TrueAllele® Validation on Identifiler® Plus Mixture Data

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

Cybergenetics, Pittsburgh, PA

August 21, 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

DNA is a common form of forensic evidence found at many crime scenes that can be used to identify perpetrators or others associated with crimes as well in missing persons or paternity cases. Often, crime scene DNA consists of a mixture of two or more contributors. A DNA mixture can pose challenges for interpretation since there are more genotype possibilities, which leads to uncertainty.

Laboratories conduct short tandem repeat (STR) testing on DNA mixtures to express the allelic peaks that are present in the genetic types of the contributors to the mixture. Currently, STR kits that use 15 STR loci plus a gender-determining marker are used for most DNA casework in the United States.

Identifiler® Plus is one 15 STR locus kit that was developed by Life Technologies (Grand Island, NY). The loci tested are Amelogenin, D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, VWA, TPOX, D18S51, D5S818, and FGA. The Identifiler® Plus kit is said to have a greater sensitivity, cleaner baseline, and better performance on mixtures, as well as being able to overcome higher levels of PCR inhibition.

STR testing of DNA mixtures produces quantitative patterns of peaks at the different loci tested. The peak heights and their patterns correspond to the sum of the contributing genotypes to a mixture. There is natural variation in the peak heights and patterns from Polymerase Chain Reaction (PCR) amplification effects and artifacts.

In order to aid in the interpretation of DNA mixtures with natural variation of the peak data, computer programs that use mathematical models to predict DNA patterns and their variation can be used. These computers are useful where manual interpretation may not be informative or yield results. Some programs use hierarchical Bayesian modeling and Markov chain Monte Carlo (MCMC) to solve probability equations with many variables.

Computer interpretation of DNA mixtures has advantages over human review when considering sensitivity, specificity, and reproducibility. The computer is able to get more information from the same data, while reducing false matches, quantifying exclusions, and providing consistent results between independent analyses.

TrueAllele® Casework is a computer system used to interpret DNA mixtures that was developed about twenty years ago by Cybergenetics (Pittsburgh, PA). TrueAllele has been used in criminal casework since 2009, with over 200 TrueAllele reports issued in the United States and internationally for a variety of criminal cases. TrueAllele reports have been used in criminal proceedings as evidence, and TrueAllele has withstood admissibility hearings in three states as well as in the United Kingdom. There have also been over twenty validation studies done to test the TrueAllele system's sensitivity, specificity, and reproducibility on a variety of data sets and types.

This TrueAllele Casework validation study assesses the system's performance on DNA mixtures of known composition developed by the Louisiana State Police Crime Laboratory that were amplified using the Identifiler® Plus STR kit. Three mixture sets were created from either two or three contributors in known mixing ratios at 5, 10, and 15 second injection times. The reliability of the TrueAllele system was established for these mixtures by assessing sensitivity, specificity, and reproducibility using the likelihood ratio match statistic, which is a measure of identification information.

Materials and Methods

STR data

The Louisiana State Police Crime Laboratory created a total of 15 different mixtures of known composition. Two groups of mixtures (Mix1 and Mix2) were created from two contributors in known proportions. There was one three-contributor group (Mix3) as well, for a total of three mixture groups. For each mixture group, the samples were amplified

twice and injected at five, ten, and fifteen seconds. Table 1 shows the individuals and mixing proportions used to create each mixture sample.

The mixture and reference samples were amplified using the Identifiler® Plus STR panel. The ABI 3130xl genetic analyzer was used to size separate the fluorescently labeled amplicons, and this produced electropherogram data recorded as .fsa files. The laboratory sent these data files and descriptions to Cybergenetics in June of 2014.

Genotype inference

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

Match information

To quantify the strength of match between two genotypes, TrueAllele makes a comparison between the two genotypes, relative to a population genotype. This match comparison is made only after the contributor genotypes have been objectively inferred from the data.

The match information for each known contributor of a mixture is calculated as a likelihood ratio (LR). The logarithm of the LR, or log(LR), is an additive measure of information, expressed in “ban” units. The log(LR) can be used to quantify the sensitivity, specificity, and reproducibility of the computer’s genotyping and match results.

Processing

The .fsa data files sent to Cybergenetics were processed through the TrueAllele Casework Visual User Interface (VUIer™) Analyze module, in order to quality check and quantitate the data peaks present. The quality-checked peaks 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 interpretation requests for each sample in the VUIer Request module. The Mix1 and Mix2 samples were processed assuming two unknown contributors. The Mix3 samples were processed assuming three unknown contributors. All mixture requests were processed with burn-in and read-out sampling times of 100K/100K. All results were run in duplicate, and additional replicates were run as needed.

