

EuroForMix: An open source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts

Abstract

We have released a software named *EuroForMix* to analyse STR-DNA profiles in a user-friendly graphical user interface. The software implements a model to explain the allelic peak height on a continuous scale in order to carry out weight-of-evidence calculations for profiles which could be from a mixture of contributors. Through a properly parameterized model we are able to do inference on mixture proportions, the peak height properties, stutter proportion and degradation. In addition, *EuroForMix* includes models for allele drop-out, allele drop-in and sub-population structure. *EuroForMix* supports two inference approaches for likelihood ratio calculations. The first approach uses maximum likelihood estimation of the unknown parameters. The second approach is Bayesian based which requires prior distributions to be specified for the parameters involved. The user may specify any number of known and unknown contributors in the model, however we find that there is a practical computing time limit which restricts the model to a maximum of four unknown contributors.

EuroForMix is the first freely open source, continuous model (accommodating peak height, stutter, drop-in, drop-out, population substructure and degradation), to be reported in the literature. It therefore serves an important purpose to act as an unrestricted platform to compare different solutions that are available. The implementation of the continuous model used in the software showed close to identical results to the R-package *DNAmixtures*, which requires a HUGIN Expert license to be used. An additional feature in *EuroForMix* is the ability for the user to adapt the Bayesian inference framework by incorporating their own prior information.

1 Introduction

Since the inception of forensic DNA profiling, the interpretation of the crime-scene evidence has been challenging. The technology of polymerase chain reaction (PCR) has made it possible to amplify small amounts of cell material

in order to quantify the amount of DNA represented at short tandem repeats (STR). An ideal amplification would perfectly quantify the amount of DNA from each contributor(s) inherent to a given sample. In reality, the amplification processes is a series of stochastic events which varies the allelic and stutter peak heights that contribute to the profile.

The method for quantifying the weight-of-evidence that a candidate S is a contributor to the trace sample is based on specifying the proposition set

$$\begin{aligned} H_p &: \text{“Individual } S \text{ contributes to the trace sample”} \\ H_d &: \text{“Individual } S \text{ does not contribute to the trace sample”} \end{aligned} \quad (1)$$

where H_p is the prosecution hypothesis while H_d is the defense hypothesis. In the first step, the fit between the hypotheses and the trace samples are quantified by the corresponding likelihood functions. In the second step, the likelihood functions are compared through the likelihood ratio (LR) statistic. Often, trace samples are complicated and challenging to analyze due to artefacts and being comprised of a mixture of several contributors. To analyze such samples, there are two alternative kinds of mathematical models: a) semi-continuous and b) continuous. The continuous models keep the quantified intensities along with the qualitative DNA information, while the semi-continuous only uses the qualitative DNA information(i.e. the allele information only). However, at this moment, all of the available open source software are semi-continuous whereas all of the continuous software are commercial. The lack of availability of continuous models imposed by commercial constraints has inhibited comparative studies. Commercial continuous software include: *STRmix*[14], *TrueAllele*[11] and *DNAmixtures*[6]. Other continuous models are considered in Cowell et al. [4], Cowell et al. [5] and Puch-Solis et al. [13]. *DNAmixtures* adapts the “HUGIN Expert System”[10] to efficiently compute the likelihood function using a probabilistic expert system[9]. In order to eliminate unknown parameters, *DNAmixtures* maximizes the likelihood function using numerical restricted optimization routines. Note that a license of “HUGIN Expert System” is required to take full advantage of *DNAmixtures*. *STRmix* and *TrueAllele* are based on a Bayesian approach through specifying prior distributions on the unknown model parameters. They use Markov Chain Monte Carlo (MCMC) methods to calculate marginalized likelihood expressions by simultaneously sampling over the discrete set of genotypes for the unknown contributors specified in the model, and the unknown parameters. Last, notice that some of the software requires independent calibration data in order to specify some of the parameters in the model.

We have developed a new user-friendly software in a graphical user interface, *EuroForMix*, which is included in the R-package *euroformix* and is freely accessible at the site www.euroformix.com. *EuroForMix* implements an extended version of the model proposed by Cowell et al. [6], which assumes the peak heights to be gamma distributed with mixture proportions, stutter proportion and peak height mean and variation as unknown parameters. The extension presented here includes models for allele drop-in, degradation and sub-population structure. Similar to the method used by *DNAmixtures*, *EuroForMix* contains a maximum likelihood approach for handling unknown parameters. In addition,

EuroForMix also includes a Bayesian framework. In both cases unknown DNA profiles are treated as latent variables and are summed out in the likelihood calculations. Our method differs from *STRmix* and *TrueAllele* in that we compute the marginalized likelihood expressions using exact methods without any need for MCMC sampling. However, as an additional optional tool available within *EuroForMix*, the posterior distributions of the unknown parameters can be efficiently explored using MCMC sampling over the parameter space. Table 1 compares the properties of different software.

Property	<i>STRmix</i>	<i>TrueAllele</i>	<i>DNAmixtures</i>	<i>EuroForMix</i>
Calibration	Required	Required	Not possible	Possible
GUI	Yes	Yes	No	Yes
Drop-in	Yes	Yes	No	Yes
Stutter	Yes	Yes	Yes	Yes
Degradation	Yes	Yes	No	Yes
Coancestry coefficient	Yes	Yes	No	Yes
Peak height distribution	Log-normal	Truncated normal	Gamma	Gamma
Inference approach	Bayes	Bayes	ML	ML Bayes

Table 1: The table compares properties between existing software using the continuous model. “Calibration” refers to incorporation of validation data from the user laboratory to specify some of the parameters involved. “GUI” refers to whether the software is presented through a graphical user interface. “Drop-in”, “stutter” and “degradation” refer to whether such properties are modeled. “Coancestry coefficient” refers to whether the model implements F_{st} . “Peak height distribution” refers to the assumed distribution family for the allelic peak heights. “ML” refers to the maximum likelihood approach, while “Bayes” refers to the Bayesian approach.

The focus of this article is to demonstrate the use of *EuroForMix*, and to illustrate how the unknown parameters in the continuous model are effectively estimated, either through a maximum likelihood or through a Bayesian approach. *EuroForMix* also has the option of using calibration data for fixing parameters or using informative priors, but this is a decision for the user to make.

In the next section we illustrate the functionality of the software and introduce guidelines to carry out different types of interpretation and validation using a variety of probabilistic methods. In section 3 we show how these are applied to a real case example. In section A in Supplementary material we provided a simulation study to show that the results obtained from a special case of the continuous model in *EuroForMix* are consistent with those obtained from *DNAmixtures*. The demonstration that consistency is obtained between two independently written programs is an important feature of the validation. In section E in Supplementary material we provided another simulation study to show the practical limitations of *EuroForMix*.

2 Interpretation using *EuroForMix*

2.1 Data

The trace samples are short tandem repeats (STR) allele data with corresponding peak height intensities measured in relative fluorescence units (RFU). The allelic peak heights are proportionate to the quantities that originate from the contributing individuals. An individual has either a homozygous genotype (two identical alleles) or a heterozygous genotype (two different alleles). For an ideal PCR-process, the peak heights belonging to a heterozygous genotype are expected to be equally large, while the peak height of a homozygous genotype is expected to be twice as large as the alleles from a heterozygous genotype. In reality, these peak heights are stochastic and will vary between PCR-processes. A detection threshold is usually introduced to remove as much background “noise” as possible without removing alleles that truly originate from contributors. If the amount of DNA is small or influenced by degradation, then an allele may fall below the detection threshold. This event is expressed as allele drop-out. A stutter effect occurs when the DNA originally expressed as allele a loses one tandem repeat in the PCR-amplification process which then is added to the peak height intensity at allele position $a - 1$.

2.1.1 Data import and view in *EuroForMix*

The user interface of *EuroForMix* makes it easy to import, view and edit imported trace samples and reference profiles, which may be exported from software such as GeneMapper® and Peak Scanner™. Before proceeding with more advanced evaluations, the user may at any time view imported trace sample data in an electropherogram (epg) like format, along with the imported and selected reference profiles. This makes it possible to directly compare presence/absence of the alleles in the reference profile and the trace sample. Exploring the alleles in the genotypes with their corresponding peak heights is always an important consideration to do before carrying out the statistical analysis.

2.1.2 An illustration of complex trace sample data

Figure 1 shows a subset of two real trace samples, where the allele information and corresponding peak heights are presented for four markers. The upper panel in Figure 1 shows the first type of trace sample where we have labeled the alleles for reference profile of individual D1 as “1”. The lower panel in Figure 1 shows another trace where we labeled the alleles for reference profiles of the individuals P1 and P2 as “1” and “2”, respectively. In the upper panel we observe that allele 17 from the reference at marker D18S51 is absent in the trace sample. We also observe that allele 16 in marker D3S1358 is present in the trace sample but not in the reference. In the lower panel, allele 35 in marker D21S11 and alleles 11 and 14 in marker D18S51, are all present in reference P2, but absent in the trace sample. Allele 14 which is present in marker D3S1358 is not present in any of the references. The former phenomena can be explained as an allele drop-out, typically observed with low level DNA, while the latter phenomenon can be attributed to a stutter effect or a drop-in effect (given that the presented

references are the only true contributors). The illustration shows why trace samples with small amounts of DNA are difficult to analyze.

