JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, July 2015, Vol. 60, No. 4
doi: 10.1111/1556-4029.12788
Available online at: onlinelibrary.wiley.com

PAPER

CRIMINALISTICS

Mark W. Perlin,[1] Ph.D., M.D.; Jennifer M. Hornyak,[1] M.S.; Garett Sugimoto,[2] M.S.; and Kevin W.P. Miller,[2] Ph.D.

TrueAllele® Genotype Identification on DNA Mixtures Containing up to Five Unknown Contributors*

ABSTRACT: Computer methods have been developed for mathematically interpreting mixed and low-template DNA. The genotype modeling approach computationally separates out the contributors to a mixture, with uncertainty represented through probability. Comparison of inferred genotypes calculates a likelihood ratio (LR), which measures identification information. This study statistically examined the genotype modeling performance of Cybergenetics TrueAllele® computer system. High- and low-template DNA mixtures of known randomized composition containing 2, 3, 4, and 5 contributors were tested. Sensitivity, specificity, and reproducibility were established through LR quantification in each of these eight groups. Covariance analysis found LR behavior to be relatively invariant to DNA amount or contributor number. Analysis of variance found that consistent solutions were produced, once a sufficient number of contributors were considered. This study demonstrates the reliability of TrueAllele interpretation on complex DNA mixtures of representative casework composition. The results can help predict an information outcome for a DNA mixture analysis.

KEYWORDS: forensic science, DNA mixture, genotype modeling, validation study, likelihood ratio, probabilistic genotyping

Deoxyribonucleic acid (DNA) evidence is the forensic gold standard (1). Millions of short tandem repeat (STR) (2) genotypes have been assayed for forensic comparison. The principles of STR interpretation are clearest on pristine, single source items containing abundant DNA (typically about 1 ng). A definite genotype can first be inferred, and then compared with another definite genotype, in order to compute a random match probability (RMP) statistic relative to a "random" population genotype. This is certainly the situation when comparing the pristine DNA of individual reference items.

However, crime laboratories today process DNA evidence that is far less pristine. The biological evidence can be mixed (containing two or more contributors), lower level (having under 200 pg of DNA [3]), or degraded. In some forensic DNA laboratories, the majority of evidence items are mixtures, possibly low level, that often contain three or more contributors. The manual "threshold-based" data interpretation procedures (4), originally developed for pristine samples, are not as effective on mixed DNA data (5).

Computer interpretation methods that use more of the quantitative STR peak height data (rather than thresholds) have been used for twenty years (6). Basic "mixture deconvolution" of forensic DNA mixture data into possible contributor genotypes is performed by other software applications such as Applied Biosystems' Genemapper® ID-X and NicheVision Forensics' ArmedXpert™. Qualitative allele "dropout" methods put a probability to unobserved peak data, as in David Balding's likeLTD (7) and Adele Mitchell's FST (8) software programs.

The "genotype modeling" method goes further and strives to preserve DNA identification information by explaining the observed STR data in terms of adding together contributor genotypes (9,10). This method develops Bayesian probability model equations that can explain the data and (when the solution space becomes vast) uses statistical search methods to solve the equations. Such computer systems include DNAmixtures (11) and related efforts (12), MixSep (13), STRmix (14), and TrueAllele® Casework (15,16).

Cybergenetics TrueAllele Casework system separates complex mixture data into its component genotypes. For each contributor, at each locus, a genotype and its uncertainty is described by a probability distribution over allele pair possibilities. This genotype summarizes the data's identification information and imparts to DNA mixtures the original simplicity of single source interpretation. For example, the match statistic resembles RMP, as inferred genotypes are compared with one another.

Previous TrueAllele validation studies have been published. Two-person mixtures of known composition have been examined for their information response, with varying amounts of template DNA (17) and on small quantities using joint interpretation (18). Over 150 casework mixture items containing 2, 3, or 4 contributors have been analyzed for match information across a broad range of mixture weights and quantities, with comparison made to human review methods (15,16,19). However, there has not yet been a study of known mixtures with up to five unknown contributors, where the mixture weights reflected realistic casework instead of simple integer ratios.

[1]Cybergenetics, 160 North Craig Street, Suite 210, Pittsburgh, PA 15213.
[2]Kern Regional Crime Laboratory, 1215 Truxton Avenue, Bakersfield, CA 93301.

*Presented the 66th Annual Meeting of the American Academy of Forensic Sciences, February 17-22, 2014, in Seattle, WA.

Received 6 Mar. 2014; and in revised form 8 Aug. 2014; accepted 15 Aug. 2014.

TABLE 1—*Study design. Five known references were used to randomly create ten mixture samples having 2, 3, 4, and 5 contributors. The mixture weights are shown.*

Reference	Sample	Two	Three	Four	Five	Sample	Two	Three	Four	Five
1	1	0.4674	0.5568	0.1628	0.0346	6	0.0891		0.0489	0.4786
2			0.3064	0.0274	0.0150				0.0475	0.0720
3		0.5326			0.4852			0.3106	0.8976	0.1236
4			0.1368	0.0876	0.2238		0.9109	0.4711	0.0060	0.0782
5				0.7222	0.2413			0.2184		0.2477
1	2		0.2322	0.5367	0.6423	7			0.4350	0.4159
2				0.3430	0.0530		0.5650		0.2087	0.0392
3		0.1770			0.0498			0.5224		0.2900
4			0.2731	0.0746	0.0730			0.2162	0.3385	0.0751
5		0.8230	0.4948	0.0457	0.1820		0.4350	0.2614	0.0179	0.1798
1	3	0.0989	0.4115		0.0930	8	0.1116	0.5774	0.0077	0.4932
2				0.4382	0.0085		0.8884	0.0728	0.0869	0.0400
3				0.1322	0.0586			0.3498	0.5230	0.0655
4			0.2702	0.3969	0.3854				0.3824	0.0876
5		0.9011	0.3182	0.0327	0.4545					0.3136
1	4			0.0271	0.2781	9				0.1197
2					0.2149		0.0585	0.6270	0.0043	0.0802
3		0.3855	0.9438	0.4956	0.2442				0.7007	0.4272
4		0.6145	0.0515	0.4117	0.1468		0.9415	0.0619	0.1750	0.2290
5			0.0047	0.0656	0.1159			0.3111	0.1201	0.1438
1	5				0.0444	10	0.6840		0.3718	0.4555
2		0.8866	0.2749	0.0397	0.3963		0.3160	0.1522	0.2480	0.1777
3			0.1044	0.3229	0.0009			0.3252		0.0197
4		0.1134		0.3603	0.1278				0.1198	0.0317
5			0.6208	0.2771	0.4306			0.5227	0.2604	0.3154

TABLE 2—*Mixture weight variation. The average standard deviation is shown for three concordant methods of computing mixture weight.*

Contributors	*N*=	Human Scoring	Genotype Unknown	Genotype Known
2	20	0.03285	0.02859	0.01944
3	30		0.07699	0.02390
4	40		0.11543	0.01894
5	50		0.15075	0.02221

This study explores the strengths and limitations of DNA interpretation using the TrueAllele Casework system on laboratory-synthesized mixtures of known composition. Mixtures having 2, 3, 4, and 5 contributors are tested, having both high and low DNA amounts. A randomized study design ensures realistic simulation of real-world casework evidence. DNA match information is used throughout to assess interpretation results.

Materials

Randomized Design

A validation study helps establish the reliability of a method, and its suitability for forensic application. DNA mixture evidence contains contributions from two or more individuals in random, unknown proportions. Most mixture studies use integer mixture ratios, providing a convenient simplification for laboratory sample assembly. While these integral ratios may suffice for manual interpretation, computer modeling can extract more information from quantitative data. Therefore, randomized mixture ratios were used in this study to more realistically represent actual casework evidence.

There were four mixture groups, corresponding to 2, 3, 4, or 5 contributors. Within each group, ten mixtures were constructed from five known reference samples. The contributors included in each mixture were determined by randomly selecting DNA references. The mixture weights of the contributors in each mixture item were randomly drawn from a uniform distribution, computed by Dirichlet sampling. The four mixture groups, each containing ten items, yielded a total of 40 randomized DNA mixture items (Table 1).

