PEELING AWAY UNCERTAINTY: A PROBABILISTIC APPROACH TO DNA MIXTURE DECONVOLUTION

by

Hajara S. Chaudhry

A Thesis
Submitted to the
Forensic Science Forensic Research Committee
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Master of Science
Forensic Science, Forensic Biology Analysis

Primary Research Advisor
Dr. Mark R. Wilson
Assistant Professor
GMU Forensic Science Program

Secondary Research Advisor
Dr. Jack Ballantyne
Professor, Department of Chemistry
University of Central Florida

Associate Director of Research
National Center for Forensic Science

GMU Graduate Research Coordinator
Dr. Joseph A. DiZinno
Associate Professor
GMU Forensic Science Program

Fall Semester 2020
George Mason University
Fairfax, VA

ACKNOWLEDGEMENTS

I would like to thank my parents for supporting me throughout my academic career. I would also like to thank Dr. Diep Ca and Dr. Laura Haubrick for motivating me to pursue a double-major in Biology during my undergraduate studies from Shenandoah University. Dr. Mark R. Wilson and Dr. Jack Ballantyne served as members of my research committee and were of invaluable support by providing exceptional guidance throughout the course of my thesis. Lastly, I would like to thank William Allan from Cybergenetics who generously provided software training and oversaw the analysis of results and John Donahue from the Beaufort County Sheriff's Office who provided the dataset used in this study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND/OR SYMBOLS

Base Pair ..... bp
Capillary Electrophoresis ..... CE
Deoxyribonucleic Acid ..... DNA
International Society for Forensic Genetics ..... ISFG
Known Reference ..... K
Kullback-Leibler ..... KL
Likelihood Ratio ..... LR
Logarithm Base 10 of Likelihood Ratio ..... log(LR)
Markov Chain Monte Carlo ..... MCMC
Microliter ..... μL
Nanogram ..... ng
National Institute of Standards and Technology ..... NIST
Polymerase Chain Reaction ..... PCR
Questioned Evidence ..... Q
Random Population ..... R
Relative Fluorescent Unit ..... RFU
Short Tandem Repeat ..... STR
Subject Reference ..... S
Visual User Interface ..... VUIer

ABSTRACT

PEELING AWAY UNCERTAINTY: A PROBABILISTIC APPROACH TO DNA MIXTURE DECONVOLUTION

Hajara S. Chaudhry, M.S.

George Mason University, 2020

Thesis Director: Dr. Mark R. Wilson

Mixture deconvolution involves the ability to reliably decipher and separate component genotypes of individual contributors at each tested genetic marker. The ultimate objective of this study is to develop an understanding of the integrated framework for attesting the value of using known samples when appropriate to decrease uncertainty in mixture deconvolution by leveraging more of the available genotyping data and observing the impact genotype conditioning has on multiple-contributor mixtures and resulting LRs. In this study known mixtures containing two, three, four, and five contributors were separated in iterative analyses through the assumption of contributors using provided known reference samples, a process referred to as genotype peeling or genotype conditioning. To direct the order of genotype conditioning, contributor mixture weights were estimated as all contributors to the mixture were assumed by mixture weight. Conditioning by match statistic was directed without genotype assumptions, where all

contributor genotypes were inferred solely on STR peak height data. Subsequent analyses of each mixture item were conducted, in which, the order of contributors was assumed from highest to lowest based on mixture weight as well as match statistics by utilizing a probabilistic program, TrueAllele®, developed by Cybergenetics. The study demonstrates how genotype conditioning effects mixture deconvolution and resulting match statistics by also considering mixture weight and the number of contributors to a mixture. The results of this study demonstrate that it is possible to generate more informative statistics by refocusing probability distributions for each contributor to the original mixture, leading to refined LRs and reduced uncertainty.

Keywords: forensic science, genotype conditioning, DNA mixture deconvolution, probabilistic genotyping, likelihood ratio, Bayesian framework

INTRODUCTION

The capabilities of individual forensic laboratories across the nation vary in their ability to process complex mixtures, ranging from manual and probabilistic approaches. Complex mixtures overpower interpretation procedures in forensic laboratories using traditional manual approaches simply because the technology falls short due to the high levels of uncertainty present, affecting the reliability in determining contributor genotypes. The prevalence of uncertainty can yield an inconclusive result, since it is the only scientifically responsible conclusion to circumvent the chance of inappropriately including or excluding a potential contributor, leading to perpetrators that remain unidentified. Moreover, recent studies demonstrate continuous models that incorporate a Bayesian framework overcome this issue by comparing experimental data against random expectations, where minor contributors with less pronounced profiles can be distinctly and properly accounted for. To further refine match statistics for each contributor, mixture deconvolution performed in subsequent analyses of the same data can be conducted using provided known reference samples in a process referred to as genotype peeling or genotype conditioning. As contributors are assumed in iteration, the sensitivity of residual contributors improves since the genotyping system can more accurately detect true contributors to a mixture. The genotype conditioning interpretation technique has the potential to satisfy this need by maximizing the information obtained from complex

mixtures while maintaining a high standard of reliability and reproducibility. Such an initiative is crucial as DNA samples are often mixed and contributors may be present in such low quantities that the reliability of typing under these conditions is often compromised, limiting the ability to draw informative conclusions that generate essential investigative leads. Genotype conditioning serves as a valuable advantage in these instances by improving sensitivity as layers of uncertainty are peeled away based on mixture weight and match statistics, allowing for full use of the available experimental data.

Mixture Deconvolution

Current Methods

Short tandem repeats (STRs) are small stretches of DNA sequences of three to five nucleotides in length that are repeated numerous times in tandem with the number of repeats varying among individuals, making STRs effective for human identification purposes. Forensic DNA analysis targets a core set of STR markers located in non-coding regions of the human genome permitting genetic information at respective loci to be compared. The evaluation of STRs has been conducted routinely through capillary electrophoresis (CE) and is currently the gold standard in the forensic DNA community. DNA templates are amplified with polymerase chain reaction (PCR) in multiplex, where primers contain a fluorescent label that is incorporated into all PCR products during amplification. The PCR products become size-separated by traveling through the capillary array in the CE platform, in which, smaller molecules migrate faster than larger molecules. Upon reaching the detection window of the capillary array, molecules are

detected as a laser strikes the fluorescent label located on each PCR amplicon. CE translates machine measured DNA-molecule migration times into DNA fragment lengths, which are visualized in an electropherogram where peak heights correspond to fluorescent intensities measured in relative fluorescence units (RFUs) representing each fragment length. Evidentiary data is loaded into the genotyping software along with an allelic ladder supplied with a forensic commercial kit, both containing the same internal size standard to enable the correlation of typing results. The allelic ladder is a collection of the common alleles present in the population for each individual STR marker, serving as a standard for allele designation. It is constructed by combining locus-specific PCR products from individuals within the population, representing the variation of alleles for each locus used in STR profiling. Kit-specific bins and panels define the allele repeat number for each STR locus and define the STR loci present in the kit, respectively. When combined with the allelic ladder data file, bins and panels provide genotyping software the capability to transform DNA size information into an STR allele repeat number for each observed peak. By comparison to an allelic ladder, length variant STR alleles are then cumulated into pairs as genotypes at each tested genetic marker to form a contributor profile.

By following a set of interpretation rules called match criteria, the DNA profile interpretation process is conducted to determine the source of the genetic information. An overview of the steps in the DNA interpretation process is described in Figure 1. The questioned evidence sample (Q) and the known reference sample (K) are both processed from peak to profile. DNA profiles produced can then be compared in the interpretation

process and conclusions can be made based on this comparison. An analyst must make decisions based on electropherogram data to differentiate between allele peaks, noise, or artifacts as well as whether alleles can be confidently paired to form genotypes, which are combined to create a contributor profile from each locus. Additionally, it must be determined whether the data are too weak or too complex to be reliably interpreted and if overall data quality is appropriate for obtaining reliable results (Butler, 2015). If a match between the evidence sample and the known reference sample is determined, the match probability is calculated to assess the strength of the match or the weight of evidence, which can be statistically expressed as a Likelihood Ratio (LR). The match statistic provides a comparison of the probabilities of the evidence under two mutually exclusive hypotheses, representing how probable the link between the evidence and the suspect is than coincidence. A strong DNA typing result is denoted by large LR values, indicating that the comparison between the Q and K samples provides support for inclusion. Whereas LR values less than one denote that the comparison provides support for exclusion.

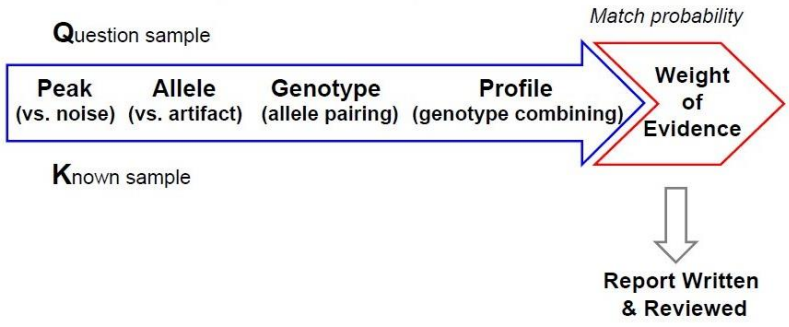

Figure 1. DNA Interpretation Process. The questioned evidence sample (Q) and the known reference sample (K) are processed from peak to profile and then compared. The match probability is calculated to assess the weight of evidence and a report is written describing the results obtained. A technical peer-review is conducted before the report is finalized and released by the laboratory. (Butler, 2015)

STR data analysis is typically conducted by implementing interpretation thresholds that are intended to cut out baseline noise and other PCR artifacts to facilitate and streamline analysis. An analyst reviews electropherogram data containing peak height information subjectively by abiding by these thresholds to manually designate alleles. Interpretation methods that use thresholds to make decisions are referred to as binary and instill an all-or-nothing approach, in which alleles and their cumulative genotypes as pairs are either included (probability of one) or excluded (probability of zero). Threshold-based analysis methods are based on qualitative peak height information and introduces subjectivity when applied to the data. By only reviewing a subset of the evidentiary data, all genotyping information is not put to full use and valuable genotyping data below established thresholds are consequently discarded. Therefore, there is a danger that contextual bias can yield a DNA interpretation that is not objective. Data selection can overstate the probative value of a match, leading to potentially inaccurate results that mislead juries.

In forensic DNA laboratories, many evidence samples are mixtures often containing low-template DNA and consist of more than three contributors, termed complex mixtures. Mixture interpretation involves the designation of alleles and possible genotypes from each contributor followed by an assessment of the statistical weight of evidence after the comparison between the questioned evidence profile and a known profile can be made. As more contributors are present, the number of possible allele combinations increase. While there are numerous possible alleles at each STR locus, some alleles are more common than others, thus are more likely to be present and hence

shared between two or more individuals. Figure 2 illustrates the potential allele overlap between contributors leading to stacking of signal, making it challenging to unambiguously distinguish contributor genotypes because an analyst cannot state with absolute certainty that a particular peak belongs to a particular contributor or not.

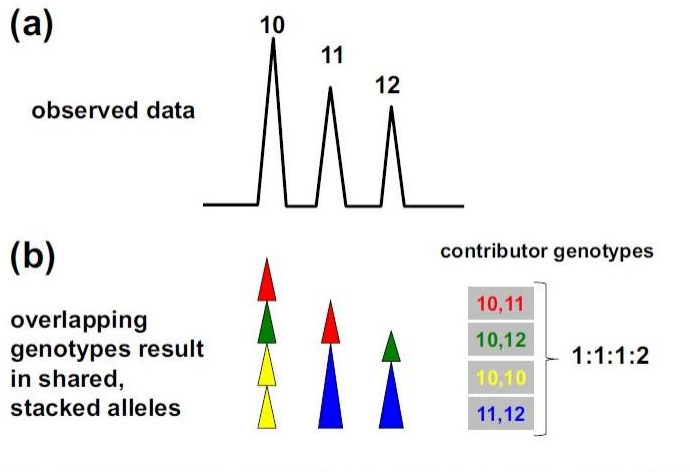

Figure 2. Allele Stacking. (a) Hypothetically observed data from a complex mixture at a single STR locus, exhibiting alleles 10, 11, and 12. (b) A depiction of allele stacking from a set of contributor genotypes present in a 1:1:1:2 mixture weight ratio. (Butler, 2015)

Stochastic effects, such as allelic dropout and heterozygote peak height imbalance, are referred to as characteristics that occur during PCR amplification as a result of the variation in random sampling. Stochastic effects are characteristic of low-template DNA and regarded as non-reproducible since they manifest as a fluctuation of results between replicate analyses of the amplified data. Since PCR is a random enzymatic process, competition for DNA polymerase favors alleles present in greater amounts. Some PCR primers may fail to amplify a particular allele due to possible variation in the STR flanking regions where PCR primers are designed to anneal. Less

amplified alleles tend to correlate with low-template contributors and are prone to stochastic effects such as allelic dropout, where one or more alleles are absent from a locus or falls below threshold limits, further complicating allele designation. There is a general trend between DNA template amounts and its effect on heterozygote peak height imbalance, leading to allelic dropout as described in Figure 3. As DNA template levels decrease, the peak height imbalance of heterozygous alleles increases until one of the alleles fails to be amplified, resulting in severe peak height imbalance where an allele drops out. PCR reactions are often performed with 1ng or less of total DNA, thus complex mixtures are likely to contain low-template DNA for one or more contributors. Each additional contributor to a mixture signifies a dilution of one or more contributors into the stochastic zone, where allelic dropout is evident. This phenomenon becomes problematic especially in the interpretation of complex mixtures, where multiple peaks are exhibited at each locus and results in uncertainty in the designation of allele pairs corresponding to contributor genotypes.