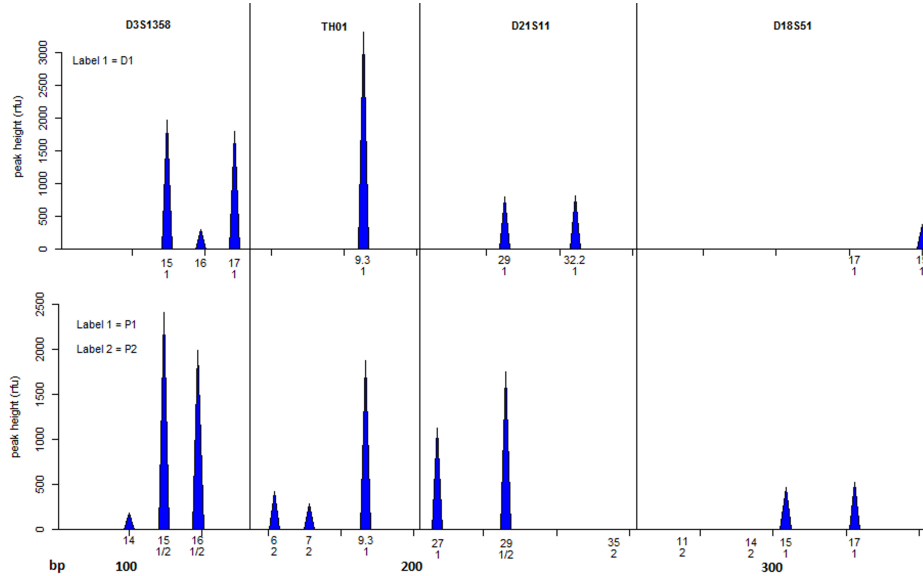


Figure 1: The upper panel shows a trace sample where the reference D1 is labeled as “1”. The lower panel shows another trace sample, where the references P1 and P2 are labeled as “1” and “2”, respectively along the horizontal axis. The labels above the peak heights are the marker names. The horizontal axis is fragment length given in number of base pairs (bp).

2.2 Theoretical considerations

Our aim is to quantify the weight-of-evidence that a candidate S is a contributor to a trace sample E . In order to do this we define the proposition set as specified in equation (1), where H_p is the prosecution hypothesis and H_d is the defense hypothesis.

In order to link a hypothesis H to the observed data, we specify a probabilistic model which gives the probability of observing sample E given that hypothesis H is true, $P(E|H)$. One way to compare the two rivaling hypotheses H_p and H_d is through the likelihood ratio (LR)

$$LR = \frac{P(E|H_p)}{P(E|H_d)}. \quad (2)$$

This quantity will evaluate how many times more likely it is to observe the sample given that H_p is true compared to the alternative that H_d is true.

As shown in Figure 1, there is a possibility that multiple individuals contribute to the same trace sample. If we believe that there are two contributors, the probabilistic model can be modified by adding an extra unknown individual.

2.2.1 The probabilistic model for peak heights

Let m be one of the M markers of the trace sample. The allele outcome vector \mathbf{A}_m is defined to be all alleles observed within a database, plus new alleles which are discovered when new individuals are typed. The peak heights at marker m is given through the vector \mathbf{Y}_m which quantifies intensities of the alleles given in \mathbf{A}_m . Also let $g_{m,k} = (g_{m,k,1}, g_{m,k,2})$ be the possibly unknown alleles for contributor k , out of K contributors, at marker m . We define the proportional amount of DNA from contributor k to allele a considered at marker m as

$$\alpha_{m,a}(g_{m,k}) = \pi_k \sum_{a' \in g_{m,k}} \mathbb{I}(a' = a) \quad (3)$$

where $\mathbb{I}(x)$ is the indicator function which is one if x is true and zero otherwise, and the mixture proportion parameter $\pi_k \in [0, 1]$ is the proportion of total DNA amount from contributor k with restriction $\sum_{k=1}^K \pi_k = 1$.

Assumption 1: The peak height contribution $Y_{m,a,k}$ from contributor k at allele a for marker m follows a gamma distribution where the expected contribution and variation is linearly dependent on the amount of k 's DNA. More specific, we assume $E[Y_{m,a,k}|g_{m,k}, \boldsymbol{\theta}] = \mu \alpha_{m,a}(g_{m,k})$ and $Var[Y_{m,a,k}|g_{m,k}, \boldsymbol{\theta}] = (\mu\sigma)^2 \alpha_{m,a}(g_{m,k})$ to be the model parametrization with parameters given as $\boldsymbol{\theta}$.

From assumption 1 it follows that the contributing peak height for contributor k at allele a for marker m is modelled as

$$p(Y_{m,a,k}|g_{m,k}, \boldsymbol{\theta}) = \text{Gamma}(\sigma^{-2} \alpha_{m,a}(g_{m,k}), \mu\sigma^2) \quad (4)$$

where the first and second argument is the shape and scale parameter respectively in the gamma distribution so that $\mu = E[Y_{m,a,k}|\alpha_{m,a} = 1]$ and $\sigma = \sqrt{Var[Y_{m,a,k}|\alpha_{m,a} = 1]}\mu^{-1}$ are parameters that directly interpret the expectation and coefficient of variation of a full heterozygote contributing allele peak height.

Assumption 2: The contributions from different contributors are independent.

By adding together peak height contributions over the K contributors for allele a at marker m , assumption 2 enables us to express the peak height density as

$$p(Y_{m,a}|\mathbf{g}_m, \boldsymbol{\theta}) = \text{Gamma}\left(\sigma^{-2} \sum_{k=1}^K \alpha_{m,a}(g_{m,k}), \mu\sigma^2\right) \quad (5)$$

where $\mathbf{g}_m = (g_{m,1}, \dots, g_{m,K})$ is the set of genotypes for the K contributors.

2.2.2 Drop-out model

To remove noise and artificial peak heights from the amplification process, a detection threshold T is specified so that alleles with corresponding peak heights below this threshold are removed. When the amount of DNA from a contributor is small, the corresponding peak heights also typically tend to be small and

may even be below the threshold. We define $Y_{m,a}^* = Y_{m,a}$ if $Y_{m,a} \geq T$ and $Y_{m,a}^* = 0$ otherwise. Typically $Y_{m,a}^*$ is the observed variable since a threshold is already applied on the original peak heights. For a marker m , the density of the truncated peak heights for allele a is given as

$$p(Y_{m,a}^* | \mathbf{g}_m, \boldsymbol{\theta}) = \begin{cases} p(Y_{m,a} | \mathbf{g}_m, \boldsymbol{\theta}) & \text{if } Y_{m,a} \geq T. \\ \int_0^T p(Y_{m,a} = x | \mathbf{g}_m, \boldsymbol{\theta}) dx & \text{if } Y_{m,a} = 0. \end{cases} \quad (6)$$

Notice that the drop-out probability for an allele is directly related to the probability of observing a peak height between zero and T given the conditional contributors and model parameters.

2.2.3 The likelihood function

Assumption 3: Genotypes are independent between markers.

Assumption 4: Peak heights are conditionally independent given the genotypes and model parameters.

For a given hypothesis H , assumption 3 assumes $\mathbf{g}_1, \dots, \mathbf{g}_M$ to be independent such that $p(\mathbf{g}_1, \dots, \mathbf{g}_M | H) = \prod_m p(\mathbf{g}_m | H)$. Assumption 4 allows us to take the product between alleles when conditioning on the genotypes and model parameters. With these assumptions, the likelihood function for the observed trace sample E , applied with a threshold T , can be calculated as

$$p(E | H, \boldsymbol{\theta}) = \prod_{m=1}^M \sum_{\substack{g_{m,k} \in \mathbb{Q}_m \\ k = 1, \dots, K}} \left(p(\mathbf{g}_m | H) \prod_{a \in \mathbf{A}_m} p(Y_{m,a}^* | \mathbf{g}_m, \boldsymbol{\theta}) \right) \quad (7)$$

where the set $\mathbb{Q}_m = \{(a, b) : a, b \in \mathbf{A}_m\}$ gives all possible genotype combinations at marker m . Note that some of the $p(\mathbf{g}_m | H)$ terms might be zero, or one depending whether H implies that some of the contributors are known.

2.2.4 Model for unknown genotypes

When the genotype of contributor k is unknown at marker m , prior information is used to specify $p(g_{m,k})$. A standard method is to assume Hardy Weinberg equilibrium (HWE) and use $p(g_{m,k}) = 2^{\mathbb{1}(a \neq b)} p_a p_b$, where p_a and p_b are estimated relative allele frequencies based on samples from the specified population.

An extension of the HWE assumption is to include the coancestry coefficient F_{st} , which we assume as a fixed value in order to take sub-population structure into account[7]. To take the sub-population structure into account, the probability of observing allele a , p_a , is updated to:

$$p'_a = \frac{u_a F_{st} + (1 - F_{st}) p_a}{1 + (v - 1) F_{st}} \quad (8)$$

where u_a denotes the number of times allele a has previously been typed and v denotes the number of previously typed alleles. By typed we mean that the allele has been observed for an individual, either within samples of the population or within the contributor(s) considered.

Assumption 5: The genotypes between contributors are independent given the coancestry coefficient.

For a given hypothesis H and marker m , assumption 5 assumes $g_{m,1}, \dots, g_{m,K}$ to be independent given F_{st} such that $p(\mathbf{g}_m|H, F_{st}) = \prod_{k=1}^K p(g_{m,k}|H, F_{st})$, which is used as model for $p(\mathbf{g}_m|H)$ in equation (7).

2.2.5 Stutter model

For a given marker m , the set of alleles given in a database, \mathbf{A}_m , is extended to include alleles $(a - 1)$ for all $a \in \mathbf{A}_m$, to assimilate potential stuttered alleles. For an allele $a \in \mathbf{A}_m$, let the peak height originating from the contributors, not being stutters, be given as $Y_{m,a}^C$, and the peak height originating by stutters from contributors at allele $a + 1 \in \mathbf{A}_m$ be given as $Y_{m,a}^S$. We can extend the continuous model given in equation (5) to accommodate stutters via the stutter proportion parameter $\xi = \frac{E[Y_{m,a}^S|\mathbf{g}_m, \theta]}{E[Y_{m,a}^S|\mathbf{g}_m, \theta] + E[Y_{m,a+1}^C|\mathbf{g}_m, \theta]}$ [8]. This parameter is defined to be the fraction of the expected contributing peak height at allele $a + 1$ which is moved to allele a . Consequently we let the distribution for $Y_{m,a}^C$ be given as $p(Y_{m,a}^C|\mathbf{g}_m, \theta) = \text{Gamma}(\sigma^{-2}(1 - \xi)\alpha_{m,a}(g_{m,k}), \mu\sigma^2)$, where $\alpha_{m,a}$ is the expression given in equation (3) and the distribution for $Y_{m,a}^S$ be given as $p(Y_{m,a}^S|\mathbf{g}_m, \theta) = \text{Gamma}(\sigma^{-2}\xi\alpha_{m,a+1}(g_{m,k}), \mu\sigma^2)$, provided that $a + 1 \in \mathbf{A}_m$.

