Federal Award Number: 15PNIJ-21-GG-02710-SLFO

Project Title: Implementation and Use of Continuous Probabilistic Genotyping Software to Improve the Interpretation of Forensic DNA Mixtures

Principal Investigator: Michael Coble, Professor, Center for Human Identification, University of North Texas Health Science Center at Fort Worth, Center for Human Identification, Fort Worth, TX 76107
Tel: 817-735-0228; email: michael.coble@unthsc.edu

Co-Principal Investigator: Tim Kalafut, Assistant Professor,
Department of Forensic Science
College of Criminal Justice
Sam Houston State University
Chemistry & Forensic Science Building, Rm. 221B
Huntsville, TX 77340
Tel: 936-294-2536 (o) email: tim.kalafut@shsu.edu

Recipient Organization: University of North Texas Health Science Center

3500 Camp Bowie Blvd., Fort Worth, TX 76107

Project/Grant Period: 01-01-2022 - 12-31-2024

Award Amount: $377,311

Summary of the project

Major goals and objectives

Forensic DNA laboratories in the United States have undergone a transformational shift in how they interpret and report DNA mixtures. Starting in 2010, the Scientific Working Group on DNA Analysis Methods (SWGDAM) recommended the use of a stochastic threshold to establish a stochastic threshold (e.g. 150 rfu) where allelic drop out was possible when peaks below this threshold were observed in a DNA profile [1]. For the most part, laboratories would present the statistical assessment of the evidence profile using a modified Random Match Probability (mRMP) or the Combined Probability of Inclusion (CPI). This approach is often referred to as the "Binary" method of DNA mixture interpretation, where all acceptable genotypes (RMP) or alleles (CPI) have equivalent impact (equal weights) on the calculation.

Over the last decade, probabilistic methods of complex DNA mixture interpretation have been embraced and adopted by the forensic DNA community. Probabilistic Genotyping (PG) Systems use molecular modeling and statistical algorithms to provide weight to genotype combinations to best explain the evidence. In short, better fitting genotypes have a higher weight, affecting the assigned likelihood ratio (LR) than poorer fitting options. Unlike the binary method of interpretation, PG does not use a stochastic threshold. Rather, the software models and evaluates the evidence probabilistically. The statistical strength of the evidence is presented as a Likelihood Ratio (LR), which is a component of Bayes' Theorem [2].

In the move from binary to probabilistic methods of interpretation, forensic DNA analysts have had to become familiar with the use of the LR which is not a probability in the traditional sense of a mRMP or CPI. Instead, the LR is a ratio of two conditional probabilities

which represents the strength (or weight) of the evidence given two competing hypotheses (e.g. the Person of Interest (POI) is a contributor to the evidence profile) versus an alternative hypothesis (e.g. the POI is not a contributor to the evidence profile.

One PG software is STRmix (ESR, New Zealand) and is by far the most widely used software among forensic DNA laboratories in the U.S. We have found that many laboratories in the U.S. seem hesitant to report low LRs, even though they may have reported RMP or CPI statistics of similar magnitude. Many of these laboratories have developed an LR reporting threshold where LRs that fall in the range of 1/1000 (0.001) to 1000 are reported as "inconclusive." We have observed this range cover LRs from 1/10,000 (0.0001) to 10,000.

The major goals of this project aimed to expand the scientific knowledge for analyzing complex DNA mixtures and address current challenges faced by the analyst in court. We wanted to research and evaluate methods to help increase the confidence needed for analysts to report low LRs from complex and challenging DNA mixtures. We focused on the PG software STRmix, but feel the same concepts can be applied to other continuous PG software programs that generate weighted genotype probabilities as well. Two software programs, Advent*LR* [3] and DBLR [4], present a visual analysis of non-contributor testing which can assist the analyst to convey the strength of low LRs to the trier of fact.

Research questions

The research questions posed by this project fall into two main categories. The first objective is evaluating the use of two "database" software programs to assist the analyst in the reporting of low LRs to the trier of fact without the need to use an inconclusive zone. The questions we wish to address with the first objective was to test the two database software programs with challenged samples to evaluate their use in forensic practice. Once the software programs were evaluated, we next wanted to conduct an interlaboratory study where laboratories, with no prior experience in either software, are given a set of mixtures and asked to report the outcome of the complex mixture using their standard SOP and reporting language. After training for both software programs, the laboratories were asked to interpret a different set of similar mixtures and use both database software programs to report their results. We also monitored consistency within and between laboratories when reporting low LRs to see if there was an improvement in reporting after using the database software programs. The first objective has three specific goals:

Goal #1: Evaluation of two software programs, DBLR and Advent*LR*, for use in forensic practice.

Goal #2: Improve consistency and reporting within a laboratory for mixture interpretation.

Goal #3: Improve the analysts' ability to convey the reliability of low LRs to the end-user of DNA results (law enforcement, lawyers, judges, and juries).

The second objective of this study was to test a proof of concept for searching a DNA database (e.g. CODIS) without the need to summarize results. The challenge we wished to address is that current recommendations for searching an evidentiary profile deconvolved by STRmix are that

alleles at a particular locus must be observed in greater than 99% of the weighted genotypes. Technically, if an allele is observed in only 98% of the weighted genotype combinations, then it would be ineligible for searching in CODIS. There are a number of freely available spreadsheet programs (e.g. COSTaR, Brian Burritt) that can summarize the STRmix output and automate a report that can be uploaded to CODIS for searching. To accomplish this goal, we will use a set of mixtures to generate a CODIS input file based on the criteria of the current rules for CODIS using a 99% weight for alleles from a STRmix deconvolution. We will compare the effectiveness of this method compared to a direct search of database of mock CODIS offenders in a probabilistic fashion using the weights generated from STRmix without the necessity to summarize the data. The second objective has one specific goal:

Goal #4: Test a proof of concept for searching a DNA database directly with genotype weights, and with no need to summarize results.