STR Data

STR mixture data were developed from the known DNA samples according to the experimental design (Table 1). DNA templates were amplified using an Applied Biosystems (Foster City, CA) Identifiler® Plus STR panel at two different DNA concentrations (1 ng and 200 pg). The PCR products were detected on an Applied Biosystems 3130xl Genetic Analyzer, with the higher concentration injected for 5 sec, and the lower amount for 10 sec. (The lower amount was also injected for just 5 sec, but the 10 sec data were more informative.)

Methods

Genotype Inference

TrueAllele Casework has a hierarchical probability model that describes STR data (17). In this Bayesian model (20), the prior genotype probability comes from population allele prevalence, while the likelihood function compares linear combinations of contributor genotypes (with experimental distortion) to observed STR data patterns. The computer uses Markov chain Monte Carlo (MCMC) statistical search (21) to sample from the joint posterior probability distribution. The posterior genotype probability is reported for each contributor at every locus. To eliminate examination bias, where conclusions can be affected by knowledge of a comparison reference (22),

TrueAllele objectively infers genotypes solely from the evidence data.

Electronic data (.fsa) files were processed through the TrueAllele Casework system, and interpretation requests were formed that assumed 1, 2, 3, 4, 5, or 6 contributors. One, two, and three contributor requests were processed with a burn in time of 100,000 MCMC cycles, and sampled from the joint posterior distribution for 100,000 cycles. Requests having four or more contributors were burned in and sampled for twice as many cycles. All requests were run in duplicate, and further replicated as needed, possibly with longer run times.

Match Statistic

Comparing two genotypes relative to a population produces a likelihood ratio (LR) (23). The LR is unaffected by prior beliefs about guilt or innocence and focuses on how well the evidence data support an identification hypothesis. A better mathematical model can elicit more identification information from the same data and (through an inferred evidence genotype) produce a more accurate LR (24). The LR is a Bayes factor that considers the effect of evidence on changing the odds of an identification, commonly used in forensic science to assess the probative force of a DNA match (25). The base

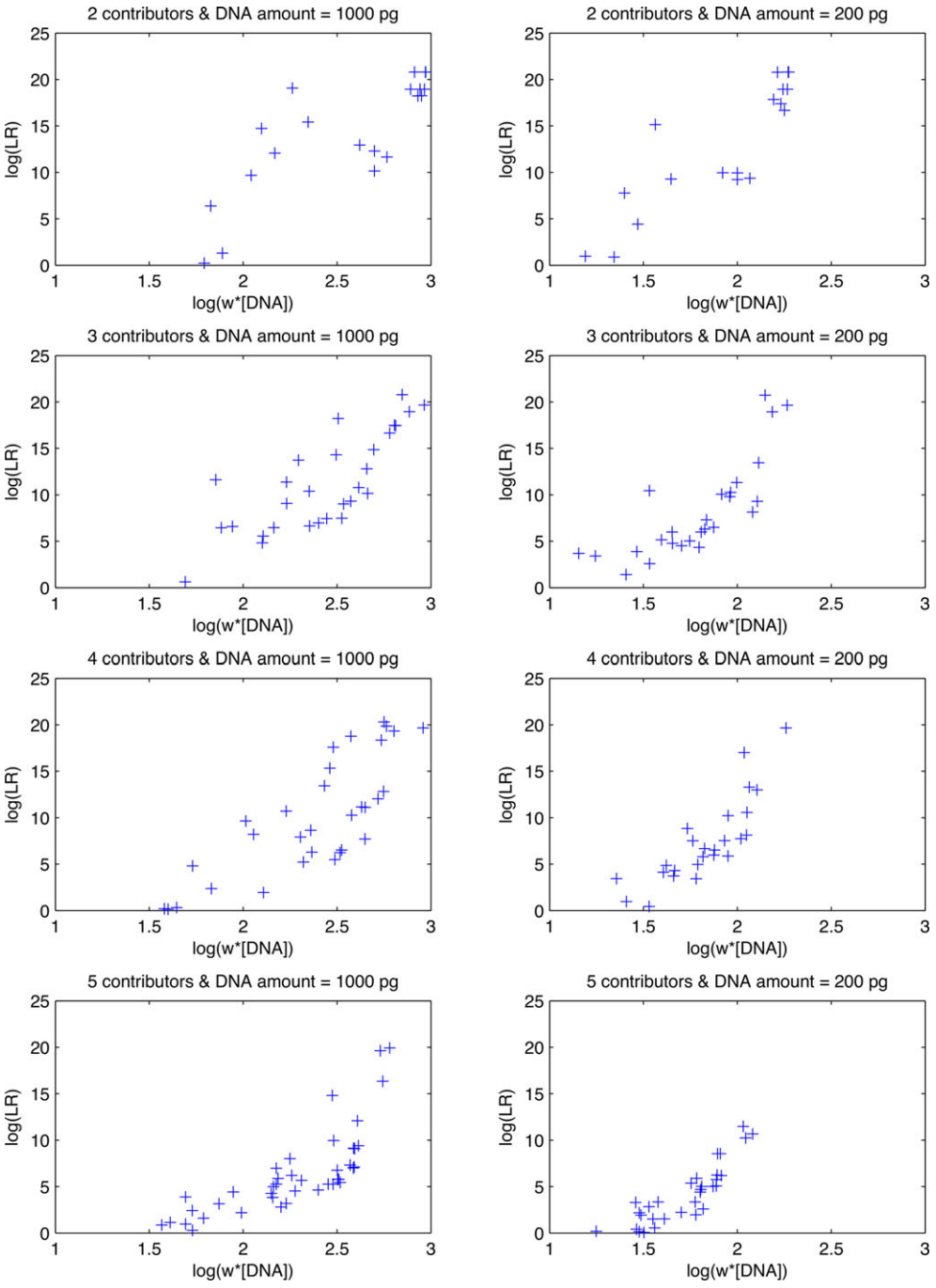

FIG. 1—*DNA information vs. amount. Scatterplots of TrueAllele-inferred log(LR) versus known DNA contributor amount shown for different numbers of contributors (2, 3, 4, and 5 individuals) and DNA amounts (1 ng and 200 pg). Only match results having positive log(LR) are displayed.*

TABLE 3—*Regression coefficient estimates. Log–log scatterplot regression line parameters of LR versus DNA contributor amount (pg). The x-intercept log(w·[DNA]) value is calculated as "−y-intercept/slope".*

Contributors	DNA (pg)	*N*=	Slope	y-intercept	x-intercept
2	1000	20	11.4148	−14.8765	1.3033
3	1000	29	11.9879	−17.8749	1.4911
4	1000	31	12.9912	−20.7610	1.5981
5	1000	41	10.3856	−17.1034	1.6468
2	200	18	15.4039	−16.8288	1.0925
3	200	26	14.0801	−17.0204	1.2088
4	200	25	17.1104	−23.9083	1.3973
5	200	31	13.2820	−18.6383	1.4033

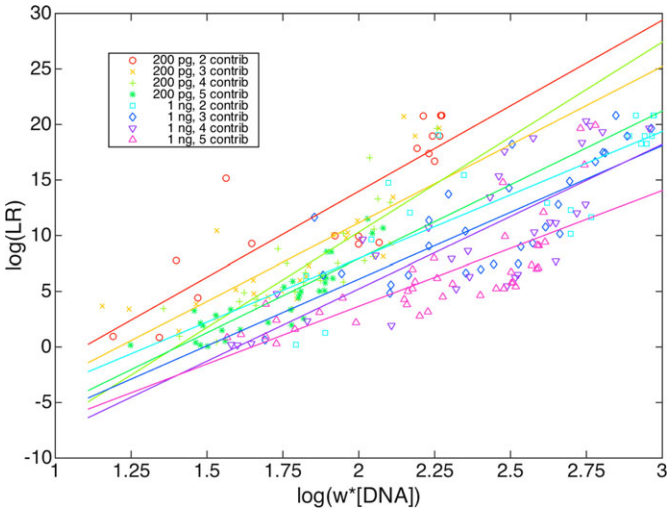

FIG. 2—*Information change regression slopes. Scatterplots of log(LR) vs. DNA amount are shown for eight different groups: 2, 3, 4, or 5 contributors, and either 1 ng or 200 pg of DNA. The scatterplots and regression lines are overlain to show their similar slope behavior.*

TABLE 4—*Analysis of covariance for regression slope. The last column in the ANCOVA gives the statistical significance of the interaction term "ncon*DNA".*

Source	d.f.	Sum Sq	Mean Sq	*F*	*p* > *F*
ncon	7	1731.33	247.33	24.99	
DNA	1	3647.95	3647.95	368.57	
ncon*DNA	7	78.03	11.15	1.13	0.3478
Error	205	2029.02	9.90		

ten logarithm of the LR, "log(LR)" or "weight of evidence", is a standard additive measure of information change, expressed in "ban" units (26).