Assumption 6: The peak heights $Y_{m,a}^C$ and $Y_{m,a}^S$ are independent given the genotypes and model parameters.

From assumption 6 we have that $Y_{m,a} = Y_{m,a}^C + Y_{m,a}^S$ is gamma distributed with the density given in equation (5) but with $\alpha_{m,a}$ modified to:

$$\alpha'_{m,a}(g_{m,k}) = (1 - \xi)\alpha_{m,a}(g_{m,k}) + \xi\mathbb{I}(a + 1 \in \mathbf{A}_m)\alpha_{m,a+1}(g_{m,k}) \quad (9)$$

2.2.5.1 Illustration of the stutter model

Consider marker D3S1358 in the upper panel of Figure 1. Here we observed the peak heights y_{15}^* , y_{16}^* and y_{17}^* . Assume $\mathbf{A} = (14, 15, 16, 17)$ to be the possible allele outcome in the population. If we condition on the genotype $g = \{15, 17\}$ assuming $K = 1$, then $\alpha'_{14}(g) = \xi$, $\alpha'_{15}(g) = (1 - \xi)$, $\alpha'_{16}(g) = \xi$ and $\alpha'_{17}(g) = (1 - \xi)$. This means that, conditioned on the model, allele 15 and 17 gives away some of its peak height as stutter to alleles 14 and 16 respectively, while the peak height of allele 16 is the stutter peak height received from allele 17. An explanation why allele 14 is not observed, is because stochastic effects have resulted in a stutter that is below the detection threshold (and therefore not recorded). Also consider the lower panel of Figure 1. If we condition on the two genotypes $(g_1, g_2) = (\{15, 15\}, \{15, 16\})$ assuming $K = 2$, it follows that $\alpha'_{14}(g) = \xi(2\pi_1 + \pi_2)$, $\alpha'_{15}(g) = 2(1 - \xi)\pi_1 + \pi_2$, $\alpha'_{16}(g) = (1 - \xi)\pi_2$ and $\alpha'_{17}(g) = 0$.

2.2.6 Degradation model

Consider both trace samples in Figure 1. Here, the summed peak heights for each marker clearly decreases as a function of the fragment length for both of the samples. This phenomenon is called degradation and is explained by DNA-strand breakage. The probability of degradation occurring within the shorter

DNA fragments is less than the probability of degradation occurring within the longer DNA fragments[16]. For some trace samples, the trend of degradation is very clear as alleles with high molecular weights (i.e. long fragment length) have smaller peak heights than alleles with low molecular weights. For such cases the fitted continuous model may not be adequate to explain the observed peak heights, unless a degradation module is included. We extend the continuous model to take into account the model introduced by Tvedebrink et al. [16] who demonstrated a degradation model with the parameter “probability of no breakage” per base of a given DNA fragment length.

We model the degradation by scaling $\alpha_{m,a}$ in equation (5) for allele a at marker m with the expression $\beta^{\frac{f_{m,a}-125}{100}}$ where β becomes an unknown slope parameter and $f_{m,a}$ is the fragment length in base pair (bp) of an allele a at marker m . A shift of fragment length 125 bp was selected as a baseline to make the parameter estimates of μ and σ explainable at fragment length equal 125 bp. The scaling of 100 makes the parameter β an exponential expression of the “probability of no breakage”, see Tvedebrink et al. [16].

2.2.7 Allele drop-in model

Allele drop-in may occur in the trace sample. This is where a DNA fragment from cellular material “falls” into the tube before or during the amplification process and causes a contaminant allele to be visualized[2]. Following Puch-Solis [12] and Taylor et al. [14], *EuroForMix* can also accommodate allele drop-in. We assume allele drop-in to be all alleles with corresponding peak heights above the detection threshold which are not explained by conditioned contributors or stutters. The probability of the event of having allele a as a drop-in at a particular marker is given as $C \times p_a$ where C is the probability for a drop-in event for a particular marker and p_a is the allele frequency for allele a . We assume that the drop in peak height Y_a is exponential distributed $Exp(\lambda) = h(Y_a|\lambda)$, where $\lambda > 0$ is a specified rate parameter. For λ strictly positive, the drop-in model ensures that smaller peak heights are always more likely to occur than larger peak heights for allele drop-in. When the detection threshold T is used to truncate the peak heights, the drop in peak height is distributed as $Y_a^* = Y_a|Y_a \geq T \sim T + Exp(\lambda)$ such that $Y_a^* - T \sim Exp(\lambda)$. Hence we specify the drop-in model as

$$p(Y_a^* | a \text{ is a drop-in}, \lambda) = Cp_a h(Y_a^* - T | \lambda). \quad (10)$$

From this, the extension of the probabilistic model given in equation (7) will be as follows: $p(Y_{m,a}^* | \mathbf{g}, \boldsymbol{\theta})$ is exchanged entirely with the drop-in model in equation (10) if allele a is a drop-in. For the situation of no allele drop-in, the product over all alleles is scaled with $(1 - C)$.

2.3 Inference of the continuous model

After the model specification, we now return to the likelihood ratio (LR) expression given in equation (2), where we are interested in calculating $P(E|H_p)$ and $P(E|H_d)$. Given the observed trace sample, the probabilistic model given in equation (7) is provided as a likelihood function depending on the model parameters, $\boldsymbol{\theta}_p$ under hypothesis $H = H_p$, and $\boldsymbol{\theta}_d$ under hypothesis $H = H_d$. In

this subsection we will present methods to accomodate the parameters θ_p and θ_d in order to calculate a value for LR . Table 2 shows a summary of the model parameters, which are either assumed known or unknown. We present two inference methods, a frequentist and a Bayesian inference approach to calculate the likelihood ratio value given in equation (2).

Parameter	Prior	User specified	Default
K (number of contributors)	fixed	$K \in \{1, \dots\}$	$K = 2$
μ (expected peak height)	$Uniform[0, \mu_1]$	$\mu_1 > 0$	$\mu_1 = 20000$
σ (cv of peak heights)	$Uniform[0, \sigma_1]$	$\sigma_1 > 0$	$\sigma_1 = 1$
\mathbf{m} (mixture proportion)	$Dirichlet(K, \mathbf{1})$	Cannot change	
ξ (stutter proportion)	$beta$ -distribution	unknown or fixed $\xi \in [0, 1]$	$beta(1, 1)$
β (degradation slope)	$Uniform(0, 1)$	unknown or fixed $\beta = 1$	$\beta = 1$
C (drop-in probability)	$Uniform(0, C_1)$	$C_1 \in (0, 1]$ or fixed $C \in [0, 1]$	$C = 0$
λ (parameter in drop-in model)	fixed	$\lambda > 0$	$\lambda = 0.01$
F_{st} (coancestry coefficient)	fixed	$F_{st} \in [0, 1]$	$F_{st} = 0$

Table 2: The table explains the model parameters and shows how prior information can be specified by the user. μ is the expectation of heterozygous peak heights, σ is the coefficient of variation of heterozygote peak heights. Any unknown parameters are elements in the model parameter set θ .

To reduce the calculation time of the likelihood function, we followed the method carried out by Taylor et al. [14] and Puch-Solis et al. [13]. An approximation to the likelihood function is achieved by grouping together alleles as a compound allele. Details are given in section B in the Supplementary material.

2.3.1 Frequentist inference

The frequentist inference is based upon maximizing the likelihood function with respect to the unknown parameters θ in order to obtain the maximum likelihood estimate

$$\hat{\theta} = \arg \max_{\theta} p(E|H, \theta). \quad (11)$$

Doing so under each of the hypotheses we obtain:

$$LR_F = \frac{p(E|H_p, \hat{\theta}_p)}{p(E|H_d, \hat{\theta}_d)} \quad (12)$$

which is the likelihood ratio quantity. Notice that this approach chooses the parameters which provide the best fit to the observed peak heights. Hence the uncertainty of the parameters are not taken into account in this likelihood ratio quantity.

2.3.2 Bayesian inference

The Bayesian inference approach requires a specification of prior density functions on the model parameters, given the considered hypothesis H . Table 2 shows possible specifications for the priors of the individual parameters in *EuroForMix*, which we assume a priori to be independent of each other. It follows

from the Bayesian framework that

$$p_B(E|H) = \int p(E|H, \boldsymbol{\theta})p(\boldsymbol{\theta}|H)d\boldsymbol{\theta} \quad (13)$$

which is the integral over the model parameters. Doing this under each of the hypotheses provides

$$LR_B = \frac{p_B(E|H_p)}{p_B(E|H_d)} \quad (14)$$

which is the Bayes factor used as a value for the likelihood ratio quantity in a Bayesian setting. We used the R-package *cubature*[3] which makes it possible to provide efficient multivariate integration over the model parameter space. This method differs from the MCMC approach in that the method is numerical, hence it is possible to reflect the accuracy of LR_B through δ , the relative error for an integral. An estimated error-interval of LR_B is provided in brackets as

$$\left[\widetilde{LR}_B \frac{(1 - \delta)}{(1 + \delta)}, \widetilde{LR}_B \frac{(1 + \delta)}{(1 - \delta)} \right] \quad (15)$$

where \widetilde{LR}_B is the value obtained through the numerical integration and δ is specified by the user.