Reporting

After TrueAllele processing was finished, the inferred mixture evidence genotypes were compared to the known reference genotypes relative to the United States FBI African American, Caucasian, and Hispanic ethnic populations to calculate match statistics. A co-ancestry coefficient of 1% was used. This calculation was done in the VUIer Report module. The inferred contributor genotype matching to each known reference was chosen based on match statistic, KL, and mixture weight.

The reported match statistic was the minimum value of the three populations was recorded for each replicate. Sensitivity, specificity, and reproducibility were assessed for each mixture group and injection time. Both amplifications were considered as part of the contributor set.

A total of 180 genotype comparisons were made from the 15 mixture items from the different mixture sets, amplifications, and three injection times.

Results

Sensitivity

Sensitivity describes the extent to which a true contributor is correctly included in a DNA mixture. In this study, sensitivity statistics were calculated using the average match statistic between two concordant, independent computer runs. The count, minimum, average, median, standard deviation, and maximum match statistic values for each contributor group and injection time were calculated (Table 2a). In addition, the match statistics were binned by log(LR) value and plotted in a frequency distribution (Figure 1). The number of false exclusions was also recorded (Table 2b). The average match information for the two contributors at all three injection times (5 sec, 10 sec, and 15 sec) was around 15 ban (a quadrillion), with five false exclusions. For three contributors at all injection times, the average match information was around 6 ban (a million), with no false exclusions observed. The overall false exclusion rate was 3%.

Specificity

Specificity describes the extent to which a true non-contributor is correctly excluded from a DNA mixture. Specificity statistics were calculated by comparing the inferred genotype for each calculated match from the first replicate against 10,000 randomly generated profiles from a population. The statistics and counts for the non-matching log(LR) values were recorded (Table 3a). The number of false positives was also recorded (Table 3b). The United States FBI African American, Caucasian, and Hispanic populations were used for a total of 30,000 comparisons for each evidence genotype. Frequency histograms were also produced (Figure 2).

There were 4,860,000 comparisons for the two contributor group (all injection times), and a total of 540,000 comparisons for the three contributor group (all injection times). The average non-match information for two contributors was around -25 ban and

around -15 ban for three contributors. The overall false positive rate was 0.001%. Only two false positives were observed with a log(LR) value greater than 2 ban.

Reproducibility

Reproducibility describes how well a method produces identical results on independent analyses. Since TrueAllele uses MCMC statistical sampling, some variation can be expected between computer runs. This variation can be quantified by calculating the within-group standard deviation for a data set.

To examine reproducibility, a comparison was made between the log(LR) values obtained between independent, replicate runs on the same data (Figure 3). The within-group standard deviations were calculated and recorded (Table 4). The within-group standard deviations were less than 0.3 ban for two contributors and less than 0.2 ban for three contributors. Overall, the runs differed by less than a factor of 2.

Conclusion

DNA mixtures are a common form of evidence in criminal cases. TrueAllele computer interpretation of mixtures preserves the identification information present in the data. This validation study examined mixtures of known composition having two or three contributors at five, ten, and fifteen second injection times that were amplified using the Indentifiler® Plus multiplex kit. On this mixture data set, 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 Identifiler® Plus, and mixture samples processed by the Louisiana State Police Crime Laboratory.

Table 1: Design. The mixture samples for each set were constructed as outlined in the table. Each mixture sample was created using different individuals (Contributor) at different mixing proportions (Ratio). In addition, each sample was amplified twice and injected for 5, 10, and 15 seconds. Note: One of the amplifications for the Mix2_1to4 sample did not amplify properly, and so was not used in this study.

