Document Title:	**Advancing Probabilistic Approaches to Interpreting Low-template DNA Profiles and Mixtures: Developing Theory, Implementing Practice**
Author(s):	**Kirk E. Lohmueller, Keith Inman**
Document Number:	**251805**
Date Received:	**July 2018**
Award Number:	**2013-DN-BX-K029**

2013-DN-BX-K029 Cumulative Technical Report

"Advancing probabilistic approaches to interpreting low-template DNA profiles and mixtures: Developing theory, implementing practice"

Principal Investigators: Kirk E. Lohmueller and Keith Inman

Table of Contents

I. ABSTRACT

Samples containing low-levels of DNA and/or mixtures of DNA from multiple individuals are routinely encountered in forensic DNA casework. We will refer to these samples inclusively as "complex" samples. Such samples are challenging to interpret because of the inherent uncertainty in determining the genotypes of the contributors to the evidence profile. Such uncertainty must be taken into account when assessing the weight of such complex DNA profiles. Probabilistic approaches, all of which employ a likelihood ratio (LR) framework, have already been established to aide in the interpretation of such profiles. However, the performance of these approaches is still in the process of being evaluated. Further, the forensic DNA community remains in need of freely available, user-friendly tools to implement a probabilistic approach to the determination of the weight of evidence. The research completed in this project fills both of these voids by examining the performance of various probabilistic approaches as applied to complex DNA profiles, increasing our understanding of the capabilities and limitations of these approaches, and improving and extending *Lab Retriever*, an existing freely available program that can assist in understanding the weight of evidence of these complex DNA profiles. Specifically, we have: 1) Generated a set of low-template mixtures. This dataset was used for various parts of the project and will be made available to the community for evaluation of different statistical approaches. 2) Evaluated the performance of estimates of drop-out probabilities in mixed samples. Approaches to calculating LRs either require, or may be improved by, an estimate of the drop-out probability. We have previously shown how such estimates can be derived from single-source samples, how the estimates compare to the true ($P(D_0)$s, and how LRs computed with each compare. Here we have evaluated how those estimators perform when applied to mixtures. 3) Computed and evaluated results of LRs for complex mixtures of 2, 3, 4 and 5 contributors. The purpose of this project is to assess the information content of such complex mixtures, and the extent of support for a proposition that a specific individual is a contributor to the mixture. We have employed simulations to explore the results of performing LRs conditioned on a wide range of known non-contributors. We have found that good quality mixtures of up to 4 contributors still contain substantial information regarding the contributors. This type of information can be used by forensic casework laboratories to help inform both their policies and their procedures. The results of these experiments also will greatly assist in understanding the meaning of weaker LRs obtained for suspected contributors. 4) Evaluated the role of stutter in mixture interpretation. In particular, we have explored how stutter and a minor contributor with an allele in the stutter position of a major contributor peak combine to determine the final peak height. We found that they often do not combine in an additive manner. 5) Extended our user-friendly, freely-available software program, *Lab Retriever,* to help analysts perform complex LR calculations for low-template DNA mixtures. This work has increased the knowledge base upon which forensic analysts may rely to interpret complex DNA

profiles and provided new resources and tools that will aide forensic DNA analysts in accurately and efficiently assessing the weight of such evidence.

II. EXECUTIVE SUMMARY

Samples containing low-levels of DNA and/or mixtures of DNA from multiple individuals (hereafter referred to as "complex samples") are routinely encountered in forensic DNA casework [1]. Such samples are challenging to interpret because of the inherent uncertainty in determining the genotypes of the contributors to the evidence profile. Such uncertainty must be taken into account when assessing the weight of such complex DNA profiles. Probabilistic approaches, all of which employ a likelihood ratio (LR) framework, have already been established to aide in the interpretation of such profiles [2–13]. Over the past few years, research evaluating the performance of these approaches has begun to emerge, however, the efficacy, accuracy, and reliability for more complex samples remains a work in progress [14–31]. This particular issue has received increased scrutiny in recent years and was highlighted in the recent PCAST report [32] and so has become of even greater urgency. Further, although the forensic DNA community now has a choice of several commercial software programs, a strong need still exists for freely available, open-source, user-friendly tools to implement a probabilistic approach to the determination of the weight of evidence. First, not all users can meet the large funding requirements of the commercial programs; second, the black-box commercial programs must be tested against similar programs for which the computer code is available to test their veracity [29,30,33].

The research that we have completed over the course of this grant examines the performance of a probabilistic approach as applied to complex DNA profiles, increasing our understanding of the capabilities and limitations of these approaches. It also improves and extends *Lab Retriever*, an existing freely available program that can assist in understanding the weight of evidence of these complex DNA profiles. Specifically, we have: 1) Generated a set of low-template mixtures. This dataset has been used in other parts of the project and will be made available to the community for evaluation of different statistical approaches. 2) Evaluated the performance of estimates of drop-out probabilities in mixed samples. Approaches to calculating LRs either require, or may be improved by, an estimate of the drop-out probability [34–39]. We had previously shown how such estimates can be derived from single-source samples, how the estimates compare to the true ($P(D_0)$s, and how LRs computed with each compare [39]. Here we have evaluated how those estimators perform when applied to mixtures. We found that the relationship between $P(D_0)$ and the average RFUs of the contributor of interest were similar in single-source samples and mixtures. Additionally, we computed LRs for the true contributors and found that the LRs computed using a $P(D_0)$ estimated from single-source samples were similar to those found using the true $P(D_0)$). This work suggests a robustness to the estimates of $P(D_0)$, and that estimates from single-source samples can be applied to mixtures. 3) Computed and evaluated results of LRs for simulated complex mixtures of 2, 3, 4, and 5 contributors. The purpose of this

part of the project was to assess the maximum information content of such complex mixtures, and the extent of support for a proposition that a specific individual is a contributor to the mixture. The information content of such mixtures can inform laboratory policy and procedures for interpreting such samples. We have employed simulations to explore the effects of calculating LRs conditioned on a wide range of known non-contributors. We found that LRs could reliably distinguish true contributors from known non-contributors, even for a 5 contributor mixture exhibiting a high level of allele sharing. Moreover, we showed that LRs have a low false negative rate, with true contributors to a complex mixture generating LRs < 1 less than 0.1% of the time for 5 person mixtures in which the numerator includes at least one unknown contributor. Similarly, LRs had very low false positive rates, with known non-contributors generating LRs > 1 in only 0.054% of the 14,000 replicates. Further, they occurred only with 4 and 5 person mixtures with one or more unknown contributor in the numerator. However, individual loci often showed LRs <1 for true contributors. These results will greatly assist in understanding the meaning of weaker LRs obtained for suspected contributors, as well as provide guidance on the advisability of performing an LR when only a few loci produce typing results. 4) Evaluated the role of stutter in mixture interpretation. In particular, we have explored how stutter and a minor contributor with an allele in the stutter position of a major contributor peak combine to determine the final peak height. Very few published works address this important issue [40–42]. Here we have found that minor donor alleles in a stutter position to a major donor peak *can* be elevated above the normally expected stutter peak height. However, this is not always true; some [minor + stutter] peaks are below the lowest stutter peaks detected. The average RFU for the stutter peaks in all template amounts was about 50, while the average RFU for the [minor + stutter] peaks was 76 RFUs. However, the range of variation in peak height was greater for the [minor + stutter] peaks than for either the minor peaks or the stutter peaks alone. This signals that something unusual is occurring during PCR that is not easily explained. At the very least, it is clear that such a peak is not merely the direct sum of a real human DNA allele and stutter. 5) Extended our user-friendly, freely available, open-source software program, *Lab Retriever,* to help analysts perform LR calculations for complex DNA mixtures. This program is based on the approach first suggested by Balding and Buckleton [5]. By incorporating a computationally efficient dynamic programming algorithm, the program can now perform calculations for hypotheses involving up to 4 unknown contributors in the denominator within seconds. The program has also been completely rewritten to replace the original GUI code, which had become deprecated. This will ensure that the program will continue to work into the reasonable future on current and future computer platforms. Additional improvements include the ability to add and choose among any user-defined population database, the ability to choose a user-defined co-ancestry coefficient (θ or F_{ST}), the ability to choose a user defined "α" term (relevant to handling homozygotes) and a transparent export of both the numerator and denominator of the LR to provide more information to the user and

assist in research efforts. The framework to handle replicate samples has been developed and future work will connect it with a GUI for easy user access. This work has increased knowledge regarding the interpretation of complex DNA profiles and added new features to a freely-available software program that continues to aide forensic DNA analysts in accurately and efficiently assessing the weight of such evidence.

The results generated from this project are being disseminated to the community using a variety of mechanisms. First, we have generated a series of 819 low-template mixtures using the Identifiler® Plus typing system. Because we believe that the raw data will be useful to the forensic community for subsequent evaluation of different programs to interpret challenging profiles, the electronic raw data files, as well as our analyzed data, will be made available, at a minimum, via the SCIEG website. Secondly, we are publishing our results in scientific journals. We have already published our simulation study on the information content of complex mixtures [31], as well as a white paper that provides the details of the *Lab Retriever* software [10]. Additional publications on the dataset, drop-out probabilities, and stutter results are being prepared. We have given over 20 talks and presentations during the course of this work on the interpretation of challenging DNA profiles and the use of *Lab Retriever*. Finally, we continue to host the freely available, open-source program, *Lab Retriever*, on the SCIEG website (www.scieg.org). This tool is freely available to the community and has been implemented or is in the process of being validated by at least 6 government and private forensic laboratories of which we are aware. We anticipate continued use of the program and will continue to support it.

III. OVERVIEW

While forensic DNA typing has historically been touted as the gold-standard for forensic identification, interpretation of complex profiles remains challenging. Difficulties in interpretation, and in estimating the weight of the evidence, can arise when considering mixtures of DNA from complex samples. These types of samples are challenging because it is often not possible to conclusively infer the genotypes of the contributors directly from the evidentiary sample. For mixtures, the moderately polymorphic nature of the short tandem repeat (STR) systems developed for forensic use exacerbates the ambiguity, as contributors often share alleles at individual loci. For LT-DNA samples, additional ambiguity arises from the fact that certain alleles from the contributors may not be detected in the evidence profile due to stochastic effects inherent in amplifying low level samples [5,38,43]. This phenomenon is called allelic drop-out. This uncertainty in determining the genotypes of the contributors must be taken into account when assessing the weight of the evidence. For LRs in particular, difficulty in determining the number of contributors adds another element of ambiguity. The theoretical framework for a likelihood ratio (LR) approach to assess the weight of these ambiguous samples has been in place for some years, but two key limitations exist that had prevented it from being widely implemented in forensic casework: 1) absence of empirical testing to validate these

approaches when applied to different types of complex DNA samples, and 2) unavailability of affordable, transparent, and user-friendly software to forensic practitioners to perform these complex calculations.

The LR approach provides a natural method to assess the weight of forensic DNA evidence [44–48]. Briefly, the LR framework compares the probability of observing the evidence under two different competing hypotheses. In the simplest case, one hypothesis (call it H_1) is that the suspect left the DNA evidence. Another (call it H_2) is that an unknown individual left the evidence. The LR is then the ratio of the probability of the evidence given H_1 to the probability of the evidence given H_2. LRs >1 provide support for H_1 and LRs <1 provide support for H_2.

LRs can be used to assess the weight of complex DNA profiles. They have previously been extended to consider mixtures by including more complex hypotheses [2,49–51] and to accommodate allelic drop-out [3–5,36,37,43,52,53] by including a model of such when computing the probability of the evidence. For complex DNA profiles in particular, the probability of the evidence under the hypothesis that the suspected contributor left the evidence may be <1.

While the theoretical foundation for the interpretation of complex DNA profiles has clearly been established, the performance of these approaches when applied to challenging samples for which ground truth is known has only in recent years begun to receive comparatively more scrutiny [14–31]. For example, only in the summer of 2015 did ASCLAD/LAB begin to require laboratories to perform and document such validation for mixed samples [54]. The summer of 2015 also saw publication of the first SWGDAM guidelines for validation of probabilistic genotyping software [33]. Additionally, accessible software had, until recently, been unavailable to the community to perform the calculations described above. Recent years have seen an explosion of probabilistic genotyping software, both commercial and freely-available open source solutions [2,4,6,7,9–11,13,18,28]. We are pleased that *Lab Retriever*, the software supported by this grant, has become the choice of many laboratories in the U.S. opting for open-source software. At the beginning of the project, we aimed to address specific voids in the field that existed at that time. In the following sections, we detail how our work has begun to fill in these deficits.

IV. Main Findings

A. Aim 1: Generation of low-template DNA profiles.

1. Background and Significance

One way to assess the performance of a particular method to estimate the weight of evidence for a complex DNA profile is to apply the method to similarly complex profiles for which the genotypes of the true contributors are known. Methods that perform well should 1) give large LRs when the true contributor is compared to the evidence profile and 2) give

small LRs (<<1) when a random individual (i.e. a known non-contributor) is compared to the evidence profile. By examining the distributions of LRs produced under these two scenarios, it is possible to compare different approaches to assess the weight of the evidence and to determine under what conditions, or using which approach, comparison of known non-contributors might result in LRs >1.

Prior to this project, we had applied this type of simulation to assess the performance of the Balding and Buckleton LR [5] approach using empirically estimated drop-out probabilities [39]. Specifically, we used 60 low-template single-source DNA profiles generated from two individuals by John Butler's group at NIST as evidence profiles [55]. We then compared both known contributors and known non-contributors to each of the low-template profiles. We found that for single-source samples, the Balding and Buckleton approach (later implemented in *Lab Retriever*), performed satisfactorily, and that drop-out probabilities estimated using average peak heights worked as well as the true benchmark probabilities. Here we applied this same type of analysis to a far greater number of mixed DNA profiles.

Application of this type of analysis to mixed DNA profiles requires a comprehensive dataset of mixtures reflecting both adequate template and LT situations. At the time we proposed this work, such a dataset did not yet exist in the public domain. Since that time, one such collection of profiles has been made available by Robin Cotton from Boston University. While that set of profiles possesses utility for certain applications, it lacks certain characteristics, such as a sufficient number of amplification replicates, enough different profile combinations, and mixtures designed to address particular and/or extreme situations required to address the specific questions that we wanted to pose. Importantly, validation studies of other probabilistic genotyping systems have relied heavily on adjudicated cases, rather than on laboratory generated mixtures for which ground truth is known [56]. While using case profiles sounds appealing because they mimic the types of samples actually encountered in casework, ground truth is not known for critical parameters, including the genotypes of the true contributors, the proportions in which they exist in the mixture, and whether drop-out has occurred. For all of these reasons, the use of casework profiles, regardless of the outcome, is an unreliable indicator of ground truth, and therefore unsuitable for validation of probabilistic approaches to weight of evidence. Mitchell et al. [3] and Perlin et al. [23] described an analysis of mixed DNA samples. However, their data are not publicly available to other researchers and cannot be used to compare different statistical approaches. While additional studies based on physical samples for which ground truth is known have recently been published [4,8,14,21,26,27,40–42] , those data have also not been made publicly available, and/or are based on European systems which are slightly different than those used in the U.S. [11,17,24,28]. Thus, it was clear that a critical need existed for a thoughtfully constructed large dataset of mixed DNA samples, comprising 2, 3 or 4 contributors, encompassing

varying mixture ratios and DNA concentrations, at with at least 5 replicate amplifications of each sample, that could be made available for use by the forensic DNA community.

2. Materials and methods

Overview

Here we generated a series of 819 mixtures, many of which were low-template for which drop-out was a possibility. We also created dilutions series of 3 of the single-source samples used to create the mixtures. Sample collection, extraction, and typing were performed at Cal State University East Bay.

Design of the study

The parameters typically considered to contribute the greatest difficulty in assessing the weight of evidence in complex samples include allelic drop-out, the number of contributors, shared alleles, mixture ratios, and stutter. In order to investigate the effect of these parameters in Aims 2 and 4, we produced a well-characterized set of mixed samples. These samples consisted of known types combined in defined ratios. We varied the total amount of DNA, the number of contributors, the mixture ratios, and the genotypes of the contributors to produce mixtures of two, three and four individuals to create the easiest and the most difficult mixtures to assess. This gave us a broad range of peak heights, masked alleles in stutter positions, and allele sharing to test the performance of LRs both with and without drop-out. The samples containing multiple individuals also exhibit various masking scenarios, either involving shared allelic peaks or stutter peaks. Tables 1-4 contain the complete matrix of the samples that we generated. Table 5 shows the number of peaks from one contributor that fall in the stutter positions of the second contributor.

Each sample was amplified 5 times. The total number of mixed samples generated from this matrix was 819. The total number of alleles detected was over 31,500.

Sample Collection

All samples were collected with informed consent and were anonymized to ensure the privacy of the contributing subjects in accordance with the UCLA and CSUEB IRB. Two sterile buccal swabs were collected from each of eight student volunteers. Once collected, the swabs were allowed to air dry, sealed in a manila envelope and stored frozen until used. The genotypes of the 8 individuals are shown in Table 1.

Sample Extraction

A separate sterile scalpel was used to bisect each swab and one half of the swab was placed into a sterile microfuge tube. The Qiagen DNEasy kit was used to extract each sample, following the manufacturer's protocol. Briefly, each sample was mixed with PBS, proteinase K, and AL Buffer and incubated for 10 min at 56°C. Each sample was mixed with 200µL ethanol, and then placed into a DNEasy spin column. The samples were centrifuged at 8,000 rpm for 1 min., and the eluate discarded. The samples were washed once with

AW1 buffer at 8,000 rpm for 1 min., and again with AW2 buffer at 14,000 rpm for 3 min., discarding the eluate each time. DNA was eluted from the column with Buffer AE, incubating the spin tube for 1 min at RT, and then centrifuging for 1 min at 8,000 rpm. Samples were stored frozen when not in use.