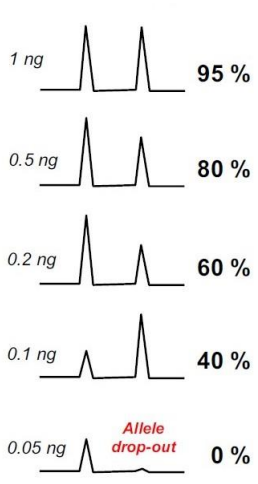

Figure 3. Peak Height Imbalance Leading to Allelic Dropout.

A display of hypothetical heterozygous alleles showing a general trend that peak height imbalance increases as DNA template amounts decrease, eventually leading to allelic dropout. When allelic dropout occurs, a false homozygote is observed. Hence, an analyst may not be able to accurately determine a contributor genotype. Peak height ratios are shown as a percent of the shorter allele to the taller allele. The concentrations of DNA templates and respective peak height ratios are displayed in descending order, from top to bottom. (Butler, 2015)

Furthermore, threshold-based methods tend to work best with pristine samples such as profiles containing a single contributor, termed single-source profiles, or simple two-contributor mixtures in which stochastic effects are minimal and DNA is present in optimal amounts. As uncertainty arises due to an indication of multiple contributors especially with low-template DNA amounts, the robustness of the threshold-based method is inadequate because, consequently, peaks from low-template contributors often do not meet desired thresholds and therefore become completely disregarded. Hence, manual interpretation methods originally developed for pristine samples are not as effective on mixture data and interpretation errors are more likely to occur by simplifying complex data from mixed evidentiary samples (Perlin et al., 2015). The variation in random sampling during PCR amplification introduces uncertainty in the interpretation of complex mixtures largely because of the characteristics and nature of such mixtures, affecting reliable interpretation. Since more uncertainty in the data exists near a threshold boundary than further from it, an advantage in probabilistic methods is that the results reflect interpretation uncertainty.

The Mixture Issue

As DNA typing technologies and STR multiplex chemistries become more sensitive, laboratories are becoming increasingly overwhelmed with low-quantity and low-quality samples as well as complex mixtures involving more than two contributors. In 2013, the National Institute of Standards and Technology (NIST) conducted an inter-laboratory study involving the interpretation of complex mixtures by sending mixture data to various labs (Butler et al., 2018). Results of the study concluded that a high level

of variation is present in complex mixture deconvolution across different laboratories as well as within a particular laboratory, specifically, the results indicated variation in the inclusion and exclusion of contributors. In the study, 100 participants examined a three-contributor mixture that did not contain a particular suspect. Seventy groups falsely included this suspect, whose DNA was not present in the mixture, representing a 70% false match rate and provided irrelevant DNA match statistics that ranged from 9 to 344,000 (Perlin, 2016). Twenty-four labs found the comparison inconclusive and only six groups correctly excluded the suspect contributing to a 6% accuracy rate, with one of the groups using Cybergenetics TrueAllele method (Perlin, 2016). These results were showed as evidence of the dangers of misinterpreting mixtures, initiating the involvement of the International Society for Forensic Genetics (ISFG) to concur recommendations for the interpretation of complex mixtures.

Scientists feared that the inaccurate interpretation of complex mixtures would wrongfully include or exclude an individual, which would result in wrongful convictions. Therefore, many forensic laboratories halted the interpretation of complex mixtures containing three or more contributors due to the variation associated in subjectively interpreting profiles. Because of this, some laboratories only separate two-person mixtures if a profile subtraction can be conducted, where the profile for one contributor is known and is subtracted from the evidentiary mixture to simplify the interpretation of the genotypes present. Upon the encounter of a mixture involving three contributors or more, the evidentiary sample is often reported as inconclusive without any evaluation whatsoever, hence any individual within such mixtures evades identification. By omitting

informative genotyping data, an inconclusive result can hinder investigative leads that could implicate the guilty or exonerate the innocent.

TrueAllele

Bayesian Framework

Probabilistic genotyping refers to the use of biological modeling and statistical theory to infer genotypes and then calculate resulting LRs for each contributor to a mixture. TrueAllele is a probabilistic genotyping program developed by Cybergenetics to mitigate stochastic effects and capture the uncertainty of DNA profile interpretation using probabilities. Foundationally, continuous models are designed to address stochastic effects by modeling the sources of variation, eliminating the subjectivity and inconsistency that can result from the manual designation of alleles. Continuous models treat data as varying continuously by modeling uncertainty and measuring parameter values based on their means and variances. The Bayesian framework incorporates data models and handles the inherent uncertainty using a conditional approach. This mathematical framework involves three interrelated components: posterior probabilities representing the outcome, prior probabilities serving as the context, and centering around a likelihood function that separates each outcome when applied to the data. A representation of these components and their interrelationships are illustrated in Figure 4.

Prior probabilities are calculated using a particular population's allele frequencies and express the probability of each genotype within that population. A likelihood function is applied to the data and is combined with prior probabilities to assign a likelihood to each possible genotype. Posterior probabilities arise as the output from a

likelihood function acting on the priors. Using the data and a likelihood function, priors then become transformed into posteriors and are redistributed into all the possible genotype outcomes as probabilities of each possible genotype. An LR is the match statistic that is calculated as a ratio of the posterior and prior probabilities.

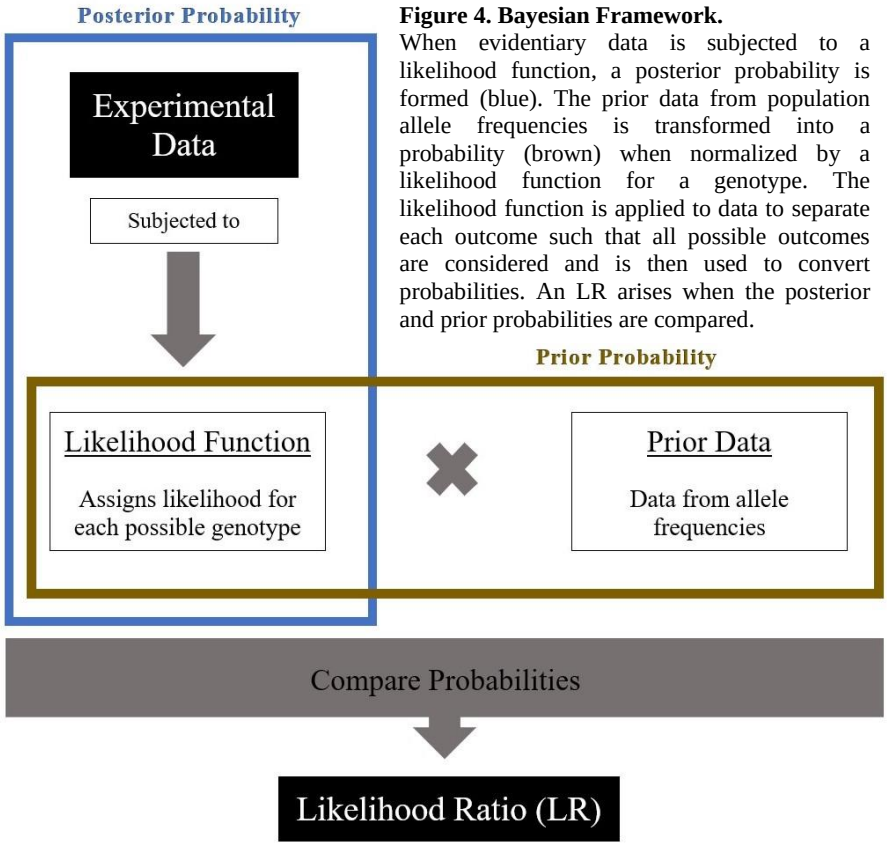

Figure 4. Bayesian Framework.
When evidentiary data is subjected to a likelihood function, a posterior probability is formed (blue). The prior data from population allele frequencies is transformed into a probability (brown) when normalized by a likelihood function for a genotype. The likelihood function is applied to data to separate each outcome such that all possible outcomes are considered and is then used to convert probabilities. An LR arises when the posterior and prior probabilities are compared.

In terms of mixture deconvolution, a posterior probability is calculated for each genotype and contributor at each locus as the likelihood function is updated with data. By utilizing statistical sampling, the experimental data transitions from an initial state of knowledge to a modified state and hypotheses are either accepted or rejected based on

how well they explain the experimental data. Since all possible genotypes are considered objectively, the framework obviates the need for data thresholds as used in manual interpretation methods. By expressing all possible genotype outcomes as distinct probabilities, continuous methodologies leverage more of the available genotyping data and have proven to produce more informative statistics with respect to contributors and non-contributors to the DNA evidence (Wilson, 2018). In doing so, probabilistic genotyping software allows for better interpretation of profiles and provides substantial advantages for interpreting complex DNA mixtures.

Markov Chain Monte Carlo

TrueAllele uses a hierarchical Bayesian probability model that combines genotype alleles, accounts for artifacts, measures variance to explain STR data, and derive parameter values and their uncertainty (Perlin, 2011). To accommodate these factors, TrueAllele is constructed using continuous models and employs statistical sampling to explore the probability space. TrueAllele's statistical sampling method, Markov Chain Monte Carlo (MCMC), is a probability tool used to determine the best possible explanation of the experimental data out of all possibilities by searching through a hierarchical alignment of all the relevant variables. MCMC samples from a posterior probability distribution and explores explanatory patterns in a high-dimensional probability space, where all variables are visited in each sampling cycle. Upon entering probability space, Markov chains begin in an initial state and then move to other states using probability calculations. Transitions to other states are either accepted or rejected and occur based on the ratio between the probability of the next state and the current

state, referred to as the transition probability. If this probability is higher in the next state, the computer moves to that region of total space. Since MCMC explores all possible patterns that explain the data based on relative probabilities, Markov chains may still transition if the probability of the next state is less than the current state in some instances.

TrueAllele determines the path that best explains the data as a whole and collectively applies that pattern to each locus. Local probability choices lead to a global solution that saturates the entire probability space resulting in a joint probability distribution, where all the variables visited are contained in a high-dimensional space. To solve for these variables, the dimensionality is reduced through marginalization to separate out each variable. This is done by summing up each variable across all dimensions over all the other variables, resulting in a reduction to a single dimension. By marginalizing the joint probability distribution, a probability distribution is obtained, where TrueAllele separates the data into its component variables including each objectively inferred genotype for every locus from each separated contributor.

Computing Likelihood Ratios

After genotype inference, a match statistic is calculated between a separated evidence genotype and a reference genotype, relative to a population. To assess the TrueAllele genotype separation, electropherogram data can be viewed in conjunction with the genotype information and match statistics at each locus, as shown in Figure 5. The Data View pane displays the original STR electropherogram data of the sample of interest at a particular locus. Allele peaks are depicted by fragment size in base pairs (bp)

and their corresponding RFUs. The original STR data is inspected to ensure that the necessary interpretation requests were performed and may provide an explanation of any possible source of ambiguity in the inferred genotypes. Genotype probability distributions and match statistics can be viewed when a questioned evidence sample (Q), subject reference (S), and a random population database (R) are selected.

