of the test, which are shown in Tables 5 and 6, are different from the previous results. In particular, for some of the higher DNA input masses, the hypothesis that the noise distribution is independent of the locus is no longer rejected. However, for the PowerPlex 16 HS kit, there are still DNA input masses and dye colours for which the hypothesis of independence is rejected, and thus a simple per-dye colour noise model is not sufficient. For that reason, we choose to perform a per-locus analysis of the noise.

3.3. Quantisation

As already mentioned in Section 2.2, the measurement values y_i^p, $i = 1, \ldots, I$, that are given by the GeneMapper ID-X v1.1.1 software, and consequently the noise measurement values, are non-negative integers. However, the true measurement values, being light intensities, are supposed to be non-negative real numbers. That is, the measurement values y_i^p, $i = 1, \ldots, I$, that are available for our analysis are quantised versions of the true but unknown values $\tilde{y}_i^p$, $i = 1, \ldots, I$. In the statistical literature this is also known as grouped data. We model this reduction in information by $y_i^p = Q\tilde{y}_i^p$, where Q denotes the quantisation operator that is defined as

$$Qx = \left\lfloor x + \frac{1}{2} \right\rfloor. \tag{1}$$

$\lfloor x \rfloor$ denotes the largest integer smaller than or equal to x.

Since the true but unknown noise values are non-negative real numbers, we choose to model the noise by continuous instead of discrete random variables. However, in order to account for the quantised measurement values, we need to quantise the continuous random variables. The influence of the quantisation operator on the distribution of a random variable is as follows. If X is a continuous random variable with cumulative distribution function (CDF) $F_X(x) = \mathbb{P}(X \leq x)$, then the CDF of the quantised random variable QX is given by

$$F_{QX}(x) = \mathbb{P}(QX \leq x) = \mathbb{P}\left(X \leq \lfloor x \rfloor + \frac{1}{2}\right) = F_X\left(\lfloor x \rfloor + \frac{1}{2}\right).$$

3.4. Baseline noise distribution

To identify a simple parametric class of distributions that is consistent with the non-zero noise measurement values, we use the statistical G-test ([25], p. 53). The G-test is a goodness-of-fit test that is used to test the hypothesis that some observed data are drawn from a distribution. For fixed locus and fixed DNA input mass, we conduct the test for the quantised log-normal, quantised Gaussian, and quantised gamma distribution classes.

Each of these classes is a two-parameter class, providing a succinct description which is essential for calibration and, ultimately, to facilitate inference. Gaussian random variables, while commonly employed for noise characteristics, suffer the drawback of possibly taking both negative and positive values, which is inappropriate in the present application. Both log-normal and gamma random variables only take non-negative values and are right-skewed. The log-normal class, in particular, has often been found to well represent biological quantities [27].

We explain the test for the log-normal reference distribution class, as other reference distributions are treated analogously.

The log-normal cumulative distribution function (CDF) with parameters ν and τ is given by

$$F_{\nu,\tau}(x) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\ln(x) - \nu}{\tau\sqrt{2}}\right)\right], \tag{2}$$

Table 6
PowerPlex 16 HS kit with $N-2$ stutter removed: G-test of independence for loci. The column p contains the p-value (in %) of the G-test, the column p_{HB} the p-value (in %) after Holm–Bonferroni correction, and the column Rej. the rejection decision according to the Holm–Bonferroni correction with significance level 5%. Excluding $N-2$ stutter reduces the number of rejections.

DNA input (ng)	Blue			Green			Yellow		
	p	p_{HB}	Rej.	p	p_{HB}	Rej.	p	p_{HB}	Rej.
0.0078	39.4	100.0	No	29.3	100.0	No	3.3	13.4	No
0.0156	61.9	100.0	No	0.3	1.4	Yes	85.7	85.7	No
0.0313	18.4	73.8	No	72.3	100.0	No	9.9	29.7	No
0.0625	0.9	4.3	Yes	89.9	89.9	No	1.6	8.1	No
0.125	72.0	72.0	No	34.4	100.0	No	30.1	60.1	No
0.25	0.0	0.2	Yes	0.1	0.5	Yes	0.0	0.0	Yes

Table 7

Estimates $\hat{v}$ and $\hat{\tau}$ of the parameters v and τ of the log-normal distribution obtained by the maximum-likelihood parameter estimation from data with a DNA input mass of 0.25 ng.

Locus	IDplus		PowerPlex	
	$\hat{v}$	$\hat{\tau}$	$\hat{v}$	$\hat{\tau}$
CSF1PO	1.12	0.47	2.07	0.67
D13S317	1.55	0.52	2.22	0.63
D16S539	1.52	0.50	2.50	0.95
D18S51	2.01	0.43	1.90	0.74
D19S433	2.09	0.41	–	–
D21S11	1.19	0.50	1.77	0.67
D2S1338	1.54	0.57	–	–
D3S1358	1.76	0.59	2.46	0.94
D5S818	2.13	0.35	2.30	0.83
D7S820	1.10	0.40	2.14	0.52
D8S1179	1.34	0.67	1.85	0.71
FGA	2.03	0.38	2.02	1.06
Penta D	–	–	1.47	0.41
Penta E	–	–	1.42	0.47
TH01	1.38	0.52	1.92	0.54
TPOX	2.08	0.40	1.77	0.62
vWA	1.97	0.36	2.21	0.65

where

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2)\mathrm{d}t$$

denotes the error function. If X is a log-normal random variable, then the parameters v and τ are the mean and standard deviation of the associated Gaussian distribution, i.e. of $\log(X)$. The CDF of a quantised log-normal distributed random variable with

parameters $v = \hat{v}$ and $\tau = \hat{\tau}$ is given by

$$F^{\mathrm{q}}_{\hat{v},\hat{\tau}}(x) = F_{\hat{v},\hat{\tau}}\left(\lfloor x \rfloor + \frac{1}{2} \right). \tag{3}$$

3.4.1. Maximum-likelihood parameter estimation

In a first step, we determine the maximum-likelihood estimate of the parameters v and τ of the log-normal distribution from our data. Since our data, i.e., the available noise measurement values, are quantised, we cannot use the usual maximum-likelihood estimators for the parameters of the log-normal distribution, but need to take the binning into account in the estimation. In our case, the bins are given by $(0.5, 1.5], (1.5, 2.5], (2.5, 3.5], \ldots$. For $k \in \mathbb{N}$, let $(\underline{x}_k, \bar{x}_k]$ denote the k-th bin and $o_k = |\{j = 1, \ldots, |\mathcal{N}^+| : n_j^+ \in (\underline{x}_k, \bar{x}_k]\}|$ the observed number of non-zero noise measurement values in bin k, which is the number of noise measurement values equal to k. We obtain the maximum-likelihood estimates $\hat{v}, \hat{\tau}$ of the parameters v, τ by maximising numerically the log-likelihood function

$$l(v, \tau) = C + \sum_{k=1}^{B} o_k \log(F_{v,\tau}(\bar{x}_k) - F_{v,\tau}(\underline{x}_k)), \tag{4}$$

where C denotes a constant that does not depend on the parameters v and τ. $B = \max\{k \in \mathbb{N} : o_k \neq 0\}$ denotes the number of bins that are considered in (4). To solve the optimisation problem, the Nelder–Mead method in SciPy 0.13.2, Python 3.3.3 (64 bit) is used [28].

The estimates $\hat{v}$ and $\hat{\tau}$, obtained from data with a DNA input mass of 0.25 ng, are listed in Table 7 for the different loci. Further, for the loci CSF1PO, D3S1358, D5S818, and TPOX, the estimates $\hat{v}$ and $\hat{\tau}$ are plotted as a function of the DNA input mass in Fig. 2. For
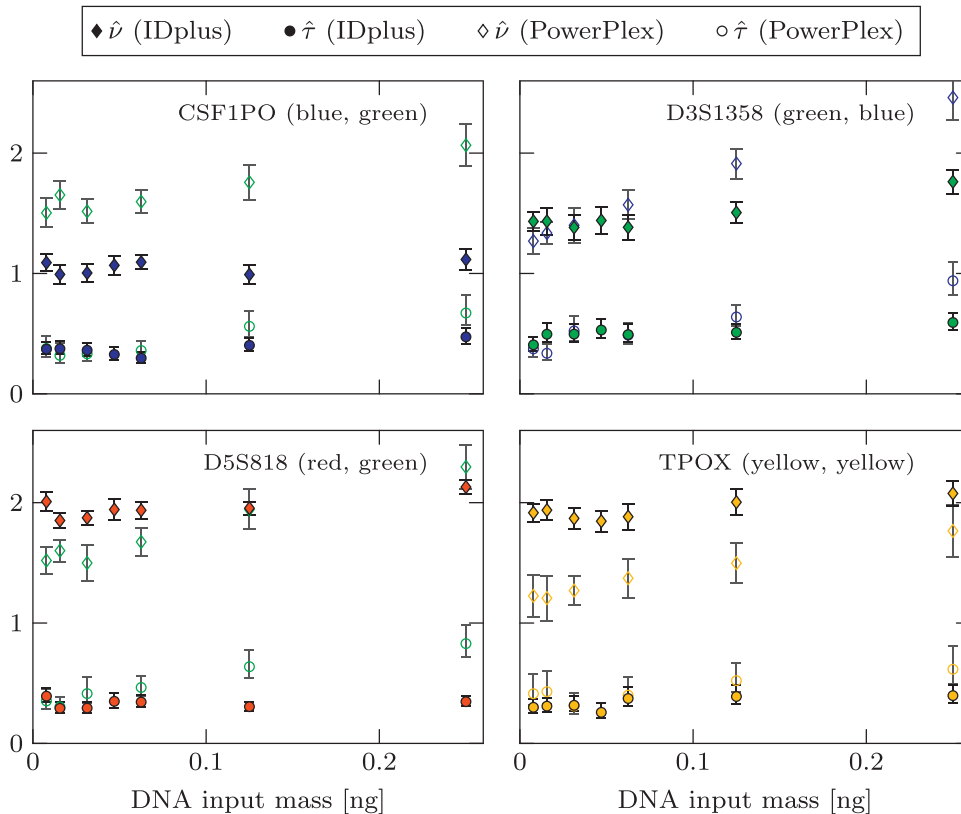

Fig. 2. Maximum-likelihood estimates $\hat{v}$ and $\hat{\tau}$, the two parameters of the best-fit log-normal distribution, as a function of the DNA input mass for different loci. The locus and the dye colours of the kits for that locus, where the first colour is for the Identifiler Plus kit and the second for the PowerPlex 16 HS kit, are given in the upper right corner of each subplot. For the Identifiler Plus kit, the parameters are almost constant over the given DNA input mass range, whereas for the PowerPlex 16 HS kit, there is a slight increase. Colours reflect the fluorescent dye colour at that locus for that kit.

the Identifiler Plus kit, the parameters are almost constant over the DNA input mass range of 0.0078 ng to 0.25 ng for most of the loci, whereas for the PowerPlex 16 HS kit, a slight increase can be noticed. This again suggests that there may be a dependence of the noise distribution on the DNA input mass, a phenomenon that we already observed in Section 3.2.

