2020

Evaluation of the Performance of Probabilistic Genotyping Software on Complex Mixture Samples

Kristen Newland

West Virginia University, klnewland@mix.wvu.edu

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Part of the [Computational Biology Commons](#)

Recommended Citation

Newland, Kristen, "Evaluation of the Performance of Probabilistic Genotyping Software on Complex Mixture Samples" (2020). *Graduate Theses, Dissertations, and Problem Reports*. 7720.

<https://researchrepository.wvu.edu/etd/7720>

This Thesis is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Thesis in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Thesis has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

Evaluation of the Performance of Probabilistic Genotyping Software on Complex Mixture Samples

Kristen Newland, B.S.

**Thesis submitted to the Eberly College at West Virginia University
in partial fulfillment of the requirements for the degree of
Masters of Science in Forensic and Investigative Science**

Tina Moroose, M.S., Chair
Casey Jelsema, PhD
Keith Morris, PhD

Department of Forensic and Investigative Science

**Morgantown, West Virginia
2020**

**Keywords: Mixture interpretation, Likelihood ratio, DNA analysis, NOCIt, TrueAllele,
GeneMapper ID-X, Expert systems
Copyright 2020 Kristen Newland**

ABSTRACT

Evaluation of the Performance of Probabilistic Genotyping Software on Complex Mixture Samples

Kristen Newland, B.S.

Probabilistic genotyping systems use sophisticated and complex algorithms to aid DNA analysts with complex mixture interpretation. These systems are used to generate a likelihood ratio (LR) which is used to provide the analyst with the statistical weight of the match between an evidence profile and a known contributor. Complex mixtures occur from samples consisting of more than two contributors and samples that have experienced degradation or consist of low template quantity.

This study was performed to compare the performance of three mixture analysis software tools on the same set of samples to determine the capabilities of each software, false inclusions, and false exclusions. Three probabilistic genotyping systems were evaluated GeneMapper ID-X™ v1.4, NOCI/CEESI, and TrueAllele® using 750 samples provided from the PROVEDIt database. The sample set consisted of 100 single-source samples and 650 samples containing 2-, 3-, 4-, and 5-person samples with template quantities ranging from 0.016 to 1 ng of DNA. Each software was evaluated using the recommended default settings. Each sample was analyzed using the true number of contributors and determined the LR value for each of the contributors present in the sample. To determine the number of false inclusions each sample was analyzed using two incorrect individuals. NOCI performed better than GeneMapper ID-X™ in regard to determining the number of contributors. However, the NOCI tool despite being more accurate than GeneMapper ID-X™ experienced a decrease in accuracy as the number of contributors increased. TrueAllele® reported the fewest number of false inclusions across one to five contributor sample sets, but it also reported a higher number of false exclusions when compared to CEESI. GeneMapper ID-X™ is only able to compute LR values for 1- and 2-person samples, and it will only produce the LR if the software determines that the selected individual is present within the sample which caused it to report the highest number of false exclusions. There was no statistical difference in the LR values obtained from CEESI and GeneMapper ID-X, but CEESI and GeneMapper ID-X™ did produce statistically higher LR values when compared to TrueAllele® for the 1- and 2-person samples. With the exception of the three contributor samples CEESI did not produce statistically higher LR values than TrueAllele®.

This study shows the value of comparing different probabilistic genotyping systems to determine which software tool is the most accurate for casework samples. Future studies should be performed to determine the optimal settings for each software and compared to other systems on the market such as STRMix™. This study shows the value of using statistical algorithms to aid in mixture sample deconvolution to help improve accuracy.

Table of Contents

Section	Page Numbers
Cover page	i
Abstract	ii
Introduction	1-7
Methods	8-9
Results	10-18
Discussion	19-21
Works Cited	22-23

Introduction

In forensic casework DNA evidence can consist of a wide variety of biological samples. In many cases those biological samples originate from one contributor, however there are instances where mixture samples are encountered. A DNA mixture is a biological sample that has more than one contributor present (Bieber et al. 2016). A mixture sample is present once a DNA profile has been generated and there are more than three peaks at two or more loci within the profile or when the differences within the peak heights are greater than the laboratory's defined peak height ratio. Complex mixture samples occur when there is the presence of more than two contributors, or if the sample has experienced a high degree of degradation (Bieber et al. 2016).

For the past several decades DNA analysis has been perceived as the “gold standard” in forensic science (Buckleton and Curran 2008). The analytical scheme and interpretation methods for single-source samples have been highly regarded and viewed as objective. With mixture DNA samples on the other hand, the interpretation becomes more subjective as the mixture becomes more complex. Based on the recommendations from the President's Council of Advisors on Science and Technology (PCAST) report scientists are working to develop a standard method to improve mixture interpretation (Buckleton and Curran 2008). These recommendations have also led to the creation of several statistical software programs to aid in the interpretation of mixture samples. Probabilistic genotyping systems are designed to translate the data in electropherograms into useful interpretation data to the analyst (Butler, 2015). There are currently more than 13 available software tools that labs have the option to choose from (Haned and Gill 2011). These tools have different capabilities not only in the interpretation model, but also in regard to the statistical computations that are used to determine the Likelihood Ratio (LR).

Following guidelines 3 and 4 from the Scientific Working Group on DNA Analysis Methods (SWGDM) guidelines all interpretation software must undergo validations that assess sensitivity, specificity, accuracy, and precision on a variety of casework type samples with ranging DNA quantities and mixture ratios (SWGDM, 2015). Standard 6 from the Federal Bureau of Investigation Quality Assurance Standards (FBI QAS, 2020) also state that all analysts must undergo training and competency testing prior to using any interpretation software on casework samples and making definitive interpretations. Any interpretations and conclusions reached by the analyst must also undergo a review (FBI QAS, 2020). However, it is important to examine different software tools to evaluate how a software's capabilities and parameters affect the overall sample interpretation.

This study sought to evaluate the performance of three mixture analysis software tools on the same set of samples to determine the strengths and weaknesses of each software, and how they compare to each other when analyzing the same set of samples. It is not financially feasible for forensic laboratories to procure several software tools to evaluate each one. Examining each software is very time consuming. Samples were provided from the Project Research Openness for Validation with Empirical Data (PROVEDIt) database and consisted of one to five contributors with template quantities ranging from 0.03 ng to 0.75 ng. Computations were performed using the NIST population datasets (STRBase, 2017).

Background

Examining Mixtures

Mixture samples can often pose interpretation difficulties for an analyst. Reviewing profile data becomes increasingly difficult when more individuals have contributed to the mixture, since the number of possible genotype combinations that explains the evidence sample increases. Samples that have a low template quantity or if the sample has experienced a high level of degradation also cause issues with interpretation (Bieber et al. 2016).