Research design, methods, analytical and data analysis techniques

For Objective 1, Goal 1, we used the GlobalFiler (Thermo Fisher) samples from the PROVEDIt database [5] analyzed on the 3500 platform with 29 PCR cycles and a 15 second injection. Single source samples along with 2-person, 3-person, and 4-person mixtures were analyzed with ArmedXpert v.3.1.14 running an embedded copy of OSIRIS v2.16. Since the focus of this study was to examine DNA profiles with “low” Likelihood Ratios (LRs), we selected a set of 163 profiles with LRs ranging from 0.07 to 9216 (Table 1).

Group	Samples	LR Range
1p	65	102 - 9216
2p	24	5 - 8593
3p	36	0.36 - 6131
4p	38	0.07 - 1282
TOTAL	163	0.07 - 9216

Table 1. Number of PROVEDIt samples and LR range for each group analyzed in this study.

The STRmix file created from ArmedXpert was analyzed with STRmix v.2.9(.1) [6]. The STRmix deconvolution file was uploaded into AdventLR v2.9 [3] or DBLR v1.2.24 [4] for an analysis of 10,000 and 100,000 simulations, respectively. DBLR generates both H1 true and H2 true contributor distributions. AdventLR generates an H2 true contributor distribution. AdventLR uses a static database, while DBLR generates a new database of non-contributor profiles for each comparison. All LRs are point source LRs using the NIST Caucasian population, FST=0.01.

For Objective 2, Goal 1, we tested a proof of concept to potentially improve the ability to identify investigative leads using DBLR compared to CODIS. Currently laboratories consider any allele(s) with a weight greater than 99% probability to be eligible for upload to CODIS. Programs such as COSTaR (Brian Burritt) can automate the export of a CODIS file to search the CODIS database. Alleles with weights less than 99% are deemed ineligible for searching in CODIS.

For this goal, we first developed a database of 25 million random DNA profiles to mimic the current size of CODIS. The mixtures used for this study were taken from the study of

Boodoosingh et al. [6] which consisted of 155 mixtures from 81 contributors resulting in 6,120 deconvolutions and a total of 9,755 contributor slots altogether through STRmix. We added the 81 contributors to the 25 million random contributors in the "offender database."

Expected applicability of the research

In the last decade, forensic DNA laboratories have rapidly adopted Probabilistic Genotyping (PG). The statistic reported from a mixture analyzed with PG software is the Likelihood Ratio. Prior to the introduction of PG, LRs were used in about 10-15% of forensic laboratories in the United States. We have provided trainings to new users of PG, and have observed that there is sometimes a hesitancy to report low LRs over the concern about the "reliability" of the statistic for a given contributor. This is really better described as a question about how to report LRs of low magnitude. While PG methods in general, and STRmix in particular, have shown improvement in statistics reported both within and between laboratories, that is not true for LRs that are small, say between 1 and 10,000.

In general, there are two schools of thought on this issue. The first is to simply report the calculated LR, as anything else is now putting a threshold into a fully continuous system. With this approach, the traditional statements of "inclusion/exclusion" are avoided by the expert. Advocates of this approach claim that this is a 'pure' form of the LR, and whether the LR supports the POI as the source of the DNA profile, or an unknown person as the source (LR<1), the jury is the ultimate decision maker and should get the data "as is".

Critics of this approach point out this may be doing a disservice to a jury of laypersons, as the role of an expert witness is to help explain difficult concepts. Perhaps this is an inference that experts themselves do not know what to do either, and is "passing the buck" to the jury in order to avoid an uncomfortable examination on the witness stand.

We note that some laboratories attempt to sidestep this issue altogether by using an "inconclusive range" where LRs between, for example, 0.001 and 1,000 are determined to be "inconclusive" or "uninformative." We have seen laboratory protocols that parse the issue based on the number of contributors. For example, LRs between 0.001 and 1,000 are considered inconclusive for partial single source profiles or two-person mixtures, but for three- or four-persons the range is from 0.0001 to 10,000. Typically, laboratories base these thresholds on non-contributor testing, with the range defined by the largest non-contributor LR seen during their validation studies. Critics of this inconclusive range approach point out that such ranges are arbitrary, and that this approach is heavily dependent on the size of the non-contributor database used in validation testing. Turing's rule says if you test a database of about 1 million persons, an LR of 1 million for a non-donor is not unexpected [7]. In addition, these pre-determined thresholds have nothing to do with the DNA mixture quantity and quality of the sample in question.

We expect this research to assist the forensic scientist to feel more comfortable with reporting low LRs using non-contributor testing software like Advent*LR* and DBLR without the need for using an inconclusive range.

We also expect the research of testing mixtures directly against a database of persons of interest to greatly increase investigative leads compared to the current rules based on allele weights of >99% at a given locus. Again, it is not feasible at the moment to replace CODIS with DBLR, but this proof-of-concept may be useful to show the power of directly searching a database against the mixture weights rather than a summary of the weights, and may be able to identify POIs from mixtures better than methods used today.

Participants and other collaborating organizations

Michael Coble (UNTHSC - CHI)

Tim Kalafut (SHSU)

Curt Schuerman, (USACIL)

Changes in approach from original design and reason for change, if applicable

Minor modifications were made to the project; the original proposal called for using the STRmix determined template as a diagnostic for whether a component of the mixture was eligible for comparison. We felt that this would add an extra layer of complexity and perhaps lose the focus of the interlaboratory study.

Outcomes

Activities/accomplishments

1. Both Advent*LR* and DBLR were shown to be fit for purpose when used for non-contributor testing.

2. The use of non-contributor testing improved reporting concordance across participating laboratories for LRs of small magnitudes.

3. The use of a probabilistic approach to search offender databases shows promise to generate more investigative leads than the current approach based on a 99% weight of alleles.

Results and findings

Objective 1, Goal 1

We compared the concordance between the two software programs (Table 2) for the 163 samples from the ProvedIt database for a total of 223 comparisons. The reason for the additional comparisons (223 rather than 163) is that in a number of the mixtures, more than one contributor could be used for a comparison (e.g. a 3-person mixture of a major:trace:trace could provide two trace contributors with low-level LRs. As a threshold, we required the LR for the POI to be above a 99.9 percentile LR generated from the non-contributor database.