Sample	Ratio	Set	ncon	Contributor 1	Contributor 2	Contributor 3
Mix1_19to1	19:1	Mix1	2	Mix1F1	Mix1M1	
Mix1_1to1	1:1	Mix1	2	Mix1F1	Mix1M1	
Mix1_1to19	1:19	Mix1	2	Mix1F1	Mix1M1	
Mix1_1to4	1:4	Mix1	2	Mix1F1	Mix1M1	
Mix1_1to9	1:9	Mix1	2	Mix1F1	Mix1M1	
Mix1_4to1	4:1	Mix1	2	Mix1F1	Mix1M1	
Mix1_9to1	9:1	Mix1	2	Mix1F1	Mix1M1	
Mix2_19to1	19:1	Mix2	2	Mix2F1	Mix2M1	
Mix2_1to1	1:1	Mix2	2	Mix2F1	Mix2M1	
Mix2_1to19	1:19	Mix2	2	Mix2F1	Mix2M1	
Mix2_1to4	1:4	Mix2	2	Mix2F1	Mix2M1	
Mix2_1to9	1:9	Mix2	2	Mix2F1	Mix2M1	
Mix2_4to1	4:1	Mix2	2	Mix2F1	Mix2M1	
Mix2_9to1	9:1	Mix2	2	Mix2F1	Mix2M1	
Mix3_1to1to1	1:1:1	Mix3	3	Mix3M1	Mix3M2	Mix3M3

Table 2: Sensitivity. Statistics were calculated for 2 and 3 contributors at 5, 10, and 15 second injection times. Table (a) shows the number of comparisons as well as the log(LR) minimum, mean, median, standard deviation, and maximum values. Table (b) shows the number of false exclusions occurring in each log(LR) bin where "0" indicates the interval [0,1).

(a) Summary statistics

ncon	**2**			**3**		
injection	**5 sec**	**10 sec**	**15 sec**	**5 sec**	**10 sec**	**15 sec**
N=	54	54	54	6	6	6
minimum	-5.249	-4.906	-1.564	3.801	4.108	4.356
mean	14.815	15.084	15.298	6.250	6.389	6.349
median	17.468	17.293	17.676	6.461	6.588	6.449
maximum	22.194	22.194	22.194	8.477	7.881	8.002
std dev	6.827	6.321	5.950	1.680	1.461	1.425

(b) False exclusions

ncon	**2**			**3**		
injection	**5 sec**	**10 sec**	**15 sec**	**5 sec**	**10 sec**	**15 sec**
-1	1	0	0	0	0	0
-2	0	0	1	0	0	0
-3	1	0	0	0	0	0
-4	0	0	0	0	0	0
-5	0	1	0	0	0	0
-6	1	0	0	0	0	0
Total	*3*	*1*	*1*	*0*	*0*	*0*

Table 3: Specificity. Statistics were calculated for 2 and 3 contributors at 5, 10, and 5 second injection times across all three United States FBI ethnic populations. Table (a) shows the number of comparisons along with the log(LR) minimum, mean, median, maximum, standard deviation, mu, and sigma values. Table (b) shows the number of false inclusions occurring in each log(LR) bin where "0" indicates the interval [0,1).

(a) Summary statistics

ncon	**2**								
injection	**5 sec**			**10 sec**			**15 sec**		
ethnicity	**BLK**	**CAU**	**HIS**	**BLK**	**CAU**	**HIS**	**BLK**	**CAU**	**HIS**
N =	540,000	540,000	540,000	540,000	540,000	540,000	540,000	540,000	540,000
minimum	-30.000	-30.000	-30.000	-30.000	-30.000	-30.000	-30.000	-30.000	-30.000
mean	-23.942	-23.178	-23.686	-24.122	-23.323	-23.831	-23.881	-23.086	-23.569
median	-24.862	-24.146	-24.702	-25.149	-24.251	-24.816	-24.908	-24.190	-24.713
maximum	2.447	3.412	1.721	0.566	0.872	1.154	0.429	1.462	0.874
std dev	5.206	5.523	5.392	5.069	5.433	5.322	5.267	5.661	5.577
positive	8	12	8	1	2	4	1	7	2
mu	-30.000	-30.000	-30.000	-30.000	-30.000	-30.000	-30.000	-30.000	-30.000
sigma	7.987	8.777	8.303	7.762	8.612	8.148	8.074	8.936	8.513

ncon	**3**								
injection	**5 sec**			**10 sec**			**15 sec**		
ethnicity	**BLK**	**CAU**	**HIS**	**BLK**	**CAU**	**HIS**	**BLK**	**CAU**	**HIS**
N =	60,000	60,000	60,000	60,000	60,000	60,000	60,000	60,000	60,000
minimum	-30.000	-27.847	-30.000	-30.000	-27.958	-30.000	-30.000	-28.026	-30.000
mean	-16.973	-14.537	-15.433	-17.031	-14.627	-15.516	-16.996	-14.548	-15.434
median	-16.995	-14.650	-15.458	-17.045	-14.733	-15.575	-17.008	-14.670	-15.470
maximum	0.353	0.979	1.226	0.116	0.851	0.206	0.353	0.873	0.508
std dev	4.127	3.989	3.991	4.125	3.983	3.984	4.128	3.985	3.994
positive	3	3	7	1	3	3	3	3	7
mu	-16.985	-14.538	-15.435	-17.044	-14.628	-15.518	-17.008	-14.549	-15.436
sigma	4.146	3.991	3.995	4.145	3.984	3.987	4.147	3.987	3.998

