New York State TrueAllele® Casework Developmental Validation

Prepared by:

Barry W. Duceman, PhD
Biological Science, New York State Police, Albany, NY 12226

Mark W. Perlin, PhD, MD, PhD
Cybergenetics, Pittsburgh, PA 15213

Jamie L. Belrose, MS
Northeast Regional Forensic Institute, Albany, NY 12222

19 February 2010

CONFIDENTIAL

DO NOT DISTRIBUTE

Executive Summary

Hypothesis: In today's forensic laboratory, a greater range of crime classifications are being considered for DNA analysis (e.g., property crimes) and ever more challenging evidence items are being submitted (e.g., low copy number). Many labs have responded to this increase in submissions by introducing automation into their workflow. The resultant increase in analytical capacity, in turn, has created bottlenecks at data interpretation and case file technical review. To resolve these bottlenecks, the New York State Police Forensic Investigation Center has undertaken to test whether Cybergenetics TrueAllele® Casework statistical modeling system is capable of unattended STR DNA data review and interpretation.

Materials: In this validation study, we re-analyzed 41 cases (all adjudicated, except for two proficiency tests) previously analyzed by competent DNA analysts at the Forensic Investigation Center. To span the range of interpretation challenges commonly encountered in forensic casework, these 41 cases were relatively equally distributed between sexual assaults containing victim and suspect reference samples along with various bodily swabs, with more complex multiple-victim homicides involving upwards of 30 evidence items. The study has 368 items covering most commonly submitted evidence including vaginal swabs, anal swabs, oral swabs, penile swabs, dried secretions, blood stains, semen stains, weapons, cigarette butts, condoms, human hair, bite marks, and fingernail scrapings.

Methods: The NYSP generated the original electronic data files generated using Applied Biosystems genetic analyzers. Cybergenetics uploaded these files to the TrueAllele Casework

system. The data and results were then available for retrospective comparison to the data gleaned from the corresponding case reports issued by the NYSP. The authors evaluated genotype concordance in all 368 items of evidence, monitored TrueAllele Casework's ability to separate mixture genotypes across a range of mixing weights and complexities, evaluated the mixture weight percentages determined by TrueAllele Casework, and compared the statistical match weight of evidence obtained by traditional means (NYSP protocol) to those calculated by the software.

Results: In this study, we first **e**xamined 4,958 alleles in 202 single-source profiles in 41 previously examined cases, and found the genotypes inferred by the statistical computing system to be in complete concordance. The results of computer genotype inference over a wide range of mixture items, whether in complex or difficult cases, were in accord with those determined using standard validated procedures at the State Police Crime lab. Without any knowledge of the STR profile of the suspect, the computer system more effectively ascertained the profile of the perpetrator and, commonly, provided more profile information than the standard non-automated manual process. The computer automatically provided likelihood ratios and, in every case examined, preserved more identification information as measured by comparison of likelihood ratios.

Conclusion: The statistical TrueAllele system conveniently utilizes STR DNA data accepted from in-house genetic analyzers. Moreover, as intended, the system has demonstrated the potential to relieve bottlenecks due to increased automation. The interpretation of STR DNA data by TrueAllele probability modeling offers enhanced objectivity through reduced examiner

bias in forensic DNA casework. The system allows the genetic testing laboratory workflow to be designed so that there is no previous exposure of the reporting analyst(s) to the DNA profiles of a suspect or pool of suspects until the laboratory report is prepared. The system achieves greater resolution in statistical inference of mixture genotypes than current standard practices. Most importantly, the system offers increased statistical strength of match.

Table of Contents

Introduction

DNA evidence resides at the center of modern criminal justice, and it is used to help apprehend, convict and exonerate suspects. An ideal DNA system would provide identification information with speed, accuracy and objectivity. These desirable features are already found in the data generation process, in which a DNA laboratory transforms biological specimens into quantitative computer signals. However the second phase – data interpretation – is still largely conducted by a human review process.

With pristine DNA data (e.g., reference samples), human review can work well. But DNA casework evidence is usually not pristine. Extracted under real-world conditions, DNA evidence is often mixed (having multiple contributors), damaged (by heat or bacteria) or low template (thus hard to discern with any certainty).

Uncertain DNA data may suggest multiple genotype possibilities, thereby reducing identification information. Human review of uncertain DNA can be a time-consuming process that does not fully elicit all the information that the data contain. Moreover, human comparison of DNA evidence and suspect may not be entirely objective.

Computer interpretation of DNA evidence can overcome these issues. Specifically, it is:

- *Fast*, with parallel computers turning out solutions every few minutes;
- *Accurate*, able to employ mathematical models that fully preserve all of the identification information residing in the DNA data; and

• *Objective*, interpreting evidence without ever seeing a single suspect genotype.

Such computer processing can effectively handle the mixed, damaged and low level DNA evidence that currently consume much of the human review effort.

To properly use such a computer system, it is essential to know its capabilities and limitations. For example, how well does it handle two, three or more unknown contributors? How damaged or low level can the DNA be? Can independent evidence be mathematically combined to make a more informative identification?

This validation study determined the applicability of Cybergenetics TrueAllele® Casework, a commercial computer system for the mathematical interpretation of DNA evidence. We analyzed 368 items of anonymized adjudicated evidence from the NYSP FIC, including 88 mixture samples. We found that the DNA match information calculated by the TrueAllele computer exceeded the reported human review score in every case. We also measured the reproducibility of TrueAllele inference using match information.

We begin this report by describing methods for the computer interpretation of quantitative DNA mixture data, the TrueAllele Casework system and the validation metrics that we used. We then describe the case materials used in the study, and our item classification approach. We provide an illustrative case example for each class of mixture item. We present our validation results, quantifying the efficacy and reproducibility of TrueAllele Casework interpretation on this mixture data set.

Methods

Interpreting uncertain DNA evidence

A definite genotype can be determined when a person's DNA produces clean data. However, when the data signals are less definitive, or when there are multiple contributors to the evidence, uncertainty arises. This uncertainty is expressed in the resulting genotype, which may describe different genetic identity possibilities. Such genotype uncertainty may translate into reduced identification information when comparison is made with a suspect.

The DNA identification task can thus be understood as a two-step process:

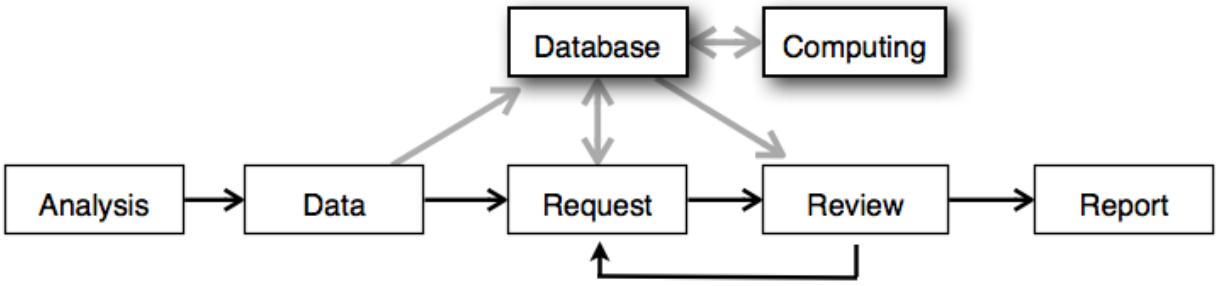
1. objectively *inferring genotypes* from evidence data, accounting for allele pair uncertainty using probability, and
2. subsequently *matching genotypes*, comparing evidence with a suspect relative to a population, to express the strength of association using probability.

The match strength is reported as a single number, the likelihood ratio (LR), which describes the gain in identification information produced by having examined the DNA evidence.

The TrueAllele® system is a computer implementation of this two-step objective genotype inference approach. TrueAllele infers genotypes from DNA data through mathematical modeling (1, 2). To capture all the identification information present in the data, the system represents genotype uncertainty using probability. These uncertain genotypes are stored on a

TrueAllele database so that they can be compared with suspects for investigative and evidential identification. The TrueAllele user asks interpretation questions of DNA data, visually reviews the computer's answers, and generates match reports to use in court.

The TrueAllele technology



Cybergenetics TrueAllele technology for automated interpretation and reporting of DNA evidence is based on biology, mathematics and computation (3). This section describes the TrueAllele workflow from a system and user perspective.

Analysis

The DNA interpretation process requires quality-checked quantitative data. The TrueAllele Analysis computer starts with the laboratory's original electronic DNA files, and works with the user to check and quantify these raw data signals, in order to produce interpretation-ready data. For each 96-well plate of DNA samples and controls, Analysis applies multiple rules to the signals to ensure that good data move forward on to interpretation. The computer gives the lab

feedback about any data issues that it finds. The process is fast, taking a few minutes of user time for a typical DNA plate.

To assess the DNA data signals in Analysis, a user opens a folder of electronic DNA sequencer files. He then asks the TrueAllele computer to check the DNA sizing calibration data, and looks for any problems with these (and other) control samples. Man-machine communication is exchanged visually, with the user pointing his mouse at the screen to explore an issue, and the computer responding by rendering a data image or figure that focuses on the user's question. After the computer has processed the peak events in the DNA data signals, the user has a data file of quality-checked quantified peaks ready for the database.

Data

After peaks are quality checked in the Analysis phase, we can view them in the TrueAllele Data interface. This gives the user another opportunity to review the peaks before interpretation. The TrueAllele computer can signal the presence of any possible artifacts in the data, so that the user can evaluate the peak and take action upon it if necessary. Once the quality-checked peaks have been reassessed, they are ready for upload to a TrueAllele database, and then used in TrueAllele interpretation.

To upload quality-checked quantified peaks into a database, the user opens a "Visual User Interface for easy review" (VUIer™) Data window. He first connects to a TrueAllele database that will store the data. After opening the file created in the Analysis phase, the data peaks

appear on the screen as intuitive visually rendered signals. Each data injection is shown within its own track. The user can ask the computer to show possible lingering data artifacts, along with pertinent data information. This annotating information is stored with the peak file on the database. When the data review is complete, the user uploads the peak data to the database, making it available for creating TrueAllele interpretation requests.

Database

The uploaded DNA data reside on a TrueAllele PostgreSQL relational database. The database is like an electronic filing cabinet that permits information retrieval simultaneously from multiple file folders. The database provides persistent and secure storage for all the information needed by the TrueAllele user and system. The (over seventy five) database tables provide quantitative DNA data, TrueAllele interpretation questions, the computed results, and supporting information, such as population frequencies. The database also helps administer system activities, and supports the monitoring expert system that coordinates the system.

The user logs on to a TrueAllele database to initiate processing or to review results. The user's interactions are mediated through the TrueAllele VUIer software installed on their computer. The VUIer database client exchanges DNA case data with the database, and presents information visually on the computer screen. All the user modules (e.g., Data, Request, Review, Report) automatically generate database queries and DNA visualizations through the VUIer. Typical displayed case information includes DNA data, genotype probability, mixture weight distribution and match rarity likelihood ratio values.

Request

Once the data are on the database, we can ask DNA interpretation questions that the TrueAllele computer can solve for us. Each question involves one or more DNA evidence items, and can be run under different problem solving conditions. (Example conditions are how many unknown contributor genotypes to find, how much computer time to use, or whether to account for degraded DNA.) While a victim reference may be optionally included in a question, for total objectivity a suspect genotype is never used. Questions can be asked one at time, in duplicate for reproducibility, or in batches of a hundred or more. Regardless, once a question has been posed, the statistical calculating is done entirely by computer.

To ask interpretation questions in a case, the user opens a VUIer Request window. After connecting to her evidence database, she selects the DNA data that she wants to use. These data images appear visually in the interface, with each signal in its own track. She then forms visual DNA items (each corresponding to an evidence sample) from the track signals. Finally, she makes each case interpretation request by indicating one or more DNA items, and setting optional problem solving parameters. Once she is satisfied with her questions, the user uploads her interpretation requests to the TrueAllele database for computer processing.

Computing

After the user has posed DNA interpretation questions, a TrueAllele server interpretation computer automatically retrieves a request and its data from the database. TrueAllele interpretation uses all the data to infer a genotype distribution and mixture weight for each DNA contributor. To infer a genotype distribution, the computer explores various peak patterns to statistically model the data. Throughout this modeling process the computer considers many different variables, such as genotype, mixture weight, stutter and preferential amplification. As a result, the reported genotype distribution reflects how well a set of proposed patterns fit the data. Patterns that closely fit the data receive higher probabilities, and patterns that do not receive lower probabilities. A separate TrueAllele server computer then matches the inferred genotype distribution against provided references, and calculates a likelihood ratio statistic.

The TrueAllele parallel compute servers can process multiple requests at the same time. For example, solving a DNA interpretation question in duplicate creates two independent calculations, establishing statistical reproducibility. We routinely run 24 parallel TrueAllele processes on our system, each one working on a different case. A typical DNA mixture takes about an hour or so to solve, so the overall throughput can be quite high (e.g., over 300 cases a day). When the problem solving is done, the computer stores its results (inferred genotype distributions, mixture weights, likelihood ratios, etc.) on the database for downstream review.

Review

Once the requests have finished processing, we can review the computer interpretation results. During this review process, we can see several aspects of the DNA case. For example, we can examine a contributor's genotype probability distribution, either visually or in a table. It is this key genotype variable, and its probability uncertainty, that establishes genetic identity. With multiple DNA contributors, we can visually review mixture information with informative pictures of mixture weight probability. The quantitative match information can be seen visually at the different genetic loci.

The user first opens a VUIer Review window, and selects a request from the database. A Profile window appears, visually displaying computed genotype probability distributions. From here, the user can navigate to other windows, including ones for the original Data and the Mixture separation. When TrueAllele finds a match between an evidence contributor and a suspect, the Match window and tables show quantitative LR match information. An Explain window visually explains the computer's reasoning. A user can always ask more questions by exiting Review and returning to the Request module, where he can create new TrueAllele interpretation questions.