2.3.3 Deconvolution

Deconvolution is a technique to assign the genotypes of unknown contributors under a specific given hypothesis. Cowell et al. [6] and Tvedebrink et al. [15] considered a continuous model in order to search for the most likely genotypes. This technique is useful to identify genotypes from potential perpetrators of crime that have not been identified by other means. The assigned genotypes may be compared against national DNA databases to identify potential suspects[1]. The flexibility of the continuous model enables *EuroForMix* to accommodate allele drop-out, stutters, degradation and allele drop-in in the deconvolution procedure.

Let a suggested set of combined genotypes over all markers be given as $G = (\mathbf{g}_1, \dots, \mathbf{g}_M)$ and $p(E|H, \hat{\boldsymbol{\theta}})$ be the likelihood function from equation (7) inserted with the maximum likelihood estimates of the model parameters under hypothesis H with $\hat{\boldsymbol{\theta}}$ given in equation (11). The probability of G is then provided as

$$p(G|E, H, \hat{\boldsymbol{\theta}}) = \left(\prod_{m=1}^M p(\mathbf{g}_m|H) \prod_{a \in \mathbf{A}_m} p(Y_{m,a}^*|\mathbf{g}_m, \hat{\boldsymbol{\theta}}) \right) / p(E|H, \hat{\boldsymbol{\theta}}) \quad (16)$$

This is the frequentist approach where each marker can be independently calculated given the maximum likelihood estimate. *EuroForMix* produces a ranked table of the most probable marker-combined genotype profiles of the unknown individuals, from highest to lowest probabilities.

A Bayesian framework allows the calculation of the posterior probabilities unconditioned on the parameter $\boldsymbol{\theta}$, $p(G|E, H) = \int p(G|E, H, \boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$. This is not yet implemented, but will be in a future update.

2.4 Model selection

In this section we establish a framework to select a model to calculate the LR quantity in equation (2). We can either follow a frequentist framework using the maximum likelihood values or a Bayesian framework using marginal probabilities of the observations. Deciding whether to include (or not include) stutters, degradation and allele drop-in into the model falls under the model selection framework. We also consider the number of contributors K in the model as unknown. We perform the model selection under a setting H including both hypotheses, that is taking into account the uncertainty of whether S is a contributor or not.

The possible models discussed in section 2.2 enables comparisons of different combinations of the following situations:

- No stutter ($\xi = 0$) versus stutter ($\xi \in (0, 1)$).
- No degradation ($\beta = 1$) versus degradation ($\beta \in (0, 1)$).
- No allele drop-in ($C = 0$) versus allele drop-in ($C \in (0, 1)$).
- The number of contributing individuals K .

A variant of these combinations can be represented as model candidate \mathcal{M}_s and the likelihood notation in equation (7) is extended to be explicitly given as $p(E|H, \boldsymbol{\theta}, \mathcal{M}_s)$. When a drop-in model is considered (i.e. $C > 0$), the model for allele drop-in peak heights assumes $\lambda = \lambda_0$ fixed.

We consider a model search strategy starting with the simplest model. The complexity is increased by adding more parameters to the model. The search stops when it is unlikely that a more complicated model performs better than one already found.

2.4.1 A frequentist model selection approach

To choose the final model in the frequentist framework we use the Akaike information criterion (AIC), given as $AIC_s = -2\hat{l}_s + 2n_s$ for candidate model \mathcal{M}_s , where $\hat{l}_s = \max_{\boldsymbol{\theta}} \log p(E|H, \boldsymbol{\theta}, \mathcal{M}_s)$ is the maximum log-likelihood value and n_s is the number of elements in $\boldsymbol{\theta}$. This criterion favors the model candidates which fits the data best, but penalized for the number of parameters to avoid overparameterizing. The model candidate \mathcal{M}_s with smallest AIC is chosen as the final model, $\mathcal{M}_F = \arg \min_s AIC_s$.

2.4.2 A Bayesian model selection approach

For a selected model \mathcal{M}_s , the Bayesian framework requires specification of priors on the parameters $\boldsymbol{\theta}$ to allow calculation of the marginal probability of the trace sample, $p_B(E|H, \mathcal{M}_s)$, which is an extended notation of that given by equation (13). The final model using the Bayesian model criterion would be to select the model \mathcal{M}_s with highest marginal probability, $\hat{\mathcal{M}}_B = \arg \max_s p_B(E|H, \mathcal{M}_s)$.

So far we have considered the number of contributors K to the trace sample

as part of the model selection problem. In the Bayesian framework, K can be given a prior with probabilities $p(K = k)$ for $K \in \mathbb{K}$, where the set \mathbb{K} is chosen by the user. The extension then becomes $p_B(E|H) = \sum_{k \in \mathbb{K}} p_B(E|H, K = k)p(K = k|H)$, where we have extended the expression in equation (13) to make it explicitly conditioned on K .

2.5 Model evaluation

The last step in the model inference procedure is to check whether the fitted continuous model is reasonable for the peak heights above the detection threshold T . We perform model evaluation under a setting H including both hypotheses. With $\hat{\theta}$ as the maximized argument under H from equation (11), the conditional observed cumulative probability for each allele is defined as (Cowell et al. [6]):

$$P_{m,a}(y_{m,a}) = P(Y_{m,a} \leq y_{m,a} | H, \hat{\theta}, \mathbf{y}_{(m,-a)}, Y_{m,a} \geq T). \quad (17)$$

Here $y_{m,a}$ is the observed peak height for allele a at marker m and $\mathbf{y}_{(m,-a)}$ are the other observed alleles in the same marker which exceed T . *EuroForMix* provides a Probability-Probability (PP) plot between all the cumulative probabilities $\{P_{m,a}(y_{m,a})\}_{y_{m,a} \in \mathcal{Y}_m}^{m=1, \dots, M}$ and the standard uniform distribution to check whether the fitted continuous model is a reasonable assumption for the observed peak heights $\mathbf{y}_1, \dots, \mathbf{y}_M$. Deviation against this assumption indicates whether the continuous model should be changed or improved.

3 A practical real case example

3.1 Data

The following example originates from a DNA-transfer experiment where the individuals P1 and P2 are known beforehand to have left epithelial cells on a wooden item. The considered STR sample was provided by the laboratory of the Norwegian Institute of Public Health (NIPH). The sample was amplified using the PowerPlex ESX 17 System kit (Promega) with 17.5 μL template and the standard 30 cycle amplification protocol on a GeneAmp PCR System 9700 (Applied Biosystems). The sample was injected into the Applied Biosystems 3500xl Genetic Analyzer at 1.2 kV for 10 s. The result was analyzed in the GeneMapper® ID-X Software v.1.2 (Applied Biosystems) and the detection threshold for alleles was given as $T = 150$ RFU (note that the 3500xl instrument is much more sensitive compared to the 3130xl). The peak height and allele information for the sample is given in Table 3 and plotted in an epg-like format in Figure S1 in Supplementary material. The epg can be used as a reference to explore reasonable model assumptions.

3.2 Interpretation

The aim of this example was to quantify the weight-of-evidence that the P2 (known) candidate was a contributor to the trace sample, where we treated P1 as unknown (as this was a controlled experiment, the identity of all donors is known). Hence we specify the prosecution and defense hypotheses in the proposition set as in equation (1) where the candidate is P2. We used the

marker	allele	height	P1	P2	marker	allele	height	P1	P2
D3S1358	13	0	0	0	D22S1045	14	0	0	0
D3S1358	14	178	0	0	D22S1045	15	1017	1	2
D3S1358	15	2405	1	1	D22S1045	16	610	1	0
D3S1358	16	1982	1	1	VWA	13	0	0	0
TH01	5	0	0	0	VWA	14	1250	1	0
TH01	6	419	0	1	VWA	15	440	0	1
TH01	7	282	0	1	VWA	17	1232	1	1
TH01	8.3	0	0	0	D8S1179	9	0	0	0
TH01	9.3	1871	2	0	D8S1179	10	206	0	1
D21S11	26	0	0	0	D8S1179	12	0	0	0
D21S11	27	1128	1	0	D8S1179	13	352	0	1
D21S11	28	0	0	0	D8S1179	14	978	1	0
D21S11	29	1750	1	1	D8S1179	15	827	1	0
D21S11	34	0	0	0	FGA	20	0	0	0
D21S11	35	0	0	1	FGA	21	664	1	0
D18S51	10	0	0	0	FGA	22	714	1	1
D18S51	11	0	0	1	FGA	24	0	0	0
D18S51	14	0	0	1	FGA	25	0	0	1
D18S51	15	467	1	0	D2S441	8	0	0	0
D18S51	16	0	0	0	D2S441	9	200	0	0
D18S51	17	524	1	0	D2S441	10	3362	1	0
D10S1248	12	0	0	0	D2S441	11	1168	0	2
D10S1248	13	1856	1	2	D2S441	13	0	0	0
D10S1248	14	155	0	0	D2S441	14	3693	1	0
D10S1248	15	1045	1	0	D12S391	17	0	0	0
D1S1656	11	0	0	0	D12S391	18	297	0	1
D1S1656	12	1140	1	0	D12S391	17.3	0	0	0
D1S1656	14	0	0	0	D12S391	18.3	1446	1	0
D1S1656	15	601	0	1	D12S391	19	751	0	1
D1S1656	15.3	0	0	0	D12S391	20	0	0	0
D1S1656	16	488	0	1	D12S391	21	171	0	0
D1S1656	16.3	155	0	0	D12S391	22	1370	1	0
D1S1656	17.3	1877	1	0	D19S433	12	0	0	0
D2S1338	16	0	0	0	D19S433	13	1157	1	0
D2S1338	17	290	0	1	D19S433	14	781	0	2
D2S1338	18	0	0	0	D19S433	14.2	0	0	0
D2S1338	19	619	1	0	D19S433	15.2	922	1	0
D2S1338	20	259	0	1	SE33	26.2	0	0	0
D2S1338	22	0	0	0	SE33	27.2	0	0	1
D2S1338	23	649	1	0	SE33	28.2	0	0	0
D16S539	8	0	0	0	SE33	29.2	221	0	1
D16S539	9	217	0	1	SE33	30.2	473	1	0
D16S539	10	312	0	1	SE33	32.2	0	0	0
D16S539	11	743	1	0	SE33	33.2	570	1	0
D16S539	12	619	1	0					

Table 3: The table shows the trace sample considered in the example. Each row denotes an allele with corresponding peak height for a given marker m . The P1 and P2 columns denote the number of corresponding alleles per reference genotype that they have in their genotype (1 is a heterozygote contribution and 2 is a homozygote contribution). In this example we treat the profile of P1 as unknown.

