2023

Comparing likelihood ratios of degraded DNA mixture profiles using DNA-view mixture solution

https://hdl.handle.net/2144/48190

BOSTON UNIVERSITY

ARAM V. CHOBANIAN & EDWARD AVEDISIAN SCHOOL OF MEDICINE

Thesis

COMPARING LIKELIHOOD RATIOS OF DEGRADED DNA MIXTURE PROFILES USING DNA-VIEW MIXTURE SOLUTION

by

CAMERON ALEXANDRE FILIPE

B.S., University of New Haven, 2021

Submitted in partial fulfillment of the
requirements for the degree of
Master of Science
2023

Approved by

First Reader

Robin W. Cotton, Ph.D.
Associate Professor of Anatomy & Neurobiology, Program Director
Biomedical Forensic Sciences

Second Reader

Charles H. Brenner, Ph.D.
Senior Research Fellow
University of California, Berkeley

Third Reader

Caitlyn Taveira, M.S.
Forensic Scientist II
Nassau County Office of the Medical Examiner

ACKNOWLEDGMENTS

First and foremost, I would like to thank Dr. Robin Cotton, my thesis advisor, for her invaluable time, assistance, knowledge, and support throughout the research process. I will never be able to convey in words how much I have appreciated your guidance and encouragement throughout my research, as well as my comprehensive time in the program.

Additionally, I would like to thank Dr. Charles Brenner for his valuable input and insight into this topic, but most importantly, for graciously allowing me to use his software, DNA-View® Mixture Solution™, for this research. I appreciate how your viewpoints have helped me tailor my research into a comprehensive assessment of probabilistic genotyping and its power in the sub-discipline of DNA mixture interpretation.

I would also like to thank Caitlyn Taveira for her guidance and assistance in the development of this research. Her input as an active practitioner in the field of forensic DNA analysis was critical to the formulation of my thesis, and I appreciate the valuable time she gave to assist me in this process.

I would like to express my heartfelt gratitude to the faculty, staff, and fellow students of the Biomedical Forensic Sciences program. I would especially like to thank Joseph Juodvalkis, whose own research with Mixture Solution served as a foundation for my own, and especially for his help in preliminarily acclimating me to the software.

Finally, and most importantly, I would like to thank my parents Marco and Rachael Filipe, my sisters Sydney and Lexie Filipe, my girlfriend Mackenzie Mitchell, and all of my friends and family who have provided unconditional support during my entire academic career, especially during my graduate study and the writing of my thesis.

COMPARING LIKELIHOOD RATIOS OF DEGRADED DNA MIXTURE PROFILES USING DNA-VIEW MIXTURE SOLUTION

CAMERON ALEXANDRE FILIPE

ABSTRACT

Interpreting DNA profiles manually can potentially call into question subjectivity between analysts who may interpret specific results differently. There are multiple features of a DNA profile that can complicate interpretation, which include allelic dropout and drop-in, allele sharing, and polymerase chain reaction (PCR) artifacts, which can all confound manual interpretation of DNA profiles. Difficulties in interpretation of DNA profile evidence can also be caused by degradation of the DNA itself, which can be caused by various environmental factors. Over the last 15 years, developments in DNA profile interpretation using probabilistic genotyping software (PGS) have been made in order to assist in the complicated task of interpreting and deconvoluting a challenging mixture. Among these PGSs is DNA-View® Mixture Solution™, a continuous-model program that is based on stochastic modeling.

In this research, Mixture Solution was used to provide statistical analyses on DNA mixtures that were subject to various levels of degradation, through the assignment of a likelihood ratio (LR) to the mixture profile. The LR would either support the hypothesis that the person of interest (POI) contributed to the mixture, or support the contrary hypothesis, that the POI was not one of the contributors. Mixtures were prepared at four different contributor ratios with varying combinations of three levels of degradation: no degradation, partial degradation, and full degradation, using controlled heating to systematically degrade the DNA template prior to amplification. Using two hypothesis tests, Mixture Solution was used to compute LRs for each of the mixtures with a variety of defined POIs.

Results showed that Mixture Solution successfully generated appropriate LRs for all 20 mixtures in this research, with no Type I errors that falsely excluded a known contributor from a mixture via an LR less than one. Even with the greatest level of degradation and at the most disproportionate ratio of contributors, Mixture Solution was able to assign an LR to each contributor that confirmed their presence in the mixture. When the DNA of a POI was subjected to degradation, decreases in the LR values were observed when compared to the values computed for undegraded DNA from the same POI. However, in all mixtures, Mixture Solution was able to assign an LR with "moderate support" or higher to each of the POIs in both tests, regardless of the level of degradation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AT	Analytical threshold
CPI	Combined probability of inclusion
DI	Degradation index
DNA	Deoxyribonucleic acid
FBI	Federal Bureau of Investigation
H_d	Defense hypothesis
H_p	Prosecution hypothesis
LR	Likelihood ratio
MAC	Maximum allele count
MCMC	Markov-chain Monte Carlo
NoC	Number of contributors
PHR	Peak height ratio
PCR	Polymerase chain reaction
PGS	Probabilistic genotyping software
POI	Person of interest
RFU	Relative fluorescent unit
STR	Short tandem repeat
SWGDAM	Scientific Working Group on DNA Analysis Methods
U.S.	United States

1. INTRODUCTION

1.1 Manual DNA Mixture Interpretation

The use of deoxyribonucleic acid (DNA) evidence in legal proceedings is a strong tool of forensic science when the evidence has high probative value and strong statistical significance. However, the analysis of DNA that has been isolated from biological samples obtained from evidence may not always provide results that are easy to interpret, which includes samples with DNA mixtures. The Scientific Working Group on DNA Analysis Methods (SWGDAM) has defined a mixture as a DNA sample that originates from more than one individual, characterized by more than two autosomal alleles present at a particular genetic locus on an electropherogram (1). When this occurs, it is pertinent for the analyst to determine how many individuals may have contributed DNA to the overall mixture sample. Among forensic laboratories, most methods of manual DNA mixture interpretation began with how many individuals may be present in a sample by utilizing the maximum allele count (MAC) method (2). MAC is performed by identifying the locus that contains the most alleles across the entire electropherogram, and then using that number of alleles to determine the minimum number of contributors (NoC) to the mixture (1,2). For example, if the maximum count of alleles at a particular locus on an electropherogram is six, then the minimum NoC is three, since a single individual cannot exhibit more than two alleles at a particular locus.

In the case of two-person mixtures, once the MAC is determined, peak height values in relative fluorescent units (RFU) are then used to determine whether the mixture is distinguishable. Distinguishable mixture samples exhibit a major and/or minor DNA component(s) based on peak height ratio (PHR). Contrastingly, indistinguishable mixture samples are those where the PHR cannot be used to definitively differentiate between the contributors of the alleles and hypothesize two unique genotype sets (1). When a contributor genotype is known, it can assist by in essence designating those alleles to that known contributor, which can allow for easier deconvolution of the remaining alleles. However, this is not always possible. As depicted in Figure 1, genotypes for both the major and minor contributors can be proposed due to the fact that the (a) and (c) alleles and the (b) and (d) alleles, respectively, share similar peak heights. Conversely, in Figure 2, genotypes for both the major and minor contributors cannot be defined due to similarities in peak heights of various alleles.

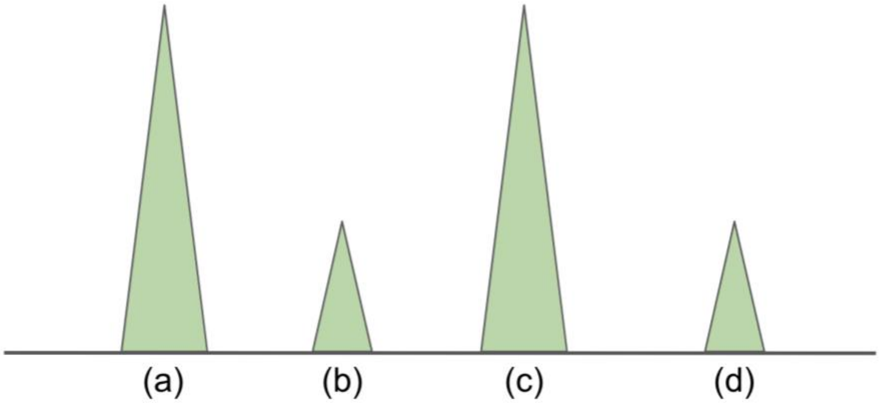


Figure 1. Schematic diagram of a locus in a distinguishable two-person DNA mixture.

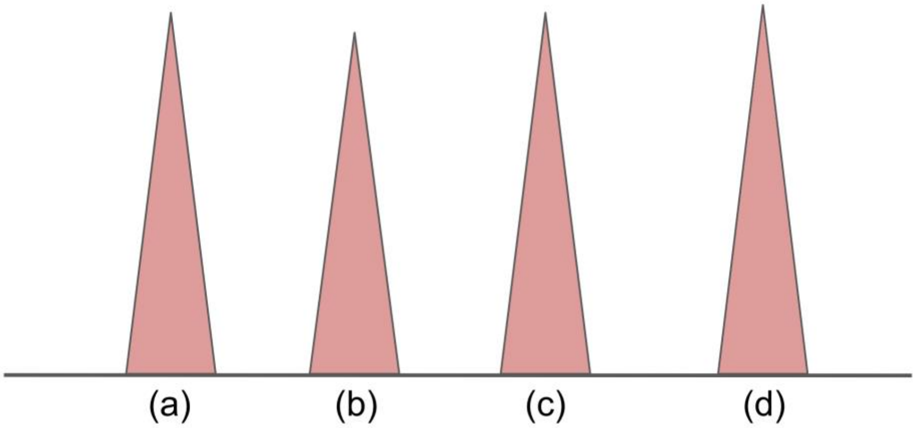


Figure 2. Schematic diagram of a locus in an indistinguishable two-person DNA mixture.

For mixtures with greater than two contributors, which would be characterized by more than four alleles at one or more loci, proposing potential genotypes for each contributor will be more difficult. Studies have shown that when using MAC, failure to correctly identify the NoC can occur in mixtures with more than two contributors (2,3). In one instance, Paoletti et al. concluded that up to 3% of three-person mixtures can be mischaracterized as two-person mixtures, and a staggering 76% of four-person mixtures can potentially be misidentified as either two- or three-person mixtures (3). It has even been suggested that manual interpretation of DNA mixtures with three or more contributors should not be attempted due to its high error rate as a result of the number of possible contributor combinations (4).

1.1.1 Complications to Manual Interpretation

Interpreting DNA profiles manually can potentially call into question subjectivity between analysts who may interpret specific results differently. There are multiple factors of a DNA profile that can complicate interpretation. These include allelic dropout and drop-in, which is the loss or gain of genetic information at a particular locus. Allele sharing between contributors to a mixture can mask certain genetic characteristics that can cloud interpretation and prevent deconvolution. Also, artifacts such as stutter, pull-up, and minus A can confound manual interpretation of DNA profiles.

A common challenge faced by DNA analysts when performing manual DNA mixture interpretation is the presence of allelic dropout and drop-in. Allelic dropout can occur in a DNA profile as a result of low quantity or quality DNA being amplified, and subsequently separated through capillary electrophoresis. As a result, certain alleles exhibited by any of the actual contributors to the DNA mixture can be omitted, or "drop out", from the DNA profile due to low fluorescent signal, resulting in a partial DNA profile (5). Conversely, drop-in can occur due to the presence of trace amounts of extraneous DNA that can result in an allele call that does not reflect the characteristics of any of the actual contributors to the overall mixture (5). Allelic drop-in is less commonly observed than allelic dropout, however, drop-in has been observed in studies using non-template control samples and is most commonly observed in the D18S51 and D8S1179 loci (6).

Allelic dropout and drop-in can complicate manual deconvolution of DNA mixture profiles due to the inherent loss of critical data, or inclusion of extraneous information. In either case, interpretation may be hindered. True contributors to the mixture can potentially be excluded due to dropout, or the inclusion of an extraneous number of alleles that have dropped-in may misguide analysts to include a potential true non-contributor or exclude a true contributor.

Allele sharing between contributors in a DNA mixture can also prevent deconvolution of the mixture. For example, if a major contributor to a DNA mixture exhibits an 11 and 12 allele at a particular locus, and a minor, low-template contributor exhibits a 12 and 13 allele at the same locus, only three peaks will appear. The 12 allele exhibited by the minor contributor becomes masked by the higher signal 12 allele of the major contributor. As a result, the 13 allele cannot be associated with another allele, as the second allele could be an 11, a 12, or another allele that could have dropped out of the profile.

Polymerase chain reaction (PCR) artifacts can also cause difficulty when manually interpreting DNA mixtures. Stutter, the most common PCR artifact, poses particular issues when attempting to manually deconvolute and interpret a DNA mixture. Stutter is an unavoidable product of PCR that is the result of slippage of the DNA polymerase enzyme on the template strand during the reaction, and varies by locus in relation to the length of the uninterrupted repeat sequence (7). This creates artifacts that are not reflective of any biological material in the sample. Stutter can vary in size due to the strand where the slippage of DNA polymerase takes place: in tetranucleotide repeat loci, a stutter artifact can be exhibited four base pairs shorter than the true allele due to a deletion of a repeat, or four base pairs longer than the true allele due to the insertion of a repeat (8). This can create complications in analysis since the difference of four base pairs can lead to the false detection of a different allele, as well as potentially masking minor contributor alleles that fall in a stutter position of another true allele.

During binary modeling interpretations, DNA analysts need to carefully consider all of the potential factors affecting the DNA data and allele designations. While analysts are likely aware of the possible presence of these hindrances, as well as being able to properly characterize and diagnose them, some shortcomings can potentially arise when a complicated mixture profile is comprised with a multitude of extraneous information, or lack thereof, in order to properly deconvolute such a mixture.

1.2 Probabilistic Genotyping Software

While manual interpretation of DNA profile data, is subject to SWGDAM guidelines, the conclusions that can be drawn are not always be reproducible as demonstrated through inter-laboratory and intra-laboratory studies (9). In order to articulate the associations of DNA profiles from an evidentiary item and a reference item, analysts must provide statistical significance to these associations. Experts in the fields of forensic DNA analysis and statistics underwent development of a system that can provide these statistical figures, as well as aid in the deconvolution and interpretation of complex DNA profile data. In 2007, just over two decades after the discovering that DNA can be used as a tool for human identification, the first probabilistic genotyping software (PGS) was developed (5,10). Probabilistic genotyping utilizes semi-continuous or fully continuous models, where the weight of a genotype set can fall anywhere on a continuum, or a scale from 0 to 1. Binary modeling is used during manual DNA interpretations and is defined by having one of two values, 0 or 1, i.e., the allelic peak is either included (weight of 1) or excluded (weight of 0) in the interpretation of a DNA profile. Probabilistic genotyping has proven to provide more information during DNA profile deconvolutions through the use of statistical theory, probability distributions, and computer algorithms to assist DNA analysts in the interpretation of DNA profiles (11).

The use of PGS has developed over the past decade and a half, and in many cases, shows significant improvement over manual interpretation methods. PGS eliminates inconsistencies and variability that may arise between the interpretation by different analysts. Reliability is increased due to computational automation and lack of user interference, which decreases the potential for error that may occur during manual interpretation. Unlike traditional manual approaches, PGS can assist in the deconvolution of mixtures by assigning likely genotypes to possible contributors using population databases that are uploaded in the software. Most importantly, the ability for PGS to interpret low-template DNA within highly complex mixtures in a time-efficient manner makes it highly desirable across forensic DNA laboratories on an international scale (12).