We observed that both programs agreed an average of 92.8% of the time. Nearly one-half (7/16) of the disagreements occurred with single-source samples. We attribute this to the way the database search LRs are displayed to the user in Advent*LR* compared to DBLR. Advent*LR* allows the user to see if the LR of the POI is the same as the 99.9 percentile LR from the database. DBLR does not easily show this to the user. We made the decision to call a "tie" – meaning the POI LR is the same value as the 99.9 percentile LR – as suitable for reporting as inclusionary when Advent*LR* was used. It is unknown if the POI LR was the same as the 99.9 percentile LR when using DBLR, so the count of an "inclusion" is different between the two tools in this specific circumstance. The greatest disagreement was observed in single-source samples

where DBLR gave a 99.715 percentile and Advent*LR* gave a 99.960 percentile. This is the equivalent of the POI LR in DBLR as the 12th largest LR from the database and (tied for) the 10th highest LR in the Advent*LR* non-contributor database.

Group	Samples	Comparisons	Disagreements	% Agreement
1p	65	65	7	89.2
2p	24	29	3	89.7
3p	36	56	2	96.4
4p	38	73	4	94.5
TOTAL	**163**	**223**	**16**	**92.8**

Table 2. Number of PROVEDIt samples compared, disagreements, and % agreement between the 99th percentile from DBLR and AdventLR.

The differences in database sizes and static vs. dynamic non-contributor profiles appear to have only a minimal impact on the value of the 99.9%ile LR, and with the exception of the "tie" scenario, the differences are related to the profiles used in the Advent*LR* database (static) and DBLR (different profiles generated each time).

Objective 1, Goal 2

For Objective 1, Goal 2, we recruited 10 laboratories to participate in an interlaboratory study to see if the use of AdventLR or DBLR could help with reporting low LRs. Ten labs were able to provide results for this study. The laboratories participated in two rounds using the low contributor mixtures tested in Objective 1, Goal 1. We provided each lab with two sets of

mixtures (Set A and Set B) consisting of eight mixtures, and we asked that the sets be provided to different people with in the laboratories. We asked, and laboratories agreed, for participants to not look at the other set of mixtures not assigned for the first round. For example, in round one Analyst A tested the eight mixtures in Set A, while Analyst B tested the eight mixtures in Set B. We asked for the analysts to report exactly as their protocol says, and use the reporting statements, LR, verbal scale, etc…

For round two, each analyst received training for AdventLR and DBLR. The analysts then analyzed the opposite sets (e.g. Analyst A tests Set B, while Analyst B tests Set A). This time, the analysts were asked to create a hybrid reporting statement (using their normal reporting language, but now report based on two software tools used in this study).

Table 3 shows the reporting of the low LRs using their standard protocol:

Sample Set	Case #	Internal Protocol				
		Unsuitable	Exclusionary	Inc	Uninformative	Inclusionary
A	C1	8		1		3
A	C2	6				6
A	C3	1		1		10
A	C4		1		1	10
A	C5	1		1		10
A	C6	2	1	1		8
A	C7	4	1	1		6
A	C8	3	2	1		6
B	C1	8				5
B	C2	8				5
B	C3	5				8
B	C4	1	1			11
B	C5	1		2		10
B	C6	2	2	1	1	7
B	C7	5	1	1		6
B	C8	4	1	1	2	5

Table 3. Results from Round 1 of the interlaboratory study for reporting low LRs using the laboratory standard protocol. Results from the ten labs on samples are categorized as being reported as “unsuitable”, “exclusionary”, “Inconclusive” (Inc), “Uninformative”, or “Inclusionary”. Sets A and B represent the two different data sets with mock “case” samples 1-8.

The results of Table 3 show that reporting of LRs tend to be inconsistent between the laboratories. We noted very minor differences within laboratories, which may be a result of the normal STRmix variation when run multiple times, as each analyst ran each mixture in STRmix.

After the training in both software tools, we noted an improved consistency in reporting (Table 4)

Sample Set	Case #	Internal Protocol					AdventLR testing (10,000)				DBLR (100,000)			
		Unsuitable	Exclusion	Inc	Uninformative	Inclusion	Exclusion	Inc-NCT	Uninformative	Inclusion	Exclusion	Inc-NCT	Uninformative	Inclusion
A	C1	8		1		3		1		11		9		4
A	C2	6				6				12				13
A	C3	1		1		10				12				13
A	C4		1		1	10				11			1	12
A	C5	1		1		10				12				13
A	C6	2	1	1		8				12				13
A	C7	4	1	1		6	1			11				13
A	C8	3	2	1		6	1			11	1			12
B	C1	8				5		1		11		8		4
B	C2	8				5				12				12
B	C3	5				8				12				12
B	C4	1	1			11				12				12
B	C5	1		2		10				12				12
B	C6	2	2	1	1	7	1		1	10	1		1	10
B	C7	5	1	1		6				12				12
B	C8	4	1	1	2	5	1		3	8	1		3	8

Table 4. Results from Rounds 1 (Internal Protocol) and Round 2 (AdventLR testing and DBLR) of the interlaboratory study for reporting low LRs. A new category for AdventLR and DBLR, "Inc-NCT" was added to include inconclusive results from Non-Contributor Testing (NCT).

Objective 1, Goal 3

Table 4 shows greater consistency and improved reporting after applying Advent*LR* and DBLR testing. For Objective 1, Goal 3, in follow on discussions with the laboratories participating in the interlaboratory study, the analysts reported an improved comfort in reporting low LRs and the use of non-contributor testing to visualize these results.

Objective 2, Goal 1

Currently CODIS uses a strategy of searching at least 8 of original 13 core loci, and must meet a threshold (Moderate Match Estimate (MME)) of 10 million. The MME is only calculated on the original CODIS core 13 loci. We first generated a tool that would allow us to assess the 6,120 deconvolutions from STRmix as being eligible to upload to CODIS. If the alleles present in the contributor were at least 99% for at least 8 of the original 13 loci, and the MME of the at least 8/13 loci is 10 million, then these loci can be searched. For the mixtures used in this study only 37.6% of the donor slots qualify as "CODIS Eligible Contributors" (9,755 total contributor slots). (The total number of contributor slots across the 6120 deconvolutions is 9755.) These are the "donors" that would be eligible for upload to CODIS using this 99% weighting approach. Over 60% of the donors in the dataset would be ineligible for searching (Figure 1).