A competent TrueAllele user reviewed the computer-inferred genotype and match results. Because of genotype uncertainty, a contributor may match more than one reference. Using the study design information, each contributor genotype inferred from a mixture item was paired with a unique known reference. Other useful pairing information included the expected contributor genotype LR value (Kullback–Leibler divergence, or "KL") (27), LR match statistics, and the mixture weights.

Match statistics were calculated relative to the United States Federal Bureau of Investigation allele databases for African American, Caucasian, and Hispanic populations (28). The most conservative LR value among these populations was used. The

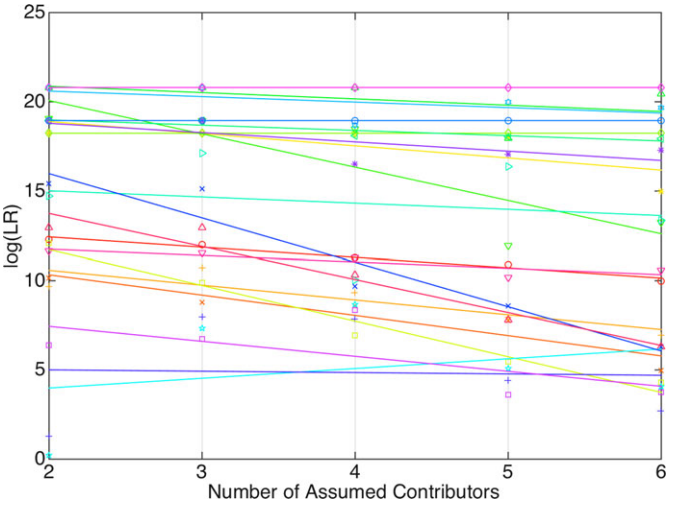

FIG. 3—*Information with excess contributors (two-person mixtures). In separate computer runs, TrueAllele assumed 2, 3, 4, 5, or 6 unknown contributors and inferred log(LR) match statistics. For each mixture component, the regression line and data points are shown under these five different contributor assumptions.*

reported log(LR) was the average of two independent computer runs, where all contributor match values were within one ban and the genotypes were concordant. On average, 3.1 computer runs were conducted per sample. The co-ancestry coefficient (theta value) was set to 1% (29).

Results

Mixture Weight

The mixture weight (*w*) of each item's contributor had a predetermined design value (Table 1), but was subject to laboratory variation (e.g., pipetting, volumes, quantification). As the study relates other variables to *w*, it was important to obtain an accurate mixture weight estimate. Therefore, empirical methods based on observed data, rather than expected design values, were used to estimate *w* for the items.

First, TrueAllele estimated mixture weights in the usual casework manner, without making any genotype assumptions. That is, all variables (including *w* and the genotypes) were estimated solely from the quantitative STR peak height data (15).

Next, the TrueAllele system used the known contributor genotypes as provided input when estimating mixture weight. That is, the genotypes were assumed, but the other variables (including *w*) were estimated based on the data and that genotype knowledge (10). As this approach starts with more information, it can produce more precise results.

Finally, mixture weights were manually calculated for all the two contributor items. Within each item, loci were identified where the two contributors had nonoverlapping alleles. The allele peak heights from these loci were entered into an Excel® (Microsoft®, Redmond, WA) spreadsheet that found each contributor's mixture weight mean and standard deviation.

There was a strong pairwise association (r^2 = 0.999) between all three data-derived contributor *w* values for an item, whether calculated by TrueAllele or a person. However, less association (r^2 = 0.907) was found between the data-derived mixture weights and the experimental design values. The TrueAllele

TABLE 5—*Contributor sufficiency. How TrueAllele behaves when assuming more than the known number of contributors. For each mixture sample having a known number of contributors (known), TrueAllele processed the data assuming up to six total contributors. This produced a group of log(LR) values for each sample's contributor. A linear model* $y = (\alpha + \alpha i) + (\beta + \beta i)\, x + \varepsilon$ *was fitted to the data, where x is the assumed number of contributors, y is the log(LR) information obtained, there are average α and group αi y-intercepts, average β and group βi slopes, and ε is the error. The table shows the average β slope values for each number of known contributors.*

Known	Slope β	SE	*p*-value
2	−0.6653	0.1120	1.5401×10^{-7}
3	−0.8501	0.1151	6.6154×10^{-10}
4	−1.3025	0.2930	1.2587×10^{-4}
5	−0.2598	N/A*	N/A*

*Five contributors provided only two points per line (assuming 5 or 6), which was insufficient for some statistical estimates.

calculations that used both the observed data and known genotypes gave the most precise mixture weights (Table 2). With two contributors, for example, the average mixture weight standard deviation was 0.0194. These minimum variance mixture weight values, inferred by TrueAllele with all genotypes known, were used in this study.

Information Response

TrueAllele's inferred identification information varies with contributor DNA amount in a predictable way (17). A scatterplot of log(LR) information (*y*-axis) as a function of a contributor's log(*w*·[DNA]) quantity (*x*-axis) is roughly linear. Linear regression of a scatterplot permits examination of many match results within a single analysis, and lets each contributor in a group of mixture items be considered separately.

There are expected deviations from linearity in some situations. First, when mixture weights are equal, peak height data do not help uniquely assign alleles to a particular genotype. This inherent genotype ambiguity impedes contributor separation, diffusing probability across multiple allele pair possibilities. Such genotype probability diffusion at equal mixture weights reduces the LR, as seen in Figure 10 of (15). Second, once there is sufficient contributor DNA to achieve the RMP maximum value, additional DNA cannot further increase the LR beyond this limit. Thus, at high DNA amounts there is an information saturation, where the LR plateaus instead of continuing to linearly increase, as seen in Figure 7 of (17).

Scatterplots of log(LR) information versus log(*w*·[DNA]) contributor quantity were developed from the mixture contributors