Over 200 forensic DNA laboratories in the United States (U.S.) have begun to use PGS in their casework. This includes laboratories where PGS is currently in use, as well as laboratories that are undergoing internal validation of such software (5,13). Currently, the most commonly employed PGS in forensic laboratories are STRmix™, TrueAllele®, and EuroForMix (14–16). Among the three, STRmix™ is most widely used in the U.S., having been validated and now currently in use in 79 U.S. forensic DNA laboratories (17).

1.2.1 Significance of Likelihood Ratios

Primarily, the statistical value assigned to information collected using PGS is the likelihood ratio (LR). The LR is a comparison between the probabilities of two explanations for the observed DNA data that via the legal parties' hypotheses (18). In DNA mixture interpretation using PGS, those hypotheses are presented by the prosecution and the defense. For example, in a DNA mixture sample recovered from biological evidence found at a crime scene, the prosecution may hypothesize that the DNA mixture contains DNA from a suspected perpetrator, along with either a known victim or some other unknown individual(s). This is referred to as the prosecution hypothesis, and is abbreviated H_p. On the other hand, the defense, whose position in the criminal justice system is to zealously advocate for their client, may hypothesize that the contributors to the mixture are a specific number of unknown persons, explicitly excluding the suspected perpetrator. This hypothesis is the defense hypothesis, which is abbreviated as H_d. The LR value represents a comparison between the likelihood of the two competing hypotheses. The LR equation compares the two hypotheses and how their probabilities, *P*, relate to each other given the evidence, *E*:

$$\mathrm{LR} = \frac{P\,(E \mid \mathrm{H_p})}{P\,(E \mid \mathrm{H_d})}$$

The probability of the more likely hypothesis will be a greater number than that of the other hypothesis being theorized. Therefore, an LR greater than one supports the H_p, and an LR less than one supports the H_d. Depending on the more probable hypothesis, the LR can be a substantially large number, or a substantially small fractional value. There have been verbal scales developed by many different entities in order to translate the LR into words, which suggest various levels of support, or lack of support, for the LR (19,20). The verbal scale suggested by the Federal Bureau of Investigation (FBI) and SWGDAM is shown in Table 1. An LR of one would be uninformative due to equal probabilities of both hypotheses, whereas an LR of one million or greater would provide very strong support in favor of the H_p. For example, the observed DNA results are one million times more probable if the DNA originated from the suspected perpetrator than if the DNA originated from an unknown, unrelated individual. This scale also provides verbal equivalents for LRs ranging from one to over one million.

Table 1. SWGDAM-recommended verbal scale for the interpretation of LR values (20).

LR for H_p Support and 1/LR for H_d Support	Verbal Qualifier
1	Uninformative
2-99	Limited Support
100-9,999	Moderate Support
10,000-999,999	Strong Support
$\geq$ 1,000,000	Very Strong Support

The use of the LR for statistical analysis can be compared to the random match probability (RMP), which for a single-source sample, is the inverse of the LR (5). While the RMP represents statistical significance of evidence as a “one out of many” statement, the LR uses the same value as a factor of likelihood. For example, a statement explaining an RMP would read as follows: “the probability of observing the same DNA profile from an individual other than the suspected perpetrator is one in one million”. Conversely, the LR statement with the same statistical data and significance would read as follows: “based on the evidence, it is one million times more likely if the DNA originated from the suspected perpetrator than if the DNA originated from an unknown, unrelated person”.

Additionally, using LRs provides important advantages over another statistical method, the combined probability of inclusion (CPI). This interpretation method is performed by calculating the sum of the squares of the allele frequencies of each of the observed alleles in a DNA mixture, then determining the product of each of the squared frequencies at each locus (21). The inverse of the product across all loci is the probability that a random person would be included as a contributor to the profile. For example, the formula for CPI when alleles a, b, and c are present in the mixture is as follows:

$$a^2 + b^2 + c^2 + 2ab + 2ac + 2bc = (a + b + c)^2$$

The CPI method does not require an assumption to be made about the NoC, however, it does not consider all of the available evidence in determining statistical significance. Most importantly, a suspect's genotype is not required, and therefore not taken into account, when performing a CPI calculation (22). This method also does not take dropout into consideration. This affects the probative value of the statistics when all of the information from the DNA profile (e.g., peak height) is not taken into consideration. The LR allows for all genetic information to be considered when making such a statistical analysis, and while it may require some inferences as it relates to the NoC, its statistical value is highly representative of the evidence and its genetic characteristics.

1.3 Degradation of DNA

Difficulties in interpretation of DNA profile evidence can also be caused by degradation of the DNA itself. Degradation of DNA is caused by random cleavage of the DNA strand, which can be caused by humidity, exposure to microorganisms, or ultraviolet light (23,24). While DNA can be degraded due to many environmental causes, DNA can also degrade over longer periods of time in a laboratory setting as a consequence of improper storage methods. Degradation can impact all DNA found in biological evidence as time passes. Larger DNA fragments are more susceptible to degradation of the template than smaller DNA fragments (25). As a result, allelic dropout is more likely to take place at the loci with larger amplicon sizes (26). Some PCR primer sets have been developed to reduce amplicon size by placing the primers at the end of each repeat, thus minimizing the effects of degradation (27). However, degradation is not wholly avoidable, and can still pose a threat to DNA when the biological material is subjected to the elements or improperly stored.

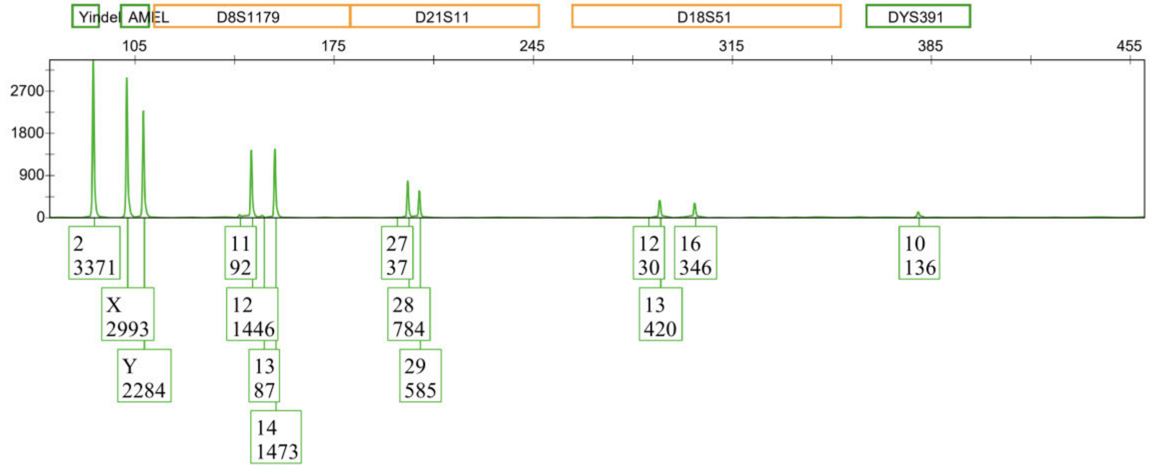


Figure 3. Green panel of a DNA electropherogram exhibiting the “ski-slope” effect, which is characteristic of DNA degradation.

Figure 3 illustrates the effects of degradation on a DNA profile. The loci with smaller amplicon sizes, in base pairs, have alleles with appropriate, expected peak heights. This is due to the signal of these alleles being less affected by degradation than the signal of the alleles that have larger base pair sizes. As the base pair size increases, the signal in

RFU decreases substantially, and may even fall below an established analytical threshold (AT), resulting in allelic or total locus dropout. This visual pattern of substantial loss of fluorescent signal is referred to as the “ski-slope” effect, named for the sharp curved shape that the peak heights create as the base pair size of the alleles increase (28).

Presence of the “ski-slope” effect is characteristic of DNA degradation, and it can impose difficulties on DNA interpretation, especially interpretation of DNA mixtures. The low-signal allelic peaks at larger-sized loci can potentially be misinterpreted as a minor contributor to the DNA mixture, or as some other source of low-template DNA that has been introduced to the sample, which can further complicate DNA mixture interpretation and deconvolution of contributor genotypes.

1.4 DNA-View® Mixture Solution™ Software

Since the advent of PGS in 2007, many different expert systems have been developed for the purpose of statistically analyzing DNA mixture data. DNA-View® Mixture Solution™, developed by Dr. Charles Brenner, is a PGS system utilizing a fully continuous-model based on stochastic modeling. This allows for the program to detect allelic dropout and its statistical weight when analyzing an overall profile (29). Additionally, Mixture Solution is based on a statistical modeling feature that differs from some of the other major PGS currently in use. These major PGS, including STRmix™, TrueAllele®, and EuroForMix, all use the Markov-chain Monte Carlo (MCMC) approach, which estimates genotype sets through repeated simulations, or iterations, to fit the observed DNA data (30). Mixture Solution does not employ MCMC methods, and rather uses a direct computation approach to determine an LR for a specific set of hypotheses (31). Among the benefits of this approach is the employment of a “stopping” feature, which is when the program recognizes that computation has proceeded to a point where an accurate result has been found, which is not commonly found in MCMC-based PGS, and is unreliable when it is used. This allows for Mixture Solution to comprehensively, yet concisely, analyze DNA mixture data.

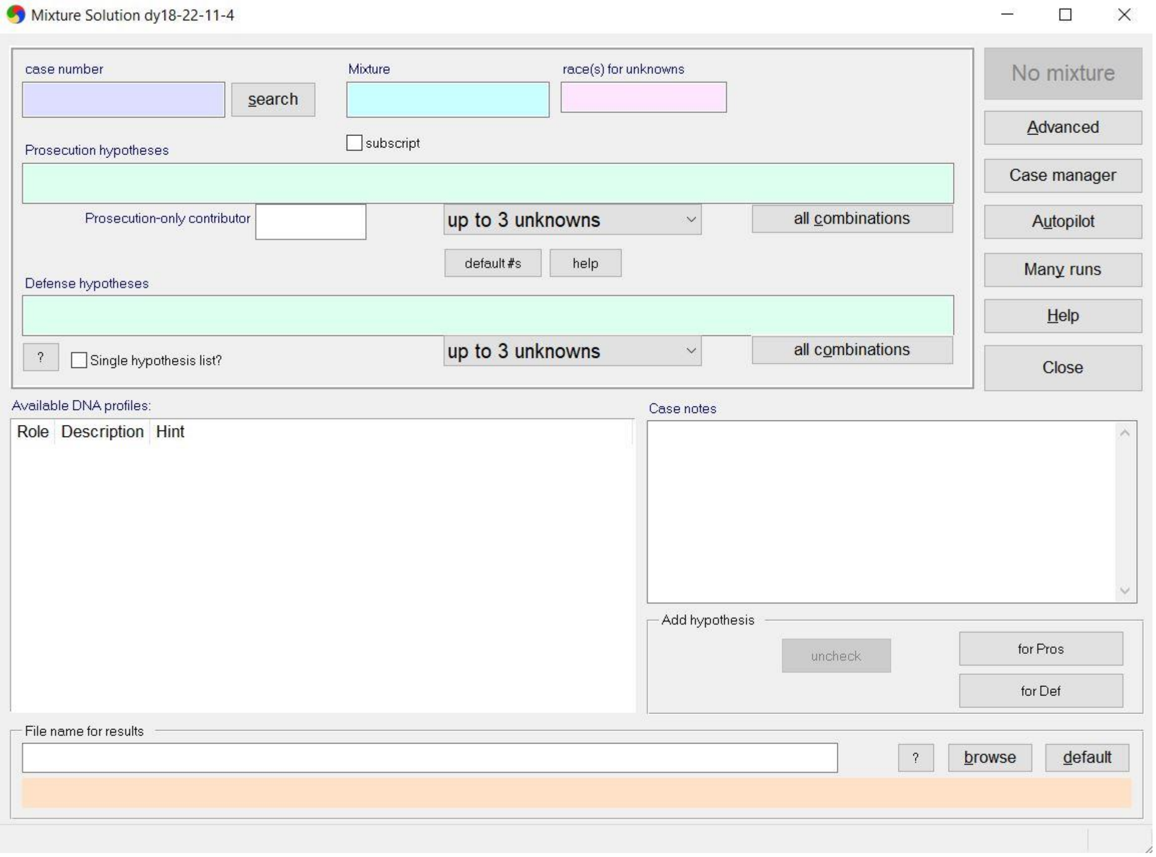


Figure 4. Main menu of Mixture Solution™ depicting the input fields of the program, including for the case information, the mixture in question, and the hypotheses being considered.

Although a shortcoming of LR calculations is the necessity to make an assumption of the NoC, Mixture Solution allows the analyst to enter the number of contributors, or a maximum number of potential unknown contributors, from which the program will create a most probable hypothesis given all of the NoC possibilities. These entries can either be set to the same value for both the prosecution and defense hypotheses or can be set to different values for each of the two hypotheses. The reports that Mixture Solution generates at the conclusion of each computation will indicate the LR for each potential NoC considered, including the estimations that were ultimately not preferred. Mixture Solution is able to consider other minor factors when calculating an LR that include proportions of known and unknown contributors, as well as an estimate of the number of the unknown contributors and their respective genotypes (32).

1.5 Project Goal

DNA-View® Mixture Solution™ is based on stochastic modeling, which can take allelic dropout into account when determining an LR for a DNA mixture. Among the complicating factors that will affect any deconvolution process is degradation (31). Since degradation can significantly impact the analysis of DNA mixture profile data, it would be useful to forensic DNA laboratories if a PGS can be used to assist in the detection of degradation via the production of an accurate LR that will not falsely exclude a degraded contributor.

In this project, three-person mixtures with varying degrees of degradation and contributor ratios were prepared in order to determine how well Mixture Solution could compute accurate LRs for each of the contributors to the mixture, regardless of their overall quantity or degree of degradation. Dropout is considered by Mixture Solution in its computation, and some levels of degradation can lead to allelic dropout. Mixture Solution is able to approximate the "ski-slope" effect of degradation as proportional to the value of one over the base pair size of the allele.

An LR, in theory, would decrease when the DNA of the person of interest (POI) was subjected to conditions that led to degradation of the DNA strands. This can be inferred since the DNA peak heights of the POI will decrease as the size of the loci increase on the electropherogram. Thus, the evidence profile will contain less information as compared to the same DNA in an undegraded state. The average peak heights of undegraded DNA would be relatively consistent across all loci. We expect the LR for a degraded sample will favor the H_p if the POI was present in some form, as long as the alleles contributed by the POI are observable at most loci.

Therefore, the purpose of this study was to assess the capabilities of Mixture Solution in its ability to draw statistical conclusions regarding the presence of contributors in three-person mixture samples at varying levels of degradation and to compare LRs that are representative of the DNA in each mixture sample.

2. MATERIALS AND METHODS

2.1 Samples

Extracted genomic DNA from three individuals (one female, two males) purchased from the Coriell Institute for Medical Research, Camden, NJ, were used to prepare project samples. A 1:10 dilution of each extract was used for real-time PCR amplification.

2.2 Quantification

Human DNA concentrations were measured using the Quantifiler Trio™ DNA Quantification Kit (Applied Biosystems, Foster City, CA) using a 7500 Real-Time PCR System (Applied Biosystems, Foster City, CA) following the manufacturer’s protocol and a calibrated virtual standard curve. Quantification of the degraded samples was performed in duplicate.

