# TrueAllele® Casework Validation on PowerPlex® 21 Mixture Data

Jennifer Hornyak, MS, William P. Allan, MS
and Mark W. Perlin, PhD, MD, PhD

Cybergenetics, Pittsburgh, PA

2 March 2014



*Corresponding Author:*

Mark W. Perlin, PhD, MD, PhD
Chief Scientific Officer
Cybergenetics
160 North Craig Street, Suite 210
Pittsburgh, PA 15213 USA
(412) 683-3004
(412) 683-3005 FAX
perlin@cybgen.com

## Table of Contents

## Introduction

In forensic science, DNA can be used in identifying perpetrators of a crime as well as in missing persons or paternity cases. Many times, DNA is found at crime scenes. Sometimes the DNA is single source, but, often, the DNA found exists as a mixture. A DNA mixture contains two or more contributors and can pose challenges for its interpretation since there are more genotype possibilities than in single source examination.

Short tandem repeat (STR) testing is done on DNA mixtures in order to express the allelic peaks that make up the genotypes of the contributors to the mixture. Currently, STR kits using 15 STR loci plus a sex determining marker are widely used. However, with current manual interpretation guidelines, crime laboratories are unable to interpret many mixtures using these kits. Therefore, forensic testing companies are working to develop STR kits that have 20 or more loci to increase the combined discriminating power of the loci and allow for a greater range of locus options for interpretation using manual methods.

PowerPlex® 21 is a 21 locus PCR amplification kit developed by the Promega Corporation (Madison, WI, USA). This kit incorporates 20 STR loci and the Amelogenin sex typing marker. The STR loci included in this kit are D3S1358, D1S1656, D6S1043, D13S317, D16S539, D18S51, D2S1338, CSF1PO, TH01, vWA, D21S11, D7S820, D5S818, TPOX, D8S1179, D12S391, D19S433, FGA, Penta D, and Penta E. PowerPlex 21 is more tolerant to DNA inhibition and has more discriminating power than the 16 locus kits. Recently, the New South Wales Forensic and Analytical Science

Service (FASS) has been using the PowerPlex 21 kit in their forensic DNA casework process for producing STR data in criminal cases.

STR testing of DNA mixtures produces quantitative patterns of peaks at the different loci. The heights of the peaks and their patterns correspond to the sum of the contributing genotypes where each allele appears in roughly the same proportion to its contributors DNA amount. There is natural variation in these peak heights from Polymerase Chain Reaction (PCR) artifacts and random amplification effects. Other variation can arise from baseline noise, degraded or inhibited DNA template, and pipetting error among other factors.

In order to overcome some of these PCR and STR variations, mathematical models have been developed to predict the DNA mixture patterns and their statistical variation. Many of these techniques use hierarchical Bayesian modeling to solve the probability equations and are done by computer. Some systems use Markov chain Monte Carlo (MCMC) sampling to solve problems that have many variables.

Computer interpretation of DNA mixtures has several potential advantages relative to manual review of the same data regarding sensitivity, specificity, and reproducibility. The computer is able to extract more information from the data, reduce false matches, quantify exclusion, and provide consistent results between independent analyses of the same data.

TrueAllele® Casework is a computer system developed to interpret DNA mixtures. TrueAllele was developed fifteen years ago by Cybergenetics (Pittsburgh, PA, USA) and has been used in criminal casework since 2009. Over a hundred TrueAllele reports have been issued in the United States and internationally for a variety of criminal cases. These reports have been used in criminal proceedings as evidence to help convict

criminals, exonerate the innocent, and obtain guilty pleas. TrueAllele has withstood admissibility hearings in three states in the United States as well as in the United Kingdom. Many validation studies have been done to test the TrueAllele system's sensitivity, specificity, and reproducibility on a variety of data types.

This TrueAllele Casework validation study assesses the system's performance on laboratory synthesized DNA mixtures of known composition amplified using the PowerPlex 21 STR panel. Three groups of mixtures were created from two, three, or four contributors in known mixing ratios. The reliability of the TrueAllele genotyping system was established for these PowerPlex 21 mixtures by assessing sensitivity, specificity, and reproducibility using a likelihood ratio match statistic, which is a standard measure of identification information.

## Methods and Materials

### *STR data*

In this study, there were three DNA sample groups that were comprised of different numbers of contributors: two, three, and four. Different individuals were used by FASS to construct these mixtures in known ratios. Table 1 shows the individuals and the mixing proportions used to create the mixture samples for each contributor group. The two and three contributor samples were all created at two DNA concentrations: 1 nanogram (ng) and 0.5 ng. The four contributor mixtures were created with a DNA concentration that ranged from 0.5 ng to 0.65 ng depending on the mixture sample.

For its own internal validation purposes, the FASS laboratory amplified the mixture samples using the Promega PowerPlex 21 STR panel. The fluorescently labeled amplicons were size separated on their AB 3500xl genetic analyzer, producing electropherogram data that were recorded as .hid files. The laboratory sent these data files to Cybergenetics in February of 2014. Reference data were also provided.

*Genotype inference*

The TrueAllele system uses a hierarchal Bayesian probability modeling approach when solving for the genotypes in a DNA mixture problem. TrueAllele considers all of the STR data and many other variables when solving DNA mixtures. Using MCMC statistical sampling, the system infers a probability distribution for the genotypes, mixture weights, and other variables based on the data. The TrueAllele inference is objective, considering only the data without seeing a subject reference during the inference process, and thorough, considering tens of thousands of possibilities for each variable in the mixture problem.