Quantitation

The DNA quantity of the samples was estimated using the Quantifiler® Human DNA Quantification Kit (Life Technologies, Foster City, CA) on either a Bio-Rad Laboratories Opticon® 2 or an ABI Prism® 7000 Real Time PCR instrument, following each manufacturer's protocols. Instrument data was interpreted by either Opticon 3 quantitation software or SDS software.

Mixture preparation

Dilutions (500, 100, 50, 30, 10 and 5 pg) of three single-source samples were prepared to estimate the probability of drop-out.

Five replicates of two, three and four person mixtures were prepared in Tris/EDTA buffer (TE^{-4}). The combinations prepared are summarized in Tables 2-4.

Two person mixtures were prepared in ratios of 1:1, 2:1, 4:1, and 9:1. Each mixture was amplified using a total DNA input of 500, 100, 50, and 30 pg (Table 2).

The two-person mixture set combinations were designed to produce extreme examples of one of two parameters, heterozygosity or overlap in stutter positions, as follows:

Max # homozygous peaks/Max # heterzygous peaks
Max # heterozygous peaks/Max # homozygous peaks
Max # homozygous peaks/Max # homozygous peaks
Max # heterozygous peaks/Max # heterozygous peaks
Maximizing # of peaks in stutter position for minor donor
Minimizing # of peaks in stutter position for minor donor

Three person mixture ratios were: 1:1:1, 2:2:1, 6:3:1, 8:1:1. Each was amplified using total input DNA of 500, 100, 50, and 30 pg (Table 3).

The three person mixture sets were designed to produce extreme combinations of heterozygosity:

Max # homozygous peaks/Max # heterzygous peaks
Max # heterozygous peaks/Max # homozygous peaks
Max # homozygous peaks/Max # homozygous peaks
Max # heterozygous peaks/Max # heterozygous peaks

The four person mixture sets were all ratios of 1:1:1:1, with no specific goal of stutter or heterozygosity. Again, each was amplified using total input DNA of 500, 100, 50, and 30 pg (Table 4).

STR amplification and detection

Thermal cycling was performed using the GeneAmp® PCR system 9700 (Life Technologies). For STR amplification, 25μL volumes of each ratio/template amount/replicate were prepared in 96-well plates. Replicates were prepared using the AmpFlSTR® Identifiler® Plus Amplification Kit (Life Technologies) following the manufacturer's protocols [6]. PCR conditions for the IdentifilerPlus Kit were 11 min at 95°C for initial incubation, 28 cycles of 20 sec at 94°C for denaturation, 3 min at 59°C for annealing, and 30 min at 60°C for final extension.

STR detection and typing

Typing of PCR products, including the kit positive control and one or more negative control samples, was carried out in 96-well plates on a 3130 Genetic Analyzer (Life Technologies). Each samples was injected for 5 sec at 3kV and data collected for 12 – 18 minutes (depending on the temperature in the laboratory). The data were analyzed using GeneMapper® ID-X (Life Technologies).

Color-specific analytical thresholds were established on a per-run basis by estimating the baseline as 2x the maximum noise peak from one or more of three negative PCR reaction controls run with each amplification plate. The range of analytical thresholds for all of the runs calculated in this way are presented in Table 6.

3. Results

The typing data will be made publicly available in the near future. The availability will be posted on the SCIEG web site (the non-profit entity that houses *Lab Retriever*) as well as on the university web sites of Professors Lohmueller and Inman. Announcements will also be disseminated through various professional news outlets.

B. Aim 2: Evaluation of estimated drop-out probabilities in mixed samples.

1. Background and Significance

We have shown in previously published work [39] that a logistic regression approach [34,35] can be used to model the relationship between allelic drop-out and peak heights in an evidence profile. This logistic equation can then be used to estimate the drop-out probability for an evidentiary profile based on the relevant peak heights. In that

foundational work, we fit the logistic curve to single-source LT-DNA profiles with 10, 30, and 100 pg of DNA.

We then evaluated the performance of our estimate of the drop-out probability [39]. To do this, we compared the LRs calculated using the estimated drop-out probabilities to the LRs calculated using the true drop-out probabilities. These calculations were performed for 2 different scenarios for each low-template profile: 1) the hypothesized contributor was the true contributor, and 2) the hypothesized contributor was a random individual simulated from a population allele frequency database (i.e. a known non-contributor). Importantly, we found that, using the Balding and Buckleton [5] approach, LRs computed with an empirically estimated probability of drop-out were similar to the LRs calculated when using the true drop-out probability [39]. Additionally, we found that >99.4% of LRs were ≤1 when comparing the LT-DNA profiles to known non-contributors, suggesting that this approach correctly provides support for the hypothesis that random individuals are not contributors to the LT-DNA samples [39].

While these results are certainly encouraging, further work was required to assess the performance of the estimated drop-out probabilities when applied to mixtures. It was unclear whether these estimates would show the same level of robustness in mixed samples as they showed in the single-source samples. Before we began this project, no systematic evaluation of the performance of estimated drop-out probabilities for low-template mixtures had been published.

2. Results

In order to fill the gap in knowledge described above, we conducted a detailed analysis of allelic drop-out using the data generated in Aim 1 of this project.

Characterization of drop-out and estimation of drop-out probabilities

We first used logistic regression to model the relationship between the proportion of alleles that dropped-out from each profile and the average RFUs of the peaks in that profile. When considering single-source LT samples, little drop-out was observed when the average RFUs were >150 RFUs (Figure 1). It is noteworthy that, for the single-source data we analyzed data from three different individuals, each with different multi-locus genotypes, the logistic regression curves fit to the data from each of the three individuals appear to be similar to each other (Figure 1). This suggests that the drop-out probability is not substantially influenced by the particular alleles carried by different individuals.

We next performed a similar analysis for 2 to 4-person mixtures. Importantly, when computing the average RFUs and the proportion of alleles that dropped out, we considered just heterozygous peaks unique to the specified contributor(s). Furthermore, we did not include peaks in stutter positions for the logistic regression analysis because such peaks may contain contributions from both stutter and alleles from the low-level contributor (further explored in Aim 4). When examining the mixed samples, little drop-out is observed

in profiles that show an average minor contributor peak height >150 RFUs (Figure 2). Below about 150 RFU, drop-out decreases in direct proportion to the average peak height of the profile. We note that the logistic regression curves for the different number of contributors appear qualitatively similar to each other, though there is a small amount of variability (Figures 2 and 3). This finding argues that, for a given average peak height of a particular contributor, drop-out in a mixture is not substantially different than what occurs in single-source samples at the ratios and DNA concentrations assessed here. Table 7 shows the logistic regression parameters inferred from different subsets of the data.

Evaluation of estimated drop-out probabilities for mixtures

The analysis described above modeled the relationship between drop-out and peak heights using logistic regression. This logistic regression model can be used to estimate the probability of drop-out ($P(D_0)$) from an evidentiary sample using the average peak heights from the evidentiary sample. Before this approach can confidently be applied to mixed samples, we need to evaluate its performance. To do this, we examined the LRs obtained when using the true contributor (TC) for each of the evidence profiles. The LRs were computed using the command line version of *Lab Retriever* assuming a drop-in probability of 1%, θ (F_{ST}) of 0.01 and using the allele frequencies from NIST [57]. For each LT profile, we computed several LRs. First, we computed the LR using the benchmark $P(D_0)$. This is the true $P(D_0)$ and is simply the proportion of alleles (0,0.5, or 1) that actually dropped out at a particular locus. This would not be known for an actual case sample, but, by design, this information is known for these mixtures analyzed in this project. Second, we computed a LR using $P(D_0)$ estimated by applying the logistic regression model to data from single-source samples. This is meant to mimic the straightforward practice of using single-source sample validation data to fit the logistic regression model and then applying that model to more complex profiles. Third, we estimated $P(D_0)$ using the logistic regression model fit to the actual type of mixture sample. For example, for 3-person mixtures, we used the logistic regression fit to the 3-person mixtures. Importantly, to avoid over-fitting, 3 replicates of each profile were used to estimate the parameters of the logistic regression model of drop-out (Table 7) and the remaining 2 replicates were used to evaluate the performance of the LR.

Focusing on the 2-person mixtures, we compared the log(LRs) for TCs when estimating the $P(D_0)$ using the three different approaches described above (Figures 4 and 5). Here we considered the following hypotheses in the LR: H1: Suspected contributor + 1 unknown contributor; H2: 2 unknown contributors. Overall, the LRs estimated using the 3 different $P(D_0)$s are remarkably similar to each other (Figures 4 and 5). However, for some profiles, the log(LRs) calculated from the estimated P(D_0)s were smaller than those computed using the benchmark values of $P(D_0)$. When the TC was used as the major contributor, this effect was more pronounced for the log(LRs) computed using $P(D_0)$ estimated from the single-source samples than from those using the 2-person mixtures

(Figure 4). When the suspected contributor was the minor contributor, this effect was more pronounced for the $P(D_0)$ estimated from the 2-person mixtures than those using the single-source samples (Figure 5). When the TC was the major contributor, all LRs computed using the estimated $P(D_0)$s were never more than 2 orders of magnitude larger than those using the true benchmark $P(D_0)$ (Figure 4). When the TC was the minor contributor, 99% of the LRs computed using the estimated $P(D_0)$s were no more than 2 orders of magnitude larger than those using the true benchmark $P(D_0)$ (Figure 5).

Turning to the 3-person mixtures, we considered the following hypotheses in the LR: H1: Suspected contributor and 2 unknown contributors; H2: 3 unknown contributors. Generally, we also found that the log(LRs) estimated using the three types of $P(D_0)$ in this situation were quite similar to each other (Figures 6 and 7). Note that for certain samples, the log(LR) calculated using $P(D_0)$ estimated from the single-source samples tend to be smaller than those from the benchmark $P(D_0)$. For 95% of the cases when the TC is the major contributor, the LRs computed using the estimated $P(D_0)$ were not more than 2 orders of magnitude larger than those calculated using the benchmark $P(D_0)$ (Figure 6). Importantly, when the TC is the minor contributor, 99% of the LRs computed using the single-source $P(D_0)$ were not more than an order of magnitude larger than the LR produced using the benchmark $P(D_0)$ (Figure 7). Importantly, cases where the LRs computed using estimated $P(D_0)$s were more than an order of magnitude larger than those computed using the true $P(D_0)$ occurred for log(LRs) in the 15-20 range.

For the 4-person mixtures, we considered the following hypotheses in the LR: H1: Suspected contributor and 3 unknown contributors; H2: 4 unknown contributors. In general the patterns observed for the 4-person mixtures mimicked those for the 2 and 3-person mixtures. We found that log(LR)s computed using the 3 different values of $P(D_0)$ were similar to each other (Figure 8).

We note that for 17 2-person mixtures and 14 3-person mixtures, the log(LR) for the true contributor is <0 (Figures 5-7), suggesting that the true contributor is not present in the mixture. This tends to occur more often for the 2 and 3-person mixtures when the true contributor is the minor contributor. Because we observe this trend regardless of which way we estimate $P(D_0)$, and see the same pattern even when using the benchmark $P(D_0)$, this effect is likely not caused by imprecise or biased estimation of $P(D_0)$. To explore this effect further, we performed a re-analysis of the 17 2-person mixtures giving log(LR)s<0 for true contributors using a range of $P(D_0)$ from 0 to 0.9 (Figure 9). For the majority of the cases, regardless of which $P(D_0)$ was used, the log(LR) remained <0. In a couple of cases, the log(LR) increased above 0, but only to a maximum of 0.25. This suggests that improved estimation of $P(D_0)$ will not improve this situation. Rather, our results point out a limitation of the current version of *Lab Retriever* that uses the same $P(D_0)$s for all contributors to the mixture when the different contributors may have different $P(D_0)$. Using the same $P(D_0)$ for all contributors of the mixture can yield log(LR)s that are too small in this situation. Figure

10 shows how the log(LR) behaves as a function of the benchmark $P(D_0)$. For 2 and 3-person mixtures, when $P(D_0)$ was greater than 0.6, the log(LRs) start to move <0. This may be a useful heuristic to determine *a priori* whether a LR calculation in *Lab Retriever* may be too small. It is important to note that these experiments do not include an assumed contributor to the mixture. If, however, case circumstances suggest it is appropriate to assume a contributor, then, in the case of a 2-person mixture, the drop-out probability will only be applied to the single unknown contributor, possibly mitigating this problem. Future developments of *Lab Retriever* will consider distinct drop-out probabilities for the different contributors.

To assess the ability of *Lab Retriever* to correctly exclude individuals who did not contribute to a mixture, we compared the mixtures generated from Aim 1 to a series of known non-contributors (KNCs). These known non-contributors were simulated from an allele frequency database (see below for Aim 3 as well as Lohmueller et al. [39] and Marsden et al. [31]). We then computed log(LRs) comparing each of these 10,000 known non-contributors to each of our mixtures using *Lab Retriever* (Table 8). For these comparisons, we used the true benchmark $P(D_0)$. Overall, we find that log(LR)s>0 occur quite rarely, <1% of the time. They occur more often (0.96%) for 4-person mixture comparisons and less often for 2 and 3-person mixtures. This is not surprising given the higher risk of a coincidental inclusion based on unrelated alleles in a higher order mixture. However, even those comparisons with log(LRs)>0 tend to be fairly close to one. Less than 0.0052% were larger than 100. The maximum log(LRs) across each set of comparisons was 6516, which occurred for a 3-person mixture.

3. Conclusions

We assessed patterns of allelic drop-out in 819 low-template mixed samples of 1-4 contributors. Overall, we found that, given the average RFUs, drop-out tended to behave similarly, regardless of the number of contributors in the sample. Further, for a given mixture, using single-source samples to estimate $P(D_0)$ tended to yield log(LR)s for the true contributor that were similar to the log(LR)s obtained using the true benchmark $P(D_0)$. Further, using the single-source $P(D_0)$, if anything, tended to understate the weight of the evidence in a limited number of cases. Approximately 95% of the LRs computed using the single-source $P(D_0)$ were not more than 2 orders of magnitude greater than the log(LR)s using the benchmark $P(D_0)$. This finding suggests that $P(D_0)$ estimated using logistic regression modeled on single-source samples performs well and does not carry a risk of overstating the strength of the evidence when applied to more complex samples. Further, we can reliably exclude known non-contributors in >99% of the comparisons considered.

C. Aim 3: Compute and evaluate results of Likelihood Ratios (LR) for complex mixtures.

1. Background and Significance

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Even without the added complication of potential drop-out in LT samples, at least two factors may reduce the information content of multi-contributor mixed samples First, many of the alleles at a particular locus may be present in the evidence sample. As such, the ability to exclude people as contributors to the mixture is diminished. Second, two or more contributors to the mixture may share the same alleles. Thus, it is more difficult to infer both the number of donors to a mixture, as well as the genotypes of the true contributors of the mixture, directly from the evidentiary sample. This effect further reduces the ability to distinguish between the proposition that a particular person is a contributor to the mixture and the proposition that the individual is not a contributor. If complex mixtures routinely produce LRs around 1, then this suggests that they contain little information, and thus, may not be worth the time and effort to interpret them. Additionally, if, due to the fact that many of the alleles at a locus are present in the evidence profile, known non-contributors produce LRs >1, then this suggests that finding an LR >1 may not be very meaningful evidence.

Before we started this project, a small body of work suggested that, especially for mixtures, some proportion of true non-contributors will generate LRs >1 [3,39]. If complex mixtures generate LRs <1 (i.e. suggesting that the hypothesized contributor is not a true contributor to the evidence) even when all of the suspected contributor's alleles are present in the evidence profile, then this may be important evidence that should be carefully considered in the case. Intuitively, if the suspected contributor's alleles were all present in the evidence sample, one might expect that the LR would be ≥1. However, Weir et al. [50] showed that this is not always the case. Specifically, they showed that if all of the alleles at the locus were detected in the evidence profile, and the hypothesis in the numerator included at least one unknown contributor, the resulting LR could be <1 if the suspected contributor in the numerator carries common alleles at the locus. We are not aware of any follow-up work in the literature to assess how often this effect would be expected to occur in different types of mixtures or with different genotypes for the hypothesized contributor. While the example presented in the Weir study [50] used loci contained in the historical Polymarker genetic typing kit, all of which contain only 2-3 alleles, the situation can be extended to STR megaplex mixtures containing 3-4 people. Thus, complex mixtures of >2 individuals provide ample opportunity to produce LRs <1, depending on the nature of the particular profile, the number of alleles at a particular locus, and the allele frequencies.

Therefore, further work was required to assess how often LRs fall below 1, the magnitude of the difference, and even if an LR of 1 should be considered the default neutral point. Additionally, we wanted to assess the distribution of LRs in complex mixtures when the hypothesized contributor is a true contributor and when the hypothesized contributor is a known non-contributor. To fill this void, we have conducted a simulation study to

investigate the performance of LRs in complex mixtures of up to 5 contributors. These results will enable a more accurate interpretation of LRs produced for complex mixtures.

2. Results

The objective of Aim 3 was to assess how well LR approaches perform at distinguishing true contributors (TC) from known non-contributors (KNC) for complex DNA mixtures with 2-5 contributors. To achieve this, we simulated a set of 10,000 complex mixture replicate sets assuming no drop-out. Individual genotype profiles were simulated using the program R, by sampling two alleles for each locus from a multinomial distribution with the parameters 2 and p, where p is the vector of allele frequencies for a specific locus. For the purposes of these simulations, we used Caucasian allele frequencies generated by the NIST group [58]. For each replicate in our simulation, we simulated a set of six individual genotypes. The first five individuals were used as contributors to create the evidence mixtures and are hereafter referred to as C1, C2, C3, C4 and C5 (Table 9). Specifically, the two person mixture was created by combining the genotypes for C1 and C2 generated for that replicate, a three person mixture was created by combining C1, C2, C3 for that replicate, and so on, up to a five person mixture. As such all mixtures were created assuming no drop-out. The sixth individual (C6) was simulated to represent a non-contributor (discussed below), and thus was not included in any of the mixtures.