2.3 DNA Degradation

Prior to mixture preparation, DNA extracts from each of the three individuals were degraded through a controlled heating procedure (33). To induce degradation, the DNA samples were initially diluted in deionized water to a concentration of 10 ng/µL. Eighty microliters of each sample was incubated at 95°C at multiple time intervals using a Veriti™ Thermal Cycler (Applied Biosystems, Foster City, CA). Samples were incubated at time intervals ranging from 15 to 120 minutes, but the samples incubated at 15 and 30 minutes were chosen for analysis and will be referred to in the results as “partial” degradation (15 minutes) and “full” degradation (30 minutes). Degradation indices (DI) were calculated during quantitation of the samples and are listed in Table 2.

Table 2. Degradation indices for each contributor used in the experimental mixtures.

Contributor	DI at 0 Minutes	DI at 15 Minutes	DI at 30 Minutes
Female	0.81	--	3.01
Male A	0.87	1.29	1.82
Male B	1.42	--	6.53

2.4 Assessment of Control and Degraded Samples

Prior to the preparation of mixture samples, the degraded and undegraded DNA extracts were amplified using the GlobalFiler™ PCR Amplification Kit (Applied Biosystems, Foster City, CA) with a target of 1 ng, according to the manufacturer’s protocol and recommendations. Capillary electrophoresis was performed using 1 µL of amplified product on the SeqStudio™ (Applied Biosystems, Foster City, CA) according to the manufacturer’s protocol with a 10-second injection time. GeneMapper ID-X™ v1.6 was used to analyze and display DNA profiles using an AT of 30 RFU.

2.5 Mixture Sample Preparation

Twenty mixture samples were prepared using DNA extracts from the three individuals. The mixtures were prepared according to the specifications in Table 3, which outlines the ratio of contributors and the level of degradation for each mixture.

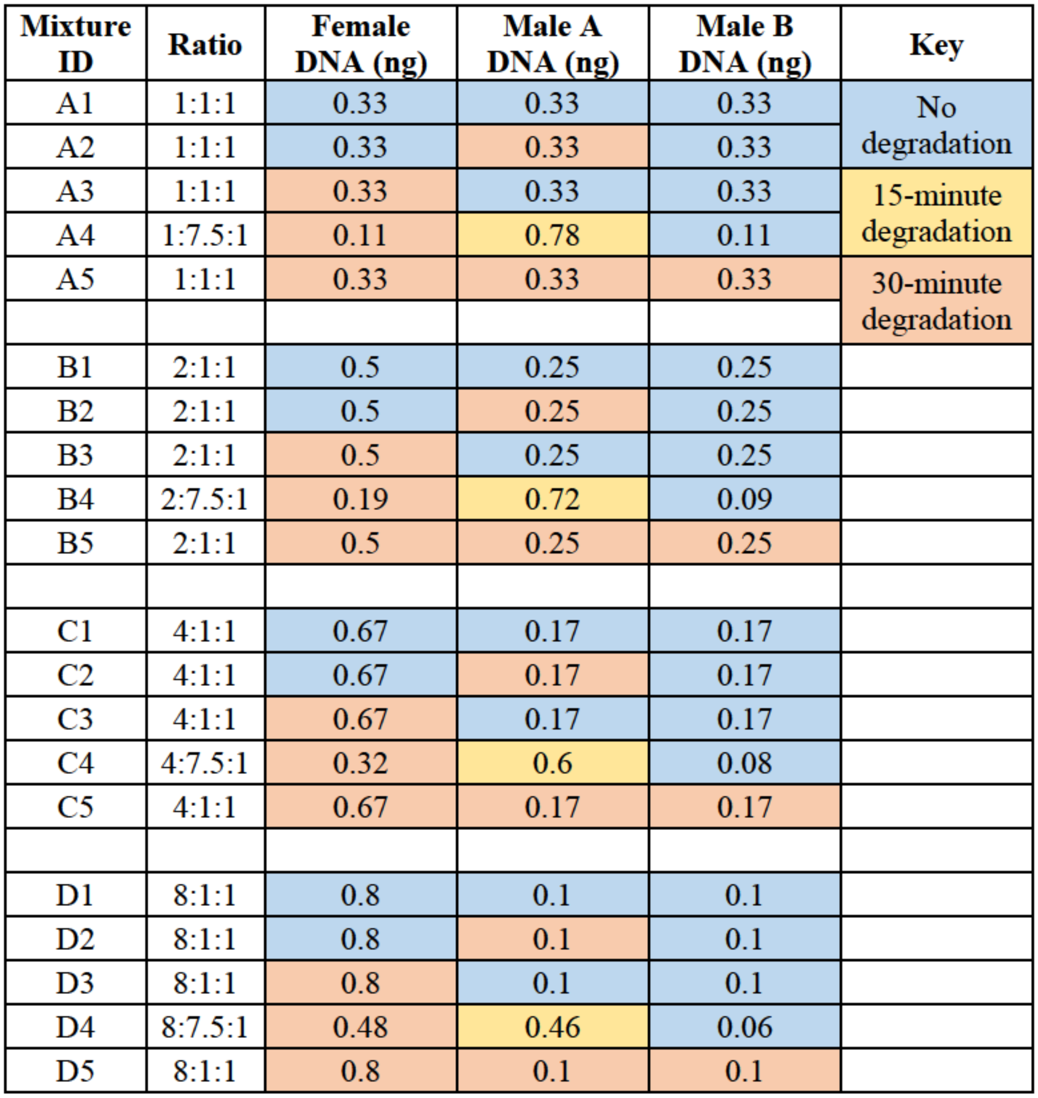
The mixtures were prepared in four groups with five samples per group, denoted using the identifiers A1-A5, B1-B5, C1-C5, and D1-D5, where the letter represents the ratio of contributors, and the number represents the combination of levels of degradation.

The final DNA concentration for the mixtures ranged from 0.07 to 0.44 ng/μL. The intended contributor ratios were: 1:1:1 for the A group, 2:1:1 for the B group, 4:1:1 for the C group, and 8:1:1 for the D group. Note that the ratio for the fourth mixture in each group (A4, B4, C4, and D4) does not follow the intended pattern due to an unrecognized quantification error. The mixture ratios in Table 3 show the corrected ratio for each of these mixtures.

The level of degradation of each contributor was also varied within each group of five mixtures. The first mixture in each series was prepared using undegraded extracts of each contributor and was used for analysis as a control. The second mixture in each group was prepared with the DNA extract of one male contributor degraded for 30 minutes, while undegraded extracts of the female contributor and the second male contributor were used. The third mixture in each group was prepared with the DNA extract of the sole female contributor degraded for 30 minutes, along with the undegraded DNA extracts from each of the two male contributors. The fourth mixture in each group was prepared with the DNA extract of the sole female contributor degraded for 30 minutes, the DNA extract of one male contributor degraded for 15 minutes, and the undegraded DNA extract of the second male contributor. Finally, the fifth and final mixture in each group was prepared with full degradation of all three DNA extracts.

Each mixture was prepared to a final volume of 30 μL with a total quantity of 3 ng of DNA. A final quantification was performed on the mixtures with the Quantifiler Trio™ DNA Quantification Kit in duplicate. Each mixture was amplified using the GlobalFiler™ PCR Amplification Kit at a template target of 1 ng. Capillary electrophoresis was performed using 1 µL of the amplified product on the SeqStudio™ with a 10-second injection time. Using GeneMapper ID-X™ v1.6, electropherograms were generated with an AT of 30 RFU. A genotype table containing the allele calls and peak heights, in RFU, was exported from the GeneMapper ID-X™ software to use with Mixture Solution.

Table 3. Specifications for the contributors in each sample group. The table shows the DNA mass and degradation time for each contributor in each mixture.



Mixture ID	Ratio	Female DNA (ng)	Male A DNA (ng)	Male B DNA (ng)	Key
A1	1:1:1	0.33	0.33	0.33	No degradation
A2	1:1:1	0.33	0.33	0.33	
A3	1:1:1	0.33	0.33	0.33	15-minute degradation
A4	1:7.5:1	0.11	0.78	0.11	
A5	1:1:1	0.33	0.33	0.33	30-minute degradation
B1	2:1:1	0.5	0.25	0.25	
B2	2:1:1	0.5	0.25	0.25	
B3	2:1:1	0.5	0.25	0.25	
B4	2:7.5:1	0.19	0.72	0.09	
B5	2:1:1	0.5	0.25	0.25	
C1	4:1:1	0.67	0.17	0.17	
C2	4:1:1	0.67	0.17	0.17	
C3	4:1:1	0.67	0.17	0.17	
C4	4:7.5:1	0.32	0.6	0.08	
C5	4:1:1	0.67	0.17	0.17	
D1	8:1:1	0.8	0.1	0.1	
D2	8:1:1	0.8	0.1	0.1	
D3	8:1:1	0.8	0.1	0.1	
D4	8:7.5:1	0.48	0.46	0.06	
D5	8:1:1	0.8	0.1	0.1	

2.6 DNA-View® Mixture Solution™

The DNA-View® Mixture Solution™ PGS, software version dy18-22-11-4, was used to generate LRs and most favorable hypotheses for manually-selected persons of interest (POI) in each of the prepared mixture samples. All mixtures were analyzed using Mixture Solution at an AT of 30 RFU. The capabilities of Mixture Solution were assessed through the performance of two different sets of proposed hypotheses for each contributor, each set comprising a test of the program. The program was tested with hypotheses that included unknown contributors, as well as with hypotheses that included some of the known contributors to the mixtures. Upon completion of Mixture Solution's computation, the program designates the most favorable H_p and the most favorable H_d. The LR is then determined based on the comparative ratio between the likelihood for each favored hypothesis.

The first test, Test 1, defined the considered hypotheses using only the POI and some number of unknowns. The H_p input was to consider the POI as a potential contributor to the mixture, with the addition of up to three unknown people as the remaining contributors. The H_d for all samples considered up to three unknown contributors to each mixture. The number of unknowns for both the H_p and H_d was selected as "up to three unknowns" since this is Mixture Solution's default setting for mixture analysis and LR computation.

The second test, Test 2, defined the considered hypotheses using the POI and the two other known contributors to the mixture. The H_p input was to consider the POI as a potential contributor to the mixture, the other two known people who contributed their

DNA to the mixture, and up to three other unknown contributors. For the defense, the H_d input was to consider the two known contributors, excluding the POI, and up to three unknown contributors as having potentially contributed their DNA to the mixture.

The “Autopilot” feature of Mixture Solution was used in order to perform computation on all of the mixture samples and the varying hypotheses based on the specific test with minimal manual user intervention. A script was created using Microsoft Excel to provide a framework which instructed the program of the user’s defined parameters for computation. These parameters included AT, stochastic ratio, integration precision, and theta correction. Each single computation, which was defined by the mixture identifier and the POI for that computation, was replicated ten times to simulate ten trials of each computation. Following analysis, an average LR was calculated for each computation, and all of the LRs reported in the results section represent this average. Mixture Solution also estimated contributor proportions for each mixture and are shown in the results section.

3. RESULTS AND DISCUSSION

3.1 Proportions of Contributors

As part of its interpretation output, the final DNA-View® Mixture Solution™ report calculated the likelihood ratio and generates an estimated proportion of contributors based on peak height differences in the DNA profile. Figures 5 through 8 are representations of the estimated proportion of the three contributors for each of the mixtures analyzed by Mixture Solution. Table 3 defines the intended proportions of contributors that were used in mixture preparation. However, minor variability in concentration results from quantification, mathematical rounding, and potential for pipetting imprecision can be factors in variation between the intended and calculated contributor proportions.

The contributor ratios that were calculated by Mixture Solution for mixtures A1, B1, C1, and D1 were within 14% of the intended proportions. As the level of degradation increased for a POI, the estimated contributor proportion of that POI generally decreased, where the other contributors to the mixture had an increase in the estimated proportion. This can be attributed to the fact that the degradation was performed on the contributors individually and prior to amplification, and not on the mixture as a whole. As a result, amplification of degraded DNA may lead to decreased yield of PCR product which can vary per contributor (34).

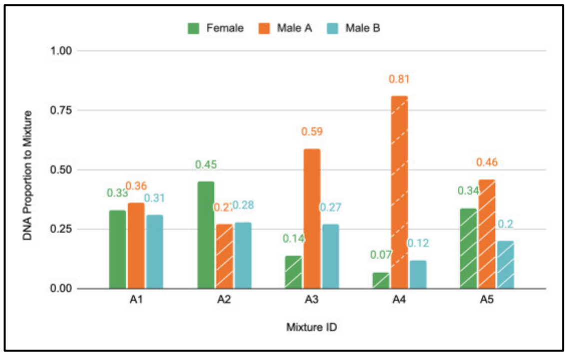


Figure 5: Mixtures with 1:1:1 contribution ratio

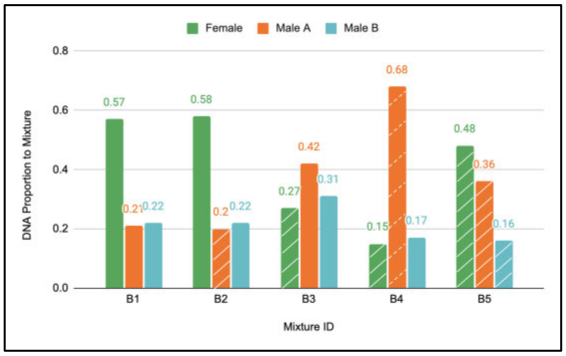


Figure 6: Mixtures with 2:1:1 contribution ratio

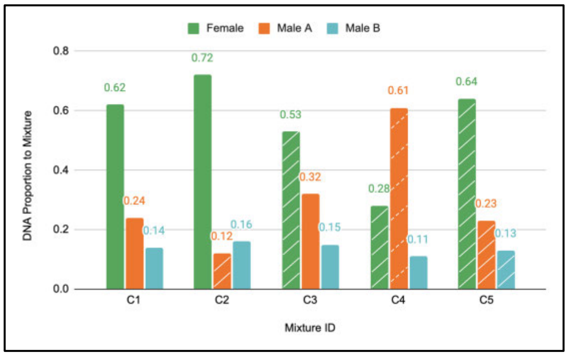


Figure 7: Mixtures with 4:1:1 contribution ratio

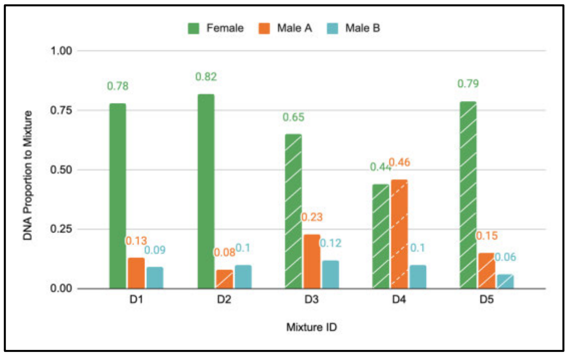


Figure 8: Mixtures with 8:1:1 contribution ratio

Figures 5-8. Bar graphs representing the average contributor proportions that Mixture Solution estimated for each mixture in Test 2. Solid diagonal lines represent fully degraded DNA, and dotted diagonal lines represent partially degraded DNA. The fourth mixture in each group (A4, B4, C4, and D4) has a ratio that is different from the other mixtures in the same group, as defined in Table 3. Note that Y-axis scales differ across figures.