*Match information*

In order to quantify the strength of match between two genotypes (i.e., evidence and reference, or evidence and evidence), TrueAllele makes a comparison between the two genotypes, relative to a population genotype. This match comparison is made only after the computer has objectively inferred the contributor genotypes to a mixture problem.

The match information for each known contributor in a mixture set was calculated as a likelihood ratio (LR). The logarithm of the LR, or log(LR), is a standard additive measure of information expressed in “ban” units. The log(LR) can be used to quantify the sensitivity, specificity, and reproducibility of TrueAllele's genotyping and match results. The computer can show its calculated LR values as numbers, words or pictures, providing multiple ways to explain the match statistics.

*Processing*

The .hid data files were processed through the TrueAllele Casework Visual User Interface (VUIer™) Analyze module, in order to quality check and quantitate the data peaks. The quality checked data were then uploaded to a TrueAllele database in the Data module.

A trained TrueAllele analyst downloaded both the mixture and reference data from the database and created the interpretation requests for each sample in the VUIer Request module. Each mixture sample was processed assuming the same number of unknown contributors as known from the study design. All mixture requests were processed at least twice with burn-in and read-out sampling times of 100,000 MCMC cycles. The reference requests were processed with burn-in and read-out sampling times of 500 cycles.

*Reporting*

After TrueAllele processing was complete, the inferred evidence genotypes were compared to the known reference genotypes relative to the Australian National Asian and Caucasian populations in order to calculate the log(LR) match statistics. This comparison was done in the VUIer Report module. The inferred contributor genotype's corresponding known reference was identified by the largest match score for that comparison. A total of 168 genotype comparisons were made from 62 mixture items across all contributor groups (Table 2).

The reported match statistic was the average log(LR) value of the two independent, replicated computer runs. The minimum match statistic of the two Australian National populations was recorded. A co-ancestry coefficient of 1% was used. Sensitivity, specificity, and reproducibility were then assessed for each contributor group (2, 3, and 4).

## Results

### *Sensitivity*

Sensitivity measures the extent to which an interpretation method correctly includes a true contributor. TrueAllele's log(LR) sensitivity was evaluated separately for the two, three, and four contributor mixture groups.

TrueAllele's log(LR) match frequency distributions are shown for each contributor group in Figure 1. For each contributor group, the majority of the log(LR) match values fell to the right of zero information. This indicates high match sensitivity, with few false

exclusions. As shown, when the number of contributors increased (Figure 1, a, b, & c), their distributions shifted to the left.

Table 3 shows the sensitivity statistics for each contributor group. For the two contributor mixtures, the average log(LR) value was around 21 ban (a sextillion) with a standard deviation of 6.8 ban. With three contributors, the average log(LR) information fell to about 14 ban (a hundred trillion) with a standard deviation of 9.2 ban. For four contributor mixtures, the average log(LR) value was around 6 ban (a million) with a standard deviation of 7.7 ban. The leftward shift of information relates to the greater uncertainty present with more contributors to a mixture and is expected. There were some negative log(LR) values observed (Figure 1b & 1c, Table 3a row “min”), which indicate that a reference was falsely excluded as a contributor to its known mixture.

The number of false exclusions for each contributor group are shown in Table 3b. There were no false exclusions seen for two contributor mixtures. For three contributor mixtures, there were 5 total false exclusions out of the 72 genotype comparisons, giving a false exclusion rate of 6.94%. With four contributors, there were 10 false exclusions out of 40 genotype comparisons, which gave a false exclusion rate of 25%. Most of the false exclusions were the minor contributor portion of the known mixture where there can be more uncertainty.

*Specificity*

Specificity measures the extent to which an interpretation method correctly excludes a non-contributor. TrueAllele’s specificity was examined by looking at negative log(LR) distributions, which indicate the degree of exclusion. These distributions were created

by comparing the inferred evidence genotype for each calculated match (from the first replicate computer run) against 10,000 randomly generated genotypes from a population and recording the count of the non-matching log(LR) values. The Australian National Asian and Caucasian populations were used for a total of 20,000 comparisons for each evidence genotype.

Figure 2 shows the negative match information distributions for the inferred evidence genotypes for the two, three, and four contributor groups. There is a highly negative distribution (to the left of zero) for each contributor group, which indicates that TrueAllele non-contributor match statistics generally support true exclusions.

The summary specificity statistics for TrueAllele are shown in Table 4a. The average log(LR) value was around -30 ban for two contributors, -21 ban for three contributors, and -16 ban for four contributors. Referring to previous validation study results, as the number of loci increases, so does TrueAllele's exclusionary power.

TrueAllele showed a very low false inclusion rate across all the contributor mixture groups. With two contributor mixtures, there were 43 false positive scores out of the 1,120,000 genotype comparisons across both population groups tested, for an error rate of 0.00384%. For the three contributor mixtures, there were 1,207 false inclusions out of 1,440,000 genotype comparisons for an error rate of 0.0838%. With four contributor mixtures, 939 false positive scores were found out of 800,000 total comparisons, giving an error rate of 0.117%. The total false inclusion rate across the entire data set (all contributor groups) was 0.0651%.

The false inclusions all had log(LR)'s less than 5. Out of 3,360,000 total comparisons, there were just 2 false inclusions with an LR in the tens of thousands ($4 < \log(LR) < 5$). The false inclusion rate for matches at this level was 0.0000595%, which is

under one in a million. The majority of the false positives (1510 out of 2189) had a LR less than ten.

