JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, November 2011, Vol. 56, No. 6
doi: 10.1111/j.1556-4029.2011.01859.x
Available online at: onlinelibrary.wiley.com

PAPER

CRIMINALISTICS

Mark W. Perlin,[1] M.D., Ph.D.; Matthew M. Legler,[1] B.S.; Cara E. Spencer,[1] M.S.; Jessica L. Smith,[1] M.S.; William P. Allan,[1] M.S.; Jamie L. Belrose,[2] M.S.; and Barry W. Duceman,[3] Ph.D.

Validating TrueAllele® DNA Mixture Interpretation*,†

ABSTRACT: DNA mixtures with two or more contributors are a prevalent form of biological evidence. Mixture interpretation is complicated by the possibility of different genotype combinations that can explain the short tandem repeat (STR) data. Current human review simplifies this interpretation by applying thresholds to qualitatively treat STR data peaks as all-or-none events and assigning allele pairs equal likelihood. Computer review, however, can work instead with all the quantitative data to preserve more identification information. The present study examined the extent to which quantitative computer interpretation could elicit more identification information than human review from the same adjudicated two-person mixture data. The base 10 logarithm of a DNA match statistic is a standard information measure that permits such a comparison. On eight mixtures having two unknown contributors, we found that quantitative computer interpretation gave an average information increase of 6.24 log units (min = 2.32, max = 10.49) over qualitative human review. On eight other mixtures with a known victim reference and one unknown contributor, quantitative interpretation averaged a 4.67 log factor increase (min = 1.00, max = 11.31) over qualitative review. This study provides a general treatment of DNA interpretation methods (including mixtures) that encompasses both quantitative and qualitative review. Validation methods are introduced that can assess the efficacy and reproducibility of any DNA interpretation method. An in-depth case example highlights 10 reasons (at 10 different loci) why quantitative probability modeling preserves more identification information than qualitative threshold methods. The results validate TrueAllele® DNA mixture interpretation and establish a significant information improvement over human review.

KEYWORDS: forensic science, expert system, DNA mixture interpretation, genotype, validation study, quantitative data, STR analysis, likelihood ratio, Bayesian model, MCMC computation

Human review of DNA mixtures can often provide identification evidence in criminal investigations (1). However, a genetic calculator can improve this review in several ways. First, computers can increase human *productivity* by enabling each analyst to review more cases, which helps eliminate DNA case backlogs. Also, genetic calculators can often extract more identification *information* from quantitative data, inferring a strong match statistic from weak DNA signals. Finally, computers can guarantee *objectivity* by first inferring an unknown genotype from DNA evidence and only afterward matching that genotype to a suspect genotype.

Cybergenetics TrueAllele® genetic calculator is a statistical computer system that productively solves multiple DNA casework problems in parallel (2). The TrueAllele (TA) calculator infers an informative genotype using all the available quantitative DNA data; with uncertain genotypes, it determines the probability of each allele pair. The calculator commits to this genotype answer and only subsequently makes an objective comparison with a suspect genotype, forming a likelihood ratio (*LR*) weight of evidence statistic relative to a reference population.

In contrast, human DNA review takes a more qualitative approach. The two most commonly used DNA mixture methods in U.S. crime laboratories are combined probability of inclusion (CPI) and combined likelihood ratio (CLR) (3). Both approaches apply thresholds to the DNA data that cut off quantitative information. While CLR considers the victim genotype, CPI makes no use of this case information. Moreover, qualitative review may produce a less objective genotype when inference refers to a known suspect genotype (4).

Previous reports suggest that quantitative genetic calculator DNA mixture interpretation can be more effective than qualitative manual review of the same data. The NIST 05 interlaboratory DNA mixture study showed that careful consideration of signal peak heights and the victim reference could give a match statistic (10^{14}) that was 10 billion times greater than CPI (10^4) (5). Quantitative computer interpretation of a DNA mixture using Markov chain Monte Carlo (MCMC) methods (6) gave a match score (10^{16}) that was 10 million times greater than a list of equally likely allele pairs (10^9) (7). In a recent homicide case, TrueAllele interpretation of a two-person mixture inferred a genotype for the 7% minor contributor that produced an *LR* (10^{11}) that was 10 million times more informative than CPI (10^4) (7). These comparisons invite a more thorough investigation into the relative efficacy of quantitative and qualitative DNA mixture interpretation methods.

An earlier study examined 40 two-person mixture samples of known genotype composition, having varying mixture weights and DNA dilutions (8). We compared quantitative TrueAllele

[1]Cybergenetics Corp., 160 North Craig Street, Suite 210, Pittsburgh, PA 15213.
[2]Northeast Regional Forensic Institute, University at Albany, 1400 Washington Avenue, Albany, NY 12222.
[3]New York State Police Forensic Investigation Center, 1220 Washington Avenue, Albany, NY 12226.

*Presented at the 62nd Annual Meeting of the American Academy for Forensic Sciences, February 22–27, 2010, in Seattle, WA.

†Dr. Mark Perlin is a shareholder, officer, and employee of Cybergenetics in Pittsburgh, PA, a company that develops genetic technology for computer interpretation of DNA evidence. Cybergenetics manufactures the TrueAllele® Casework system, which is one of the methods described in the article.

Received 27 April 2010; and in revised form 23 Sept. 2010; accepted 8 Oct. 2010.

interpretation to qualitative threshold-based methods, using log(*LR*) information as our metric. We demonstrated a sensitivity level of 15 picograms (pg) of culprit DNA for quantitative computer review, relative to a less sensitive 150 pg for qualitative human review. Having already established an information gap between different mixture interpretation methods on laboratory synthesized data, we turn here to a comparison study on casework data.

We collected 16 anonymized and adjudicated two-person case mixture samples for which DNA match statistics had been reported. Half of these cases used available victim information to determine a CLR statistic, while the other half did not, and so reported a CPI statistic. The forensic literature has arguments supporting various mixture methods such as CLR (9,10) or CPI (3,11), but provides little direct comparison data. In this work, we study such mixture case data and make comparisons between quantitative computer-inferred match information and reported qualitative human-inferred match scores.

All of these DNA match statistics, whether qualitative CPI and CLR, or quantitative TrueAllele scores, are *LR*s (7). Moreover, the logarithm of the *LR* is a standard measure of information (12). Therefore, to compare the *efficacy* of two DNA mixture interpretation methods, we examine the pairwise differences of their respective log(*LR*) statistics on the same case, and the average of these differences (13). We assess the *reproducibility* of a mixture interpretation method by determining the within-case standard deviation when duplicate interpretations have been made on the same case (13). These efficacy and reproducibility measures can help determine the reliability of a mixture interpretation method, as well as its suitability for forensic use.

We begin by describing the methods and materials that we used, including computer inference, forensic data, and comparison statistics. We next provide a motivating case example that illustrates relative efficacy and reproducibility of different mixture interpretation approaches for both "with victim" (computer vs. CLR) and "without victim" (computer vs. CPI) situations. We then present our comparison results for the eight CLR cases and the eight CPI cases. Finally, we discuss the need for rigorous scientific validation of DNA interpretation methods, what we learned from the study, and its potential impact on society. The appendices describe the statistical modeling and computation.

Interpretation Methods

Genotype Inference

The probability distribution of any random variable can be determined by Bayes theorem (14), which decomposes the calculation into a *prior* probability and a *likelihood* function (15). At a particular genetic locus l, there is a fixed, finite set of possible allele pair values X. Suppose that Q is a questioned genotype of one of the (one, two, three, or more) contributors to DNA mixture evidence. The prior genotype probability $\Pr\{Q = x\}$ is our belief that questioned genotype Q has the allele pair value x in set X before we examine the evidence data. This genotype prior is well estimated by the population frequency of allele pair x using the product rule.

The likelihood function assesses a genotype candidate value to determine how well it explains the observed data. The likelihood is larger when the quantitative data are better accounted for by a predicted peak height pattern based on the allele pair value. For the ith data observation $d_{l,i}$ at locus l, the likelihood function for genotype Q is the probability $\Pr\{d_{l,i} \mid Q = x, \ldots\}$ of the data conditioned on genotype value x, where "…" denotes the other model variable values.

Combining the prior genotype probability together with I independent genetic data observations, we can compute the *posterior* genotype probability using Bayes theorem as the product of prior probability and joint likelihood

$$\Pr\{Q = x|d_{l,1}, d_{l,2}, ..., d_{l,i}, ...\} \propto \Pr\{Q = x\} \cdot \prod_{i=1}^{I} \Pr\{d_{l,i}|Q = x, ...\} \quad (1)$$

The proportionality "$\propto$" indicates that the product is normalized by dividing by the total data probability $\sum_{x \in X} \left[\Pr\{Q = x\} \cdot \prod_{i=1}^{I} \Pr\{d_{l,i}|Q = x, ...\}\right]$, after considering all possible allele pairs $x \in X$, to produce a genotype probability distribution that adds up to one. When we have a definite belief (say, over 99.5%) that there is only one feasible allele pair, for clarity we shall round this genotype probability value up to one.

Short tandem repeat (STR) mixture data are inherently quantitative, because peak heights appear in rough proportion to the relative amounts of each contributor genotype present in the data (16,17). To make full use of the data and extract the most identification information, the likelihood function must be a quantitative model such as a multivariate normal (18) or gamma (19) distribution. We describe in the Appendix the quantitative Bayesian model equations, and how to compute their solution.

Qualitative mixture interpretation offers an approximation that can be performed by hand, instead of using a genetic calculator. The CPI and CLR methods also use a genetic population prior. However, rather than using quantitative peak height information, qualitative methods truncate the data by applying a *threshold* that reduces the peaks to qualitative all or none "allele" events. In this study, qualitative review applied a predetermined relative fluorescence unit (rfu) threshold value to the STR peak data. The resulting qualitative likelihood function then tests for set inclusion of allele pair candidate values in a set of putative alleles. The likelihood thus becomes a list of allele pair possibilities, each having equal weight (7). As threshold-based methods do not use all the data, we expect them to be less informative than quantitative mixture interpretation (10).

In many cases, it is appropriate to assume the presence of the victim's DNA in a mixture sample. With quantitative inference methods, this assumption reduces the number of unknown contributors by one (say, from two to one), thereby simplifying the genotype inference problem. With qualitative approaches, considering the victim can reduce the list of allele pair possibilities, which concomitantly increases belief in the remaining candidates.

Match Strength

Once a questioned genotype Q has been inferred from the evidence, we want to compare it with a suspect genotype S relative to a reference population genotype R to assess match strength. This assessment is performed using an *LR* (12), expressed in the odds form of Bayes theorem that factors out prior beliefs about guilt or innocence. The *LR* is the gain in identification information resulting from having observed evidence data. If H is the identification hypothesis that the suspect contributed their DNA to the evidence data d, then the *LR* is given by

$$LR = \frac{O(H|d)}{O(H)} \quad (2)$$

In *LR* Eq. 2, $O(H)$ denotes the prior odds $\Pr\{H\}/\Pr\{\overline{H}\}$ of hypothesis H, and $O(H|d)$ the posterior odds $\Pr\{H|d\}/\Pr\{\overline{H}|d\}$,

where $\bar{H}$ is the alternative hypothesis that someone other than the suspect was the contributor.

All the DNA match statistics used in this two contributor mixture case study are *LR*s. The TrueAllele match score TA1 is the *LR* when inferring one unknown contributor genotype using the victim genotype, and TA2 is the TrueAllele *LR* when inferring two unknown contributor genotypes. The reported CLR is already in *LR* form, while CPI is an *LR* (7) that can be obtained from a reported combined probability of exclusion by subtracting from one and taking the reciprocal. The *LR* can be computed as a ratio of probability-weighted likelihoods (Appendix)

$$LR = \frac{\sum_{x \in X} \lambda_Q(x) \cdot s(x)}{\sum_{x \in X} \lambda_Q(x) \cdot r(x)} \tag{3}$$

In *LR* Eq. 3, $\lambda_Q(x)$ is the likelihood function of questioned evidence genotype Q. The posterior probability mass functions (pmf) $r(x)$ and $s(x)$ are for genotypes R and S, respectively. CPI and CLR are special cases of Eq. 3 where $\lambda_Q(x)$ assigns equal likelihood to each feasible allele pair, and $s(x)$ places all its mass on one allele pair, yielding an *LR* statistic having a one divided by a sum of genotype population frequencies. Although we could adjust Eq. 3 for population substructure (7,20) by introducing a co-ancestry coefficient θ, the reported CPI and CLR statistics did not make use of this correction, and so we do not introduce it into this comparison study.