(b) False inclusions

ncon	**2**								
injection	**5 sec**			**10 sec**			**15 sec**		
ethnicity	**BLK**	**CAU**	**HIS**	**BLK**	**CAU**	**HIS**	**BLK**	**CAU**	**HIS**
0	7	10	6	1	2	3	1	6	2
1	0	1	2	0	0	1	0	1	0
2	1	0	0	0	0	0	0	0	0
3	0	1	0	0	0	0	0	0	0
Total	*8*	*12*	*8*	*1*	*2*	*4*	*1*	*7*	*2*

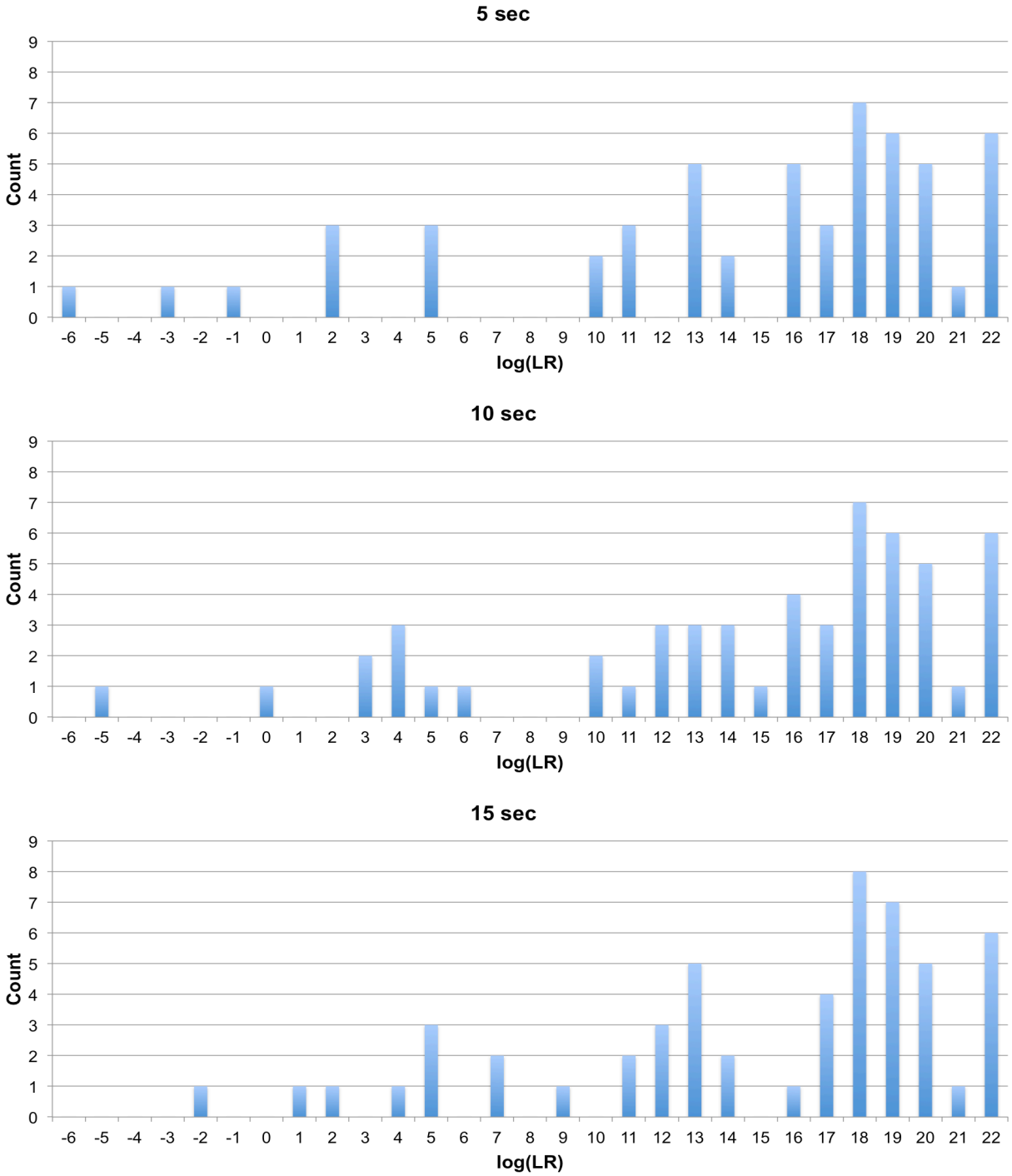
ncon	**3**								
injection	**5 sec**			**10 sec**			**15 sec**		
ethnicity	**BLK**	**CAU**	**HIS**	**BLK**	**CAU**	**HIS**	**BLK**	**CAU**	**HIS**
0	3	3	6	1	3	3	3	3	7
1	0	0	1	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0
Total	*3*	*3*	*7*	*1*	*3*	*3*	*3*	*3*	*7*