*Reproducibility*

Reproducibility measures how precisely the computer can repeat its results between independent runs on the same mixture data. To assess TrueAllele's reproducibility, comparison was made between the log(LR) values obtained between duplicate TrueAllele runs on the same data.

Figure 3 shows the reproducibility scatterplots for each contributor group. The points indicate the log(LR) value from an initial computer run (x-axis) and the duplicate computer run (y-axis) for each genotype match. The data points for each contributor group (Figure 3 a, b, & c) reside near the equal information line (i.e., x = y). Thus the analyses appear visually to be reproducible. As contributor number increased from two to three, and from three to four, the scatterplot width increased, indicating that the results were more reproducible with fewer contributors.

To quantify the reproducibility between computer runs, the within-group standard deviations were calculated for each contributor group. Table 5 shows these values along with the total mean and standard deviation for each contributor group. The within-group standard deviation for the two contributor group was 0.379 ban. For three contributors, this value was 0.702 ban. With four contributors, the within-group standard deviation was 0.764 ban. Since the within-group standard deviations across each contributor group (2, 3, and 4) are less than 1 ban, the TrueAllele interpretation method

shows good reproducibility of its match statistic log(LR) values regardless of contributor assumption or match score (positive or negative).

## Conclusion

DNA mixtures are found at many crime scenes, and can be significant to a criminal investigation. TrueAllele computer interpretation of mixtures can help preserve the identification information present in the data. This validation study examined mixtures of known composition having two, three, and four contributors that were amplified using the PowerPlex 21 multiplex kit. On these mixtures, Cybergenetics TrueAllele Casework system was found to be sensitive, specific, and reproducible.

The results presented in this study show how TrueAllele can be a useful, accurate, and reliable tool for interpreting DNA mixtures in forensic casework. The results also validate TrueAllele's accuracy and reliability for interpreting DNA mixture data amplified using PowerPlex 21, and mixture samples processed by the FASS laboratory.

## Tables

**Table 1: Study Design**. Information regarding the construction of the mixture samples for the two, three, and four contributor groups. Each mixture sample was created using different individuals (Contributor 1, 2, 3, or 4) at different mixing proportions (Ratio).

| Sample | Ratio | ncon | Contributor 1 | Contributor 2 | Contributor 3 | Contributor 4 |
|---|---|---|---|---|---|---|
| A359953 | 1:10 | 2 | A353007 | A352979 | | |
| A359954 | 1:10 | 2 | A353007 | A352979 | | |
| A359955 | 1:5 | 2 | A353007 | A352979 | | |
| A359956 | 1:5 | 2 | A353007 | A352979 | | |
| A359957 | 1:2 | 2 | A353007 | A352979 | | |
| A359958 | 1:2 | 2 | A353007 | A352979 | | |
| A359959 | 1:1 | 2 | A353007 | A352979 | | |
| A359960 | 1:1 | 2 | A353007 | A352979 | | |
| A359961 | 2:1 | 2 | A353007 | A352979 | | |
| A359962 | 2:1 | 2 | A353007 | A352979 | | |
| A359963 | 5:1 | 2 | A353007 | A352979 | | |
| A359964 | 5:1 | 2 | A353007 | A352979 | | |
| A359965 | 10:1 | 2 | A353007 | A352979 | | |
| A359966 | 10:1 | 2 | A353007 | A352979 | | |
| A359967 | 1:5:10 | 3 | A353007 | A352979 | A352990 | |
| A359968 | 1:5:10 | 3 | A353007 | A352979 | A352990 | |
| A359969 | 5:1:10 | 3 | A353007 | A352979 | A352990 | |
| A359970 | 5:1:10 | 3 | A353007 | A352979 | A352990 | |
| A359971 | 10:5:1 | 3 | A353007 | A352979 | A352990 | |
| A359972 | 10:5:1 | 3 | A353007 | A352979 | A352990 | |
| A359973 | 10:1:5 | 3 | A353007 | A352979 | A352990 | |
| A359974 | 10:1:5 | 3 | A353007 | A352979 | A352990 | |
| A359975 | 1:10:5 | 3 | A353007 | A352979 | A352990 | |
| A359976 | 1:10:5 | 3 | A353007 | A352979 | A352990 | |
| A359977 | 5:10:1 | 3 | A353007 | A352979 | A352990 | |
| A359978 | 5:10:1 | 3 | A353007 | A352979 | A352990 | |
| A403357 | 6:2:1:1 | 4 | A322020 | A273168 | A119720 | A274985 |
| A403359 | 4:4:1:1 | 4 | A322020 | A273168 | A119720 | A274985 |
| A403360 | 5:5:1:1 | 4 | A025331 | A320212 | A111626 | A274983 |
| A403362 | 8:2:1:1 | 4 | A025331 | A320212 | A111626 | A274983 |
| A403363 | 1:1:1:1 | 4 | A113187 | A273762 | A020024 | A279459 |
| A403364 | 12:8:1:1 | 4 | A113187 | A273762 | A020024 | A279459 |
| A403365 | 7:1:1:1 | 4 | A111135 | A274280 | A276880 | A281527 |
| A403366 | 4:4:4:1 | 4 | A111135 | A274280 | A276880 | A281527 |
| A403367 | 4:3:2:1 | 4 | A069768 | A275503 | A315860 | A276291 |
| A403368 | 6:4:2:1 | 4 | A069768 | A275503 | A315860 | A276291 |