As we are working with adjudicated cases, throughout this study we shall assume that the identification hypothesis H is true. Note that interpretation methods that better preserve DNA identification information will more accurately concentrate evidence genotype probability on a perpetrator allele pair. This concentration yields higher *LR*s when H is true and lower *LR*s when the alternative nonidentification hypothesis is true. Thus, preserving identification information will hinder the guilty and benefit the innocent. Conversely, methods that use less of the available data (e.g., not using some loci or a known victim genotype) are not information "conservative." Such less informative methods disperse genotype probability, forming weaker *LR*s that benefit the guilty and hinder the innocent (21).

This study examines the efficacy and reproducibility of different interpretation methods in extracting identification information from DNA mixture data. The *LR* logarithm is a standard measure of information (22) that we have on hand for every interpretation result. We therefore use the base 10 logarithm $\log_{10}(LR)$ as our information measure.

Mixture Weight

There is some amount of DNA from each contributor present in an evidence mixture sample. The proportions of each contributor in the sample give the DNA template mixture weight vector, whose components add up to one. Each STR locus experiment j on the DNA template measures mixture weight. These measurements are conditionally independent, given the template weight. The dependence of the observed peak height data on mixture weight is expressed through the likelihood function $\Pr\{d_j \mid W = w,\ldots\}$, now rewritten to emphasize the mixture weight variable W. Here, w is a mixture weight vector, and "…" includes genotype and other values (18).

Combining these independent likelihood values together with a (say, uniform) prior probability $\Pr\{W = w\}$ using Bayes theorem gives the continuous posterior probability distribution

$$\Pr\{W = w | d_1, d_2, \ldots, d_J, \ldots\} \propto \Pr\{W = w\} \cdot \prod_{j=1}^{J} \Pr\{d_j | W = w, \ldots\} \tag{4}$$

Mixture weight Eq. 4, and its hierarchical refinements (6), can be solved using MCMC computation, as described in the Appendix.

Data Uncertainty

There are two sources of data uncertainty affecting our confidence in a peak height measurement. Variation in polymerase chain reaction (PCR) *amplification* and template sampling accounts for the different peak height patterns that are seen in multiple experiments on the same DNA template. A greater mass of a DNA fragment yields a more confident peak height y having a lower coefficient of variation. So probability modeling (23) and empirical observation (24) have us scale the amplification variance with the peak height as $y \cdot \sigma^2$ to account for stochastic effects.

There is also a signal *detection* variance that arises from the DNA sizing instrument. This baseline variation can be modeled by a constant background variance τ^2, which helps account for dropout alleles. The use of two independent variance components $y \cdot \sigma^2$ and τ^2 is in the spirit of current human review practice (3), which may set separate peak height thresholds for an "inclusion" (amplification) and "exclusion" (detection), respectively. Unlike human review, however, statistical computing is able to infer this peak variation directly from the quantitative evidence data (25).

The variance parameters have probability distributions determined by a prior and a likelihood. Data variance priors $\Pr\{\sigma^2 = s^2\}$ and $\Pr\{\tau^2 = t^2\}$ can be modeled using an inverse gamma distribution (25). The likelihood $\Pr\{d_j | \sigma^2 = s^2, \tau^2 = t^2, \ldots.\}$ of observing peak heights d_j at locus experiment j describes the probability of the independent data peak events given the data uncertainty variances and other parameters (genotype, mixture weight, …).

We can combine the prior variance probability together with the likelihoods of these J independent quantitative peak experiments. Bayes theorem then produces the posterior probability variance distributions

$$\Pr\{\sigma^2 = s^2 | d_1, d_2, \ldots, d_j, \ldots\} \propto \Pr\{\sigma^2 = s^2\} \cdot \prod_{j=1}^{J} \Pr\{d_j | \sigma^2 = s^2, \ldots\}$$

$$\Pr\{\tau^2 = t^2 | d_1, d_2, \ldots, d_j, \ldots\} \propto \Pr\{\tau^2 = t^2\} \cdot \prod_{j=1}^{J} \Pr\{d_j | \tau^2 = t^2, \ldots\} \tag{5}$$

The data uncertainty Eq. 5 can be solved using Metropolis-Hastings (26,27) statistical search (Appendix).

Materials

Computer Software

Cybergenetics TrueAllele Genetic Calculator uses a fully Bayesian model of the STR data generation process, based on genotype, mixture weight, and data certainty probability distributions (Eqs [1], [4], and [5], respectively). The calculator accounts for PCR stutter (17,28), relative amplification and other experiment factors. Conditioning on the observed quantitative STR data, the TrueAllele computer explores the model's parameter space using MCMC

statistical search to determine the posterior probability distribution for every variable (Appendix). Results for variables of interest, such as genotypes and mixture weights, are reported as probability distributions (2). Match statistic results are reported as *LR*s (7,29). The TrueAllele Visual User Interface (VUIer™) program lets a user visually explore their STR data and computed results (e.g., genotypes, matches, and mixture weights), as well as conduct "what-if" analyses.

In this study, when the victim genotype was known, we had the TrueAllele system solve for one unknown contributor genotype, sampling for 24,000 cycles following initial burn-in to equilibrium (30). When no genotype was assumed, we asked TrueAllele to solve for two unknown contributor genotypes, with 48,000 sampling cycles after burn-in. To assess the reproducibility, the computer solved all cases in duplicate. Markov chain convergence was assessed using the Gelman-Rubin statistic (31) and by visual inspection of the mixture weight chain. Match information was computed as $\log_{10}(LR)$, where the *LR* compared the inferred unknown genotype to a suspect reference, relative to a population (Appendix).

Mixture Cases

We reviewed 166 mixture samples that occurred in 40 adjudicated cases and one proficiency test conducted in the New York State Police (NYSP) Forensic Investigation Center (FIC) in Albany, NY. About half of these mixtures were not included in this mixture study because they had a clear major contributor whose genotype could be reported using a single source match statistic, such as conditional match probability. The computer found 86 mixture items that matched a suspect (with an *LR* score), while the laboratory reported a numerical match score for 26 of these items.

To make a numerical comparison, we focused on just the 26 items whose human review produced a quantifiable match score. We identified 16 samples that had an unambiguous two-person mixture that had been reported out using a CPI or CLR statistic computed in the CODIS Popstats module. Eight of these statistics were CPI, while the other eight were CLR. These 16 two-person mixture samples comprise our study data.

Population Databases

The NYSP computes match statistics relative to four population databases: African American (BLK), Caucasian (CAU), Southeast Hispanic (SEH), and Southwest Hispanic (SWH). They use the standard FBI allele databases (32) for these populations. All match statistics reported in this study, whether from the original case or computed in TrueAllele, were calculated using these population databases.

Validation Methods

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, using a population prior and a likelihood function. Computer-based modeling methods (6,33), such as the TrueAllele system (2), employ a quantitative likelihood function that compares proposed patterns with STR peak heights.

A qualitative binary method, such as CPI or CLR, forms a genotype list of length N that contains reportable allele pairs, with each one assigned an equal likelihood (7). An *LR* compares this evidence genotype to a suspect genotype, relative to a population genotype, through their pmfs to obtain match information (Appendix). Thus, the *LR* provides a universal mechanism for comparing match information between mixture genotypes inferred by different methods, relative to the same suspect and population.

The log(*LR*) is a standard additive measure of information (12,22). The match statistics TA1, TA2, CLR, and CPI can all be viewed as *LR*s (7). Therefore, we can compare the relative efficacy of two mixture interpretation methods by examining the difference in their $\log_{10}(LR)$ scores. For a set of cases, we can also look at the mean value of these information differences. Statistical significance between the pairwise differences can be measured using a *t*-test.

In this study, we are chiefly concerned with comparing differences in identification information between quantitative and qualitative mixture interpretation methods. When the victim genotype is known and used, the information difference between computer and human interpretation for one inferred genotype is log(TA1) − log(CLR). This information difference is the same as the logarithm of the *LR* information gain log(TA1/CLR). When the victim is not available for genotype inference, two unknown genotypes are inferred, and this information improvement becomes log(TA2) − log(CPI).

Reproducibility

An important aspect of scientific reliability is a method's reproducibility (34). The reproducibility of a set of measurements is conventionally reported as the standard deviation of these numbers (35). 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*) values. From this set of information measurements, we can assess the method's reproducibility by computing a standard deviation.

To sharpen the reproducibility estimate of a mixture interpretation method, we use more cases. The "within-case" standard deviation σ_w (36) describes the method's reproducibility over a population of mixture cases (13). We can compute σ_w as the root mean square deviation of replicated $\log_{10}(LR)$ information scores, relative to the mean value within each case (36), as shown in Eq. 6.

$$\sigma_w^2 = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J_i} (s_{ij} - \bar{s}_i)^2}{\sum_{i=1}^{I} J_i} \qquad (6)$$

Here, I is the number of cases, J_i is the number of independent interpretations of the ith case, s_{ij} is the $\log_{10}(LR)$ score of the jth interpretation of the ith case, and $\bar{s}_i$ is the mean score of the s_{ij} values within the ith case.

Case Example

To illustrate our comparison measures of efficacy and reproducibility, we examine two mixture samples from a sexual assault case in some detail. In this case, buccal reference samples from the victim (item G) and suspect (A) were obtained. The items of evidence

we consider here are a two-person vaginal swab (F) mixture, which the laboratory reported using a CLR statistic, and a two-person anal swab (E) mixture on which CPI was performed by a vendor laboratory. We begin with the simpler CLR analysis that assumes the known victim contributor genotype and infers the genotype of the unknown second contributor.

One Unknown Contributor

We start by examining the quantitative peak height data for evidence item F (Table 1, column F). The four equal peak heights at loci D18 and vWA suggest a 50:50 two-person mixture. The 2:1:1 peak height ratios at loci having three major peaks (e.g., D13, D16, D21, D3, D5, and TPOX) further support this hypothesis. We see that the victim alleles (Table 1, column G) are included in mixture item F.

TABLE 1—*Data for case example.*

Peak Height Data		A	F	E	G
Experiment	Allele	Suspect	Evidence	Evidence	Victim
CSF1PO	10	791	554	431	661
	12	550	556	333	706
D13S317	8		220	87	796
	12	480	303	90	574
	13	588	183	34	
D16S539	9		385	293	719
	11	884	658	371	816
	13	899	361	104	
D18S51	12		291	318	746
	14		222	184	652
	15	604	219	38	
	18	552	262	20	
D21S11	28		186	139	970
	29	508	538	168	603
	35	603	175	34	
D3S1358	14		437	202	813
ProfilerPlus	15	1006	678	261	707
	17	813	374	46	
D3S1358	14		331	229	789
Cofiler	15	918	472	411	595
	17	789	324	53	
D5S818	9	684	203	21	
	11		239	142	814
	12	737	488	172	726
D7S820	10	468	420	94	412
ProfilerPlus	11	448	352	88	538
D7S820	10	570	351	166	570
Cofiler	11	393	301	151	514
D8S1179	13		283	169	640
	14		249	215	666
	15	1259	461	136	
FGA	21		43	41	
	22		511	321	1620
	23	527	235	49	
	27	563	186	74	
TH01	7		710	563	1422
	8	735	334	53	
	9	682	286	68	
TPOX	8		289	252	889
	9	666	558	259	801
	10	688	284	54	
vWA	14	826	285	118	
	15	831	347	93	
	17		414	323	702
	18		436	252	802

Quantitative allele information is shown in separate columns for the suspect (A), evidence (F and E), and victim (G) samples. The peak height (as rfu) is given in each locus allele row for all the samples.

Dual human review of 50:50 mixture F was performed by the NYSP FIC using the victim reference G. The inferred CLR genotypes had one or three possibilities at each locus, with each potential unknown allele pair allocated equal likelihood, so that the posterior probability was proportional to the population genotype frequency (Table 2, columns CLR and Pop). The *LR* of the evidence genotype (CLR) was computed relative to the suspect (A) and four populations (Table 3, column CLR). We conservatively show the BLK population, as it gave the smallest match score. The FIC did not determine an *LR* score for D7 or CSF because at each locus, the victim and suspect genotypes were equal (i.e., had the same two alleles), and the NYSP does not report unknown genotypes at loci where there is no indication of a second contributor.

After uploading the case data to the TrueAllele database, we set up interpretation requests for inferring the victim, evidence and suspect genotypes (Fig. 1). The evidence request was run in duplicate, conditioning on the data (Table 1, column F) and the victim genotype. The computer searched for the unknown second contributor genotype, without any knowledge of a suspect genotype. The inferred definite evidence genotype (Table 2, column TA1, runs 1 and 2) was compared with the suspect (A) and population (BLK) genotypes to find the *LR* that the suspect was present in the mixture (7). The *LR* scores at each locus, and the joint match statistics are shown (Table 3, column TA1, runs 1 and 2). TrueAllele inferred the mixture weight of the unknown second contributor to be 48.2% (48.6% in the duplicate run) with a standard deviation of 2.7%.