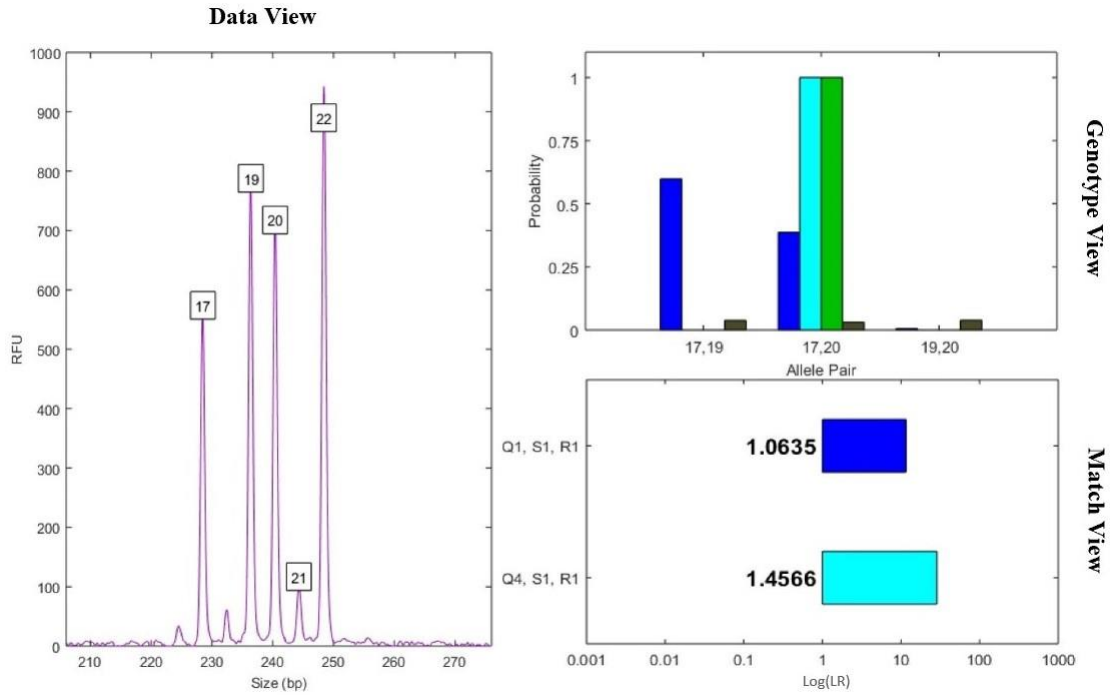

Figure 5. Genotype Inference and Computing LRs. Data View (left): Original electropherogram data of a two-contributor mixture with a 1:1 mixture weight ratio displayed as size (bp) vs. RFU. Alleles 17, 19, 20, 21, and 22 are exhibited at a peak height cutoff of 100 RFU. Genotype View (top, right): The evidence (blue), reference (green), and population (brown) genotypes for one contributor are shown. The evidence (cyan) depicts the improved probability of the reference genotype (17, 20) after one contributor was assumed by genotype conditioning. Match View (bottom, right): Subject reference (S1) has a log(LR) value of 1.0635 when matched against the inferred contributor (Q1) shown in blue. Subject reference (S1) displays an improved log(LR) of 1.4566 when matched against the inferred contributor (Q4) after one contributor was assumed in genotype peeling, shown in cyan.

In the Genotype View pane genotyping information is displayed as a histogram at each locus, where the x-axis shows all possible allele pairs for the genotype probability distribution and the y-axis represents the probability associated with each possible allele pair. In the histogram, blue bars represent posterior probabilities of Q genotypes inferred by TrueAllele and brown bars represent the prior probabilities of each possible genotype defined by the selected population. The green bar indicates the genotype probability distribution of the subject reference. Posterior probabilities are divided by the prior probabilities for each possible allele pair to form a ratio of probabilities and a locus LR is calculated. Since LR values can be rather large, it is often preferable to report them in terms of their logarithm (log) value, in which, a single log(LR) unit is referred to as ban.

In the Match View pane, the x-axis shows the relative match statistics that can be viewed on a logarithmic or linear scale. The y-axis exhibits the specific combinations of Q, S, and R used in each LR calculation and refers to the selected genotypes. A blue bar extended to the right indicates a positive match score displaying log(LR) values greater than zero, which equates to LR values greater than one. A red bar extended to the left indicates a negative match score displaying log(LR) values less than zero and equate to LR values less than one. For a particular contributor, the calculated log(LR) values at each locus are summed together to form the genotype log(LR). This calculation is conducted for each contributor to the mixture, in which genotype LRs from different individuals can then be compared and considered as inclusionary or exclusionary. The comparison between posterior and prior probabilities establishes the weight of evidence for statistical assessment and represents the amount of information present. Positive

log(LR) values provide support for inclusion and display a gain in information as the experimental data transitions from an initial state of knowledge (prior) to a modified state (posterior). Likewise, negative log(LR) values provide support for exclusion and display no information gain. Additionally, log(LR) values closer to zero can be considered as inconclusive. Identification information is summarized as a single number, which facilitates comparisons and courtroom presentations. The objectivity derived from the automated process ensures consistency, leading to more reliable and enhanced LRs.

Genotype Conditioning. To further refine the LR from mixed evidentiary samples, subsequent analyses of the same data can be conducted using known reference samples in a process referred to as genotype peeling or conditioning. TrueAllele's algorithm is set to assume that a known profile is present and genotype inference is repeated under these new conditions. Known genotypes serve as parameters that remain constant throughout a computer run, facilitating genotype inference for residual contributors. Therefore, uncertainty in mixture deconvolution is reduced with each iterative analysis since probability distributions become refocused on these residual contributors. In sequential genotype peeling, the initial run separates the mixture and component genotypes are inferred. Peeling is then performed iteratively in subsequent analyses to capture and peel away each layer of uncertainty by refocusing probability distributions. In the following run, TrueAllele assumes the first genotype and infers the remaining unknown genotypes. After separation and genotype inference, a second contributor is selected for conditioning based on the resulting data. The third run assumes the first and second contributor genotypes and TrueAllele infers the remaining unknown

genotypes. From the resulting data in the previous run, a third contributor can then be selected for conditioning. Genotype peeling can continue until all contributors are considered. An LR match statistic can then be calculated to determine the strength of the match, measured on a logarithmic scale. A contributor's log(LR) plateaus at its maximum value when the evidence genotype probability reaches 100% (Bauer et al., 2020). In genotype peeling a log(LR) cutoff does not exist, thus an unhelpful genotype assumption may either depress a match statistic or leave it unchanged.

Objectives

This study assessed mixtures containing two, three, four, and five contributors using the TrueAllele program to address the effect of genotype peeling on mixture deconvolution and resulting LRs. Constructed using continuous models, probabilistic genotyping was incorporated to provide mixture deconvolution capabilities while enabling full use of the data. The study aims to attest the value of using known reference samples when appropriate to decrease uncertainty in mixture deconvolution by leveraging more of the available genotyping data, as well as to observe the impact that genotype conditioning has on multiple-contributor mixtures and resulting match statistics. The anticipated results include the generation of more informative statistics by refocusing probability distributions for each contributor to the original mixture, leading to refined LRs and reduced uncertainty.

Variables including mixture weight and the number of contributors to a mixture were also considered and were assessed essentially due to their behavior and impact in mixture deconvolution. The relationship between the amount of match information and

the amount of contributor DNA, measured on logarithmic scales, is linear with a predictable slope (Bauer et al., 2020). Equal mixture proportions disrupt this relationship because peak height data becomes less helpful in separating genotypes. Thus, inferred genotype information is restricted resulting in depressed match statistics. Moreover, low-template contributors behave differently than contributors present in optimal amounts. Alleles associated with low-template contributors are present in limited quantities and therefore are more susceptible to potential allelic dropout (Butler, 2015). Since the amount of contributor DNA is directly proportional to the amount of match information obtained, genotype information is limited for low-template contributors. As there are more contributors to a mixture, the amount of DNA becomes further divided for each additional contributor and the number of allele possibilities increase, resulting in more dispersed probability distributions. Evidently, mixture weight and the number of contributors must be considered to determine whether genotype conditioning can mitigate factors effecting mixture deconvolution and potentially improve resulting match statistics.

PRIOR LITERATURE

Twenty years ago, TrueAllele was developed to address the mixture issue and has been used in forensic casework since 2009. Since its development, numerous peer-reviewed validation studies have been conducted and published. Early studies addressed the comparison of manual DNA profile interpretation to computer-based interpretation, in which, computer-based analysis rapidly outpaced human interpretation analysis (Bauer et al., 2020). Additional validation studies were conducted to assess single-source DNA samples and mixtures containing up to two, three, four, or five contributors. In these studies, typical validation metrics such as sensitivity, specificity, and reproducibility were assessed. Sensitivity relates to how well the genotyping system detects true contributors to a mixture, specificity relates to how well the genotyping system rejects non-contributors, and reproducibility measures the closeness of match values on replicate independent computer runs. Low-template DNA, joint data analysis, and contributor number were listed as other variables and features that were assessed. Both laboratory-generated DNA mixture data and casework DNA mixture data were tested in these studies.

In a recent study, a ten-contributor mixture was analyzed to assess the reliability and validity of TrueAllele computer interpretation on complex mixtures, providing an extension to previous validation work by examining STR data from laboratory-generated

mixtures containing up to ten unknown contributors (Bauer et al., 2020). Validation metrics including sensitivity, specificity, and reproducibility were measured based on log(LR) match information as used in previous validation studies. Additionally, the influence of contributor number and the effect of MCMC statistical sampling duration were examined based on these metrics. Lastly, the impact of genotype conditioning and how the number of input data peaks affects match statistics was examined. In this study, mixture composition was constructed randomly and analyzed by two independent TrueAllele interpretation groups. The mixture was conditioned based on mixture weight and the results demonstrated a general increase in log(LR) values of minor contributors (Bauer et al., 2020). The average specificity of the inferred genotypes significantly improved and yielded more exclusionary log(LR) values regarding the non-contributor distribution, such that, the number of non-contributor positives reduced two-fold. It was found that earlier peeling rounds improved sensitivity and produced refined evidence genotypes for true contributors. However, residual data from minor contributors with mixture proportions below 5% showed less specificity and produced more log(LR) false positives over 3 ban units (Bauer et al., 2020). Once majority of the data was accounted for, additional rounds of peeling were less informative and produced less focused genotypes.

This study focused on a single complex mixture containing ten-contributors. Contributors within the mixture consisted of varying mixture weights ranging from high-template to low-template amounts. Genotype conditioning was conducted based on mixture weight, where a single method of peeling was employed for validation and

reliability purposes. There is still a critical need for a more comprehensive study that evaluates the effect of genotype conditioning on mixture deconvolution and resulting match statistics. To properly evaluate the effects of genotype conditioning, a larger sample size consisting of mixtures with varying compositions and contributor number must be assessed. This assessment advocates for the exploration of different routes of genotype conditioning such as mixture weight, as previously conducted, and other routes based on match statistics.

METHODS AND MATERIALS

STR Data

The STR mixture dataset was obtained from the Forensic DNA Laboratory in the Beaufort County Sheriff's Office and were developed from known reference samples according to the experimental design (Table 1). DNA templates were amplified in 25μL reactions for 28 cycles using an Applied Biosystems® (Foster City, CA) GlobalFiler® PCR amplification kit at a DNA concentration of 1ng. PCR products were size-separated and detected using an Applied Biosystems® 3500xl Genetic Analyzer. The dataset consisted of four mixture groups corresponding to 2, 3, 4, and 5 contributors, and within each group five mixtures were composed randomly out of 26 total known reference samples. The four mixture groups each containing five items, yielded a total of 20 randomized DNA mixture samples. Randomized mixture ratios were used in this study to realistically represent actual casework evidence. The contributors included in each mixture were determined by randomly selecting known references, and mixture weights of the contributors in each mixture were also randomly drawn.

Genotype Conditioning

Electronic data files were processed through TrueAllele® (Pittsburgh, PA), where interpretation requests were formed that assumed 0, 1, 2, 3, 4, or 5 contributors. Requests were processed with a burn-in and read-out time of 25,000 MCMC cycles and were run

Table 1. Experimental Design.