Report

After the interpretation requests have been processed by the computer and reviewed by the analyst, we are ready to generate reports for court presentation. TrueAllele generates the

customizable report automatically based on user selected options. A typical report consists of an evidence interpretation summary, lab information, a match rarity statement and detailed locus results. The reported match statistic incorporates appropriate population allele frequencies, and can apply a coancestry coefficient (theta) for a statistic with population substructure.

For automatic report generation, the user opens the VUIer Report window. After connecting to a TrueAllele database, the user downloads genotypes of interest: evidence contributors, suspect references, and population frequencies. The probability distributions of each genotype are displayed together visually in the VUIer Report window. The user can review different matches of evidence contributors to suspect references, and generate a report for any match. She can export her report from VUIer as a text document, and import it into a spreadsheet program.

Comparison metrics

There are different ways to compare DNA inference results. Some comparisons are qualitative, such as examining genotype or mixture weight probability distributions. Results can be compared quantitatively using the LR DNA match association score.

Likelihood ratio match information

The identification information of the STR data is captured in a single number, the LR that describes the gain in information resulting from having looked at the data (4). We use this

generally accepted LR gold standard for quantifying the efficacy and reproducibility of DNA interpretation methods, and making comparisons between methods. Scientists generally use the logarithm (powers of ten, or "order of magnitude") of the LR as the information measure.

With a strong DNA match of evidence to suspect relative to a population, the LR is typically over a million ($\log(LR) > 6$). The LR can reach into quintillions ($\log(LR) > 18$) when using 13 STR loci on unambiguous DNA data. A mismatch will have a very small LR under 1, typically less than a quadrillionth ($\log(LR) < -15$). Thus the LR numerically quantifies the extent of match, based on DNA data (5, 6). This numerical LR presentation of scientific evidence is more precise than using qualitative words like "inclusion" or "exclusion" (7), binary decisions that are perhaps best left to the trier of fact.

The LR logarithm pervades all of our study Results. For method *efficacy*, we use the log(LR) to measure the identification information inferred from the DNA data. For *relative efficacy* results, we again use the LR, since the information gain of one method over another is just the logarithm of their LR ratios. We quantify *reproducibility* through log(LR) variation within a case. The *productivity* of DNA processing can be assessed by the probability that a sample will produce a reportable LR match score.

Genotype probability distribution

Uncertainty in DNA evidence translates into genotype uncertainty. All mixture interpretation methods report out a list of allele pair possibilities, associating them with probabilities. It can be instructive to examine and compare these inferred genotype probability distributions.

In our Case Examples, we will look directly at the genotype likelihood or probability distribution resulting from different interpretation methods on mixture data. For more quantitative assessment, it is useful to translate these genotype probabilities into LRs and determine their match information. That translation reduces the genotype's multidimensional probability distribution over allele pairs at multiple loci into a single information number, the log(LR).

Mixture weight probability distribution

Individuals contribute their DNA to a mixture item in a certain proportion, or "mixture weight". Analysts often try to infer this mixture weight (mean and variance) for the DNA template from some genetic loci using quantitative peak height data and contributor genotypes. Computers can use all of the loci to infer the mixture weight of the DNA template (2).

In the Results section, we shall use mixture weight as an auxiliary variable that helps us assess the complexity of a mixture item. A small mixture weight can mean a lower quantity of contributor DNA, which may reduce match information through stochastic effects. In a 50:50

mixture that is interpreted without a victim reference, more allele pair combinations are possible, and this genotype uncertainty can decrease match information.

Validation methods

This report centers on the scientific validation of the TrueAllele Casework DNA interpretation system (3). The statistical approach uses DNA match information as the key metric (8), since that is the single measure of association used by law enforcement and the courts (9).

Efficacy

The outcome of any genotype inference from evidence data is a probability distribution over allele pair values at each locus. These probabilities arise from Bayesian inference (10), using a population prior and a likelihood function.

Computer-based modeling methods (11, 12), such as the TrueAllele system (13), employ a quantitative likelihood function that compares proposed patterns with STR peak heights. Human review of DNA mixtures is commonly done qualitatively using the Combined Probability of Inclusion (CPI) or Combined Likelihood Ratio (CLR) (14). A qualitative binary method, such as CPI or CLR, forms a genotype list of length N that contains reportable allele pairs, each one assigned a likelihood of $1/N$ (9).

A LR compares this evidence genotype to a suspect genotype, relative to a population genotype, through their probability distributions to obtain match information (9). Thus the LR provides a universal mechanism for comparing match information between genotypes inferred by different mixture interpretation methods, relative to the same suspect and population (15).

The log(LR) is a standard measure of information that describes how much information was gained in a hypothesis *H* by observing some data (4, 16). The LR is often defined as the ratio of posterior odds to prior odds, as follows

(1) $$LR = \frac{O(H|data)}{O(H)}$$

In forensic DNA identification, hypothesis *H* is that the suspect contributed to the DNA data. The prior odds of the hypothesis is denoted as $O(H)$ before seeing any DNA data, and the posterior odds as $O(H|data)$ after having examined the data.

(The "odds", as in a game of chance, is the ratio of two probabilities – that of some event occurring or hypothesis being true, relative to that of the opposite alternative. In football, for example, the hypothesized event may be that a team will win a game against a particular opponent. The "odds" then quantifies our belief in the team's chance of winning. A LR would describe how much information we gain about the team's chance of winning from some new data that we learn about the situation.)

All currently reported match statistics (e.g., TrueAllele, kinship, CLR, CPI) can be viewed as LRs (9). Therefore, we can compare the relative efficacy of two mixture interpretation methods

by examining the difference in their log(LR) scores. For a set of cases, we can also look at the mean value of these information differences.

In this project we will compare differences in identification information between quantitative and qualitative mixture interpretation methods (3). When the victim genotype is known and used, the difference is log(TrueAllele) – log(CLR). When the victim is not available for genotype inference, this information difference is log(TrueAllele) – log(CPI).

Through these measures of efficacy, the validation study can verify that the TrueAllele system extracts at least as much information as current manual review methods. Moreover, the efficacy measures can quantify the extent of additional information that the computer is able to derive from the data.

Reproducibility

An important aspect of scientific reliability is a method's reproducibility (17). The reproducibility of a set of measurements is conventionally reported as the standard deviation of these numbers (18). Any mixture interpretation method applied to some DNA data will infer a genotype, which yields a single information log(LR) measurement when compared with a suspect and population. Independent interpretations using the same method on the same DNA mixture data, relative to the same suspect and population, produce a set of log(LR) information values. From this set of measurements, we can assess the method's reproducibility by computing a standard deviation of the inferred match information for a case.

To sharpen the reproducibility estimate of a mixture interpretation method, we use more cases. The "within-case" standard deviation σ_w (19) describes the method's reproducibility over a population of mixture cases (8). We can compute σ_w as the root mean square deviation of replicated log(LR) information scores, relative to the mean value within each case (19), as

$$\sigma_w^2 = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J_i} \left(s_{ij} - \overline{s}_i \right)^2}{\sum_{i=1}^{I} J_i} \qquad (2)$$

Here, I is the number of cases, J_i is the number of independent interpretations of the i^{th} case, s_{ij} is the log(LR) score of the j^{th} interpretation of the i^{th} case, and $\overline{s}_i$ is the mean score of the s_{ij} values within the i^{th} case.

Through these measures of match information reproducibility, the validation study can quantify the reliability of the TrueAllele system under different casework situations. This quantification is done by assessing reproducibility on subgroups of DNA items of differing sample complexity (number of contributors, mixture weight, DNA amount, DNA degradation, etc.).

Case Materials

Evidence items

In this validation study, we re-analyzed 41 cases (all adjudicated, except for two proficiency tests) previously analyzed by competent DNA analysts at the Forensic Investigation Center. To span the range of interpretation challenges commonly encountered in forensic casework, these 41 cases were relatively equally distributed between sexual assaults containing victim and suspect reference samples along with various bodily swabs, with more complex multiple-victim homicides involving upwards of 30 evidence items (Table 1A). The 368 study items were derived from 206 distinct biological source samples. These samples cover most of the commonly submitted evidence sources, including vaginal swabs, anal swabs, oral swabs, penile swabs, dried secretions, blood stains, semen stains, weapons, cigarette butts, condoms, human hair, bite marks, and fingernail scrapings (Table 1B).

The TrueAllele-inferred mixture weights of these evidence items were broadly distributed between 0 and 1 (Figure 1). We wanted to study actual mixtures, and not consider nonmixture samples or blanks. We therefore focused on the 88 evidence items that had a mixture weight between 0.05 and 0.95 (inclusive).

Classification

Evidence items were classified as simple, intermediate or complex. There were about twice as many items in each of the simple and complex categories than there were in the intermediate category (Table 1C). The number and nature of the contributors to the items were tabulated, as shown (Table 2).

Simple. All the simple mixture case items had two contributors (Table 2, simple). For most of these cases, one contributor was known and the task was to infer the unknown second contributor. There were no low template DNA sources. The data showed clear major and minor contributors. The DNA sources in the simple category were primarily from sexual assault differential extractions.

Intermediate. Mixture items in the intermediate category were more challenging. Some items were derived from low template DNA sources, while others contained contributors in approximately equal 50:50 mixture weights. Some mixture items contained two or three contributors, with multiple unknown contributors appearing in about half the items (Table 2, intermediate).

Complex. Mixture items in the complex category had two or three contributors, often with multiple unknown contributors (Table 2, complex). The STR data showed peak imbalance, and contributors with approximately equal 50:50 mixture weights (i.e., no clear major or minor

distinction). Some samples were amplified from low template DNA. Several of these cases also had multiple suspect or victim references.

Case Examples

We present three illustrative mixture items, one in each category: simple, intermediate and complex. The samples become progressively more interesting, and respectively entail one, two and three unknown contributors.

Simple

In this simple case, the data suggest that there are two contributors. The evidence is from a sexual assault, so one contributor is known to be the victim. This constraint fixes one genotype in the mixture, so the task is then to infer the one unknown contributor genotype. With just one genotype to infer, the problem is similar to a single source scenario. A highly informative genotype would lead to a large match score.

Mixture. With the genotype of the known contributor fixed, inferring the contributor mixture weight is straightforward for a computer. The sharp bell curve shown indicates good genotype separation, with a relatively certain mixture weight of 60% and a standard deviation of 3% (Figure 2A).

Genotype. A genotype probability distribution was inferred for the unknown contributor (Figure 2B). At each locus, there is single definite allele pair having probability one. Since the victim's genotype was known, the unknown major contributor has a definite genotype (no probability uncertainty), as with a single source profile.

Explanation. Using the Explain window in the TrueAllele interface, we can see how the computer models the data. The peak data is modeled here with proposed genotype pattern at locus D7 (Figure 3A). We know the [10, 12] victim genotype. The proposed [8, 8] allele pair for the unknown contributor genotype produces a pattern that closely fits the data (Figure 3B). No other allele pair candidate can properly account for the quantitative peak data, assuming a 60% mixture weight and a known first victim contributor.

Information. With a unique genotype possibility for the unknown contributor at every locus, the match strength was 247 quintillion, or 20.4 log units of information. The strong match score here is the same as it would be for a (single source) random match probability, since the inferred genotype is definite.

Intermediate

In this intermediate case, the data suggest a DNA mixture of two unknown individuals. The task here then is to infer a separate genotype probability distribution for each of the two unknown contributors. The computer must also infer each contributor's mixture weight. Inferring two unknown contributor genotypes can be more complex than inferring one (as in the previous

problem), and may yield less identification information. We quantify our uncertainty in the inferred genotypes and mixture weight using probability.

Mixture. For this two contributor case, the computer inferred mixture weights of 64% (Figure 4A, major contributor shown in blue) and 36% (minor contributor, orange). The uncertainty introduced through having a second unknown genotype was translated into a somewhat broader mixture weight probability distribution, with a standard deviation of 4.3% (Figure 4A).

Genotype. The computer inferred two genotypes (Figure 4B, blue major and orange minor). Each genotype shows the allele pair probability distribution at every locus. At most loci (e.g., FGA), there is more than one possibility, and the allele pair probabilities are shown. Note that for each contributor, at most loci, the inferred genotype suggests that one of these allele pair candidates is quite probable, as indicated by a long probability bar that approaches 1.

Explain. The likelihood of a genotype candidate expresses how well its model fits the data. The 64% major contributor allele pair [10 12] (Figure 5A, blue), together with the 36% minor [11 13] (orange) at locus D5, form a pattern (Figure 5B, gray) that closely fits the quantitative STR data. Therefore this genotype value pair has a high likelihood of being correct. However, most proposed genotype values do not fit the data all that well. A typical unlikely minor contributor allele pair [9 13] (Figure 5C, orange) forms a pattern (Figure 5D, gray) that is discordant with the quantitative data. The hypothesized [9 13] predicts a "9" allele peak where there is none, and does not account for the "11" allele peak. The evident mismatch between model and data makes this allele pair choice extremely unlikely, both to our eye and to TrueAllele's likelihood function.

Information. DNA data uncertainty affects the inferred genotype, which in turn affects the identification information. In this case, with two unknown contributors, the inferred probabilities were relatively high at the matching allele pairs, though less certain than in the simple case. Therefore, the match rarity between the major contributor and suspect decreased two log units to 18.1, or 1.52 quintillion.

Complex

In this complex case, the data suggest that there are three contributors to the DNA mixture sample. None of these contributors are known. The key task is to infer a genotype for each contributor. With more contributors to infer, we expect the genotype probability distributions to be less certain. Such less definite genotypes should yield less match information.

Mixture. Looking at the mixture weights, we show the inferred probability distribution for each contributor (Figure 6A). We see the weight distributions of the most abundant 60% major contributor (blue), the intermediate 27% minor contributor (green), and the lowest 12% minor contributor (orange). The observed mixture weight at each locus has stochastic variation, so the template's mixture weight data uncertainty is represented through probability. Here, the mixture weights have highly confident standard deviations of around 3%.