The computer extracted more identification information from evidence item F ($\log(LR) = 17.65$) than did the human review ($\log(LR) = 13.52$) (Table 3, joint *LR* information, columns TA1 and CLR). As the *LR* calculations use the same suspect and population for comparison, the only difference was in the inferred genotype (7). So the explanation for the four orders of magnitude (10^4) *LR* increase seen in the TA1/CLR ratio resides entirely in the genotype probability distributions (Table 2, column F). With quantitative data from a well-amplified 50:50 mixture, and a known victim genotype, a genetic calculator can often infer a definite second genotype, as in this case. The resulting probability distribution then places all of its belief in one allele pair at each locus. In this case, the resulting inferred evidence genotype is the same as the suspect's genotype.

Threshold-based CLR mixture interpretation does not use all the peak height information. Therefore, many practitioners conservatively infer a single allele pair only when the evidence shows two nonvictim alleles. With three alleles, CLR infers three equally likely allele pairs from the thresholded allele peak data. Loci with matching victim and suspect genotypes were not included in the laboratory's reported statistic, a policy that can further reduce CLR information.

However, quantitative peak height data (not used by CLR) can often indicate just one definite genotype solution. The TrueAllele VUIer Explain window provides a "what-if" analysis that lets a user explore alternative genotype values and mixture weights. Looking at locus D3, we see that of the three CLR reported allele pairs, only the [15 17] possibility realistically fits the quantitative data (Fig. 2). The other two candidates do not adequately explain the peak heights, and so the likelihood function of Eq. 1 assigns them essentially zero probability. Therefore, with Bayes theorem having duly considered but "eliminated the impossible" (37), TrueAllele assigns the remaining allele pair [15 17] a rounded probability of one at locus D3 (Table 2, column F).

TABLE 2—*Genotypes for case example.*

Genotype		Evidence F			Evidence E			Sus	Pop
			TA1			TA2			
Locus	Allele Pair	CLR	Run 1	Run 2	CPI	Run 1	Run 2	A	BLK
CSF1PO	10, 10	0.224			0.224	0.387	0.374		0.074
	10, 12	0.499	1	1	0.499	0.610	0.624	1	0.163
	12, 12	0.277			0.277				0.090
D13S317	8, 8					0.002			0.001
	8, 12					0.011	0.012		0.035
	8, 13	0.145				0.111	0.164		0.009
	12, 12					0.030	0.024		0.234
	12, 13	0.711	1	1		0.646	0.615	1	0.121
	13, 13	0.144				0.191	0.176		0.016
D16S539	9, 9				0.075				0.039
	9, 11				0.260				0.117
	9, 13	0.313			0.138				0.066
	11, 11				0.225				0.087
	11, 13	0.542	1	1	0.239	0.972	0.972	1	0.097
	13, 13	0.144			0.064	0.024	0.024		0.027
D18S51	12, 12					0.102	0.082		0.003
	12, 15					0.850	0.868		0.020
	12, 18					0.008	0.011		0.016
	14, 14					0.011	0.008		0.004
	14, 15					0.006	0.006		0.022
	15, 18	1	1	1		0.013	0.015	1	0.045
D21S11	28, 28					0.010	0.009		0.047
	28, 29					0.084	0.079		0.083
	28, 35	0.475				0.044	0.042		0.012
	29, 29					0.103	0.100		0.036
	29, 30						0.003		0.069
	29, 35	0.508	1	1		0.752	0.759	1	0.011
	35, 35	0.017							0.001
D3S1358	14, 14				0.034				0.015
	14, 15				0.180				0.072
	14, 17	0.219			0.120				0.049
	15, 15				0.241	0.034	0.021		0.086
	15, 17	0.586	1	1	0.320	0.962	0.977	1	0.118
	17, 17	0.195			0.106				0.041
D5S818	9, 9	0.025							0.000
	9, 11	0.452				0.013	0.010		0.007
	9, 12	0.523	1	1		0.210	0.187	1	0.010
	11, 11					0.037	0.046		0.069
	11, 12					0.139	0.150		0.187
	12, 12					0.598	0.604		0.127
D7S820	10, 10	0.340			0.340	0.030	0.023		0.105
	10, 11	0.486	1	1	0.486	0.802	0.824	1	0.145
	11, 11	0.174			0.174	0.165	0.152		0.050
D8S1179	13, 13				0.156	0.500	0.571		0.050
	13, 14				0.306	0.030	0.021		0.149
	13, 15	0.444			0.172				0.096
	14, 14				0.150	0.028	0.030		0.112
	14, 15	0.434			0.168	0.019	0.014		0.143
	15, 15	0.122	1	1	0.047	0.421	0.361	1	0.046
FGA	21, 22					0.014	0.014		0.057
	21, 23					0.259	0.171		0.032
	21, 27					0.176	0.115		0.006
	22, 23					0.027	0.037		0.057
	22, 27					0.011	0.011		0.010
	23, 23					0.010	0.004		0.016
	23, 27	1	1	1		0.496	0.638	1	0.006
	27, 27						0.003		0.000
TH01	7, 7				0.249				0.194
	7, 8				0.256	0.009	0.010		0.164
	7, 9				0.244	0.328	0.364		0.128
	8, 8				0.066				0.034
	8, 9	1	1	1	0.126	0.662	0.623	1	0.054
	9, 9				0.060				0.021
TPOX	8, 8				0.408				0.136
	8, 9				0.314	0.013	0.018		0.134
	8, 10	0.678			0.148	0.275	0.273		0.069
	9, 9				0.060				0.033

Continued

TABLE 2—*Continued.*

Genotype		Evidence F			Evidence E				
			TA1			TA2		Sus	Pop
Locus	Allele Pair	CLR	Run 1	Run 2	CPI	Run 1	Run 2	A	BLK
9, 10	0.261	1	1	0.057	0.702	0.702	1	0.034	
10, 10	0.061			0.013				0.009	
vWA	14, 14				0.017				0.004
	14, 15	1	1	1	0.063	0.992	0.997	1	0.031
	14, 17				0.097				0.024
	14, 18				0.068				0.018
	15, 15				0.058				0.056
	15, 17				0.178				0.087
	15, 18				0.125				0.064
	17, 17				0.136				0.034
	17, 18				0.191				0.050
	18, 18				0.067				0.018

The inferred genotype probability distributions are shown for evidence samples F and E, along with the suspect and BLK population reference genotypes. Allele pair values are shown up to a cumulative probability of 0.99. When no genotype was inferred at a locus, its probability entry was left blank. Mixture sample F was interpreted using combined likelihood ratio (CLR), and so the feasible allele pairs were assigned one or three equal likelihoods, producing probabilities proportional to the population prior. As the victim was known, and the mixture peak height data was informative, the TA1 computer method inferred a unique (maximally informative) genotype. Without a victim profile, both human (combined probability of inclusion, CPI) and computer (TA2) interpretations of mixture E, produced uncertain genotypes, represented as probability distributions. Note that the computer's TA2 probability is higher than the human CPI probability at every suspect-matching allele pair.

TABLE 3—*Likelihood ratios for case example.*

Likelihood Ratio (*LR*)	*LR* of F and A with BLK			*LR* of E and A with BLK		
	CLR	TA1		CPI	TA2	
Locus	Consensus	Run 1	Run 2	Consensus	Run 1	Run 2
CSF1PO		6.17	6.17	3.06	3.76	3.89
D13S317	6.83	8.27	8.27		6.16	5.79
D16S539	5.26	10.34	10.34	2.31	10.04	9.54
D18S51	22.97	22.59	22.59		0.77	1.89
D21S11	42.78	93.54	93.56		77.26	75.11
D3S1358	4.88	8.52	8.52	2.63	8.20	7.83
D5S818	57.68	100.88	100.81		10.51	10.37
D7S820		6.93	6.93	3.33	5.59	5.67
D8S1179	3.53	21.85	21.85	1.68	14.41	12.61
FGA	180.18	178.84	178.84		93.96	87.59
TH01	18.54	18.62	18.62	1.68	11.44	11.19
TPOX	8.98	29.59	29.59	2.41	22.22	21.41
vWA	31.75	31.92	31.92	2.58	31.67	30.59
Joint	3.34×10^{13}	4.55×10^{17}	4.56×10^{17}	1.09×10^{3}	7.27×10^{13}	1.13×10^{14}
Information	13.52	17.65	17.66	3.04	13.86	14.05

The *LR* scores are shown for the inferred evidence genotypes, relative to the suspect and BLK population genotypes. The first three columns show the inferred genotype F locus *LR*s for combined likelihood ratio (CLR) and duplicate TA1 runs, where the victim genotype is known. The last three columns give locus *LR*s of the inferred genotype E for combined probability of inclusion (CPI) and duplicate TA2 runs, where two unknown contributors were inferred; the *LR* is for the matching contributor genotype. The last two rows give the joint *LR*s and their base 10 logarithms.

Two Unknown Contributors

We now turn to the quantitative peak height data for evidence item E (Table 1, column E). The four allele data of locus vWA suggests the presence of two unknown contributors: one major and one minor. The victim genotype is not used in this interpretation because it was not available to the vendor analyst who conducted the manual mixture interpretation. Therefore, the task is to infer the genotypes of both unknown contributors and then compare them with the suspect.

Human review of two-person mixture E was performed by a vendor laboratory using an allele inclusion analysis. This CPI approach examined all peaks with heights of at least 50 rfu, considering them to be alleles. There were five loci (D13, D18, D21, D5, and FGA) at which some peaks had heights that were under threshold, but matched the suspect; these loci were not used in computing the CPI statistic. (Allele dropout *LR* methods that might incorporate such low peak data (9) were not yet approved for forensic DNA interpretation use in NYS.)

At the remaining loci, a PI match statistic was computed using the population frequencies of I observed alleles as

$$PI = \frac{1}{\left(\sum_{i=1}^{I} f_i\right)^2} \tag{7}$$

where f_i is the observed frequency of allele i at locus l. The PI Eq. 7 can be viewed as comparing an inferred genotype Q having

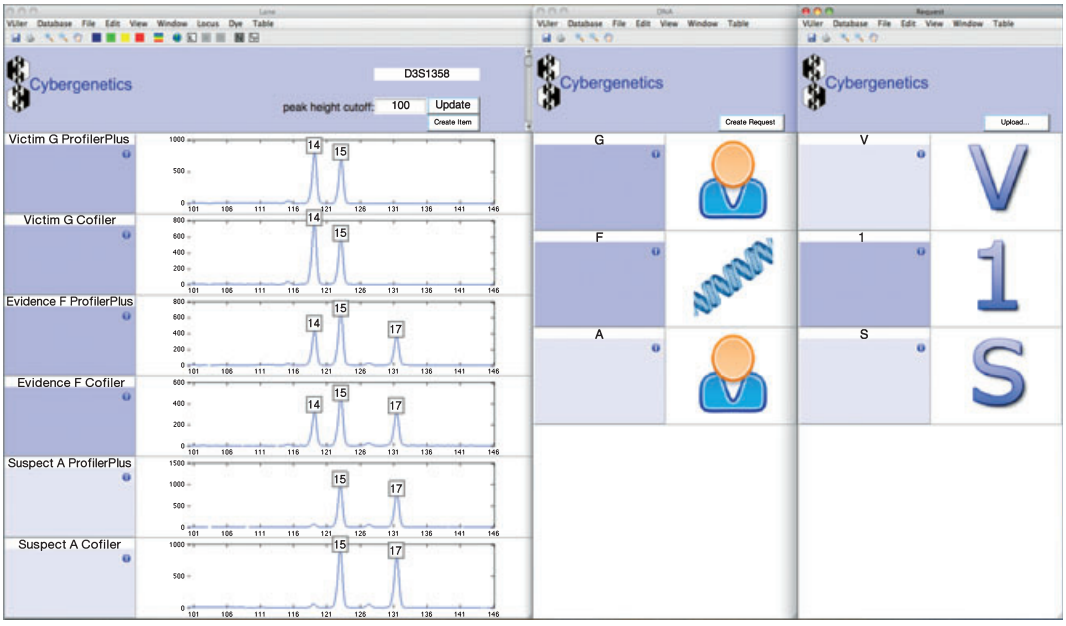

FIG. 1—*Setting up interpretation requests. The TrueAllele VUIer Request module shows annotated data signals (left column), DNA items (middle), and interpretation requests (right). The user proceeds from left to right, grouping short tandem repeat (STR) data rows to form DNA items, and then grouping these items to form requests. Request "1" (right column, middle row) sets up a one unknown contributor genotype inference. When the user selects Request "1" (right), the VUIer highlights the DNA items involved—victim reference G and mixture evidence F (middle)—along with the STR data used (left column, first four rows).*

$N = I(I+1)/2$ equally likely (included) allele pairs to a matching suspect genotype *S*, relative to a population genotype *R*. Substituting these genotypes *Q*, *R,* and *S* into Eq. 2 shows that PI is an *LR* (7). Therefore, CPI can be compared with other *LR* match statistics to measure information differences.