An examiner can use peak information to evaluate the peak height ratios to deconvolute major and minor contributors within the mixture sample. In the case of a simple mixture, it can either be separated through the use of peak height information or deducing a known contributor from a two-contributor mixture. After the sample has been separated then it can be treated the same as a single source sample and a Random Match Probability (RMP) is performed on the sample (Butler, 2010).

Peak information can also be used to determine the number of contributors present in the sample by examining the number of peaks at each locus. The presence of four alleles is often indicative of a two-contributor mixture, whereas more than five alleles is indicative of three or more contributors (Clayton et al. 1997). This can help the analyst explain the potential genotypes present within the sample, as well as be useful for statistical interpretation. The number of contributors is one of the assumptions for likelihood ratio calculations (Biedermann et al. 2012).

Determining the number of contributors becomes increasingly more difficult when taking into account allelic dropout and allele sharing. If a sample has experienced a lot of dropout then it can reduce the number of alleles at each locus and then make it appear as if there are fewer contributors present. The same thing can be said in regard to allele sharing. If two contributors share the same two alleles then it still appears as if only one person contributed to that particular sample (Biedermann et al. 2012).

Interpretation Methods for Mixtures

The Scientific Working Group on DNA Analysis Methods (SWGDM) is a group of scientists that have set the standards for all DNA analysis and interpretation methods for DNA laboratories at the federal, state and local levels in the United States (Butler, 2010). These standards help ensure that all samples that are submitted to a crime laboratory are being processed and examined using validated techniques and objective methods. Mixture interpretation presents difficulty not only due to the increased number of allele calls, but having to account for artifacts that may be present within the mixture and for allelic dropout that can occur with low template and degraded samples (Swaminathan et. al, 2016).

Combined Probability of Inclusion and Combined Probability of Exclusion

In most countries to report the value of the evidence from a mixture sample the laboratory calculates the combined probability of inclusion (CPI) which is the amount of a population that can be included as a potential contributor to the mixture that is present within an evidence sample (Butler, 2010). This statistic is computed by calculating the sum of the allele frequencies at each locus (Equation 1) and then multiplying across loci to obtain the CPI statistic for the mixture (Equation 2) .

$$P_{locus} = PI = (p_1 + p_2 + \dots p_c)^2 \quad (1)$$

$$CPI = P_{locus1} \times P_{locus2} \times \dots P_{locus(N)} = P_{Mixture} \quad (2)$$

The value of $P_{Mixture}$ refers to the probability that an unrelated individual in the population is the contributor to the sample in regards to all of the loci present in the sample. N is for the number of loci present in the stain

Another statistic that can be computed is the combined probability of exclusion (CPE). This statistic is reported as a percentage and is providing the percentage of individuals that can be excluded as a contributor to the sample. This calculation is shown in Equation 3 (Butler, 2010).

$$CPE = 100 \times (1 - P_{Mixture}) \quad (3)$$

These two probability statistics are obtained from the interpretation model that factors in the sum of allele frequencies that are observed at each locus. A suspect's profile is used for the exclusion portion of analysis, but is not used in the calculation of the statistic (Bieber et al. 2016). This calculation is unable to account for missing data such as allele sharing and allelic dropout. The calculation only counts each allele that is visible once, and because of this it is unable to account for contributors sharing alleles. This interpretation method does not make the best use of the data.

Based on SWGDAM guidelines there are two models that can be used to evaluate the samples the restricted or unrestricted models. As seen in Figure 1, the restricted model allows the analyst to use peak height information to help aid in the overall interpretation of the sample, whereas the unrestricted allows for all potential genotype combinations to be considered this is most often used when there is a distinct major/minor contributor to the sample. The restricted model allows for a more concise and accurate interpretation of the sample since it is eliminating unlikely genotype combinations based on the peak height information present in the sample. The restricted approach can be more reliable since it utilizes peak information and pairs alleles together that have closer RFU values, which eliminates the unlikely combinations. Using an unrestricted LR model can include genotypes in the calculation that would be excluded by an analyst during manual interpretation, and because of this the unrestricted approach has the potential to less accurately reflect the data present in the mixture sample as opposed to using the restricted LR model (SWGDAM, 2017).

Figure 1. SWGDAM mixture interpretation guidelines. This figure displays the difference between using a restricted vs unrestricted model to evaluate mixture sample data (SWGDAM, 2017) .

Likelihood Ratios

Mixture interpretation has begun to shift away from exclusion probabilities and towards likelihood ratios (LR) due to their more complex and informative framework (Perlin et al. 2015). They are able to evaluate each sample under an inclusionary and an exclusionary hypothesis to evaluate the strength of each using conditional probabilities (equation 4). Likelihood ratio calculations take into account each potential combination of genotypes present in the mixture sample. Likelihood ratio calculations are dependent on determining the number of contributors present within the sample, and you are unable to have contributors from different sub-populations. Obtaining a LR value greater than one shows support for the inclusionary hypothesis indicating that it is more likely that the contributor of interest is present within the mixture as opposed to an unknown unrelated individual (Egeland et. al, 2003).

$$LR = \frac{\Pr(H_p)}{\Pr(H_d)} \quad (4)$$

examining a one-contributor sample seen in Figure 2. The prosecutor's hypothesis or the numerator is going equal $\Pr=1$, since the prosecutor is saying that based on the data that they are in there sample. The defense however is stating that it is not the suspect so the analyst has to explain the alleles a and b. Following Hardy-Weinberg results in the LR formula seen in equation 5. This is the same as performing the random match probability (Butler, 2015).

$$LR = \frac{1}{2(p_a p_b)} \quad (5)$$

Figure 2. Image of an electropherogram displaying a one-contributor locus at the D3 locus.

If an analyst was examining the two-contributor mixture seen in Figure 3. using an unrestricted approach. The analyst would have to explain all of the potential genotype combinations. The selected hypothesis would dictate the structure of the LR formula. The prosecutor is stating that the suspect (a,b) is in the mixture ($\Pr=1$), but now they have to account for this unknown individual (c,d). Since there are two unexplained alleles following Hardy-Weinberg it is explained by using $2p_c p_d$. The defense is stating that the suspect isn't in the profile and then the second set of alleles also needs explained. Since all of the potential combinations have to be considered to explain the mixture sample then you end up with $24p_a p_b p_c p_d$. The numerator cancels out and results with a final equation of:

$$LR = \frac{1}{12(p_a p_b)} \quad (6)$$

Figure 3. Image of an electropherogram displaying a mixture at the D3 locus

Probabilistic Genotyping Software

In the past decade the use of algorithms and computer software has become increasingly utilized to aid in the deconvolution of DNA mixtures obtained from casework samples. It helps apply a more objective and consistent approach to mixture analysis by automating the calculations determining the

probability that a contributor is present in the sample (Haned and Gill, 2011). However, there is variation in these techniques depending on the developer and how the software was designed to work.