3.4.2. Results of the G-test

Having determined $\hat{v}$ and $\hat{\tau}$, we apply the G-test, which is described in Appendix D. As reference probability distribution, we use the distribution of a quantised log-normal distributed random variable with parameters $v = \hat{v}$ and $\tau = \hat{\tau}$, the CDF of which is given by (3). In Figs. 3 and 4 we see, as a solid line, the empirical CDF

$$F^{E}(x) = \frac{1}{|\mathcal{N}^{+}|} \sum_{i=1}^{|\mathcal{N}^{+}|} \mathbf{1}_{\{n_i^{+} \leq x\}} \qquad (5)$$

of the non-zero quantised noise measurement values in $\mathcal{N}^{+}$, and, as a dashed line, $F^{q}_{\hat{v},\hat{\tau}}(x)$, i.e., the CDF of a quantised log-normal distributed random variable with parameters $v = \hat{v}$ and $\tau = \hat{\tau}$ for the loci CSF1PO, D3S1358, D5S818, and TPOX and data obtained from samples with a DNA input mass of 0.25 ng. Fig. 3 contains the data for the Identifiler Plus kit and Fig. 4 the data for the PowerPlex 16 HS kit.

The result of the G-test is a p-value, which is a measure of the quality of a fit. We select a significance level of 5%, so that if the p-value is smaller than 5%, the hypothesis that the samples are taken from the reference distribution is rejected. If the p-value is larger than 5%, the hypothesis is not rejected.

Tables 10 and 11 give the p-values for the different distributions, loci, and DNA input masses. Table 10 contains the data for the Identifiler Plus kit and Table 11 the data for the PowerPlex 16 HS kit. For the Identifiler Plus kit and a DNA input mass of 0.25 ng,

except for locus D18S51, the p-values for the log-normal distribution are larger than 5%. This shows that the log-normal model provides good statistical consistency with the data. In contrast, all p-values for the Gaussian distribution are smaller than 5%, which is significant for the rejection of the hypothesis that the non-zero quantised noise measurement values follow a Gaussian distribution. For small DNA input masses, the gamma distribution class shows consistency with the data that is comparable to the log-normal distribution class. For large DNA input masses, there are more rejections for the gamma distribution class. For the PowerPlex 16 HS kit, the results are less clear; however, the log-normal distribution class still gives the best fit. The last row in both of the tables gives the number of rejections after a Holm–Bonferroni correction across the loci. Qualitatively, the results remain unchanged with this correction.

3.4.3. Visual comparison of the three distributions

In Figs. 5 and 6 we provide a visual comparison of the fit of the log-normal, the Gaussian, and the gamma distribution, where the parameters of the three distributions are obtained, as described in Section 3.4.1, from the data, by a maximum-likelihood estimation. We see the normalised histogram of the non-zero quantised noise measurement values together with the pseudo probability density functions (PDF) of the three distributions. By pseudo PDF we mean the function that is constant over the intervals $(k - 1/2, k + 1/2]$, $k \in \mathbb{N}$, and has a value equal to the probability mass of the corresponding distribution in each interval. Out of the three distributions, the log-normal distributions (dashed line) visually give the best fit.

3.4.4. Influence of N − 2 stutter on the quality of fit

In this section we study the influence of $N - 2$ stutter on the outcome of the G-test. To this end, we remove all noise peaks in
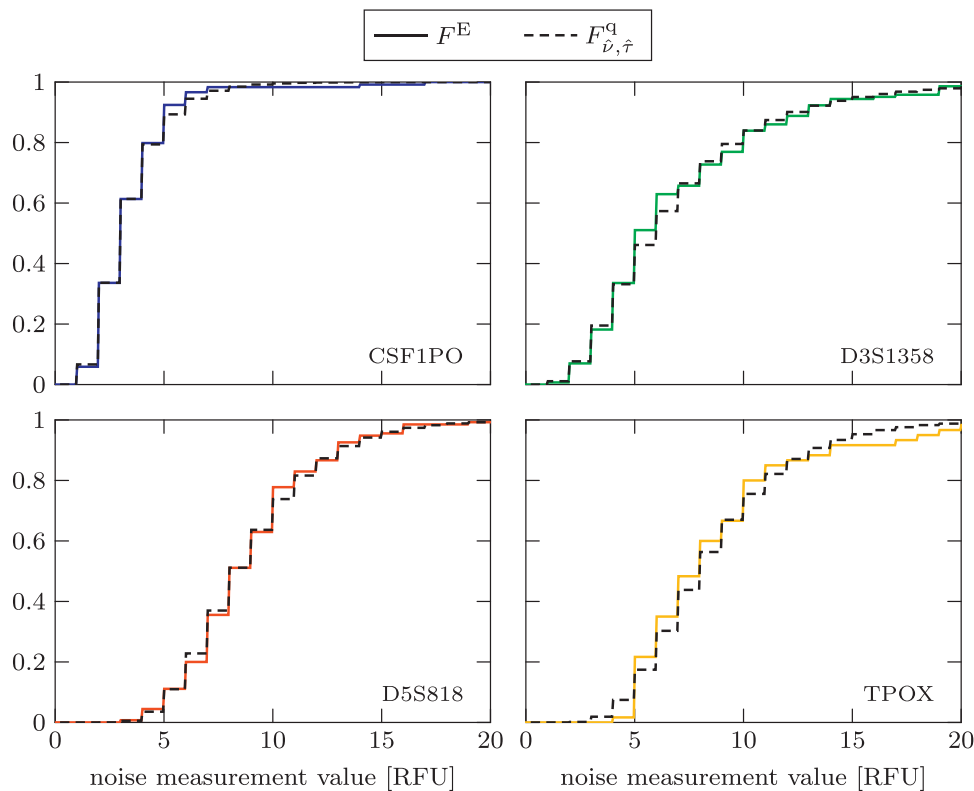