Norwegian Population frequencies typed with ESX 17¹. If alleles in the trace sample or reference samples with corresponding stutter alleles were not recorded in the population, these alleles were inserted and given a frequency equal to the value of the smallest observed allele frequency in the dataset. For the Bayesian approach we required a relative error $\delta = 0.1$ to calculate the integrals.

Notice that a comparison of the alleles of the candidate P2 to the trace sample showed that there were five alleles from P2 not observed in the sample. Hence

¹The relative allele frequencies can be found in the textfile ESX17_Norway.csv under the installation folder of euroformix: 'R-folder'\library\euroformix\tutorialdata\FreqDatabases

the only explanation for P2 being a contributor (under H_p) was because of allele drop-out due to small amount of DNA being present.

3.2.1 Model selection

We followed the model search strategy as mentioned in section 2.4 and assumed no sub-population structure ($F_{st} = 0$). If a model with allele drop-in is considered, $\lambda = 0.01$ is fixed which means peak heights around 500 RFU are very unlikely to be drop in compared to smaller peak heights. Table 4 shows the results from the model selection where it is shown that the model with $K = 2$ contributors including stutter, degradation and no drop-in gave the final model for both the frequentist and Bayesian framework. The result from the model selection coincides with the inspection of markers with many observed alleles (i.e. D1S1656, D2S1338, D16S539 and D8S1179). Here it is clear that the sample is most likely to be a two-person mixture profile where the remaining alleles are explained as stutters.

K	stutt	deg	dropin	AIC	p_B
1	0	0	1	998.34	3.50e-215
1	0	1	1	983.81	1.71e-212
1	1	0	1	993.72	1.69e-214
1	1	1	1	982.53	1.04e-212
2	0	0	1	964.24	2.16e-207
2	0	1	1	953.54	3.49e-205
2	1	0	0	948.31	1.59e-204
2	1	1	0	934.69*	5.13e-202*
3	0	0	0	958.93	1.96e-206
3	0	1	0	947.54	4.94e-204
3	1	0	0	950.31	1.54e-209
3	1	1	0	936.21	2.92e-208

Table 4: The table shows the frequentist and Bayesian model selection scores based on AIC and marginal probability, respectively. The star denotes the final model which is the model with smallest AIC value (frequentist) or largest marginal probability (Bayesian). K is number of contributors, Stutt, deg and dropin are indicators for whether stutter, degradation or allele drop-in is assumed in the model. For the drop-in model $\lambda = 0.01$ was assumed fixed.

3.2.2 Weight-of-evidence assuming Hardy Weinberg Equilibrium

We used the final model from the model selection to evaluate the trace sample with the LR formula provided in equation (2). In this section we assumed no sub-population effect by fixing $F_{st} = 0$. Figure 2 shows the maximum likelihood based LR which gives $LR_F = 10^{10.45}$, clearly supporting H_p . From the figure we see that under H_p the mixture proportion of P2 was estimated as $\pi_1 = 0.23$, the expectation as $\mu = 1947$ and the coefficient of variation as $\sigma = 0.28$. With no degradation ($\beta = 1$), the model denoted the drop-out probability for a heterozygous allele of P2 as 0.084. In contrast, with the estimated degradation slope $\beta = 0.676$, the drop-out probability at fragment length 240 becomes 0.24 and further increased to 0.37 for fragment length 300. This means that it was likely that the unobserved alleles from P2 had dropped out since these alleles all have large fragment lengths.

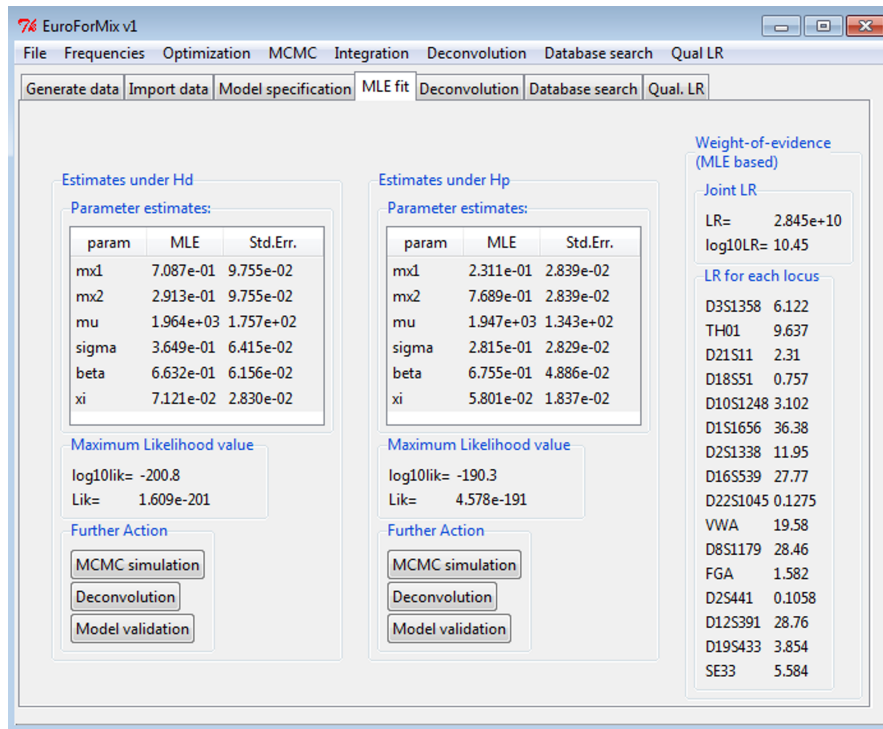


Figure 2: Screen-shot to show the results of the maximum likelihood estimation, based on the continuous model incorporating stutter (with stutter proportion ξ) and degradation (with degradation slope β), assuming $K = 2$ contributors. Under H_d , the mixture proportions $mx1$ and $mx2$ belong to the unknown contributors which are decreasingly sorted. Under H_p , $mx1$ belongs to P2 and $mx2$ belongs to an unknown contributor. Further analysis can be carried out under “Further action”- including model validation, deconvolution and MCMC simulation to get the posterior densities of the parameters.

Notice that the LR values in D22S1045 and D2S441 (Figure 2) are small since the observed peak heights of these two markers are smaller and greater, respectively, compared to what is expected (see Figure S2 in Supplementary material). We checked the performance of the fitted continuous gamma model by creating a Probability-Probability (PP) plot. The linear trend in Figure 3 showed that the assumed model followed the observed peak heights quite well. However, there seems to be more points below the line than above which indicated that some of the observed peak heights are lower than that assumed by the model. For the Bayesian approach the likelihood ratio becomes $LR_B = 10^{8.884} [10^{8.798}, 10^{8.969}]$. This is much smaller compared to the maximum likelihood approach, since the uncertainty of the model parameters and the non-symmetry in the likelihood are taken into account.

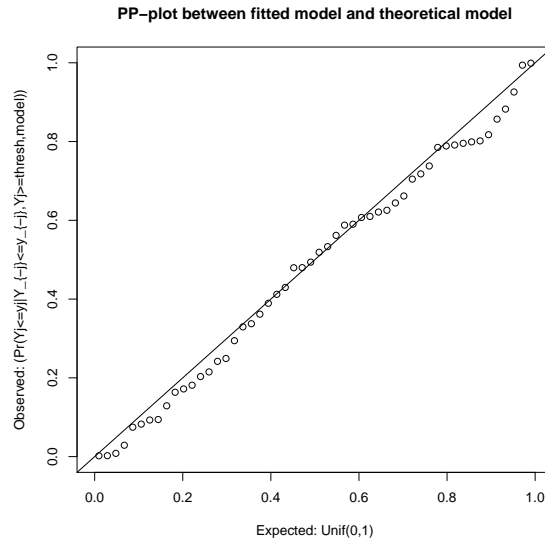


Figure 3: The plot shows the performance check of the maximum likelihood fitted model. The y-axis represents the observations which are inserted into the fitted model compared with the theoretical assumption which is given at the x-axis.

3.2.3 Weight-of-evidence assuming sub-population structure

We considered the situation where the contributors to the trace sample could belong to a sub-population of the Norwegian population. We considered the sub-population model from section 2.2.4 where the coancestry coefficient F_{st} was introduced to decide how divergent the allele frequencies in sub-populations may be from the allele-frequencies of the general Norwegian population. The likelihood ratio was calculated as a function of F_{st} for both the maximum likelihood approach and the Bayesian approach. Table 5 showed the reduction of the likelihood ratio as the value of F_{st} increased. Here it is seen that the weight-of-evidence clearly decreases, but that H_p is still clearly supported, even for large values of F_{st} .

F_{st}	0	0.01	0.02	0.03	0.04	0.05	0.06	0.08	0.1
$\log_{10}LR_F$	10.45	10.11	9.81	9.52	9.26	9.02	8.80	8.40	8.04
$\log_{10}LR_B$	8.88	8.54	8.22	7.94	7.69	7.40	7.15	6.84	6.50

Table 5: The table shows the resulting likelihood ratio quantities for the maximum likelihood approach LR_F and the Bayesian approach LR_B as a function of the coancestry coefficient F_{st} .