There are three types of models that can be utilized—continuous, deconvolution and semi-continuous. A continuous model utilizes all of the data present in the sample and relies on peak height information to calculate the likelihood ratio that is assigned to each of the genotypic combinations. Most of the systems that fall under this category use a Bayesian approach (see equation 1) and use approximation algorithms such as Markov Chain Monte Carlo (MCMC). These types of algorithms perform integrations over all of the parameters specified in the software (Haned and Gill, 2011).

Sampling algorithms are beneficial especially with Bayesian models. Markov Chain Monte Carlo is a computer driven method that combines two principles: Monte Carlo and Markov chain. Monte Carlo is used to examine random samples from a normal distribution and calculate the mean from those samples. The Markov chain principle is where the randomly selected samples are produced by a sequential process, or chain where each random sample is used to help generate the next random sample. The Markov property states that even though a new sample depends on the one before, the new samples do not depend on any of the samples before the previous sample (van Ravenzwaaij, et al. 2018). Due to the random sample selection one of the drawbacks of using a sampling algorithm is that analyzing the same sample twice can result in a slightly different value, since the sampling process is not identical.

Semi-continuous systems do not rely on peak heights or mixture ratios so they don't utilize all of the data in the sample, but these types of software systems are able to account for allelic drop in and drop out probabilities. A deconvolution system can also be classified as a semi-continuous or continuous system, depending on how it works. GeneMapper™ ID-X is a deconvolution system that uses peak information to separate out contributors, but does not utilize an approximation algorithm or more advanced parameters when performing computations (Haned and Gill, 2011).

GeneMapper™ ID-X v. 1.4

GeneMapper™ ID-X is a genotyping software that was developed by Applied Biosystems. This software is more commonly used for viewing genetic profiles once they have finished capillary electrophoresis. GeneMapper™ ID-X has a variety of interpretation tools including a Mixture Analysis Tool that is a deconvolution tool that relies on a threshold-based method so that it can utilize peak height information to evaluate the data in the sample (Hansson and Gill, 2011). The Mixture Analysis Tool segregates the samples into three categories—one contributor, two contributors, and three or more contributors. After the samples have been segregated, this tool is able to calculate the LR statistic for one and two contributor samples and then computes the CPI/CPE for three or more contributor samples. It uses the defined Peak Height Ratio rule and the Residual rule to display all of the potential genotypic combinations at each locus (Marshall University, 2008).

Once the Mixture Analysis Tool is open, the user is able to customize the analysis settings for examining the peak height information. From there, the user can select the samples of interest from the project that is open and then the software segregates the samples. The user then selects the desired statistical computations for the populations of interest.

NOCI/CEESIt

NOCI/CEESIt is a software that was developed in a collaborative event between Boston University, Rutgers University, and the Massachusetts Institute of Technology. This is a two-part software that can compute the number of contributors in the software and then calculate the LR for a selected contributor using .csv files to utilize a multi-threaded processing to use multiple available cores on a computer (Swaminathan et al. 2016). This software requires a calibration from a single source

sample to model peak information to train the software to evaluate the sample files for both portions of the software. The calibration set must consist of samples from several individuals with at least 5 template quantities. The calibration set creates a zip file that is then run with every set of samples being analyzed with the software. This software allows you to upload your own population databases (Swaminathan et al. 2016).

The NOCIt portion of the software is used to determine the number of contributors within a mixture sample. NOCIt computes the probability for the number of individuals in a sample using the Monte Carlo (MC) sampling process to create a uniform distribution to model the ratio of the “n” contributors. Using this process the probability is computed based on the genotypes for the “n” contributors in the sample using the allele frequencies (Alfonse et al. 2017). The model’s parameters are able to account for baseline noise, stutter, and allelic drop-in and drop out, peak heights, and potential sample degradation. A uniform priori distribution is assumed to determine the posteriori probability that “n” number of contributors are present in the sample where “n” is between zero to five individuals (Alfonse et al. 2017).

The CEESIt portion of the software uses a continuous model that determines the likelihood that a selected contributor is present within the mixture sample. The model incorporates parameters to account for noise, allelic dropout and artifacts such as stutter. The algorithm assumes a constant mixture ratio for all of the loci in the mixture sample (Swaminathan et al. 2016). The user selects the population of interest and the number of contributors for the selected sample. Then the user can import known genotypes into the software and select them from the dropdown list. The user also has the capability to select known contributors to the sample.

TrueAllele® Casework System

TrueAllele® is a mixture interpretation software that relies on MCMC and Bayesian probability modeling to determine the likelihood ratio match statistic. This modelling uses peak information, artifacts, and accounts for variance to explain the data present in the mixture. The MCMC statistical sampling method is used to compute the Bayesian equations to solve for each joint posterior probability and gives the marginal distribution for the explanatory variables by considering tens of thousands of possible values for each variable. This software evaluates if a known contributor is present within the mixture sample by comparing a known reference sample to one of the separated unknown genotypes compared to a relative population of individuals (Bauer et. al, 2020).

The match strength is mathematically determined at reference points before making a comparison to a known sample. The software creates a distribution of match strength values for the user to examine the amount of genotype uncertainty for the sample. A mostly positive distribution is indicative of a matching genotype, whereas a distribution that is centered at the left of zero is more indicative that the genotype is not a match (Perlin et al. 2015).

The software user begins sample analysis by uploading the .fsa or .hid files into the Analysis Module. From here the user can select the necessary amplification template and the format for the genetic analyzer that was used. This module allows the software and the user to check the size standard to ensure that it was properly tracked. Then the controls and ladder peaks are analyzed for potential issues and control checks. The software then completes the process of converting the .fsa or .hid file into the .gel file that is used for the rest of the interpretation process.

Once a .gel file has been created it can be uploaded and continue through the TrueAllele® analytical scheme. For deconvolution to begin, a request file is created by distinguishing the sample type—evidence versus a known sample. From there the user has the capability to select the number of contributors and adjust settings such as how many cycles the computer should take to solve the problem

and if the sample is degraded. TrueAllele® is capable of resolving six samples at once and depending on how many cycles the computer is set to run, it takes a few hours to process the samples. If one sample finishes early it moves onto the next sample in queue to keep six samples running at a time.

Once the sample has been processed through the software's algorithm it can then be further examined in the report module. In this module this is where the user selects the known genotypes of interest and the populations of interest and then the log(LR) score is generated for the sample. The user also has the option of view the log(LR) for each of the loci in the sample. This can help the user determine which loci contained the most information (Perlin et al. 2016).

PROVEDIt

Samples were obtained from the PROVEDIt (Project Research Openness for Validation with Empirical Data) dataset. The PROVEDIt database is a large-scale database that contains over 25,000 STR profiles synthetically created to vary in quality and quantity to mimic those that can be obtained from crime scenes (Alfonse et al. 2017). The collection of profiles includes one contributor samples along with two to five person mixtures of varying contributor ratios. Samples were obtained from whole blood that were extracted using phenol chloroform, quantitated using Quantifiler® Duo, and were amplified using either Globalfiler®, PowerPlex®16 HS, or Identifiler® Plus with target quantities ranging from 0.007 to 1 ng. The samples were then injected onto either the Applied Biosystems® 3500 Genetic Analyzer for 5, 15, and 25 seconds or on the Applied Biosystems® 3130 Genetic Analyzer for 5, 10, and 20 seconds (Alfonse et al. 2017).