Genotype. For each contributor, the system inferred a genotype (Figure 6B). The most abundant major contributor (greatest mixture weight) produced the most definite probability distribution

(blue). As the amount of a contributor in the mixture decreases, its genotype distribution grows more diffuse (green). The lowest minor contributor has the most uncertain probability distribution (orange). This probability diffusion trend is clearly seen at locus D7, proceeding across the figure from greater confidence (left) to less (right).

Explain. The Explain interface shows data patterns, together with genotype models (Figure 7A). Note that each color corresponds to the allele pairs of a contributor at the CSF locus. A predicted model pattern that closely fits the observed quantitative data has a high likelihood. However, patterns from other allele pair alternatives might also fit the data, as reflected in the diffuse genotype distributions (Figure 6B). A better fit to the data tends to produce a higher allele pair probability (Figure 7B).

Information. A genotype having many likely allele pair possibilities can reduce match information. The major contributor inferred genotype here matched a reference with a high LR strength of 513 quadrillion (17.7 log units), relative to an ethnic population database. The intermediate minor 27% contributor genotype had a LR of 372 billion (11.6 log units). The minor 12% contributor genotype LR match strength was 186 billion (11.3 log units). The more diffuse inferred genotype probability distributions of the two minor contributors produced match strengths below a trillion. The three LR scores that TrueAllele inferred from this three unknown contributor mixture item are lower than in the intermediate case, but are larger than the world's population and quite useful for human identification.

Validation Results

There is often little agreement on the posterior probability distribution of a genotype inferred from DNA mixture data (20). However, for any genotype, its log(LR) against a known suspect relative to a reference population provides a standard measure of identification information. Since TrueAllele computes LRs, as do human mixture interpretation methods (e.g., CPI, CLR), we can quantitatively assess and compare DNA interpretation methods.

Efficacy

The log(LR) provides a measure of efficacy – how well an interpretation method works, in terms of the quantitative identification information it can extract from DNA data. Here we use the LR to describe the absolute efficacy of TrueAllele interpretation, as well as its relative performance when compared with human review.

For *absolute* efficacy, we compare the log(LR) information scores of simple, intermediate and complex cases in this data set (Figure 8). We see that, on average, simpler cases (blue) have more identification information (16.28 log units), intermediate cases (green) less information (13.14 log units), and complex ones (orange) even less (11.87 log units) (Table 3).

For *relative* efficacy, it would be helpful to know whether the TrueAllele computer is as (or more) informative than human review. We can determine this by comparing computer and

human review log(LR) statistics that were produced by analyzing the same DNA mixture data. In this study, we used mixture case items with two contributors to effect this comparison.

When a known victim reference genotype is available, the task is to infer the one unknown second contributor. To do this, a person often uses a CLR mixture interpretation method. For comparison on eight two person mixtures, we had the TrueAllele system look for one unknown genotype, assuming the victim reference. We see that in every case, the computer inferred more information on duplicate runs (Figure 9, blue, green) than a person using CLR did (orange) from the same data. On these eight cases, the average log(LR) match score improvement was 4.67 log units, or about 50 thousand.

Without a known victim reference genotype, one must infer both unknown contributors. This is often done by a person by applying the CPI method. We identified eight appropriate two person mixture items for comparison. We ran the TrueAllele system in duplicate on these data as a two unknown contributor problem, without any references. In every case, TrueAllele probability modeling (Figure 10, blue, green) inferred more information than human CPI review (orange) from the same data. On this set of eight items, the average log(LR) match score improvement was 6.24 log units, or over one million.

Reproducibility

We measured interpretation reproducibility by examining the variation of log(LR) match score resulting from duplicate computer runs on the evidence item. These variation data can be

combined across many case items using equation (2) to compute the within-group standard deviation.

Qualitatively, we see that one unknown mixture problems are more reproducible than are two unknown problems (Figures 9 and 10). Across the spectrum of all 88 mixture items, we see that more informative results tend be more reproducible (Figure 11). In this figure, the two replicate TrueAllele runs (blue, green) are shown sorted by log(LR) information. As information decreases (from left to right), we see more divergence between an item's two inferred match information values.

Quantifying reproducibility by item complexity, we confirm that greater data certainty and LR scores produce (on average) more reproducible interpretations (Table 3). The simple items have a within-group standard deviation of 0.102. The intermediate samples double that variation to 0.255, while complex ones double the spread yet again to 0.437.

The observed with-in group standard deviations (Table 3) are all less than the population sampling variation of around one log unit found in reported LRs. We conclude that TrueAllele mixture interpretation is highly reproducible in all the data situations that we examined.

Productivity

Not all items of evidence in current laboratory practice yield a reportable quantitative DNA match statistic. Instead, an analyst may sometimes report out an item in "consistent with",

"cannot be excluded" or "insufficient" language, without attaching a LR number. The TrueAllele computer, though, has no choice but to infer a genotype that can later be used in an LR match comparison. If the data are uninformative, that fact is reflected in a diffuse genotype distribution and a low LR score.

How informative are TrueAllele's LR values in each sample item category? The computer is programmed to report a LR statistic on every genotype comparison that it makes. We see that on this data set, TrueAllele yielded an average LR match values of about 10^{16} (10 quadrillion) for simple items, 10^{13} (10 trillion) for intermediate ones, and 10^{12} (trillion) for complex cases (Table 3). So, regardless of item complexity, TrueAllele mixture interpretation usually produces an informative result on each item that it examines.

How often does human review produce a numerical LR match value? For the 35 simple items, human review in this study yielded a LR value (e.g., CPI, CLR) 49% of the time (Table 2, number and fraction reported by lab with LR statistic). The yield was less for the 20 intermediate items, with just 25% generating a numerical match score. With the 33 complex cases, only 21% of items produced a LR score. Overall, the laboratory's human review reported a numerical DNA match statistic for 29 of the 88 mixture samples, for a frequency of one statistic produced for every three items (33%).

TrueAllele interpretation always produces a LR match score for an item. However, with human review, a lab must analyze three mixture items (on average) in order to produce one match statistic. A TrueAllele-based interpretation process might therefore enable a laboratory to

consistently process fewer items of evidence. Such a sample volume reduction might accelerate turnaround time, and consume fewer reagents with less expenditure of human effort.

Other comparisons

We computed the mixture weight manually on spreadsheets for two person mixtures using those loci where the contributor allele sets did not overlap. At each locus, the contributor peak height sum was divided by the allele peak height total to estimate mixture weight. Examining the differences between TrueAllele-computed mixture weight (using all loci in a probability model computation) and these rougher human spreadsheet estimates, we see good agreement between the values (Figure 12).

Using the TrueAllele-computed log(LR) DNA match information scores, we see a distribution of DNA match information values with a median value centered around 15 log units (Figure 11). Using the TrueAllele-computed mixture weight (for the suspect-matching contributor), we see a roughly uniform distribution of mixture weight across the items (Figure 1). Pairing each item's log(LR) match information together with its mixture weight, scatter plots show (for each of the three classifications) how increasing item complexity is associated with reduced match information, and somewhat lower mixture weight (Figure 13 and Table 3).

Discussion

Science (21) and the law (22) often require forensic expert testimony to have a sound scientific basis. To demonstrate the reliability of DNA testing, forensic scientists conduct extensive validations of their STR data generation methods (23-25). Given the wide disparities found in DNA mixture interpretation results (26) and the ongoing controversy surrounding mixture interpretation methods (27, 28), clearly these methods should similarly be subject to scientific scrutiny. However, most mixture interpretation methods have not been validated to determine their efficacy and reproducibility. Without such rigorous validation, though, mixture interpretation may be subject to challenge in court (29).

Two analysts may independently review the same mixture data and arrive at different allele (or genotype) lists (26). It can be hard to quantify these qualitative discrepancies, or to make comparisons between different methods. Fortunately, in a validation study, the match statistic provides a single number that captures the identification information extracted from the data, relative to a known subject and a reference population. For DNA mixture interpretation methods currently in use (including CPI, CLR and TrueAllele), these match rarity numbers are all LRs (9). Since log(LR) is a standard measure of information (4), these numbers can be compared both within and between case interpretations to form the basis of a quantitative statistical validation study (8).

The advent of genetic calculators enables a computer interpretation of DNA evidence. While computers have inferred genotypes from genetic data for quite some time (30), they have only

recently been used for forensic identification (2, 31, 32). Computers offer three principal advantages in the interpretation process:

- *Productivity*. Computer review can help the analyst conduct rapid and accurate DNA data review (33). Reliable computing can eliminate the (often time-consuming) human review of cases that are impossible to solve, infer genotypes from extremely difficult mixture samples, and accelerate the processing of straightforward data.
- *Information*. Human review typically makes simplifying assumptions that can discard considerable identification information contained in the DNA evidence (27). A computer can use a statistical model to fully examine the quantitative peak height data.
- *Objectivity*. Human mixture interpretation methods sometimes use the suspect genotype to help infer or report results (34). A mathematically programmed computer can infer a genotype directly from the evidence data without using any suspect information, and then compute a match LR statistic from this genotype.

There is currently some controversy regarding the manual interpretation of uncertain DNA evidence. Some scientists dispute the proper way to qualitatively examine DNA mixtures (20, 27, 35, 36), with particular concern about stochastic effects and setting thresholds. However, a quantitative data variance model (3) can determine the probability distributions of the peak data. In this way, the TrueAllele computer system exploits stochastic effects for more informative genotype inference, and obviates the need for thresholds.

Forensic scientists also debate ways to objectively examine DNA evidence (37-40). The concern is that prematurely exposing a human examiner to a suspect profile can introduce observer bias.

The TrueAllele method, however, uses a two-step probability approach: first inferring genotypes from the evidence, and only afterwards making any LR comparison with the suspect. This "parallel unmasking" of independent evidence and suspect genotypes entirely eliminates any such objectivity concern.

In this report, we validated the TrueAllele genetic calculator for DNA mixture interpretation using statistical measures of efficacy and reproducibility based on log(LR) match information. When a victim reference was available, the computer was four and a half orders of magnitude more efficacious than human review on the same data. Without a victim reference, the average efficacy of the computer increased to six orders of magnitude. The computer methods were highly reproducible, as measured by within-case log(LR) standard deviation on duplicate runs.

Scientifically validated computer systems that can reliably solve DNA mixture cases could have a positive impact on criminal justice. For the forensic scientists and their laboratory, a computer assistant can help reduce the time, cost and uncertainty of DNA mixture review. Moreover, when testifying in court, scientists who report on match results using validated mixture interpretation methods will be less subject to challenge. By extracting (on average) a million times more identification information than the prevalent inclusion method from the same DNA evidence, quantitative computer interpretation provides the police with greater investigative power, the prosecutor with greater evidentiary power, and the defense with greater exculpatory power. Widespread deployment of these objective, information-rich computer-based productivity tools may help society by enhancing public safety.

References

1. Perlin MW, Lancia G, Ng S-K. Toward fully automated genotyping: genotyping microsatellite markers by deconvolution. Am J Hum Genet. 1995;57(5):1199-210.

2. Perlin MW, Szabady B. Linear mixture analysis: a mathematical approach to resolving mixed DNA samples. Journal of Forensic Sciences. 2001;46(6):1372-7.

3. Perlin MW, Sinelnikov A. An information gap in DNA evidence interpretation. PLoS ONE. 2009;4(12):e8327.

4. Good IJ. Probability and the Weighing of Evidence. London: Griffin, 1950.

5. Evett IW, Buffery C, Willott G, Stoney D. A guide to interpreting single locus profiles of DNA mixtures in forensic cases. J Forensic Sci Soc. 1991 Jan-Mar;31(1):41-7.

6. Lindley DV. A problem in forensic science. Biometrika. 1977;64(2):207-13.

7. Aitken CG, Taroni F. A verbal scale for the interpretation of evidence. Science & Justice. 1998;38(4):279-81.

8. Perlin MW. Scientific validation of mixture interpretation methods. Promega's Seventeenth International Symposium on Human Identification, 2006 Oct 10-12; Nashville, TN. 2006.

9. Perlin MW, Kadane JB, Cotton RW. Match likelihood ratio for uncertain genotypes. Law, Probability and Risk. 2009;8(3):289-302.

10. O'Hagan A, Forster J. Bayesian Inference. Second ed. New York: John Wiley & Sons, 2004.

11. Cowell RG, Lauritzen SL, Mortera J. Probabilistic modelling for DNA mixture analysis. Forensic Science International: Genetics Supplement. 2008;1(1):640-2.

12. Curran J. A MCMC method for resolving two person mixtures. Science & Justice. 2008;48(4):168-77.

13. Perlin MW. Simple reporting of complex DNA evidence: automated computer interpretation. Promega's Fourteenth International Symposium on Human Identification, 2003; Phoenix, AZ. 2003.

14. SWGDAM. Short Tandem Repeat (STR) interpretation guidelines (Scientific Working Group on DNA Analysis Methods). Forensic Sci Commun (FBI). 2000 July;2(3).

15. Buckleton JS, Triggs CM, Walsh SJ, editors. Forensic DNA Evidence Interpretation. Boca Raton, FL: CRC Press, 2004.

16. Savage LJ. The Foundations of Statistics. New York: John Wiley & Sons, 1954.

17. Faigman DL, Kaye DH, Saks MJ, Sanders J. Science in the Law: Standards, Statistics, and Research Issues: West Group, 2002.

18. Snedecor GW, Cochran WG. Statistical Methods. Eighth ed. Ames: Iowa State University, 1989.

19. Scheffé H. The Analysis of Variance. New York: John Wiley & Sons, 1959.

20. Gill P, Brenner CH, Buckleton JS, Carracedo A, Krawczak M, Mayr WR, Morling N, Prinz M, Schneider PM, Weir BS. DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. Forensic Science International. 2006;160:90-101.

21. National Research Council. Strengthening Forensic Science in the United States: A Path Forward. Washington, DC: National Academies Press, 2009.