The human-inferred PI evidence genotype *Q* allocates equal likelihood 1/*N* to each individual allele pair, producing a posterior probability proportional to the population prior (Table 2, column E, CPI). With just two allele peaks (CSF, D7), each allele pair is assigned likelihood 1/3. With three peaks (D16, D3, D8, TH01, and TPOX), the PI likelihood is 1/6. Four alleles (vWA) generate 10 possibilities, each receiving from PI a likelihood of 1/10. CPI only reports a statistic at suspect-matching loci that have all allele peaks over a preset threshold (e.g., D21's allele 35 is under threshold, and so that locus is not included in the combined PI). The resulting CPI *LR* scores are shown (Table 3, column CPI), giving a combined *LR* product of 1090.

We also analyzed this sample as a two unknown case in the TrueAllele system. The computer solved for the two genotypes and the mixture weighting, along with other variables, using MCMC search of the hierarchical Bayesian model (Appendix). The genotype results, run in duplicate, are shown for the nonvictim minor contributor (Table 2, column E, TA2, runs 1 and 2). (The inferred major contributor genotype perfectly matched the victim, and is not shown.) We see that the allele pair probabilities at each locus within a run are not equal. The mixture weight of the suspect-matching minor contributor was 23.4%, with a standard deviation of 4.1% (Fig. 3). The *LR* was around 100 trillion (Table 3, column TA2, runs 1 and 2), reproduced in two independent computer genotype inferences.

The computer-inferred TA2 *LR* information of 10^{14} for evidence E is 11 orders of magnitude greater than the human CPI *LR* result of 10^3. There are several reasons for this 100 billion-fold TA2/CPI ratio information improvement.

1. By using *quantitative* peak height data, the computer can analyze the identification information in every peak, including those that are below the human review threshold. The system models amplification variance $y \cdot \sigma^2$, which accounts for stochastic effects at each peak of height *y*. Introducing a detection variance τ^2 permits meaningful comparisons down to a peak height of zero (e.g., complete allele dropout). For example, at locus D13, a matching allele 13 peak with (under threshold) height 34 rfu precludes any CPI analysis (Tables 1 and 2, column E). But computer modeling can account for this greater data peak uncertainty and assign a probability of over 60% to the suspect's [12 13] genotype (Table 2, column E), yielding an *LR* locus contribution of 6 (Table 3, column TA2).
2. The computer searches for two unknown *genotypes* that can explain the quantitative data, thereby separating the data into two different contributor genotypes. For example, at locus vWA, TrueAllele uses the quantitative peak height pattern (two low, two high) (Table 1) to determine a suspect-matching contributor [14 15] allele pair and a second victim-matching contributor [17 18] genotype value with over 99% certainty (Table 2). But CPI allocates only a 10% likelihood to the [14 15] suspect genotype, because it treats equally all 10 pairwise combinations of the four alleles (Table 2). The more definite TrueAllele-inferred genotype has an *LR* of 30, while CPI's less distinct genotype distribution produces an *LR* of only three (Table 3).
3. Calculating the *mixture weight* variable facilitates the genotype separation in this 75:25 situation. This informative use of mixture weight is seen at locus D16, where there are three alleles of varying peak heights (Table 1, column E). Weighting the contributor allele pairs [9 11] and [11 13] (matching victim and suspect, respectively) in a 3:1 proportion produces a pattern that fits the observed data (Fig. 4). The computer assigns

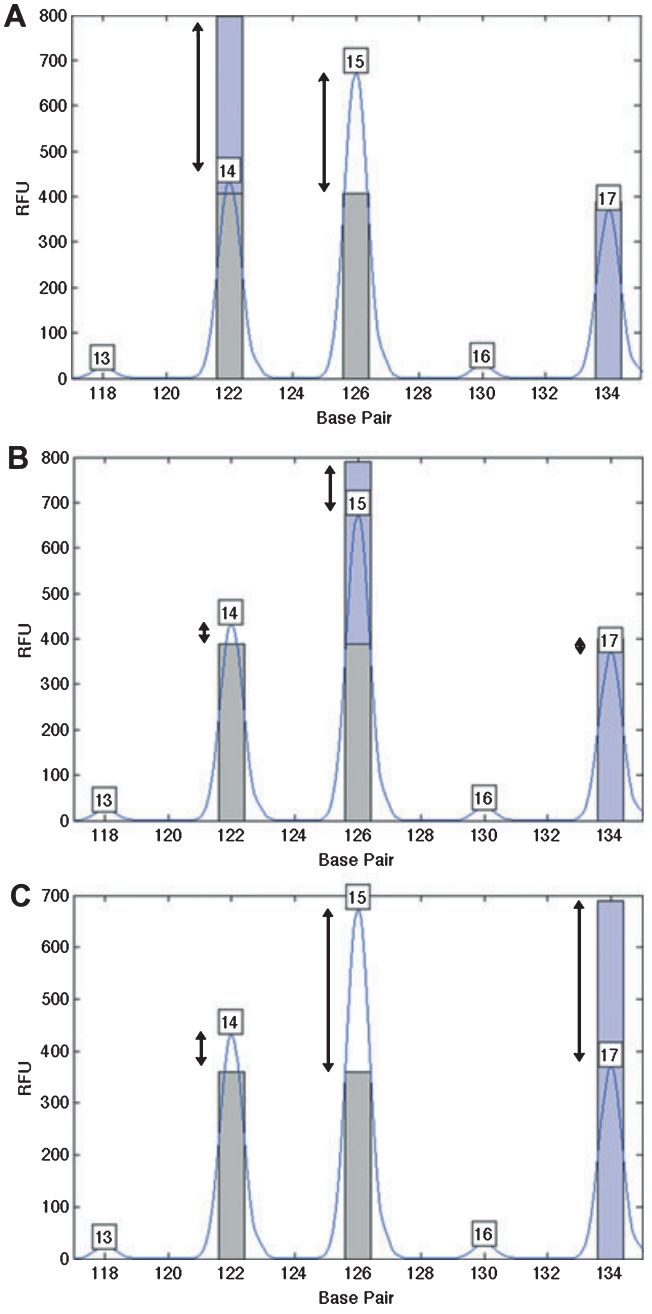

FIG. 2—*Explaining interpretation results. The quantitative data peak TA1 method inferred only one genotype possibility [15 17] for the unknown second contributor at locus D3 of evidence sample F. However, the qualitative combined likelihood ratio (CLR) method inferred three possibilities ([14 17], [15 17], [17 17]), assigning them each one an equal likelihood. The TrueAllele Explain window shows why the probability model considers some allele pairs to be more likely than others, relative to the quantitative data. The victim genotype allele (gray) and unknown contributor alleles (blue) combine in proportion to their mixture weights to form a base pair versus rfu pattern (bars) that can be compared with the short tandem repeat (STR) data (curve). Comparison (A) for genotype value candidate [14 17] shows large deviations between peak and model heights at alleles 14 and 15, so this is a highly unlikely possibility. Similarly, Explain window comparison (C) for candidate [17 17] is extremely unlikely, because of large data and model disparities at alleles 15 and 17. The visual comparison (B) for candidate [15 17] shows a close fit of the model pattern to the data signal at all alleles. As there is no other genotype value with a good quantitative fit, TrueAllele assigns value [15 17] a probability of one. The qualitative CLR method treats the data alleles 14, 15, and 17 as having indistinguishable peak heights, and so it cannot differentiate between the three allele pairs with obligate allele 17. Therefore, CLR assigns each pair an equal likelihood of 1/3.*

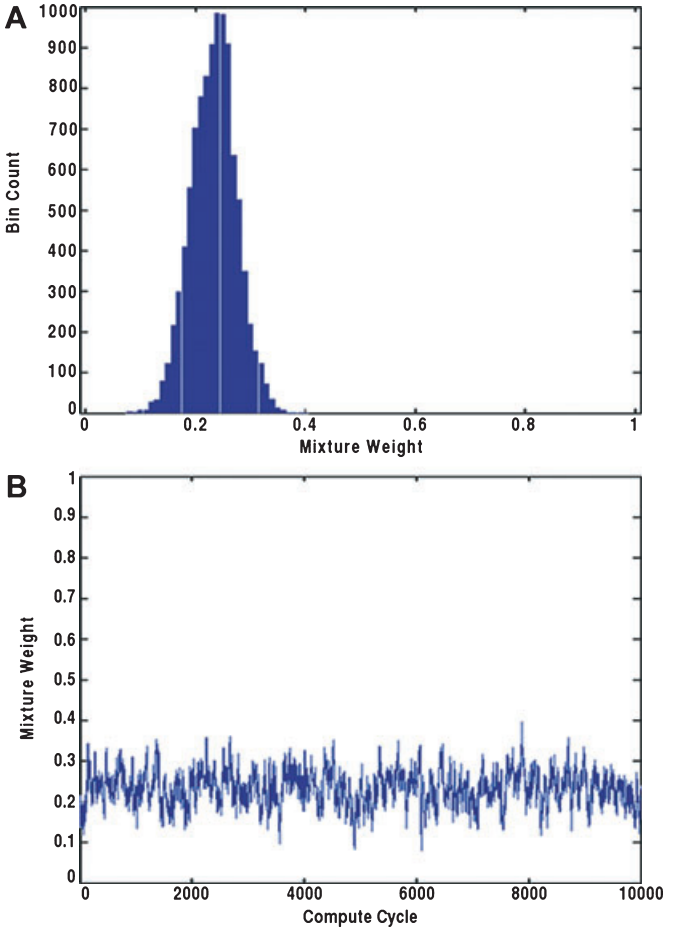

FIG. 3—*Mixture weight interpretation results. The TrueAllele VUIer Mixture Weight window shows (A) a histogram that gives the probability distribution of mixture weight. This histogram bins the (B) different template mixture weight values visited during the Markov chain problem solving, which is viewable as a chain history. The mixture weight distribution of the unknown contributor has a mean of 23.4% and a standard deviation of 4.1%.*

this explanatory genotype combination a probability of 97.2% (Table 2), giving a locus *LR* of 10 (Table 3). CPI, on the other hand, does not employ quantitative mixture weight and only gives an *LR* of 2.3 (Table 3).