3.2 The Degradation Function

While degradation of DNA cannot easily be defined in a quantitative manner, it can be assessed by observing changes in a DNA profile relative to an undegraded DNA profile, primarily through the observation of the presence and intensity of the ski-slope effect. With degradation, peak heights tend to decrease substantially as DNA amplicon sizes increase (26). Figure 9 depicts a distribution of all short tandem repeat (STR) alleles for the male A contributor at the three different levels of degradation used in this study. These levels of degradation include no degradation, partial degradation, and full degradation, defined as

0 minutes, 15 minutes, and 30 minutes of exposure to heat, respectively. Homozygous loci were treated as two alleles, by halving the peak height of the homozygous allele and indicating it as two individual data points.

As expected, the alleles observed in the samples not subjected to degradation did not exhibit a strong correlation of decay between peak height and base pair size compared to partially and fully degraded samples. Among all alleles observed in the undegraded male A single source profile, thirteen alleles had peak heights greater than the maximum peak height observed in the partially degraded male A single source profile. For the partially degraded samples, over 85% of the alleles yielded peak heights that were greater than the average peak height of the profile.

In the profile of the partially degraded DNA sample, the loss of allelic signal was observed in the allele distribution although most of the alleles' peak heights fell in the same range as the peak heights of the alleles observed in the DNA that was not subjected to degradation. Only four alleles in the partially degraded DNA profile had peak heights that were less than the minimum peak height observed in the undegraded DNA profile. No allelic dropout was observed in the partially degraded DNA profile. Reflected by the peak height of the alleles in the partially degraded profile and the substantial best fit curve of the data, the DNA of male A exposed to 15 minutes of heat resulted in the greatest substantial decay of peak height across the profile compared to no degradation and full degradation samples. However, peak heights in this profile are still large enough to be comparable to a pristine DNA profile, which would not affect the mixture interpretation.

Finally, the fully degraded DNA profile also exhibited a substantial loss of RFU across alleles, specifically targeting larger DNA fragments. None of the alleles in the fully degraded DNA profile yielded a peak height greater than the corresponding allele in the no degradation or partial degradation samples. No allelic dropout occurred in the single-source fully degraded profile of male A, however, the average peak height of alleles in this profile was over four times less than that of the average peak height of the alleles in the partially degraded DNA profile, and over seven times less than that of the average peak height of the alleles in the DNA profile not subjected to any degradation.

However, a factor within the DNA profiles that was not impacted by increasing the level of degradation was the average peak height ratio. The average peak height ratio of the alleles at heterozygous loci in the fully degraded sample demonstrated the lowest heterozygosity of approximately 0.85. The partially degraded sample had the greatest average peak height ratio of approximately 0.89. The average peak height ratio in the undegraded DNA profile was approximately 0.87. Although larger size amplicons are more susceptible to damage as a result of degradation, the individual alleles within each locus tended to lose signal frequency at the same rate regardless of level of degradation.

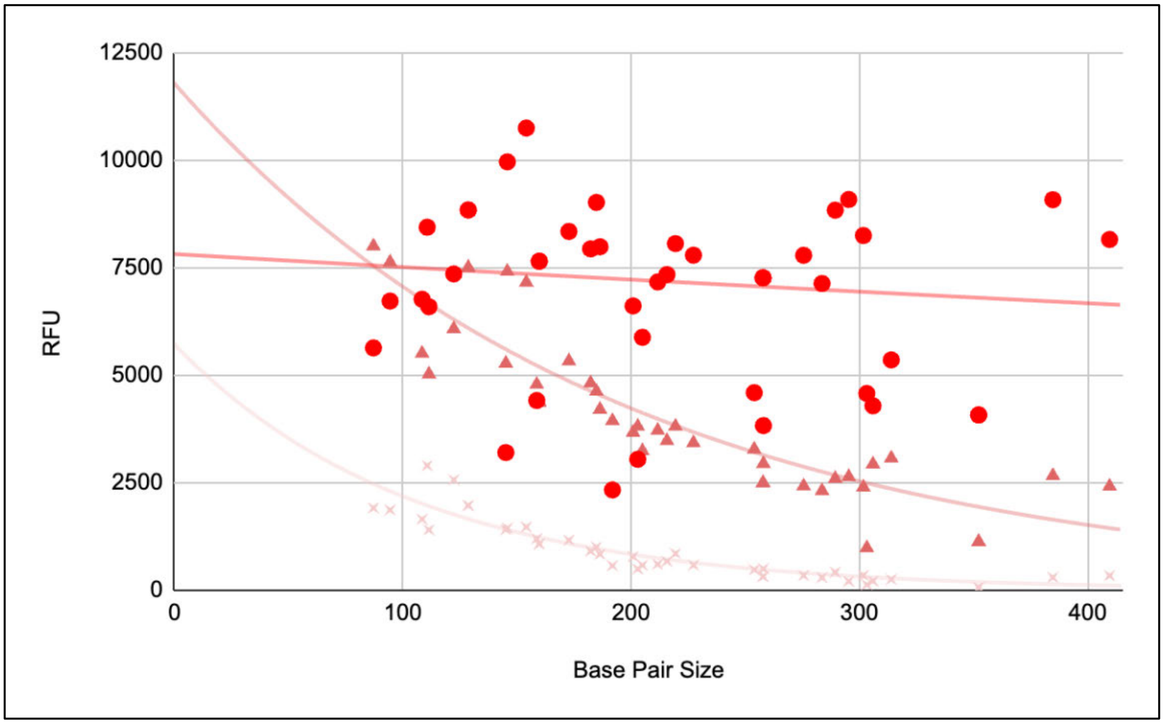


Figure 9. Distribution of size and peak height of alleles exhibited by male A at different levels of degradation. Red circular points represent undegraded DNA, lighter red triangle data points represent partially degraded DNA, and lightest red X shaped data points represent fully degraded DNA.

Figures 10-12 represent the ski-slope effect observed in the profile of full DNA degradation for each of the three contributors. The peak heights of all autosomal STR alleles in the DNA profile were plotted against the DNA fragment sizes (in base pairs), independent of locus or dye channel. Concordantly to the data in Figure 9, the peak heights of all homozygous alleles were halved in Figures 10-12 and indicated as two individual data points for a consistent comparison.

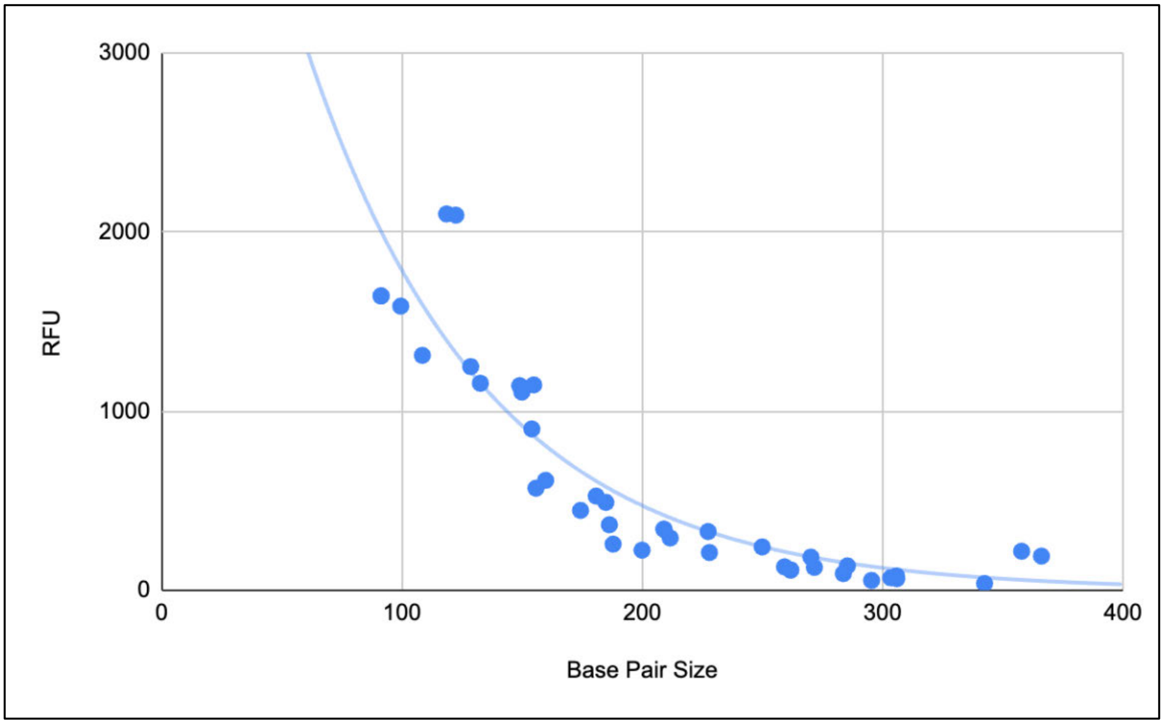


Figure 10. Representation of ski-slope effect observed in fully degraded DNA of the female contributor across all alleles in a single-source profile**.**

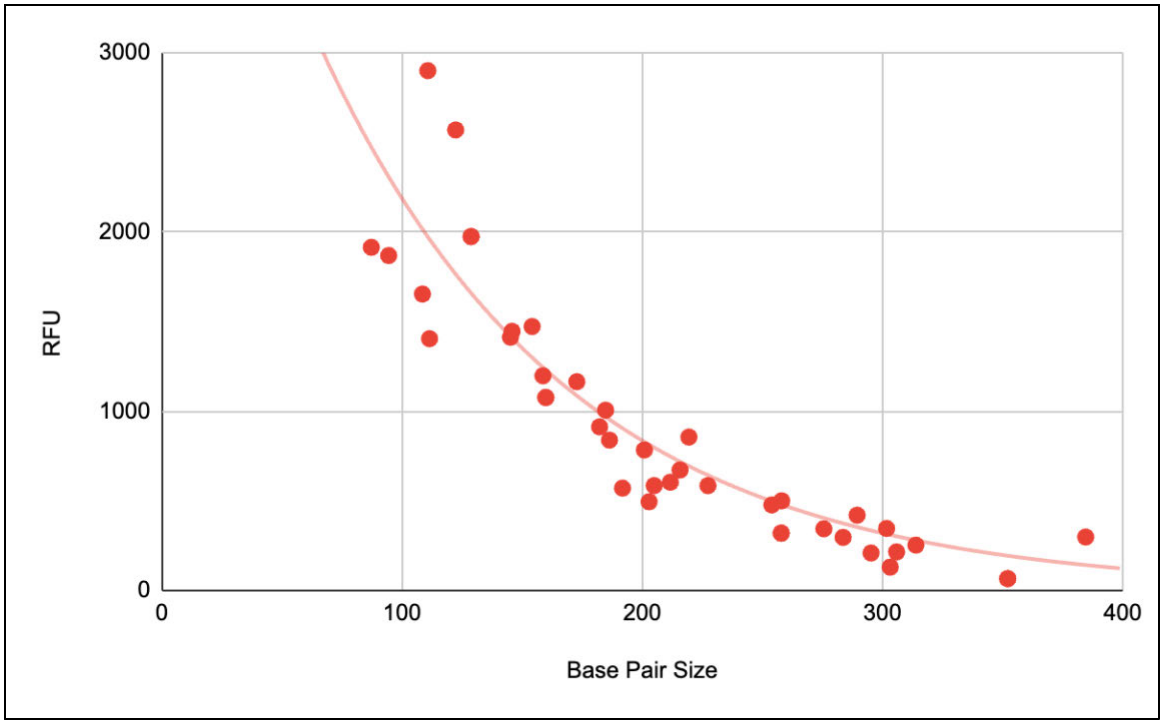


Figure 11. Representation of ski-slope effect observed in fully degraded DNA of male A across all alleles in a single-source profile**.**

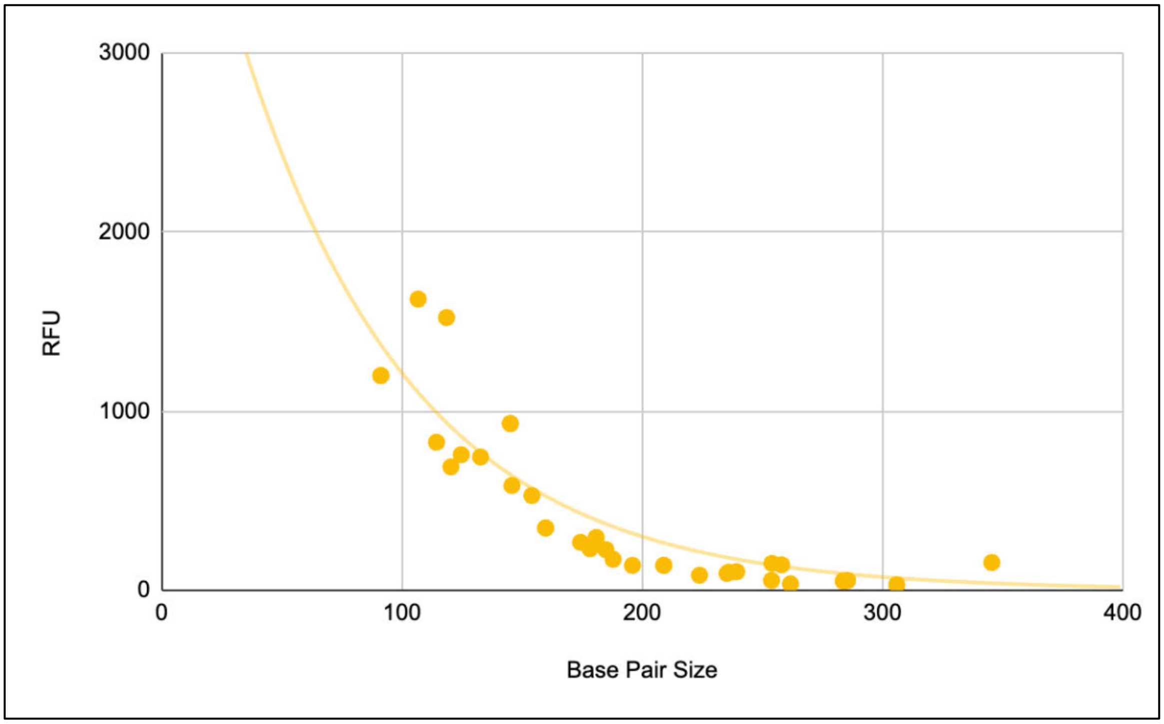


Figure 12. Representation of ski-slope effect observed in fully degraded DNA of male B across all alleles in a single-source profile**.**

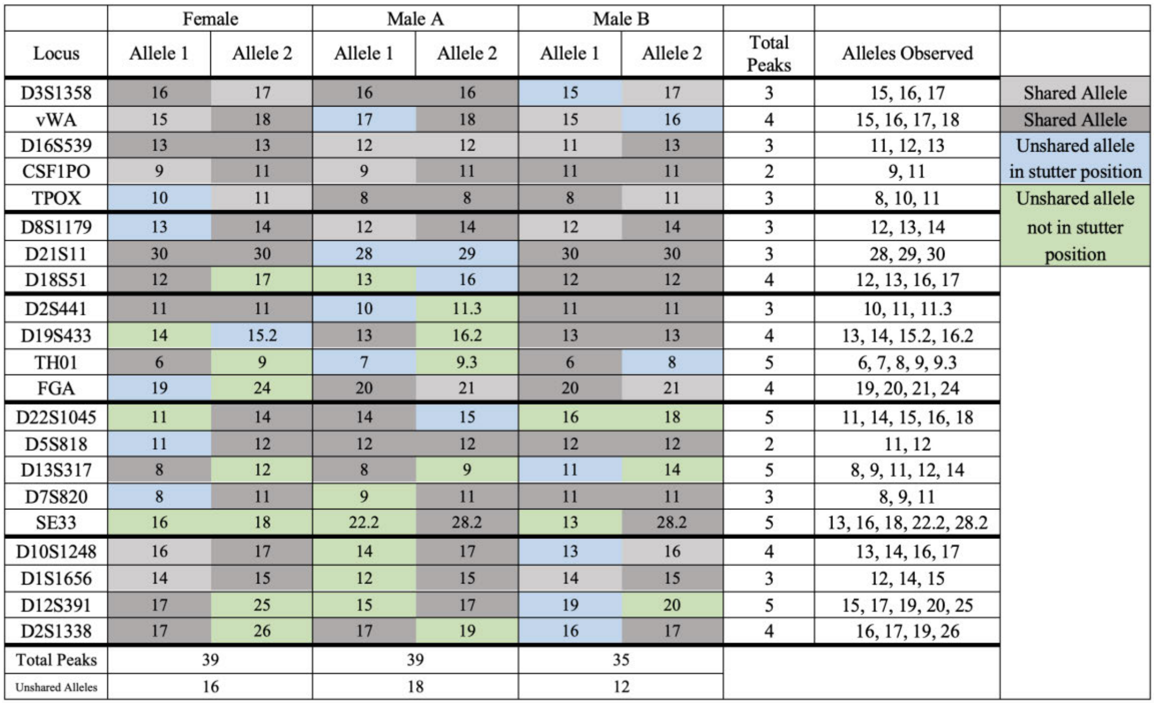
With regards to allelic dropout, male B exhibited the most dropout. In the single-source fully degraded sample, five alleles dropped out. This includes a single homozygous 11 allele at the CSF1PO locus, both 8 and 11 alleles at the TPOX locus, the single allele at the Y chromosome locus DYS391, and a 16 allele at the D2S1338 locus. Multiple alleles at larger sized loci had an RFU that was just above the AT. The female contributor only lost two alleles due to dropout, resulting in total locus dropout at the TPOX locus. In all cases of dropout from these two contributors, the dropout occurred at the largest or second-largest locus in its specific dye channel. This is consistent with the damaging effects of DNA degradation of DNA of larger amplicons. The fully degraded DNA profile of male A did not exhibit allelic dropout at any locus. While Mixture Solution uses stochastic modeling that can predict and account for dropout, the increased number of alleles that have dropped out from a degraded profile can potentially result in a reduced LR.