22. Faigman DL, Kaye DH, Saks MJ, Sanders J. Science in the Law: Forensic Science Issues: West Group, 2002.

23. DNA Advisory Board. Quality assurance standards for forensic DNA testing laboratories and for convicted offender DNA databasing laboratories. Forensic Sci Commun (FBI). 2000;2(3).

24. Krenke B, Tereba A, Anderson S, Buel E, Culhane S, Finis C, Tomsey C, Zachetti J, Masibay A, Rabbach D, Amiott E, Sprecher C. Validation of a 16-locus fluorescent multiplex system. J Forensic Sci 2002;47(4):773-85.

25. Sgueglia J, Geiger S, Davis J. Precision studies using the ABI prism 3100 genetic analyzer for forensic DNA analysis. Anal Bioanal Chem. 2003;376(8):1247-54.

26. Butler JM, Kline MC. NIST Mixture Interpretation Interlaboratory Study 2005 (MIX05), Poster #56. Promega's Sixteenth International Symposium on Human Identification, 2005; Grapevine, TX. 2005.

27. Budowle B, Onorato AJ, Callaghan TF, Manna AD, Gross AM, Guerrieri RA, Luttman JC, McClure DL. Mixture interpretation: defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework. J Forensic Sci. 2009;54(4):810-21.

28. Gill P, Curran J, Elliot K. A graphical simulation model of the entire DNA process associated with the analysis of short tandem repeat loci. Nucl Acids Res. 2005;33(2):632-43.

29. Roberts v. United States. District of Columbia Court of Appeals; 2007.

30. Heuch I, Li F. PEDIG - A computer program for calculation of genotype probabilities, using phenotypic information. Clinical Genetics. 1972;3:501-4.

31. Bill MR, Gill P, Curran J, Clayton T, Pinchin R, Healy M, Buckleton J. PENDULUM - A guideline based approach to the interpretation of STR mixtures. Forensic Science International. 2005;148(2-3):181-9.

32. Evett IW, Gill P, Lambert JA. Taking account of peak areas when interpreting mixed DNA profiles. J Forensic Sci. 1998;43(1):62-9.

33. Varlaro J, Duceman B. Dealing with increasing casework demands for DNA analysis. Profiles in DNA. 2002;5(2):3-6.

34. Thompson WC. Painting the target around the matching profile: the Texas sharpshooter fallacy in forensic DNA interpretation. Law, Probability and Risk. 2009(in press).

35. Budowle B, Chakraborty R, van Daal A. Authors' response to commentary by Gill and Buckleton. J Forensic Sci. 2010;55(1):269-72.

36. Gill P, Buckleton J. Commentary on: Budowle B, Onorato AJ, Callaghan TF, Della Manna A, Gross AM, Guerrieri RA, Luttman JC, McClure DL. Mixture interpretation: defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework. J Forensic Sci 2009;54(4):810-21. J Forensic Sci. 2010;55(1):265-8.

37. Budowle B. Authors' response to commentary by E. Krane, Simon Ford, Jason R. Gilder, Keith Inman, Allan Jamieson, Roger Koppl, Irving L. Kornfield, D. Michael Risinger, Norah Rudin, Marc Scott Taylor, William C. Thompson. J Forensic Sci. 2010;55(1):275-6.

38. Budowle B, Bottrell M, Bunch S, Fram R, Harrison D, Meagher S, Oien C, Peterson P, Seiger D, Smith M, Smrz M, Soltis G, Stacey R. A perspective on errors, bias, and interpretation in the forensic sciences and direction for continuing advancement. J Forensic Sci. 2009;54(4):798-809.

39. Krane DE, Ford S, Gilder JR, Inman K, Jamieson A, Koppl R, Kornfield IL, Risinger DM, Rudin N, Taylor MS, Thompson WC. Sequential unmasking: a means of minimizing observer effects in forensic DNA interpretation. J Forensic Sci. 2008;53(4):1006-7.

40. Krane DE, Ford S, Gilder JR, Inman K, Jamieson A, Koppl R, Kornfield IL, Risinger DM, Rudin N, Taylor MS, Thompson WC. Commentary on: Budowle B, Bottrell MC, Bunch SG, Fram R, Harrison D, Meagher S, Oien CT, Peterson PE, Seiger DP,Smith MB, Smrz MA, Soltis GL, Stacey RB. A perspective on errors, bias, and interpretation in the forensic sciences and direction for continuing advancement. J Forensic Sci 2009;54(4):798-809. J Forensic Sci. 2010;55(1):273-4.

Figure Legends

Figure 1. *Item mixture weights.* The distribution of mixture weight over the case items is shown as a histogram.

Figure 2. *Simple example: infer.* The TrueAllele VUIer interface shows (A) the mixture weight posterior probability distribution as a histogram over the interval [0, 1]. The VUIer also shows (B) the genotype posterior probability distribution of the allele pairs at each locus as a horizontal bar chart.

Figure 3. *Simple example: explain.* The TrueAllele VUIer shows an explanation for the observed data peaks. (A) In Model view, the known victim allele pair (gray) is added to the proposed unknown contributor allele pair (blue) in a mixture weight proportion to form a pattern. (B) In Pattern view, the model pattern is expanded to include variables for PCR artifacts and other second order effects.

Figure 4. *Intermediate example: infer.* With two unknown contributors, the VUIer Review mode (A) Mixture window shows template mixture weight probability distributions for the major (blue) and minor (orange) contributors. (B) We also see in Review the genotype probability distributions for each contributor (blue, orange) at every locus.

Figure 5. *Intermediate example: explain.* The VUIer Explain window shows (A) the weighted genotype allele pair candidates from the major (blue) and minor (orange) contributors. (B) The

expanded peak pattern includes other STR variables, and fits the observed data peaks well. Also shown is an example of a poor candidate genotype, using the Explain window's (C) genotype view to see an unlikely minor contributor [9 13] allele pair (orange). We see that (D) the pattern produced (gray) does not fit the data peaks well.

Figure 6. *Complex example: infer.* There are three unknown contributors in this DNA item. (A) The VUIer Mixture window shows mixture weight probability distributions for the major (blue), greater minor (green) and lesser minor (orange) contributors. (B) The VUIer Profile window shows the posterior genotype probability distributions for each contributor (blue, green, orange) at every locus.

Figure 7. *Complex example: explain.* The VUIer Explain window shows how (A) the weighted genotype allele pair candidates from the major (blue) and two minor (green, orange) contributors help explain the data. (B) The peak pattern includes other explanatory STR variables.

Figure 8. *Item information.* The distribution of item match information (as inferred by TrueAllele) is shown as a histogram of log(LR) counts for each complexity category. The simple items (blue) are distributed more to the right than the intermediate items (green). The leftmost distribution is for the complex items (orange), which tend to be less informative.

Figure 9. *Match information comparison with one unknown contributor*. The log(LR) match information values are shown for each of the eight one unknown cases. The results for duplicate

LR1 computer runs (blue, green) and the reported CLR value (orange) are sorted by information divergence.

Figure 10. *Match information comparison with two unknown contributors*. The log(LR) information values are shown for each of the eight two unknown cases. The results for the duplicate LR2 computer runs (blue, green) and the reported CPI value (orange) are sorted by information divergence.

Figure 11. *Duplicate item match information*. All the items are sorted by descending match information. For each item, the $\log_{10}$(LR) information values of the independent first (blue) and second (green) TrueAllele computer runs are shown. We observe a median information value of 15 log(LR) units, and that variation between runs increases with decreasing identification information.

Figure 12. *Mixture weight comparison*. The mixture weights of two contributor items were determined by two different quantitative allele peak methods. A histogram of the differences is shown. Human calculation was done using a spreadsheet that used peak heights from alleles that could be separated by contributor. TrueAllele used all of the peak height data at all loci in a Markov chain computation. We observe that the human approximation is in reasonable agreement with the computer solution.

Figure 13. *Mixture weight vs. information*. For each category, a scatterplot is shown of mixture weight vs. log(LR) for all items in that category. We see that (A) simple items tend to have more

identification information and higher mixture weights than other items. The (B) intermediate items have less information and a lower mixture weights, while the (C) complex items have the least information, on average.

Table Legends

Table 1. *Case items*. Subtotals organized by (A) type of crime, (B) source of biological sample and (C) classification.

Table 2. *Item complexity*. The rows are organized into three groups, based on the simple, intermediate and complex classification. In the left half of the table, the number of contributors is shown, arranged by total, known and unknown contributor columns. The right half shows the number of items, giving the total number, how many of those items were reported out by the human review laboratory with a match statistic, and the corresponding fraction. (Since the TrueAllele computer always generates a match statistic when making a genotype comparison, its fraction would always be 1.)

Table 3. *Efficacy and reproducibility*. The rows organize the evidence items by their classification as simple, intermediate or complex. The columns are for the average log(LR) match information and mixture weight in each category.

Research Papers

We have attached three relevant scientific articles written by Dr. Mark Perlin, along with a contextually useful newsletter.

Linear Mixture Analysis. This 2001 peer reviewed Journal of Forensic Sciences paper presents the linear model used for modeling quantitative STR peak data. The key idea is that the amount of allele DNA that goes into a PCR amplifier is roughly proportional to the observed peak heights. When mathematically described, that insight permits accurate determination of genotypes from DNA mixture data.

Match Likelihood Ratio. This 2009 peer reviewed Law, Probability and Risk paper introduces the match LR (MLR) approximation to the standard LR. It shows how genotypes can be first inferred, and then subsequently matched, giving useful examples. The MLR provides a simple explanation of the LR. All of these MLR ideas can be applied to the exact LR form that we used in our validation study.

Information Gap. This 2009 peer reviewed PLoS ONE validation paper compares TrueAllele probability computer modeling with current human review methods of DNA mixtures. The paper shows that while human review can reliably proceed down to about 100 pg of DNA, mathematical computation can extend this range down to about 10 pg of DNA. That is, DNA laboratories produce data that contain considerable identification information that is often discarded. The paper also describes the use of TrueAllele in the Foley homicide case.

The Foley Case. This 2009 Cybergenetics newsletter describes the Pennsylvania homicide Commonwealth vs. Foley case. In the 2009 criminal trial, TrueAllele computer interpretation of DNA mixture was introduced into court as evidence after a Frye hearing that established the general acceptance of the underlying principles in the relevant scientific community. While the CPI method produced a LR score of 13,000, computer review of the same data using probability modeling yielded a LR score of 189 billion. These results are consistent with our observations in this validation study – computer review is typically a million times more informative than human review of the same DNA mixture evidence.

Figure 1

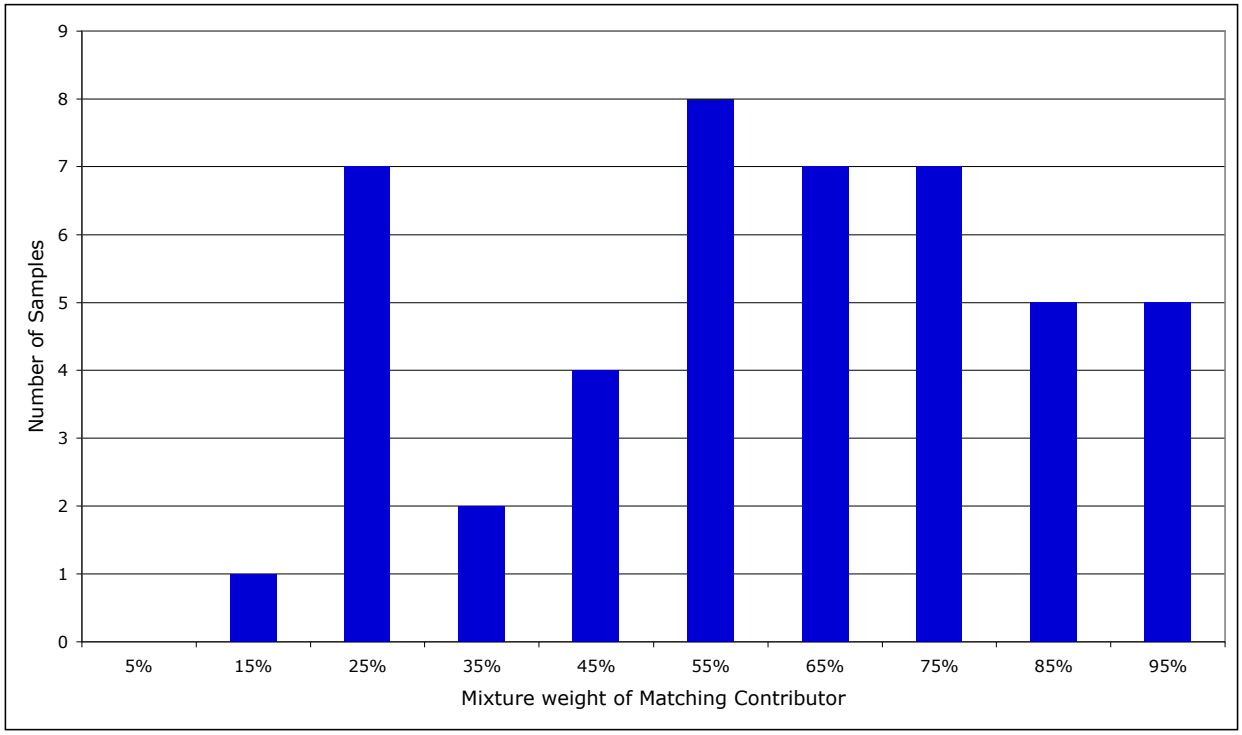
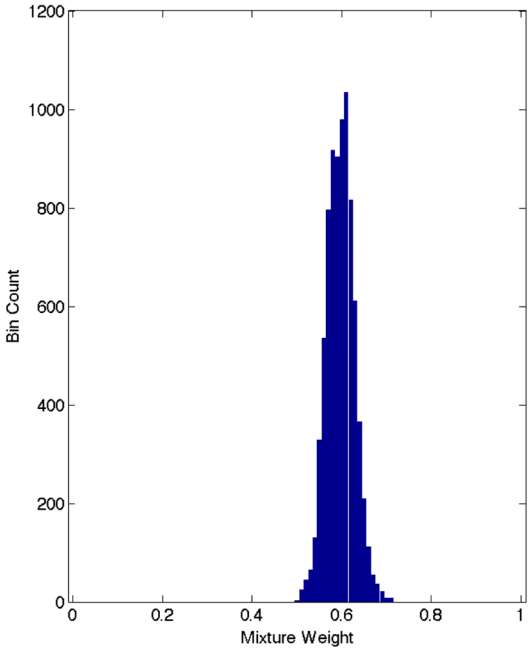