			Expected Weight				
Mixture	***n* con**	**Contributor IDs**	**1**	**2**	**3**	**4**	**5**
A	2	75 25	0.54	0.46			
B	2	29 71	0.73	0.27			
C	2	73 72	0.84	0.16			
D	2	24 27	0.91	0.09			
E	2	26 36	0.98	0.02			
F	3	22 37 25	0.48	0.46	0.06		
G	3	34 23 21	0.51	0.28	0.21		
H	3	40 24 35	0.67	0.22	0.11		
I	3	29 58 74	0.79	0.13	0.08		
J	3	30 31 72	0.91	0.07	0.02		
K	4	38 35 71 37	0.34	0.26	0.21	0.19	
L	4	30 28 38 29	0.46	0.36	0.11	0.07	
M	4	32 58 72 40	0.53	0.25	0.19	0.03	
N	4	26 31 74 25	0.61	0.26	0.09	0.04	
O	4	22 36 34 23	0.79	0.14	0.06	0.01	
P	5	25 26 58 30 34	0.31	0.21	0.20	0.15	0.13
Q	5	31 36 32 28 27	0.40	0.38	0.09	0.07	0.06
R	5	37 24 73 23 75	0.47	0.27	0.13	0.09	0.04
S	5	22 35 72 40 74	0.69	0.17	0.07	0.05	0.02
T	5	71 29 21 39 38	0.83	0.08	0.04	0.03	0.02

in duplicate with additional replicates at higher cycles as needed for reproducibility. For two-contributor mixtures, the initial interpretation request was formed without the assumption of known reference samples and contributor genotypes were inferred solely based on the STR data. This initial request established the major contributor genotype, determined based on the highest mixture weight. In the next request the major contributor was assumed, and the genotype of the minor contributor was inferred by TrueAllele. This request was conducted again with the assumption of the minor contributor, in which, TrueAllele inferred the major contributor genotype. Lastly, an interpretation request was formed with the assumption of both the major and the minor contributor. This process of

genotype conditioning was applied to all two-contributor mixtures, such that all combinations of contributors were peeled. For mixtures containing three or more contributors, two genotype conditioning approaches were employed when appropriate for comparison: one based on mixture weight and the other based on match statistics.

Peeling by Mixture Weight

Each contributor within a mixture had a corresponding mixture weight that was predetermined by the experimental design, as shown in Table 1. Since laboratory generated data is subject to variation in pipetting or during amplification, this may cause the observed mixture proportions to differ from the designed values. Because this study relates mixture weight to other variables, it was important to obtain accurate mixture weight estimates. Therefore, the observed mixture weight values obtained through empirical methods were used instead of the designed values. To accurately estimate mixture proportions of each contributor to the mixture, TrueAllele used the known contributor genotypes as provided input. Therefore, when conditioning based on mixture weight, the initial request was formed with zero unknowns. That is, all contributor genotypes were assumed but variables including mixture weight were estimated based on the STR data and given genotype knowledge. After processing, TrueAllele separated each contributor genotype and estimated component mixture proportions summing to one. Using the data resulting from the initial run, the contributors were ranked from highest to lowest in terms of mixture weight. In the following request, the contributor with the highest mixture weight was assumed by inputting the respective reference profile and TrueAllele then inferred genotypes for the remaining unknown contributors, which

demonstrated the first round of peeling for this path. For the second round of peeling, the contributor with the highest mixture weight used in the previous run, and the next highest mixture weight were both assumed and TrueAllele inferred the remaining unknown genotypes. Genotype conditioning continued until all contributors within the mixture were assumed in rank order based on mixture weight.

Peeling by Match Statistic

When conditioning based on match statistics, the initial interpretation request was formed without the assumption of contributor genotypes. Variables including mixture weight and contributor genotype were inferred solely from quantitative STR peak height data. After processing, TrueAllele separated each contributor genotype and calculated each contributor's match statistic relative to a population. Based on the data resulting from this initial request, the contributor with the highest match statistic displayed as log(LR) values was assumed by inputting the respective reference profile. TrueAllele then inferred the remaining unknown contributor genotypes, which demonstrated the first round of peeling. For the second round of peeling, the contributor with the highest match statistic used in the previous run and the next highest match statistic were both assumed, and TrueAllele then inferred the remaining unknown genotypes. Genotype conditioning continued until all contributors were assumed in rank order based on match statistics.

DATA ANALYSIS AND INTERPRETATION

Data Generation

In TrueAllele, samples are processed on two parallel processing computers, where one server interprets the uploaded experimental data using MCMC statistical sampling to separate contributor genotypes based on probabilities by considering thousands of possible values for each variable. Results from this server are conveyed to a database server and the user can then view these results to work out match statistics through the Visual User Interface (VUIer^{TM} Software). The software version numbers for the genotyping server were 3.25.5840.1 and 3.3.5926.1 for the VUIer client. Data generation included the processing of the prepared STR mixtures by uploading the experimental data to TrueAllele's genotyping server to separate contributor genotypes and calculate respective match statistics, recorded in log(LR) ban units. These results were then conveyed to TrueAllele's database server and accessed using the VUIer client, where interpretation requests were made. To calculate match statistics, inferred genotypes were compared with known reference genotypes relative to the NIST 1036 African American, Caucasian, and Hispanic ethnic populations (Hill et al., 2013). The co-ancestry coefficient was set to 1%. The system has no analytical threshold and signals were used above 10 RFU, a level within baseline noise.

Mixtures were conditioned in gradual iterations using provided known reference samples. To determine the relationship between genotype conditioning and resulting match statistics, the log(LR) values of residual contributors were compared before and after each round of peeling within a single mixture. To assess the impact of mixture weight on genotype conditioning, match statistic comparisons were made across mixtures containing the same number of contributors. Mixtures containing three or more contributors were conditioned both on mixture weight and match statistics when appropriate to determine which method produced more refined log(LR) values. The two approaches could only be compared when the order of contributors based on mixture weight and that of match statistics were different, diverging into two routes of peeling. If the order of contributors regarding these two factors were the same, the routes directly overlap therefore pursuing both routes is not necessary.

In this study, reproducibility was assessed by examining the closeness of match values from replicate computer runs and the genotype concordance. The variation in the amount of information expressed as log(LR) values was documented, where genotypes derived from multiple independent computer runs were compared with the same corresponding reference genotype. Upon variation in the log(LR) values of greater than 3 ban units between replicate computer runs occurred, an additional replicate was run to address such variation. If genotype concordance was not improved by the additional replicate, the sample was run with longer cycle runtimes. To assess sensitivity or how well the genotyping system detects true contributors to a mixture, the log(LR)

distribution of inferred genotype comparisons to true contributors were examined. The match statistic progression with successive peeling rounds were analyzed.

Bayesian Metrics

The Kullback-Leibler (KL) divergence is a mathematical statistic used to measure the similarity or difference between any two probability distributions. In the context of Bayesian inference, the KL divergence is a measure of data information gain from the prior probability distribution to the posterior probability distribution, in which the posterior distribution is updated with priors using Bayes' theorem. The amount of information lost is expressed when the prior probability distribution is used to approximate the posterior. A common goal in the Bayesian experimental design is to maximize the expected KL divergence between the posterior and prior distributions. The statistic was used in this study to provide an estimate of how much more information is gained about the experimental data after computer inference (posterior) than before computer inference (prior). Higher KL values provide insight to how informative a genotype may be as measured by the difference between the posterior and prior (population) distributions, whereas lower KL values indicate less differences and hence a potentially less informative genotype. Low-template DNA profiles tend to have lower KL values due to the high levels of uncertainty present. TrueAllele calculates the KL after performing a Bayesian analysis on the dataset for each inferred contributor in a mixture. The KL then estimates the match information present in a contributor's probabilistic genotype. Thus, KL can be used to predict the match statistic and the potential informativeness of a separated profile in instances where a reference sample is not

available and allows for uploading profiles to a DNA database for comparison (Donahue & Perlin, 2019).

The Gelman-Rubin statistic is used in MCMC algorithms and measures the convergence of Markov chains to a stationary distribution within an acceptable error, relative to the amount of statistical sampling cycles implemented. The statistic is a useful tool in monitoring Markov chains before inducing any specific decisions regarding the kinds of inferences that can be made from the model. Therefore, it can be used to reliably establish genotype concordance, a key factor in demonstrating the reproducibility of the genotype inference. Concordance is observed when multiple, independent computer runs produce similar posterior genotype probabilities for a contributor and allows for the reporting of those results. Since genotype inference is a random process and runs are independent from one another, some variation is expected between replicate computer runs due to Markov chains beginning in different areas of probability space. This statistic becomes important to show that the same result is inferred in a concordant manner and demonstrates the reproducibility of the process. Indications of concordance are observed when the similar probabilities are seen between independent computer runs. Due to the nature of the data, however, the inference process may not produce concordant genotypes.

The need for improvement of concordance is indicated when genotype probabilities do not share similar amounts of probability and have different allele pair possibilities. The lack of concordance can be caused by several factors such as the need for additional sampling, degradation or low-template DNA amounts, or an incorrect

contributor assumption. To potentially improve genotype concordance, there are several options for additional processing that may help to resolve the issue. Additional replicates can be conducted if independent replicates produce different results. The additional replicate can be used to act as a tiebreaker if the new replicate is concordant with one of the initial runs. Complex mixture samples can be challenging to resolve, and dozens of variables must be considered. In such cases, a shorter cycle runtime may not provide enough time for the inference process to have fully explored the solution. Also, cases with very little data information, severe degradation, or rare alleles may require additional cycles to fully explore all possibilities.

RESULTS AND DISCUSSION

Two-Contributor Mixtures

The first mixture group containing five items labeled A through E, were each composed of two contributors and peeling results are displayed in Tables 2 through 5. Each mixture item was conditioned by conducting all combinations of peels, where 0, 1, or 2 contributors were assumed. As genotype assumptions were made throughout peeling rounds, the log(LR) values of assumed contributors were greyed out and match statistics for residual contributors were compared as peeling progressed. All interpretation requests were initially run at a burn-in and read-out time of 25,000 MCMC cycles. Within an interpretation request, TrueAllele separated the mixture into component genotypes, estimated the mixture weight and its standard deviation, measured the divergence as KL, and provided match statistics in log(LR) ban units relative to a population for each contributor in the mixture.

Table 2. Two-Contributor Peeling Results: Mixture A

					Log(LR)	
Evidence	Contributor	Weight	Std Dev	KL	25	75
Mix A	1	0.43	0.019	27.15	12.45	
	2	0.57	0.019	27.71		11.15
Mix A+75	1	0.55	0.022	29.40		27.44
	2	0.45	0.022	30.44	28.66	
Mix A+25	1	0.45	0.023	30.44	28.66	
	2	0.55	0.023	29.40		27.44
Mix A+25+75	1	0.45	0.019	30.44	28.66	
	2	0.55	0.019	29.40		27.44

In Mixture A both contributors exhibited similar mixture weights, a factor that is known to depress match statistics since contributor genotypes are more difficult to separate (Bauer et al., 2020). By assuming a contributor, the log(LR) value for the remaining contributor increased significantly over two-fold as shown in Table 2.

Table 3. Two-Contributor Peeling Results: Mixture B and C

					Log(LR)	
Evidence	Contributor	Weight	Std Dev	KL	29	71
Mix B	1	0.76	0.021	28.54	26.40	
	2	0.24	0.021	27.84		23.13
Mix B+29	1	0.76	0.021	28.54	26.59	
	2	0.24	0.021	28.37		23.21
Mix B+71	1	0.22	0.014	30.86		28.38
	2	0.78	0.014	30.46	26.37	
Mix B+29+71	1	0.76	0.018	28.54	26.59	
	2	0.24	0.018	30.86		28.38
					Log(LR)	
Evidence	Contributor	Weight	Std Dev	KL	72	73
Mix C	1	0.17	0.027	25.24	23.83	
	2	0.83	0.027	29.01		27.52
Mix C+73	1	0.83	0.020	29.02		27.52
	2	0.17	0.020	26.68	23.37	
Mix C+72	1	0.17	0.017	30.29	28.49	
	2	0.83	0.017	29.02		27.52
Mix C+72+73	1	0.17	0.019	30.29	28.49	
	2	0.83	0.019	29.02		27.52

For Mixture B and Mixture C, the genotyping program had no issue resolving these mixtures without integrating known profile assumptions since contributor proportions largely differed at approximately 75/25 and 83/17, respectively. Therefore, the log(LR) values without genotype assumptions for each contributor were highly informative and remained consistent as contributors were peeled, displayed in Table 3.