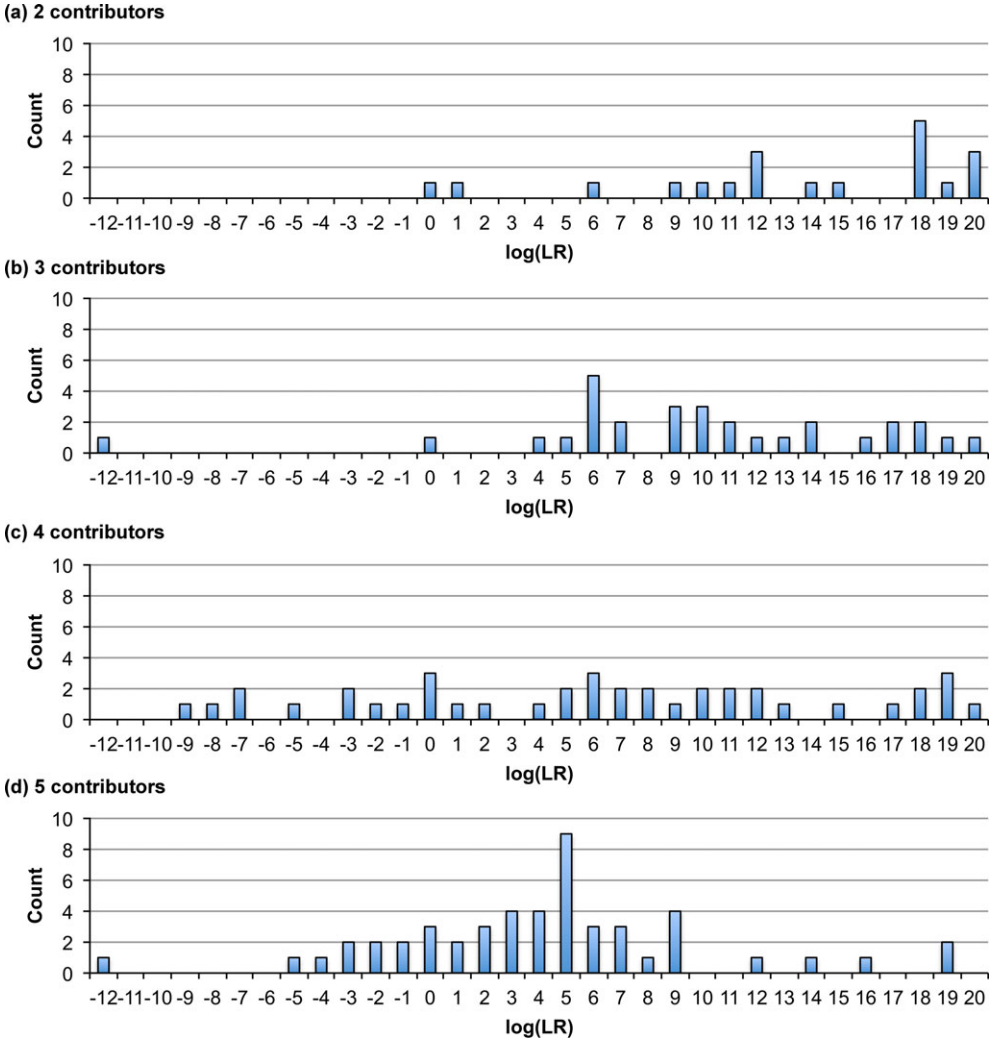

FIG. 4—*Sensitivity (1 ng). Histograms of the log(LR) distribution for mixtures having (a) 2, (b) 3, (c) 4, and (d) 5 contributors. Average replicate log(LR) scores were used.*

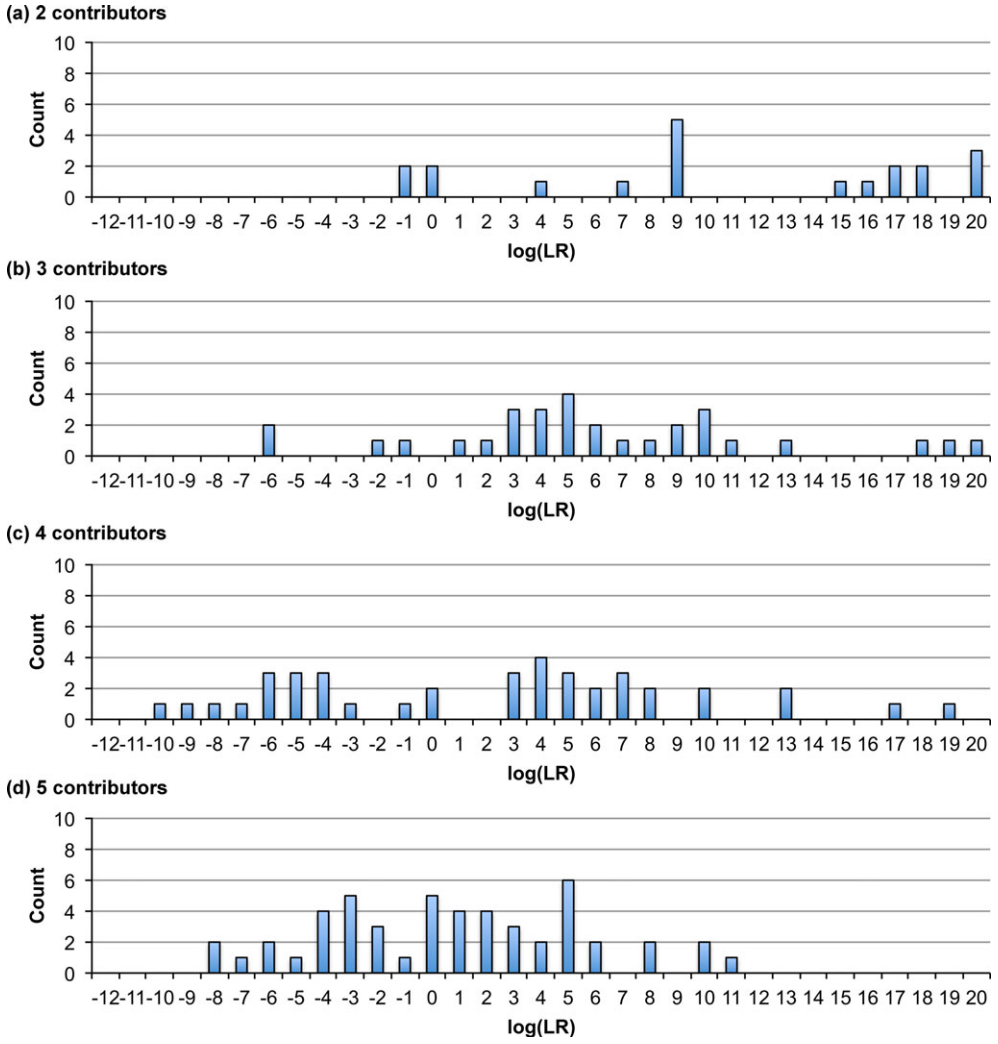

FIG. 5—*Sensitivity (200 pg). Histograms of the log(LR) distribution for mixtures having (a) 2, (b) 3, (c) 4, and (d) 5 contributors. Average replicated log (LR) scores were used.*

using their weight, quantity, and log(LR) values (Fig. 1). The scatterplots of positive match results were roughly linear ($r^2 = 0.505$), and for two contributors showed the expected log(LR) reductions for equal contributor weights and high DNA amounts. The average regression slope across all groups was 13.33 log(LR)/log(DNA), with a standard error of 0.74. This slope value means that a 10-fold change in contributor DNA amount yields about a trillion-fold change in LR (Table 3).

Interpretation Invariance

There were eight test groups, two for DNA quantity (high, low) and four different contributor numbers (2, 3, 4, and 5 individuals). The slope parameter describes an important aspect of interpretation behavior, namely how contributor DNA amount affects match information. Finding similarity in the slope parameter between the groups' regression results would suggest that TrueAllele's interpretation behavior is relatively invariant across these conditions. Such interpretation invariance would show that TrueAllele behaves consistently, regardless of the number of contributors or amount of DNA.

Consider, for example, the interpretation of a two-person high-template mixture, relative to that of a five-person low-template mixture. The peak height data for these two situations would look entirely different. On average, there is more identification information in a 1 ng two-person mixture than in a 200 pg five-person mixture, as seen in the 4 ban difference in respective y-intercept values of −14.9 and −18.6 (Table 3). But their respective slopes of 11.4 and 13.3 are similar, indicating a consistent information response to changes in contributor DNA amount.

Analysis of covariance (ANCOVA) was used to test this similarity hypothesis. The covariate was the slope of a regression line (Fig. 2). The null hypothesis was that the slopes (across the eight groups) were the same. To reject the null hypothesis, there would need to be a significant difference between the slopes. (The intercept values were expected to differ, as each DNA mixture group had its own average identification information.)

The eight groups showed different intercept values (Table 3), expressing group differences in DNA detectability (x-intercept) and identification information (y-intercept). There was no significant difference in regression line slope ($p = 0.3478 > 0.05$), and so the null hypothesis could not be rejected (Table 4). Table 3 indicates the slope invariance across four different contributor numbers (2, 3, 4, and 5) and DNA template amounts (1 ng and 200 pg). This invariance shows that TrueAllele's overall information response to DNA data does not significantly depend on a particular mixture's number of contributors or template amount.