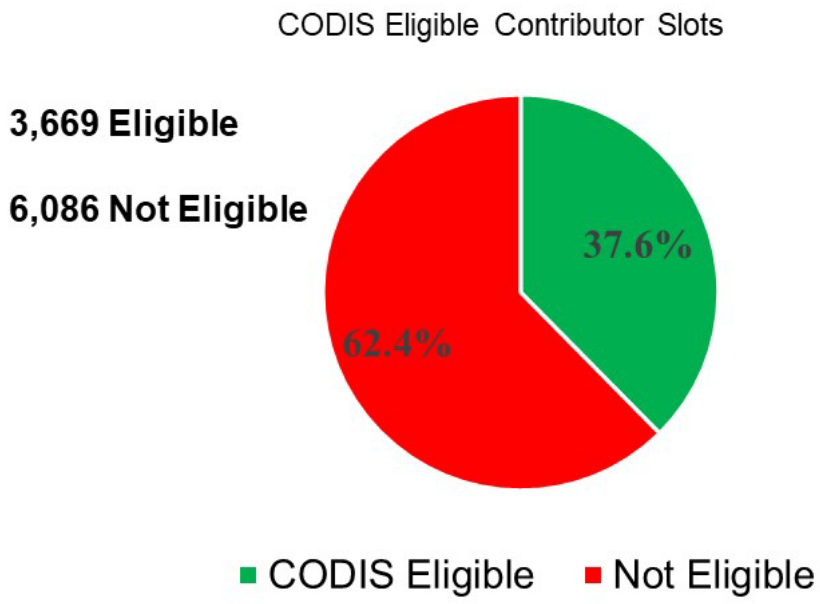


Figure 1. The total number of contributor slots that would qualify for CODIS eligibility, or not qualify, for the mixtures tested in this study.

For the second part of this proof of concept, we searched directly each individual of the mock offender database (>25 million individuals) against each contributor slot from the STRmix deconvolution. For example, a 4-person mixture would have 4 contributor slots that all 25 million+ offenders would be tested to generate an LR. We used a threshold of 10 million to be concordant with the MME threshold used by CODIS. In other words, if an "offender" in the database returned an LR of 10 million or greater, this would be noted as a "hit" or an association. LRs less than 10 million were ignored.

Using this "Direct Searching" approach (Figure 2), we found that 82.1% of true contributors could be identified out of the 9,775 available contributor slots. (The seeded ground truth donors returned LRs greater than 10 million.)

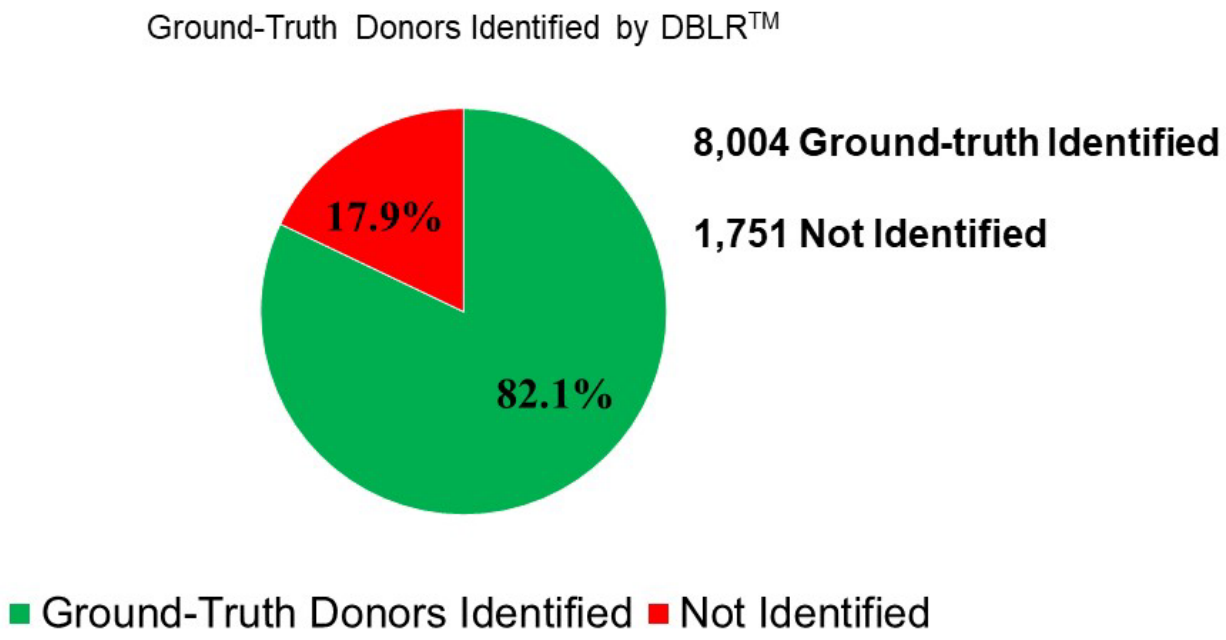


Figure 2. The total number of ground truth donors that could be identified using a direct searching method with the software DBLR for the mixtures tested in this study.

Searching a large database of over 25 million individuals with an LR threshold of 10 million, one would expect to observe non-contributor "fortuitous" associations. In fact, according to "Turing's Rule" [7], one should expect 25 million/10 million about 2.5 non-donor hits per search. Of all of the searches conducted, only 178 non-donors returned LRs greater than 10 million across more than 240 billion searches.

Limitations

It is not feasible at the moment to replace CODIS with DBLR at the national level, but this proof of concept may be useful to show the power of directly searching a database against the mixture weights rather than a summary of the weights, and may be able to identify POIs from mixtures better than methods used today. We envision that this could be easily implemented at the local (LDIS) or State (SDIS) level for now.

Artifacts

List of products (e.g., publications, conference papers, technologies, websites, databases), including locations of these products on the Internet or in other archives or databases

Posters

"An Exploration of Two Software Programs to Convey the Strength of Low Likelihood Ratios" was given at the 2023 International Symposium for Human Identification.