Then for each mixture set, we computed LRs (using the freely available program DNAMIX, http://genomine.org/dnamix/index.html) when: 1) a TC was compared to the mixture, and 2) a KNC was compared to the mixture. The TC was one of the multilocus genotypes that went into the mixture, in each case the sample designated C1. For each simulated mixture, we compared one TC (C1) and one KNC (C6), varying the number of assumed donors. In all, 280,000 LRs were computed for this portion of the study (10,000 replicates x 14 distinct hypotheses x 2 (TC or KNC in the numerator).

We found that LRs could reliably distinguish true contributors from KNC, even with a 5 contributor mixture in which a high level of allele sharing exists (Figure 11). Moreover, we showed that LRs have a low false negative rate, with TCs to a complex mixture rarely generating a LR < 1 (7 of 140,000 replicates), and then only for complex mixtures with 5 contributors in which the numerator hypothesis includes one or more unknown contributor (Table 10). Similarly, LRs had very low false positive rates, with KNCs generating LR > 1 in only 75 of the 140,000 replicates (0.054%), and only with 4 and 5 person mixtures with one or more unknowns in the numerator (Table 10).

For some complex samples, not all loci will amplify to produce detectable alleles (locus dropout). We wanted to know the effect that a reduced number of loci would have on the overlap of LR distributions for TCs and KNCs in such a circumstance. We assessed this situation based on a 9 locus Profiler Plus profile. Similar to the 15 locus complete profile (Figure 11), separation in the distribution of LRs from TCs and KNCs was good

when based on a 9 locus profile (Figure 12). Notably, however, a greater overlap in the distributions was observed, as TC LR values were lower, and KNC values higher, than for the 15 locus profiles. This is particularly apparent as the number of contributors and number of unknowns in the numerator increases. For example, 88% of replicates for a 3-person mixture with 1 unknown in the numerator yielded LR > 1 million when using a complete 15 locus profile compared with 22% from a 9 locus profile. By contrast, 83% of replicates for 5 person mixtures with 2 unknowns in the numerator yielded LR > 1000 with a complete 15 locus profile compared with just 37% from a 9 locus profile. In addition, the number of TCs with LR < 1 (64/140,000) and the number of KNCs with LR > 1 (751/140,000) was an order of magnitude higher with the 9 locus profiles compared to the complete profiles.

How commonly do TC yield LR < 1 at individual loci?

As discussed above, we found that 7/140,000 replicates yielded LRs < 1 for TCs. All of these examples were associated with 5 person mixtures with 1 or more unknown contributors in the numerator. We investigated how frequently LRs < 1 were found at *individual loci* when based on a 15 locus profile. While hypotheses with 0 unknowns always gave a LR > 1 at all loci, for other hypotheses a large proportion of replicates generated a locus with a LR < 1. For example, ~ 45% of replicates for 2-person mixtures and ~95% of replicates from 5 person mixtures showed at least one locus with a LR < 1 (Table 11). These results suggest that a potential contributor should not be excluded on the base of a single locus.

3. Conclusions

Overall, we find that simulated mixtures containing all alleles of the input profiles, and including up to 5 contributors, provide a substantial amount of information. It is possible to distinguish between true contributors and non-contributors quite reliably, using a LR framework that does not include information about peak heights. As such, while overlapping of alleles (allele sharing, sometimes called "stacking") may complicate the deconvolution of mixtures, it does not negate the numerical weight of such evidence. Of course, complications such as degradation and allelic drop-out, which were not examined in this part of the study, will complicate the interpretation, and likely decrease the strength of the evidence. As such, this work provides information to establish a ceiling for the expected LR under optimal conditions. In this way, our work provides information to laboratories to assist with optimizing their policies and procedures.

The results of this part of the project were published in Marsden et al. (2016) [31]. Please see that publication for additional technical details and discussion of our findings.

D. Aim 4: Evaluation of the role of stutter in mixture interpretation.

1. Background and Significance

A characteristic of Short Tandem Repeat (STR) *in vitro* amplification is the loss or gain of repeat units [60,61]. The frequency of these events is inversely proportional to the repeat length. Thus dinucleotide repeats show the highest occurrence of "stutter" peaks and the frequency decreases as the repeat length increases to tri-, tetra, and penta-nucleotide repeats. The most prevalent event is the loss of one repeat unit, however the gain of one repeat unit is observed at a lower frequency. Likely the loss or gain of additional repeat units also occurs, but at a frequency that is not usually detected under conditions typical of forensic DNA typing.

While stutter peaks pose no problem to interpreting single-source profiles, they can confound the interpretation of mixtures, in particular those in which a minor component is present in the same RFU range as the stutter peaks of a major contributor [40–42]. It is often assumed that for such overlapping alleles, the contribution of the minor allele and the stutter peak are directly additive. Little work has been performed to address this important issue and to investigate the relative contributions of a minor allelic peak and a stutter peak to the total height (RFU), but see [40–42].

We have included mixtures designed specifically to address this question (Table 5).

2. Results

As a simple first pass, data from all template amounts were used and alleles representing three different conditions were identified:

1. Alleles solely from the minor donor
2. Peaks due to stutter only
3. Peaks from the minor donor that occur in stutter positions for the major donor [minor+ stutter].

Only the 9:1 mixtures produced peaks of similar height in both the stutter peaks and minor donor alleles. The average peak height for the minor donor alleles was 55 ($\sigma = 39$), while the average height for stutter peaks was 50 ($\sigma = 33$). The distribution of the heights for both types of peaks appears visually similar (Figure 13). In addition to the concordance of means of the peak heights, the full distributions of the stutter and minor peaks appear to be similar to each other (Figure 14, compare the red and black lines). Further, shorter alleles tend to show elevated stutter relative to longer alleles (Figure 13).

For those peaks from the minor donor that occur in stutter positions of alleles from the major donor, a clear increase in RFU is seen in some, but not all, samples (Figure 15). More important, RFU values for the [minor + stutter] peaks are clearly more variable than for either alone, exhibiting values both lower and higher than either the minor or stutter peaks show separately. The average RFU of these peaks is 76 ($\sigma = 66$). The [minor+stutter] peak height distribution has a longer tail of high RFU values relative to the minor peak height distribution and the stutter peak height distribution (Figure 14). Given that the

medians of all three distributions are similar (45, 46, and 49 RFUs for the minor, stutter, and [minor+stutter] distributions, respectively), the long tail is driving the higher average RFUs for the [minor+stutter] peaks. Another way to visualize this difference is to plot the peak height ratios of the stutter only and the [minor + stutter] peaks (Figure 16; trend lines added). The peak height ratio is higher for the [minor + stutter] peaks, although the degree to which it is elevated is quite variable. The trend lines appear to differ between the two sets of peaks.

If stutter and minor peaks were truly additive, then the heights of the [minor+stutter] peaks could be modeled by summing the heights of the stutter and the minor peaks together. To test the additive model, we randomly sampled 47 minor peaks and 47 stutter peaks (we have 47 [minor+stutter] peaks in the distribution shown in blue in Figure 17). The height of a stutter peak was added to each minor allele peak, giving a distribution of 47 heights that each are the sum of a stutter peak and a minor peak. We did this 1000 times, combining different sets of 47 peaks together. We find that the [minor+stutter] peak distribution (blue curve in Figure 17) does not fall within the range of the distribution from what is predicted when stutter and minor allele peaks are additive (gray curves in Figure 17). This allows us to reject a model where the heights [minor+stutter] peaks are simply the sum of the minor peak heights and the stutter peak heights. The lack of fit of the additive model is especially pronounced in the low-template range (<150 RFUs). For higher peak heights >150 RFUs, the additive model is more consistent with the observed heights of the [minor+stutter peaks] (note the blue curve falls within the bounds of the gray lines in Figure 17 for peak heights >150 RFUs).

3. Conclusions

Our data demonstrate that minor donor alleles in a stutter position of a major donor peak *can* be elevated above the normally expected stutter peak height. However, this is not always the case; some [minor + stutter] peaks fall below the lowest stutter peaks detected. The average RFU of the stutter peaks for all template amounts was about 50, while the average RFU for the [minor + stutter] peaks was 76. However, the range of variation was greater for the [minor + stutter] peaks than for either the minor peaks or the stutter peaks alone. This signals that something unusual is occurring during PCR that is not readily explained. At the very least, it is clear that such a peak cannot be assumed to be the direct sum of a true allele and stutter. Because the fluorescent signal of a stutter peak and a real allele is expected to be additive, our rejection of an additive model argues that some other aspect of the typing process is responsible for the apparent lack of additivity. One possibility is that, when present together, the PCR amplification of the true allele, and the generation of the N-1 stutter product become dependent in some as yet undefined way.

It is possible that the wide range of [minor + stutter] peak heights is related to a difference in the repeat sequence between the stutter peak and the minor donor allele, as outlined in Bright et al. 2013 [40]. That possibility was not investigated in this study.

For mixed samples, a common practice among practitioners is to subtract some statistical value related to expected stutter (maximum stutter or average stutter, for example) from the peak in a stutter position, and designate that peak as representing a true DNA allele present in the DNA source if the remaining RFU value is above the analytical threshold. The data developed here suggests that this is unwise, as the height of a peak resulting from a true minor allele + stutter cannot be assumed to be directly additive.

E. Aim 5: Development, distribution and support of *Lab Retriever.*

1. Background and Significance

While the theoretical foundation for calculating LRs for complex DNA samples has existed for some time, at the time our project began, no freely available, transparent, and user-friendly software was available that forensic DNA analysts could use to perform the calculations. Some programs provide a graphical user interface, but are opaque, difficult to use and/or require a substantial monetary investment (e.g. TrueAllele®). Others, while somewhat more affordable, only deconvolve mixtures without considering the probability of drop-out, and do not employ a probabilistic approach (e.g. Armed Xpert™.) Still others are freely available, but require expertise in computer programming to use (e.g. the original LRMix, likeLTD, DNAmixtures, European Forensic Mixtures). In recognition of this void, we developed *Lab Retriever*. *Lab Retriever* provides a graphical user interface (GUI) to the algorithms described by Balding and Buckleton [5]. As such, this program provides a user-friendly environment in which forensic analysts can calculate LRs incorporating a P(D_0) for complex samples (Figure 18).

2. Results

During this project, we have made a number of substantive improvements to our *Lab Retriever* program. While the program is based on R code originally published by Balding and Buckleton [5], *Lab Retriever* includes a number of important technical

considering hypotheses involving multiple unknown contributors. In this work, the graphical user interface (GUI) has been completely rewritten to replace the original GUI code (Tide-SDK), which had become deprecated. This will ensure that the program will continue to work into the reasonable future on current and future computer platforms.

contributors in the denominator. The original algorithm used a series of “for” loops to iterate through the genotypes of all possible pairs of contributors. As this is extremely computationally intensive, extending the program to handling more than 2 unknown

contributors would be difficult. However, we have overcome this technical challenge by replacing the nested "for" loops with a computationally efficient dynamic programming

the denominator within seconds. The details of this algorithm may be found in the Supplementary Notes of our publication on *Lab Retriever* [10]. We now report a number of additional improvements that have been incorporated. These include the ability to add and choose among any user-defined population database, the ability to choose a user-defined

of the LR to provide more information to the user and assist in research efforts. This work has increased knowledge regarding the interpretation of complex DNA profiles and added

analysts in accurately and efficiently assessing the weight of such evidence.

We have published detailed description of the methodology used in *Lab Retriever* in *BMC Bioinformatics* [10]. The program is freely available for download from the website (www.scieg.org) and the code, which follows GNU licensing, is available on GitHub (https://github.com/SCIEG/LabRetriever.)

3. Conclusions

At present, a number of software solutions are available to aid in the interpretation of challenging forensic DNA profiles. These programs use different summaries of the DNA profile data, make different modeling assumptions, require different computational resources, and have different levels of accessibility and transparency in licensing. To this end, we continue to develop and support *Lab Retriever*, a freely available open-source software solution to aid forensic analysts. *Lab Retriever* has a number of attractive features including an open source platform, a basis in published algorithms, flexibility in terms of the hypotheses that can be considered, nearly instantaneous run-time, and a friendly graphical user interface.

At present a number of laboratories are using *Lab Retriever* for casework. These labs include:

Denver Police Department
Jefferson County Regional Crime Laboratory
Cayman Islands Health Authority
Lake County Crime Laboratory
Genetic Technologies, Inc.
Bode Technology
Laboratorio Biologia e Genetica Forense

Additionally, Johnson County Sheriff's Office, and Anne Arundel Co Police Department are working on validating the program.

Thus, we believe that *Lab Retriever* is and will continue to be a useful tool for the forensic community.

Publications resulting from this grant:

Inman K, Rudin N, Cheng K, Robinson C, Kirschner A, Inman-Semerau L, Lohmueller KE. Lab Retriever: a software tool for calculating likelihood ratios incorporating a probability of drop-out for forensic DNA profiles, *BMC Bioinformatics*, 2015; 16: 298.

Haned H, Gill P, Lohmueller K, Inman K, Rudin N. Validation of probabilistic genotyping software for use in forensic DNA casework: Definitions and illustrations. *Sci Justice*, 2016; 56(2):104-8.

Marsden CD, Rudin N, Inman K, Lohmueller KE. An assessment of the information content of likelihood ratios derived from complex mixtures. *Forensic Sci Int Genet*, 2016; 22:64-72.

Publications in progress:

Inman K, Rudin N, Marsden CD, Lohmueller KE. An analysis of the effect of stutter on the interpretation of complex DNA profiles.

Marsden CD, Rudin N, Inman K, Lohmueller KE. Comprehensive assessment of the probability of drop-out in low-template DNA mixtures.

Rudin N, Inman K, Marsden CD, Lohmueller KE, An assessment of the risk of frank mistyping in low-template samples

Inman K, Rudin N, Lohmueller KE. Validation of empirical analytical thresholds using two different methods.

Inman K, Rudin N, Marsden CD, Lohmueller KE. Comparison of four different open-source probabilistic genotyping programs using a curated, publically-available dataset of complex samples.

Presentations:

Lab Retriever Training provided to:

Denver Police Department Crime Laboratory (CO)
Jefferson County Regional Crime Laboratory (CO)
NMS Laboratory (PA)
Westchester County Crime Laboratory (NY)
Santa Clara County Crime Laboratory (CA)
San Francisco City and County Crime Laboratory (CA)
Oregon State Police – Portland Metro Forensic Laboratory (OR)

Carabinieri Scientific Investigations Department (RIS), Rome, Italy
Cayman Islands Health Services Authority Forensic Science Laboratory
Allegheny County Office of the Medical Examiner (PA)
Johnson County Sheriff's Office Crime Laboratory (KS)

Presentations on the interpretation of challenging DNA profiles at national and international meetings:

Haned, H, Gill, P., Lohmueller, K., Inman, N., Rudin, N., Validation of probabilistic genotyping software for use in forensic DNA casework, AAFS, New Orleans, LA, February 2017.

Norah Rudin. A story of samples and statistics: The history of a forensic sample, the history and current state of forensic DNA interpretation and statistics in the U.S. Isaac Newton Institute for Mathematical Sciences, Cambridge, England, November 2016.

Keith Inman. A comparison of complex profiles analyzed with different software tools. Isaac Newton Institute for Mathematical Sciences, Cambridge, England, November 2016.

Clare Marsden, Norah Rudin, Keith Inman, Kirk Lohmueller, 2015. Defining the limits of forensic DNA profile interpretation: An assessment of the information content inherent in complex mixtures., NIST International Symposium on Forensic Science Error Management, Washington D.C.
http://scieg.org/uploads/Inman_AAFS_Presentation_Final.pdf

Clare Marsden, Norah Rudin, Keith Inman*, Kirk Lohmueller. Defining the limits of forensic DNA profile interpretation: An assessment of the information content inherent in complex mixtures. Platform presentation at the AAFS meeting, Orlando, FL, February 2015. (*Presenter)
http://scieg.org/uploads/Inman_AAFS_Presentation_Final.pdf

Clare Marsden*, Norah Rudin, Keith Inman, Kirk Lohmueller. Defining the limits of forensic DNA profile interpretation: An assessment of the information content inherent in complex mixtures. Platform presentation at the California Association of Criminalists, Rohnert Park, CA October 2014. (*Presenter)

Keith Inman*, Norah Rudin*, Kirk E Lohmueller*. Probabilistic software workshop: Lab Retriever. Workshop presenter at the 25th International Symposium on Human Identification, Phoenix, AZ, September 2014. (*Presenter)
(http://scieg.org/uploads/ISHI_2014_Lab_Retriever_FINAL.pdf)

Keith Inman. A practical solution to training U.S. forensic DNA practitioners on implementing probabilistic approaches to weighting forensic DNA evidence. Platform presentation at the International Conference on Forensic Inference and Statistics, the Netherlands, August 2014.
(http://scieg.org/uploads/ICFIS_Inman_2014_Posting.pdf)

Norah Rudin*, Keith Inman, Kirk E Lohmueller. Lab Retriever: A software tool to estimate the quantitative evidential value of complex DNA profiles. Platform presentation at the International Conference on Forensic Inference and Statistics, the Netherlands, August 2014. (*Presenter)
(http://scieg.org/uploads/ICFIS_NR_2014_for_scieg_web_site.pdf)

Keith Inman*, Norah Rudin*, Kirk E Lohmueller*. Calculating Likelihood Ratios Incorporating a Probability of Drop-out using the free program Lab Retriever. Workshop at the 24th International Symposium on Human Identification, Atlanta, GA, October 2013. (*Presenter)
(http://scieg.org/uploads/Promega_2013.pdf)

Select presentations at regional forensic science meetings and workshops:

1) California Association of Criminalists, May, 2014
2) Mid-Atlantic Association of Forensic Science, May 2014
3) North-Eastern Association of Forensic Science, May 2014
4) A special week-long workshop organized by the Midwestern Association of Forensic Science, June 2014
5) Southern Association of Forensic Science, September 2014
6) Association of Forensic DNA Administrators and Analysts, January 2016, July 15, 2016

South-Western Association of Forensic Science, September 2016

References:

[1] K.E. Lohmueller, N. Rudin, Calculating the weight of evidence in low-template forensic DNA casework, J. Forensic Sci. 58 Suppl 1 (2013) S243–249. doi:10.1111/1556-4029.12017.