3.2.4 Deconvolution

In this subsection we will assume that the profile of individual P1 is known beforehand and that this individual is a contributor to the trace sample. We attempted to assign the genotype profile for the minor contributor in the trace

sample where we used the final model from the model selection which suggested two contributors. Hence P1 will be one of the contributors and the last will be an individual with an unknown genotype profile which was assigned using the frequentist method described in section 2.3.3. The resulting ranked assigned profile table (see Figure S3 in Supplementary material) showed that the probability of the top rank profile was $1.224e-05$ with the next ranking profiles approximately the same. These small differences in probability indicate that the full unknown joint genotype profile is difficult to resolve for some markers. Note here that allele 99 represents all non-present alleles (i.e. potential allele drop-outs). Comparing the top rank profile with the profile of individual P2, we found that 13 markers fully matched at both alleles and that 3 markers partially matched at one allele. We also correctly classified that allele drop-out (the missing allele cannot be identified of course) had occurred for both alleles at locus D18S51 and only one allele at D21S11, FGA and SE33. In comparison, an experienced reporting officer, who used expert opinion to blindly deconvolve the mixture, reported the markers D3S1358, D18S51 and FGA as inconclusive (could not assign any alleles) and the markers D21S11, D10S1248, D22S1045, vWA, D2S441 and SE33 as partially inconclusive (he could only assign one allele at each locus). The other markers were assigned correctly by him.

4 Summary and discussion

In this article it has been demonstrated with real data how the *EuroForMix* software can be used for DNA interpretation utilizing peak heights. In particular we have implemented an extended version of the continuous model as presented by Cowell et al. [6] to take into account allele drop-in, degradation and sub-population structure. The Bayesian inference approach was introduced as an alternative to the maximum likelihood approach, where its accuracy was quantified. The Bayesian framework enables the user to specify his/her own priors in order to take into account the uncertainty of the parameter.

The interpretation requires a probabilistic model which is automatically inferred using the observed trace sample. Such models are based on peak height information, which makes it possible to take into direct account the uncertainty of allele drop-out, allele drop-in, stutters and degradation. In order to select a probabilistic model, both a frequentist and a Bayesian model selection framework was introduced, where Akaike information criterion (AIC) was used for the former, and the integrated likelihood with respect to the model parameters was used for the latter. After a final model was selected, a model validity plot was applied in order to investigate the properties of the selected probabilistic model. Finally, the likelihood ratio (LR) value for a given candidate was calculated in order to investigate whether the data supported the hypothesis that he was a contributor to the trace sample or not.

In the real experiment example we considered a major-minor sample which was challenging since the candidate, the minor contributor, had five alleles dropped out. From the model selection, for both frameworks, the final model was a model with two contributors which included degradation and stutter. The LR values were investigated with the amount of sub-population structure through

the coancestry coefficient F_{st} . By considering the maximum likelihood approach we obtained values between $LR_F(F_{st} = 0) = 10^{10.45}$ and $LR_F(F_{st} = 0.1) = 10^{8.04}$. In comparison, for the Bayesian approach we obtained values between $LR_B(F_{st} = 0) = 10^{8.88}$ and $LR_B(F_{st} = 0.1) = 10^{6.50}$, which was about two orders of magnitudes lower than the maximum likelihood approach. The reason for this is that the Bayesian approach takes into account the uncertainty of the parameters which means it should be a more reliable framework than using the maximum likelihood approach. Finally, using the deconvolution method described, we assigned the full joint genotype profile of the minor contributor conditioned on the major profile as a “known” individual. Choosing the top ranked candidate from the method gave 13 full and 3 partial marker matches.

5 Acknowledgements

We would like to thank Ane Elida Fonnelløp for typing the trace sample used in the provided example and Oskar Hansson, Hinda Haned, Corina Benschop and Thore Egeland for useful comments. We also would like to thank the two anonymous reviewers for improving the paper. ØB and PG have received funding support from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement number 285487 (EUROFORGEN-NoE).

6 Supplementary material

Supplementary material can be found in the online version.

References

- [1] M. Bill, P. Gill, J. Curran, T. Clayton, R. Pinchin, M. Healy, and J. Buckleton. Pendulum-a guideline-based approach to the interpretation of str mixtures. *Forensic Science International*, 148:181–189, 2005.
- [2] J. Buckleton and P. Gill. Low copy number. In *Forensic DNA Evidence Interpretation*, CRC Press, chapter 8, pages 275–297. Boca Raton, Florida, 2005.
- [3] C code by Steven G. Johnson and R by Balasubramanian Narasimhan. *cubature: Adaptive multivariate integration over hypercubes*, 2013. URL <http://CRAN.R-project.org/package=cubature>.
- [4] R. G. Cowell, S. L. Lauritzen, and J. Mortera. A gamma model for DNA mixture analysis. *Bayesian Analysis*, 2(2):333–348, 2007.
- [5] R. G. Cowell, S. L. Lauritzen, and J. Mortera. Identification and separation of DNA mixtures using peak area information. *Forensic Science International*, 166:28–34, 2007.
- [6] R. G. Cowell, T. Graversen, S. L. Lauritzen, and J. Mortera. Analysis of forensic DNA mixtures with artefacts. *Appl. Statist.*, 64(1):1–32, 2015.

- [7] J. Curran, P. Gill, and R. Bill, M. Interpretation of repeat measurement DNA evidence allowing for multiple contributors and population substructure. *Forensic Science International*, 148:47–53, 2005.
- [8] P. Gill and J. Buckleton. Biological basis for DNA evidence. In *Forensic DNA Evidence Interpretation*, CRC Press, chapter 1, pages 21–23. Boca Raton, Florida, 2005.
- [9] T. Graversen and S. Lauritzen. Computational aspects of DNA mixture analysis. *Statistics and Computing*, 25(3):527–541, 2015. ISSN 0960-3174. doi: 10.1007/s11222-014-9451-7. URL <http://dx.doi.org/10.1007/s11222-014-9451-7>.
- [10] HUGIN Expert A/S. URL <http://www.hugin.com/>.
- [11] M. W. Perlin, M. M. Legler, C. E. Spencer, J. L. Smith, W. P. Allan, J. L. Belrose, and B. W. Duceman. Validating TrueAllele DNA Mixture Interpretation. *Journal of Forensic Sciences*, 56:1430–1447, 2011.
- [12] R. Puch-Solis. A dropin peak height model. *Forensic Sci. Int. Genet.*, (11):80–84, 2014.
- [13] R. Puch-Solis, L. Rodgers, A. Mazumder, S. Pope, I. W. Evett, J. Curran, and D. Balding. Evaluating forensic DNA profiles using peak heights, allowing for multiple donors, allelic dropout and stutters. *Forensic Science International: Genetics*, 7:555–563, 2013.
- [14] D. Taylor, J. A. Bright, and J. Buckleton. The interpretation of single source and mixed DNA profiles. *Forensic Science International: Genetics*, 7:516–528, 2013.
- [15] T. Tvedebrink, P. S. Eriksen, H. S. Mogensen, and N. Morling. Identifying Contributors of DNA Mixtures by Means of Quantitative Information of STR Typing. *Journal of Computational Biology*, 18:887–902, 2011.
- [16] T. Tvedebrink, P. S. Eriksen, H. S. Mogensen, and N. Morling. Statistical model for degraded DNA samples and adjusted probabilities for allelic drop-out. *Forensic Science International: Genetics Supplement Series*, 3: 489–491, 2011.

Supplementary material

A Comparison with *DNAmixtures*

In this section we will compare the likelihood values between *EuroForMix* and *DNAmixtures* by randomly generating single source stains and two- and three-person mixtures. Note that in order to get the same results as *DNAmixtures* we let alleles with the same whole integer (e.g. allele 30, 30.2, 30.3 become allele 30) become one allele (summing their relative frequencies). Also, no degradation and sub-population structure are assumed ($\beta = 1$, $F_{st} = 0$). We used the Norwegian ESX 17 population frequencies, considering 16 markers, where new observed alleles were assigned as the smallest observed frequency. For a given number of contributors K , three random crime samples were simulated from the continuous model with parameters equal to $\sigma = 0.2$ and $\xi = 0.1$. The detection threshold used in the simulation and models was $T = 150$ RFU.

In the comparison we let *DNAmixtures* be optimized (with the true parameters as start values in the optimizer). The optimized parameters values $\theta = (\pi_1, \dots, \pi_K, \mu, \sigma, \xi)$ were later inserted into the likelihood function implemented in *EuroForMix* to check that the same likelihood values were obtained. The comparison tables are shown where the values of θ are represented (5 significant digits) together with the log-likelihood values. Notice here that $\pi_K = 1 - \sum_k^{K-1} \pi_k$.

The sampled evidences were evaluated by conditioning on different types of hypotheses. We let the code xKyU denote that the model conditions on the hypothesis that the x first true references from the sampled stain are considered as known contributors, while y is number of unknown contributors in the same hypothesis. The script for carrying out the comparison can be found at www.euroformix.com/validation. The comparison table in Table S1 shows the likelihood values for each of the sampled evidence up to 12 decimals for each possible considered hypothesis. By comparison we found that the log likelihood values from *EuroForMix* and *DNAmixtures* are always identical up to 11 decimals.