4. The TrueAllele-inferred genotypes are not limited to having equal likelihood values. Instead, their distributions assign *higher likelihoods* to more likely genotype values that better fit the quantitative data. This genotype inference result is seen at locus TPOX, where matching allele pair [9 10] has a probability of 70%, allele pair [8 10] a chance of 27%, and value [8 9] only 2% (Table 2). The objectively inferred genotype happens to place greater weight on the matching allele pair, so the *LR* is 21 (Table 3). CPI, however, distributes its likelihood equally across the six allele pair possibilities (Table 2). This less informative genotype distribution reduces the *LR* 10-fold to just 2.4 (Table 3).
5. To be *objective*, the computer solves for genotypes at *all* loci, without having any knowledge of a suspect genotype. While severe artifacts (e.g., spikes) may occasionally render a locus experiment unusable, generally all the data are used. In item E, locus D21 has a suspect-matching 35 allele with a peak height of 34 rfu (Table 1). As this matching peak is below threshold, CPI does not use this locus. But TrueAllele must—it models

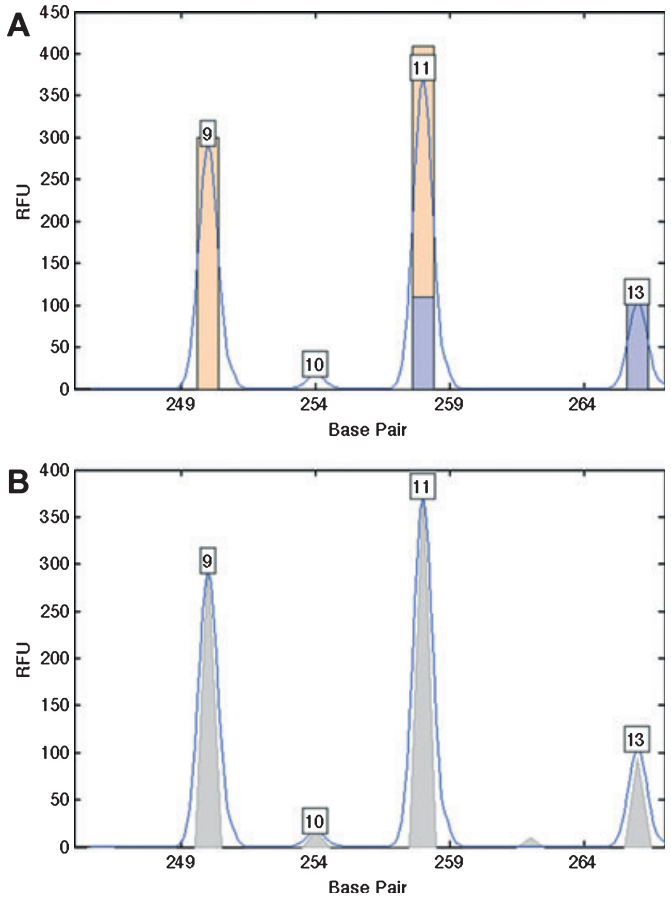

FIG. 4—*Mixture model fits the data. TrueAllele VUIer Explain views are shown for locus D16 of sample E. (A) The weight combination of first (blue) and second (orange) contributor genotype alleles forms a pattern that fits the observed data. (B) A model pattern (gray) is shown which further accounts for PCR stutter and relative amplification effects, providing a closer fit to the data.*

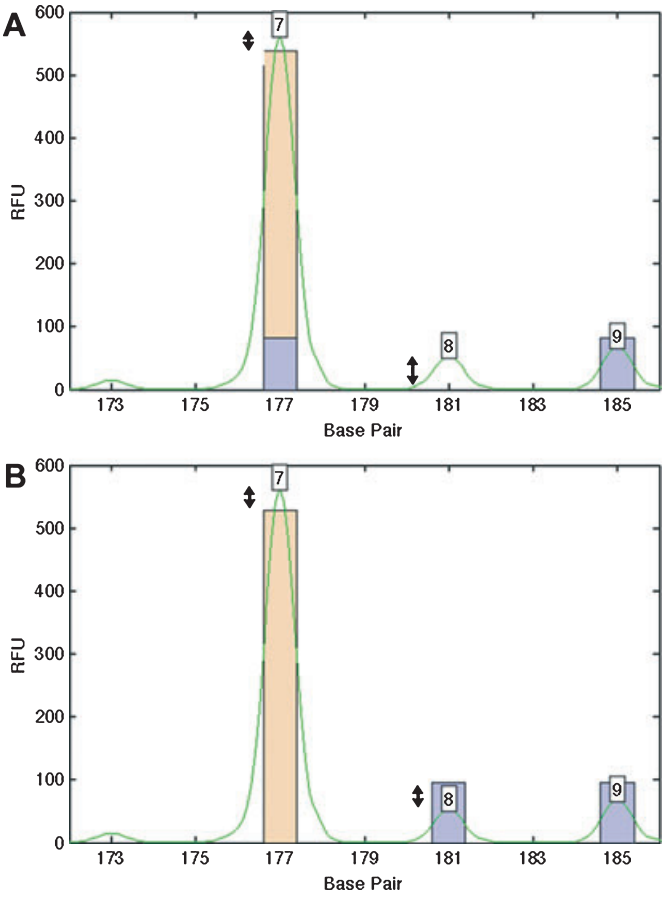

FIG. 5—*Explaining the genotype possibilities. The TrueAllele VUIer Explain window shows two explanations at locus TH01 of mixture sample E. The second contributor's genotype is inferred to be [7 7] (orange). But there are two genotype possibilities for the suspect-matching first contributor (blue). (A) Genotype candidate [7 9] does not quantitatively account for the allele 8 peak as well as (B) candidate [8 9]. Therefore, allele pair [8 9] has a higher probability.*

quantitative data under all possible two contributor genotype scenarios, inferring probabilities for all allele pairs. It assigned a 75% probability to [29 35] (Table 2), which matched the suspect with an *LR* of 75 (Table 3). Not all loci need increase the weight of evidence, as we shall see next.

6. A valid genotype probability inference specifies a *prior* distribution from Eq. 1 that comes into play when the data are missing or uninformative. Locus D18 has two low-level non-victim data peaks: allele 15 with peak height 38 rfu, and allele 18 at 20 rfu (Table 1). These low peaks appear in some low likelihood genotype values, so the population frequency prior is helpful here. The matching [15 18] allele pair has only a probability of 1.3% (run 1), with an *LR* of 0.77 (Tables 2 and 3). This *LR* value is less than one and so decreases the weight of evidence. CPI could not provide any statistic at locus D18 as the two matching allele peaks are below threshold.
7. TrueAllele uses a quantitative *likelihood function* that compares data peaks with a genotype model pattern. This comparison is seen at locus TH01, where there is a tall victim allele peak at 7, and two smaller allele peaks at 8 and 9 (Table 1). The computer assigned a 2/3 probability to allele pair [8 9] and half that probability to [7 9] (Table 2). Looking at the data peaks and hypothesized model patterns (Fig. 5), we see that neither candidate gives a perfect fit to the data. However, candidate [8 9] accounts for the low peak at allele 8, which [7 9] does not, and so the computer gives that allele pair a higher probability, yielding an *LR* information value of 11 (Table 3). CPI has a less informative likelihood function that qualitatively assigns the same weight to every individual allele pair and a zero to the others (7). With six possibilities, dividing the likelihood into six equal parts irrespective of the observed quantitative peak heights, CPI reduces the *LR* match score 10-fold to 1.7 (Tables 2 and 3).
8. Bayes Theorem requires us to consider *all allele pairs* (15) when inferring a mathematically valid genotype. This is because in Eq. 1, every genotype candidate $x \in X$ (even an apparently unlikely one) enters into the denominator's total probability. At locus D5, allele 9 has a peak height of only 21 rfu (Table 1). Well below the 50 rfu threshold, this suspect-matching allele is invisible to qualitative review, and so CPI is silent about the locus (Tables 2 and 3). However, TrueAllele must assess each genotype candidate relative to the quantitative data and so discovers that suspect-matching allele pair [9 12] has a probability of about 20%, which contributes a D5 locus *LR* of 10 (Tables 2 and 3).
9. Efficient statistical inference has us include *all data* affected by a variable in a likelihood function (15). At locus D3, we have independent experiments on item E from both the ProfilerPlus

and Cofiler STR panels (Table 1). Therefore, we form the product of likelihoods, together with the genotype prior, to infer the genotype via Eq. 1 (Table 2). We see the fit of the genotype patterns to the two independent D3 experiments (Fig. 6). Note that the three allele ProfilerPlus experiment shows high background peaks, while the Cofiler experiment has a taller middle peak (411 rfu, instead of 261 rfu). Neither data experiment provides a perfect fit, so each one in isolation would confer uncertainty to the genotype. However, taken together, a joint likelihood function over all the data gives a probability of 97% to matching allele pair [15 17] (Table 3). This TA2 genotype yields an *LR* of eight. As CPI cannot statistically assess quantitative data, it just distributes equal likelihood over its six hypothesized D3 genotype values, for an *LR* of 2.6 (Tables 2 and 3).

10. The joint *LR* is formed by multiplying together the independent locus *LR*s (1). Considering *all loci*, TrueAllele forms a product of these 13 numbers as 10^{14}, or 100 trillion (Table 3). In contrast, the CPI method does not use five loci because of data peaks that are below their reporting threshold, a practice that may not always be conservative (38). CPI multiplies together the item's eight remaining locus *LR* scores to obtain a joint *LR* of 10^3, or 1000. Objectively using all 13 loci (regardless of low peaks), TrueAllele computes an *LR* that is 10^{11} (i.e., $10^{14}/10^3$, or 100 billion) times greater than CPI's *LR* score.

To summarize, the computer can mathematically set up a probability problem, and solve it using statistical search. By modeling the STR molecular process (Appendix), the solution can account for the contributor genotypes, the DNA template composition, PCR artifacts, and the statistical certainty of data peaks (e.g., stochastic effects and allele dropout). Moreover, the Bayesian reasoning quantitatively assesses all genotype possibilities. But CPI's qualitative approach just looks at thresholded peaks and does not invoke these biological or mathematical modeling considerations. Therefore, it cannot extract comparable identification information from low-level, minor contributor DNA evidence such as item E.

Results

We present the validation results for adjudicated mixture samples having two contributors, describing both efficacy and reproducibility in terms of log(*LR*) match information. We first examine eight cases with the victim reference considered (TA1 vs. CLR) and afterward turn to interpretation without using a victim reference (TA2 vs. CPI).

One Unknown Contributor

Cybergenetics used the TrueAllele system to interpret two-person mixture evidence items from eight different adjudicated cases, using the victim reference genotype and solving for the unknown second contributor genotype. Comparison in TrueAllele of each inferred genotype with its respective suspect genotype, relative to four ethnic subpopulations, produced *LR* match scores. The *LR* hypothesis here is that the unknown contributor is the suspect; the alternative is that he is not. The $\log_{10}$(TA1) match information values for these one unknown solutions are shown for duplicate computer runs of the eight cases. We conservatively show just the minimum *LR* value among the four subpopulation statistics (Table 4, TA reps 1 and 2). For reference, we also show the TrueAllele-inferred mixture weight of the unknown second contributor.

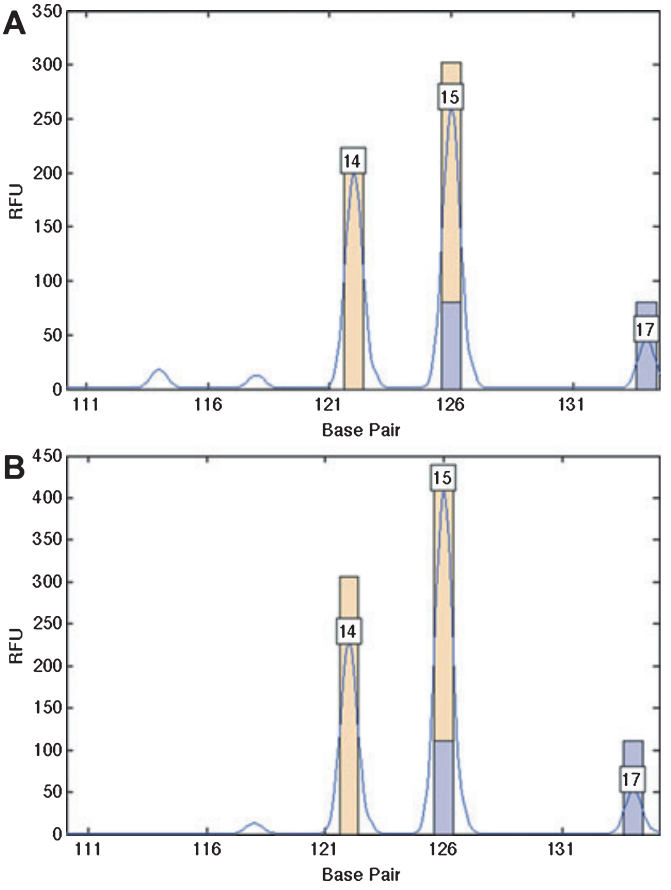

FIG. 6—*Using all the data. Two independent short tandem repeat experiments were performed at locus D3 of sample E, one in the (A) ProfilerPlus panel and another using (B) Cofiler. While neither experiment has a perfect fit to the genotype model, considering both experiments together in a joint likelihood function produces a more informative genotype.*

The NYSP FIC laboratory had previously analyzed the same eight evidence items. They similarly used the victim reference to help infer an unknown second contributor genotype. The analyst then used the CODIS Popstats software to compare each inferred evidence genotype with a suspect, relative to four subpopulations, and determine their reported CLR statistic. We show the CLR value corresponding to the smallest TrueAllele population statistic (Table 4, CLR).

Comparing log(*LR*) match information values, we see that the two repetitions of the TrueAllele TA1 computer runs agree well with each other (Fig. 7; Table 4). In every case, the computer's TA1 match score improves on the CLR value from human review. The average information improvement $\log_{10}$(TA1/CLR) between the computer and human interpretation is 4.67 log units, or about 50,000-fold (Table 4, information gain). This difference is statistically significant, with a *t*-test showing a *p*-value <0.001.

These information differences vary somewhat according to mixture weight (Fig. 8). A small contribution of the unknown fraction (around 20%) indicates that there is less perpetrator DNA present in the sample. The resulting data uncertainty correctly reduces the *LR*, because it is based on a less certain genotype. At higher mixture weights (over 40%), we see that the information improvement remains steady at about 4.5 log units.