Oral presentations

Tim Kalafut, An Early Look at an Inter-Laboratory STRmix Concordance Study; Association of DNA Analysts and Administrators Summer Meeting, Houston, TX – July 2022

Tim Kalafut*, Michael Coble; A Picture Is Worth 10,000 LRs: The Evaluation and Implementation of Tools to Improve Interpretation and Reporting of Mixtures Using Probabilistic Genotyping, American Academy of Forensic Sciences 76th Annual Scientific Meeting, Denver, CO – February 2024

Masters student theses

Safia Boodoosingh, Evaluation of Inter-laboratory Probabilistic Genotyping Parameters for Improved Forensic DNA Database Searching and Interpretation of DNA Mixtures, Master Thesis 2022

Cesar Garza-Sanchez, Probabilistic tools for improved database searching and investigative leads, Master Thesis 2023

Publications

S. Boodoosingh, H. Kelly, J.M. Curran, T. Kalafut, An inter-laboratory comparison of probabilistic genotyping parameters and evaluation of performance on DNA mixtures from different laboratories, Forensic Sci. Int. Genet. 71 (2024) 103046. https://doi.org/10.1016/j.fsigen.2024.103046.

Three manuscripts are currently in progress.

Archived research data

NACJD

Dissemination activities

See the **Artifacts** section.

References

1. Scientific Working Group on DNA Analysis Methods. SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories2010.

2. Bright J-A, Coble M. Forensic DNA Profiling A Practical Guide to Assigning Likelihood Ratios. Boca Raton: CRC Press; 2020.

3. Schuerman C, Kalafut T, Buchanan C, Sutton J, Bright JA. Using the Nondonor Distribution to Improve Communication and Inform Decision Making for Low LRs from Minor Contributors in Mixed DNA Profiles. J Forensic Sci. 2020;65(4):1072-84.

4. Kelly H, Kerr Z, Cheng K, Kruijver M, Bright J-A. Developmental validation of a software implementation of a flexible framework for the assignment of likelihood ratios for forensic investigations. Forensic Sci Int Rep. 2021;4.

5. Alfonse LE, Garrett AD, Lun DS, Duffy KR, Grgicak CM. A large-scale dataset of single and mixed-source short tandem repeat profiles to inform human identification strategies: PROVEDIt. Forensic Science International: Genetics. 2018;32:62-70.

6. Boodoosingh S, Kelly H, Curran JM, Kalafut T. An inter-laboratory comparison of probabilistic genotyping parameters and evaluation of performance on DNA mixtures from different laboratories. Forensic Science International: Genetics. 2024;71:103046.

7. Taylor D, Buckleton J, Evett I. Testing likelihood ratios produced from complex DNA profiles. Forensic Science International: Genetics. 2015;16:165-71.

Appendix 1. Protocol for using the DBLR software to present potential contributor and non-contributor results with low Likelihood Ratios

Research and Evaluation of the Implementation and Use of Continuous Probabilistic Genotyping Software to Improve the Interpretation of Forensic DNA Mixtures

Federal Award Number: 15PNIJ-21-GG-02710-SLFO

CoPIs:

Michael Coble, PhD University of North Texas Health at Fort Worth

Tim Kalafut, PhD, Sam Houston State University

1. Introduction

Probabilistic Genotyping has extended the number of complex and low-level evidentiary samples submitted for interpretation compared to threshold-based "binary" methods [1]. All PG software systems utilize the Likelihood Ratio (LR) to report the strength or weight of the evidence using a Bayesian framework.

Although LRs were recommended as an appropriate statistic for DNA mixtures by the DNA Advisory Board in 2000 [2], most laboratories in the United States used modified Random Match Probability and the Combined Probability of Inclusion to report their statistical results. Butler et al. [3] estimated in 2018 that approximately 15% of the U.S. were reporting LRs.

As more laboratories brought LRs online through the validation and implementation of PG software, each new laboratory to LRs were faced with properly understanding LRs themselves, and explaining the meaning of an LR to non-scientists in the legal system (juries, judges, and lawyers). With uncertainty associated with reporting low LRs, some laboratories developed an "Inconclusive Zone" where LRs that fall within a range surrounding an LR of 1.0 were simply deemed to be "inconclusive." For example, any resulting LR from 0.001 (1/1,000) to 1,000 would be given an inconclusive finding. Statistical purists believe that the LR should be reported no matter the number – that at its heart, the LR is an evaluation of the evidence. To the Defense, an LR less than 1.0 means the evidence, when evaluated under two competitive hypotheses, favors an alternative POI and therefore is not "inconclusive."

This research evaluated the use of two software programs: DBLR [4] and Advent*LR* [5] to conduct non-contributor testing from the results of STRmix. These programs provide a visual of the non-contributor analysis that could be a part of the interpretation report, or even presented to a jury. This protocol is for using DBLR, which provides both H1 true potential contributors and H2 true non-contributors. Advent*LR* was developed by the Defense Forensic Science Center (DFSC) and is freely available [5]. DFSC has developed a protocol for Advent*LR*, and they are willing to share with forensic laboratories upon request.

The goal of this research was to present a protocol for using non-contributor testing to increase confidence in presenting low LRs from PG interpretations. If the non-contributor testing gave LRs greater than 99.9% of the LRs at the sub-source LR from STRmix, then the results should be reported, and the visual of this distribution could be used to explain the LR.

2. Scope

These procedures apply to DNA personnel who interpret nuclear STR profiles and perform interpretation and statistical analyses with STRmix.

3. Equipment

DNA profiles interpreted with an analysis software (e.g. GeneMapper IDX, Osiris, GeneMapper). Analysis of the profile with STRmix (ESR) and DBLR v.1.2 and above (ESR) are required. For this protocol, DBLR version 1.3.35 was used.