Fig. 3. Identifiler Plus kit: Empirical CDF F^{E} of the non-zero quantised noise measurement values (——) and CDF of the quantised log-normal random variable $F^{q}_{\hat{v},\hat{\tau}}$ (- - - -) with parameters $v = \hat{v}$ and $\tau = \hat{\tau}$ for a DNA input mass of 0.25 ng. The close match between the empirical CDF F^{E} and $F^{q}_{\hat{v},\hat{\tau}}$ is visible. Colours (top left: blue, top right: green, bottom left: red, bottom right: yellow) reflect the fluorescent dye colour at that locus.
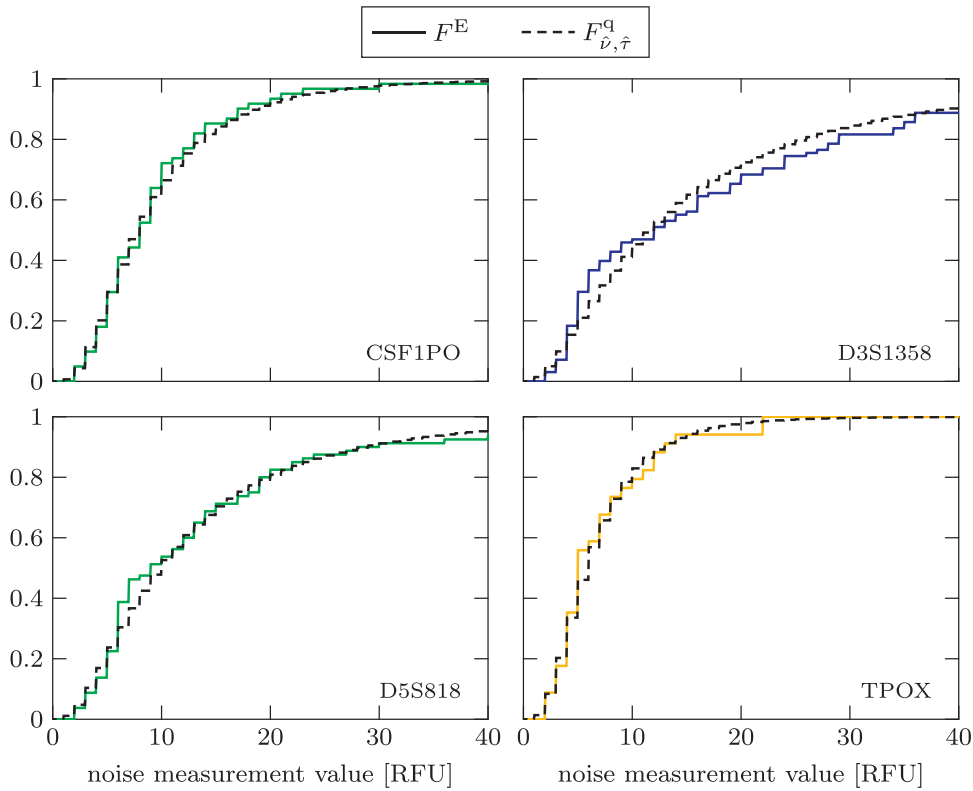
Fig. 4. PowerPlex 16 HS kit: Empirical CDF F^{E} of the non-zero quantised noise measurement values (——) and CDF of the quantised log-normal random variable $F^{\mathrm{q}}_{\hat{\nu},\hat{\tau}}$ (- - - -) with parameters $\nu = \hat{\nu}$ and $\tau = \hat{\tau}$ for a DNA input mass of 0.25 ng. The close match between the empirical CDF F^{E} and $F^{\mathrm{q}}_{\hat{\nu},\hat{\tau}}$ is visible. Colours (top left: green, top right: blue, bottom left: green, bottom right: yellow) reflect the fluorescent dye colour at that locus.
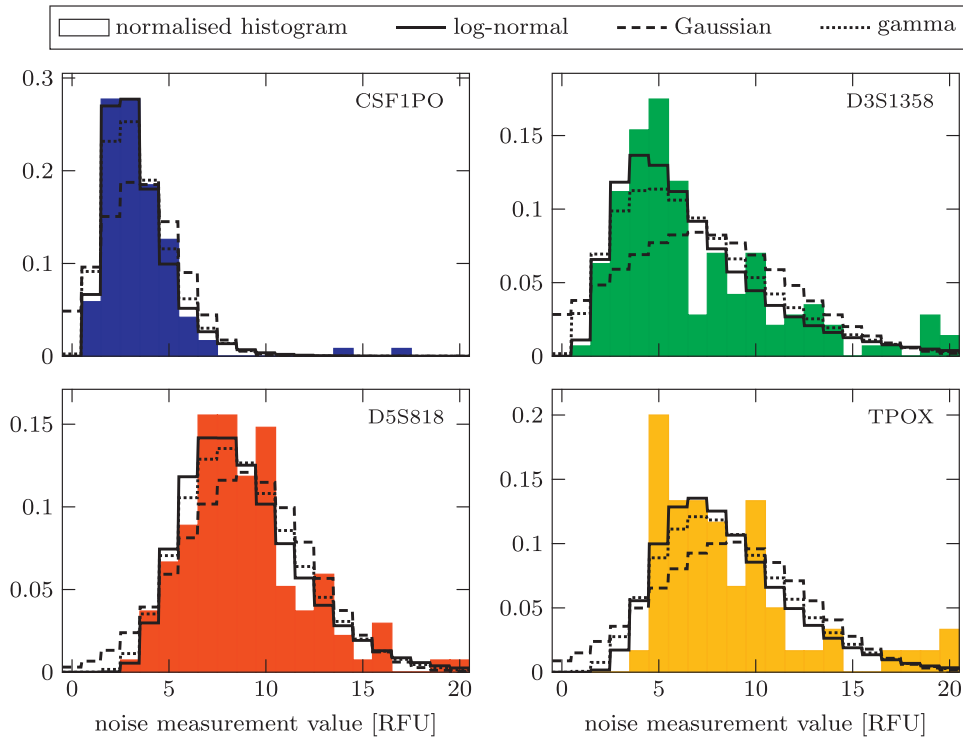
Fig. 5. Identifiler Plus kit: Normalised histogram of the non-zero quantised noise measurement values (▢) and pseudo PDF of a quantised log-normal distributed random variable (——), a quantised Gaussian distributed random variable (- - - -), and a quantised gamma distributed random variable (········) for a DNA input mass of 0.25 ng. The parameters of the distributions are determined by the maximum-likelihood estimators. The log-normal distributions visually give the best fit. Colours (top left: blue, top right: green, bottom left: red, bottom right: yellow) reflect the fluorescent dye colour at that locus.
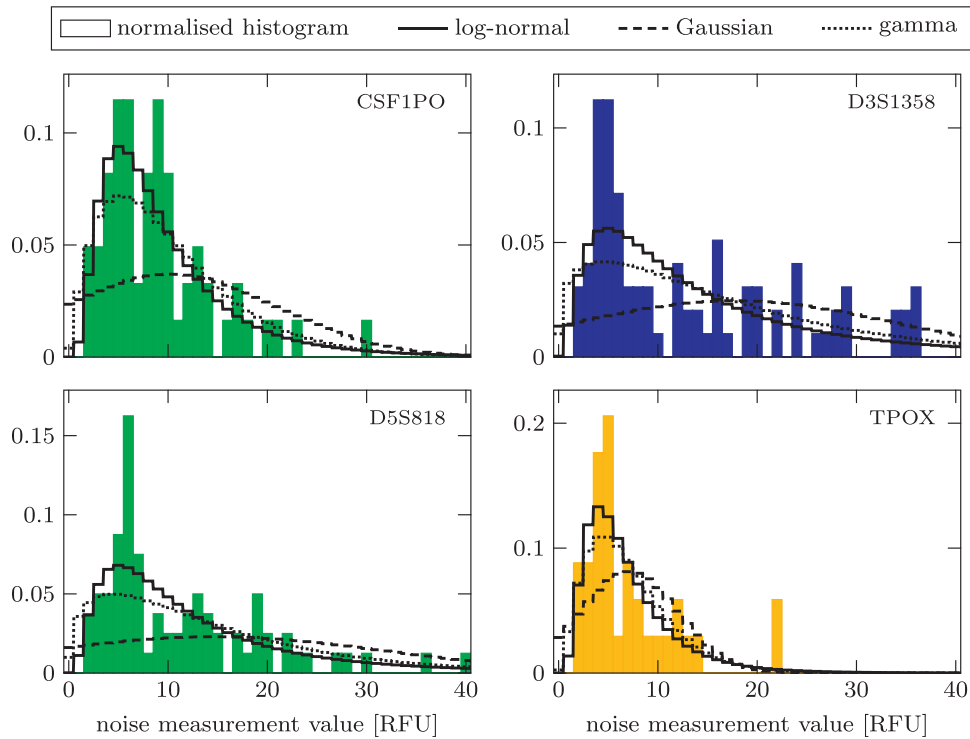
Fig. 6. PowerPlex 16 HS kit: Normalised histogram of the non-zero quantised noise measurement values (▭) and pseudo PDF of a quantised log-normal distributed random variable (——), a quantised Gaussian distributed random variable (- - - -), and a quantised gamma distributed random variable (·········) for a DNA input mass of 0.25 ng. The parameters of the distributions are determined by the maximum-likelihood estimators. The log-normal distributions visually give the best fit. Colours (top left: green, top right: blue, bottom left: green, bottom right: yellow) reflect the fluorescent dye colour at that locus.

$N - 2$ stutter location and redo the analysis as described before. We do not report all results, but only those that are unexpected.

It has to be noted that removing the noise peaks in $N - 2$ stutter location leads to a further reduction of the amount of data, as can be seen in Table 8. In particular, for the PowerPlex 16 HS kit there are certain DNA input mass and locus combinations for which there is insufficient data to perform the G-test.

First, we observe a change in the DNA input mass dependence of the parameter estimates $\hat{v}$ and $\hat{\tau}$. Unlike the slight increase that is observed in Section 3.4.1, here the parameter estimates are virtually constant, as shown in Fig. 7 for the loci CSF1PO, D3S1358,

Table 8
Empirical frequency p_z of the noise measurement values being zero and the number of non-zero noise measurement values $|\mathcal{N}^+|$ for the Identifiler Plus kit and the PowerPlex 16 HS kit for data with a DNA input mass of 0.25 ng and with $N - 2$ stutter removed.

Locus	IDplus		PowerPlex	
	p_z	$|\mathcal{N}^+|$	p_z	$|\mathcal{N}^+|$
CSF1PO	0.88	90	0.86	31
D13S317	0.86	59	0.77	39
D16S539	0.89	76	0.82	35
D18S51	0.90	288	0.83	151
D19S433	0.90	140	–	–
D21S11	0.87	318	0.82	221
D2S1338	0.88	108	–	–
D3S1358	0.89	91	0.84	49
D5S818	0.84	106	0.82	42
D7S820	0.87	113	0.82	29
D8S1179	0.85	124	0.85	53
FGA	0.89	434	0.86	198
Penta D	–	–	0.88	55
Penta E	–	–	0.87	91
TH01	0.92	93	0.87	31
TPOX	0.90	46	0.92	18
vWA	0.89	131	0.80	76

D5S818, and TPOX. One potential explanation of this observation could be that the increase of the parameter estimates towards the higher DNA input masses is caused by $N - 2$ stutter products.

The normalised histogram of the non-zero noise measurement values with $N - 2$ stutter removed and the pseudo PDFs of the quantised log-normal, Gaussian, and gamma distributions with parameters determined by the maximum-likelihood estimators are visualised in Figs. 8 and 9. Removing the $N - 2$ stutter peaks leads to a reduction in the number of large noise measurement values, as can be seen clearly at locus D3S1358 for both kits by comparing Fig. 5 with Fig. 8 and Fig. 6 with Fig. 9.

However, the quality of the fits is in general not significantly improved, an observation confirmed by the results of the G-test, given in Tables 12 and 13. Comparing Table 10 with Table 12 and Table 11 with Table 13, we see a slight improvement for the high DNA input masses, but also a slight deterioration for the low DNA input masses. While for the PowerPlex 16 HS kit the number of rejections for the log-normal distribution class after Holm–Bonferroni correction is reduced for the high DNA input masses

Table 9
Choice of the analytical threshold (AT) for a Gaussian noise model and for a log-normal noise model (AT_{ln}), second and third columns. Also shown is log base 10 of the resulting probability that a single noise measurement exceeds that threshold ($\log(p_{meas})$), column four, and the log base 10 of the probability that at least one of 50 noise measurements, as seen, approximately, per profile in our data, exceeds the threshold ($\log(p_{profile})$), column five.

k	AT	AT_{ln}	$\log(p_{meas})$	$\log(p_{profile})$
3	$\mu + 3\sigma$	$\exp(\nu + 3\tau)$	−2.8697	−1.1850
4	$\mu + 4\sigma$	$\exp(\nu + 4\tau)$	−4.4993	−2.8007
5	$\mu + 5\sigma$	$\exp(\nu + 5\tau)$	−6.5426	−4.8437
6	$\mu + 6\sigma$	$\exp(\nu + 6\tau)$	−9.0059	−7.3069
7	$\mu + 7\sigma$	$\exp(\nu + 7\tau)$	−11.8928	−10.1939
8	$\mu + 8\sigma$	$\exp(\nu + 8\tau)$	−15.1764	−13.4775

Table 10
Identifiler Plus kit: *p*-values of the G-test in %. In the last row, the number of rejections after a Holm–Bonferroni correction across the loci with significance level 5% is given. The Gaussian distribution class exhibits the most rejections. We have the fewest rejections for the log-normal distribution class.