Number of Samples
0
1
2
3
4
5
6
7
8
9
5%
15%
25%
35%
45%
55%
65%
75%
85%
95%
Mixture weight of Matching Contributor

Figure 2

A


1200
1000
800
600
400
200
0
Bin Count
0
0.2
0.4
0.6
0.8
1
Mixture Weight

B

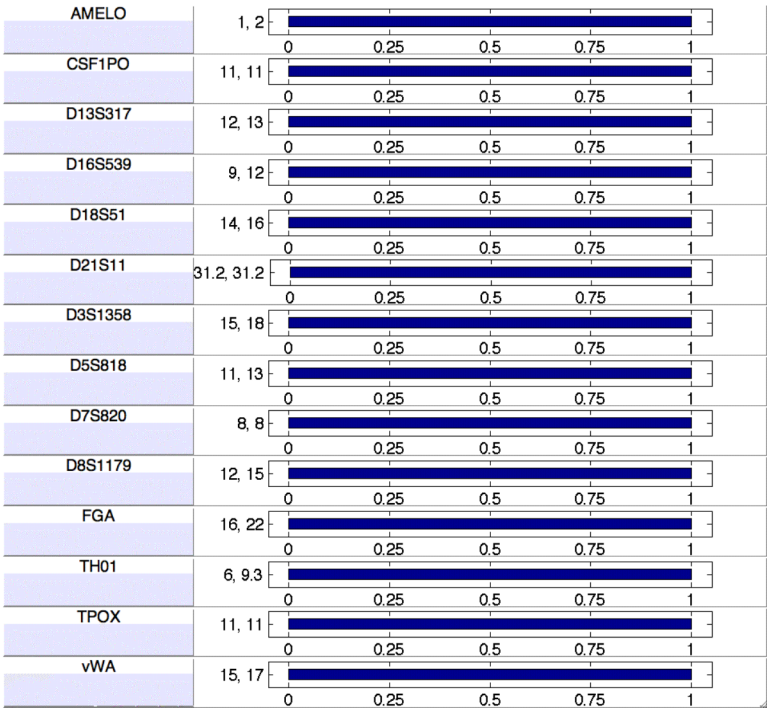
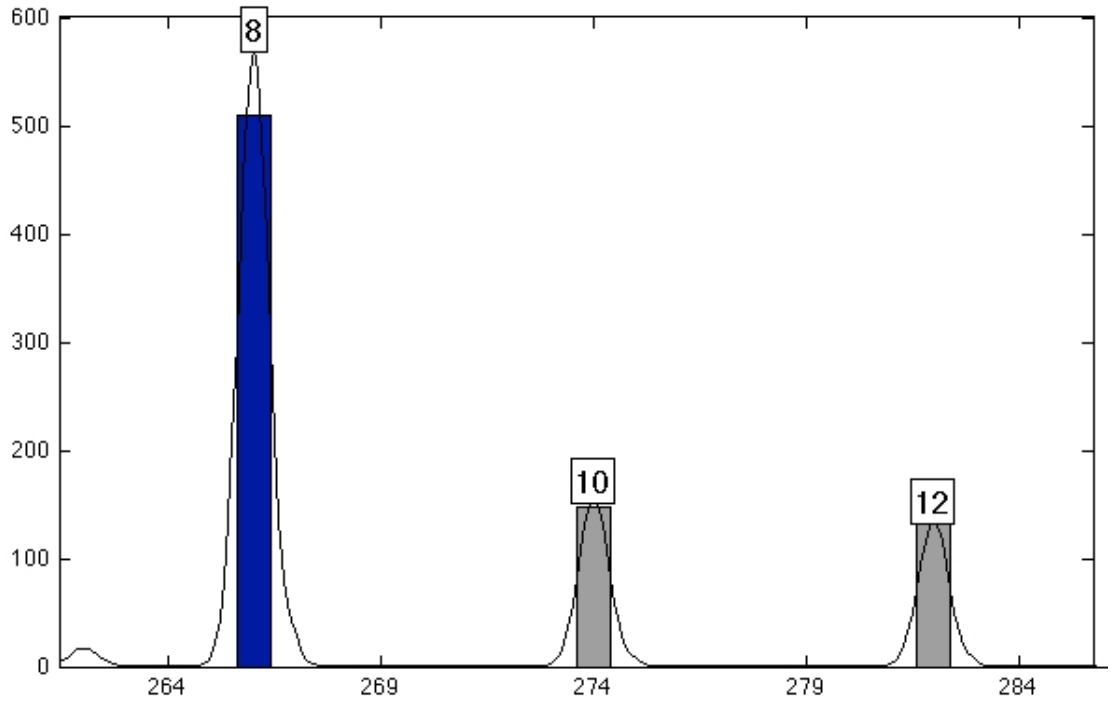

AMELO
1, 2
CSF1PO
11, 11
D13S317
12, 13
D16S539
9, 12
D18S51
14, 16
D21S11
31.2, 31.2
D3S1358
15, 18
D5S818
11, 13
D7S820
8, 8
D8S1179
12, 15
FGA
16, 22
TH01
6, 9.3
TPOX
11, 11
vWA
15, 17
0
0.25
0.5
0.75
1

Figure 3

A


600
500
400
300
200
100
0
8
10
12
264
269
274
279
284

B

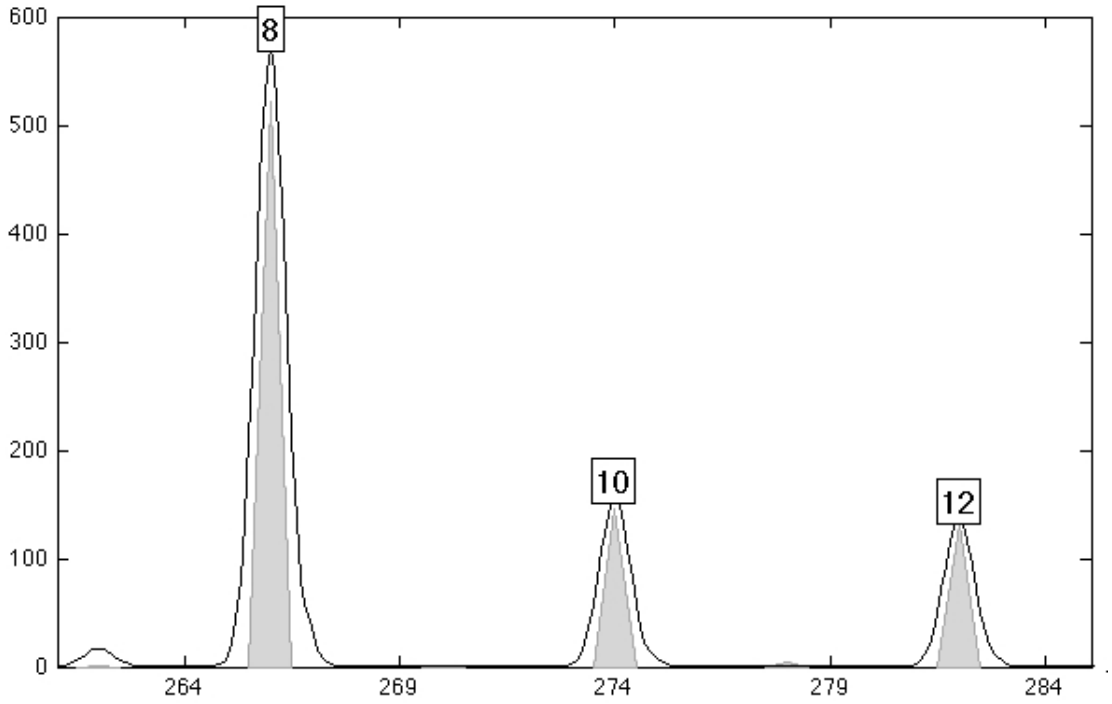
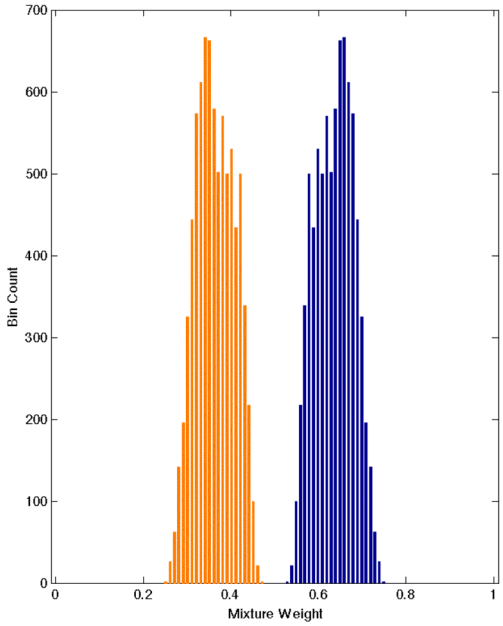

600
500
400
300
200
100
0
8
10
12
264
269
274
279
284

Figure 4

A


Bin Count
Mixture Weight
0
100
200
300
400
500
600
700
0
0.2
0.4
0.6
0.8
1

B

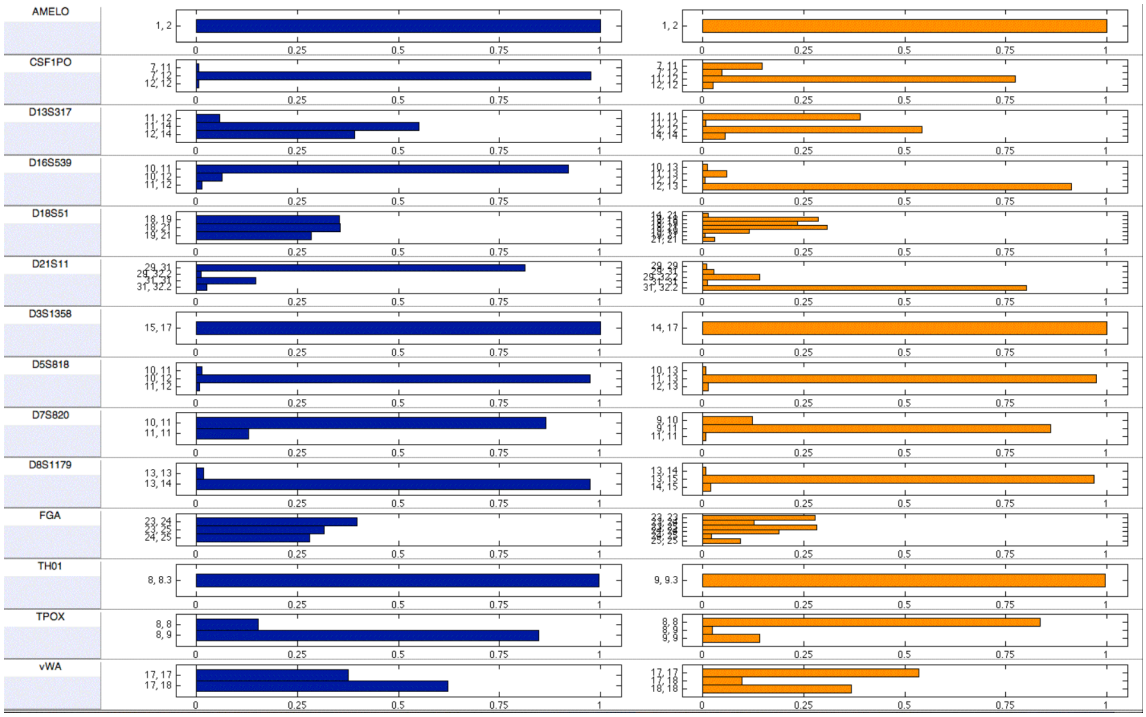
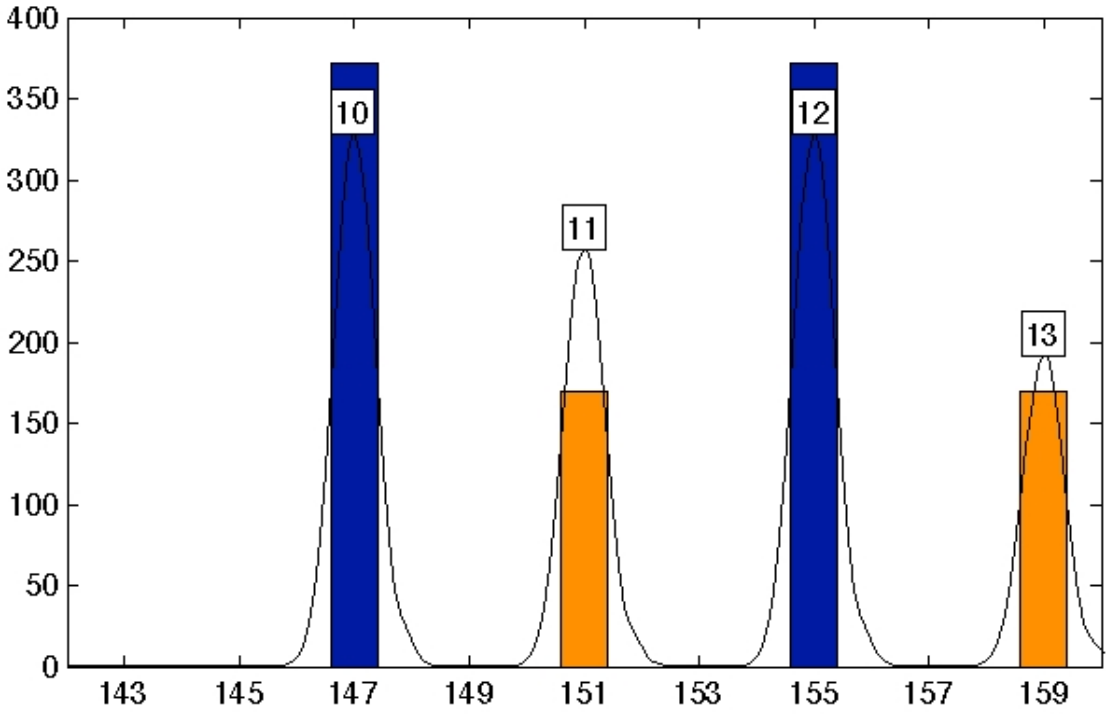