4. Procedure

4.1 Preliminary Analysis

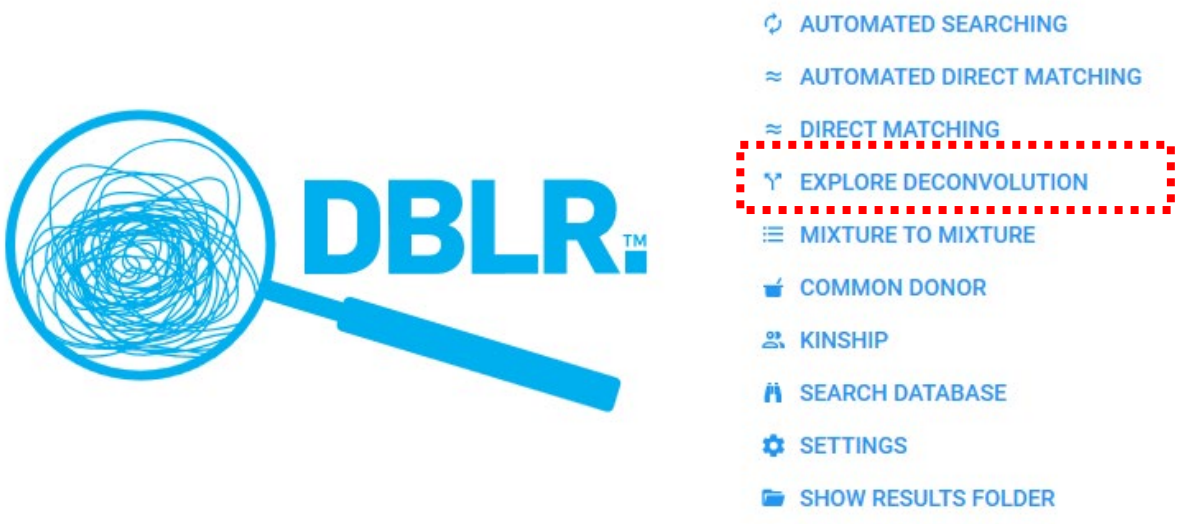
For this particular example, a low-level single-source profiles was selected from the ProvedIt dataset. In STRmix, the resulting sub-source LR was 338 which falls within the range of 0.001 – 1,000 and would be deemed as “inconclusive” by some laboratory protocols.

PER LOCUS LIKELIHOOD RATIOS

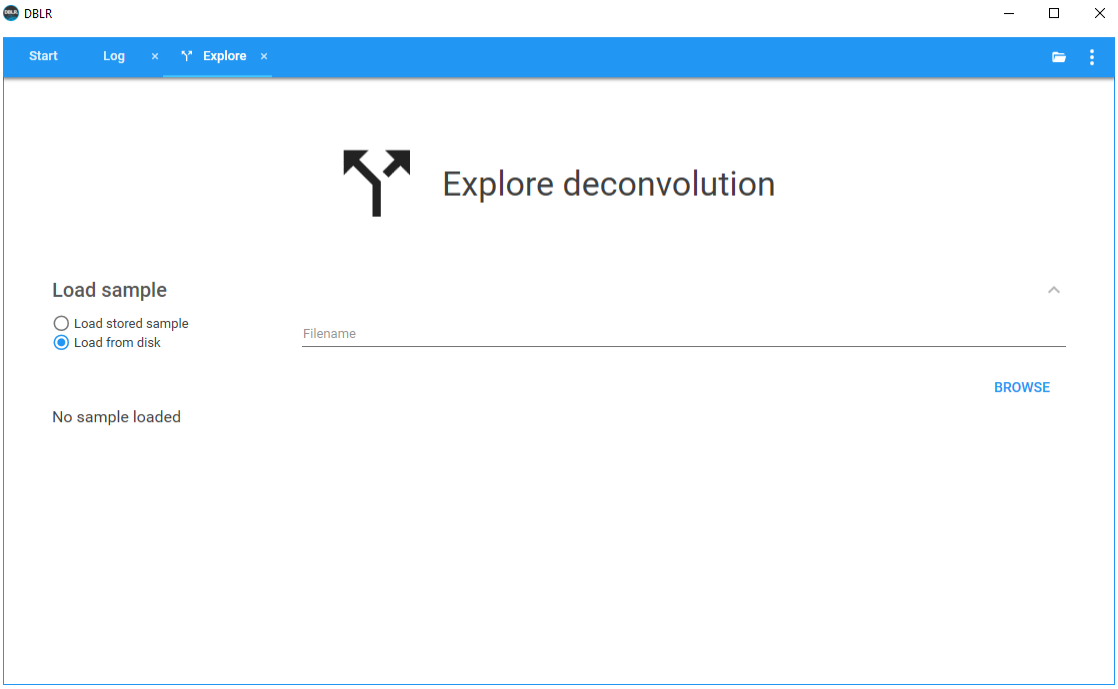
	FBI_EXTENDED_CAUC 0.01b(1.0, 1.0)		
LOCUS	Pr(E|Hp)	Pr(E|Hd)	LR
D3S1358			
vWA	2.3540E-1	1.5013E-1	1.5680E0
D16S539			
CSF1PO			
TPOX			
D8S1179	3.6175E-1	8.1464E-2	4.4406E0
D21S11			
D18S51			
D2S441			
D19S433			
TH01			
FGA			
D22S1045	2.2862E-1	4.7010E-3	4.8632E1
D5S818			
D13S317			
D7S820			
SE33			
D10S1248			
D1S1656			
D12S391			
D2S1338			
SUB-SUB-SOURCE LR			3.3861E2
SUB-SOURCE LR			3.3861E2
99% 1-SIDED LOWER HPD INTERVAL			2.2629E2