Table 1. Mixture Ratios used for the samples in the PROVEDIt database. The mixture samples were created using the following ratios after being quantified using Quantifiler® Duo. The ratios dictated the amount of DNA contributed by each of the individuals present in the mixture.

2 Contributor	3 Contributor	4 Contributor	5 Contributor
1:1	1;1;1	1:1:1:1	1:1:1:1:1
1:2	1;2:1	1:1:2:1	1:1:2:1:1
1:4	1:4:1	1:1:4:1	1:1:4:1:1
1:9	1:9:1	1:1:9:1	1:1:9:1:1
	1:2:2	1:2:2:1	1:1:2:2:1
	1:4:4	1:4:4:1	1:1:4:4:1
	1:9:9	1:9:9:1	1:1:9:9:1
		1:4:4:4	1:2:2:2:1
			1:4:4:4:1
			1:9:9:9:1

Interpretation Methods

Sample Set

The sample set used for this study consisted of 750 samples that contained 1, 2, 3, 4, and 5 contributors that were obtained from the PROVEDit dataset. All 750 samples were treated as unknown samples that were used on each of the three probabilistic genotyping software systems. The extraction procedure consisted of using phenol/chloroform purification and alcohol precipitation on whole blood samples. The samples were then quantified using Quantifiler® Duo. Mixtures were created using set mixture ratios (see Table 1) and then using target concentrations of 0.5, 0.25, 0.125, 0.063, 0.031, 0.016, and 0.007 ng for the final sample concentrations (Alfonse et. al, 2017). The samples were then amplified with Globalfiler® Amplification kit and injected on the Applied Biosystems® 3500 Genetic Analyzer for 15 seconds.

GeneMapper™ ID-X

The mixture samples were examined by uploading the .hid files into the software. Prior to analyzing them in the Mixture Analysis Tool, mixture samples were analyzed using a threshold of 50 RFU and single-source samples were analyzed using 250 RFU. Once the samples were uploaded to the Mixture Analysis Tool they were automatically segregated into 1, 2 or 3 or more contributor tabs. Known contributors were compared to the sample profile for the 1 and 2 contributor samples that had been correctly segregated. Along with evaluating the correct known contributors, two incorrect contributors were selected per sample to evaluate how often false inclusions were selected. When a contributor was able to be pulled out of the mixture, the LR was calculated in reference to the NIST Caucasian population (STRbase, 2017).

NOCIt/CEESIt

Calibration

NOCIt requires a calibration in order to analyze any samples. The calibration was performed using the Globalfiler® kit file. The calibration set consisted of 75 single-source samples that had five template quantities: 0.5, 0.25, 0.125, 0.063, and 0.031 ng. Single-source samples that were used during the calibration weren't used in the testing set. The calibration set is used to inform the software so that the software functions at a more accurate and optimal level since it provides a model of what data should look like at various DNA concentrations to model allelic dropout (Alfonse et al. 2017).

Testing Set

In NOCIt, all of the sample runs were performed using three population datasets—African American, Caucasian, and Hispanic using the .csv files obtained from the PROVEDit dataset. The samples were analyzed using the software developers' default settings for the software with the exception of the refinement time limit which was set to 18,000 seconds to go along with the standard error tolerance of 0.05. This is the tolerance level for the estimated standard error of the a posteriori probability distribution that is used for termination of the process. NOCIt first performs an initial Monte Carlo sampling for each number of contributors from 0 to the maximum number of contributors (Swaminathan et al. 2015). The maximum number of contributors was set to 5 for all of the analyzed samples. After this initial sampling, NOCIt selectively performs more Monte Carlo sampling until either the estimated standard error of the a posterior probability for examined number of contributors is below this 0.05, or if the running time exceeds the 18,000 second refinement time limit that was used for this study.

After all of the samples were run in NOCI, the samples were analyzed in the CEESI portion of the software. Due to time constraints samples were run using their true number of contributors for only the NIST Caucasian population. After importing the samples into the software, a person of interest was selected for each sample and the threshold was set to 1 RFU to capture all of the data. All of the samples were processed analyzing one of the contributors at a time. After all known contributors had been assessed, two contributors that were not present in the sample were compared to determine if any potential false inclusions occurred.

TrueAllele® Casework System

The mixture samples were processed using the .hid files in the TrueAllele® software. Requests were created for each sample using the true number of contributors. Samples were viewed manually to evaluate sample quality to determine factors such as degradation and allelic drop out. Samples that did not appear to have experienced severe dropout were processed with the recommended Screen settings of 5,000 burn in MCMC cycles and 5,000 read out cycles. Samples that appeared to have diminished quality underwent 10,000 burn in MCMC cycles and 10,000 read out cycles along with adjusting settings to account for degradation and dropout to enhance the accuracy of casework systems cycling parameters so that it could account for a more complicated mixture interpretation (Perlin et al, 2015).

Once sample processing had finished, each sample was compared to the known contributor profile along with two false inclusions relative to the NIST Caucasian reference population to assess the log(LR) value.

Statistical Analysis

For this study the statistical analysis consisted of a linear mixed-effect model logistic regression computed in R Studio (R Core Team R) to make comparisons of the systems calculated log(LR) values that were computed using the NIST Caucasian dataset. This statistical model is similar to linear regression, and is used to account for variation occurring from fixed effects that can be explained from the independent variable and variation that occurs from random effects that cannot be explained by the independent variable. This model factored in the sample concentration and the ratio of contribution for the 750 samples on the probabilistic genotyping systems.

Accuracy was assessed for each software by determining false inclusion rates for incorrectly chosen contributors and was evaluated in terms of percentages for each software, after each sample was subjected to two incorrect contributors. To determine the accuracy of the NOCI/CEESI tool it was evaluated using three NIST reference populations—African American, Caucasian, and Hispanic (STRbase, 2017). For each sample, the true number of contributors was compared to the predicted number of contributors.

Results

Segregation Accuracy

NOCIt was evaluated using three NIST populations to determine the accuracy for the segregation portion of the software (Table 2). As expected, the most complex mixtures containing five contributors displayed the lowest accuracy. However, unexpectedly, there was a variation in accuracy across the three populations. For the African American population, the software was most accurate when examining three contributor samples with an accuracy of 87.6%, whereas the Hispanic population was the most accurate evaluating two contributor samples with 90.9% accuracy. Additionally, variation was observed as a large decrease in the accuracy values between the populations for the five contributor samples. Specifically, there was over a 20% decrease in accuracy when comparing the Hispanic population to both the African American and Caucasian populations (Table 2).