Locus	0.0078 ng			0.0156 ng			0.0313 ng			0.047 ng		
	log-n.	Gauss	Gamma	log-n.	Gauss	Gamma	log-n.	Gauss	Gamma	log-n.	Gauss	Gamma
CSF1PO	84.1	1.7	80.2	10.9	0.0	3.4	32.8	4.4	63.1	9.0	33.0	25.9
D2S1338	43.8	0.0	35.9	63.1	0.0	55.2	69.9	0.0	20.9	22.2	0.0	8.3
D3S1358	0.8	0.0	1.4	13.5	1.9	32.6	27.1	0.2	34.5	3.3	0.0	7.4
D5S818	10.3	2.6	23.2	22.7	15.0	29.6	5.1	7.6	9.5	2.0	15.3	6.3
D7S820	37.3	0.0	29.9	50.0	0.3	56.6	0.2	0.0	1.4	16.6	5.8	56.9
D8S1179	19.7	1.0	46.4	4.3	0.5	13.6	8.9	0.0	0.0	30.1	0.3	38.5
D13S317	36.4	30.4	67.7	54.2	11.1	67.1	23.5	23.2	60.4	56.0	0.0	28.1
D16S539	28.0	1.7	47.7	42.5	4.7	61.7	14.3	0.4	13.5	0.1	0.7	0.3
D18S51	32.8	0.0	42.8	18.0	0.1	56.4	0.6	0.1	4.4	11.0	30.8	61.0
D19S433	8.5	14.6	40.0	12.0	7.3	34.3	0.0	1.1	0.1	23.8	10.2	47.3
D21S11	63.3	0.0	12.3	3.1	0.0	4.7	5.0	0.0	1.1	49.3	0.0	6.5
FGA	31.8	0.0	64.1	7.8	0.0	9.7	3.0	0.1	36.3	20.5	0.1	46.0
TH01	0.9	0.1	2.7	0.9	0.8	3.2	47.6	0.9	47.4	7.8	0.0	0.0
TPOX	0.5	1.4	1.0	5.1	3.9	6.5	27.5	76.5	51.2	3.3	1.4	2.7
vWA	17.1	65.4	52.4	25.8	27.6	60.8	0.8	43.5	12.4	7.3	0.4	7.2
# Rej.	0	7	0	0	7	0	2	8	2	1	9	2

Locus	0.0625 ng			0.125 ng			0.25 ng		
	log-n.	Gauss	Gamma	log-n.	Gauss	Gamma	log-n.	Gauss	Gamma
CSF1PO	15.2	1.4	12.2	10.3	0.0	1.9	53.7	0.0	14.1
D2S1338	89.1	0.8	84.1	53.7	0.0	32.9	7.8	0.0	4.7
D3S1358	28.2	0.0	22.9	92.2	0.0	45.0	19.2	0.0	1.2
D5S818	39.1	90.8	79.0	4.0	82.8	25.2	42.0	1.9	43.8
D7S820	1.4	0.6	8.4	57.1	0.0	28.9	76.4	0.0	23.9
D8S1179	12.1	0.0	4.1	34.0	0.0	0.9	20.3	0.0	0.0
D13S317	5.7	0.0	0.0	41.1	2.6	48.3	58.2	4.6	86.5
D16S539	80.6	0.0	50.9	63.8	0.0	47.8	29.4	0.0	16.1
D18S51	30.8	0.0	75.7	8.3	0.0	0.3	4.2	0.0	5.0
D19S433	6.1	0.0	2.2	1.3	1.7	5.6	45.6	0.0	17.7
D21S11	39.6	0.0	20.7	2.3	0.0	0.0	14.4	0.0	0.1
FGA	50.8	0.0	91.8	76.7	0.0	90.4	17.8	0.0	0.8
TH01	6.8	0.6	11.1	38.4	0.0	15.6	12.1	0.0	11.4
TPOX	40.8	21.0	46.5	42.2	43.6	60.4	38.1	2.2	31.1
vWA	6.6	0.0	0.2	51.1	3.6	75.4	54.6	1.2	54.6
# Rej.	0	13	2	0	10	2	0	12	2

of 0.125 ng and 0.25 ng, the number of rejections for the DNA input mass 0.0156 ng is increased. For the Identifiler Plus kit, the number of rejections for the log-normal distribution class after Holm–Bonferroni correction stays zero for the high DNA input masses of 0.0625 ng, 0.125 ng, and 0.25 ng. However, for the DNA input mass of 0.0313 ng the number of rejections is increased from one to two and for the DNA input mass of 0.0078 ng from zero to two.

Reclassifying peaks in the $N-2$ position as stutter rather than noise eliminated to a high degree the increase of the parameter estimates for the log-normal distribution. Although this result

Table 11
PowerPlex 16 HS kit: *p*-values of the G-test in %. Cells with "–" indicate an insufficient number of samples. In the last row, the number of rejections after a Holm–Bonferroni correction across the loci with significance level 5% is given. The Gaussian distribution class exhibits the most rejections. We have the fewest rejections for the log-normal distribution class.

| Locus	0.0078 ng			0.0156 ng			0.0313 ng			0.0625 ng			0.125 ng			0.25 ng		
	log-n.	Gauss	Gamma	log-n.	Gauss	Gamma	log-n.	Gauss	Gamma	log-n.	Gauss	Gamma	log-n.	Gauss	gamma	log-n.	Gauss	Gamma
D3S1358	24.0	3.9	22.3	43.7	41.6	68.6	16.3	0.0	10.3	70.2	0.0	37.1	26.3	0.0	11.3	0.4	0.0	0.1
TH01	31.2	9.2	34.2	53.3	0.0	3.5	–	–	–	7.6	14.3	17.0	26.9	0.1	17.7	7.6	0.0	2.9
D21S11	0.3	0.2	3.2	0.6	1.8	12.7	42.0	0.0	21.7	0.9	0.0	0.0	19.1	0.0	0.0	0.2	0.0	0.0
D18S51	0.1	0.0	0.0	17.4	0.0	0.8	0.4	0.0	0.0	12.0	0.0	28.5	6.3	0.0	0.0	0.0	0.0	0.0
Penta E	26.4	0.0	8.6	54.2	1.5	53.3	52.3	6.7	61.1	46.0	1.0	33.0	10.6	0.0	0.7	4.1	0.0	0.0
D5S818	13.8	27.0	23.3	57.1	67.0	77.6	38.6	48.5	62.7	4.2	0.0	0.4	33.1	0.0	2.8	2.1	0.0	0.1
D13S317	–	–	–	–	–	–	16.0	24.8	23.1	1.2	0.0	0.2	9.5	0.1	7.8	21.5	0.0	17.0
D7S820	35.5	7.4	27.8	–	–	–	32.6	50.1	46.3	0.4	0.0	0.0	20.3	0.0	7.8	34.3	0.5	36.4
D16S539	–	–	–	7.4	0.0	1.4	24.5	1.0	24.7	18.3	0.0	8.3	87.5	0.0	57.7	0.5	0.0	0.4
CSF1PO	19.2	24.3	30.8	24.8	16.6	28.1	5.0	13.6	9.9	27.2	0.5	16.2	20.5	0.0	1.4	68.7	0.0	20.3
Penta D	87.9	18.9	90.6	15.7	9.1	27.5	64.7	1.3	52.3	5.2	0.1	3.0	80.1	0.4	52.7	61.1	0.0	24.8
vWA	0.9	35.4	6.9	71.5	23.6	71.8	3.8	0.0	0.1	49.6	1.0	44.7	80.3	0.0	85.4	81.0	0.0	19.5
D8S1179	0.4	0.0	0.0	37.8	0.0	13.5	81.6	0.0	20.6	3.5	0.0	2.5	13.7	0.0	2.8	49.4	0.0	50.7
TPOX	–	–	–	7.2	30.2	18.0	–	–	–	–	–	–	7.9	0.0	5.2	12.1	0.0	6.0
FGA	12.2	0.0	1.4	0.4	0.0	0.0	1.1	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
# Rej.	3	5	2	0	5	1	0	6	3	0	13	5	1	15	3	4	15	7
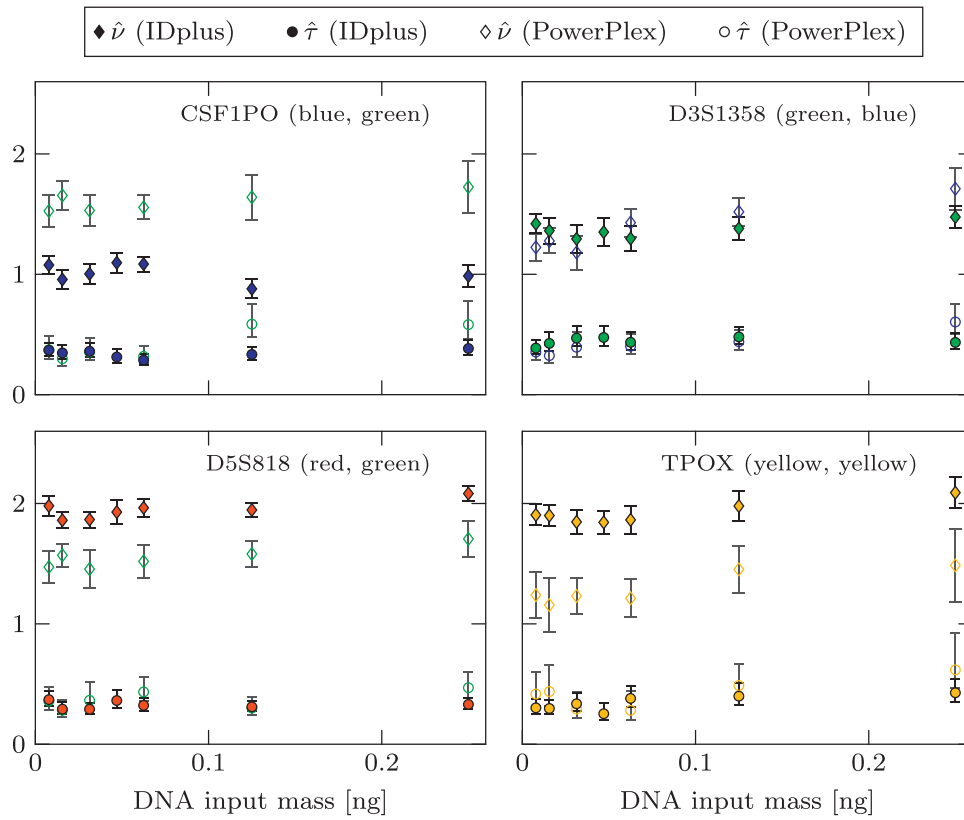
Fig. 7. Maximum-likelihood estimates $\hat{\nu}$ and $\hat{\tau}$ from data with $N - 2$ stutter removed as a function of the DNA input mass for different loci. The locus and the dye colours of the kits for that locus, where the first colour is for the Identifiler Plus kit and the second for the PowerPlex 16 HS kit, are given in the upper right corner of each subplot. For both kits the parameters are almost constant over the given DNA input mass range. Colours reflect the fluorescent dye colour at that locus for that kit.
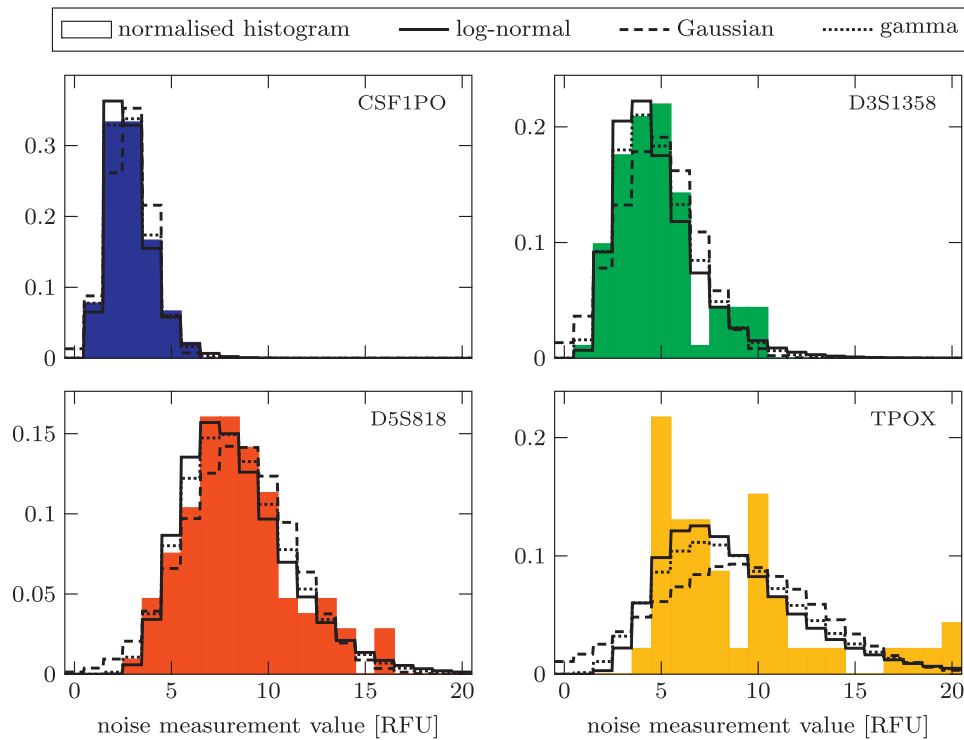