4.2 DBLR Start Page

4.2.1 On the DBLR start page, click on the “Explore Deconvolution” option of the software

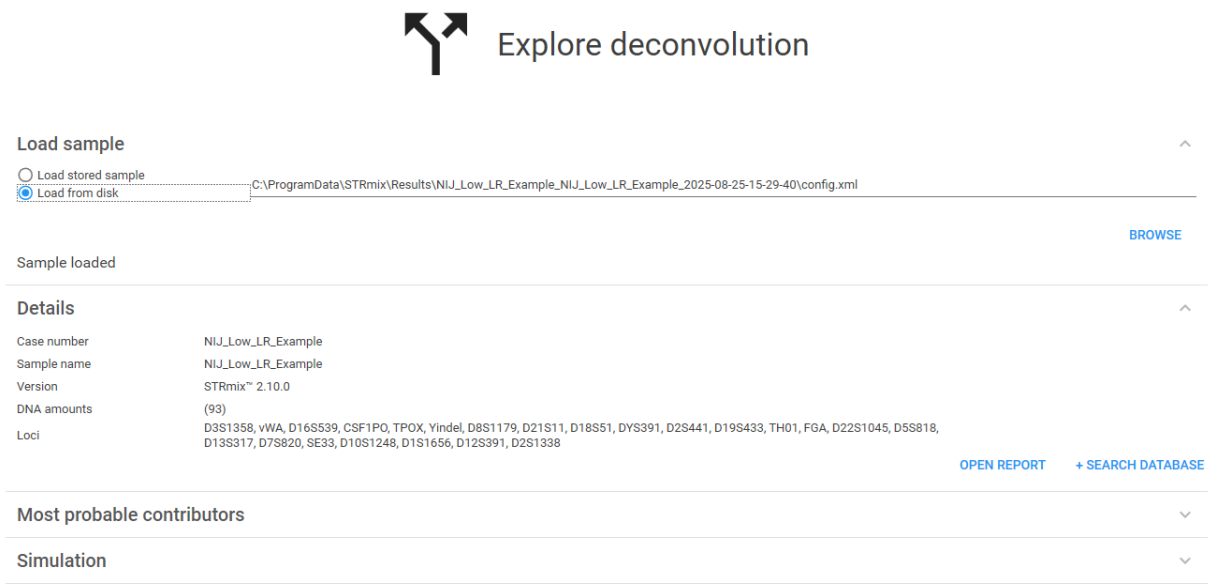


4.2.2 To load a sample, open the Results folder of the deconvolution from the STRmix sample to interpret.



4.2.3 Select the STRmix deconvolution folder for your sample/mixture, drag and drop this into the “Explore Deconvolution” page (above)

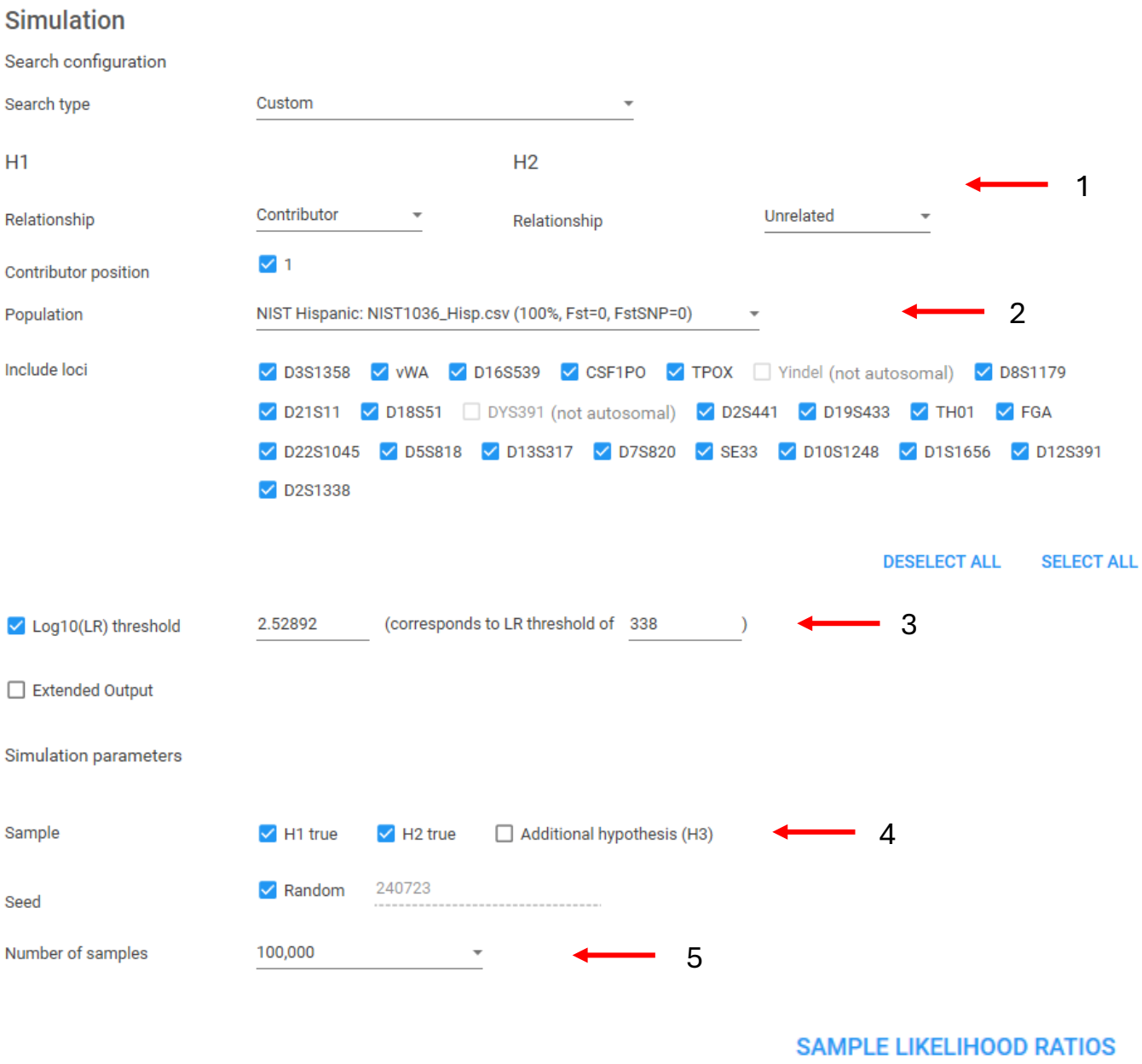
4.2.4 Once the file has been loaded, the screen should present this:



4.2.5 The details of the sample loaded from the STRmix deconvolution are presented in the window. There are two functions that can be performed: Most probable contributors and ***Simulation***.

The Simulation option will facilitate the H1 true and H2 true contributor testing.

4.2.6 Select Simulation and the following screen will open:



4.2.7 the following options can be selected:
In line 1, the analyst can select the relationship being simulated, here, H1 is a simulation of a potential contributor to the evidence profile, H2 is a simulation of an unrelated individual contributing to the evidence.