[2] M.W. Perlin, A. Sinelnikov, An information gap in DNA evidence interpretation, PloS One. 4 (2009) e8327. doi:10.1371/journal.pone.0008327;

[3] A.A. Mitchell, J. Tamariz, K. O'Connell, N. Ducasse, Z. Budimlija, M. Prinz, T. Caragine, Validation of a DNA mixture statistics tool incorporating allelic drop-out and drop-in, Forensic Sci. Int. Genet. 6 (2012) 749–761. doi:10.1016/j.fsigen.2012.08.007;

[4] D. Taylor, J.-A. Bright, J. Buckleton, The interpretation of single source and mixed DNA profiles, Forensic Sci. Int. Genet. 7 (2013) 516–528. doi:10.1016/j.fsigen.2013.05.011.

[5] D.J. Balding, J. Buckleton, Interpreting low template DNA profiles, Forensic Sci. Int. Genet. 4 (2009) 1–10. doi:10.1016/j.fsigen.2009.03.003.

[6] D.J. Balding, Evaluation of mixed-source, low-template DNA profiles in forensic science, Proc. Natl. Acad. Sci. U. S. A. (2013). doi:10.1073/pnas.1219739110.

[7] P. Gill, H. Haned, A new methodological framework to interpret complex DNA profiles using likelihood ratios, Forensic Sci. Int. Genet. 7 (2013) 251–263. doi:10.1016/j.fsigen.2012.11.002.

[8] R. Puch-Solis, L. Rodgers, A. Mazumder, S. Pope, I. Evett, J. Curran, D. Balding, Evaluating forensic DNA profiles using peak heights, allowing for multiple donors, allelic dropout and stutters, Forensic Sci. Int. Genet. 7 (2013) 555–563. doi:10.1016/j.fsigen.2013.05.009.

[9] C.D. Steele, D.J. Balding, Statistical Evaluation of Forensic DNA Profile Evidence, Annu. Rev. Stat. Its Appl. 1 (2014) 361–384. doi:10.1146/annurev-statistics-022513-115602.
[10] K. Inman, N. Rudin, K. Cheng, C. Robinson, A. Kirschner, L. Inman-Semerau, K.E. Lohmueller, Lab Retriever: a software tool for calculating likelihood ratios incorporating a probability of drop-out for forensic DNA profiles, BMC Bioinformatics. 16 (2015) 298. doi:10.1186/s12859-015-0740-8.
[11] R.G. Cowell, T. Graversen, S.L. Lauritzen, J. Mortera, Analysis of forensic DNA mixtures with artefacts, J. R. Stat. Soc. Ser. C Appl. Stat. 64 (2015) 1–48. doi:10.1111/rssc.12071.
[12] Ø. Bleka, G. Storvik, P. Gill, EuroForMix: An open source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts, Forensic Sci. Int. Genet. 21 (2016) 35–44. doi:10.1016/j.fsigen.2015.11.008.
[13] H. Swaminathan, A. Garg, C.M. Grgicak, M. Medard, D.S. Lun, CEESIt: A computational tool for the interpretation of STR mixtures, Forensic Sci. Int. Genet. 22 (2016) 149–160. doi:10.1016/j.fsigen.2016.02.005.
[14] T.W. Bille, S.M. Weitz, M.D. Coble, J. Buckleton, J.-A. Bright, Comparison of the performance of different models for the interpretation of low level mixed DNA profiles, Electrophoresis. 35 (2014) 3125–3133. doi:10.1002/elps.201400110.
[15] J.-A. Bright, I.W. Evett, D. Taylor, J.M. Curran, J. Buckleton, A series of recommended tests when validating probabilistic DNA profile interpretation software, Forensic Sci. Int. Genet. 14C (2014) 125–131. doi:10.1016/j.fsigen.2014.09.019.
[16] H. Kelly, J.-A. Bright, J.S. Buckleton, J.M. Curran, A comparison of statistical models for the analysis of complex forensic DNA profiles, Sci. Justice. 54 (2014) 66–70. doi:10.1016/j.scijus.2013.07.003.
[17] L. Prieto, H. Haned, A. Mosquera, M. Crespillo, M. Alemañ, M. Aler, F. Alvarez, C. Baeza-Richer, A. Dominguez, C. Doutremepuich, M.J. Farfán, M. Fenger-Grøn, J.M. García-Ganivet, E. González-Moya, L. Hombreiro, M.V. Lareu, B. Martínez-Jarreta, S. Merigioli, P. Milans Del Bosch, N. Morling, M. Muñoz-Nieto, E. Ortega-González, S. Pedrosa, R. Pérez, C. Solís, I. Yurrebaso, P. Gill, Euroforgen-NoE collaborative exercise on LRmix to demonstrate standardization of the interpretation of complex DNA profiles, Forensic Sci. Int. Genet. 9 (2014) 47–54. doi:10.1016/j.fsigen.2013.10.011.
[18] R. Puch-Solis, T. Clayton, Evidential evaluation of DNA profiles using a discrete statistical model implemented in the DNA LiRa software, Forensic Sci. Int. Genet. 11 (2014) 220–228. doi:10.1016/j.fsigen.2014.04.005.
[19] D. Taylor, Using continuous DNA interpretation methods to revisit likelihood ratio behaviour, Forensic Sci. Int. Genet. 11 (2014) 144–153. doi:10.1016/j.fsigen.2014.03.008.
[20] D. Taylor, J.-A. Bright, J. Buckleton, J. Curran, An illustration of the effect of various sources of uncertainty on DNA likelihood ratio calculations, Forensic Sci. Int. Genet. 11 (2014) 56–63. doi:10.1016/j.fsigen.2014.02.003.
[21] S.A. Greenspoon, L. Schiermeier-Wood, B.C. Jenkins, Establishing the limits of TrueAllele® Casework: A validation study, J. Forensic Sci. 60 (2015) 1263–1276. doi:10.1111/1556-4029.12810.
[22] M.W. Perlin, K. Dormer, J. Hornyak, L. Schiermeier-Wood, S. Greenspoon, TrueAllele Casework on Virginia DNA mixture evidence: Computer and manual interpretation in

72 reported criminal cases, PLoS ONE. 9 (2014) e92837. doi:10.1371/journal.pone.0092837.
[23] M.W. Perlin, J.M. Hornyak, G. Sugimoto, K.W.P. Miller, TrueAllele® genotype identification on DNA mixtures containing up to five unknown contributors, J. Forensic Sci. 60 (2015) 857–868. doi:10.1111/1556-4029.12788.
[24] H. Haned, C.C.G. Benschop, P.D. Gill, T. Sijen, Complex DNA mixture analysis in a forensic context: evaluating the probative value using a likelihood ratio model, Forensic Sci. Int. Genet. 16 (2015) 17–25. doi:10.1016/j.fsigen.2014.11.014.
[25] C.D. Steele, M. Greenhalgh, D.J. Balding, Verifying likelihoods for low template DNA profiles using multiple replicates, Forensic Sci. Int. Genet. 13 (2014) 82–89. doi:10.1016/j.fsigen.2014.06.018.
[26] D. Taylor, J. Buckleton, I. Evett, Testing likelihood ratios produced from complex DNA profiles, Forensic Sci. Int. Genet. 16 (2015) 165–171. doi:10.1016/j.fsigen.2015.01.008.
[27] D. Taylor, J. Buckleton, Do low template DNA profiles have useful quantitative data?, Forensic Sci. Int. Genet. 16 (2015) 13–16. doi:10.1016/j.fsigen.2014.11.001.
[28] Ø. Bleka, C.C.G. Benschop, G. Storvik, P. Gill, A comparative study of qualitative and quantitative models used to interpret complex STR DNA profiles, Forensic Sci. Int. Genet. 25 (2016) 85–96. doi:10.1016/j.fsigen.2016.07.016.
[29] M.D. Coble, J. Buckleton, J.M. Butler, T. Egeland, R. Fimmers, P. Gill, L. Gusmão, B. Guttman, M. Krawczak, N. Morling, W. Parson, N. Pinto, P.M. Schneider, S.T. Sherry, S. Willuweit, M. Prinz, DNA Commission of the International Society for Forensic Genetics: Recommendations on the validation of software programs performing biostatistical calculations for forensic genetics applications, Forensic Sci. Int. Genet. 25 (2016) 191–197. doi:10.1016/j.fsigen.2016.09.002.
[30] H. Haned, P. Gill, K. Lohmueller, K. Inman, N. Rudin, Validation of probabilistic genotyping software for use in forensic DNA casework: Definitions and illustrations, Sci. Justice J. Forensic Sci. Soc. 56 (2016) 104–108. doi:10.1016/j.scijus.2015.11.007.
[31] C.D. Marsden, N. Rudin, K. Inman, K.E. Lohmueller, An assessment of the information content of likelihood ratios derived from complex mixtures, Forensic Sci. Int. Genet. 22 (2016) 64–72. doi:10.1016/j.fsigen.2016.01.008.
[32] President's Council of Advisors on Science and Technology (U.S.), & United States. Executive Office of the President, Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods, (2016). https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf.
[33] Scientific Working Group on DNA Analysis Methods, Guidelines for the validation of probabilistic genotyping systems, (2015). http://media.wix.com/ugd/4344b0_22776006b67c4a32a5ffc04fe3b56515.pdf.
[34] T. Tvedebrink, P.S. Eriksen, H.S. Mogensen, N. Morling, Estimating the probability of allelic drop-out of STR alleles in forensic genetics, Forensic Sci. Int. Genet. 3 (2009) 222–226. doi:10.1016/j.fsigen.2009.02.002.
[35] T. Tvedebrink, P.S. Eriksen, M. Asplund, H.S. Mogensen, N. Morling, Allelic drop-out probabilities estimated by logistic regression--further considerations and practical implementation, Forensic Sci. Int. Genet. 6 (2012) 263–267. doi:10.1016/j.fsigen.2011.06.004;

[36] J. Buckleton, J. Kelly, J.A. Bright, D. Taylor, T. Tvedebrink, J.M. Curran, Utilising allelic dropout probabilities estimated by logistic regression in casework, Forensic Sci Int Genet. 9 (2014) 9-11.
[37] R. Hedell, C. Dufva, R. Ansell, P. Mostad, J. Hedman, Enhanced low-template DNA analysis conditions and investigation of allele dropout patterns, Forensic Sci. Int. Genet. 14 (2015) 61–75. doi:10.1016/j.fsigen.2014.09.008.
[38] M.D. Timken, S.B. Klein, M.R. Buoncristiani, Stochastic sampling effects in STR typing: Implications for analysis and interpretation, Forensic Sci. Int. Genet. 11 (2014) 195–204. doi:10.1016/j.fsigen.2014.03.015.
[39] K.E. Lohmueller, N. Rudin, K. Inman, Analysis of allelic drop-out using the Identifiler and PowerPlex 16 forensic STR typing systems, Forensic Sci Int Genet. 12C (2014) 1–11. doi:10.1016/j.fsigen.2014.04.003.
[40] J.-A. Bright, D. Taylor, J.M. Curran, J.S. Buckleton, Developing allelic and stutter peak height models for a continuous method of DNA interpretation, Forensic Sci. Int. Genet. 7 (2013) 296–304. doi:10.1016/j.fsigen.2012.11.013.
[41] J.-A. Bright, J.M. Curran, Investigation into stutter ratio variability between different laboratories, Forensic Sci. Int. Genet. 13 (2014) 79–81. doi:10.1016/j.fsigen.2014.07.003.
[42] S. Manabe, Y. Hamano, C. Morimoto, C. Kawai, S. Fujimoto, K. Tamaki, New stutter ratio distribution for DNA mixture interpretation based on a continuous model, Leg. Med. Tokyo Jpn. 19 (2016) 16–21. doi:10.1016/j.legalmed.2016.01.007.
[43] P. Gill, J. Whitaker, C. Flaxman, N. Brown, J. Buckleton, An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA, Forensic Sci. Int. 112 (2000) 17–40.
[44] J. Buckleton, A Framework for Interpreting Evidence, in: J. Buckleton, C.M. Triggs, S.J. Walsh (Eds.), Forensic DNA Evid. Interpret., CRC Press, Boa Raton, FL, 2005: pp. 27–63.
[45] I.W. Evett, B.S. Weir, Interpreting DNA Evidence: Statistical Genetics for Forensic Scientists, Sinauer Associates, Sunderland, MA, 1998.
[46] D.J. Balding, Weight-of-Evidence for Forensic DNA Profiles, John Wiley & Sons Ltd, West Sussex, England, 2005.
[47] B. Budowle, A.J. Onorato, T.F. Callaghan, A. Della Manna, A.M. Gross, R.A. Guerrieri, J.C. Luttman, D.L. McClure, Mixture interpretation: defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework, J. Forensic Sci. 54 (2009) 810–821. doi:10.1111/j.1556-4029.2009.01046.x.
[48] Scientific Working Group on DNA Analytical Methods, SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories, 2010. http://www.fbi.gov/about-us/lab/codis/swgdam.pdf (accessed January 1, 2010).
[49] I.W. Evett, C. Buffery, G. Willott, D. Stoney, A guide to interpreting single locus profiles of DNA mixtures in forensic cases, J. Forensic Sci. Soc. 31 (1991) 41–47.
[50] B.S. Weir, C.M. Triggs, L. Starling, L.I. Stowell, K.A. Walsh, J. Buckleton, Interpreting DNA mixtures, J. Forensic Sci. 42 (1997) 213–222.
[51] T. Wang, N. Xue, J.D. Birdwell, Least-square deconvolution: a framework for interpreting short tandem repeat mixtures, J. Forensic Sci. 51 (2006) 1284–1297. doi:10.1111/j.1556-4029.2006.00268.x.

[52] J. Buckleton, P. Gill, Low Copy Number, in: J. Buckleton, C.M. Triggs, S.J. Walsh (Eds.), Forensic DNA Evid. Interpret., CRC Press, Boa Raton, FL, 2005: pp. 275–297.
[53] P. Gill, A. Kirkham, J. Curran, LoComatioN: a software tool for the analysis of low copy number DNA profiles, Forensic Sci. Int. 166 (2007) 128–138. doi:10.1016/j.forsciint.2006.04.016.
[54] ASCLSD/LAB, Board of Directors Interpretations on DNA Testing Services, (2015). http://www.ascld-lab.org/board-of-directors-interpretations-on-dna-testing-services/ (accessed October 31, 2016).
[55] J.M. Butler, C.R. Hill, Scientific issues with analysis of low amounts of DNA, 13 (2010) May 2, 2010.
[56] M.W. Perlin, M.M. Legler, C.E. Spencer, J.L. Smith, W.P. Allan, J.L. Belrose, B.W. Duceman, Validating TrueAllele(R) DNA mixture interpretation, J. Forensic Sci. 56 (2011) 1430–1447. doi:10.1111/j.1556-4029.2011.01859.x;
[57] C.R. Hill, D.L. Duewer, M.C. Kline, M.D. Coble, J.M. Butler, U.S. population data for 29 autosomal STR loci, Forensic Sci. Int. Genet. 7 (2013) e82–e83. doi:10.1016/j.fsigen.2012.12.004.
[58] J.M. Butler, R. Schoske, P.M. Vallone, J.W. Redman, M.C. Kline, Allele frequencies for 15 autosomal STR loci on U.S. Caucasian, African American, and Hispanic populations, J. Forensic Sci. 48 (2003) 908–911.
[59] J. Buckleton, J. Curran, A discussion of the merits of random man not excluded and likelihood ratios, Forensic Sci. Int. Genet. 2 (2008) 343–348. doi:10.1016/j.fsigen.2008.05.005.
[60] P. Gill, R. Sparkes, C. Kimpton, Development of guidelines to designate alleles using an STR multiplex system, Forensic Sci. Int. 89 (1997) 185–197. doi:10.1016/S0379-0738(97)00131-X.
[61] P.S. Walsh, N.J. Fildes, R. Reynolds, Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA, Nucleic Acids Res. 24 (1996) 2807–2812.

Figures and tables

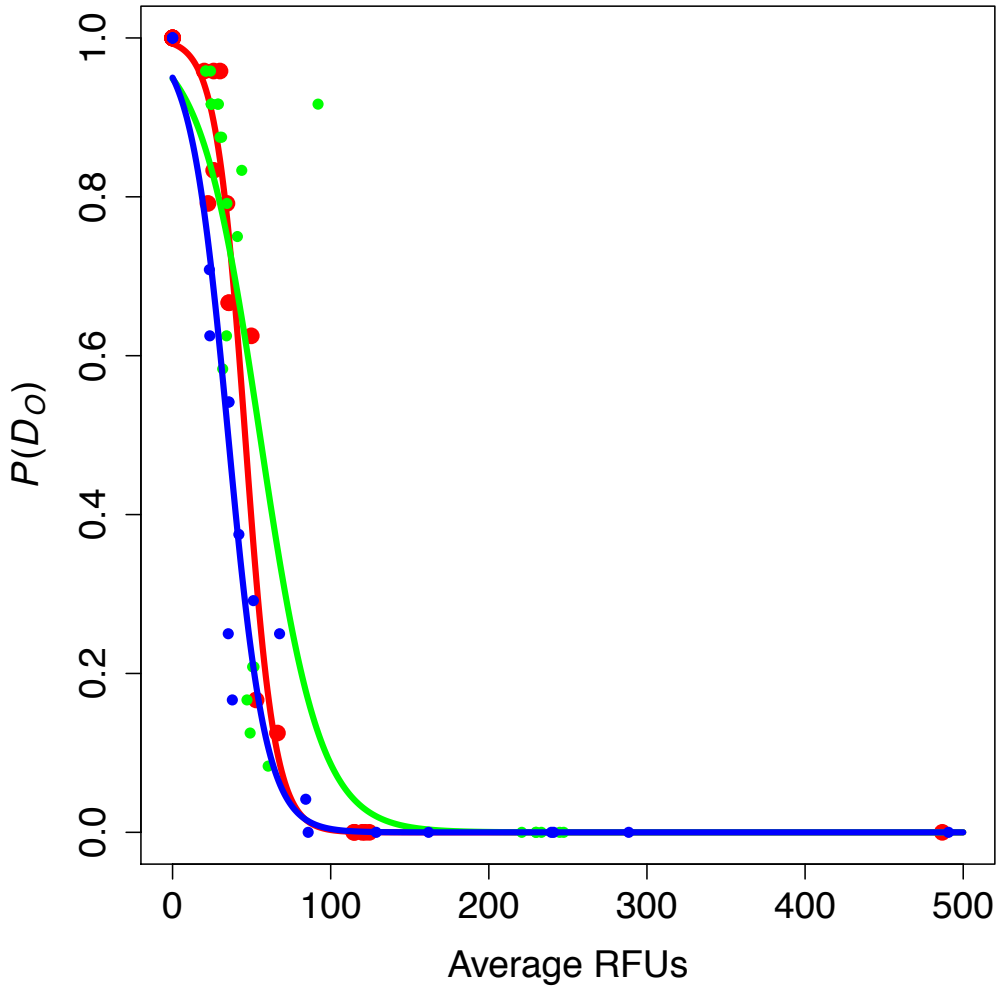

Figure 1: Probability of drop-out vs. average RFUs for the single-source (SS) profiles. Each color denotes one of the three different individuals used to generate the dilution series. Solid lines denote the logistic regression curves fit to the data (see Table 7 for the parameters of the model).