**Table 2: Item and Genotype Totals.** The item and genotype totals, along with the overall totals, for each contributor group are listed. Overall, there were 168 genotype comparisons from 62 mixture samples.

| ncon | Item total | Genotype total |
|---|---|---|
| 2 | 28 | 56 |
| 3 | 24 | 72 |
| 4 | 10 | 40 |
| Overall | 62 | 168 |

**Table 3: Sensitivity.** Statistics were calculated for the 2, 3, and 4 contributor groups. Table (a) shows the count as well as the log(LR) minimum, mean, median, standard deviation, and maximum for the different contributor groups in ban units. Table (b) shows the number of false exclusions that occurred in the indicated log(LR) intervals (i.e., "-1" indicates the interval [-1, 0]).

(a) Summary statistics

| **ncon** | **2** | **3** | **4** |
|---|---|---|---|
| N = | 56 | 72 | 40 |
| min | 4.949 | -3.214 | -7.944 |
| mean | 20.855 | 14.451 | 5.709 |
| median | 21.261 | 16.916 | 5.969 |
| std dev | 6.800 | 9.173 | 7.669 |
| max | 29.445 | 28.334 | 24.537 |

(b) False exclusions

| **ncon** | **2** | **3** | **4** |
|---|---|---|---|
| -1 | 0 | 2 | 2 |
| -2 | 0 | 1 | 2 |
| -3 | 0 | 1 | 1 |
| -4 | 0 | 1 | 2 |
| -5 | 0 | 0 | 1 |
| -6 | 0 | 0 | 1 |
| -7 | 0 | 0 | 0 |
| -8 | 0 | 0 | 1 |
| *Total* | *0* | *5* | *10* |

**Table 4: Specificity.** Statistics were calculated for each contributor group across the two Australian National populations. Table (a) shows the number of comparisons as well as the log(LR) minimum, mean, maximum, and standard deviation values expressed in ban units. Table (b) shows the number of false inclusions occurring in the indicated log(LR) interval (i.e., “0” indicates the interval [0, 1]).

(a) Summary statistics

| **ncon** | **2** | | **3** | | **4** | |
|---|---|---|---|---|---|---|
| ethnicity | ASN | CAU | ASN | CAU | ASN | CAU |
| N = | 560,000 | 560,000 | 720,000 | 720,000 | 400,000 | 400,000 |
| min | -40.000 | -40.000 | -40.000 | -40.000 | -40.000 | -40.000 |
| mean | -29.542 | -30.916 | -20.703 | -21.858 | -16.096 | -15.424 |
| max | 2.559 | 3.255 | 4.933 | 3.723 | 3.780 | 3.954 |
| std | 8.178 | 7.958 | 10.209 | 10.240 | 8.553 | 8.587 |

(b) False inclusions

| **ncon** | **2** | | **3** | | **4** | |
|---|---|---|---|---|---|---|
| ethnicity | ASN | CAU | ASN | CAU | ASN | CAU |
| 0 | 17 | 14 | 527 | 297 | 331 | 324 |
| 1 | 7 | 3 | 161 | 112 | 93 | 119 |
| 2 | 1 | 0 | 50 | 32 | 25 | 34 |
| 3 | 0 | 1 | 11 | 15 | 5 | 8 |
| 4 | 0 | 0 | 2 | 0 | 0 | 0 |
| *Total* | *25* | *18* | *751* | *456* | *454* | *485* |

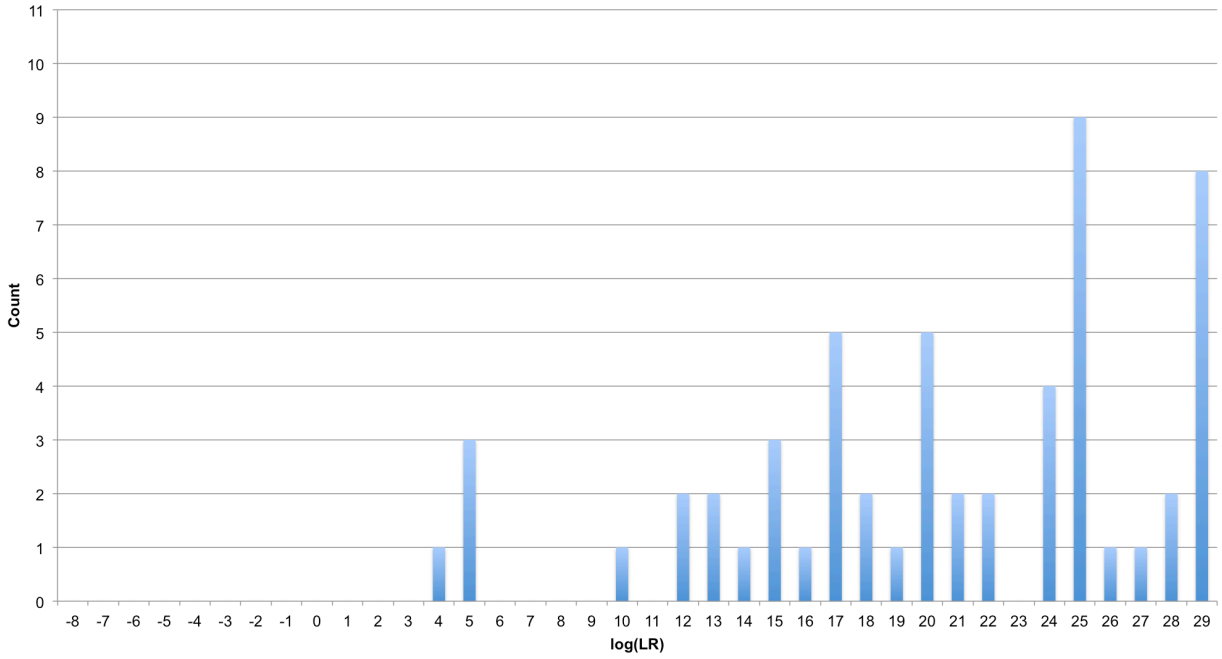
**Table 5: Reproducibility.** For each contributor group, the log(LR) mean (μ), standard deviation (σ), and within-group standard deviation ($\sigma_w$) measure of reproducibility are shown. The reproducibility results were calculated from two independent TrueAllele runs for each sample.