Table 4. Two-Contributor Peeling Results: Mixture D

					Log(LR)	
Evidence	Contributor	Weight	Std Dev	KL	24	27
Mix D	1	0.03	0.012	11.75		14.52
	2	0.97	0.012	30.29	29.00	
Mix D+24	1	0.92	0.015	30.29	29.01	
	2	0.08	0.015	21.24		19.29
Mix D+27	1	0.07	0.013	31.73		29.02
	2	0.93	0.013	30.29	29.01	
Mix D+24+27	1	0.92	0.012	30.29	29.01	
	2	0.08	0.012	31.73		29.02

In Mixture D, the minor contributor was apportioned an 8% mixture weight, estimated by assuming both contributor genotypes shown in Table 4. By assuming the major contributor, the log(LR) of the minor contributor increased by 5 ban units.

Table 5. Two-Contributor Peeling Results: Mixture E

					Log(LR)	
Evidence	Contributor	Weight	Std Dev	KL	26	36
Mix E	1	0.95	0.031	28.94	27.20	
	2	0.05	0.031	12.89		
Mix E+26	1	0.99	0.007	28.95	27.20	
	2	0.01	0.007	7.41		
Mix E+36	1	0.01	0.006	34.48		32.66
	2	0.99	0.006	28.95	27.20	
Mix E+26+36	1	0.99	0.006	28.95	27.20	
	2	0.01	0.006	34.48		32.66

In Mixture E, the minor contributor was attributed an estimated 1% mixture weight, thus allele detection was limited as seen in Table 5. Consequently, additional sampling time of 50,000 MCMC cycles was explored. At 25,000 cycles, the log(LR) value for the minor contributor was less than 1 ban without genotype assumptions and remained unchanged after the major contributor was peeled. At 50,000 cycles the minor contributor resulted in a log(LR) of 2 ban without genotype assumptions and remained

the same after the next peeling round. Upon observing STR peak height data, the minor contributor exhibited allelic dropout of one allele in the following three locations: D21S11, D2S441, and D3S1358. In low-template contributors, allelic dropout can be expected and is especially apparent in high molecular weight loci, therefore, these results are expected.

Three-Contributor Mixtures

The second mixture group containing five items labeled F through J were each composed of three contributors. Each mixture item was conditioned based on mixture weight where 0, 1, 2, or 3 contributors were assumed as shown in Tables 6 through 10. Mixture F was composed of two contributors with similar mixture weights as well as a 4% contributor, described in Table 6. When the contributor with the highest mixture weight was assumed, the resulting log(LR) for reference 37 increased by 9 ban and 3 ban for reference 25. In the second round of peeling, the log(LR) of the remaining contributor increased by 4 ban in increment.

Table 6. Two-Contributor Peeling Results: Mixture F

					Log(LR)		
Evidence	Contributor	Weight	Std Dev	KL	22	25	37
Mix F	1	0.32	0.009	19.46			16.01
	2	0.67	0.009	23.77	17.68		
	3	0.01	0.008	8.71		2.54	
Mix F+22	1	0.54	0.008	28.33	27.12		
	2	0.01	0.007	4.49		5.10	
	3	0.45	0.008	25.70			25.34
Mix F+22+37	1	0.52	0.052	28.33	27.12		
	2	0.43	0.051	27.61			25.90
	3	0.05	0.044	9.48		9.39	
Mix F+22+37+25	1	0.55	0.036	28.33	27.12		
	2	0.04	0.018	30.45		28.66	
	3	0.42	0.035	27.61			25.90

Mixture G exhibited uncertainty in genotype separation in the first round of peeling where residual contributors matched both inferred genotypes. Match statistics for all three contributors were initially informative before peeling, displaying a minimum of 9 ban units shown in Table 7. In the first peeling round, an increase in match statistics were observed for both remaining contributors, in which, the log(LR) of reference 23 improved by 3 ban and reference 21 improved by 5 ban. In the second round of peeling, the log(LR) of reference 21 incrementally increased by 6 ban.

Table 7. Three-Contributor Peeling Results: Mixture G

					Log(LR)		
Evidence	Contributor	Weight	Std Dev	KL	21	23	34
Mix G	1	0.53	0.040	28.77			26.48
	2	0.18	0.026	16.65	9.68		
	3	0.29	0.035	20.89		13.49	
Mix G+34	1	0.52	0.029	33.92			30.70
	2	0.30	0.040	23.34	1.04	16.83	
	3	0.18	0.030	19.09	14.65	6.22	
Mix G+34+23	1	0.30	0.036	30.91		28.39	
	2	0.52	0.039	33.92			30.70
	3	0.18	0.019	23.69	20.28		
Mix G+34+23+21	1	0.19	0.018	27.28	26.34		
	2	0.29	0.022	30.91		28.39	
	3	0.52	0.023	33.92			30.70

For Mixture H, the genotyping program had no issue resolving the mixture without integrating known profile assumptions, which was expected due to contributor proportions that largely differed, seen in Table 8. Match statistics for all three contributors were highly informative before peeling, displaying a minimum log(LR) value of 20 ban. Match statistics for residual contributors generally remained unchanged throughout successive peeling rounds with slight increases of 1 ban.

Table 8. Three-Contributor Peeling Results: Mixture H

					Log(LR)		
Evidence	Contributor	Weight	Std Dev	KL	24	35	40
Mix H	1	0.29	0.039	24.96	25.19		
	2	0.58	0.041	27.73			26.70
	3	0.12	0.024	22.33		20.11	
Mix H+40	1	0.62	0.040	29.16			27.44
	2	0.27	0.038	26.81	26.14		
	3	0.10	0.021	20.99		20.15	
Mix H+40+24	1	0.27	0.034	30.29	29.01		
	2	0.63	0.033	29.16			27.44
	3	0.11	0.019	23.80		21.20	
Mix H+40+24+35	1	0.29	0.019	30.29	29.01		
	2	0.10	0.014	33.62		31.02	
	3	0.61	0.022	29.16			27.44

In Mixture I, uncertainty was exhibited in the separation of genotypes between the two minor contributors both before genotype conditioning was initiated as well as in the first round of peeling, described in Table 9. Without genotype assumptions, match statistics were informative for all three contributors, displaying a minimum log(LR) value of 12 ban. In the first round of peeling, a slight decrease of 1 ban in the log(LR) value was observed in residual contributors. In the second round of peeling, the log(LR) of the remaining contributor increased by 6 ban.

Table 9. Three-Contributor Peeling Results: Mixture I

					Log(LR)		
Evidence	Contributor	Weight	Std Dev	KL	29	58	74
Mix I	1	0.18	0.056	17.99		18.51	3.92
	2	0.12	0.045	14.54		10.23	12.52
	3	0.70	0.036	28.35	25.57		
Mix I+29	1	0.72	0.038	28.54	26.59		
	2	0.13	0.054	14.14		15.97	11.09
	3	0.15	0.058	15.11		17.91	9.48
Mix I+29+58	1	0.71	0.035	28.54	26.59		
	2	0.19	0.034	28.95		27.42	
	3	0.10	0.023	18.61			17.21
Mix I+29+58+74	1	0.69	0.025	28.54	26.59		
	2	0.20	0.023	28.95		27.42	
	3	0.11	0.016	27.40			26.20

Mixture J consisted of a 7% mixture weight contributor and the contributor apportioned the least amount of DNA showed a mixture proportion of 1%, shown in Table 10. As seen in the table, the KL statistics reflect the inability of the program to effectively separate these trace contributors. Once majority of the data is accounted for by peeling the contributor with the highest mixture weight, a slight increase in log(LR) is observed in reference 31. Match statistics for reference 72 fluctuated throughout peeling rounds, where a 1 ban decrease was observed in the first peeling round and a 1 ban increase in the second peeling round. Thus, genotype assumptions in successive peeling rounds were unhelpful for the 1% contributor.

Table 10. Three-Contributor Peeling Results: Mixture J

					Log(LR)		
Evidence	Contributor	Weight	Std Dev	KL	30	31	72
Mix J	1	0.01	0.007	5.03			2.32
	2	0.88	0.008	32.65	30.27		
	3	0.11	0.008	17.98		8.53	
Mix J+30	1	0.87	0.007	32.65	30.27		
	2	0.01	0.006	5.96			1.00
	3	0.12	0.007	19.08		9.66	
Mix J+30+31	1	0.93	0.014	32.65	30.27		
	2	0.06	0.010	29.22		27.29	
	3	0.01	0.016	4.99			2.08
Mix J+30+31+72	1	0.92	0.018	32.65	30.27		
	2	0.07	0.018	29.22		27.29	
	3	0.01	0.008	30.28			28.49

Four-Contributor Mixtures

The third mixture group containing five items labeled K through O were each composed of four contributors and genotype peeling results are displayed in Tables 11 through 15. All items within the mixture set were conditioned based on mixture weight

and Mixture M was additionally conditioned by match statistics. Interpretation requests were formed in which 0, 1, 2, 3, or 4 contributor genotypes were assumed.

Table 11. Four-Contributor Peeling Results: Mixture K

					Log(LR)			
Evidence	Contributor	Weight	Std Dev	KL	35	37	38	71
Mix K	1	0.31	0.149	9.61	12.10	2.74	13.45	4.83
	2	0.07	0.114	4.39	5.53	3.22	2.73	4.79
	3	0.28	0.101	10.92	12.80	5.04	13.03	5.88
	4	0.34	0.082	12.27	12.23		15.48	4.30
Mix K+38	1	0.40	0.034	33.47			29.67	
	2	0.01	0.015	3.85		1.16		1.39
	3	0.20	0.020	18.47	5.15			2.82
	4	0.40	0.039	19.69	13.03			
Mix K+38+35	1	0.28	0.036	33.62	31.02			
	2	0.37	0.042	33.47			29.67	
	3	0.15	0.043	15.16		10.94		14.96
	4	0.20	0.051	16.57		9.82		16.75
Mix K+38+35+71	1	0.28	0.026	33.62	31.02			
	2	0.37	0.031	33.47			29.67	
	3	0.19	0.031	30.86				28.38
	4	0.16	0.028	22.19		18.26		
Mix K+38+35+71+37	1	0.28	0.023	33.62	31.02			
	2	0.15	0.022	27.61		25.90		
	3	0.38	0.022	33.47			29.67	
	4	0.19	0.024	30.86				28.38

In Mixture K, high levels of genotype separation uncertainty were prevalent before peeling, where all four contributors exhibited positive match statistics for multiple inferred genotypes as seen in Table 11. The mixture proportions of the two major contributors differed by an estimated 10% and the two minor contributors differed by only 4%. Due to this similarity in mixture proportions between contributors, the level of genotype separation uncertainty is expected. In the first round of peeling, the log(LR) for reference 35 increased slightly and a decrease in log(LR) for references 37 and 71 was observed since unhelpful genotype assumptions may either reduce match statistics or

leave them unchanged. As peeling progressed, match statistics of residual contributors improved significantly, and the level of uncertainty in genotype separation reduced but was still apparent. In the second round of peeling, the match statistics for reference 37 improved by 9 ban after the first peeling round and improved by 5 ban compared to before peeling was conducted. For reference 71, match statistics improved by 14 ban from the first peeling round and improved by 11 ban compared to before peeling was applied. In the third round of peeling, the log(LR) of the remaining contributor, reference 37, increased incrementally by 8 ban.

Table 12. Four-Contributor Peeling Results: Mixture L

					Log(LR)			
Evidence	Contributor	Weight	Std Dev	KL	28	29	30	38
Mix L	1	0.08	0.035	11.95		2.48		11.00
	2	0.42	0.045	18.72	12.57		17.68	
	3	0.02	0.028	3.59		3.10		1.07
	4	0.48	0.046	17.61	11.93		17.55	
Mix L+30	1	0.51	0.009	32.65			30.27	
	2	0.13	0.009	15.85				2.37
	3	0.35	0.008	25.48	24.46			
	4	0.01	0.006	5.69		1.76		
Mix L+30+28	1	0.43	0.031	30.03	27.71			
	2	0.42	0.030	32.65			30.27	
	3	0.04	0.051	5.26		3.96		3.96
	4	0.12	0.033	16.25				13.13
Mix L+30+28+38	1	0.41	0.018	30.03	27.71			
	2	0.44	0.018	32.65			30.27	
	3	0.12	0.014	33.46				29.67
	4	0.02	0.034	4.91		5.99		
Mix L+30+28+38+29	1	0.41	0.025	30.03	27.71			
	2	0.06	0.018	28.54		26.59		
	3	0.42	0.030	32.65			30.27	
	4	0.12	0.021	33.46				29.67

In Mixture L, contributors also exhibited similar mixture proportions in which the two major contributors were relatively the same and the two minor contributors differed

by only 6% as shown in Table 12. Therefore, uncertainty in genotype separation was evident however, the log(LR) values were highly informative and above 10 ban units, excluding the contributor with the lowest mixture proportion of 6%. In the first round of peeling, the log(LR) value for reference 28 increased two-fold. As uncertainty in genotype separation reduced in the first peeling round, the log(LR) of reference 38 reduced from 11 ban units to 2 ban units. Match statistics for reference 29 fluctuated as peeling progressed. In the second round of peeling, the two major contributor genotypes were assumed and the log(LR) for reference 38 jumped to an informative 13 ban, although, some genotype separation uncertainty was still observed. Once this genotype was accounted for in the third peeling round, a small increase of 2 ban was observed in reference 29 resulting in a more informative log(LR) value of nearly 6 ban.