Table 2. Using 750 samples from the PROVEDIt database, the accuracy of the NOCIt portion of the NOCIt/CEESIt software was investigated to determine how well the software determined the number of contributors for three NIST populations. NOCIt was less accurate as the number of contributors increased.

Contributor Number	Population		
	African American	Caucasian	Hispanic
1	78%	86%	89%
2	84%	88.6%	90.9%
3	87.6%	83.8%	86.9%
4	82.5%	87.3%	78.3%
5	54.5%	56.4%	30.7%

GeneMapper™ ID-X (v1.4) was able to accurately segregate 87.9% of the mixture samples, and 76% of single source samples using a threshold of 50 RFUs. For the 138 two contributor samples that were correctly classified, only 48 successfully had a LR statistic calculated. For 38 of those samples, the Mixture Analysis Tool calculated a LR statistic for only one of the two contributors in the two contributor mixture, therefore 38 LR statistics were calculated. As for the other 10 samples, a LR statistic could be computed for both contributors, therefore calculating 20 LR statistics. The Mixture Analysis Tool will not compute the likelihood ratio for a contributor that is not included in the sample, so this is why the tool was only able to compute the LR statistic for 48 of the 138 two contributor samples correctly classified, and therefore a total of 58 LR statistics calculated.

Table 3. Evaluation of the GeneMapper™ ID-X (v1.4) Mixture Analysis Tool sample segregation ability. The Mixture Analysis Tool was more accurate in segregating mixture samples than single source samples.

	1 contributor	2 contributor	3 or more	Total
Incorrect	24	37	23	73
Correct	76	138	463	534
Accuracy	76%	78.86%	95.3%	
# of Potential LR	200	350	NA	71
LR Calculated	23	58	NA	71

Comparison of LR Statistics

For the three probabilistic genotyping systems, the log of the likelihood ratios were calculated for 100 single source samples. Unlike in casework where the true number of contributors is not truly known,

all sample LR values were computed for this study using the true number of contributors due to time restraints. With this in mind, Figure 3 shows that overall CEESIt produced statistically higher LR values than TrueAllele® on the single sample set at the 95% CI, but not when compared to GeneMapper™ ID-X, $p=0.0002902$ and $p=0.58428278$ respectively. GeneMapper™ ID-X also computed higher LR values than TrueAllele ($p=0.03060$). The majority of the contributors from the 100 single source samples were correctly identified, except CEESIt incorrectly determined the contributors for three samples from the 100 samples. TrueAllele®, had a lower misclassification rate since it only misidentified one individual from the 100 analyzed samples. Due to software restrictions with calculating LR statistics for non-contributors, GeneMapper™ ID-X v 1.4 only provided data for 23 of the 100 samples, but all of them correctly identified the individual in the sample.

Figure 3. Log(LR) values for the single source samples from the PROVEDIt database computed using the NIST Caucasian dataset with the three probabilistic genotyping systems. TrueAllele® produced lower log values with lower template quantity, whereas CEESIt and GeneMapper™ ID-X did not seem to be impacted by lower template quantity.

Not only is it important to not incorrectly exclude a contributor from an evidence sample, it is more important to not falsely include a contributor. Therefore, for any interpretation technique, the error rates should be determined. In this study, each probabilistic genotyping system compared each sample profile to two incorrect contributor profiles in order to evaluate how often false inclusions occurred. As expected, as the number of contributors increased, more individuals were incorrectly included as contributors to the sample. Of the three systems, TrueAllele® was the most accurate when it came to excluding the noncontributors from the sample for the one and two contributor samples with a maximum false inclusion rate of 4% as opposed to CEESIt and its 10.6% false inclusion rate. Additionally, it was also more accurate for the four and five contributor samples when compared to the results obtained from CEESIt (Table 4).

Table 4. False inclusion occurrences within each system after a comparison was performed for each sample and two noncontributor profiles. The highest number of false inclusions occurred using the CEESIt software.

Probabilistic Genotyping System	Number of False Inclusions				
	1	2	3	4	5
GeneMapper™ ID-X v1.4	3	3	NA	NA	NA
NOCIt/CEESIt	2	2	10	11	16
TrueAllele®	0	1	6	6	6
Total Number of Samples	100	175	160	165	150
Total Number of LRs	100	350	480	660	750

Probabilistic genotyping systems often compute the LR values on the log scale. If a negative log value is obtained, it indicates that the contributor was not determined to be present within the mixture sample. The false exclusions that were determined by each of the systems for the sample sets can be found in Table 5. There were less false exclusions with fewer contributors. Overall, GeneMapper™ ID-X had more false exclusions where the software would not produce a LR statistic for the 1 and 2-contributor samples. TrueAllele® had more false exclusions when compared to CEESIt with the exception of the five contributor samples (Table 5). The false exclusion rate for the mixture samples increased from 20.2% to 22.2% whereas CEESIt had a larger increase when comparing the rate for four and five contributor samples.

Table 5. Number of false exclusions within each system after all true contributors were compared to the sample profiles. With the exception of the 1-contributor and five-contributor samples TrueAllele® produced the most false exclusions on the sample set.

Probabilistic Genotyping System	Number of False Exclusions				
	1	2	3	4	5
GeneMapper™ ID-X v1.4	67	302	NA	NA	NA
NOCIt/CEESIt	3	2	5	37	171
TrueAllele®	1	71	67	145	167
Total Number of Samples	100	175	160	165	150
Total Number of LRs	100	350	480	660	750

Linear Mixed-Effects Model

In order to account for the ratio of contribution and sample concentration, a linear mixed-effects model was used to determine the effect that each system had on the computed LR values for each sample from the mixture sets. Each software was evaluated on how well it correctly identified all of the individuals present in the mixture samples for the different sample concentration ratios. The majority of the samples that GeneMapper™ ID-X could compute an LR for were for samples with ratios of 1:4 and 1:9. (Figure 4). As seen in Table 6, TrueAllele® computed significantly lower LR values when compared to CEESIt and even GeneMapper™ ID-X ($p < 0.0001$), and also had a larger range of computed values when compared to the other two systems. There was not an observable difference between the values computed for CEESIt and GeneMapper™ ID-X ($p = 0.8391$). It was observed that as template quantity for the contributors increased, then the log LR value increased as well for each software, with the exception for CEESIt when computing the LR for samples with a ratio of 1:4 where there was a decrease.

GeneMapper™ ID-X was only included for the one and two contributor sample set since it is unable to compute a LR value for samples with three or more contributors. So for the interpretation of three or more contributors, GeneMapper™ ID-X v1.4 is not included for comparison purposes.

Figure 4. Log(LR) values for the two contributor mixture samples from the PROVEDIt database computed using NIST Caucasian dataset with the three probabilistic genotyping systems. TrueAllele® computed lower log LR values especially at lower template quantities.