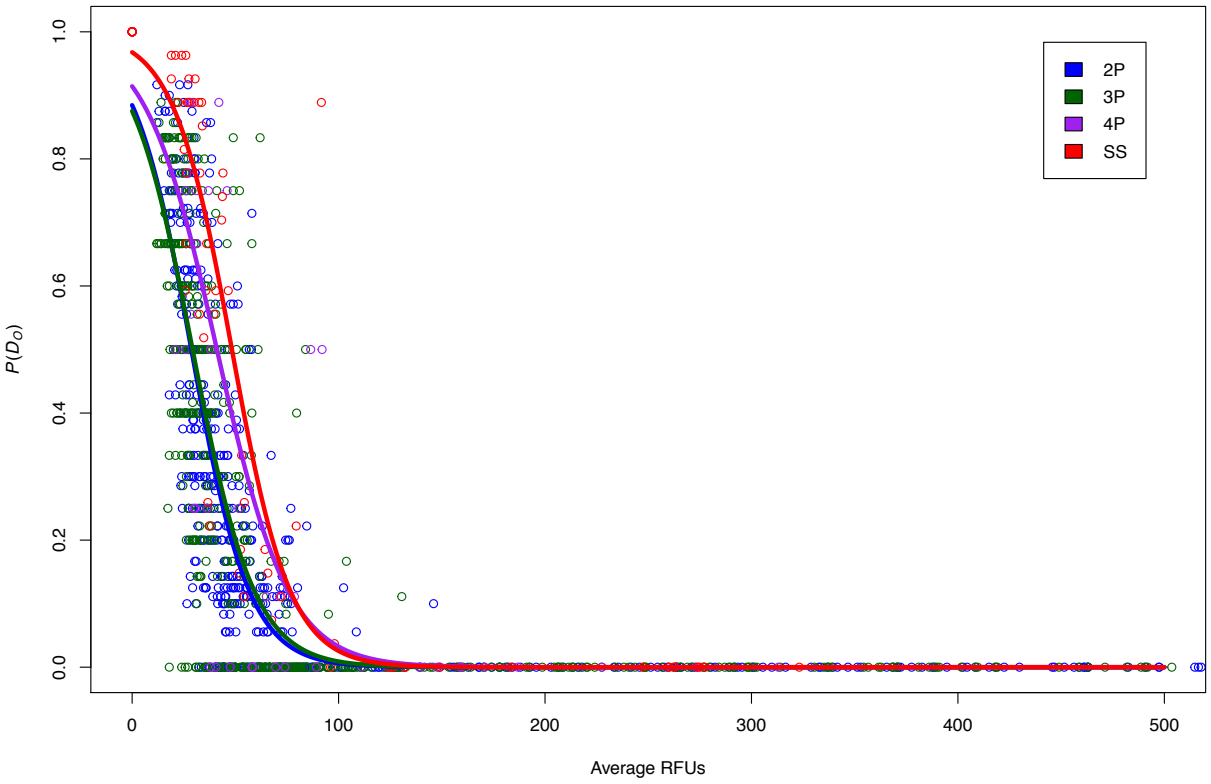

Figure 2: Probability of drop-out vs. average RFUs for profiles with different numbers of contributors. Colors denote the different number of contributors to each mixture. 2P denotes 2-person mixtures. 3P denotes 3-person mixtures. 4P denotes 4-person mixtures. SS denotes single-source samples. Solid lines denote the logistic regression curves fit to the data (see Table 7 for the parameters of the model).

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

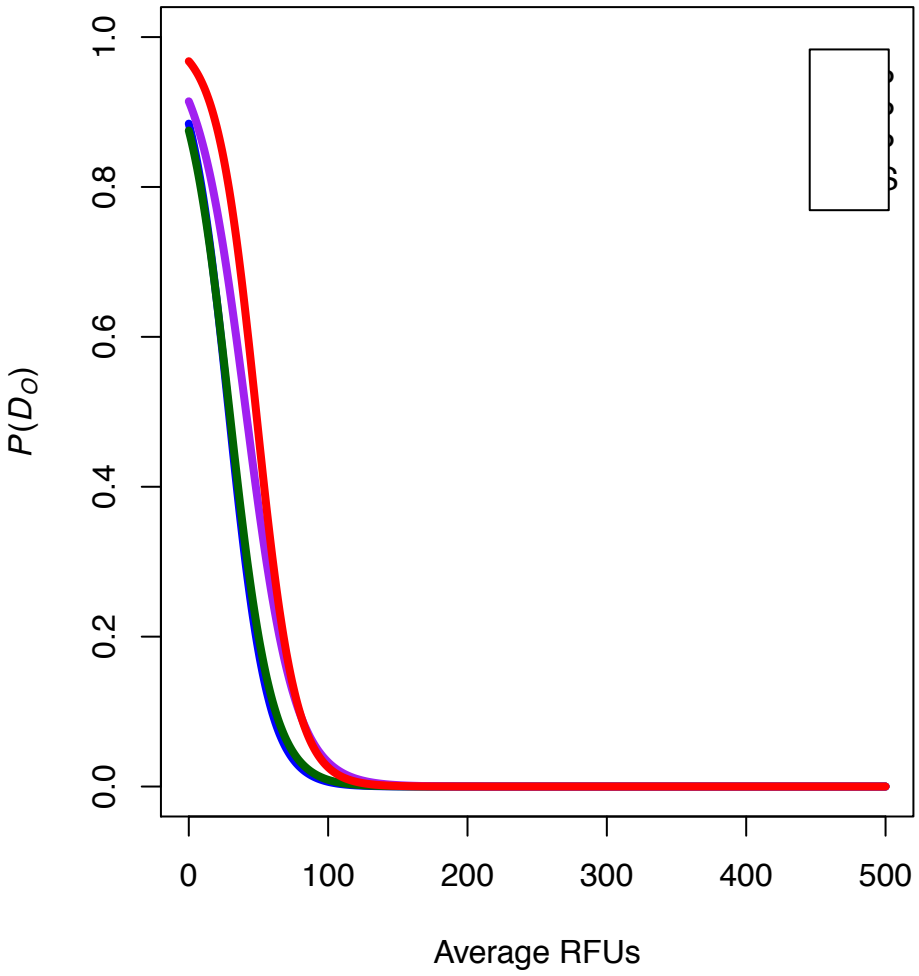

Figure 3: Logistic regression curves for profiles with different numbers of contributors. Colors denote the different number of contributors to each mixture. 2P denotes 2-person mixtures. 3P denotes 3-person mixtures. 4P denotes 4-person mixtures. SS denotes single-source samples. See Table 7 for the parameters.

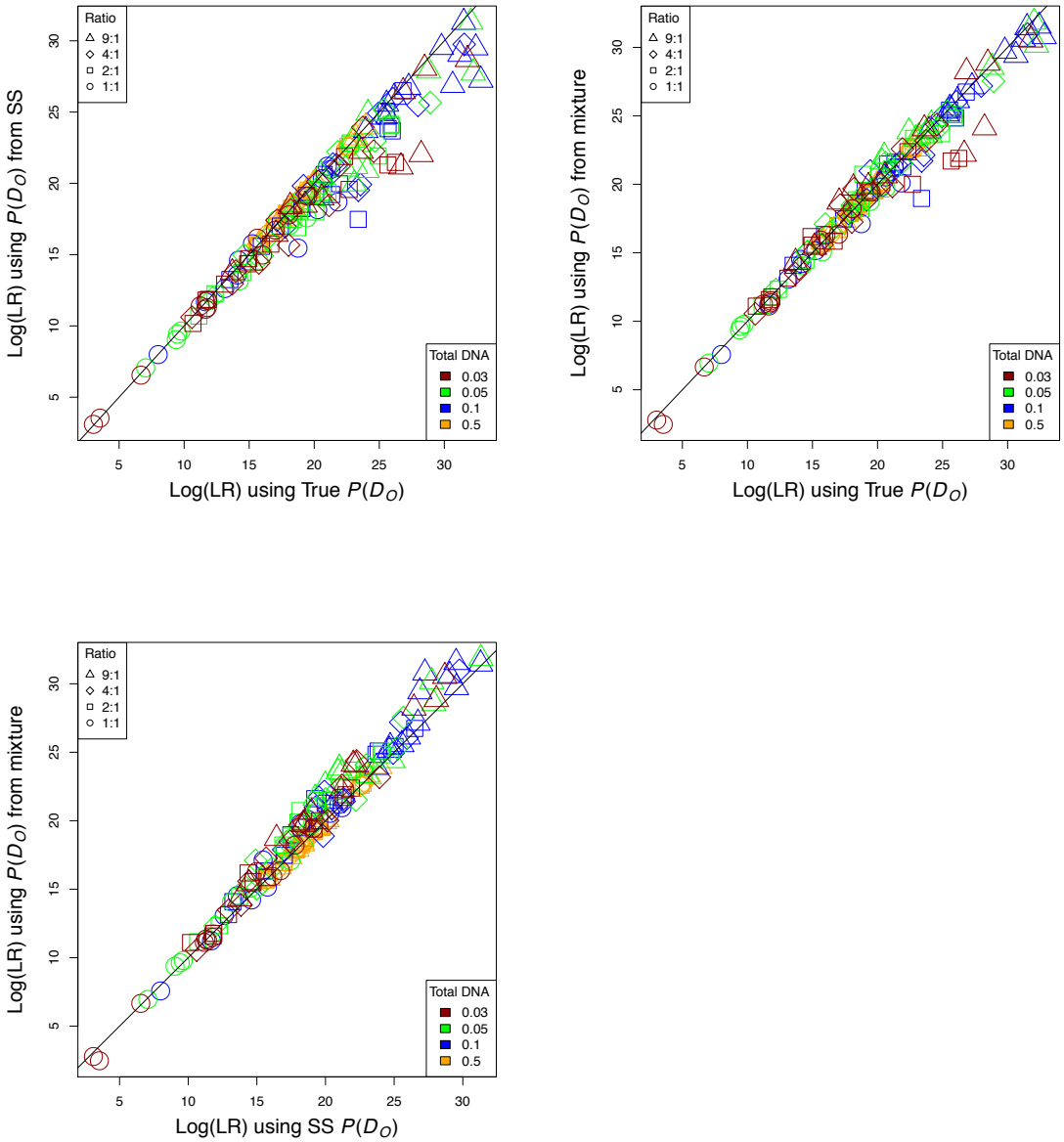

Figure 4: Comparison of the log(LRs) for 2-person mixtures where the suspected contributor is the true major contributor to the mixture. Log(LRs) were computed using 3 different ways to estimate $P(D_0)$. Each plot compares log(LRs) from two of these approaches. The solid line is the diagonal. Different colors denote different quantities of DNA and shapes denote different mixture ratios. SS denotes single-source.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

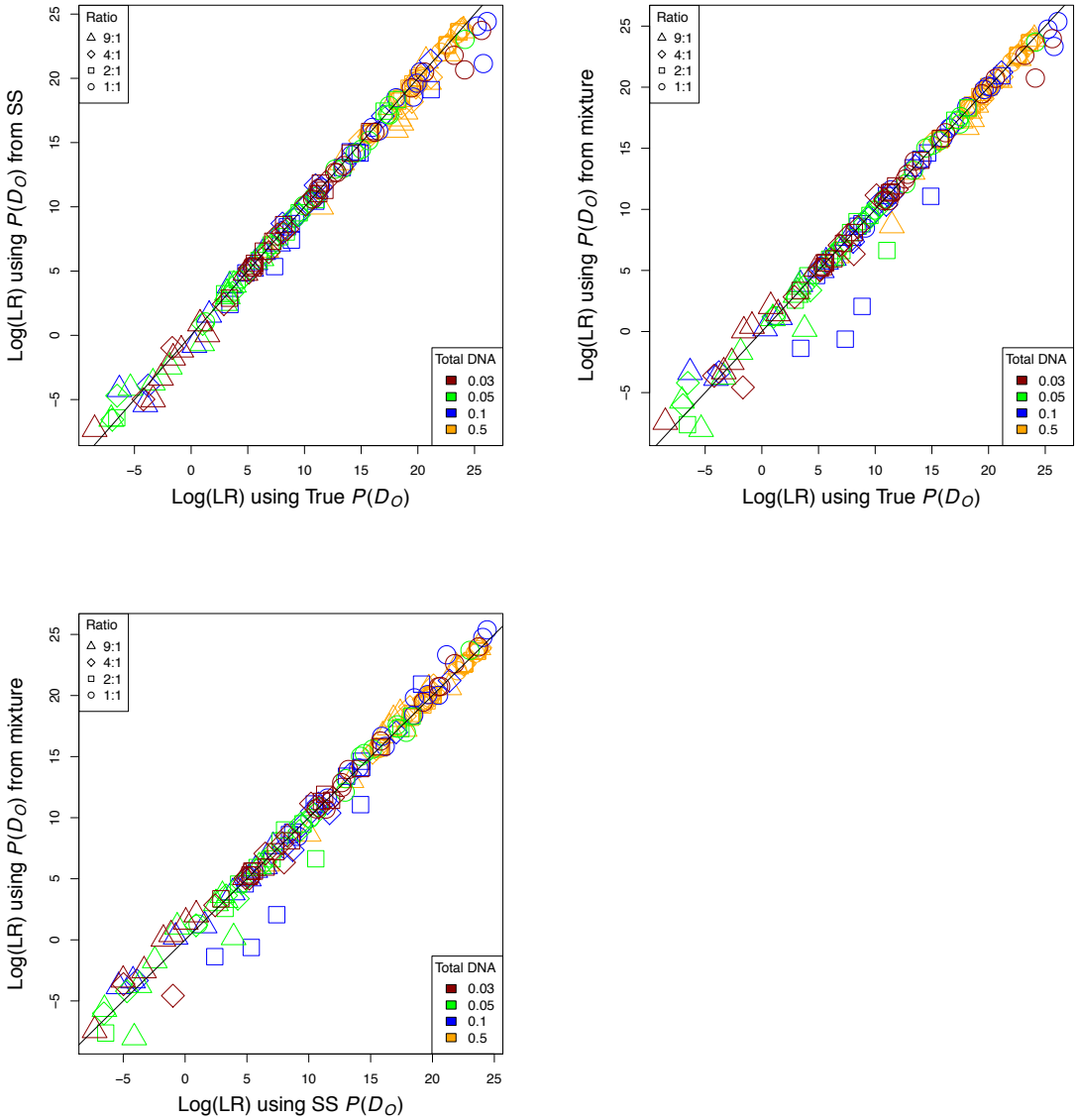

Figure 5: Comparison of the log(LRs) for 2-person mixtures where the suspected contributor is the true minor contributor to the mixture. Log(LRs) were computed using 3 different ways to estimate $P(D_0)$. Each plot compares log(LRs) from two of these approaches. The solid line is the diagonal. Different colors denote different quantities of DNA and shapes denote different mixture ratios. SS denotes single-source.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