TABLE 6—*Sensitivity. Sensitivity statistics were calculated for the eight groups (quantity and contributor number) as the average of two replicate log(LR) values. (a) The minimum, mean, maximum, and standard deviation (ban) use the smallest values across three ethnic populations. (b) The number of false exclusions are binned by log(LR) value (rows), with a total of 59 events.*

	1 ng				200 pg			
ncon	2	3	4	5	2	3	4	5
(a) Summary statistics								
N=	20	30	40	50	20	30	40	50
Min	0.219	−11.422	−8.994	−11.315	−0.722	−5.970	−9.719	−7.883
Mean	14.084	10.476	6.789	4.723	11.388	6.656	2.691	1.276
SD	6.209	6.542	8.375	5.716	7.572	6.323	7.258	4.725
Max	20.799	20.789	20.304	19.923	20.799	20.723	19.665	11.483
	1 ng				200 pg			
log(LR)	2	3	4	5	2	3	4	5
(b) False exclusions								
−1			1	2	2	1	1	1
−2			1	2		1		3
−3			2	2			1	5
−4				1			3	4
−5			1	1			3	1
−6						2	3	2
−7			2				1	1
−8			1				1	2
−9			1				1	
−10							1	
−11								
−12		1		1				
Total	0	1	9	9	2	4	15	19

TABLE 7—*Sensitivity varies with mixture weight. The true inclusion rate (one minus the false exclusion rate) based on positive log(LR) counts is shown for mixture weight ranges. There were a total of 280 observations, divided equally between the 1 ng and 200 pg DNA levels.*

N=	Mixture Range, %	1 ng, %	200 pg, %
4	0–1	0	0
20	1–5	40	0
17	5–10	82	24
33	10–25	100	91
39	25–50	100	100
25	50–100	100	100
140			

Contributor Sufficiency

Each assumed unknown genotype provides another dimension that can explain the data. When too few contributors are assumed, genotype inference can be restricted. This restriction artificially sharpens genotype (and match) results for major contributors, and dissipates minor contributor genotypes (and matches). With a surplus of assumed (relative to actual) contributors, there is sufficient genotype dimensionality to resolve a mixture.

Every mixture item in this study was synthesized with a known number of contributing individuals. TrueAllele processed each 1 ng mixture over a full range (1, 2, 3, 4, 5, and 6) of assumed unknown contributors, that is, the one correct value and five alternative values. Duplicate log(LR) results for an inferred genotype were averaged together. Figure 3 shows match information regression lines for two-person mixtures (one line for each mixture contributor) as the number of contributors assumed by the computer's interpretation is varied.

Once TrueAllele had assumed a sufficient number of contributors (i.e., at least as many as the actual number), the match results remained consistent (Fig. 3). With an excess of assumed contributors, the log(LR) scatterplot values generally decreased. Most of the information slopes were negative, suggesting that match statistic decreases as excess assumed contributors are added.

For all actual contributor numbers (2, 3, 4, and 5), linear regression showed negative average slopes for interpretations that had an excess of contributors (Table 5). Note that the negative slope values were sufficiently greater than their standard errors to be statistically different from zero ($p < 0.01$). The slope magnitudes were small, with values ranging from −1.30 to −0.26, indicating little average reduction in log(LR). Thus, assuming extra contributors in TrueAllele preserves the average match result, without overstating the match statistic.

Inclusion Distribution

Sensitivity measures the extent to which a mixture interpretation method includes a contributor. The log(LR) measures the degree of match between a genotype inferred from an evidence item and the genotype of an individual who has contributed to that item, relative to a population genotype. Previous studies have shown that this match information (in ban units) correlates with how much of that contributing individual's DNA (on a logarithmic scale) is present in the item (17).

Sensitivity was determined for each of the eight test groups. Figure 4 shows the log(LR) frequency distribution for each match of the high DNA quantity group (1 ng) for separate contributor numbers (2, 3, 4, or 5), while Fig. 5 shows the distribution for the low DNA quantity (200 pg) groups. The bar charts show a leftward shift as contributor number increases, indicating a decrease in average identification information. Using less DNA (200 pg vs. 1 ng) further reduced the log(LR) score.

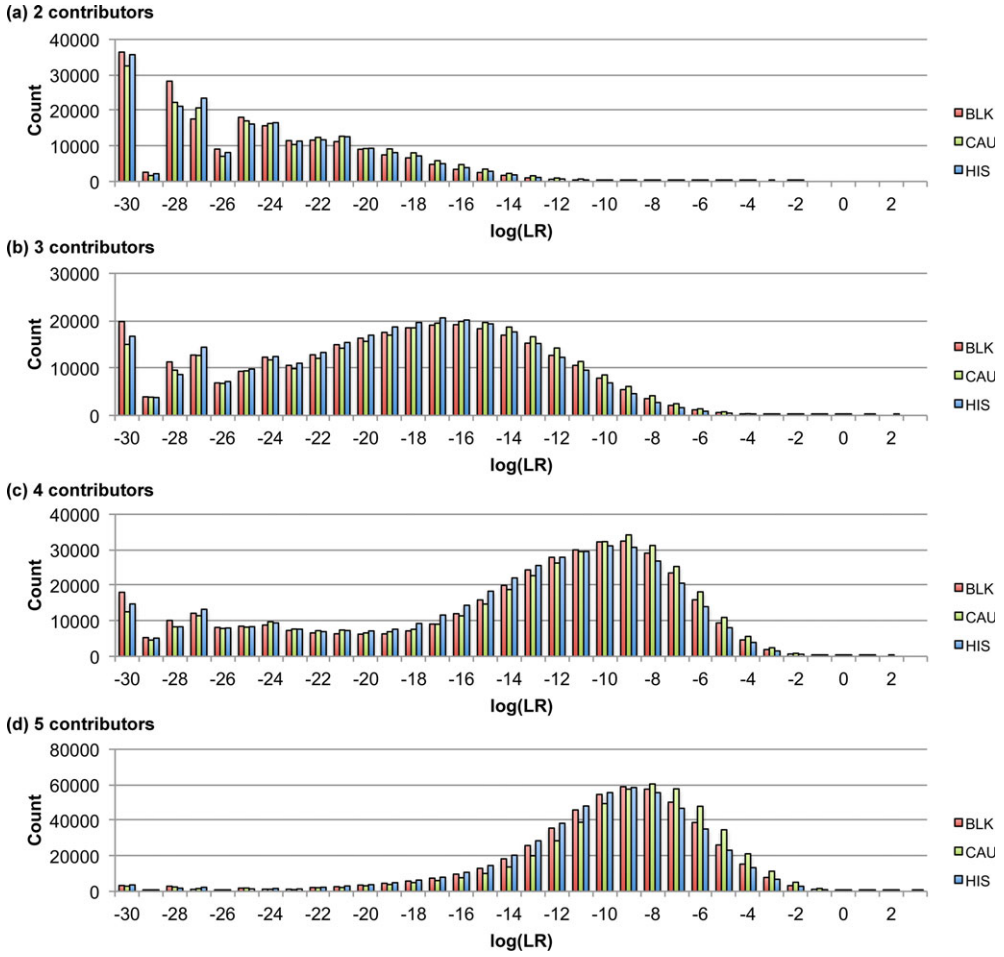

FIG. 6—*Specificity (1 ng). The log(LR) specificity distribution for mixtures having (a) 2, (b) 3, (c) 4, and (d) 5 contributors. The LRs were computed relative to 10,000 randomly generated profiles across the FBI African American (BLK, red), Caucasian (CAU, green), and Hispanic (HIS, blue) populations.*

These trends are quantified in Table 6a. The mean identification information for 1 ng mixtures decreased steadily from 14.084 ban with two contributors to 4.732 ban with five contributors. This decrease reflects the reduced amount of DNA in each contributor, as well as the uncertainty in separating their genotypes. The maximum values show that major contributors can produce definite genotypes that preserve all match strength, even with four other contributors present. The minimum values show more exclusion of known contributors with increasing contributor number. With a lower 200 pg template, the trends are similar, but start at a lower log(LR) level.