In case 1A, there was a greater difference in the amount of preserved information—11 log units (100 billion improvement)—despite the presence of a 60% major contributor (Table 4, Case 1A, information gain). This outlier occurred because the mixture

TABLE 4—*Information comparison with one unknown contributor.*

		One Unknown Log(*LR*) Information			
Case	Weight	TA rep 1	TA rep 2	CLR	Gain
1A	0.590	20.38	20.34	8.20	11.31
1B	0.423	20.02	20.01	14.91	4.51
1C	0.584	18.63	18.64	15.03	3.61
1D	0.748	15.30	15.30	10.53	4.77
1E	0.486	17.66	17.66	13.52	4.13
1F	0.199	13.83	13.87	12.24	1.00
1G	0.556	17.43	17.46	11.87	5.52
1H	0.213	15.26	15.45	12.85	2.51
		17.31	17.34	12.40	4.67

The table shows $\log_{10}(LR)$ match information results for eight victim-known DNA mixture cases. Genotypes were inferred from two-person mixture data using a victim reference. Corresponding *LR*s were computed relative to a known suspect and a reference population. Each row corresponds to one case item and is identified in the first label column. The second label column gives the computed mixture weight of the unknown contributor. The likelihood ratio (*LR*) is conservatively computed relative to the ethnic subpopulation that produces the lowest TrueAllele score. The first two columns show the TrueAllele replicate runs using the TA1 method. The next two columns show the reported combined likelihood ratio (CLR) match results and the difference between TrueAllele and CLR match information. The last row provides column averages for each method.

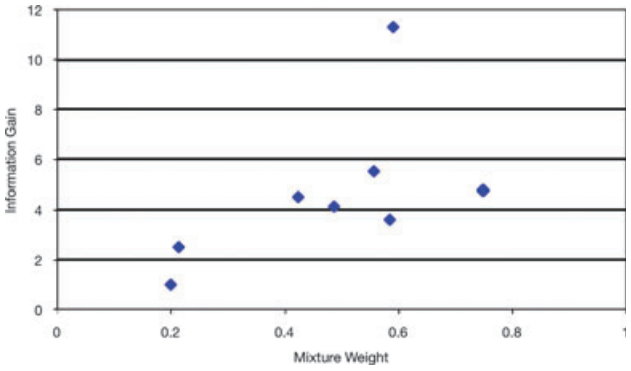

FIG. 8—*Match information improvement by mixture weight for one unknown contributor. In this scatterplot, each point's* x *coordinate is the mixture weight of the unknown contributor, while the* y *coordinate is the information improvement expressed as* $\log_{10}(TA1/CLR)$. *Results are shown for eight combined likelihood ratio (CLR) cases, each conservatively computed at the ethnic subpopulation with the smallest* LR *score.*

data had five loci at which just one or two allele peaks were above threshold. All of these alleles matched the victim genotype, without developing any potential unknown alleles. As no obligate alleles could be assigned, a CLR was not reported at these loci (3). But the continuous TA1 method was able to use all the peak data and infer a genotype and *LR* at every locus. The TA1 information differences between the two independent replicate TrueAllele runs of each case are very small (Fig. 7; Table 4, TA reps 1 and 2). Indeed, the average within-case square deviations of Eq. 6 yield a standard deviation of 0.036 log units (Table 6). This small standard deviation indicates that on this validation data set, the TrueAllele one unknown mixture interpretation method is highly reproducible.

Two Unknown Contributors

For the eight case items that did not have a victim reference, Cybergenetics ran TrueAllele to interpret the evidence by searching for the two unknown contributor genotypes. The statistical computation represented uncertain genotypes using probability distributions that assigned higher weight to locus allele pair combinations that better explained the quantitative STR data. Following genotype inference, the computer compared the two inferred genotypes to a suspect genotype, and computed *LR*s relative to four ethnic subpopulations. The *LR* here measured the weight of evidence for the identification hypothesis that the suspect contributed to the evidence. The $\log_{10}$(TA2) match information scores are shown for duplicate computations of the eight cases relative to the subpopulation that gave the smallest score (Table 5, TA reps 1 and 2). We also show the TrueAllele computed mixture weight of the suspect-matching contributor.

Before the start of this study, the NYSP laboratory had conducted a dual human review of these two-person mixtures. As no victim reference was available, they examined thresholded peak height data, enumerating alleles at every locus to (implicitly) infer a contributor genotype. A human analyst then entered these alleles into CODIS Popstats software to compare the inferred evidence genotype with a suspect, relative to four subpopulations, and calculate CPI statistics. We show these CPI scores, conservatively using

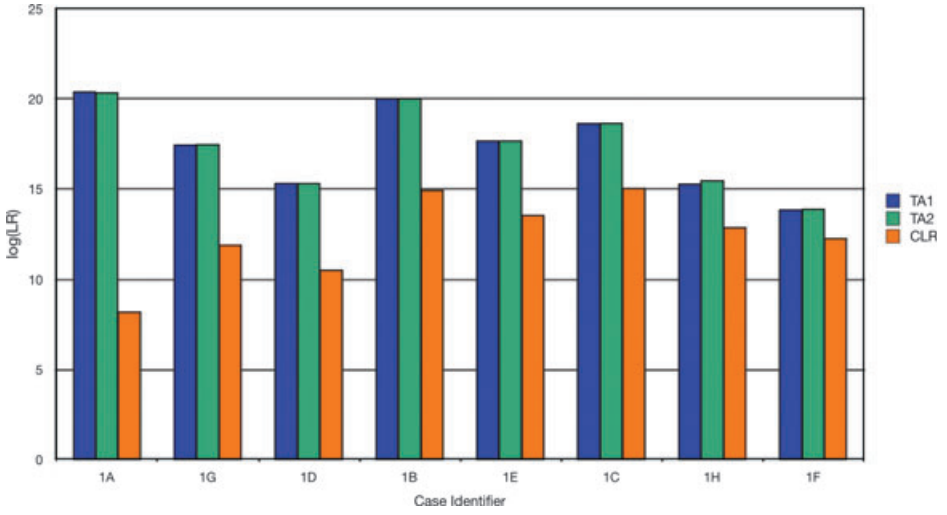

FIG. 7—*Match information comparison with one unknown contributor. The* $\log_{10}$(LR) *information values are shown for every case, with results for two replicate TA1 computer runs (blue, green) and the reported combined likelihood ratio value (orange). The cases are sorted by descending information difference.*

TABLE 5—*Information comparison with two unknown contributors.*

		Two Unknown Log(*LR*) Information			
Case	Weight	TA Rep 1	TA Rep 2	CPI	Gain
2A	0.652	17.89	17.54	7.01	10.49
2B	0.379	15.23	14.91	7.83	6.84
2C	0.333	14.74	14.63	5.87	8.22
2D	0.594	12.88	13.12	5.49	7.51
2E	0.520	10.11	9.73	7.53	2.32
2F	0.403	10.32	10.96	6.86	3.79
2G	0.464	9.07	8.71	6.24	2.65
2H	0.385	16.16	16.17	7.24	8.06
		13.30	13.22	6.76	6.24

The table shows $\log_{10}(LR)$ match information results for eight DNA mixture cases, without using a victim reference. Genotypes were inferred from two-person mixture data, and the corresponding *LR*s were computed relative to a known suspect and a reference population. The case item of each row is identified in the first label column; the second label column is the mixture weight. The likelihood ratio (*LR*) is conservatively computed relative to the ethnic subpopulation that produces the lowest TrueAllele score. The first two columns show the TrueAllele replicate runs using the TA2 (two unknown) method. The remaining columns show the reported combined probability of inclusion (CPI) match results and the difference between TrueAllele and CPI. The last row provides column averages for each method.

the population that minimized the TrueAllele score for each item (Table 5, CPI).

There is good agreement between the log(TA2) scores of the two replicate TrueAllele computer runs (Fig. 9; Table 5). Compared with the log(CPI) values from the reported inclusion probability results, the computer is more informative in every case, with an average match information gain $\log_{10}(TA2/CPI)$ of 6.24 log units, or 1.72 million (Table 5, CPI). Examination of pairwise differences by a *t*-test shows a statistically significant improvement with a *p*-value <0.01.

These information differences are markedly affected by mixture weight composition (Fig. 10). With an imbalanced DNA mixture, the computer can productively use mixture weight to separate out the two contributor genotypes, giving an average information increase over CPI of about eight log units. But there is a smaller information advantage with a 50:50 mixture, where we see a gain of only two and half log units. As most cases are not 50:50 mixtures, quantitative statistical interpretation has an average gain of about six log units.

Duplicate computer solutions of the same case let us quantitate the reproducibility of the TrueAllele two unknown TA2 mixture interpretation method. We see that the two log(*LR*) values are similar for each ethnic subpopulation within a case (Fig. 9; Table 5, TA reps 1 and 2). We use Eq. 6 to calculate a within-case standard deviation of 0.175 log(*LR*) units (Table 6). This relatively small variation shows that on this validation data set, the TrueAllele two unknown mixture interpretation method is highly reproducible.

Information Comparison

For one unknown contributor, TrueAllele solutions of two-person mixtures using a victim reference on our eight cases had an average log(*LR*) match information efficacy of 17.33 and a within-case reproducibility of 0.036 (Table 6, column 1). By comparison, the reported CLR human review score on the same cases averaged only 12.66 log units. The TrueAllele genotype inference and match *LR* were more informative in every case, with an average improvement of 4.67 log units.

With two unknown contributors, without using a victim reference, TrueAllele inferred both genotypes, matching the suspect with an average log(TA2) information efficacy of 13.26 (Table 6, column 2). On this second set of eight cases, the reproducibility was a within-case standard deviation of 0.175 log units. On these same cases, relative to the same populations, the reported human CPI value had an average logarithm of only 7.03. Thus, the computer showed an average match information improvement over manual review of 6.24 log units and was more informative in every case.

Using a victim genotype generally retained more identification information than when not using such a reference (Table 6). On average, the computer information $\log_{10}(TA1/TA2)$ increased 4.07 log units by using a victim (Table 6, first row). An even greater information gain $\log_{10}(CLR/CPI)$ of 5.63 log units was seen between the reported CLR and CPI (Table 6, second row).

Overall, TA1 computer inference with a victim reference was the most effective mixture interpretation method, averaging 17.33 log units. At the low end, CPI human interpretation without using

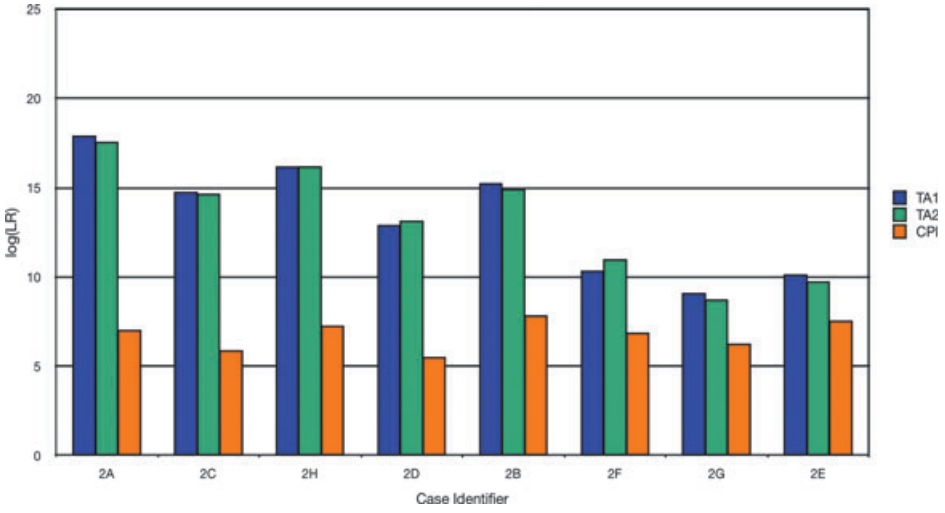

FIG. 9—*Match information comparison with two unknown contributors. The log_{10}(LR) information values are shown for every case, with results for the two replicate TA2 computer runs (blue, green) and the reported combined probability of inclusion (CPI) value (orange). The cases are sorted by descending information difference.*

TABLE 6—*Summary information comparison.*

Log(*LR*)	One Unknown (with Victim)	Two Unknown (without Victim)
TrueAllele computer	log(TA1) = 17.33 σ = 0.036	log(TA2) = 13.26 σ = 0.175
Reported match	log(CLR) = 12.66	log(CPI) = 7.03
Information gain	$\log\left(\frac{TA1}{CLR}\right) = 4.67$	$\log\left(\frac{TA2}{CPI}\right) = 6.24$

The average $\log_{10}(LR)$ information results are summarized and compared. The first column shows results for the one unknown contributor mixture problem, while the second column is for two unknown contributors. The rows give TrueAllele computer results, the reported match scores, and the net information gain. For the replicated computer results, we also provide the within-case standard deviation as a measure of reproducibility. CLR, combined likelihood ratio; CPI, combined probability of inclusion; *LR*, likelihood ratio.

a victim genotype was least effective, averaging only 7.03 log units. The information difference $\log_{10}$(TA1/CPI) of 10.30 between these two extremes is 10 orders of magnitude (10 billion), reflecting how effectively the different mixture interpretation methods make use of DNA data.