AMELO
CSF1PO
D13S317
D16S539
D18S51
D21S11
D3S1358
D5S818
D7S820
D8S1179
FGA
TH01
TPOX
vWA

Figure 5

A


400
350
300
250
200
150
100
50
0
143
145
147
149
151
153
155
157
159
10
11
12
13

B

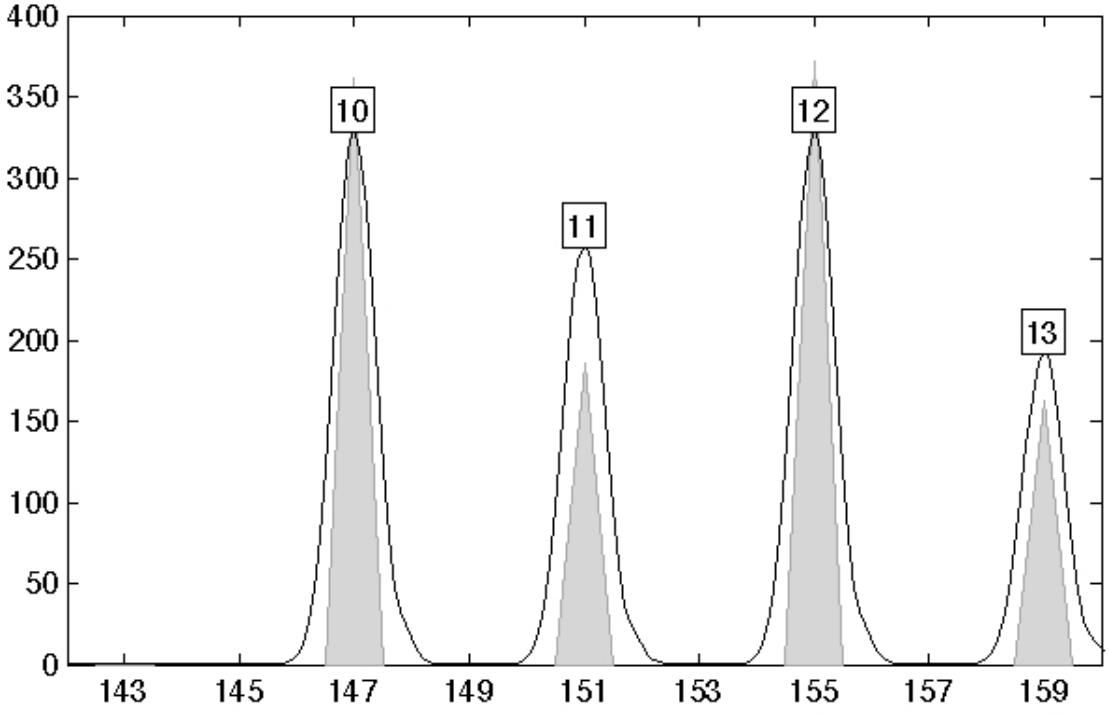
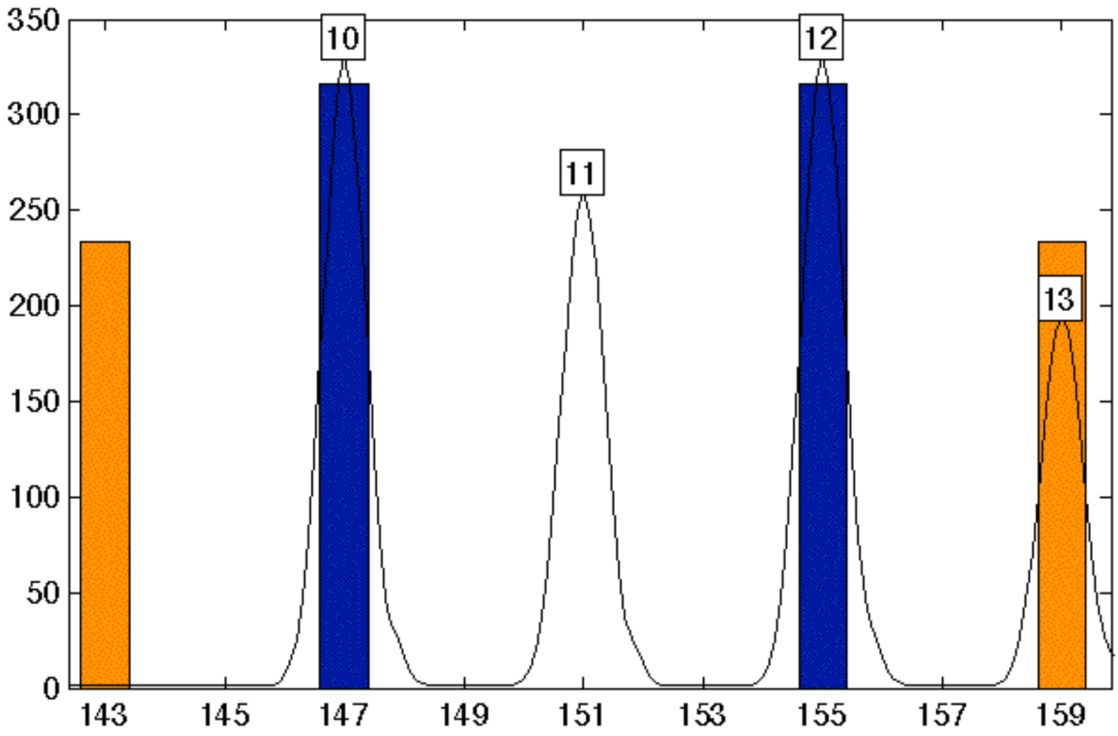