False exclusions increased with contributor number (Table 6b, Figs 4 and 5). The table rows stratify the false exclusion events by ban value. With 1 ng DNA, false exclusions with 2 or 3 contributors were rare (2%), but became more common (20%) with 4 or 5 contributors. There were more false exclusions when there was less DNA (200 pg), consistently increasing from 10% for two contributors to 38% with five contributors. There were a total of 59 false exclusions, of 280 observations (21%).

The true inclusion rate (i.e., 1 – false exclusion rate) was estimated as a function of mixture weight for common ranges used in forensic practice (Table 7). For full DNA amounts of 1 ng, mixture weights above 10% always gave a positive match result (no false exclusions), regardless of the number of contributors. This success rate fell to 82% in the 5–10% range, and down to 40% in the 1–5% range. With low-template amounts (200 pg), a positive identification was always made with a mixture weight over 25%. While the inclusion rate was 91% in the 10–25% mixtures, it dropped to 24% with 5–10% mixtures, and no matches were found below 5%. There were no inclusions when the mixture weight was under 1% ($N = 4 + 4$, for 1 ng and 200 pg).

Exclusion Distribution

Specificity measures the extent to which a mixture interpretation method excludes a noncontributor. The log(LR) measures the degree of exclusion (relative to a population) through the magnitude of a negative match value. A mismatch can occur between two genotypes, one inferred from an evidence item, and another from an individual who may have not contributed their DNA to that item. Previous studies have shown that such mismatches generally produce negative log(LR) numbers, with occasional positive values near zero (16).

To assess specificity, each inferred evidence genotype (using the first replicate) was compared with 10,000 genotypes that were randomly generated from an ethnic allele frequency distribution. This comparison was performed three times, once for each ethnic population.

Specificity was determined for each of the eight mixture subgroups. Figure 6 shows the empirical log(LR) distribution for mismatch with high DNA levels (1 ng) for each contributor number (2, 3, 4, or 5). Similarly, Fig. 7 shows the mismatch distribution for low DNA levels (200 pg). The figures show

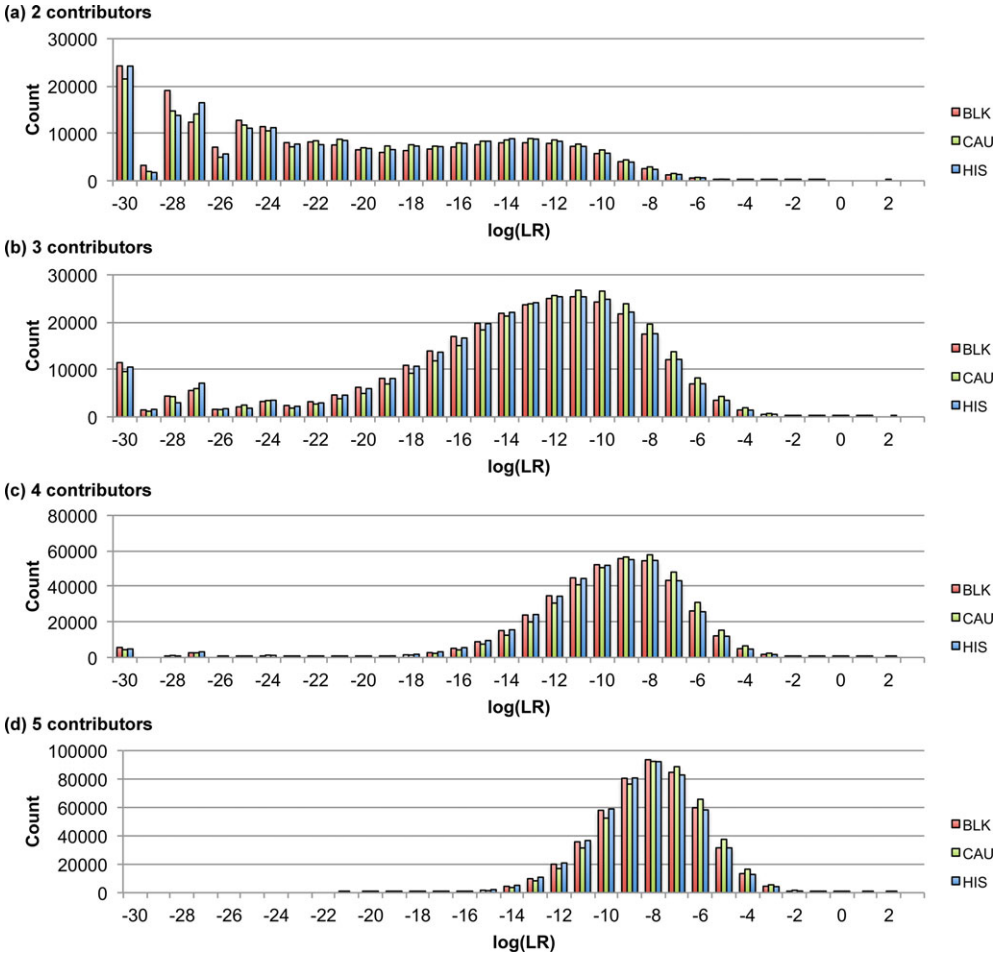

FIG. 7—*Specificity (200 pg). The log(LR) specificity distribution for mixtures having (a) 2, (b) 3, (c) 4, and (d) 5 contributors. The LRs were computed relative to 10,000 randomly generated profiles across the FBI African American (BLK, red), Caucasian (CAU, green), and Hispanic (HIS, blue) populations.*

TABLE 8—*Specificity. Specificity statistics were calculated for the eight groups (quantity and contributor number). (a) The minimum, mean, maximum, and standard deviation log(LR) values were averaged across three ethnic populations. (b) The total number of false inclusions is shown for each group, binned by log(LR) value (rows).*

	1 ng				200 pg			
ncon	2	3	4	5	2	3	4	5
(a) Summary statistics								
N=	600,000	900,000	1,200,000	1,500,000	600,000	900,000	1,200,000	1,500,000
Min	−30.000	−30.000	−30.000	−30.000	−30.000	−30.000	−30.000	−20.143
Mean	−23.904	−18.339	−13.878	−9.429	−20.247	−13.507	−9.517	−7.636
SD	4.608	5.990	7.183	4.536	6.821	5.986	4.048	2.218
Max	−1.514	1.511	2.140	3.202	0.410	1.878	2.006	1.671
	1 ng				200 pg			
log(LR)	2	3	4	5	2	3	4	5
(b) False inclusions								
0	0	18	142	1071	0	36	152	123
1	0	6	37	200	0	16	22	18
2	0	1	7	24	2	1	3	4
3	0	0	0	6	0	0	0	0
Total	0	25	186	1301	2	53	177	145

shrinkage toward zero information, as contributor number increases, for both high and low DNA amounts (1 ng and 200 pg).