Prior and Posterior Probability

In principle, more informative priors could be obtained through a laboratory-specific calibration. A Bayesian framework, though, permits the use of generic prior probabilities. Therefore, such calibration is not necessary here and was not performed in this study.

In forming a posterior probability, the likelihood function addresses the observed data, whereas the prior probability does not. With STR data, the likelihood component typically overwhelms the prior contribution. For example, in the two unknown contributor case item 2A, the prior amplification variance σ^2 had a mean value of two. But after likelihood examination of the STR peak height data, the average σ^2 parameter value increased to 8.61.

A reporting level can affect how much of the posterior genotype probability distribution is displayed. A genotype credible level lists all allele pairs, starting from the most probable, until the aggregate probability reaches or exceeds the reporting level. For example, in the two unknown contributor case example E, locus D21 shows six allele pairs at a 99% cumulative level (Table 2, locus D21S11, Evidence E, TA2). However, at a 99.9% credible level, run 1 lists seven new allele pairs and run 2 has 13 new ones, with their intersection containing values [28 30], [29 32.2] and [35 35]. At a 99.99% level, run 1 lists a total of 21 allele pairs, while run 2 lists 26 values. This probability diffusion is expected at higher levels, given the inherent uncertainty of the nonvictim allele peak 35, which has a height of only 34 rfu.

The preceding example highlights the importance of the *LR*, which is unaffected by reporting level. The *LR* measures the likelihood concentration at a suspect genotype value, relative to its dispersal over a population genotype distribution. As such, the *LR* (unlike lists of alleles or allele pairs) is less subject to arbitrary reporting thresholds.

Discussion

Science (39) and the law (40) prefer forensic expert testimony that has a sound scientific basis. To demonstrate the reliability of DNA testing, forensic scientists conduct extensive validations of their STR data generation methods (41–43). Given the wide disparities found in DNA mixture interpretation results (5) and the ongoing controversy surrounding mixture interpretation methods (3,44), 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 (45).

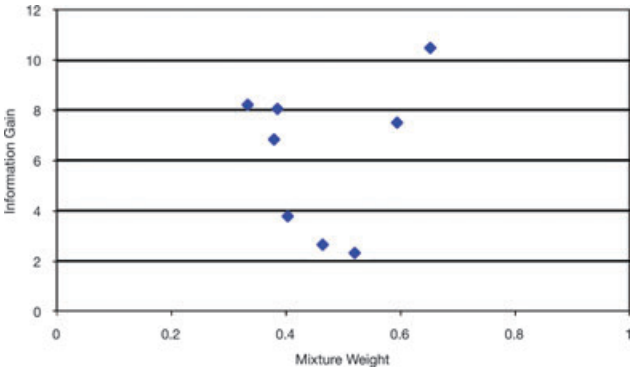

FIG. 10—*Match information improvement by mixture weight for two unknown contributors. In this scatterplot, each point's x value is the mixture weight of the matching unknown contributor, while the y value is the information improvement expressed as* $\log_{10}$*(TA2/CPI). Results are shown for eight combined probability of inclusion (CPI) cases, each conservatively computed at the ethnic subpopulation that produces the minimal* LR *match value.*

Two analysts may independently review the same mixture data and arrive at different allele (or genotype) lists (5). It can be hard to quantify these qualitative discrepancies or to make comparisons between different methods. Fortunately, in a validation study, the *LR* 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), their match statistic numbers are all *LR*s (7). As log(*LR*) is a standard measure of information (12), these numbers can be compared both within and between case interpretations to form the basis of a quantitative statistical validation study (13).

The advent of genetic calculators enables a computer interpretation of DNA evidence. While computers have been inferring genotypes from genetic data for quite some time (46), they have only recently been used for forensic identification (18,47,48). Computers offer three principal advantages in the interpretation process:

- *Productivity*. Computer review can help the analyst conduct rapid and accurate DNA data review (49). 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 (3). 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 (4). A mathematically programmed computer can infer a genotype directly from the evidence data without using any suspect information and then afterward 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 (3,10,50,51), with particular concern about stochastic effects and setting thresholds. However, a quantitative data variance model (Eqs [5], [10], [11], and [12]) 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 (52–55). 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 afterward making any *LR* comparison with the suspect. This "parallel unmasking" of independent evidence and suspect genotypes eliminates entirely any such objectivity concerns.

The recently released mixture interpretation guidelines (56) of the Scientific Working Group on DNA Analysis Methods (SWGDAM) specifically address stochastic effects and objectivity. One suggested SWGDAM human review approach is to introduce a "stochastic threshold," which can lead to a less informative result. For example, in the two unknown contributor case item E, applying a typical stochastic threshold of 150 rfu would eliminate the suspect's alleles at every locus, producing no CPI match score, with $\log(LR) = 0$. However, SWGDAM also approves of using a validated computer system for probabilistic genotypes (56, paragraph 3.2.2), as described in this article. Whereas the stochastic threshold procedure would reduce item E's DNA identification information 1000-fold from a reported CPI of 10^3 to an uninformative match score of 10^0, the probabilistic genotype approach instead increased the reported *LR* 100 billion-fold to 10^{14}.

The DNA subcommittee of the NYS Commission on Forensic Science has approved further "validation studies of TrueAllele by the NSYP" on active property crime cases. This ongoing evidence study is applying both quantitative TrueAllele and qualitative threshold methods to the same DNA data, comparing how well they each preserve DNA identification information. The low amount of (often mixed) DNA found in property crime evidence should be particularly amenable to a quantitative probabilistic genotype approach that uses all the locus data.

In this study, 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. The computer could extract more information from a 50:50 mixture when it used a victim reference. Without a victim reference, the computer TA2 method extracted less information from 50:50 mixtures than from imbalanced mixtures, though considerably more than the human CPI method. Having a victim reference enabled the computer TA1 method to extract comparable identification information from both 50:50 and imbalanced mixtures.

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.

Acknowledgments

The authors would like to thank the NYSP FIC analysts who helped identify the mixture cases and supply the case data, including Elizabeth Staude, Melissa Lee, and Shannon Morris. We appreciate the legal counsel of Steve Hogan at the NYSP, who provided guidance for handling case information. Cybergenetics developer Meredith Clarke provided software support in our use of the TrueAllele VUIer program. Careful readings by Jack Ballantyne, Robin Cotton, and Joanne Sgueglia improved the manuscript. We very much appreciate the thoughtful questions and suggestions posed by two anonymous reviewers, which helped clarify the presentation.

References

1. Evett IW, Weir BS. Interpreting DNA evidence: statistical genetics for forensic scientists. Sunderland, MA: Sinauer Assoc, 1998.
2. Perlin MW. Simple reporting of complex DNA evidence: automated computer interpretation. Proceedings of the Fourteenth International Symposium on Human Identification; 2003 Sept 30–Oct 3; Phoenix, AZ. Promega, 2003, http://www.promega.com/es/products/pm/genetic-identity/ishi-conference-proceedings/14th-ishi-oral-presentations/ (accessed June 11, 2011).
3. Budowle B, Onorato AJ, Callaghan TF, Manna AD, Gross AM, Guerrieri RA, et al. 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.
4. Thompson WC. Painting the target around the matching profile: the Texas sharpshooter fallacy in forensic DNA interpretation. Law, Probability and Risk 2009;8(3):257–76.
5. Butler JM, Kline MC. NIST mixture interpretation interlaboratory study 2005 (MIX05). Proceedings of the Sixteenth International Symposium on Human Identification; 2005 Sept 26–29; Grapevine, TX. Promega, 2005, http://www.promega.com/es/products/pm/genetic-identity/ishi-conference-proceedings/16th-ishi-poster-abstracts/ (accessed June 11, 2011).
6. Curran J. A MCMC method for resolving two person mixtures. Sci Justice 2008;48(4):168–77.
7. Perlin MW, Kadane JB, Cotton RW. Match likelihood ratio for uncertain genotypes. Law, Probability and Risk 2009;8(3):289–302.
8. Perlin MW, Sinelnikov A. An information gap in DNA evidence interpretation. PLoS ONE 2009;4(12):e8327.
9. Balding DJ, Buckleton J. Interpreting low template DNA profiles. Forensic Sci Int Genet 2009;4(1):1–10.
10. 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;160:90–101.
11. 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.
12. Good IJ. Probability and the weighing of evidence. London, UK: Griffin, 1950.
13. Perlin MW. Scientific validation of mixture interpretation methods. Proceedings of the Seventeenth International Symposium on Human Identification; 2006 Oct 10–12; Nashville, TN. Promega, 2006, http://www.promega.com/es/products/pm/genetic-identity/ishi-conference-proceedings/17th-ishi-oral-presentations/ (accessed June 11, 2011).
14. Feller W. An introduction to probability theory and its applications, 3rd edn. New York, NY: John Wiley & Sons, 1968.
15. O'Hagan A, Forster J. Bayesian inference, 2nd edn. New York, NY: John Wiley & Sons, 2004.

16. 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.
17. 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.
18. Perlin MW, Szabady B. Linear mixture analysis: a mathematical approach to resolving mixed DNA samples. J Forensic Sci 2001; 46(6):1372–7.
19. Cowell RG, Lauritzen SL, Mortera J. A gamma bayesian network for DNA mixture analysis. Bayesian Anal 2007;2(2):333–48.
20. 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.
21. Buckleton J, Curran J. A discussion of the merits of random man not excluded and likelihood ratios. Forensic Sci Int Genet 2008;2(4):343–8.
22. Savage LJ. The foundations of statistics. New York, NY: John Wiley & Sons, 1954.
23. Stolovitzky G, Cecchi G. Efficiency of DNA replication in the polymerase chain reaction. Proc Natl Acad Sci USA 1996;23:12947–52.
24. Stenman J, Orpana A. Accuracy in amplification. Nat Biotechnol 2001;19(11):1011–2.
25. Gelman A, Carlin JB, Stern HS, Rubin D. Bayesian data analysis. Boca Raton, FL: Chapman & Hall/CRC, 1995.
26. Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 1970;57:97–109.
27. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. J Chem Phys 1953;21(6):1087–92.
28. Bakker SC, Sinke RJ, Pearson PL. Differences in stutter intensities between microsatellites are related to length and sequence of the repeat. In: Bakker SC, editor. Unravelling the genetics of schizophrenia and ADHD. Utrecht, the Netherlands: University of Utrecht, 2005;81–94.
29. Aitken CG, Taroni F. Statistics and the evaluation of evidence for forensic scientists, 2nd edn. Chicester, UK: John Wiley & Sons, 2004.
30. Gilks WR, Richardson S, Spiegelhalter DJ, editors. Markov chain Monte Carlo in practice. London, UK: Chapman and Hall, 1996.
31. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. Stat Sci 1992;7(4):457–511.
32. 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.
33. Cowell RG, Lauritzen SL, Mortera J. Probabilistic modelling for DNA mixture analysis. Forensic Sci Int Genet Suppl 2008;1(1):640–2.
34. Faigman DL, Kaye DH, Saks MJ, Sanders J. Science in the law: standards, statistics, and research issues. St. Paul, MN: West Group, 2002.
35. Snedecor GW, Cochran WG. Statistical methods, 8th edn. Ames, IA: Iowa State University, 1989.
36. Scheffé H. The analysis of variance. New York, NY: John Wiley & Sons, 1959.
37. Kadane JB. Bayesian thought in early modern detective stories: Monsieur Lecoq, C. Auguste Dupin and Sherlock Holmes. Stat Sci 2009;24(2):238–43.
38. Curran JM, Buckleton J. Inclusion probabilities and dropout. J Forensic Sci 2010;55(5):1171–3.
39. National Research Council. Strengthening forensic science in the United States: a path forward. Washington, DC: National Academies Press, 2009.
40. Faigman DL, Kaye DH, Saks MJ, Sanders J. Science in the law: forensic science issues. St. Paul, MN: West Group, 2002.
41. 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).
42. Krenke B, Tereba A, Anderson S, Buel E, Culhane S, Finis C, et al. Validation of a 16-locus fluorescent multiplex system. J Forensic Sci 2002;47(4):773–85.
43. 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.
44. Gill P, Curran J, Elliot K. A graphical simulation model of the entire DNA process associated with the analysis of short tandem repeat loci. Nucleic Acids Res 2005;33(2):632–43.
45. *Orlando Roberts v. United States*: District of Columbia Court of Appeals, February 15, 2007, http://caselaw.findlaw.com/dc-court-of-appeals/1086885.html (accessed June 11, 2011).
46. Heuch I, Li F. PEDIG—a computer program for calculation of genotype probabilities, using phenotypic information. Clin Genet 1972;3: 501–4.
47. Bill MR, Gill P, Curran J, Clayton T, Pinchin R, Healy M, et al. PENDULUM—a guideline based approach to the interpretation of STR mixtures. Forensic Sci Int 2005;148(2–3):181–9.
48. Evett IW, Gill P, Lambert JA. Taking account of peak areas when interpreting mixed DNA profiles. J Forensic Sci 1998;43(1):62–9.
49. Varlaro J, Duceman B. Dealing with increasing casework demands for DNA analysis. Profiles in DNA 2002;5(2):3–6.
50. Budowle B, Chakraborty R, van Daal A. Authors' response to commentary by Gill and Buckleton. J Forensic Sci 2010;55(1):269–72.
51. 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.
52. 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.
53. Budowle B, Bottrell M, Bunch S, Fram R, Harrison D, Meagher S, et al. A perspective on errors, bias, and interpretation in the forensic sciences and direction for continuing advancement. J Forensic Sci 2009; 54(4):798–809.
54. Krane DE, Ford S, Gilder JR, Inman K, Jamieson A, Koppl R, et al. Sequential unmasking: a means of minimizing observer effects in forensic DNA interpretation. J Forensic Sci 2008;53(4):1006–7.
55. Krane DE, Ford S, Gilder JR, Inman K, Jamieson A, Koppl R, et al. 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.
56. SWGDAM. Interpretation guidelines for autosomal STR typing by forensic DNA testing laboratories. Washington, DC: Scientific Working Group on DNA Analysis Methods (SWGDAM), 2010.
57. Buckleton JS, Triggs CM, Walsh SJ, editors. Forensic DNA evidence interpretation. Boca Raton, FL: CRC Press, 2004.
58. Wang T, Xue N, Birdwell JD. Least-square deconvolution: a framework for interpreting short tandem repeat mixtures. J Forensic Sci 2006;51(6):1284–97.
59. Sisson SA. Genetics: genetics and stochastic simulation do mix! Am Stat 2007;61(2):112–9.
60. Thomas DC, Gauderman WJ. Gibbs sampling methods in genetics. In: Gilks W, Richardson S, Spiegelhalter D, editors. Markov chain Monte Carlo in practice. Boca Raton, FL: Chapman and Hall, 1995;419–40.