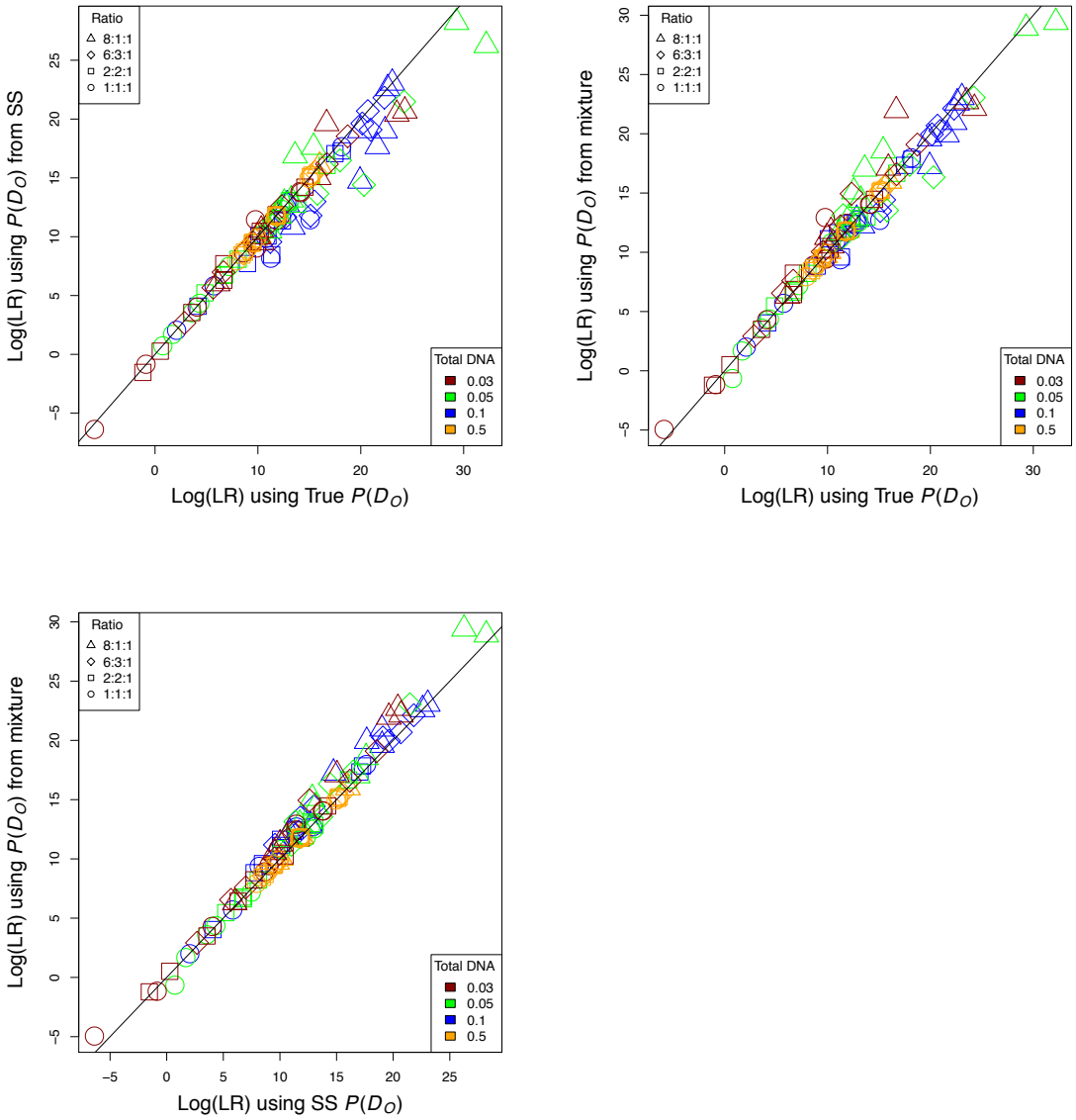

Figure 6: Comparison of the log(LRs) for 3-person mixtures where the suspected contributor is the true major contributor to the mixture. Log(LRs) were computed using 3 different ways to estimate $P(D_0)$. Each plot compares log(LRs) from two of these approaches. The solid line is the diagonal. Different colors denote different quantities of DNA and shapes denote different mixture ratios. SS denotes single-source.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

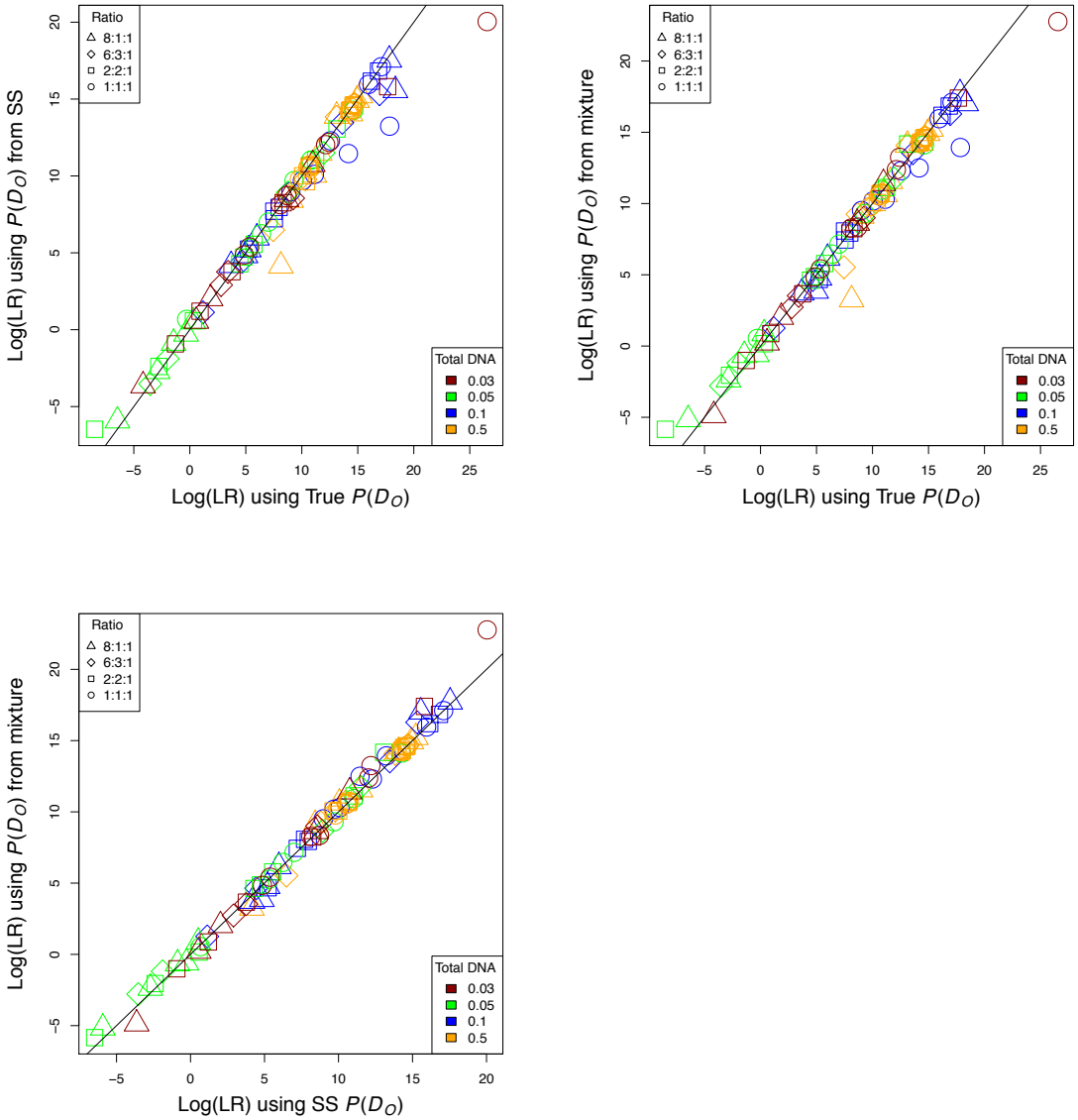

Figure 7: Comparison of the log(LRs) for 3-person mixtures where the suspected contributor is the true minor contributor to the mixture. Log(LRs) were computed using 3 different ways to estimate $P(D_0)$. Each plot compares log(LRs) from two of these approaches. The solid line is the diagonal. Different colors denote different quantities of DNA and shapes denote different mixture ratios. SS denotes single-source.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

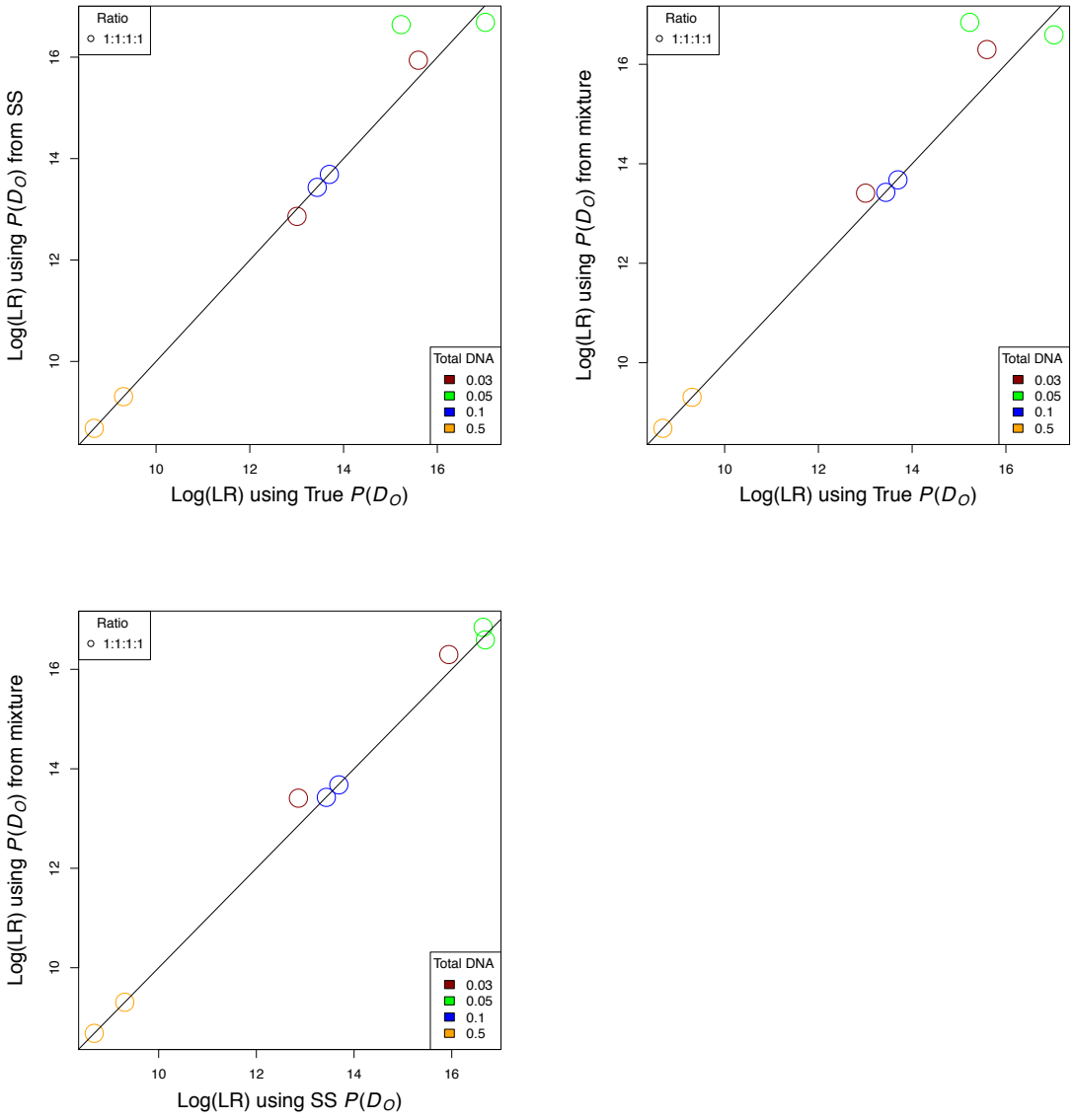

Figure 8: Comparison of the log(LRs) for 4-person mixtures where the suspected contributor is the true contributor to the mixture. Log(LRs) were computed using 3 different ways to estimate $P(D_0)$. Each plot compares log(LRs) from two of these approaches. The solid line is the diagonal. Different colors denote different quantities of DNA and shapes denote different mixture ratios. SS denotes single-source.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

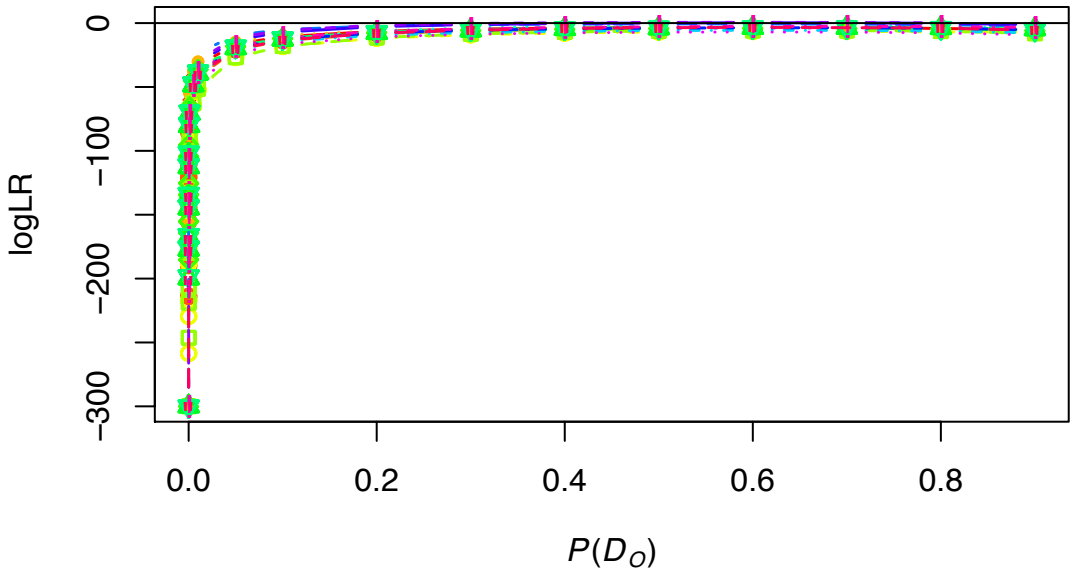

Figure 9: Log(LR) for a range of $P(D_O)$s where the suspected contributor is the true minor contributor for 17 2-person mixtures that gave a log(LR)<0. When the log(LR) was 0, it was set to -300 for plotting purposes. Note that for most of these mixtures, the log(LR) is <0 across the entire range of drop-out probabilities. Different colored curves denote the different samples.

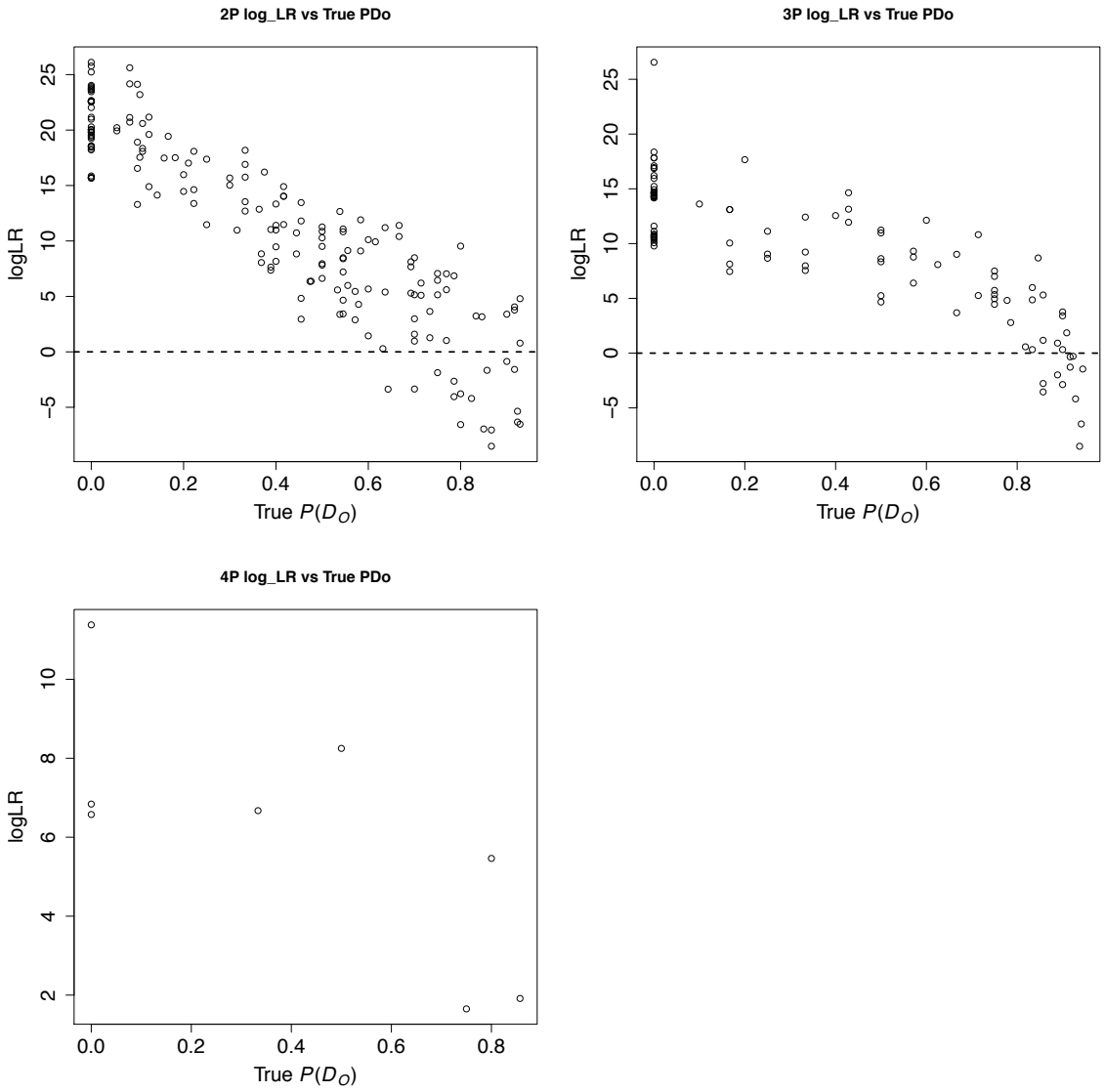

Figure 10: Log(LRs) decrease with increasing $P(D_O)$. Left: 2-person mixtures. Right: 3-person mixtures. Bottom: 4-person mixtures. The $P(D_O)$ on the x-axis is the true $P(D_O)$ for the contributor of interest. In all panels, the suspected contributor is the true minor contributor. The LRs compare the probability of the evidence assuming the suspected contributor and 1 (or more) unknown contributors vs. 2 (or more) unknown contributors. For mixtures with $P(D_O)$>0.6, there is a non-negligible probability of a log(LR)<0 for a true contributor.