Mixture M was conditioned by mixture weight as well as by match statistics, where genotype conditioning routes diverge in the first and second rounds of peeling. Genotype peeling results are displayed in Table 13 and interpretation requests conducted based on match statistics are indicated by an asterisk. The two major contributors consisted of similar mixture proportions differing by an estimated 3%. Uncertainty in genotype separation was evident however, the log(LR) values were highly informative at 10 ban and above excluding the contributor with the lowest mixture proportion. The contributor that apportioned the least amount of DNA showed a mixture proportion of 2% and match statistics generally centered around 2 ban throughout successive peeling rounds, regardless of the conditioning method. In the first round of peeling by mixture weight, reference 32 was assumed and the log(LR) of reference 58 increased two-fold.

The log(LR) value of reference 72 remained the same. When peeling by match statistics, reference 72 was assumed and the log(LR) of reference 32 increased by 3 ban. The log(LR) of reference 58 remained the same. In the second round of peeling based on mixture weight, contributors 32 and 58 were assumed and a small increase of 2 ban was observed in reference 72. When peeling by match statistics, the genotypes of contributors 72 and 58 were assumed and the log(LR) of reference 32 increased two-fold.

Table 13. Four-Contributor Peeling Results: Mixture M

					Log(LR)			
Evidence	Contributor	Weight	Std Dev	KL	32	40	58	72
Mix M	1	0.13	0.041	10.66	1.29		4.70	15.60
	2	0.48	0.066	18.42	10.32		11.63	
	3	0.24	0.084	11.45	10.37		10.56	9.25
	4	0.15	0.200	6.28	7.90	2.09	7.11	2.48
Mix M+32	1	0.39	0.006	27.77	26.18			
	2	0.18	0.006	24.45				15.85
	3	0.42	0.006	28.07			23.53	
	4	0.01	0.005	5.62				
Mix M+72*	1	0.20	0.006	30.28				28.49
	2	0.01	0.006	5.32		1.51		
	3	0.41	0.008	17.51	13.19		11.79	
	4	0.37	0.007	17.68	12.10		12.78	
Mix M+32+58	1	0.40	0.060	27.77	26.18			
	2	0.38	0.041	28.95			27.42	
	3	0.10	0.133	8.61		2.27		16.26
	4	0.11	0.062	11.63				17.49
Mix M+72+58*	1	0.36	0.026	28.95			27.42	
	2	0.16	0.013	30.28				28.49
	3	0.45	0.035	27.19	26.00			
	4	0.02	0.029	6.39		2.08		
Mix M+32+58+72	1	0.44	0.006	27.77	26.18			
	2	0.39	0.006	28.95			27.42	
	3	0.17	0.006	30.28				28.49
	4	0.01	0.005	5.56				
Mix M+32+58+72+40	1	0.42	0.020	27.77	26.18			
	2	0.02	0.011	29.16		27.44		
	3	0.39	0.024	28.95			27.42	
	4	0.17	0.026	30.28				28.49

In Mixture N, the contributor that apportioned the least amount of DNA showed a mixture proportion of an estimated 1%, displayed in Table 14. Match statistics for this contributor were uninformative at 1 ban and generally remained constant before conditioning and throughout successive peeling rounds. Without genotype assumptions, high levels of uncertainty in genotype separation between contributors was present and reduced in successive peeling rounds as more contributor profiles were assumed. In the first round of peeling, the contributor with the highest mixture weight was assumed resulting in an increase of 5 ban units for reference 31 and the match statistics of the other residual contributors relatively remained the same. In the second round of peeling, the log(LR) of reference 74 increased by 3 ban.

Table 14. Four-Contributor Peeling Results: Mixture N

					Log(LR)			
Evidence	Contributor	Weight	Std Dev	KL	25	26	31	74
Mix N	1	0.11	0.147	4.58	1.22	7.24	2.46	4.75
	2	0.33	0.062	16.35		13.06	18.60	
	3	0.45	0.131	17.62		22.31	4.74	
	4	0.11	0.059	8.45		2.42		9.88
Mix N+26	1	0.55	0.065	28.95		27.20		
	2	0.10	0.087	6.97	1.42		2.88	10.85
	3	0.29	0.045	22.02			23.95	
	4	0.06	0.070	5.38	2.54			8.04
Mix N+26+31	1	0.60	0.007	28.95		27.20		
	2	0.30	0.007	29.22			27.29	
	3	0.01	0.006	8.43				
	4	0.09	0.007	16.85				13.21
Mix N+26+31+74	1	0.56	0.006	28.95		27.20		
	2	0.31	0.006	29.22			27.29	
	3	0.12	0.006	27.40				26.20
	4	0.01	0.005	7.16				
Mix N+26+31+74+25	1	0.01	0.008	30.45	28.66			
	2	0.59	0.018	28.95		27.20		
	3	0.27	0.019	29.22			27.29	
	4	0.12	0.019	27.40				26.20

In Mixture O, the mixture proportion of the contributor that was attributed the least amount of DNA was also estimated to be about 1% as seen in Table 15. The match statistics for this contributor remained uninformative with log(LR) values less than one before and after successive peeling rounds. Uncertainty in genotype separation was observed periodically, due to the two minor contributors with similar mixture weights, differing by only an estimated 7%. Excluding the 1% minor contributor, match statistics of the other inferred genotypes were informative before peeling rounds and improved as peeling progressed. In the first round of peeling, the contributor with the highest mixture weight was assumed and the log(LR) values of reference 34 and reference 36 increased by 5 ban and 2 ban, respectively. In the second round of peeling, the log(LR) of reference 34 improved incrementally by 2 ban.

Table 15. Four-Contributor Peeling Results: Mixture O

					Log(LR)			
Evidence	Contributor	Weight	Std Dev	KL	22	23	34	36
Mix O	1	0.11	0.009	17.00			8.95	
	2	0.01	0.016	5.25				
	3	0.22	0.012	25.42				17.44
	4	0.66	0.009	27.87	26.85			
Mix O+22	1	0.67	0.024	28.33	27.12			
	2	0.04	0.048	6.12			9.23	
	3	0.21	0.023	27.65				19.48
	4	0.08	0.021	13.08			13.00	1.67
Mix O+22+36	1	0.54	0.163	28.33	27.12			
	2	0.17	0.015	34.48				32.66
	3	0.06	0.045	10.54			15.23	
	4	0.23	0.151	18.03	21.46		9.04	
Mix O+22+36+34	1	0.24	0.070	28.33	27.12			
	2	0.08	0.020	33.91			30.70	
	3	0.19	0.019	34.48				32.66
	4	0.49	0.070	27.95	27.03			
Mix O+22+36+34+23	1	0.72	0.015	28.33	27.12			
	2	0.01	0.005	30.91		28.39		
	3	0.08	0.010	33.91			30.70	
	4	0.19	0.014	34.48				32.66

Five-Contributor Mixtures

The fourth mixture group containing five items labeled P through T were each composed of five contributors and genotype conditioning results are displayed in Tables 16 through 20. The mixture items were conditioned based on mixture weight and when genotype conditioning routes diverged, mixture items were also conditioned based on match statistics. Interpretation requests were formed that assumed 0, 1, 2, 3, 4, or 5 contributors and requests conducted based on match statistics are labeled with an asterisk.

In Mixture P, match statistics for inferred genotypes before peeling were uninformative and centered around 1 ban, excluding the contributor with the highest mixture weight displaying an informative log(LR) of 11 ban as shown in Table 16. In the first round of peeling, the contributor with the highest mixture weight was assumed and match statistics for residual contributors improved. The log(LR) values for references 25, 26, and 34 each increased by 4 ban and the log(LR) value for reference 30 increased by 3 ban. Uncertainty in genotype separation was observed before peeling was initiated and increased in the first peeling round since contributors exhibited similar mixture proportions. The two contributors that consisted of the least amount of DNA, references 30 and 34, were estimated to have the same mixture weight of 11% and the mixture weights of references 25 and 26 differed by an estimated 4%.

In the second round of peeling, the log(LR) value of reference 26 improved significantly by 9 ban and the log(LR) value of reference 30 increased by 4 ban, in increment. The match statistic for reference 34 slightly decreased due to contributors consisting of similar mixture weights. In the third round of peeling, match statistics for

Table 16. Five-Contributor Peeling Results: Mixture P

					Log(LR)				
Evidence	Contrib.	Weight	Std Dev	KL	25	26	30	34	58
Mix P	1	0.39	0.007	15.65					11.71
	2	0.18	0.007	16.87	1.45				
	3	0.01	0.005	3.92					
	4	0.10	0.006	13.61					
	5	0.33	0.007	15.98		1.33			8.44
Mix P+58	1	0.16	0.077	28.95					27.42
	2	0.16	0.037	10.97	1.07	1.49	3.43	4.56	
	3	0.40	0.073	16.81	2.91				
	4	0.05	0.082	3.09	1.86		2.14	3.31	
	5	0.23	0.053	11.17	5.36	5.41			
Mix P+58+25	1	0.29	0.008	30.45	28.66				
	2	0.30	0.008	28.95					27.42
	3	0.15	0.008	17.80		1.40	7.55	1.80	
	4	0.01	0.006	3.43					
	5	0.25	0.008	21.33		14.34			
Mix P+58+25+26	1	0.24	0.048	30.45	28.66				
	2	0.23	0.031	28.95		27.20			
	3	0.31	0.053	28.95					27.42
	4	0.12	0.045	14.18			12.97	10.24	
	5	0.09	0.064	9.42			13.06	10.63	
Mix P+58+25+26+30	1	0.24	0.037	30.45	28.66				
	2	0.20	0.038	28.95		27.20			
	3	0.13	0.021	32.65			30.27		
	4	0.30	0.038	28.95					27.42
	5	0.14	0.044	19.05				17.79	
Mix P+25+26+58+30+34	1	0.24	0.031	30.45	28.66				
	2	0.20	0.021	28.95		27.20			
	3	0.11	0.018	32.65			30.27		
	4	0.11	0.019	33.91				30.70	
	5	0.33	0.035	28.95					27.42

residual contributors improved significantly. The log(LR) values for reference 30 increased by a 6-ban increment and a 9-ban increment for reference 34. Genotype separation uncertainty was still observed; however, it was eliminated in the following round of peeling upon the genotype assumption of the contributor with a similar mixture weight. As a result, the match statistics for the remaining contributor became more refined and improved by an additional 7 ban units in the fourth peeling round. Figure 6

provides a visual representation of the observed increases in the log(LR) values for reference 34 in Mixture P as genotype conditioning progressed.