Fig. 8. Identifiler Plus kit with $N - 2$ stutter removed: Normalised histogram of the non-zero quantised noise measurement values (▢) and pseudo PDF of a quantised log-normal distributed random variable (——), a quantised Gaussian distributed random variable (- - - -), and a quantised gamma distributed random variable (·········) for a DNA input mass of 0.25 ng. The parameters of the distributions are determined by the maximum-likelihood estimators. The log-normal distributions visually give the best fit. Colours (top left: blue, top right: green, bottom left: red, bottom right: yellow) reflect the fluorescent dye colour at that locus.
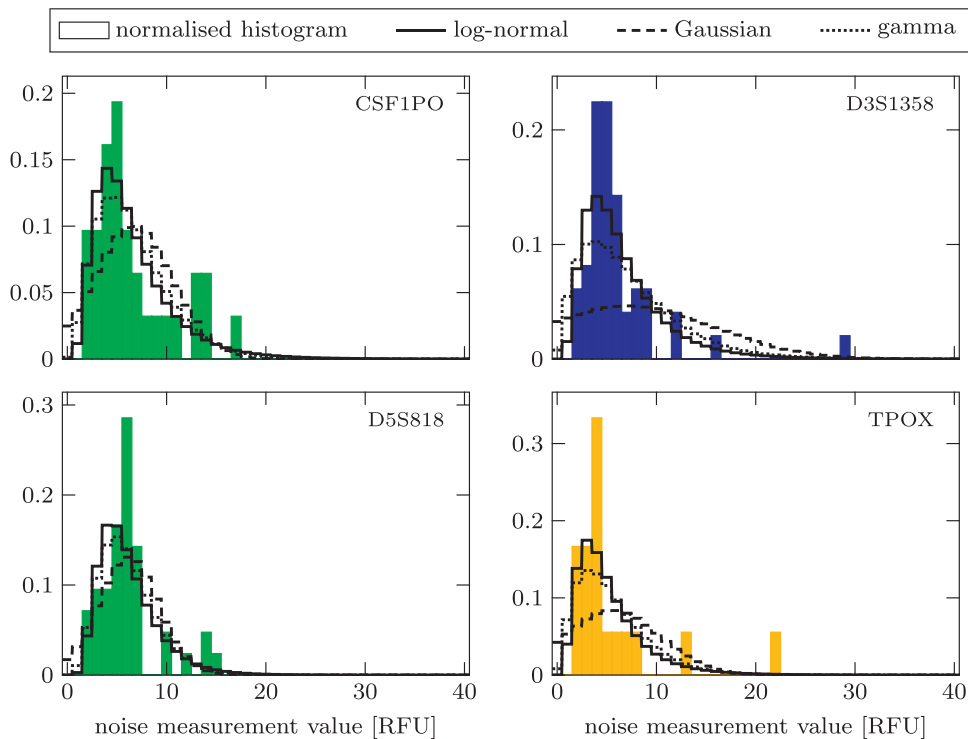
Fig. 9. PowerPlex 16 HS kit with $N - 2$ stutter removed: Normalised histogram of the non-zero quantised noise measurement values (▢) and pseudo PDF of a quantised log-normal distributed random variable (——), a quantised Gaussian distributed random variable (- - - -), and a quantised gamma distributed random variable (·········) for a DNA input mass of 0.25 ng. The parameters of the distributions are determined by the maximum-likelihood estimators. The log-normal distributions visually give the best fit. Colours (top left: green, top right: blue, bottom left: green, bottom right: yellow) reflect the fluorescent dye colour at that locus.

would advocate the explicit treatment of $N - 2$ stutter in noise analysis, the results of the G-test suggest that this is only necessary for DNA input masses larger than 0.047 ng.

4. Discussion

Based on the data from 643 profiles for the Identifiler Plus kit and 303 profiles for the PowerPlex 16 kit, we perform statistical tests for the suitability of the log-normal, Gaussian, and gamma distribution classes to describe baseline noise. Overall, the log-normal distribution class provides good statistical consistency with the data and so can be employed to summarise succinctly peak-height distributions through a small number of parameters. The Gaussian distribution class, which is commonly used to describe noise, gives the worst fit. This suggests replacing Gaussian noise models, as employed, for example, in [17], with those based on log-normal distributions in order to enhance the accuracy of deductions in the analysis of mixtures.

We further observe a dependence of the noise distribution on the locus, in particular if $N - 2$ stutter is not removed. Thus, a per-dye colour description of noise seems to be inadequate and a per-locus description is needed.

Often negatives, that is samples with no DNA included, are used to study the properties of baseline noise. Our results indicate that noise obtained from negatives is not representative of the noise when the samples contain a significant amount of DNA. If $N - 2$ stutter is not removed, the DNA input mass has an influence on the noise distribution, especially for the PowerPlex 16 HS kit. Thus, it is advisable to study noise and, consequently, for laboratories that wish to employ an analytical threshold (AT) to ensure a reasonable level of protection against the false detection of noise, to infer the AT from data with a DNA input mass comparable to the DNA input mass of the samples to be analysed.

Prior to the present study, the AT would typically be determined by assuming the noise is generated from a Gaussian distribution. We describe that process alongside an equally simple, equivalent procedure based on modelling the noise as log-normally distributed. In advance, we note that for the same probabilistic level of protection against noise, the log-normal AT, AT_{ln}, will typically be larger than that suggested by the Gaussian AT.

Determination of the AT via the following approach has been described in both the forensic and analytical chemistry literature [6,18,29–32] based on a Gaussian noise assumption. One estimates the mean μ and standard deviation σ of the non-zero noise observations, circa 50 per profile in our data, picks a number of standard deviations k, and sets

$$AT = \mu + k\sigma.$$

Thus defined, the probability that a non-zero noise measurement exceeds AT, and thus the level of protection against noise, does not depend on μ or σ, but rather on k:

$$p_{\text{meas}} = \mathbb{P}(\text{Norm}(\mu, \sigma) > AT) = \frac{1}{2}\left(1 - \text{erf}\left(\frac{k}{\sqrt{2}}\right)\right), \qquad (6)$$

where erf is the error function, and $\text{Norm}(\mu, \sigma)$ denotes a Gaussian random variable with mean μ and variance σ^2. This exceedance probability is tabulated in Table 9 for a range of values of k.