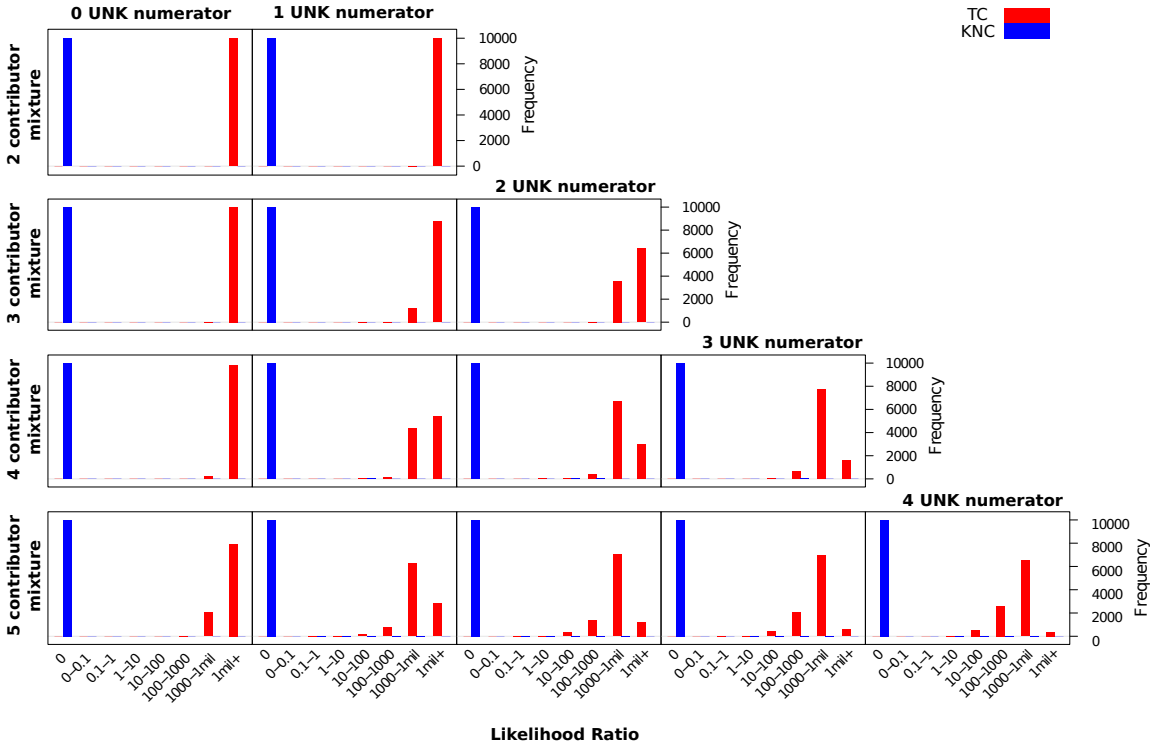

Figure 11: Distribution of LRs for simulated mixtures. TC denotes true contributor (red) while KNC denotes a known non-contributor (blue). Rows denote the total number of contributors in the mixture while columns denote the number of unknowns in the numerator. The denominator always contains one additional unknown contributor. Overall, note the good separation of LRs between TCs and KNCs.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

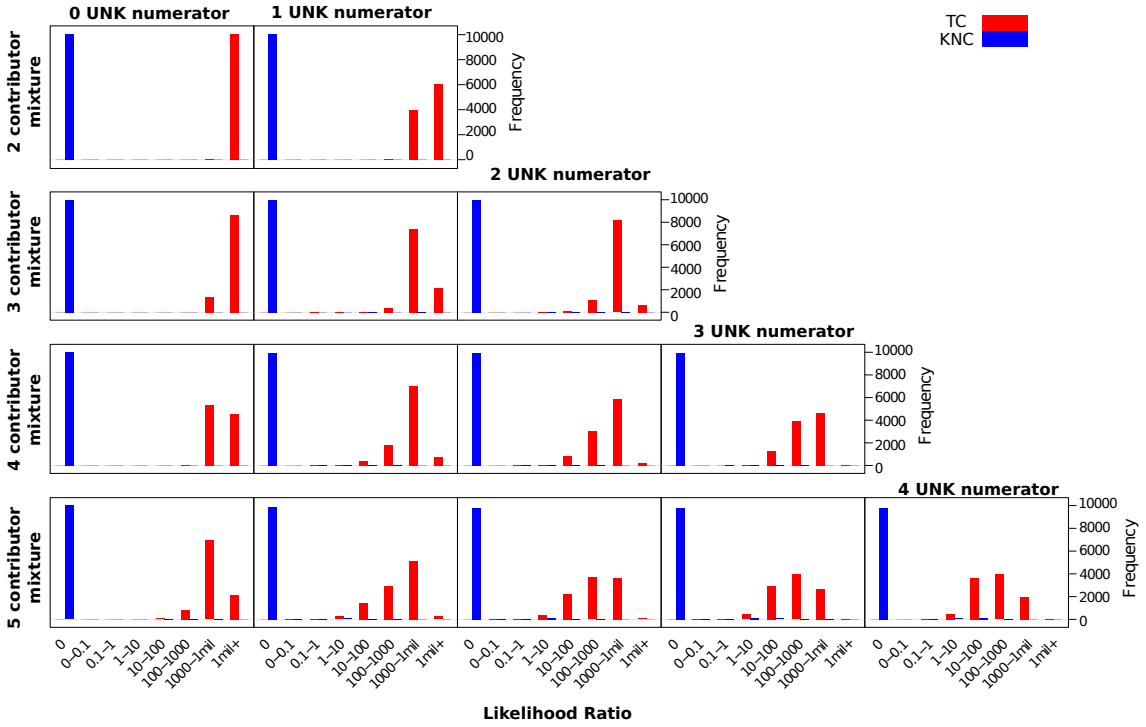

Figure 12: Distribution of LRs for simulated mixtures including only 9 loci. TC denotes true contributor (red) while KNC denotes a known non-contributor (blue). Rows denote the total number of contributors in the mixture while columns denote the number of unknowns in the numerator. The denominator always contains one addition unknown contributor. Overall, the separation of LRs between TCs and KNCs is not as good as with 15 loci.

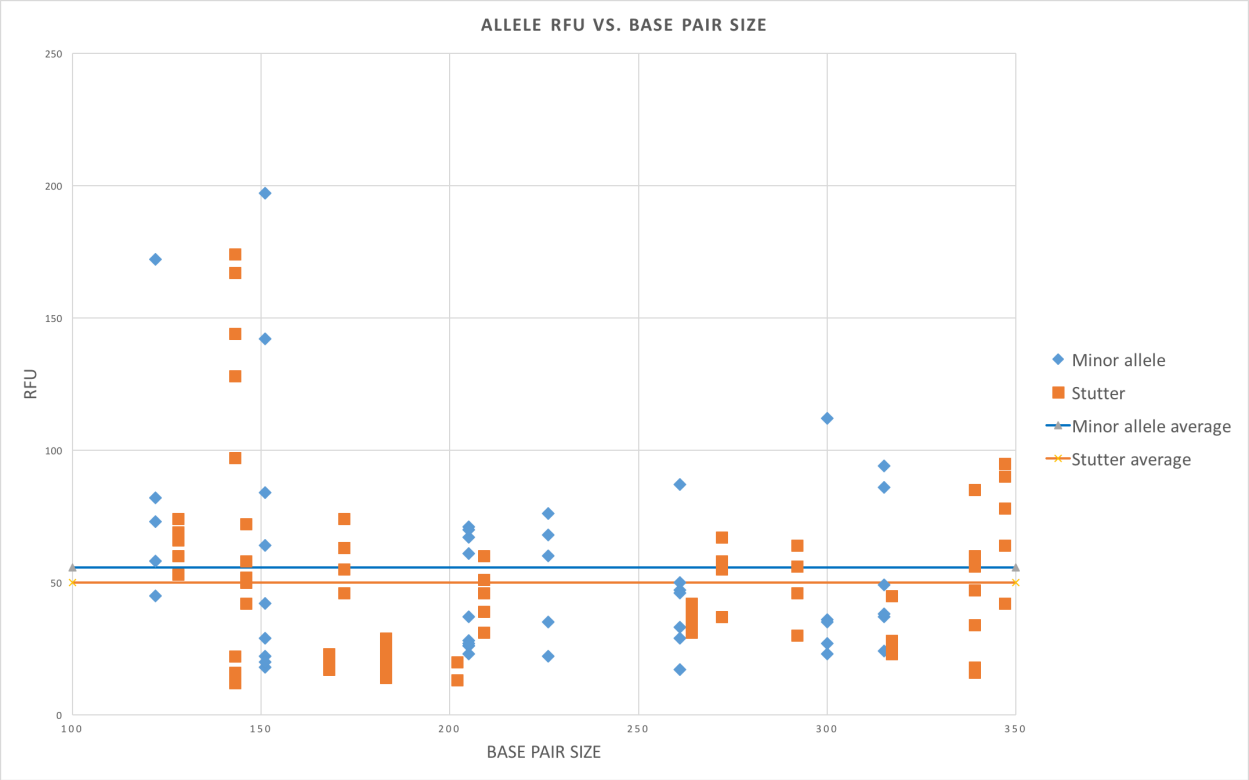

Figure 13: Peak heights of stutter peaks and minor allele peaks as a function of the size of the allele. Note that the points from stutter and minor alleles appear to have similar heights to each other.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

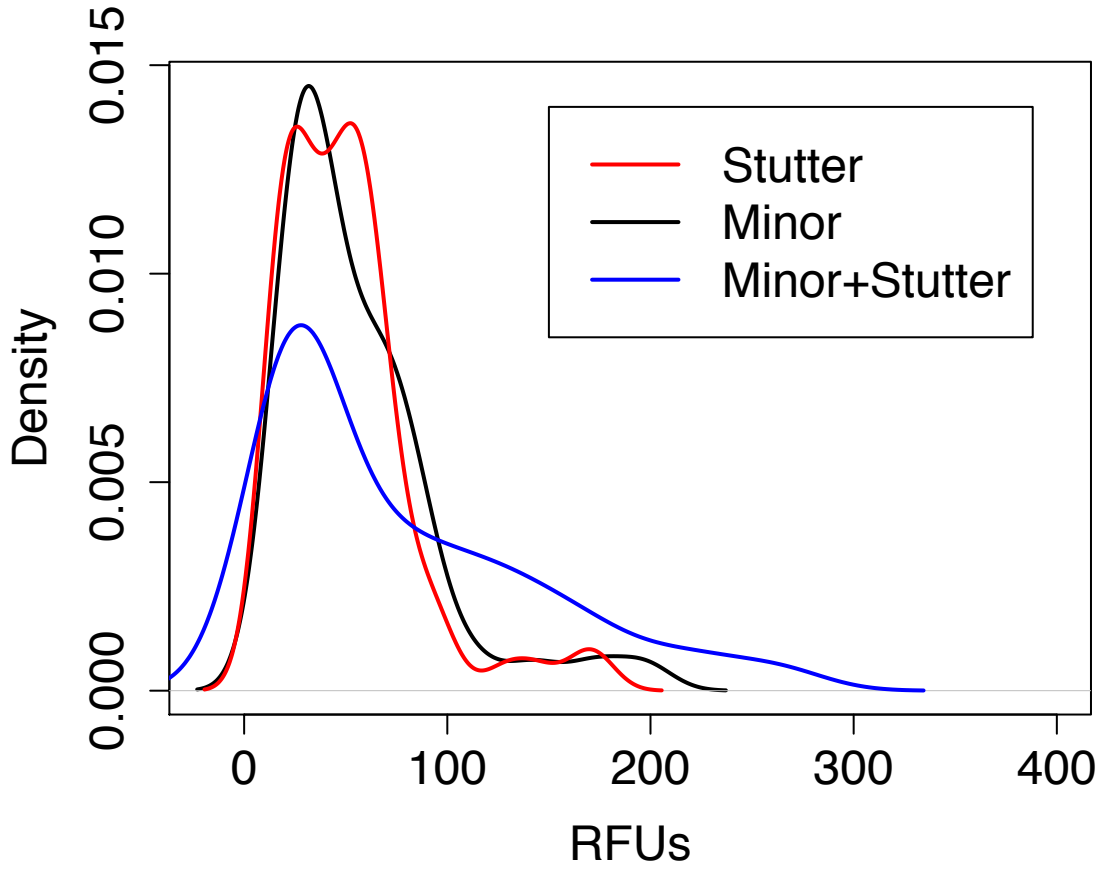

Figure 14: Distributions of the peak heights for stutter, minor, and [minor + stutter] peaks in the 9:1 2-person mixtures. A total of 74, 47, and 124 peaks are in the stutter, minor, and [minor + stutter] distributions, respectively.

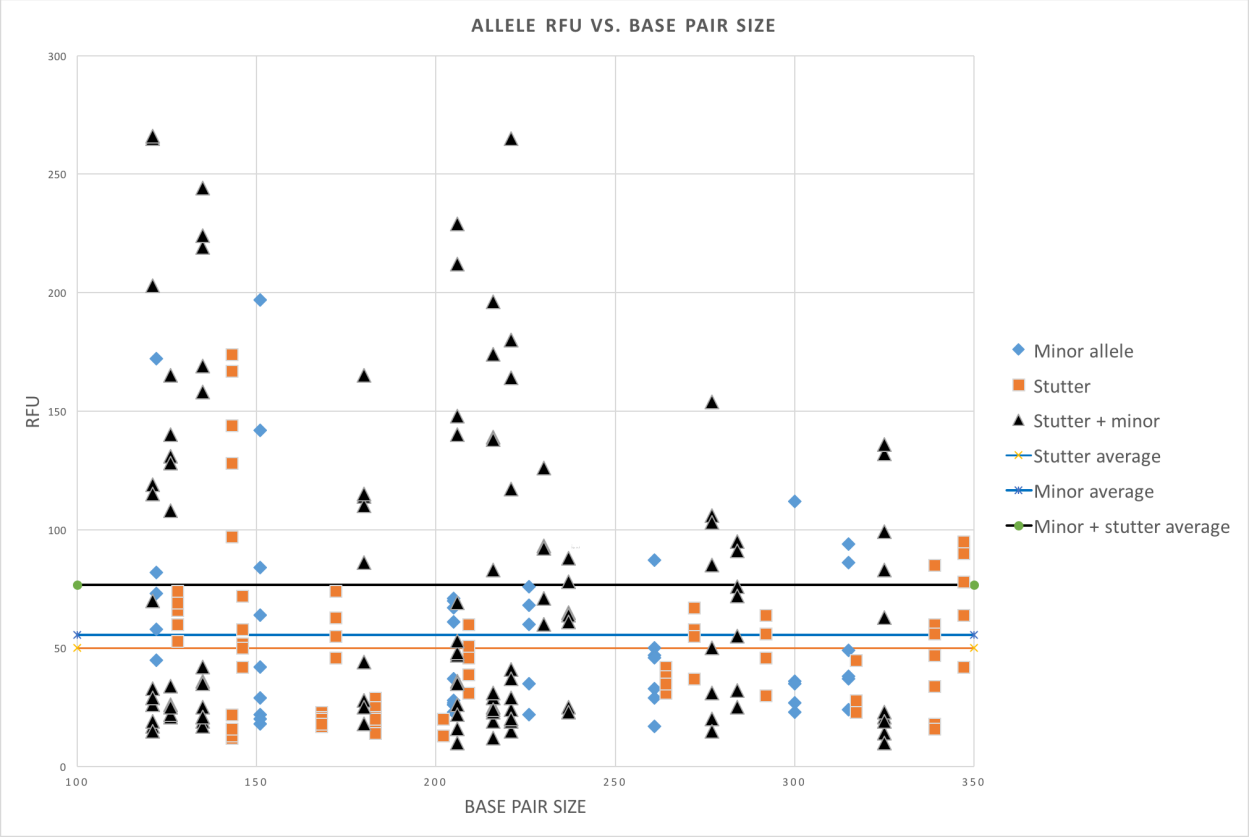

Figure 15: Peaks heights of positions where minor alleles overlapped with stutter peaks (black). Note that the positions where stutter and minor peaks overlap have a higher average height. However, the variance in RFU is much larger for these peaks compared to those with just stutter or minor alleles.

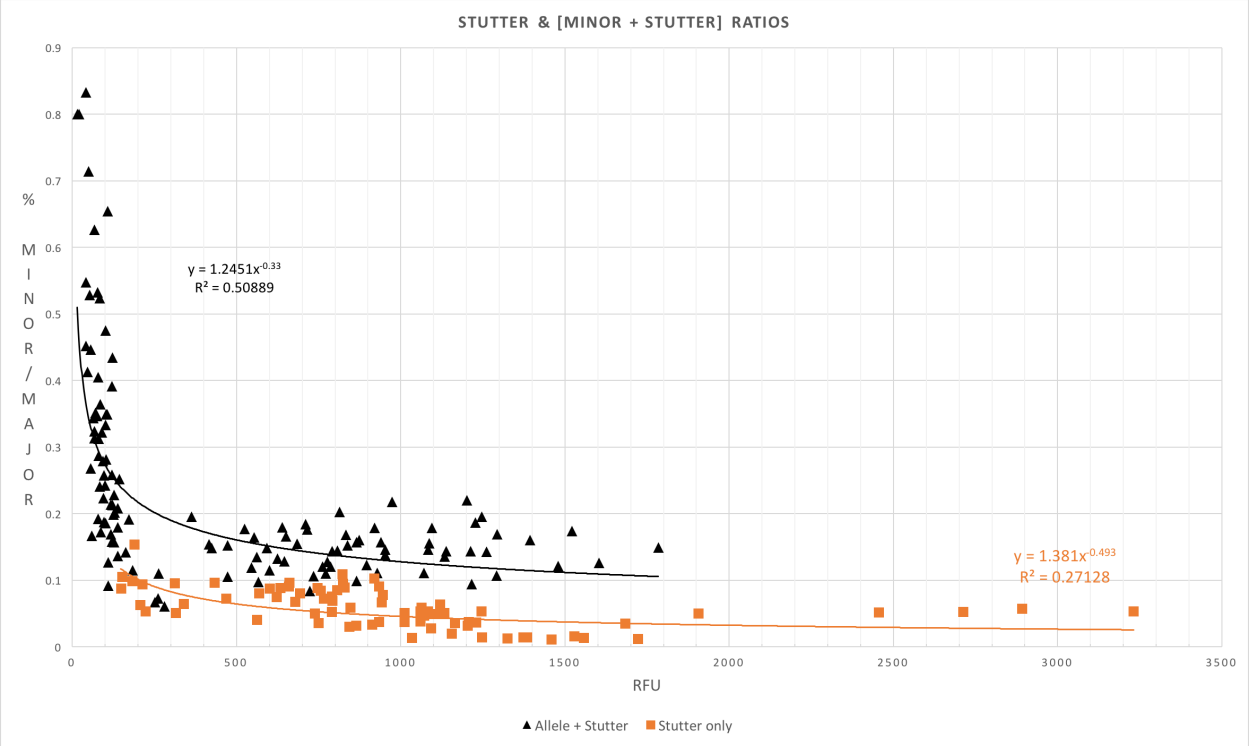

Figure 16: Peak height ratios of the stutter only and the [minor + stutter] peaks. Note the elevated peak height ratio for the [minor + stutter peaks]. However, the extent of this elevation is variable.