3.3 Allele Sharing

One complicating factor of DNA mixture interpretation is the possibility of allele sharing between contributors. For example, low signal RFU of a minor contributor may be masked if that allele is shared with the major contributor to the mixture. Table 4 compares the genotypes of the three contributors used to prepare the mixtures for this study, and highlights allele shares that may potentially complicate mixture deconvolution. These shares were classified separately depending on whether the shared allele is located in a stutter position relative to a true allele exhibited by any of the contributors. This table also indicates the expected total number of peaks and the allele calls that should be observed at each locus in the mixtures.

Among the three contributors, male A exhibited the greatest number of unique alleles, with eighteen unshared alleles. This was followed by the female contributor with sixteen unshared alleles, and lastly, by male B, with twelve unshared alleles. Over half of the female's and male A's unshared alleles did not fall in a stutter position, thus a majority of each of their unshared alleles would not be subjected to masking by a potential stutter peak at that base pair location. However, seven of male B's twelve unshared alleles were located in stutter positions relative to other alleles, which could lead to potential masking in the final mixture profile.

Table 4. Genotypes of individual mixture contributors. Includes indications of allele sharing according to the key at right of the table.



	Female		Male A		Male B				
Locus	Allele 1	Allele 2	Allele 1	Allele 2	Allele 1	Allele 2	Total Peaks	Alleles Observed	
D3S1358	16	17	16	16	15	17	3	15, 16, 17	Shared Allele
vWA	15	18	17	18	15	16	4	15, 16, 17, 18	Shared Allele
D16S539	13	13	12	12	11	13	3	11, 12, 13	Unshared allele in stutter position
CSF1PO	9	11	9	11	11	11	2	9, 11	
TPOX	10	11	8	8	8	11	3	8, 10, 11	Unshared allele not in stutter position
D8S1179	13	14	12	14	12	14	3	12, 13, 14	
D21S11	30	30	28	29	30	30	3	28, 29, 30	
D18S51	12	17	13	16	12	12	4	12, 13, 16, 17	
D2S441	11	11	10	11.3	11	11	3	10, 11, 11.3	
D19S433	14	15.2	13	16.2	13	13	4	13, 14, 15.2, 16.2	
TH01	6	9	7	9.3	6	8	5	6, 7, 8, 9, 9.3	
FGA	19	24	20	21	20	21	4	19, 20, 21, 24	
D22S1045	11	14	14	15	16	18	5	11, 14, 15, 16, 18	
D5S818	11	12	12	12	12	12	2	11, 12	
D13S317	8	12	8	9	11	14	5	8, 9, 11, 12, 14	
D7S820	8	11	9	11	11	11	3	8, 9, 11	
SE33	16	18	22.2	28.2	13	28.2	5	13, 16, 18, 22.2, 28.2	
D10S1248	16	17	14	17	13	16	4	13, 14, 16, 17	
D1S1656	14	15	12	15	14	15	3	12, 14, 15	
D12S391	17	25	15	17	19	20	5	15, 17, 19, 20, 25	
D2S1338	17	26	17	19	16	17	4	16, 17, 19, 26	
Total Peaks	39		39		35				
Unshared Alleles	16		18		12				

3.4 Likelihood Ratio Comparisons

Tables 5 and 6 summarize the LRs computed by Mixture Solution for each of the POIs, reflecting changes in contributor proportions and level of degradation. The color scheme used in both Tables 5 and 6 is the same as the color scheme defined in Table 3 and identifies the level of degradation of each contributor. Table 5 lists the LRs that were computed using the parameters set for Test 1, and Table 6 lists the LRs that were computed using the Test 2 parameters.

All LRs determined by Mixture Solution favored the H_p, which is directly concordant to accurate associations with the given POIs to the mixture. As a result, no Type I errors were observed with Mixture Solution's computations for the data analyzed.

To reflect the SWGDAM verbal scale of interpretation for LR values, all LRs that are in the “very strong support” range have been listed in scientific notation, while any LR value that falls below this range was listed in standard notation. Mixture Solution was able to compute an LR with “very strong support” for 83% of the mixtures in Test 1.

Generally, the LR increased in Test 2 compared to those assigned to Test 1, irrespective of the contributor designated as the POI, the ratio of contributors, or the level of degradation. These results were expected due to the inclusion of known contributors in both hypotheses allowing Mixture Solution to condition those genotypes when performing the computation, thus, eliminating uncertainty in the conditioned contributor’s designated alleles. As a result, the number of mixtures that were assigned an LR in the “very strong support” range increased to 95% for Test 2. Also, in most cases, the assigned LR decreased when the POI contributed degraded DNA to the mixture. The decrease in signal or complete loss in peak detection complicates associations of the POI to the mixture when compared to the use of undegraded DNA. These general findings are consistent with prior studies using PGS to analyze degraded DNA samples (35,36).

Table 5. Likelihood ratios for each contributor in Test 1 where the H_p was defined as the mixture containing the DNA of the POI and up to three unknown people, and where the H_d was defined as the mixture containing the DNA of up to three unknown people, excluding the POI as being a potential contributor to the mixture.

(H_p) = POI + x unknowns, (H_d) = all unknowns			
	POI		
Mixture ID	**Female**	**Male A**	**Male B**
A1	1.54E+16	7.08E+13	1.24E+14
A2	1.07E+20	1.28E+10	9.09E+11
A3	831000	1.10E+21	2.02E+12
A4	6520	1.91E+29	1.96E+12
A5	1.00E+09	1.55E+15	12200
B1	1.01E+26	2.24E+14	4.77E+11
B2	7.21E+26	2.64E+09	2.49E+11
B3	8.94E+08	1.14E+14	1.24E+11
B4	30200	1.99E+26	5.88E+09
B5	1.14E+16	3.88E+11	20500
C1	3.59E+29	7.69E+12	8.36E+08
C2	1.49E+31	2.41E+09	4.99E+12
C3	7.94E+11	9.68E+10	471000
C4	7.25E+12	6.48E+23	143000
C5	2.47E+23	2.49E+12	1.10E+06
D1	2.39E+31	1.15E+14	3.93E+07
D2	3.21E+31	1.69E+07	6.04E+08
D3	1.62E+16	9.15E+08	246000
D4	1.21E+13	6.50E+16	118000
D5	3.91E+26	5.17E+13	2950

Table 6. Likelihood ratios for each contributor in Test 2 where the H_p was defined as the mixture containing the DNA of the POI and two known people, and where the H_d was defined as the mixture containing the DNA of the other two known people, excluding the POI, and one unknown person as a third contributor.

(Hp) = POI + 2 knowns, (Hd) = 2 knowns + 1 unknown			
Mixture ID	**Female LR**	**Male A LR**	**Male B LR**
A1	7.21E+30	3.14E+29	1.80E+27
A2	7.51E+29	5.71E+22	9.36E+21
A3	1.51E+13	8.38E+28	1.44E+21
A4	207.8	6.89E+29	6.39E+10
A5	5.15E+12	1.60E+21	4.87E+06
B1	4.59E+31	7.11E+25	4.65E+17
B2	2.07E+31	2.25E+18	8.04E+17
B3	4.15E+19	1.05E+26	5.77E+20
B4	3.34E+06	1.28E+29	2.18E+11
B5	3.77E+21	5.02E+20	1.06E+07
C1	8.37E+31	2.20E+24	5.01E+14
C2	8.33E+31	2.38E+14	7.01E+17
C3	3.69E+20	7.06E+20	1.21E+10
C4	1.27E+18	7.04E+28	1.54E+07
C5	1.71E+26	3.22E+18	4.24E+08
D1	7.92E+31	1.91E+18	6.25E+10
D2	9.30E+31	9.19E+08	3.86E+10
D3	1.41E+23	4.19E+16	1.00E+09
D4	6.81E+22	7.83E+25	368700
D5	2.15E+27	4.35E+16	150620

Figures 13-15 show the differences in LR that were computed using the parameters set in Test 1 for each POI. Each data point is labeled with the mixture ID defined for each of the mixtures in Table 3. Generally, a greater LR was assigned for a POI who contributed undegraded DNA to the mixture compared to when the POI contributed degraded DNA to the mixture. Additionally, computed LRs generally tended to increase for a particular POI when the quantity of that person's DNA input was increased, even when the increased amount of DNA was degraded.

Figure 13 compares the LRs assigned by Mixture Solution to the female contributor. When comparing samples with equivalent levels of degradation, all LR values increased when the amount of the female contributor's DNA input was increased. For the mixtures where the DNA of the female contributor was not degraded, the LR values were greater than that of those mixtures which contained degraded DNA. In the undegraded control mixtures, represented in blue in Figure 13, the LR assigned to the female contributor was slightly less than the LR assigned to the mixtures where only male A contributed degraded DNA to the mixture, represented in red. This was observed in the 1:1:1, 2:1:1, and 4:1:1 mixtures, however, when the ratio was increased to the maximum 8:1:1, the LRs for the female contributor in mixtures D1 and D2 became almost identical. This is likely due to the female contributor emerging as the major contributor without subjectivity in allele assignment, thus reaching an "LR ceiling", since a maximum log(LR) value of 31 was observed in this data.

For the mixtures where the female contributed degraded DNA, the LR values never exceeded the LRs corresponding to the same undegraded DNA input amount. However, they steadily increased as the input amount of female DNA increased. The lowest LR values for the female contributor were assigned when the female DNA degraded and the DNA of at least one other contributor was undegraded. This includes some LR calculations that fell below the "very strong support" threshold. It was expected that the assigned LR values for the female contributor would be greater than the other two contributors, as a result of being the major contributor in most mixtures, which was true in most cases.

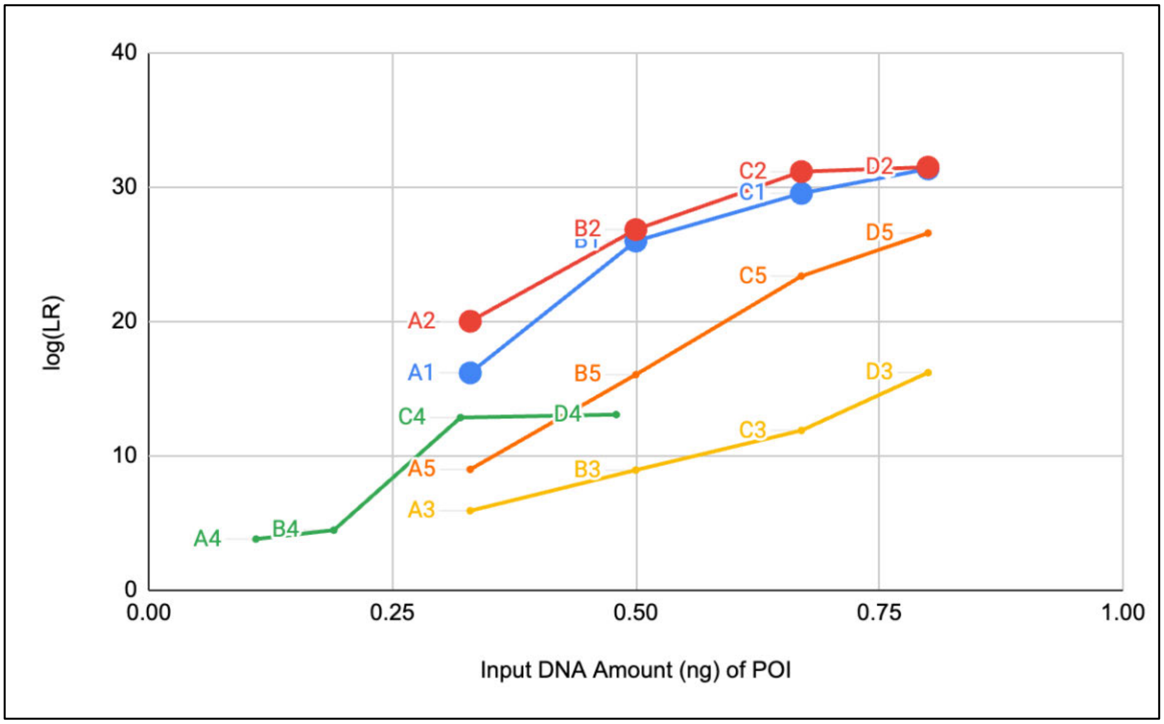


Figure 13. LR calculations in Test 1 where the female contributor was the POI. Small data points represent the POI was a degraded contributor, and large data points represent the POI contributed undegraded DNA.

Figure 14 compares LRs assigned by Mixture Solution to the male A contributor. Male A was the only contributor who contributed DNA at three different levels of degradation, where in addition to DNA with no degradation and full degradation, the fourth set of mixtures in each group included partially degraded DNA from male A. In the mixtures where male A contributed undegraded DNA, the LR values decreased as the mass of male A decreased.

For the two sets of mixtures where male A contributed fully degraded DNA, the LR values for male A as the POI were less than most other LRs for different mixtures with undegraded DNA from male A. In the set of mixtures where male A was the only degraded contributor, the computed LR values were the least among all of the mixtures with male A as the POI. However, in the mixture set where the DNA of all contributors were fully degraded, the LRs were comparable with the control mixtures, however, decreased as the input amount increased. This intersecting of values is consistent with the effects of input amount countering the effects of DNA degradation.