To identify an AT based on a log-normal noise model that gives an identical level of protection, one estimates the mean ν and variance τ of the natural-log (log base e) of the non-zero measurements either directly or, for example, by the methods described earlier in this paper that take the quantisation from the GeneMapper ID-X v1.1.1 software into account. The appropriate log-normal AT is then

$$AT_{ln} = \exp(\nu + k\tau).$$

Table 12
Identifiler Plus kit with $N - 2$ stutter removed: p-values of the G-test in %. In the last row, the number of rejections after a Holm–Bonferroni correction across the loci with significance level 5% is given. The Gaussian distribution class exhibits the most rejections.

Locus	0.0078 ng			0.0156 ng			0.0313 ng			0.047 ng		
	log-n.	Gauss	Gamma	log-n.	Gauss	Gamma	log-n.	Gauss	Gamma	log-n.	Gauss	Gamma
CSF1PO	47.9	0.7	30.9	17.7	0.4	11.7	40.2	6.2	65.4	16.3	32.3	32.3
D2S1338	61.8	0.1	50.3	29.0	0.0	26.3	26.3	0.0	2.1	13.5	0.0	5.6
D3S1358	7.3	0.3	7.9	4.1	3.7	10.1	70.8	1.6	72.3	2.3	0.3	4.6
D5S818	0.3	9.9	2.7	9.3	3.1	9.6	10.0	10.5	14.7	0.6	5.7	1.9
D7S820	34.8	0.0	33.5	59.0	0.5	59.4	0.3	0.1	2.2	20.0	4.7	61.1
D8S1179	39.6	0.3	42.5	11.6	1.8	27.6	36.6	0.0	0.7	14.5	0.2	13.3
D13S317	51.6	25.2	68.3	21.4	32.9	42.1	0.5	3.9	1.9	27.8	0.0	6.7
D16S539	23.7	3.6	49.2	44.8	3.5	55.4	29.5	7.6	43.2	0.2	0.8	0.6
D18S51	34.2	0.0	38.2	17.2	0.2	58.5	0.2	0.6	3.6	22.9	41.9	81.0
D19S433	7.7	58.8	48.5	10.6	4.5	25.9	0.0	1.4	0.2	10.1	4.3	23.6
D21S11	70.5	0.0	27.2	18.3	0.0	12.4	6.9	0.0	1.2	28.1	0.0	2.8
FGA	37.0	0.0	64.1	6.9	0.0	6.9	2.1	0.4	34.8	9.1	0.1	29.3
TH01	1.1	0.2	3.2	4.4	2.8	11.2	50.0	1.3	51.2	57.1	0.0	22.4
TPOX	0.2	1.8	0.6	22.8	12.7	23.4	16.4	33.7	27.9	5.5	3.3	4.9
vWA	22.3	68.2	58.5	0.6	10.8	4.6	0.6	50.8	10.0	5.8	0.2	4.8
# Rej.	2	9	0	0	5	0	3	5	1	1	8	0

Locus	0.0625 ng			0.125 ng			0.25 ng		
	log-n.	Gauss	Gamma	log-n.	Gauss	Gamma	log-n.	Gauss	Gamma
CSF1PO	12.6	0.9	8.0	6.1	0.1	3.7	28.2	4.6	58.2
D2S1338	62.3	10.5	88.0	56.9	0.9	77.9	15.9	0.3	28.1
D3S1358	29.6	19.0	64.5	22.5	0.0	7.6	0.9	0.9	1.6
D5S818	45.0	80.5	74.2	15.1	87.4	51.5	47.1	35.7	71.6
D7S820	1.2	0.1	3.6	53.9	0.3	42.9	26.3	0.6	27.6
D8S1179	8.0	0.0	2.5	46.1	0.0	11.0	86.7	0.0	12.7
D13S317	8.3	0.0	0.0	46.9	4.0	45.5	21.9	3.7	40.2
D16S539	83.6	0.0	44.3	37.5	1.9	43.8	50.9	0.5	40.1
D18S51	23.5	0.1	74.9	6.3	0.0	0.3	2.3	0.0	8.0
D19S433	3.3	0.0	1.2	1.8	1.8	6.0	18.7	0.0	5.1
D21S11	44.5	0.0	32.9	25.1	0.0	0.8	60.3	0.0	6.9
FGA	58.3	0.0	92.2	48.1	0.0	85.5	28.7	0.0	3.9
TH01	7.4	1.9	15.4	35.1	0.0	12.4	63.7	0.0	45.7
TPOX	26.9	10.6	29.2	44.7	26.4	53.5	9.3	1.9	11.2
vWA	11.3	0.0	0.7	42.3	2.6	64.1	15.3	4.9	21.1
# Rej.	0	9	1	0	8	1	0	10	0

Thus defined, for any given k (integer or otherwise), the probability that a log-normal noise measurement exceeds $AT_{\ln}$ is the same as in the Gaussian model as the following holds

$$\mathbb{P}(\exp(\mathrm{Norm}(\nu, \tau)) > AT_{\ln}) = \mathbb{P}(\mathrm{Norm}(\nu, \tau) > \nu + k\tau)$$
$$= \frac{1}{2}\left(1 - \mathrm{erf}\left(\frac{k}{\sqrt{2}}\right)\right),$$

the same value as in Eq. (6). Thus replacing AT by $AT_{\ln}$, one obtains the same level of protection as in the originally assumed Gaussian model, but with the log-normal nature of the noise taken into account.

The primary quantitative difference between the two noise models is that large noise values are more likely with the log-normal model. Thus, while for a given k the resulting AT and $AT_{\ln}$

Table 13
PowerPlex 16 HS kit with $N - 2$ stutter removed: p-values of the G-test in %. Cells with "–" indicate an insufficient number of samples. In the last row, the number of rejections after a Holm–Bonferroni correction across the loci with significance level 5% is given. The Gaussian distribution class exhibits the most rejections. We have the fewest rejections for the log-normal distribution class.