In line 2, the analyst can select a population frequency database for the testing.

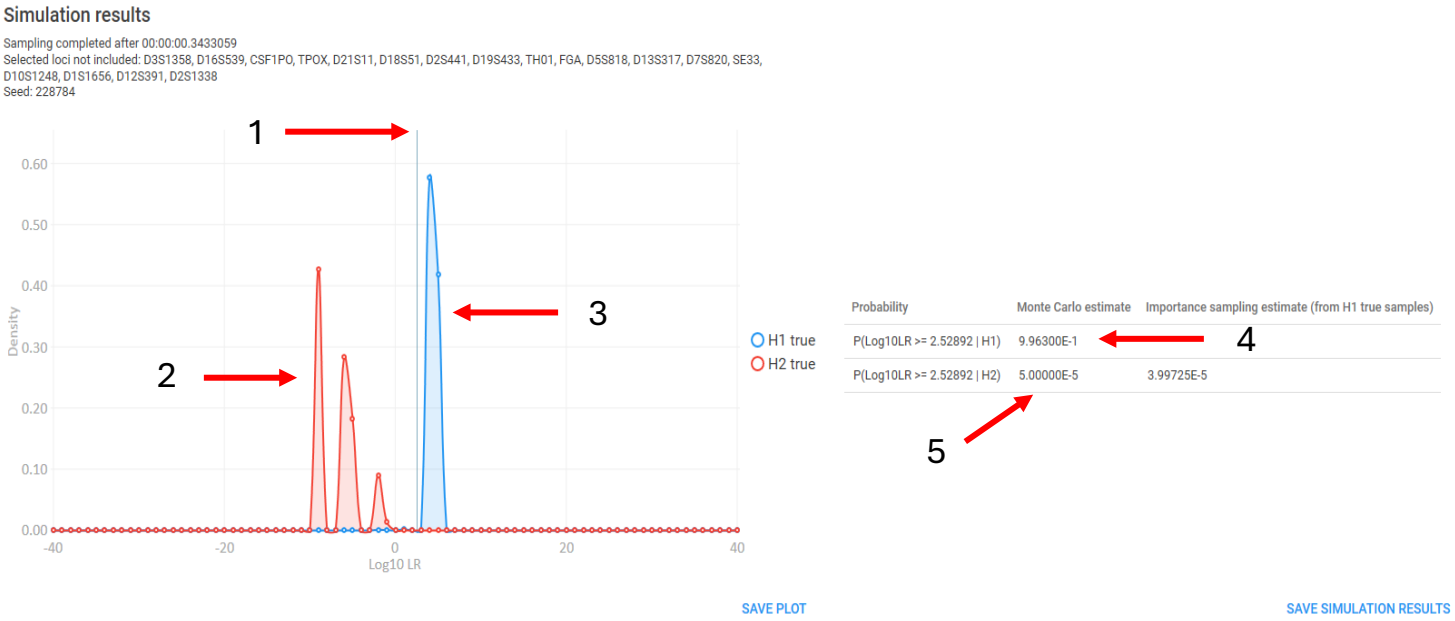
In line 3, the analyst enters the LR of the POI (the default is 0, which is an LR of 1). In this example, the POI LR of 338 = Log_{10}LR of 2.52892

In line 4, any additional hypotheses (H3) can be tested.

In line 5, the number of simulations to conduct can be selected from 10,000; 100,000; 1,000,000; or 10,000,000. For this example, a simulation of 100,000 individuals was analyzed.

Selecting "Sample Likelihood Ratios" will start the simulation.

4.2.8 At the end of the simulation, plots of H1 True (in blue) and H2 true (in red) are presented:



4.2.9 The key points in the plot are highlighted above

In line 1, the $\log_{10}$LR of the POI = 2.52892 threshold is noted as the vertical line, distributions to the left of this line plot H2 true simulations, while distributions to the right of this line plot H1 true simulations.

In line 2, the H2 true distributions are plotted with their density on the y-axis.

In line 3, a single H1 true distribution is plotted with its density (here, most of the $\log_{10}$LRs range from 3-6 bans (i.e. LRs from 10^3 to 10^6)).

In line 4, the probability of observing an H1 true $\log_{10}$LR greater than or equal to 2.52892 is 0.9963 (i.e. more than 99.63% of the H1 true profiles were greater than the POI $\log_{10}$LR = 2.52892).

In Line 5, the probability of observing an H2 true $\log_{10}$LR greater than or equal to the $\log_{10}$LR of the POI is 5.00000E-5 or 0.00005. It is

easier to represent this by determining the probability of non-contributors less than 1.0 (1 – 0.00005 = 0.99995). In other words, greater than 99.995% of H2 true contributors gave LRs less than the STRmix sub-source LR for the POI of 338 (or, a $\log_{10}$LR = 2.52892).

5. References

1. Coble MD & Bright J-A (2019) Probabilistic genotyping software: an overview. Forensic Science International: Genetics 38:219-224. https://doi.org/10.1016/j.fsigen.2018.11.009

2. DNA Advisory Board (2000) Statistical and population genetics issues affecting the evaluation of the frequency of occurrence of DNA profiles calculated from pertinent population database(s). Forensic Sci. Comm. 2(3). Available at https://archives.fbi.gov/archives/aboutus/lab/forensic-science-communications/fsc/july2000/dnastat.htm.

3. Butler JM, Kline MC, Coble MD (2018) NIST interlaboratory studies involving DNA mixtures (MIX05 and MIX13): variation observed and lessons learned. Forensic Science International: Genetics 37: 81-94.

4. Kelly H, Kerr Z, Cheng K, Kruijver M, Bright JA (2021) Developmental validation of a software implementation of a flexible framework for the assignment of likelihood ratios for forensic investigations. FSI: Reports 4: 100231.

5. Schuerman C, Kalafut T, Buchanan C, Sutton J, Bright JA. (2020) Using the Nondonor Distribution to Improve Communication and Inform Decision Making for Low LRs from Minor Contributors in Mixed DNA Profiles. J Forensic Sci. 65(4): 1072-1084.