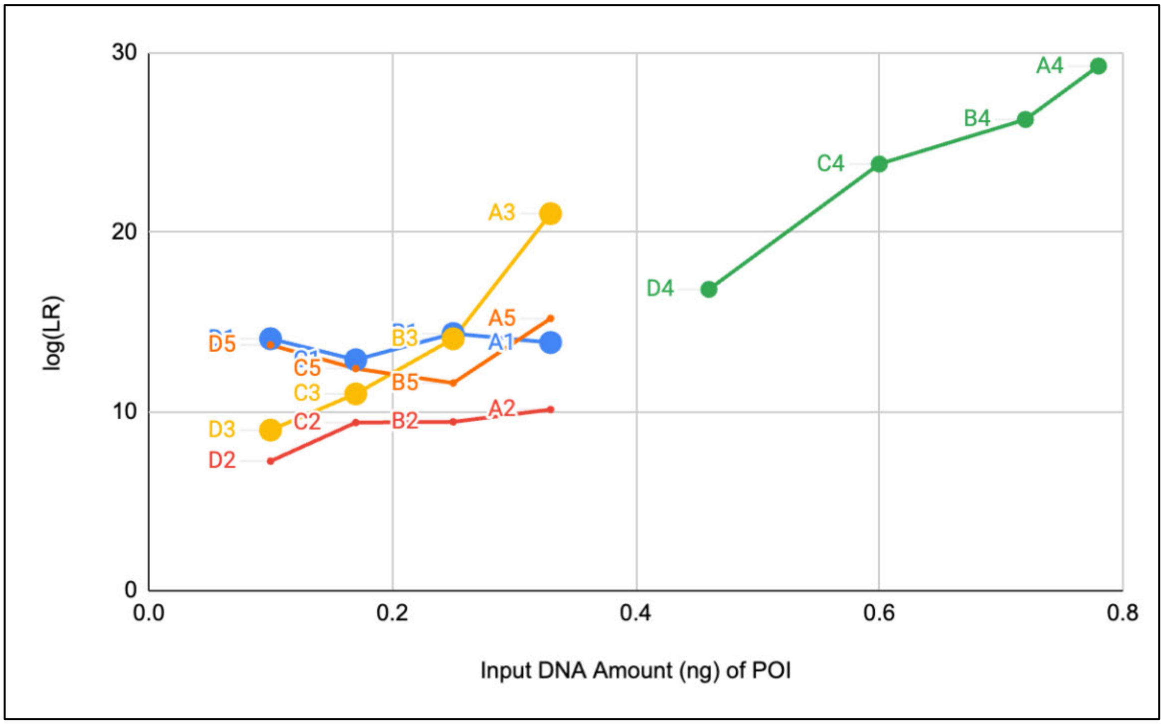


Figure 14. LR calculations in Test 1 where male A was the POI. Smallest data points represent the POI was a fully degraded contributor, largest data points represent the POI contributed undegraded DNA, and the intermediate size data points represent the POI was a partially degraded contributor.

Figure 15 compares LRs assigned by Mixture Solution to the male B contributor. Overall, the LR values assigned to male B when designated as the POI were less than that of the other two contributors. This is mostly unexpected since male B only contributes degraded DNA in one set of mixtures. However, this is likely related to male B having the least number of shared alleles of the three contributors. The degraded DNA of male B also showed the most allelic dropout in the degradation study as seen with the single-source fully degraded male B profile.

For the control mixtures, the LRs for male B steadily increased as input amount increased, as shown in the blue line in Figure 15. However, when other contributors were masked by degradation, in the third and fourth sets of mixtures (represented by yellow and green in Figure 15, respectively), the assigned LRs for male B as the POI remained less than the LRs assigned to male B for the control mixtures. It would be expected that contributors with undegraded DNA would be assigned higher LRs when other contributors are masked by either degradation, allele sharing, or dropout.

The only mixture set where male B contributed degraded DNA was the fifth set, where all contributors were fully degraded. With the exception of mixture C5, the assigned LRs to male B were the least in comparison to the mixtures where undegraded DNA was used. The result in C5 could be due to errors in sample preparation or variation in DNA quantification values that may have resulted in differing DNA inputs to the mixture than intended. These minor variations in DNA input, of either degraded DNA or undegraded DNA, can impact peak height, and subsequent downstream analysis, up to and including the use of PGS for mixture interpretation.

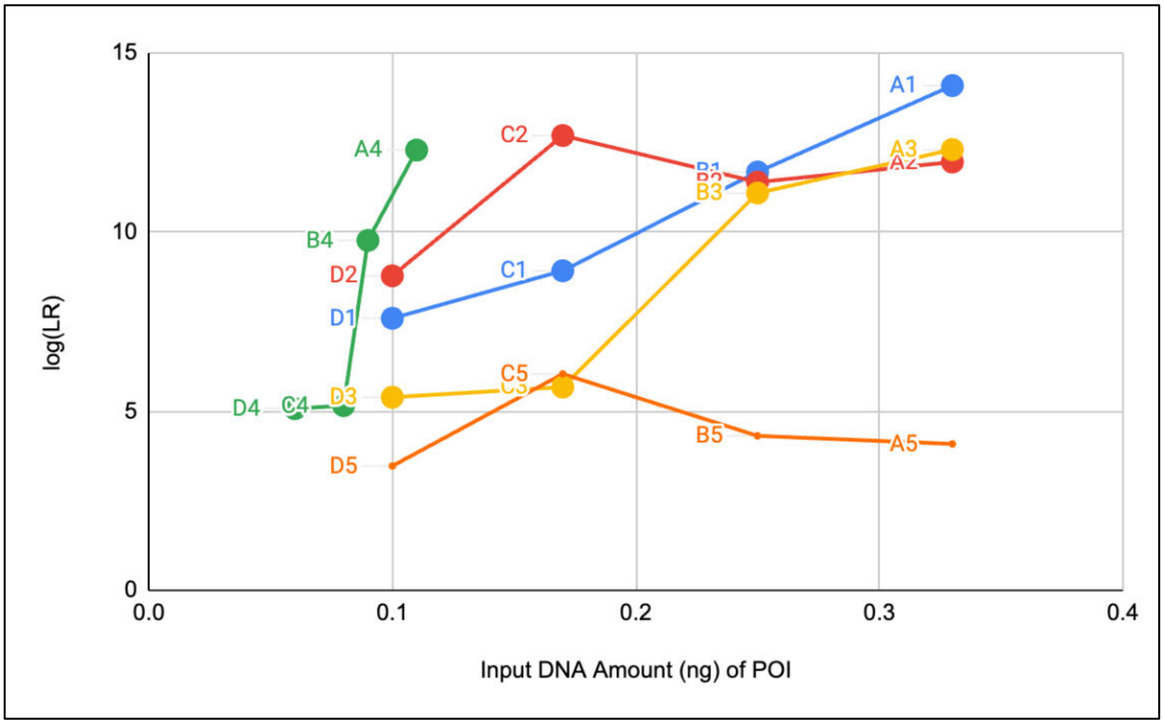


Figure 15. LR calculations in Test 1 where male B was the POI. Small data points represent the POI was a degraded contributor, and large data points represent the POI contributed undegraded DNA.

Figures 16-18 are graphical representations of the differences in LR that were computed using the parameters set in Test 2, which included the identity of known contributors in both the H_p and the H_d. The most obvious observation is that the LR values for corresponding mixtures increased when known contributors were defined in the hypotheses, which occurred in all but one of the mixtures. Similar to that of the LR values computed in Test 1, LRs remained the same or increased as DNA input amount increased. Additionally, LRs assigned to undegraded DNA contributors continued to exceed those assigned to contributors with degraded DNA.

Figure 16 compares LRs calculated under the Test 2 parameters for the female contributor. While LRs for the sets of mixtures where the female contributed undegraded DNA generally remained constant as the DNA input amount increased, almost entirely overlapping, the mixtures where the female contributor was fully degraded increased with DNA input. The curves incorporating the data points followed a similar pattern to the assigned LRs under the parameters and hypotheses for Test 1.

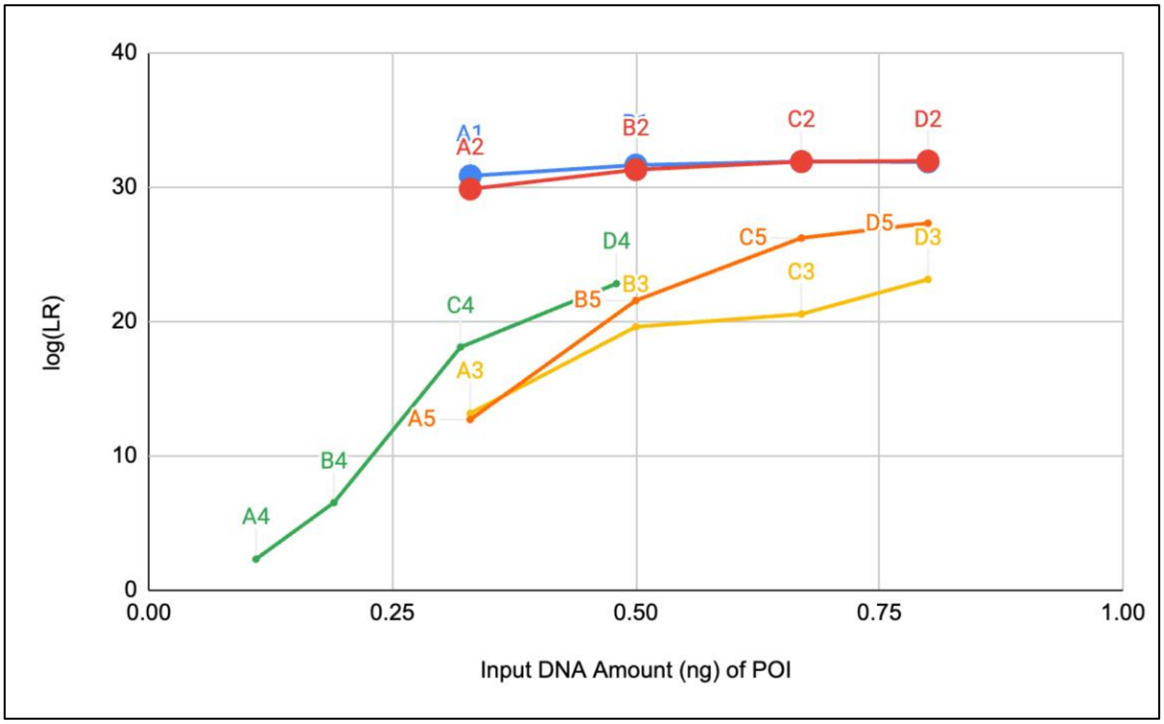


Figure 16. LR calculations in Test 2 where the female contributor was the POI. Small data points represent the POI was a degraded contributor, and large data points represent the POI contributed undegraded DNA.

Figure 17 compares LRs computed under the Test 2 parameters for male A. As in Test 1, the LR values increased consistently with the increase of the DNA input amount. Mixture sets where male A contributed undegraded DNA produced greater LR values than the sets where fully degraded DNA was contributed. Each of the LRs in these undegraded sets increased on a logarithmic scale with an increase in the DNA input amount. The LRs of the mixture sets where male A contributed degraded DNA also increased relative to increasing DNA input amount, however, remained less than that of the other mixture sets where undegraded DNA was contributed by male A.

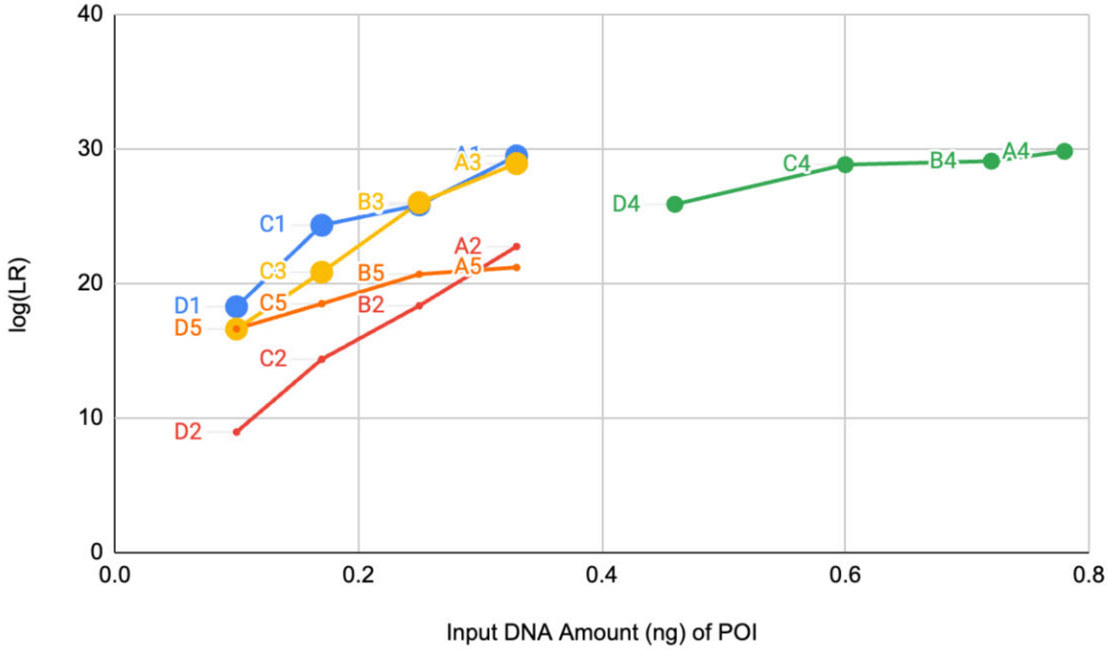


Figure 17. LR calculations in Test 2 where male A was the POI. Smallest data points represent the POI was a fully degraded contributor, largest data points represent the POI contributed undegraded DNA, and the intermediate size data points represent the POI was a partially degraded contributor.

Figure 18 depicts comparisons between the Test 2 LRs assigned to male B. All mixture sets, except for those that contained fully degraded DNA from all three contributors, exhibited a logarithmic increase in LRs as DNA input increased. Comparatively to Test 1, mixture C5 is the only sample to yield an LR for male B slightly greater than the LR for the undegraded DNA. When excluding outlier C5, the assigned LRs remain fairly constant as the quantity of DNA for male B increases.