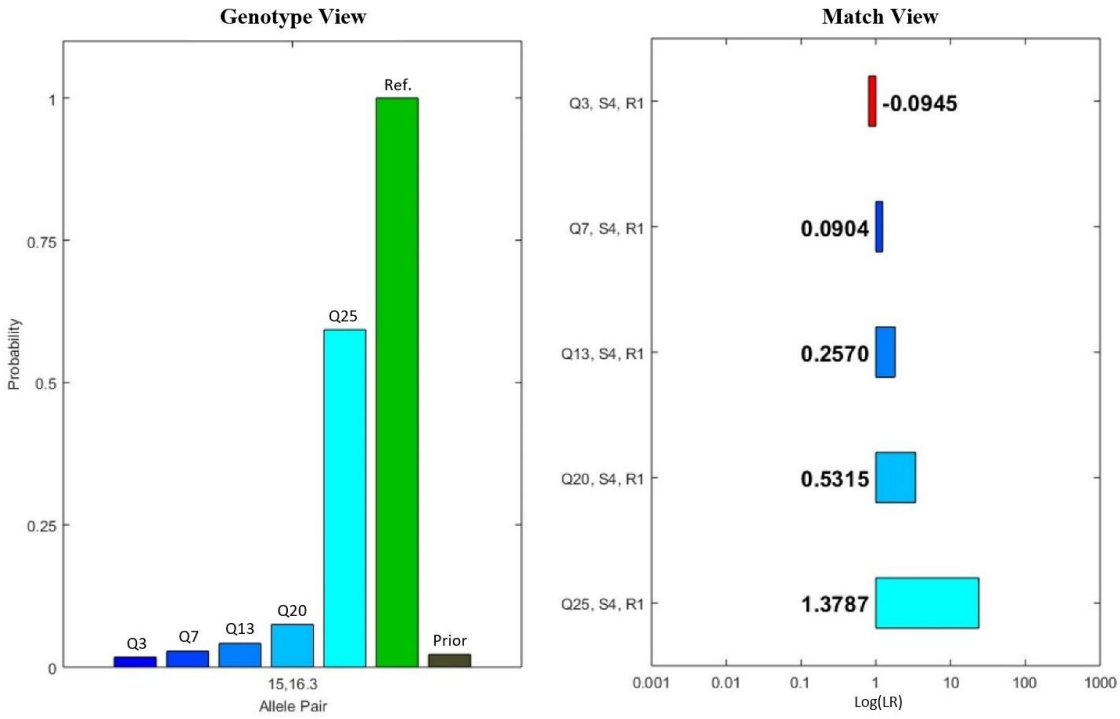

Figure 6. Log(LR) Increases for Reference 34 in Mixture P at D1S1656. The questioned evidence corresponds to the inferred genotype probabilities before conditioning (Q3) followed by a total of four successive peeling rounds (Q7, Q13, Q20, Q25) for subject reference 34 (S4). Genotype View (left): The probabilities of the questioned evidence (shades of blue), reference (green), and population (brown) for genotype (15, 16.3) are shown. The blue bars represent the posterior probabilities, and the brown bar represents the prior probability. Each posterior probability value is divided by the prior probability value to produce the genotype log(LR) for this locus, shown in Match View (right). Log(LR) values show incremental improvements as peeling rounds progressed at this genetic marker.

In Mixture Q, the mixture weights of the two high-template contributors were within close range of each other and differed by an estimated 5%, shown in Table 17. The three remaining contributors consisted of low-template DNA amounts where one contributor consisted of a 6% mixture weight and the other two contributors were each

apportioned 1%. Genotype separation uncertainty is expected and prevalent due to the similarity of mixture proportions between multiple contributors. Therefore, all interpretation requests within Mixture Q were processed with a higher burn-in and read-out time of 50,000 MCMC cycles. Mixture Q was conditioned by mixture weight as well as by match statistics, where genotype conditioning routes diverge in the first and third rounds of peeling. Interpretation requests conducted based on match statistics are indicated by an asterisk.

Before peeling, the match statistics of the high-template contributors were informative, displaying log(LR) values of 9 ban for reference 31 and 16 ban for reference 36. References 28 and 32 were each apportioned a 1% mixture weight and match statistics were uninformative for these contributors centering around 1 ban, which remained consistent throughout all rounds of peeling regardless of the peeling approach used. The log(LR) of reference 27 was about 4 ban units before conditioning, which fluctuated as different contributors were peeled in both conditioning routes. In the first round of peeling by mixture weight, reference 31 was assumed and the log(LR) value of reference 36 increased nearly two-fold. Although there was a slight increase in the log(LR) for reference 32, the match statistic was still uninformative. Upon conditioning by match statistics, reference 36 was assumed and the log(LR) for reference 31 improved over two-fold by 15 ban units. The log(LR) for reference 27 decreased slightly in both conditioning routes and genotype separation uncertainty was still apparent. In the third round of peeling by match statistics, the log(LR) for reference 27 was 5 ban, 1 ban greater than before peeling was conducted.

Table 17. Five-Contributor Peeling Results: Mixture Q

					Log(LR)				
Evidence	Contrib.	Weight	Std Dev	KL	27	28	31	32	36
Mix Q	1	0.35	0.021	18.50			9.37		16.65
	2	0.15	0.030	14.24				1.71	
	3	0.39	0.031	17.87			9.30		16.22
	4	0.03	0.018	7.61					
	5	0.08	0.039	9.57	4.72				
Mix Q+31	1	0.37	0.009	29.23			27.29		
	2	0.05	0.009	8.42	3.66				
	3	0.44	0.009	30.96					31.20
	4	0.12	0.008	14.78	1.73			2.18	
	5	0.01	0.007	3.14					
Mix Q+36*	1	0.45	0.009	34.48					32.66
	2	0.18	0.009	15.08					
	3	0.01	0.007	5.27	1.36				
	4	0.34	0.009	23.10			24.48		
	5	0.02	0.009	4.95					
Mix Q+36+31	1	0.44	0.007	29.23			27.29		
	2	0.39	0.007	34.48					32.66
	3	0.01	0.005	2.42				1.14	
	4	0.15	0.007	16.84					
	5	0.01	0.006	3.70					
Mix Q+36+31+27	1	0.01	0.007	31.73	29.02				
	2	0.36	0.008	29.23			27.29		
	3	0.45	0.008	34.48					32.66
	4	0.13	0.008	17.10					
	5	0.06	0.008	9.73					
Mix Q+36+31+32*	1	0.42	0.020	29.23			27.29		
	2	0.01	0.007	27.77				26.18	
	3	0.42	0.017	34.48					32.66
	4	0.03	0.053	6.31	4.16	1.50			
	5	0.12	0.033	13.23	5.32				
Mix Q+36+31+32+27	1	0.06	0.019	31.73	29.02				
	2	0.42	0.038	29.23			27.29		
	3	0.01	0.010	27.77				26.18	
	4	0.39	0.032	34.48					32.66
	5	0.12	0.027	16.49		1.07			
Mix Q+31+36+32+28+27	1	0.06	0.007	31.73	29.02				
	2	0.01	0.006	30.04		27.71			
	3	0.43	0.007	29.23			27.29		
	4	0.01	0.007	27.77				26.18	
	5	0.48	0.007	34.48					32.66

In Mixture R, the mixture weights of the two contributors that make up the larger portion of the mixture were within close range of each other and differed by an estimated

5% as described in Table 18. Of the three remaining contributors, two contributors were each apportioned a 13% mixture weight and one contributor consisted of a 1% mixture weight. Genotype separation uncertainty is expected and prevalent because of this similarity of mixture proportions between two or more contributors. Mixture R was conditioned by mixture weight as well as by match statistics, where genotype conditioning routes diverged in the first and third rounds of peeling. Interpretation requests conducted based on match statistics are indicated by an asterisk. Before genotype conditioning, match statistics for the two contributors that make up the larger portion of the mixture (references 24 and 37) were informative and each displayed log(LR) values of 12 ban. The match statistics for the two 13% contributors (references 23 and 73) each displayed log(LR) values of 3 ban. The log(LR) value for the 1% contributor rested at values less than 1 ban and did not show improvement in subsequent peeling rounds.

In the first round of peeling by mixture weight, reference 37 was assumed and the match statistic for reference 24 improved significantly by 11 ban. Match statistics for reference 23 slightly improved by 2 ban, resulting in a log(LR) of 5 ban. The log(LR) value for reference 73 remained steady at 3 ban in both conditioning routes. Upon peeling by match statistics, reference 24 was assumed and the log(LR) values for reference 37 improved by 10 ban. The match statistics for the three remaining contributors were the same in ban units as compared to before peeling was employed. As majority of the genotyping data was accounted for in the second round of peeling, an improvement in match statistics for the residual 13% contributors was observed. The log(LR) for

reference 23 increased by 5 ban and 7 ban compared to the previous peeling round by mixture weight and match statistics, respectively. The log(LR) for reference 73 increased

Table 18. Five-Contributor Peeling Results: Mixture R

					Log(LR)				
Evidence	Contrib.	Weight	Std Dev	KL	23	24	37	73	75
Mix R	1	0.19	0.007	11.66		3.53	6.15	3.30	
	2	0.33	0.007	14.79		11.66	12.29		
	3	0.14	0.007	12.21	3.29			3.22	
	4	0.01	0.005	3.83					
	5	0.34	0.007	15.18		12.90	10.92		
Mix R+37	1	0.29	0.028	27.61			25.90		
	2	0.01	0.023	5.38	2.25				
	3	0.22	0.013	14.32		3.92		3.57	
	4	0.12	0.032	12.78	5.95				
	5	0.35	0.013	22.50		23.81			
Mix R+24*	1	0.32	0.008	30.29		29.01			
	2	0.09	0.007	13.14	3.74				
	3	0.16	0.007	16.45				3.30	
	4	0.43	0.007	24.44			22.15		
	5	0.01	0.005	3.43					
Mix R+37+24	1	0.33	0.007	30.29		29.01			
	2	0.41	0.008	27.61			25.90		
	3	0.01	0.006	3.95					
	4	0.15	0.007	15.85	4.46			9.93	
	5	0.11	0.007	13.32	10.41			8.20	
Mix R+37+24+73	1	0.38	0.005	30.29		29.01			
	2	0.39	0.005	27.61			25.90		
	3	0.13	0.005	29.02				27.52	
	4	0.01	0.004	8.47					
	5	0.09	0.005	20.51	10.41				
Mix R+24+37+23*	1	0.09	0.021	30.91	28.39				
	2	0.38	0.027	30.29		29.01			
	3	0.33	0.093	27.61			25.90		
	4	0.03	0.049	6.86				4.89	
	5	0.17	0.069	19.04				15.11	
Mix R+37+24+73+23	1	0.09	0.005	30.91	28.39				
	2	0.35	0.005	30.29		29.01			
	3	0.44	0.005	27.61			25.90		
	4	0.12	0.005	29.02				27.52	
	5	0.01	0.004	4.87					
Mix R+37+24+73+23+75	1	0.13	0.011	30.91	28.39				
	2	0.34	0.008	30.29		29.01			
	3	0.39	0.009	27.61			25.90		
	4	0.13	0.013	29.02				27.52	
	5	0.01	0.005	29.40					27.44

by about 6 ban units from both conditioning approaches conducted in the previous round. In the third round of peeling by mixture weight, match statistics for residual contributors remained unchanged. Upon peeling by match statistics, the log(LR) value of reference 73 increased by 6 ban, incrementally.

Mixture S consisted of three low-template contributors with mixture weights less than 10%, of which, reference 72 attributed 8%, reference 40 attributed 6%, and reference 74 attributed 1% as described in Table 19. With low-template contributors consisting of relatively similar mixture proportions, genotype separation uncertainty is expected and prevalent before conditioning and in the first peeling round. Before genotype conditioning, match statistics for the two contributors that composed the larger portion of the mixture were more informative. The match statistics for reference 22 showed a log(LR) value of 18 ban and 11 ban for reference 35. The 8% contributor displayed a log(LR) value of 5 ban and was more informative than the 6% and 1% contributors, whose log(LR) values were both less than 1 ban.

In the first round of peeling, the log(LR) value for reference 35 improved by 9 ban. Match statistics for reference 72 (8% mixture weight) and reference 40 (6% mixture weight) slightly improved to 4 ban and 6 ban respectively, however additional rounds of peeling were not helpful. This was evident in the second peeling round, where a decrease in log(LR) values for references 72 and 40 were observed. The match statistics for the 1% contributor, reference 74, displayed log(LR) values less than 1 ban throughout majority of peeling rounds and showed a log(LR) value of 1 ban in the fourth round of peeling, remaining uninformative.