These trends are quantified in Table 8. The mean values showed roughly equal specificity across the three different ethnic groups (Tables S1 and S2). At 1 ng (Table 8a), there was

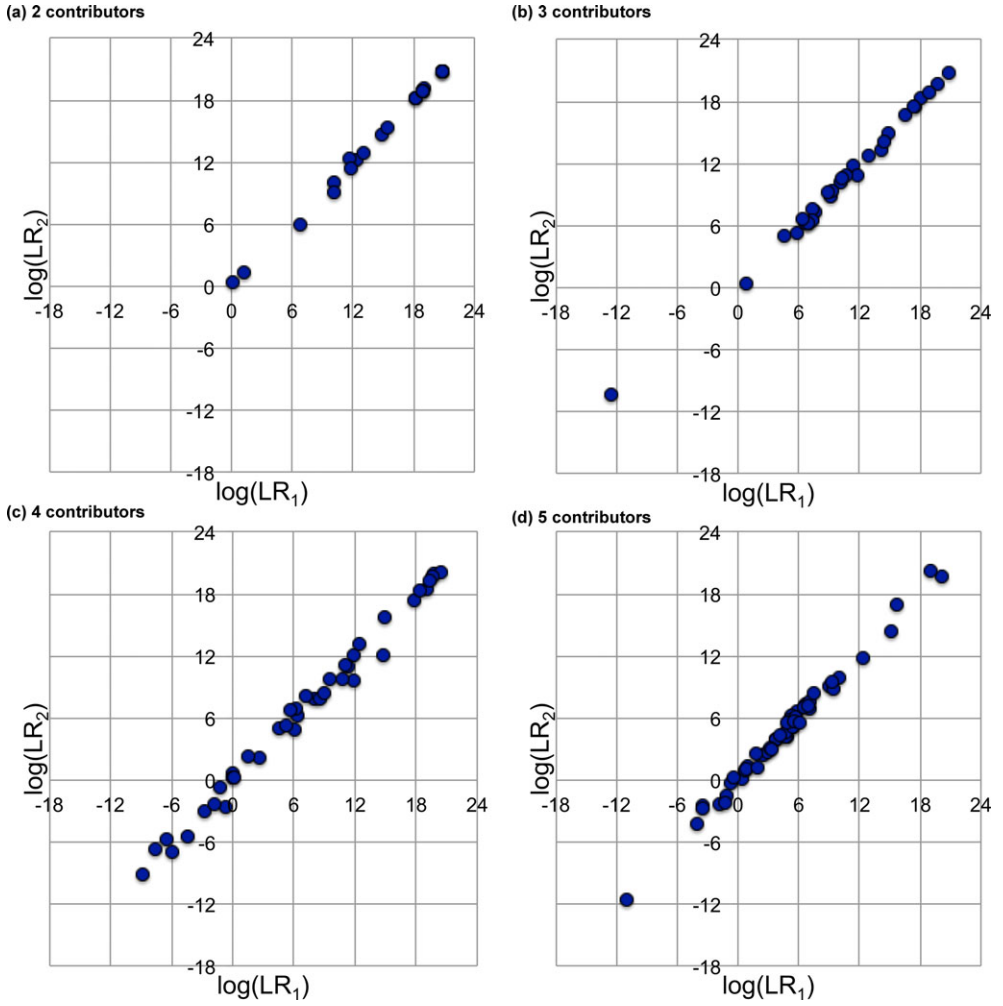

FIG. 8—*Reproducibility (1 ng). Scatterplots of paired log(LR) values for duplicate computer runs on the same mixture sample. The mixtures had (a) 2, (b) 3, (c) 4, and (d) 5 contributors. Each point shows the first (LR_1) and second (LR_2) replicates.*

shrinkage toward zero information when proceeding from two contributors (−24 ban) to five (−9 ban). The lower DNA amount (200 pg) showed the same progression, but the reduced genotype information was already closer to zero: −20 ban for two contributors, increasing to −7 with five contributors.

For two contributors, false inclusions were rarely seen (Table 8b, Figs 6 and 7), with none occurring at 1 ng and just 2 events at 200 pg (N = 600,000). The table rows stratify the false inclusion events by ban value. The false inclusion level increased with contributor number, reaching a maximum rate of 0.0867% with five contributors in 1 ng of DNA (1301 events of 1,500,000 comparisons). The other seven subgroups had appreciably lower error rates (Table 8b). There were few false matches beyond an LR of 10, and essentially none (six events in 8,400,000) with an LR > 1000.

Reproducibility Comparison

The reproducibility of a DNA interpretation method describes how well a match statistic is independently replicated on the same data. Once two (or more) interpretations have been made on the same data group, an interpretation method's reproducibility can be quantified using a within-group standard deviation. This statistic measures the log(LR) variation (about the average interpretation result) for each mixture contributor within the group.

There is expected interpretation variation arising from the MCMC statistical sampling. Scatterplots show that when genotypes are concordant, so too are the DNA match statistics (Figs 8 and 9). Each point gives the pair of log(LR) values from two concordant computer runs, independently run using the same parameter settings. As these points line up along the 45 degree equi-information line, TrueAllele's reproducibility is visually evident.

Table 9 gives the within-group standard deviation (σ_w) values for each group. Small σ_w values were found in all eight subgroups, never exceeding half a ban. These small σ_w values quantitatively confirm TrueAllele's reproducibility. In forensic practice, two independent computer runs on an evidence item can provide reporting confidence.

Conclusions

The computer interpretation of DNA evidence is a 21st century necessity. With ever-increasing numbers of STR loci, DNA mixtures having three or more contributors, low-level or degraded samples, and the potential for subjective examination bias (22,30), human analysts cannot be expected to fully process

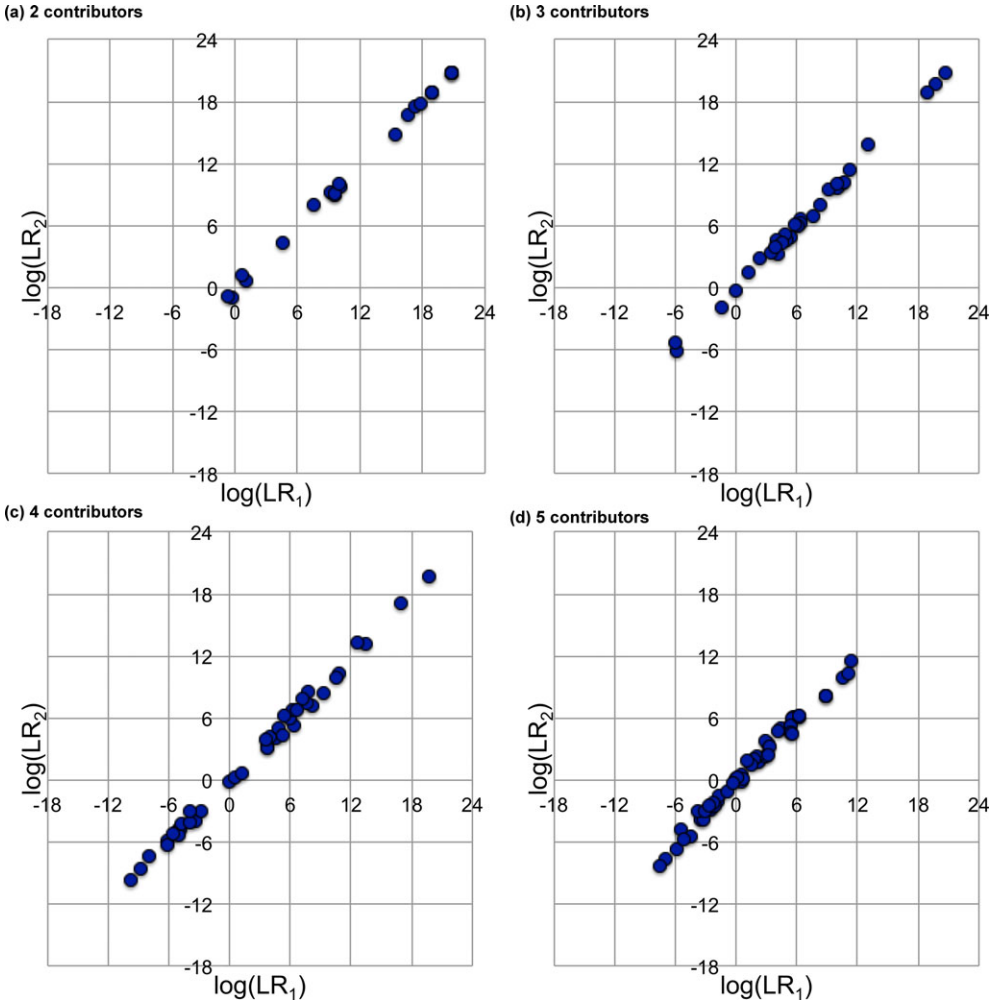

FIG. 9—*Reproducibility (200 pg). Scatterplots of paired log(LR) values for duplicate computer runs on the same mixture sample. The mixtures had (a) 2, (b) 3, (c) 4, and (d) 5 contributors. Each point shows the first (LR_1) and second (LR_2) replicates.*

TABLE 9—*Reproducibility. The table shows the within-group standard deviation σ_w (ban) for each of the eight test groups, at both 1 ng and 200 pg DNA template amounts.*

ncon	1 ng	200 pg
2	0.189	0.171
3	0.281	0.205
4	0.430	0.255
5	0.287	0.254

all the data. Such thorough and objective mathematical DNA mixture interpretation is the province of machines (31).