Single-source stain (K=1) with $\mu = 200$

	<i>DNAmixtures</i>	<i>EuroForMix</i>	μ	σ	ξ
E1(1K0U)	-130.49052633191	-130.49052633191	200.0	0.2	0.1
E1(0K1U)	-171.261610009232	-171.261610009232	200.0	0.2	0.1
E2(1K0U)	-130.452181039142	-130.452181039142	200.0	0.2	0.1
E2(0K1U)	-175.250747463902	-175.250747463902	200.0	0.2	0.1
E3(1K0U)	-130.700911474164	-130.700911474164	200.0	0.2	0.1
E3(0K1U)	-179.675139974559	-179.675139974559	200.0	0.2	0.1

Two-person mixtures (K=2) with $\mu = 2500$ and $m_1 = 0.9$

	<i>DNAmixtures</i>	<i>EuroForMix</i>	m1	μ	σ	ξ
E1(2K0U)	-415.511575927926	-415.511575927926	0.90277	2477.35889	0.12543	0.10333
E1(1K1U)	-440.012776443395	-440.012776443395	0.91883	2486.27513	0.13554	0.10988
E1(0K2U)	-485.116377602029	-485.116377602029	0.91883	2486.27230	0.13554	0.10988
E2(2K0U)	-437.996105475833	-437.996105475833	0.87656	2394.87353	0.18024	0.10382
E2(1K1U)	-465.780902545946	-465.780902545946	0.87951	2400.37008	0.17808	0.11074
E2(0K2U)	-518.586686971514	-518.586686971514	0.87951	2400.36966	0.17808	0.11075
E3(2K0U)	-433.643227926185	-433.643227926185	0.91109	2384.77826	0.12932	0.10748
E3(1K1U)	-458.705779978583	-458.705779978583	0.90930	2388.84276	0.13214	0.11029
E3(0K2U)	-511.604121759721	-511.604121759722	0.90931	2388.84072	0.13215	0.11029

Three-person mixtures (C=3) with $\mu = 5000$ and $(m_1, m_2) = (0.6, 0.3)$

	<i>DNAmixtures</i>	<i>EuroForMix</i>	m1	m2	μ	σ	ξ
E1(3K0U)	-583.089927798733	-583.089927798732	0.58313	0.32130	5082.01259	0.11275	0.10151
E1(2K1U)	-623.069145322142	-623.069145322141	0.58567	0.32439	5080.98020	0.12075	0.10813
E1(1K2U)	-662.761440697117	-662.761440697116	0.57166	0.33105	5081.36505	0.10575	0.10480
E1(0K3U)	-692.355343509357	-692.355343509358	5.9543e-01	3.1728e-01	5.0826e+03	8.6744e-02	1.0361e-01
E2(3K0U)	-599.664027562726	-599.664027562726	0.60111	0.29599	4985.57972	0.18048	0.10699
E2(2K1U)	-617.69683952578	-617.696839525781	0.57751	0.25805	4977.95000	0.14675	0.10716
E2(1K2U)	-649.811404686069	-649.811404686068	0.57299	0.26916	4980.03152	0.13630	0.11070
E2(0K3U)	-681.724194926551	-681.72419492655	0.61305	0.23166	4981.26517	0.12217	0.10977
E3(3K0U)	-614.039509472884	-614.039509472883	5.6347e-01	3.2398e-01	4.8694e+03	1.3483e-01	9.9929e-02
E3(2K1U)	-650.84283554739	-650.84283554739	0.57955	0.31307	4867.94637	0.14258	0.11022
E3(1K2U)	-698.596535698722	-698.596535698723	0.56473	0.31811	4867.80505	0.14615	0.12143
E3(0K3U)	-734.443561553084	-734.443561553084	0.59049	0.29989	4868.43967	0.13953	0.12000

Table S1: The table compares the log-likelihood values for *EuroForMix* and *DNAmixtures* for the inserted parameter value θ based on three independent samples given as single-source stains or two- or three-person mixtures for different specified hypotheses. For K=1, *EuroForMix* obtained identical values as for *DNAmixtures*. For K=2, *EuroForMix* obtained identical values as for *DNAmixtures* except for E3(1K1U) and E3(0K2U) which were different in 12th decimal. For K=3, *EuroForMix* obtained identical values as for *DNAmixtures* except for E1(3K0U), E1(2K1U), E1(1K2U), E2(2K1U), E2(1K2U), E2(0K3U), E3(3K0U), E3(1K2U) which was different in 12th decimal.

B Technical details of model and inference

Approximation of the likelihood function using a compound allele

Some alleles who have a peak height below the detection threshold T give important information to the model parameters. Consider the situation where a peak height is presented above T for allele a , but not for $a - 1$ and $a + 1$. Assuming the stutter-model, a contributor with allele a will expect to add some stutter to allele $a - 1$. However, since the peak height at allele $a - 1$ is below the threshold, this indicates a small stutter proportion. Hence the information about the non-present peak height at allele $a - 1$ is important. For a contributor at allele $a + 1$ we expect some stutter proportion to allele a . However, since the peak height at allele $a + 1$ is below the threshold, the contributing stutter from $a + 1$ to a is expected to be less than ξT . Hence the information about potential stutters from alleles falling below the threshold is not important. Based on these arguments, we make an approximation of the likelihood function in equation (7) by redefining the set of alleles given in the database, \mathbf{A}_m , as

$$\mathbf{A}'_m = \mathbf{S}_m \cup Q_m \quad (18)$$

where $\mathbf{S}_m = \{a \in \mathbf{A}_m : Y_{m,a} \geq T\} \cup \{a - 1 \in \mathbf{A}_m : Y_{m,a} \geq T\}$ is the set of alleles with peak heights above the threshold and their corresponding potential stutter alleles, and Q_m is any of the remaining alleles in the set $\mathbf{A}_m \setminus \mathbf{S}_m$, grouped together as a compound allele. As used by Taylor et al. [14] and Puch-Solis et al. [13], we let the allele frequency of Q_m be given as $1 - \sum_{a \in \mathbf{S}_m} p_a$. For the degradation model, Q_m , is assigned to have fragment length $\max_a f_{m,a}$, the maximum fragment length at (undegraded) marker m .

The likelihood function for independent replicated samples

For this particular case we assumed that the sample has been retyped (after extraction) such that stain samples are independent replicates which are assumed to contain the same contributors and satisfy the same model assumptions (same model properties across replicates). Consider the number of replicates as R , and let the observed peak height for allele a at marker m for replicate r be given as $Y_{r,m,a}^*$. Then the probabilistic model given in equation (7) is extended to

$$p(E|H, \theta) = \prod_{m=1}^M \sum_{\substack{g_{m,k} \in Q_m \\ k = 1, \dots, K}} \left(p(\mathbf{g}_m|H) \prod_{r=1}^R \prod_{a \in \mathbf{A}_m} p(Y_{r,m,a}^* | \mathbf{g}_m, \theta) \right) \quad (19)$$

C Peak height summary of the trace sample

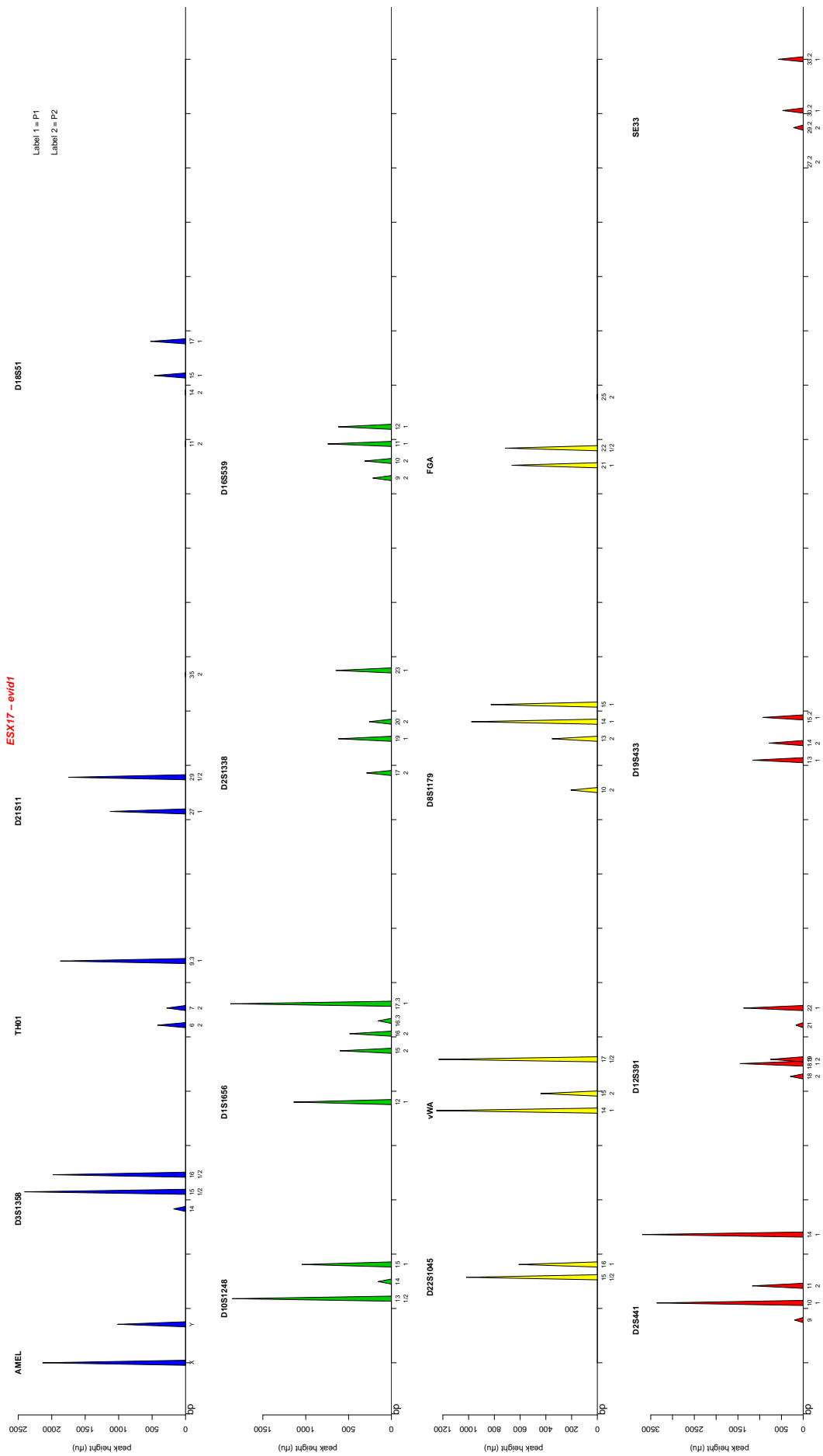


Figure S1: The plot shows the peak heights of the trace sample provided in the example, in a epg-like format.