| Locus	0.0078 ng			0.0156 ng			0.0313 ng			0.0625 ng			0.125 ng			0.25 ng		
	log-n.	Gauss	Gamma	log-n.	Gauss	Gamma	log-n.	Gauss	Gamma	log-n.	Gauss	Gamma	log-n.	Gauss	Gamma	log-n.	Gauss	Gamma
D3S1358	29.9	11.3	32.4	4.6	30.3	12.3	1.9	8.4	4.8	44.3	0.0	14.7	30.0	0.1	18.2	2.6	0.0	0.0
TH01	–	–	–	12.9	0.0	0.5	–	–	–	0.7	1.8	1.3	32.1	2.7	32.0	27.1	0.0	9.6
D21S11	0.1	1.2	3.0	0.4	1.8	9.4	29.0	1.7	49.6	2.4	0.0	0.0	51.1	0.0	4.9	0.3	0.0	0.0
D18S51	0.0	0.0	0.0	64.7	0.0	10.2	2.2	0.0	0.1	10.6	0.6	34.2	35.1	1.3	84.8	0.2	0.0	0.0
Penta E	8.2	0.0	1.6	38.0	1.1	38.5	16.1	3.5	23.0	29.9	0.4	15.2	58.8	0.2	30.9	20.5	0.0	5.2
D5S818	3.7	3.7	5.2	32.7	53.4	45.7	14.0	34.3	24.8	1.7	0.0	0.2	32.9	17.6	32.0	1.0	0.1	1.2
D13S317	–	–	–	–	–	–	10.9	20.7	15.4	–	–	–	–	–	–	12.9	0.0	6.1
D7S820	–	–	–	–	–	–	–	–	–	–	–	–	24.3	6.7	40.4	7.8	0.3	7.6
D16S539	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.5	0.0	0.0
CSF1PO	8.5	5.6	10.3	30.9	8.8	23.6	4.7	12.6	8.7	24.6	36.9	36.3	4.0	0.0	0.1	25.8	0.6	22.8
Penta D	13.3	18.1	25.6	12.6	9.0	18.2	86.8	12.8	84.3	4.7	0.1	2.5	80.4	0.2	43.9	38.3	0.0	12.3
vWA	0.8	24.1	4.5	56.9	17.3	64.4	40.9	1.5	22.7	2.1	13.4	6.5	39.6	8.0	73.7	61.9	0.0	23.6
D8S1179	2.0	0.0	1.3	41.9	0.0	18.6	75.0	0.0	23.9	0.9	0.0	0.7	41.8	0.0	12.9	11.0	0.3	19.7
TPOX	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
FGA	5.5	0.0	0.5	0.1	0.0	0.0	5.0	0.0	0.0	2.2	0.0	0.1	7.9	0.0	1.1	4.1	0.0	0.0
# Rej.	2	4	2	2	4	2	0	3	2	0	8	3	0	7	1	2	14	5
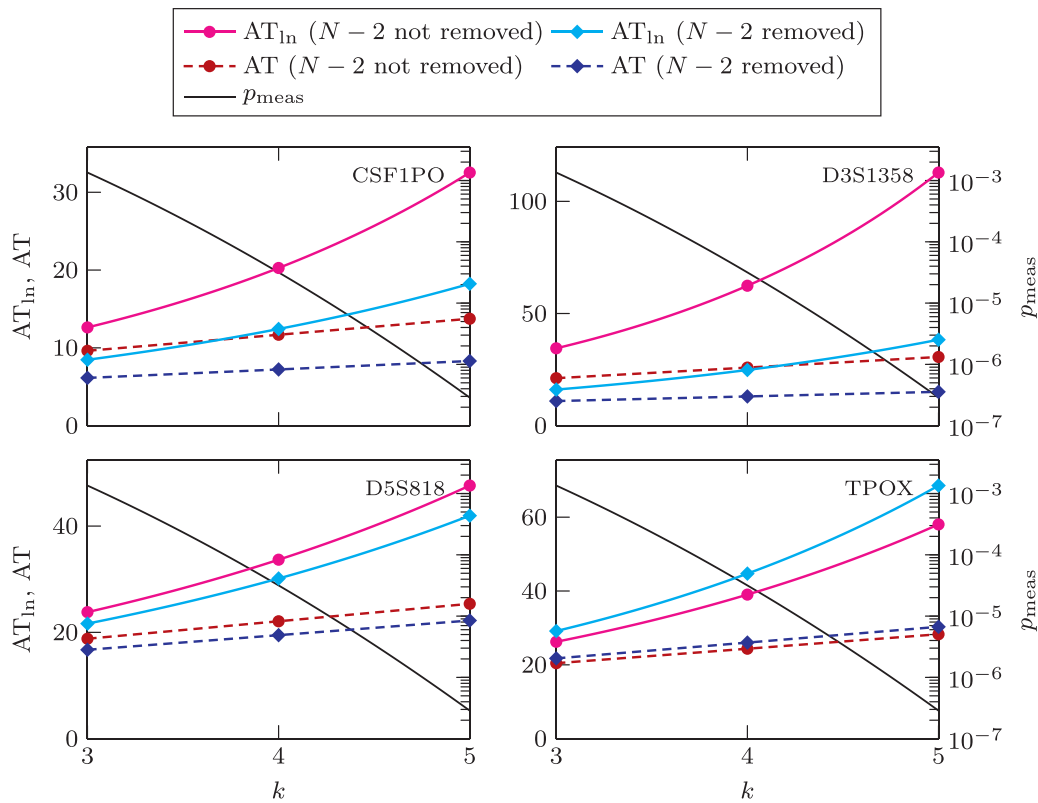
Fig. 10. Identifiler Plus kit: analytical threshold using the log-normal noise model (AT_{ln}) and the Gaussian noise model (AT) as a function of k and the corresponding level of protection, described by p_{meas}. The left ordinate of each subplot is for AT_{ln} and AT, whereas the right ordinate is for p_{meas}.
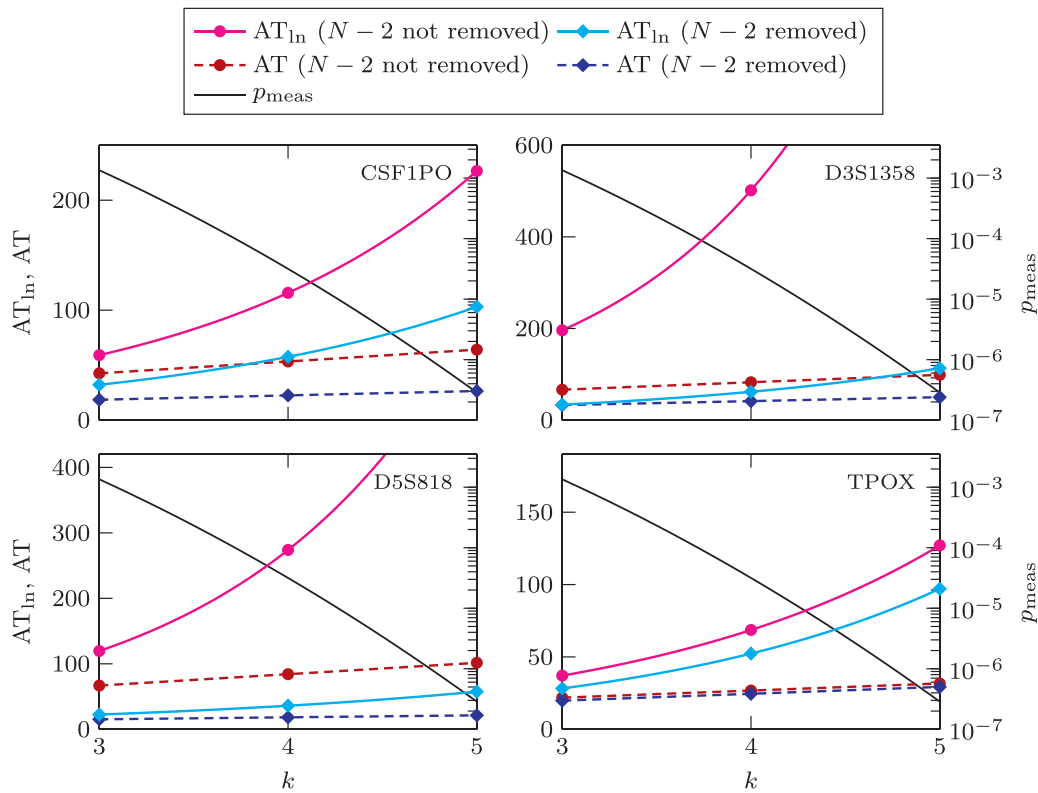

Fig. 11. PowerPlex 16 HS kit: analytical threshold using the log-normal noise model (AT_{ln}) and the Gaussian noise model (AT) as a function of k and the corresponding level of protection, described by p_{meas}. The left ordinate of each subplot is for AT_{ln} and AT, whereas the right ordinate is for p_{meas}.

provide the same level of probabilistic protection against noise, the thresholds themselves behave differently as a function of k. The AT based on the Gaussian noise model depends linearly on k, but the AT based on the log-normal noise model depends exponentially on k. As a result, in order to obtain a higher level of protection against noise, substantially larger ATs are truly required than one would deduce from the ill-fitting Gaussian noise model.

That feature is illustrated in Figs. 10 and 11, which show both the usual AT and AT_{ln} for a range of values of k. As k increases, to provide protection against the log-normally distributed noise, which was found to be appropriate for this data, the AT_{ln} grows substantially larger than the AT determined by the traditional procedure.

For CSF1PO, D5S818, and TPOX, moving to the more accurate log-normal noise model results in a modest increase in the AT. For D3S1358, however, it results in a significant increase. Furthermore, Figs. 10 and 11 also illustrate the impact of the $N-2$ stutter position being classified as noise on AT and AT_{ln}. This argues, perhaps, in favour of continuous interpretation methods that explicitly include noise and stutter artefacts over methods that attempt to filter it.

Regarding the choice of k, under either model, $k = 3$ leads to a probability of $p_{meas} = 0.0013$ that a single non-zero noise measurement value is larger than the AT. This may be considered tolerable, but given that each profile contains circa 50 non-zero noise measurements, it would suggest that over 6% of profiles would be expected to exhibit at least one noise measurement value that is larger than the AT (see the final column in Table 9).

Alternatively, one may set a level of protection such that less than 1% (or another pre-defined percentage) of profiles contain at least one noise measurement value that exceeds the AT. In this instance, assuming each profile contains roughly 50 non-zero noise measurement values, the AT under either model would be determined by adding $k = 4$ standard deviations to the mean. In this case, circa 1 of 631 profiles (probability $10^{-2.801}$) would exhibit noise that is larger than the AT.

In any case, it is to be noted that the relationship between the number of standard deviations from the mean at which the AT is set and the probability that a profile exhibits a noise measurement value that exceeds the AT is sensitive, as depicted in Table 9. For this reason, utilising an arbitrarily chosen large k such as 10 is not recommended. Note that, even at $k = 7$, the probability of observing a profile with noise which is larger than AT is roughly 10^{-10} (1 in 10 billion), which may be excessively conservative and result in high non-detection rates [5].

Acknowledgements

Appendix A. Binning for the G-test of independence

In this section we provide a description of the procedure that was used to bin the data for the G-test of independence. The data are binned so that the observed frequency in each bin is larger than or equal to 5, simultaneously for all loci $l = 1, \ldots, L$. Let $\mathcal{N}_l^+ =$

$(n_{l,1}^+, n_{l,2}^+, \ldots, n_{l,|\mathcal{N}_l^+|}^+)$ denote the collection of all non-zero noise measurement values for locus l, and let

$$F_l^E(x) = \frac{1}{|\mathcal{N}_l^+|} \sum_{i=1}^{|\mathcal{N}_l^+|} \mathbf{1}_{\{n_{l,i}^+ \leq x\}}$$

be the empirical distribution for locus l. We construct the bins iteratively. Let $x_0 = 0$. Assume that we already have determined x_k, then x_{k+1} is the smallest natural number such that

$$|\mathcal{N}_l^+|(F_l^E(x_{k+1} + 1/2) - F_l^E(x_k + 1/2)) \geq 5 \qquad (7)$$

for all $l = 1, \ldots, L$. The construction ends if there is no such number. Let x_K denote the last number found by this procedure. Then the K bins are given by $(x_0 + 1/2, \ x_1 + 1/2], \ (x_1 + 1/2, \ x_2 + 1/2], \ \ldots, (x_{K-1} + 1/2, \infty)$.

Appendix B. G-test of independence

The G-test of independence is shortly described here for completeness. Assume that the contingency table has L rows and K columns, and let $\mathcal{N}_l^+ = (n_{l,1}^+, n_{l,2}^+, \ldots, n_{l,|\mathcal{N}_l^+|}^+)$ denote the collection of all non-zero noise measurement values for locus l. Then the test statistic is given by

$$G = 2 \sum_{l=1}^{L} \sum_{k=1}^{K} o_{lk} \ln \left(\frac{o_{lk}}{e_{lk}} \right),$$

where $o_{lk} = |\{i = 1, \ldots, |\mathcal{N}_l^+| : n_{l,i}^+ \text{ is in bin } k\}|$ is the observed frequency in row l and column k, i.e., the number of non-zero noise measurement values from locus l whose height falls in bin k, and

$$e_{lk} = \frac{(\sum_{m=1}^{L} o_{mk})(\sum_{n=1}^{K} o_{ln})}{\sum_{m=1}^{L} \sum_{n=1}^{K} o_{mn}}$$

is, under the null hypothesis, the maximum-likelihood estimate of the expected number of observations in row l and column k.

The p-value is computed from a chi-squared distribution χ_d^2 with $d = (L-1)(K-1)$ degrees of freedom, according to

$$p - \text{value} = 1 - \chi_{(L-1)(K-1)}^2 (G).$$

Appendix C. Holm–Bonferroni correction

Let M denote the number of hypotheses in the compound test. We order the p-values in increasing order. Let $p_1, \ldots, p_M$ be the ordered p-values and $H_1, \ldots, H_M$ be the corresponding hypotheses. For a given significance level $\alpha = 0.05$, let k be the minimal index such that $p_k > \alpha/(M + 1 - k)$. If $k = 1$ then we reject none of the hypotheses. If no such k exists, then we reject all of the hypotheses. Otherwise, we reject the hypotheses $H_1, \ldots, H_{k-1}$ and do not reject $H_k, \ldots, H_M$.