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

Appendix

Appendix A: Hierarchical Bayesian Model

We model the quantitative data at STR locus l (of L loci) using several variables. Data vector $\mathbf{d}_l$ forms a pattern that maps DNA product lengths into their observed quantitative peak heights. With K contributors to the data, we represent the kth contributor genotype parameter at locus l as a vector $\mathbf{g}_{k,l}$, where the DNA length entries contain allele counts that sum to one (18). A heterozygote genotype vector $\mathbf{g}_{k,l}$ contains two 1/2 entries, while a homozygote has a single one entry; all other vector entries are zero (17). The mixture weight parameter at locus l is a vector $\mathbf{w}_l$ whose K

contributor components sum to one, so that $\sum_{k=1}^{K} w_{k,l} = 1$. The total DNA quantity at locus l is given by mass parameter m_l. A quantitative linear model of data pattern $\mathbf{d}_l$ at locus l has an expected vector value μ_l given by the weighted genotype sum

$$\mu_l = m_l \cdot \sum_{k=1}^{K} w_{k,l} \cdot \mathbf{g}_{k,l} \tag{8}$$

Additional model variables can include PCR stutter, relative amplification, DNA degradation, and dye separation (57).

A hierarchical model of mixture weight at every locus provides a better fit to the data (6). We therefore draw each individual locus weight $\mathbf{w}_l$ as a hierarchical prior from a common DNA template mixture weight $\mathbf{w}$ using a truncated (simplex) multivariate normal distribution as

$$\mathbf{w}_l \sim N_{[0,1]^{K-1}}(\mathbf{w}, \psi^2 \cdot I) \tag{9}$$

The mixture weight covariance is an identity matrix scaled by a mixture variance ψ^2.

We write the peak data covariance matrix Σ_l as

$$\Sigma_l = \sigma^2 \cdot V_l + \tau^2 \tag{10}$$

where σ^2 is amplification dispersion, τ^2 is detection variation, and V_l is a diagonal matrix $diag(\mathbf{d}_l)$ of peak heights. We linearly model the data vector $\mathbf{d}_l$ using a truncated ($\geq \mathbf{0}$) multivariate normal distribution N_+ of the mean vector μ_l and covariance matrix Σ_l (18) as

$$\mathbf{d}_l \sim N_+(\mu_l, \Sigma_l) \tag{11}$$

Other square deviation data models can be used (47,58), as well as nonnormal distributions (19).

To infer the pmf $q(x)$ of genotype $\mathbf{g}_{k,l}$, we form the joint probability distribution of Eq. 1 over all the relevant random variables (25). The likelihood function elements $\Pr\{d_{l,i} | \mathbf{g}_{k,l} = x, ...\}$ are given by Eq. 11. The prior probability assignments are given in Eqs (9) and (12).

$$\begin{aligned} \mathbf{g}_{k,l} &\sim \begin{cases} f_i^2, & i = j \\ 2 f_i f_j, & i \neq j \end{cases} \\ \mathbf{w} &\sim Dir(\mathbf{1}) \\ m_l &\sim N_+(5000, 5000^2) \\ \sigma^{-2} &\sim Gam\,(10, 20) \\ \tau^{-2} &\sim Gam\,(10, 500) \\ \psi^{-2} &\sim Gam\,(1/2, 1/200) \end{aligned} \tag{12}$$

The genotype prior probability $\Pr\{\mathbf{g}_{k,l} = x\}$ at allele pair $x = [i\ j]$ is a product of population allele frequencies $\{f_i\}$. The template mixture weight $\mathbf{w}$ is assigned a uniform prior probability over the K contributor simplex. The locus mass m_l prior is a (nonnegative) truncated normal distribution on feasible total peak rfu values. The data variation parameters σ^2 and τ^2 have inverse gamma prior probability distributions, as does the mixture variance ψ^2.

Appendix B: Statistical Computing

Our goal is to determine uncertain genotype Q, described by its pmf $q(x)$ for each contributor at every locus. We described the posterior probability distributions of the key random variables Q, $\mathbf{w}$, σ^2, and τ^2 in Eqs ((1), (4), and (5). We use the joint probability distribution over all the data and variables to compute Q (15).

The joint probability distribution is fully specified (59,60) as the product of the likelihood and prior distributions, given in Eqs (11) and (12). Using a Metropolis-Hastings sampler (26), we iteratively draw from the posterior probability distributions of variables $\{\mathbf{g}_{k,l}\}$, $\{\mathbf{w}_l\}$, $\{m_l\}$, $\mathbf{w}$, σ^2, τ^2, and ψ^2 using MCMC computer methods (2,6). Once beyond the initial burn-in phase, the Markov chain produces samples from the joint posterior probability distribution (15). Marginalizing these posterior samples to each genotype random variable $\mathbf{g}_{k,l}$ for contributor k at locus l, we obtain the desired posterior probability functions $q(x)$ for genotype Q.

Appendix C: Likelihood Ratio for Uncertain Genotypes

The likelihood ratio (LR) is the information gained in the hypothesis H odds by having observed data (12)

$$LR = \frac{O(H | d_Q, d_R, d_S)}{O(H)} \tag{13}$$

Here, hypothesis H is that the suspect contributed to the DNA evidence, and the DNA data comprises the questioned evidence d_Q, the reference population allele frequencies d_R and suspect profile d_S. Standard Bayesian rearrangements (29) tell us that the LR can also be written as the ratio of conditional probabilities

$$LR = \frac{\Pr\{d_Q | H, d_R, d_S\}}{\Pr\{d_Q | \overline{H}, d_R, d_S\}} \tag{14}$$

where $\overline{H}$ is the alternative hypothesis that someone else contributed to the evidence (*Proof A*).

Suppose that there is uncertainty in the evidence genotype Q having pmf $q(x)$ or in suspect genotype S with pmf $s(x)$. Then, this genotype uncertainty is expressed in the LR as

$$LR = \frac{\sum_{x \in G} \lambda_Q(x) \cdot s(x)}{\sum_{x \in G} \lambda_Q(x) \cdot r(x)} \tag{15}$$

where $\lambda_Q(x)$ is the likelihood function of the evidence genotype Q and $r(x)$ is the pmf of reference population genotype R (*Proof B*). Although this LR shares many useful features of the match LR approximation (7), this exact LR equation uses likelihood function $\lambda_Q(x)$ instead of posterior probability $q(x)$. When genotype Q is inferred using a population prior R, likelihood λ_Q and posterior q are easily interconverted by renormalizing with prior r, because $q(x) \propto \lambda_Q(x) \cdot r(x)$.

Proof A.

We begin in the conventional way by expanding the LR odds definition (13) into probability ratios

$$\begin{aligned} LR &= \frac{O(H | d_Q, d_R, d_S)}{O(H)} \\ &= \frac{\Pr\{H | d_Q, d_R, d_S\} / \Pr\{\overline{H} | d_Q, d_R, d_S\}}{\Pr\{H\} / \Pr\{\overline{H}\}} \end{aligned}$$

Rearranging denominators we have

$$= \frac{\Pr\{H|d_Q, d_R, d_S\}/\Pr\{H\}}{\Pr\{\overline{H}|d_Q, d_R, d_S\}/\Pr\{\overline{H}\}}$$

By Bayes theorem, the posterior probability of H can be interchanged with its likelihood, renormalizing appropriately. Doing this separately for numerator and denominator, we obtain

$$= \frac{\Pr\{d_Q|H, d_R, d_S\}/\Pr\{d_Q\}}{\Pr\{d_Q|\overline{H}, d_R, d_S\}/\Pr\{d_Q\}}$$

Canceling out the total probability factors $\Pr\{d_Q\}$ yields the desired Eq. 14.

Proof B.

We start from Eq. 14 with the conditional probability form of the *LR*

$$LR = \frac{\Pr\{d_Q|H, d_R, d_S\}}{\Pr\{d_Q|\overline{H}, d_R, d_S\}}$$

Using the law of total probability (or, "extending the conversation"), we consider every possible allele pair $x \in G$ for genotype Q.

$$= \frac{\sum_{x\in G} \Pr\{d_Q|H, d_R, d_S, Q = x\} \cdot \Pr\{Q = x|H, d_R, d_S\}}{\sum_{x\in G} \Pr\{d_Q|\overline{H}, d_R, d_S, Q = x\} \cdot \Pr\{Q = x|\overline{H}, d_R, d_S\}}$$

The likelihood's probability of the data $\Pr\{d_Q|\ldots\}$ is unaffected by hypothesis H or $\overline{H}$. In the numerator, the evidence genotype Q under hypothesis H that the suspect contributed to the evidence, $\Pr\{Q = x|H, \ldots\}$ becomes the suspect's genotype S. Similarly, in the denominator the genotype Q under hypothesis $\overline{H}$ that someone else contributed $\Pr\{Q = x|\overline{H}, \ldots\}$ becomes the population genotype R. We therefore derive the ratio

$$= \frac{\sum_{x\in G} \Pr\{d_Q|d_R, d_S, Q = x\} \cdot \Pr\{S = x|d_R, d_S\}}{\sum_{x\in G} \Pr\{d_Q|d_R, d_S, Q = x\} \cdot \Pr\{R = x|d_R, d_S\}}$$

Eliminating noninfluential conditioning variables, we then have that

$$= \frac{\sum_{x\in G} \Pr\{d_Q|Q = x\} \cdot \Pr\{S = x|d_S\}}{\sum_{x\in G} \Pr\{d_Q|Q = x\} \cdot \Pr\{R = x|d_R\}}$$

Substituting in our notation for the evidence likelihood $\lambda_Q(x)$, and posterior pmfs for the suspect $s(x)$ and population $r(x)$ genotypes, we obtain the desired result, Eq. 15.