Table 19. Five-Contributor Peeling Results: Mixture S

					Log(LR)				
Evidence	Contrib.	Weight	Std Dev	KL	22	35	40	72	74
Mix S	1	0.29	0.007	16.51	13.34				
	2	0.01	0.005	3.74					
	3	0.12	0.007	15.81				5.24	
	4	0.21	0.007	15.90	2.98	11.81			
	5	0.38	0.008	20.33	18.81				
Mix S+22	1	0.62	0.029	28.33	27.12				
	2	0.02	0.029	3.93			1.78	1.71	
	3	0.21	0.038	20.21		20.18			
	4	0.10	0.053	9.60		11.00	3.59	6.18	
	5	0.05	0.057	5.35		3.07	4.20	6.13	
Mix S+22+35	1	0.18	0.018	28.33	27.12				
	2	0.20	0.008	33.62		31.02			
	3	0.01	0.005	3.72					
	4	0.12	0.009	20.53				4.90	
	5	0.50	0.020	27.27	26.77				
Mix S+22+35+72	1	0.12	0.008	28.33	27.12				
	2	0.21	0.008	33.62		31.02			
	3	0.08	0.008	30.29				28.49	
	4	0.58	0.013	27.82	26.65				
	5	0.01	0.005	4.29					
Mix S+22+35+72+40	1	0.69	0.010	28.33	27.12				
	2	0.17	0.011	33.62		31.02			
	3	0.05	0.008	29.16			27.44		
	4	0.08	0.007	30.29				28.49	
	5	0.01	0.004	4.56					1.17
Mix S+22+35+72+40+74	1	0.66	0.028	28.33	27.12				
	2	0.19	0.022	33.62		31.02			
	3	0.06	0.021	29.16			27.44		
	4	0.08	0.019	30.29				28.49	
	5	0.01	0.009	27.40					26.20

Mixture T was composed of three low-template contributors with corresponding mixture weights of less than 10% and produced uninformative match statistics. With low-template contributors consisting of relatively similar mixture proportions, genotype separation uncertainty is expected and prevalent throughout conditioning. Therefore, all interpretation requests within Mixture T were processed with a higher burn-in and read-out time of 50,000 MCMC cycles. Mixture T was conditioned by mixture weight as well

Table 20. Five-Contributor Peeling Results: Mixture T

					Log(LR)				
Evidence	Contrib.	Weight	Std Dev	KL	21	29	38	39	71
Mix T	1	0.76	0.052	27.65					23.37
	2	0.08	0.025	8.15		7.09			
	3	0.05	0.046	5.48		5.56			
	4	0.04	0.061	5.06		1.85			
	5	0.07	0.032	7.84		6.98			
Mix T+71*	1	0.27	0.009	30.54					28.22
	2	0.01	0.007	7.01					
	3	0.52	0.009	29.80					27.83
	4	0.13	0.008	14.85		11.96			
	5	0.07	0.008	11.35		4.39			
Mix T+29	1	0.07	0.008	28.54		26.59			
	2	0.01	0.007	6.46					
	3	0.85	0.008	30.33					28.22
	4	0.06	0.008	13.49					
	5	0.01	0.008	4.79					
Mix T+71+29*	1	0.08	0.035	28.54		26.59			
	2	0.79	0.036	30.54					28.22
	3	0.05	0.059	5.45	1.25			1.98	
	4	0.04	0.050	4.42	1.21			1.66	
	5	0.05	0.059	5.54	1.04			1.40	
Mix T+29+39	1	0.07	0.032	28.54		26.59			
	2	0.01	0.011	31.12				28.32	
	3	0.03	0.030	5.70					
	4	0.01	0.007	5.56					
	5	0.88	0.025	28.93					28.17
Mix T+71+29+39	1	0.09	0.032	28.54		26.59			
	2	0.01	0.009	31.12				28.32	
	3	0.73	0.207	30.54					28.22
	4	0.05	0.059	5.35	1.85				
	5	0.12	0.195	7.95	2.24				14.66
Mix T+71+29+39+21	1	0.01	0.005	27.27	26.34				
	2	0.11	0.014	28.54		26.59			
	3	0.01	0.005	31.12				28.32	
	4	0.09	0.012	30.54					28.22
	5	0.77	0.022	30.53					28.22
Mix T+71+29+21+39+38	1	0.15	0.004	27.27	26.34				
	2	0.32	0.003	28.54		26.59			
	3	0.03	0.004	33.46			29.67		
	4	0.27	0.004	31.12				28.32	
	5	0.24	0.004	30.54					28.22

as by match statistics, where genotype conditioning routes diverge in the first and second rounds of peeling. Interpretation requests conducted based on match statistics are indicated by an asterisk in Table 20. Match statistics for the two contributors that were apportioned the highest mixture weight were informative, displaying log(LR) values of 23 ban for reference 71 and 7 ban for reference 29 before genotype conditioning was initiated.

In the first round of peeling by match statistics, reference 71 was assumed and the log(LR) of reference 29 improved by 4 ban. When peeling by mixture weight, reference 29 was assumed and the log(LR) of reference 71 improved by 5 ban. Match statistics for references 21, 38, and 39 exhibited log(LR) values of less than one before conditioning as well as after the first peeling round, regardless of the conditioning approach performed. In the second round of peeling by match statistics, references 71 and 29 were assumed. The log(LR) values for reference 39 remained less than 1 ban throughout all runs and references 21 and 39 displayed log(LR) values of 1 ban. An improvement in match statistics was not observed for any contributor when peeling by mixture weight. In the third peeling round, the match statistic for reference 21 slightly improved, exhibiting a log(LR) of 2 ban.

When observing STR peak height data, it was determined that Mixture T was subjected to human error. The genotype for reference 71 showed an allele pair of (23.2, 33) at locus SE33, however, allele 23.2 was absent from the electropherogram data, resulting in a large negative match statistic at this genetic location. This instance was rather odd, since reference 71 is the major contributor and having a major allele dropout

is unusual. On the other hand, the data itself displayed a major contributor with allele pair (26.2, 33) that should have been designated as reference 71. It was presumed that a mistype had occurred due to human error when entering the genotype for reference 71. This assumption was confirmed with the Forensic DNA Lab in the Beaufort County Sheriff's Office that had provided the original mixture data. In TrueAllele, the genotype for reference 71 was then updated and new interpretation requests were formed with this updated genotype for analysis. Match statistics were then recalculated, and these results were used for comparisons. Typically, human error causes adverse effects that can be reflected in the experimental data and indicated when measured results deviate from the true or desired value. In instances where true values are unknown, human error becomes easily overlooked and ultimately effects identification, which emphasizes the importance of accuracy and reliability in complex mixture interpretation to minimize potential sources of human error.

CONCLUSION

The purpose of any forensic analysis and interpretation is to derive informative conclusions that generate investigative leads, which remains the focus for genotype conditioning applications. Due to the high amounts of uncertainty associated with mixture samples, many laboratories have halted the interpretation of complex mixtures involving three or more contributors. The research results reported herein demonstrate additional capabilities in the area of mixtures, tests the waters for what probabilistic genotyping can accomplish, and shows the importance and value of using known samples when appropriate to decrease uncertainty in mixture deconvolution. Empirical testing serves as the basis for scientific and legal reliability (Bauer et al., 2020). Thus, validation studies testing laboratory-generated data, as described in this study, and casework field data are both necessary. Developing novel methods to analyze complex mixtures efficiently would continue to benefit the forensic science community in understanding and applying these methods to feasibly generate higher match statistics than traditional mixture deconvolution procedures used in forensic DNA laboratories. These kinds of validation studies can lead to casework practices that are expected to result in higher identification rates.

In this study, DNA mixtures consisting of 2, 3, 4, and 5 contributors were interpreted probabilistically by incorporating genotype conditioning methods based on

mixture weight and match statistics. Overall, adding information to mixture samples is beneficial in mixture deconvolution and increases in resulting LRs was repeatedly observed upon genotype conditioning. The results of this study conclude that genotype conditioning was helpful in mixture deconvolution, especially in contributors that were apportioned similar mixture weights. Without peeling, match statistics of contributors within such mixtures were initially depressed and subsequently improved when a contributor with a similar mixture proportion was assumed by genotype conditioning.

Contributors that were attributed less than a 3% mixture weight were unaffected by genotype conditioning and relative match statistics generally remained constant and uninformative throughout all rounds of peeling. However, improvements in match statistics of contributors with mixture weights between 3% and 10% were observed and were more pronounced in two and three-contributor mixtures. These improvements in log(LR) values became less prominent as the number of contributors increased across mixture sets, yet they were still informative. As the number of contributors increased, the levels of uncertainty in genotype separation were observed, in which, contributors exhibited positive match scores for multiple inferred genotypes. This level of uncertainty is expected and prevalent due to the similarity of mixture proportions between two or more contributors. Genotype separation uncertainty can be caused by several factors such as the presence of low-template contributors and contributors exhibiting similar mixture weights. It is possible that in such instances the computer program may require additional MCMC cycles to explore all variables due to the nature of the sample.

Since only four mixture items diverged into two routes of peeling, it was difficult to determine whether conditioning by mixture weight or by match statistic was more beneficial in producing consistently higher log(LR) values. This factor should be addressed in further studies to determine which approach may be more favorable and under what circumstances. Future directions of this project may also include the assessment of mixtures greater than five contributors, as well as mixtures containing limited quantities of DNA from touch or degraded samples. Additionally, the impact of genotype conditioning on contributors consisting of less than 10% mixture weight should be assessed through the exploration of higher MCMC cycles to ensure that the program is provided enough time to determine the solution as well as to further ascertain the circumstances under which genotype peeling becomes more or less appropriate.

As technology evolves the range of its techniques and discoveries widen, therefore it is incumbent upon the forensic community to adopt new methodologies that corroborate justice by their ability to interpret challenging samples efficiently in an effort to continue to exonerate the innocent and confidently identify criminals. The growing demands for evidence processing in recent years have clearly shown that labs must move towards more efficient and sophisticated methods and procedures for mixture deconvolution. Scientific and technological advances made possible through strong support for research will be essential in ensuring that forensic DNA laboratories keep pace with increasing demand. When implemented after careful consideration and evaluation, genotype peeling can be used to maximize the information obtained from complex mixtures while maintaining a high standard of reliability and reproducibility. As

more forensic laboratories are incorporating probabilistic methods into their workflow, there has been an increase in the ability to successfully separate complex mixtures where valuable conclusions could be drawn and reported. Improving the accuracy and scope of forensic DNA interpretation methods minimizes human error and ultimately strengthens the administration of justice.

REFERENCES

Bauer, D. W., Butt, N., Hornyak, J. M., & Perlin, M. W. (2020). Validating TrueAllele® Interpretation of DNA Mixtures Containing up to Ten Unknown Contributors. *Journal of Forensic Sciences*, 65(2), 380–398. https://doi.org/10.1111/1556-4029.14204.

Butler, J. M. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*. Boston, MA: Elsevier.

Butler, J. M., Kline, M. C., & Coble, M. D. (2018). NIST Interlaboratory Studies Involving DNA Mixtures (MIX05 and MIX13): Variation Observed and Lessons Learned. *Forensic Science International: Genetics*, 37, 81–94. https://doi.org/10.1016/j.fsigen.2018.07.024

Donahue, J. & Perlin, M. "Genotype Information Criteria for Forensic DNA Databases", *American Academy of Forensic Sciences 71st Annual Meeting*, Baltimore, MD, 23-Feb-2019.

Perlin, M.W. “Failing to Interpret DNA Mixture Evidence.” Cybergenetics White Paper, July 2016.

Perlin, M. W. (2011). Combining DNA Evidence for Greater Match Information. *Forensic Science International: Genetics Supplement Series*, 3(1), E510-E511. https://doi.org/10.1016/j.fsigss.2011.09.112.

Perlin, M. W., Hornyak, J. M., Sugimoto, G., & Miller, K. W. (2015). TrueAllele® Genotype Identification on DNA Mixtures Containing up to Five Unknown Contributors. *Journal of Forensic Sciences*, 60(4), 857–868. https://doi.org/10.1111/1556-4029.12788.

Perlin, M. W., Legler, M. M., Spencer, C. E., Smith, J. L., Allan, W. P., Belrose, J. L., & Duceman, B. W. (2011). Validating TrueAllele® DNA Mixture Interpretation. *Journal of Forensic Sciences*, 56(6), 1430–1447. https://doi.org/10.1111/j.1556-4029.2011.01859.x

Wilson, M. R. Touch DNA, Direct Amplification, Continuous Modeling, and Joint Likelihood Approaches to Forensic DNA Mixtures. (2018). Retrieved From: https://outlook.office.com/mail/sentitems/id/AAQkAGZiNDU4ZTQxLTQyODctNDdhMC05ZWVkLTk5OGY4MWVlYzMxZQAQAMNuOAF97IdMt%2Bq%2BBmZYHCc%3D/sxs/AAMkAGZiNDU4ZTQxLTQyODctNDdhMC05ZWVkLTk5OGY4MWVlYzMxZQBGAAAAAABKYQWaz4dCRYLv6N3TekCGBwCvRgxhYA2%2FSrLjCbeelyOdAAAAAAEMAACvRgxhYA2%2FSrLjCbeelyOdAAELZ7pUAAABEgAQAM01GVLOyqlIuqbis5Jsneg%3D.

BIOGRAPHY

Hajara S. Chaudhry graduated from Lake Braddock Secondary School located in Burke, Virginia, in 2013. She received a Bachelor of Science in Biology and a Bachelor of Science in Chemistry from Shenandoah University located in Winchester, Virginia, in 2017. She received her Master of Science in Forensic Science, with a concentration in Forensic Biology Analysis from George Mason University in Fairfax, Virginia, in 2020.