Table 4: Reproducibility. The mean (μ), standard deviation (σ) and within-group standard deviation (σ_w) measure of reproducibility are shown for both two and three contributors at 5, 10, and 15 second injection times.

ncon	**2**			**3**		
injection	**5 sec**	**10 sec**	**15 sec**	**5 sec**	**10 sec**	**15 sec**
μ	14.815	15.084	15.298	6.251	6.389	6.349
σ	6.799	6.297	5.926	1.611	1.402	1.368
σ_w	0.216	0.268	0.235	0.169	0.151	0.146

Figure 1: Sensitivity. Histograms show the log(LR) genotype match distribution for (a) 2 contributor mixtures at 5, 10, and 15 seconds and (b) 3 contributors mixtures at 5, 10, and 15 seconds.

(a) 2 contributors



(b) 3 contributors

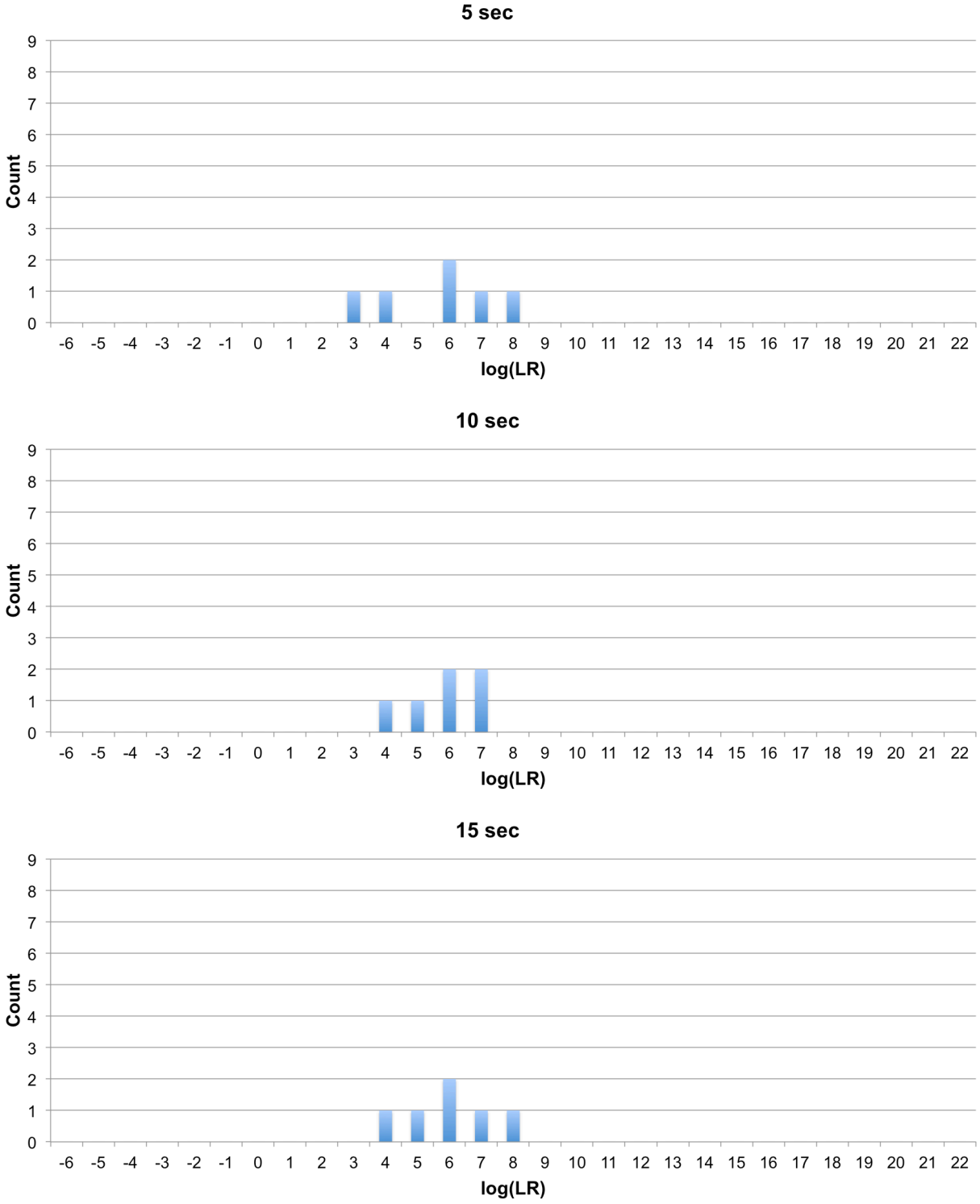
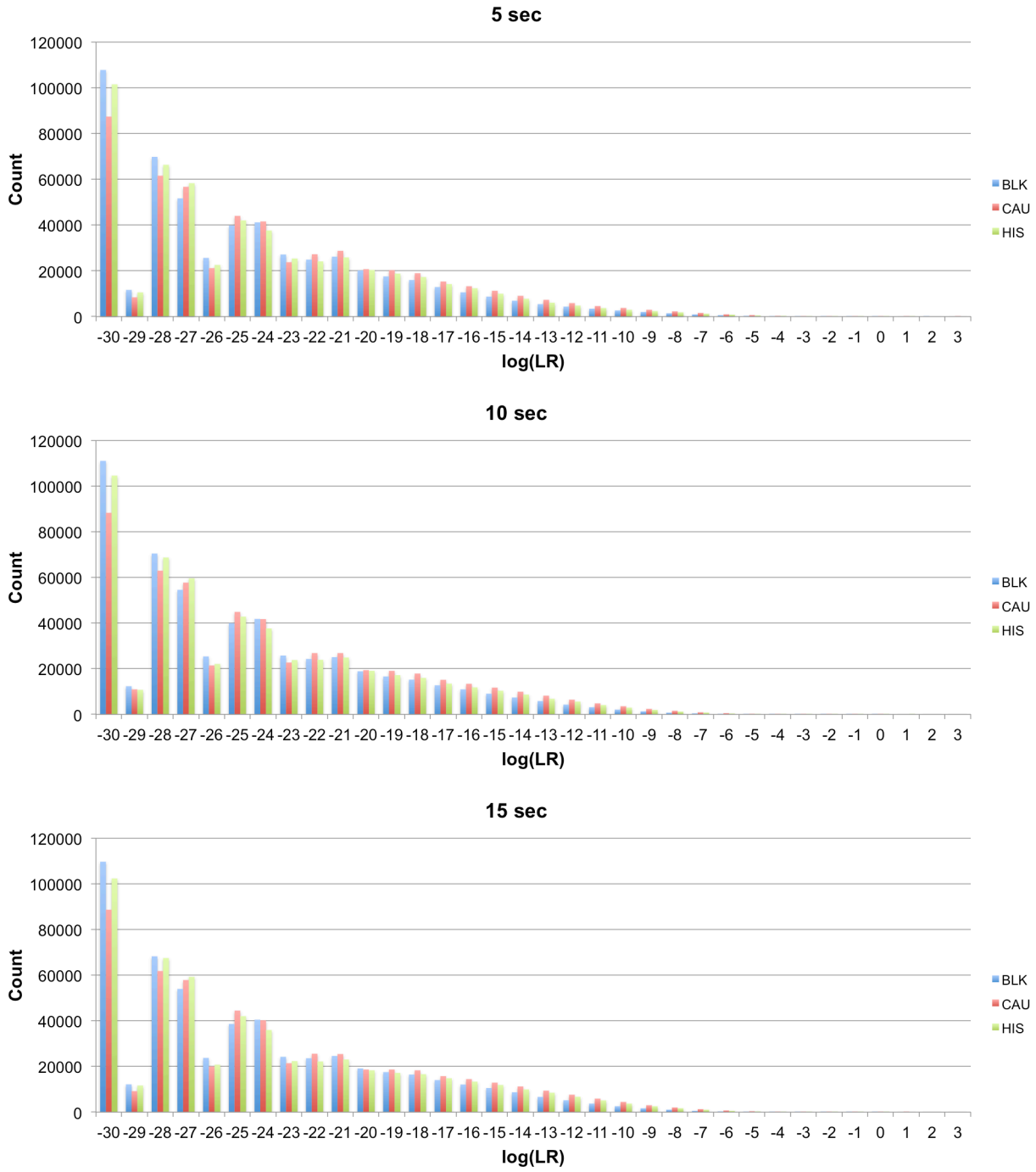