400
350
300
250
200
150
100
50
0
143
145
147
149
151
153
155
157
159
10
11
12
13

Figure 5, continued

C



D

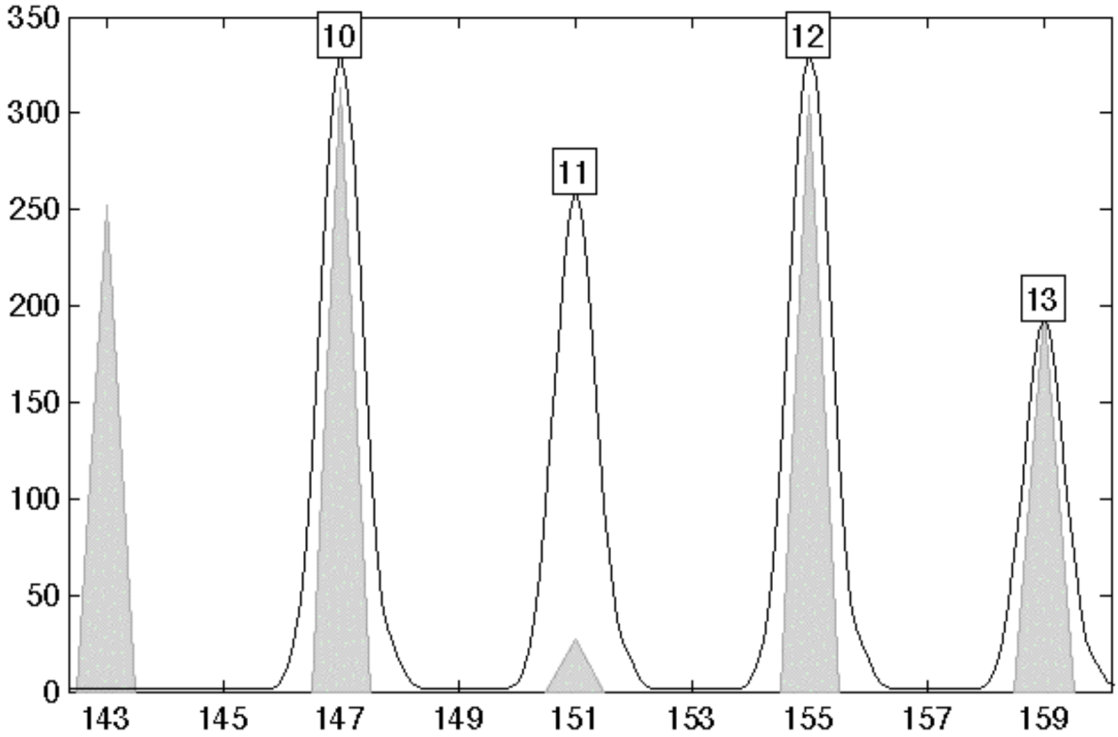


Figure 6

A

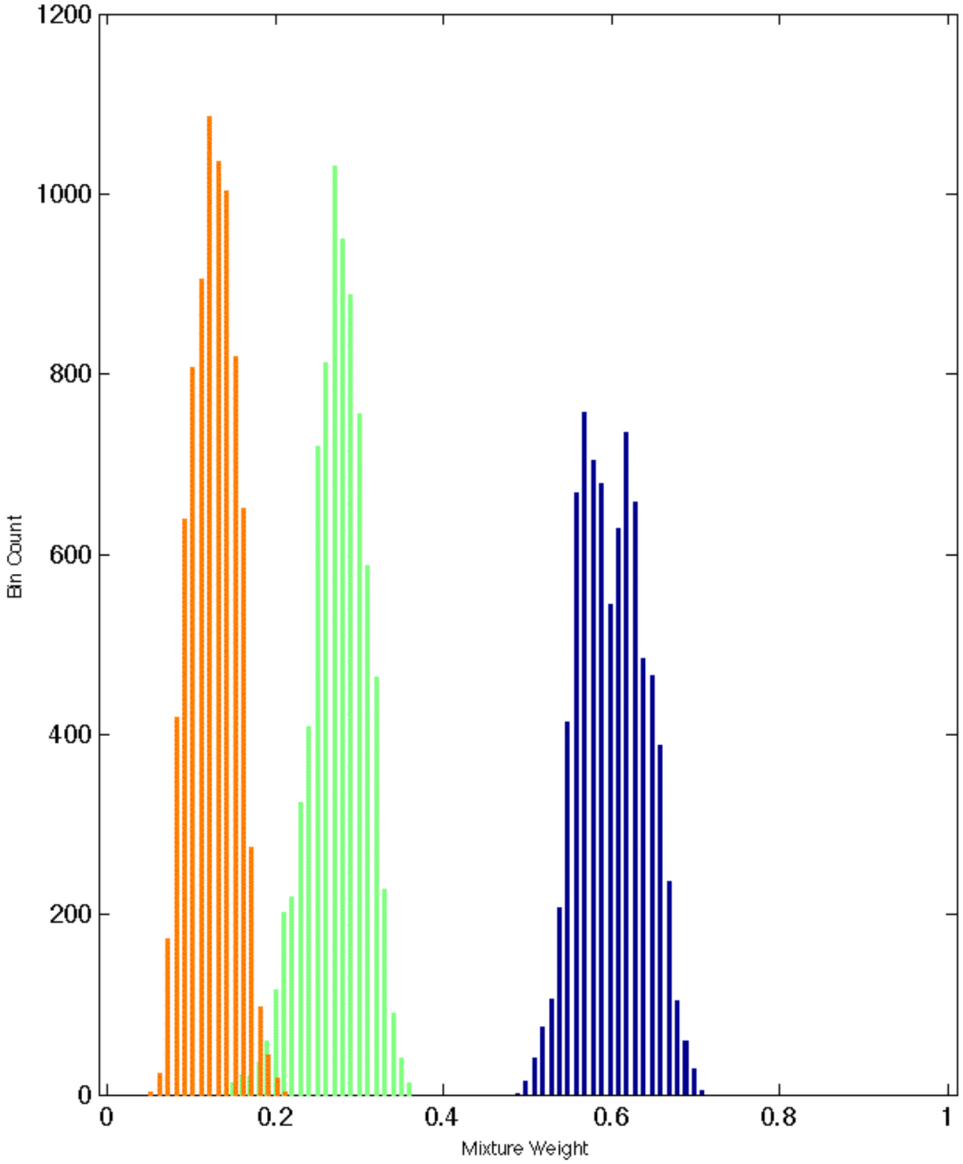

1200
1000
800
600
400
200
0
Bin Count
0
0.2
0.4
0.6
0.8
1
Mixture Weight

Figure 6, continued

B

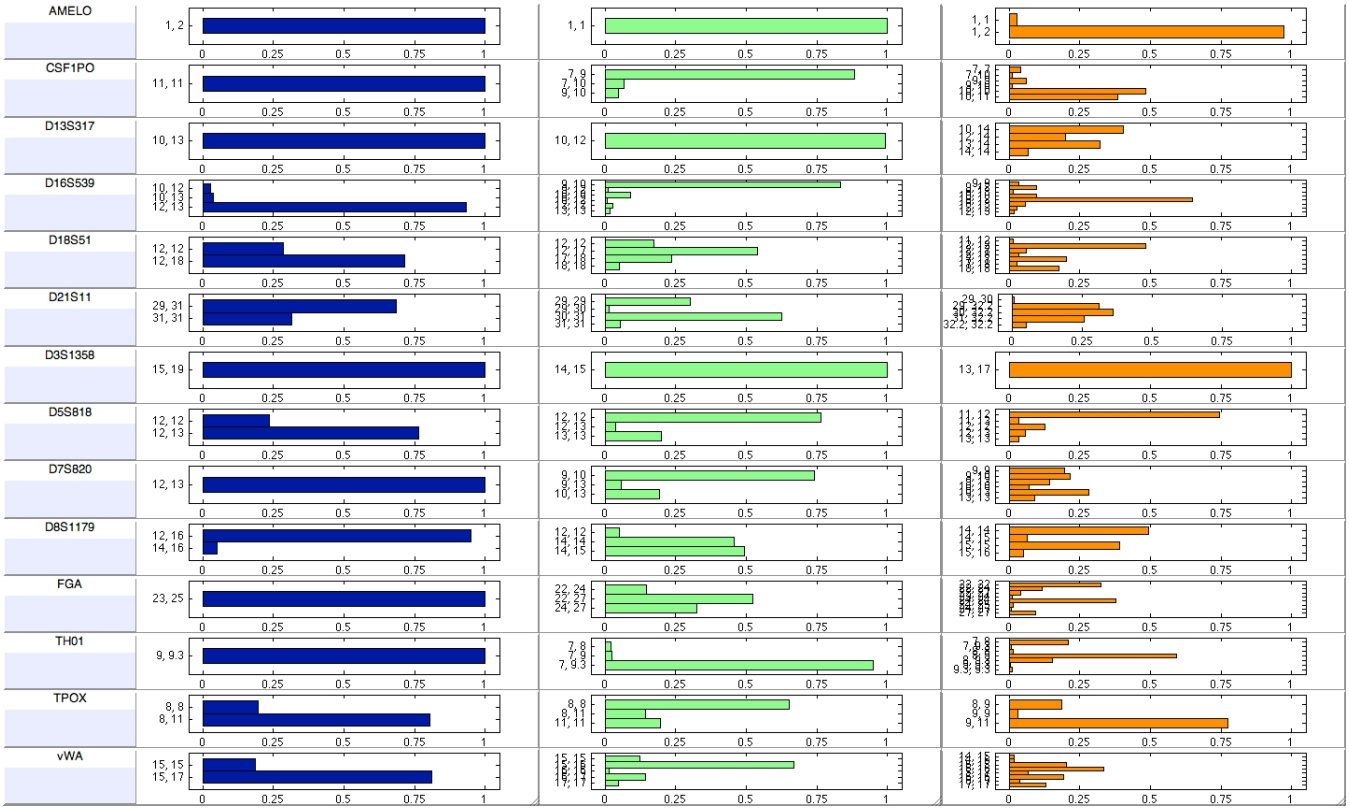
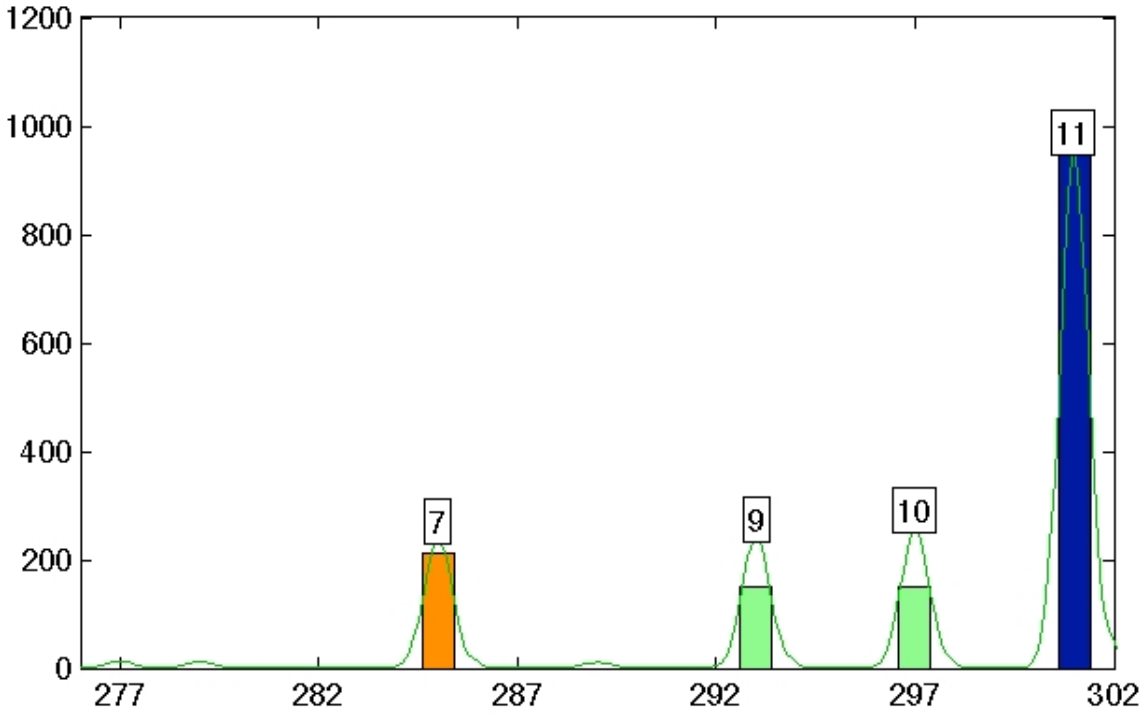