Table 6. Comparison statistics for the three probabilistic genotyping software programs when calculating the LR value for the two contributor samples from the PROVEDIt database. There was no statistically significant difference between CEESIt and GeneMapper™ ID-X, but a statistical difference was found between GeneMapper™ ID-X and TrueAllele® and CEESIt and TrueAllele® for this sample set.

Contrast	Estimate	SE	df	t. ratio	p.value
CEESIt – GeneMapper™ ID-X	1.113481	1.973026	629.58	0.564	0.8391
CEESIt – TrueAllele®	15.828250	1.055879	543.60	14.991	<.0001
GeneMapper™ ID-X- TrueAllele®	14.714769	1.973121	629.47	7.458	<.0001

As seen in both Figure 5 and Table 7, the analysis of the three contributor samples found a similar trend that TrueAllele® computes statistically significant lower LR values when compared to CEESIt ($p < 0.0001$) for three contributor samples. CEESIt produced higher LR values, where the average calculated LR in CEESIt was 19.239, 95% CI[18.129,20.348], which was higher than the values obtained from TrueAllele®, which for the three contributor samples averaged 5.78, 95% CI[4.67, 6.895]. Overall, as the quantity of DNA increased for a contributor as did the value of the computed LR for each system.

Figure 5. Log(LR) values for the three contributor mixture samples from the PROVEDIt database computed using NIST Caucasian dataset using CEESIt and TrueAllele®. TrueAllele® computed lower log LR values especially at lower template quantities.

Table 7. Comparison statistics for the three contributor samples for the log LR values after being evaluated on CEESIt and TrueAllele®. There was a significant difference in the LR values produced from both systems.

Contrast	Average	SE	df	Lower CI	Upper CI	t. ratio	p.value
CEESIt	19.239092	0.564277	363.37	18.1294332	20.348752	17.857	<0.0001
TrueAllele®	5.784972	0.5642647	363.37	4.675337	6.894606		

As for the four contributor samples, CEESIt computed significantly larger LR values for the 165 samples in the four-contributor set when compared to TrueAllele® ($p < 0.0001$), as seen in Table 8. However, for both systems as template quantity increased there was an observable and significant increase in the LR value calculated by each system (Figure 6).

Figure 6. Log(LR) values for the four contributor mixture samples from the PROVEDIt database computed using NIST Caucasian dataset using CEESIt and TrueAllele®. TrueAllele® computed lower log LR values compared to CEESIt especially at lower template quantities.

Table 8. Comparison statistics for the four contributor samples for the log LR values after being evaluated on CEESIt and TrueAllele®. There was a significant difference in the LR values produced from both systems.

Contrast	Average	SE	df	Lower CL	Upper CL	t. ratio	p.value
CEESIt	13.787379	0.6932127	209.07	12.420797	15.153962	18.408	<0.0001
TrueAllele®	3.101122	0.6932127	209.07	1.734539	4.467704		

For the five-contributor sample set, there was a large range of variation found between the two systems that resulted in an insignificant difference between the calculated LR statistics ($p=0.7657$). As shown in Figure 6, the obtained values were inconsistent for both programs for the different ratios and template DNA quantities. There was no distinctive trend for either software. For both CEESIT and TrueAllele®, the average calculated LR was negative, 95% CI[-1.87, 1.45] and CI[-2.56, 0.76], respectively). Both CEESIt and TrueAllele® experienced a larger quantity of misclassifications despite the template quantity for the contributors. TrueAllele®, however, did produce higher LR values when the template quantity increased (Figure 7).

Figure 7. Log(LR) values for the five contributor mixture samples from the PROVEDIt database computed using NIST Caucasian dataset using CEESIt and TrueAllele®. TrueAllele® computed lower log LR values compared to CEESIt, especially at lower template quantities.

Table 9. Table 6. Comparison statistics for the five contributor samples for the log LR values after being evaluated on CEESIt and TrueAllele®. There was not a significant difference in the LR values produced from both systems.

Contrast	Average	SE	df	Lower CL	Upper CL	t. ratio	p.value
CEESIt	0.1922	0.9600	519.07	-1.7066	2.0911	0.298	0.7657
TrueAllele®	-0.1199	0.8957	519.07	-1.8952	1.6552		

Swaminathan et al. (2016) evaluated the effect of contributor number on the calculated LR statistic for the CEESIt software. This evaluation was also performed in this study for both TrueAllele® and CEESIt using one random sample from each contributor category. The selected sample contained a contributor with a template quantity of 0.125 ng. GeneMapper™ ID-X was not included in this evaluation since it was not able to compute the LRs for samples where individuals contributed 0.125ng of DNA for both the 1- or 2-contributor sample sets. TrueAllele® computed a decreasing LR statistic as the number of contributors increased despite the same template quantity, with the exception of the 5-contributor sample. On the other hand, CEESIt had no distinctive trend between the contributor number and the calculated LR statistic (Figure 8).

Figure 8. Comparison LR values obtained from an individual with a 0.125 template quantity for 1-,2-,3-,4-, and 5- contributor samples. The 2-contributor sample had a template quantity of 0.25 ng for a 1:1 mixture ratio. The contributor from the 3-contributor sample came from a sample with a mixture ratio of 1;2;1 and a template quantity of 0.5 ng. For the 4-contributor sample, it came from one of the minor contributors of a mixture with a template quantity of 0.75 and a 1;2;2;1 mixture ratio. Then the final data point from the 5-contributor sample was provided from a contributor from a sample with a template quantity of 0.75 ng and a 1:1:2:1:1 mixture ratio. There is no consistent trend for the samples computed with CEESIt, however, for TrueAllele® as the number of contributors increases the LR decreases despite the contributors template quantity with the exception of the 5-contributor sample.

Sample Processing

GeneMapper™ ID-X computes the LR statistic since there are no cycling parameters. The samples are loaded into the Mixture Analysis Tool, it automatically segregates the samples, and then the analyst can select the sample and perform the desired statistic.

Even though all of the samples that were run through the NOCIt portion of the NOCIt CEESIt software were ran using a selected maximum number of five contributors, the run time varied based on the number of contributors present within the sample, where the longest run time was observed for the three contributor samples (Table 10)

Table 10. Average sample processing time for each sample run through NOCIt for the three population datasets

Contributor Number	Run Time (min)
1	56.70
2	41.96
3	88.03
4	70.93
5	69.53

The amount of time for sample interpretation for a software programs is dependent on the cycling parameters. The running time of CEESIt increases as the number of contributors increases. To compute the LR statistics for all 100 samples, it took approximately 7 minutes, where the

average run time was about 6.5 seconds each. As the contributor number increased, the amount of time to complete the computations for one person of interest increased as well (Table 11). The five contributor samples took the longest time to complete with an average of 3,105.88 minutes or 51.76 hours to pull out one person of interest. CEESIt has been implemented to employ multi-threaded processing to take advantage of multiple cores when available to process the samples as fast as possible (Swaminathan et. al 2016).