| **ncon** | **2** | **3** | **4** |
|---|---|---|---|
| μ | 20.855 | 14.451 | 5.708 |
| σ | 6.780 | 9.168 | 7.659 |
| $\sigma_w$ | 0.379 | 0.702 | 0.764 |

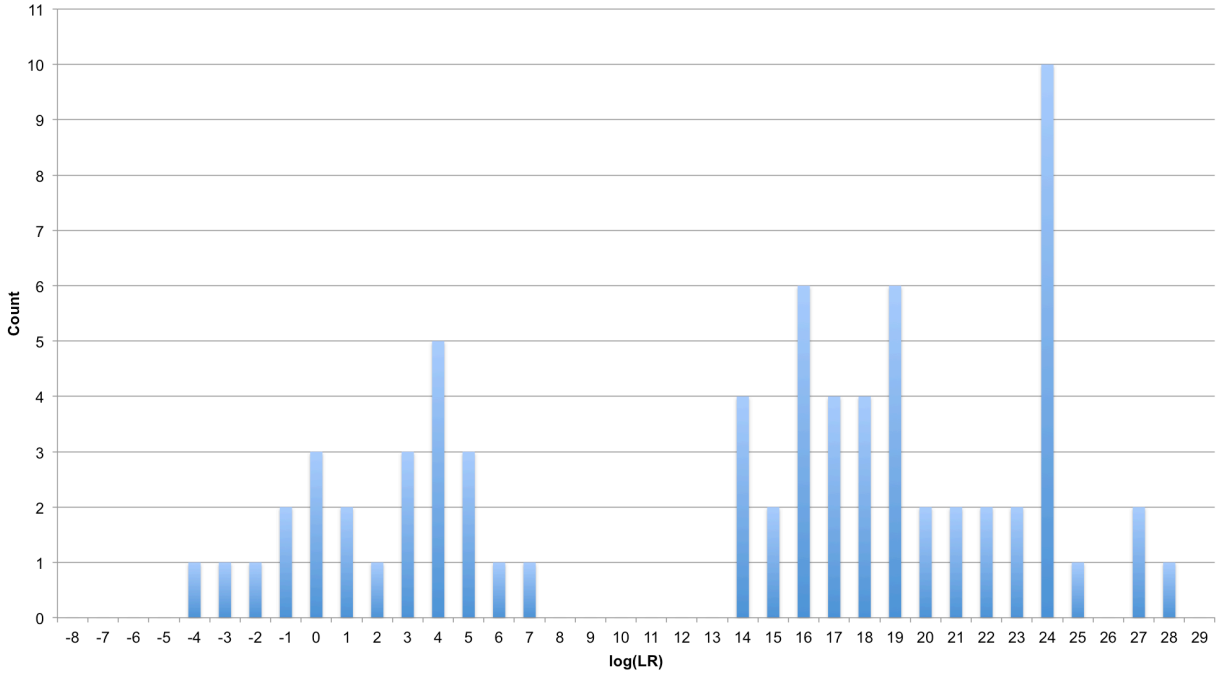
## Figures

**Figure 1: Sensitivity.** Histograms show the log(LR) genotype match distributions for (a) 2 contributor mixtures, (b) 3 contributor mixtures, and (c) 4 contributor mixtures. The x-axis displays the log(LR) bins while the y-axis displays the number of match statistics in that bin.