Figure 7

A


1200
1000
800
600
400
200
0
277
282
287
292
297
302
7
9
10
11

B

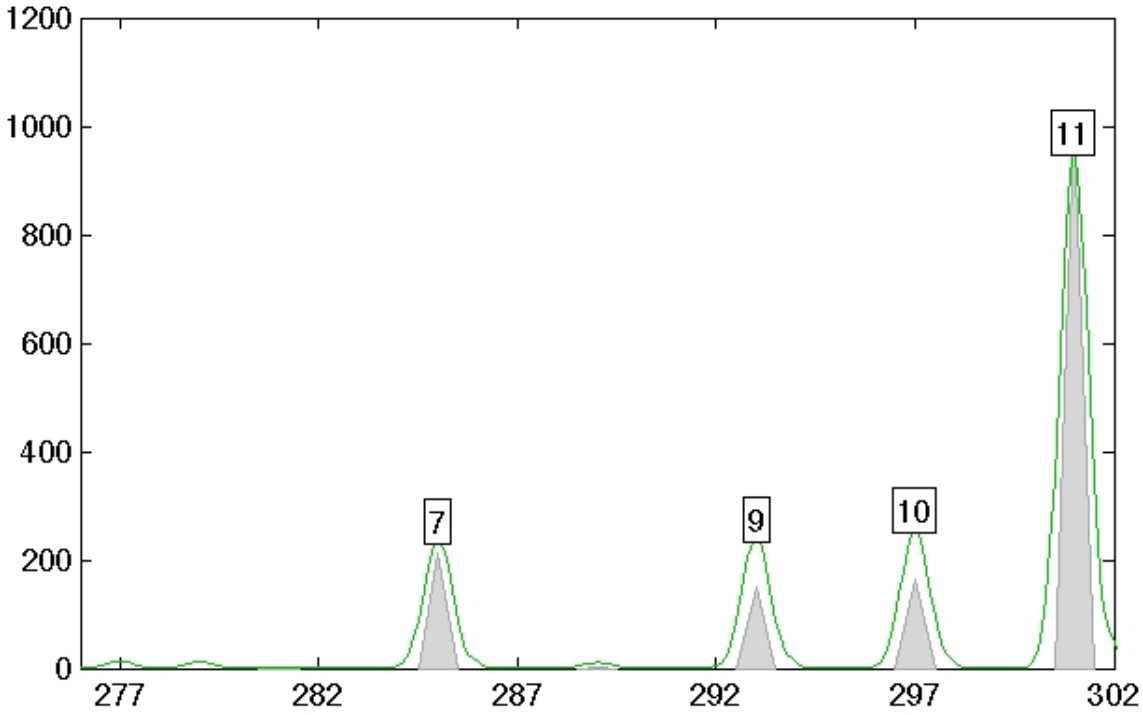

1200
1000
800
600
400
200
0
277
282
287
292
297
302
7
9
10
11

Figure 8

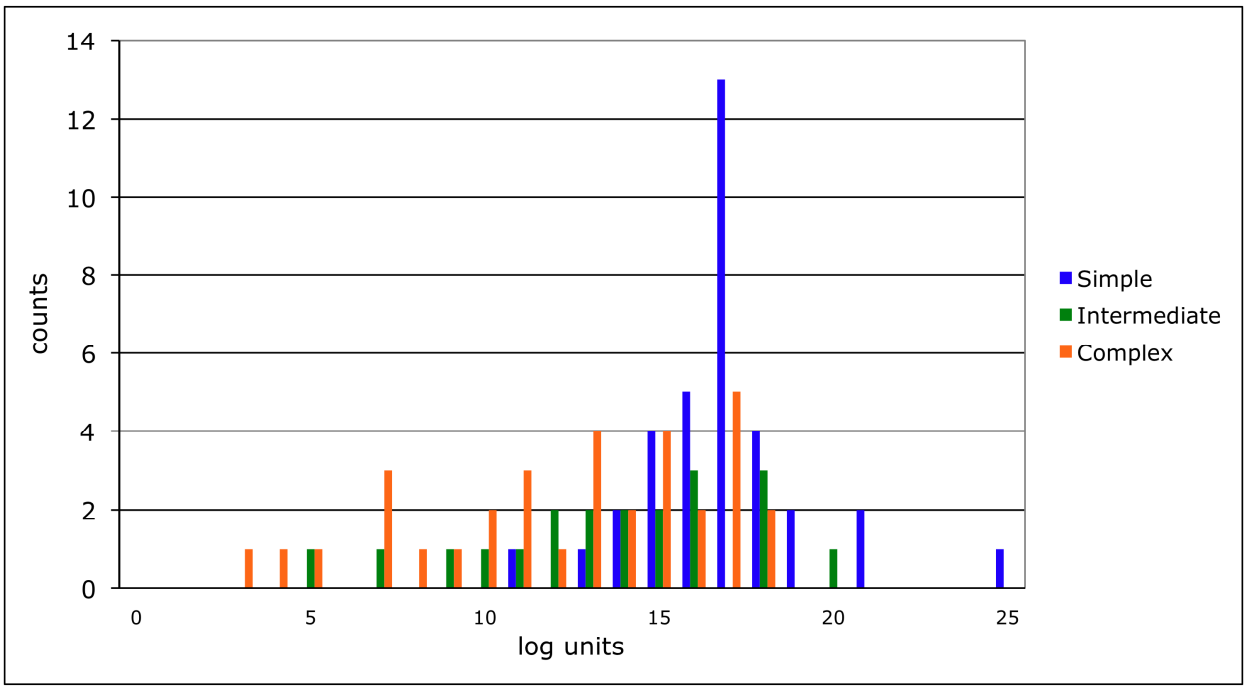

14
12
10
8
6
4
2
0
counts
0
5
10
15
20
25
log units
Simple
Intermediate
Complex

Figure 9

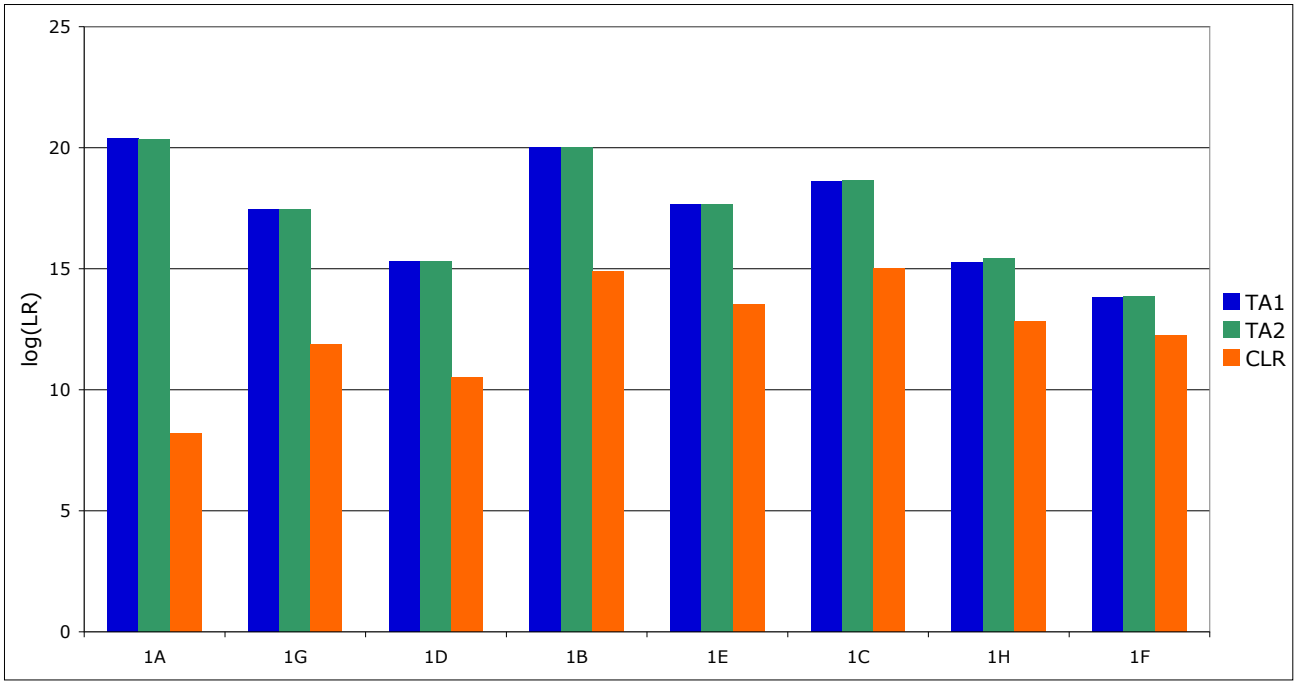


Figure 10

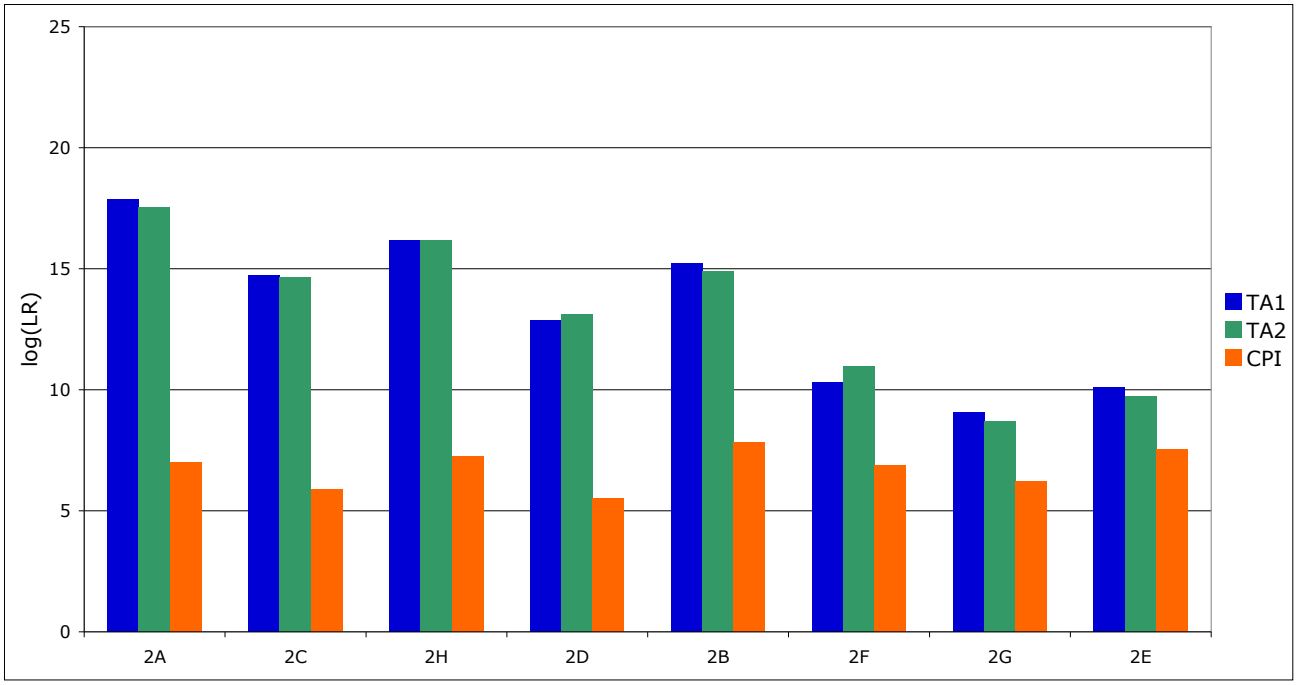


Figure 11

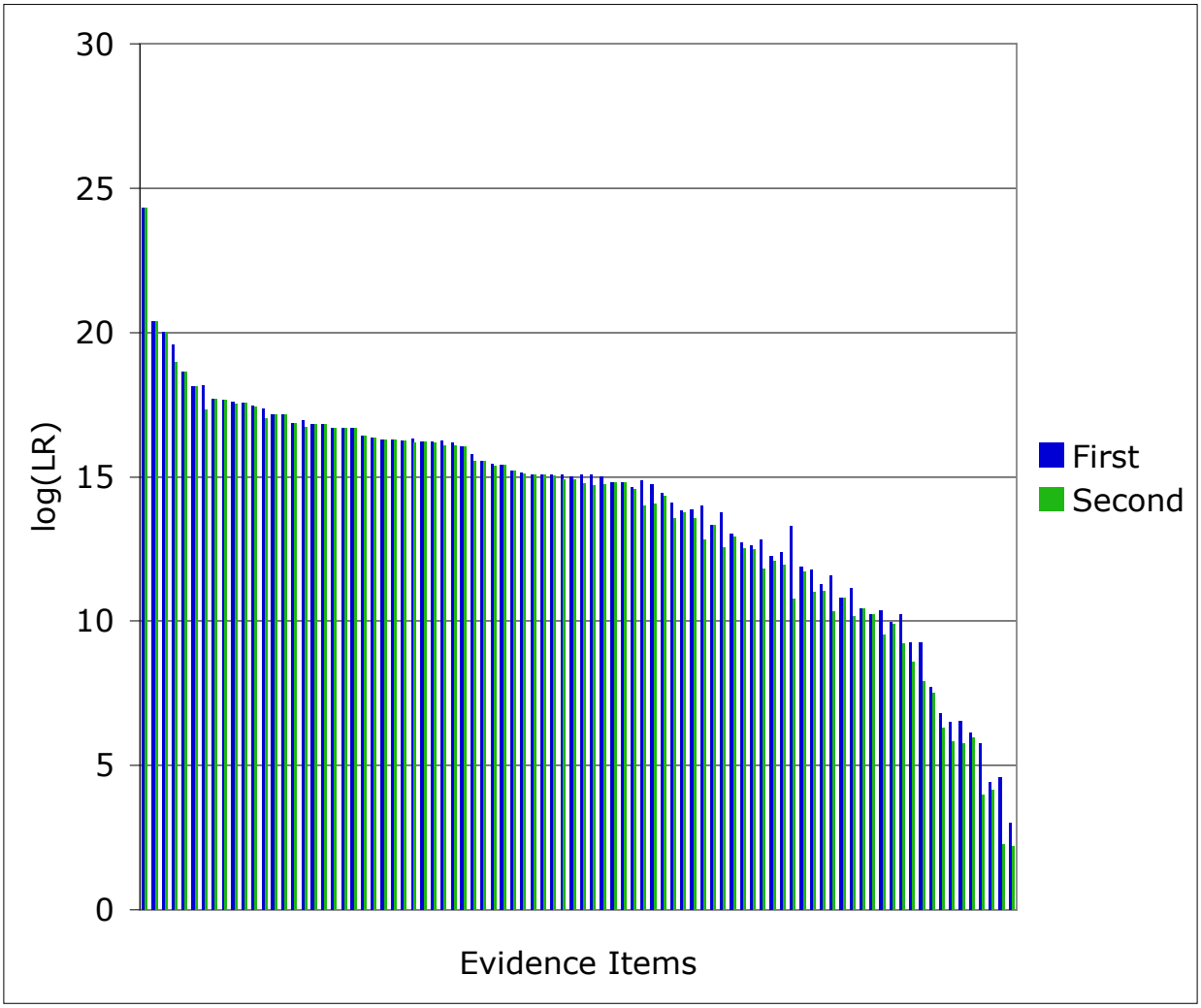


Figure 12

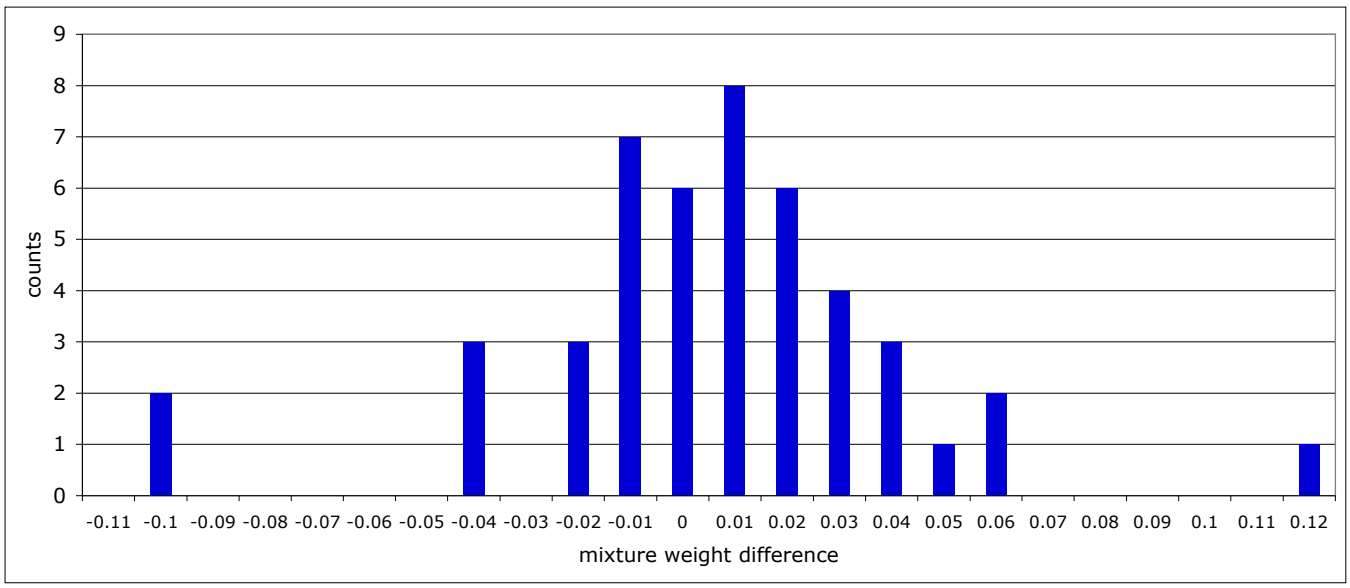
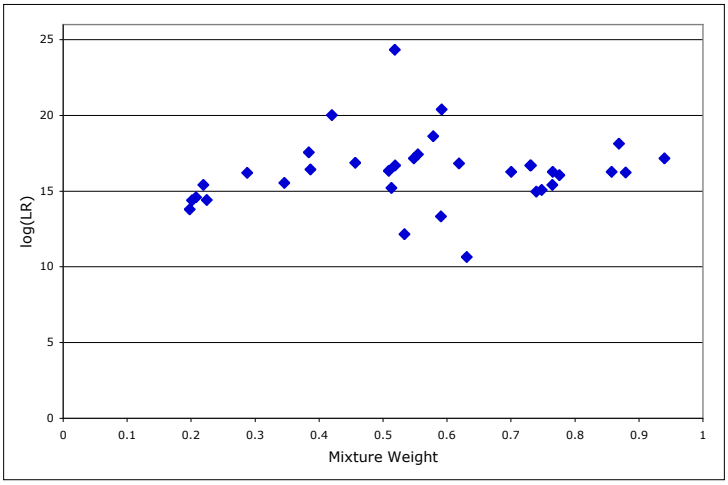
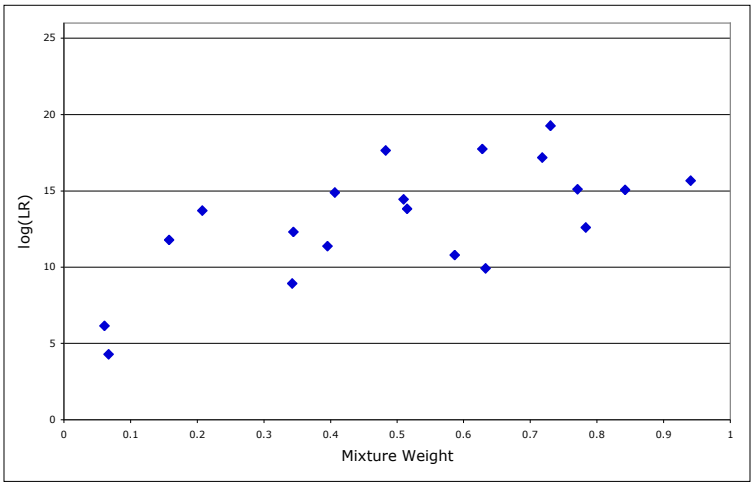

9
8
7
6
5
4
3
2
1
0
counts
-0.11 -0.1 -0.09 -0.08 -0.07 -0.06 -0.05 -0.04 -0.03 -0.02 -0.01 0 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 0.09 0.1 0.11 0.12
mixture weight difference

Figure 13

A



B



C

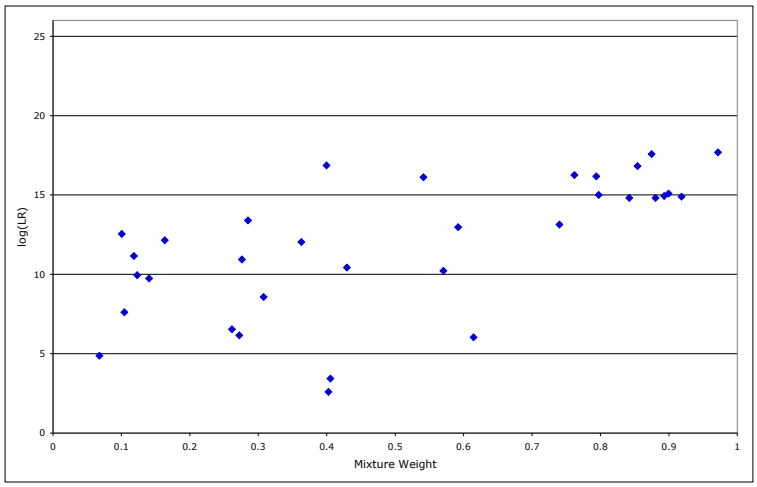


Table 1

A

Type	
sexual assault	32
homicide	3
assault	4
miscellaneous	2
	41

B

Source	
vaginal swab	17
anal swab	7
penile swab	1
semen stain	39
clothing item	10
bedding item	3
weapon	11
cigarette butt	2
condom	1
dried secretion	8
hair	3
bite mark	2
fingernail	9
blood stain	69
miscellaneous	24
	206

C

Category	
simple	35
intermediate	20
complex	33
	88

Table 2

	contributors			items		
complexity	total number in DNA item	how many knowns	how many unknowns	total number	number reported by lab with LR statistic	fraction reported by lab with LR statistic
simple	2	1	1	30	15	0.500
	2	0	2	5	2	0.400
				35	17	0.486
intermediate	2	1	1	10	2	0.200
	2	0	2	7	2	0.286
	3	1	2	1	0	0.000
	3	0	3	2	1	0.500
				20	5	0.250
complex	2	1	1	8	2	0.250
	2	0	2	17	0	0.000
	3	1	2	5	3	0.600
	3	0	3	3	2	0.667
				33	7	0.212
				88	29	0.330

Table 3

	Information average	Within-group standard deviation
Simple	16.28	0.102
Intermediate	13.14	0.255
Complex	11.87	0.437