Figure 2: Specificity. Histograms show the log(LR) genotype match distribution for (a) 2 contributor mixtures at 5, 10, and 15 seconds and (b) 3 contributor mixtures at 5, 10, and 15 seconds, relative to ten thousand randomly generated profiles. Each ethnic population is depicted in a different color.

(a) 2 contributors



(b) 3 contributors

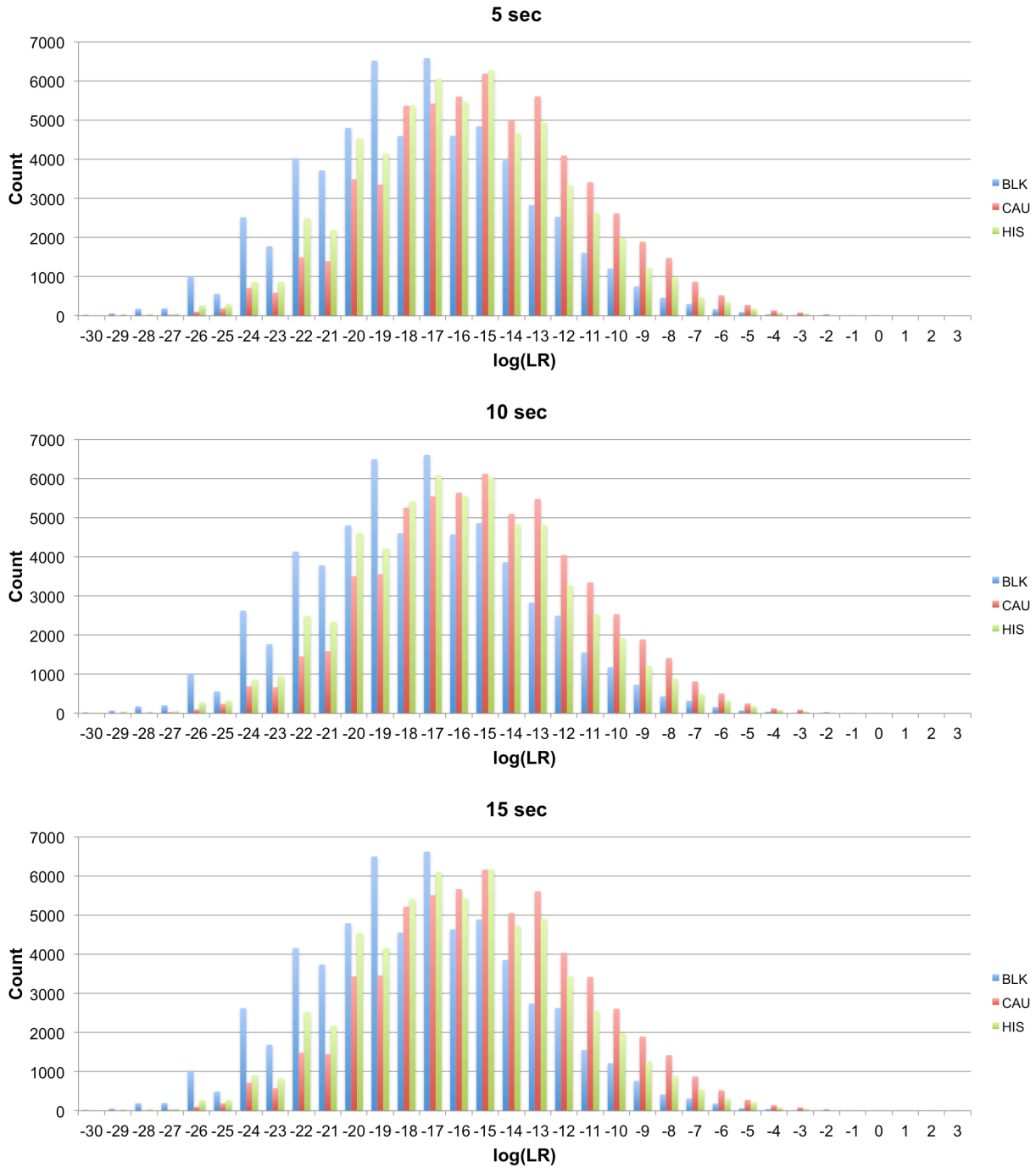