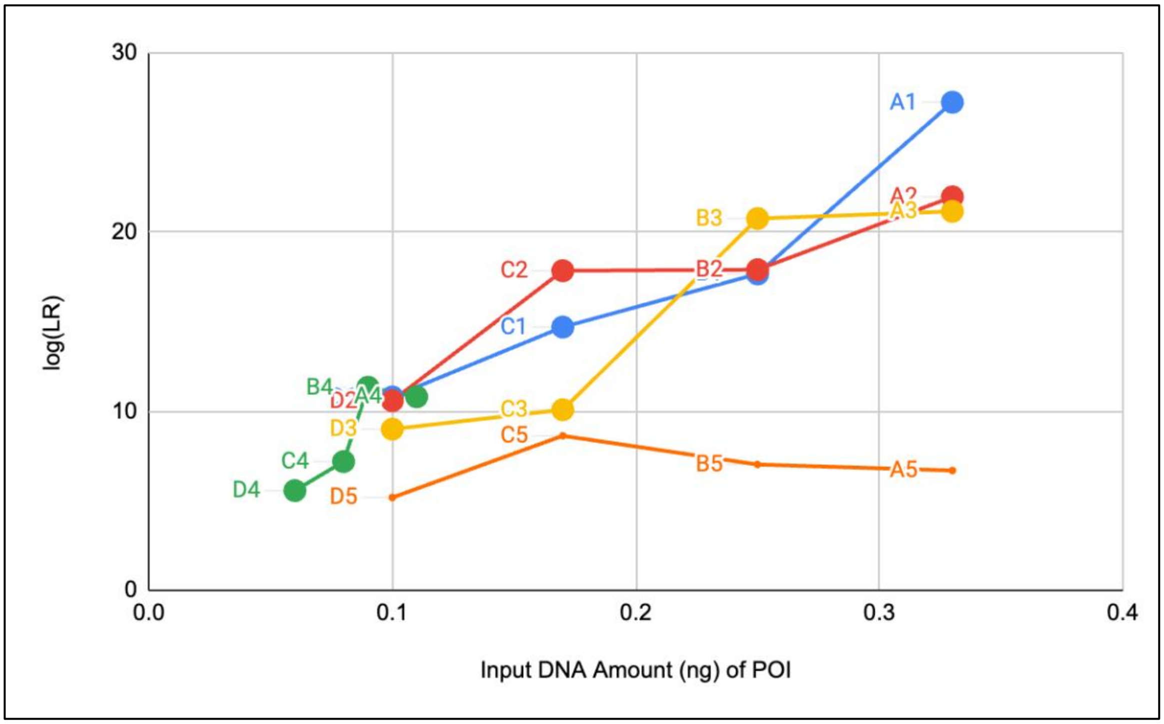


Figure 18. LR calculations in Test 2 where male B was the POI. Small data points represent the POI was a degraded contributor, and large data points represent the POI contributed undegraded DNA.

Figure 19 represents three selected loci each from mixtures C1 and C5. Allele sharing coupled with reduced peak heights due to degradation in mixture C5 makes manual interpretation of these mixtures difficult. However, Mixture Solution was able to assign LRs to each POI with “very strong support” in both tests. Peak height ratios of the alleles at these loci between the two mixtures remained relatively constant, as well as in the entire profile, and no dropout occurred at these loci in mixture C5.

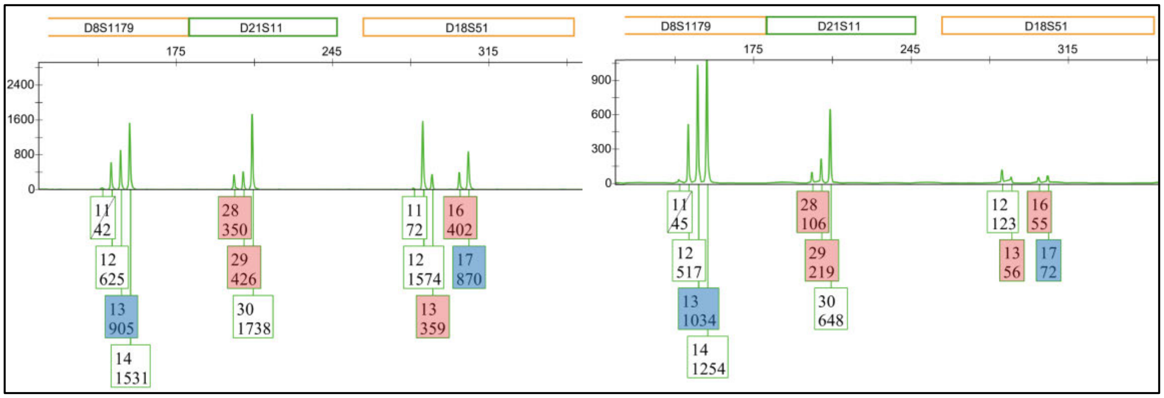


Figure 19. Three selected loci from mixtures C1 and C5. Mixture C1 is depicted on the left side of the figure, and Mixture C5 is depicted on the right side of the figure. Alleles highlighted in blue represent alleles unique to the female. Alleles highlighted in red represent alleles unique to male A. Allele calls with a strikethrough were determined to be stutter. Note that Y-axis scales differ for each mixture profile.

Figures 20 and 21 show changes in the log(LR) for each of the three contributors as POI (36). Figure 20 compares the LRs that were computed using the hypotheses defined for Test 1, and Figure 21 compares the LRs that were computed using the Test 2 hypotheses. An overall comparison of all LR values affirms that in most cases, the LRs assigned to a POI who contributed undegraded DNA to the mixture were greater than the POIs who contributed fully degraded DNA.

The data in Figure 21 representing Test 2 follows a similar distribution to the data observed in Figure 19, with some differences. In Test 2, there appears to be more log(LR) values greater than 20 with approximately 50% of LR values exceeding that level. In most cases, the LR values assigned to each of the POIs followed expected trends as they related to the input amount of DNA, level of degradation, effects of allele sharing, dropout, and estimated contributor proportions.

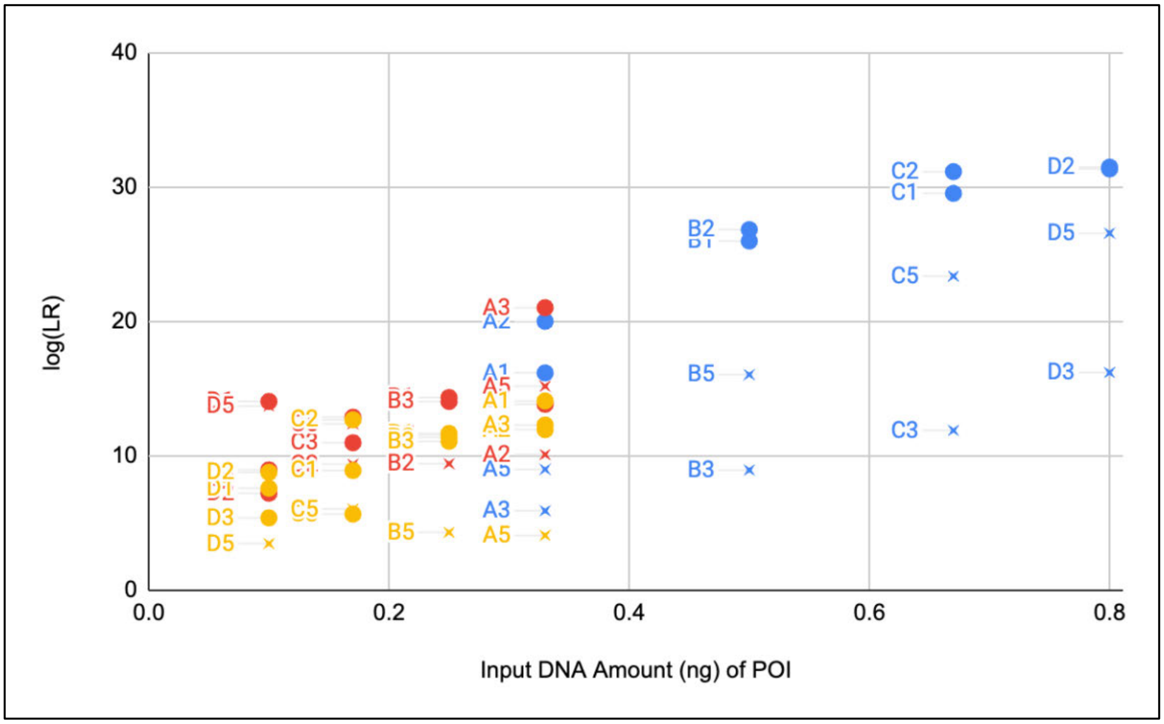


Figure 20. Plot of log(LR) for each mixture and each POI in Test 1. Some overlap occurs but each column of points represents a particular input DNA amount. Circular data points represent undegraded DNA and X shaped data points represent degraded DNA. Blue data points signify the female contributor, red data points signify male A, and yellow data points signify male B.

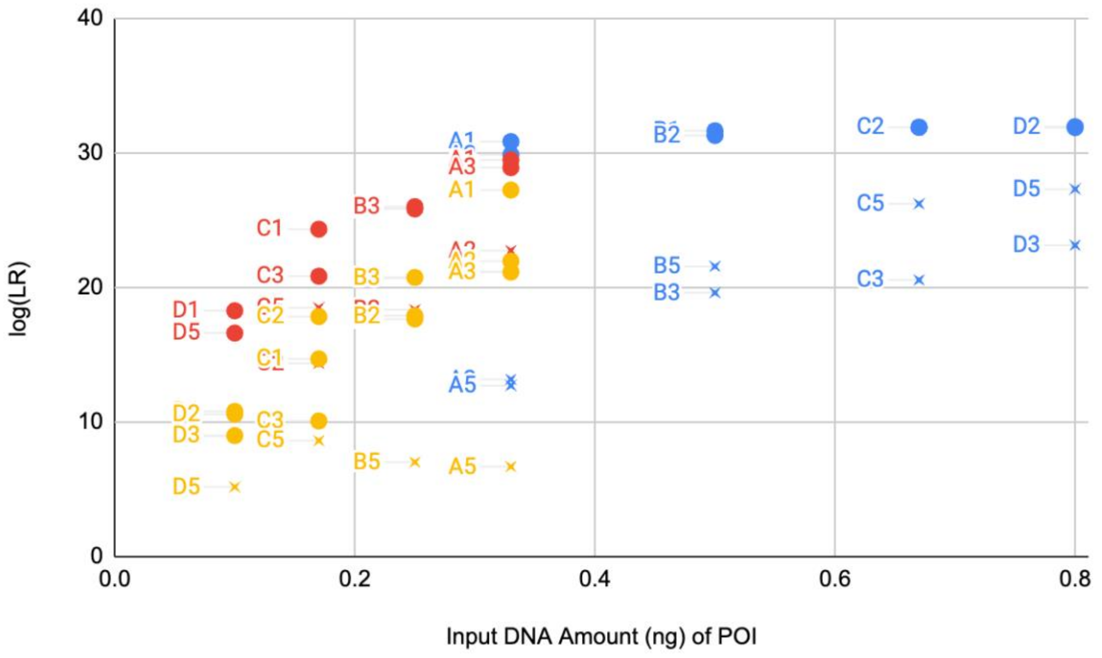


Figure 21. Plot of log(LR) for each mixture and each POI in Test 2. Some overlap occurs but each column of points represents a particular input DNA amount. Circular data points represent undegraded DNA and X shaped data points represent degraded DNA. Blue data points signify the female contributor, red data points signify male A, and yellow data points signify male B.

3.5 Number of Contributors

In order for Mixture Solution to assign an LR to a set of hypotheses, it must first determine the most favorable hypothesis for both the prosecution and defense prior to comparing the individual likelihoods. These favored hypotheses are a direct reflection of the NoC that Mixture Solution suggests for each DNA mixture. Table 7 summarizes the most favorable H_p that was determined by Mixture Solution, to compare with the most favorable H_d to subsequently assign an LR value.

In Test 2, where the remaining known contributors were included in hypothesis comparison along with the POI, Mixture Solution was able to correctly assign a NoC of three in all samples. However, in Test 1, there were some instances where Mixture Solution incorrectly assigned a NoC of two, which was not consistent with the true NoC. This was apparent in some or all of the computations for twelve out of the 20 mixtures analyzed by Mixture Solution in Test 1 where only the POI was defined as a potential contributor.

These NoC values are assigned based on the “cost” value that Mixture Solution calculates. In Mixture Solution, “cost” is defined as the negative logarithm of the likelihood. Therefore, the smallest cost value represents the best estimate of the NoC. Since cost is on a logarithmic scale, a cost difference of 1 and 2 represents an LR difference of 10 and 100, respectively, and so on. A cost will be calculated for each hypothesis, unless the calculation performed with one fewer number of unknown contributors produces a worse cost value, in which case, the program will assign a cost of “worse”. The “worse” cost indicates that the calculation for that hypothesis was skipped to save time if the program’s maximum likelihood proportion analysis encounters an unknown contributor with an estimated contribution of 0% (37). If a cost difference for the hypothesis with one fewer unknown contributor is less than one, a numerical value will be reported, but if it is greater than one, the cost will be reported as “worse”.

Profile data, including peak height balance and dropout, was manually analyzed in the twelve mixtures that were assigned an incorrect NoC one or more times. Using the MAC method, eight of these twelve mixtures would be considered a two-person mixture due to the dropout that occurs at loci where more than four alleles were expected to be observed. In all of these mixtures, Mixture Solution assigned a cost of zero to the hypothesis that the NoC was two, and assigned a “worse” cost to the hypothesis that the

NoC was three. This indicates that the log(LR) value was decreased by greater than one, asserting the low likelihood that the NoC of these mixtures was three.

For the remaining four mixtures that were incorrectly assigned a NoC of two, profile data was sufficient to produce a NoC of three using the MAC method since more than four alleles were present at one or more loci. Mixture Solution did not assign a “worse” cost for the three-contributor hypothesis for any of these four mixtures; rather, a numerical value less than one. This indicates that the log(LR) values for the two- and three-contributor hypotheses were within one, signifying a much smaller statistical difference between the two LRs, and that an estimate of three-person mixture is highly plausible, just slightly less likely than a NoC of two.

One of the two mixtures that Mixture Solution suggested a NoC of two and assigned a “worse” cost was Mixture A4. Mixture A4 has a contributor ratio of 1:7.5:1 where the female contributor contributed fully degraded DNA, male A contributed partially degraded DNA, and male B contributed undegraded DNA. Figure 22 is the red dye channel from the mixture profile as viewed on GeneMapper ID-X™ v1.6. In this dye channel, the degraded DNA of the female contributor in Mixture A4 exhibited dropout at three loci. The 12 allele at the D13S317 locus, the 8 allele at the D7S820 locus, and both 16 and 18 alleles at the SE33 locus drop out of the profile. Neither male A nor male B exhibited any dropout in the red dye channel. All four of the female contributor’s alleles that dropped out of the red dye channel were unique to the female contributor.

Mixture Solution has an option which users can enable to generate visual aids of the results of the LR computation, explaining potential deconvolution where applicable through colored bar graphs representing contributors overlayed over an illustration of peaks as if they were being viewed on the electropherogram. Figure 23 is the red dye channel portion of the visual aid generated by Mixture Solution for Mixture A4. A user can compare the visual aid with the electropherogram in order to assist in the deconvolution and make assumptions of the NoC to the mixture. Colored bars will appear behind the peaks to represent the supplied genotypes of the reference contributors included in the hypothesis. An absence of bars represents estimated contribution from an unknown person.

Table 7. Most favorable H_p for each contributor in Tests 1 and 2. The designation "u" represents an unknown individual (e.g., 2u is two unknown individuals). Hypotheses for Test 1 shaded in red represent an incorrect hypothesis predicted by DNA-View® Mixture Solution™.