To be forensically useful, interpretation methods must be fully tested on realistic data. When software programs cannot robustly resolve challenging mixtures, their casework applicability becomes limited (e.g., DNAMIX, I-3, LoComatioN, LSD, PENDULUM). For over 10 years, TrueAllele has been extensively assessed in validation studies performed by crime laboratories and Cybergenetics, with publication in peer-reviewed journals (15–19).

This TrueAllele validation study used randomly generated DNA mixtures of known composition that were representative of actual casework. The samples contained up to five contributors, for both high- and low-template amounts. The study assessed the efficacy of the computer's genotype modeling, as quantified by LR.

The computer's mixture weight values were found to be reliable. The computed match information varied with DNA quantity in a predictable way that did not significantly depend on contributor number or template amount. Excess assumed contributors did not materially affect the conclusions.

The match statistic determination of inclusion and exclusion gave reproducible match values. The system was highly sensitive, preserving considerable identification information. It was also extremely specific, providing large exclusionary match statistics. Error rates were determined for false inclusions and exclusions. Inclusion accuracy was tabulated as a function of mixture weight.

This in-depth experimental study and statistical analysis establish the reliability of TrueAllele for the interpretation of DNA mixture evidence over a broad range of forensic casework conditions.

Conflict of Interest

Dr. Mark Perlin is a shareholder, officer, and employee of Cybergenetics, Pittsburgh, PA. Jennifer Hornyak is an employee of Cybergenetics. Garett Sugimoto and Dr. Kevin Miller are employees of the Kern Regional Crime Laboratory, a government

agency that provides expert DNA testimony in criminal cases and uses the TrueAllele Casework system.

References

1. Lynch M. God's signature: DNA profiling, the new gold standard in forensic science. Endeavour 2003;27(2):93–7.
2. Edwards A, Civitello A, Hammond H, Caskey C. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am J Hum Genet 1991;49:746–56.
3. Budowle B, Eisenberg A, van Daal A. Validity of low copy number typing and applications to forensic science. Croat Med J 2009;50 (3):207–17.
4. Scientific Working Group on DNA Analysis Methods (SWGDAM). Short Tandem Repeat (STR) interpretation guidelines. Forensic Sci Commun 2000;2(3).
5. Scientific Working Group on DNA Analysis Methods (SWGDAM). Interpretation guidelines for autosomal STR typing by forensic DNA testing laboratories. http://www.fbi.gov/about-us/lab/codis/swgdam-interpretation-guidelines, 2010.
6. 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.
7. Balding DJ. Evaluation of mixed-source, low-template DNA profiles in forensic science. Proc Natl Acad Sci USA 2013 Jul 23;110(30):12241–6.
8. Mitchell AA, Tamariz J, O'Connell K, Ducasse N, Budimlija Z, Prinz M, et al. Validation of a DNA mixture statistics tool incorporating allelic drop-out and drop-in. Forensic Sci Int Genet 2012;6(6):749–61.
9. Gill P, Sparkes R, Pinchin R, Clayton TM, Whitaker JP, Buckleton J. Interpreting simple STR mixtures using allele peak area. Forensic Sci Int 1998;91:41–53.
10. Perlin MW, Szabady B. Linear mixture analysis: a mathematical approach to resolving mixed DNA samples. J Forensic Sci 2001;46 (6):1372–7.
11. Cowell R, Graversen T, Lauritzen S, Mortera J. Analysis of DNA mixtures with artefacts. J R Stat Soc Ser C Appl Stat 2015;64(1):1–48.
12. Puch-Solis R, Rodgers L, Mazumder A, Pope S, Evett I, Curran J, et al. Evaluating forensic DNA profiles using peak heights, allowing for multiple donors, allelic dropout and stutters. Forensic Sci Int Genet 2013;7 (5):555–63.
13. Tvedebrink T, Eriksen PS, Mogensen HS, Morling N. Identifying contributors of DNA mixtures by means of quantitative information of STR typing. J Comput Biol 2012;19(7):887–902.
14. Taylor D, Bright J-A, Buckleton J. The interpretation of single source and mixed DNA profiles. Forensic Sci Int Genet 2013;7(5):516–28.
15. Perlin MW, Legler MM, Spencer CE, Smith JL, Allan WP, Belrose JL, et al. Validating TrueAllele® DNA mixture interpretation. J Forensic Sci 2011;56(6):1430–47.
16. Perlin MW, Belrose JL, Duceman BW. New York State TrueAllele® Casework validation study. J Forensic Sci 2013;58(6):1458–66.
17. Perlin MW, Sinelnikov A. An information gap in DNA evidence interpretation. PLoS ONE 2009;4(12):e8327.
18. Ballantyne J, Hanson EK, Perlin MW. DNA mixture genotyping by probabilistic computer interpretation of binomially sampled laser captured cell populations: combining quantitative data for greater identification information. Sci Justice 2013;53(2):103–14.
19. Perlin MW, Dormer K, Hornyak J, Schiermeier-Wood L, Greenspoon S. TrueAllele® Casework on Virginia DNA mixture evidence: computer and manual interpretation in 72 reported criminal cases. PLoS ONE 2014;9(3):e92837.
20. Gelman A, Carlin JB, Stern HS, Rubin D. Bayesian data analysis. Boca Raton, FL: Chapman & Hall/CRC, 1995.
21. Gilks WR, Richardson S, Spiegelhalter DJ, editors. Markov Chain Monte Carlo in practice. London, U.K.: Chapman and Hall, 1996.
22. Dror IE, Hampikian G. Subjectivity and bias in forensic DNA mixture interpretation. Sci Justice 2011;51(4):204–8.
23. Perlin MW. Explaining the likelihood ratio in DNA mixture interpretation. Promega's Twenty First International Symposium on Human Identification; 2010 Oct 11-14; San Antonio, TX. Madison, WI: Promega, 2010.
24. Gill P, Brenner CH, Buckleton JS, Carracedo A, Krawczak M, Mayr WR, et al. DNA commission of the International Society of Forensic Genetics: recommendations on the interpretation of mixtures. Forensic Sci Int 2006;160(2–3):90–101.
25. Aitken CG, Taroni F. Statistics and the evaluation of evidence for forensic scientists, 2nd edn. Chicester, U.K.: John Wiley & Sons, 2004.
26. Good IJ. Probability and the weighing of evidence. London, U.K.: Griffin, 1950.
27. MacKay DJ. Information theory, inference and learning algorithms. Cambridge, U.K.: Cambridge University Press, 2003.
28. Budowle B, Moretti T, Baumstark A, Defenbaugh D, Keys K. Population data on the thirteen CODIS core short tandem repeat loci in African Americans, U.S. Caucasians, Hispanics, Bahamians, Jamaicans, and Trinidadians. J Forensic Sci 1999;44(6):1277–86.
29. Balding DJ, Nichols RA. DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. Forensic Sci Int 1994;64(2–3):125–40.
30. Curran JM, Buckleton J. Inclusion probabilities and dropout. J Forensic Sci 2010;55(5):1171–3.
31. Perlin MW. Cybergenetics TrueAllele technology enables objective analysis of previously unusable DNA evidence. MathWorks Newsletters: Technical Article 2013;1–4.

Additional information and reprint requests:
Mark W. Perlin, Ph.D., M.D.
Cybergenetics, Corp.
160 North Craig Street, Suite 210
Pittsburgh, PA 15213
E-mail: perlin@cybgen.com

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1 Specificity (1 ng). The statistics for specificity were calculated for each contributor group across all three FBI ethnic populations.

Table S2 Specificity (200 pg). The statistics for specificity were calculated for each contributor group across all three FBI ethnic populations.