Table 11. Average sample processing time for each sample run through CEESIt for the Caucasian population dataset to pull out one contributor and for entire sample processing. As the number of contributors in the sample increased the amount of time to deconvolute the sample increased as well to account for all of the potential genotypic combinations at each locus.

Contributor Number	Run Time per contributor (min)	Run Time per sample (min)
2	35.87	71.75
3	45.48	136.44
4	311.85	1247.40
5	3105.88	15529.44

TrueAllele utilizes a different processing method where the software continuously runs six samples at one time and once it has finished with one of the six samples it immediately converges onto the next sample in queue. The amount of sample processing time depends on the number of MCMC cycles that the user specifies. The samples that were run for this study that utilized 5,000 MCMC cycles were completed in approximately three and a half hours for the six samples. For samples run using the 10,000 MCMC cycles, it averaged approximately 8 hours to complete all six samples.

Discussion

This study evaluated 750 files from the PROVEDIt database on GeneMapper™ ID-X, CEESIt, and, TrueAllele® using the Caucasian population from the NIST dataset. These files consisted of 1-, 2-, 3-, 4-, and 5-contributor samples with template quantities between 0.016 and 0.75 ng. For the NOCIIt portion of the NOCIIt/CEESIt software the African American, Caucasian and Hispanic population datasets from NIST were used (STRBase, 2017).

GeneMapper™ ID-X software interface is very conservative and automatically rejects an individual as a contributor if there is any allelic dropout in the sample since it computes a more restricted set of genotype combinations for the contributors (Hansson and Gill, 2011). It is true that it is better to be conservative to avoid false inclusions of individuals, but the Mixture Analysis Tool should have its parameters to better account for allele sharing and allelic dropout to reduce the number of false exclusions. Updated parameters could also help increase both the segregation aspect and computing capability for calculating an LR statistic.

In a previous study, NOCIIt was found to be more accurate than using a maximum likelihood estimator and maximum allele count for 1-, 2-, 3-, 4-, and 5-contributor samples. For, the 5 contributor samples that were analyzed, NOCIIt performed comparatively to a maximum likelihood estimator (Alfonse et. al 2017). Similarly, to Alfonse et al. (2017), this study also observed a decrease in the accuracy of the software as the number of contributors increased. This is most likely due to allele sharing between contributors, and increased artifacts such as stutter. This study the accuracy decreased from This portion of the software is beneficial since the LR calculation requires the assumption of the number of contributors. With the use of this tool the analyst can have a better idea of the number of contributors to calculate the LR statistic. When using the TrueAllele® Casework system you can create multiple requests using different contributor values to help with sample interpretation, but having a statistical software help aid with this process is beneficial and has the potential to reduce the amount of time for sample interpretation.

Instances where negative log values were calculated occurred more often where there was a low quantity from a contributor that resulted in allelic dropout and/or artifacts such as stutter. In contrast to Swaminathan et. al (2016), Figure 8 shows that there was an inconsistent trend in the value of the LR when comparing the same contributor template quantity for one to five contributor samples, with the exception of four contributor samples. This trend was observed on the calculated LR values that were equally as high and in some instances higher for the 4-contributor samples compared to the 3-contributor samples of equal template DNA quantity (Figure 5 and Figure 4 respectively). Across each mixture sample set CEESIt did compute higher LR values compared to TrueAllele® indicating a higher degree of certainty when indicating the person of interest as a contributor to the sample.

TrueAllele® has been repeatedly validated for use on mixtures containing five contributors and has even been validated for samples containing 10 contributors (Bauer et. al, 2020). Those validation studies have found TrueAllele® to reliably deconvolute mixture samples. The lower LR values obtained in this study could be attributed to the number of MCMC cycles used during this study. Increasing the number of cycles could improve the LR values obtained since longer MCMC sampling has been found to improve the program's ability to deconvolute samples and make genotype inferences (Bauer et. al, 2020). TrueAllele® computed fewer false inclusions when compared to CEESIt, however TrueAllele® also reported more false exclusions.

The LR value comparison was performed with randomly selected samples, it is possible the randomly selected sample could have been an outlier to the sample set which could account for the lack of

trend observed in the CEESIt software and the increased LR value for TrueAllele® software. The segregation accuracy of GeneMapper™ ID-X was lower for one contributor samples as opposed to two contributor samples. Since the samples were analyzed at 50 RFU it is possible that the artifacts present in the sample were above that threshold and included at each locus resulting in the incorrect segregation of the samples. This could also explain the accuracy increase from one-contributor to two-contributor samples as well. NOCIt experienced a decrease in accuracy as the contributor number increased, but it also experienced variation across each population. This variation can be explained by the Monte Carlo sampling algorithm. Variation is to be expected when using a sampling algorithm depending on the random set of samples in the distribution that are used (van Ravenzwaaij, et al. 2018).

All three software programs were user friendly and utilize step by step interfaces to guide the analyst through the analytical scheme of the software. However, TrueAllele® does require more preparation because there are more quality control checks within the analytical scheme of the software. Before samples are uploaded to the data module of the software, a request is created, and an LR statistic is calculated. On the other hand, CEESIt requires less preparation per run since the calibration file that is used to inform the model's parameters for processing samples would be created for the validation of the software, and then it can then be used for all future samples along with the created population datasets. However, the NOCIt/CEESIt computational tool takes up nearly all of the computer's processing power while TrueAllele® takes up minimal processing. GeneMapper™ ID-X only requires the analyst to validate the set peak height parameters that are going to be used for analysis, and it easily walks the analyst through the mixture interpretation process.

Probabilistic genotyping systems are software programs that have been created using algorithms to compute a LR value for selected samples to aid analysts with DNA interpretation. Software programs that often utilize a Bayesian approach help to factor in parameters such as allelic dropout and artifacts so that they can focus on the important factor at hand, which are the genotypes present within the sample (Bauer et al, 2020). Evaluating the posterior genotype probability against the prior probability produces the LR, which is useful to determine the presence or absence of an individual within an evidence sample, and provides the statistical weight to the evidence sample.

This study compared TrueAllele®, CEESIt, and GeneMapper™ ID-X to determine how sample interpretation differs when the same set of samples are analyzed using these different methods. The results from this study indicated that even though both TrueAllele® and CEESIt are continuous systems, the cycling parameters and analytical scheme have an impact on the final interpretation result. This was also observed in a study conducted by Swaminathan et al. (2018) where four model variants affected sample interpretation and found that the parameters used for analysis can have a significant impact the calculated LR value and the overall sample interpretation.

Overall, CEESIt computed significantly larger LR values indicating higher certainty of the presence of the selected individual within samples. However, it also had a higher number of false inclusions when compared to TrueAllele®. The argument can be made that it is better to have a higher false negative rather than a higher false inclusion to prevent the incarceration of an innocent individual. Additionally, based on the results from the study, the Mixture Analysis Tool in GeneMapper™ ID-X, which operates as a deconvolution system, is user friendly, but it should only be used as an interpretation aid to the analyst due to its limited capabilities, and all data should be reviewed before making final interpretation decisions.