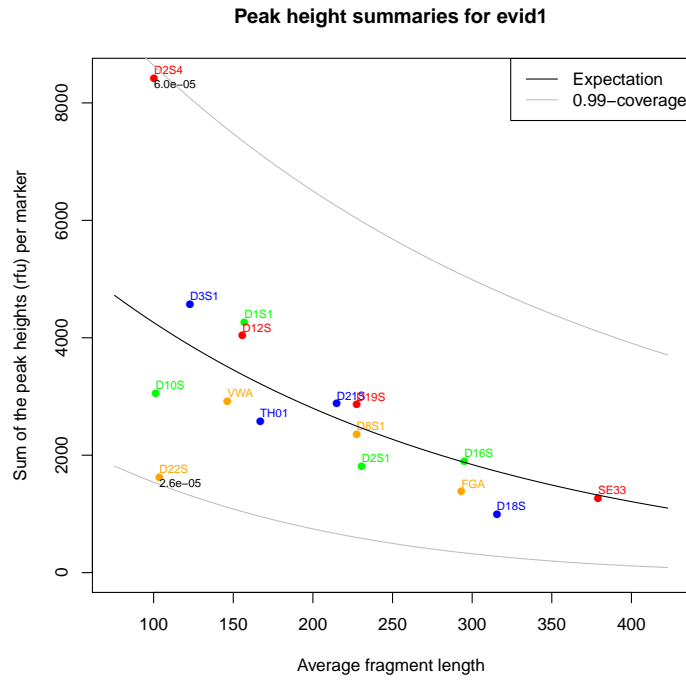


Figure S2: The plot shows the sum of the peak heights y per marker as a function of average fragment length (per marker) x for the evidence provided in the example. The lines show the expectation, the 0.005- and 0.995-quantiles from the maximum likelihood fitted model of the underlying model $y \sim \text{gamma}(2\sigma^{-2}\beta^x, \mu\sigma^2)$ using all data. The values belonging to the points of D2S441 and D22S1045 are the probability of observing a more extreme value than observed. Each probability is based on the maximum likelihood fitted model where the corresponding marker is left out from the data.

D Resulting deconvolution table

Rank	D3S1358	TH01	D21S11	D18S51	D10S1248	D1S1656	D2S1338	D16S539	D22S1045	VWA_C	D8S1179	FGA_C2	D2S441	D12S391	D19S433	SE33_C2	probability
1	15/16	6/7	29/99	99/99	13/14	15/16	17/20	9/10	15/16	15/17	10/13	22/99	11/14	18/19	14/14	29.2/99	1.224e-05
2	15/16	6/7	29/99	99/99	13/14	15/16	17/20	9/11	15/16	15/17	10/13	22/99	11/14	18/19	14/14	29.2/99	1.222e-05
3	15/16	6/7	29/99	15/99	13/14	15/16	17/20	9/10	15/16	15/17	10/13	22/99	11/14	18/19	14/14	29.2/99	1.203e-05
4	15/16	6/7	29/99	15/99	13/14	15/16	17/20	9/11	15/16	15/17	10/13	22/99	11/14	18/19	14/14	29.2/99	1.201e-05
5	15/16	6/7	29/99	99/99	13/14	15/16	17/20	9/10	15/99	15/17	10/13	22/99	11/14	18/19	14/14	29.2/99	1.191e-05
6	15/16	6/7	29/99	99/99	13/14	15/16	17/20	9/11	15/99	15/17	10/13	22/99	11/14	18/19	14/14	29.2/99	1.188e-05
7	15/16	6/7	29/99	15/99	13/14	15/16	17/20	9/10	15/99	15/17	10/13	22/99	11/14	18/19	14/14	29.2/99	1.17e-05
8	15/16	6/7	29/99	15/99	13/14	15/16	17/20	9/11	15/99	15/17	10/13	22/99	11/14	18/19	14/14	29.2/99	1.168e-05
9	15/16	6/7	29/99	14/99	13/14	15/16	17/20	9/10	15/16	15/17	10/13	22/99	11/14	18/19	14/14	29.2/99	1.024e-05
10	15/16	6/7	29/99	14/99	13/14	15/16	17/20	9/11	15/16	15/17	10/13	22/99	11/14	18/19	14/14	29.2/99	1.022e-05
11	15/16	6/7	29/99	99/99	13/14	15/16	17/20	9/10	15/16	15/17	10/13	21/99	11/14	18/19	14/14	29.2/99	1.002e-05
12	15/16	6/7	29/99	99/99	13/14	15/16	17/20	9/11	15/16	15/17	10/13	21/99	11/14	18/19	14/14	29.2/99	9.998e-06
13	15/16	6/7	29/99	14/99	13/14	15/16	17/20	9/10	15/99	15/17	10/13	22/99	11/14	18/19	14/14	29.2/99	9.957e-06
14	15/16	6/7	29/99	14/99	13/14	15/16	17/20	9/11	15/99	15/17	10/13	22/99	11/14	18/19	14/14	29.2/99	9.937e-06
15	15/16	6/7	29/99	17/99	13/14	15/16	17/20	9/10	15/16	15/17	10/13	22/99	11/14	18/19	14/14	29.2/99	9.932e-06
16	15/16	6/7	29/99	17/99	13/14	15/16	17/20	9/11	15/16	15/17	10/13	22/99	11/14	18/19	14/14	29.2/99	9.912e-06
17	15/16	6/7	29/99	15/99	13/14	15/16	17/20	9/10	15/16	15/17	10/13	21/99	11/14	18/19	14/14	29.2/99	9.846e-06
18	15/16	6/7	29/99	15/99	13/14	15/16	17/20	9/11	15/16	15/17	10/13	21/99	11/14	18/19	14/14	29.2/99	9.826e-06
19	15/16	6/7	29/99	99/99	13/14	15/16	17/20	9/10	15/99	15/17	10/13	21/99	11/14	18/19	14/14	29.2/99	9.745e-06

Figure S3: The plot shows the 19 first ranked genotypes with corresponding probabilities for the unknown profile under a model with two contributors, but where individual P1 is known to be one of the contributors. In addition the model assumes degradation and stutters, but does not assume allele drop-in or sub-population structure. Allele “99” represents any allele not presented in the sample (a compound allele).

E Experimental results

To check the practical performance of *EuroForMix*, we simulated three random DNA profiles where one, two, three and four individuals contributed. The purpose of the experiment was to discover the time taken for the methods in this article to return results for different scenarios. The experiment was carried out by first simulating alleles for K number of contributors using the population frequencies and the corresponding allelic peak heights using the model from section 2.2 with specified model parameters. We then considered the first sampled contributor as the person of interest S . To calculate the likelihood ratio (LR) we compare hypotheses H_p :“ S contributed to the sample” versus H_d :“ S did not contribute to the sample”. We calculated both the maximum likelihood based LR , LR_F , and the Bayesian based LR , LR_B . All samples were simulated with peak height expectation $\mu = K * 1000$, peak height coefficient of variation $\sigma = 0.2$, stutter proportion $\xi = 0.07$ and degradation slope parameter $\beta = 0.7$. Each sample was generated with different specified mixture proportions (π_1, \dots, π_K) given in Table S2. For the maximum likelihood estimations we required 5 random start points for the optimization. For the integrals we required relative error $\delta = 0.2$. The upper boundary of the parameters in the uniform priors were 10000 for μ , 0.5 for σ and 0.2 for ξ . For the model we assumed K number of contributors, detection threshold $T = 150$ RFU, no drop-in ($C = 0$) and no sub-population structure ($F_{st} = 0$). The degradation and stutter model was considered with the corresponding parameters β and ξ treated as unknown. The total computing time to achieve a LR value was registered with an Intel Core i7-2600 3.4 GHz processor.

From Table S2 it can be observed how the calculation time grows exponentially with number of contributors, and how the Bayesian based LR tends to be more time-consuming than the maximum likelihood based LR .

K	$\pi_1/\dots/\pi_K$	$\log_{10}LR_F$	time (min)	$\log_{10}LR_B$	time (min)
1	1	22.92	0.03	22.92	0.02
1	1	22.75	0.01	22.43	0.01
1	1	22.41	0.01	22.21	0.02
2	0.5/0.5	13.42	1.2	13.02	2.7
2	0.25/0.75	13.26	0.6	13.09	2.5
2	0.1/0.9	8.24	0.6	7.61	1.8
3	0.5/0.25/0.25	14.68	113.9	13.83	1051.6
3	0.25/0.5/0.25	6.83	102.8	4.97	179.8
3	0.1/0.5/0.4	6.93	71.72	5.9	559.1
4	0.5/0.2/0.2/0.1	18.46	7395.9	15.72	11470.3
4	0.3/0.4/0.2/0.1	9.55	7089.8	8.15	11982.9
4	0.1/0.5/0.2/0.2	4.74	9134.99	0.42	12716.9

Table S2: The table shows the resulting likelihood ratio quantities for the maximum likelihood approach, LR_F , and the Bayesian approach, LR_B , for different simulated DNA samples. K is the true and assumed number of contributors to the sample, π_x is the specified mixture proportions for contributor x (here π_1 is the mixture proportion of P). The time is total number of minutes taken to calculate the LR quantity.