Appendix D. Goodness of fit G-test

In order to apply the G-test, we bin the positive real numbers so that the observed frequency in each bin is larger than or equal to 5. We use the same binning procedure that was used in Section 3.2, and which is described in Appendix A, with the difference that here, we only have one empirical distribution. Let $(\underline{x}_1, \bar{x}_1], (\underline{x}_2, \bar{x}_2], \ldots, (\underline{x}_K, \infty)$ be the bins that are produced by this procedure,

where K denotes the number of bins. For notational convenience, we set $\bar{x}_K = \infty$.

Let $F_{\hat{v},\hat{\tau}}$ be the distribution of a log-normal distributed random variable as given in (2) and $F_{\hat{v},\hat{\tau}}^{\mathrm{q}}$ the distribution of a quantised log-normal distributed random variable as given in (3), where the parameters $\hat{v}$ and $\hat{\tau}$ are the maximum-likelihood estimates. Further, let F^{E} be the empirical distribution of the non-zero quantised noise measurement values, as defined in (5). Then, the G-test statistic is given by

$$G = 2\sum_{k=1}^{K} o_k \ln\left(\frac{o_k}{t_k}\right),$$

where $o_k = |\mathcal{N}^+|(F^{\mathrm{E}}(\bar{x}_k) - F^{\mathrm{E}}(\underline{x}_k))$ denotes the observed frequency and $t_k = |\mathcal{N}^+|(F_{\hat{v},\hat{\tau}}^{\mathrm{q}}(\bar{x}_k) - F_{\hat{v},\hat{\tau}}^{\mathrm{q}}(\underline{x}_k)) = |\mathcal{N}^+|(F_{\hat{v},\hat{\tau}}(\bar{x}_k) - F_{\hat{v},\hat{\tau}}(\underline{x}_k))$ the expected number of observations in the k-th bin under the null hypothesis. The last equality is due to the fact that our bins are of the form $(k-1/2, l+1/2]$, $k, l \in \mathbb{N}$, $k \leq l$, which is compatible with the quantisation operator (1). The p-value is computed from a chi-squared distribution χ_d^2 with $d = K - 3$ degrees of freedom, reduced from $K - 1$ to $K - 3$, as we have estimated two parameters from the data [33,34], according to

$$p - \text{value} = 1 - \chi_{K-3}^2(G).$$

References

[1] J.M. Butler, Forensic DNA Typing, 2nd ed., Elsevier Academic Press, 2005.
[2] P. Gill, R. Puch-Solis, J. Curran, The low-template-DNA (stochastic) threshold – its determination relative to risk analysis for national DNA databases, Forensic Sci. Int.: Genet. 3 (2) (2009) 104–111, http://dx.doi.org/10.1016/j.fsigen.2008.11.009.
[3] B. Budowle, A.J. Onorato, T.F. Callaghan, A.D. Manna, A.M. Gross, R.A. Guerrieri, J.C. Luttman, D.L. McClure, Mixture interpretation: defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework, J. Forensic Sci. 54 (4) (2009) 810–821, http://dx.doi.org/10.1111/j.1556-4029.2009.01046.x.
[4] R. Puch-Solis, A. Kirkham, P. Gill, J. Read, S. Watson, D. Drew, Practical determination of the low template DNA threshold, Forensic Sci. Int.: Genet. 5 (5) (2011) 422–427, http://dx.doi.org/10.1016/j.fsigen.2010.09.0015.
[5] 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 (6) (2012) 723–728, http://dx.doi.org/10.1016/j.fsigen.2012.06.012.
[6] 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 (1) (2013) 120–129, http://dx.doi.org/10.1111/1556-4029.12008.
[7] Scientific Working Group on DNA Analysis Methods, SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories, http://swgdam.org/Interpretation_Guidelines_January_2010.pdf (accessed September 2014).
[8] D. Taylor, J. Buckleton, Do low template DNA profiles have useful quantitative data? Forensic Sci. Int.: Genet. 16 (2015) 13–16, http://dx.doi.org/10.1016/j.fsigen.2014.11.001.
[9] P. Gill, R. Sparkes, R. Pinchin, T. Clayton, J. Whitaker, J. Buckleton, Interpreting simple STR mixtures using allele peak areas, Forensic Sci. Int. 91 (1) (1998) 41–53, http://dx.doi.org/10.1016/S0379-0738(97)00174-6.

[10] R.G. Cowell, S.L. Lauritzen, J. Mortera, Identification and separation of DNA mixtures using peak area information, Forensic Sci. Int. 166 (1) (2007) 28–34, http://dx.doi.org/10.1016/j.forsciint.2006.03.021.
[11] J.M. Curran, A MCMC method for resolving two person mixtures, Sci. Justice 48 (4) (2008) 168–177, http://dx.doi.org/10.1016/j.scijus.2007.09.014.
[12] T. Tvedebrink, P.S. Eriksen, H.S. Mogensen, N. Morling, Identifying contributors of DNA mixtures by means of quantitative information of STR typing, J. Comput. Biol. 19 (7) (2012) 887–902, http://dx.doi.org/10.1089/cmb.2010.0055.
[13] D. Taylor, J.-A. Bright, J. Buckleton, The interpretation of single source and mixed DNA profiles, Forensic Sci. Int.: Genet. 7 (5) (2013) 516–528, http://dx.doi.org/10.1016/j.fsigen.2013.05.011.
[14] J.-A. Bright, D. Taylor, J.M. Curran, J.S. Buckleton, Developing allelic and stutter peak height models for a continuous method of DNA interpretation, Forensic Sci. Int.: Genet. 7 (2) (2013) 296–304, http://dx.doi.org/10.1016/j.fsigen.2012.11.013.
[15] M.W. Perlin, A. Sinelnikov, An information gap in DNA evidence interpretation, PLoS ONE 4 (12) (2009) e8327, http://dx.doi.org/10.1371/journal.pone.0008327.
[16] M.W. Perlin, M.M. Legler, C.E. Spencer, J.L. Smith, W.P. Allan, J.L. Belrose, B.W. Duceman, Validating TrueAllele® DNA mixture interpretation, J. Forensic Sci. 56 (6) (2011) 1430–1447, http://dx.doi.org/10.1111/j.1556-4029.2011.01859.x.
[17] H. Swaminathan, C.M. Grgicak, M. Médard, D.S. Lun, NOCIt: A computational method to infer the number of contributors to DNA samples analyzed by STR genotyping, Forensic Sci. Int.: Genet. 16 (2015) 172–180, http://dx.doi.org/10.1016/j.fsigen.2014.11.010.
[18] J.R. Gilder, T.E. Doom, K. Inman, D.E. Krane, Run-specific limits of detection and quantitation for STR-based DNA testing, J. Forensic Sci. 52 (1) (2007) 97–101, http://dx.doi.org/10.1111/j.1556-4029.2006.00318.x.
[19] O. Hansson, P. Gill, T. Egeland, STR-validator: an open source platform for validation and process control, Forensic Sci. Int.: Genet. 13 (0) (2014) 154–166, http://dx.doi.org/10.1016/j.fsigen.2014.07.009.
[20] M.W. Perlin, G. Lancia, S.-K. Ng, Toward fully automated genotyping: genotyping microsatellite markers by deconvolution, Am. J. Hum. Genet. 57 (5) (1995) 1199–1210.
[21] Applied Biosystems, AmpFlSTR Identifiler Plus PCR amplification kit user guide, https://tools.lifetechnologies.com/content/sfs/manuals/cms_076395.pdf (accessed January 2015).
[22] Promega Corporation, PowerPlex 16 HS System, https://www.promega.com//media/files/resources/protocols/technicalanuals/101/powerplex.
[23] A.J. Gibb, A.-L. Huell, M.C. Simmons, R.M. Brown, Characterisation of forward stutter in the AmpFlSTR SGM Plus PCR, Sci. Justice 49 (1) (2009) 24–31, http://dx.doi.org/10.1016/j.scijus.2008.05.002.
[24] J.M. Butler, Advanced Topics in Forensic DNA Typing: Interpretation, Academic Press, 2014.
[25] J.H. McDonald, Handbook of Biological Statistics, 3rd ed., Sparky House Publishing, 2014.
[26] S. Holm, A simple sequentially rejective multiple test procedure, Scand. J. Stat. 6 (2) (1979) 65–70, http://www.jstor.org/stable/4615733.
[27] E. Limpert, W.A. Stahel, M. Abbt, Log-normal distributions across the sciences: keys and clues, BioScience 51 (5) (2001) 341–352, http://dx.doi.org/10.1641/0006-3568(2001)051[0341:LNDATS]2.0.CO;2.
[28] E. Jones, T. Oliphant, P. Peterson, et al., SciPy: Open Source Scientific Tools for Python, http://www.scipy.org/ (accessed February 2015).
[29] L.A. Currie, Limits for qualitative detection and quantitative determination. application to radiochemistry, Anal. Chem. 40 (3) (1968) 586–593, http://dx.doi.org/10.1021/ac60259a007.
[30] H. Kaiser, Report for analytical chemists. II. Quantitation in elemental analysis, Anal. Chem. 42 (4) (1970) 26A–59A, http://dx.doi.org/10.1021/ac60286a027.
[31] L.A. Currie, The measurement of environmental levels of rare gas nuclides and the treatment of very low-level counting data, IEEE Trans. Nucl. Sci. 19 (1) (1972) 119–126, http://dx.doi.org/10.1109/TNS.1972.4326496.
[32] L.A. Currie, Detection and quantification limits: origins and historical overview, Anal. Chim. Acta 391 (2) (1999) 127–134, http://dx.doi.org/10.1016/S0003-2670(99)00105-1.
[33] H. Cramér, Mathematical Methods of Statistics, Princeton University Press, Princeton, NJ, 1946.
[34] W.G. Cochran, The χ^2 test of goodness of fit, Ann. Math. Stat. 23 (3) (1952) 315–345, http://dx.doi.org/10.1214/aoms/1177729380.