Many future studies can be built off of this work. For this study the default settings were used to analyze the samples, but these parameters could be manipulated more to determine the most effective parameters to analyze casework type samples. Evaluating different MCMC cycles in TrueAllele® would be beneficial to determine the best settings for analyzing various casework type samples. TrueAllele® has

additional features in the review module that could have been evaluated such as its determination for the mixture weights in the sample, and the systems deconvolution capabilities (Perlin et al. 2015). Different allele frequencies can result in different final LR values so it would be worth evaluating different population datasets to determine how it affects overall sample interpretation. I used the NIST population dataset so it would be worth comparing it to the FBI population datasets.

The results from this study suggest that future studies should consist of more comparisons between different continuous systems such as TrueAllele® and STRMix® which are two competing commercial probabilistic genotyping software systems that are available to DNA laboratories to use on casework. Following recommendations from Bauer et al. (2020) more MCMC cycles should be used when making the comparison between STRMix and TrueAllele®. Based off the performance of the NOCI/CEESIt software it is also worth evaluating further and comparing to other probabilistic genotyping software. A new study can be conducted to evaluate performance and determine the software programs strengths and weaknesses, which can provide DNA laboratories with important information to help them select the best system and to help ensure that it is being used effectively and accurately by the laboratory. The use of probabilistic genotyping software help to provide a more accurate and objective approach to mixture interpretation which follows the recommendation from both the President's Council of Advisors on Science and Technology (PCAST) report published in 2016 and the National Academy of Science (NAS) report published in 2009. As mixture interpretation becomes more complex, whether from an increase in the number of contributors or low template quantities which can result in allelic dropout, these software programs can help to provide a useful aid to mixture interpretation. It is important to use reliable and thoroughly investigated techniques when analyzing forensic DNA evidence. It is important for analysts to utilize a software that has a low false inclusion rate and a low false negative rate to ensure only those who actually provided DNA to a mixture sample are being implicated.

Works Cited

1. Alfonse, Lauren E., et al. 2016. "Inferring the Number of Contributors to Complex DNA Mixtures Using Three Methods: Exploring the Limits of Low-Template DNA Interpretation." *Journal of Forensic Sciences*, 62(2):308–316.
2. Bauer DW, Butt N, Hornyak J, Perlin MW. 2020 "Validating TrueAllele" Interpretation of DNA Mixtures Containing up to Ten Unknown Contributors" *Journal of Forensic Sciences*, 65(2):380–398.
3. Bieber, Frederick R., et al. 2016. "Evaluation of Forensic DNA Mixture Evidence: Protocol for Evaluation, Interpretation, and Statistical Calculations Using the Combined Probability of Inclusion." *BMC Genetics*. 17(1).
4. Budowle B, Onorato AJ, Callaghan TF, Manna AD, Gross AM, Guerrieri RA, Luttman JC, Mcclure DL. 2009. Mixture Interpretation: Defining the Relevant Features for Guidelines for the Assessment of Mixed DNA Profiles in Forensic Casework. *Journal of Forensic Sciences*. 54(4):810–821.
5. Butler JM. 2015. Advanced topics in forensic DNA typing: interpretation. Amsterdam: Elsevier;
6. Butler JM. 2010. Fundamentals of forensic DNA typing. Amsterdam: Academic Press/Elsevier.
7. Clayton T, Whitaker J, Sparkes R, Gill P. 1998. Analysis and interpretation of mixed forensic stains using DNA STR profiling. *Forensic Science International*. 91(1):55–70.
8. Curran JM, Triggs CM, Buckleton JS, Weir BS. 1999. Interpreting DNA mixtures in structured populations. *J Forensic Sci*. 44(5):987–95.
9. Egeland T, Dalen I, Mostad PF. 2003. Estimating the number of contributors to a DNA profile. *Int J Legal Med* 117(5):271–5.
10. Hansson O, Gill P. 2011 Evaluation of GeneMapper™ID-X Mixture Analysis tool. *Forensic Science International: Genetics Supplement Series*.;3(1).
11. Kelly H, Bright J-A, Buckleton JS, Curran JM. A comparison of statistical models for the analysis of complex forensic DNA profiles. *Science & Justice*. 2014;54(1):66–70.
12. Federal Bureau of Investigation (2020) Quality assurance standards for forensic DNA testing laboratories, *Forensic Science Communications* [Online]. (July 2020). Available: <https://www.fbi.gov/file-repository/quality-assurance-standards-for-forensic-dna-testing-laboratories.pdf/view>
13. National Research Council (U.S.). 1996. The Evaluation of Forensic DNA Evidence. Washington, D.C.: National Academy Press. 1-138
14. Gianelli, P.C, Perlin, Mark. 2015. "Inclusion probability for DNA mixtures is a subjective one-sided match statistic unrelated to identification". *Journal of Pathology Informatics*. (6);59.
15. Marshall University Forensic Science Center. 2008. "The Use of GeneMapper™ID-X Software v1.1 for the Analysis of Mixed DNA Samples. National Institute of Justice.
16. Perlin MW, Dormer K, Hornyak J, Schiermeier-Wood L, Greenspoon S. 2014. TrueAllele" Casework on Virginia DNA mixture evidence: computer and manual interpretation in 72 reported criminal cases. *PLoS ONE*. 9(3)
17. PROVEDIt. Laboratory for Forensic Technology Development & Integration. [accessed 2018 Apr 25]. <https://lftdi.camden.rutgers.edu/provedit/>
18. R Core Team R: A Language and Environment for Statistical Computing
19. R Foundation for Statistical Computing, Vienna, Austria (2016). ISBN: 3-900051-07-0.
20. Swaminathan H, Grgicak CM, Médard M, Lun DS. 2015. "NOCIt: A Computational Method to Infer the Number of Contributors to DNA Samples Analyzed by STR Genotyping". *Forensic Science International: Genetics*. 16:172-80.
21. Swaminathan H, Qureshi MO, Grgicak CM, Duffy K, Lun DS (2018) Four model variants within a continuous forensic DNA mixture interpretation framework: Effects on evidential inference and reporting. *PLoS ONE* 13(11).

22. SWGDAM. 2017 SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories.
23. SWGDAM. 2015. SWGDAM Guidelines for the Validation of Probabilistic Genotyping Systems.
24. STRBase. 2017 miniSTR Homepage. [accessed 2018 May 9]. https://strbase.nist.gov/tri_tab.htm
25. Thompson, W. C., Ford, S., Doom, T. E., Raymer, M. L., & Krane, D. E. (2003). Evaluating Forensic DNA Evidence: Essential Elements of a Competent Defense Review. *The Champion*, 27 (16), 16-25.