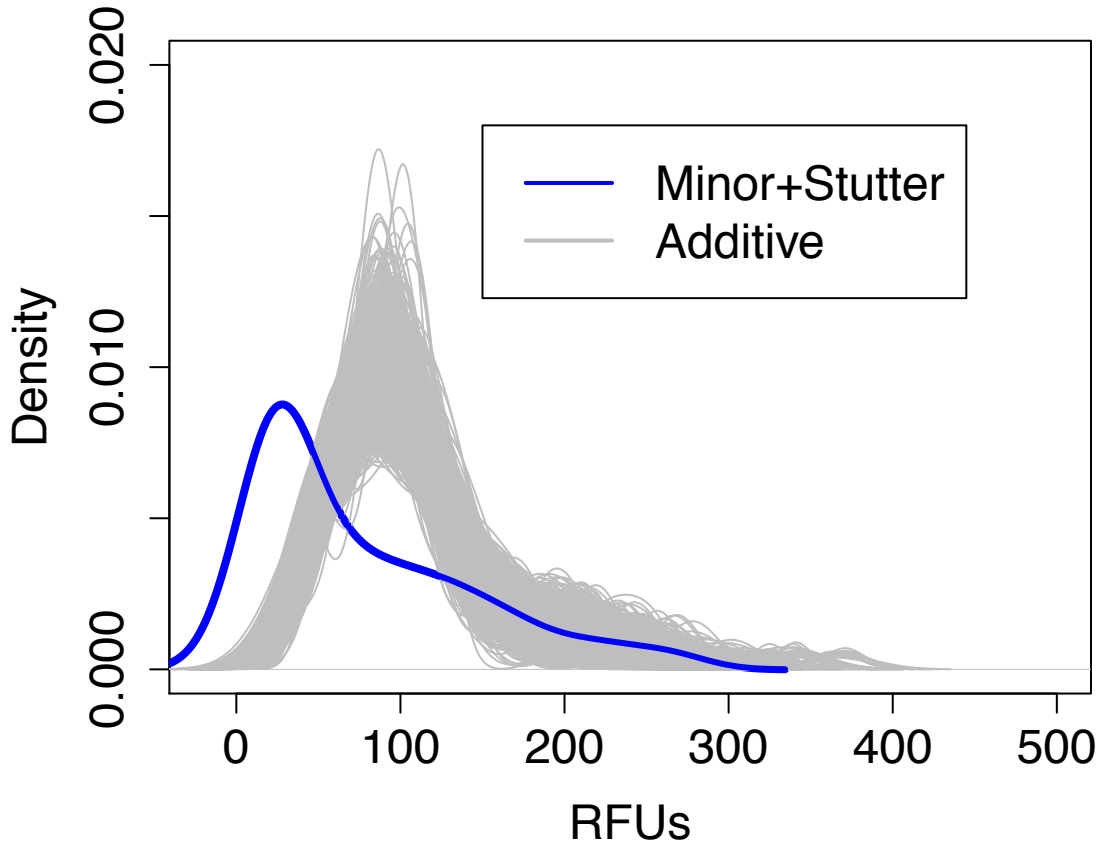

Figure 17: The distribution of [minor + stutter] peak heights is not consistent with the distribution of the sum of the minor peaks and the stutter peaks. Blue shows the empirical distribution of the [minor + stutter] peak heights. Each gray curve represents the distribution of the heights of 47 peaks, where each peak is the sum of a random minor allele and a random stutter peak. If the heights of [minor + stutter] peaks behaved in an additive manor, the blue distribution would fall within the range of the gray distributions.

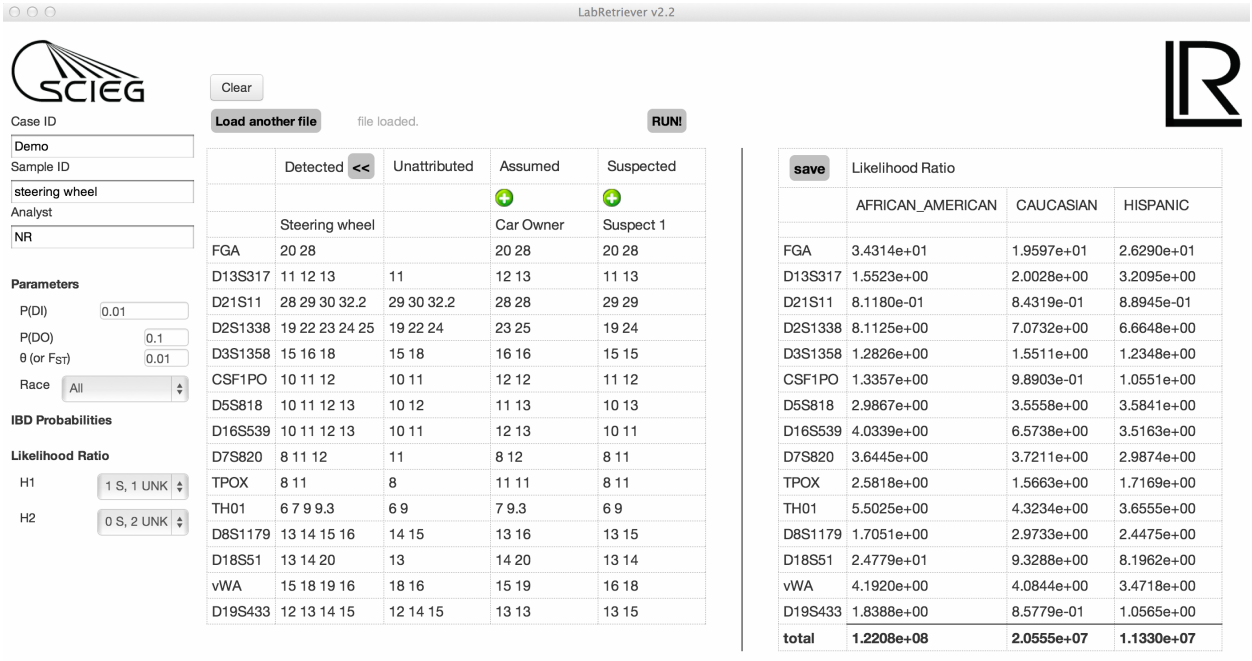

Figure 18: Graphical user interface for *Lab Retriever*. On the leftmost side, users have the opportunity to select the drop-in probability, drop-out probability, co-ancestry adjustment, population allele frequency group and the hypotheses in the LR. The middle section shows the evidence profile and the profile of the suspected contributors. The right section shows the LR calculated for each locus, as well as for the entire profile for three major US population groups.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table 1: Genotypes of the 8 individuals from which the mixtures were generated.

Sample	*A-365**	*B-483*	*C-497*	*D-555*	*E-681*	*F-788*	*G-805*	*H-985*
D8S1179	13,15	13	13,14	13	13,14	12,16	15	14,15
D21S11	28,30	26,29	31.2,32.2	30,32.2	29,31.2	28,30	29	30,30.2
D7S820	11	10,12	11	12,13	7,11	10,11	12	9,10
CSF1PO	10,12	11,12	10,11	10,12	10,11	10,12	11,12	10,11
D3S1358	14,15	14,15	15	15,18	14,17	17	16,19	15
TH01	6,9.3	9.3,7	6,7	6,9.3	9.3	6,9.3	9,10	7,8
D13S317	10,12	11,15	8,13	10,13	8,12	10,11	9,11	10,11
D16S539	9,12	12,13	11,13	9,11	9,11	9,12	9,12	11,14
D2S1338	20,21	16	22,25	24,26	17,24	18,20	18,24	17,20
D19S433	14	13,14	14,15	15,15.2	13,14	13,14	14.2,16.2	15,16.2
vWA	17,18	16,17	16	16,18	16,17	17	14,16	17
TPOX	8,10	8,10	8,11	8,12	8,11	8	9,11	8
D18S51	13,18	12	16	13,15	12,16	14,15	13,16	13,17
AMEL	X,Y	X,X	X,X	X,X	X,Y	X,Y	X,Y	X,X
D5S818	11,12	12	11	11,12	11	13,14	7,9	11,12
FGA	19,24	21,25	22,24	19,22	20,21	21,24	22,23	23,25

* - Letter = Used to code sample; Number = anonymization code

Table 2: Total DNA, mixture ratio, and individual DNA contribution for 2-person mixtures

2 person		
Total DNA (ng)	**Mix ratio**	**Major ng/minor ng**
0.5	1:1	0.25/0.25
	2:1	0.33/0.17
	4:1	0.4/0.1
	9:1	0.45/0.05
0.1	1:1	0.05/0.05
	2:1	0.067/0.033
	4:1	0.08/0.02
	9:1	0.09/0.01
0.05	1:1	0.025/0.025
	2:1	0.033/0.017
	4:1	0.04/0.01
	9:1	0.045/0.005
0.03	1:1	0.015/0.015
	2:1	0.02/0.01
	4:1	0.024/0.006
	9:1	0.027/0.003

For each 2-person mixture:	**Major donor/ minor donor**	
	Max # homozygous peaks /Max # heterzygous peaks	C-497 /E-681
	Max # heterozygous peaks/ Max # homozygous peaks	E-681 /C-497
	Max # homozygous peaks/ Max # homozygous peaks	C-497 /B-483
	Max # heterozygous peaks/ Max # heterozygous peaks	E-681 /D-555
	Maximizing # of peaks in stutter position for minor donor	E-681 /D-555
	Minimizing # of peaks in stutter position for minor donor	A-365 /D-555

Table 3: Total DNA, mixture ratio, and individual DNA contribution for 3-person mixtures

3 person		
Total DNA (ng)	**Mix ratio**	**Major ng/minor ng/minor ng**
0.5	1:1:1	0.165/0.165/0.165
	2:2:1	0.2/0.2/0.1
	6:3:1	0.3/0.15/0.05
	8:1:1	0.4/0.05/0.05
0.1	1:1:1	0.033/0.033/0.033
	2:2:1	0.04/0.04/0.02
	6:3:1	0.06/0.03/0.01
	8:1:1	0.08/0.01/0.01
0.03	1:1:1	0.0099/0.0099/0.0099
	2:2:1	0.012/0.012/0.006
	6:3:1	0.018/0.009/0.003
	8:1:1	0.024/0.003/0.003

For each 3-person mixture:	**Major donor //2 minor donors**	
	Max # homozygous peaks //Max # heterzygous peaks	C-497//D-555/E-681
	Max # heterozygous peaks //Max # homozygous peaks	E-681//C-497/B-483
	Max # homozygous peaks //Max # homozygous peaks	C-497//B-483/H-985
	Max # heterozygous peaks //Max # heterozygous peaks	D-555//E-681/A-365

Table 4: Total DNA, mixture ratio, and individual DNA contribution for 4-person mixtures

4 person		
Total DNA (ng)	**Mix ratio**	**Amounts of DNA (ng per donor)**
0.5	1:1:1:1	0.125/0.125/0.125/0.125
0.1	1:1:1:1	0.025/0.025/0.025/0.025/
0.05	1:1:1:1	0.0125/0.0125/0.0125/0.0125/
0.03	1:1:1:1	0.0075/0.0075/0.0075/0.0075/
	4 person samples	985/805/788/483

Table 5: Numbers of peaks in stutter positions between pairs of contributors

	A-365	*B-483*	*C-497*	*D-555*	*E-681*	*F-788*	*G-805*	*H-985*
A-365		8/9	5/8	***3/6***	9/5	6/4	6/7	10/8
B-483			8/8	6/8	9/6	9/6	4/6	6/8
C-497				5/6	9/3	9/3	3/9	3/12
D-555					***13/4***	7/5	6/8	3/8
E-681						3/9	3/9	3/9
F-788							7/6	10/6
G-805								9/4

Sample mixtures can be constructed in two ways:

1. *Samples in the top row as the major donor and samples in the first column as the minor;*
2. *Samples in the first column as the major donor and samples in the top row as the minor.*

The number to the left of the "/" denotes the number of peaks in the minor donor in stutter position to a peak from the major donor in condition 1. The number to the right of the "/" denotes the number of peaks in the minor donor in stutter position to a peak from the major donor in condition 2. For example, when sample B-483 is the major donor and A-365 is the minor, 8 peaks in A-365 are in stutter positions to peaks in B-483, while 9 peaks in B-483 are in stutter position when A-365 is the major and B-483 is the minor.

The samples indicated in bold/italic/underline were chosen to use in the stutter experiment.

Table 6: Run-specific analytical thresholds

Dye	Minimum	Maximum
Blue	10	20
Green	12	18
Yellow	16	24
Red	18	30

Table 7: Logistic regression parameter estimates from different types of samples

Sample	Intercept	Slope
Single-source	3.40412105	-0.07042581
2-person mixture	2.04937860	-0.06985159
3-person mixture	2.28445971	-0.06933856
4-person mixture	2.92274972	-0.06864481

Table 8: Distribution of the log(LRs) for empirical mixtures where the suspected contributor did not contribute to the mixture (i.e. is a known non-contributor).

Mixture	Number of comparisons	% log(LR)>0	% log(LR)>1	% log(LR)>2	Max log(LR)
2 person	1,810,000	0.0835	0.0143	0.0023	1403
3 person	1,000,000	0.3890	0.0536	0.0052	6516
4 person	80,000	0.9583	0.1288	0.0050	1290

Table 9: Details of hypotheses investigated when calculating LRs for different mixtures.

Hypothesis*	***Total # of contributors to the mixture***	***Contributors conditioned under H1 *****	***Contributors conditioned under H2***
h21	*2*	*C1, C2*	*C2 + 1 UNK*
h22	*2*	*C1*	*2 UNK*
h31	*3*	*C1, C2, C3*	*C2, C3 + 1 UNK*
h32	*3*	*C1, C2 + 1 UNK*	*C2 + 2 UNK*
h33	*3*	*C1 + 2 UNK*	*3 UNK*
h41	*4*	*C1, C2, C3, C4*	*C2, C3, C4 + 1 UNK*
h42	*4*	*C1, C2, C3 + 1 UNK*	*C2, C3 + 2 UNK*
h43	*4*	*C1, C2 + 2 UNK*	*C2 + 3 UNK*
h44	*4*	*C1 + 3 UNK*	*4 UNK*
h51	*5*	*C1, C2, C3, C4, C5*	*C2, C3, C4, C5 + 1 UNK*
h52	*5*	*C1, C2, C3, C4 + 1 UNK*	*C2, C3, C4 + 2 UNK*
h53	*5*	*C1, C2, C3 + 2 UNK*	*C2, C3 + 3 UNK*
h54	*5*	*C1, C2 + 3 UNK*	*C2 + 4 UNK*
h55	*5*	*C1 + 4 UNK*	*5 UNK*

** Hypotheses were named as follows: h[number of contributors in the mixture][number of unknown contributors in H2]. In other words, h21 means [2 contributors to the mixture][1unknown contributor in H2].*
*** C1 represents the hypothesized contributor, i.e. the conditioned contributor for whom the weight of evidence is being assessed. In order to calculate LR for a known non-contributor, C1 was replaced with KNC. UNK = an unknown contributor.*

Table 10: LR values derived for different mixtures and hypotheses (see Table 9 for notation) for TC and KNC with a complete 15 locus profile.

	Percentage of replicates where TC LR			Percentage of replicates where KNC LR	
	> 1	**> 1000**	**> 1 million**	**> 1**	**>1000**
LR					
h21	100	100.00	100.00	0	0
h22	100	100.00	99.80	0	0
h31	100	100.00	99.98	0	0
h32	100	99.95	87.72	0	0
h33	100	99.80	63.96	0	0
h41	100	100.00	97.75	0	0
h42	100	98.23	54.36	0.01	0
h43	100	96.01	29.43	0.02	0
h44	100	93.47	15.97	0.02	0
h51	100	99.91	79.31	0	0
h52	99.99	90.69	28.03	0.09	0.01
h53	99.97	82.81	12.38	0.19	0.02
h54	99.97	75.38	6.00	0.21	0.01
h55	100	68.54	3.47	0.21	0.01
Total	99.99	93.20	55.58	0.05	0.004

Table 11: Proportion of replicates generating LR < 1 for TC at 1, 2,3,4 and 5 or more loci by hypothesis based on a complete 15 locus profile (see Table 9 for hypothesis notation).

	TC has LR < 1 at				
Hypothesis	1+ loci	2+ loci	3+ loci	4+ loci	5+ loci
h21	0	0	0	0	0
h22	44.15	10.12	1.49	0.17	0.04
h31	0	0	0	0	0
h32	77.22	40.95	15.64	4.01	0.73
h33	73.74	36.97	12.41	2.98	0.52
h41	0	0	0	0	0
h42	89.18	63.29	34.84	13.83	3.87
h43	89.65	62.55	33.04	13.15	4.05
h44	86.22	55.94	26.9	9.88	2.5
h51	0	0	0	0	0
h52	93.56	72.32	44.74	21.52	8.27
h53	95.65	79.28	52.93	28.31	11.39
h54	94.78	76.66	49.63	24.9	9.89
h55	93.44	72.73	43.83	21.09	7.4