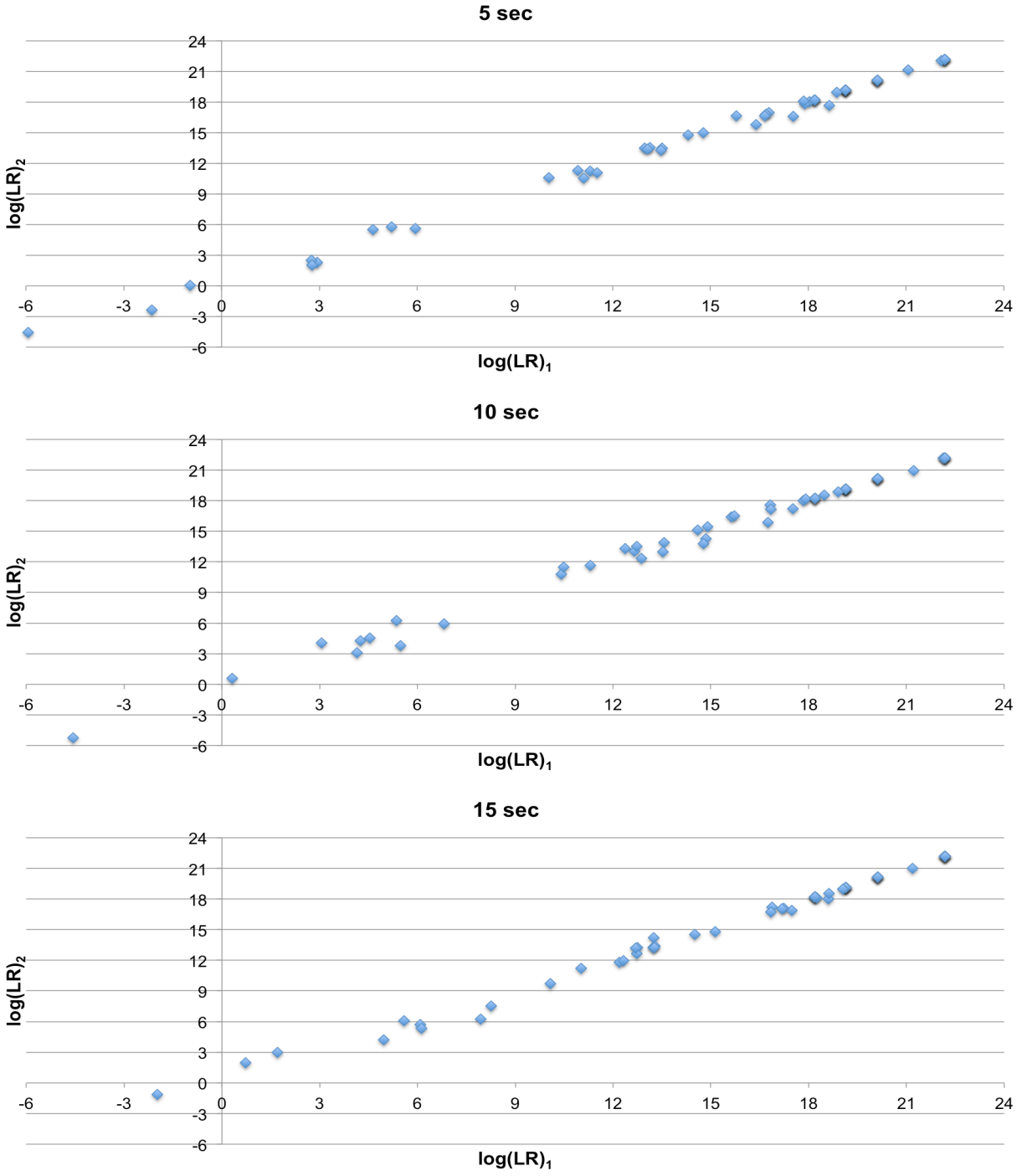


Figure 3: Reproducibility. The scatterplots show log(LR) genotype match values for duplicate computer runs on the same evidence for (a) 2 contributor mixtures at 5, 10, and 15 seconds and (b) 3 contributor mixtures at 5, 10, and 15 seconds. Each point depicts the two match values on the first (x) and second (y) run.

(a) 2 contributors



(b) 3 contributors