(a) 2 contributors




(b) 3 contributors



## (c) 4 contributors

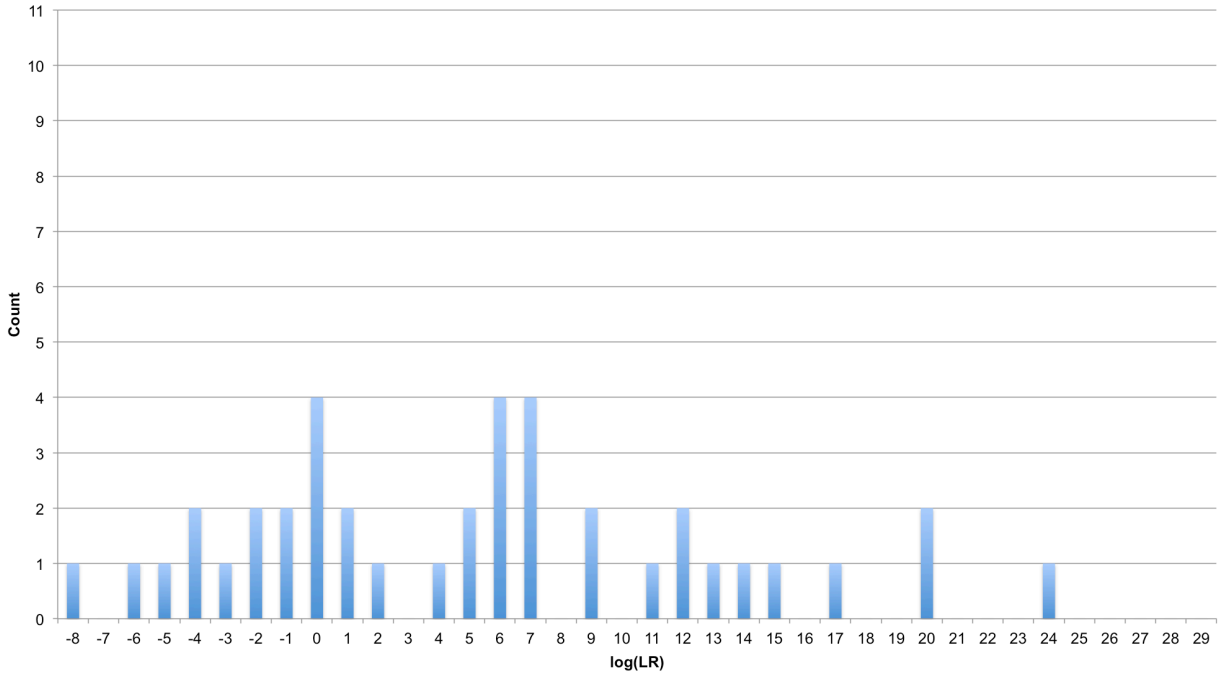
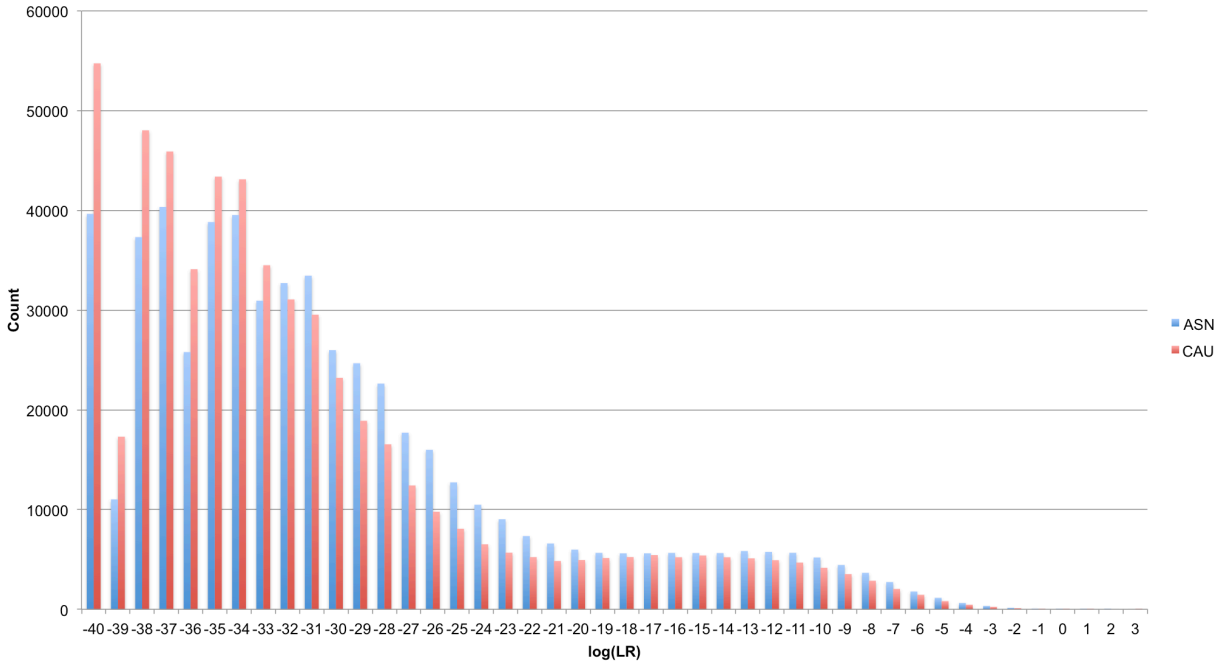
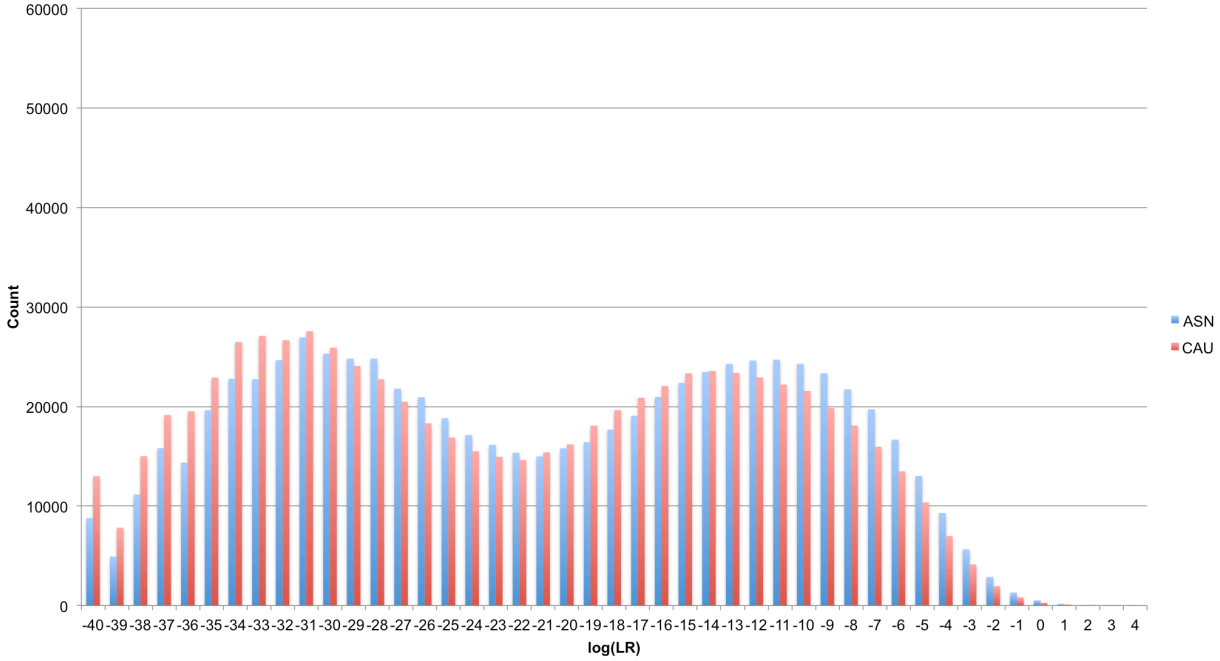