	Test 1			**Test 2**		
Mix. ID	**Female**	**Male A**	**Male B**	**Female**	**Male A**	**Male B**
A1	F + 2u	M_A + 2u	M_B + 2u	F + M_A + M_B	F + M_A + M_B	F + M_A + M_B
A2	F + 2u	M_A + 2u	M_B + 2u	F + M_A + M_B	F + M_A + M_B	F + M_A + M_B
A3	F + 2u	M_A + 2u	M_B + 2u	F + M_A + M_B	F + M_A + M_B	F + M_A + M_B
A4	F + 1u	M_A + 1u	M_B + 1u	F + M_A + M_B	F + M_A + M_B	F + M_A + M_B
A5	F + 1u	M_A + 1u	M_B + 2u	F + M_A + M_B	F + M_A + M_B	F + M_A + M_B
B1	F + 1u	M_A + 2u	M_B + 2u	F + M_A + M_B	F + M_A + M_B	F + M_A + M_B
B2	F + 2u	M_A + 2u	M_B + 2u	F + M_A + M_B	F + M_A + M_B	F + M_A + M_B
B3	F + 2u	M_A + 2u	M_B + 2u	F + M_A + M_B	F + M_A + M_B	F + M_A + M_B
B4	F + 1u	M_A + 1u	M_B + 1u	F + M_A + M_B	F + M_A + M_B	F + M_A + M_B
B5	F + 1u	M_A + 2u	M_B + 2u	F + M_A + M_B	F + M_A + M_B	F + M_A + M_B
C1	F + 1u	M_A + 1u	M_B + 2u	F + M_A + M_B	F + M_A + M_B	F + M_A + M_B
C2	F + 1u	M_A + 2u	M_B + 2u	F + M_A + M_B	F + M_A + M_B	F + M_A + M_B
C3	F + 2u	M_A + 2u	M_B + 2u	F + M_A + M_B	F + M_A + M_B	F + M_A + M_B
C4	F + 1u	M_A + 1u	M_B + 2u	F + M_A + M_B	F + M_A + M_B	F + M_A + M_B
C5	F + 2u	M_A + 2u	M_B + 2u	F + M_A + M_B	F + M_A + M_B	F + M_A + M_B
D1	F + 2u	M_A + 1u	M_B + 2u	F + M_A + M_B	F + M_A + M_B	F + M_A + M_B
D2	F + 1u	M_A + 2u	M_B + 2u	F + M_A + M_B	F + M_A + M_B	F + M_A + M_B
D3	F + 2u	M_A + 2u	M_B + 2u	F + M_A + M_B	F + M_A + M_B	F + M_A + M_B
D4	F + 1u	M_A + 1u	M_B + 2u	F + M_A + M_B	F + M_A + M_B	F + M_A + M_B
D5	F + 1u	M_A + 1u	M_B + 2u	F + M_A + M_B	F + M_A + M_B	F + M_A + M_B

The only artifactual peak in the red dye channel of Mixture A4 is the 13 peak at the D221045 locus; all other alleles observed are true alleles. With near total dropout of the female contributor from this channel, Mixture Solution was not able to consider the female contributor as a possible contributor to the mixture. The only allele unique to the female contributor that was observed in this channel was an 11 allele at the D22S1045 locus. Therefore, it is highly unlikely, given all of these factors, that Mixture Solution would consider the female contributor a true contributor to Mixture A4 as a second unknown without being explicitly included in the proposed hypothesis. This is also true of the other contributors being defined as the POI for this mixture, where there is not enough genetic information present to suggest that more than two contributors are present in the mixture, which is a disadvantage of degradation to DNA profile analysis.

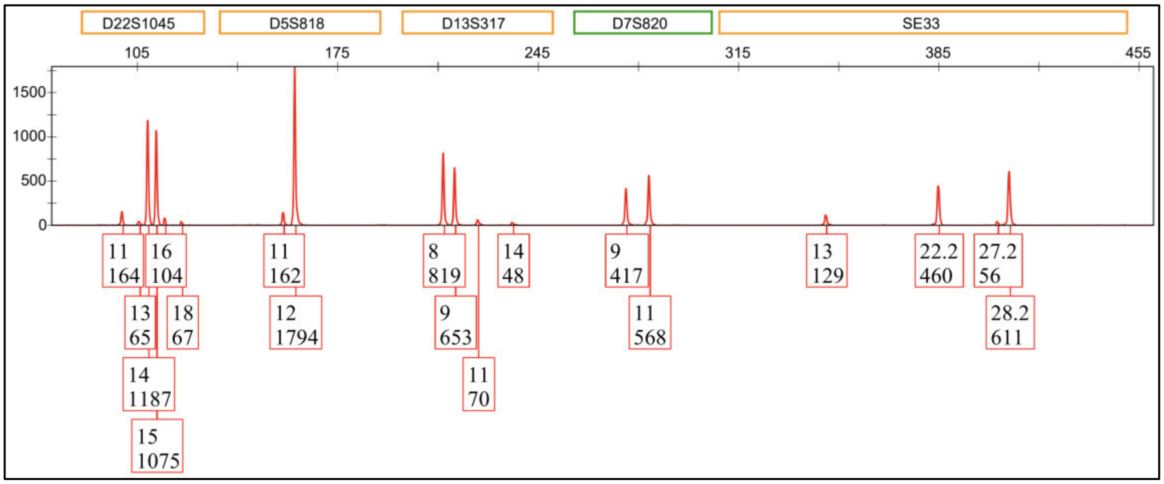


Figure 22. Red dye channel of Mixture A4 for which an incorrect NoC was assigned by Mixture Solution in all cases in Test 1.

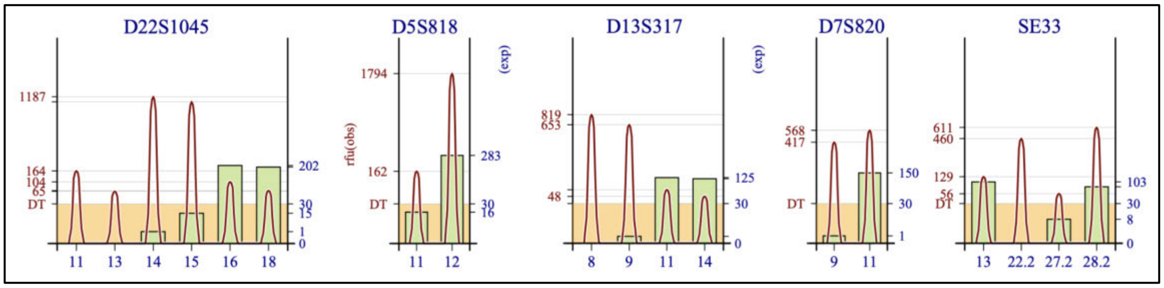


Figure 23. Visual aid of a Test 1 computation of Mixture A4 generated by Mixture Solution when male B was defined as the POI. Green bars represent predicted contribution of male B.

4. CONCLUSIONS

When DNA-View® Mixture Solution™ was used to assign LRs for mixtures containing DNA that was subjected to degradation, anticipated trends were present in the data from the computations. When the DNA of a POI was subject to degradation, decreases in the LR values were observed when compared to the LR values for undegraded DNA from the same POI. However, in all mixtures, Mixture Solution was able to assign an LR with "moderate support" or higher to the POIs in both tests, whether or not other contributors were included as knowns in the H_p and H_d and regardless of the level of degradation.

In some cases, increased DNA input negated the effects of degradation on the overall computation of the LR. Generally, as the template of DNA increased, LR values increased, irrespective of DNA quality. However, allele sharing and dropout had strong effects on the LR values in all mixtures with or without DNA degradation. Although degradation can increase the probability for allelic or locus dropout, the presence of more unique alleles can lead to a higher assigned LR for a contributor whose DNA may have been subjected to degradation.

Mixture Solution successfully generated appropriate LRs for all 20 mixtures tested in this research project, with no Type I errors that falsely excluded a known contributor from a mixture with an LR less than one. Even for the fully degraded samples at the most disproportionate ratio of contributors and low-template mass of approximately 0.1 ng, Mixture Solution was able to assign an LR to each contributor that confirmed their presence in the mixture. The user-friendly interface and the ability to perform computations rapidly would be important to any user. Results from analysis of the 20 three-person mixtures with and without degradation of one or two of the three contributors demonstrates that Mixture Solution computes likelihood ratios that are consistent with the observed relative changes in the DNA profiles of the 20 three-contributor mixtures.

LIST OF JOURNAL ABBREVIATIONS

Aust J Forensic Sci	Australian Journal of Forensic Sciences
Forensic Sci Int	Forensic Science International
Forensic Sci Int Genet	Forensic Science International. Genetics.
Int J Legal Med	International Journal of Legal Medicine
J Forensic Sci	Journal of Forensic Sciences
J Pathol Inform	Journal of Pathology Informatics
Sci Justice	Science and Justice

BIBLIOGRAPHY

1. SWGDAM interpretation guidelines for autosomal STR typing by forensic DNA testing laboratories.

2. Haned H, Pène L, Lobry JR, Dufour AB, Pontier D. Estimating the number of contributors to forensic DNA mixtures: does maximum likelihood perform better than maximum allele count? J Forensic Sci. 2011;56(1):23–8. doi: 10.1111/j.1556-4029.2010.01550.x.

3. Paoletti DR, Doom TE, Krane CM, Raymer ML, Krane DE. Empirical analysis of the STR profiles resulting from conceptual mixtures. J Forensic Sci. 2005;50(6):1361–6.

4. Lynch PC, Cotton RW. Determination of the possible number of genotypes which can contribute to DNA mixtures: Non-computer assisted deconvolution should not be attempted for greater than two person mixtures. Forensic Sci Int Genet. 2018;37:235–40. doi: 10.1016/j.fsigen.2018.09.002.

5. Coble MD, Bright J-A. Probabilistic genotyping software: An overview. Forensic Sci Int Genet. 2019;38:219–24. doi: 10.1016/j.fsigen.2018.11.009.

6. Moore D, Clayton T, Thomson J. A comprehensive study of allele drop-in over an extended period of time. Forensic Sci Int Genet. 2020;48:102332. doi: 10.1016/j.fsigen.2020.102332.

7. Butler J. Advanced topics in forensic dna typing: interpretation. Boston, MA: Elsevier, 2014.

8. D’Angelo O, Vandepoele ACW, Adelman J, Marciano MA. Assessing non-LUS stutter in DNA sequence data. Forensic Sci Int Genet. 2022;59:102706. doi: 10.1016/j.fsigen.2022.102706.

9. Buckleton JS, Bright J-A, Cheng K, Budowle B, Coble MD. NIST interlaboratory studies involving DNA mixtures (MIX13): A modern analysis. Forensic Sci Int Genet. 2018;37:172–9. doi: 10.1016/j.fsigen.2018.08.014.

10. Gill P, Jeffreys AJ, Werrett DJ. Forensic application of DNA ‘fingerprints.’ Nature. 1985;318(6046):577–9. doi: 10.1038/318577a0.

11. SWGDAM guidelines for the validation of probabilistic genotyping systems.

12. Forensic science regulator guidance software validation for DNA mixture interpretation. FSR-G-223 issue 2.

13. Kruijver M, Bright J-A. A tool for simulating single source and mixed DNA profiles. Forensic Sci Int Genet. 2022;60:102746. doi: 10.1016/j.fsigen.2022.102746.

14. Ward D, Henry J, Taylor D. Analysis of mixed DNA profiles from the RapidHIT^{TM} ID platform using probabilistic genotyping software STRmix^{TM}. Forensic Sci Int Genet. 2022;58:102664. doi: 10.1016/j.fsigen.2022.102664.

15. Stokes NA, Stanciu CE, Brocato ER, Ehrhardt CJ, Greenspoon SA. Simplification of complex DNA profiles using front end cell separation and probabilistic modeling. Forensic Sci Int Genet. 2018;36:205–12. doi: 10.1016/j.fsigen.2018.07.004.

16. Bleka Ø, Eduardoff M, Santos C, Phillips C, Parson W, Gill P. Open source software EuroForMix can be used to analyse complex SNP mixtures. Forensic Sci Int Genet. 2017;31:105–10. doi: 10.1016/j.fsigen.2017.08.001.

17. John Buckleton. https://johnbuckleton.wordpress.com/strmix/ Accessed 7 Mar 2023.

18. Gill P, Brenner CH, Buckleton JS, Carracedo A, Krawczak M, Mayr WR, et al. DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. Forensic Sci Int. 2006;160(2–3):90–101. doi: 10.1016/j.forsciint.2006.04.009.

19. Haned H. Likelihood ratios in practice. Netherlands Forensic Institute.

20. Lien E. A very brief introduction on the SWGDAM recommendations on reporting likelihood ratios.

21. Perlin MW. Inclusion probability for DNA mixtures is a subjective one-sided match statistic unrelated to identification information. J Pathol Inform. 2015;6(1):59. doi: 10.4103/2153-3539.168525.

22. Bille TW, Weitz SM, Coble MD, Buckleton J, Bright J-A. Comparison of the performance of different models for the interpretation of low level mixed DNA profiles: Nucleic acids. Electrophoresis. 2014;35(21–22):3125–33. doi: 10.1002/elps.201400110.

23. Bright J-A, Taylor D, Curran JM, Buckleton JS. Degradation of forensic DNA profiles. Aust J Forensic Sci. 2013;45(4):445–9. doi: 10.1080/00450618.2013.772235.

24. Hansson O, Egeland T, Gill P. Characterization of degradation and heterozygote balance by simulation of the forensic DNA analysis process. Int J Legal Med. 2017;131(2):303–17. doi: 10.1007/s00414-016-1453-x.

25. Ginart S, Caputo M, Corach D, Sala A. Human DNA degradation assessment and male DNA detection by quantitative-PCR followed by high-resolution melting analysis. Forensic Sci Int. 2019;295:1–7. doi: 10.1016/j.forsciint.2018.11.013.

26. Bender K. Preparation of degraded human DNA under controlled conditions. Forensic Sci Int. 2004;139(2–3):135–40. doi: 10.1016/j.forsciint.2003.10.003.

27. Chung DT, Drábek J, Opel KL, Butler JM, McCord BR. A study on the effects of degradation and template concentration on the amplification efficiency of the STR miniplex primer sets. J Forensic Sci. 2004;49(4):1–8. doi: 10.1520/JFS2003269.

28. Nicklas JA, Noreault-Conti T, Buel E. Development of a real-time method to detect DNA degradation in forensic samples. J Forensic Sci. 2012;57(2):466–71. doi: 10.1111/j.1556-4029.2011.02001.x.

29. Lucassen A, Ehlers K, Grobler PJ, Brenner CH. Evaluating Mixture Solution™—rapid and non-MCMC probabilistic mixture analysis. Int J Legal Med. 2021;135(6):2275–84. doi: 10.1007/s00414-021-02680-2.

30. Greenspoon SA, Schiermeier-Wood L, Jenkins BC. Establishing the limits of TrueAllele® casework: a validation study. J Forensic Sci. 2015;60(5):1263–76. doi: 10.1111/1556-4029.12810.

31. Brenner CH. Mixture Solution™ overview.

32. Brenner CH. There’s DNA everywhere: an opportunity for APL.

33. Sharma V, van der Plaat DA, Liu Y, Wurmbach E. Analyzing degraded DNA and challenging samples using the ForenSeq™ DNA Signature Prep kit. Sci Justice. 2020;60(3):243–52. doi: 10.1016/j.scijus.2019.11.004.

34. Holmes AS, Houston R, Elwick K, Gangitano D, Hughes-Stamm S. Evaluation of four commercial quantitative real-time PCR kits with inhibited and degraded samples. Int J Legal Med. 2018;132(3):691–701. doi: 10.1007/s00414-017-1745-9.

35. Holland MM, Tiedge TM, Bender AJ, Gaston-Sanchez SA, McElhoe JA. MaSTR™: an effective probabilistic genotyping tool for interpretation of STR mixtures associated with differentially degraded DNA. Int J Legal Med. 2022;136(2):433–46. doi: 10.1007/s00414-021-02771-0.

36. Duke KR, Myers SP. Systematic evaluation of STRmix™ performance on degraded DNA profile data. Forensic Sci Int Genet. 2020;44:102174. doi: 10.1016/j.fsigen.2019.102174.

37. Brenner CH. Compact summary of hypotheses. Mixture Solution™ help file.

CURRICULUM VITAE

[redacted]

[redacted]