**Figure 2. Specificity.** Histograms show the log(LR) genotype match distributions for (a) 2 contributor mixtures, (b) 3 contributor mixtures, and (c) 4 contributor mixtures, relative to 10,000 randomly generated profiles. Each ethnic population is shown in a different color.

(a) 2 contributors




(b) 3 contributors



## (c) 4 contributors

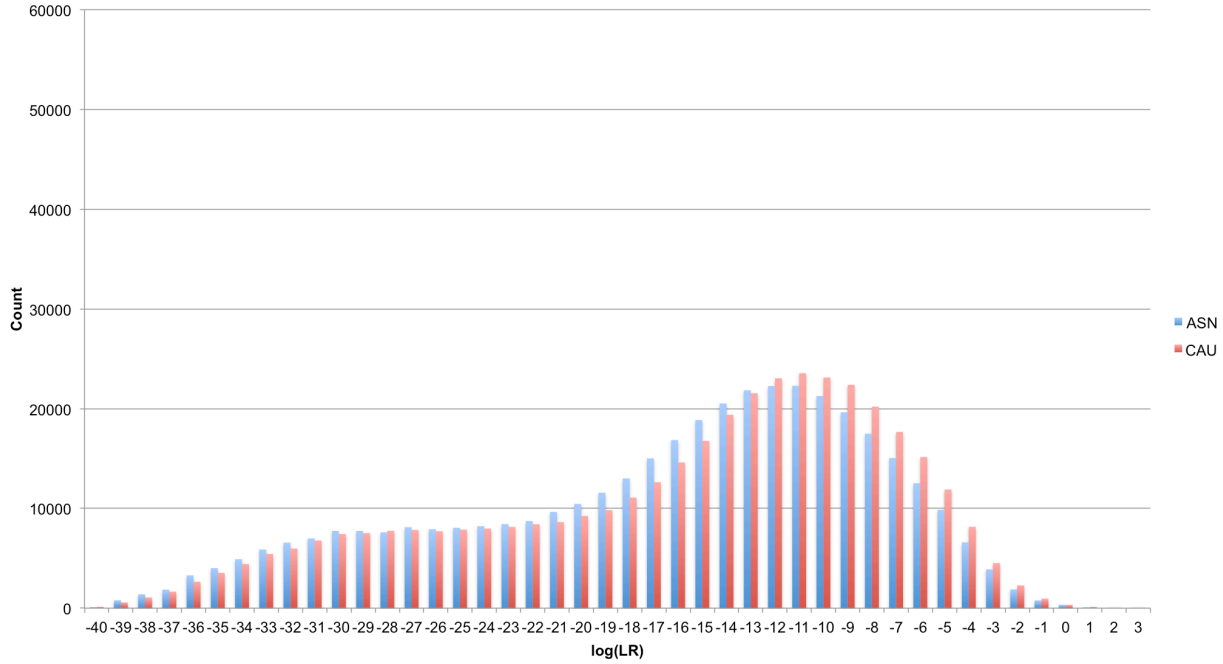
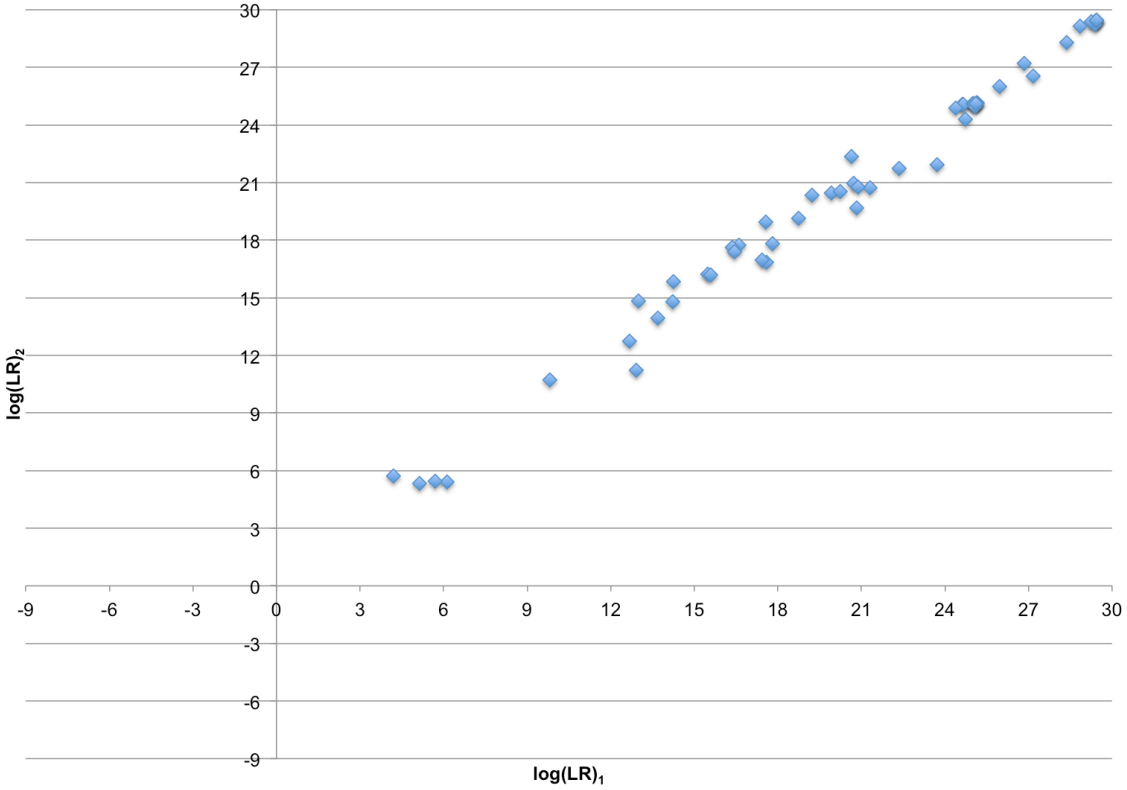
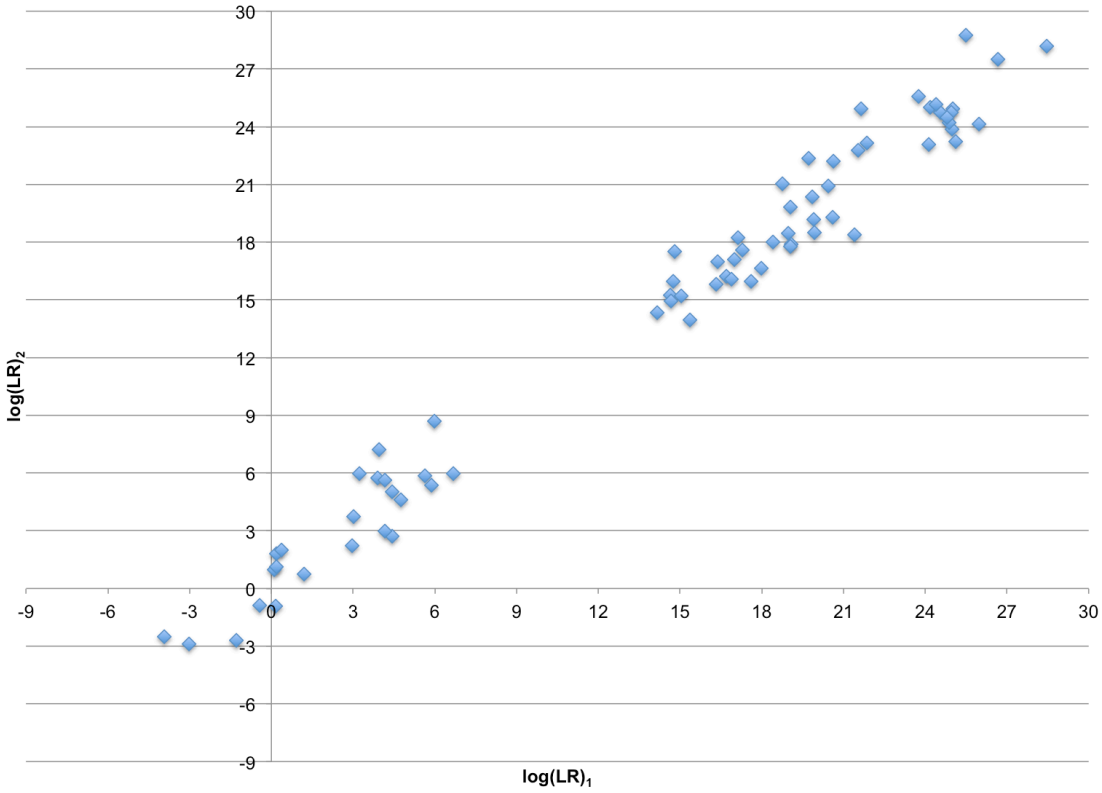


**Figure 3. Reproducibility.** Scatterplots show the log(LR) genotype match values for 2 independent computer runs on the same mixture sample for (a) 2 contributor mixtures, (b) 3 contributor mixtures, and (c) 4 contributor mixtures. Each point depicts the two match values from the first (x) and second (y) computer run.

(a) 2 contributors



(b) 3 contributors




(c) 4 contributors